سلسلة الاحصاء والبحث للعلوم السلوكية والنفسية

باستخدام حزم البرامج الاحصائية

الجزء الاول

ما بعد اختبارات الدلالة الاحصائية

فترات الثقة، وحجم التاثير، والقوة الاحصائية، وما وراء التحليل

"SPSS, G-Power, & CMA"

Beyond Statistical Significance Testing

Confidence Intervals, Effect Size, Statistical power, & Meta-analysis

عبد الناصر السيد عامر

كلية التربية– جامعة قناة السويس

2020 م

إهداء السلسلة

الي روح امي وابي رحمة الله عليهما

الي الاستاذ الدكتور/ عبد العاطي الصياد الذي تتلمذت علي يديه

الي ابنائي احمد فاطمه رقيه

الي زوجتي امدها الله بالصحة والسعادة والعافية

فهرس محتويات الجزء الاول

إهداء

فهرس محتويات الجزء الاول

مقدمة السلسلة

شكر وتقدير

مقدمة الجزء الأول

الفصل الأول: المفاهيم الأساسية

الفصل الثانى: الإحصاء وعملية البحث (خطوات البحث)

الفصل الثالث: المتغيرات والبيانات

الفصل الرابع: أساسيات التعامل مع برنامج SPSS

الفصل الخامس: الثبات

الفصل السادس: التوزيعات التكرارية وتمثيلها بيانيًا

الفصل السابع: مقاييس النزعة المركزية

الفصل الثامن: مقاييس التشتت والالتواء والتفرطح والمنحنى الاعتدالي

الفصل التاسع: اختبارات الفروض الصفرية وفترات الثقة

الفصل العاشر: حجم التأثير والقوة الإحصائية

الفصل الحادى العاشر: العينات

الفصل الثانى عشر: اختبارات الفروض للتحقق من الاعتدالية

الفصل الثالث عشر: معامل ارتباط بيرسون الخطى ومعامل الارتباط الجزئي

الفصل الرابع عشر: معامل الارتباط الثنائى الأصيل

الفصل الخامس عشر: الانحدار البسيط

الفصل السادس عشر: اختبار z و T لاختبار فروض لمتوسط عينة واحدة

الفصل السابع عشر: اختبار فروض لإحصائيات أخرى (معامل الارتباط – النسبة)

الفصل الثامن عشر: اختبار T المرتبطة

الفصل التاسع عشر: اختبار T المستقلة

الفصل العشرون: تحليل التباين أحادى الاتجاه أو البسيط بين المجموعات

الفصل الحادى والعشرون: المقارنات المتعددة بين متوسطات المعالجات أو التحليل البعدي

الفصل الثانى والعشرون: ما وراء التحليل

المراجع

مقدمة السلسلة

يستخدم الباحثون الإحصاء فى العلوم السلوكية والنفسية والتربوية كوسيلة للتعامل مع بياناتهم فى بحوثهم ودراساتهم للتحقق من فروضهم البحثية، وتعاملت كثيرًا من المجلدات الإحصائية فى البيئة العربية مع الأساليب الإحصائية بصورة منعزلة عن المنهجية البحثية وهذا ما حاولت تفاديه فى هذا الكتاب حيث تناولت عرض الأساليب الإحصائية الاستدلالية فى اطار قضية أو مشكلة بحثية نفسية أو اجتماعية أو تربوية، وعرضت الخطوات البحثية لكل اختبار إحصائى من حيث السؤال البحثى وفرض البحث، والمنهج البحثى، والنموذج الإحصائى، والاختبار الإحصائى المناسب، وتناولت كل اختبار الإحصائى فى ضوء خطوات اختبارات الفروض التى ضمنتها فى سبعة خطوات هى الفروض الإحصائية الصفرى والبديل، والاختبار الإحصائى المناسب ومسلماته، ومستوى الدلالة الإحصائية وقاعدة القرار، والحسابات لكل اختبار، والقرار، حساب حجم التأثير لكل اختبار فى ضوء مؤشرات عديدة، وتقدير القوة الإحصائية فى ضوء برنامج G-Power وكيفية عرض نتائج الاختبار فى تقرير البحث فى ضوء توصيات الجمعية النفسية الامريكية APA.

في هذه السلسلة تحول الاهتمام من اختبارات الفروض الصفرية التي تعاني الكثير من الانتقادات والمحددات منها ثنائية القرار في ضوء $0.05 \geq P$ حيث تؤدي الي دلالة او عدم دلالة احصائية بمعني هل يوجد تاثير من عدمه؟، وهذا لا يؤدي الي تقدم العلم وتطوره، وانتقلت الي اصلاح الاحصاء reform Statisitics من خلال اعطاء مدي من القيم لقبول او رفض الفرض الصفري من خلال فترات الثقة، وكذلك الانتقال من وجود او عدم وجود تاثير الي مقدار وحجم هذا التاثير من خلال المكمل الاساسي حجم التاثير، وكذلك الانتقال من الدلالة الاحصائية للدراسة الواحدة الي استخدام ما وراء التحليل للاجابة عن الاسئلة البحثية من خلال اعطاء مؤشر تلخيصي للدراسات في التراث التي تناولت العلاقة بين متغيرين في الدراسة، وعليه استخدام فترات الثقة وحجم التاثير كمكملين لاختبارات الفروض الصفرية بجانب مؤشر هام للدقة الاحصائية وهو القوة الاحصائية؛ واستخدام ما وراء التحليل كبديل لاختبارات الفروض الصفرية

يمكن ان نطلق عليه تجديد او اصلاح الاحصاء كما اشار اليه (2013) Kline او الاحصاء الجديدة New Statistics كما اشار اليه (2012) Cumming وهذا ما حاولت التركيز عليه في هذه السلسلة وهذه المؤشرات والاجراءات الاحصائية تُوصي بها الجمعية النفسية الامريكية(APA) منذ اكثر من عشرين عاماً وكذلك المجلات النفسية والتربوية وفي كل التخصصات.

تناولت القوة الإحصائية كجزء ضرورى لاختبارات الفروض الصفرية وذلك لأن صناعة القرار تعتمد فى المقام الأول على القوة الإحصائية للاختبار حيث يرى المتخصصون ألا تؤخذ نتائج الاختبارات فى الاعتبار عند صناعة القرار إذا انخفضت القوة الإحصائية عن الحد الأدنى المقبول لها.

وفى مقابل القطبية فى صناعة القرار الإحصائى القائمة على اختبارات الفروض الصفرية الإحصائية، عرضت مدخل بديل لهذا وهو فترات الثقة حيث تمدنا بمدى له حد أدنى وحد أعلى وبينهما مجموعة من قيم معلم المجتمع المتضمن فى الفرض الصفرى. وكذلك تم عرض ما وراء التحليل باعتبارها أحد الطرق البديلة لاختبارات الدلالة الإحصائية للإجابة على أسئلة الدراسة. وبعد تناول الاختبار أو الأسلوب الإحصائى وحساباته اليدوية، عرضت كيفية إجرائه فى البرنامج الإحصائى Version SPSS 23 بكل خطواته من إدخال البيانات وخطوات تنفيذه فى البرنامج وكيفية تفسير النتائج وحساب القوة الإحصائية من خلال برنامج G-Power.

وأطلقت على هذ السلسلة "**الإحصاء والبحث للعلوم السلوكية والنفسية باستخدام حزم البرامج الإحصائية**" لأنها تضمنت التحليل الإحصائى للاختبارات الإحصائية لبيانات قضايا ومشكلات بحثية فى مجال علم النفس والتربية والعلوم السلوكية فى إطار التصميم البحثى وكذلك تضمنت تحليل البيانات فى أحد البرامج الإحصائية الكمبيوترية المتخصصة مثل SPSS أو غيره.

وتضمنت هذه السلسلة ثلاثة أجزاء كالاتي:

الجزء الأول: ما بعد اختبارات الدلالة الإحصائية : فترات الثقة، وحجم التأثير، والقوة الاحصائية وما وراء التحليل

"SPSS, G-Power, & CMA"

الجزء الثانى: تحليل البيانات اللابارامترية فى ضوء حجم التأثير "SPSS"

الجزء الثالث: تحليل البيانات المتعددة والمتدرجة فى ضوء حجم التأثير

"SPSS & LISREL"

وقد غطت هذه السلسلة معظم الاختبارات البارامترية للمتغير الواحد وللنموذج البسيط (متغير مستقل واحد ومتغير تابع واحد) Univariate، وللنموذج المتعدد Multiple (اكثر من متغير مستقل ومتغير تابع وحيد)، وللنموذج المتدرج Multivriate (متغير مستقل واحد فاكثر واكثر من متغير تابع) (الصياد، 1985) والاختبارات اللابارامترية التى يتناولها الباحث فى العلوم السلوكية والنفسية والتربوية بل معظم التخصصات الأكاديمية واعتمدت فى إعداد هذه السلسلة على كتب ومراجع إحصائية متخصصة وكذلك على خبرتى التدريسية لمقررات القياس والإحصاء فى كليتى لمدة تزيد على خمسة وعشرون عامًا وخبرتى البحثية فى مجال علم النفس والقياس والاحصاء، وكذلك على الاستشارات واستفسارات الباحثين أثناء تحليل بياناتهم.

وتمتاز هذه السلسلة بالآتى:

- الاعتماد على قضايا ومشكلات بحثية واقعية مأخوذة من مراجع متخصصة ولكن تمت معالجتها برؤية مختلفة.
- عدم الاقتصار على اختبارات الدلالة الإحصائية فى صناعة القرار بل تضمين فترات الثقة باعتبارها مدخلًا بديلًا لاختبارات الفروض الصفرية.
- الربط بين الاختبارات الإحصائية والتصميمات البحثية والتجريبية.
- معالجة بيانات الاختبارات والأساليب فى برنامج SPSS .
- حساب حجم التأثير والقوة الإحصائية لكل اختبار باعتبارهما من مؤشرات الدقة الإحصائية.
- التأكيد على تفسير وعرض النتائج فى تقرير البحث فى ضوء توصيات APA.

- عرض أساليب إحصائية متقدمة تناسب طبيعة الظواهر النفسية والاجتماعية مثل تحليل الانحدار المتعدد وتحليل التغاير وتحليل التباين العاملي والمتدرج والتحليل العاملى بنوعيه.
- عرض أسلوب ما وراء التحليل بمفاهيمه وخطوات إجرائه وتنفيذه فى برنامج Comperhensive Meta Analysis (CMA).

واتمنى أن اكون قدمت عملًا مفيدًا للطلاب والباحثين فى كافة التخصصات ليساعدهم على تصميم بحوثهم وتحليل بياناتهم بكل موثوقية وبقليل من الخوف والرهبة من التعامل مع البيانات وبعيدًا عن التعقيدات الحسابية.

والله الموفق والمستعان

أ. د. عبد الناصر السيد عامر

أستاذ القياس والإحصاء النفسى

كلية التربية– جامعة قناة السويس

adr.abdenasser@yahoo.com

شكر وتقدير

أتقدم بالشكر والتقدير إلى كل من ساعدنى فى إنتاج هذه السلسلة، وأخص بالذكر أستاذى معالى الأستاذ الدكتور عبد العاطى الصياد أستاذ مناهج البحث والقياس والتقويم والإحصاء الذى تتلمذت على يديه وتعلمت منه الكثير فى هذا الشأن اطال الله فى عمره وامده بالصحة والسعادة، كما أشكر الزميل الدكتور محمود على موسى المدرس بقسم علم النفس التربوى كلية التربية جامعة قناة السويس، والأستاذ تامر محمد جابر المصمم التعليمى الإلكترونى بجامعة قناة السويس الذين ساعدونى فى توفير الكثير من الكتب التى استفدت منها كثيرًا فى كتابة هذه السلسلة، كما اتقدم بالشكر الاستاذة الاء نور الدين الشرقاوي المدرس المساعد بقسم علم النفس التربوي بكلية التربية جامعة قناة السويس علي مراجعتها وتنسيقها ومساعدتها في نشر الكتب الكترونياً، وكما لا يفوتني الي ان اتقدم بالشكر والتقدير الي الزميل الفاضل الاستاذ فريد كمال معلم خبير اللغة الانجليزية بادارة فاقوس التعليمية محافظة الشرقية علي المراجعة اللغوية الرصينة للكتاب.

وفى النهاية أتقدم بالشكر والعرفان بالجميل لأفراد أسرتى وفى مقدمتهم زوجتى التى تحملت معى الكثير من المتاعب وساعدت على توفير المناخ الذى ساعدنى على إنتاج هذا العمل.

مقدمة الجزء الأول

تمثل الاختبارات الإحصائية البارامترية أهمية كبيرة في البحث النفسي والاجتماعى والتربوى حيث توظف الدراسات والبحوث هذه الاختبارات بدرجة كبيرة ودائمًا يفضل الباحث هذه النوعية من الاختبارات لأنها أكثر الفة بها، كما انها تناسب كثيرًا من البيانات، ولكن يوجد محدد رئيسي في استخدامها وهو أن الباحثون لا يتحققون من مسلمات هذه الاختبارات خاصة العشوائية والاعتدالية. وهذا ما حاولت اتجنبه في عرض اختبارات الفروض لهذه الاختبارات. وتناول هذا الجزء اختبارات نموذج المتغير الواحد واختبارات النموذج البسيط حيث انها تمثل 73.1% من الاختبارات الاحصائية المستخدمة في البحث النفسي والتربوي العربي(عامر، 2012).

وبعد عرض الاختبار أو الأسلوب الإحصائى وحساباته اليدوية، عرضت كيفية إجرائه في البرنامج الإحصائى SPSS Version 23 بكل خطواته من إدخال البيانات وخطوات تنفيذه في البرنامج وكيفية تفسير النتائج. ويتضمن هذا الجزء المفاهيم الأساسية التى تساعد الباحثون على انجاز دراساتهم مثل المتغيرات، دور الإحصاء فى خطوات البحث، وكيفية انتقاء العينات وتحديد حجمها فى ضوء طبيعة الاختبارات المختلفة، وأساسيات التعامل مع برنامج SPSS.

وتناولت مقاييس الإحصاء الوصفى مثل مقاييس النزعة المركزية والتشتت وكيفية عرض البيانات بيانيًا باعتبارها الخطوة الأولى للتعامل مع البيانات فهى تهدف إلى عرض وتنظيم وتلخيص البيانات بيانياً وعددياً لمحاولة فهمها بصورة أكثر وضوحًا وعرضت لكيفية حساب المتوسط المهذب للتعامل مع البيانات المتطرفة. وعرضت اختبارات الفروض الصفرية وخطواتها ابتداءً من الفروض الإحصائية انتهاءً بصناعة القرار، ودور حجم التأثير والقوة الإحصائية كمؤشرات للدقة الإحصائية، وعرضت لموضوع العينات انواعها وكيفية انتقائها وتحديد حجمها فى برنامج G-Power. واعتمدت فى هذا الجزء على مداخل حديثة لصناعة القرار اكدت عليها معظم المؤسسات البحثية والمجلدات العلمية ومنها الاعتماد على حجم التأثير لما له من أهمية فى صناعة القرار حيث يؤكد على القيمة العملية التطبيقية لنتائج الدراسات وليس

كما هو سائد الاعتماد على الدلالة الإحصائية التى لها محددات عديدة من أهمها ثنائية أو قطبية القرار وتأثرها بحجم العينة. تناولت القوة الإحصائية كجزء ضرورى لاختبارات الفروض الصفرية وذلك لأن صناعة القرار تعتمد فى المقام الأول على القوة الإحصائية للاختبار حيث يرى المتخصصون ألا تؤخذ نتائج الاختبارات فى الاعتبار عند صناعة القرار إذا انخفضت القوة الإحصائية عن الحد الأدنى المقبول لها. كذلك تم عرض لاستراتيجية Bootstrap باعتبارها الية لصناعة القرار فى أحجام عينات متولدة من قاعدة البيانات المتاحة وحساب فترات الثقة لهذه العينات للوصول إلى صناعة قرار فى ضوء عدد كبير من العينات.

كما تضمن الكتاب بديل هام لاستخدام الدلالة الإحصائية وهو أسلوب ما وراء التحليل البديل المفضل للإجابة عن الأسئلة البحثية للعلاقة بين متغيرين بدون إجراء الدراسات الأولية التى تعتمد على انتقاء عينة من الأفراد وبناء المقاييس ،إنما عينته هى الدراسات السابقة من ثم فلا حاجة للطريقة النمطية لإجراء البحوث الأولية خاصة إذا كان التراث البحثى يتضمن تراكم هائل من الدراسات الأولية ذات النتائج المتعارضة وعلى ذلك فهو يسهم فى تصميم بحوث جديدة وعليه فإنه طريقة منهجية جديدة لإجراء الدراسات حيث عينته هى الدراسات السابقة. واستحدث أسلوب ما وراء التحليل كرد فعل للانتقادات الموجهة لاختبارات الدلالة الإحصائية حيث لا تساعد على صناعة قرارات فعالة ولا تساعد على تقدم العلوم النفسية، أيضًا نتيجة للانتقادات الموجهة للطريقة التقليدية الكيفية السردية فى مراجعة الدراسات السابقة حث تؤدى إلى استنتاجات مختلفة وتعتمد على ذاتية الباحث خاصة لنتائج الدراسات المتعارضة.

وتناولت خطوات اختبارات الفروض الصفرية للاختبارات الإحصائية البارامترية ومنها العلاقة والانحدار البسيط والاختبارات الفارقة مثل T لعينة واحدة والمرتبطة والمستقلة، وتحليل التباين البسيط وكذلك إجراء المقارنات البعدية المقرونة بتحليل التباين، وإجراء هذه الاختبارات فى برنامج SPSS، وكذلك حساب القوة الإحصائية من خلال برنامج G-Power.

ويمتاز هذا الجزء بالتالى:

- الاعتماد على قضايا ومشكلات بحثية واقعية مأخوذة من مراجع متخصصة ولكن تمت معالجتها برؤية مختلفة.
- عدم الاقتصار على اختبارات الدلالة الإحصائية فى صناعة القرار بل تضمين فترات الثقة باعتبارها مدخل بديل لاختبارات الفروض الصفرية.
- الربط بين الاختبارات الإحصائية والتصميمات البحثية والتجريبية.
- معالجة بيانات الاختبارات والأساليب فى برنامج SPSS مع تفسير مخرجاته.
- حساب حجم التأثير والقوة الإحصائية لكل اختبار باعتبارهما من مؤشرات الدقة الإحصائية.
- التأكيد على تفسير وعرض النتائج فى تقرير البحث فى ضوء توصيات APA.

ويساعد هذه الجزء الباحثين في كل التخصصات خاصة فى مجال العلوم النفسية والاجتماعية والسلوكية على تصميم بحوثهم وانتقاء الأسلوب الإحصائى المناسب لفروض دراسته حتى نصل إلى قرارات تتسم بالموثقية والدقة الإحصائية.

واتمنى أن اكون قدمت عملًا مفيدًا للطلاب والباحثين فى كافة التخصصات مجال الإحصاء يساعدهم على تصميم بحوثهم وتحليل بياناتهم بكل موثوقية وبقليل من الخوف وبعيدًا عن التعقيدات الحسابية، وكذلك التعامل مع القضايا والظواهر والمشكلات النفسية والاجتماعية بما يناسبها من الأساليب الإحصائية المناسبة للوصول إلى موثوقية وطمأنينة فى صناعة القرارات.

والله الموفق والمستعان

أ.د. عبد الناصر السيد عامر

الفصل الأول

مفاهيم أساسية

Basic Concepts

منذ اهتمامى باعمل الاحصائي وتحليل البيانات فإن زملائى الاساتذة فى الكلية يرسلون لى طلاب الماجستير والدكتوراه لإعطائهم بعض النصائح المتعلقة بالاختبارات الإحصائية المناسبة لمشاريع بحوثهم، ويأتى الباحث ويقول" أنا لا أعرف أى شئ فى الإحصاء" على الرغم مما درسه فى مرحله الدراسات العليا من مقررات إحصائية وبالطبع هؤلاء الطلاب الباحثون يعرفون ماذا يريدون إنجازه ولكن كيفية تحليل بياناتهم واختيار الأسلوب الإحصائى الأمثل يبدو مفقودًا لديهم. وأصبح العالم اليوم عبارة عن أمواج من البيانات وأصبح صانعى القرارات مطالبين بكيفية إدارة واستخدام المعلومات التى تولدها البيانات. وأصبح الباحثون يعتمدون على طرق إحصائية لإدارة وتحليل كم هائل من البيانات لصناعة قرارات تستخدم فى عملية التغير، وعملية إدارة وتحليل البيانات هى صلب العمل الإحصائى.

تحمل كلمة الإحصاء **Statistics** معانى مختلفة للأشخاص، فالمدرس فى الفصل يراها مجموعة من المعادلات والقوانين والرسومات لتنظيم ولتفسير درجات تحصيل طلابه، وخبير الأرصاد الجوية يراها إحصائيات لتوقعات عن حالة الجو سواء درجات حرارة منخفضة أو مرتفعة، وعالم الرياضيات يراها كتخصص يهتم بالاستدلال الرياضى للقوانين والمعادلات الإحصائية، بينما يراها الباحث فى العلوم النفسية والاجتماعية باعتبارها وسيلة يستخدمها فى دراساته لتحليل بياناته والوصول إلى قرار يخص أسئلة بحثه. ويعتبر الرياضيون الإحصاء كتخصص أساسى ولذلك تسمى الإحصاء الأساسى **Basic Statistics**، فى حين يعتبرها السلوكيون والنفسيون أسلوبًا تحليليًّا تطبيقيًّا يستخدم فى حل المشكلات والقضايا النفسية والسلوكية ولذلك تسمى الإحصاء التطبيقى **Applied Statistics** وهذا هو محور اهتمام هذا الكتاب الذى يتناول الأساليب والاختبارات الإحصائية الشائعة فى البحوث النفسية والاجتماعية، وذلك من منظورين: الأول الحسابات والمعادلات الرياضية فى ضوء المنهجية البحثية، والثانى من خلال

إجراء الاختبار فى البرنامج الإحصائى **Statistical Package for Social Sciences (SPSS).**

مفهوم الإحصاء

ماذا تعنى الإحصاء للباحث فى العلوم الاجتماعية والنفسية والتربوية؟ بكل تأكيد هى مجموعة من الطرق والأساليب والاختبارات يطبقها الباحث لتحليل بياناته فى محاولة لفهمها وتفسيرها، فعلى سبيل المثال أراد باحث أن يحدد مستوى الأمن النفسى أو الدافعية لمجموعة من الأفراد أو للعينة المختارة فإنه يقوم بإعداد اداة قياس لجمع بيانات أو استجابات الأفراد ثم معالجتها باستخدام الاختبارات أو الأساليب الإحصائية لتحديد مستوى الامن النفسى أو الدافعية. والإحصاء مجموعة من النظريات والإجراءات والطرق تطبق بهدف فهم وتفسير البيانات، ويشير Duun (2001) إلى أن الإحصاء تركز على الطرق المناسبة لجمع وتبويب وتحليل وتفسير البيانات وهذه الطرق المستخدمة لفحص البيانات توضع فى مجموعة من القواعد والإجراءات يطلق عليها الإحصاء.

ويشير Kirk (2008) إلى أن للإحصاء أربعة معانى فى ضوء السياق التى تطرح فيه وهى البيانات، وظائف البيانات مثل المتوسط والمدى، أساليب جمع وتفسير البيانات لصناعة قرارات، علم ابتكار وتطبيق الأساليب الإحصائية. ويرى Wilcox (2009) أن الإحصاء علم وفن استخلاص المعلومات من البيانات المقاسة مع التأكيد على حل مشكلات العالم الحقيقى، ويرى Aron, Coups, & Aron (2013) أن الإحصاء أحد فروع الرياضيات التى تهتم بتنظيم وتحليل وتفسير البيانات، ويعرفها Howell (2014) بانها مجموعة من الإجراءات والقواعد التى تستخدم لاختصار أحجام كبيرة من البيانات إلى نسب معينة وتسمح لنا بعمل استنتاجات. واستخدامنا للإحصاء هو لتلخيص واعطاء صورة لخصائص محتويات مجموعة البيانات لتحديد الانماط فى هذه البيانات.

ويرى Hinkle, Wiersma, & Jurs (1994) أن العلوم الاجتماعية والسلوكية ما كنت أن توجد بهذا الوضع الان بدون الإحصاء.

وللأساليب الإحصائية ثلاثة وظائف أساسية (Howitt & Cramer, 2011):

1. تمدنا بالطرق التى تساعد على تلخيص وتنظيم البيانات الخام المجمعة من مصادر متنوعة سواء كانت مقاييس أو استبيانات أو بيانات ارشيفية وهذا يتحقق من خلال عرض الأشكال والرسومات البيانية التى تلخص البيانات مما يساعد على تيسير فهمها أو تلخيص البيانات المتعددة فى مؤشرات إحصائية بسيطة وهذا الفرع من الإحصاء يسمى الإحصاء الوصفي**Descriptive statistics**.

2. تمدنا بالطرق التى تساعدنا على صناعة القرار فى البحوث حيث تيسر وتوفر الجهد على الباحثين عند دراسة مجتمعات دراستهم من خلال العينات ثم تسعى إلى تعميم النتائج من العينة إلى المجتمعات التى سحبت منها وهذا الفرع من الإحصاء يسمى الإحصاء الاستدلالى **Inferential statistics**.

3. فى ظل وجود كم هائل من البيانات فإن الإحصاء توفر عدة أساليب تتميز بالقوة فى تبسيط بيانات المتغيرات واختصارها لعدد أقل من العوامل ولذلك فهى مدخل لتبسيط البيانات **Data simplification**.

ويستخدم الباحثون الإحصاء للتحقق من اسئلتهم البحثية وامدادهم بأدلة كمية إمبريقية لتدعيم فروضهم أو النظريات التى يتبنوها فى دراساتهم وعلى ذلك تستخدم الإحصاء لتقويم وتفسير السلوك فى صور كمية (رقمية).

فرعى الإحصاء

تصنف الإحصاء حسب الهدف من إجراء الدراسة الى:

الإحصاء الوصفى: تستخدم الإحصاء الوصفى عندما يكون الهدف وصف مجموعة من البيانات فقط وتستخدم لتصنيف وتلخيص البيانات فى صورة مبسطة كتقدير المتوسط أو الانحراف المعيارى للبيانات، فعلى سبيل المثال المدرس فى الفصل يبنى اختبار تحصيلى لقياس مهارات اللغة العربية ثم يطبقه ويصححه ويعطى درجات للطلاب بعد ذلك يقوم بعرض بتمثيل الدرجات بيانيًا من خلال الرسومات وتقدير متوسط التحصيل وتحديد نسب النجاح والرسوب فى الفصل. والإحصاء الوصفى مجموعة من الإجراءات تهدف إلى عرض ووصف وتلخيص البيانات بهدف تيسير فهمها والوصول لانطباعات مبدئية عنها بالتالى فإن الإحصاء الوصفى ذلك الفرع من

الإحصاء الذى يهتم بوصف محتويات البيانات لرسم صورة عامة عنها، بينما التعمق فى فحص البيانات قبل استخدام الأساليب الإحصائية المتقدمة تسمى تحليل البيانات الاستكشافى **Exploratory data analysis (EDA)**، ومن أهم الأساليب الإحصائية التى تقع فى الإحصاء الوصفى مقاييس النزعة المركزية ومقاييس التشتت والتوزيعات التكرارية والرسومات البيانية وغيرها.

الإحصاء الاستدلالى: بعد وصف البيانات وأصبحنا على وعى بماذا تقول على مستوى وصفى سطحى نهتم بعد ذلك بالإحصاء الاستدلالى وهى تتكون من مجموعة من الإجراءات لعمل تعميمات **Generalization** واستنتاجات واختبارات فروض عن الحالة فى المجتمع عن طريق دراسة العينات التى يتم سحبها من هذا المجتمع، فمثلًا فى تصميم التجارب نحن ندرك أنه من المستحيل تطبيق التجربة على المجتمع ككل ولظروف خاصة بعملية الضبط والتكلفة والوقت يتم سحب عينات من المجتمع الأصلى.

متوسط مفهوم الذات فى المجتمع ككل يشار اليه بالمعلم **Parameter(μ)**، ونفس مؤشر المتوسط لمفهوم الذات لبيانات العينة يطلق عليها إحصائية)**Statistic** ($\bar{X}$) ومن الأفضل تقدير المعلم للمجتمع ككل. والإحصائية هى تخمينات يمكن أن تكون تعبيرًا صادقًا وحقيقيًا للمجتمع إذا كانت العينة ممثلة تمثيلًا جيدًا للمجتمع أو تخمينات غير حقيقية إذا كانت العينة غير ممثلة للمجتمع، وبالتالى يتم استنتاج معالم المجتمع من خلال ما نعرفه من إحصائيات العينة. فالإحصاء الاستدلالى هو مجموعة من الطرق لعمل استدلالات عن خصائص المجتمع من خلال معرفة خصائص العينة.

ومن الخطأ القول إن الإحصاء الوصفى والإحصاء الاستدلالى تصنيفين متمايزين ومختلفين عن بعضها البعض فكلاهما مكمل للآخر بمعنى أن الإحصاء الوصفى هى ضرورية للإحصاء الاستدلالى، فالإحصاء الاستدلالى هى وصفية أيضًا لأن البيانات التى تقدر على مستوى العينة هى وصفية للعينة ولكن الإحصاء الاستدلالى تمتلك وظيفة اضافية إلى جانب الوصف وهى الوصول إلى تعميمات من العينة للمجتمع، ودائمًا الإحصاء الوصفى يتناول متغير واحد على حدة بينما الإحصاء الاستدلالى يتناول دراسة متغير واحد فاكثر وعليه **لإحصاء استدلالى بدون إحصاء وصفى.**

وحدث تقدم ونمو فى العمل الإحصائى فى أواخر القرن التاسع عشر على يد Galton من خلال اكتشافه لأسلوب تحليل الانحدار، وعمل Karl Pearson فى المطابقة البارامترية للتوزيعات الاحتمالية. وحث تطور للعمل الإحصائى من خلال استخدام تقريب كأى تربيع فى تطبيقات الزراعة والبيولوجى والوراثة. وتبع ذلك أفكار الإحصاء لاستدلالى فى القرن العشرين خاصة من 1890 حتى 1950 على يد الإخوة Karl & Egon Pearson وفيشر Fisher و Neyman Jerzy الذين اسسوا المنهجية الإحصائية والنظرية التأسيسية خاصة اثناء الحرب العالمية الثانية. ومن الجامعات التى اسهمت بدور كبير فى نمو النظرية الإحصائية جامعة North Carolina وفيها العلماء Hotelling و Cox، وجامعة California at Berkeley وفيها العالم Neyman وجامعة Prinston من خلال العالم John Tucky بالولايات المتحدة الامريكية حيث امتلكت مراكز بحثية إحصائية متميزة

فوائد الإحصاء

للإحصاء فوائد عديدة منها:

- وصف وتلخيص البيانات فى صورة مؤشرات كمية.
- عرض البيانات فى صورة أشكال ورسومات بيانية حتى يسهل فهمها.
- صياغة واختبارات الفروض.
- تساعد على صناعة القرارات المتعلقة بالقضايا النفسية والتربوية والاجتماعية.
- تساعد بالتنبؤ بما يحدث فى مستقبل الظواهر بما تتضمنها من متغيرات فى كل التخصصات الانسانية والطبيعية.
- تحلل البيانات.
- تسهم بدور أساسى فى خطوات الطريقة العلمية ابتداء من وضع الفروض، اختيار العينات، تحليل البيانات، وعرض النتائج.
- تساعد على دراسة المجتمعات الكبيرة من خلال انتقاء العينات مما يوفر الوقت والجهد والتكلفة ويشار إلى هذا بالوظيفة الاقتصادية للإحصاء.
- تسهم بدور فعال فى تحليل البيانات Data analysis من خلال توفير مجموعة كبيرة من الاختبارات الإحصائية.

- تفسير الدراسات البحثية من خلال عرض وتفسير النتائج.
- تقييم الإحصائيات المستخدمة فى حياتنا اليومية.
- ترشد للتصميم التجريبى والبحثى.

هل الإحصاء تكذب؟

يمكن للباحث أن يصل إلى قرارات إحصائية خاطئة من خلال استخدام الإحصاء وهذا ربما لا يعود إلى الإحصاء بل يعود إلى سوء استخدام الإحصاء، كأن يعتمد الباحث على اختبار إحصائى غير مناسب، أو لم تتحقق مسلمات استخدامه فكأن من الأفضل عدم الاعتماد عليه، أو يعتمد الباحث على أحجام عينات صغيرة بالتالى ربما لا يصل إلى دلالة إحصائية ولكن فى الحقيقة على مستوى المجتمع توجد اختلافات، أو يتعامل الباحث مع مقاييس ضعيفة المصداقية، أو يتعامل مع بيانات منخفضة الثبات. وكما نعلم أن العمل الإحصائى يعتمد على بيانات واهم شئ يتوفر فى هذه البيانات جودتها بمعنى ثباتها يكون مرتفع والتعامل مع بيانات ثباتها أقل من **0.70** دائمًا يؤدى إلى نتائج مشوهه وذلك لأن القياسات فى مجملها كانت لأخطاء وليس للظاهرة الحقيقية المراد قياسها.

فالأساليب الإحصائية تكون مفيدة للباحثين الذين يحسنون استخدامها فالبداية بالسؤال وتصميم الدراسة الجيد، وفى تحليل البيانات الإحصاء هى اداة مساعدة للوصول إلى استنتاجات وقرارات سليمه بقدر الامكان، وعلى ذلك فإن الإحصاء لا تكذب انما سوء استخدامها يؤدى إلى قرارات خاطئة.

الفصل الثانى

الإحصاء وعملية البحث (خطوات البحث)

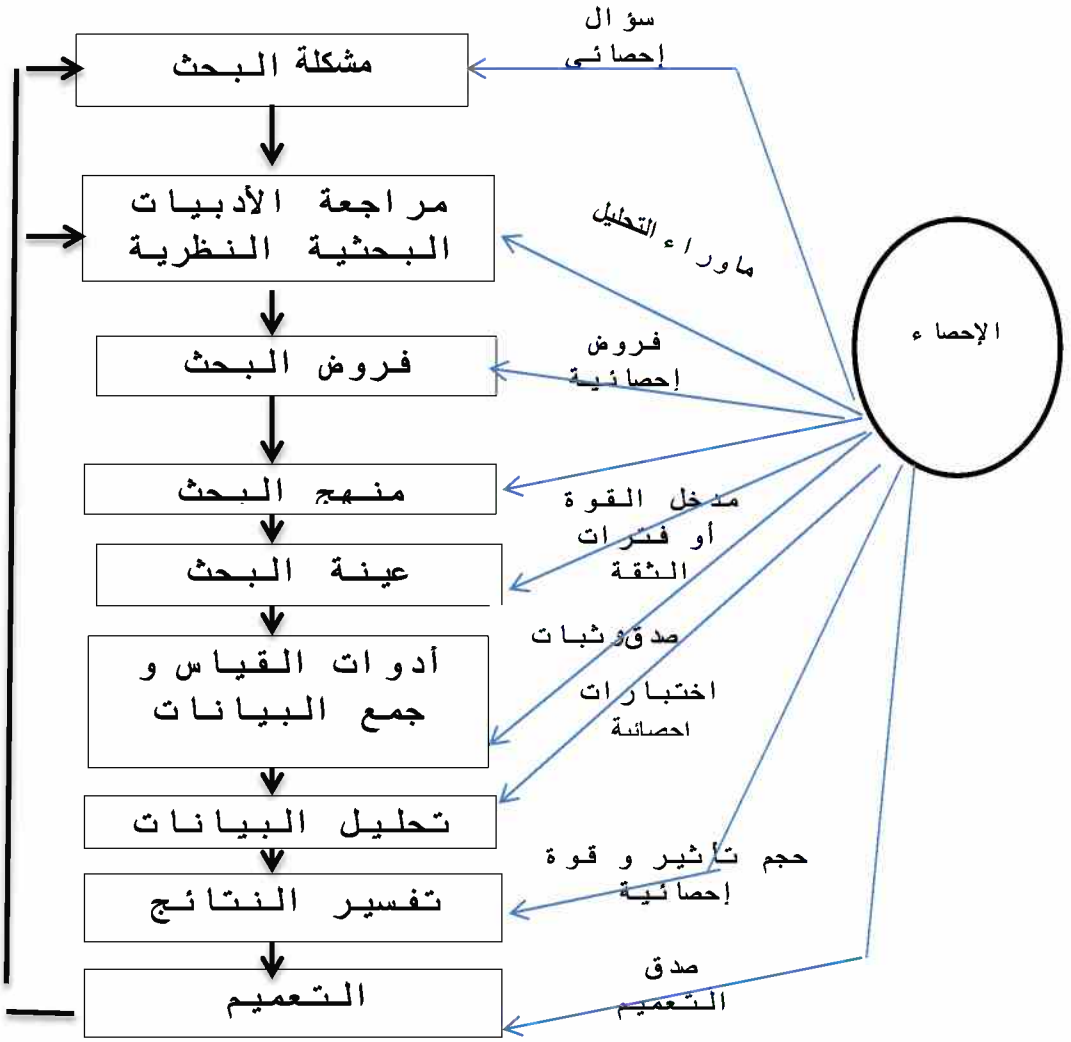
Statistics and Research Process

يختلف الباحثون حول مدى إسهام الإحصاء فى خطوات أو عمليات البحث، فيرى البعض أن دورها يقتصر على التحليل الإحصائى للبيانات، بينما يرى البعض الآخر أن دورها يتخطى ذلك بكثير حيث تسهم بدور فعال فى مراجعة الدراسات السابقة وصياغة الأسئلة والفروض وتفسير النتائج. وتلعب الإحصاء دورًا كبيرًا فى كل خطوات عملية البحث التى هى بمثابة سلسلة متفاعلة من الخطوات للوصول إلى حلول أو قرارات لمشكلة ما كما هو واضح فى شكل (1.2).

1. مشكلة البحث: فيها يتم ترجمة إحساس الباحث حول قضية أو ظاهرة ينوى دراستها إلى أسئلة بحثية **Research questions** بحيث تكون قابلة لاختبار صحتها والتحقق منها وهذه الأسئلة تعكس متغيرات الدراسة أو الظاهرة، فمثلًا تناول باحث دراسة اثر عادات الاستذكار على التفكير التأملى بالتالى طرح سؤال البحث "ما أثر عادات الاستذكار على التفكير الابتكارى؟" أو اهتم باحث بدراسة العلاقة بين إدمان الفيس بوك والوحدة النفسية فسؤاله "ما طبيعة العلاقة بين إدمان الفيس بوك والوحدة النفسية؟". وفى هذه المرحلة يبدو أن دور الإحصاء محدودًا إلى حدًا ما ولكن يمكن أن تساعد الباحث على ترجمة السؤال البحثى إلى سؤال إحصائى **Statistical question**، مثلًا هل يهتم الباحث بدراسة متوسط التفكير الابتكارى أو وسيطه أو تباينه؟، ويترجم السؤال البحثى إلى "ما أثر أو فعالية عادات الاستذكار على تنمية متوسط درجات التفكير الابتكارى. وتتنوع الأسئلة البحثية الى:

- **فروق Difference:** هل توجد فروق بين الذكور والإناث فى الابتكارية؟
- **علاقة Association:** هل توجد علاقة بين الاتجاه نحو المدرسة والعنف المدرسى؟
- **تنبؤ Prediction:** هل الذكاء منبئ بالابتكارية ؟

- **حسن مطابقة Goodness of fit**: هل إحصائيات العينة (التكرارات المشاهدة) تتطابق مع معالم المجتمع؟
- **بنية عاملية Factorial structure**: ما طبيعة البناء العاملى لمفردات مقياس الشخصية؟



شكل (1.2) : عمليات البحث ودور الإحصاء.

2. مراجعة الدراسات السابقة (الأدبيات البحثية) والنظريات: بعد تحديد المتغيرات المراد دراستها يقوم الباحث بالبحث عن الدراسات السابقة التى تناولت هذه المتغيرات من خلال تحديد الكلمات المفتاحية **Key words** والبحث فى قواعد البيانات المتخصصة، والباحث يكون لديه كمًا هائلًا من الدراسات والبحوث، وفى هذه المرحلة تلعب الإحصاء دورًا هامًا حيث تساعد الباحث على قراءة الدراسة السابقة قراءة منهجية وإحصائية جيدة، ويجب أن يكون لدى الباحث بعض المفاهيم الإحصائية مثل العلاقة، الفروق، التنبؤ، الانحدار، تحليل التباين، التحليل البعدى، مستوى الدلالة الإحصائية، حجم التأثير، القوة الإحصائية، و... إلخ.

والسائد فى مراجعة الدراسات السابقة أن يقوم الباحث بمراجعتها مراجعة كيفية سردية **Narrative qualitative review** وفيها يقوم بتفسير وقراءة الدراسة وقد يصل إلى استنتاجات يشوبها الذاتية حيث يمكن لباحثان أن يصلوا إلى استنتاجات مختلفة لنفس الدراسة ولكن الطريقة الأكثر موضوعية لمراجعة الدراسات السابقة هى المراجعة الكمية **Quantitative review** أو ما يطلق عليها البعض ما وراء التحليل **Meta-analysis** وهى استراتيجية منهجية لمراجعة الدراسات السابقة كميًا والوصول إلى متوسط حجم التأثير أو الارتباط بين متغيرين ما وهذا الاستخدام غائب فى مراجعة الدراسات السابقة فى البيئة العربية، وحساب هذا الأسلوب يعتمد فى المقام الأول على الإحصاء مثل مؤشرات حجم التأثير.

3. فروض البحث Research hypotheses: بعد صياغة أسئلة البحث ومراجعة الأدبيات البحثية يبدأ الباحث فى صياغة فروض بحثه وهى عبارة عن حلول أو توقعات محتملة لطبيعة العلاقة أو التأثير بين متغيرين فأكثر ولكن من الضرورى أن يتم صياغة الفرض فى ضوء النظرية أو التراث البحثى للظاهرة. فإذا كان سؤال البحث: هل توجد فروق بين الذكور والإناث فى الابتكارية؟ فإن الباحث أمام أربع إجابات (فروض) محتملة لهذا السؤال هى:

- توجد فروق بين الذكور والإناث فى الابتكارية.
- توجد فروق بين الذكور والإناث فى الابتكارية لصالح الإناث.

- توجد فروق بين الذكور والإناث فى الابتكارية لصالح الذكور.
- لا توجد فروق بين الذكور والإناث فى الابتكارية.

فالباحث لديه أربع إجابات متوقعة ولكن عليه أن يختار الفرض الذى يتفق مع الدراسات السابقة أو النظرية، وفى هذه الخطوة تلعب الإحصاء دورًا فى غاية الأهمية حيث تسهم فى ترجمة فرض البحث إلى الفرض الإحصائى **Statistical hypothesis** وهو التعبير الإحصائى لفرض البحث فى ضوء مؤشرات إحصائية (متوسط، تباين) ليبين طبيعة الأسلوب الإحصائى المناسب، ويصاغ فى صورة فرضين إحصائيين أحدهما صفرى والآخر بديل. فالفرض الإحصائى يحدد طبيعة المؤشر الإحصائى (المعلم) المستخدم لاختبار الفروض سواء المتوسط أو الوسيط أو التباين أو غيرها، واستخدام الإحصاء فى التحقق من الفروض يطلق عليها **اختبارات الفروض الصفرية** Null hypothesis testing(NHT).

4. **التصميم البحثى أو منهج البحث Research Design**: فى العلوم النفسية والاجتماعية لا بد من استخدام منهج البحث أو طريقة البحث Research method وفيها يحدد الباحث طبيعة المنهج أو المدخل العام الذى يتبعه للتحقق من فروض بحثه، وهذا يتحدد من خلال طبيعة أسئلة البحث المطروحة، وبدوره منهج البحث يحدد الإجراءات والخطوات التى يتبعها الباحث من اختيار العينة وتطبيق أدوات القياس والإجراءات المتبعة لجمع البيانات. والإحصاء وتصميم البحث متلازمان ومتداخلان ولفهم التوظيف الإحصائى يجب عرض أهم التصميمات البحثية ولكننا فى هذا الصدد نركز على أهم طرق البحث الإمبريقية الشائعة الاستخدام فى البحث النفسى والاجتماعى وهى:

البحث التجريبى Experimental method

يهدف إلى محاولة الوصول إلى التحقق من وجود علاقة السبب والتأثير Cause-effect بين متغيرين من خلال تجربة أحدهما يطلق عليه المتغير المستقل أو المتغير التجريبى أو متغير المعالجة وهو يخضع لسيطرة وتحكم الباحث ويجب أن يمتلك على الأقل مستويين وتسمى شروط أو عوامل Factors، أحدهما شرط المعالجة

Treatment والآخر شرط الضبط Control، بينما المتغير الآخر يسمى بالمتغير التابع أو المتغير المحك Criterion أو الناتج Outcome ويشار إليه بنتائج أو مخرجات التجربة مثل التحصيل والابتكارية والعنف. ولكى ندرس السببية من خلال تجربة فلا بد أن تتضمن إجراءات صارمة منها:

- المعالجة لمتغير تجريبى مراد دراسته Manipulation.
- الضبط لكل المتغيرات الدخيلة التى يمكن أن تؤثر فى المتغير التابع.
- العشوائية Randomization فى اختيار عينة البحث وكذلك عشوائية فى توزيع المشاركين على مجموعات التجربة بطريقة تحدث درجة كبيرة من التكافؤ Equivalence.
- المقارنة بين درجات المعالجات المختلفة وهى المجموعات التجريبية (التى تتلقى المعالجة) والمجموعة الضابطة (لا تتلقى المعالجة أو تتلقى مستوى آخر تقليدى من المعالجة)، فالفرق بين نواتج المعالجات دليل على أن المعالجة (المتغير المستقل) المسبب للتغيرات فى درجات المتغير التابع.

وعلى ذلك فأى تجربة تتضمن أربع مكونات أساسية لاستنتاج السببية هى المعالجة، العشوائية، الضبط، والمقارنة. فمثلًا إستراتيجية التدريس هى (المعالجة) أو المتغير المستقل (تقليدية-عصف ذهنى) والتحصيل تابع أو ناتج، العلاج السلوكي(المعالجة) بينما خفض الاكتئاب تابع (ناتج)، إستراتيجية طرح التساؤل الذاتى (المعالجة) بينما الفهم القرائى تابع(الناتج). وعليه فإن معالجة مستويات المتغير المستقل (عالي-منخفض أو معالجة -لا معالجة) وتوزيع أفراد العينة عشوائيًا على مجموعات أو مستويات المعالجة وضبط المتغيرات الدخيلة هى سمات البحث التجريبى. والمنهج التجريبى يتضمن تصميمات عديدة منها:

1. تصميم بين الأفراد أو المجموعات Between-subjects design: السمة الأساسية لهذا التصميم هو المقارنة بين مجموعات مختلفة من الأفراد وتوزيعهم عشوائيًا على مستويات ومعالجات المتغير المستقل وبالتالى توجد مجموعة أو مجموعات تجريبية Experimental group(s) ومجموعة ضابطة Control group ويطلق

عليــه التصــميم التجريبــى للقياســات المســتقلة Independent measures experimental Design وفيـه تـتم المقارنـة بـين المجموعـات المختلفـة وتحديـد الفـروق بينهما وفيما يلى عرض لأهم التصميمات التجريبية (Shadish & Cook (2002:

- تصميم المجموعة التجريبية الواحدة والمجموعة الضابطة وقياس بعدى:

R X O1 (مجموعة تجريبية)

R O2 (مجموعة ضابطة)

R العشـوائية،X المعالجــة، O_1 قيـاس بعـدى للمجموعـة التجريبيـة، O_2 قيـاس بعـدى للمجموعة الضابطة.

- تصميم المجموعتين التجريبيتين (المعالجتين) وقياس بعدى:

R XA O_1
R XB O_2

- تصميم مقارنة معالجتين (مجموعتين تجريبيتين) ومجموعة ضابطة وقياس بعدى:

R XA O_1

R XB O_2

R O_3

- تصميم المجموعة الضابطة وقياس قبلى وبعدى:

R O_1 X O_2
R O_3 O_4

- تصميم المعالجات البديلة مع قياس قبلى:

R O1 XA O2
R O_3 XB O_4

- تصميم المعالجات المتعددة مع المجموعة الضابطة و قياس قبلى :

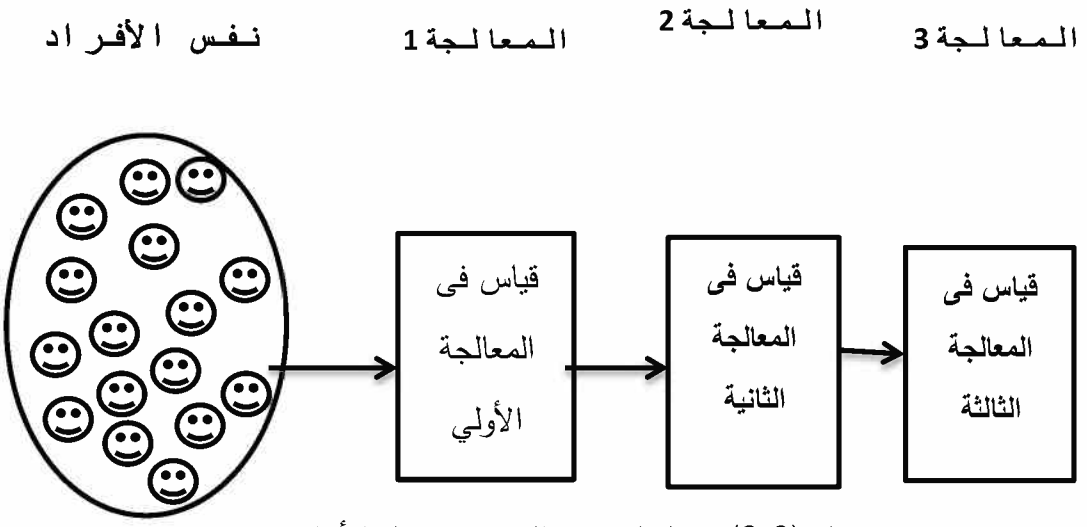
RO_1 XA O_2
R O_3 XB O_4
R O_5 O_6

ويمكن تعميمه لأكثر من ثلاث مجموعات:

R	O_1	XA	O_2
R	O_3	XB	O_4
R	O_5	XC	O_6
R	O_7		O_8

ولكن التصميم الأكثر شيوعًا فى البحوث النفسية والتربوية والسلوكية هو تصميم المجموعة الضابطة والتجريبية وقياس قبلى وبعدى، وكذلك تصميم المعالجات المتعددة مع المجموعة الضابطة وقياس قبلى. واستخدام الإحصاء فى هذه النوعية من التصميمات يستخدم لدراسة الفروق بين القياسات البعدية للمجموعات. ومن أهم الأساليب الإحصائية المستخدمة اختبار T للعينات المستقلة وكذلك تحليل التباين أحادى الاتجاه (ANOVA) مصحوبًا باختبارات المقارنات المتعددة البعدية Post-hoc لتحديد أى المجموعات التى أحدثت الفروق مثل اختبارات شيفيه، وتوكى ونيومان-كولز وغيرها. وكذلك يمكن استخدام اختبارات الإحصاء اللابارامترى فى حالة عدم توافر مسلمات الإحصاء البارامترى (الاعتدالية وتجانس التباينات) مثل اختبار مان-ويتنى Mann-Whitney (M.W) واختبار كروسكال- ولاس (K.W) وغيرها.

2. تصميم داخل الأفراد أو تصميم القياسات المتكررة Repeated measures designs within-subject: وفى هذا التصميم تستخدم مجموعة واحدة من الأفراد ويطبق عليها كل معالجات أو مستويات المتغير المستقل ثم قياس أو ملاحظة أداء كل فرد فى كل المعالجات ويطلق عليه تصميم القياسات المتكررة لأنه يحدث إعادة للقياسات على نفس الأفراد ولكن لظروف معالجات مختلفة وهذا التصميم سائد أيضًا فى الدراسات غير التجريبية مثل الدراسات الطولية التى تتناول التغيرات للمتغيرات عبر فترة زمنية طويلة ويمكن تمثيله كالآتى:



شكل (2.2): بنية التصميم التجريبى داخل الأفراد.

وهذه النوعية من التصميمات نادرة الاستخدام فى البحوث النفسية والتربوية والاجتماعية، والمعالجة الإحصائية لهذا النوع من خلال تحليل التباين للقياسات المتكررة Repeated ANOVA أو الاختبار اللابارامترى المقابل فريدمانFriedman أو اختبار T المرتبطة والمقابل اللابارامترى مان – ويتنى أو اختبار الاشارة.

المنهج شبه التجريبى Quasi expremintal method

فى التصميمات شبه التجريبية يهدف الباحث إلى دراسة أثر المتغير المستقل على المتغير التابع فى تجربة تحدث فى مواقف الحياة الطبيعية مثل المدرسة أو المستشفى حيث إن إجراءات الضبط فى التصميم التجريبى الحقيقى غير متوفرة وكذلك لا يتوفر فيها التوزيع العشوائى للأفراد على ظروف المعالجات المختلفة، وهذه التصميمات سائدة فى العلوم النفسية والاجتماعية والتربوية حيث إجراءات الضبط تبدو عملية صعبة جدًا نظرًا لأن الظاهرة النفسية ذات متغيرات متعددة ومعقدة ومتفاعلة ولذلك فإن هذه التصميمات تعانى من نقص الصدق الداخلى Internal validity وهو الادعاء بأن التغير فى المتغير التابع يرجع ويرجع فقط إلى المتغير التجريبى وبالتالى فالتجارب التى تجرى فى واقعنا التربوى فى المدارس يغلب عليها التصميم شبه التجريبى ويطلق عليها تجارب الميدان أو الحقلField experiments وعلى الرغم من ذلك فإن التصميمات شبه التجريبية لها درجة صدق خارجى External validityعالية حيث يعايش الباحث الظاهرة كما تحدث فى الواقع بالتالى فإن نتائج التجربة لها قدرة تعميمية عالية.

وتوجد العديد من التصميمات شبه التجريبية كما أوردها Shadish & Cook (2002):

- **تصميمات شبه تجريبية بدون مجموعات ضابطة**

أ. تصميم المجموعة الواحدة وقياس بعدى فقط:

$$\text{x} \quad O_1$$

بـ. تصميم المجموعة الواحدة وقياس قبلى وبعدى:

$$O_1 \quad X \quad O_2$$

ج. تصميم المجموعة الواحدة وقياس قبلى ثنائى:

$$O_1 O_2 \quad X \quad O_3$$

د. تصميم المعالجة المحذوفة أو المستبعدة:

$$O_1 \quad X \quad O_2 \, O_3 \quad * \quad O_4$$

هـ. تصميم المعالجة المتكررة:

$$O_1 \quad X \quad O_2 * O_3 \quad X \; O_4$$

- **تصميمات شبه تجريبية مع وجود مجموعة ضابطة بدون قياسات قبلية:**

أ- تصميم القياس البعدى فقط مع مجموعات غير متكافئة :

$$N_R \qquad \text{X} \; O_1$$
$$N_R \qquad O_2$$

N_R تعنى لاعشوائية Non- randomization

ب- تصميم القياس البعدى مع قياس قبلى لعينة مستقلة:

$$N_R \text{ ---- } O_1 \quad \text{X} \quad O_2$$
$$N_R \text{ ---- } O_3 \qquad O_4$$

- **تصميمات شبه تجريبية تستخدم مجموعات مقارنة وقياس قبلي- بعدى:**

أ- تصميم المجموعة الضابطة غير المعالجة مع قياس قبلى وبعدى:

$$\text{N}_\text{R}\text{O}_1 \quad \text{X} \quad \text{O}_2$$
$$N_R \quad O_3 \qquad\qquad \text{O}_4$$

ب- تصميم المجموعة الضابطة غير المعالجة وقياس قبلى وبعدى مزدوج :

$$\text{N}_\text{R}\text{O}_1\text{O}_2 \; \text{X} \quad \text{O}_3$$
$$\text{N}_\text{R}\text{O}_4\text{O}_5 \qquad \text{O}_6$$

جـ- تصميم المجموعة الضابطة غير المعالجة وقياسات بعدية وقبلية مزدوجة:

$$\text{N}_\text{R} \quad \text{O}_1 \; \text{X} \; \text{O}_2\text{O}_3$$

$$N_R \quad O_4 \quad O_5 \quad O_6$$

د- تصميم المجموعة الضابطة غير المعالجة مع قياسات قبلية وبعدية باستخدام المجموعة الضابطة المعالجة المعكوسة:

$$N_R \begin{matrix} O_1 & X & O_2 \\ \hline O_3 & X & O_4 \end{matrix}$$

- ***تصميم السلاسل الزمنية: يتضمن:***

أ- تصميم السلاسل الزمنية البسيطة:

$$O_1 O_2 O_3 \; O_4 \; O_5 X \; O_6 \; O_7 \; O_8 \; O_9 \; O_{10}$$

ب- تصميم السلاسل الزمنية مع مجموعات ضابطة غير معالجة و غير متكافئة:

$$O_1 O_2 O_3 \; O_4 \; O_5 \; X \; O_6 \; O_7 \; O_8$$
$$O_9 O_{10} O_{11} \; O_{12} \; O_{13} \quad O_{14} \; O_{15} \; O_{16}$$

ولكن يبدو أن أكثر التصميمات شبة التجريبية الشائعة الاستخدام فى البحوث النفسية والتربوية هى تصميم المجموعة الواحدة وقياس قبلى وبعدى وكذلك تصميم القياس البعدى مع مجموعات غير متكافئة وتصميم المجموعة الضابطة غير المعالجة وقياس قبلى وبعدى.

والتصميمات التجريبية الأساسية تتناول أثر متغير مستقل واحد بأكثر من مستوى على متغير تابع واحد ويطلق عليها تصميمات العامل الوحيد Single factor designs وهذا النوع من التصميمات هى الأكثر استخدامًا فى علم النفس والتربية ولكن أحيانًا يستخدم الباحثون تصميمات معقدة Complex designs حيث يتناول معالجة متغيرين مستقلين فأكثر فى نفس التجربة وهذه النوعية يطلق عليها التصميمات العاملية Factorial Designs لأن المتغيرات المستقلة يطلق عليها عوامل ويوجد فيها نوعين من التأثيرات هى التأثيرات الرئيسية Main effects وتأثيرات التفاعل Interactions effects.

المنهج الارتباطى Correlational method or design

الهدف من إستراتيجية البحث الارتباطى هو دراسة ووصف العلاقة أو الارتباط بين المتغيرات ولا يحاول تفسير العلاقة ولا يهتم بأى محاولة للضبط أو التجريب أثناء دراسة المتغيرات كما هو سائد فى البحث التجريبى. فالبحث الارتباطى هو أحد الطرق للوصف فى صورة كمية موضحًا إلى أى درجة واتجاه العلاقة بين المتغيرات ويعتبره

البعض أحد أشكال المنهج الوصفى ولكنه منهجية مستقلة بذاتها. ويجب التأكيد أن اكتشاف العلاقة الارتباطية لا يعنى وجود ارتباط سببى فالارتباطية لا تعنى بالضرورة سببيةCorrelation doesn't mean Causality ولكن وجود علاقة ارتباطية يمكن أن تكون بداية أساسية لدراسات تنبئية ولدراسات تحليل المسار Path analysis. والبحث الارتباطى يعطى الوضع عن المتغيرات التى نبحث عنها والباحث لا يؤثر على الأحداث، والقياسات تكون غير متحيزة وهذا مظهر مهم للصدق الواقعى أو البيئى Ecological validity.

المنهج الوصفى Descriptive method

يستخدم هذا المنهج لوصف أو فحص الوضع الحالى لظاهرة ما، ويتضمن جمع البيانات لاختبار فروض أو الإجابة عن أسئلة تخص الوضع الحالى لمتغير من المتغيرات كما يحدث فى الواقع. وهذه المنهجية البحثية تستخدم بصورة شائعة فى تقدير وقياس الاتجاهات والآراء والمعلومات الديموجرافية، ويتم جمع البيانات من خلال الاستبيانات والمقابلات والملاحظات، وعليه فإن البحث الوصفى يهدف إلى وصف متغير وحيد بهدف الحصول على توصيفات له.

وهذه الإستراتيجية على جانب كبير من الأهمية فى الدراسات الاستكشافية وتتناول الظواهر الجديدة فى المجتمع مثل القضايا المرتبطة باستخدام وسائل التواصل الاجتماعى كالفيسبوك مثل: ما نسبة استخدام الفيسبوك بين طلاب المراحل التعليمية المختلفة أو الاتجاه نحو المناهج الجديدة فى المدارس.

المنهج السببى المقارن - Causal comparative method or Post facto comparative

يشير هذا المنهج إلى الدراسات التى تفحص علاقات السبب والتأثير (النتيجة) عن طريق الملاحظة لحالات أو وقائع معينة وعلى ذلك فإن المتغير المستقل حدث بالفعل أو موجود بالفعل ويريد الباحث معرفة عواقبه الحاضرة. ويرى البعض أن فلسفة البحث السببى المقارن تقترب من البحث شبه التجريبى ولكن المتغير المستقل تصنيفى بمستويين (ذكر – أنثى، مرتفع– منخفض) موجودة بالفعل ويريد دراسة الفروق بينهما على متغير تابع ما ولكن البحث السببى المقارن عكس البحث التجريبى حيث فى

البحث التجريبي يتحكم الباحث في المتغير المستقل بإجراءات الضبط، بينما البحث السببى المقارن المتغير المستقل موجود بالفعل بدون إجراء تجربة وإجراءات ضبط. وكأمثلة عليه:

- الفروق بين الذكور والإناث فى العدوان.
- الفروق بين المدخنين وغير المدخنين فى الإصابة بالسرطان.

ويلجأ الباحث إلى البحث السببى المقارن لدراسة ظاهرة ما عندما يتعذر دراستها تجريبيًا لاعتبارات أخلاقية.

وفيما يلى ملخص التصميمات البحثية والاختبارات الإحصائية المصاحبة لها:

الجدول (1.2) : الملامح الأساسية للتصميمات البحثية والاختبارات الإحصائية المصاحبة لها

	خصائصه:	الاختبارات الإحصائية
التصميم التجريبي	-معالجة للمتغير المستقل -توزيع عشوائى للأفراد على المجموعات. -ضبط صارم. -مقارنة بين المجموعات	-اختبار T المستقلة -اختبار تحليل التباين (*ANOVA*) -اختبار مان-ويتني(*M.W*) -*MANOVA* -تحليل التبان المتعدد -تحليل التغاير
التصميم شبه التجريبي	- معالجة للمتغير المستقل - توزيع غير عشوائى للأفراد على المجموعات -مقارنة بين المجموعات	-اختبار Tالمستقلة -اختبار T المرتبطة -ANOVA -اختبار مان-ويتني(*M.W*) -اختبار ويلكوكسون

التصميم الارتباطى	- تحديد درجة التغاير أو الارتباط بين المتغيرات -لا يمكن استنتاج السببية من الارتباطات -تمهد للدراسات التنبؤية و الدراسات السببية	-معامل ارتباط بيرسون(r) -معامل ارتباط سيرمان(rs) -تحليل الانحدار -تحليل المسار

5.العينـــة أو المشـــاركون Sample: يـتم انتقـاء العينـة بصـورة عشـوائية مـن مجتمـع الدراسـة، والمسـلمة الأساسـية فـى الإحصـاء الاسـتدلالى هـى العشـوائية. والإحصـاء ليسـت مقطوعـة الصـلة بهـذا الجانـب فـى البحـث لأن العشـوائية تـتم فـى ضـوء قـوانين الاحتمـالات، وتلعـب الإحصـاء دورًا كبيـرًا فـى اختيـار العينـة فـى ضـوء مـدخل القـوة الإحصـائية باسـتخدام التحليـل القبلـى للقـوة الإحصـائية لتحديـد حجـم العينـة المناسـب وكـذلك فـى ضـوء مـدخل فترات الثقة (للمزيد انظر العينات).

6. إعـــداد أدوات القيـــاس وجمـــع البيانـــات: الإجابـة المنظمـة لأسـئلة البحـث تتطلـب قياسـات صـادقة وفـى هـذه الخطـوة يسـهم علـم القيـاس النفسـى بإسـهامات فعالـة حيـث يقـدم أسـس ومبـادئ إعـداد أدوات القيـاس مثـل المقـاييس والاختبـارات والاسـتبيانات ومـن شـروط أداة القيـاس الجيـدة هـى الصـدق والثبـات. فالصـدق يعطـى ادلـة علـى مصـداقية بنـاء الأداة، أمـــا الثبـــات يعطـــى موثوقيـــة للبيانـــات المـــراد تحليلهـــا ومعرفـــة أخطـــاء القيـــاس Measurement errors. فـى هـذه الخطـوة تلعـب الإحصـاء دورًا كبيـرًا بـالغ الأهميـة وتقدم معادلات لحساب أدلة الصدق ومؤشرات الثبات.

7.المعالجـــة الإحصـــائية وتحليـــل البيانـــات: تـتم المعالجـة الإحصـائية فـى ضـوء فـروض البحـــث وكـــذلك الفـــروض الإحصـــائية وطبيعـــة المتغيـــرات المقاسة(متصـــلة- منفصـــلة)، وحجـم العينـة وأهـداف التحليـل (فـروق- علاقـة- تنبـؤ)، وطبيعـة توزيـع البيانـات (اعتداليـة- ملتويـة) كـل هـذه العوامـل تحـدد نوعيـة الإحصـاء المسـتخدمة سـواء كانـت إحصـاء بـارامترى Parametric statistics أو إحصـاء لابـارامترى Non-Parametric statistics ، وكذلك استخدام خطوات اختبارات الفروض الصفرية.

8. تفسير النتائج والمناقشة Interpretation and Discussion : بعد المعالجة الإحصائية وما تمدنا به من قرارات ونتائج يأتى دور الباحث فى تفسير النتائج وما إذا كانت الفروق تعود إلى الصدفة أم إلى المعالجة المستخدمة وقد أوضحت الجمعية النفسية الأمريكية (APA) أن تفسير النتائج لا بد أن يكون مصحوبًا بمؤشرى الدقة الإحصائية وهما حجم التأثير وتقدير القوة الإحصائية.

9.التعميم Generalization: بعد تفسير النتائج يقوم الباحث بإصدار العديد من التوصيات المرتبطة بنتائج دراسته ولكن لكى يصدر الباحث توصيات قابلة للتعميم فلا بد من التأكد من صدق التعميم Cross- validation لدراسته إذا ما أعيدت على عينات أخرى من نفس المجتمع وهوما يطلق عليه بالدراسة التتبعية.

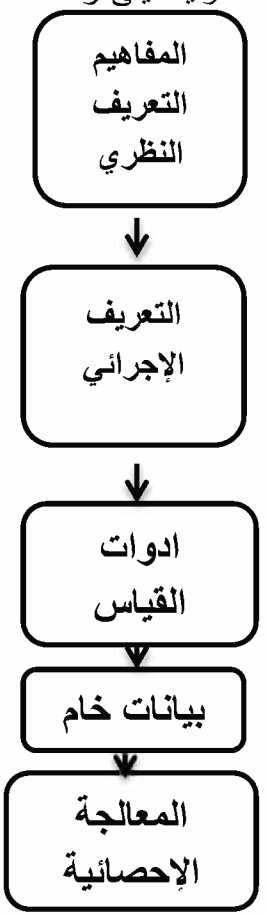
وفى ضوء ذلك تؤدى الإحصاء دورًا أساسيًا فى كل خطوات وعمليات البحث المختلفة ابتداءً من صياغة المشكلة وانتهاءً بالتعميم وهذا عكس ما يعتقده البعض بإن الإحصاء هى وسيلة تستخدم فقط فى تحليل البيانات.

الفصل الثالث

المتغيرات والبيانات

Variables and Data

تتعامل الإحصاء مع البيانات Data وهذه البيانات تأتى من مقاييس أو استبيانات أو اختبارات بهدف قياس متغيرات ما. وفيما يلى رحلة الحصول على البيانات:



الشكل (1.3): رحلة الحصول على البيانات.

فى أى دراسة يتعامل الباحث مع مجموعة بيانات Data set يحصل عليها من مرحلة جمع البيانات Data collection وفيها يمتلك أى فرد أو حالة فى عينة الدراسة مجموعة بيانات تعكس مجموعة متغيرات مثل الجنس – التحصيل – الذكاء .... إلخ، حيث يطلق على بيانات متغيرات مثل الجنس، موقع السكن، التخصص، الوظيفة، والحالة الاجتماعية بالبيانات الكيفية Qualitative data ويسميها البعض بالبيانات الأساسية أو الديموجرافية وهى تقاس بصورة كيفية تصنيفية ويمكن أن يكون

المتغير الكيفى ثنائى التصنيف Dichotomous مثل الجنس (ذكر- أنثى) أو متعددة التصنيفات مثل الديانة (مسلم- مسيحى - يهودى)، ويطلق على بيانات متغيرات مثل الابتكارية، الاتجاه، العنف، ومفهوم الذات بالبيانات الكمية Quantative data وهى تقاس بصورة كمية أو رقمية ويتم تحليلها بإجراءات إحصائية معينة.

والمتغير صفة أو ظاهرة أو سمة لها حالات أو قيم متعددة ومختلفة وتتغير من وقت لآخر، وبعد أن يطرح الباحث أسئلة بحثه ثم يقوم بترجمة المفاهيم النظرية المجردة إلى مفاهيم إجرائية Operational concepts وهى عبارة عن متغيرات قابلة للقياس.

تصنيف المتغيرات

تصنف المتغيرات Variables فى ضوء:

أولًا: مستوى القياس الذى تنبع منه: وفى هذا الإطار تصنف المتغيرات حسب مستوى القياسLevel of measurement أو مقياس البيانات Data scale هرميًا من الاقل رقيًا (الأسمى) إلى الاكثر رقيًا (النسبى):

1. مستوى القياس الاسمى Nominal scale: فى هذا المستوى البيانات عبارة عن تصنيفات مثل الديانة، الجنس، الوظيفة، التخصص، وغيرها. فمثلًا التخصص يتضمن عدة مستويات مثل الطب، الهندسة، الآداب، العلوم، والزراعة. والمقاييس الاسمية تعطى بيانات تصنيفية Categorical data ويوضع الفرد فى أحد التصنيفات ويمكن أن تكون تصنيفين فأكثر. والتوظيف الإحصائى فى هذا المستوى محدود لأنها بيانات كيفية ولا تأخذ قيم رقمية والارقام فى هذا المستوى للتكويد Coding وذلك بهدف ادخالها لبرنامج الكمبيوتر، فالجنس يتضمن ذكر = 1 ، أنثى = 2 ، فالرقم لا يعنى أن الانثى ضعف الذكر. والمتغيرات الاسمية يتم تكويدها بالأرقام وهى تستخدم لخلق متغير Dummy variable وهى اعتباطية ولا معنى لها ويمكن أن يكون التكويد حرفى مثل A, B, C.

والبيانات الاسمية لا تسمح باستخدام الإحصاء البارامترى، ودائمًا الإحصائيات عبارة عن العدد فى كل تصنيف وحساب التكرارات والنسبة المئوية فى كل تصنيف وكذلك المنوال كمؤشر للنزعة المركزية (القيمة الاكثر شيوعًا أو تكرارًا).

2. مستوى القياس الرتبى Ordinal scale: هذا المستوى فى غاية الأهمية لترتيب المتغيرات الكيفية مثل ترتيب الطلاب من حيث إنجازهم للواجبات المدرسية أو التقدير فى المدرسة أو الترتيب فى الاسرة أو الرتبة العسكرية (لواء –عميد – عقيد) والحالة الاقتصادية (مرتفع – متوسط– منخفض) وبالتالى يتضمن هذا المستوى التصنيف ثم الترتيب من الرتبة الأعلى إلى الاقل أو العكس. ويرى Miller (2014) أن الارقام فى هذا المستوى لها معنى وتعطى الارقام فى شكل رتب. ويمكن أن تسأل فرد ما عن درجة تفضيله لشيء معين فى ضوء مقياس يتكون من عشر تفضيلات حيث (10) تفسر على انها أكثر التفضيلات ايجابية و(1) يعكس أقل التفضيلات ايجابية. وإذا طلب من فرد ترتيب تفضيلاته لأطعمة مختلفة، فيختار (10) حيث تعكس أكثر الاطعمة تفضيلًا، بينما(1) أقل الأطعمة تفضيلًا.

وعلى ذلك فالهدف من هذا المقياس هو الترتيب، والترتيب لا يعطى معلومات عن مقدار الرتب. مثال: **على أطول من محمد**، ولكن ليس بمقدار كمى ولكن فى ضوء ظاهر السمة، والترتيب للأشياء يقع على متصل وفيها يحدد موقع التصنيف على هذا المتصل. واستخدام مؤشرات المتوسط والانحراف المعيارى غير مناسبة فى هذا المستوى ويمكن استخدام التكرارات والنسب المئوية وكذلك يعتبر الوسيط مقياس مناسب للنزعة المركزية للدرجات المرتبة ذلك لأن الوسيط يتطلب فقط نفس العدد من القيم تقع فوق أو تحت النقطة المنتصفية Midpoint. ولذلك فإن الإحصاء اللابارامترى هو المناسب مع بيانات مستوى القياس الرتبى.

ويمكن القول إن المقاييس النفسية التى يتم بنائها فى ضوء مقياس ليكرت تقع ضمن مستوى القياس الرتبى وتتم معالجة البيانات الرتبية كأنها بيانات فترية باستخدام الإحصاء البارامترى.

3. المقيـاس الفتـرى Interval scale: يتضـمن التعبيـر الكمـى ويسـتخدم الارقـام للتعبيـر عـن الصـفات أو الخصـائص، والفتريـة الحقيقيـة تتضـمن توزيـع متصـل وعلـى ذلـك فـإن الفتـرات بـين الارقـام لهـا معنـى، فدرجـة الحـرارة (45) هـى أقـل خمـس درجـات مـن درجـة الحـرارة (50)، ودرجـة العدوانيـة تتـراوح مـن (20) إلـى (80)، والدرجـة (60) أكثـر بعشـر وحـدات مـن الدرجـة (50) ولكـن (40) لـيس هـى ضـعف (20).البيانـات علـى مقيـاس ليكـرت الخماسـى أو السـباعى أو التسـاعى يمكـن التعامـل معهـا علـى أسـاس بيانـات فتريـة. والدرجـة علـى متصـل السـمة لهـا معنـى مختلـف فالدرجـة التحصـيلية بـين 9,10 تختلف عن الدرجة التحصيلية بين 1,2.

والـدرجات علـى المقيـاس الفتـرى تعـالج الفـروق فـى الـدرجات كأنهـا مسـافات متسـاوية بـين هـذه الـدرجات، ولا يوجـد صـفر كنقطـة بدايـة، والصـفر لـيس نقطـة حقيقيـة علـى المقيـاس بـل هـى افتراضـية ويسـمى صـفر نسـبى أو اعتبـارى Relative zero حيـث انعـدام الصـفة علـى المسـتوى المقـاس لا يعنـى انعـدامها علـى المسـتوى الحقيقـى. والعلاقـة بين الدرجة المقاسة X والدرجة الحقيقية T :

$$T= X + E$$

- E خطأ القياس

وإذا كانت $X = 0$ فان:

$$T= E$$

وبيانـات العلـوم النفسـية والاجتماعيـة تحلـل علـى أسـاس أنهـا بيانـات فتريـة ويعتبـر اسـتخدام إجـراءات الإحصـاء البـارامترى لتمثيـل البيانـات الفتريـة محـل تسـاؤل لأسـباب عديـدة أهمهـا ان اعتداليـة البيانـات الفتريـة هـى محـل جـدل. عمومًـا متغيـرات العلـوم النفسـية والاجتماعيـة التـى تقـاس بطريـق غيـر مباشـر بصـورة كميـة تقـع تحـت هـذا المسـتوى. وعلـى ذلـك فـإن القياسـات الفتريـة تسـعى إلـى التصـنيف والترتيـب مـن الاقـل إلـى الاكبـر وكـذلك توضـح النسـب بـين المجموعـات المتكافئـة مـن الارقـام ويمكـن اسـتخدام إحصـائيات مثـل المتوسـط والانحـراف المعيـارى وهـذه ضـرورية للإحصـاء البـارامترى مثـل اختبـارات T و ANOVA و r حيـث تفتـرض اعتداليـة البيانـات. وعلـى الـرغم ان الـبعض ينـادى بضـرورة

معالجة بيانات العلوم الاجتماعية والنفسية باستخدام إحصاء لابارامترى على أساس ان البيانات المتولدة من مقاييس ليكرت هى فى طبيعتها رتبية أكثر من كونها فترية.

4. المقياس النسبى Ratio scale: يستخدم البيانات الكمية ولها فترات متساوية بين الارقام ويتوفر الصفر الحقيقى حيث انعدام الصفة على المستوى المقاس يفيد بانعدامها على المستوى الحقيقى، وايضا فالصفر يعنى لا شئ فالدخل والطول والمسافة والوقت هما متغيرات نسبية، فالفرق بين الوزن 40 و 50 هو نفسه الفرق بين 60 و70. ويمكن القول إن الاستجابة A هى تساوى مرتين أو ثلاث مرات الاستجابة C ولكن هذا غير متاح فى مستوى القياس الفترى لعدم توافر الصفر الحقيقى فلا نستطيع القول إن درجة طالب 100 فى الامتحان تساوى ضعف طالب آخر درجته 50.

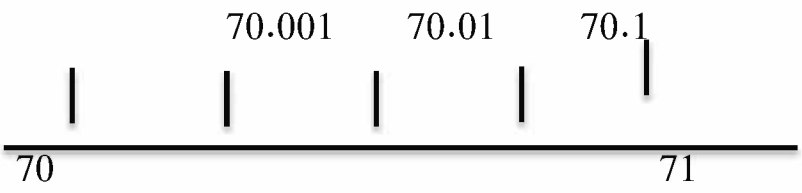
والعلاقة بين X و T كالآتى:

$$T= X$$

ويمكن إجراء العمليات الحسابية الاربعة على القيم مباشرة وكذلك على الفترات بين النقاط على المقياس. وعلى ذلك فهذا المستوى يهدف إلى التصنيف (الأسمى) والترتيب (ترتيبى) والكشف عن النسب بين أى فترتين أو قيمتين ويتوفر له الصفر المطلق Absolute Zero. البيانات فى العلوم النفسية والسلوكية والاجتماعية لا تخضع لهذا المستوى لأنها لا تلبى متطلباته ويستخدم المتوسط والانحراف المعيارى كمعالم للمجتمع وعليه فإن الإحصاء البارامترى المناسب لبيانات هذا المستوى.

ثانيًا: تصنيف المتغيرات حسب اتصاليتها الى:

1. متغيرات متصلة (مستمرة) Continuous variable: تعنى ان عدد القيم الواقعة بين أى قيمتين (نقطتين) على متصل السمة لا نهائى وغير قابل للعد، فمثلًا وزن 70 كجم ووزن 71 كجم تم تمثيلهما على متصل ما كالآتى:



وإذا كانت اداة قياس تتميز بالحساسية الشديدة فإن عدد الأوزان الواقعة بين القيمتين 70, 71 لا نهائى، والمتغيرات المتصلة غير قابلة للعد نظريًا، ودائمًا المتغيرات الفترية والنسبية هى متغيرات متصلة. وكأمثلة على المتغيرات المتصلة: العمر، الوزن، الطول، المسافة، والأمن النفسى. ولكن عدد أفراد الأسرة ليس متغيرًا متصلًا على الرغم أنه نسبى.

2. المتغيرات المتقطعة (المنفصلة) Discrete variables: هى المتغيرات القابلة للعد مثل متغير الجنس (ذكر وأنثى) أو متغير موقع السكن (الريف- البدو - الحضر) بمعنى لا توجد قيم بين أى قيمتين للمتغير، فمثلًا بين الذكور والريف لا يوجد شئ آخر بينهما، والمتغيرات الاسمية مثل الجنس، الديانة، التخصص فى الكلية، والوظيفة هى متغيرات متقطعة فى طبيعتها وقابلة للعد ويطلق عليها أيضًا **متغيرات تصنيفية Categorical variable**. وهى متغيرات كيفية وقد تكون غير رتبية مثل التقدير فى الكلية وقد تكون كمية مثل عدد أفراد الأسرة حيث هى متغيرات محدودة المدى.

وتصنف حسب وظيفتها فى التصميم التجريبى أو التصميم البحثى كالآتى:

1. **المتغير المستقل Independent variable:** هو المتغير الخاضع لسيطرة الباحث أو سيطرة المجرب، وهو خاصية أو سمة تؤثر فى الناتج أو التأثيرات ، ويمكن للمتغير المستقل ان يؤثر فى المتغير التابع من خلال متغير وسيط أو ناقل بالتالى يكون التأثير غير مباشر، وتسمى المتغيرات المستقلة متغيرات منبئة، معالجات، ومتغيرات سابقة، فمثلًا طريقة التدريس أو التعزيز، وما وراء المعرفة كلها متغيرات يسيطر عليها الباحث ومستقلة للتحصيل، فالباحث الذى يدرس أثر المتغير (X) على متغير آخر (Y) فإن المتغير (X) مستقل ولا بد أن يكون سابق فى الحدوث زمنيًا عن المتغير (Y) ويسمى المتغير المستقل أحيانًا فى التصميمات التجريبية بمتغير المعالجة Treatment variable، وفى الدراسات التنبؤية يسمى المتغير المنبئ. وإذا أراد الباحث أن يتنبأ بالنجاح الأكاديمى فى الجامعة من درجات الثانوية العامة فى هذه الحالة فإن امتحانات الثانوية العامة متغير مستقل، أى أن المتغير المستقل هو السبب لحدوث تغير فى متغير آخر تالٍ له. الدراسة أو البرنامج هو متغير مستقل. و**متغيرات**

المعالجـــة هـــى نـــوع مـــن المتغيـــرات المســـتقلة وتســـتخدم فـــى البحـــوث التجريبيـــة والمعالجـــات التربويـــة، فيعطـــى المعالجـــة لمجموعـــة بينمـــا يســـتبعدها مـــن المجموعـــة الأخـــرى. ويصـــبح الســـؤال مـــا إذا كانـــت المجموعـــة التـــى تلقـــت المعالجـــة تتفـــوق أو تختلـــف علـــى المجموعـــة التـــى لـــم تتلـــق المعالجـــة. بالتـــالى فـــإن هـــذا المتغيـــر تصـــنيفى يمكـــن أن يكـــون بمســـتويين فـــأكثر. وفـــى التـــراث البحثـــى تســـمى هـــذه المتغيـــرات بالمعالجـــات أو المســـتويات أو ظـــروف المعالجـــة. والمتغيـــرات المســـتقلة فـــى التربيـــة وعلـــم الـــنفس كثيـــرة مثـــل برنـــامج علاجـــى أو تـــدريبى، مهـــارات مـــا وراء المعرفـــة، أثـــر عـــادات العقـــل، أثـــر عـــادات الاســـتذكار، عـــدد ساعات الاستذكار، العلاج السلوكى، العلاج العقلانى، وغيرها.

2. المتغيـــر التـــابع Dependent variable: هـــو خاصـــية أو ســـمة تعتمـــد علـــى أو تتـــأثر بـــالمتغير المســـتقل ويكتـــب فـــى تقـــارير البحـــوث تحـــت مســـمى النـــاتج، التـــأثير، المحـــك، ويمكـــن للدراســـة الواحـــدة ان تتضـــمن أكثـــر مـــن متغيـــر تـــابع ويمكـــن أن يكـــون قياســـات كميـــة متصـــلة أو قياســـات كيفيـــة منفصـــلة. هـــو المتغيـــر النـــاتج ويعتمـــد علـــى ســـلوك المســـتجيبين ويتغيـــر حدوثـــه نتيجـــة تغيـــر المتغيـــر المســـتقل أى أنـــه النتيجـــة، فالنجـــاح الأكـــاديمى فـــى الجامعـــة هـــو المتغيـــر التـــابع وهـــو المتغيـــر الـــذى يـــأتى بعـــد المتغيـــر المســـتقل أو يحـــدث معـــه فـــى نفـــس اللحظـــة، ونـــواتج الدراســـة (البيانـــات) هـــى متغيـــرات ناتجـــة تابعـــة. وامثلـــة للمتغيـــرات التابعـــة فـــى التربيـــة وعلـــم الـــنفس التحصـــيل، تنميـــة الابتكاريـــة، تنميـــة الاتجـــاه، خفـــض القلـــق، تنمية الذكاء الاجتماعى، المهارات الإدارية، المناخ الأسرى، خفض الاكتئاب.

3. المتغيـــر الضـــابطة Control variable : هـــى متغيـــرات مســـتقلة ولكـــن يضـــبط اثرهـــا اثنـــاء التجربـــة، فمـــثلا اراد باحـــث دراســـة اثـــر التماســـك الاســـرى علـــى الاكتئـــاب، فلدراســـة هـــذا فإنـــه توجـــد متغيـــرات اخـــرى تـــؤثر علـــى الاكتئـــاب مثـــل المســـتوى الاقتصـــادى للأســـرة والوحـــدة النفســـية وهـــى متغيـــرات تـــؤثر علـــى النـــواتج فـــلا بـــد أن يضـــبطها الباحـــث ويســـتبعد اثرهـــا عنـــد تصـــميمه للتجربـــة مـــن خـــلال قياســـها. ومـــن امثلـــة هـــذه النوعيـــة مـــن المتغيـــرات الحالـــة الاقتصـــادية الاجتماعيـــة، النـــوع، الجـــنس، والـــذكاء. ويمكـــن ضـــبط هـــذه المتغيـــرات إحصـــائيًا وإذا كانـــت متغيـــرات متصـــلة يطلـــق عليهـــا تغـــايرات Covariates مـــن خـــلال أســـاليب إحصائية مثل تحليل التغاير Analysis of Covariance

4. **المتغير الوسيط Mediating variable**: هى متغيرات تتأثر بالمتغير المستقل وتؤثر على المتغير التابع، بمعنى انها تنقل اثر متغير مستقل إلى المتغير التابع ويقيسها الباحث وتدخل فى المعالجة الإحصائية، وهى متغيرات شائعة فى النماذج السببية وتلعب دور المتغير المستقل والتابع فى نفس الوقت. وكأمثلة للمتغيرات الوسيطة البناء المعرفى والذكاء وغيرها وهى متغيرات فترية. ويرى Creswell (2012) ان المتغيرات الوسيطة هى متغيرات متداخلة **Intervenig** حيث انها تختلف عن المتغيرات التابعة وأى نوع من المتغيرات المستقلة، فباستخدام تفكير السبب والنتيجة، فإن بعض العوامل تتداخل بين المتغير المستقل والتابع وتؤثر فى المتغير التابع بل ينقل أثر المتغير المستقل إلى المتغير التابع، بالتالى تتوسط العلاقة بين المستقل والتابع. وفى بعض الدراسات الكمية يمكن ضبطها باستخدام أساليب إحصائية.

5. **المتغير الدخيل Confoudnig**: هى متغيرات غير مقاسة وغير ملاحظة وتلعب دورًا فى التصميم التجريبى أو فى العلاقة السببية بين المتغير المستقل والتابع ويطلق عليها بالمتغيرات المشوشة أو Spurious variables ولا يمكن استبعادها من هذه المتغيرات وتؤثر فى طبيعة العلاقة بين المتغيرات المستقلة والتابعة.

ويمكن أن يسأل الباحث عدة تساؤلات عن طبيعة هذه المتغيرات كالتالى:

- ما هو النواتج فى دراستى التى تحتاج لتفسير؟ (متغيرات تابعة).
- ما هى المتغيرات أو العوامل التى تؤثر فى الناتج؟ (متغيرات مستقلة).
- ما هى المتغيرات التى احتاج لقياسها والتى من المحتمل أن تشكل عوامل أساسية فى تأثيرها على النواتج وليس متغيرات أخرى؟ (المتغيرات الضابطة).
- ما هى المتغيرات التى ربما تؤثر على نواتج الدراسة ولا يمكن قياسها؟ (المتغيرات الدخيلة).

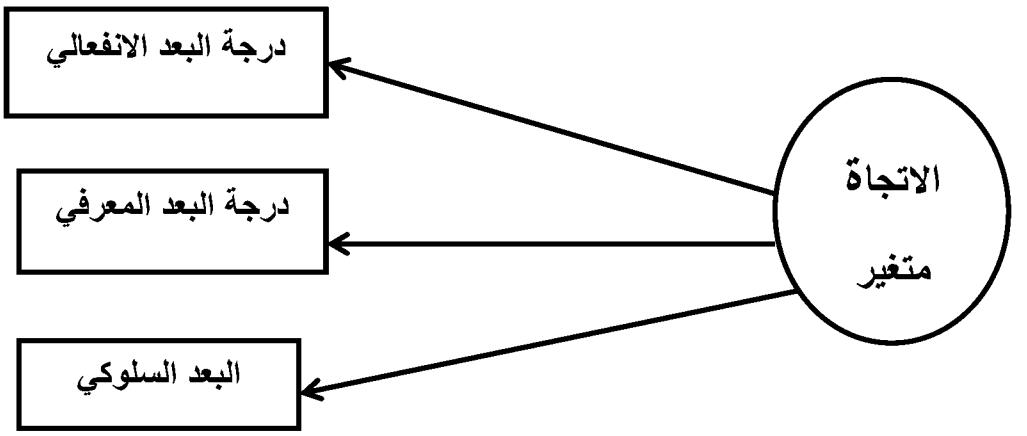
6. **المتغيرات المتفاعلة Moderatig variables**: هذه المتغيرات تستحق الانتباه وغالبًا ما تستخدم فى التصميمات والتجارب التربوية، هذه المتغيرات هى جديدة وتتولد

نتيجة التفاعل مع متغيرات اخرى فمثلًا عن طريق تفاعل مستويات المتغيرات المستقلة المعالجة ويطلق عليها تأثير التفاعل Interaction effect ومثال لهذا: الطلاب الذكور مع طريقة المناقشة يحصلوا على درجات تحصيلية أكثر من الطالبات الإناث مع المحاضرة.

وتصنف حسب التعامل معها فى التحليلات الإحصائية الى:

1.المتغيرات المقاسةObserved variables: وهذا ما يحدث فى معظم المتغيرات التى يتم الاستدلال عليها مباشرة من ادوات القياس مثل الذكاء والدافعية حيث إن الدرجة الكلية فى مقياس الذكاء أو فى مقياس الدافعية تسمى الدرجة المقاسة بعض النظر عن كون المقياس مكون من بعد واحد أو من ابعاد عديدة.

2. المتغيرات الكامنة أو الخفية Latent variables : وهى تعكس البناء التحتى للسمة موضع القياس، وهذا يعكس المتغيرات غير المتضمنة فى التحليل الإحصائى الذى يعتمد على الدرجة الكلية للمقياس واحيانًا يتعامل الباحث مع درجة كل بعد على حدة، ولكن إذا تمكن الباحث من تجميع كل الأبعاد فى مكون افتراضى فإنه يسمى متغير كامن أو عامل، وكمثال عليه فالاتجاه له ثلاثة ابعاد البعد الانفعالى والبعد السلوكى والبعد المعرفى فإذا تعامل الباحث مع المكونات الثلاثة فى مكون واحد كالآتى:



الشكل (3.2): العلاقة بين الابعاد الفرعية المقاسة والمتغيرات الكامنة.

فيقال إن الأبعاد الثلاثة كونت عامل (متغير كامن).

طبيعة البيانات

تدخل البيانات فى كل مجالات الحياة، فالتلميذ فى المدرسة له ملف يتضمن بيانات عديدة عن عمره واسرته وميلاده والامراض إلى اصيب بها فى الطفولة وتحصيله فى المراحل الدراسية السابقة، ويعتبر هذا الملف أو الاستمارة كنز بياناتى، فالبيانات هى جمع ومفردها Datum وتكتب فى تقرير البحث (حللت البيانات Data were analysed.....)، والباحث عند إجراء وإتمام دراسته فإنه يمر بمرحلة جمع البيانات من خلال قياسات أو استبيانات أو مقابلات وغيرها وهذه الادوات امثلة واضحة لجمع البيانات.

ويتم جمع البيانات من الأفراد وهؤلاء يعتبروا وحدات تحليل Units of analysis (ملاحظات – أفراد – حالات)، والخاصية التى تم قياسها أو جمع بيانات لها تسمى متغير وهى خاصية تأخذ قيم مختلفة، والاستجابات المرتبطة بالأفراد على هذه الخصائص (المتغيرات) يطلق عليها قيم Values، بينما أى قيمة لفرد ما على السمة أو الخاصية تسمى درجة أو قيمة Score.

وعندما يتم التعامل مع وحدات التحليل وهى المتغيرات والقيم معا تسمى بيانات ويؤكد Diamantopoulos & Schlegelmilch (2003) إلى أنه عندما نشير إلى البيانات فإنها تتضمن:

أ. الموضوع الذى نهتم بدراسته (يتم وصفه بمجموعة من المتغيرات).

ب. المستجيبين (وحدة التحليل).

ج. الاستجابات المرتبطة بالموضوع (القيم).

على ذلك تكون البيانات لمتغيرات حيث تتنوع قيم المتغير للخاصية بمعنى يأخذ أكثر من قيمة، ولكن إذا اخذت كل وحدات التحليل (الأفراد) نفس القيمة فى هذه الحالة يطلق عليها ثابت Constant. والامثلة على المتغيرات كثيرة فى العلوم النفسية والاجتماعية مثل الذكاء، التنمر، الوحدة النفسية، أساليب التفكير، المخاطرة، والشخصية، الامن المجتمعى، إدمان الإنترنت، وغيرها.

وعندما يتم اتحاد أو جمع وحدات التحليل (الأفراد) مع المتغيرات ووضع القيم الخاصة بالأفراد على المتغيرات فإن هذا يسمى بمجموعة البيانات Data set، وعندما

يتم عرض هذا الدمج أو الجمع فى صفوف وأعمدة فى تمثيل أو عرض بصرى يطلق عليه مصفوفة البيانات كالآتى:

الجدول (1.3): مثال لمصفوفة بيانات لمجموعة من المتغيرات والأفراد.

م	وحدات التحليل	الجنس	العمر	دخل الاب (جنيه)	التحصيل (50)	الذكاء
1	**على**	**ذكر**	**15**	**1200**	**40**	**105**
2	**دينا**	**انثى**	**16**	**200**	**35**	**99**
3	**احمد**	**ذكر**	**17**	**700**	**30**	**90**
.	.	.	.	.	.	.
45	**هانى**	**ذكر**	**20**	**1500**	**25**	**100**

وبالتالى إذا وجد عدد n من الوحدات (الأفراد) وعدد m من المتغيرات فإنها تكون مصفوفة البيانات كما فى الجدول السابق.

والاستفادة الرئيسية من اعاده ترتيب أو وضع البيانات فى شكل مصفوفة البيانات هى ان تصبح المنظومة البيانية ذات تمثيل بصرى، وتعتبر مصفوفة البيانات نقطة البداية للتحليل وطبيعتها وتنظيمها يمكن أن يساعد فى تحديد طبيعة الأساليب الإحصائية المستخدمة. وإذا تأملنا إلى صفوف المصفوفة (n) فإنها تشير إلى كم أو عدد الوحدات (الأفراد) فى التحليل وهذا يعتبر حجم العينة أو المجتمع Sample or Population size، وعدد الأعمدة (m) يشير إلى عدد المتغيرات التى استخدمت أو قيست بالفعل على عينة البحث.

وبمعرفة عدد المتغيرات يمكن تطبيق الأساليب أو التكنيكات الإحصائية فاذا كانت:

- عدد المتغيرات التابعة= **1** فإننا نطلق عليها أساليب تحليل بيانات إحصائية Univariate
- عدد المتغيرات التابعة2 > فإنها تستخدم أساليب تحليل بيانات متدرجة Multivariate
- عدد المتغيرات فى التحليل=**2** فإننا نطلق عليها أساليب تحليل بيانات Bivariate

وطبيعة قيم المتغيرات تحدد من خلال مستوى القياس، وتجمع البيانات من الأفراد من خلال صيغ عديدة قد تكون توجيه أسئلة للأفراد أو عبارات تقريرية وهذا فى صورة

استبيانات Questionnaires وقد تكون الاستبيانات صيغتها موحدة بمعنى تقدم الأفراد بنفس النمط من العبارات أو قد تتضمن أسئلة أو عبارات مختلفة مثل أسئلة لها اختيارات (بدائل) أو أسئلة مفتوحة وفى هذا تسمى استجابات متعددة Multiple responses.

انواع البيانات

توجد أنواع متعددة من البيانات، قسمها Diamantopoulos & Schlegelmilch (2003) إلى:

اولًا: فى ضوء معناها: وتقسم الى:

أ- **حقائق** Facts: وهى عباره عن خصائص أو مواقف موجودة منذ زمن بعيد مثل العمر والجنس والدخل والديانة وغيرها ويطلق عليها فى الدراسات بالبيانات الأساسية أو البيانات الديموجرافيةDemographic data.

ب- **معارف:** وهى معارف الفرد عن قضية ما أو مخترع ما أو صاحب انجاز رياضى ما وغيرها.

ج- **اتجاهات واراء** Attitudes and opinions : حيث يشير إلى وجهات النظر أو التفضيلات أو مشاعر الأفراد نحو ظاهرة معينة مثل اتجاه التلاميذ نحو المدرسة أو رؤية الأفراد عن المناخ المدرسى أو المعتقدات السياسية أو الاتجاهات الشرائية وغيرها من القضايا والمشكلات الحياتية.

د- **الدوافع أو البواعث للأفراد** Motives: وهى القوى الداخلية (الرغبات – الحاجات – البواعث – الدوافع) التى تحرك أو تدفع الفرد نحو سلوكيات فى اتجاه معين، وعلى الرغم ان الدوافع من المظاهر المعقدة الا أنه من الضرورى محاولة التعرف عليها لأنها تبين لماذا يسلك الأفراد سلوكيات معينة دون غيرها؟ وكذلك تهتم بالتفسير لتفضيلات الأشخاص لأشياء دون غيرها.

و- **النوايا** أو **النيات** Intentions: تهتم بماذا يدور فى نفوس وعقول الأفراد فماذا يريد الفرد أن يأتى بسلوك ما مستقبلًا، ماذا يرغب الطالب ان يتخصص فى الجامعة؟

ثانيًا: فى ضوء مصدرها: تقسم البيانات فى ضوء من أين حصلنا عليها وفى هذا الإطار يجب التميز بين البيانات الثانوية Secondary data والبيانات الأولية Primary data.

- **البيانات الأولية:** تجمع لأغراض محددة وواضحة عند الباحث، فمثلًا اراد باحث دراسة العلاقة بين أساليب التفكير والتحصيل فهدفه جمع البيانات عن هذين المتغيرين لدراسة العلاقة بينهما وذلك لاستكمال مشروع بحثه ومثل هذه النوعية من البيانات تجمع ما خلال مسح Survey باستخدام أساليب مختلفة سواء استبيانات أو مقابلات أو عبر البريد أو من تجارب Experiments (فى معمل أو فى المواقف الطبيعية) أو من طرق الملاحظةObservation methods باستخدام اجهزة فنية أو تسجيل الملاحظات أو غيرها.
- **البيانات الثانوية:** هى البيانات التى لم تجمع مباشرة للدراسة ولكنها تجمع لأغراض اخرى من مصادر اخرى مثل الإحصائيات المطبوعةPublished statistics عن طريق الحكومات أو المؤسسات أو الجمعيات العلمية أو المؤسسات البحثية أو التقارير السنوية Annual reports التى تعطيها الشركات التجارية أو المنظمات أو ملخصات البحوث Research abstract التى تعطيها الدوريات أو المجلات العملية أو خدمات قواعد البيانات Database services التى تمدنا بمعلومات عبر الكمبيوتر من مختلف دول العالم مثل قاعده البيانات ERIC أو Academic elliet research حيث تساعد فى نقل مجموعات البيانات الكترونيًا من موقع إلى آخر عبر الدول.

ثالثًا: فى ضوء بعد الوقت: وهذا مرتبط بفتره زمنية سواء كانت فى فترة واحدة أو عدد من المراحل أو الفترات الزمنية. والبيانات فى هذ الإطار تصنف إلى بيانات عبر مستعرضة Cross-sectional data تطبق فى فترة زمنية واحدة أو بيانات طولية Longitudinal data للمتغيرات موضوع الاهتمام فى فترة من الزمن وهذا مرتبط بتطبيق الادوات فى فترات زمنية متعددة (عبر الزمن). والبيانات المستعرضة تعطى نظرة عامه للمتغيرات موضع اهتمام الدراسة مثل التعرف على تحصيل التلاميذ فى كل المواد الدراسية فى شهر اكتوبر لسنة ما ولكن فى البيانات الطولية نتتبع تحصيل

الثانويــة العامــة منــذ عــام 2000 حتــى الآن، وهــذا يعــرف تصــميم القياســات المتكــررة Repeated measures design ومثل تتبع تحصيل التلاميذ فى منهج العلوم عبر شــهور الســنة وهــى تتبــع نفــس المتغيــر عبــر فتــرات زمنيــة مختلفــة ومما يلــى مثــال توضيحى:

الجدول (2.3): مثال لبيانات مستعرضة ولبيانات طولية.

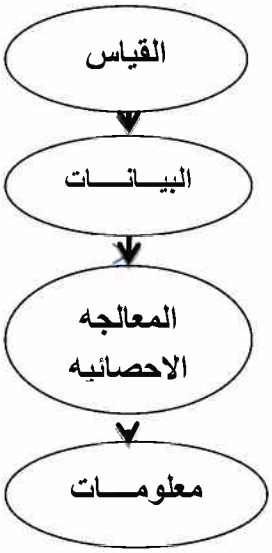
المتغيرات	البيانات		
	A	B	C
تحصيل العلوم فى اكتوبر	10		10
تحصيل العربى اكتوبر	20		11
تحصيل الرياضيات اكتوبر	30		12
تحصيل العلوم مارس		15	15
تحصيل العربى مارس		25	10
تحصيل الرياضيات مارس		20	7

ومجموعة البيانات A , B هى مثال لبيانات مستعرضة وكل قيمة تعطى نظرة على المتغيرات مثل 10 ، 20 فى نفس الفترة الزمنية، فالبيانات A تمثل التحصيل فى اكتوبر فى مواد مختلفة، و B تمثل التحصيل فى نوفمبر بينما مجموعة البيانات C مثال لبيانات طولية وتتضمن القياسات المتكررة عبر الزمن للمتغيرات فهى تعطى لمحة عن التحصيل فى اكتوبر ومارس معًا، وهنا يمكن المقارنة بين تحصيل المواد المختلفة عبر اوقات زمنيه مختلفة وأيضًا تحصيل المادة الواحدة عبر فترات زمنية مختلفة.

البيانات والمعلومات Data and Information

يرى البعض أن البيانات والمعلومات مترادفان، لكن الأمر ليس بهذه البساطة، فالبيانات هى المادة الخام وإذا تمت معالجتها وتهذيبها باستخدام عمليات حسابية أو إحصائية فإنها تمدنا بمعلومات فى صورة قرارات، والمعلومات هى منتج للبيانات وعلى ذلك فإن معالجة مصفوفة البيانات بما تتضمنه من Raw data بيانات خام (بيانات المقاييس والاختبارات) تعطى المعلومات، ويمكن من خلال فحص مصفوفة البيانات ان نستخلص القليل من المعلومات ولكنه غير كافٍ خاصة إذا كنا فى مجال الوصول

إلـى قـرارات حـول ظـاهره معينـة، ومعالجـة البيانـات الخـام تكـون باسـتخدام أسـاليب وتقنيـات الإحصاء الوصفى والاستدلالى. ويمكن عرض العلاقة بين البيانات والمعلومات كالآتى:



الشكل (3.3): العلاقة بين القياس والبيانات والمعلومات.

الفصل الرابع

أساسيات التعامل مع برنامج SPSS

يعتبر برنامج SPSS أحد أهم البرامج الشائعة الاستخدام فى تحليل بيانات العلوم الاجتماعية والنفسية والسلوكية بالإضافة إلى برامج أخرى مثل ,SAS, Minitab, R, S-PLUS, STATA EXCEL وغيرها، إلا أنه أكثرهم يسرًا وسهولة فى استخدامه. ويهتم البرنامج بحساب مؤشرات الإحصاء الوصفى من مقاييس النزعة المركزية ومقاييس التشتت والعلاقة والرسومات البيانية وكذلك كل الأساليب الإحصائية الاستدلالية البسيطة والمتدرجة بالإضافة إلى دور البرنامج فى حساب مصداقية المقاييس وثبات البيانات وحساب مؤشرات المفردات واستخدام مداخل تحليل الفقرات Itemanalysis وذلك من خلال معاملات الصعوبة، معاملات التمييز، تشبعات المفردات، ومعاملات الارتباط الداخلية للمفردات.

ويتميز برنامج SPSS فى سهولة التعامل معه ووجود وظائف وخصائص لإجراء معالجة البيانات خصوصًا فى اختبارات دلالات الفروق، اضافة إلى تفوق البرنامج فى وجود طرق لمعالجة البيانات الغائبة بمجموعة من الطرق سواء بالحذف أو التعويض.

الكمبيوتر والإحصاء

وجود الكمبيوتر جعل الإحصاء سهلة الاستخدام ولكن يبقى دور الباحث فى تفسير المخرجات، ووجود الكمبيوتر لتنفيذ الأساليب الإحصائية جعل إجراءات العمليات الحسابية المعقدة لا ضرورة لها. وقدم Wilcox (2009) تصنيف للبرامج الإحصائية حسب جودتها ودقتها ومدى مسايرتها للتطورات الحديثة فى مجال الاستدلال الرياضى كالتالى:

برامج ممتازة: مثل برنامج R وهو متاح مجانًا ومتوفر على الموقع: http:/cran.R-project.org حى يتضمن كل الطرق الحديثة التى طورت فى السنوات الاخيرة وكذلك الأساليب الكلاسيكية. وبرنامج S-Plus وهذا البرنامج مشابهة فى جودته لبرنامج R.

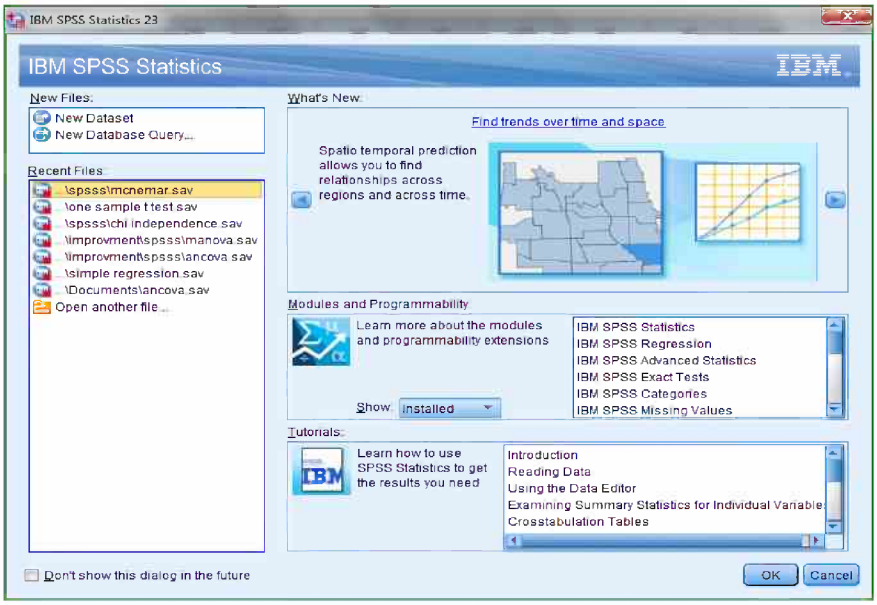
برامج جيدة جدًا: مثل برنامج SAS.

برامج جيدة: مثل برنامج Minitab

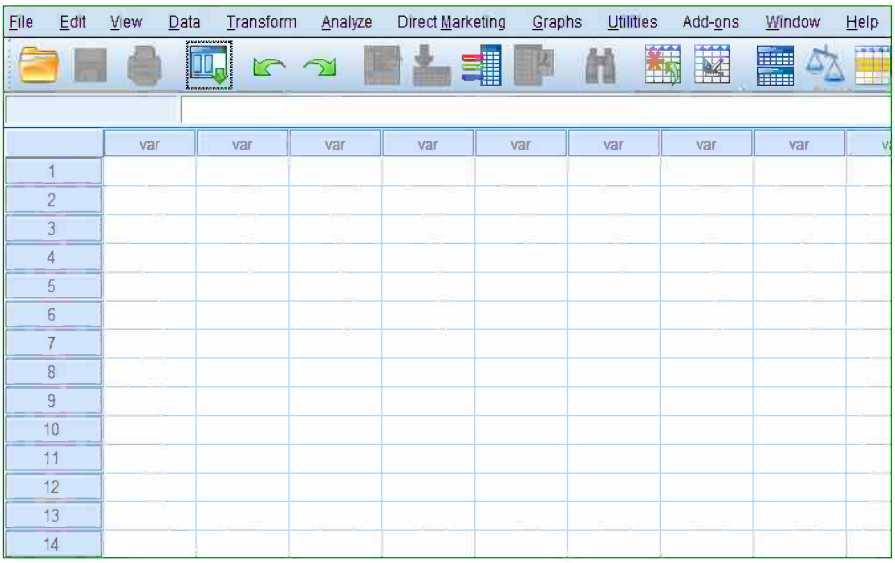
بـرامج غيـر مرضـية: مثـل برنـامج SPSS وهـو أكثـر البـرامج شـهرة واسـتخدامًا وكـذلك برنامج Excel.

فتح البرنامج

1.اضغط Start→ All programs → IBM SPSS23 تظهر الشاشة الآتية:



2.اضغط Cancel او OK تظهر الشاشة الافتتاحية للبرنامج كالآتى:



وهى عبارة عن صفوف وأعمدة حيث تشكل الأعمدة المتغيرات Variables بينما الصفوف تمثل الحالات أو الأفراد Cases.

ويلاحظ فى الشاشة السابقة فى صفحة العمل تسمية الأعمدة باسم Var كاختصار لمصطلح Variable. ويختلف العمل فى برنامج SPSS عن نظيره EXCEL فى وجود تبويبين بأسفل الصفحة أحدهما يعرض صفحة البيانات أو العمل وتسمى Data view والآخر يعرض تسمية وتكويد المتغيرات وتحديد معالمها وتسمى Variable view.

قوائم البرنامج

اولًا: قائمة File: اضغط علىFile تظهر الأوامر الآتية:

New: إنشاء ملف بيانات ويمكن أيضًا إنشاء ملف أوامرSyntax.

Open: فتح ملف سبق حفظه قد يكون بيانات أو مخرجات أو أوامر.

Save: حفظ ملف سواء بيانات أو مخرجات.... إلخ للمرة الأولى.

Save as: حفظ ملف سبق حفظه وتغيير وضع حفظه فى موضع آخر.

Exit: الخروج من البرنامج.

ثانيًا: قائمة Edit: تتضمن أوامر Paste, Copy , Cut.

ثالثًا: قائمة View : تستخدم لإظهار أو إخفاء Status Bar أو شريط القوائم أو إظهار fonts أو Gride lines .

رابعًا: قائمة البيانات Data view: وهى تتعامل مع البيانات مثل ترتيب الحالات Sort cases واختيار حالات Select cases، وجعل المتغيرات حالات والحالات متغيرات Transpose وغيرها وتحديد خصائص المتغيرات ووزن الحالات وغيرها.

خامسًا: قائمة التحويل Transform: تستخدم فى تحوير قيم متغير إلى قيم أخرى وفى إنشاء متغير جديد وحساب متغير جديد من خلال امرCompute وإعادة ترميز قيم متغير والتعامل مع البيانات الغائبة.

سادسًا: قائمة التحليل Analyze: وهى القائمة الرئيسية للبرنامج وتتضمن معظم التحليلات والأساليب الإحصائية البسيطة والمعقدة.

سابعًا: قائمة The Direct Marketing: قائمة متخصصة جدًا وتسمح لمستخدمى البرنامج تصميم قوائم لأهداف خاصة ولن تستخدم فى هذا الكتاب.

سابعًا: قائمة الرسومات Graphs menu: وتستخدم لرسم المدرج والخط والمنحنى **وBar** وغيرها من الرسومات والأشكال البيانية.

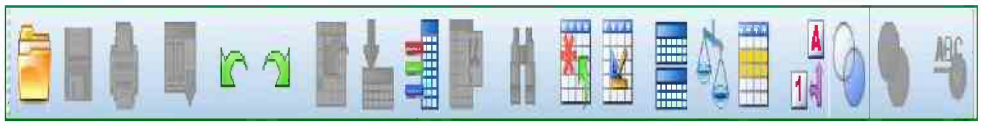
ثامنًا: قائمة Utilities Menu: تستخدم لاكتشاف معلومات عن المتغيرات والملفات ويمكن من خلالها تحديد واستخدام مجموعات من المتغيرات حيث كل متغير يتم تحديد مسماه وتعريفه.

تاسعًا: قائمة المساعدة Windows: وهى مثل أى قوائم فى أى برنامج.

عاشرًا: قائمة Help: تساعد على الحصول على مساعدة عن موضوع معين فى البرنامج.

شريط Tool Bar: يتضمن عدة أيقونات كالآتى:

-



- فتح ملف بيانات سبق حفظه Open data document:

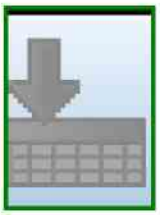
- حفظ بيانات جديدة Save this document:

- طبع ملف Print :

- العودة إلى تغير سابق Undo a user action:

- العودة إلى تغير لاحق Redo a user action:

- الذهاب إلى حالة معينة برقمها Go to case :



- الذهاب إلى متغير ما Go to variable:

- إعطاء معلومات عن متغير Variables:

- إجراء إحصاء وصفى Run descriptive statistics:

- إيجاد قيمة معينة Find:

- حشر حالات فى ملف البيانات Insert cases:

- حشر متغير جديد فى ملف البيانات Insert variable:

- تجزئة ملف البيانات Split file:

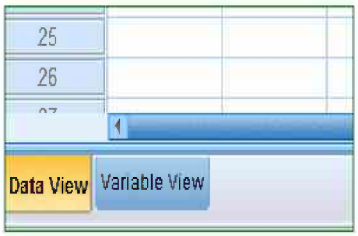
- وزن حالات Weight cases :

- اختيار حالات بمحكات معينة Select cases:

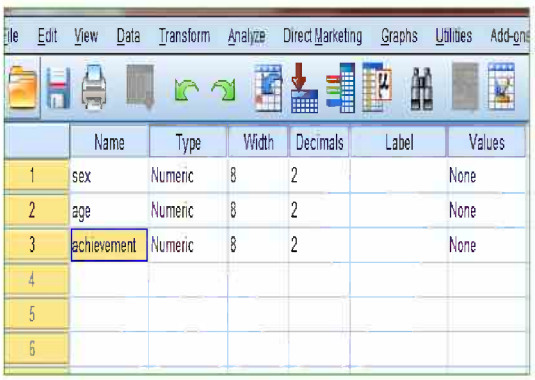
- إظهار كل المتغيرات Show all variables:

تسمية المتغيرات Naming Variable

اتبع الخطوات الآتية:



1. فى الجزء الأسفل على يسار الشاشة الافتتاحية اضغط على **Variable view**تظهر الشاشة الآتية:



حيث توجد مسميات عديدة وهى **Width Type , Name,** .... إلخ، وكل صف يعكس متغير، والأعمدة تمثل خصائص المتغيرات.

2. تحت عمود **Name** فى الخلية الأولى (الصف الأول – العمود الأول) اكتب مسمى المتغير وليكن الجنس واكتبه **Sex** أو أى مسمى آخر، ثم اضغط **Enter** أو السهم، ثم اكتب مسمى المتغير الثانى وليكن العمر **Age** فى الخلية الثانية ثم اضغط **Enter** أو السهم، ثم اكتب مسمى المتغير الثالث وهكذا.

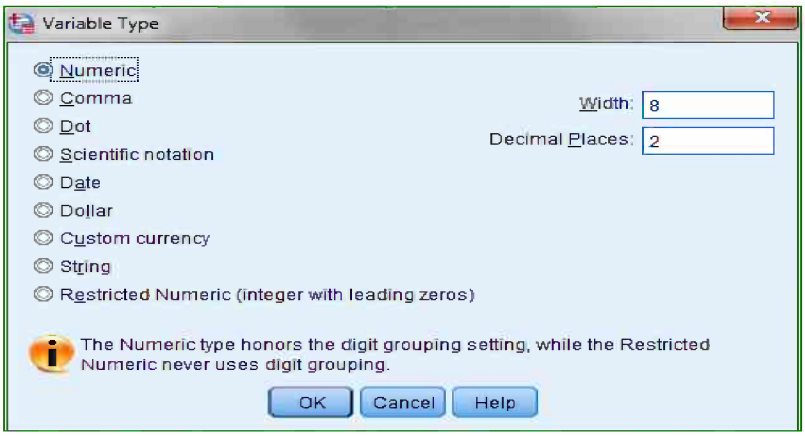
ويجب مراعاة الآتى:

أ – مسمى المتغير لا يزيد عن **32** حرف فى النسخ الحديثة بينما فى نسخة **12** لا يزيد عن **16** حرف، أما فى نسخ **11** ، **10** لا يزيد عن ثمانية أحرف.

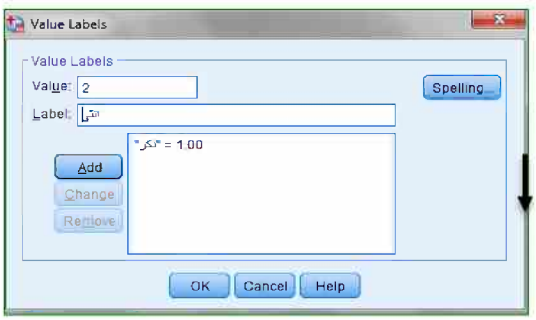
ب – لا يتضمن الاسم مسافة.

خصائص المتغيرات

- **Type** : يحدد نوع البيانات المدخلة للمتغير سواء أرقام أو حروف أو رمز علمى أو عملة مالية أو تاريخ أو مقطع **text** واضغط على العلامة [...] لتحديد نوع المتغير تظهر الشاشة الآتية:



- **Width**: يحدد أقصى عدد الحروف أو المسافة فى كل خلية وتتراوح من **1** حتى **40** حرف أو رمز اعتمادًا على نوع البيانات والموجودة فى البرنامج حيث يمكن زيادتها أو تقليلها عن خلال الاسهم فى الخلية.
- **Decimals**: تحدد عدد الأرقام بعد العلامة العشرية أو النقاط وتتراوح من **10** إلى **16** ويمكن تغيرها من خلال الأسهم.
- **Label**: يسمح بتوصيف كل متغير حيث الجنس ذكر= **1**، أنثى= **2** فيمكن تحديد ذلك فى **Label** من خلال كتابة الجنس.
- **Value**: تسمح بتحديد وتوصيف كود المتغيرات التصنيفية، فمثلًا لتحديد توصيف كود الجنس اضغط على العلامة الزرقاء تظهر الشاشة الآتية:



– اكتب **1** فى المستطيل امام**Value** ثم أمام

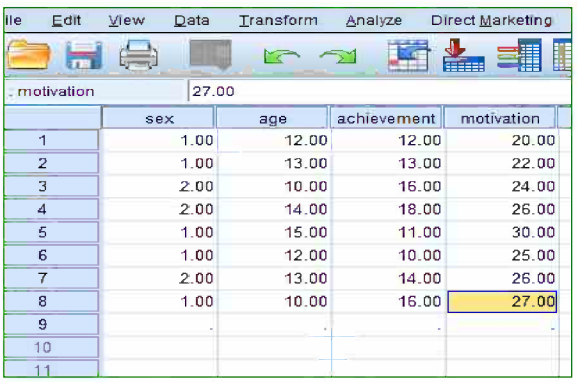
Label اكتبها ذكر أو**Male** ثم اضغط على **Add** وهكذا للقيمة **2** أمام **Value** ثم أمام **Label** اكتبها أنثى ثم اضغط على**Add** ثم اضغط **Ok**.

- **Missing**: هذا العمود يحدد البيانات الغائبة وكيف التعامل معها.
- **Measure**: هذا العمود يحدد مستوى القياس فاضغط على وحدد نوع مستوى القياس للمتغير سواء كانت اسمى أو ترتيبى أو قياسى (فترى – نسبى).
- **Role**: يحدد دور المتغير سواء كان مستقل **IV** أو تابع **DV** أو كلاهما.
- **Align**: يحدد كيف تظهر البيانات فى الخلية فى الوسط أم على الشمال أم على يمين الخلية.
- **Columns**: يحدد عدد المسافات بكل متغير فى شاشة البيانات.

وأهم هذه الأعمدة هى **Type , Value , Measure** حيث تستخدم بصفة دائمة، وعليه تم إدخال أربعة متغيرات هى **Sex** والتحصيل **Achievement** والدافعية **Motivation** كما فى الشاشة الموضحة أسفل.

إدخال البيانات

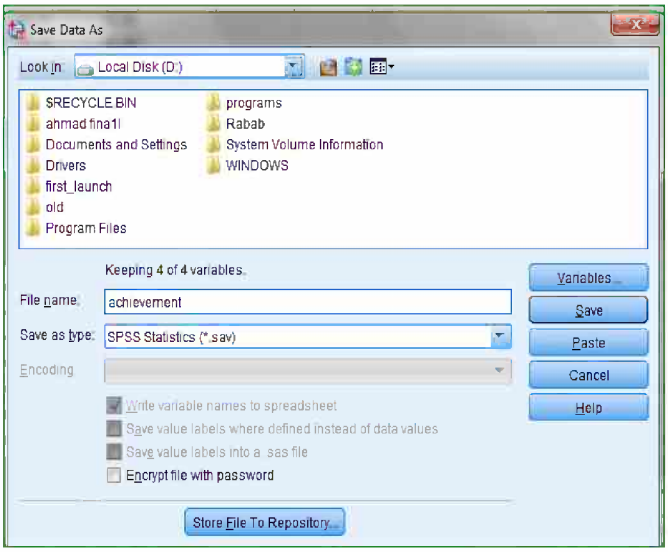
بعد تسمية المتغيرات اضغط على أيقونة **Data view** تظهر شاشة البيانات حيث يتضمن كل عمود متغير.



	sex	age	achievement	motivation
1	1.00	12.00	12.00	20.00
2	1.00	13.00	13.00	22.00
3	2.00	10.00	16.00	24.00
4	2.00	14.00	18.00	26.00
5	1.00	15.00	11.00	30.00
6	1.00	12.00	10.00	25.00
7	2.00	13.00	14.00	26.00
8	1.00	10.00	16.00	27.00
9	.	.	.	.
10				
11				

ويمكن إدخال بيانات المتغير (العمود) فاضغط بالماوس على الخلية الأولى ثم اكتب قيمة المتغير وليكن **1** ثم اضغط **Enter** ينتقل إلى الحالة الثانية (الصف الثانى) اكتب **1** وهكذا، أما إذا تم الإدخال للحالة وهذا هو الأفضل لقواعد البيانات الكبيرة فاستخدم الأسهم الأربعة فى **Key board** للانتقال إلى الخلايا عبر المتغيرات المختلفة للحالة الواحدة، وإذا كتبت قيمة خطأ وتريد تغييرها فاضغط على الخلية بالماوس ثم اكتب القيمة الجديدة مباشرة.

وبعد ذلك أحفظ ملف البيانات من خلال اضغط **File** ثم **Save** يعطى الشاشة:



اكتب مسمى الملف فى مستطيل **File name** وحدد موضع الحفظ ثم اضغط **Save** وإذا أردت تغير موضع حفظ ملف البيانات اضغط على **Save As** ثم اكتب مسمى الملف ومسار حفظه (**D** أو **C** أو **Desktop** أو..) ثم اضغط **Ok**.

فتح ملف سبق حفظه

1. اضغط **Data** ← **File** ← **Open**
2. إذا ظهر لك مسمى الملف فى البرنامج اضغط عليه مرتين بالماوس أو اضغط عليه بالماوس ثم اضغط **Ok**.

حشر حالة فى ملف البيانات

أثناء إدخال البيانات قد ينسى الباحث حالة معينة ولإدخالها فى ملف البيانات:

1. اضغط **Edit** ثم اضغط **Insert cases**
2. اضغط بالماوس على الحالة **3** تظهر كالآتى:

2	1.00	13.00	13.00	22.00
3	.	.	.	.
4	2.00	10.00	16.00	24.00

حشر متغير فى ملف البيانات

1. اختار موضع المتغير المراد إدخاله أو حشره ودائمًا اختار الخلية على اليمين، فمثلًا يرغب الباحث فى حشر متغير الذكاء **Intelligence** بعد متغير **Age** فأضغط بالماوس على متغير **Achievement**.
2. اضغط قائمة **Edit** ثم اضغط **Insert variables** تظهر الشاشة:

age	VAR00002	achievement
12.00	.	12.00
13.00	.	13.00
.	.	.
10.00	.	16.00
14.00	.	18.00
15.00	.	11.00
12.00	.	10.00
13.00	.	14.00
10.00	.	16.00

حذف حالة

لحذف حالة فى ملف البيانات:

1. اضغط على الصف الخاص بالحالة.
2. اضغط **Edit** ثم اضغط **Clear** أو **Del** فى لوحة المفاتيح أو اضغط على يمين الماوس ثم اضغط **Clear**.

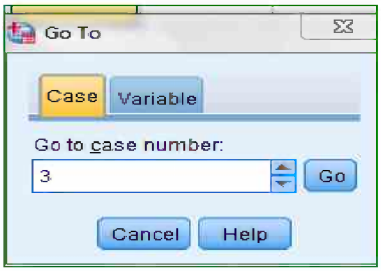
حذف متغير

لحذف متغير من ملف البيانات:

1. اضغط على العمود الخاص بالمتغير.

2.اضغط قائمة **Edit** ثم اضغط **Clear** أو مفتاح **Del** فى لوحة المفاتيح.

للذهاب إلى حالة بعينها



1.اضغط على قائمة **Edit**

2.اضغط **Go to case** تظهر الشاشة:

3. اكتب رقم الحالة فى مربع **Go to case**

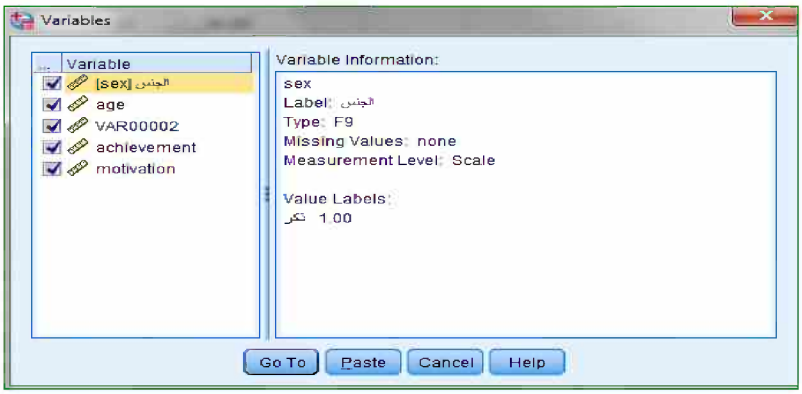
number لتكن **500**

4. اضغط **Go**

إعطاء نظرة عن المتغيرات المتضمنة فى ملف البيانات

1. اضغط **Utilities**

2. اضغط **variables**



البحث عن متغير ما فى ملف البيانات

1. من قائمة **Edit** اضغط **Go tovariable**:

2. اكتب مسمى المتغير فى مربع **Go to variable**

3. اضغط **Go**

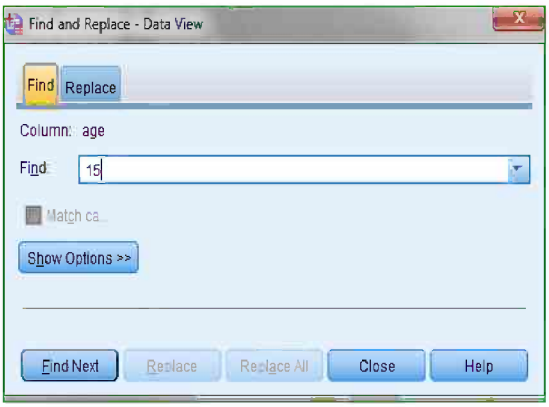
الذهاب إلى المتغير فى قاعدة البيانات وهذا مهم فى قواعد البيانات الكبيرة.

البحث عن قيمة معينة

1. اضغط على العمود الذى يمثل المتغير مثل **Age**.

2. اضغط قائمة **Edit**

3. اضغط **Find** تظهر الشاشة:



4. اكتب القيمة **15** مثلًا التى تمثل العمر

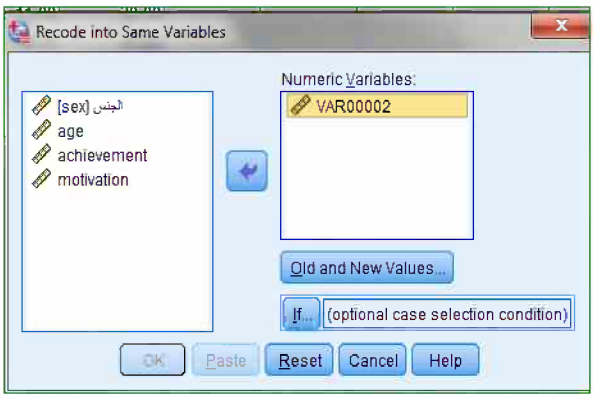
5. اضغط **Find Next** يقوم البرنامج بتحديد القيمة للمتغير.

إعادة ترميز البيانات

أحيانًا عند إدخال بيانات مفردات المقاييس فإن بعض العبارات تكون موجب الصياغة وبعض العبارات سالبه الصياغة ويريد الباحث اعادة ترميز تصحيح العبارات السلبية فمثلًا القيمة **5** تصبح **1** ، القيمة **1** تصبح **5** وهكذا ولتنفيذ ذلك اتبع الآتى:

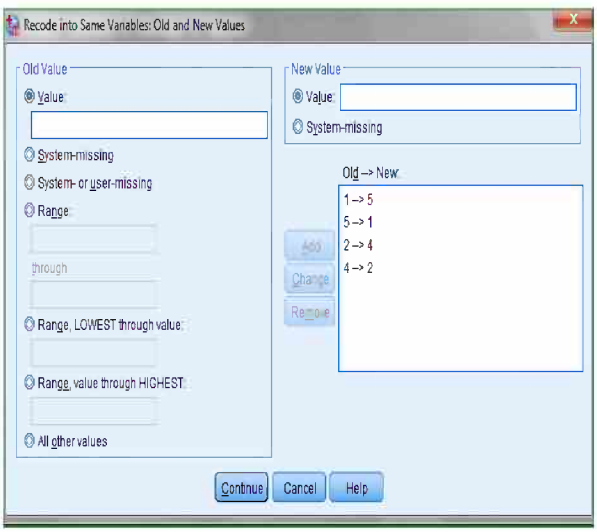
1. اضغط قائمة **Transform**

2. اضغط **Recode into SameVariables** تظهر الشاشة الآتية:



3. انقل المتغير المراد ترميز بعض مفرداته إلى مربع **VariablesNumeric**

4. اضغط **Old and New Variables** تظهر الشاشة الآتية:



5. اكتب تحت مستطيل **Value** القيمة المراد ترميزها **5** مثلًا ثم فى **New Valu** اكتب القيمة الجديدة **1**

6. اضغط **Add** وهكذا لباقى القيم المراد تكويدها

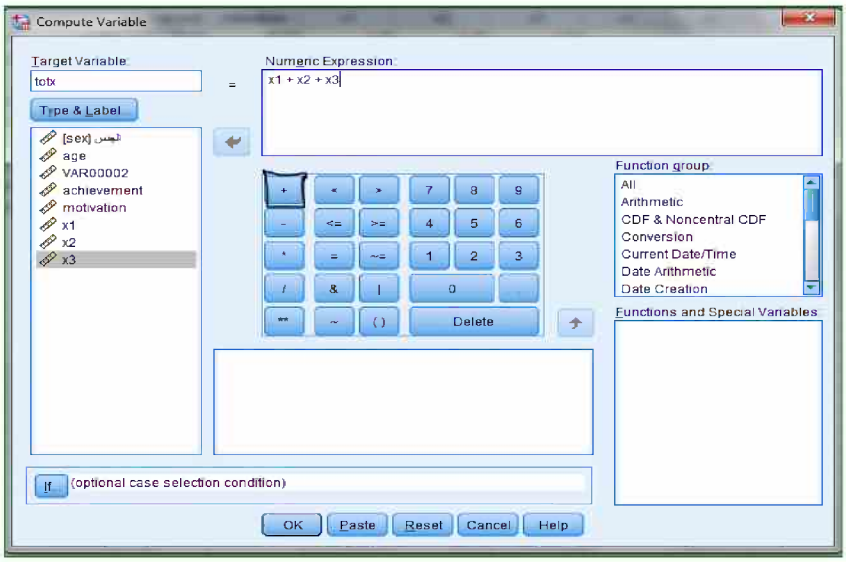
7. اضغط **Continue** ثم اضغط **Ok**

إنشاء متغير جديد

يقوم الباحث بإدخال درجات مفردات

المقاييس أو الاختبارات فى ملف البيانات ولكن المقياس يتكون من إبعاد فرعية ويرغب فى حساب مجموع كل بعد على حده بالإضافة إلى الدرجة الكلية للمفردات الفرعية:

1. اضغط قائمة **Transform**
2. اضغط **Compute Variable** تظهر الشاشة الآتية:



3. تحت مستطيل **Target Variable** اكتب مسمى المتغير الجديد أو مجموع البعد مثلًا **totx**
4. اضغط على المتغير **X1** مثلًا وانقله بالسهم إلى مربع **Numeric Expression** ثم اضغط علامة (+) (بالماوس).
5. اضغط على المتغير **X2** فى مربع **Type & Lable** وانقله إلى مربع **Numeric Expression** ثم اضغط بالماوس على علامة (+) وهكذا.
6. بعد نقل المتغيرات أو المفردات المكونة للبعد اضغط **Ok** تكون شاشة البيانات:

achievement	motivation	x1	x2	x3	totx
12.00	20.00	1.00	1.00	2.00	4.00
13.00	22.00	2.00	1.00	1.00	4.00
.	.	3.00	1.00	3.00	7.00
16.00	24.00	2.00	2.00	1.00	5.00
18.00	26.00	3.00	2.00	2.00	7.00
11.00	30.00	1.00	3.00	3.00	7.00
10.00	25.00	2.00	3.00	2.00	7.00
14.00	26.00	3.00	1.00	1.00	5.00
16.00	27.00	2.00	3.00	2.00	7.00

يظهر عمود جديد فى ملف البيانات تحت مسمى **totx**

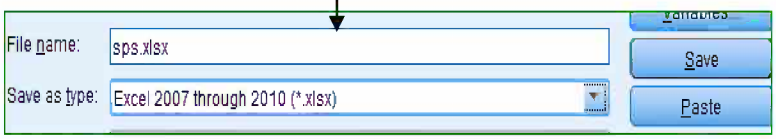
حفظ الرسم البيانى

1. اضغط **File** ثم **save**
2. اكتب مسمى الملف **line** مثلًا فى مربع **File Name** لاحظ أن امتداد الملف **SPV**.
3. اضغط **Ok** أو **Save**

تصدير ملف بيانات SPSS إلى Excel

1. اضغط **File** ثم **Save as**

2. أمام **Save as type** اضغط على السهم واختر امتداد حفظ الملف فى **Excel** وهو **.xlxs**



3. اضغط **Save** على ذلك يتم حفظ الملف البيانات كملف بيانات فى **Excel** فى برنامج **Excel**، ويمكن حفظ ملف البيانات **SPSS** فى أى برنامج آخر مثل:

البرنامج	الامتداد
Excel	**.xls, .xlsx.**
Lotus	**.W**
SAS	**.sd , .ssd**
Stata	**.dta**
Systat	**.syd, .sys.**

استيراد ملف من برنامج آخر فى برنامج SPSS

1. اضغط **File** ثم اختار **Open**

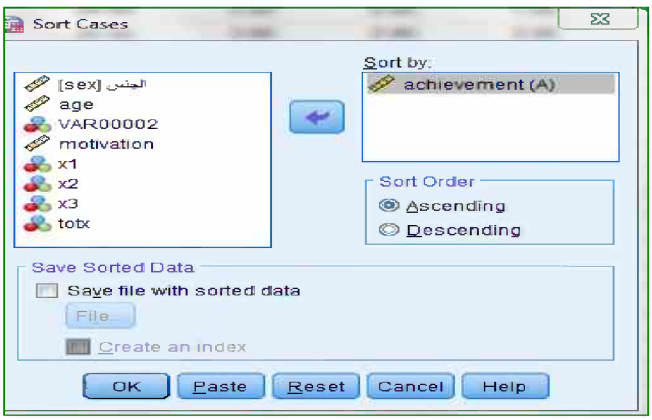
2. حدد ملف البيانات فى **Excel** واختار امتداد **Save** من الملف وهى ملف **SPSS**.

ترتيب البيانات Sorting Data

حيث يرتب قيم المتغيرات تنازلى أو تصاعدى فى ملف البيانات كالآتى:

1. اضغط قائمة **Data**

2. اختار **Sort cases** تظهر الشاشة الآتية:



3. انقل المتغير المراد ترتيب قيمه وليكن achievement.

4. حدد نوع الترتيب تصاعدى Ascending أو تنازلى Descending

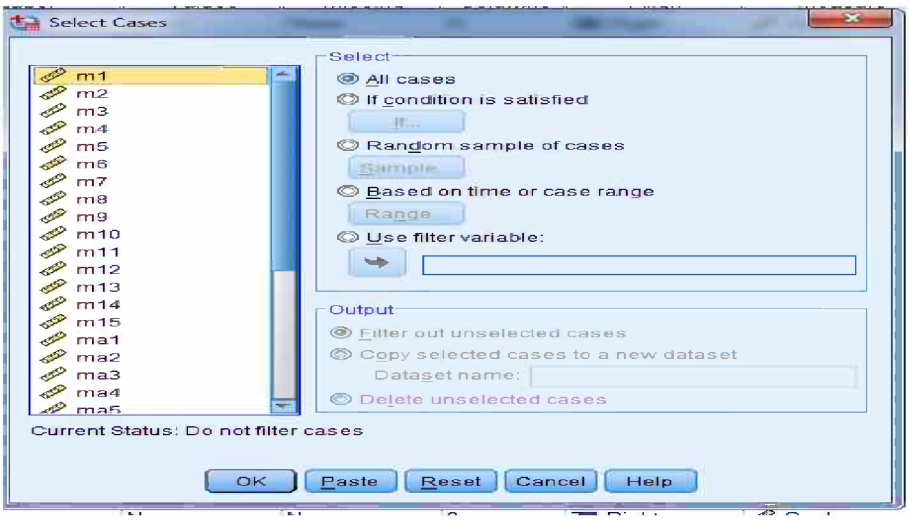
5. اضغط **Ok**.

لاحظ الترتيب يتم لكل المتغيرات فى ملف البيانات.

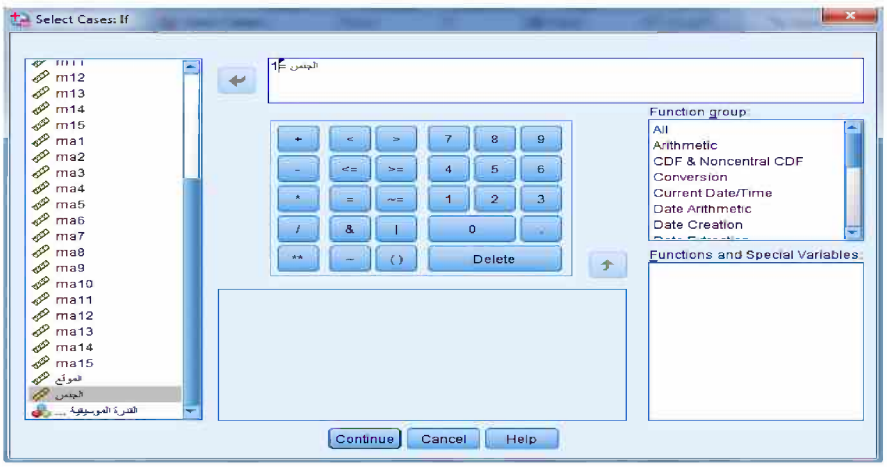
اختيار حالات Select cases

يـتم اختيـار عينـة مـن الحـالات موضـع الدراسـة كاختيـار طـلاب مـن شـعبة اللغـة الفرنسـية مـثلًا لاختبـار فـرض إحصـائى علـيهم، أو اختيـار طـلاب مـن عينـة الـذكور مـن العينة الكلية، ويتم إجراء الأمر كالآتى:

1.افتح قائمة **Data** واختار أمر **Select cases** فتظهر النافذة التالية:



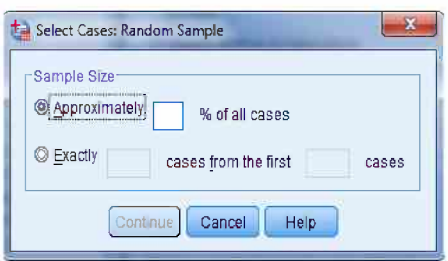
2.اضـغط اختيـار **If condition is satisfied** فيقـوم باختيـار أحـد تصـنيفات أفـراد العينة وليكن الجنس= 1(الذكور).



3.اضغط على **Continue** ثم **OK**.

واختيار **All cases** يتم تنفيذ الأمر أو التحليل لبيانات العينة كلها.

أمـا اختيـار **Random sample of**



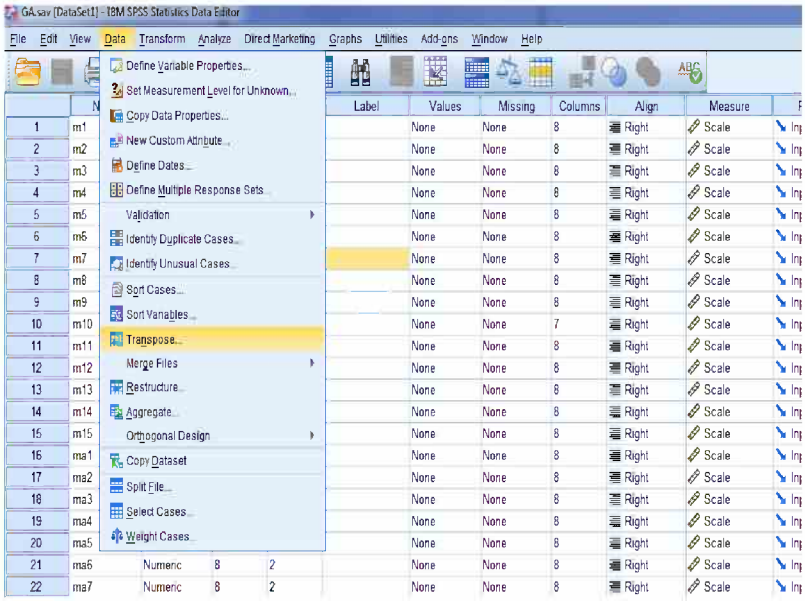
cases فيستخدم بغرض اختيار عينة عشوائية من ملف البيانات بغرض انجاز البحوث والضغط على **Random sample cases** تظهر النافذة التالية:

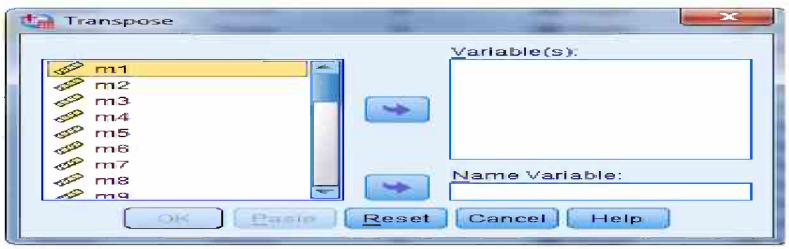
اكتب نسبة العينة المطلوبة. أمام**Approximately** ثم اضغط **Continue** ثم **Ok**.

التحوير Transpose

يستخدم هذا الأمر لتحويل المتغيرات فى ملف البيانات إلى حالات والحالات إلى متغيرات أى تدوير مصفوفة البيانات الخام وذلك كالآتى:

1. اضغط قائمة البيانات **Data** ثم اختار أمر **Transpose**فتظهر النافذة التالية:





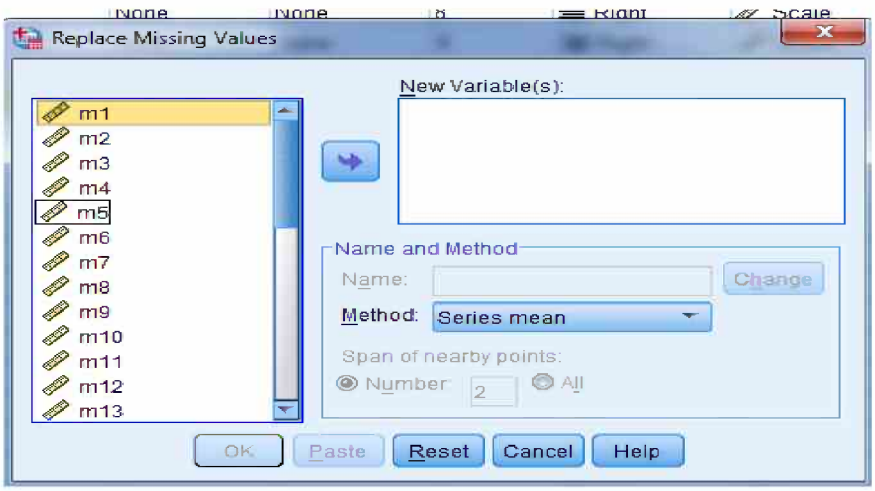
2. انقل المتغيرات المراد تدويرها إلى مربع**Variables**

3. اضغط على **Ok**.يتم تنفيذ الأمر واظهار المصفوفة الجديدة للبيانات الخام فى ملف البيانات **SPSS data**.

استبدال البيانات الغائبة

ويتم إجراء الأمر من خلال:

1. اضـغط قائمـة **Transform** واختـار **data missing replace** تظهـر النافـذة التالية:



2. انقـل المتغيـرات المـراد معالجتهـا مـن أثـر البيانـات الغائبـة إلـى مربـع **new variables**

3. اضـغط علـى وظيفـة Method فتظهـر قائمـة تحتهـا الطـرق التـى تقـوم بمعالجـة البيانـات المفقـودة أو الغائبـة احـدد طريقـة معالجـة البيانـات الغائبـة ولـيكن اسـتبدالها بمتوسط المتغير.

4. اضغط **Ok** وبالرجوع إلى ملف البيانات يلاحظ استبدال البيانات المفقودة بمتوسط درجات المتغير.

الفصل الرابع

الثبات

Reliability

بعد جمع البيانات وادخلها لا بد من تقدير جودة البيانات من خلال تقدير الثبات، ولكى يكون المقياس ثابت فإن الدرجات التى يمدنا بها تمثل أو تعكس الحالة الحقيقية للمتغير المراد تقديره وأى تغير فى الدرجة المقاسة لا بد أن يقابلها تغيير فعلى حقيقى فى البناء التحتى للسمة المقاسة. والمقياس تام الثبات Perfect reliability يجب أن يكون انعكاسًا تمامًا للدرجة الحقيقية وليس شئ آخر وهذا صعب المنال فى القياسات النفسية والاجتماعية لأن الدرجة المقاسة تتأثر ببعض المتغيرات والعوامل الخارجية ولذلك يمكن القول بأن الثبات هو" نسبة التباين التى تعزو إلى الدرجة الحقيقية للمتغير الكامن". وكل طرق تقدير الثبات تتضمن تقدير الدرجة الحقيقية للمتغير وتحديد ما النسبة درجة المقياس المقاسة التى تعكسها الدرجة الحقيقية.

الهدف من القياس هو الحصول على بيانات وقبل النظر إلى كيف نقيس فلننظر أولًا إلى ماذا نقيس، فالقياس دائمًا منصبًا حول مفهوم أو بناء أو ظاهرة، والمفهوم تعبيرات مجردة ونتعرف عليه من خلال ملاحظات وفى البحث العلمى نتكلم أيضًا عن أبنية أو مفاهيم Constructs. وفى تعريف المفهوم لا بد أن نفرق بين التعريف المفاهيمى النظرى A Conceptual definition ويهدف إلى تعريف المفهوم فى ضوء مفاهيم اخرى فمثلًا يعرف البعض الذكاء إنه القدرة على التكيف مع البيئة المحيطة. أما التعريف الإجرائى Operational definition فيعرف المفهوم من خلال إجراءات أو عمليات قياسية، فالتعريف الإجرائى يهدف إلى ترجمة المفهوم إلى احداث ملاحظة أو مقاسة عن طريق تحديد ماذا على الباحث عمله لقياس المفهوم.

والباحث دائما يبحث عن إجابات واضحة عن أسئلته المراد دراستها ويريد أن يتأكد من مظاهر وخطوات بحثه صادقة وثابته وكلًا من الثبات Reliability والصدق Validity لهما تطبيقات متعددة فى البحث.

فى هذا الإطار نتطرق إلى قضية الثبات للقياسات وقبل التطرق إلى مناقشة هذا لا بد من تناول موضوع هام وهو:

أخطاء القياس

فى الواقع المثالى عندما نقيس خاصية أو صفة نفترض أن الدرجة المتحصل عليها تعكس وتناظر الدرجة الحقيقة لهذه الخاصية. بكلمات اخرى نحن نفترض تناظر تام بين الدرجة المقاسة أو الملاحظة **Observed(X)** والدرجة الحقيقية **True Score(T)** وخطأ القياس **Measurement Error(E)** فى هذه الحالة يساوى صفر. ولكن فى الواقع العملى خاصة فى قياسات العلوم النفسية والاجتماعية يبدو هذا بعيد المنال حيث إن خطأ القياس موجود فى قياساتنا.

فإذا كان خطأ القياس كبير نسبيًا فنلاحظ وجود فروق جوهرية من قياس إلى آخر وتصبح القياسات للأداة عبر الزمن غير ثابتة وفى هذه الحالة نحن لا نستطيع أن نثق فى المقياس وحده ليمدنا بمؤشر دقيق للدرجة الحقيقية للفرد ولذلك عدم الاتساق فى القياس يأتى من الأخطاء التى يمكن أن تأتى من مصادر عديدة منها.

وقضية الثبات قائمة على مسلمة وهى أن المتغير الذى نقيسه يتميز بالثبات أو الاستقرار، فعلى سبيل المثال الذكاء لا يتغير بصورة جوهرية من يوم إلى آخر ولكن مستواه يتميز بالثبات. وعندما نقيس متغير مثل الذكاء فإن القياسات تتضمن عنصر من الخطأ، وعلى ذلك فإن الدرجة المقاسة أو المشاهدة للذكاء كما نحصل عليها من أداة القياس تتجزأ إلى الدرجة الحقيقية الفعلية الكامنة للذكاء ولكنها تتأثر بمجموعة متنوعة من العوامل الأخرى مثل الحالة المزاجية والتعب والصحة العامة ودرجة التخمين على الأسئلة التى لا يعرفها الفرد وعوامل موقفية أخرى خاصة بظروف ادارة الاختبار .....إلخ، وهذه العوامل تشكل الخطأ المرتبط بالدرجة المقاسة وتسمى خطأ القياس Measurement error ، وتسمى هذه الأخطاء المرتبطة بعدم فهم الأسئلة أو التخمين وعوامل موقفيه أخرى بالخطأ العشوائى ويمكن أن يحدث لو أن الباحث طبع الأسئلة بصورة سيئة أو أساء تسجيل اجابات الأفراد، وهذا الخطأ العشوائى يؤثر على درجات المتغير المقاس، وهذه النوعية من الأخطاء تزيد من درجة بعض الأفراد أو

تـنقص درجـات الآخـرين، وهـذه الزيـادة والنقصـان يحـدث بينهمـا تـوازن ويمكـن أن يمحـو أحدهما الآخر.

مفهوم الثبات

يعـرف بالثبـات حيـث إن الدرجـة القياسـية تعكـس تمامًـا الدرجـة الحقيقيـة (**X = T**) أى أن ثبـات القياسـات يعنـى أن الدرجـة الحقيقيـة تسـاوى أو تنـاظر تمامًـا الدرجـة المقاسـة وفـى هـذه الحالـة يقـال ان الاداة تامـة الثبـات. أو إلـى أى مـدى أن القياسـات حـرة خاليـة مـن كـلًا من اخطاء القياس.

ونموذج القياس الأساسى يتضمن الآتى:

الدرجة الحقيقية **(T)** = الدرجة الملاحظة **(X)** + خطأ القياس **(E)** .

$$T = X + E\ (R + S)$$

وحيث إن $R = 0$ الخطأ العشوائى، إذًا:

$$X = T + S$$

إلى أى مدى تتطابق الدرجة المقاسة مع الدرجة الحقيقية يعرف بالثبات.

المحـك الأساسـى لتقـويم جـودة أدوات القيـاس هـو الثبـات ويقـال أن أداة القيـاس تتميـز بالثبـات لـو أعطـت نتـائج متماثلـة (أو قريبـة للتماثـل) عنـدما تطبـق عـدة مـرات علـى نفـس الأفراد تحـت نفـس الظـروف، علـى سـبيل المثـال لـو طبقنـا اختبـار ذكـاء علـى مجموعـة مـن الأفـراد اليـوم وطبقنـا نفـس الاختبـار علـى نفـس الأفـراد تحـت نفـس الظـروف الأسـبوع القـادم فـنحن سـوف نحصـل علـى درجـات متقاربـة أو هـى نفسـها فـى التطبيقـين، بالتـالى فـإن الثبـات هـو اسـتقرار أو اتسـاق القياسـات المتولـدة مـن أداة قيـاس محـددة عبـر الـزمن علـى نفـس الأفـراد تحـت نفـس الظـروف، وعلـى ذلـك فالثبـات هـو اتسـاق أو اسـتقرار القيـاس ويعنـى الاتسـاق للـدرجات المتحصـل عليهـا مـن تطبيـق إلـى اخـر. فـإذا كـان الاختبـار ثابـت فـنحن نتوقـع أن التلاميـذ الـذين يحصـلون علـى درجـات عاليـة فـى التطبيـق الأول سـوف يحصـلون علـى درجـات عاليـة فـى التطبيـق القـادم، ويمكـن للـدرجات المتحصـل عليهـا مـن الأداة أن تتميز بالثبات.

والثبات للمقياس يتراوح فى المدى من الصفر (لا ثبات) إلى 1.00(تام الثبات)، فالثبات يكون صفر عندها يكون التباين الكلى المقاس لا يناظر أى من تباين الدرجة الحقيقية بمعنى أن القياسات لا تعكس السمة المراد قياسها إنما يعكس أخطاء قياس ويكون القياسات عديمة القيمة. وعندما يساوى الثبات الواحد الصحيح فإن التباين الكلى المقاس يناظر تمامًا تباين الدرجة الحقيقية ويقال إن المقياس تام الثبات ولا يوجد أخطاء قياس على الإطلاق وهذا صعب المنال فى قياسات العلوم الانسانية والسلوكية. ويقال ان الاداة تتمتع بالثبات وتصلح للاستخدامات البحثية فى صناعة قرار والحصول على استدلالات منها إذا كان على الأقل 50% من التباين الكلى فى الدرجات يعكس تباين الدرجة الحقيقية وقيمة ثبات 0.90 تشير إلى ان 90% من تباين فى درجات الاختبار ترجع إلى التباين الحقيقى أو السمة الكامنة المراد قياسها والباقى 10% يعكس تباين الخطأ.

وعمومًا فى العلوم النفسية والسلوكية فإن أى إجراء قياس يتضمن قدرًا من الخطأ وكمية ومقدار لخطأ تحدد ثبات القياس، فعندما يكون الخطأ كبير فإن الثبات يكون منخفضًا، وعندما يكون صغيرًا يكون الثبات عاليًا.

طرق تقدير الثبات:

معاملات الثبات يمكن تصنيفها فى ثلاثة تصنيفات عريضة كالتالى:

- المعاملات المشتقة من ادارة نفس الاختبار فى مناسبات مختلفة (ثبات الاختبار واعادته).
- المعاملات المشتقة من الصيغ المتوازية أو المتكافئ (ثبات الصيغ المتكافئة).
- المعاملات المشتقة من ادارة الاختبار من تطبيق واحد (معامل الاتساق الداخلى).

بالإضافة إلى ذلك يوجد معامل رابع للثبات مشتق من ثبات المقدرين عندما يتم حساب ثبات بين مقدرين مختلفين لنفس مفردات الاداة

وفيما يلى طرق تقدير الثبات:

الاختبار وإعادته Test – Retest Reliability

تشير إلى اتساق استجابات الأفراد على المقياس عبر الزمن، بمعنى آخر إلى أى درجة الدرجات على نفس المتغير ترتبط ببعضها فى حالة تطبيقها أكثر من مرة عبر الزمن تحت نفس الظروف. وهى بذلك تتضمن تطبيق نفس الاختبار مرتين فاكثر على نفس الأفراد بحيث يوجد فترة زمنية بين مرتى التطبيق ولتكن أسبوع فأكثر، ومعامل الارتباط بين درجات مرتى التطبيق يسمى معامل الثبات وهو حساس لخطأ القياس الناتج عن معاينة الزمن بمعنى الفترة الزمنية بين مرات التطبيق. ويجب التأكيد أن الارتباط المرتفع يشير إلى أن القياسات أو الدرجات عبر مرتى التطبيق متماثلة أو متقاربة وهذا يعنى أن الأداة على درجة عالية من الثبات وهذا يدل على أن الدرجة المقاسة تعكس الدرجة الحقيقية الفعلية للأفراد وأن أخطاء القياس تكون محدودة. ولمعظم المقاييس يفضل ان يكون الثبات على الأقل 0.80. ومعامل الثبات (معامل الارتباط) يتأثر بطول الفترة الزمنية بين مرتى التطبيق فكلما زادت الفترة الزمنية فمن المحتمل أن تكون قيمة الثبات منخفضة لاحتمال حدوث تغيرات فى استجابات الأفراد وهذا يكون واضح بالنسبة للأداء على بعض المهارات المعرفية مثل القدرة الحسابية أو اللغوية أو غيرها حيث يمكن أن يتذكر الأفراد كيف كانت استجاباتهم فى التطبيق الأول ويحاولون تعديل أو تصحيح بعض الأخطاء التى ارتكبوها ويتجنبوها وتعرف هذه المشكلة بتأثيرات إعادة الاختبار Retesting effects ولكن بعض السمات أو الخصائص الشخصية مثل تقدير الذات والذكاء أكثر اتساقًا وثباتًا ولكن متغير الحالة المزاجية لا يتميز بالاستقرار والثبات لفترة طويلة. وثبات الاختبار وإعادته غير مناسب وغير واقعى فى الأغراض التربوية.

واستقرار الدرجات عبر فترة زمنية شهر أو شهرين تعتبر دليلًا كافيًا لثبات الاختبار وإعادته ولو كانت الفترة الزمنية بين مرتى ادارة الاختبار قصيرة وليكن ساعات أو ايام فإن معامل الثبات يحدث له تضخم نتيجة لتأثيرات الذاكرة والممارسة من الإدارة الأولى للاختبار ولو كانت الفترة الزمنية طويلة فإن معامل الثبات يحدث له انكماش. ومعامل

الثبات المتحصل عليه من هذه الطريقة من خلال معامل الارتباط بين درجات القياسات يطلق عليه معامل الاستقرار Stability coefficient.

حقيقة أن مشكلة تأثيرات إعادة الاختبار تحدث لأن بعض الأفراد يعتقدوا أن الباحث يريد منهم التعبير أو الاستجابة بآراء مختلفة فى التطبيق الثانى أو بعض الأفراد يحاولون إعطاء إجاباتهم السابقة تمامًا بغض النظر عن مدى تغيرها عبر الزمن حتى يبدو أنه صادق ويتجنب الظهور بعدم الاتساق وهذا يؤدى إلى زيادة الثبات بصورة اصطناعية. ويمكن تجنب ذلك من خلال جعل الفترة الزمنية بين مرتى التطبيق طويلة مثلًا شهر فأكثر. وحقيقة استخدام طريقة الاختبار وإعادته تبدو غير مناسبة فى الحقل التعليمى التربوى المدرسى حيث تتطلب تطبيق الاختبار مرتين وهذا يمكن أن يكون مدعاة للسخرية من قبل الطلاب، كما ان تطبيق الاختبار مرتين يستلزم مدة زمنية طويلة بمعنى حصتين من المدرس وهذا غير واقعى لطبيعة عمل المدرس حيث يكون ملتزم بجدول وموضوعات محددة زمنيًا. وأن تقدير الثبات باستخدام هذه الطريقة يبدو ذو فعالية فقط إذا كانت السمات المقاسة ليس من المتوقع أن تتغير من مرتى القياس مثل الذكاء ولكنه تكون أقل فعالية للخصائص الاقل استقرارًا مثل سمات الشخصية والحالة المزاجية.

الصيغ المتكافئة أو المتوازية Equivalent forms method or parallel

تستخدم عند تطبيق صورتين متكافئتين تمامًا من الاختبار أو الأداة حيث يلجأ الباحث إلى بناء أداتين متكافئتين فى الخصائص الكيفية من حيث إن الأهداف التى تقيسها مفردات الصيغة الأولى هى نفسها الأهداف التى تقيسها الصيغة الثانية وعدد الأسئلة متساوٍ، وكذلك توفر للصيغتين نفس الخصائص الكمية (بعد التطبيق) حيث متوسط الدرجات والانحراف المعيارى والتباين ومستوى الصعوبة والتميز متساوٍ فى الصيغتين، وعلى ذلك كان الاختبار ومكافئه يقيس نفس المحتوى يطلق عليها بطريقة الصيغ البديلة أو المتوازية.

وعندما يتم تطبيق الصيغة الأولى من الاختبار على الأفراد ثم بعدها بأسبوع على الأقل تطبق الصيغة المكافئة ويقدر معامل الارتباط بيرسون بين الدرجات في الصيغتين فإن معامل الارتباط هو معامل الثبات ويطلق عليه معامل الاستقرار (فترة زمنية) والتكافؤ (صيغ متكافئة) Stability and equivalence coefficient . وإذا تم تطبيق الصيغة الأولى على مجموعة من الأفراد وكذلك تطبيق الصيغة المكافئة بحيث لا توجد فترة زمنية بين تطبيق الصيغتين فإن معامل الثبات يطلق عليه معامل التكافؤ Equivalence coefficient . ومعامل الارتباط العالى أو المرتفع يشير أو يقدم دليلًا قويًا إلى أن الصيغتين يقيسا نفس الشيء ومن محددات هذه الطريقة هى صعوبة بناء اختبار بديل أو مكافئ حيث أنها عملية فى غاية الصعوبة إذا أخذنا فى الاعتبار الشروط الكيفية والكمية لعملية التكافؤ. ولكن يمكن اعتبار هذه الطريقة للتغلب على محددات طريقة الاختبار واعادته من حيث وجود فترة زمنية يمكن خلالها أن يحدث عوامل نضج ونمو بالتالى تؤثر على استجابات الأفراد.

الاتساق الداخلى Internal consistency

يهتم الاتساق الداخلى بالتجانس بين مفردات داخل المقياس، فالمقاييس فى ضوء النظرية الكلاسيكية تهدف إلى قياس ظاهرة وحيدة كما ذكرنا أثناء الحديث عن مسلمات النظرية الكلاسيكية بوجود ارتباطات منطقية بين المفردات والمتغير الكامن وإذا توفرت هذه الارتباطات القوية نتوقع أن الارتباطات بين المفردات قوية على الرغم أننا لا نستطيع ملاحظة العلاقات بين المتغير الكامن والمفردات ولكن يمكن أن نستعيض عنها بالعلاقة القوية بين المفردات وعلى ذلك فالمقياس المتسق داخليًا لا بد أن تكون العلاقات بين مفرداته قوية، العلاقات القوية بين المفردات Inter-item correlations تفترض أن كل المفردات تقيس نفس الشيء بالتالى علاقات قوية بين المفردات والمتغير الكامن، بالتالى المقياس أحادى البعد Unidimenstional للمقياس متعدد الأبعاد(كل بعد كأنه اختبار) يتكون من مفردات ترتبط ارتباطًا عاليًا مع بعضها البعض(DeVellis, 2017).

للمفاهيم المعقدة مثل الذكاء أو الشخصية تقاس باستخدام اختبار أو استبيان حيث يتكون من مفردات عديدة والفكرة هى أن المفردة بحد ذاتها أو السؤال غير كافٍ لإعطاء صورة كاملة أو قياس كامل للبناء والمثال الشائع هو استخدام امتحانات عديدة لقياس الأداء فى مقرر ما، والقياس النهائى لكل فرد يتحدد عن طريق اضافة أو متوسط استجابات الأفراد عبر مجموعة كبيرة من المفردات والمسلمة الأساسية فى هذه العملية هو أن أى مفردة (أو مجموعة من المفردات) تقيس جزء من البناء أو المفهوم ككل ولذلك لا بد أن يوجد بعض الاتساق بين الدرجات المتحصل عليها من مفردات مختلفة، وعلى ذلك فإن الفكرة أنه لا بد من وجود بناء تحتى مشترك لكل المفردات التى تقيس المفهوم بمعنى أن كل المفردات لا بد أن تكون أحادية البعد Dimensionality لأن كل المفردات فى المقياس تقيس نفس المتغير لذا فهى تعطى نتائج متسقة أو متشابهة.

ولتقدير ثبات الاتساق الداخلى فإنه يستلزم تطبيق الاختبار على الأفراد مرة واحدة ومعامل الثبات المحسوب يسمى معامل الاتساق الداخلى Internal consistency coefficient وتوجد عدة طرق لتقدير معامل الاتساق الداخلى أهمها:

- **ثبات التجزئة النصفية Split half reliability** : القضية فى الصور المكافئة هى صعوبة الحصول على صيغ مكافئة تقيس نفس الشىء من هنا تأتى طريقة التجزئة النصفية وهى تطبيق المقياس مرة واحدة ثم يُقسم إلى نصفين وتتضمن تقدير الارتباط بين الدرجة الكلية على نصف الاختبار والدرجة الكلية على النصف الآخر، حيث يطبق الاختبار مرة واحدة وبعد تطبيق الاختبار يتم تقسيمه إلى نصفين عشوائيًا والطريقة الأكثر استخدامًا فى الحسابات اليدوية هو النصف الأول يتضمن درجات الأسئلة الفردية والنصف الثانى يتضمن درجات الأسئلة الزوجية حيث لا يتم اخذ الجزء الأول فى نصف والجزء الثانى فى النصف الآخر لأنه قد يحدث تأثير لعامل التعب على أداء الفرد فى الجزء الثانى أو تظهر تأثير الممارسة على الأداء فى النصف الثانى بينما فى البرامج الإحصائية مثل SPSS يتم انتقاء مفردات كل نصف بطريقة عشوائية بحتة ثم حساب معامل الارتباط بين درجات النصفين وهذا

المعامل يطلق عليه معامل الاتساق الداخلى وهذا المعامل يشير إلى أى درجة نصف الاختبار يعطى نفس النتائج للنصف الثانى. ونلاحظ يتم تقسيم الاختبار إلى نصفين ولكن لو اعتبرنا ان المفردات كلها معًا فسوف يكون الاختبار أكثر ثباتًا بمعنى عند التطبيق على الأفراد تم اعتباره اختبار واحد وعند تقدير ثبات التجزئة النصفية تم اعتباره اختبارين وهذا محدد رئيسى فى هذه الطريقة معامل الارتباط بين نصفى الاختبار يعطى تقدير أقل من القيمة الحقيقية لمعامل الثبات ولذلك يتم تصحيح معامل الارتباط بين نصفى الاختبار باستخدام صيغة سبيرمان- براون وهذا التصحيح كالآتى:

$$r = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}}$$

- $r_{\frac{1}{2}\frac{1}{2}}$ معامل الارتباط بين نصفى الاختبار.

فإذا كان $r_{\frac{1}{2}\frac{1}{2}} = 0.56$ فإن ثبات الدرجات على الاختبار ككل:

$$r = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}} = \frac{2 \times 0.56}{1 + 0.56} = \frac{1.12}{1.56} = 0.717$$

وهذا يبين أن معامل الثبات المصحح أكبر من معامل الثبات المقدر من خلال معامل الارتباط بين نصفى الاختبار. وليس شرط أن يكون عدد الأسئلة فى كل نصف متساويًا.

- معادلة كودر- ريتشاردسون KR-21 : من أهم الطرق المستخدمة لتقدير معامل الاتساق الداخلى حيث تتطلب معلومات أساسية عن الاختبار وهى عدد مفردات الاختبار والمتوسط والتباين للدرجة الكلية. وهذه الطريقة تتطلب أن تكون معاملات الصعوبة للمفردات متساوية وهذا شرط يصعب تحقيقه فى معظم مواقف القياس وتتحدد المعادلة كالآتى:

$$KR21 = \frac{K}{\mathrm{K} - 1}\left[1 - \frac{\bar{x}_T(K - \bar{x}_T)}{KS_T^2}\right]$$

- K عدد الأسئلة (المفردات).
- $\bar{x}_T$ متوسط الدرجة الكلية.
- S_T^2 تباين الدرجة الكلية.

فإذا كان $K = 50$ ، $\bar{\bar{X}} = 40$ ، $s^2 = 16$ فإن :

$$Reliability = \frac{50}{49}\left[1 - \frac{40(50-40)}{50 * 16}\right] = 1.02\left[1 - \frac{400}{800}\right]$$

$$=1.02[1-0.5]=0.51$$

بالتالى فإن الثبات = 0.51 ثبات منخفض.

- **معادلة كودر- ريتشاردسون 20** : تتحدد المعادلة فى ضوء الصيغة الآتية:

$$KR20 = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum p_i q_i}{s_T^2}\right]$$

- K عدد المفردات
- p_i معامل الصعوبة (عدد الاستجابات الصحيحة على الفردة مقسومًا على عدد الأفراد الذين حاولوا الاستجابة على المفردة).
- q_i معامل السهولة (١- معامل الصعوبة).
- s_T^2 تباين الدرجة الكلية (مجموع درجات الأفراد على المفردات).

وهذه الصيغة لا تتطلب تساوى معاملات الصعوبة للمفردات وتستخدم معادلات كودر- ريتشاردسون 20 ،21 فى حالة المفردات ثنائية التصحيح (0 ، 1) فقط.

مثال (فى: Reynolds & Livingiston, 2014) فيما يلى بيانات ستة طلاب على اختبار مكون من خمسة مفردات موضوعية تصحيحها (0, 1):

	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
Student 1	1	0	1	1	1	4
Student 2	1	1	1	1	1	5
Student 3	1	0	1	0	0	2
Student 4	0	0	0	1	0	1
Student 5	1	1	1	1	1	5
Student 6	1	1	0	1	1	4
p_i	0.8333	0.5	0.6667	0.8333	0.6667	$SD^2 = 2.25$
q_i	0.1667	0.5	0.3333	0.1667	0.3333	
$p_i \times q_i$	0.1389	0.25	0.2222	0.1389	0.2222	

- Total score الدرجة الكلية، p عامل الصعوبة، q معامل السهولة (1- p).

فمثلًا معامل الصعوبة للمفردة الأولى Item 1 كالآتى:

$$p_1 = \frac{n}{N} \times 100 = \frac{5}{6} = 0.833$$

- n عدد الطلاب الذين اجابوا اجابة صحيحة على المفردة.
- N عدد الطلاب الذين حاولوا الاجابة على المفردة.

ومعامل السهولة:

$$q_1 = 1 - p_1 = 1 - 0.883 = 0.1667$$

اذًا حاصل ضرب معامل الصعوبة فى السهولة:

$$q_1 p_1 = 0.883 \times 0.166 = 0.1389$$

ومعامل الصعوبة للمفردة الثانية:

$$p_2 = \frac{n}{N} \times 100 = \frac{3}{6} \times 100 = 0.5$$

ومعامل السهولة للمفردة الثانية:

$$q_2 = 1 - p_2 = 1 - 0.5 = 0.5$$

حاصل ضرب معامل السهولة فى الصعوبة:

$$q_2 p_2 = 0.5 \times 0.5 = 0.25$$

وهكذا لكل المفردات بالتالى مجموع حاصل ضرب معامل الصعوبة فى السهولة لكل المفردات كالتالى:

$$\sum p_1 q_1 = (0.8333 \times 0.1667) + (0.5 \times 0.50) + (0.6667 \times 0.333) + (0.8333 \times 0.1667) + (0.6667 \times 0.3333)$$
$$= 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222 = 0.972$$

اما تباين الدرجة الكلية يقدر كالتالى:

- حساب المتوسط:

$$\overline{X_T} = \frac{4+5+2+1+5+4}{6} = \frac{21}{6} = 3.5$$

- حساب التباين كالتالى:

$$S_T^2 = \frac{(4-3.5)^2 + (5-3.5)^2 + (2-3.5)^2 + (1-3.5)^2 + (5-3.5)^2 + (4-3.5)^2}{6}$$
$$= 2.25$$

وبالتطبيق فى معادلة كودر–ريتشاردسون 20 :

$$KR20 = \frac{K}{K-1}\left[1 - \frac{\sum p_i q_i}{s_T^2}\right] = \frac{5}{5-1}\left[1 - \frac{0.972}{2.25}\right]$$
$$= 1.25\left[\frac{1.278}{2.25}\right]$$
$$= 1.25(0.568)$$
$$= 0.71$$

ويقدر الثبات باستخدام معادلة كودر–ريتشاردسون 20 كالتالى:

$$KR21 = \frac{K}{K-1}\left[1 - \frac{\bar{x}_T(K - \overline{\bar{x}_T})}{KS_T^2}\right] = \frac{5}{5-1}\left[1 - \frac{3.5(5-3.5)}{5 \times 2.25}\right]$$
$$= 1.25\left(1 - \frac{5.25}{12.5}\right)$$
$$= 1.25 \times 0.58$$
$$= 0.696$$

لاحظ أن الفروق بين الثبات باستخدام الصيغتين طفيف وهذا يرجع إلى ان مسلمة صيغة 21 الخاصة بتساوى معاملات الصعوبة لم تتحقق بصورة تامة. وفى كلا الحالتين الثبات تقريبًا 0.70 وهذا يمثل الحد الأدنى لمعامل الثبات لأداة القياس.

- **المعامـل ألفـا Alpha Coefficient** : ابتدعـه (1951) Cronbach ويرمـز لـه بـالرمز α وهـو أحـد الطـرق الأكثـر انتشـارًا لتقـدير معامـل الاتسـاق الـداخلى، وتعتبـر صـيغ كـودر– ريتشاردسـون 20 و 21 حالـة خاصـة مـن ألفـا للمتغيلاررالابـؤ ~؟٤ فئـة لطريقـة ألفـا عنـدما تكـون الأسـئلة موضـوعية تصـحيحها(0 ،1) وكمـا نعلـم ان التبـاين لمجموعة من درجات المفردات يرجع إلى مصدرين من التباين هما:

– الاختلاف الفعلى عبر الأفراد فى الظاهرة التى يقيسها مفردات المقياس (التباين الحقيقى فى المتغير الكامن).

– الخطأ وهو التباين البواقى غير المفسر فى درجات المقياس (مثل أن تتضمن المفردة معنيين).

ويقـوم معامـل ألفـا بتجزئـة التبـاين الكلـى بـين مجموعـة مـن المفـردات إلـى مكـونين أحـدهما التبـاين الحقيقيـى والآخـر تبـاين الخطـأ وقيمـة ألفـا تعبـر عـن التبـاين الفعلـى الكلـى بـين المفردات التى تعزى إلى المتغير الكامن أو البناء التحتى للسمة وليس إلى شئ اخر.

وتتحدد قيمة المعامل ألفا كرونباخ بالصيغة الآتية:

$$\alpha = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum si^2}{s_T^2}\right]$$

- K عدد المفردات.
- $\sum si^2$ مجموع تباينات درجات الأسئلة أو المفردات.
- s_T^2 تباين مجموع الدرجات للأسئلة.

وهـذا المقيـاس للثبـات يمكـن التعبيـر عنـه فـى ضـوء الارتباطـات حيـث هـو عبـارة عـن متوسـط الارتباطـات الداخليـة بـين كـل المفـردات ببعضـها الـبعض أو بكلمـات أخـرى هـو يمـدنا بمتوسـط كـل معـاملات الثبـات باسـتخدام التجزئـة النصـفية الممكنـة. ولـذلك فـإن قيمـة المعامـل ألفـا قائمـة علـى متوسـط كـل معـاملات الارتبـاط الداخليـة بـين المفـردات وعـدد المفـردات فـى المقيـاس وهـذا يعنـى أنـه كلمـا زادت عـدد المفـردات فـى المقيـاس ارتفعـت قيمة معامل الثبات ويمكن التعبير عنه بالصيغة الآتية (DeVellis, 2017):

$$\alpha = \frac{K\bar{r}}{1 + (K-1)\bar{r}}$$

- $\bar{r}$ متوسط الارتباطات الداخلية بين المفردات المتضمنة فى المقياس.

وتتميز صيغة α كرونباخ بالتعقيد فى حساباتها ولذلك فهى متوفرة فى برنامج SPSS و SAS وغيرها . وإذا كانت معادلات كودر- ريتشاردسون 20 ، 21 تصلح للأسئلة(0,1) فإن المعامل ألفا كرونباخ يصلح للمفردات ثنائية التصحيح (1 ,0) وللمفردات المتدرجة Polytomousley .

مثال: اختبار مكون من خمسة مفردات طبق على ستة طلاب وتراوحت درجات المفردات من 1 إلى 5 والدرجات كالتالي(Reynolds & Livingiston, 2014):

	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
Student 1	4	3	4	5	5	21
Student 2	3	3	2	3	3	14
Student 3	2	3	2	2	1	10
Student 4	4	4	5	3	4	20
Student 5	2	3	4	2	3	14
Student 6	2	2	2	1	3	10
SD_i^2	0.8056	0.3333	1.4722	1.5556	1.4722	SD^2 = 18.81

- حيث SD_i^2 او S_i^2 تباين المفردات، SD^2 تباين الدرجة الكلية.

مجموع تباينات المفردات كالتالى:

$$\sum SD_i^2 = SD_1^2 + SD_2^2 + SD_3^2 + SD_4^2 + SD_5^2$$

$$= 0.8056 + 0.3333 + 1.4722 + 1.5556 + 1.4722$$

$$= 5.63889$$

وتباين الدرجة الكلية:

$$SD_T^2 = 18.81$$

اذًا المعامل ألفا كالتالى:

$$\alpha = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum S_i^2}{s_T^2}\right]$$

$$= \frac{5}{5-1}\left(1 - \frac{5.63889}{18.81}\right)$$
$$= 1.25(1 - 0.29978)$$
$$= 1.25(0.70)$$
$$= 0.875$$

ويرى Sijtsma (2009) أن المعامل ألفا ليس الطريقة الأمثل لتقدير الثبات لأنه يمثل الحد الأدنى للثبات الفعلى لمجموعة من المفردات ويشير إلى ان فكرة الاتساق الداخلى غامضة ومبسطة عن الاتساق الداخلى للمقياس، فالاتساق يدور حول البناء العاملى لمجموعة من المفردات واستخدام ألفا لهذا الغرض تبسيط كبير، كما أن ألفا صمم لبيانات متصلة وحسابه يعتمد على معامل ارتباط بيرسون بالتالى يفترض بيانات فترية ولكن Gadermann, Guhn, & Zumbo (2012) لاحظ أن معظم الاستجابات على مفردات العلوم الاجتماعية والنفسية تعتمد على بيانات رتبية وليست فترية مثل(موافق بشدة، موافق، إلى حدًا ما، غير موافقو غير موافق على الإطلاق) وفى هذه المقاييس لا نستطيع افتراض تساوى الفترات بمعنى أن الفرق بين استجابة موافق بشدة وموافق ليس هو نفس الفرق بين غير موافق بشدة وغير موافق، بالتالى عدم تساوى الفترات يقود إلى تقدير غير دقيقى للمعامل ألفا واقترحوا المعامل ألفا الرتبى Ordinal alpha حيث لا يفترض القياسية الفترية ومدخلهم استبدل معامل ارتباط بيرسون بمعامل الارتباط الرباعى Tetrachoric (بين المفردات تصحيحها 1 ,0) أو معامل ارتباط Polychoric(اختيارات استجابات أو تصنيفات رتبية) ولأنهما يفترضوا ان البيانات ذات بناء تحتى متصل. واستخدام ألفا بصورتها الأصلية لبيانات رتبية يعطى تقلص لقيمة معامل الثبات بالتالى يؤدى إلى حدوث تحيز مقارنة باستخدام ألفا الرتبية (Gadermann et al., 2012).

ولاحظ Dunn, Baguley, & Brunsden (2014) ان الصعوبات الرئيسية مع استخدام المعامل ألفا كمقياس للاتساق الداخلى تتلخص فى الآتى:

- اعتماده على مسلمات من الصعب تحقيقها وهى افتراض أن كل المفردات متساوية من حيث الجودة والدقة كمؤشرات للمتغير الكامن أو البناء الوحيد.

- عدم تحقق هذه المسلمات يسبب تضخم أو تقلص لمعامل ألفا وهذ من الصعب تحقيقه اذ تختلف المؤشرات أو المفردات من حيث جودتها من مفردة إلى اخرى وكذلك يمكن عدم تحقق الاحادية البعد للمقياس خاصة إذا كان المقياس يتضمن مجموعة من المفاهيم المتمايزة مثل الشخصية أو الاتجاهات وغيرها من المفاهيم بشكل تام. فمثلًا مقياس القلق فإنه يتضمن بعض المفردات التى تبدو أنها تتداخل مع الاكتئاب وهذا لا يؤدى إلى أن المفردات تستخلص نفس القدر من تباين المتغير الكامن وهذه أهم مسلمات النظرية الكلاسيكية لأن أحادية البعد للمفردات تفترض ان التباين المشترك بين كل مفردتين هو نفسه لمعامل ارتباط المفردة بالمتغير الكامن وعلى الرغم أن المعامل ألفا هو الحد الأدنى للثبات إلا أنه يحدث له تضخم نتيجة زيادة عدد مفردات المقياس
- ألفا بعد حذف المفردة Alpha if item deleted والمدخل العام المستخدم لتحديد تأثير حذف المفردة من المقياس على الثبات الإجمالى لا يعكس هذا التأثير على ثبات مجتمع للمفردات وذلك للوصول إلى أفضل مجموعة من المفردات الممثلة لبناء التحتى. فيمكن أن تزداد ألفا لعدة أسباب وليس نتيجة زيادة فى الدرجة الحقيقية وهذا الحذف قائم على أساس ان كل المفردات لها نفس القدر المتساوى من تباينات الأخطاء ولكن هذا لا يمكن تحقيقه فى تطبيقات ألفا ولكن أخطاء القياس الواقعة على المفردات مختلفة عن بعضها البعض، بالتالى فإن حذف المفردة ربما لأنها لا تمتلك تباين من الدرجة الحقيقية ولكنها أقل خطأ تباين.
- عرض تقدير المعامل ألفا فى قيمة تقديرية واحدة بدون فترات الثقة حول هذا التقدير فإنه من الصعب الثقة فى هذا التقدير وهذا ما دعا كثير من الباحثين إلى إجراء اختبارات الفروض حول المعامل الفا.

علاج محددات ألفا

أحد الطرق للتغلب على أحد المحددات السابقة هو إجراء اختبارات فروض إحصائية حول تقدير المعامل ألفا مقرونة بفترات الثقة وهذا يسمح بتقدير المعامل فى حدود خطأ محددة مسبقًا وليكن 95% ثقة حيث يقع معامل الثبات الحقيقى بين حدى الثقة الأعلى

والأدنى، فمثلًا لو أن الثبات 0.85 له 95% حدى الثقة 0.80 و 0.90 فإن هذا يشير إلى ان احتمال 95% إلى ان قيمة الثبات الحقيقى تقع داخل فترة الثقة هذه واحتمال ان يكون خارج حدى الثقة 5% وأوصى Dunn et al. (2014) إلى تحديد فترات الثقة لمعامل الثبات باستخدام Bootstrap .

ولتقدير فترات الثقة يستلزم تحديد مجموعات من البيانات مشتقة من عينات من نفس المجتمع للحصول على توزيع من قيم للمعلم المراد تقديره، فمثلًا إذا أردنا تقدير فترات الثقة حول متوسط الطول للأفراد فى مدينة ما فنقوم بسحب عينات كثيرة من الأفراد من المدينة ثم نقدر متوسط طول كل عينة ثم ننشئ المنحنى لتوزيع هذه المتوسطات، والمتوسط العام (متوسط المتوسطات) Grand Mean يكون هو المتوسط الحقيقى للطول فى المدينة ونتوقع ان الاختلافات بين متوسط العينات يحدث بطريقة عشوائية لأن كل العينات اشتقت من نفس المجتمع ويكون أن يكون توزيع المتوسطات توزيعًا اعتداليًا بالتالى نستطيع تحديد 95% من متوسطات الأطول بحد أدنى وحد أعلى. وإذا كان متوسط الطول 160 فيمكن إيجاد الحد الأعلى والأدنى لـ 95% فترة الثقة حول هذه القيمة. بالمثل فإن استراتيجية Bootstrap لها نفس العملية مع اختلاف جوهرى وهو بدلًا اخذ عينات عديدة من نفس المجتمع فإنه يولد مجموعة من العينات الفرعية من بيانات عينة وحيدة التى جمعها الباحث بالتالى تتعامل مع هذه البيانات باعتبارها المجتمع ويمكن حساب المعامل ألفا لكل عينة ثم اعداد توزيع لقيم ألفا لتحديد فترات الثقة ويمكن أن يتم فى برامج إحصائية مثلSPSS , R , وغيرها. بل يمكن اختبارات فروض حول قيمة ألفا يحددها مسبقًا فى ضوء الدراسات السابقة.

المعامل أوميجا (ω) Omega Coefficient

سبق وذكرنا أنه يمكن تنقيح استخدام ألفا باستخدام معاملى الارتباط Tetrachoric/Polychoric أو باستخدام فترات الثقة، ولكن كثيرًا من الباحثين الخبراء يقترحون مدخل بديل أكثر جدوى للتغلب على محددات المعامل ألفا وهو المعامل اوميجا **(ω)** لتقدير معامل الثبات وقائم على فكرة المعامل ألفا وكلاهما يحدد

الثبات بانه نسبة تباين الدرجة الحقيقية إلى التباين الملاحظ الكلى ولكن المعامل اوميجا يستخدم طريقة مختلفة لحساب التباينات المناسبة، فبينما المعامل ألفا يشتق تقديرات التباين من مصفوفة التغاير أو الارتباط لمفردات المقياس، فإن اوميجا يستخدم مصفوفة تشبعات المفردات Factor Loadings على العامل العام Common factor التى يشترك عليها المفردات من خلال التحليل العاملى فأحد استخداماته تقدير العلاقات بين المفردات والعامل الذى يمثل البناء التحتى للسمة المقاسة وتسمى التشبعات، ويقيس التباين الكلى إلى تباين يرجع إلى المتغير الكامن أو السمة الحقيقية والجزء الآخر يرجع إلى الخطأ شرط ان تكون مفردات المقياس أحادية البعد ويعتبر المعامل أوميجا أكثر تفضيلًا من المعامل ألفا ولكنه أقل استخدامًا نتيجة للتعقيدات الحسابية عكس المعامل ألفا الذى يقدر باستخدام الورقة والقلم عن طريق حساب تباين الدرجة الكلية وحساب تباينات درجات المفردات((DeVellis, 2017). ويتحدد بالصيغة الآتية (Finch & French, 2019) :

$$\omega = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\sigma_X^2} = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\left[\left(\left(\sum_{i=1}^{I} a_i\right)\right)^2 \left(\sum_{i=1}^{Ii} \psi_i^2\right)\right]}$$

- a_i تشبع المفردة بالعامل.
- σ_X^2 تباين الدرجة الكلية
- ψ_i^2 تباين الخطأ للمفردة المتحصل عليها من نموذج التحليل العاملى خاصة التوكيدى.

مثال: يمكن حساب المعامل اوميجا لعشرة مفردات وتباين الدرجة الكلية=10، وكانت تشبعات المفردات بالعامل كالآتى:

$$\omega = \frac{(\sum_{i=1}^{I} a_i)^2}{\sigma_X^2}$$

$$= \frac{.80 + 0.91 + 0.85 + 0.94 + 0.89 + .92 + 0.90 + 0.88 + 0.97 + 0.91}{10} = \frac{8.97}{10}$$

$$= 0.897$$

لاحظ تم الاعتماد على قيمة التشبع بدون تربيعها على اعتبار أنها مقدار للتباين المفسر للعامل في إسهامه في المفردة. وكلما زاد مقدار ارتباطات المفردات بالعامل زادت قيمة الثبات والعكس صحيح.

ألفا الطبقية أو لمقياس مكون من أبعاد فرعية Stratified Alpha

وفي حالة وجود مقياس مكون من مقاييس فرعية(ابعاد) مثل مقاييس الذكاء أو الشخصية فيوجد مدخل معيارى لتقدير معامل ألفا للمقياس ككل واقترح Felt & Brennan (1989) الصيغة الآتية لتقدير ألفا لمقياس مكون من ابعاد فرعية لمفردات تقيس نفس البناء التحتى للمفهوم كالتالى:

$$\alpha_{stratified} = 1 - \frac{\sum_{m=1}^{M} \sigma_m^2 (1 - \alpha_m)}{\sigma_X^2}$$

- σ_m^2 تباين المقياس الفرعى m
- α_m المعامل ألفا للمقياس أو البعد الفرعى m
- σ_X^2 تباين الدرجة الكلية

مثال: لدينا مقياس يتكون من اربعة ابعاد. وتباينات الابعاد الفرعية ومعامل ألفا لها كالآتى:

$$\sigma_1^2 = 2.3, \sigma_2^2 = 1.8, \quad \sigma_3^2 = 2.5\,, \sigma_4^2 = 1.9$$
$$\alpha_1 = 0.79, \alpha_2 = 0.82, \alpha_3 = 0.85, \alpha_4 = 0.74$$
$$\sigma_X^2 = 35$$

وعليه فإن معامل ثبات للمقياس ككل كالتالى:

$$\alpha_{stratified} = 1 - \frac{(2.3 \times 0.79) + (1.8 \times 0.82) + (2.5 \times 0.85) + (1.9 \times o.74)}{35} =$$

$$1 - \frac{1.82 + 1.48 + 2.13 + 1.41}{35} = 1 - \frac{6.84}{35} = 0.81$$

ولكن إذا وجدت فروق جوهرية كبيرة بين تشبعات المفردات بالعوامل فإن المعامل ألفا الطبقى يعطى تقديرات منخفضة لمعامل الثبات الكلى للاختبار وتكون هذه الصيغة فعالة عندما تكون تشبعات المفردات قريبة من بعضها البعض.

طرق بناء فترات الثقة لمعامل الثبات الفا

فى الجزء السابقة تناولنا كيفية الحصول على قيمة وحيدة تصف الثبات فى العينة ويراه الباحثون بأنه أفضل تقدير لمعامل الثبات. ولكن ليس هذا المهم انما الاهم هو كيفية الحصول غلى معامل الثبات على مستوى المجتمع أو بكلمات اخرى إجراء استدلال إحصائى لمعامل الثبات على مستوى المجتمع. فمثلًا قدر باحث معامل الثبات ألفا لمقياس الشخصية المكون من 20 مفردة فى ضوء مقياس ليكرت وطبق على 200 من الطلاب وقيمة معامل الثبات ألفا هى 0.75، وطبق الباحث نفس المقياس على عينة عشوائية اخرى من الطلاب وقدر نعامل الثبات ألفا ولنها لم تكن تمامًا القيمة 0.75 بل اختلفت عنها بالزيادة أو النقصان. ولو تم تقدير معامل الثبات من العديد من العينات العشوائية من نفس المجتمع وبنفس العدد بالتالى تعطى مدى محتمل من لقيم لثبات المجتمع (Finch & French, 2019). ومن الطرق الاستدلال الإحصائى هو بناء فترات الثقة لاختبار فروض الدلالة الإحصائية حول تقدير معامل الثبات ألفا المقدر من العينة.

بناء فترات الثقة بناءًا على ألفا المحورة

توجد طرق عديدة لحساب حدى الثقة الأعلى والأدنى للمعامل ألفا كرونباخ، أحد هذه الطرق تحوير ألفا إلى تحوير فيشر Z تمامًا مثل تحوير معامل الارتباط إلى تحوير Fisher Z لبناء فترات الثقة واختبارات فروض حولها، وهذا يعتمد على ان تحوير ألفا ذات توزيع اعتدالى، وتحوير فيشر لمعامل ألفا يأخذ الصيغة الآتية:

$$Z_\alpha = 0.5ln\left[\frac{1+\alpha}{1-\alpha}\right]$$

- ln تحوير لوغاريتمى.

وتباين هذا التقديرلـ Z كالتالى:

$$S^2_{Z_\alpha} = \frac{1}{N-3}$$

- N حجم العينة الكلى.

ويقدر الحد الأدنى والأعلى لفترات الثقة لـ Z باستخدام تحوير فيشر كالتالى (Finch & French, 2019):

$$Z_{Lower} = Z_{\alpha} - Z_{CV}\sqrt{\frac{1}{N-3}}$$

$$Z_{Upper} = Z_{\alpha} + Z_{CV}\sqrt{\frac{1}{N-3}}$$

- Z_{CV} قيمة Z الجدولية أو الحرجة عند 0.05(1.96) أو 0.01(2.58)

مثال: لمقياس مكون من 20 مفردة وطبق على 200 فرد والمعامل ألفا 0.85 فإن تحوير Z للمعامل ألفا كالتالى:

$$Z_{\alpha} = 0.5ln\left[\frac{1+0.85}{1-0.85}\right] = 0.5\ln\left(\frac{1.85}{0.15}\right) = 0.5\ln(12.33) = 0.5(2.51) = 1.25$$

وعليه فإن 95% فترات الثقة، الحد الأدنى كالتالى:

$$Z_L = Z_{\alpha} - Z_{CV}\sqrt{\frac{1}{N-3}} = 1.25 - 1.96\sqrt{\frac{1}{200-3}} = 1.25 - 1.96(0.07) = 1.25 - 0.14 = 1.11$$

والحد الأعلى كالتالى:

$$Z_u = Z_{\alpha} + Z_{CV}\sqrt{\frac{1}{N-3}} = 1.25 + 1.96\sqrt{\frac{1}{200-3}} = 1.25 + 1.96(0.07) = 1.25 + 0.14 = 1.39$$

و95% حدى الثقة يتضمن كل قيم معاملات ألفا المقدرة فى المجتمع. ولتفسير قيمة Z فإنها تحور مرة اخرى إلى المعامل ألفا كالتالى (Finch & French, 2019):

$$\alpha = \frac{e^{2Z_{\alpha}} - 1}{e^{2Z_{\alpha}} + 1}$$

وتقدر معامل ألفا الأدنى كالتالى:

$$\alpha_L = \frac{e^{2Z_L} - 1}{e^{2Z_L} + 1} = \frac{e^{2(1.11)} - 1}{e^{2(1.11)} + 1} = \frac{9.21 - 1}{9.21 + 1} = \frac{8.21}{10.21} = 0.80$$

وتقدر الحد الأعلى للمعال ألفا كالتالى:

$$\alpha_U = \frac{e^{2Z_U} - 1}{e^{2Z_U} + 1} = \frac{e^{2(1.39)} - 1}{e^{(1.39)} + 1} = \frac{16.12 - 1}{16.12 + 1} = \frac{15.12}{17.12} = 0.88$$

وبناءً على هذه النتائج فإن 95% فترات الثقة تراوحت من 0.80 إلى 0.88 متضمنة القيمة 0.85 لتقدير الثبات من العينة.

يوجد مدخل آخر للحصول على فترات الثقة للمعامل ألفا باستخدام منهجية أو استراتيجية Bootstrap لـ Efron (1979) وهو منهجية إحصائية عامة قائمة على اعادة المعاينة بالاستبدال(توليد عينات جديدة من قاعة البيانات الاصلية) ويستخدم فى اغراض عديدة أهمها الحصول على الخطأ المعيارى وفترات الثقة لإحصائيات عديدة خاصة عندما لا تتحقق المسلمات التحتية للطرق الإحصائية واهمها توفر الاعتدالية(Wilcox, 2012) وتوجد ثلاثة مداخل لاستخدام منهجية البوتستراب للحصول على فترات الثقة للمعامل ألفا واحد هذه الطرق تطبق كالتالى (Finch & French, 2019):

- حساب قيمة المعامل ألفا لبيانات العينة.
- اختار عشوائيًا مجموعة من الأفراد من العينة المقدر منها معامل الثبات الفا، مع تكرار هذه العملية للحصول على عينات جديدة ويمكن أن تصل إلى 10000
- حساب المعامل ألفا لكل عينة مولدة باستخدام هذه الاستراتيجية.
- حساب التوزيع العينى الإحصائى للمعامل ألفا المتولدة من العينات.
- حساب فترات الثقة للمعامل الفا.

وتوجد العديد من الطرق لتقدير فترات الثقة للمعامل ألفا منها Bonnet (2002) و Felt (1965) و Maydeu-Oliver, Coffman, & Harmann(2007) و Iacobucci & Duhachek(2003) غيرها(للمزيد انظر: Finch & .French, 2019) والتراث بحاجة إلى دراسات للمفاضلة بين هذه الطرق لتقدير فترات الثقة للمعامل الفا.

فـى كـل الأحـوال فـإن المعامـل ألفـا أكثـر اسـتخدامًا علـى الـرغم مـن محدداتـه ولكنـه الحـد الأدنـى لمعامـل الثبـات فمـع الأخـذ فـى الحسـبان فـى بنـاء واختيـار المفـردات التـى يجـب أن تكـون أحاديـة البعـد يكـون اسـتخدامه مفيـد فـى تقـدير معامـل الثبـات للأبنيـة النفسـية والاجتماعيـة ولكـن يمكـن أن تكـون الفـروق بـين المعامـل ألفـا والمعامـل أوميجـا لا تـؤثر فـى القيمـة القراريـة لاسـتخدام أداة القيـاس فـيمكن أن تكـون ألفـا 0.80 والمعامـل اوميجـا 0.82 فـإن ثبـات الاختبـار جيـد فـى الحـالتين، ولكـن المعامـل أوميجـا أكثـر دقـة مـن المعامل ألفا (DeVellis, 2017).

- **ثبـات المقـدرين** Inter-rater reliability: ويطلـق عليـة ثبـات الملاحظـين inter-observer reliability أو ثبـات المحكمـين inter-judge reliability يتضـمن تقـدير مقـدار الاتسـاق بـين بـاحثين أو ملاحظـين أو مصـححين أو مقـدرين الـذين يلاحظـوا ويسـجلوا سـلوكيات الأفـراد. ففـى بعـض المواقـف البحثيـة فـإن المقـدرين يلاحظـوا السـلوكيات ويضـعوا تقـديرات أو أحكـام. فمـثلًا إذا طلـب مـن بـاحثين أو مقـدرين تقـدير سـلوك العدوانيـة للأطفـال فـى المدرسـة، فكـل محكـم أو مقـدر يضـع تقـديرات أو أحكـام عـن العدوانيـة ولكـن اسـتخدام احكـام ملاحـظ واحـد تعتبـر غيـر ثابتـة وعلـى ذلـك فـإن ثبـات المقـدرين يشـير إلـى أى درجـة يوجـد اتفـاق أو اتسـاق بـين تقـديرات الملاحظـين. فـإذا أعطـى مقـدر للسـلوك العـدوانى 15 وأعطـى الآخـر 18 فـإن الفـرق بينهما يعكس خطأ القياس.

وتوجـد طـريقتين لتقـدير ثبـات المقـدرين فـإذا كـان المقـدرين يسـجلوا مـا إذا كـان سـلوك الأطفـال (متـوفرة – غيـر متـوفرة) مثـل متعـاون، متنـافس، عـدوانى فـإن المؤشـر المعـروف هـو معامـل كابـا (k) Cohen's kappa ويسـتخدم لقيـاس الاتفـاق بـين المحكمـين وتتـراوح قيمتـه مـن 0 إلـى 1.00. وإذا كـان المقـدرون يقـدرون السـلوك بصـورة كميـة علـى مقيـاس تـدرجى أو تصـحيح اختبـارات بحيـث يعطـى درجـة كليـة فيـتم تقـدير معامـل الارتبـاط بـين درجـات المصـححين أو الملاحظـين عبـر كـل الأفـراد أو المسـتجيبين ويجـب أن يزيـد الارتبـاط عـن 0.70 بينمـا يـرى الـبعض أن يكـون الارتبـاط المرغـوب بـين المصـححين على الأقل 0.80.

اختيار معامل الثبات المناسب

لو صمم الاختبار لتطبيقه فى أكثر من فترة زمنية لنفس الأفراد مثل تطبيقه فى القياس القبلى والبعدى فإن تقدير الثبات باستخدام الاختبار واعادته أو الاختبار ومكافئه من الطرق المفضلة لتقدير الثبات لأنهما حساسين لأخطاء القياس الناتجة عن معاينة الوقت ولو استخدم الاختبار للتنبؤ بأداء الأفراد على محك أو متغير متنبأ به يفضل استخدام هاتين الطريقتين.

اذا صمم الاختبار لتطبيقه أو ادارته مرة واحدة فقط فإن تقدير معامل الاتساق الداخلى اختيار مفضل شرط وجود تجانس بين فقرات المقياس، فطريقة التجزئة النصفية تقدر خطأ التباين الناتج عن معاينة المحتوى فى حين ان المعامل ألفا و KR-20 تقدر خطأ التباين الناتج عن معاينة المحتوى وتجانس المحتوى، وطريقة ألفا و KR-20 حساسين لعدم تجانس المحتوى بالتالى يتطلبوا تجانس مفردات الاختبار. فاذا كان لدينا اختبار لقياس مهارة الجمع يتضمن20 مفردة بالتالى فإن المعامل ألفا يعطى تقدير جيد لمعامل الثبات، ولكن لدينا اختبار حساب يتضمن مهارة الجمع والطرح معًا فإن محتوى الاختبار غير متجانس بالتالى فإن ألفا وKR-20 يعطوا تقدير منخفض للثبات ولكن الاختيار المفضل فى هذه الحالة التجزئة النصفية (Reynolds & Livingston, 2014) لأن الهدف من التجزئة النصفية المقارنة بين صيغتين متكافئتين. فمثلًا يقيس الاكتئاب فمن المتوقع ان يكون المحتوى متجانس بالتالى يفضل استخدام ألفا أو KR-20، ولكن لو ان الاختبار يقيس الحالة المزاجية متضمن مفردات عن القلق ومفردات عن الاكتئاب ومفردات عن الغضب فهذه الحالة المقياس ذات محتوى غير متجانس ويفضل استخدام التجزئة النصفية.

تقويم معاملات الثبات

يوجد بعض الإرشادات تمدنا ببعض التوجيهات لتقويم معامل الثبات حددها ((Reynolds & Livingston, 2014:

- لـو ان نتـائج الاختبـار تسـتخدم لصـناعة قـرارات هامـة مثـل الشـهادات أو الانتقـاء أو التسكين فمن المنطقى ان يكون معامل الثبات 0.90 أو حتى 0.95.
- لاختبارات الشخصية والتحصيل يفضل ان يكون الثبات 0.80 فاعلى.
- لاختبـارات التحصـيل التـى يعـدها مـدرس الفصـل والاختبـارات التحصـيلية المسـتخدمة فى المسح والتشخيص يجب أن يكون معامل الثبات على الاقل 0.70.

الـبعض يقتـرح ان معامـل الثبـات 0.60 يكـون مقبـول لأبحـاث المجموعـات وتقويمـات الأداء والمقـاييس الإسـقاطية ولكـن إشـكالية هـذه القيمـة تعنـى ان 40% مـن التبـاين المقاس للسمة أو المفهوم يرجع إلى الخطأ العشوائى.

العوامل المؤثرة فى معامل الثبات

يتأثر ثبات المقاييس بعدة عوامل أهمها:

- زيـادة عـدد المفـردات (اسـتخدام قياسـات متعـددة) فـى المقيـاس، فزيـادة قيمـة الثبـات تـرتبط بزيـادة عـدد مفـردات المقيـاس. ولكـن هـذه المسـلمة ليسـت صـحيحة فـى كـل الاحـوال ويجـب تهـذيب العبـارة الآتيـة" زيـادة عـدد مفـردات المقيـاس يترتـب عليهـا زيـادة ثبـات المقيـاس" إلـى العبـارة الآتيـة " المفـردات عاليـة الجـودة السـيكومترية مـن يـؤدى إلى ثبات عالٍ بغض النظر عن زيادة عدد المفردات.
- معامـل الصـعوبة والتمييـز: فـالمفردات السـهلة جـدًا أو الصـعبة جـدًا تـؤدى إلـى ثبـات مـنخفض وينطبـق هـذا علـى معامـل التمييـز فـالمفردات غيـر المميـزة بـين منخفضـى ومرتفعى القدرة أو السمة تؤدى إلى ثبات منخفض.
- تباينـات العينـة Samplining variation: يكـون الثبـات مـنخفض كلمـا كـان أفـراد العينـة أكثـر تجانسًـا بغـض النظـر عـن السـمة المقاسـة، لاحـظ أن مقـام طـرق الثبـات سـواء ألفـا أو كـودر– ريتشاردسـون 20 وغيرهـا يعتمـد علـى تبـان الـدرجات وكلمـا زاد التباين زاد معامل الثبات المقدر.

مثـال: مقيـاس مكـون مـن 5 مفـردات وكانـت تباينـات المفـردات كالتـالى (Finch & French, 2019):

$$S_1^2 = 0.4,\ S_2^2 = 0.6,\quad S_3^2 = 0.5,\quad S_4^2 = 0.7, S_5^2 = 0.3$$

وتباين الدرجة الكلية=7.2 بالتالى المعامل ألفا كالتالى:

$$\alpha = \frac{5}{5-1}\left[1 - \frac{0.4 + 0.6 + 0.5 + 0.7 + 0.3}{7.2}\right] = 1.25\left[1 - \frac{2.5}{7.2}\right] = 0.83$$

ولكن إذا انخفض تباين الدرجة الكلية واصبح 4.2 وبقى تباينات المفردات كما هو فإن المعامل ألفا كالتالى:

$$\alpha = \frac{5}{5-1}\left[1 - \frac{0.4 + 0.6 + 0.5 + 0.7 + 0.3}{4.2}\right] = 1.25\left[\frac{2.5}{4.2}\right] = 0.51$$

كما هو واضح انخفض الثبات من 0.83 إلى 0.51 نتيجة تجانس الدرجات فى للمجموعة الثانية (التباين = 4.2).

- إجراءات إدارة وتطبيق الأداة الموضوعية والمعيارية، فكلما كانت إجراءات التطبيق من توفر الوقت المناسب والمكان المناسب والمناخ المناسب ووضوح تعليمات الاختبار يؤدى إلى زيادة ثبات الأداة. وحاول بقدر الإمكان تنبه على المستجيبين أن يأخذوا الموضوع بجدية لأن هذا متعلق بمشروع البحث الذى فى ضوء نتائجه يتم صناعة قرارات.
- تصحيح الأداة بعناية، لأن وجود أخطاء فى عملية التصحيح يقلل من قيمة معامل الثبات.
- وضوح أسئلة المقياس أو الاختبار حيث تكون جيدة الصياغة بعيدة بقدر الإمكان عن الصيغة النفى فإن ذلك يزيد من قيمة الثبات.
- إجراء اختبار استطلاعى Pilot test : الاختبار الاستطلاعى يتضمن تطبيق الاستبيان أو المقياس على مجموعة صغيرة من الأفراد للحصول على فكرة عامة عن كيفية تفاعل الأفراد على الاختبار قبل التطبيق النهائى على العينة. وبعد جمع البيانات من الاختبار الاستطلاعى يستطيع الباحث تعديل بعض عبارات المقياس ويمكن أن تسأل الأفراد حتى يعطوا آرائهم وانطباعاتهم عن المقياس وعباراته ويمكن استخدام هذه الأفكار فى تغير صيغة بعض المفردات الغامضة وأيضًا لتغير كيفية الاستجابة عليها.
- تنوع بدائل الاستجابات على المفردات فوضع سبعة بدائل للاستجابة أفضل من أربعة أو خمسة بدائل لأن ذلك يساعد فى الحصول على تباينات أوسع لاستجابات

الأفراد وعلى ذلك يؤدى إلى زيادة قدرة المفردة على إظهار مدى أوسع من الفروق الفردية وهذا بدوره يزيد من قيمة الثبات.

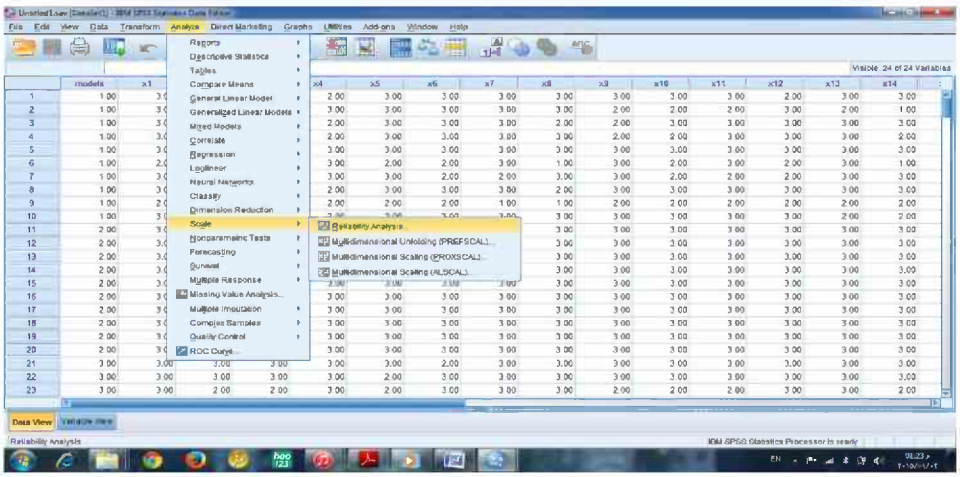
بقدر الإمكان كلما ذلك ممكن استخدم مقاييس موجودة بالفعل وتتميز بالصدق والثبات بدلًا من بناء مقاييس جديدة لأن الصدق والثبات لهذه المقاييس تم تقديره.

إجراء تحليل الثبات فى SPSS:

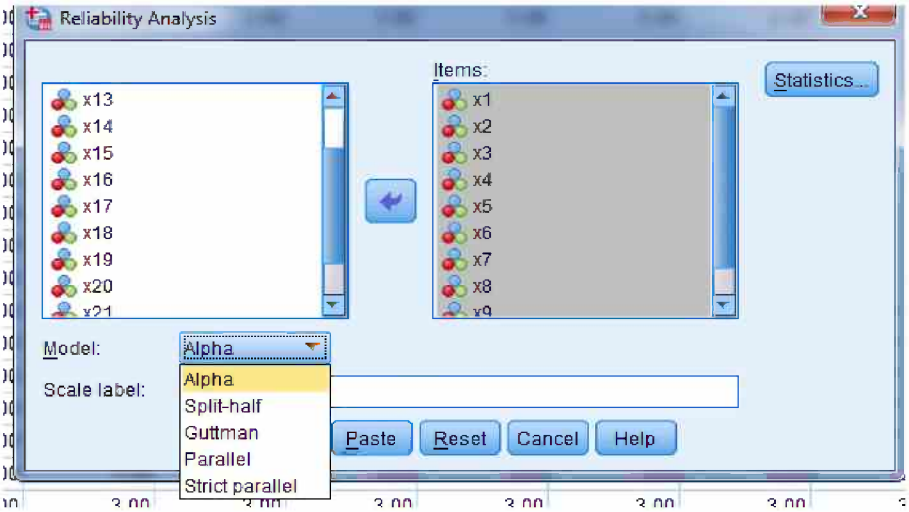
ويستخدم هذا الأمر لتحليل ثبات الاختبارات التحصيلية والنفسية والمعرفية، بمختلف أنواع الاستجابة المستخدمة فيها المتصلة Scale، أو الأسمية Nominal، أو الثنائية Binary أو الرتبية فى ضوء مقياس ليكرت.

ويتم تنفيذ الأمر من خلال:

- أضغط Analyze، ثم اختر Scale، ثم اختر Reliability analysis،كما يلى:

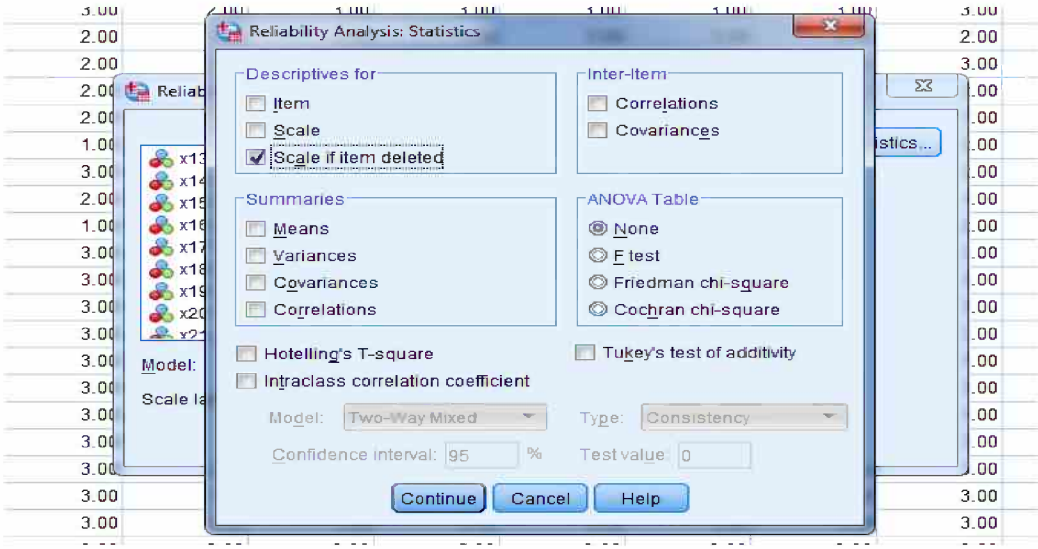


وعندئذ تظهر شاشة الآتية:



والتـى تحتـوى مـربعين أحـدهما بـه المتغيـرات المتضـمنة فـى مصـفوفة البيانـات الخـام، والأخـرى مسـتطيل يـتم نقـل المتغيـرات المـراد تحليلهـا ويسـمى Items، وتحتـوى النافـذة علـى مربـع يسـمى Model يـتم اختيـار طريقـة تحليـل الثبـات اختـر طريقـة ألفـا كرونبـاخ لحساب ثبات المقياس أو الاختبار النفسى المكون من عشرة مفردات.

- اضغط على اختيار Statistics على يمين الشاشة تظهر الشاشة الفرعية:



-واضغط على اختيار Scale if item deleted وذلك لحساب معامل ثبات المقياس عند استبعاد كل مفردة على حدة.

- وللتعـرُّف علـى خصـائص المفـردات مثـل معـاملات الارتبـاط والتغـاير بـين كـل مفـردة مـن مفـردات الاختبـار تـيم اختيـار correlations لتقـدير معـاملات الارتبـاط أو covariances لتقدير التغاير.

-اضغط Continue ثم Ok

-المخرج:

Reliability

[DataSet1] E:\رضوي\Untitled1.sav

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	309	99.7
	Excluded[a]	1	.3
	Total	310	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.818	10

ويوضـح الجـدول الأول عـدد الحـالات الفعليـة التـى اسـتجابت علـى مفـردات المقيـاس ككـل وهـى Valid "309" ونسـبتهم "99.7%" مـن إجمـالى العينـة الكليـة، أمـا الأفـراد الـذين لـم يعطـوا اسـتجابات علـى مفـردات المقيـاس Excluded وعـددهم فـرد واحد بنسبة "٠.3% من إجمالى العينة الكلية.

ويوضـح الجـدول الثـانى قيمـة معامـل ثبـات ألفـا كرونبـاخ للمقيـاس ككـل Cronbach's Alpha وبلغـت قيمتـه ".808"، بينمـا بلـغ عـدد المفـردات الكلـى للمقيـاس N of items العدد "10" مفردات.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
x1	23.4239	10.693	.496	.804
x2	23.6117	10.472	.452	.807
x3	23.6764	10.142	.490	.803
x4	23.6602	10.342	.487	.803
x5	23.5922	10.190	.517	.800
x6	23.5275	10.341	.471	.805
x7	23.6958	10.057	.505	.801
x8	23.7443	9.749	.558	.795
x9	23.6958	9.998	.529	.799
x10	23.7023	9.898	.502	.802

ويتضح من الجدول السابق:

- Scale mean if item deleted متوسط المقياس فى حالة استبعاد كل مفردة على حدة وكلما زاد المتوسط فإن هذا يعنى أن المفردة لا يمكن استبعادها وأنها تمثل البنية الصحيحة فى ضوء إطار نظرى منطقى استندت وصيغت فى ضوئه المفردات.
- Scale variance if item deleted تباين المقياس فى حالة استبعاد كل مفردة على حدة وكلما قل التباين فإن هذا يعنى أن استبعاد المفردة يؤدى إلى انخفاض تباين الأفراد على مفردات الاختبار.
- Corrected item total correlation معامل الارتباط المصحح للمفردة على حدة ويجب استبعاد المفردات التى يقل معامل ارتباطها المصحح عن القيمة "0.3".
- Alpha if item deleted معامل الثبات للمقياس بعد استبعاد كل مفردة على حدة ويجب استبعاد المفردات التى يزيد معامل الثبات المقياس بعد حذفها عن معامل الثبات الكلى "818.".

الفصل السادس

التوزيعات التكرارية وتمثيلها بيانيًا

Frequencies distributions

فى هذا الفصل نعرض نوعين من إجراءات وصف وتلخيص البيانات وهى التوزيعات التكرارية وتمثيلها بيانيًا، فإجراءات القياس تولد مجموعة كبيرة من البيانات أو الدرجات. والقضية التى تواجه الباحث هى كيفية تنظيم هذه الدرجات فى صيغة قابلة للفهم وفى نمط يكون قابل للتواصل عبر الباحثين وهذه هى الوظيفة الأساسية للإحصاء الوصفى وهى تبسيط وتنظيم عرض البيانات. ويوجد نوعين من الإجراءات الوصفية التى تستخدم لهذا الغرض وهى الطرق الرقمية والطرق البيانية، فالطرق الرقمية تعرف بالتوزيعات التكرارية، بينما الطرق البيانية تعرف بالعروض والرسومات البيانية.

التوزيع التكرارى Frequency distribution

هو وسيلة لتنظيم البيانات فى ترتيب منطقى حتى يسهل تفسيرها، وتعتبر الخطوة الأولى لتحليل البيانات، ويعرف Pagano (2013) التوزيع التكرارى بأنه عرض للقيم أو الدرجات(X) وتكرار حدوثها(F) فى جدول وتنظم فى ترتيب منطقى حيث تكون القيمة الدنيا فى أسفل أو نهاية التوزيع والقيم العظمى فى بداية التوزيع.

ويعرفها(2014) Gravetter & Wallnan بأنه جدولة منظمة لمجموعة من الأرقام المفردة. ويعرفها kirk (2008) بانه جدول يعرض فئات متكافئة والتكرار المناظر لقيم الدرجات ويؤكد ان الفئات المتكافئة Equivalence classes يمكن أن تكون:

أ- درجة مفردة مثل القيمة 14.

ب- تجمع لمجموعة من القيم أو الدرجات من الدرجة 5 حتى 9 مثلًا.

جـ - تصنيف كيفى (أفراد ذوو تقدير جيد).

والفئات المتكافئة للتوزيع التكرارى تسمى فترات الفئات Class intervals وإذا كانت فترات الفئة درجات مفردة فيطلق عليها التوزيع التكرارى غير المجمع Ungrouped ويكون للدرجات الخام وإذا كانت الفترات تتضمن قيمتين فأكثر فيطلق عليها التوزيع

التكـــرارى المجمع(الفئـــات) Grouped. وعليـــه فـــإن التوزيـــع التكـــرارى هـــو الجدولـــة أو التبويب لمجموعة من الدرجات بحيث تعرض القيم وعدد مرات تكرارها فى التوزيع.

التوزيع التكرارى غير المجمع للمتغيرات الكمية

يـتم فيـه وضـع قـيم المتغيـر X فـى عمـود وتعـرض فيـه القـيم مـن الاكبـر إلـى الاصـغر وبجانب كل قيمة يوجد تكرارها (مرات حدوثها).

مثال: فيما يلى درجة 20 طالب فى امتحان مقرر الإحصاء وهى:

8, 9, 8, 7, 10, 9, 6, 4, 9, 8, 7, 8, 10, 9, 8, 6, 9, 7, 8, 9

ولبناء التوزيع التكرارى يستلزم:

1. تحديداكبر قيمة فى التوزيع X=10 واصغر قيمة X 4 = وعلى ذلك فالعمود الأول يكون فيه قيم X مرتبة من 10 إلى 4.
2. رصد تكرار كل درجة فى العمود الثانى فمثلًا الدرجة 10 تكررت مرتين:

X	F	X F
10	2	20
9	5	45
8	7	56
7	3	21
6	2	16
5	0	0
4	1	4

والملاحـظ فـى التوزيـع أنـه يوجـد درجتـين لهمـا أكبـر تكـرار فـى التوزيـع وهمـا 8 ، 9 بينمـا الدرجـة 4 حصـل عليهـا طالـب واحـد فقـط والدرجـة 5 لـم يحصـل عليهـا أى طالـب، ومجموع التكرارات هى تعبر عن عدد الأفراد أو عدد الدرجات فى التوزيع بالتالى:

$$\sum F = N$$

ومجموع درجات X يعبر عنها كالآتى:

$$\sum X = \sum XF = 10\times2 + 9\times5 + 8\times7 + 7\times3 + 5\times0 + 4\times1 = 146$$

مثال: فيما يلى درجات لعشرة طلاب فى مقرر الرياضيات:

5 , 4 , 4 , 3 , 2 , 3 , 1 , 3 , 2 , 2

والتوزيع التكرارى كالتالى:

X	F	P (F/ N)	p% (100)
5	1	1/10 = 0.1	10 %
4	2	2/10 = 0.2	20%
3	3	3/10 = 0.3	30%
2	3	3/10 = 0.3	30%
1	1	1/10 = 0.1	10%

وفى هذا التوزيع يمكن إضافة عمود ثالث وهو يمثل التكرار النسبى Relative frequency لكل درجة فى التوزيع بمعنى كم فى المائة من الأفراد حصلوا على الدرجة 4 . ويقدر كالآتى:

$$P = F / N$$

حيث F تكرار الدرجة، N العدد الكلى من الأفراد.

وللدرجة 4 فإن النسبة كالآتى:

$$P = 2/10 = 0.2$$

ويظهر هذا الناتج فى صورة كسر، وبالإضافة إلى التكرار F والتكرار النسبى Pيفضل وصف درجات التوزيع فى ضوء نسب مئوية، فيفضل البعض القول بأن 10% من طلاب الفصل حصلوا على الدرجة 5 او 4 وهكذا ولحساب النسبة المئوية المرتبطة بكل درجة كالآتى:

$$\text{Percentage} = P(100) = \frac{f}{N} \times 100$$

ولذلك فإن النسبة المئوية للدرجة 3 هى:

$$=3/10 \times 100= (0.3 \times 100) = 30\ \%$$

ولذلك يضاف عمود رابع فى جدول التوزيع التكرارى ليعبر عن النسبة المئوية. وعليه فإن التوزيع التكرارى يتكون من عمودين أساسين هما الدرجة وتكرارها ويمكن إضافة عمودين آخرين هما التكرار النسبى والنسبة المئوية.

التوزيعات التكرارية المجمعة Grouped Frequency distribution

عندما يوجد مجموعة كبيرة من البيانات أو الدرجات فمن الضرورى بناء جدول تكرارى مجمع لأن عرضها فى جدول تكرارى غير مجمع عملية صعبة وكذلك يوجد كثير من القيم لا يحصل عليها أى فرد بالتالى يكون تكرارها صفر وهذا يجعل من الصعب عرض شكل التوزيع بيانيًا.

ويقصد بالتوزيع التكرارى المجمع هو عرض قيمتين أو أكثر فى فترة فئوية Class Interval وهى تجمع لعدد من الدرجات ويطلق على التوزيع التكرارى فى هذه الحالة Frequency distribution of grouped scores.

فاعتبر ان مجموعة كبيرة من الدرجات وكانت أكبر قيمة 96 أصغر قيمة 41 فإن مدى هذه الدرجات 55=(41-96) وعند عرض هذه الدرجات فى توزيع تكرارى غير مجمع فإنها تأخذ 55 صف وهذا يعتبر طويل ويسبب الملل والارتباك وهذا يتناقض مع هدف التوزيعات التكرارية تبسيط عرض البيانات فى صورة منظمة وعلى ذلك يلجأ الباحث إلى تجميع الدرجات فى فئات ويتم عرض الفئات فى عمود بدلًا من عرض الدرجات المفردة وتوجد العديد من الإرشادات تساعد الباحثين على بناء التوزيعات التكرارية المجمعة وهذه توجيهات نسبية تختلف من باحث إلى آخر وليست صارمة لأن بناء الفئات فيه شيء من الذاتية.

إرشادات لبناء الفئات

1. التوزيع التكرارى المجمع يجب أن يتضمن من 10 إلى 20 فئة لأن وجود عدد كبير من الفئات يصبح عملية صعبة ويدعو للملل وهذا يتناقض مع الهدف من الفئات فى المقابل وجود عدد محدود من الفئات 5 , 4 مثلا فانك تفقد معلومات كثيرة حول توزيع الدرجات وعلى ذلك فإن عدد 10 فئات هى قاعدة عامة وعلى ذلك فإن عرض جدول فئات فى تقرير علمى يجب أن يكون عشر فئات على الاقل وهذه القاعدة يؤكد عليها(Gravetter & Wallnau,2014; Pagano, 2013).

ويرى(2008) Kirk إلى أنه يجب أن تكون عدد الفئات من 10 إلى 20 فئة ما لم يوجد عدد محدود من القيم فى التوزيع فمثلا توزيع به 25 درجة فقط فلا داعى لتمثيله بعشر فئات ويمكن الاقتصار على 5, 6 , 7 فئات.

2. **طول أو حجم** size or width كل فئة يجب أن يتضمن عدد معقول من القيم على سبيل المثال 3 , 5 ,10 وذلك للمتغيرات الكمية ويمكن تحديد حجم الفئة من الصيغة الآتية:

$$\text{عدد الفئات} = \frac{Range(\text{المدى})}{}$$

3. كل الفئات يكون لها نفس المسافة أو الطول.

4. لفئات البيانات الكمية يجب عدم وجود فجوات بين الفئات أى عدم وجود فئات تكرارها صفر.

5. ابدأ الجدول التكرارى بالفئة التى تتضمن الدرجات العظمى بمعنى اجعلها الفئة الأولى من أعلى وعلى الرغم أن (2008) Kirk يرى وجود خلاف فى هذا الشأن حيث يرى فريق آخر أن نبدأ الفئات من أعلى بأقل الدرجات ولكنه يؤيد أن تتضمن الفئة الأولى أعلى درجة وعموما كثيرًا من برامج الكمبيوتر تضع الدرجة العليا فى أسفل التوزيع ولكن هذه القاعدة غير ذات أهمية للمتغيرات الكيفية.

6. تأكد عدم وجود تداخل بين الفئات واستخدام فئات بفترات متساوية لأن استخدام فئات بفترات غير متساوية يخلق اشكاليات تفسيرية.

مثال: قام معلم بإعداد اختبار تحصيلى فى مقرر العلوم وطبقه على 70 طالب وكانت درجاتهم كالآتى:

95	57	76	93	86	80	89
76	76	63	74	94	96	77
65	79	60	56	72	82	70
67	79	71	77	52	76	68
72	88	84	70	83	93	76
82	96	87	69	89	77	81
87	65	77	72	56	78	78
58	54	82	82	66	73	79
86	81	63	46	62	99	93
82	92	45	76	40	74	67

المطلوب عرض هذه الدرجات فى توزيع تكرارى مجمع؟

الخطوات:

1. حدد مدى التوزيع أو الدرجات: Range =99 - 46 = 53

2. حدد طول الفئة وليكن عرض الدرجات في10 فئات:

$$i = \frac{\text{المدى}}{\text{الفئاتعدد}} = \frac{53}{10} = 5.3$$

وعند وجود كسور اتبع التقريب وعلى ذلك فإن $5.3 \simeq 5$

3. حدد حدود كل فترة فئوية، ضع الفئة التى تتضمن أقل قيمة(الفئة الدنيا) أسفل التوزيع.

لتحديد الحد الأدنى للفئة الدنيا يوجد متطلبين:

ا- يجب أن تتضمن أقل قيمة فى التوزيع.

ب- أدنى قيمة فى التوزيع هى 46 وبالتالى يجب أن يكون الحد الأدنى أصغر من هذه القيمة كان يكون 45 وعليه فإن طول الفئة 5 وبالتالى فإن الفئة الدنيا هى 45-49 فى هذا فإن عدد الفئات يزيد عن 10 ويصبح 11 فئة حتى يتم تضمين كل درجات التوزيع.

4. اعرض الفئات فى عمود ثم وضع علامات بشكل مستطيل Tally يعبر عن مرات تكرارات الدرجات فى التوزيع ثم يعبر عن هذه العلامات فى ضوء تكرار F فى العمود الثالث

وعلى ذلك يكون جدول الفئات كالتالى:

class interval الفئات	Real limits الحقيقة	العلامات	F
95-99	94.5-99.5	////	**4**
90-94	89.5-94.5	////- /	**6**
85-89	84.5-89.5	-//// //	**7**
80-86	79.5-84.5	//// ////	**10**
75-79	74.5-79.5	-//// ////-////-	**16**
70-74	69.5-74.5	////-////	**9**
65-69	64.5-69.5	-//// //	**7**
60-64	59.5-64.5	////	**4**
55-59	54.5-59.5	////	**4**
50-54	49.5-54.5	//	**2**
45-49	44.5-49.5	/	**1**
			∑F = N = 70

ومنتصف الفئة هو متوسط الحدود العليا والدنيا، فالفئة 99-95 فإن منتصفها:

$$=\frac{99+95}{2} = \frac{194}{2} = 97$$

وهكذا لباقى الفئات. لاحظ فى هذا الجدول تم عرض الحدود الحقيقة Real limits للفئة وذلك لأن المتغيرات المتصلة تمتلك عدد لانهائى من القيم أو النقاط وعندما تقاس فإن نتيجة القياسات تناظر فترات على متصل السمة وليس نقاط مفردة فلو حصلت على درجة 99.5 فأين يتم وضعها فى الفئة الأولى أم الثانية، ولذلك تم إنشاء الحدود الحقيقة للفئات حتى يتم تضمين أى قيمة مثل 55.5 ، 70.5 وهكذا، وعلى ذلك فحدود الفئة الأولى 99-95 تسمى حدود ظاهرية Apparent limits، بينما 94.5- 99.5 تسمى حدود حقيقة. والحد الحقيقى الأدنى للفئة هو أصغر قيمة تقع فى الفئة بالمثل الحد الحقيقى الأعلى هو أكبر قيمة تقع فى الفئة وبعمل الحدود الحقيقة تصبح الفئات متصلة ولا يوجد فجوة بين فئة والفئة التى تليها.

التعبير عن البيانات فى ضوء فئات ينتج عنه عدم إعطاءنا صورة دقيقة عن التوزيع حيث يحدث فقد للمعلومات خاصة إذا كان حجم الفئات كبيرا. فمثلا فى الفئة 89-85 يوجد سبع تكرارات ولكننا لا نعرف ما هى الدرجات بالضبط التى أخذت هذه التكرارات فهل هى 89 , 85 , 86 , 87 .

التكرار النسبى والتكرار التراكمى والتوزيع المئوى التراكمى

التوزيع التكرارى النسبى يشير إلى نسبة العدد الكلى من الدرجات فى كل فئة، والتوزيع التكرارى التراكمى Cumulative frequent distribution يشير إلى عدد الدرجات التى تقع تحت الحد الحقيقى الأعلى لكل فئة، والتوزيع المئوى التراكمى Accumulative Percentage distribution تشير إلى النسبة المئوية للدرجات التى تقع تحت الحد الحقيقى الأعلى لكل فئة.

الفترية الفئوية	F	التكرار النسبي	التكرار التراكمي	التراكمى المئوى %
95-99	4	4/70= 0.06	70	7/70 100%
90-94	6	0.09	66	49.29
85-89	7	0.010	60	85.71
80-84	10	0.14	53	75.71
75-79	16	0.23	43	61.43
70-74	9	0.13	27	38.57
65-69	7	0.10	18	25.71
60-64	4	0.06	11	15.71
55-59	4	0.6	7	10.00
50-54	2	0.3	3	4.29
45-49	1	0.01	1	1.43
	70	1.00		

فالجدول السابق يعبر عن التوزيع التكرارى فى ضوء التوزيع التكرارى النسبى ويتم الحصول عليه من خلال خارج قسمة التكرار المقابل لكل فئة على العدد الكلى للدرجات:

$$Relative\ F = \frac{F}{N}$$

فمثلًا الفئة(45−49) تكرارها1 وعليه فإن التكرار النسبى لها:

$$Relative\ F = \frac{F}{N} = \frac{1}{70} = 0.01$$

وهكذا للفئة (56−59) فإن تكرارها النسبى:

$$Relative\ F = \frac{7}{70} = 0.10$$

وهذا الإحصاء مفيد فى أنه يخبرنا عن نسبة الدرجات فى كل فئة.

أما التكرار التراكمى لكل فترة نحصل عليه عن طريق جمع تكرار الفئة وتكرارات كل الفئات التى تقع تحتها فمثلا التكرار التراكمى للفئة 60−64 هى: 4 + 2 + 4 =11

1+4 وهو يخبرنا بان 11 طالب حصلوا على درجات تقع فى الفئة 60−64 فأقل بمعنى 11 طالب حصلوا على درجة 64 فأقل.

اما التكرار التراكمى المئوى لكل فئة يتم حسابه عن طريق قسمة التكرار التراكمى للفئة إلى العدد الكلى للأفراد مضروبا فى 100 %

$$\text{Cum } \% = \frac{cum\ F}{N} \times 100$$

- $cum\ F$ التكرار التراكمى

وللفئة 60–64 فى التكرار التراكمى المئوى كالآتى :

$$\text{Cum } \% = \frac{11}{70} \times 100 = 15.71$$

ويمكن إضافة عمود آخر يمثل التكرار النسبى المئوى. والتكرار التراكمى والتراكمى المئوى مفيد فى حساب المئينيات الرتب المئينية.

التمثيل البيانى للبيانات

العرض البيانى لبيانات كيفية(اسمية أو رتبية)

اولًا: الوصف والتمثيل البيانى للمتغيرات الكيفية

1. **رسومات Bar و Pie:** يتم وصف بيانات المتغيرات الكيفية من خلال التكرارات والنسب المئوية حيث يتم وضع الفئات على المحور السيني (X) والتكرارات على المحور الصادي (Y) وشكل الـ .Bar(▭) الرأسى وطوله يناظر عدد القيم أو التكرارات والمسافة بين الأعمدة تؤكد على عدم اتصالية المتغير.

وفيما يلى بيانات لمتعير كيفى وهو عدد الطلاب فى تخصصات كلية الطب:

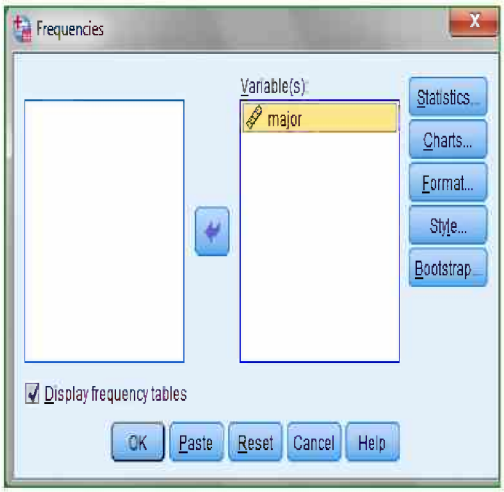
% F	Prop F	F	التخصص
42	0.42	92	باطنه
15	0.15	33	جراحة
38	0.38	85	أطفال
5	0.05	11	جلدية
100	1.00	221	المجموع

لعرض التوزيع التكرارى فى SPSS:

أولًا: إدخال البيانات:1. اضغط Variable view

2. اكتب مسمى المتغير major مثلًا تحت عمود Name ويرمز متغير التخصص باطنة = 1 ، جراحة = 3، أطفال = 2، جلدية = 4.

3. اضغط علىdata view ثم قم بإدخال البيانات فى عمود واحد.



ثانيًـــا: تنفيـــذ الامـــر: 1. اضـــغط علـــى Analyze ثـــم اختـــار Descriptives ثـــم اضغط على Frequencies يعطى الشاشة:

2. انقـل المتغيـــرmajor مـن خـلال الضـغط علـى السـهم ⇒ إلـى مربـع Variable، ثـم اضغط على OK.

ثالثًا: المخرج: أعطى البرنامج إحصائياتStatistics لمتغير major كالتالى:

Frequencies

Statistics

major

N	Valid	221
	Missing	0

major

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	92	41.6	41.6	41.6
	2.00	85	38.5	38.5	80.1
	3.00	33	14.9	14.9	95.0
	4.00	11	5.0	5.0	100.0
	Total	221	100.0	100.0	

- حجم العينة الكلى 221 =**Valid** بينما لا توجد بيانات غائبة **Missing**
- العمـود الأول هـو القـيم(الكـود) وهـى أربعـة تخصصـات كمـا تـم ترميزهـا، فمـثلًا الكـود (1) تخصـص باطنـة يوجـد أمامـه التكـرار (العمـود الثانى) **Frequency**=92 فردًا
- النسبة المئوية **Percent** :

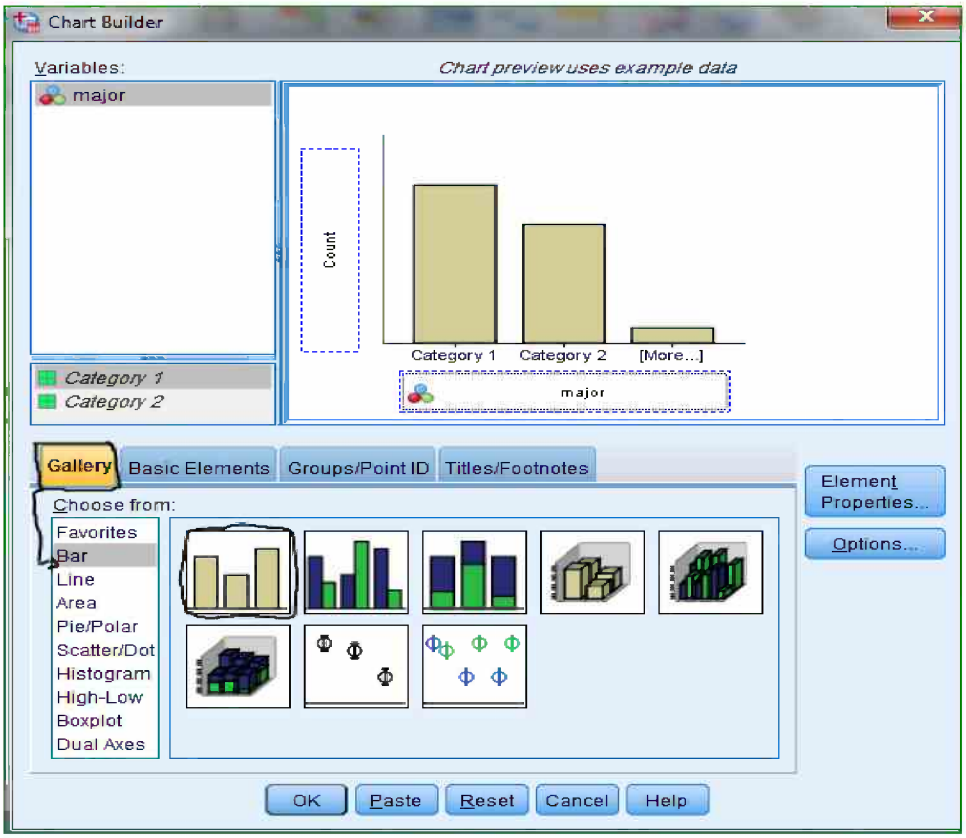
$$Percent = \frac{92}{221} \times 100 = 41.6$$

إذا 41.6% من طلبة الطب فى تخصص باطنة و 38.5 % تخصص جراحة.

- **Valid percent** النسـبة المئويـة الحقيقـة أو الصـادقة وقيمتهـا مثـل Percent ولكنها تختلف عنها إذا وجدت بيانات مفقودة.
- **Cumulative percent** النسبة المئوية التراكمية وهى مفيدة حيث 80.1% من أفراد العينة هم فى تخصص الباطنة والجراحة.

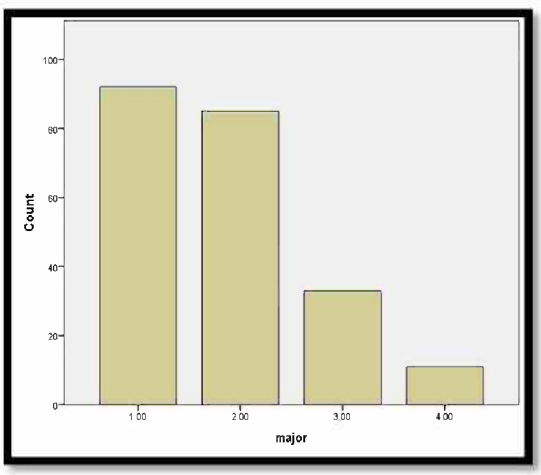
التمثيل البيانى فى ضوء Bar

1. اضغط على Graphs واختار Builder Chart يعطى شاشة بها خيارين Ok و Define variable properties اضغط على OK تظهر النافذة الآتية:



2. اضغط على Gallery توجد قائمة بها أشكال أو رسومات بيانية كثيرة.

3. اضغط على اختيار Bar على شمال الشاشة أسفلfrom Choose ثم اضغط مرتين متتاليين Simple Bar(المظلل) يظهر أعلى الشاشة.



4. فى مربع Chart preview uses أعلى الشاشة اضغط على متغير major ثم اسحبه إلى محور Axis

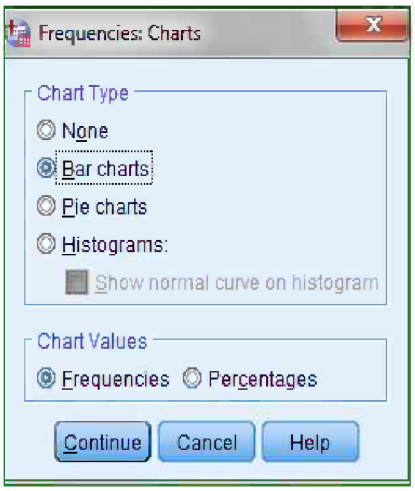
5. اضغط OK يظهر **المخرج:** يتضح من المخرج ان التخصص واحد له أعلى التكرارات وهو 97، يليه تخصص 2 ثم تخصص 3 وأقل التخصصات هى الجلدية 4.

ويمكن تنفيذ شكل Bar كالآتى:

1. اضغط على Analyze ثم اختر statistics Descriptive ثم اضغط على Frequencies

2. انقل المتغير major الى مربع Variables

3. اضغط على اختيار Charts تظهر الشاشة الآتية:



4. اضغط Bar charts

5. فى المربع Values Chart يظهر Percentage,Frequencies، اضغط على Frequencies ويمكن أن تختار النسبة المئوية.

6. اضغط على Continueلترجع إلى شاشة الأمر الرئيسية ثم اضغط OK.

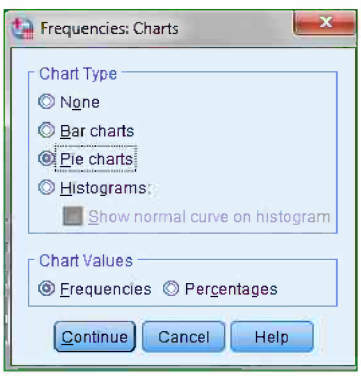
المخرج: سبق عرضه.

شكل Pie الدائرى

1.اضغط على Analyze ثم اختار statistics Descriptive ثم اضغط على Frequencies

2. انقل المتغير major الى مربع Variables

3. اضغط على اختيار Charts تظهر الشاشة التالية:

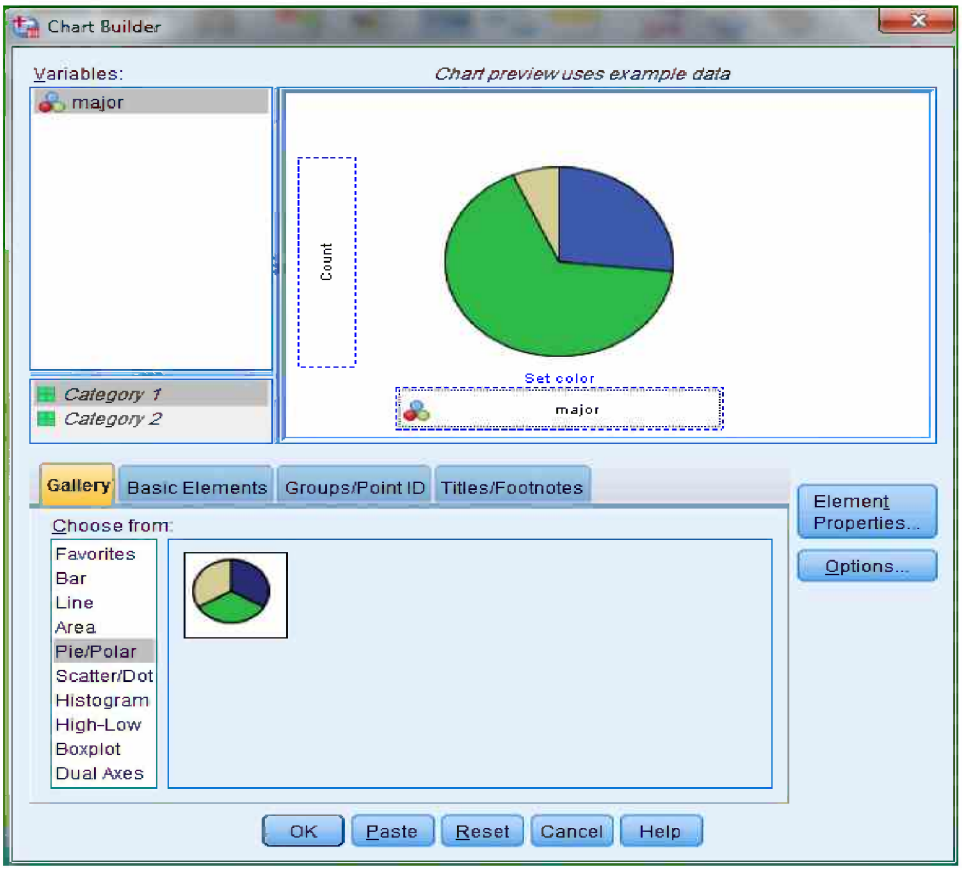


4. اختار charts Pie يمكن أن تختار عرض شكل Pie فى ضوء التكرارات أو فى ضوء النسبة المئوية فاختار النسبة المئوية.

ويمكن تنفيذ pie من قائمة graphs :

1. اضغط على قائمة Graphs ثم اختار Chart Builder

2. تظهر شاشة أخرى بها اختارين OK، variable Define اضغط على OK يعطى الشاشة التالية:

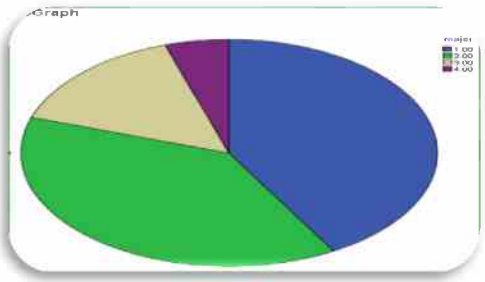


3. فـي وسـط الشاشـة يسـارًا اختـار Gallery فـي كـل الاحـوال هـي نشـطة فـي البرنـامج إذا لم تختارها ثم اضغط على Pie / polar

4. اضغط على الشكلPie في يمين الشاشة تظهر في الجزء الأعلى من الشاشة.

5. اسحب المتغير major من مربع Variables وضعه في محورِ Axis X

6.اضـغط علـى OK يعطـي شـكلPie كالآتي:



عرض شكل Bar للتعبير عن متوسطات مجموعات مستقلة

بفرض لدينا بيانات لمتغيرين تولدت من تصميم تجريبي ذو ثلاث مجموعات والمتغيرين هما:

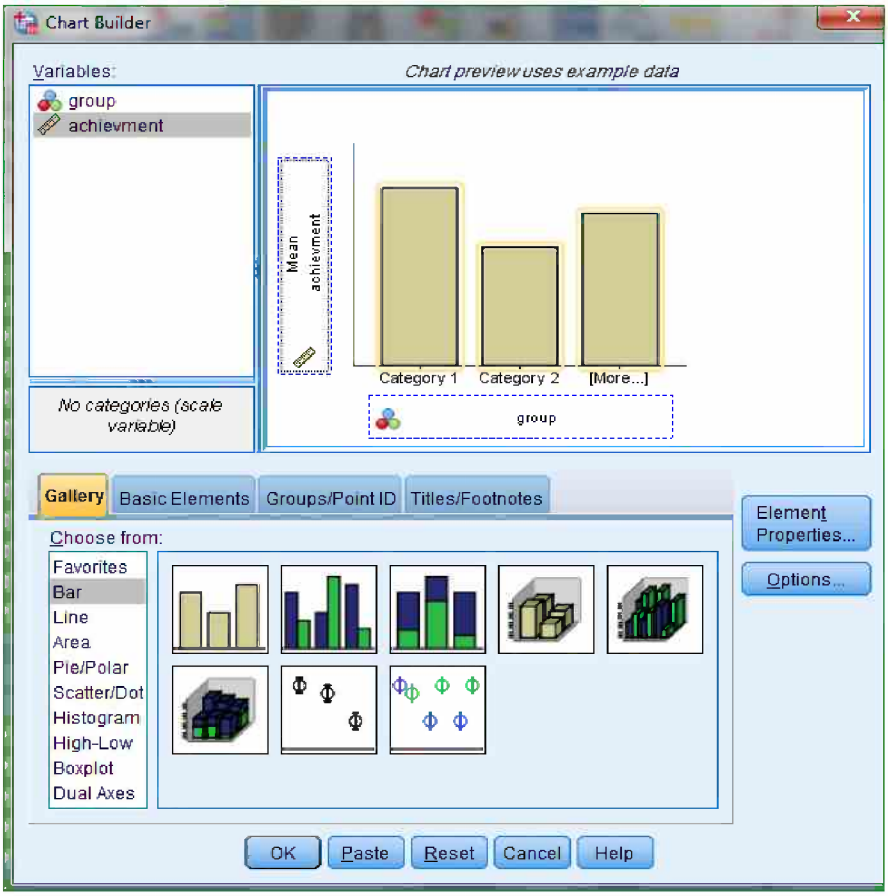
1. المجموعة group: ثلاثة مجموعات وتم تكويدهم 3,2,1 (مثل بيانات ANOVA أو حتى اختبار T المستقلة).

2. المتغير التابع achievement: التحصيل للمجموعات الثلاثة وللتعبير عن متوسطات المجموعات الثلاثة فى شكل Bar اتبع الخطوات الآتية:

1. اضغط على قائمة Graphs ثم اختار Chart Builder

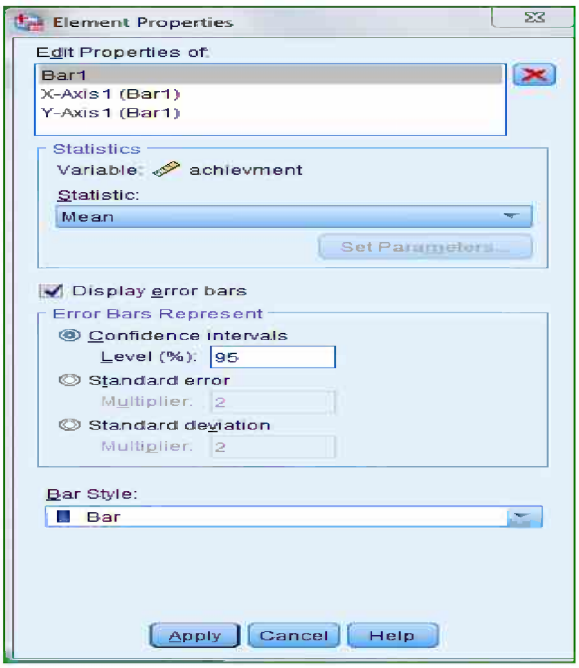
2. تظهر شاشة بها اختيارين احدهما OK اضغط على OK

3. اضغط علىGallery وتحتها عدة اختيارات اضغط على Bar تظهر عدة أشكال أو رسومات بيانية:



4. اضغط مرتين متتالين على أول شكل Simple Bar يظهر فى الجزء الأعلى من الشاشة رسم بيانى عبارة عن إحداثيات السينى X Axis وإحداثى Y Axis فالإحداثى X يوضع عليه المتغير المستقل (group) والإحداثى Y يوضع عليه المتغير التابعY(achievement)

5. على يمين الشاشة يوجد اختيار Element properties اضغط عليه يعطى نافذة فرعية كالتالى:



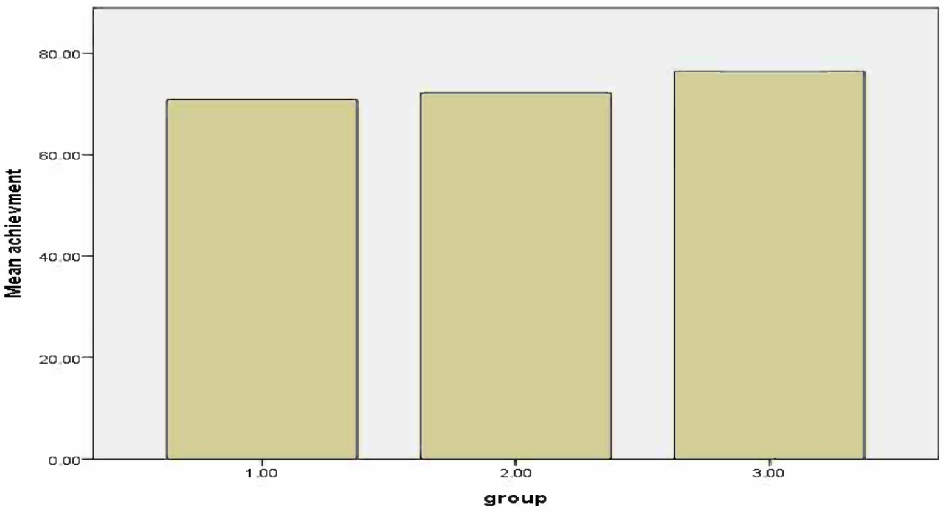
وهذه الشاشة أو المربع الحوارى يتضمن ثلاثة ملامح رئيسية هى:

- Statistics تحتها عدة بدائل اضغط على السهم يعطى اختيارات اضغط علي Mean.
- اضغط على Display error bars وهذا لأخطاء تظهر 95% فترة الثقة ويمكن تغيرها 90 مثلًا ويمكن عرض الانحراف المعيارى أو الخطأ المعيارى.

– يوجد اختيار Bar Style والبرنامج يحددها Bar إذا لم تختار أى بديل آخر ولذلك اضغط على البديل Bar إذا لم تظهر.

6. اضغط علي Apply أسفل الشاشة الفرعية.

7. اضغط OK يظهر المخرج والشكل يعرض المتوسطات للمجموعات الثلاثة:



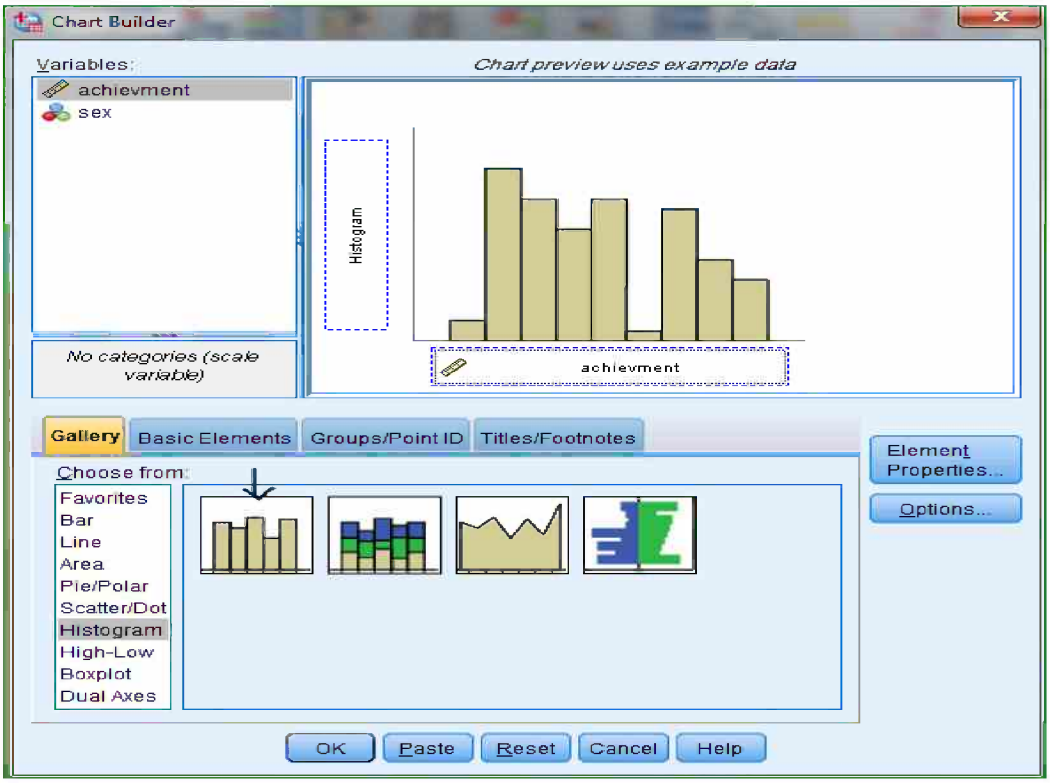
العرض البيانى للمتغيرات المتصلة (الفترية والنسبية)

اولًا: المدرج التكرارى Histogram: حيث توضع القيم أو الدرجات على المحور السينى والتكرارات على المحور الصادى وطول المستطيل يناظر تكرار كل درجة أوكل

فئـة ولعـرض المـدرج التكـرارى لمتغيـر واحـد لبيانـات التحصـيل فـى SPSS اتبـع الخطـوات الآتية(لمتغير التحصيل) :

1. اضـغط علـى قائمـة Graphs ثـم اختـار Chart Builder تظهـر رسـالة فـى مربـع اضغط على OK

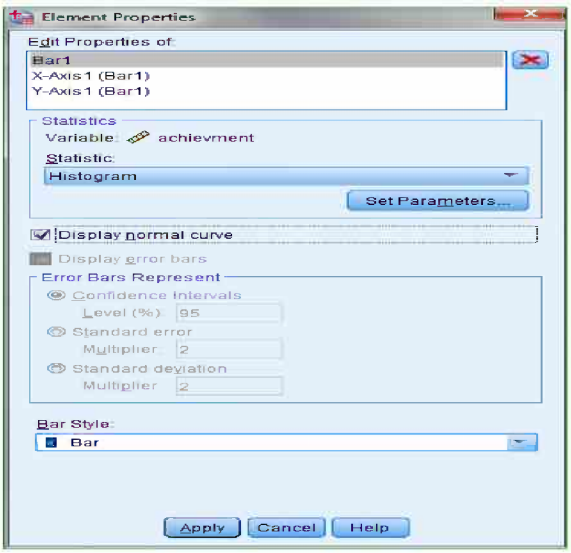
2. اضـغط علـى Gallery ثـم اضـغط علـيHistogram يظهـر لـك مجموعـة مـن الرسـومات علـى يمـين المربـع اضـغط علـى الشـكل الأول وهـو يعبـرعن Simple Histogram تظهر الشاشة الآتية:



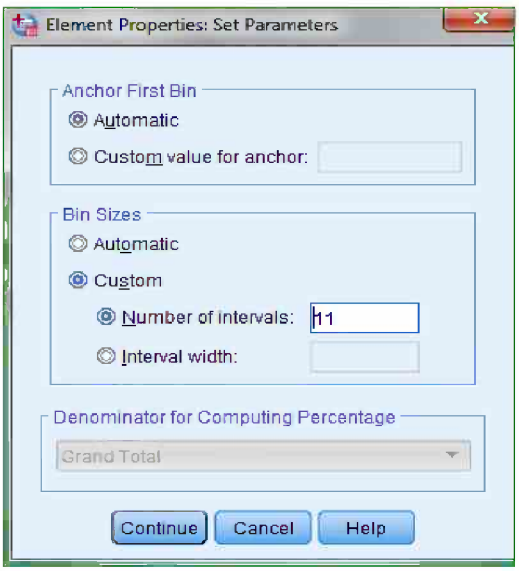
3. فـى الجـزء الأعلـى مـن الشاشـةdata example uses preview Chart يظهـر رسـم المـدرج التكـرارى اسـحب متغيـر Achievement فـى مربـع variables وضـعة أسـفل المدرج على محور Xعلى يمين الشاشة.

4. اضغط على Element properties على يمين الشاشة:

5.اضغط على السهم في مربع Statistics يعطي عدة بدائل اختار Histogram وهي تعتبر default



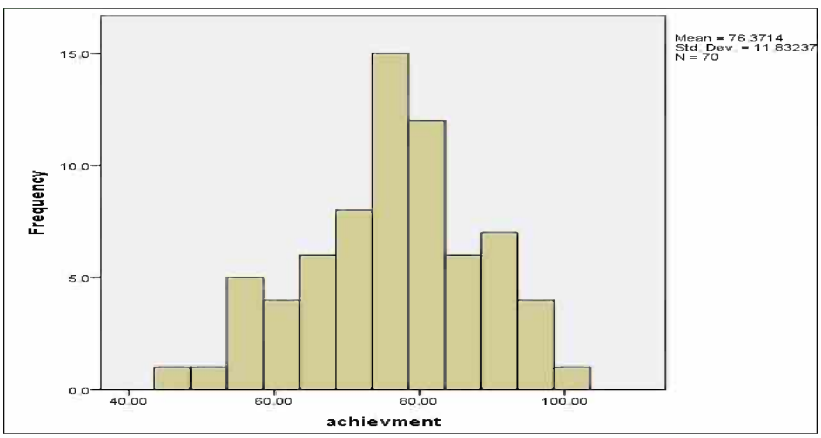
6. اضغط على Set parameters تظهر النافذة الآتية:



7. في مربع Bin sizes حيث يمكن أن نترك للبرنامج حرية في عمل الفئات أو يمكن تحديدها وذلك بالضغط على اختيار Custom ومنها يتم تحديد عدد الفئات مثلا 11 فئة كما في المثال أو يمكن تحديد طول الفئة بدلا من عدد الفئات وفي هذه الحالة اضغط على Custom ثم حدد طول الفئة في Interval width بـ 5 أو 10 أو 15 .

8. اضغط Continue ترجع إلى الشاشة الرئيسية، ثم اضغط على Apply

9. اضغط على OK يظهر المخرج الآتي:

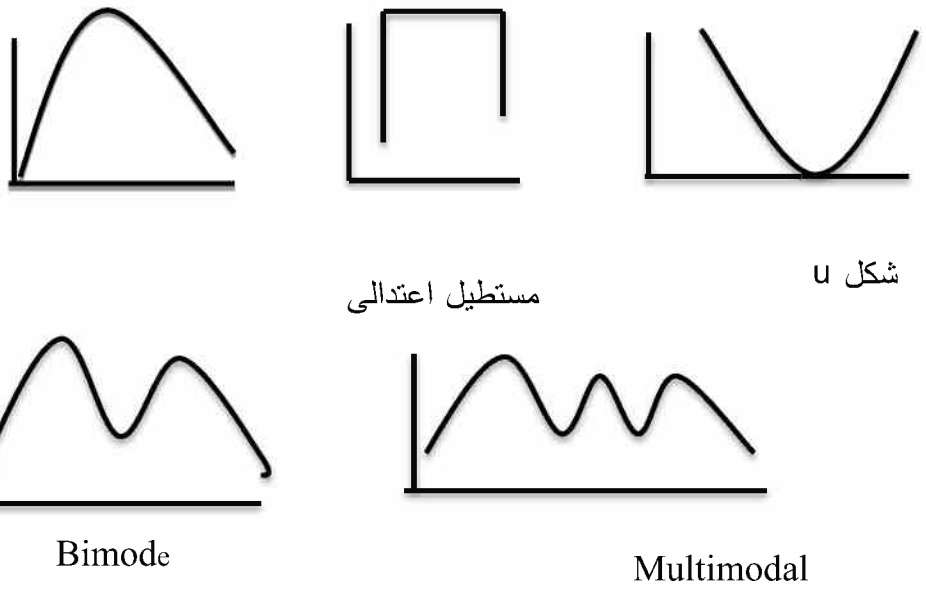


حيث إن غالبية الدرجات فى وسط المنحنى، وفى يمين المدرج عرض متوسط الدرجات Mean = 76.37 والانحراف المعيارى Std. DV= 11.832 وعدد الطلاب=70 ويمكن تنفيذ المدرج بدون تحديد عدد فئات ونترك الحرية للبرنامج لتحديدها ولتنفيذها اتبع كل الخطوات السابقة ولكن فى نافذة properties Element اضغط على Display normal curve ولا تقوم بتحديد عدد فئات أو طولها.

أشكال المنحنيات

تأخذ التوزيعات التكرارية أشكال عديدة مختلفة وتصنف إلى:

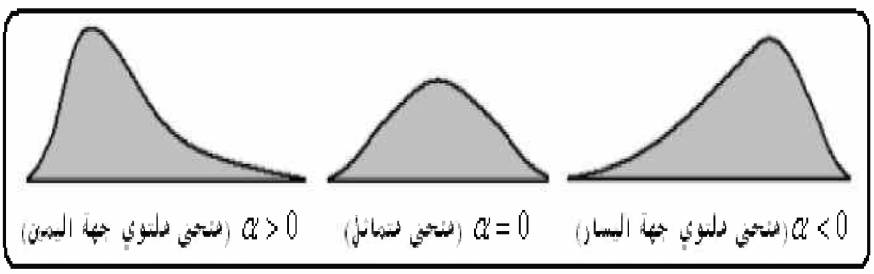
1. التوزيعات المنتظمة أو المتماثلةSymmetrical distribution :

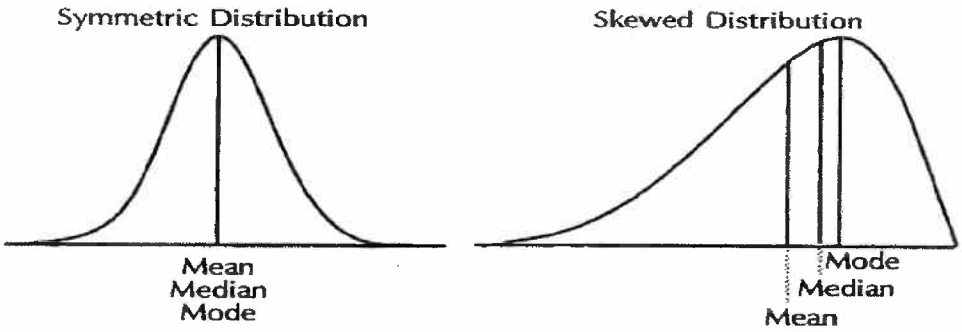


الشكل(1.6): امثلة لتوزيعات منتظمة.

وبالتالى لو رسمت خط رأسى فى المنتصف يكون أحد جانبى المنحنى صورة مصغرة للجانب الآخر.

2. **التوزيعات الملتوية:** فالتوزيع المنتظم عندما تقع الدرجات تقريبًا بالتساوى على جانبى المنحنى بينما التوزيع الملتوى كما فى الأشكال الآتية:





الشكل(2.6): توزيعات ملتوية واعتدالى.

وعندما يكون المنحنى ملتويًا التواءً موجبًا فإن معظم الدرجات تقع فى جانب المنحنى الأيسر حيث القيم المنخفضة على المحور الأفقى وذيل المنحنى يكون ناحية الجانب الأيمن حيث القيم المرتفعة ويكون:

المتوسط > الوسيط > المنوال

بينما الالتواء السالب فإن معظم الدرجات تكون فى جانب المنحنى الايمن حيث القيم المرتفعة وذيل المنحنى يكون ناحية اليسار حيث القيم المنخفضة ويكون:

المتوسط < الوسيط < المنوال

والتوزيع الذى يشبه مع الجرس يسمى منحنى اعتدالى Normal curve وهو منحنى ذو منوال واحد Unimodal وهو منحنى منتظم مع قمة مدببة بدرجة متوسطة ويكون تقريبًا:

المتوسط = الوسيط = المنوال

وفى التوزيعات الملتوية بدرجة متوسطة (Stuart & Ord, 1994):

المنوال- المتوسط ≅ 3(الوسيط- المتوسط)

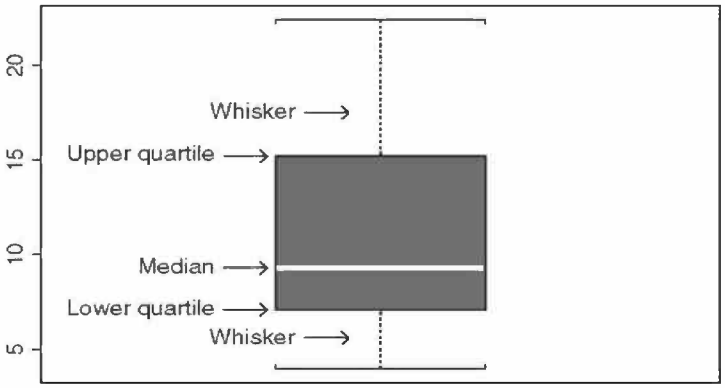
3. **التوزيع المتفرطح Kurtosis**: التوزيع المتفرطح هو يتشابه مع المنحنى الاعتدالى ولكن قمة المنحنى أكثر تفرطحًا أو تدببًا والتفرطح تأتى من الكلمة اللاتينية Kyros وهى تعنى منحنى.

وعند تمثيل التوزيع التكرارى فى شكل منحنى تكرارى، قد يكون هذا المنحنى منبسط أو مدبب، فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى ويقل فى طرفيه يكون المنحنى مدببا، وعندما يتركز عدد أكبر على طرفى

المنحنى ويقل بالقرب من المنتصف يكون المنحنى مفرطحًا أو منبسطًا.

شكل Boxplot

المدرج التكرارى لا يعطى توصيف دقيق عن مدى وجود قيم متطرفة فى التوزيع من عدمه وللذلك يتم عرض شكل Boxplot. فقام John Tuckey بتطوير عرض بيانى للبيانات وكذلك لوصف انتشار التوزيع فى شكل Boxplot أحد العروض البيانية لتمثيل المتغيرات المتصلة وكذلك للكشف عن القيم المتطرفة فى التوزيع ويتم عرضه فى شكل مستطيل:



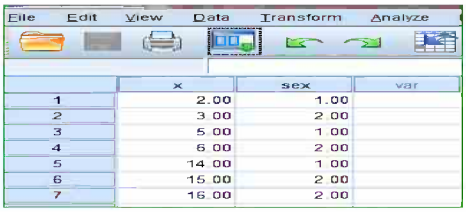
وإذا كان خط الوسيط داخل الصندوق قريبًا من الإرباعى الأعلى دل ذلك على ان التوزيع ذو التواء سالب، وإذا كان قريب من الإرباعى الأدنى دل ذلك على ان التوزيع ذو التواء موجب، والخط السميك داخل الصندوق هو قيمة الوسيط أو المئينى 50% بمعنى القيمة التى يقع فوقها % 50 من القيم وتحتها 50% من القيم والخطوط الراسية خارج الصندوق تسمى Whiskers وهى من أسفل تعكس القيم الدنيا ومن أعلى القيم العظمى وأيضًا تعرض القيم المتطرفة حيث القيم المنخفضة والمرتفعة فى البيانات يجب أن تكون أقل أو أعلى 1.5 طول الصندوق سواء من الإرباعى الأعلى أو الأدنى.

وتعرف القيمة المتطرفة Outlier هى أكبر من 1.5 طول الصندوق بعيدًا عنه بينما القيمة أو الحالة شديد Extreme case بانها تبعد 3.5 طول الصندوق.

مثال: فيما يلى درجات 15 طالب فى متغير ما وهى كالتالى:

CNO	X	
1	2	
2	3	
3	5	
4	6	الإرباعى الأولى = 6
5	14	
6	15	
7	16	
8	16	الوسيط = 16
9	16	
10	21	
11	22	
12	23	الإرباعى الأعلى = 23
13	24	
14	33	
15	45	

تنفيذ الشكل البيانىBoxplot للمتغير X فى SPSS

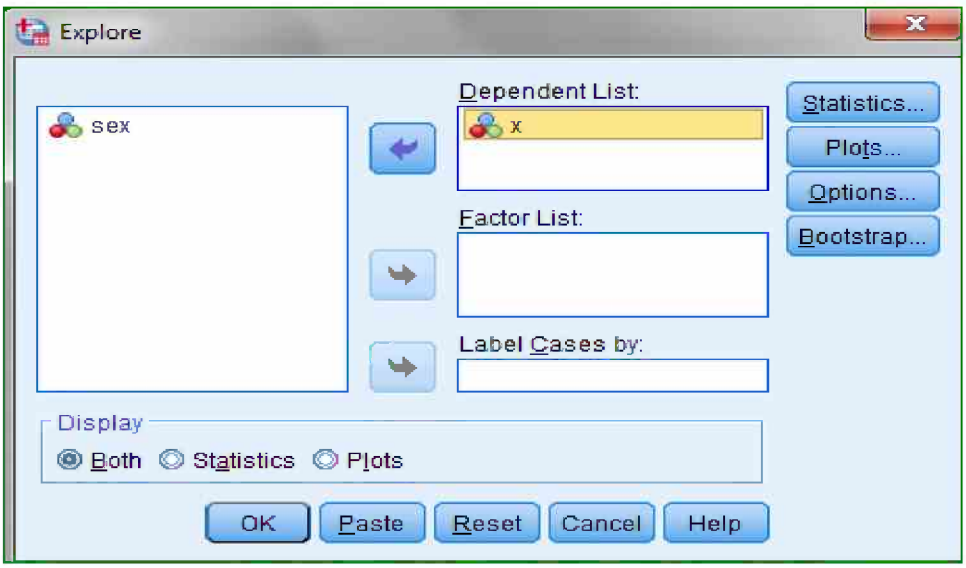


اولا: إدخال البيانات:1. اضغط variable view وفى عمود name اكتب مسمى المتغير x

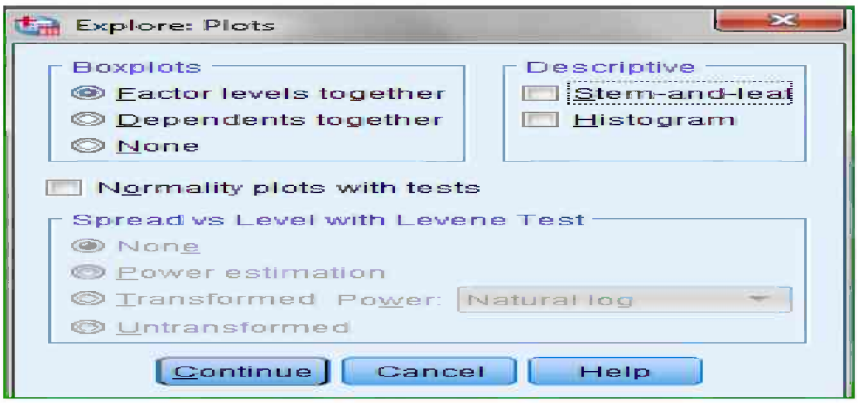
2. اضغط data view ابداء فى إدخال البيانات

ثانيًا: تنفيذ الأمر:

1.اضغط على Analyze ثم Descriptive statistics ثم Explore الشاشة:

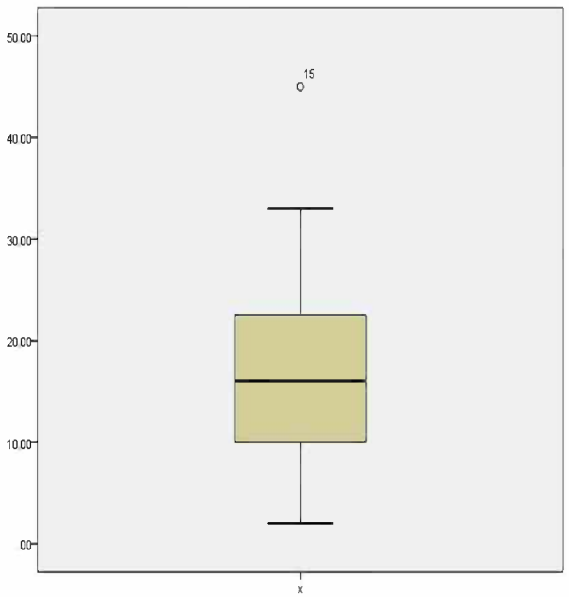


2. انقل المتغير X إلى مربع Dependentlist اضغط على اختيار Plots تظهر النافذة الآتية:



3. فى مربع Boxplot اضغط على اختيار Factors levels.

4. اضغط Continue ترجع إلى شاشة الأمر الرئيسية، ثم اضغط OK:



لاحظ أن الخط فى وسط المستطيل يمثل الوسيط بينما الإرباعى الأعلى يناظر 23 والإرباعى الأدنى 6 وان القيمة المناظر للخط السميك T(33) هى قيمة كبيرة فى التوزيع ولكنها ليست متطرفة وأعطى البرنامج العلامة (O) تعبر عن قيمة متطرفة هى تناظر قيمة الحالة (15) وهى 45 وإذا أعطى البرنامج العلامة (*) فهى تعبر عن قيمة متطرفة جدًا أى تعادل 3 مرات المدى الربيعى.

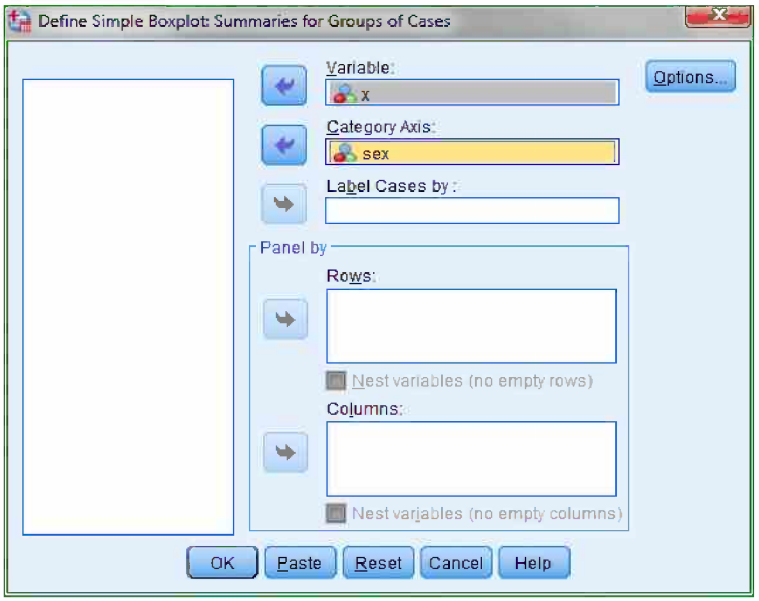
ولتنفيذ Boxplot لدرجات متغير كمى عبر مستويات متغير كيفي

1. اضغط على قائمة Graphs ثم اضغط legacy Dailogs واختار Boxplot تظهر الشاشة الآتية:

2. اضغط على Simple ثم اضغط على

Summaries for groups of cases

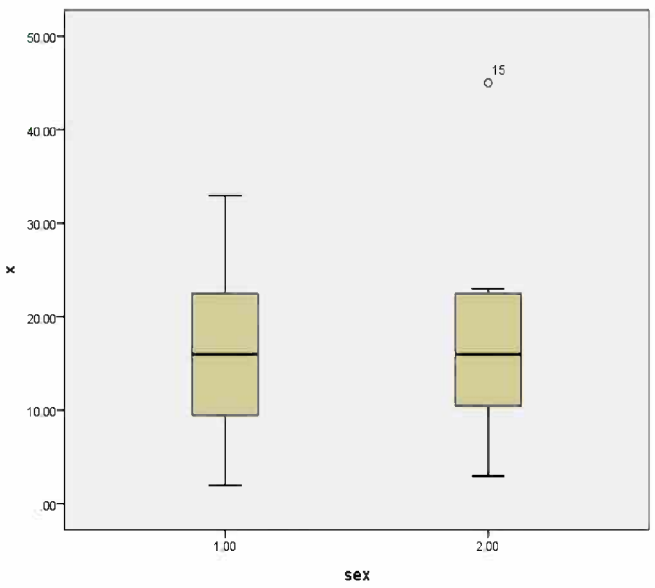
3. اضغط على Define تظهر الشاشة الآتية:



4. اضغط على متغير X وانقله↵إلى مربع Variable

5. اضغط علىSex ثم اضغط على↵ وانقله إلى مربع Axis Category ثم اضغط OK

ثالثًا: المخرج:



من الواضح أنه لا يوجد قيمة متطرفة في شكل Boxplot للذكور، بينما يوجد قيمة متطرفة لدرجات الإناث للحالة 15،لاحظ أن Boxplot هو عرض بصرى لتوزيع المتغير فى ضوء مؤشرات إحصائية عديدة وداخل الصندوق أو المستطيل يمكن الوصول إلى استنتاج حول تشتت البيانات فالصندوق الواسع يعبر عن تباين عالٍ بينما الصندوق الضيق يعبر عن تباين منخفض للبيانات.

شكل Stem and leaf

كمـا سـبق ان المـدرج التكـرارى يعـرض البيانـات فـى ضـوء قـيم مجمعـة فـى فئـات، وهـذا العـرض يفقـد كثيـرًا مـن المعلومـات الخاصـة بالـدرجات المفـردة داخـل الفئـات وعلـى الجانـب الآخـر فـإن التوزيـع التكـرارى للـدرجات المفـردة (لا يسـتخدم الفئـات) يبعـث علـى الملـل إذا كـان حجـم البيانـات كبيـرة، والمـدخل البـديل لتجنـب عيـوب المـدرج التكـرارى والتوزيـع التكـرارى هـو عـرض شـكل الورقـة أو الجـذع والسـاق وهـو مـدخل عـام لتحليـل البيانـات ويعرف (EDA) Exploratory Data Analysis وهو عرض بيانى آخر لتمثيل البيانـات وهـذا النـوع مـن العـرض مفيـد خاصـة لقواعـد البيانـات الصـغيرة لأن كـل درجـة يـتم تـدوينها فـى الشـكل البيـانى وبنـاء العـرض البيـانى للسـاق والورقـة يتطلـب تجميـع القـيم فـى Stems(تجمعـات) وبعـد ذلـك يـتم عرضـها عبـر كـل صـف ولتوضـيح عـرض هـذا الشـكل فيما يلى مجموعة من البيانات كالآتى:

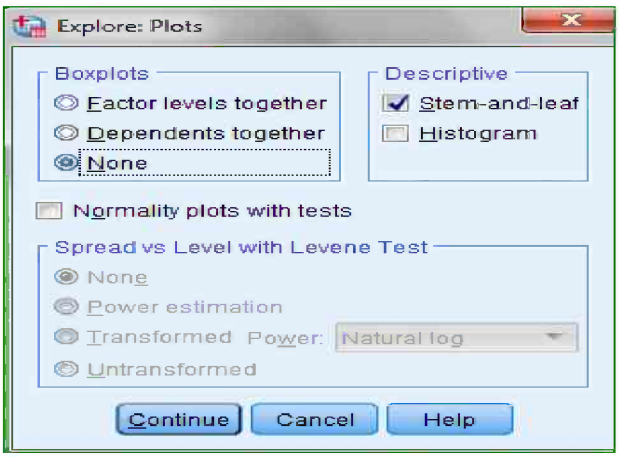
12, 33, 24, 42, 11, 14 , 23, 32, 46,10, 16, 26, 29, 15, 30, 47, 52, 44, 19

ولتنفيذ ذلك خلال SPSS اتبع الخطوات الآتية:

اولًا: إدخـــال البيانـــات:1.اضـغط view variable وفـى عمـود Name اكتـب اسـم المتغير X.

2. اضغط data view ابداء فى إدخال البيانات.

ثانيًـا: تنفيـذ الامـر:1. اضـغط Analyze ثـم Descriptive Statistics اضـغط Explore تظهر الشاشة (كما فى Boxplot).



2. انقل المتغير X إلى مربع Dependent list ثم اختار plots تظهر:

3. فى مربع Descriptive اضغط على Stem and leaf

4. اضغط Continue ثم أضغط OK

ثالثًا: المخرج:

```
x

x Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

     7.00        1 .  0124569
     4.00        2 .  3469
     3.00        3 .  023
     4.00        4 .  2467
     1.00        5 .  2

 Stem width:     10.00
 Each leaf:       1 case(s)
```

فالصف الأول 0124569. 1 فالقيمة قبل النقطة أو العمود (.) هى Steam بينما القيم بعد النقطة هى Leaf وهذا يعنى ان القيم التى تبدأ بالواحد هى 11,12,14 ،10 وهكذا وتكرارها 7 قيم، بينما القيم التى تبدأ بـ 2 هى 3, 4, 6, 9 وهكذا.

ولبناء شكل الورقة والساق توضع القيمة 12 مع الساق1 وتوضع القيمة 50 مع 5، ويتم وضع اول صف بالقيم التى تبدأ بـ 1 وهى (10 , 11 , 12 , 14 , 15 , 16 , 19) واربع قيم تبدأ مع 2 وهى (23 , 24 , 26 , 29) و ثلاث قيم تبدأ بـ 3 وهى (30 , 32 , 33) وهكذا. ونلاحظ فى الصف الأول 1.0 لها أعلى تكرار 7 وعليه يمكن استنتاج ان المنحنى ملتوى التواء موجب لأن معظم القيم تكون منخفضة. وأسفل الشكل نلاحظ أن طول المسافة Stem width هى 10 بينما أى رقم فى leaf يمثل الحالة أو الحالات.

ولكن ماذا لو كان Stem width = 1 فإن الصف الأول قيمة تكون كالآتى , 1.4 1.5 , 1.6 , 1.9 , 1.0 , 1.1 , 1.2 وهكذا. ويشار للقيم 1 , 2 , 3 , 4 , 5 شمال النقاط الراسية بـ leading digits وهى تكون الساق Stem أو المحور الرأسى فمثلا القيم 10 , 11 , 12 فإن وحدات الكسر Unitdigit هى 0 , 1 , 2.

حقيقة يتشابه هذا العرض مع المدرج التكرارى فى إعطاء صورة عامة عن التوزيع.

الفصل السابع

مقاييس النزعة المركزية

Measures of Central Tendency

فى الفصل السابق تناولنا تنظيم البيانات فى ضوء التوزيعات التكرارية للبيانات غير المجمعة (الدرجات الخام) أو للبيانات المجمعة(الفئات)، وكذلك تم عرضها بيانيًا لإعطاء صورة عن شكل توزيع الدرجات سواء كان منحنى اعتدالى أو ملتوى أو متفرطح ولكن هذا ليس كافيًا لوصف بيانات أى توزيع حيث أنه لا يعطى وصف كمى ولا يعطى مؤشرات كمية للتعبير عن طبيعة التوزيع، فافترض أن باحث أراد تحديد ما إذا كان توجد فروق بين الذكور والإناث فى الاستعداد الرياضى فيصبح لديه مجموعتين من الدرجات أحداهما للذكور والأخرى للإناث، فكيف تتم المقارنة بين توزيع الدرجات بينهما ولإجراء هذه المقارنة فإننا نحتاج إلى تكميم هذه التوزيعات ويتم هذا من خلال حساب متوسط الدرجات لكل مجموعة ثم المقارنة.

وهذه النوعية من المقاييس أو المؤشرات يطلق عليها مقاييس النزعة المركزية Measures of Central tendency ويشار إليها أحيانًا بمقاييس الموقع location وهى تعرض مجموعة من القياسات تعكس نقطة تمركز التوزيع وتختلف هذه المؤشرات من حيث توظيفها فى ضوء طبيعة البيانات واعتداليتها ومدى وجود قيم متطرفة ولكنها تحاول ان تخبرنا بعض الشيء عن أين يقع مركز التوزيع أو تحاول ان تخبرنا عن القيمة التى يقترب أو ينحو إليها معظم درجات التوزيع، ويعرفها Gravetter & Wallanw (2014) بأنها مقياس إحصائى يحاول تحديد قيمة وحيدة عادة تقع فى مركز التوزيع وهى قيمة نموذجية تعكس مجموعة من الدرجات.

مقاييس النزعة المركزية

تتضمن مقاييس النزعة المركزية ثلاثة مؤشرات هى المتوسط والوسيط والمنوال:

المتوسط Mean

يعرف أيضًا بالمتوسط الحسابى Arithmetic averge ويحسب عن طريق مجموع كل الدرجات فى التوزيع مقسوما على عدد هذه الدرجات ويرمز بالمتوسط فى المجتمع بالرمز μ وتنطق ميو وتسمى معلم Parameter ومتوسط العينة بالرمز $\bar{X}$ وتنطق X

بـارا أو M وتسـمى إحصـائية Statistic ولكـن معظـم المجلـدات الإحصـائية تسـتخدم $\bar{X}$ ليعبـر عـن متوسـط العينـة بينمـا فـى الدراسـات المنشـورة دائمًـا يسـتخدم الرمـز M ويقـدر كالآتى :

$$M = \bar{X} = \frac{\sum X_i}{n}$$

$$= \frac{x_1 + x_2 + x_3 + \cdots \ldots \ldots + x_n}{n} = \frac{\sum x_n}{n}$$

- $\sum X_i$ مجموع قيم X
- n عدد القيم

بينما متوسط المجتمع يقدر كالآتى :

$$\mu = \frac{\sum X_i}{N}$$

حيث:

- $\sum$ (سيجما) مجموع.
- n أو N عدد الدرجات أو القيم فى التوزيع.

مثال: للدرجات الآتية: 3 , 5 , 12 , 5

$$\bar{X} = \frac{3+5+12+5}{4} = \frac{25}{4} = 6.25$$

مثال: للدرجات الآتية : 20, 22, 28, 30, 37, 38

$$\bar{X} = \frac{20+22+28+30+37+38}{6} = \frac{175}{6} = 29.17$$

المتوسط الموزون أو العام Weighted mean or overall mean

هـذا المؤشـر أو المقيـاس يحـدد متوسـط مجموعـة درجـات لعينـات مختلفـة مـن حيـث العـدد أو عينـات غيـر متسـاوية الأفـراد بكـل تأكيـد لـنفس المتغيـر بالتـالى متغيـر واحـد يقـاس علـى عينتين فأكثر غير متساوية الحجم وتكون صيغته كالآتى:

$$\overline{X_w} = \frac{\sum \bar{X}_i n_i}{\sum n}$$

$$\overline{Xw} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 \ldots\ldots\ldots\ldots + ni\bar{x}_1}{n_1 + n_2 + n_3 + \ldots\ldots\ldots\ldots + n_i}$$

- $\bar{X}$ متوسط كل عينة.
- n حجم كل عينة.

$A \quad \pi$
$= \quad r^2$

مثال: التحصيل فى ثلاثة عينات مختلفة كالآتى:

العينة	$\bar{X}$	n
الريف	46	12
الحضر	58	14
البدو	73	30

لاحظ أن حجم العينة فى كل تصنيف مختلف وإذا تم حساب المتوسط العادى للعينات الثلاثة يكون كالآتى:

$$\overline{X_w} = \frac{46 + 58 + 73}{12 + 14 + 30} = 59$$

فان المتوسط العام للمتوسطات الثلاثة معًا 59 وهذا غير صحيح لأنه لم يأخذ فى الاعتبار عدد الدرجات التى أسهمت فى حساب كل متوسط على حدة وعليه فإن المتوسط الموزون كالآتى:

أ – يقدر المجموع الموزون كالآتى:

$= (46 \times 12) + (58 \times 14) + (73 \times 30) = 3554$

ب– قسمة المجموع الموزون على أحجام العينات كالآتى:

$$\overline{X_w} = \frac{3554}{12 + 14 + 30} = 63.5$$

لاحظ أن المتوسط الموزون 63.5 بينما المتوسط الحسابي 59.0 وعليه فإن المتوسط الموزون ينحو نحو متوسط العينة الأكبر حجمًا (أو العينة الأكثر وزنًا) وهذا هو الواقع الحقيقى وهذا يجعل المتوسط الموزون أكثر دقة لحساب متوسط عينات غير متساوية الأحجام وعلى ذلك فالمتوسط الموزون يستخدم فى حالة وجود عينة طبقية.

مثـال أخـر: اجـرى أسـتاذ امتحـان فـى علـم الـنفس لشـعبتين وكـان متوسـط الشـعبة الأولـى 90 وعـدد الطـلاب 20، ومتوسـط الشـعبة الثانيـة 70 وعـدد الطـلاب 40 وعليـه فـإن المتوسط العام للشعبتين معا كالآتى:

$$\bar{X}_w = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\bar{X}_w = \frac{20(90) + 40(70)}{20 + 40} = 76.67$$

نلاحظ أن المتوسط العام قريب من متوسط درجات الشعبة الثانية الأكبر حجمًا.

خصائص المتوسط

- **المتوسـط حسـاس لكـل قيمـة فـى التوزيـع:** عنـد حسـاب المتوسـط فإننـا نضـيف كـل القـيم فـى المعادلـة ولـذلك فـأى تغيـر فـى أى درجـة يسـبب تغيـر فـى قيمـة المتوسـط وهـذه الخاصية غير متوفرة فى حساب المنوال والوسيط.
- مجموع انحرافات القيم عن متوسطها يساوى صفرا ويرمز لها $\sum(X - \bar{X}) = 0$ ويمكن توضيح ذلك بالمثال الآتى:

X	$\bar{X}$	$X - \bar{X}$
1	3	1-3= -2
2	3	2-3 = -1
3	3	3-3 = 0
4	3	4-3 = 1
5	3	5-3 = 2

وعليه فإن:

$$\bar{X} = \frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

$$\sum(x - \bar{X}) = -2-1+0+1+2 = 0$$

وهـذا يرجـع إلـى حقيقـة أن قيمـة المتوسـط هـى نقطـة التـوزان Balance point فـى التوزيع.

- المتوسـط شـديد الحساسـية للقـيم شـديدة التطـرف Exterm values أو Outlier وهـذا مـن حقيقـة هـى أن المتوسـط يعتبـر النقطـة أو القيمـة المحوريـة أو التـى يتمركـز حولهـا القـيم فإذا كان لدينا الشكل الآتى:

X	2	4	6	8	10

$\overline{X} = 6.00$

فالمتوسط 6 يتمركز حوله درجات التوزيع وعليه فإضافة قيمة متطرفة وليكن 50 إلى التوزيع السابق فيصبح المتوسط كالآتى:

$$\overline{X} = \frac{2+4+6+8+10+50}{6} = 13.33$$

وعليه يحدث تغير دراماتيكى لقيمة المتوسط من 6 إلى 13.33 علما بان هذه القيمة لا تعكس تمامًا حقيقة درجات التوزيع ولكن حدث ذلك نتيجة وجود قيمة متطرفة.

- مجموع مربعات انحرافات القيم عن متوسطها فى التوزيع تعطى قيمة أقل من مجموع انحرافات القيم عن أى قيمة أخرى فى التوزيع $\sum(X-\overline{X})^2$ بمعنى أقل قيمة.

مثال:

X	$(X-3)^2$	$(X-4)^2$	$(X-5)^2$	$(X-6)^2$
2	$(2-3)^2=1$	$(2-4)^2=4$	$(2-5)^2=9$	$(2-6)^2=16$
4	1	0	1	4
6	9	4	1	0
8	25	16	9	4
	36	24	20	24

وحيث $\overline{X}=5$ كما هو واضح ان مجموع مربعات انحرافات القيم عن المتوسط – $(X \sum \overline{X})^2$ هو أقل قيمة 20 مقارنة بمجموع مربعات انحرافات القيم عن أى قيمة أخرى فى التوزيع. وهذه الخاصية على درجة كبيرة من الأهمية فى الإحصاء الاستدلالى بمعنى تحت كل الظروف فإن المتوسط هو أقل عرضة أو تأثرًا لتباينات درجات العينة وهذه الخاصية أيضًا موجودة لمؤشرى الوسيط والمنوال ولكن المتوسط هو أقل مؤشرات النزعة المركزية تأثرًا بالاختلافات فى درجات العينة بكل تأكيد ما عدا وجود قيم شديدة التطرف فى التوزيع وهذا ما يجعل المتوسط ملك الإحصاء الاستدلالى لأنه أكثر استقرارًا بالتغيرات التى تحدث فى قياسات الظاهرة. ولكن تذكر أن المتوسط يستخدم للمتغيرات الكمية التى يعبر عنها رقميًا وأيضًا لحسابه تستخدم كل الدرجات فى التوزيع ولذلك فهو يعطى صورة جيدة للقيم فى التوزيع وأيضًا من أهم مميزات المتوسط أنه يدخل فى حساب مؤشرات إحصائية أخرى مثل الانحراف المعيارى والتباين ولذلك فإنه

يلعب دورًا فى غاية الأهمية فى الإحصاء الاستدلالى وعلى ذلك فهو أفضل مؤشرات النزعة المركزية على الإطلاق ولكن فى مواقف معينة يفضل استخدام الوسيط والمنوال.

المتوسط المهذب ($\bar{X}_T$) Trimmed means

هو متوسط يقدر من البيانات حيث يستبعد نسبة مئوية من نهايتى التوزيع، فعلى سبيل المثال إذا كان 100 ملاحظة أو درجة ونريد حساب10%المتوسط المهذب بمعنى نستبعد 10% من درجات طرفى التوزيع يستبعد 10% من درجات العليا و 10% من الدرجات الدنيا ويقدر المتوسط للدرجات الباقية وهذا يعطى تقدير جيد للمتوسط للمجتمع فإذا وجدت عينة لها درجات ذات تشتت أو اختلافات عالية حيث يوجد مجموعة درجات عالية جدًا أو منخفضة جدًا فإن متوسط العينة لا يكون تقدير جيد لمتوسط المجتمع ولكن بحذف القيم المتطرفة يعطى متوسط أكثر استقرارًا والسبب الآخر لحساب المتوسط المهذب وهو التغلب على محددات المتوسط وهو التعامل مع التوزيعات الملتوية وقد ايد ذلك بشدة Wilcox (2009) وكذلك يدعم Howell (2014) ويجب على الباحثين التعامل مع هذا المؤشر لأنه حديث فى التراث البحثى.

مثال: احسب المتوسط المهذب 10% للبيانات الآتية:

37, 14, 26, 17, 21, 43, 25, 6, 9, 11

الحل: 1. ترتيب القيم تصاعديًا:

6, 9, 11, 14, 17, 21, 25, 26, 37, 43

2. وعليه يوجد عشر درجات ويتم حذف 10%من الدرجات القصوى و 10% من الدرجات الدنيا حيث:

$$\frac{10 \times 10}{100} = 1$$

وعليه يتم استبعاد الدرجتين 6 و43 .

3. حساب المتوسط:

$$= \frac{9 + 11 + 14 + 17 + 21 + 25 + 26 + 37}{8} = 20$$

ويقدر المتوسط المهذ ب 20% كالآتى:

$$\frac{20 \times 10}{100} = 2$$

وبالتالى يتم استبعاد أعلى قيمتين وأدنى قيمتين وعليه يكون عدد القيم المتبقية هى 6 كالآتى: 11, 14, 17, 21, 25, 26

ويكون:

$$\bar{X} = \frac{11+14+17+21+25+26}{6} = 19$$

وبفرض أنه يوجد 99 درجة فى التوزيع ونرغب فى حساب المتوسط المهذب 20% فإن الصيغة هى :

عدد القيم المحذوفة =0.2n = 0.2 x 99 = 19.8 = 1.9%.

Awinsorized Mean

عندما يتم حساب المتوسط المهذب 10% تم استبعاد 10% من القيم الدنيا والعليا من الدرجات ولكن عند حساب $X_{Awin}10\%$ يتم حذف القيمتن 43 , 6 ويتم استبدالهما بالقيمة التى تلى 6 وكذلك القيمة التى تسبق 43 كالتالى:

6, 9, 11, 14, 17, 21, 25, 26, 37,43

يصبح التوزيع كالآتى:

9, 9, 11, 14, 17, 21, 25, 26, 37, 37

ويصبح المتوسط:

$$10\%\bar{X}_w = \frac{9+9+11+14+17+21+25+26+37+37}{10} = 20.6$$

أما المتوسط $20\%\bar{X}_w$ يتم من خلال استبعاد أدنى قيمتين وكذلك أقصى قيمتين كالآتى:

11, 14, 17, 21, 25, 26

واذا استُبدلت أدنى قيمتين ب 11 وأقصى قيمتين بـ 26 ويصبح التوزيع كالآتى :

11, 11, 11, 14, 17, 21, 25, 26, 26,26

ويكون المتوسط كالآتى:

$20\%\ X_w = 18.8$

وهذه التهذيبات للمتوسط لمعالجة التوزيعات الملتوية وكذلك القيم المتطرفة فى التوزيع.

وهذ المتوسطات المهذبة تقترب من فلسفة Bootstrap.

الوسيط Median

هـو المقيـاس الثـانى للنزعـة المركزيـة والهـدف مـن الوسـيط هـو تحديـد النقطـة الوسـيطة أو المنتصـفة فـى التوزيـع Midpoint ، ويعرفـه (Gravtter & Wallanu (2014، بانـه إذا كانـت درجـات التوزيـع موضـوعة فـى ترتيـب مـن الاصـغر إلـى الاكبـر فهـو النقطـة المنتصـفية فـى التوزيـع، أو بكلمـات اخـرى هـو النقطـة علـى المقيـاس التـى تحتهـا 50% مـن درجـات التوزيـع وفوقهـا 50% مـن الـدرجات، ويعـرف أيضًـا بالنقطـة التـى تقسـم الـدرجات إلـى مجمـوعتين متسـاوين الحجـم أو التكـرار أو العـدد وحسـاب الوسـيط يعتمـد على عدد الدرجات.

ويـرى Pagano (2013) بـان الوسـيط هـو المئينـى خمسـين P50 ويرمـز لـه بالرمز (Md) ورمزه واحد سواء كان للعينة أو للمجتمع.

حساب الوسيط

اولًا: عدد الدرجات فى التوزيع فردى Odd : فلحساب الوسيط لا بد من:

1. ترتيب الدرجات تصاعدى أو تنازلى
2. حساب رتبة الوسيط وهى:

$$Md = \frac{n+1}{2}$$

- n عدد الدرجات.

3. تحديد الدرجة التى تعادل رتبة الوسيط وهى قيمة الوسيط

مثال: للتوزيع الآتى: 1 ,2, 5 , 4 , 8 , 7 , 9

لحساب الوسيط:

1. ترتيب الدرجات تصاعدى أو تنازلى:

1,2, 4 , 5 , 7 , 8 , 9

2. تحديد رتبة الوسيط:

$$Md = \frac{7+1}{2} = 4$$

3. تحديد الدرجة التى رتبتها 4وهى الدرجة 5 إذا Md = 5

ثانيًا: حساب الوسيط إذا كان عدد القيم زوجى Even

مثال: للتوزيع الآتى: 10 , 3 , 5 , 15 , 20 , 9

لحساب الوسيط اتبع الخطوات الآتية:

1. ترتيب الدرجات تصاعدى أو تنازلى:

3, 5 , 9 , 10 , 15 , 20

2 . تحديد رتبة الوسيط وفى هذه الحالة فإن للوسيط رتبتين كالآتى:

$$d1 = \frac{n}{2} = \frac{6}{2} = 3$$

$$d2 = \frac{n}{2} + 1 = \frac{6}{2} + 1 = 4$$

3. تحديد الدرجات المقابلة للرتب 3 , 4 هى الدرجة 9 والدرجة 10

4. حساب متوسط الدرجتان: $Md = \frac{9+10}{2} = \frac{19}{2} = 9.5$

إذا القيمة 9.5 يقع تحتها ثلاث درجات وهى تعادل 50% من قيم التوزيع

خصائص الوسيط

1. الوسيط أقل تأثرًا بالدرجات المتطرفة مقارنة بالمتوسط فمثلًا الوسيط للأتى:

5, 8, 9, 15, 16

5, 8, 9, 15, 206

فان الوسيط للتوزيعين 9 وهذا يعنى أن الوسيط لم يأخذ فى اعتباره أثناء حسابه القيم المتطرفة مثل 206 عكس المتوسط الذى تأثر بها تأثرًا شديدًا وتكون قيمته 48.2 وهى لا تعكس التوزيع على الإطلاق حيث إنها بعيدة عن أربع قيم فى التوزيع وهى 5, 8, 9, 15 وهى أيضًا بعيدة عن القيمة المتطرفة 206. وعلى ذلك فإن الوسيط له أهمية كبيرة أثناء التعامل مع خصائص أو صفات تتضمن وجود قيم متطرفة. فمثلًا فى تجارب عندما تكون القياسات بالثانية يحدث أن أداءات بعض الأفراد بها تطرف شديد للدرجات مثل: 30, 1.7, 6, 5, 1.3, 1.2 ، وعليه فإن استخدام الوسيط يعتبر ضرورى فى هذه الحالة وبالتالى عندما تكون التوزيعات ملتوية سالب أو موجب يفضل استخدام الوسيط كمؤشر للنزعة المركزية وليس المتوسط.

يوجد مواقف يستخدم فيها الوسيط كبديل جيد للمتوسط وهى:

- الدرجات المتطرفة أو التوزيعات الملتوية.

- فـى مواقـف بحثيـة لا تأخـذ فيهـا المتغيـرات درجـات كـان يتعامـل مـع رتـب مـثلا وفـى هـذه الحالـة لا يمكـن اسـتخدام المتوسـط ويسـتخدم الوسـيط كبـديل جيـد بمعنـى القياسـات تكـون نابعـة مـن مسـتوى القيـاس الرتبـى ويكـون مـن الأفضـل إجـراء اختبارات فروض حول الوسيط وليس المتوسط.
- فـى توزيعـات النهايـات المفتوحـة Open – ended distribution مثـل التوزيـع الآتى:

X	F
5 فاكثر	3
4	2
3	2
2	3
1	6

وعليـه فـلا يسـتطيع الباحـث حسـاب المتوسـط ولكـن يمكـن حسـاب الوسـيط لأنـه لا تـدخل النهاية المفتوحة 5 فأكثر فى حساباته وحيث يكون الوسيط = 1.5.

2. الوسـيط أكثـر عرضـة للاختلافـات أو تباينـات العينـة مـن المتوسـط ولكنـه أقـل عرضـة لاختلافات العينة من المنوال.

المنوال Mode (MO)

هـو القيمـة الأكثـر حـدوثًا أو تكـرارًا أو شـيوعًا فـى التوزيـع وعلـى العكـس فـإن المتوسـط وأحيانًـا الوسـيط قـد يكونـوا قـيم ليسـت فـى التوزيـع وميـزة المنـوال أنـه يظهـر القيمـة التـى يحصـل عليهـا أكبـر عـدد مـن الأفـراد وهـذا المؤشـر يتعلـق أكثـر بالمتـاجر أو حمـلات البيـع للأزيـاء أو غيرهـا حيـث يركـز الأفـراد علـى الـزى الأكثـر اسـتخدامًا ويعتبـر أسـهل مقـاييس النزعة المركزية حسابًا.

مثال : للدرجات الآتية: 5 , 15 , 9 , 8, 5, 12, 11, 6, 5

فان المنوال يساوى 5 لأن كان لها أعلى تكرار ثلاث مرات.

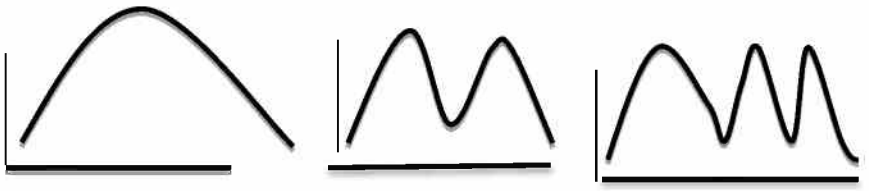
ويقـال إن التوزيـع فـى هـذه الحالـة ذو منـوال واحـد Unimodal ويمكـن للتوزيـع أن يكـون له أكثر من منوال مثل التوزيع الآتى:

1 , 2 , 3 , 3 , 3 ,3, 4 , 5 , 7 , 7 , 7 , 7 , 8 , 1

وعلى ذلك فيوجد منوالين هما 3 , 7 ويقال إن التوزيع ذو منولين Bimodal.

وعلى الرغم من سهوله حسابه إلا أن استخدامه فى العلوم النفسية والسلوكية محدودًا لأنه لا يتميز بالاستقرار من عينة إلى أخرى وغالبا ما يوجد أكثر من منوال للتوزيع.

وكذلك يوجد توزيعات متعددة المناويل Multimodal



Unimodal Bimodal Multimodal

مقاييس النزعة المركزية والمنحنى المنتظم

إذا كان التوزيع أحادى المنوال وذو توزيع منتظم أو متماثل فان:

المتوسط = الوسيط = المنوال

وهذا يتحقق فى المنحنى الاعتدالى، إذا كانت قيم المتوسط والوسيط والمنوال مختلفة فإن التوزيع غير منتظم. وإذا كان التوزيع ملتويًا فإن المتوسط لا يساوى الوسيط لأن المتوسط يتأثر بالدرجات المتطرفة والمتوسط يكون قريبًا من الدرجات المتطرفة ولذلك فإن التوزيع الملتوى التواء سالب يكون المتوسط أكبر من الوسيط وعلى العكس التوزيع الملتوى التواء موجب فإن المتوسط يكون أقل من الوسيط.

اختيار مقياس النزعة المركزية المناسب

اختيار مقياس النزعة المركزية فى المواقف الآتية:

المتوسط: درجات ذات توزيع اعتدالى، ومتغيرات فترية على الأقل.

الوسيط: درجات ذات توزيع ملتوٍ ومتغيرات رتبية وأيضًا فترية ونسبية بعد تحويلها إلى رتب وتوزيع ذات النهايات المفتوحة – وجود قيم متطرفة فى التوزيع.

المنوال : متغيرات من مستوى قياس اسمى ومتغيرات منفصلة ويستخدم كوصف للمنحنى بجانب المتوسط والوسيط ، تذكر ان المنوال يحدد موقع القيمة العليا Peak أو القمم فى الشكل البيانى سواء كان مدرج أو منحنى ولكن إذا كانت لديك مجموعة بيانات وكان متوسطها = 72 ومنوالها = 80، فمن الأفضل استحدام المتوسط فقط لأنه يعطى صورة صادقة للتوزيع.

وفيما يلى ملخص للمواقف التى يستخدم فيها مؤشرات النزعة المركزية:

المؤشر	شكل التوزيع	مستوى القياس
المتوسط	اعتدالي	فترى ونسبي
الوسيط	ملتوي	ترتيبي
المنوال	نموذجى Modal	اسمي

كتابة مقاييس النزعة المركزية فى تقارير البحوث وفقا لـ APA

تستخدم مقاييس النزعة المركزية بصورة شائعة فى أبحاث العلوم النفسية والاجتماعية والسلوكية لتلخيص ووصف نتائج الدراسة البحثية، فعلى سبيل المثال يكتب الباحث متوسطات عينتين أو أكثر لوصف الفروق بين المعالجتين أو كتابة الوسيط ويتم عرض هذه المقاييس فى صورة جداول وكذلك عرضها على الرسومات.

فى تقرير النتائج فإن كثير من المجلات يستخدم إرشادات الجمعية النفسية الأمريكية APA كما هو معروض فى كتابها السنوي(2010) والجمعية استخدمت الرمز (M) للتعبير عن متوسط العينة وغالبا ما تستخدم الدراسات العبارة الآتية:

أظهرت المجموعة التجريبية أن متوسط التحصيل (M=30) مقارنة بالمجموعة الضابطة (M = 25) وان وسيط التحصيل Md=27 للمجموعة التجريبية مقارنة بوسيط التحصيل للمجموعة الضابطة Md=20

أو يمكن عرضها كالآتى:

	التجريبية	**الضابطة**
إناث	20	15
ذكور	25	17

إجراء مؤشرات النزعة المركزية فى برنامج SPSS

مثال: فيما يلى الوقت الذى استغرقه 20 طالبًا لكتابة عشرة استخدامات للكمبيوتر فى اختبار الابتكارية :

41	46	38	115	64	36	122
65	48	90	80	59	80	100
123	87	132	80	51	143	

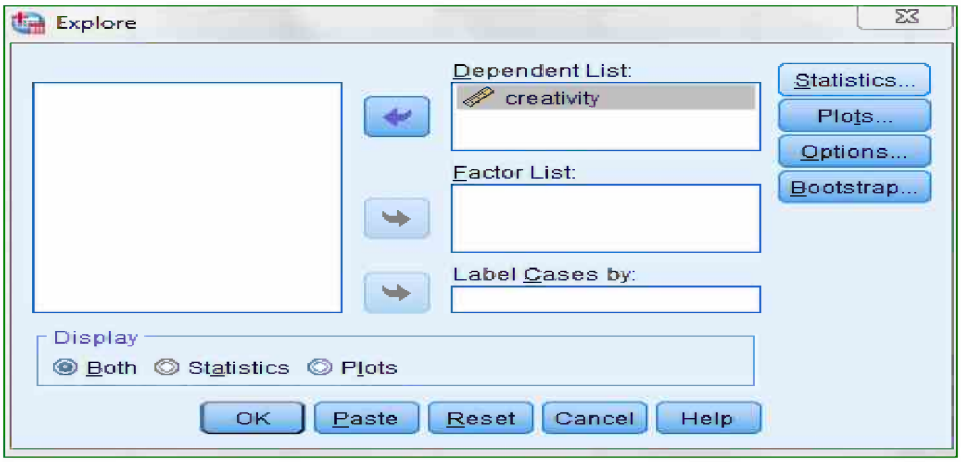
لإجراء مقاييس النزعة المركزية اتبع الآتى:

أولًا: إدخـــال البيانـــات: 1. بعـد فـتح البرنـامج اضـغط علـــيview أسـفل Variable

الشاشة على اليمين ثم اكتب اسم المتغير Creativity تحت عمود Name

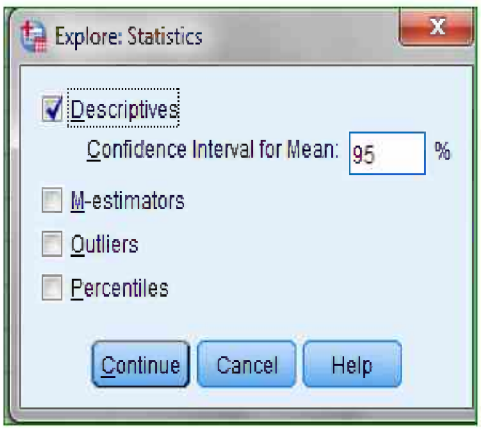
2.اضـــغط علـــى أيقونـــةview تظهـــر شاشـــة البيانـــات حيـــث يشـــغل متغيـــر Creativity عمود واحد ثم ابدأ فى إدخال البيانات.

ثانيًا: تنفيذ الأمر:1- اضغط على Analyze ثم Descriptive Statistics ثم اضغط على Explore تظهر الشاشة الآتية :



2. انقل متغير creativity إلى مربع Dependent list من خلال السهم ↪

3. اضـــغط علـــى اختيـــار Statistics تظهـــر الشاشة:



4. اضغط على Descriptive

5. اضـــغط علـــــيContinue لتعـــود إلـــى الشاشة الأمر الرئيسية ثم اضغط على OK

ثالثًا: تفسير المخرج:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
creativity	20	100.0%	0	0.0%	20	100.0%

Descriptives

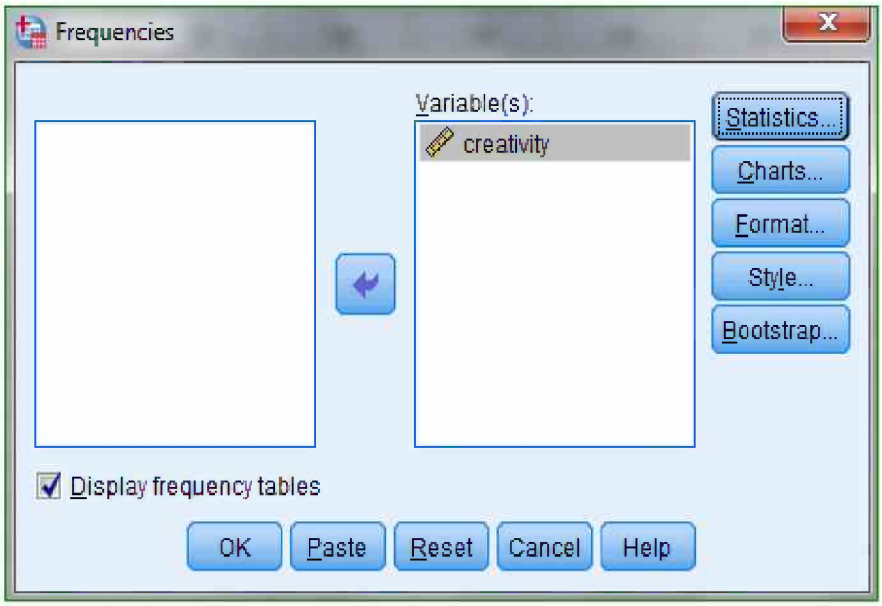
			Statistic	Std. Error
creativity	Mean		80.0000	7.47346
	95% Confidence Interval for Mean	Lower Bound	64.3579	
		Upper Bound	95.6421	
	5% Trimmed Mean		78.9444	
	Median		80.0000	
	Variance		1117.053	
	Std. Deviation		33.42234	
	Minimum		36.00	
	Maximum		143.00	
	Range		107.00	
	Interquartile Range		62.50	
	Skewness		.415	.512
	Kurtosis		-1.001-	.992

أعطى إحصائيات وصفية للمتغير حيث إن المتوسط = 80 والوسيط = 80، وهذا يعطى انطباع ان التوزيع اعتدالى، لاحظ أن المخرج أعطى المتوسط المهذب5% Trimmed Mean وهذا معناه أن البرنامج استبعد 5% من القيم العليا و5% من القيم الدنيا وأصبحت قيمته 78.949 ونلجأ إلى هذا المؤشر عندما يوجد قيم متطرفة فى التوزيع، كما اعطى البرنامج مؤشرى الالتواء=0.415 والتفرطح Kurtosis = 1.001 وهذا يشير إلى وجود التواء موجب بدرجة خفيفة ولكن ما زال المنحنى اعتدالى ولكن التفرطح قيمته زادت عن الواحد وهى قيمة منخفضة أيضًا حيث القيم المرتفعة للتفرطح تزيد عن 7وعندئذ يقال أن التوزيع متفرطح.

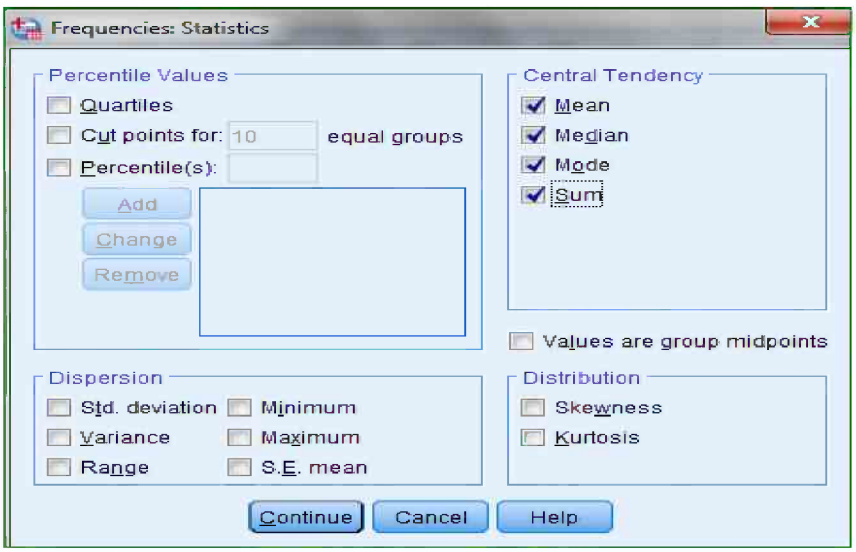
ولكن المخرج السابق لم يعطى المنوال Mode.

ويمكن تنفيذ مقاييس النزعة المركزية:

1.اضغط علىAnalyze ثم Descriptive Statistics ثم اضغط على Frequencies تظهر الشاشة:



2. انقل متغير Creativity إلى مربع Variable(s) ثم اضغط `Statistics:



3. على يمين الشاشة فى مربع Central Tendency اضغط علي , Mean ,Sum , Median اضغط على Mode ثم اضغط Continue. ويمكن عدم تنشيط Display frequency table فى حالة عدم الحاجة أليها وذلك فى شاشة الأمر الرئيسية.

- **المخرج:**

```
FREQUENCIES VARIABLES=creativity
  /STATISTICS=MEAN MEDIAN MODE SUM
  /ORDER=ANALYSIS.
```

Frequencies

Statistics

creativity

N	Valid	20
	Missing	0
Mean		80.0000
Median		80.0000
Mode		80.00
Sum		1600.00

valid– N: عـدد القـيم الفعليـة فـى التحليل = 20

Missing–: القيم المفقودة = 0

Mean – المتوسط = 80:

Median- الوسيط= 80

Mode - المنوال= 80

Sum - مجموع الدرجات= 1600

والمتوسط = الوسيط = المنوال وعليه فتوزيع الدرجات اعتدالى. ثم عرض البرنامج التوزيع التكرارى وهو مكونة من خمسة أعمدة:

Creativity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	36.0	1	5.0	5.0	5.0
	38.0	1	5.0	5.0	10.0
	41.0	1	5.0	5.0	15.0
	46.0	1	5.0	5.0	20.0
	48.0	1	5.0	5.0	25.0
	51.0	1	5.0	5.0	30.0
	59.0	1	5.0	5.0	35.0
	64.0	1	5.0	5.0	40.0
	65.0	1	5.0	5.0	45.0
	80.0	3	15.0	15.0	60.0
	87.0	1	5.0	5.0	65.0
	90.0	1	5.0	5.0	70.0
	100.	1	5.0	5.0	75.0
	115.	1	5.0	5.0	80.0
	122.	1	5.0	5.0	85.0
	123.	1	5.0	5.0	90.0
	132.	1	5.0	5.0	95.0
	143.	1	5.0	5.0	100.0
	Tota l	20	100.0	100.0	

العمود الأول: Valid القيم أو الدرجات مرتبة تصاعديًا من الأصغر للأكبر.

العمود الثانى: Frequency وهو تكرار كل قيمة أو درجة فى التوزيع حيث الدرجة 36 حصل عليها طالب واحد بالتالى تكرارها1 أما الدرجة 80 فحصل عليها ثلاث طلاب وهكذا.

العمود الثالث: النسبة المئوية Percent النسبة المئوية للذين حصلوا على درجة معينة فمثلا الدرجة 36 حصل عليها 5%، 15%حصلوا على الدرجة 80.

العمود الرابع: النسبة المئوية المصححة Valid Percent : عمود Percent يكون مقامهN للطلاب الحاضرين والغائبين اما فىpercent Valid يكون المقام عدد الطلاب الحاضرين فقط ونسبتها هى نفس النسبة percent إذا لم يوجد بيانات غائبة.

العمود الخامس: التكرار التراكمى المئوى التصاعدى Cumulative percent وهى يبدأ بأول نسبة مئوية مصححة 5.0 ثم يجمع النسبة التى يليها مباشرة 5.0 + 5.0 = 10 وهكذا و 60% من الطلاب حصلوا على الدرجة 80 فأقل وعليه % 35 من الطلاب حصلوا على الدرجة 59 فأقل وهكذا.

الفصل الثامن

مقاييس التشتت والالتواء والتفرطح والمنحنى الاعتدالي

Measures of Dispersion, Skeweness, Kurtosis, & Normal curve

فى الفصل السابق تناولنا مقاييس عديدة مرتبطة بمركز التوزيع ومهما يكن فإن القيمة المنتصفية للتوزيع سواء المتوسط أو الوسيط أو المنوال لا تعطى صورة كاملة عن التوزيع، وعليه فنحن بحاجة إلى مقاييس إضافية لتوضح إلى أى درجة بيانات أو قياسات الأفراد تتبعد أو تنحرف عن هذه القيمة المتوسطة، فالمتوسط ربما يعكس الموقع العام لمعظم الدرجات، والمقاييس التى توضح كيفية انتشار أو ابتعاد درجات الأفراد حول هذه القيمة المتوسطة وكذلك حول بعضها البعض تسمى مقاييس الاختلاف أو التشتت أو الانتشار وهذه النوعية من المقاييس على درجة كبيرة من الأهمية لسببين هما:

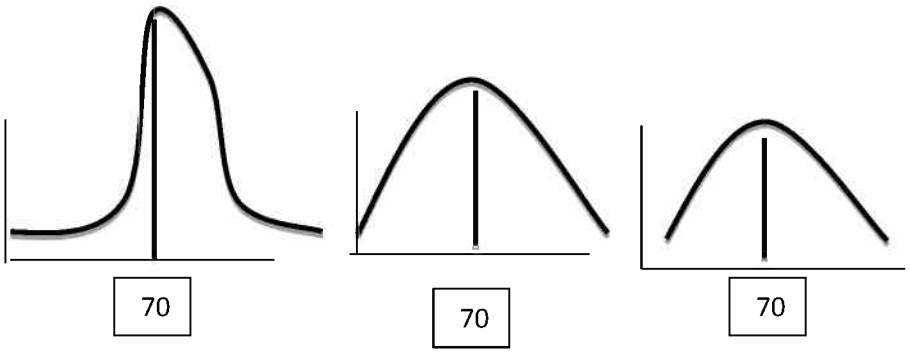
1. انتشار أو تشتت البيانات متطلب أساسى لكثير من الاختبارات الإحصائية الاستدلالية مثلr, T, F.

2. استخدام مقاييس التشتت فى تقيم البرامج فى التصميمات التجريبية، فعند تقيم برنامج فالباحث ليس مهتم فقط بالقيمة المتوسطة لكن أيضًا بكيف أن درجات الأفراد على المتغير تختلف عن بعضها البعض وكذلك عن المتوسط وأيضًا لمعرفة ان تأثير البرنامج هو نفسه أم مختلف لمستويات مختلفة من الطلاب أو الأفراد أو عبر مجموعات مختلفة.

ومصطلح Variability له نفس المعنى فى الإحصاء، فنحن نطلق على مفهوم ما بالمتغير وهذا يعنى أنه له حالات مختلفة وليس حالة واحدة وهكذا بالنسبة لدرجات التوزيع فإذا كانت كلها لها نفس القيمة بالتالى لا يوجد اختلاف أو تنوع (8 , 8 , 8)، وإذا وجدت فروق صغيرة بين الدرجات بالتالى فالاختلاف يكون صغير , 7 , 8 ,8)

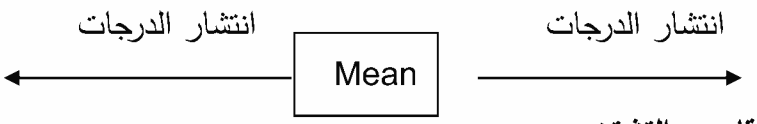
(9، وإذا وجدت فروق كبيرة فإن الاختلاف يكون كبير، وعلى ذلك فإن الاختلاف يعطى مؤشر كمى للفروق بين الدرجات ويصف مدى انتشار أو تجمع الدرجات وقيمته تتراوح من 0 إلى ∞ +

وفيما يلى الشكل الآتى:



الشكل (1.8): توزيعات مختلفة درجات التحصيل لنفس المتوسط.

فالأشكال الثلاثة لها نفس المتوسط ولكن شكل انتشار الدرجات مختلف، وعليه فإن مؤشر النزعة المركزية غير كافٍ لوصف توزيع معين، ومقاييس التشتت تعكس انتشار الدرجات حول نقطة مركزية ويعبر عنها فى ضوء مسافة Distance على المحور الأفقى أو محور X



وفيما يلى مقاييس التشتت:

المدى Range

يعرف بأنه الفرق بين الدرجة العليا أو العظمى والدرجة الدنيا أو الصغرى، وهو المسافة التى يغطيها الدرجات فى التوزيع من أصغر قيمة إلى أكبر قيمة ويرمز له بالرمز R

$$R = X_{Max} - X_{Min}$$

- X_{Max} هى أكبر قيمة فى التوزيع
- X_{Min} هى أصغر قيمة.

وتكتب المعادلة السابقة فى ضوء الحدود الحقيقية للدرجات كالآتى:

$$R = X_{UL} - X_{LL}$$

- X_{UL} حيث الحد الحقيقى الأعلى
- X_{LL}الحد الحقيقى الادنى

ويمكن حساب المدى من الصيغة الآتية:

$$R = X_{Max} - X_{Min} + 1$$

والصيغة الأولى هى التى تستخدم أثناء عرض الكتاب بينما الصيغة الثانية هى المستخدمة فى برامج الكمبيوتر.

مثال: 2 , 3 , 5 , 8 , 10

$$R = 10 - 2 = 8$$

ولكن إذا اعتمدنا على صيغة المدى من الحدود الحقيقة فان:

$$R = 10.5 – 1.5 = 9$$

مثال:

18, 12, 28, 15, 20

$$R = 28 – 18 = 16$$

ويفضل(2014) Gravetter & Wallanw صيغة الحدود الحقيقة حيث يعتمد عليها برنامج SPSS وأن هذه الصيغة تصلح للمتغيرات المتصلة.

وإذا كانت الدرجة القصوى لاختبار الذكاء 144 والدنيا 76 فإن المدى:

$$R = 144.5 – 75.35 = 69$$

والمدى 69 لدرجات الذكاء هو المسافة على المحور السينى X أو الخط الأفقى الذى يتضمن 100% من الدرجات.

والمدى سهل حسابه ولكنه يعطى مؤشر نسبى ليس بالثقة الذى نعتمد عليه لأن المدى يقيس انتشار الدرجات المتطرفة فقط وليس انتشار الدرجات بين القيمتين الصغرى والعظمى، وعليه فإن حسابه يعتمد على قيمتين فقط وليس كل درجات التوزيع ولذلك لا يستخدم بدرجة كبيرة فى الدراسات وحيث لا يتميز بالاستقرار من عينة إلى أخرى Sampling stability وليس له معنى للبيانات الكيفية حيث لا يمكن تقديره، وعليه فهذا مقياس غير صادق للتشتت، ومن عيوبه تأثره بالقيم المتطرفة.

الانحراف المعيارى Standard deviation (SD)

يعتبر الانحراف المعيارى أكثر مقاييس التشتت استخدامًا وأهمية حيث يعتبر بجانب المتوسط الملوك العظماء فى الإحصاء الوصفى والعمود الفقرى للإحصاء الاستدلالى البارامترى، والانحراف المعيارى يستخدم المتوسط للتوزيع كنقطة مرجعية يقيس الاختلاف أو التشتت اخذًا فى اعتباره المسافة بين كل درجة فى التوزيع والمتوسط. وكما سبق عرضه فى خصائص المتوسط أن مجموع الدرجات الانحرافية= صفر

$$\sum(X - \bar{X}) = 0$$

ويرمز للانحراف المعيارى لبيانات العينة بالرمز S أوSD ولبيانات المجتمع بالرمز σ

ويقدر الانحراف المعيارى فى المجتمع بالصيغة الآتية:

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

مثال: 3,4 ,5 ,6 ,7

1 . حساب متوسط درجات المجتمع μ:

$$\mu = \frac{3+4+5+6+7}{5} = \frac{25}{5} = 5$$

2 . حساب الدرجات الانحرافية (المسافة المعيارية بين الدرجة ومتوسط المجتمع):

X	$X - \mu$	$(X-\mu)^2$
3	3-5= -2	4
4	4- 5=-1	1
5	5-5 = 0	0
6	6-5 = 1	1
7	7-5 = 2	4

والدرجة الانحرافية تمدنا بمعلومات عن الإشارة (− ، +) والقيمة، فالإشارة تخبرنا عن الموقع من المتوسط سواء كانت فوقه (+) أو تحته (−) والقيمة تعبر عن المسافة الفعلية من المتوسط، فعلى سبيل المثال الدرجة 6 فإن اشارة الدرجة الانحرافية موجبة بالتالى تقع فوق المتوسط، وعليه فإنها تبتعد بمقدار نقطة أو وحدة عن المتوسط. ولأن الهدف هو حساب المسافة المعيارية من المتوسط فالخطوة التالية حساب متوسط الدرجات الانحرافية وعليه فإن متوسطها$(\frac{\Sigma(x-\bar{X})}{N})$ يساوى صفر.

وحيث إن:

$$\sum(X - \bar{X}) = -2 -1+5+1+ 2$$
$$= -3 + 3 = 0$$

وعلى ذلك فإن الدرجات الانحرافية متوسطها صفر وعلى ذلك فإنه لا قيمة لها كمؤشر للانتشار أو للاختلاف. وبما أن متوسط الدرجات الانحرافية لا يعمل جديًا كمؤشر

للتبـاين أو الاخـتلاف لأنـه دائمـا يسـاوى صـفر وهـذا نتيجـة للقـيم الموجبـة والسـالبة حيـث تلغـى أحـدهما الأخـرى ولـذلك فـإن الحـل هـو تربيـع كـل درجـة انحرافيـة كمـا هـو واضـح فـى العمود الثالث.

3. حساب الانحراف المعيارى من الصيغة:

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

$$= \sqrt{\frac{10}{5}} = 1.41$$

ومعادلـة حسـاب الانحـراف المعيـارى مـن العينـة S ويسـتخدم لتقـدير الانحـراف المعيـارى فـى المجتمـع ولكـن صـيغة الانحـراف المعيـارى للعينـة يـتم القسـمة علـى $n-1$ بـدلا مـن n لاعطاء تقدير دقيق وغير متحيز لـ σ وتصبح المعادلة:

$$S = \sqrt{\frac{(X-\bar{X})^2}{n-1}}$$

وهـذه الصـيغة المسـتخدمة فـى إجـراءات الإحصـاء الاسـتدلالى ولكـن إذا كـان الانحـراف المعيـارى لبيانـات العينـة للوصـف فقـط فالمقـام يكـون n لاحـظ اسـتخدمنا N للمجتمـع و n لبيانات المجتمع ويمكن أن تستخدم كلا من N , n بنفس المعنى.

مثال: لدرجات خمسة طلاب الآتية:

X	$X - \bar{X}$	$(X - \bar{X})^2$
2	2 – 6 = - 4	16
4	-2	4
6	0	0
8	2	4
10	4	16

فإن:

$$\bar{X} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$$

وبما أن مجموع انحرافات القيم عن متوسطها:

$$\sum(X - \bar{X}) = -4-2-0+2+4+0 = 0$$

ومجموع مربعات انحرافات القيم عن متوسطها:

$$= \sum (X - \bar{X})^2 = 16 + 4 + 0 + 4 + 16 = 40$$

إذا:

$$S = \sqrt{\frac{(X-\bar{X})^2}{n-1}} = \sqrt{\frac{40}{5-1}} = \sqrt{10} = 3.16$$

ويمكن حساب الانحراف المعيارى من الدرجات الخام بصيغة أخرى اعتمادًا على الدرجات الخام كالآتى:

$$S = \sqrt{\frac{SS}{n-1}}$$

- SS مجموع المربعات.

$$SS = \sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

- $\sum X^2$ مجموع مربعات X
- $(\sum X)^2$ مربع مجموع X

وفيما يلى المثال التالى:

X	X^2
10	100
12	144
13	159
15	225
18	324
20	400
22	484
25	625
∑X	$\sum X^2$=
135	2471

$$SS = \sum X^2 - \frac{(\sum x)^2}{n} = 2471 - \frac{(35)^2}{8}$$
$$= 2471 - 2278.125$$

$$= 192.875$$

$$S = \sqrt{\frac{ss}{n-1}} = \sqrt{\frac{192.875}{8-1}} = 5.25$$

خصائص الانحراف المعيارى

1. يمدنا بمقياس لانتشار للدرجات بالنسبة للمتوسط وهنا يختلف عن المدى حيث يعطى مقياس مطلق للانتشار بين قيمتين الأعلى والادنى. والانحراف المعيارى هو متوسط المسافة لانحرافات الدرجات عن المتوسط، والتوزيعات الاعتدالية تعتمد على المتوسط والانحراف المعيارى حيث إن 99.7% من البيانات تقع داخل ثلاث وحدات انحرافات معيارية 3SD(S) من المتوسط.

2. الانحراف المعيارى حساس لكل درجة فى التوزيع، فإذا كانت الدرجات قريبة من المتوسط فإن الانحراف المعيارى يكون صغيرًا والعكس إذا كانت الدرجات تبتعد بعيدًا عن المتوسط فإن الانحراف المعيارى يكون كبيرًا. والتوزيع ذو انحراف معيارى كبير يكشف عن مدى واسع من الفروق الفردية بين الأفراد.

3. الانحراف معيارى أكثر استقرارًا مقارنة بالمدى، حيث بحساب الانحراف المعيارى لعينات كثيرة فإن الانحراف المعيارى للعينات تختلف بدرجة قليلة من عينة إلى أخرى وهذا عكس المدى حيث توجد فروق كبيرة لمدى العينات المختلفة.

4. الانحراف المعيارى دائمًا يأخذ قيم موجبة $S(SD) > 0$ والقيم السالبة لا معنى لها.

والخصائص السابقة تجعل الانحراف المعيارى مقياس جيد للتعبير عن تشتيت الدرجات مقارنة بالمدى.

التباين Variance

التباين هو مربع الانحراف المعيارى للبيانات العينة(غير متحيز):

$$S^2 = \frac{\sum(X - \bar{X})^2}{n-1}$$

ويمكن كتابة المعادلة بالصيغة الآتية:

$$S^2 = \frac{\sum(X - \bar{X})^2}{n}$$

يستخدم التباين للوصف أو للإحصائيات الوصفية للعينة وايضًا لاستخدامه فى الإحصاء الاستدلالى.

وصيغة التباين فى المجتمع :

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

- μ متوسط درجات المجتمع

والتباين لا يستخدم بصورة كبيرة فى الإحصاء الوصفى لأنه يعبر عن الوحدات التربيعية للقياس ولكنه يستخدم بدرجة متزايدة فى الإحصاء الاستدلالى.

مثال: فيما يلى مجموعة من الدرجات:

X	X^2
4	16
6	36
5	25
11	121
7	49
9	81
7	49
3	9
$\sum X = 52$	$\sum X^2 = 386$

حساب مجموع المربعات (SS):

$$SS = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$= 386 - \frac{(52)^2}{8} = 386 - 338 = 48$$

$$=386-338=48$$

$$S^2 = \frac{SS}{n-1} = \frac{48}{8-1} = 6.86$$

والانحراف المعيارى:

$$S = \sqrt{S^2} = \sqrt{6.86} = 2.62$$

تذكر أن الصيغة السابقة للتباين (n-1) لبيانات العينة حيث يعطى تقدير دقيق غير متحيز لتباين المجتمع، واعلم أن التباين والانحراف المعيارى شديد الحساسية للقيم

المتطرفة ولذلك هذا من أهم الأسباب الرئيسية لأن الإحصاء البارامترى لا يحبذ التعامل مع التوزيعات الملتوية والمتطرفة.

الارباعيات Quartiles

الارباعيات Quartiles تقسم التوزيع إلى أربعة أجزاء متساوية، والمئينات **Percentiles** تقسم التوزيع إلى 100 جزء متساوى، يوجد أربعة إرباعيات هى الإرباعى الأول **Q1** وهو يقابل المئينى **P25** ، والإرباعى الثاني**Q2** وهو يقابل المئينى **P50**، والإرباعى الثالث **Q3** وهو يقابل المئينى **P75**، والإرباعى الرابع **Q4** وهو يقابل المئينى **P100**، وعلى ذلك فإن الإرباعيات تقسم التوزيع إلى أربعة أجزاء وكل جزء يتضمن 25% من البيانات ويطلق على الإرباعى الأول بالإرباعى الادنى**Lower quartile** والإرباعى الثانى بالوسيط والإرباعى الثالث يسمى الإرباعى الأعلى **Upper quartile**

ولتقسيم البيانات إلى أربعة أجزاء اتبع الآتى:

1. تحديد الوسيط لكل البيانات Q2.

2. تحديد الوسيط للدرجات تحت **Q2** فهذا الإرباعى الأدنى.

3. تحديد الوسيط للدرجات فوق **Q2** فهذا يكون الإرباعى الأعلى.

مثال: 98, 93, 96, 88, 92, 87, 85, 87, 84, 76, 80, 68, 74, 62

لحساب الإرباعيات اتبع الآتى:

1. ترتيب القيم تصاعدى:

62 68 74 76 80 84 85 87 87 88 92 93 96 98

2. حساب رتبتى الوسيط:

$$d1 = \frac{n}{2} = \frac{14}{2} = 7$$

$$d2 = \frac{n}{2} + 1 = \frac{14}{2} + 1 = 8$$

اذا القيمتين المقابلتين لرتبتى الوسيط هى 85 و 87 اذا:

$$Q_2 = \frac{85 + 87}{2} = 86$$

والقيمة 86 تقسم التوزيع إلى قسمين متساوين.

3. للقيم أقل من 86 يوجد سبعة قيم وهى 85 , 84 , 80 , 76 , 74 , 68 , 62

حساب رتبة الوسيط لهم (عدد القيم فردى):

$$= \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اذا كانت القيمة المقابلة للرتبة 4 هى 76 إذا قيمة الإرباعى الأدنى **Q1** = 76

وهى القيم التى يقع تحتها % 25 وفوقها % 75 من القيم .

4. للقيم فوق الوسيط **Q2** (86) وهى:

87, 87, 88,92,93,96, 98

حساب رتبة الوسيط :

$$= \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اذا القيمة التى ترتيبها 4 هى 92 إذا هى قيمة الإرباعى الأعلى وعليه **Q3** = 92

وهى القيمة التى يقع تحتها 75% من القيم وفوقها 25% من القيم.

نصف المدى الربيعيQ

هو نصف المسافة بين قيمة الإرباعى الأدنى **Q1** و قيمة الإرباعى الأعلى **Q3**

$$\frac{\textbf{Q3 - Q1}}{\textbf{2}}$$

ويطلق عليه الانحراف الربيعى Quartile Deviation والقيمة الصغيرة تشير إلى تشتت محدود للدرجات ولان **Q** يستبعد 25% من الدرجات الدنيا والعليا فى التوزيع ولذلك فإنه يستبعد القيم المتطرفة ان وجدت فى التوزيع وعلى ذلك فإن المدى الربيعى من مقاييس التشتت الجيدة.

وعليه نفس التوزيع أو المثال السابق فان:

$$= \frac{Q_3 - Q_1}{2} = \frac{92 - 76}{2} = \frac{16}{2} = 8$$

لاحظ أن نصف المدى الربيعى يعالج عيب المدى وهو أنه يعتمد على القيم المتطرفة الدنيا والعليا فى التوزيع عند حسابه ولذلك فهو مؤشر أكثر ثباتًا واستقرارًا من المدى. ودائما يوضع نصف المدى الربيعى بجانب الوسيط لاعطاء وصف كامل عن التوزيع

Q ± Md و يعطى قيمتين أو نقطتين يقع بينهما 50% من القيم إذا كان التوزيعات منتظمة واستخدام Q و Md فى الإحصاء الاستدلالى محدود.

المئينات Percentile (P%)

هى نقطة على المحور الأفقى السينى **X** تحدد نسبة مئوية معينة للدرجات تقع تحتها فمثلا P_R يشير إلى النسبة المئوية من الدرجات التى تقع تحت نقطة مئوية معينة R. فمثلا المئينى 50(P_{50}) وهى القيمة التى تقع تحتها 50% من القيم وهى قيمة الوسيط و P_{25} هى قيمة الإرباعى الأدنى و P_{75} الإرباعى الأعلى وP_{60} هى القيمة أكثر يقع تحتها 60% من القيم. ولحساب المئينات فيما يلى المثال الآتى لدرجات 70 طالبًا فى امتحان الإحصاء (Pango ,2013)

Class	F	Cum F	Cum %
95-99	4	70	70/70 × 100 = 100
90-94	6	66	49.29
85-89	7	60	85.71
80-89	10	53	75.71
(74.5 – 79.5) 75-79	16	54	61.43 →P_{50}
70-74	9	27	38.57
(64.5 – 69.5) 65-69	7	18	25.71→ P_{20}
60-64	4	11	15.781
55-59	4	7	10.00
50-54	2	3	4.29
45-49	1	1	1.43

يقدر المئينى من المعادلة الآتية:

$$P\% = X_L + i(\frac{N\left(P_R/100\right) - \Sigma f_b}{f_i})$$

- $P\%$ هى النقطة المئوية
- X_Lالحد الحقيقى الأدنى للفئة التى تقع فيها رتبة المئيني
- P_R رتبة المئينى.
- i طول الفئة
- f_bالتكرار التراكمى تحت الفئة التى تتضمن رتبة المئيني
- f_iالتكرار المقابل للفئة التى تتضمن رتبة المئيني
- N حجم العينة

لحساب المئينى 50 (P_{50})اتبع الخطوات الآتية:

1. حساب رتبة المئينى:

$=N\ (P_R\ /100) = 70\ (50/100) = 35$

اذًا 35 هى رتبة المئينى P_{50}.

2. حدد الحد الحقيقى الأدنى للفئة التى تقع فيها رتبة المئينى 50 (فى ضوء عمود Cum F) إذا فإن 35 تقع فى التكرار التراكمى 43 والفئة المقابلة لها 75-79 وحددها الحقيقى الأدنى (74.5)

3. طول الفئة: $i = 79.5 - 74.5 = 5$

4f_b.هى التكرار التراكمى تحت الفئة التى تتضمن رتبة المئيني=27

٥- f_iالتكرار المقابل للفئة التى تتضمن رتبة المئينى هى 16

$$P_{50} = 74.5 + 5\left(\frac{35 - 27}{16}\right)$$

$$= 74.5 + 2.5 = 77$$

اذًا القيمة 77 هى المئينى 50 بمعنى يقع تحتها 50% من القيم.

ولحساب المئيني20 كالآتى:

1. حساب رتبة المئينى 20 كالآتى:

$$P_{20} = 70(\frac{20}{100}) = 14$$

2. حدد الفئة التى تقع فيها الرتبة 14 من خلال التكرار التراكمى 69.5 – 64.5

وحدها الحقيقى الأدنى = 64.5

3. حدد طول الفئة: i = 69.5 – 64.5 = 5

4. حدد f_b وهى التكرار الذى يأتى بعد التكرار التراكمى القابل للفئة التى تقع فيها رتبة المئينى وهى 11

5. حدد f_i تكرار الفئة التى يقع فيها المئينى (عمود التكرار) وهي 7 وعليه فان:

$$P_{20} = 64.5 + 5\left(\frac{14 - 11}{7}\right)$$

$$= 64.5 + 2.14 = 66.64$$

وهى القيمة التى تقع تحتها 20% من القيم وفوقها 80 % من القيم.

معامل الاختلاف Coefficient of variation

احد الممارسات الشائعة فى استخدام الإحصاء هو مقارنة متوسطين أو مجموعتين فأكثر أو حتى متغيرين فأكثر والمقارنة بين المجموعات أو المتغيرات هى شرعية ومفيدة فى العمل الإحصائى.

مثال: طبق باحث اختبار للذاكرة طويلة الامد وأعطى بيانات لمجموعة ما حيث الانحراف المعيارى SD = 3.5 والمتوسط = 15 و طبق اختبار آخر للذاكرة طويلة الامد لنفس المجموعة وأعطى حيث $\bar{X}$ =75 و SD= 10.5, تساءل الباحث أى الاختبارين أفضل فى تقدير الذاكرة طويلة المدى وربما نستنتج أن الاختبار الثانى أفضل وذلك لأن الانحراف المعيارى له أكبر من الانحراف المعيارى للاختبار لأولK ولكن أفضل طريقة للمقارنة بين الانحرافات المعيارية على اختبارات مختلفة لها متوسطات مختلفة هو معامل الاختلاف ويقدر كالآتى:

$$CV = \frac{SD}{\bar{X}} \times 100$$

وحيث تم وضع 100 للتعبير عن النتائج فى ضوء نسبة مئوية.

وبتطبيق ذلك على المثال السابق للاختبار الأول:

$$CV = \frac{3.5}{15} \times 100 = 23.3$$

وعليه فإن الانحراف المعيارى تقريبًا23% من المتوسط.

وللاختبار الثانى:

$$CV = \frac{10.5}{75} \times 100 = 14$$

وفى هذه الحالة فإن معامل الاختلاف للاختبار الثانى أكبر من معامل الاختلاف للاختبار الأول، وعليه فإن الاستنتاج القائم على مقارنة القيم المطلقة للانحرافات المعيارية غير صحيح، حيث فى ضوء معامل الاختلاف اتضح أن الاختبار الأول أفضل من الثانى وعليه نختار الاختبار الأول عكس قرار الاعتماد على الانحراف المعيارى فقط .

مؤشر الاختلاف الكيفى The Index of Qualitative Variation (IQV)

مقاييس التشتت مثل المدى والانحراف المعيارى والتباين تقدر لبيانات فترية ونسبية ولكن ماذا لو كانت البيانات اسمية فى هذه الحالة يوجد مؤشر الاختلاف الكيفى ويقدر من المعادلة الآتية(Kirk, 2008):

$$\mathrm{I}_{QV} = \frac{\mathrm{K}(100^2 - \sum \mathrm{P}^2)}{100^2(\mathrm{K} - 1)}$$

- K عدد التصنيفات والمستويات للمتغير
- $\sum \mathrm{P}^2$ مجموع مربعات النسبة المئوية للتكرارات فى كل المجموعات.

مثال : فيما يلى توزيع الطلاب فى السنة الجامعية الأولى فى كلية التربية:

التصنيف	المسمى	عدد الطلاب	النسبة المئوية P%	P^2
1	انجليزى	218	68.1	4637.61
2	لغة عربية	55	17.2	295.84
3	رياضيات	25	7.8	60.84
4	كيمياء	10	3.1	9.61
5	تربية خاصة	12	3.8	14.44
				$\sum P^2$==5018.3

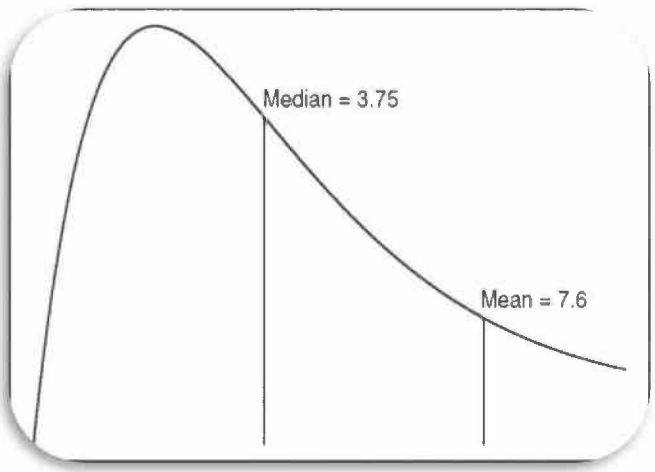
وعليه:

$$I_{QV} = \frac{5(100^2 - 5018.3}{100^2(5-1)} = \frac{24908.3}{40.000} = 0.62$$

وإذا كان I_{QV}= 0 إذا لا يوجد اختلاف، I_{QV}=1 يوجد اختلاف بدرجة تامة، وعليه يوجد تنوع بدرجة مرتفعة بين الشعب حيث تفسر قيمته تقريبًا مثل معامل الارتباط (r).

الالتواء والتفرطح Skweness and Kurtosis

لإعطاء وصف كامل للتوزيع فنحن نحتاج إلى مؤشرين هما الالتواء والتفرطح، فالالتواء يشير إلى عدم الانتظامية Asymmetry للتوزيع كالآتى:



والتفرطح يشير إلى التدبب Flatness أو القمة العالية Peakedness. وتوجد صيغ عديدة لتقدير الالتواء أهمها الصيغة الآتية:

$$SK = \frac{\frac{\sum(X-\bar{X})^3}{n}}{S^3}$$

حيث S الانحراف المعيارى ولو كان التوزيع منتظم Symmerrical فإن SK= 0 ولو كان ملتوى التواء موجب SK > 0 ولو كان ملتوى التواء سالب SK < 0

مثال: Kirk (2008)

X	$X-\bar{X}$	$(X-\bar{X})^2$	$(X-\bar{X})^3$	$(X-\bar{X})^4$
6	2	4	8	16
5	1	1	1	1
5	1	1	1	1
5	1	1	1	1
5	1	1	1	1
4	0	0	0	0
3	-1	1	-1	1
2	-2	4	-8	16
1	-3	9	-27	81
$\sum X= 36$	$\sum(X-\bar{X})=0$	$\sum(X-\bar{X})^2=22$	$\sum(X-\bar{X})^3=24$	$\sum(X-\bar{X})^4= 18$

$$\bar{X} = \frac{\sum X}{n} = \frac{36}{9} = 4$$

اذًا الانحراف المعيارى:

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{22}{9}} = 1.563$$

$$SK = \frac{\frac{-24}{9}}{(1.563)^3} = \frac{-2.667}{3.818} = -0.70$$

وعلى ذلك فإن SK = - 0.7 وهذا يشير إلى وجود التواء سالب ولكن بدرجة خفيفة وعليه فإن المنحنى ليس اعتدالى بدرجة تامة ولا يقدر الالتواء بصورة كبيرة باستخدام معادلات رياضية انما يتم الاعتماد على شكل التوزيع التكرارى أو العرض البيانى لتحديد طبيعة البيانات.

يقدر التفرطح من الصيغة الآتية:

$$Kur = \frac{\frac{\sum(X-\bar{X})^4}{n}}{S^4} - 3$$

ويقدر للبيانات السابقة كالآتى :

$$Kur = \frac{\frac{118}{9}}{(1.563)^4} - 3 = \frac{13.111}{5.968} - 3 = -0.80$$

وعليه يوجد تفرطح بدرجة قليلة فى المنحنى حيث إذا كان Kur= 0 فإن المنحنى اعتدالى التوزيع بدرجة تامة.

تنفيذ إحصائيات التشتت والالتواء والتفرطح فى SPSS

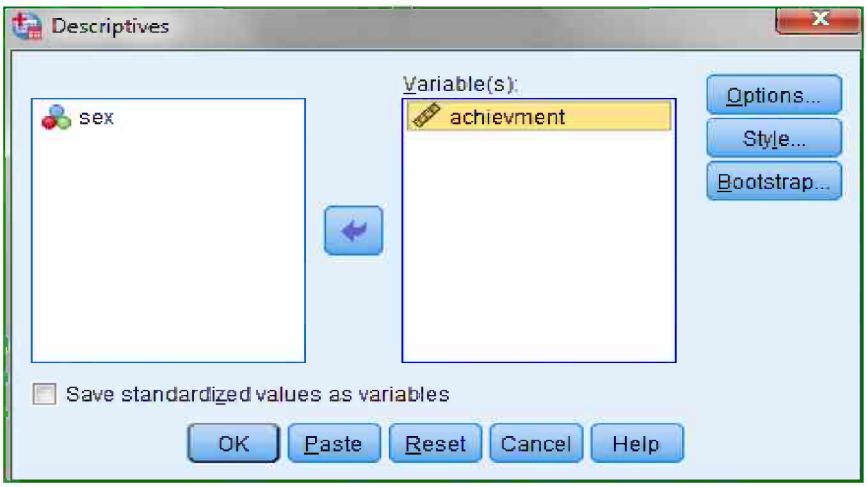
وذلك لتحصيل 70 طالب فى مقرر الإحصاء

أولًا: إدخال البيانات بعد فتح الشاشة الرئيسية

1. اضغط علىVariable view أسفل الشاشة على اليمين
2. اكتب اسم المتغير فى عمود Name: achievement
3. اضغط على أيقونة Dataview، قم بإدخال البيانات فى عمود واحد

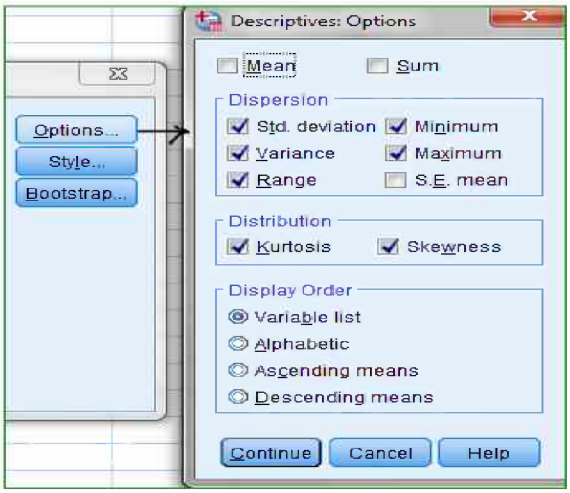
ثانيًا: تنفيذ الأوامر: يمكن تنفيذ أمر حساب مؤشرات التشتت والتفرطح والالتواء من عدة طرق أهمها:

1. اضغط علىAnalyze ثم Statistics ثم Descriptive ثم اضغط على Descriptive تظهر الشاشة الآتية:



2. انقل متغير achievement↳ إلى مربع Variable

3. اضغط على اختيار Options على اليمين تظهر الشاشة الآتية:



4. اضغط على مقاييس التشتت فى مربع Dispersion

Std .deviation: و Variance و Range

5. اضغط على مؤشرى التفرطح والالتواء فى مربع Distribution

6. اضغط على Continue ثم OK

ثالثًا: المخرج:

```
DESCRIPTIVES VARIABLES=achievment
  /STATISTICS=STDDEV VARIANCE RANGE MIN MAX KURTOSIS SKEWNESS.
```

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
achievment	70	53.00	46.00	99.00	11.83237	140.005	-.300-	.287	-.289-	.566
Valid N (listwise)	70									

كما هو واضح أعطى الإحصائيات:

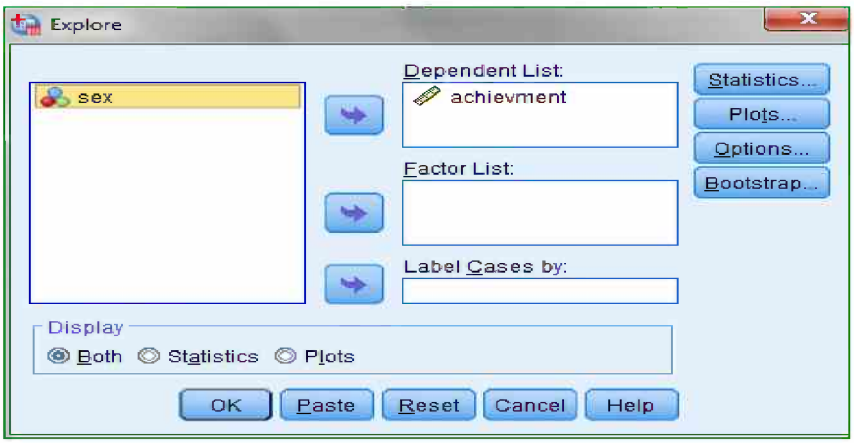
- المدى وهو الفرق بين أكبر قيمة واصغر قيمة (99-46)

-الانحراف المعيارى والتباين

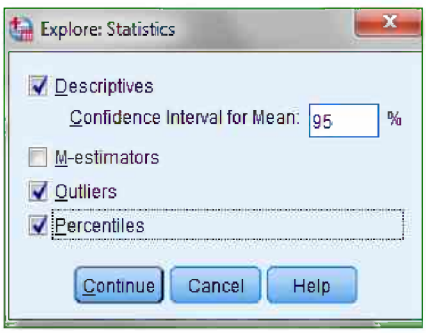
-وقيمة Kurtosis = 0.289 ,Skewness= - 0.300 أى يوجد درجة ضعيفة من الالتواء السالب ولكن فى المجمل المنحنى اعتدالى إلى حدٍّ ما كما يوجد درجة قليلة من التفرطح فى وسط المنحنى.

ولكن هذا الأمر لم يعطينا نصف المدى الربيعى. **ويمكن تنفيذ هذه المؤشرات بجانب المدى الربيعى من خلال الآتى:**

1. اضـغط علـى Analyze ثـم Statistics Descriptive ثـم اضـغط علـى Explore يعطى الشاشة الآتية:



2. انقـــل متغيـــر achievement إلـــى مربـــع Dependent list



3. اضـغط علـى اختيـار Statistics تظهـر الشاشـة الآتية:

4. اضـــغط علـــى اختيـــار Descriptive و Quitlers و Percentiles

5. اضغط على Continue ثم OK

ثالثا: المخرج:

```
EXAMINE VARIABLES=achievment
  /PLOT BOXPLOT STEMLEAF
  /COMPARE GROUPS
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Descriptives

			Statistic	Std. Error
achievment	Mean		76.3714	1.41424
	95% Confidence Interval for Mean	Lower Bound	73.5501	
		Upper Bound	79.1928	
	5% Trimmed Mean		76.6270	
	Median		77.0000	
	Variance		140.005	
	Std. Deviation		11.83237	
	Minimum		46.00	
	Maximum		99.00	
	Range		53.00	
	Interquartile Range		15.75	
	Skewness		-.300-	.287
	Kurtosis		-.289-	.566

حيث المتوسط Mean= 76.3714 والمتوسط المهذبTrimmed Mean مع استبعاد 5% من الدرجات الدنيا وكذلك القصوى= 76.62 . ونلاحظ أنه لم يتغير كثيرًا، والوسيط Median = 77 لأنه لا يوجد قيم متطرفة.

والتبـــاين Variance = 140.005، الانحـــراف المعيـــارى Std. Deviation = 11.83، المـدى Range = 53، نصـف المـدى الربيعـىInterquartile Range= 15.75. لاحـظ أن قـيم الانحـراف المعيـارى ونصـف المـدى الربيعـى متقاربـة إلـى حـد مـا ولكـن المـدى قيمتـه كبيـرة جـدًا 53 لأنـه يعتمـد علـى القـيم المتطرفـة، والالتـواء = 0.300- ، التفرطح = 0.284-

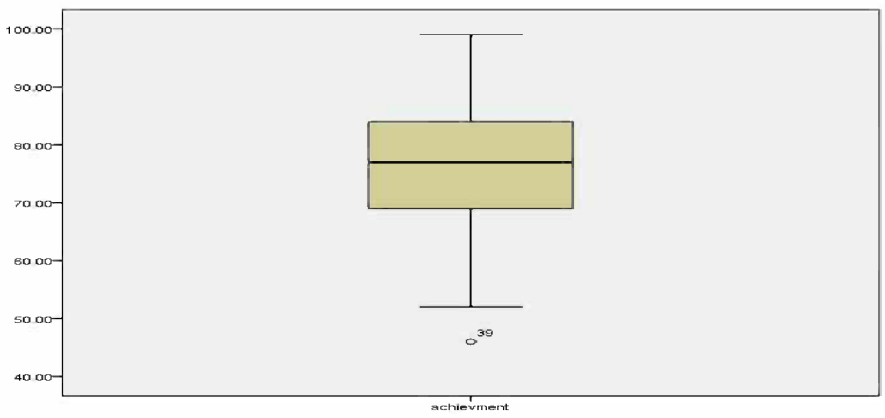
- **أعطى البرنامج المئينات كالآتى:**

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	achievment	55.1000	58.2000	68.7500	77.0000	84.5000	93.0000	95.4500
Tukey's Hinges	achievment			69.0000	77.0000	84.0000		

حيث P5 = 55.100 وهى القيمة التى تقع تحتها 5% من القيم، المئينى الأول (الإرباعى الأدنى) P25= 68.75 ، المئينى الثاني(الوسيط) P50=77، المئينى الثالث (الإرباعى الأعلى) P75= 84.00 هى القيمة التى يقع تحتها 75% من قيم التوزيع، لاحظ أن البرنامج حسب بعض المئينات وليس كلها.

وتشخيص وتحديد القيم المتطرفة من خلال شكل Boxplot كالآتى :



حيث ظهر أن الحالة 39 المناظرة للدرجة 46 هى قيمة متطرفة كما هو واضح من العلامة (0).

المنحنى الاعتدالى والدرجات المعيارية The normal Curve and Standardized Scores

افترض أن معلمك فى الفصل أخبرك بأنك حصلت على درجة ذكاء 132 فهل ستكون سعيدًا أم حزينًا؟ ولكن هذه الدرجة صعب تفسيرها وتكون عديمة المعنى ما لم توجد مجموعة معيارية Reference group تقارن فى ضوئها وبدون هذه المجموعة لا تستطيع تحديد ما إذا كانت هذه الدرجة منخفضة أو متوسطة أو مرتفعة. ولتوضيح مدى إعطاء معنى لهذه الدرجة افترض أنه يوجد توزيع يتضمن1000 درجة و توجد به الدرجة 132، فيمكن تحديد كم نسبة مئوية من الدرجات (%) أقل أو أكبر من الدرجة 132 كما سبق توضيحه من الرتبة الميئينة للدرجة.

وأن أى توزيع أو مجموعة من الدرجات يتم وصفها فى ضوء المتوسط والانحراف المعيارى حول المتوسط وبذلك سوف يتم عرض الدرجات المعيارية لوصف أى درجة فى ضوء مدى ابتعادها أعلى أو أسفل المتوسط.

الدرجة المعيارية Z Score

قبل الحديث عن الدرجة المعيارية نتحدث عن المعيارية **Standardization**

وهى تحويل الدرجات المفردة الخام إلى درجات محورة أو معيارية أو معدلة أى تحويل الدرجات من التوزيعات الاعتدالية المختلفة إلى توزيع عام مشترك اعتدالى بمعلومية المتوسط والانحراف المعيارى والمئينيات.

واحد المشاكل لعمل مقارنات لها معنى هو قياس متغيرات مختلفة على وحدات قياس مختلفة، فعلى سبيل المثال يقاس الطول بالسنتمتر بينما الوزن بالجرام وفى سبيل مقارنة الطول بالوزن فنحن نحتاج إلى أن نضع المتغيرات المختلفة على نفس المقياس المعيارى ولذلك يمكن معيارية لمتغيرات مختلفة عن طريق استخدام المتوسطات والانحرافات المعيارية لتحويل أى درجة خام إلى درجة معيارية (Z).

وعليه فالدرجة المعيارية Z هى تحدد عدد الانحرافات المعيارية لدرجة ما من المتوسط، أو هى درجة معدلة وفيها يتم تحديد كم وحدة انحراف معيارى تقع الدرجة الخام أعلى

وأسفل المتوسط. وعلى ذلك فإن الدرجة المعيارية تستخدم للتعبير عن الدرجة الخام فى ضوء المتوسط والانحراف المعيارى.

وإذا وقعت الدرجة المعيارية فوق المتوسط فإنها تكون موجبة، وإذا وقعت تحت المتوسط تكون سالبة. وتقدر الدرجة المعيارية فى المجتمع:

$$Z = \frac{X - \bar{X}}{\sigma}$$

وتقدر لبيانات العينة:

$$Z = \frac{X - \bar{X}}{S}$$

ولحساب الدرجة Z

1. حساب متوسط الدرجات
2. حساب المسافة بين كل درجة والمتوسط($X - \bar{X}$) او ($X - M$)
3. حساب الانحراف المعيارى
4. حساب الدرجة المعيارية لكل درجة من المعادلة السابقة
5. التعبير عن المسافة فى ضوء الانحراف المعيارى عن طريق قسمتها عليه.

مثال: 1 , 4 , 5 , 7 , 8

أ . تحديد متوسط الدرجات الخام :

$$\bar{X} = \frac{\sum X_i}{n} = \frac{25}{5} = 5$$

ب. حساب الانحراف المعيارى:

X	X^2
1	1
4	16
5	25
7	49
8	64
$\sum X = 25$	$\sum X^2 = 155$

وبما ان: $$SS = 155 - \frac{(25)^2}{2} = 30$$

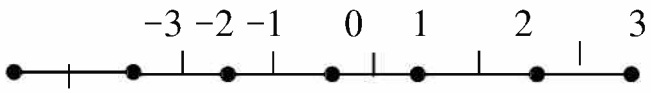
اذاً : $$S = \sqrt{\frac{30}{4}} = 2.7386$$

وعليه:

X	Z
1	$Z = \frac{X - \bar{X}}{\sigma} = \frac{1-5}{2.7366} = -1.46$
4	-0.37
5	0.00
7	0.73
8	1.10

لا تنسى إشارة Z حيث تغيرها من السالب إلى الموجب يحدث فروق كثيرة.

الدرجة Z كمقياس يمكن تمثيل الدرجة المعيارية على مقياس كالآتى:



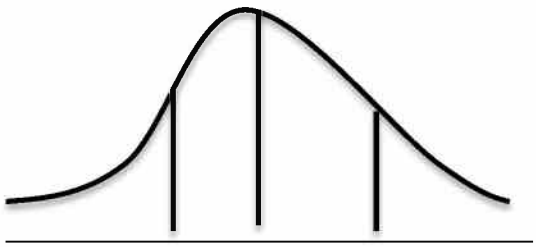
Z 1 4 5 7 8

وعلى ذلك الدرجة 8 تبتعد 1.10 وحدة معيارية أعلى المتوسط بينما الدرجة 4 تبتعد 0.37 وحدة معيارية تحت المتوسط.

ويوجد ملمحين أساسين للدرجة المعيارية Z:

الأول: أن توزيع Z دائما متوسطة – صفر وإذا كانت الدرجة فى التوزيع هى المتوسط فإنها تناظر الدرجة المعيارية صفر بمعنى ابتعادها هو صفر وحدة انحراف معيارى.

الثانى : توزيع Z دائما له انحراف معيارى مساويا للواحد الصحيح.



Z = −1 Z= 0 Z = 1

حيث Z = 1 وحدة انحراف معيارى فوق المتوسط، Z = -1 وحدة انحراف معيارى تحت المتوسط. فإذا كان متوسط درجات امتحان الإحصاء = 70، والانحراف المعيارى = 10 ، فالدرجة 80 تبتعد عشر نقاط أو وحدة واحدة انحراف معيارى فوق المتوسط

ولذلك فإن $Z_{80} = 1$ ، والدرجة 50 فإنها تبعد عشرين نقطة عن المتوسط أو وحدتين انحراف معيارى تحت المتوسط.

تحويل الدرجات Z إلى الدرجات الخام

لتحويل الدرجة المعيارية إلى درجة خام تكون الصيغة كالآتى:

$$X - \bar{X} = Z\,S$$
$$X = ZS + \bar{X}$$

فإذا كانت **Z =1** والانحراف المعيارى = 5 والمتوسط = 8 فان:

$$X = 1 \times 5 + 8 = 13$$

وفى ضوء ما سبق فان:

1. متوسط أى توزيع لـ Z هو صفر وعليه فإنه فى أى توزيع مجموع درجات Z الموجبة مساوية لمجموع درجات Z السالبة
2. الانحراف المعيارى لتوزيع Z دائمًا =1
3. لا يتغير شكل توزيع الدرجات الخام إذا عدلت إلى الدرجة المعيارية Z.

والدرجة المعيارية Z يطلق عليها الدرجات المعيارية لأنها لا تمتلك قيم معيارية للمتوسط والانحراف المعيارى ولذلك تعتبر الدرجة Z نوع من المقياس أو القياس المعيارى لأى متغير وعلى ذلك يستخدم للمقارنة بين درجات اختبارين لهما وحدات مختلفة ولذلك فهى تستخدم للمقارنة بين التفاح والبرتقال كان تقارن بين أداء طالب فى مقررين مختلفين، ولذلك نستطيع اخذ التفاحة من التوزيع الاعتدالى للتفاح وتحويلها إلىZ باستخدام المتوسط والانحراف المعيارى للتفاح ثم تحول Z إلى المئينى لتقول أن 85 % من كل التفاح المتبقى يقع فوق هذه التفاحة.

استخدام Z لعمل مقارنات

تخيل انك حضرت مقرر فى الإحصاء وزميلك حضر هذا المقرر مع أستاذ آخر أنت حصلت على درجة 92 و متوسط فصلك = 78.1 والانحراف المعيارى 12.2 وزميلك حصل على درجة 8.1 ومتوسط فصله = 6.8 والانحراف المعيارى = 0.74 (اختلاف استراتيجيات وضع الدرجات)

وعليه فإن الدرجة المعيارية لدرجتك:

$$Z_{92} = \frac{92 - 78.1}{12.2} = 1.14$$

والدرجة المعيارية لزميلك :

$$Z_{8.1} = \frac{8.1 - 6.58}{0.74} = 1.76$$

ونلاحظ أن درجتك ودرجة زميلك تقع فوق المتوسط ولذلك فإن الدرجة Z موجبة لك ولزميلك ولكن أداء زميلك أفضل من أدائك لأن الدرجة المعيارية له أعلى من الدرجة المعيارية لك. وعلى ذلك فإن Z تسمح بإجراء مقارنات بين التلاميذ الذين يتلقون امتحانات مختلفة فى مقررات مختلفة.

للدرجة Z فوائد اهمها:

1. الدرجة Z تعطى إحساس أو فهم عام لموقع الدرجة فى علاقتها بمتوسط الدرجات و الانحراف المعيارى فى التوزيع.
2. تسمح الدرجة المعيارية Z بمقارنة الدرجات من توزيعات مختلفة.
3. يمكن تحويل الدرجات Z إلى المئينات Percentiles .

ولا نشكل توزيع Z اعتدال ى فلا بد أن نتعرض للتوزيع الاعتدالى.

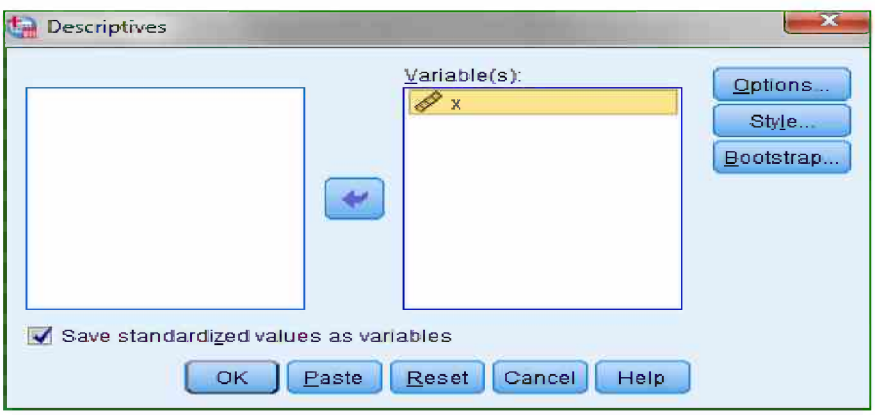
تنفيذ إحصاءZ فى برنامج SPSS (للمثال السابق)

أولًا: إدخال البيانات: 1. اضغط علىVariable view فى أسفل الشاشة على.

2. اكتب مسمى المتغيرx فى عمود بـ Name

3. اضغط علىData view

ثانيًا: تنفيذ الامر: 1. اضغط على analyze ثم اختار Descriptive statistics ثم اضغط على Descriptive تظهر الشاشة الآتية:

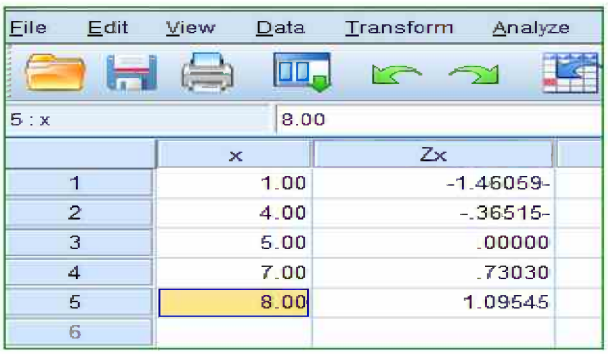


2. انقل المتغير X إلى مربع Variables (اضغط على السهم ↳)

3. اضغط على اختيار Save Standardized values as variables

4. اضغط على OK

ثالثـا: تفسـير المخـرج: يعطـى الـدرجات المعياريـة ملـف البيانـات حيـث يقـوم البرنـامج بإنشـاء عمـود تحـت مسـمى ZX لكـل درجـة خـام وبـالرجوع إلـى ملـف البيانـات مـن خـلال الضغط على قائمة Window ثم اضغط على الملف المتضمن البيانات .



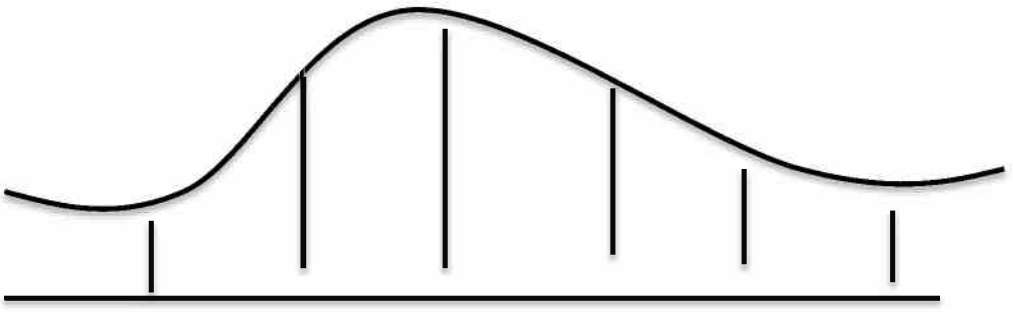
	x	Zx
1	1.00	-1.46059-
2	4.00	-.36515-
3	5.00	.00000
4	7.00	.73030
5	8.00	1.09545
6		

التوزيع الاعتدالي

هـو توزيـع نظـرى لـدرجات للمجتمـع النسـبة ويشـبه الجـرس ومنـتظم Symmetric وأحـادى المنـوال Unimodal ولـذلك فمـن السـهل معرفـة بعـض الشـيء مـن النسـبة المئويـة لأى منطقة تحت المنحنى.

ولإنشـاء المنحنـى عليـك تـذكر مـا عرضـناه فـى التوزيعـات التكراريـة وتمثيلهـا بالمـدرج التكـرارى ولـذلك فـنحن نسـتطيع تحديـد النسـبة المئويـة للمنحنـى الاعتـدالى التـى تقـع تحـت أو فـوق أى خـط رأسـى فـى هـذا المنحنـى. والإحصـائيون حـددوا نسـب معينـة تقـع داخـل أى منطقـة محـددة فـى المنحنـى الاعتـدالى وعليـك ان تضـع فـى اعتبـارك ان المنحنـى الاعتـدالى هـو منـتظم بمعنـى 50% مـن القيم تقـع تحـت المتوسـط و % 50 مـن القيم تقع فوق المتوسط واحد جانبى المنحنى هو مرآة تمامًا للأخرى للجانب الآخر.

والشكل الآتـى يوضـح المنحنـى الاعتـدالى بالنسـب التـى تقـع داخلـه (Aron et al., 2013)



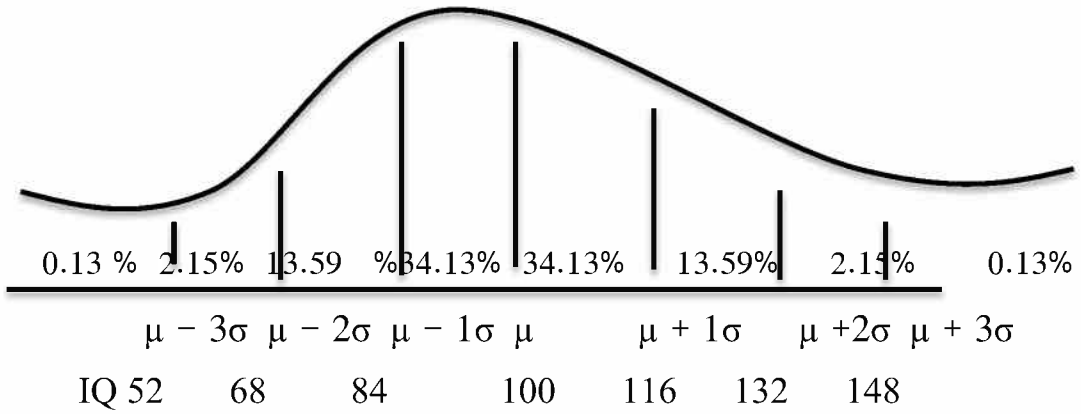
الشكل (1.7): المنحنى الاعتدالى مع النسب المئوية التقريبية للدرجات بين المتوسط ووحدات الانحراف المعيارى تحت وفوق المتوسط

ففى الشكل السابق تم تقسيم المنحنى الاعتدالى إلى نصفين متساوين تقريبا 34% من الدرجات تقع بين المتوسط (0) والدرجة المعيارية (1−) و14% من الدرجات تقع بين الدرجة المعيارية 1− والدرجة المعيارية 2− و 2% من الدرجات التى تقع بين الدرجة المعيارية 2− والدرجة المعيارية 3−. وحوالى 68% من الدرجات تقع داخل وحدة انحراف معيارى 1− و 1+ و % 96 داخل نطاق الدرجة المعيارية 2− إلى 2+ وتقريبا 100% من الدرجات 34) (34 + 34 + 14 + 14) وهذه تعتبر 3− إلى 3+ تقع فى نطاق الدرجة المعيارية (2+2+14+14+34+ إرشادات لتحديد النسبة المئوية المرتبطة بـ Z.

فعلى سبيل المثال متغير الذكاء IQ وفان متوسط IQ = 100 والانحراف المعيارى 15 وتوزيع درجات IQ اعتدالى وعليه فإن 34% من الدرجات أو الأشخاص لديهم ذكاء بين 100 المتوسط و115 فوق المتوسط و 34% من الدرجات أو الأشخاص ذكائهم يقع بين الدرجة 100 و 85 (درجة تعادل انحراف معيارى تحت المتوسط) وعلى ذلك فإن 68 من الأشخاص ذكائهم يقع بين 85 و 115 و 14% من الأشخاص ذكائهم من 115 وحدة انحراف معيارى من المتوسط و 130 (وحدتين) فوق المتوسط و 14% من الأشخاص يقع ذكائهم من 85 إلى 70 تحت المتوسط .

ولذلك % 50 و 34% و 14% هى قواعد عملية هامة للعمل مع مجموعة من الدرجات التى تتبع التوزيع الاعتدالى.

وعرض Pagano(2013) النسب تحت المنحنى بالآتى:



وذلك لمجتمع درجات الذكاء لـ 100.000 درجة حيث:

$\mu = 100$,$\sigma = 16$

و % 34.13 من الدرجات تقع بين الدرجات 100 و 116

$\mu + 1\sigma = 100 + 16 = 116$

و % 34.59 من الدرجات تقع بين الدرجات 116 و 132

$\mu + 2\sigma = 100 + 32 = 132$

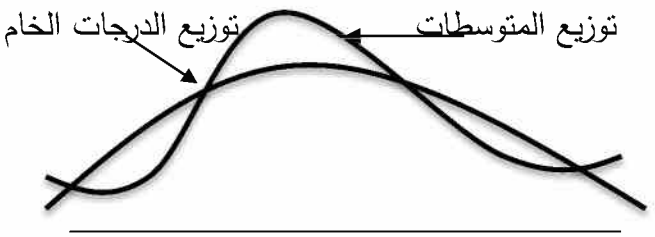
و % 2.26 من الدرجات تقع بين الدرجات 132 و 148

$\mu + 3\sigma = 100 + 48 = 148$

و % 0.13 من الدرجات التى تقع فوق 148

توزيع المتوسطات Distribution of means

هو التوزيع أو المنحنى التكرارى لعدد كبير من متوسطات عينات مسحوبة من المجتمع ويتم عرضها بيانيًا للحصول على المنحنى الذى يمثلها ويعتبر توزيع المتوسطات أكثر ثباتًا واستقرارا من الدرجات حيث له انحراف معيارى أقل من الانحراف المعيارى للدرجات الخام.



ولاحظ أن توزيع المتوسطات ضيقة وليس لها انتشار عالٍ كما فى حالة توزيع الدرجات الخام وان الخطأ المعيارى هو مسمى الانحراف المعيارى لتوزيعات المتوسطات وهو الانحراف المعيارى للمجتمع مقسوما على الجذر التربيعى لحجم العينة:

$$\sigma_{\mu(\bar{X})} = \frac{\sigma}{\sqrt{N}}$$

واذا كان الانحراف المعيارى لتوزيع مجموعة من الدرجات 5 لعينة 10 أفراد فإن الخطأ المعيارى: $= \frac{10}{\sqrt{5}} = 1.58$

ولاحظ الفرق ان الانحراف المعيارى للدرجات = 5 بينما للمتوسطات 1.58 ولذلك فإن تشتت المتوسطات أصغر من تشتت الدرجات الخام ولذلك يتم تقدير Z للمتوسطات كالآتى:

$$Z = \frac{X - \bar{X}}{\sigma_{\bar{X}}}$$

حيث:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

وهذه هى الصيغة التى تستخدم فى الفصول السابقة عند تناول اختبارات Z، T.

الفصل التاسع

اختبارات الفروض ومداخل صناعة القرار الإحصائي

Hypothesis testing & Statistical Descion Making Approaches

تسمح الإحصاء الاستدلالى للباحثين بجمع البيانات من العينة للوصول إلى استنتاجات أو نتائج أو معلومات عن السلوك فى المجتمع الذى بدوره يكون كبيرًا بدرجة لا نستطيع دراسة كل أفراده، فمثلًا متوسط العينة يكون تقدير غير متحيز للمجتمع إذا تم اختيار العينة بصورة عشوائية من المجتمع ويقال عندئذ أن قيمة المتوسط فى العينة $\overline{X}$ تعكس تقريبًا قيمة المتوسط فى المجتمع (μ)، ولذلك عندما يستخدم الباحث الإحصاء الاستدلالى عليه أن الاجابة عن أربعة أسئلة أساسية هى:

- ما هو المجتمع المراد دراسته؟
- كيف تم اشتقاق أو سحب العينة من المجتمع؟
- ما هى خصائص العينة التى تطبق عليها إجراءات وادوات القياس؟
- كيف يمكن تلخيص بيانات العينة للحصول على إحصائية معينة تستخدم لعمل استدلالات تخص معالم Parametrs المجتمع؟

ولإجراء ذلك لا بد من المرور بخطوات لمساعدة الباحثين على استخدام إحصائيات العينة للوصول إلى معالم المجتمع غير المعروفة وهذا الإجراء يسمى اختبارات الفروض، اختبار الفرض يقترح ما إذا كانت العلاقة موجودة (حقيقة) أم لا، وما إذا كان الفرق بين مجموعتين من المحتمل أن يكون للصدفة أم لا ؟

منطق اختبارات الفروض

من غير المنطقى أن يقوم باحث بملاحظة كل فرد فى المجتمع ولذلك يقوم بجمع بيانات العينة واستخدامها للإجابة عن الأسئلة البحثية التى تدور حول الحالة فى المجتمع ولذلك فإن اختبارات الفروض هى إجراء إحصائى يسمح للباحثين باستخدام بيانات العينات للوصول إلى استدلالات أو استنتاجات عن المجتمع موضع الاهتمام (Gravetter & Wallnau, 2014) . أو بكلمات أخرى هى الإجراء

الإحصائى للوصول إلى صناعة قرار لمشكلة ما يتناولها الباحث، أو هى الطريقة المستخدمة لتقيم بيانات العينة لنعلم أكثر عن خصائص المجتمع. ويعرفها Privitera (2015) بأنها الطريقة المستخدمة لاختبار ادعاء أو فرض عن المعالم فى المجتمع باستخدام بيانات العينة.

والفرض هو تفسير مفترض أو تخمين ذكى أو توقع لظاهرة ما ويختبر باستخدام إجراءات الطريقة العلمية وغالبًا يتم صياغته عن الحالة فى المجتمع، وعلى الرغم ان تفاصيل اختبار الفرض تتغير من موقف بحثى إلى آخر (الفروق غير العلاقات) ولكن عملية اختبار الفروض بخطواتها تبقى ثابتة نسبيًا لكل الاختبارات.

ويرى Howell (2013, 2014) أن اختبارات الفروض يجب أن توضح لنا حدود الثقة Confidence limits وكذلك لا بد أن تكون حول قوة الاختبار Power ويرى أن الية العمل الإحصائى هى:

- يصيغ الباحث فرض فى ضوء دراسات سابقة أو استدلال منطقى.
- يضع الباحث خصائص وشروط عن العينة (عمرها، نوعها).
- يحسب متوسط الدرجات أو أى إحصائية أخرى للمتغير التابع.
- يحسب قيمة الاختبار.
- يستخدم قيمة P value للوصول لقرار.
- إذا كانت النتائج دالة إحصائيًا يفترض الباحث أن فرضه صحيح ثم يصدر توصيات لنتائج بحثه وإذا لم تكن النتائج دالة إحصائيًا يحاول تفسيرها وربما يطالب بإجراء دراسات أخرى.

واختبار الفرض المستخدم فى متن الدراسة البحثية ويعتمد على نوع الدراسة ونوعية البيانات، وتفاصيل اختبار الفرض تتغير من موقف بحثى إلى آخر.

المدخل العام لاختبارات الفروض الصفرية

يوجد جدل كبير حول الخطوات المراد إتباعها للتحقق من الفروض خلال اختبارات الفروض فيراها البعض أربعة خطوات (Gravetter & Wallnau, 2014 ; Hinkle et al., 1994; Privitera, 2015) بينما يفترضها البعض ستة خطوات (Howell,

2013; Huck, 2012; Nolan & Heinzen, 2012) ، وسواء أربعة خطوات أو ستة فيوجد اختلاف بينهما فى ماهية هذه الخطوات، فمثلًا Howell (2013) يضع الفرض البحثى اولى الخطوات ثم الفروض الإحصائية الخطوة الثانية، بينما يضع Nolan & Heinzen (2012) تحديد المجتمع ومسلمات الاختبار أولى الخطوات ثم تليها الفروض الإحصائية، فى حين يبدأ Huck (2012) بالفرض الصفرى ثم الفرض البديل فى الخطوة الثانية.

الخطوة الأولى: الفروض الإحصائية: عملية اختبارات الفروض تبدأ بصياغة الفروض عن معالم المجتمع غير المعروفة. وفى حديثنا عن الفروض يوجد نوعان:

1. **الفروض البحثية Research hypotheses:** وهى توقع لما سوف تكون عليه الحالة فى المجتمع أو تأكيد أو صياغة عن خصائص معينة فى المجتمع مثل: للبرنامج فعالية فى خفض الاكتئاب أو درجات الذكور أفضل من درجات الإناث.
2. **الفروض الإحصائية Statistical hypotheses:** هى الترجمة الإحصائية لفرض البحث بحيث يمكن اختبار صحته أو عدم صحته إحصائيًا باستخدام تصميم إحصائى أمثل.

وإذا كانت الخصائص المراد دراستها فى المجتمع مقاسة كميًا بواسطة معالم مثل المتوسط أو التباين فإن الفرض يكون قائم على معالم ويطلق عليه فرض معلمي Parametric hypothesis، أما إذا كانت هذه الخصائص غير مكممة أو كيفية عندئذ فالفرض غير معلمى Non- Parametric Hypothesis وهذه الفروض المعلمية وغير المعلمية تعرف بالفرض الإحصائى.

وتصاغ الفروض الإحصائية فى صورة فرضين أحدهما صفرى أو العدمى (H0) Null Hypothesis والآخر بديل (H_A) Alternative Hypothesis وكليهما مكملين لبعضهما البعض. وفى الحقيقة الفرض الصفرى هو الوسيلة لاختبار فرض البحث إحصائيًا ولذلك فإن رفض الفرض الصفرى يسمح للباحثين بقبول فرض البحث أو الفرض البديل، وعليه فإن الإحصاء الاستدلالى قائم على اختبار الفروض الصفرية

ولذلك يطلق عليها اختبارات الفروض الصفرية Null Hypotheses Testing (NHT)، من ثم فعلى الباحث فى صياغة فرض البحث ان لا يكون صفريًا لأنه فى كل الأحوال سواء صاغه أو لم يصيغه فإنه يتحقق منه إحصائيًا لا محالة. ولذلك فإن بعض الباحثين يطلقوا على الفرض البديل فرض البحث مثل Nolan & Heinzen (2012).

- **الفرض الصفرى:** هو فرض العدم لا فروق، لا علاقة، لا تنبؤ، ولا أثر للتجربة على المتغير التابع، ويأخذ الفرض الصفرى الرمز H0 حيث H تشير إلى Hypothesis بينما (0) Zero تشير إلى فرض التأثير صفر Zero effect Hypothesis ، ويفترض أن يكون الفرض الصفرى حقيقى. ويتم صياغة الفرض الصفرى لرفضه، وإذا فشل الباحث فى رفض H0 فى ضوء بيانات العينة فيقال إن الباحث فشل فى رفض الفرض الصفرى وعليه يتم قبول الفرض البديل أو فرض البحث. وفى العلوم النفسية يصاغ الفرض الصفرى بالمصطلحات الآتية: لا توجد علاقة بين المتغيرات، لا توجد فروق بين المجموعات.

ولتوضيح الفرض الصفرى افترض أن باحث أجرى دراسة على مجتمع طلاب الجامعة وكان اهتمامه دراسة نسبة الذكاء لديهم وكان الاهتمام الإحصائى هو متوسط درجات الذكاء وعليه فإن الفرض الصفرى H0: μ=100، حيث يوضع الرمز H0 أولًا متبوعًا بـ : ثم رمز المعلم (μ) ثم القيمة التى افترضها الباحث المراد اختبار الفروض حولها.

وإذا كان اهتمام الدراسة اختبار شئ آخر غير المتوسط، مثل معامل الارتباط فى المجتمع (ρ) فإن الفرض الصفرى H0:ρ=0 أى لا يوجد علاقة أو ارتباط بين المتغيرين. وللباحث الحرية فى اختيار أى قيمة يريد اختبارها فى المجتمع مثلًا:

$$H0:\mu = 118,\ \mu = 100,\ \mu = 0.0$$

وليس شرطًا أن يكون المعلم مساويًا للصفر وإنما مساوى لقيمة معينة يفترضها الباحث. والقيود الإحصائية على القيمة العددية التى تظهر فى H0 هى:

- يجب أن تكون القيمة المراد اختبارها تقع فى متصل القيم الاحتمالية التى تناظر المعلم (1.00 , 0.0 , 1.00-).

- لا يمكن تثبيتها عند الحد الأعلى أو الأدنى للمعلم مثلًا:

$$H0: \rho = +1.30$$

$$H0: \rho = 1.00$$

فهذا غير مقبول لأن قيمة معامل الارتباط فى المجتمع لا تزيد عن 1.30 أو لا تصل إلى 1.00 خاصة فى العلوم النفسية والاجتماعية. فإذا صاغ باحث الفرض الصفرى كالآتى: $(\mu_1-\mu_2=0)$ $H0: \mu_1=\mu_2$، فهذا يعنى أن متوسط المجتمعين متساوٍ ولا فرق بينهما وهذا يكون مفاده أن الفروق بينهما صفرًا.

وعلى الرغم أن الباحثون لهم الحرية فى أى قيمة تختبر حولها الفروض إلا أن القيمة صفر تبقى هى الاكثر استخدامًا فى الدراسات والبحوث عند المقارنة بين مجتمعين. ويعتقد معظم الباحثون أن الفرض الصفرى يجب أن يوضع (لا فروق وتساوى صفرًا) ولكن هذا خطأ فيمكن أن يوضع الفرض الصفرى كالآتى:

$$H0: \mu 1-\mu 2 = 5$$

والفرض الصفرى قائم حول اختبار المعلم فى المجتمع من خلال بيانات العينة. ونادرًا ما يشير الباحثون أو يتحدثون عن الفرض الصفرى فى تقرير بحوثهم.

- **الفرض البديل Alternative Hypothesis:** يجب على الباحث صياغة فرض بديل مقابل للفرض الصفرى، ويعرف بأنه فرض البحث (Verma, 2004)، ويعنى وجود فروق بين معالم المجتمعات وبيانات العينة وكذلك وجود أثر للمعالجة، وجود تغيير، وجود علاقة. ويرمز له بالرمز HA أو H1 حيث H هى اختصار لـ Hypothesis و A هى اختصار لـ Alternative، وعلى ذلك ففى أى دراسة بحثية يبدأ الباحث فى صياغة فرض البحث لاختبار معلم المجتمع ويقوم بصياغة الفرض الصفرى للتحقق من فرض البحث إحصائيًا، وعلى ذلك فالفرض البديل هو أن الادعاء بأن الفروق فى النتائج بين المعالجات أو الظروف أو المجموعات ترجع إلى المتغير المستقل، فإذا كان الفرض الصفرى يتعامل مع قيمة محتملة مثلًا لمعامل ارتباط

بيرسون فى المجتمع: $\rho=0.00$:H0 ، فإن الفرض البديل يجب أن يتعامل مع نفس القيمة المفترضة فى المجتمع:

$$HA: \rho > , <, \neq$$

وإذا كان الفرض الصفرى يتعامل مع فروق بين متوسطات مجتمعين:

$$H0: \mu_1 = \mu_2$$

فإن الفرض البديل: $HA: \mu_1 \neq \mu_2 , \mu_1 > \mu_2, \mu_1 < \mu_2$

وعلى ذلك فإن H0 , H_A متماثلين فى أنهما يتعاملوا مع نفس المجتمع ونفس الاهتمام الإحصائى (المعلم) ونفس المتغيرات، الاختلاف الوحيد بينهما هو القيمة المفترضة فى المجتمع المتضمنة فى H_A تختلف عن تحديدها فى H0 فإذا كان:$\rho =0$:H0 ، فإن $\rho \neq 0$:HA، ودائمًا اختبارات الفروض تنتهى بقرار إما رفض أو الفشل فى رفض الفرض الصفرى. فإذا رفض H0ب التالى فإن HA يعكس الحالة فى المجتمع، بكلمات أخرى فإنH0 , HA وجهين متعارضين بمعنى 50% قبول H0 و50% قبول HA.

والقضية فى صياغة الفرض HA هى هل يوضع $\neq 0.1$ او > 0.10 او < 0.1 وهذا يقودنا إلى قضية الاتجاهية للفرض البديل بمعنى صياغة الفرض البديل صياغة موجهة Directional أو صياغة غير موجهة Non-directional ولتوضيح ذلك دعنا نعرض المثال الآتى:

أجرى باحث دراسة لمقارنة الذكور والإناث فى الذكاء وكان اهتمامه هو المتوسط وعلية فإن الفرض الصفرى : $H0: \mu_{\text{إناث}} = \mu_{\text{ذكور}}$

وإذا كان فرض البحث لا يوجد فروق بين الذكور والإناث فى الذكاء فإن الفرض البديل: $HA: \mu_{\text{إناث}} \neq \mu_{\text{ذكور}}$، وعندئذ يقال ان الفرض البديل بانة غير موجه. ولكن إذا كان فرض البحث توجد فروق بين الذكور والإناث فى الذكاء لصالح الذكور أو توجد فروق بين الذكور والإناث فى الذكاء لصالح الإناث عندئذ يقال أن الفرض البديل موجه حيث:

$$HA: \mu 1 > \mu 2$$

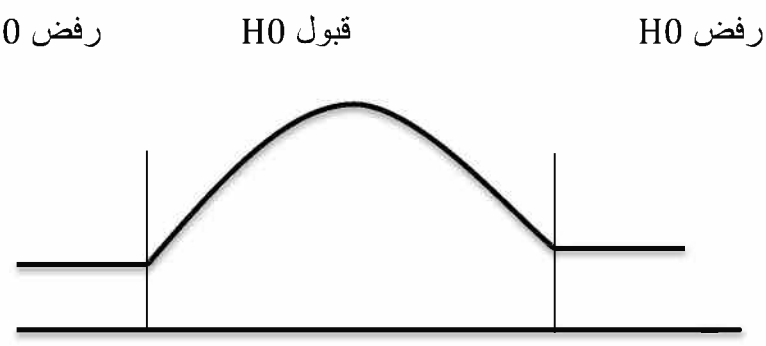
أو

$$HA: \mu 1 < \mu 2$$

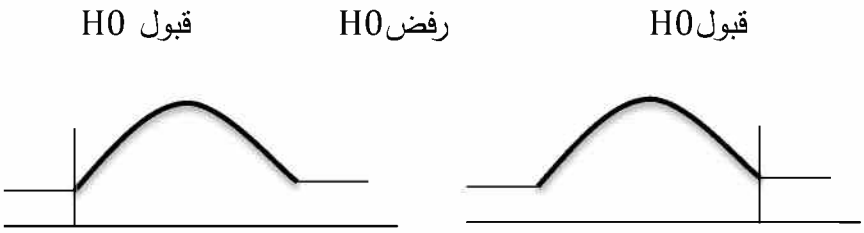
وعمليـة تحديـد طبيعـة اتجاهيـة الفـرض البـديل سـواء كـان موجهـة أو غيـر موجهـة علـى غايـة كبيـرة مـن الأهميـة فـى إجـراء اختبـارات الفـروض، ولا يمكـن صـناعة قـرار حـول H0 مـا لـم يـتم تحديـد خاصـية الاتجاهيـة سـواء موجهـة أو غيـر موجهـة. ويـرى Huck (2012) أنـه فـى معظـم الدراسـات الإمبريقيـة توضـع الفـروض البديلـة غيـر موجـه، ويمكـن عـرض فروض غير موجهة وموجهة كالآتى:

فرض بديل موجهة	فرض بديل غير موجهة	فرض صفرى
$H_A: \rho > 0.20$	$H_A: \rho \neq 0.20$	$H0: \rho = 0.20$
$H_A: \mu_1 > \mu_2$	$H_A: \mu_1 \neq \mu_2$	$H0: \mu_1 = \mu_2$

ويضـع البـاحثون H_A بصـورة غيـر موجهـة لأنهـم غيـر متأكـدين مـا إذا كانـت المعلمـة فـى الفـرض الصـفرى كبيـرة أو صـغيرة أو لصـالح أى مجتمـع، ويمكـن صـياغته بصـورة موجهـة إذا كـان هنـاك تـدعيم مـن الدراسـات السـابقة أو النظريـة. ويـرى Huck (2012) مـن غيـر المناسـب للبـاحثين فحـص البيانـات أولًا وتحديـد فـى أى اتجـاه تكـون المعلمـة فـى الفـرض البـديل وعلـى ذلـك فـالقرار فـى تحديـد اتجاهيـة الفـرض البـديل يجـب أن يكـون مسـتقل تمامًـا عـن أى حـديث عـن البيانـات أو فحصـها. والفـرض البـديل غيـر الموجهـة يطلـق عليـه البـاحثين اختبـار ذو ذيلـين Two tailed test لأن قاعـدة القـرار تكـون علـى جانبى (ذيلى) التوزيع العينى الاعتدالى.



بينما الفرض البديل الموجه يسمى اختبار ذو ذيل واحد One tailed test وقد يكون اتجاهية المعلمة لأعلى (أكبر من >) HA: μ1> μ2 عندئذ يطلق عليه اختبار ذو ذيل واحد لأعلى Upper one tailed test، وقد تكون اتجاهية المعلمة لأدنى (أصغر من <) HA: μ1 < μ2 عندئذ يطلق عليه اختبار ذو ذيل واحد لأدنى Lower one tailed test.



(A): ذو ذيل واحد لأدنى (B): ذو ذيل واحد لأعلى

على ذلك فالفرض غير الموجه يكون محدد بالعلامة (≠). وهى شائعة الاستخدام فى البحوث الاستكشافية وتصاغ عندما يعانى الباحث من ندرة فى الدراسات السابقة أو وجود تعارض بين نتائج الدراسات السابقة، بينما الفرض الموجه يكون محدد بالعلامة (> أو <) و تصاغ فى البحوث التوكيدية حيث توجد أدلة كافية لتحديد طبيعة اتجاه الفرض.

ثانيًا: اختيار مستوى الدلالة الإحصائية Selecting level of significance: بعد صياغة الفروض الإحصائية فإن الخطوة التالية هى تحديد الظروف التى من خلالها يرفض الفرض الصفرى وأحيانًا يؤجلها البعض إلى بعد مرحلة جمع البيانات وتلخيصها،واختبارات الفروض لا تسمح مطلقًا بوجود ذاتية للباحث لتؤثر على الوصول إلى القرار مما يؤدى إلى وجود تحيز فيما يخص التحقق من الفرض الصفرى. وعليه فإن الهدف تحديد أو الإعتماد على نقطة قطع لتحديد ما إذا كانت بيانات العينة متسقة أو متعارضة مع الفرض الصفرى. ومستوى الدلالة الإحصائية يعتبر بمثابة نقطة القطع الموضوعية لتحديد ما هو القرار الذى نتوصل إليه فيما يخص الفرض الصفرى، وهذا التحديد لا بد أن يتم قبل جمع البيانات وهذه نقطة القطع يتم فى ضوئها تحديد القيمة الحرجة للاختبار. على الرغم أن للباحث الحرية لاختيار أى قيمة لمستوى الدلالة الإحصائية من 0.00 إلى 1.0 ولكن معظم الباحثين يختاروا قيم صغيرة مثل 10.(

10%) ، 0.05 (5%)، 0.01 (1%) والقيمة الأكثر استخدامًا وشيوعًا في البحث النفسي والسلوكي هي 0.05، ويرى Gravetter & Wallnau (2014) أن مستوى ألفا Alpha level هي حدود أو قيمة احتمالية تفصل العينات ذات الاحتمالية العالية من العينات ذات الاحتمالية المنخفضة ولكن ما المقصود بعالية الاحتمالية ومنخفضة الاحتمالية فهذا مرتبط باختيار قيمة احتمالية محددة تعرف بمستوى الدلالة الإحصائية لاختبار الفروض. فعند استخدام α =0.05 فنحن نفصل بين معظم الـ 5% من متوسطات العينة (القيم المتطرفة) غير المتوقعة عن 95% من المتوسطات المتوقعة للعينة (القيم المركزية). والقيم غير المتوقعة Extreme values أو الشاذة كما حددت عن طريق مستوى ألفا تسمى المنطقة الحرجة Critical region، وهذه القيم الحادة في ذيلي المنحنى أو التوزيع تعرف بأنها نواتج لا تتسق مع الفرض الصفرى ومن غير المحتمل حدوثها لو أن الفرض الصفرى حقيقي. وعندما تقع متوسط بيانات عينة البحث في منطقة الرفض فإننا نستدل على أن البيانات غير متسقة مع الفرض الصفرى وبالتالى نرفضه.

ويستخدم كل الباحثون مصطلح مستوى الدلالة ويستخدم أيضًا مستوى ألفا، بينما يذكر البعض 0.05 = p وأحيانًا يشير البعض إلى أن رفض الفرض الصفرى إذا كانت(α)0.05 > p وهكذا بالنسبة لـ α =0.01 وصناعة القرار بالاعتماد على برامج الكمبيوتر يتم بمقارنة قيمة p مع قيمة α لتحديد ما إذا كان الفرض الصفرى يرفض من عدمه، وتلعب مستوى الدلالة الإحصائية دورًا هامًا في قاعدة القرار وذلك في الحصول على القيمة الحرجة Critical value للاختبار (القيمة الجدولية) من خلال الكشف في الجداول لكل اختبار محدد.

وبغض النظر عن القرار حول H0 بعد مقارنة p بـ α أو مقارنة القيمة الحرجة بالقيمة المحسوبة للإحصاء المستخدم فمن المحتمل أن يصل الباحث إلى قرار خاطئ، فاختبار فرض صفرى في ضوء فرض بديل باستخدام بيانات العينة يصاحبه نوعان من الخطأ قد يكون بسبب عدم انتقاء العينة عشوائيًا أو عدم توافر مسلمات الاختبار الإحصائى المناسب أو بسبب عوامل أخرى أثناء التحليل أو التطبيق.

الأخطاء القرارية فى اختبارات الفروض

اختبارات الفروض هى عملية استدلالية بمعنى أننا نستخدم معلومات محدودة من العينة كأساس للوصول إلى استنتاج عام والعينة هى الوحيدة التى تمدنا بكم المعلومات سواء كانت غير كاملة أو كاملة عن طبيعة المجتمع ككل. ودائمًا فى هذه الحالة يوجد احتمال أن القرار الذى نصل إليه على مستوى العينة يكون غير صحيح، وعلى الرغم أن العينة يجب أن تكون ممثلة للمجتمع إلا أنه توجد فرصة لارتكاب قرار خاطئ بالتالى عمل استنتاجات أو قرارات خاطئة على مستوى المجتمع، بكلمات أخرى الوصول إلى نتائج بحث مزيفة أو غير حقيقية، وفى اختبارات الفروض يوجد نوعين أساسيين من الأخطاء هما:

- **الخطأ من النوع الأول (α) Type I error:** من المحتمل أن بيانات العينة تقودنا إلى رفض الفرض الصفرى (عدم دلالة إحصائية) بينما فى الحقيقة (المجتمع) فإن للمعالجة أثر (دلالة إحصائية). لاحظ أنه ليس فى كل الأحوال أن تكون العينة ممثلة للمجتمع بل أحيانًا توجد عينات مختلفة عن خصائص المجتمع وهذا ما يشار إليه بخطأ المعاينة.

فافترض أن باحث اختار عينة N=25 من طلاب لديهم مهارات رياضية عالية بمعنى درجاتهم فوق المتوسط فحتى إذا كانت المعالجة ليس لها اثر فإن أفراد العينة مازالت درجاتهم مرتفعة فوق المتوسط وفى هذه الحالة يصل الباحث إلى استنتاج أن المعالجة لها أثر بينما فى حقيقة الأمر هى ليست كذلك وعلى ذلك فإن الباحث ارتكب خطأ فى صناعة القرار وهذا يطلق عليه الخطأ من النوع الأول ويحدث إذا رفض الباحث الفرض الصفرى على مستوى بيانات العينة بينما على مستوى الواقع أو المجتمع حقيقى بمعنى أن الباحث يصل إلى نتيجة أن للتجربة أثر بينما فى الواقع ليس لها هذا الأثر.

وفى كل المواقف البحثية فإن عواقب الخطأ من النوع الأول فى غاية الخطورة لأن الباحث يرفض الفرض الصفرى ويعتقد أن للمعالجة أثرًا أو توجد دلالة إحصائية بمعنى قبول الفرض البديل وهذه تكون نتيجة منشورة فى مجلة ما، وقد تكون دراسة قومية ويترتب عليها اتخاذ قرارات هامة فى العملية المؤسسات التعليمية المختلفة وهذا يعنى أن التقرير خاطئ False report، وعليه فإن الخطأ من النوع الأول يقود إلى تقارير

خاطئة في الأدبيات البحثية المرتبطة بظاهرة ما. ويمكن للباحثين أن ينشؤا نظريات ويعدوا تجارب أخرى في ضوء هذه النتائج الخاطئة وبدوره يترتب عليه إهدار للوقت والموارد.

- **الخطأ من النوع الثانى (β) Type II error:** عندما يرفض الباحث الفرض الصفرى فهذه مخاطرة الخطأ من النوع الأول، بالمثل عندما يفشل الباحث فى رفض الفرض الصفرى (عدم دلالة إحصائية) على مستوى بيانات العينة بينما هو فى الواقع مرفوض (غير صحيح) (دلالة إحصائية)، بالتالى فإن الخطأ من النوع الثانى هو الفشل فى رفض الفرض الصفرى الخاطئ، بكلمات أخرى فإن الخطأ من النوع الثانى يعنى أن للمعالجة أثر على المستوى الواقعى الحقيقى بينما الاختبار الإحصائى فشل فى الوصول إلى هذه الدلالة الإحصائية أى أن للتجربة أثر فى المستوى الحقيقى، بينما بيانات العينة أعطت عدم أثر للتجربة بمعنى قبول H0على مستوى العينة بينما هو خطأ على مستوى المجتمع. والخطأ من النوع الثانى يحدث عندما يكون متوسط العينة ليس فى نطاق منطقة الرفض لـ H0على الرغم أن للمعالجة أثر فى العينة، ويحدث هذا غالبًا عندما يكون تأثير المعالجة أو المتغير المستقل صغير نسبيًا حيث إن مقدار الأثر ليس كبير ليحرك متوسط العينة إلى منطقة الرفض. وعواقب الخطأ من النوع الثانى ليس بخطورة عواقب الخطأ من النوع الأول لأن الخطأ من النوع الثانى يعنى أن الباحث لا يصل إلى نتائج ذو قيمة (عدم الدلالة الإحصائية)، لا يترتب على ذلك تطبيق برنامج أو بناء نظريات ما.

الباحث يمكن أن يقبل هذه النتيجة ويستنتج إما أن تكون للمعالجة بلا تأثير، وإما أن يكون لها تأثير ضعيف، وعليه ليس لها دور فى التراث أو يمكن تعاد التجربة أو البحث مرة أخرى سواء بوجود تحسينات منهجية أو زيادة حجم العينة وعكس الخطأ من النوع الأول فإن الخطأ من النوع الثانى لا توضع له قيمة احتمالية أو لا يمكن التحكم فيه حيث فى الخطأ من النوع الأول يمكن التحكم فيه من خلال مستوى ألفا. ويعتمد الخطأ من النوع الثانى على مجموعة من العوامل أهمها حجم العينة وتباين المجتمع وثبات أدوات القياس وليس له قيمة احتمالية معينة. ويمكن عرض الأخطاء القرارية المصاحبة لاختبارات الفروض كالآتى:

الحالة الحقيقة (المجتمع H0) ؟

العينة / المجتمع	H0 حقيقى لا أثر للتجربة	H0 غير حقيقى أثر للتجربة
أقبل H0	لا أخطاء No errors	خطأ من النوع الثانى (بيتا β)
أرفض H0	خطأ من النوع الأول (ألفا α)	لا أخطاء No errors

واختيار 0.05 =α فإنه الفرصة لرفض الفرض الصفرى الحقيقى يساوى 5 مرات من 100مرة القرار فيهم صحيح. وإذا كانت 0.01 =α فإن الفرصة لرفض 100 بقيم مساوية مرة من مائة مرة . وعليه فإن α هى احتمالية ارتكاب الخطأ من النوع الأول.

وبعد أن يدرك الباحث أنه ضبط الخطأ من النوع الأول فربما تكون مندهش لماذا لا يختار الباحث مستوى ألفا التى تقلل احتمالية أن 100الحقيقى تتم رفضه. ويتساءل الباحث لماذا لا تكون ألفا 0.001 حيث فرصة الخطأ من النوع الأول 1 من 1000 أو 0.0001 حيث فرصة احتمال الخطأ من النوع الأول 1 من عشرة ألاف أو أقل من ذلك، للإجابة على هذا التساؤل لا بد أن نضع فى الاعتبار أن تغير قيمة ألفا له تأثير على β و α.

ويمكن تشبيهه ارتكاب الخطأ من النوع الأول والثانى لما يحدث فى المحكمة مع القاضى حيث يوجد موقفين القرار فيهما صحيح وموقفين أخرين القرار فيهما خطأ، فالقرارين الصحيحين هما المتهم برئ وهو حقيقة برئ والآخر المتهم مذنب وهو حقيقة مذنب ولكن يحدث أخطاء منها:

عندما يكون المتهم برئ فى الواقع وثم يحكم عليه القاضى بأنه مذنب بمعنى أن القاضى حكم بالإدانة على متهم وهو حقيقة برئ (الخطأ من النوع الأول) والخطأ الآخر عندما يحكم القاضى على متهم بالبراءة وهو فى الحقيقة مذنب (الخطأ من النوع الثانى) ويمكن تشبيهه ذلك بالتالى:

	الحقيقى أن المتهم	
	برئ	مذنب
حكم القاضى **برئ**	قرار صحيح(1−α)	قرار خاطئ (β)
مذنب	قرار خاطئ α	قرار صحيح (1− β) القوة الإحصائية

حيث تقليل مستوى ألفا له تأثير على خفض الاحتمال فى الوقوع فى الخطأ من النوع الأول ولكن هذه القيمة المنخفضة لـ α لها تأثير عكسى على زيادة الخطأ من النوع الثانى (β). وعليه فالباحث نادرًا ما يغير من قيمة مستوى الدلالة الإحصائية 0.05 إلى مستويات أقل (0.001) للحماية من الوقوع فى الخطأ من النوع الأول لأن هذا التغير فى α بالزيادة يؤدى إلى احتمالات عالية غير مقبولة للوقوع فى الخطأ من النوع الثانى. ولذلك فإن معظم الباحثون يفضلوا مستويات دلالة إحصائيةα متحفظة Conservative مثل 0.05 ، 0.01 وأحيانًا 0.001 لتقليل مخاطر الوصول إلى تقرير بحث خاطئ. ومستوى ألفا 0.000001 يعنى تقريبًا لا مخاطرة لارتكاب الخطأ من النوع الأول.

ولكن ماذا لو أن باحث تبنى مستوى ألفا أكبر من 0.05، مثلًا جعلها 0.10 وهذا يعطى احتمالية عالية لرفض H0. ولكن معظم الباحثون لا يستخدموا هذه المستويات العالية لأن المجتمع العلمى اعتبر أن الخطأ من النوع الأول هو الأكثر خطورة من الخطأ من النوع الثانى. وفى بعض التخصصات قليل من الباحثين يعطوا انتباه لبعض للدراسات التى ترفض H0 عند مستويات أعلى من 0.2 لأن هذه المستويات العالية من α تعتبر متساهلة. وعليه فإن معظم مستوى الدلالة الشائعة فى كثير من الدراسات هى 0.05 ويرى (2012) Huck أن بعض الباحثين يكونوا فى صالح استخدام ألفا 0.01 أو 0.001 ولكن على الجانب الآخر إذا اعتقد الباحث أن الخطأ من النوع الثانى أشد خطوة من الخطأ من النوع الأول فلا حرج فى اختيار مستويات عالية لـ α مثل 0.1 و 0.15 ولكن هذا غير شائع فى العلوم السلوكية والنفسية على الرغم أن الخطأ من النوع الثانى β يلعب دورًا هامًا فى جودة الاختبار الإحصائى حيث إنه مرتبط بمفهوم فى غاية الأهمية فى اختبارات الفروض وهو قوة الاختبار الإحصائى.

وقبل إنهاء هذه المناقشة نطرح تساؤلين فى غاية الأهمية **الأول:** هل مستوى ألفا يمكن أن يلعب دورًا فى احتمالية الوقوع فى الخطأ من النوع الثانى؟ **والثانى:** لو رفض الباحث H0 فهل مستوى ألفا يشير إلى أن احتمالية H0 حقيقى؟ حقيقة التساؤل الأول ينصب عن العلاقة بين مستوى ألفا والخطأ من النوع الثانى وسبق أن تناولتها فى سياق الحديث بينما التساؤل الثانى يخص بعد أن يصل الباحث فى صناعة القرار (الخطوة السادسة) وتم رفض H0 فليس من المناسب أن يعود الباحث مرة أخرى ويعيد النظر فى مستوى α ويفسر مستوى ألفا على أنه مؤشر الاحتمالية أن H0 صحيح، ومستوى ألفا فى الدراسة يشير فقط إلى أن 5 مرات من 100 مرة H0 يرفض إذا كان H0 حقيقى.

ويرى Verma (2004) أنه يوجد مدخلين لتقليل الخطأ من النوع الأول والثانى وهما:

المدخل الأول: زيادة حجم العينة ولكن هذه تبدو إشكالية فى بعض التصميمات البحثية خاصة التصميم التجريبى حيث يفقد الضبط أثناء التجربة.

المدخل الثانى: تحديد الخطأ لكى يكون حاد Severe وتثبيته عند مستوى مرغوب ومحاولة تقليل الخطأ الآخر بأقصى درجة ممكنة مثل 0.01 , α=0.05 وعليه فإن الخطأ من النوع الثانى يكون عند أدنى قيمته. وإذا انخفضت ألفا من الواحد إلى الصفر فإن β تزيد من الصفر إلى الواحد. ويجب تحديد مستوى الدراسة الإحصائية فى بداية الدراسة ولا يجب تغيرها بعد تحليل البيانات ويفضل أن يكون مستوى الدلالة الإحصائية واحد لكل أسئلة الدراسة.

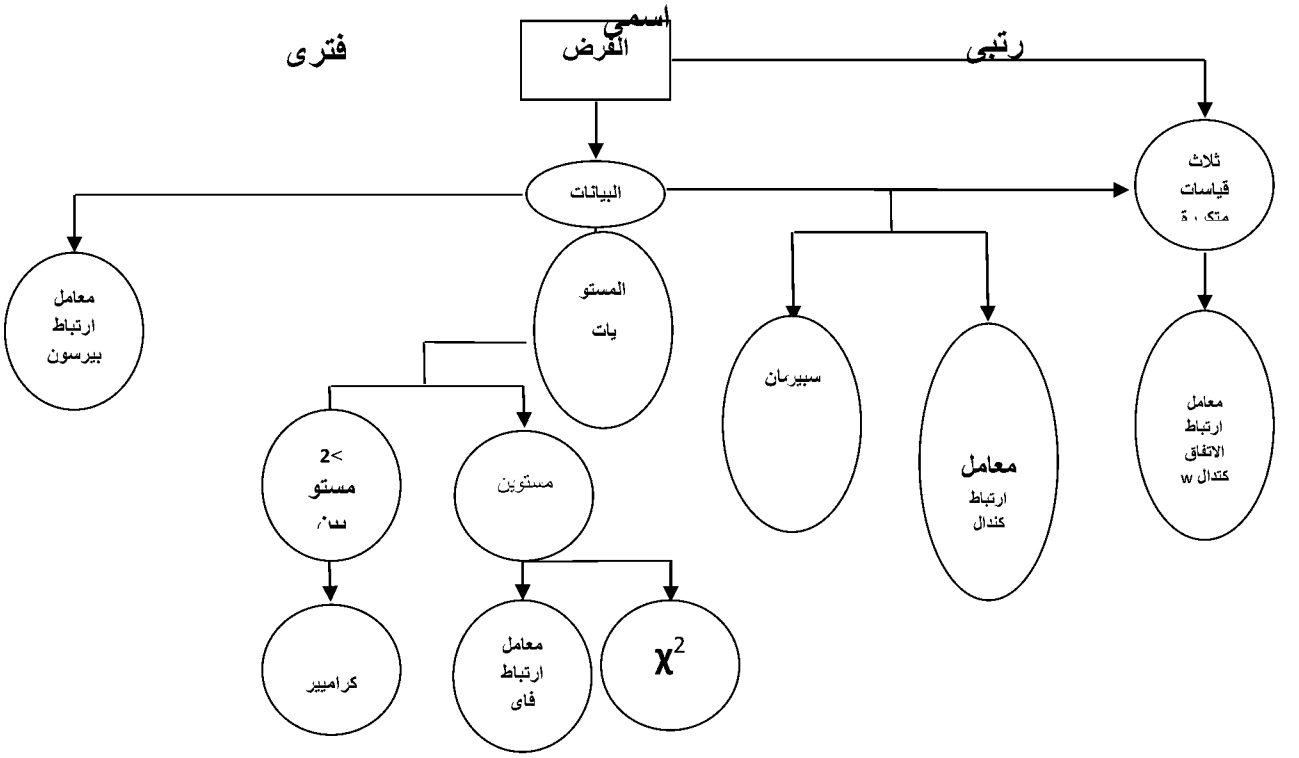
ثالثًا: اختيار الاختبار الإحصائى المناسب: الاختبار الإحصائى هو تكنيك يستخدم لاختبار فرض معين. ويعتبر اختيار الاختبار الإحصائى من أصعب وأعقد الخطوات فى عملية اختبارات الفروض والسبب بكل بساطة هو وجود العديد من الاختبارات الإحصائية المتاحة والاختيار الخاطئ للاختبار للتحقق من الفرض أو البيانات فإن بقية الخطوات اللاحقة هى مضيعة للوقت بمعنى أن النتائج لا قيمة لها.

وتوجد العديد من العوامل التى تلعب دورًا فى اختيار الاختبار الإحصائى:

1. **الهدف من التحليل أو نوع الفرض المراد اختباره:** فى استخدام الإحصاء توجد عدة أهداف للتحليل وهى:

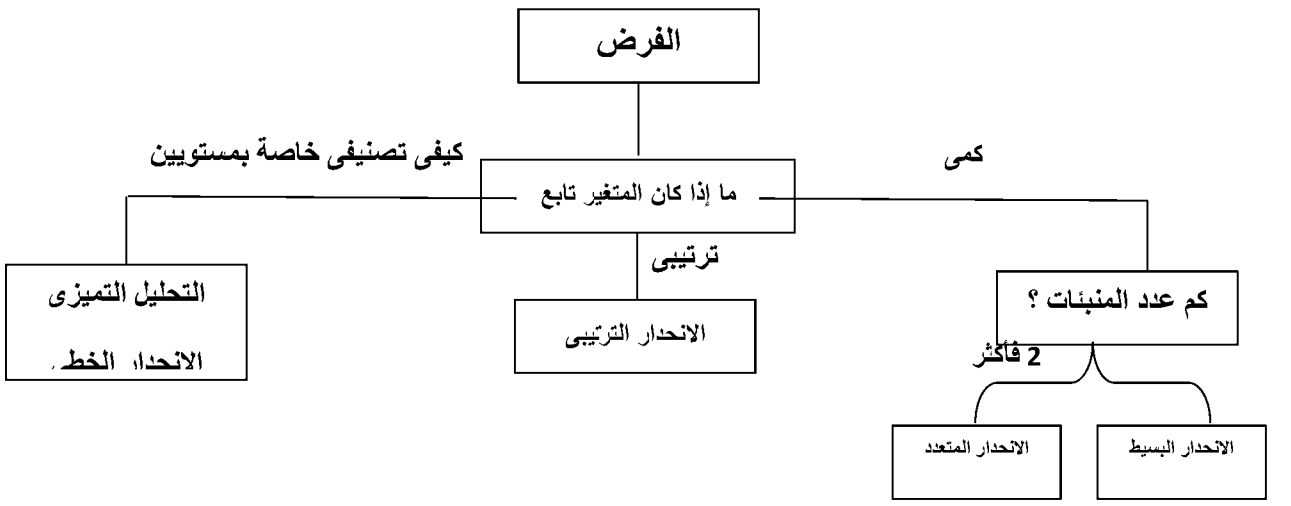
- **وصف واكتشاف إحصائيات البيانات Descriptive focus:** وفيها يهدف الباحث إلى إعطاء صورة ملخصة لبيانات العينة وكذلك تمثيل البيانات بيانيًا وهذه النوعية من الإحصاء الوصفى تتضمن النسبة المئوية (20% مدمن الفيس بوك مثلًا) أو مقاييس النزعة المركزية ومقاييس التشتت والرسومات البيانية وكل هذه الأساليب تندرج تحت الإحصاء الوصفى.
- عندما يكون هدف الباحث الاستدلال من بيانات العينة للحالة أو الوضع فى المجتمع فإن هدف الباحث هو التقدير Estimation focus وعلى ذلك فلا بد للباحث التركيز على اختبارات الإحصاء الاستدلالى أو اختبارات الدلالة الإحصائية للتحقق من الفروض أو الأسئلة الآتية:
- **الكشف عن فروق Differences :** حيث يكون اهتمام الفرض أو السؤال دراسة فروق بين مجموعتين (أو متوسطين) فأكثر، بالتالى توجد الاختبارات الفارقة البارامترية مثل T المستقلة أو المرتبطة، واختيار F المستقلة أو المرتبطة وغيرها، والاختبارات اللابارامترايــة مثل مان – ويتنى، اختبار كروسكال- والاس، فريدمان، ويلكوكسون وغيرها.
- **العلاقة بين متغيرين فأكثر:** يهدف الباحث إلى دراسة للعلاقة بين متغيرين وتوجد اختبارات عديدة مثل معامل ارتباط سبيرمان ومعامل ارتباط بيرسون ومعامل ارتباط كندال وغيرها.

وفيما يلى مخطط يوضح أهم الاختبارات العلاقة:



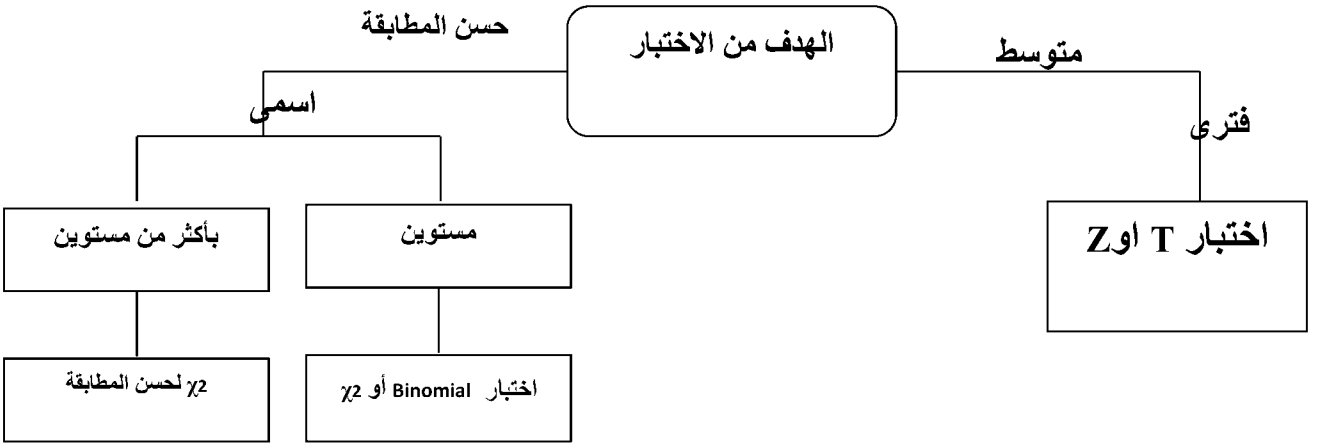
الشكل(1.9): مخطط لأهم اختبارات العلاقة.

- **عمـل تنبـوءات**Making predictions: إذا وجـد علاقـة أو ارتبـاط بـين متغيـرين فـإن الهـدف التـالى هـو دراسـة مـا إذا كـان يمكـن التنبـؤ بأحـد المتغيـرات (التـابع) مـن المتغيـر المستقل.

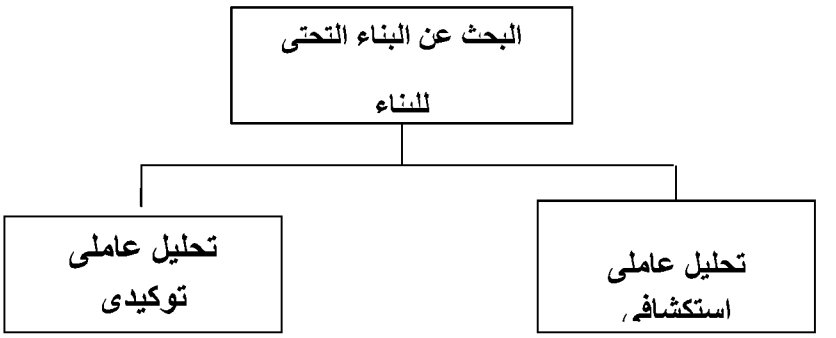


الشكل (2.9): مخطط لأهم اختبارات التنبؤ أو الانحدار.

- **دراسة مدى حسن المطابقة بين بيانات العينة والمجتمع:** بمعنى اختبار فروض حول متغير واحد.



- **البحث عن البناء التحتى للمتغيرات:** وإذا كان هدف البحث العثور عن البناء الداخلى لمفهوم ما وذلك لمجموعة فرعية من المفردات المكونة للأبعاد.



2 . خصائص التوزيع Distributional charecteristics: فالمسلمة الأساسية فى استخدام الاختبارات الإحصائية البارامترية هى أن بيانات العينة تكون مسحوبة من مجتمع ذات توزيع اعتدالى، وهذه المسلمة ضرورية لاختبار فروض حول متوسطات تباينات أو نسب أو ارتباطات ، وكذلك من أهم المسلمات الواجب توافرها هى تساوى التباينات (تجانسها) بمعنى أن تباينات المجتمعات متساوية، بالتالى فاختبارات فروض المتوسطات (T , ANOVA) بين مجموعتين فأكثر لا بد أن يتوفر لها هذه المسلمات وهذه المعالم مثل المتوسط والتباين يستخدم لها اختبارات بارامترية Parametric tests بينما فى حالة عدم توافر مسلمتى الاعتدالية وتجانس التباينات فإنه يفضل استخدام اختبارات لا بارامترية Non-parametric tests أو إحصاء

التوزيعات الحرة وإذا تحققت هذه المسلمات بدرجة متوسطة وليست كاملة فإن نتائج الاختبارات البارامترية تكون صادقة إذا كان حجم العينة كبيرًا .

3. مستوى القياس للمتغيرات: فمستويات القياس عالية الرتبة (الفترى – النسبى) تحتاج إلى إحصاء أكثر تعقيدًا وهى تختبر فروض حول معالم حيث لها متوسط وانحراف المعيارى وتباين، وهذه النوعية من مستويات القياس تعطى بيانات مترية Metric data فى هذه الحالة يستخدم الإحصاء البارامترى ، بينما مستويات القياس الأدنى رتبة (الاسمية– الرتبية) فلا يتوفر لها معالم لاختبار فروض حولها حيث تعطى بيانات لا مترية Non-metric data ولذلك يستخدم معها إحصاء لا بارامترى.

4. خصائص العينة (نوعها – حجمها): كما هو معروف ما لم تكن اختيار العينة بصورة احتمالية عشوائية فإن استخدام الإحصاء الاستدلالى يكون أقل موثوقية وذلك لأن الاختيار العشوائى للعينة يقلل من أخطاء المعاينة. اما فيما يخص حجم العينة فإن بعض الإجراءات الإحصائية خاصة البارامترية لا تؤدى بكفاءة مالم يكن حجم العينة كبيرًا ويفضل البعض أن يزيد عن 30 فردًا، فمع العينات الكبيرة يفضل استخدام الإحصاء البارامترى ولكن مع أحجام العينات الصغيرة يفضل استخدام الإحصاء اللابارامترى خاصة إذا لم تتوافر الاعتدالية للبيانات ولكن حتى مع أحجام العينات الصغيرة يستخدم الإحصاء البارامترى بفاعلية وكفاءة إذا توافرت مسلمات الاختبار خاصة الاعتدالية.

5. طبيعة القياسات Nature of measures: توضع طبيعة القياسات فى الاعتبار عند اختيار الأسلوب الإحصائى المناسب، فمثلًا إذا كانت القياسات على نفس الأفراد عبر الزمن (قياس قبلى – قياس بعدى) فانة يطلق عليها تصميم القياسات المتكررة Repeated Measure Design وهو أن نفس المفردات يقاس عليهم متغيرات مختلفة بهذه النوعية تتبع من تصميم البحوث الطولية (قبل – بعد) Longitudinal Research Design وكذلك فى الدراسات التجريبية وهذه النوعية من العينات يطلق عليها العينات المرتبطة وإذا كان قياس متغير ما على مجموعات مختلفة (قياس بعدى ضابطة –قياس بعدى تجريبية أو ذكور – إناث) فإنه يطلق عليها تصميم القياسات

المستقلةIndependent Measure Design وهو أن الأفراد المختلفين يقاس عليهم نفس الخاصية وهذه النوعية تنبع من تصميم البحوث المستعرضة Cross-section Research وكذلك من الدراسات السببية المقارنة والدراسات التجريبية وهذه النوعية من العينات يطلق عليها العينات المستقلة، وعليه فإن الأساليب الإحصائية للعينات المرتبطة غير الأساليب الإحصائية للعينات المستقلة.

6.عدد المتغيرات فى التحليل: إذا كان التحليل قائم حول اختبارات فروض لمتغير واحد أو لعينة واحدة فإن التحليل يسمى **Univariate analysis**، وإذا كان التحليل قائم على متغيرين يستخدم إحصائيات النموذج البسيط ويطلق عليها **Bivariate analysis**، أو إذا كان التحليل يتضمن متغير تابع وحيد فإنه يسمى أيضًا **Univariate analysis**. أما إذا كان التحليل يتضمن أكثر من متغير تابع حيث النموذج الإحصائى اما أن يكون النموذج المتدرج Multivariate Multiple وفيه يتعامل الباحث مع متغير واحد مستقل فأكثر وأكثر من متغير تابع مثل MANOVA أو يكون النموذج المتدرج وفيها يتعامل الباحث مع قضية تتضمن العديد من المتغيرات حيث لا يتم تحديدا أيهما تابع وأيهما مستقل ويسمى النموذج فى هذه الحالة بالنموذج المتدرج Multivariate Model مثل التحليل العاملى الاستكشافى فإن الإحصاء المستخدم فى هذه الحالات الإحصاء المتدرج **Multivariate statistics** .

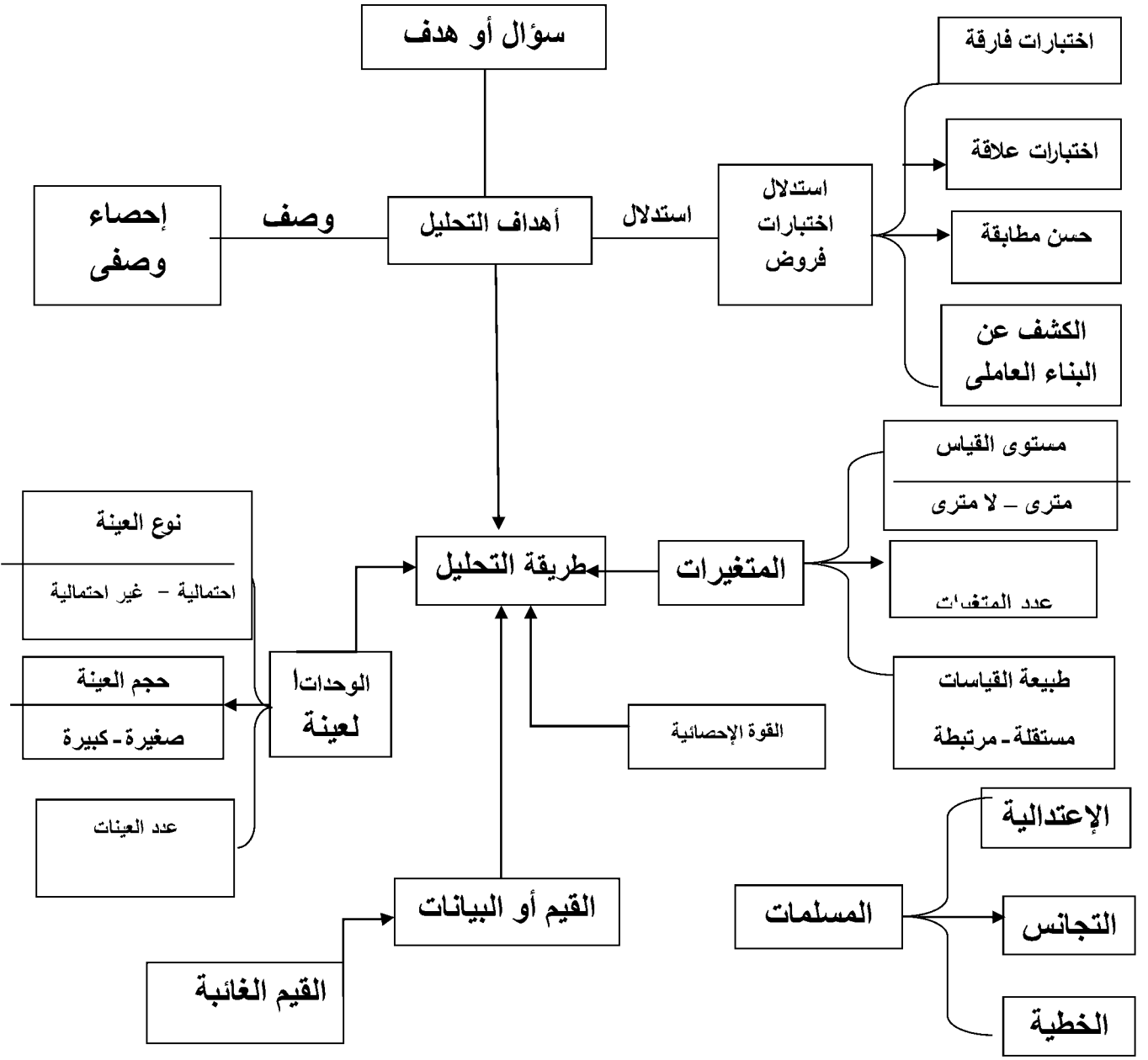
7. طبيعة البيانات المتحصل عليها من المفحوصين: القضية الأساسية هى ان المفردات التى لا يستجيب عليها المفحوصين Non-Response Item أو البيانات المفقودة Missing Data على متغير أو أكثر فى متغيرات التحليل، والأثر المباشر للبيانات الغائبة هو تقليل حجم العينة وهذا بدوره يؤثر على نوعية الإجراء الإحصائى المناسب، حيث يطبق الباحث فى دراسته على 50 فردًا وينوى استخدام استراتيجية تحليل بارامترية ولكن بعد استبعاد الأفراد الذين لديهم قيم غائبة على بعض مفردات المقياس يتوفر 20 فرد كامل البيانات فربما يفضل استخدام الإجراء اللابارامترى.

8. قوة الاختبار Power of a Test: تعرف قوة الاختبار على أنها الاحتمالية لرفض الفرض الصفرى عندما يجب أن يرفض أو قدرة الاختبار فى الحصول على

دلالـة إحصـائية وهـى موجـودة بالفعـل فـى مجتمـع الدراسـة وقـوة الاختبـار هـى متلازمـة مـع الخطأ من النوع الثانى β حيث: $P = 1-\beta$

والاختبـارات البارامتريـة أكثـر قـوة إحصـائيًا مـن الاختبـارات اللابارامتريـة حيـث إن الاختبـار البـارامترى يكـون أكثـر قـدرة علـى رفـض الفـرض الصـفرى وإن قـوة الاختبـار الإحصـائى تزيـد بزيـادة حجـم العينـة، وفـى حالـة أحجـام العينـات الصـغيرة فـإن الاختبـار البـارامترى واللابـارامترى لهمـا نفـس القـوة وإذا تـوافرت مسـلمات الاختبـار الإحصـائى البـارامترى وخاصـة الاعتداليـة فإنـه يقـال ان الاختبـار لـه صـدق إحصـائى عـالٍ High Statistical Validity هـى اسـتخدام الاختبـار مـع تـوفر أفضـل شـروط اسـتخدامه. ولكـن هـل عـدم تـوفر مسـلمات الاعتداليـة والتجـانس للتباينـات لا يسـتخدم الاختبـار البـارامترى؟ حقيقـة معظـم الاختبـارات البـارامترى تتميـز بالمناعـة أو الضـلاعة النسـبية Relative robust أى أنـه غيـر حسـاس نسـبيًا للابتعـاد الضـئيل أو المتوسـط Mild-to-Moderate عـن مسـلماته خاصـة مـع أحجـام العينـات الكبيـرة وهـذه المناعـة تكـون محـددة بكـل اختبـار بمعنـى كـل الاختبـارات البارامتريـة ليسـت لهـا نفـس الدرجـة مـن المناعـة ولكـن إذا لـم تتحقـق هـذه المسـلمات بدرجـة كبيـرة (التـواء شـديد – تفـرطح شـديد) فمـن الأفضـل عـدم الاعتمـاد علـى الاختبـار البـارامترى ويفضـل اسـتخدام نظيره اللابارامترى.

فـى ضـوء مـا سـبق فـى هـذه الخطـوة مـن اختبـارات الفـروض واسـتنادًا علـى فـرض أو سـؤال البحـث وتحديـد النمـوذج الإحصـائى المناسـب، فـإن الباحـث يسـتطيع تحديـد نـوع الإحصـاء (بـارامترى – لا بـارامترى) وبـدوره يحـدد الأسـلوب أو الاختبـار الإحصـائى المناسـب. يمكـن تلخـيص أهـم العوامـل المـؤثرة فـى اختيـار الأسـلوب الإحصـائى فـى الشكل الآتى:



الشكل (3.9): مخطط لأهم العوامل المؤثرة فى اختيار الأسلوب الإحصائى

Diamantopoulos & schlegelmilch (2003)

وبعد تحديد الاختبار فإن يجب أن تحدد قيمة درجات الحرية له.

درجات الحرية Degrees of freedom: أى اختبار إحصائى لا بد أن يكون مصحوب بدرجات الحرية الخاصة به، فتقدير أى معلم يتم من خلال مقدار معين من المعلومات أو مجموعة البيانات وعدد القطع المستقلة من البيانات أو الدرجات المستخدم فى تقدير المعلم يعرف بدرجات الحرية يرمز له بالرمز df .عمومًا فإن درجات الحرية لأى معلم تقدر باعتبارها عدد الدرجات المستقلة المتطلبة أو قطع أو

كميـة المعلومـات التـى تكـون حـرة عنـد حسـاب الإحصـائية المسـتخدمة فـإذا كـان تقـدير التبـاين $\sigma 2$ فـى المجتمـع يقـدر مـن خـلال S2 فـى العينـة مـن خـلال عينـة عشـوائية لعـدد n مـن الـدرجات وعليـة فـإن درجـات الحريـة تقـدر مـن n وفـى هـذه الحالـة (n-1) وتختلف درجات الحرية من اختبار لأخر، ففى حالة اختبار T المستقلة فان:

$$df = n1+n2-2$$

فى حالة $\chi 2$ لاستقلالية متغيرين فإن:

$$df =(c-1)(r-1)$$

- r عدد الصفوف
- C عدد الأعمدة.

لاحـظ فـى هـذه النقطـة بعـد تحديـد الاختبـار الإحصـائى ودرجـات الحريـة لا بـد للباحـث أن يتحقـق مـن مسـلمات الاختبـار الإحصـاء المسـتخدم سـواء كانـت الاعتداليـة، الاستقلالية، العشوائية، والتجانس وغيرها قبل تحليل البيانات.

وبفـرض ان باحـث اراد اختبـار فرضـية: مـا إذا كـان متوسـط التحصـيل فـى مجتمـع المدرسـة الثانويـة فـى مـادة الرياضـيات يزيـد عـن 0 ، فـإن الاختبـار المناسـب Z ولا بـد مـن التحقق من مسلمات استخدامه وهى:

- **المعاينـة العشـوائية**Random sampling: أن يـتم اختيـار مفـردات العينـة ولـتكن 200 من المجتمع وبصورة عشوائية حتى نثق فى تعميم النتائج.
- **القياسـات المسـتقلة:** العينـة تتكـون مـن قياسـات مسـتقلة حيـث كـل فـرد فـى العينـة مسـتقل عـن الآخـر، فدرجـة أحـد الأفـراد لا تتـأثر بدرجـة فـرد آخـر وهـذا يتحقـق فـى العينات العشوائية.
- **التوزيـع الاعتـدالي**Normal distribution: لتقيم فـروض اختبـار Z لا بـد مـن اسـتخدام جـداول التوزيـع الاعتـدالى لتحديـد منطقـة الـرفض، وعليـه فـإن بيانـات العينـة لا بد أن تكون اعتدالية التوزيع.

رابعًـا: قاعـدة القـرار ومنطقـة رفـض H0: أى قيمـة يـتم حسـابها مـن الاختبـار هـى مقيـاس معبـر لنتـائج الاختبـار وهـى تعكـس بيانـات العينـة ويطلـق عليهـا بالإحصـائية أو القيمـة المحسـوبة Calculated value بينمـا التوزيـع الاحتمـالى للإحصـاء المسـتخدم

يشير إلى كل قيم الاختبار عندما يكون H0 حقيقى (فى المجتمع) . وللحصول على التوزيع الاحتمالى فى المجتمع من خلال المتوسط (μ) والتباين S2 وهو توزيع تم افتراضة لدرجات كل أفراد المجتمع وعمليًا لا يمكن الوصول اليه ولكن تم افتراضة نظريًا عن طريق اخذ عينات متعددة من المجتمع وهو غالبًا يكون اعتدالى وهذا التوزيع الاحتمالى للاختبار يسمح بتقدير خطأ المعاينة. ولكثير من الاختبارات الإحصائية فإن التوزيع الاحتمالى يأخذ أشكالًا معروفة (اعتدالى مثلًا) وفى بعض الاختبارات يكون تقريبى.

فعلى سبيل المثال يستخدم اختبار Z لاختبارات فروض حول متوسط التحصيل فى مجتمع ما ويكون صيغته كالآتى:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$$

- $\bar{X}$ متوسط العينة
- μ القيمة المفترضة لمتوسط المجتمع
- σ الانحراف المعيارى للمجتمع
- N حجم العينة

لاحظ أنه يجب أن يكون فى المقام الانحراف المعيارى للمجتمع ولكن يستخدم بدلا منه S وذلك للعينات الكبيرة. وإحصاء Z يتبع التوزيع الاعتدالى المعيارى ووضع الإحصائيون جداول إحصائية لاستخدامها فى الممارسة العملية.

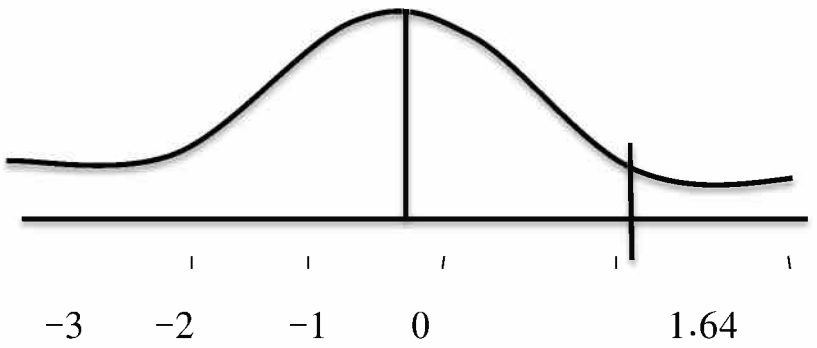
وعلينا فى هذه الحالة تحديد مدى قيم Z التى تقودنا لرفض الفرض الصفرى أو بكلمات اخرى تحديد منطقة الرفض Rejection region او منطقة الدلالة Significance region وكل القيم الأخرى التى تشير إلى قبول الفرض الصفرى وتعرف بانها منطقة قبول Acceptance region . والقيمة Z التى تفصل بين منطقة القبول ومنطقة الرفض تسمى القيمة الحرجة Critical value. وبعد تحديد منطقة الرفض يقوم الباحث بحساب قيمة إحصائية الاختبار (المحسوبة) فى ضوء بيانات العينة وإذا وقعت داخل منطقة الرفض وبالتالى نرفض H0 واذا وقعت داخل منطقه

القبول يفشل الباحث فى رفضH0. ولكن كيف يمكن تحديد منطقة الرفض ؟ فى هذه النقطة تلعب مستوى الدلالة الإحصائية دورا هامًا التى سبق تحديدها فى الخطوة الثانية، فاذا كان α=0.05 فى هذه الحالة الباحث يقسم التوزيع للاختبار إلى منطقة رفض تغطى 5% من القيم والمنطقة الأخرى تغطى 95% من القيم (منطقة قبول H0) فإذا كان الفرض الإحصائى:

$$H0: \mu \leq 3, \mu = 3,$$

$$HA: \mu > 3$$

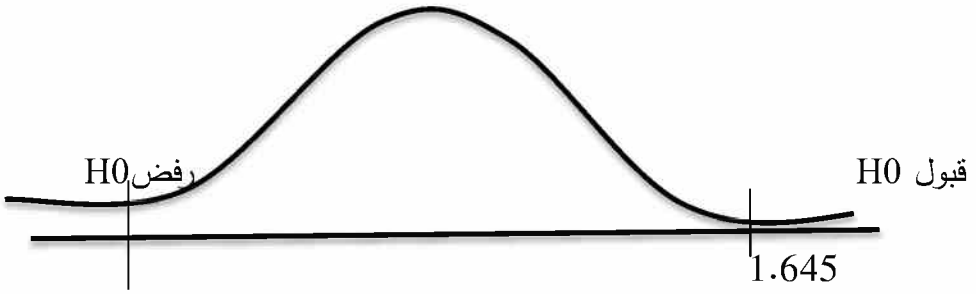
وبالكشف فى جدول Z لاختبار ذو ذيل واحد لأعلى (اليمين) وبمستوى دلاله إحصائية α=0.05 فإن Z الحرجة (الجدولية)= 1.645. وعليه فإن 5% من القيم تزيد عن 1.645 و 95% من القيم تقع أقل من 1.645 وعليه فاحتمال ان إحصائية الاختبار المحسوبة تزيد عن 1.645 عندما يكون الفرض الصفرى حقيقى هى 0.05 (مستوى الدلالة الإحصائية α) وعليه يمكن صياغة قاعدة القرار كالآتى:



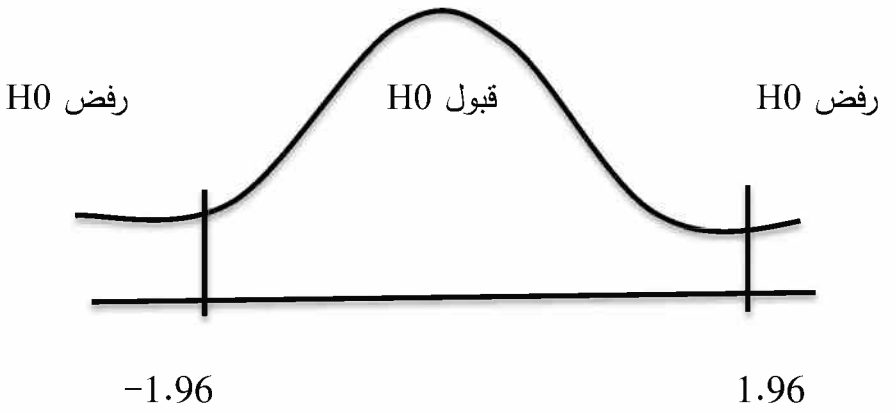
واذا كانت القيمة المحسوبة للاختبار ≤ القيمة الحرجة للاختبار، فإننا نرفض H0 بمعنى الحصول على دلاله إحصائية والعكس قبول H0 أى عدم دلاله إحصائية، وهذه القاعدة لمعظم الاختبارات الإحصائية (r, F, T) ولكن هذه القاعدة تكون العكس حيث إذا كانت القيمة الجدولية > القيمة المحسوبة يتم رفض الفرض الصفرى للقليل من الاختبارات اللابارامترية مثل مان– ويتنى و ويلكوكسون. وإذا تغيرت مستوى الدلالة الإحصائية من 0.05 إلى 0.01 سوف تتغير قاعدة القرار حيث تتغير قيمة z الحرجة إلى 2.56، ولكن ليس من المرغوب ان تغير مستوى الدلالة الإحصائية بعد تحليل البيانات، فاذا ثبت عدم

دلالتها عند 0.01 فيغيرها الباحث إلى 05. على أساس يتم رفض H0 عندها. وإذا كانت الفروض الإحصائية: $H_A: \mu < 3$, $H0: \mu = 3$

فان الاختبار ذو ذيل واحد لأدنى (لليسار) وبالكشف فى جداول z فإن $z = 1.64$ ، وعليه فمنطقة الرفض والقبول H0 كالآتى :



وهى مثل الاختبار ذو ذيل واحد لأعلى ولكن الفرق هو ان القيم الكبيرة لإحصائية الاختبار المحسوبة فى منطقة الرفض سوف تكون أصغر وليس اكبر. وإذا كان الفروض الإحصائية: $H0: \mu = 3$ ، $H_A: \mu \neq 3$، فإن الفرض البديل غير موجه أو ذو ذيلين وبالكشف فى الجدول z للاختبار ذو ذيلين $\alpha = 0.05$ فإنها تكون الحرجة z $= \pm 1.96$



لذلك توجد منطقتين للرفض كلًا منهما تشغل0.025 (25%) منطقة رفض تحت المنحنى. ولاحظ توجد قيمتين حرجتين هى أكبر من 1.96 و 1.96 - . وعلى ذلك فإن تحديد الفرض البديل يلعب دورا كبيرًا فى تحديد منطقى القبول والرفض H0، وكما هو واضح ان رفض الفرض الصفرى فى الاختبار ذو ذيل واحد اسهل من رفضه لاختبار ذو ذيلين ولذلك فإن الاختبار ذو ذيل واحد اقوى إحصائيًا من الاختبار ذو

ذيلين لأن الاختبار ذو ذيل واحد يرفضH0 عندما يكون الفرق بين العينة والمجتمع صغير نسبيًا، والاختبار ذو ذيلين يتطلب ان يكون الفرق كبيرًا نسبيًا وإذا كان الاختبار ذو ذيلين يعطى دلاله إحصائية فإن نظيرة ذو ذيل واحد يعطى دلاله إحصائية أيضًا ولكن إذا كان الاختبار ذو ذيل واحد يعطى دلالة إحصائية فإن الاختبار ذو ذيلين ليس بالضرورة ان يعطى دلاله إحصائية.

وتغير الفرض من فرض استكشافى غير موجه إلى فرض توكيدى موجه ذو ذيل واحد بعد جمع وتحليل البيانات غير مقبول خاصة عند اشتقاق نظرية. ولكن فى معظم التقارير البحثية لا ترى عبارة مثل "القيمة المحسوبة 6.3 أكبر من القيمة الحرجة 1.96وعليه فهى داله إحصائيًا ولذلك يتم رفض الفرض الصفري" ولكن فى تقارير البحوث يتم عرض القيمة المحسوبة للاختبار فقط وعليه فإن الطريقة الثانية لتقيم الدلالة الإحصائية للاختبار هى مقارنة القيمة الاحتمالية p بقيمة معينة تقع فى المدى من 0.0 إلى 1.00 ولكن هذه القيمة المعينة تكون قيمة مستوى الدلالة الإحصائية α وقيمة p تظهر فى مخرجات برامج الكمبيوتر وهى احتمالية الرفض للفرض الصفرى وهى مشابهة لمستوى الدلالة الإحصائية.

على ذلك فإن برامج الكومبيوتر يعطى قيمة p مصاحبة للقيمة المحسوبة للاختبار ويتم القرار فى ضوء مقارنة قيمة p بقيمة مستوى الدالة الإحصائية α فاذا كانت:

$$P \leq \alpha$$

يرفض الفرض الصفرى بمعنى وجود دلاله إحصائية وإذا كانت:

$$P > \alpha$$

نقبل الفرض الصفرى.

وهذا المدخل هو الشائع الاستخدام فى تقارير البحوث العلمية.

خامسًا: جمع البيانات والحسابات: فى هذه الخطوة يقوم الباحث بجمع البيانات من العينة ثم يقوم بحساب القيمة المحسوبة أو إحصائية الاختبار. وقبل حساب إحصائية الاختبار لا بد أن يتأكد الباحث من صحة إدخال البيانات ويقوم بتنظيفها Data cleaning واعدادها للمعالجة الإحصائية والتحقق من مسلمات الاختبار.

ومعادلة حساب كل اختبار تختلف من اختبار لآخر. ففى المثال الحالى لاختبار z قام الباحث بجمع بيانات تحصيل عينة مكونة من 200 طالب وقام بحساب :$\bar{X}$=3.5، S = 2.4 اذا: Z= 2.94

سادسًا: القرار والتفسير: فى نهاية عملية اختبارات الفروض امام الباحث بديلين فيما يخص H0 اما رفضه (وجود دلالة إحصائية) أو الفشل فى رفضه (عدم وجود دلالة إحصائية). ويشير الباحث إلى وجود فروق ذات دلالة جوهرية أو عدم وجود فروق حقيقية أو النتائج غير دالة إحصائيًا كلها مصطلحات يستخدمها الباحث. ويتم صناعة القرار مما يخص H0 فى ضوء مدخلين:

- مقارنة القيمة المحسوبة للاختبار بنظيرتها الحرجة: القيمة المحسوبة (2.94) > القيمة الحرجة (١.٦٤٥) وعلية يتم رفض H0 بمعنى أن متوسط التحصيل يزيد أو أكثر من 3 إذا كان الفرض البديل 3 < μ.

- حساب القيمة الاحتمالية للاختبار الإحصائى وهى قيمة p عندما يتم تحليل بيانات العينة باستخدام الكومبيوتر لتقيم الفرض الصفرى تستخدم قيمة p الاحتمالية والقيمة الصغيرة لـــ p تساعد على رفض H_0. فى هذا المثال فإن قيمة p المرتبطة باختبار z هى 0.0016:

$$0.0016 \leq 0.05$$

وبناء على ذلك يتم رفض H0 . وتكتب فى تقرير البحث كالآتى:

Z = 2.94, P.0.05 على ذلك فكلمة دال تعنى كبير أو هام أو يستحق الاهتمام ولكن فى اطار اختبارات الفروض فإنها لها معنيين مختلفين فتشير إلى أنه من غير المحتمل ان يحدث إذا كان H0 حقيقى، أو بيانات العينة غير متسقة مع ما توقعنا فى الفرض الصفرى ولكن هذا لا يعنى أن النتائج فى غاية الأهمية أو هامة Important أو وجود فروق حقيقية بين بيانات العينة و H0 بمعنى فروق كبيرة large.

والنتائج الدالة تكون نتيجة:

- جودة أسئلة البحث بمتغيراته التى تكون بمثابة الدافع للبحث.

- جودة تصميم البحث تكون بمثابة المرشد لجمع البيانات ولكن ليست كل النتائج الدالة إحصائيًا تستحق الأهمية أو تكون ذات قيمة عملية كبيرة، فعلى سبيل المثال افترض الباحث وجود علاقة بين متغيرين أى أن H0:ρ=0 ولكن بعد جمع البيانات وتحليلها تم رفض الفرض الصفرى حيث إن (r = 0.03) دالة عند مستوى 0.001 وفى هذه الحالة فلا نعتبر قيمة معامل الارتباط تستحق الانتباه حيث انها تكاد تكون صفر حيث التباين المفسر لها 0009. ومع ذلك فهى ذات دلالة إحصائية تافهة لا تستحق الوقوف عندها وهذه الدلالة جاءت نتيجة حجم العينة مثلا 20000.

بعض المحاذير

بعد عرض الخطوات الستة لإجراءات اختبارات الفروض فعليك ان تراعى الآتى:

- ان تبدأ الدراسة البحثية بجمع الادلة (البيانات) لمعرفة ما إذا كانت للمعالجة اثر من عدمه وفى هذا الاطار تبدا المحكمة بجمع الادلة لمعرفة حقيقة ما كان المتهم ارتكب "الجريمة". لاحظ أن البحث والمحكمة يحاولوا رفض الفرض الصفرى (المتهم برئ).

- إذا توافرت ادله كافية فالباحث يرفض الفرض الصفرى ويصل إلى نتيجة وهو ان للمعالجة أثر، وعليه إذا توفرت ادله كافية للقاضى فإنه يرفض H0 (المتهم برئ) ويصل إلى استنتاج ان المتهم مذنب فى ارتكاب الجريمة.

- اذا لم تتوفر ادله كافية فالباحث يفشل فى رفض H0 ويصل إلى استنتاج ان المعالجة ليس لها اثر وبالمثل ذا لم تتوافر الأدلة الكافية فإن القاضى يفشل فيه توجيه اتهام للمتهم ويبرئه.

مدخل الخطوات السبعة لاختبارات الفروض: تقدير حجم التأثير

فبالإضافة للخطوات الستة السابقة حيث ينهى الباحث هذه الخطوات اما برفض أو قبول H0 فان الباحث يرجع مرة اخرى إلى بيانات بحوثه لإجراء مهمة اضافية هو الذهاب ابعد من صناعة قرار حول H0 فى سبيل ايضاح بعض الشيء عن درجة أو قوة العلاقة أو حجم الفروق أو العلاقة بين المتغيرات.

لماذا الخطوة السابعة؟ قد يصل الباحث إلى دلالة إحصائية على الرغم من صغر حجم الفروق أو حجم العلاقة أى يصل الباحث إلى دلاله إحصائية بدون وجود للدلالة العملية أو الأهمية التطبيقية Practical significance وذلك لأن أحجام العينات الكبيرة تزيد من احتمالية رفضH0 على الرغم ان بيانات العينة تبتعد عن H0 بمقدار ضئيل جدًا وعلية تم قبول H_A بدون وجود أى قيمة عملية تطبيقية وعلى ذلك فلا بد للباحث ان يذهب خطوة ابعد من الدلالة الإحصائية وهى التعرف على القيمة العملية لهذه الدلالة.

تقدير حجم التأثير: الباحثون الذين يميزون بين الدلالة الإحصائية والدلالة العملية دائما يضيفون خطوة سابعة للصيغة الأساسية لاختبارات الفروض وهى تقدير حجم التأثير، ففى المقارنة بين طريقتين للعلاج يتم رفض الفرض الصفرى فإن حجم التأثير يسمح لنا بمعرفة ماذا كان هذا التأثير الفارق للمعالجة صغير أو متوسط أو كبير، وفى دراسة العلاقة بين متغيرين ويصل الباحث ان معامل الارتباط مرتفع يختلف عن الصفر فإن تقدير حجم التأثير يسمح لنا بالحديث عن قوة العلاقة المقاسة، بكلمات أكثر بساطة فإن حجم التأثير يقدم لنا رؤية وتفسير آخر فيما وراء الدلالة الإحصائية. وعند الحديث عن حجم التأثير لا بد الاخذ فى الاعتبار مفهوم الدلالة العملية ولا بد أن تراعى قضيتين أساسيتين:

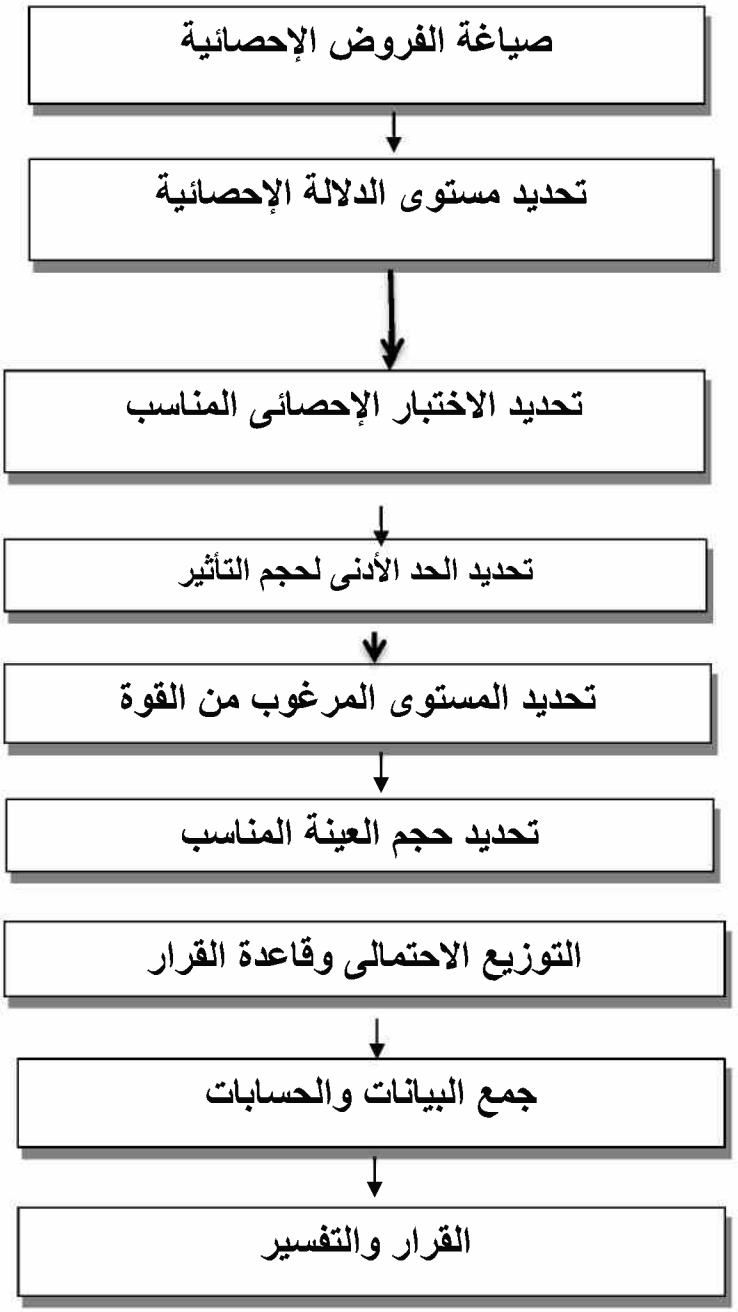
الأولى: عندما يقارن الباحث إحصائية بيانات العينة بالمعلمة فى الفرض الصفرى H0 فإن ذلك يتم فى ضوء القيمة الاحتمالية p ولكنها لا تعتبر هى الفيصل فى كل شئ ولكن لا بد من التركيز على إلى أى درجة بيانات العينة تبتعد أو تنحرف عنH0، فعلى سبيل المثال إذا تناولت دراسة العلاقة بين متغيرين وكان الفرض الصفرى H0: $\rho = 0$,فالباحث يحاول النظر إلى كم درجة قيمة r فى العينة تبتعد

عن 0.0، وفى دراسة الفروق بين مجموعتين فإن الفرض الصفرى $H0:\mu_1=\mu_2$ والباحث يفحص بيانات العينة لمعرفة إلى أى مدى الفرق بين متوسطى x_1 و x_2 يختلف عن الصفر فى الفرض الصفرى.

الثانية: الباحث يقيم الفروق المقاسة بين ملخص أو إحصائية بيانات العينة وما يفترضه فى H0 فان تحليل حجم التأثير يتضمن الحكم ما إذا كان حجم الفروق المقاسة $(\bar{X}_1-\bar{X}_2)$ فى العينة صغيرًا Tiny أو متوسط Medium أو كبير Large or Gigantic ولتحديد ذلك فإن كثير من الباحثين يضعون معايير ومحكات يوصى باستخدامها من قبل المؤسسات والهيئات الإحصائية. ولكن عند استخدام معايير موحدة ومعروفة لا بد أن يتوخى الباحث الحذر لأن هذه المعايير فى تفسير حجم التأثير ليست مناسبة لكل التخصصات ولكل المواقف البحثية ولذلك فيجب على الباحثين فى تحديد معايير حجم التأثير الصغير والمتوسط والكبير ان يكون لدية معرفة متعمقة بالنظرية فى دراسته أو بالدراسات السابقة. وعلى ذلك فبإدخال حساب حجم التأثير فى خطوات اختبارات الفرض فإنها تتكون من سبعة خطوات ويكون حجم التأثير الخطوة السابعة.

مدخل الخطوات التسعة لاختبارات الفروض Huck (2012)

على الرغم ان الكثير من الباحثين يتبعون مدخل الخطوات الستة وكذلك السبعة خطوات لاختبارات الفروض الصفرية الا أنه يوجد اتجاه ثالث يؤيد مدخل التسعة خطوات حيث يرى البعض ان خطوات اختبارات الفرض ذو التسعة خطوات وهى شاملة للمكونات الأساسية ويؤيد البعض إدخال ثلاث خطوات اخرى مرتبطة وهى حجم العينة والقوة وحجم التأثير وفيما يلى ترتيب هذه الخطوات:



الشكل (٤.٩): مدخل الخطوات التسعة لاختبار الفروض.

فالخطوات الثلاثة الأولى والثلاثة الاخيرة فى مدخل التسع خطوات هو مماثل للخطوات الستة السابقة، وهنا يتم التركيز على الخطوات من الرابعة حتى السادسة، ومن الضرورى التعرف على الفروق بين مدخل السبعة خطوات ومدخل التسعة خطوات وخاصة فيما يتعلق بحجم التأثير. ففى مدخل السبع خطوات تم حساب حجم التأثير بعد الوصول إلى قرار وذلك بالرجوع مرة اخرى إلى البيانات أى بعد مرحلة جمع البيانات

واما حسابها فى مدخل التسعة خطوات فإن الباحث يضيف ثلاثة خطوات ومنها حجم التأثير قبل مرحلة جمع البيانات. وعلى ذلك فإن حجم التأثير تعتبر قيمة محسوبة من بيانات العينة لتحديد قوة العلاقة أو حجم الفروق فى مدخل السبعة خطوات اما مدخل التسعة خطوات تأخذ شكل الافتراض فى ضوء خبرة الباحث أو قد يتم الوصول اليها أو تقديرها فى ضوء الدراسات السابقة للظاهرة وموضع الدراسة.

الخطوة الرابعة: تقدير حجم التأثير: لاحظ أن الحديث عنها فى مدخل الخطوات السبعة ذكرنا " تقدير أو حساب حجم التأثير" اما فى مدخل التسعة فإن حجم التأثير يكون محدد مسبقًا لما سوف يرغب فيه الباحث الوصول إلى حجم تأثير متوسط أو ضعيف أو كبير فيشكل الصورة التى يريد ان يرسمها هو الذى يجعلك تختار القلم أو الالوان المناسبة لذلك.

لتحديد ما نوع حجم التأثير وكيف يمكن الحصول عليه افترض أن باحث استخدم إجراء اختبارات الفروض فى دراسة مجتمع ما حيث إن البيانات لمتغير الذكاءIQ، ويكون اهتمامه الإحصائى هو دراسة متوسط المجتمع حيث يكون: H0: $\mu = 100$, HA: $\mu > 100$ ، فى هذه الدراسة المفترضة باعتبار أن الدرجة العظمى IQ =150 فربما يقرر الباحث ان يضع 110 لحجم التأثير الخام ويعلن ان الفروق الصغيرة تقع من 100 إلى 110 وان الفرق يكون هام إذا كان على الأقل أكبر من 10 نقاط من H0 (100) قبل إجراء الدراسة.

ودائمًا يأخذ حجم التأثير صورتين هما حجم تأثير الخام Raw size effect و حجم التأثير المعيارى Astandardized effect size. وتحديد حجم التأثير الخام هو الاستراتيجية الأفضل ولكنه غالبًا صعب وبينما حجم التأثير المعيارى سهل تحديده. وتحديد حجم التأثير الخام كما فى مثال نسبة الذكاء حيث وضع الباحث حجم تأثير خام 110 وكان الفرض الصفرى مساويًا لـ 100 حيث يحدد الباحث خط واضح وفاصل بين النواتج عديمة الأهمية والنواتج التى لها أهمية على متصل المتغير التابع فى الدراسة.

وكثيرًا من الباحثين الذين يستخدمون مدخل التسعة خطوات يحدد قيمة حجم التأثير المعيارى ويشير الباحث إلى معايير يفضل الوصول إليها، وفى اختيار أحد أحجام التأثير المعيارية (ضعيف – متوسط – قوى) فإن الباحث يواجه بثلاثة أسئلة:

- ما هو حجم التأثير الحقيقى الواقعى للمتغيرات فى المجتمع الواقعى؟
- هل يريد الباحث إضافة حساسية Sensitivity أو دقة عالية فى الدراسة تسمح له للكشف عن حجم التأثير القوى (المرتفع) أو المتوسط؟ ومن المهملاى دراسة ان يكون لها الحساسية العالية ان تكشف ليس فقط عن التأثيرات الكبيرة أو المتوسطة وبل أيضًا الضعيفة. ومعايير أو حدود القطع الصغيرة والمتوسطة والكبيرة تختلف من اختبار إلى آخر، فمعاييرها لاختبار T هى 0.20 و 0.50 و 0.80 للضعيفة والمتوسطة والكبيرة على التوالى، ولمعامل الارتباط r هى 0.10 و 0.30 و 0.50 على التوالى.

وفعليك ان تعرف ان خطوات اختبارات الفروض ذو الخطوات التسعة تتطلب تحديد مسبق لحجم التأثير وللباحث اختيارات اما ان يحددها بصورة درجة خام أو بصورة معيارية. وفى هذا الشأنفان معظم الباحثين يتمنون الوصول إلى حجم تأثير مرتفع ولكن هذه العملية ليست اعتباطية حيث توجد مداخل لتحديد حجم التأثير قبل إجراء الدراسة منها:

- إجراء دراسة استطلاعية لتحديد حجم التأثير من بيانات هذه العينة.
- مسح للدراسات السابقة لتحديد حجم التأثير الشائع فى هذه الدراسات، ويتم إجراء دراسة أو مراجعة كمية للدراسات السابقة" ما وراء التحليل" للوصل إلى حجم التأثير، وفى هذا الشأن يوصى بإجراء ما وراء تحليل لخمس دراسات على الأقل تناولت متغيرات البحث.
- كثير من الباحثين يحدد حجم التأثير المتوسط حسب كل اختبار مستخدم فى الدراسة.

الخطوة الخامسة: تحديد مستوى قوة إحصائية مرغوب: فى هذه الخطوة يحدد الباحث مستوى قوة إحصائية معينة مرغوب لرفضH0 وكما نعلم ان القوة الإحصائية هى مقدرة الاختبار الإحصائى للوصول أو للكشف عن الدلالة الإحصائية وهى موجودة بالفعل فى واقع الظاهرة، وتتراوح قوة الاختبار من صفر إلى الواحد الصحيح والقيم العالية تؤخذ فى الاعتبار وتعتبر مؤشرًا هامًا لجودة الدراسة ذلك لأن قوة الاختبار تكون متلازمة للخطأ من النوع الثانى ويوصى الخبراء بان الحد الأدنى للقوة الإحصائية الواجب توافرها هى 0.80 (Cohen,1988) حيث يكون للاختبار فرصة 80% لرفض الفرض الصفرى والكشف عن الفروق الحقيقية ولاحظ أن مفهومى القوة الإحصائية و α مرتبطين بالأخطاء الاستدلالية التى يمكن أن تحدث، فألفا يحدد احتمال الوقوع فى الخطأ من النوع الأول بينما القوة يحدد احتمال عدم ارتكاب الخطأ من النوع الثانى ودائما يشار إلى α بـ Type I Error risk ويشار للقوة Error risk Type II ولذلك فدائمًا فى تقارير البحوث يجب كتابة مثلًا" لاختبار ذو ذيلين وبقوة إحصائية حددت مسبقا بـ 80. عند مستوى دلالة إحصائية 05. توصل الباحث ..). فالقوة العالية فى الدراسة تكون أفضل من القوة المنخفضة ولذلك يوجد سببين يجعل الباحثين يختاروا تحليل قوة قبلى عنده 80. وليس 90. فى الدراسة النفسية والتطبيقية:

1. وضع مستويات عالية للقوة الإحصائية يضع متطلبات غير معقولة على الباحثين خاصة عند حساب حجم العينة فى الخطوة السادسة فلمستوى قوة 90. يتطلب أحجام عينات كبيرة جدًا.

2. مستويات قوة 95. أو 99. تزيد من احتمال الحصول على دلالة إحصائية حتى لو كان الفرق أو ابتعاد إحصائية العينة عنH0 ضئيل جدًا، حيث يتم رفض H0 وهو فى الحقيقة صحيح "تضخم الخطا من النوع الأول".

الخطوة السادسة: تحديد حجم العينة المطلوب لإنجاز الدراسة: فبعد تحديد الفروض الإحصائية وتحديد مستوى الدلالة الإحصائية وكذلك حجم التأثير والقوة المرغوبين، فإن الباحث يستخدم صيغ ومعادلات معينة أو برامج كومبيوتر باستخدام معلومات مسبقة لتحديد حجم العينة المطلوب ويسمى فى هذه الحالة مدخل القوة قبل إجراء الدراسة Prospective or prior power analysis وذلك بهدف تحديد حجم العينة وهذه

الخطوة هى مرحلة أساسية فى تصميم الدراسة هذا يفيد فى تقليل التكلفة والموارد حين يزيد الباحث حجم العينة بدون داعى ويوصى Cohen (1992) بأهمية إجراء هذا النوع من تحليل القوة قبل إجراء الدراسة، ويعتبر هذا أحد المداخل المستخدمة لتحديد حجم العينة ويشير الباحث إلى العبارة الآتية فى الدراسات "اوضح التحليل القبلى للقوة أنه لتحقيق مستوى قوة 0.80 مع 0.05 > P وحجم تأثير متوسط 0.30 مع ثلاثة متغيرات منبئة ان حجم العينة المطلوب 41 فردًا مثلًا. ومستوى قوه 0.80 يشير إلى ان احتمالية ارتكاب الخطأ من النوع الثانى منخفضة وهى 0.20.

واحد مميزات مدخل التسع خطوات هى تفسير النتائج فى ضوء قواعد ومعايير تم تحديدها لتقدير حجم العينة وكذلك الاقتصاد فى الوقت والتكلفة حيث تحديد حجم العينة بدون مدخل القوة قد يتطلب حجم عينة كبيرًا لا داعى له وفى هذا فإن الاختبار يصل إلى رفض الفرض الصفرى حتى لو كانت الفروق أو العلاقات ضعيفة جدًا.

واذا لم يعطى اعتبار لمفهومى حجم التأثير والقوة الإحصائية فإن الفشل فى رفض H0 يعزى لانحرافات تافهه عن H0 وقد يعزى إلى عدم حساسية الاختبار للكشف عن الفروق الحقيقة نتيجة لصغر الحجم العينة أو ان رفض H0 نتيجة فروق ضئيلة ويعطى الاختبار دلالة الإحصائية لاوجود لها على مستوى حقيقة الظاهرة لأن حجم العينة كبيرًا. واحيانًا يذكر الباحثون العبارة الآتية "المحدد الاكبر فى هذه الدراسة صغر حجم العينة وان صغر حجم العينة يقلل من قوة الاختبار الإحصائى ولذلك فإن النواتج الإحصائية تكون معرضة للخطأ من النوع الثاني"، وعلى الرغم ان قيمة التأثير صغيرة جدًا ولكنها داله إحصائية فى ضوء حجم العينة الكبيرة (2000) مثلا وهذا التأثير غير دال عمليًا. وكذلك يستطيع الباحث ان يعرف ويضبط احتمالية ارتكاب الخطأ من النوع الثانى ويصل الباحث إلى نتائج دالة إحصائية دون وجود انتقادات لحجم هذه الدلالة أو حجم العينة المستخدم وقد يستغرق الباحث وقتًا فى تصميم الدراسة لتحديد حجم العينة ولكن هذا يعطى الدراسة قدرًا كبيرًا من الموثوقية فى نتائجها.

مشاكل اختبارات الدلالة الإحصائية

يعتقد الباحث ان مستوى الخطأ من النوع الأول يظل ثابتًا ولكن بتعبير أكثر صدمة فإن نسبة الخطأ أكبر من 5% الذى يتبناها الباحث، فالكثير يعتقد ان استخدام اختبارات الدلالة كفيل ان يجعل نسبة الخطأ من النوع الأول عند 5 % أو أقل وهذا غير حقيقى(Hunter & Schmidt, 2004). فلو ان الفرض الصفرى حقيقى فى المجتمع(لا علاقة أو لا فروق) وتم رفضه فى ضوء بيانات العينة، فى هذه الحالة تم الوقوع فى الخطأ من النوع الأول(α) Type 1 error، وإذا كان الفرض الصفرى غير حقيقى على مستوى المجتمع وتم قبوله على مستوى بيانات العينة بالتالى تم الوقوع فى الخطأ من النوع الثانى(β) Type 11 error.، واختبارات الدلالة الإحصائية تحاول بقدر الامكان ان تجعل الخطأ من النوع الأول تقريبًا عند 5% ولكن هذا صعب المنال ولكن الشىء الوحيد الضامن لهذا هو هجر أو الابتعاد عن استخدام اختبارات الدلالة الإحصائية والاعتماد على استخدام فترة الثقة Confidence intervals كبديل لاختبارات الفروض الصفرية. وتوجد عدة انتقادات للدلالة الإحصائية أهمها:

- **حساسيتها لحجم العينة Sensitivity to sample size** : عندما يكون حجم العينة صغيرًا فإن التأثيرات القوية لا تكون دالة إحصائيًا(ارتكاب الخطأ من النوع الثانى)، فى المقابل عندما تكون حجم العينة كبيرًا فإن التأثيرات الصغيرة تكون دالة إحصائيًا(ارتكاب الخطأ من النوع الأول)، فمثلًا إذا كان معامل الارتباط r (20)=.40 غير دال إحصائيًا عند $P \geq 0.05$ **لاختبار** ذو ذيلين، بينما r =(1000) .07 دال إحصائيًا. وعلى الجانب الآخر إذا كانت .25 = r فإن النتائج دالة إذا كان حجم العينة n=63 وغير دالة إذا كان n=61 وعليه فإن هذا يؤدى إلى استنتاجات خاطئة.
- لا تخبر الباحث عن ماذا يريد ان يعرفه ولكنها تخلق إضاءة عن برهان احتمالى للتناقض بين الفرض الصفرى والفرض البديل.
- **نسب الخطأ والقوة Power and error rate**: الانتقاد الأساسى لاختبارات الفروض الصفرية النسبة العالية للخطأ من النوع الثانى غير المقبولة Schmidt,

(1996)، وهذا يؤدى إلى نقص القوة الإحصائية المناسبة(Cohen, 1962). والملاحظ فى التراث البحثى فى التربية وعلم النفس ان متوسط القوة فى معظم التراث البحث فى علم النفس تراوح من 0.40 إلى 0.60 بل احيانًا 0.20 فى بعض التخصصات (Hunter, 1997 , Schmidt, 1996; Sedlmeier & Gigerenzer, 1989). وتوصل Maxwell (2004) إلى ان وسيط القوة فى الدراسات غير التجريبية فقط حوالى 50. وهذا مفاده ان فرصة 50% لرفض صحيح للفرض الصفرى فى ضوء بيانات العينة و50% رفض خاطئ للفرض الصفرى. ويشير Kline(2013) فى هذه الحالة من الأفضل عدم إجراء البحث وجمع البيانات والاعتماد على مبدأ ملك أو كتابة للعملة لتحدد رفض أو قبول الفرض الصفرى.

وتوصل عامر(2014) إلى أن وسيط القوة للدراسات النفسية التربوية العربية ذات حجم التأثير الصغير 0.28 ولحجم التأثير المتوسط0.90، ولحجم تأثير كبير 0.99 و فى المجمل فإن متوسط القوة الإحصائية 72. . وفى المجمل فى التراث الأجنبى والعربى انخفضت القوة الإحصائية عن الحد الأدنى المناسب للقوة الإحصائية وهو 0.80(معيار Cohen)، ومن أهم العواقب المترتبة على الاعتماد على قوة منخفضة هو عدم القدرة على تفسير نتائج التراث البحثى لظاهرة ما، فاذا وجد تأثير حقيقى للظاهرة موضع الدراسة ولكن القوة الإحصائية 0.50 فإن هذا يعنى ان نصف الدراسات تعطى نتائج دالة بينما النصف الآخر لا يعطى دلالة إحصائية، ولو نشرت هذه الدراسات فإن نصفها يعطى نتائج ايجابية والنصف الآخر نتائج سالبة وقد عبر Simmons, Nelson,& Simonsohn (2011) عن ذلك بعبارة " علم النفس الموجب والخطأ".

وعليه فالاعتماد التقليدى على اختبارات الدلالة الإحصائية فى تفسير الدراسات يقود إلى تفسيرات خاطئة عما تعنيه نتائج الدراسات. ويرى Schmidit (1996) ان استخدام المدخل التقليدى فى تحليل البيانات يجعل من غير المحتمل الوصول إلى استنتاجات صحيحة فى معظم التخصصات. ويشير Cohen (1990) إلى ان العلوم الطبيعية

مثل الفيزياء والكيمياء لا تستخدم اختبارات الدلالة الإحصائية فى تفسير بياناتها. ويعتقد الباحثون أنه إذا كانت النتيجة دالة عند 0.05 فإن احتمالية تكرار هذه النتيجة فى الدراسات اللاحقة هى 1- 05.= 95. وهذا الاعتقاد خاطئ تمامًا. ولكن احتمال الاعادة فإن القوة الإحصائية للدراسة تقريبًا أقل من 95.

الاستنتاجات الخاطئة بعد صناعة القرار عن الفرض الصفرى

توجد استنتاجات خاطئة بعد رفض أو الفشل فى رفض الفرض الصفرى وهى كالتالى (Kline, 2013):

- **اخدوعة الاحتمالية فى مقابل الصدفة Odds against Chance Fallacy**: يوجد اعتقاد خاطئ بان القيمة الاحتمالية P تشير إلى أن نتائج الدراسة تحدث نتيجة خطأ المعاينة، ولذلك 0.05 > P تعنى يوجد احتمال أقل من 5% أن نتيجة الدراسة تكون للصدفة ويوجد مفهوم خاطئ هو ان قيم P تصنف النتائج إلى تصنيفين اما إلى نتيجة تكون للصدفة (قبول الفرض الصفرى) ونتيجة تكون للتأثيرات الحقيقية الواقعية (رفض الفرض الصفرى) وهذه الاعتقادات خاطئة. ولكن توجد عوامل اخرى تلعب دورًا فى تفسير وحساب قيمة P فى إطار التصميم البحثى للدراسة وعملية القياس وأساليب التحليل الإحصائى. وكثيرًا من الباحثين يعتقدون ان 0.05 > P لـ α=0.05 تعنى ان احتمال القرار اللازم لرفض الفرض الصفرى هو ان الخطأ من النوع الأول أقل من 0.05.
- **اخدوعة الاعادة Replicability**: يوجد اعتقاد خاطئ ان النتائج الدالة إحصائيًا هى قابلة للإعادة إذا اجريت الدراسة على عينات مماثلة اعتقادًا ان 05.> P تعنى ان احتمالية الاعادة أكبر من 95%، ولكن Greenwalw, Gonzalez, Harris, & Guthrie (1996) وجدوا علاقة منحنية بين قيم P ومتوسط القوة الإحصائية لدراسات عشوائية تم اعادتها على دراسات اخرى مماثلة لها بنفس العدد من أحجام العينات وهذا يعنى احتمال الاعادة لنفس النتائج تقريبًا 50%.
- **اخدوعة الحجم Magnitude Fallacy**: الاعتقاد بان القيمة المنخفضة لـ P تشير إلى تأثير كبير أو قوة المعالجة كبيرة حيث يصف الباحث النتائج التى P < α

بانها دالة إحصائيًا بالتالى فإن النتائج الدالة عند 0.01 أكثر حجمًا من النتائج الدالة عند 0.05 وهذا الاستنتاج غير مضمون لأن هذه الدلالة تتأثر بحجم التأثير وحجم العينة لأنه يمكن الحصول على قيم صغيرة جدا لـ P لأحجام العينات الكبيرة بغض النظر عن حجم الاثر أو قوة العلاقة.

- **اخدوعة المعنوية Meaningfulness fallacy Fallacy:** فى ضوء اخدوعة المعنى يعتقد الباحث ان رفض الفرض الصفرى يؤكد الفرض البديل وهذا يعكس خطأين. الأول رفض H_0 فى الدراسة لا يعنى تحقق ان الفرض البديل، الثانى حتى لو ان الفرض البديل الإحصائىH_1 صحيح فإن هذا لا يعنى الفرض الجوهرى من وراء H_1 حقيقى. فالدلالة الإحصائية لا تبرهن على صحة فرض معين، فبعد رفض أو إبقاء الفرض الصفرى لا يوجد امام الباحث الكثير ليفعله سوى اصدار التوصية.
- **اخدوعة الصفر واخدوعة التكافؤ Zero Fallacy and Equivalence Fallacy:** يوجد اعتقاد خاطئ بان الفشل فى رفض الفرض الصفرى يعنى ان حجم التأثير فى المجتمع يساوى صفر. بينما اخدوعة التكافؤ تحدث عند الفشل فى رفض الفرض الصفرى H_0 :μ1=μ2 فإنه يفسر على أساس ان المجتمعين متكافئين ولكن هذا اعتقاد خاطئ لأنه يمكن أن يحدث اختلاف بينهما فى التباين أو شكل التوزيع.
- **أخدوعة الجودة وأخدوعة النجاح Quality Fallacy and Success Fallacy:** الاعتقاد بأن الحصول على الدلالة الإحصائية مؤشر لجودة التصميم التجريبى أو البحثى ويشير أيضًا إلى نجاح الدراسة، فالدراسة ضعيفة التصميم أو التى تتضمن خطأ معاينة يمكن أن تقود إلى رفض غير صحيح لـ H_0 (الخطأ من النوع الأول)، وأحيانًا الفشل فى رفض الفرض الصفرى يقود إلى تولد أفكار ومناقشات جيدة خاصة عندما لا يتم تدعيمه من الباحثين الآخرين فى المجال وعليه وجود علم جيد وعليه نقص النتائج الايجابية من الدراسات يكون أكثر فائدة فى تقدم العلم.
- **اخدوعة الفشل Failure Fallacy:** يوجد اعتقاد خاطئ بان نقص الدلالة الإحصائية مؤشر لفشل الدراسة فى تحقيق اهدافها ولكن استخدام الطرق التحليلية

غيـر المناسـبة أو القـوة الإحصـائية المنخفضـة تسـبب الخطـأ مـن النـوع الثـانى ويعـزو الباحث الفشل فى رفض الفرض الصفرى إلى تصميم الدراسة غير الجيد.

- **اخدوعـــة التنقـــيح Reification Fallacy**: الاعتقـاد الخـاطئ بـان الفشـل فـى الحصـول علـى نفـس النتـائج مـن دراسـات الأخـرى لـنفس المتغيـرات هـو فشـل فـى الوصـول لـنفس القـرار عبـر دراسـات اخـرى، فالنتيجـة لا تعتبـر قابلـة للتكـرار لـو ان H_0 رفـض فـى الدراسـة الأولـى ولـيس كـذلك فـى الدراسـة الثانيـة، وهـذه الاشـكالية تتجاهـل حجـم العينـة، حجـم التـأثير، والقـوة عبـر الدراسـات المختلفـة. ومهـا اختلفـت نتـائج الدراسات حول H_0 فإن هذا فى حد ذاته دليل للإعادة Replication .
- **اخدوعـة الموضـوعية Objectivity Fallacy**: الاعتقـاد بـان اختبـارات الدلالـة هـى طريقـة موضـوعية لاختبـارات الفـروض وكـل طـرق الاسـتدلال الأخـرى ذاتيـة هـى اضـحوكة غيـر مقبولـة، فهـى ليسـت الإطـار الوحيـد فقـط لاختبـار الفـروض بـل يوجـد التقـدير البيزانـىBayesian estimation وهـى بـديل جيـد لاختبـارات الدلالـة وكـذلك فترت الثقة.
- **اخدوعـــة الثنائيـــة Sanctification Fallacy**: وتشـير إلـى اشـكالية التفكيـر التصـنيفى الثنـائى عـن متصـل قـيم P. فمـاذا لـو α= 05. وكانـت P= 0.049 فـى مقابـل P =0.051 عمليًـا متمـاثلين فـى ضـوء نـواتج الاختبـار، فالباحـث يـولى اهتمـام للأولـى (دالـة) ويتجاهـل الثانيـة(غير دالـة)، وهـذا التغيـر فـى درجـة الثقـة نتيجـة تغيـر علـى متصـل قيمـة P وعليـة القـيم الكبيـرة لـ P تنـاظر تغيـرات صـغيرة وغيـر دالـة فـى البنـاء التحتـى للمتغيـرات. لكـن قـيمP الأقـل مـن 05.٠ مثـل 048. أو 049. فهـى توصف بانها غير دالة (Kline, 2013).
- **أخدوعـــة المناعـــة Robustness Fallacy**: الاختبـارات الإحصـائية المعمليـة البارامتريـة ليسـت لـديها مناعـة أو ضـلاعة ضـد القـيم المتطرفـة أو عـدم تحقـق مسـلمات خصـائص التوزيـع مثـل الاعتداليـة خاصـة للعينـات الصـغيرة، ولكـن كثيـر مـن البـاحثين يعتقـدون العكـس والبـاحثين لا يعطـوا مؤشـرات علـى مـدى تـوافر مسـلمات الاختبـار المستخدم (Onwuegbuzie, .2002)

وعليه الاعتماد على الدلالة الإحصائية فى تفسير الدراسات يقود إلى استنتاجات خاطئة عن ماذا تعنى نتائج الدراسة، وفى الواقع فإن المدخل التقليدى (اختبارات الفروض الصفرية) فى تحليل البيانات يجعل من غير المحتمل الوصول إلى استنتاجات صحيحة فى معظم مجالات البحث (Schmidt, 1996).

حقيقة حدثت نقاشات كثيرة منها المؤيد لاختبارات الدلالة بعضها معارض لاستخدامها فى إجراءات العمل الإحصائى ونتيجة لذلك قدم Kline (2013) كتاب معنون بـ"ما بعد اختبارات الدلالة: تجديد الإحصاء للعلوم السلوكية " قدم فيه عرض لمحددات الدلالة الإحصائية وبدائلها.

والتراث البحثى فى مجال العلوم السلوكية وضع جل اهتمامه على الحصول على دلالة إحصائية وأهمل جانب آخر من الدقة الإحصائية وهو القوة الإحصائية على الرغم من تأكيد الجمعية النفسية الامريكية على استخدامها(APA, 2010) وهذا التجاهل للقوة الإحصائية نتيجة القوة الإحصائية المنخفضة فى البحوث.

وفى ضوء المراجعة السردية الكيفية التقليدية تمدنا بنتائج متضاربة والوصول إلى استنتاج هو " إجراء المزيد من الدراسات السابقة" ولكن أى نتائج جديدة سوف تعزز أو تزيد من هذا التعارض والغموض لو ان القوة منخفضة. ووصف كثيرًا من الباحثين الأخطاء المعرفية المرتبطة باختبارات الدلالة الإحصائية بانها نوعا من العبث أو عدم التعلم المتدرب عليه Trained incapacity الذى يعوق الباحثين من فهم نتائج دراساتهم، والبعض الآخر يصفها بانها أساس الفشل التربوى.

ويشير Kline (2013) إلى ان الاعتماد المتزايد على اختبارات الفروض الصفرية دمر تراثنا البحثى. وتحيز النشر للأبحاث المتضمنة الدلالة الإحصائية Publication bias for significance يقترح ان النسبة الحقيقية أو الواقعية للخطأ من النوع الأول أو الرفض غير الصحيح للفرض الصفرى تكون أعلى من المستويات التقليدية لمستوى الدلالة بالإحصائية المفترضة.

وحدث جدل ومناقشات مرتبطة بمدى جدوى الاعتماد على الدلالة الإحصائية، فالبعض نادى باستبعادها من البحوث المنشورة فى المجلات العلمية وتمت مناقشة فكرة الاستبعاد فى اعداد أو إصدارات خاصة من مجلات مثل Journal of Experimental education (Thomposon, 1993) و Psychological Science (Shrout, 1997) وغيرها وكان السؤال الرئيسى:" ماذا يكون الوضع بدون استخدام اختبارات الدلالة الإحصائية؟"، وقدم Armstrong (2007) النصيحة الآتية:

الكتابة لكتاب أو تقارير بحوث فلا تذكر اختبارات الدلالة الإحصائية. والكتابة لمجلات فلا بد للبحث عن طرق لتقليل الاضرار المحتملة جراء اختبارات الدلالة، ويجب استبعاد كلمة دلالة لأن النتائج التى ترفض الفرض الصفرى ليست بالضرورة دالة وتلك التى تفشل فى رفض الفرض الصفرى ليست بالضروة غير دالة.

ونادى بعض الباحثين بهجر أو ترك الدلالة الإحصائية فى الدراسات والبحوث امثال((Schmidt, 1996; Shmidt & Hunter, 1997) ، حيث اشاروا إلى ان الدلالة الإحصائية اعاقت تقدم المعرفة العلمية ولا تقدم اسهامات ايجابية للعلم وهى خطأ رهيب واستراتجية علمية فقيرة واحدى الاشياء السيئة التى حدثت فى تاريخ علم النفس. وهى استخدام سيئ لفترة طويلة(Cohen,1994)، بل تقود إلى صناعة قرار بحثى فقير وضعيف.

وقدمت الجمعية النفسية الامريكية فى اصدارها الخامس والسادس APA (2001, 2010) عدة توصيات فيما يخص اختبارات الدلالة اللإحصائية أهمها:

- اعرض الإحصاء الوصفى مثل المتوسطات والتباينات وحجم العينة لكل مجموعة لأن هذه المعلمات ضرورية لإجراء دراسات لاحقة ثانوية مثل ما وراء التحليل.
- اعرض أحجام التأثير كلما أمكن وعدم عرضه يعتبر قصور فى الدراسة.
- من الضرورى استخدام فترات الثقة كاستراتجية لصناعة القرار.

وشدد Rosnow & Rosenthal (1989) على ضرورة تقدير حجم التأثير عند تقدير للنتائج الدالة إحصائيًا وغير الدالة إحصائيًا وهذا عكس ما يعتقده الباحثين

من أن تقدير حجم التأثير يكون للنتائج الدالة إحصائيًا فقط . واشار Cohen (1990) إلى ان الناتج الأولى للبحث هى مقياس أو أكثر لحجم التأثير وليس قيم P الاحتمالية.

علاوة على الاعتماد الكلى على اختبارات الفروض الصفرية وتجاهل القوة الإحصائية وحجم التأثير وما ينتج عنها من استنتاجات خاطئة، توجد اشكالية اخرى هى ازمة القياس Measurement crisis فى العلوم السلوكية منذ عهود كثيرة حيث تفتقر البرامج الدراسية للطلاب فى المراحل الجامعية الأولى إلى المقررات السيكومترية ولا يتم تدريب طلاب الدراسات العليا فى برامج الماجستير والدكتوراه على كيفية اعداد المقاييس بالصورة المرجوة.

وبالإضافة إلى محددات قيمة P الاحتمالية حيث تمدنا بإجابة عن السؤال البحثى فى ضوء نعم أو لا (رفض أو الفشل فى رفض الفرض الصفرى) وهذا لا يعطى متعة فى العمل العلمى ولا تخبرنا عن المناسبة العلمية أو الدلالة الإكلينيكية أو حجم التأثير وهذا ما دعا Armstrong (2007) إلى القول بان اختبارات الدلالة الإحصائية لا تساعد على التقدم العلمى.

وفى الانتقادات للدلالة الإحصائية اوصى Cumming (2012) يجب على الباحثين اعطاء قليل من الاهتمام لقيمة P والتأكيد على عينة نتائج البحوث واشار إلى مصطلح الإحصاء الجديدة New statistics وهى ليست إحصاء جديدة على الاطلاق انما تتكون أساسا من التأكيد على استخدام حجم التأثير وفترات الثقة وهذا ما اكدت عليه من قبل الجمعية النفسية الامريكية (APA, 2010) عن حجم التأثير" لمعرفة حجم أو أهمية نتائج الدراسة يجب بقدر المكان كلما كان متاح تضمين بعض مقاييس حجم التأثير" .

والدلالة الإحصائية لا تشير مباشرة إلى حجم التأثير ولان هذا المؤشر هو إحصائية للعينة فهو تقدير تقريبى لحجم التأثير فى المجتمع بينما تقدير فترة الثقة يشير إلى أى درجة خطأ المعاينة مرتبط بالإحصائية وعلى الرغم يتم تقدير خطأ المعاينة فى

اختبارات الفروض الصفرية لكن يتم اخفائه فى حساب قيمة P ولكن مقدار من خطأ المعاينة يظهر فى ضوء الحدود الدنيا والعليا لفترات الثقة.

وعرض فترات الثقة يعكس التفكير التقديرى Estimation thinking (Cumming, 2012) واوصت الجمعية النفسية الامريكية (APA, 2010) كلما كان متاح تتم مناقشة النتائج وتفسيرها فى ضوء مستوى الدلالة وتقديرات الثقة. وفى ضوء الحديث على محددات الدلالة الإحصائية فإن التجديد الإحصائى يتطلب:

- الإقلال من استخدام اختبارات الدلالة الإحصائية كلما أمكن.
- توجيه الاهتمام إلى انواع اخرى من الإحصاء مثل حجم التأثير وفترات الثقة.
- التأكيد على أهمية الاعادة للدراسات Replication خاصة فى العلوم السلوكية.

واشار Cumming & Finch (2001) إلى ضرورة أن يفكر الباحثون تفكيرًا ما وراء تحليلًا Meta-analytic thinking الذى يركز فى المقام الأول على تراكم الادلة المتنوعة عبر الدراسات ومن أهم مظاهره:

- التقدير الدقيق لنتائج الدراسات السابقة.
- يجب أن تكون دراسة الباحث اسهمت فى تراث الظاهرة فهى وحدها غير قادرة على الوصول لحل قضية أساسية، فدراسة العينات الصغيرة غير قادرة على حل ابسط القضايا.
- الباحث يعرض نتائج الدراسات من خلال تضمينها فى دراسات مستقبلية لما وراء التحليل.
- تفسير النتائج الجديدة تتم بمقارنتها بأحجام الاثر السابقة للظاهرة موضع الدراسة.

والتفكير ما وراء التحليلى لا يركز فقط على الاختبارات الإحصائية انما يعتمد على أحجام التأثير عبر مجموعة من الدراسات الأولية المرتبطة وعليه فإن التركيز يكون على حجم التأثير وليس الدلالة الإحصائية فى الدراسات المفردة.

اختبارات الفروض باستخدام فترات الثقة Confidence intervals

حث التقرير السنوى لـ(2009) APA على استخدام فترات الثقة باعتبارها أفضل الاستراتيجيات فى تقرير البحث واوصت بشده باستخدامه كاستراتيجية لصناعة القرار وفى اختبارات الفروض ويؤيد بعض الباحثين استبدال الدلالة الإحصائية بفترات الثقة امثال (Aron et al., 2013; Cohen, 1994; shmidt, 1996)، لما لها من مميزات أهمها انها تتضمن كل المعلومات المتضمنة فى اختبارات الفروض الإحصائية وبل تعطى معلومات اضافية لتقدير مدى من القيم يمكن أن تتضمن المتوسط الحقيقى للمجتمع بمعنى انها ليست ثنائية القرار فقط (نعم – لا)، كما انها تمدنا بتقدير للتأثيرات وليس فقط ما ان كان القرار الخاص بالتأثير يختلف عن الصفر وهى تكون مفيدة عندما تكون النتائج غير دالة إحصائيًا لأنها تعطى فكره عن مدى الابتعاد عن وجود التأثير وما إذا كانت بعيدة بدرجة قليلة أم كبيرة عن وجود التأثير أو المعلمة.

ويمكن للباحثين اختبار الفروض باستخدام وسائل اخرى مثل فترات الثقة وليس عن طريق مقارنة القيمة المحسوبة بالقيمة الحرجة أو مقارنة قيمة p بـ α وعلى الرغم ان استخدم مدخل اختبارات الفروض باستخدام فترات الثقة غير مستخدم بصورة شائعة كما هو الحال لاختبارات الفروض الصفرية ولكن من المهم ان يدرك الباحثون كيفية استخدام فترات الثقة داخل اطار اختبارات الفروض. وعند استخدام فترات الثقة فكل خطوات اختبارات الفروض تبقى كما هى فيما عدا قاعدة القرار وهذا المدخل لاختبارات الفروض يتضمن صياغة الفروض الإحصائية وتحديد α والاختبارات الإحصائية والخطوة الاخيرة القرار رفض أو الفشل فى رفض H0 ومفهومى الخطأ من النوع الأول والثانى تبقى كما هى، وكذلك حجم التأثير والقوة وحجم العينة إذا ما اعتمد الباحث على مدخل الخطوات التسعة.

كما سبق ان القيمة المحسوبة والحرجة للاختبار هى قيم عددية حرة المترية (بدون وحدات قياس) وليس لها معنى مرتبط بمستويات القياس المرتبطة بالبيانات وهذه تعتبر ميزة للباحثين لاستخدام قيم حره القياس ولكنها تعطى فهم محدود لكيفية رفض أو عدم رفضH0، وميزة حدود الثقة انها تعطينا هذا الفهم ويتم تقدير حدود الثقة من بيانات أو

إحصاء العينة فبدلًا من حساب القيمة المحسوبة للاختبار فإن الباحث يحسب الفترة فى وجود مستوى دلالة إحصائية ألفا وهذه الفترة يعبر عنها بفترة 95% لألفا 0.05 أو فترة 99% لمستوى ألفا 0.01 فبدلًا من التركيز على القيمة الحرجة فإن الباحثين يتجهون إلى الفرض الصفرى ويقارن فترة الثقة بالقيمة المتضمنة فى الفرض الصفرى (H0: μ =0)او (H0: μ =5) وهكذا. وقاعدة القرار فى الخطوة النهائية هو لو ان القيمة المفترضة فى الفرض الصفري(0,5,10) وقعت خارج فترة الثقة فانة يرفض الفرض الصفريH0 وإذا وقعت داخل فترة الثقة يتم قبولH0.

ويشير Aron et al. (2013) إلى ان فترات الثقة (CI) هى مدى من الدرجات بين "الحد الأعلى والحد الأدنى " ومن المحتمل ان تتضمن متوسط المجتمع أو أى مدى من متوسطات المجتمعات الممكنة التى بدورها توجد احتمالية عالية ان تتضمن متوسط العينة. وحدى الثقة هو القيمة الدنيا والعليا لفترات الثقة ويتم عرضها فى تقرير البحث كالآتى:

$$(r = 0.27, 95\%\ CI\ (0.11, 0.42), p= 0.0012)$$

وفى هذا فإن 95% فترة الثقة تتسع من 0.11 إلى 0.42 ولان الفرض الصفرى يساوى صفر 0= μ :H0 وهذا غير متضمن داخل الفترة وعليه يتم رفض H0، وإذا كان الباحث يختبر فروض حول متوسطات :

$$H0: \mu 1 - \mu 2 = 0$$

وكانت النتائج :(Mean difference:13 , 95%CI, (-9 , 36) وعلى ذلك القيمة صفر تقع داخل مدى حدى الثقة وعليه يقبل الباحث الفرض الصفرى وهذا يعنى ان أعلى فرق بين المجموعات أو المتوسطات هو 36 أو فى فرق فى الاتجاه المعاكس هو 9- وهذا المدى من القيم لفرق المتوسطات يحدث لو ان الدراسة تم اعادتها 100 مرة . فمثلا دراسة هدفت إلى تحديد ما إذا كان تحصيل الرياضيات للذكور يختلف عن تحصيل الإناث ويختلف عن 32 درجة فيقوم الباحث بحساب فرق المتوسط عن طريق طرح متوسط درجات الإناث من متوسط الذكور وعلى ذلك فإن الفروق بين متوسط الإناث ومتوسط الذكور تسمى تقديرات النقاط Point estimates حيث انها الإحصاء

المستخلص من العينة وتستخدم لتقدير معلم المجتمع وهى تقوم عليها اختبارات الفروض ذو الست والتسع خطوات.

ولكن بدلًا من ذلك يمكن أن يستعيض الباحث بتقديرات الفترات وهى قائمة على إحصاء العينة وتمدنا بمدى معقول تدور حولها تقديرات النقاط من القيم لمعالم المجتمع وهى تتضمن متوسط المجتمع كنسبة معينة كما لو اخذنا عينات من المجتمع وحدى الثقة قائمة حول متوسط العينة. مستوى الثقة Confidence level95% هو الاكثر استخدامًا ويشير إلى ان 95% ان المتوسط يقع بين ذيلى المنحنى. وفترات الثقة تعطى معلومات مثل المعلومات التى تعطيها قيمة p بل واكثر، فافترض أن قيمة p=0.051 فما هو قرار الباحث، حقيقة معظم الباحثين يعتبروها دالة ولذلك فإن استخدام فترات الثقة تحل هذه الاشكالية.

فترة الثقة لأى إحصائية هى:

$$CI = Statistic \pm critical\ value*(SE)$$

- Statistic هى أى إحصائية محسوبة سواء متوسط أو تباين وغيرها
- SE الخطأ المعيارى للإحصائية.
- Critical value هى القيمة الحرجة المرتبطة بإحصائية الاختبار عند α، df.

وفى اطار هذا الكتاب نتعرض لكيفية حساب فترات الثقة كثيرًا ونعتمد على قيمتها فى برنامج SPSS وعمومًا تعتبر فترات الثقة اختبارات للدلالة الإحصائية للفروض الصفرية غير المدخل الكلاسيكى القائم على تقدير النقاط. ويمكن بناء فترات الثقة بدون وجود فرض صفرى ولكن لا يمكن إجراء اختبارات الفروض الإحصائية الصفرية بدون الفرض الصفرى. ومن النادر جدًا الاعتماد على فترات الثقة بدون وجود اختبارات الدلالة الإحصائية فى الدراسات النفسية والتربوية العربية.

مدخل فترات الثقة

الاستراتيجية الأخرى لتحديد وهو مدى من القيم حول ما إذا كان معلم المجتمع (المتوسط) من المحتمل ان يقع فيه. فعلى سبيل المثال إذا كانت القيمة0.0(صفر) تقع داخل الفترة فإن التقدير غير دال إحصائيًا لمستوى (α-1) . وتقدر حدود الثقة باستخدام برامج مثل Minitab, SAS, SPSS . والمؤيدين لاستراتيجية فترات الثقة

يؤكدون على انها تمدنا بمعلومات عن دقه التقديرات وحجم مسافة فترات الثقة تشير إلى الدقة (درجة الخطأ العشوائى المرتبط بها) ويمكن أن تكون التقديرات عالية الدقة إذا وقعت قيمتها داخل هذا المدى الضيق.

وبعد الدراسة فإن CI يقدم معلومات عن الدقة وتكون أكثر فائدة واستخدامًا من قيمة القوة الإحصائية. وعلى الرغم ان CI حول الفروق تكون اداة لفحص حجم الفروق ويوجد دعم قوى لاستخدام CI اما مكملًا أو بديلًا لاختبارات الفروض

وفقًا لـ(1994) Borenstein يوجد سته خطوات لبناء فترات الثقة كالآتى:

1. اختيار مستوى الثقة Confidence level ويشير إلى الاحتمالية إلى ان معالم المجتمع الحقيقى تكون داخل مدى محدود منCI ودائمًا يعبر عنها بنسبة مئوية مثل 95% مستوى ثقة و تفترض ان الاحتمالية بان يكون معلم المجتمع الحقيقى داخل فترات الثقة0.95 .
2. حساب ألفا وتشير إلى احتمالية ان معلم المجتمع الحقيقى هو خارج CI ويعبر عن ألفا فى ضوء نسبة Proportion ولذلك إذا كان مستوى الثقة 95% فإن الفا:

$$= 1 - 0.95 = 0.05$$

3. تحديد إحصائية العينة (متوسط، انحراف معيارى) وهى تكون بمثابة نقطة التقدير لمعلم المجتمع.
4. تحديد التوزيع العينى Sampling distribution للإحصائية للعينة افترض اننا سحبنا كل العينات الممكنة بأحجام n من المجتمع ثم تم تقدير الإحصائية (المتوسط – النسبة – SD) لكل عينة فإن التوزيع الاحتمالى لهذه الإحصائية يسمى التوزيع العينى.
5. فى ضوء التوزيع العينى للإحصائية تقدر القيمة لاحتمالية الأولى وهذه القيمة هى الحد الأعلى لـCI وهذه القيمة التراكمية تشير إلى الاحتمالية بان قيمة الإحصائية تقع فى المدى المحدود لها.
6. بنفس الطريقة السابقة، حاول إيجاد الاحتمالية الثانية وهى تعتبر الحد الأدنى من CI.

الاستراتيجية البوتسترابية Bootstrap

وفقا لـ Efron (1982) فإن هذا المصطلح مشتق من المقولة القديمة عن انقاذ أو اقتلاع نفسك عن طريق Bootstrap وهى قائمة على فكرة ان العينة الواحدة المتاحة تعطى موثوقية لعينات الأخرى.

استراتيجية البوتستراب Bootstrap

هى أحد أشكال أو طرق اعادة المعاينة من قاعدة البيانات الأصلية من خلال إجراءات الاستبدال ولذلك يطلق عليها إجراءات اعادة المعاينة Resampling وهذه مشابهة لإجراءات اخرى مثل استرتيجية Jackknife، واستخدمها (1979) Eforn كادوات إحصائية لتقدير الاخطاء المعيارية. وبعد اصدار كتاب (1982) Efron أصبح النشاط البحثى فى استخدام استراتيجية البوتستراب أكثر انتشارًا. وفى مؤتمر مؤسسة الإحصاء الرياضى فى مايو عام 1990 فى ولاية Michigan الامريكية حيث تم عرض تطبيقات ومحددات استراتيجية البوتستراب وفى نفس العام عقد المؤتمر الثانى فى مدينة Tier بألمانيا ناقش تطورات استخدامات استراتيجية البوتسراب وغطى هذا المؤتمر طرق المحاكاة باستخدام Monte Carlo وحدود الثقة البوتسترابية وموضوعات خاصة وتطبيقاتها ومحدداتها. وعلى الرغم ان هذه الاستراتيجية قدمت عام 1977 على يد Efron فى جامعة ستانفورد فى تقرير فنى ولكن إجراءاتها لم تنمو بسرعة الا فى بداية 1990 فى مجلدات خاصة وبدأت هذه الاستراتيجية أكثر انتشارًا فى المجتمع العلمى من خلال تطبيقاتها المختلفة على يد Efron وزملائه وعرض Efron & Tibshirani (1993) المحاولة الأولى لعرض هذه المنهجية وتطبيقاتها إلى مجتمع العلماء والباحثين الاجتماعيين. ونحن لا نعطى هذه الاستراتيجية الإحصائية اهتمامًا يذكر فى دراساتنا وأبحاثنا على الرغم من الدقة الكبيرة لها فى التحليلات الإحصائية المختلفة وتحليل البيانات وتطور التفكير الإحصائى.

وبفرض أن $\hat{\theta}_i^*$ تقدير البوتستراب من عينة بوتستراب i th وكل عينة حجمها n ونفترض ان 90% من $\hat{\theta}_i^*$ تتضمن فى 90% حدود الثقة لـ $\hat{\theta}_i^*$ وتسمى حدود الثقة البوتسترابية باستخدام طريقة فترة الثقة المئينية Percentile method confidence interval وهذا مشابه لتقديرات فترات الثقة للعينات العشوائية. وطريقة البوتسراب مشابهه تمامًا كما لو اخذت العديد من العينات العشوائية من المجتمع ولكن البوتستراب تعتمد على قاعدة البيانات المتاحة ثم تولد باقى العينات العشوائية من خلال الاستبدال

من قاعة البيانات المتاحة، وكلما كان حجم العينة المتاح فى البوتسراب كبير فإن الفروق بين تقديرات البوتسراب والعينات العشوائية المسحوبة من المجتمع يكون صغير. ولكن مع أحجام العينات الصغيرة فإن طريقة المئينية لفترات الثقة لا يكون ادائها جيد.

والمناصرين لهذه الاستراتيجية يعتقدون انها تعطى حرية من عاملين مقيدين للنظرية الإحصائية التقليدية وهما:

- مسلمة ان البيانات تتوزع وفقًا للمنحنى الجرسى أو الاعتدالى.
- الحاجة للتركيز على القياسات الإحصائية التى لها خصائص نظرية يمكن تحليها نظريًا اولًا.

وفى هذه الاستراتيجية فإن البيانات الاصلية تستخدم لتوليد عينات عشوائية باستخدام طريقة الاستبدال لإعطاء قواعد بيانات جديدة (Goodd, 1999) . وهى أحد الأساليب اللابارامترية حيث لا تعتمد على نماذج بارامترية لتوزيعات المجتمع وهذه الطرق اللابرامترية تمدنا بوسائل للحصول على تقديرات معاينة أو اختبارات فروض بدون مسلمات حول توزيعات المعاينة مثل الاعتدالية (Chernick, 2008). وتوليد العينات الفرعية باستخدام استراتيجية البوتستراب يكون تقريبى باستخدام طريقة توليد البيانات لمونت كارلو وهذا التقريب يجعل البوتستراب أكثر عملية على الرغم ان إجراءات حسابة معقدة و معقدة وهى تولد عينة عشوائية من خلال استبدال المفردات من قاعدة البيانات الاصلية بالتالى فإن البيانات المتولدة من البوتستراب هى عينة عشوائية بسيطة.

وفيا يلى مثال لتوليد عينة عشوائية بسيطة، بفرض لدينا مجتمع مكونة من ستة أفراد وهى ابجديًا A, B, C, D, E, F اردنا ان تكون العينة اربعة أفراد. بالتالى فإن عدد العينات المحتملة:

$$C(6,4) = \frac{6!}{[4!\,2!]} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{[(4 \times 3 \times 2 \times 1)][2 \times 1]} = \frac{720}{48} = 15$$

بالتالى فإن عدد العينات المحتملة المكونة من اربعة أفراد هى 15 عينة من قاعدة بيانات تتضمن ستة أفراد. هي

1. $\{A, B, C, D\}$

2. $\{A, B, C, E\}$

3. $\{A, B, C, F\}$

4. $\{A, B, D, E\}$

5. $\{A, B, D, F\}$

6. $\{A, B, E, F\}$

7. $\{A, C, D, E\}$

8. $\{A, C, D, F\}$

9. $\{A, C, E, F\}$

10. $\{A, D, E, F\}$

11. $\{B, C, D, E\}$

12. $\{B, C, D, F\}$

13. $\{B, C, E, F\}$

14. $\{B, D, E, F\}$

15. $\{C, D, E, F\}$

بالتالى لإتمام عملية توليد البيانات من قاعدة البيانات الأساسية لا بد من استخدام مولد البيانات باستخدام الكمبيوتر. وبفرض تم تحديد ستة عينات مكونة من اربعة أفراد وتم حساب معلم المجتمع من هذه العينات الستة وليكن المتوسط مثلًا.

وبفرض ان متوسطات العينات الستة كالتالى:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{26 + 17 + 45 + 70 + 32 + 9}{6} = \frac{199}{6} = 33.1667$$

وعندما تكون العينة عشوائية بسيطة فإن متوسط العينة يكون تقدير غير متحيز لمتوسط المجتمع وهذا لا يعنى ان متوسط العينة يكون مساوى لمتوسط المجتمع ولكن يعنى ان يمكن القول إن متوسط متوسطات العينات عبر كل العينات العشوائية المحتملة يكون مساوى لمتوسط المجتمع. وعندما يكون حجم المجتمع كبير بالتالى يكون حجم العينة المنتقاة كبير نسبيًا فإن عدد العينات العشوائية المتولدة المحتملة من قاعدة البيانات الأولية يكون كبير ايضًا.

عمومًا نحن نختار عينة عشوائية لحجم n مع الاستبدال Replacement من المجتمع N ، وفى تطبيقات البوتستراب فإن المجتمع للمعاينة البوتسترابية لا يكون المجتمع الفعلى موضع الاهتمام ولكن معاينة عشوائية مفترضة من المجتمع. والحقيقة ان متوسط العينات البوتسترابية يكون تقدير غير متحيز لمتوسط المجتمع (Chernick, 2008).

وتتضمن خطوات:

اولًا: بناء مجتمع افتراضى ويسمى Apseudo population تمامًا كمرأة أو يتماثل مع شكل العينة المتحصل عليها.

ثانيًا: سحب عينات محددة الحجم من المجتمع المتولد مع وجود استبدالات وكل عينة تتكون من نفس عدد المفردات ولكن ليس بالضرورة نفسها والاستبدال هو جوهرى لإعادة تشكيل العينة ويولد نفس حجم العينة عدد من المرات.

ثالثًا: إجراء التحليل باستخدام الأسلوب الإحصائى لكل عينة مسحوبة.

رابعًا: اخذ متوسط النتائج لكل العينات المسحوبة من المجتمع الافتراضى.

وفى نهاية الحديث عن اختبارات الفروض توجد بعض الملاحظات التى اود عرضها اليها كما اوردها Huck (2012) :

- يختلط الأمر لدى مستخدمى اختبارات الفروض حيث النتائج الدالة إحصائيًا وهى نتائج هامة، ففى الحقيقة عندما يعطى الاختبار نتيجة دالة فإن هذا لا يعنى بالضرورة انها نتيجة هامة أو معنوية تستحق الاعتبار فاذا قارنت وزن 20 فيل

بمتوسط وزن 20 ديك فإنك ستحصل على نتائج داله 0.00000001 > P وعلى ذلك فإن عرض النتائج الدالة لا يكون غريب لدى الآخرين فاذا كانت الفروض تافهة فإن الدلالة الإحصائية لا معنى لها.

- النتائج الدالة إحصائيًا حتى مع فروض معقولة ومنطقية لاتخبرنا عن أى شىء فيما يخص مقدار Magnitude التأثير فاذا كان الفرق فى القياس البعدى بين المجموعة التجريبية والمجموعة الضابطة دال فهذا لا يعنى بالضرورة أنه ذو دلالة عملية Practical Significance فقد يكون الفرق صغير جدًا لا يستحق الوقوف عنده وعليك ان تتذكر جيدًا ان قوة الاختبار الإحصائى تزيد كلما زاد حجم العينة ولذلك حتى إذا كان الفرق ضئيل جدًا ومع أحجام عينات كبيرة فإن الاختبار ينتج دلاله إحصائية.

- لا تتجاهل النتائج غير الدالة إحصائيًا فالفشل فى رفض الفرض الصفرى يبقى على درجة كبيرة من الإثارة، فالفشل فى رفض الفرض الصفرى يثير مناقشات وحوارات تكون مدعاة للباحثين لإجراء المزيد من الدراسات خاصة إذا كان هذا يتعارض مع نظرية متماسكة ومتفق على صلاحيتها ومصداقيتها. وعلى ذلك فالنتائج غير الدالة ليست شيئًا سيئًا، والعكس فالنتائج الدالة ليست بالضرورة جيدة.

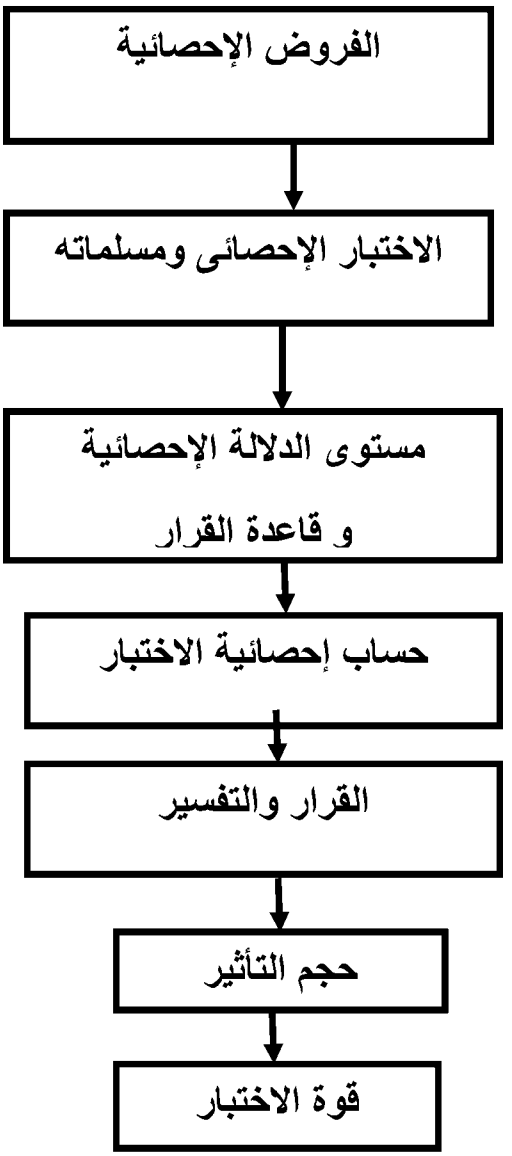
- قبل إجراء أى اختبار خاصه البارامترى فإن عدم الدلالة لاختبارات التحقق من مسلمة الاعتدالية هو أمر مفضل وذلك لأن النتائج غير الدالة تعنى توافر الاعتدالية وعلى ذلك فإن توافر هذا يعنى قبول H0 يكون من اهتمام الباحث وعليه فإن الحصول على H0 ليس أمرًا سيئًا طول الوقت.

- لا تحاول استغلال امكانيات برامج الكومبيوتر لإجراء اختبار تلو الآخر على نفس البيانات للسعى وراء الحصول على الدلالة الإحصائية.

- لا تنزعج من استخدام الاستكشاف مع عملية التحقيق. فمن غير الجائز بعد إجراء اختبارات الفروض الاستكشافية "فروض غير موجهة" ان تختبر فروض أكثر تخصيصًا "فروض موجهة" على نفس البيانات ولكن يمكن أن تختبر فروض موجهة على عينات اخرى.

- الدلالة الإحصائية لا تسمح لنا بادعاء الدلالة الجوهرية Substantive significance ومدى جوهرية النتائج فى ضوء تضمينات النظرية والممارسة العملية وليس فقط هى دالة أو غير دالة فاختبارات الدلالة هى فقط ادوات تساعدنا للتعامل مع قضايا الملائمة العملية أو النظرية وهى ذاتها ليس لها قيمة معنوية.

وبعد هذا العرض لإجراءات اختبارات الفرض فإن معد هذا الكتاب اقترح الخطوات الآتية التى يلتزم بها اثناء عرضة للاختبارات الإحصائية فى الكتاب:



الشكل (5.9): عملية اختبارات الفروض كما تصورها معد الكتاب.

وفـى هـذا النمـوذج المكـون مـن سـبعة خطـوات اعتبـرت فيـه اولـى خطواتـه هـى الفـروض الإحصـائية ثـم الاختبـار الإحصـائى ومسـلماته وتـم وضـع مسـتوى الدلالـة الإحصـائية مـع قاعـدة القـرار ،كمـا تـم اعطـاء أهميـة كبيـرة لمفهـوم الدقـة أو الجـودة الإحصـائية للاختبـار وتـم تقسـيمها لمكـونين همـا حجـم التـأثير وقـوة الاختبـار الإحصـائى وعكـس نمـوذج التسـع خطـوات عنـد (2012) Huck الـذى اسـتخدم فيـه حجـم التـأثير والقـوة الإحصـائية فـى وسـط السلسـلة أى فـى تصـميم الدراسـة أو البحـث للحصـول علـى حجـم العينـة غيـر أننـى اعتبـرت العكـس ان قـوة الإحصـائية وحجـم التـأثير فـى نهايـة السلسـلة أو مخرجـات اختبارات الفروض وذلك للحكم على الجودة الإحصائية لعملية اختبارات الفرض.

الفصل العاشر

العينات

Samples

يعتبر تحديد حجم العينة على درجة كبيرة من الأهمية وهى من الخطوات الصعبة فى التخطيط للدراسة. وتصميم المعاينةSampling design وممارسات انتقاء حجم العينة تم مناقشتها فى التراث البحثى بصورة مستفيضه. فالمجتمع هو مجموعة من الأفراد أو العناصر تشترك فى خصائص عامة. فمثلًا مجتمع يتضمن أفراد يسكنوا مدينة ما أو مجموعة من الطلاب فى مرحلة تعليمية ما مثل المرحلة الاعدادية أو الثانوية أو مجتمع الطلاب فى جامعة معينة أو مجموعة الأفراد الذين اصيبوا بمرض ما خلال الخمس سنوات الأخيرة، والباحث مهتم بدراسة خصائص المجتمع. والعينة هى مجموعة فرعية من المجتمع، والعنصر هو وحدة (التلميذ – الشركة – المدرسة) المجتمع، والمجتمع نظريًا هو مجموعات أو تجمعات من العناصر أو الأفراد أو الأحداث، وتكلفة دراسة المجتمع ككل عملية مستحيلة ومكلفة وتستغرق وقتًا طويلًا واموالًا طائلة، ولذلك يتم اختيار مجموعه فرعية من المجتمع وهذه يطلق عليها عينة Sample . والمعاينة Sampling هى استراتيجية تستخدم لاختيار العناصر أو المفردات من المجتمع. والعينة هى مجموعة جزئية أو فرعية من وحدات المجتمع تنتقى فى ضوء استراتيجية المعاينة وهى تمثيل للمجتمع، وإطار المعاينة Sampling frame هو قائمة أو تسجيلات من مفردات أو وحدات المجتمع. والهدف الأسمى من المعاينة امدادانا بأليات عملية واقتصادية لمساعدتنا على التعميم من العينة إلى المجتمع وكيفية استخدام الاستدلال الإحصائى للتعميم من العينة إلى المجتمع. ويعتبر حجم العينة من الجوانب الهامة فى تحديد الدقة الإحصائية Statistical precision لأن زيادة حجم العينة يقلل خطأ المعاينة، وحجم العينة الكبير يعطى احتمالية أكبر إلى ان تتقارب النتائج مع واقع المجتمع وعمومًا فإن تكلفة دراسة المجتمع ككل تكون مستحيلة للباحثين نتيجة الوقت والتكلفة ونتيجة لذلك يتم اختيار مجموعة فرعية من المجتمع.

والإحصاء هو علم صناعة القرار فى ظل حالات عدم التأكد (الاحتمال) ولفظ إحصائية Statistic يشير إلى المؤشر الإحصائى المحسوب من العينة بينما المؤشر المحسوب من المجتمع يسمى معلم، وعندما يستخدم الإحصاء المحسوب من العينة لتقدير معلم المجتمع فإنه يسمى مقدر Estimator وفيما يلى العلاقة بين العينة والمجتمع فى علاقتها بالإحصاء والمعلم:

الشكل (1.10): العلاقة بين المجتمع والعينة والإحصاء.

المجتمع يضم كل العناصر وعندما يكون المجتمع معروف عناصره يسمى مجتمع محدد A Finite population مثل مجتمع الصف الخامس الابتدائى فى محافظة ما وعدد التلاميذ فى هذا الصف 10000 ، وعندما يكون عدد عناصر المجتمع غير محددة فإنه يسمى مجتمع غير محدد Definite population وليس من الضرورى ان تقدير إحصائيات العينة تساوى قيم معالم المجتمع وهذا لا يكون مدعاة للدهشة وذلك لأن إحصاء العينة يأتى من مجموعة فرعية من المجتمع وحتى أن إحصائيات العينة تختلف عن إحصائيات عينة أخرى من نفس المجتمع. ودائمًا الباحث يرغب أن يختار عينة تعكس تقريبًا معالم المجتمع ولو حدث هذا يسمى تقدير غير متحيز Unbiased لمعالم المجتمع وهذا يحدث لو أن عدد أفراد العينة كبيرًا كبرًا كافيًا. والفرق بين النتيجة المتحصل عليها من العينة وذلك المتحصل عليها من المجتمع الكامل يسمى خطأ المعاينة Sampling Error وهو دالة لحجم العينة فكلما زاد حجم العينة قل خطأ المعاينة.ويرى Thompson (2003) ان العينة فى بحوث العلوم الاجتماعية لا تكون ممثلة تمثيلًا جيدًا للمجتمع. ولان المعاينة هى عمليه انتقاء عناصر المجتمع للدراسة فإن تصميم المعاينة يهتم بقضيتين هما:

- كيفية انتقاء العينة من المجتمع.
- كم عدد أفراد العينة (حجم العينة).

أنواع العينات

تنقسم العينات من حيث طريقة انتقاؤها إلى قسمين العينة العشوائية (الاحتمالية) Random Sample والعينة غير العشوائية (اللاحتمالية) Non Random Sample، والعينة الاحتمالية هى تلك التى تستخدم نظريات وقوانين الاحتمالات والإحصاء فى تعيينها، بمعنى أن العينة الاحتمالية تلك النوع من العينات التى يكون فيها اختيار أى مفردة من مفردات المجتمع عشوائيًا، والعشوائية تعنى ان أى مفردة من مفردات المجتمع لها نفس درجة الاحتمال فى ان تكون إحدى مفردات العينة المختارة. أما العينة اللاحتمالية فتحدث عندما يختار الباحث أفراد العينة غير احتماليًا كان يقصد وعن عمدًا اختيار مدرسة معينة أو قرية معينة أو مؤسسة اجتماعية معينة ولهذا فهى سميت بالعينة المقصودة أو العمدية والدراسات التى تعتمد على اختيار العينة عمديًا تعانى من القصور من حيث تعميم نتائجها.

أولًا: العينة العشوائية البسيطة Simple Random Sampe : هى أحد أنواع العينات الاحتمالية. وأساس اختبار مفردات هذه العينة ويمكن فى فرض مؤداه أن جميع مفردات المجتمع لها نفس فرص الظهور فى العينة وأن عملية اختيار مفردة من مفردات المجتمع لا تؤثر ولا تتأثر بعملية الاختيار لمفردة أخرى. هذا يعنى أن مفردات العينة مستقلة عن بعضها البعض استقلالًا إحصائيًا. ولاختيار العينة العشوائية البسيطة تقوم بعمل الآتى:

- ترقم مفردات المجتمع.
- تحديد حجم العينة المناسب.
- اختيار العينة ذات الحجم المحدد عشوائيًا بالآتى :
  - طريقة الجداول الإحصائية العشوائية.
  - طريقة البطاقات أو الكيس.

- حـــزم البـــرامج الإحصـــائية مثـــل برنـــامجSPSS (الحزمـــة الإحصـــائية للعلـــوم الاجتماعية) باستخدام أمر Simple Random Sample.

ولنأخذ مثال على انتقاء عينة عشوائية بسيطة من مجتمع ما أراد باحث دراسة اتجاهات طلاب الفرقة الأولى بكلية التربية بجامعة ما نحو مهنة التدريس. ومن هنا فإن مجتمع الدراسة هو طلاب الفرقة الأولى فى كلية التربية بالجامعة. وبعد الاطلاع على قوائم هؤلاء الطلبة اتضح أن عددهم 500 طالب وطالبة وحدد الباحث عينة حجمها100 طالب وطالبة أى ان عينة الدراسة 100 ثم قام الباحث بترقيم أفراد المجتمع من 001 إلى 500 ثم كتب 500 ورقم وعلى كل ورقة اسم كل طالب ووضعها فى كيس، ثم سحب أول ورقة فكانت رقم 350 المعروفة أنها لطالب ما من ثم يصبح احتمال كل طالب ان يكون ضمن أفراد العينة 1/500 ومن هنا فالطالب رقم 350 تم اختياره بإحتمال 1/500 ولكن إذا وضعنا الورقة المكتوب عليها اسم الطالب رقم 350 ثم سحب ورقة ثانية فكانت رقم 100 المعروفة باسم طالب ما وهكذا يتم اختيار كل طالب ثم نعيد ورقته إلى الكيس من هنا تسمى عينة عشوائية بسيطة بكسر احتمالى ثابت حيث فرص الظهور لكل طالب ان يكون ضمن أفراد العينة متساوية. ولكن إذا لم يعيد الباحث الورقة رقم 350 إلى الكيس مرة أخرى فيصبح أفراد المجتمع 499 وتكون فرص الطالب التالى فى العينة 1/499 من هنا تغير الكسر الاحتمال، فى هذه الحالة عينة احتمالية بسيطة بكسر احتمالى متغير. ومن الأفضل أن يعتمد الباحث على عينة احتمالية بكسر احتمالى ثابت.

حقيقة فى طرح المثال السابق للتوضيح فقط ويبدو ان طريقة الكيس تبدو غير عملية للمجتمعات الكبيرة ومن الأفضل الاعتماد على الجداول العشوائية أو على الكمبيوتر باستخدامه حزمة البرامج الإحصائية SPSS باستخدام امر العينات العشوائية البسيطة. وإذا كان هدف الباحث معرفة اتجاهات طلاب الفرقة الأولى ولا يهتم بدراسة الفروق بين الاتجاهات باختلاف المتغيرات الديموجرافية فالباحث ينظر إلى العينة باعتبارها متجانسة، وعليه فالعينة العشوائية البسيطة يتم الاعتماد عليها إذا كانت أفراد المجتمع المسحوب منه العينة متجانسًا بقدر الإمكان.

ثانيًا: العينة العشوائية الطبقية Stratified Random Sample: عند اختيار العينة باعتبارها عينة عشوائية بسيطة يفترض ان يوجد تجانس بين مفردات المجتمع بمعنى أن لا ينقسم المجتمع إلى طبقات متمايزة مثل ريف – حضر، أو مستوى اقتصادى عالٍ فى مقابل مستوى اقتصادى منخفض ومن هنا يصعب افتراض تجانس وحدات المجتمع، ولهذا إذا رغب الباحث فى إجراء دراسة يتوقع استخدام وتعميم نتائجها على جميع طبقات مجتمع الدراسة يجب اختيار العينة اختيارًا طبقيًا عشوائيًا بإتباع الخطوات الآتية:

1. التعرف على خصائص المجتمع من حيث المتغيرات والعوامل التى تتضمنها، ثم يقسم المجتمع إلى طبقات من حيث المتغيرات ويحدد حجم كل طبقة فى المجتمع.
2. حدد حجم العينة الكلى المراد اختيارها فى المجتمع.
3. احسب عدد الأفراد التى يجب أن تساهم بها كل طبقة فى العينة كما يلى:

أ- **التوزيع المتساوى:** تحدد حجم العينة المحسوبة من كل طبقة بالتساوى. فلو فُرض أن حجم العينة الكى 300 وأنه يوجد ثلاث طبقات متميزة فى المجتمع فإننا سنختار من كل طبقة (اختياريًا عشوائيًا باعتبار أن الطبقة تمثل مجتمع مصغر عدد من المفردات يساوى 100 ، لاحظ أنه فى هذا النوع من العينات افترضنا تساوى عدد مفردات كل طبقة فى المجتمع وهو افتراض نادرًا ما يتوفر).

ب- **التوزيع المتناسب:** لفهم هذا النوع من العينات الطبقية نأخذ المثال البحثى فى المثال السابق شرحه فى العينة العشوائية البسيطة وهى معرفة اتجاهات طلاب الفرقة الأولى بكلية تربية بجامعة ما نحو مهنة التدريس فالمجتمع يتكون من 3000 وهذا المجتمع يتكون من ثلاث طبقات:

- طبقة طلاب الريف (n1) = 2000
- طبقة طلاب الحضر (n2) = 750
- طبقة طلاب البدو (n3) = 250

لاحظ أن حجم المجتمع: $n1 + n2 + n3 = 3000$

والسؤال الذى يواجه الباحث ما العدد الذى يجب أن يختاره من كل طبقة حتى يمكن الحصول على عينة عشوائية طبقية كلية حجمها = 100

للإجابة على هذا السؤال يتبع الباحث الآتي n1 : n2 : n3

هذه النسبة هى 2000 : 750 : 250 أو بالتقريب تصل إلى 8 : 3 : 1

$1 + 3 + 8 = 12$

إذًا من المفترض ان عينة طلاب الريف $8/12 \times 100 = 67$

طلاب عينة الحضر $3/12 \times 100 = 25$

طلاب عينة البدو $1/12 \times 100 = 8$

لهذا يجب على الباحث ان يحصل على عينة عشوائية طبقية عدد 100 ويجب أن يختار عشوائيًا 67 من طبقة طلاب الريف و 25 من طبقة طلاب المدينة و 8 طلاب من طبقة طلاب البدو.

ثالثًا: العينة العشوائية المنتظمة Systematic Random: نختار العينة فى صورة عدد متساوٍ بعد بداية محددة عشوائيًا وتحديد هذا العدد المتساوى يتم فى ضوء نسبة N/n حيث N عدد مفردات المجتمع و n عدد مفردات العينة المراد انتقاؤها. فمثلًا فى المثال السابق نجد أن المجتمع 3000 وحجم العينة 100 من هنا فإنه يتم اختيار العينة فى ضوء تحديد الفترة المتساوية 3000/100 = 30 ، ونأخذ بداية فى المدى من واحد (١) إلى (30) وليكن 10 ومن هنا بعد العدد 10 يتم انتقاء المفحوص رقم 10+30=40 من هنا يتضح العينة للمفردات التى تأخذ الأرقام الآتية: 10 , 40 , 70 , 100 , 130 , .. ، أى يتم تحديد عدد بصورة عشوائية كنقطة بداية ثم بعد فترة محددة متساوية يتم اختيار المفردات الأخرى.

رابعًا: العينة التجميعية Cluster sample: تستخدم لانتقاء عينه عشوائية من مجتمع كبير جدًا أو مجتمع له تنوعات جغرافية مختلفة والهدف العام للعينة التجمعية هو تقليل التكلفة عن طريق زياده كفاءة المعاينة. يكون المجتمع عبارة عن مجموعات

هرميـــة أو تجمعـــات مـــن المجموعـــات وهـــذه المجموعـــات تختلـــف مـــن حيـــث الحجـــم والطبيعـة، فعلـى سـبيل المثـال مجتمـع تلاميـذ المدرسـة يـتم تجمعهـم فـى عـدد مـن الفصـول أو عـدد مـن المـدارس ويـتم انتقـاء العينـة مـن الطـلاب مـن هـذا المجتمـع عـن طريـق اختيـار تجمعـات مـن الطـلاب أو الفصـول أو المـدارس و لـيس كمـا يحـدث فـى العينـة البسيطة العشوائية حيث يتم انتقاء مدرسة ما.

القضـية مـع هـذا النـوع مـن المعاينـة (علـى الـرغم ان يكـون للتجمـع نفـس الفرصـة ان يـتم انتقـاؤه) هـو ان المفـردات أو العناصـر داخـل التجمعـات الكبيـرة تقـل فرصـتها ان تكـون ضمن العينة النهائية.

وفيما يلى مثال لاختيار عينة عشوائية تجمعية:

المدارس	الأولى		الثانية		الثالثة	
الفصل	**فصل 1**	**فصل 2**	**فصل3**	**فصل 4**	**فصل 5**	**فصل 6**
التلاميذ	4	4	**4**	**4**	**4**	**4**

حقيقـة مـن الصـعب تحقيـق العينـة العشـوائية لاعتبـارات اخلاقيـة والوقـت و التكلفـة ولـذلك فمـن الضـرورى اسـتخدام تكنيكـات أو أسـاليب اخـرى للمعاينـة و يطلـق عليهـا المعاينـة غيـر الاحتماليـة أو غيـر العشـوائية **Non-probability or Non-random sampling** فـى ظـل اسـتخدام هـذه الأسـاليب فـإن الفـرد فـى المجتمـع لـه فرصـة غيـر معروفـة لأن يكـون مـن ضـمن أفـراد المعاينـة المختـارة وعلـى احسـن الاحـوال فـإن العينـة غيـر العشـوائية ممكـن ان تمـدنا بعينـة ربمـا تكـون ممثلـة لخصـائص المجتمـع وعلـى اسـواء الاحـوال تعطـى عينـة غيـر ممثلـة تمثـيلًا صـادقًا للمجتمـع ويوجـد اسـتراتيجيات عديـدة لهـذه المعاينة غير العشوائية مثل:

1. العينـة المتاحـة Availabilitysample or Convenience: حيـث يـتم انتقـاء الأفـراد لسـهولة الحصـول عليهـا مـن قبـل الباحـث و الانتقـاء بهـذا الأسـلوب يـؤدى إلـى

وجود تحيز Bias في العينة، وهذه العينة يتم اختيارها مثلا من مؤسسة تربوية أو مؤسسة اكلينيكية حيث يعمل فيها الباحث.

2. **العينة المستهدفة Purposive sample:** يتم اختيار مفردات العينة بناءً على حكم و رؤية الباحث حيث يرى الباحث انها الأفضل للحصول على المعلومات ويتم انتقاء الأفراد بناء على استعدادهم للمشاركة في الدراسة أو معرفتهم لموضوع الدراسة.

3. **العينة Quota sample:** هي الصيغة أو الصورة غير الاحتمالية للعينة الطبقية حيث يتم تمثيل العينة في ضوء طبقات المجتمع التي يتكون منها ويشار اليها بالعينة الممثلة تمثيلًا جيدا لطبقات المجتمع حيث يتم انتقاء مفردات كل طبقة في ضوء حجمها ولا يوجد ضمان لفحص دقة التقديرات من كل طبقة.

4. **عينة كرة الثلج أو المتدحرجة Snowball sample** تستخدم لتحديد العينة عندما يصعب تحديد المتطوعين الأساسين للدراسة فمثلًا على سبيل المثال إذا كان من الصعب تحديد العدد المناسب من الاشخاص الصم فإن الشخص الصم المشترك أو الذى تم انتقائه يساعدنا في تحديد اشخاص صم اخرين مستعدين للمشاركة في الدراسة أى يتم تحديد أفراد العينة من خلال معرفتهم ببعضهم البعض.

مداخل تحديد حجم العينة

العينة هي جزء فقط من المجتمع والتعميم من العينة إلى المجتمع غالبًا يتضمن اخطاء وعمومًا يوجد نوعين من الاخطاء التي تحدث في عملية التعميم من العينة إلى المجتمع هما(2008) Dattalo :

- خطأ المعاينة أو الخطأ العشوائى ويكون ناتج عن حظ السحب luck of the drawوهو وجود عناصر كثيرة من نوع واحد وعناصر غير كافية لنوع آخر فكلما زاد حجم العينة فإن الخطأ المنتظم يقل وإذا كان المجتمع متجانس خطا المعاينة يكون صغيرًا.
- خطأ عدم المعاينة Non-sampling أو الخطأ المنتظم Systematic وهذا الخطأ يطلق عليه التحيز وتكون مشكلة خطيرة لأن خطأ عدم دقة المعاينة لا يمكن التحكم به عن طريق زيادة حجم العينة.

ويصنف خطا المعاينة إلى ثلاثة تصنيفات:

- تحيز الانتقاء Selection bias: وهو النزعة إلى استبعاد بعض المفردات من العينة وحدث هذا بصورة واضحة مع العينة المتاحة وفى المقابل فإن تحيز الانتقاء فى العينة العشوائية الجيدة يكون فى الحد الادنى.
- تحيز عدم الاستجابة Nonresponsive: إلى أى درجة أن المستجيبين وعدم المستجيبين يختلفون على المتغيرات موضوع الاهتمام وحدوث استقطاب Extrapolation للمستجبين تكون اشكالية كبيرة.
- تحيز الاستجابة: نتيجة اتجاهات الباحث أو الذين يقومون بالمقابلة أو مفردات عبارات الاستبيان أو مدى تشابه أو تقارب الأسئلة.

وتصميم المعاينة الفعال يتطلب توازن بين محكات هامة أهمها:

- مدى تحقيق أو انجاز اهداف الدراسة.
- مدى دقة تقديرات تباينات العينة.
- مدى إجرائيتها.
- الاقتصادية (تحقيق اهداف الدراسة بأقل تكلفة).

هذه المعايير الأربعة متصارعة مع بعضها البعض والباحثين يحاولون إحداث التوازن بينهم ويحدد(2003) Diamantopolus & Schlegelmilch اعتبارات إحصائية عند تحديد حجم العينة:

- درجة التباين فى المجتمع: فالمجتمع غير المتجانس يحتاج إلى حجم عينة أكبر لإظهار هذا التنوع، فلو أن مجتمع يتضمن 300000 فردًا ذو تنوعات مختلفة فى النوع والمستوى الاقتصادى والخلفية الأسرية فهذا يتطلب اختيار حجم عينة كبير حتى تكون العينة ممثلة لطبقات أو تنوعات المجتمع.
- درجة الدقة المرغوبة Desired degree of precision: المرتبطة بتقديرات معالم المجتمع فاذا اراد الباحث الحصول على تقديرات ذات دقة عالية فهذا يتطلب زيادة حجم العينة.

- درجة الثقة المرغوبة Confidence degree: المرتبطة بأى تقديرات للمعالم، فمعالم عند حدود ثقة 99% تحتاج إلى عينات أكبر من معالم عند حدود ثقة 95%.

وتوجد مداخل عديدة لتحديد حجم العينة وهى:

- مدخل تحليل القوة الإحصائية
- مدخل القواعد المتعارف عليها
- فترات الثقة
- مدخل طرق كمبيوترية معقدة

ويتم عرض هذه المداخل موضحًا نواحى القوة لكل مدخل علمًا بأنه لا يوجد حزمة برامج واحدة تسمح للباحثين بتحديد حجم العينة للمداخل الثلاثة ولكن من أهم هذه البرامج هو برنامج G- Power لسهوله استخدامه ومجانية الحصول عليه من خلال الموقع:

http : WWW.psycho.uniduesseldorp.de/aap/projects/gpower/

مدخل تحليل القوة الإحصائية

اشار Cook & Campbell (1979) إلى ان صدق الاستنتاج الإحصائى Statistical conclusion validity للدراسة يتحدد بعدة عوامل أهمها تصميم البحث، مستوى الدلالة الإحصائية (الخطأ من النوع الأول)، تباين المجتمع فى المتغير المحك (التابع)، حجم الفروق بين القيم الحقيقة للمعلم فى المجتمع والقيمة المحددة عن طريق الفرض الصفرى (حجم التأثير)، نوع الفرض ذو ذيل واحد (موجهة) مقابل ذو اتجاهين (غير موجهة)، نوع الاختبار المستخدم، وحجم العينة. وكلما زاد حجم العينة فإن النتائج تكون أكثر مصداقية لتمثيلها للمجتمع وتقل اخطاء المعاينة.

وتعتبر القوة الإحصائية دالة وظيفة للعوامل الآتية:

- مستوى الدلالة الإحصائية (α) أو الخطأ من النوع الأول
- الخطأ من النوع الثانى (بيتا β)
- مقدار حجم التأثير المفترض

- الانحراف المعيارى أو التباين لحجم التأثير المفترض
- حجم العينة:

والمنطق وراء تحليل القوة هو ان الدراسة لو اعتمدت على حجم عينة كبير كبر كافٍ فإنه توجد احتمالية مسبقة للحصول على تأثير أو فروق إذا كانت موجودة بالفعل وإذا كان حجم العينة صغيرًا فإن الدراسة تفشل فى الكشف عن الفروق الموجودة بالفعل فى المجتمع وعلى ذلك فالدراسة مضيعة للوقت والجهد والموارد.

ويفيد تحليل القوة الإحصائية فى انتقاء حجم العينة فى ضوء اسس منطقيه بدلا من انتقائها فى ضوء القواعد المتعارف عليها Rules of thumb التى ليس لها اسس علميه موضوعية، وفى تحليل القوة قبل إجراء الدراسة لا بد من تحديد حجم التأثير وقيمتى α ، β، تحديد اتجاه الاختبار، وعلى ذلك يمكن تقدير حجم العينة المناسب.

وفى هذا التحليل يرغب الباحث فى تحديد حجم العينة فى ضوء مستويات مقبولة من حجم التأثير و α و القوة وقد يرتضى الباحث $\alpha=0.05$ والحد المقبول المناسب من القوة هو 0.80. والقضية هنا هى كيفية تقدير حجم التأثير وفى هذا الشأن يوجد ثلاثة مداخل لتقدير حجم التأثير هى:

1. إجراء دراسة استطلاعية (n=50) مثلا وتقدير حجم التأثير للاختبار المستخدم.
2. مراجعة الدراسات السابقة وتقدير متوسط حجم التأثير للدراسات التى تناولت نفس متغيرات الدراسة ويمكن إجراء ما وراء التحليل لعدد محدود يتراوح من 5 إلى 10 دراسات للوصول إلى حجم التأثير فى الدراسات السابقة.
3. تبنى معايير حجم التأثير للاختبار المستخدم، وهذه المعايير وضعها Cohen (1988)، فمثلًا فى حاله اختبار T للعينات المستقلة القيمة d=0.2 تعبر عن حجم تأثير ضعيف، 0.50 حجم تأثير متوسط، 0.80 كبير.

مثال: حساب حجم العينة لاختبار T المستقلة

فيما يلى مثال لتحديد حجم العينة فى حاله استخدام اختبار T باستخدام تحليل القوة القبلى باستخدام برنامج G-POWER وحيث اراد باحث دراسة فعالية استراتيجية التعلم المنظم ذاتيًا على تحسين القدرة اللغوية وحدد المعالم الآتية:

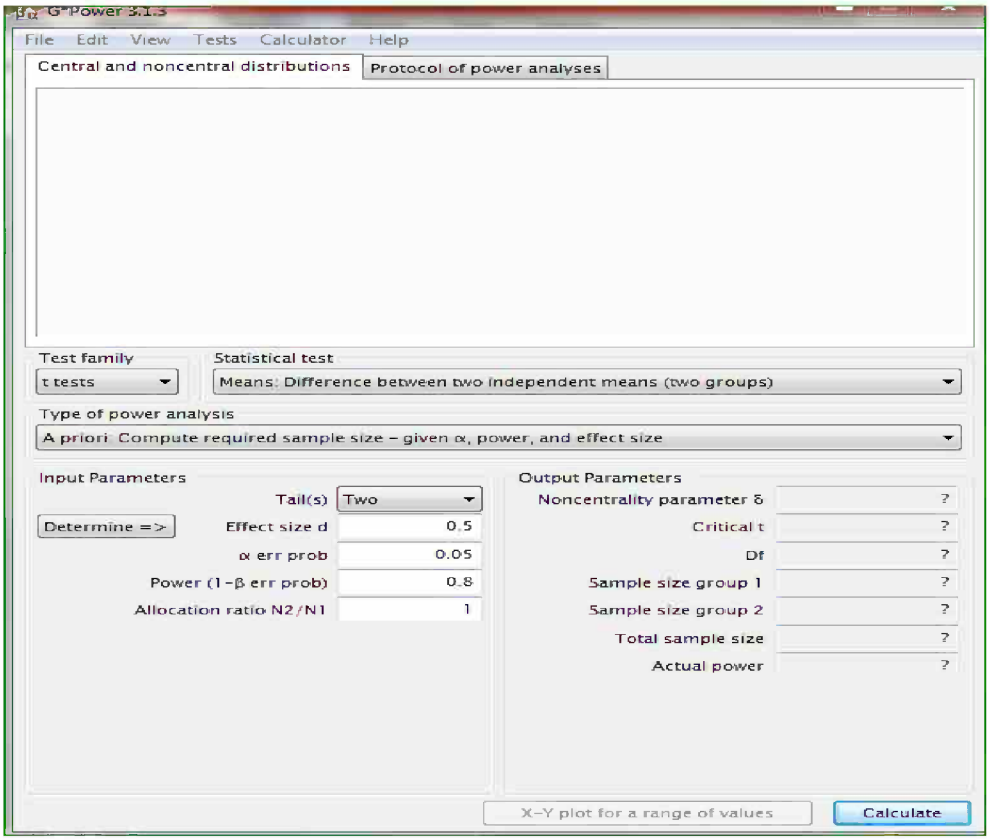
1. الاختبار هو T لعينات مستقله ذو ذيلين

2. مستوى الدلالة الإحصائيةα=0.05

3. القوه المناسبة = 0.8

4. تبنى حجم تأثير متوسط = 0.50

وفيما يلى الشاشة الافتتاحية للبرنامج حين يتم تحديد Test family وتتضمن اختبارات عديدة وهى كما هو موضح :



- اختار التحليل القبلى Apriori تحت مربع Type of power analysisحيث يهدف إلى تقدير حجم العينة المناسب.

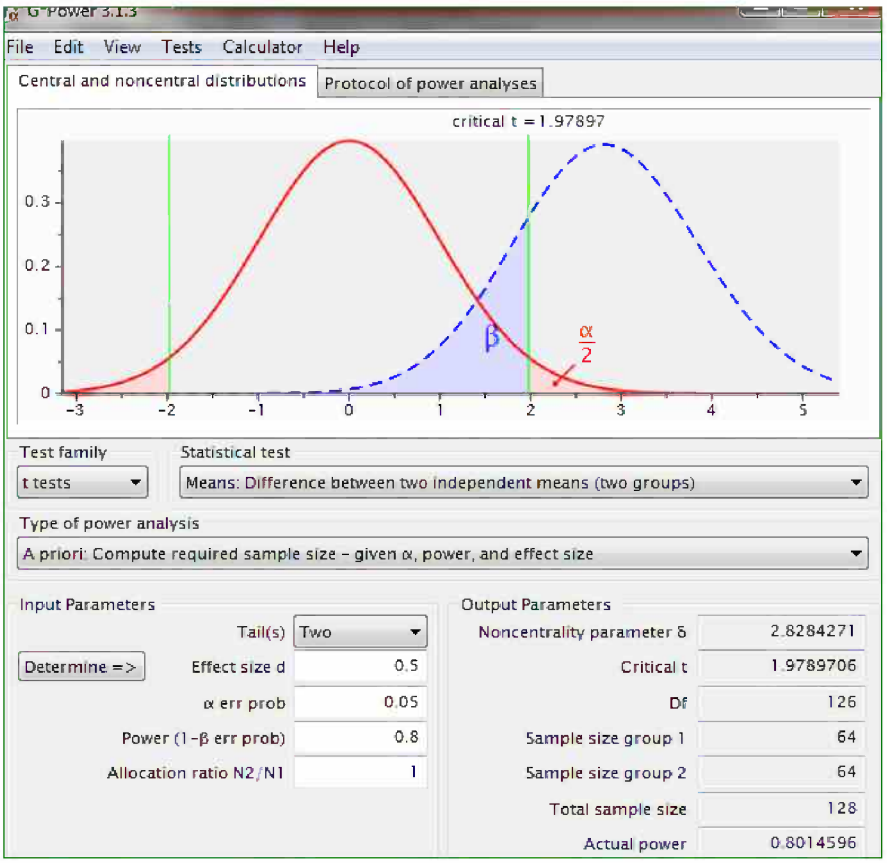
- اختار من تحت عائلة الاختبارات اختيارt-test

- من تحت Statistical test اختار نوع الاختبار المرد تحدد حجم العينة له وفى هذه الحالة هو :

– ويتضمن البرنامج معالم الإدخال Input parameters قم بإدخالها وهى: تحديد طبيعة اتجاهية الاختبار ذو ذيل واحد أو ذو ذيلين ، حجم التأثير، قيمة α، قيمة القوة P، وتحديد نسبة حجم العينة الأولى N1 الى الثانية وغالبًا تكون النسبة N1:N2 وهى 1:1

وتم إدخال المعالم السابقة فى البرنامج.

- اضغط Calculate تظهر الشاشة الآتية :



تم اعطاء قيمة T الحرجة =1.97 و Df =126 وحجم العينة الكلي=128 مقسمة إلى 64 فردًا فى المجموعة الضابطة و 64 فردًا فى المجموعة التجريبية وذلك لتحقيق قوة فعلية 0.8014

مدخل القواعد المتعارف عليها Rules of thumbs

قدم Wilson Vanvoorhis & Morgan (2007) إرشادات عملية واطار لتحديد حجم العينة فى ضوء قواعد متعارف عليها ولا بد أن تتضمن الدراسة حجم عينة مناسب لهدفها. ويجب أن تكون كبيرة كبرًا مناسبًا وذلك فى محاولة الحصول على دلالة إحصائية. فحجم عينة غير مناسب هو مضيعة للموارد والوقت وحجم عينة كبير غير ضرورى وذلك لأنها تحمل

الباحث عبء شديد. غالبًا يتم تجاهل تحديد حجم العينة فى ضوء مدخل القوة الإحصائية، ولكن لا بد على الأقل الانتباه إلى القواعد المتعارف عليها، وفيما يلى جدول يوضح ذلك:

الجدول (1.10): حجم العينة فى ضوء القواعد المتعارف عليها WilsonVanvoorhis& Morgan (2007)

الاختبار	حجم العينة المناسب
T ,ANOVA	تتضمن الخلية 30 فردا لتحقيق P=0.8 واذا قلت ليس أقل من 7 .
الارتباط والانحدار	50 فردًا على الأقل
$\chi 2$	على الأقل 20. ولا يقل حجم الخلية عن 5
التحليل العاملي	300 فاكثر جيد

وفيما يلى توضيح لذلك:

اولا: حجم العينة لاختبارات الفروق: حجم الخلية لاختبارات الفروق مثل اختبار T المستقلة و T المرتبطة وتحليل التباين الأحادى ANOVA وتحليل التباين المتعدد وتحليل التباين المتدرج MANOVA ولتحقيق حجم تأثير متوسط إلى كبير يجب أن يكون 30 فردًا فى كل خلية لأن هذا يقود إلى قوة إحصائية 80%. وفى بعض الظروف ولبعض الاسباب فإن أدنى حد لحجم العينة7 أفراد وهذا يؤدى إلى الحصول على قوة 50%، و14 فردًا لكل خلية ولحجم تأثير 0.50 يؤدى إلى قوه تقريبًا 80%. ويجب الاشارة هنا إلى ان الخلية تمثل المجموعة، فى هذه الحالة فإن اختبار T يتضمن خليتين و F يتضمن خليتين فأكثر. ويجب التأكيد على عدة امور منها:

1.المقارنة بين مجموعات قليلة (2 مثلًا) يتطلب حجم عينة مناسب وذلك للحصول على قوة مناسبة.

2. إذا كان حجم التأثير المتوقع صغير أو منخفض فإن هذا يتطلب حجم عينة أكبر لتحقيق مستوى قوة مناسب.

3.لاستخدام تحليل التباين المتعدد أو المتدرج فمن الضرورى توافر حجم عينة كبيرًا فى كل خلية (Tabachnick & Fidell, 2007).

ثانيًا: حجم العينة اللازمة لدراسة العلاقات والانحدار: يوجد العديد من الصيغ المعقدة فى هذا الشأن وعمومًا يجب استخدام حجم عينة لا تقل عن 50 فردًا للارتباط وأيضًا للانحدار ولكن يجب أن يزيد العدد فى حالة زيادة عدد المتغيرات التابعة. وقدم Green (1991) اطارا شاملًا للإجراءات المستخدمة لتحديد أحجام عينات الانحدار واقترح الصيغة الآتية:

$$N > 50 + 8m$$

حيث m عدد المتغيرات المستقلة أو المنبئات وذلك لاختبار الدلالة الإحصائية لمعامل الارتباط المتعدد (اختبار F) .

ولاختبار معالم المنبئات أو المتغيرات المستقلة (اختبارT) تستخدم الصيغة الآتية:

$$N > 104 + m$$

وقدم Harris (1985) صيغة تستخدم لتحديد عدد أفراد العينة فى تحليل الانحدار لعدد منبئات 5 فأقل فاقترح حجم العينة 50 على الأقل وهو ان العدد الكلى من أفراد العينة يساوى عدد المنبئات بالإضافة الى50 فردًا:

$$\mathrm{N} > m + 50$$

فمثلًا إذا كان تحليل انحدار يتضمن سته منبئات (متغيرات مستقلة) أو أكثر فإن الحد الأدنى المطلق هو تمثيل المتغير المستقل بعشرة أفراد وهذا مناسب.

وتوصل عامر (2007) بان تمثيل المتغير بثلاثة افراد يساوي في الدقة والجودة والمطابقة مع النموذج تمثيلة بخمسة او عشرة افراد شرط الثبات المرتفع للمتغيرات. وطرح عامر(2007) هل يلجأ الباحث إلى زيادة حجم العينة عند إجراء أسلوب الانحدار المتعدد؟، حيث لم تصل الدراسة إلى إجابة بالإيجاب أو بالنفى تماماً، ولكن يمكن القول بأنه إذا كان هدف الباحث من أسلوب الانحدار المتعدد هو تفسيرى، أى معرفة مدى الإسهام النسبى للمتغيرات المنبئة فى تفسير المتغير التابع فلا داعى لزيادة حجم العينة خاصة إذا كان عدد المتغيرات المنبئة محدوداً وتتميز بدرجة عالية من الثبات وأن النصائح بحجم عينة 200 أو 300 لا داعى لها حيث تُلقى

على الباحثين تكلفة مادية وجهد، ومن المحتمل أن تؤدى إلى وفرة فى البيانات أكثر من اللازم خاصة إذا كانت عدد المتغيرات المنبئة محدود.

ثالثًا. حجم العينة لأسلوب $\chi 2$: يستخدم لاختبار الاستقلالية للمتغيرات التصنيفية وفى هذا الأسلوب فإن أفراد العينة يوجد بينها استقلالية والتكرارات المتوقعة الصغيرة فى خلية أو أكثر تحد من الوصول إلى القوة المناسبة لأنها تضخم مستوى الخطأ من النوع الأول وحجم عينة على الأقل 20 فردًا فإن الخطأ من النوع الأول نادرًا ما يذيد عن 0.06 (Howell, 2013). والقاعدة الصارمة يجب أن لا يكون التكرار المتوقع فى الخلية أقل من 5.

رابعًا. حجم العينة للتحليل العاملى: القاعدة الجيدة للعينة فى التحليل العاملى هو 300 فردا (Tabachnick & Fidell, 2007) أو أكثر من 50 فرد لكل عامل (Pedhazur & Schmehrin,1991). وقدم Comrey & Lee (1992) القاعدة الآتية: 50 فقير جدًا، 100 حجم عينة فقير، 200 مناسب، 300 جيد، 500 جيد جدًا و 1000 ممتاز.

واشار Guadagnoli & Velicer (1988) ان الحلول العاملية مع وجود تشبعات عاليه للمتغيرات (أكبر من 0.80) لا تتطلب استخدام عينات كبيرة.

مدخل فترات الثقة

الاستراتيجية الأخرى لتحديد حجم العينة هى على أساس الحدود المرغوبة من فترات الثقة(CI)interval Confidence وهو مدى من القيم حول ما إذا كان معلم المجتمع (المتوسط) من المحتمل ان يقع فيه. فعلى سبيل المثال إذا كانت القيمة0.0(صفر) تقع داخل الفترة فإن التقدير غير دال إحصائيًا لمستوى ($1-\alpha$) .وتقدر حدود الثقة باستخدام برامج مثل Minitab SAS, SPSS, . والمؤيدين لاستراتيجية فترات الثقة يؤكدون على انها تمدنا بمعلومات عن دقه التقديرات وحجم مسافة فترات الثقة تشير إلى الدقة (درجة الخطأ العشوائى المرتبط بها) ويمكن أن تكون التقديرات عالية الدقة إذا وقعت قيمتها داخل هذا المدى الضيق. واتساع مدى فترات الثقة يشير إلى دقة ضعيفة. واختيار حدود ثقة عند99% يزيد من دقة حدود الثقة (يمتلك فرص كبيره لتضمين معالم المجتمع) ولكن يقلل من مستوى الدقة. ويمكن تقدير CI قبل إجراء الدراسة وفترات الثقة

تستخدم كمرشد لاختيار حجم العينة، وتقدر CI حتى لو لم يوجد فرض إحصائى صفرى.

وبعد الدراسة فإن CI يقدم معلومات عن الدقة وتكون أكثر فائدة واستخدامًا من قيمة القوة الإحصائية. وعلى الرغم ان CI حول الفروق تكون أداة لفحص حجم الفروق ويوجد دعم قوى لاستخدام CI اما مكملًا أو بديلًا لاختبارات الفروض ويمكن أن تستخدم لتقدير حجم العينة وقارن Orme & Hudson (1995) بين مدخل القوه الإحصائية ومدخل فترات الثقة لتقدير حجم العينة وتوصلا إلى ان استخدام مدخل CI يمدنا بحجم عينة تؤدى إلى تقديرات أكثر دقة من مدخل القوة. ويوجد معادلات وجداول لتقدير حجم العينة فى ضوء حدود فترات الثقة.

مدخل طرق كمبيوترية المعقدة Computer Intensive – methods

تناولنا تحديد حجم العينة فى ضوء القوة الإحصائية وحدود الثقة. ولكن فى بعض التصميمات الإحصائية المعقدة أو المواقف البحثية يبدو ان استخدام هذه المداخل محدودًا بالإضافة إلى عدم وجود حزمة إحصائية واحدة لتحديد حجم العينة لكل الأساليب الإحصائية ولذلك يحتاج الباحثون إلى برامج إحصائية متنوعة وتبدو انها مكلفة. وللنماذج الإحصائية المعقدة يوجد مدخل آخر لتحديد حجم العينة باستخدام استراتيجيات كمبيوترية معقدة وتسمى أيضًا استراتيجيات اعادة المعاينة Resampling strategies وهى متوفرة منذ 1950. ومصطلح الاستراتيجيات المعقدة كمبيوتريًا هى تصف مجموعة من المداخل متضمنة طرق Bootsrap ومونت كارلو Monte Carlo methos. وتعتمد استراتيجيات المعاينة العشوائية والمحاكاة الكمبيوترية للحصول على حلول تقاربية للمشكلات الرياضية أو الفيزيقية.

طرق مونت – كارلو Monte –Carlo methods

تعبير مونت كارلو (MC) يصف مجموعة من الأساليب الاستاتيكية القائمة على استخدام اعداد عشوائية والإحصاء الاحتمالى، وتعتمد على توليد البيانات وفقا لخصائص معينة وذلك لدراسة قضايا معينة فى الرياضيات. وتستخدم استراتيجيات MC لحل قضايا عديدة يصعب دراستها بالبيانات الحقيقية وتتراوح من موضوعات فى

الإحصاء والاقتصاد والفيزياء النووية. بكلمات اخرى فإن تجربة مونت كارلو هى استخدام ارقام عشوائية لحل مشكله ما. وتستخدم فى قضايا مثل فحص انظمة معقدة مثل التفاعلات بين الذرات. وتستخدم فى قضايا المشكلات النفسية مثل دراسة ما إذا كان الباحث يريد التحقق من أثر الارتباطات المرتفعة بين المنبئات فى الانحدار المقدر على دقة تقديرات معاملات بيتا أو مطابقة النموذج.

مولد الاعداد العشوائي

فى استراتيجية Bootstrap ومونت كارلو تقدر الإحصائيات من خلال عينات متعددة . ولكن Bootstrap تختار العينات من حالات أو أفراد العينة المختارة من المجتمع وتتم المعاينة مع الاستبدال بينما فى مونت كارلو يتم انتقاء العينات من توزيع احتمالى نظرى مع خصائص محددة لمعالم النموذج، وفيها يتم استخدام خصائص عينة وحيدة لبناء مجتمع كبير غير محدد ثم سحب عينات عشوائية متعددة من هذا المجتمع ولكن بمساعدة برامج الكمبيوتر تبدو عمليه سهله نسبيًا لأن معظم البرامج الإحصائية مثل SPSS, SAS تتضمن مولد الاعداد العشوائى وهو يعطى اعداد مختارة بالصدفة من توزيعات محددة مثل انتقاء مجموعات كبيرة من هذه الارقام للتحقق من قضية معينة وهذه الارقام أو الحالات النظرية يتطلب دائمًا ان تكون مستقلة وهذه العملية تسمى Pseudo-random number.

تستخدم استراتيجيات البرامج الكمبيوترية المعقدة فى الدراسات المنهجية التى تختبر أداء المقدرات الإحصائية تحت شروط مختلفة فعلى سبيل المثال البيانات المتولدة والنماذج يتم تقدير معالمها تحت شروط مثل البيانات الغائبة أو عدم تحقق شروط مثل الاعتدالية أو الخطية أو غيرها ويتم مقارنة هذه التقديرات فى ضوء مدى تحيز تقدير المعالم وتحيز الاخطاء المعيارية ومدى قدرة النموذج على اعطاء حلول متناسقة. وتستخدم استراتيجيات البرامج الكمبيوترية المعقدة بصورة محدودة فى تحديد حجم العينة ويمكن أن تستخدم أيضًا لتقدير القوة لأحجام عينات مختلفة، وعكس المداخل الأخرى المستخدمة لتقدير حجم العينة التى تستلزم مسلمات عديدة مثل توزيعات البيانات الاعتدالية، وهذه الاستراتيجيات لا تستلزم مسلمات خاصة بتوزيع البيانات ويمكن أن

تستخدم لدراسة تأثير عدم تحقق مسلمات على تقدير المعالم وبالتالى على الخطأ من النوع الأول والقوة الاحصائية وفترات الثقة.

نواحى القوه لاستراتيجيات البرامج الكمبيوترية المعقدة

يرى المؤيدين لهذه الاستراتيجيات كبديل حيوى للإجراءات الاستدلالية الكلاسيكية تتميز بالآتى:

1. الإحصاء الاستدلالى التقليدى يتطلب مسلمات توزيعية معينة مثل الاعتدالية ويكون التعميم من العينة إلى المجتمع محاط بالشكوك إذا لم تتحقق هذه المسلمات ولكن هذه الاستراتيجيات لا تتطلب توافر هذه المسلمات.

2. يتطلب الإحصاء الاستدلالى الكلاسيكى عينات عشوائية لدراسة مصداقية الاستدلال من العينة إلى المجتمع بينما إجراءات مداخل الكمبيوتر صادقة لأى نوع من البيانات سواء كانت عشوائية أو غير عشوائية.
3. تعالج أو تختبر العينات مثل المجتمعات الافتراضية لتوليد عينات كثيرة.
4. تستخدم عندما يوجد عدد كبير من المشاركين.
5. إجراءات الاستدلال الكلاسيكى لا تخبرنا عن مدى احتمالية مصداقية النتائج عبر المجتمع بينما استراتيجيات CI تساعدنا على اعادة البحث على عينات اخرى مثل استخدام استراتيجية الصدق التعميمى Cross-Validation و Bootstrap .

مظاهر الضعف أو القصور لاستراتيجيات البرامج الكمبيوترية المعقدة

1. التقدير فى هذه الاستراتيجيات يمثل اشكالية عندما تكون تباينات المجتمع غير معروفة، فى حين تفترض الاختبارات التقليدية تساوى التباينات.
2. يتشكك بعض الباحثين حول فوائد هذه الاستراتيجيات عند استخدام عينات غير عشوائية وغير ممثله للمجتمع، فلو ان العينة المستخدمة لتوليد التوزيع العينى لا تعكس خصائص المجتمع فإن صدق الاستدلال معرض للشكوك.
3. يعتقد البعض ان هذه استراتيجيات " تحاول ان تحصل على بعض الشىء من لا شىء " فنفس البيانات تستخدم مره تلو الأخرى.
4. البعض يتساءل عن مدى دقه هذه استراتيجيات ففى بعض المواقف فإن هذه الاستراتيجيات أقل دقة من الأساليب البارامترية التقليدى.

فى النهاية فإن استراتيجيات البرامج الكمبيوترية ليس علاج لكل المشاكل تمامًا مثل فهى تفترض ان الملاحظات فى العينة مستقلة وبالإضافة إلى انها تتطلب مسلمات بارامترية معينة.

التوصيات:

- انتقاء العينة فى ضوء كل المداخل بحيث يحقق قوة على الأقل 0.80.
- حجم العينة الكافى هو المكون الوحيد لجودة الدراسة.
- انتقاء العينة يتم فى ضوء جودة تصميم الدراسة وواقعيتها وثبات القياسات.
- يجب اعطاء وصف تفصيلى للعينة وذلك لفهم مجتمع الدراسة وليساعدنا ذلك على تعميم نتائج الدراسة، وعلى ذلك فلا بد أن تكون العينة ممثلة تمثيلًا جيدًا للمجتمع خاصة عند استخدام العينات غير العشوائية وهذا الوصف الكامل الوافى للعينة يساعدنا فى إجراء ما وراء التحليل للدراسات العديدة.
- يجب إجراء دراسة استطلاعية عندما نحتاج إلى تقدير تحليل القوة أو حدود الثقة.
- يجب استخدام مداخل عديدة لحساب حجم التأثير وأحجام العينات المناظرة لها.
- لاعتبارات اخلاقية والكفاءة فإن الباحث يجب أن يقيم القوة الإحصائية فى مراحل مختلفة من الدراسة والتوقف عن جمع البيانات إذا لم تتحقق القوة المرغوبة.
- اعتمد على عينات عشوائية لأن عدم العشوائية تضع قيودًا على تعميم النتائج وكذلك اعتمد على استخدام تحليل القوة وفترات الثقة لأن هذه الاستراتيجيات تفترض ان البيانات تأتى من خلال عملية عشوائية والتعميم من عينة غير عشوائية إلى المجتمع يؤدى إلى الوصول إلى استنتاجات متحيزة.

وقدم Isaac & Michaell (1995) مبررات على أهمية لاستخدام عينات كبيرة فى البحث:

- يمكن تقسيمها إلى عينات فرعية لمقارنة النتائج عبر هذه العينات.
- عندما يتم دراسة عدد كبير من المتغيرات لأنه توجد مخاطرة من البيانات المفقودة.

ومبررات لاستخدام العينات الصغيرة فى:

- فى البحوث الاستكشافية أو الدراسات الاستطلاعية.
- عندما يكون الضبط المنهجى عالى.

- عندما تحتاج إلى الدخول فى عمق الظاهرة مثل الدراسات الكيفية من خلال المقابلات.

الفصل الحادى عاشر

القوة الإحصائية وحجم التأثير

Statistical Power& Effect size

تعتبر الإحصاء العمود الفقرى لصناعة القرارات والوصول إلى المعلومات والمعارف فى العلوم النفسية والاجتماعية وغيرها. وتقوم نظرية الاستدلال الإحصائى على إستراتيجية اختبارات الفروض التى تقوم على صياغة الفروض الإحصائية وتحديد الاختبار الإحصائى المناسب ومستوى الدلالة الإحصائية ثم صناعة القرار فى ضوء قيمة p حيث إذا كانت $p>0.05$ نقبل عدم الدلالة الإحصائية (قبول الفرض الصفريH0) وإذا كانت $p \leq 0.05$ نقبل الدلالة الإحصائية (رفض الفرض الصفريH0).

ويعتبر Cohen (1988) القوة الإحصائية المؤشر الكمى للحساسية الإحصائية للبحث أو للتجربة، واعتبرها Sawyer & Ball (1981) مظهرًا أساسيًا لصدق الاستنتاج الإحصائى Statistical Conclusion Validity، ويعتبرها البعض مؤشر هام لجودة تصميم الدراسة. وعلى الرغم من أن القوة الإحصائية جزء مهم فى البحوث إلا أنه ما زال الاهتمام منصبًا على الدلالة الإحصائية (Cohen, 1990).

وقد بدأ الاهتمام حديثًا بتضمين القوة الإحصائية فى الدراسات للمساعدة فى تفسير النتائج لأنه لو فشل الاختبار فى إظهار التأثير أو العلاقة الموجودة فقد يكون هذا ليس بسبب عدم وجود تأثير ولكن ربما بسبب ضعف الاختبار فى إظهار التأثيرات الدالة إحصائيًا. ومعظم العمل الإحصائى القائم على تحليل البيانات يكون اهتمامه فى المقام الأول تقليل أو على الأقل ضبط احتمالية الوقوع فى الخطأ من النوع الأول بقدر الإمكان. وعند تصميم التجارب فإن الباحثون يميلون إلى تجاهل حقيقة فى غاية الأهمية وهو وجود الخطأ من النوع الثانى (β). فبينما يتعامل الخطأ من النوع الأول مع مشكلة الحصول على فروق وهى حقيقية غير موجودة فى المجتمع فإن الخطأ من النوع الثانى يكون بنفس القدر من الخطورة. وهو عدم وجود فروق وهى حقيقة موجودة،

والتجربة النموذجية هى التى تصل إلى وجود تأثير نبحث عنه وهو موجود بالفعل فى مجتمع الدراسة بأقل قدر من التكلفة والوقت.

مفهوم القوة الإحصائية

يعرف Pagano (2013) القوة الإحصائية للتجربة بأنها مقياس لحساسية التجربة"Sensitivity" للكشف عن التأثير الحقيقى للمتغير المستقل فنحن نقصد به ذلك التأثير الذى ينتج تغير فى المتغير التابع، ولو أن المتغير المستقل لا يؤدى إلى تغير فى المتغير التابع فنحن نقول أن المتغير المستقل ليس له تأثير حقيقى. والحصول على تأثير حقيقى يعنى رفض H0 و لذلك تعرف القوة بأنها قدرة الاختبار الإحصائى على رفض الفرض الصفرى إذا كانت للتجربة (المتغير المستقل) لها تأثير حقيقى بالفعل. وعليه فإن الاختبار القوى إحصائيًا هو ذلك الاختبار القادر على الحصول على نتائج دالة إحصائيًا (رفضH0) وهى موجودة بالفعل فى مجتمع الظاهرة. ويعرف Howel (2013) القوة بأنها الاحتمالية الصحيحة لرفض H0 وهو حقيقى مرفوض على مستوى المجتمع (الفرض البديل هو الحقيقى) و لذلك فإن:

$$P=1-\beta$$

والتجربة القوية هى التى تمتلك الفرصة الأفضل لرفض H0، وقيمة القوة الإحصائية تتراوح من الصفر حتى الواحد الصحيح ويقال إن التجربة القوية هى التى لها مستوى قوة 0.80 فأكثر. ويرى Pagano (2013) أن هذا يكون أكثر مرغوبية فى البحوث والدراسات ولكن هذا نادرًا فى العلوم السلوكية والقيم من 0.40 حتى 0.60 هى الأكثر شيوعًا ومن المهم تحديد القوة للتجربة فى ضوء:

- **عند تصميم التجربة:** لتحديد حجم العينة ويطلق عليه التحليل القبلى للقوة و التصميم.
- **بعد تحليل النتائج وتفسيرها:** وهذا مفيد فى معرفة الدقة الإحصائية للاختبار المستخدم و يعرف هذا بالتحليل البعدى Post hoc procedure ، أى بعد إجراء الاختبار.

وصناعة القرار فى نظرية اختبارات الفروض الكلاسيكية تتم بمقارنة قيمة p بمستوى الدلالة الإحصائية (α) وعلى ذلك يتم قبول الفرض الصفري(H0) أو رفضه (قبول الفرض الإحصائى البديل H_A). وقبول H0 لا يعنى عدم وجود فروق بالفعل ولكننا لم نمتلك الدليل والوسائل للرفض، وعلى ذلك فإن استخدام الإحصاء الاستدلالى يؤدى إلى الوقوع فى أخطاء قرارية على ذلك يوجد خطأين هما: الخطأ من النوع الأول (α) ويتم التحكم فى هذا الخطأ قبل إجراء الدراسة من قبل الباحث من خلال تبنى مستوى دلالة إحصائية معينة مثل 0.05 أو 0.01 و 0.001 وغيرها. والخطأ من النوع الثانى (β)، وفى الواقع يصعب التحكم فى هذا الخطأ قبل إجراء الدراسة ولا يستطيع الباحث التحكم فيه أو السيطرة عليه.

وترتبط القوة الإحصائية بالخطأ من النوع الثانى (β) وتعرف باحتمال تجنب حدوث (β) أو الوصول إلى التأثير الحقيقى باحتمال ($1-\beta$). على ذلك فالقوة الإحصائية مقياس للدرجة التى تكون من خلالها الدراسة قادرة على إظهار الفروق أو العلاقات الحقيقية الموجودة فى مجتمع الدراسة (Cohen, 1988)، وتعرف بأنها احتمالية رفض الفرض الصفرى المرفوض فى المجتمع، بمعنى تجنب حدوث الخطأ من النوع الثانى (β)، أو قدرة الاختبار على الحصول على دلالة إحصائية عندما يكون الفرض الصفرى غير حقيقى. فالقوة الإحصائية 0.80 = P تعنى وجود احتمال 80% للحصول على دلالة إحصائية وهى موجودة بالفعل بالتالى فإن فرصة ارتكاب (β) تساوى 0.2.

وعلى ذلك فإن دراسات تحليل القوة الإحصائية تجيب على تساؤل ما هى النسبة الاحتمالية لرفض H_0 ؟ ولا يوجد اتفاق بين الباحثين عن المستوى المقبول من القوة الإحصائية للدراسة. وهنا يوصى Cohen (1988) بأن على الباحثين الوصول إلى مستوى قوة لا يقل عن 0.80 حيث إنه مستوى مقبول وواقعى والقوة أقل من 0.80 تعنى ثقة منخفضة فى مصداقية القرار وعلى ذلك يكون أقصى خطأ من النوع الثانى β = 0.20، ويرى Verma & Goodale (1995) أن الخطأ α أشد خطورة مقارنة

بخطأ (β) ولتحقيق قوة معقولة لا بد أن تكون نسبة ألفا إلى بيتا هى (4 إلي1) أى أقصى ألفا يساوي0.05 وأقصى بيتا يساوي0.20.

العوامل المؤثرة فى القوة الإحصائية

القوة الإحصائية هى دالة وظيفية لثلاثة متغيرات هى مستوى الدلالة الإحصائية، حجم العينة، وحجم التأثير (Cohen, 1988; Hallahan & Rosenlhal, 1996)، وعبر عنها Cohen (1988) بالعلاقة الآتية:

$$P = F(\alpha, N, ES)$$

ويمكن عرض لأهم العوامل المؤثرة على القوة كالآتى:

1. حجم العينة(N): كلما زاد حجم العينة قلت أخطاء المعاينة بالتالى تزيد قوة الاختبار وزيادة حجم العينة يؤدى إلى أخطاء أقل وثبات أعلى واحتمالية عالية لرفض الفرض الصفرى والحصول على دلالة إحصائية ويشير Saywer & Ball (1981) أن العشوائية فى اختيار العينة وتجانسها يقلل الأخطاء المعيارية بالتالى يزيد من دقة القياسات.

2. الخطأ من النوع الأول (α): بجعل ألفا أقل صرامة تزيد قيمة القوة ، فبدلا من جعل ألفا 0.01 يمكن أن تكون 0.05 فأكثر، ذلك لأن العلاقة بين α و(β) علاقة عكسية، فإذا قلت قيمة α تزيد (β) وبدوره تقل القوة الإحصائية. ولكن إذا جعل الباحث α أكثر صرامة فإنه يتطلب زيادة حجم العينة ويرى البعض بأنه لا مانع من زيادة قيمة α حتى لو زادت عن 0.10 فى سبيل ان تكون القوة عالية (Sink & Mvududu, 2010).

3. حجم التأثير (ES): زيادة حجم الفروق بين المتوسطات أو قوة العلاقة بين المتغيرات يؤدى إلى احتمال كبير للحصول على دلالة إحصائية بالتالى إلى قوة إحصائية عالية، وذلك لأن حجم التأثير المرتفع يعنى القدرة العالية لإظهار أو الكشف عن الفروق الموجودة بالفعل فى المجتمع، وهذا يعنى رفض الفرض الصفرى والحصول على قوة عالية (Sink & Mvududu, 2010)، أى أن القوة الإحصائية تزيد عندما يكون للاختبار قدرة عالية على إظهار حجم تأثير كبير مقارنة بحجم تأثير صغير.

4. تصميم البحث: حدد (Baroudi & Orlikowski, 1989; Sink & Mvududu, 2010) عوامل تصميم البحث المؤثرة فى القوة وهى:

أ- **اختيار المتغيرات:** دقة الباحث وشموليته فى اختيار أهم المتغيرات المستقلة المؤثرة فى الظاهرة يزيد من القوة وليس اختيار متغيرات لا ترتبط بالظاهرة فإن هذا يؤدى إلى عدم دلالة إحصائية.

ب- **الثبات:** فالثبات المرتفع للقياسات يؤدى إلى قوة إحصائية عالية Zimmerman (& Williams,1989).

ج- **توزيع مفردات العينة:** فالتوزيع المتساوى على المجموعات المتعددة أو مستويات العامل يزيد القوة، وأن التوزيع غير المتساوى على مجموعات التصميم العاملى يؤدى إلى تقليص الارتباطات أو الفروق، ومن ثم تقليص حجم التأثير وهذا بدوره يؤدى إلى قوة منخفضة.

د- **المعالجة التجريبية:** إجراء معالجات وإجراءات تجريبية قوية ومعيارية وكذلك تصميمات القياسات المتكررة يؤدى إلى قوة عالية.

هـ- **الاختبار الإحصائى:**الاختبارات الإحصائية البارمترية أكثر قوة من الاختبارات اللابارامترية وذلك لأن تقطيع المتغيرات المتصلة الاعتدالية ينشأ عنها فقد كثير من المعلومات، كما أن الاختبارات الإحصائية الموجهة (ذو اتجاه واحد إلى أعلى أو أدنى) أكثر قوة من الاختبارات غير الموجهة ذو اتجاهين أو ذيلين (Sink & Mvududu ,2010). كما ان استخدام الاختبار فى المقارنة بين أكثر من مجموعتين فى تحليل واحد أفضل من المقارنة بين كل مجموعتين على حدة Omnibus test لأن هذا يزيد من القوة (Hallahan & Rosenthal, 1996).

و- **دراسات ما وراء التحليل:** دراسات ما وراء التحليل أكثر قوة وأوثق من الدراسات المفردة.

ز- استخدام بدائل الاختبارات الإحصائية الكلاسيكية: استخدام إجراءات استدلالية رياضية مثل إحصاء بيزن Bayesian أكثر مرونة لاختبار الفروض الكلاسيكية ثنائية القطبية فى القرار، ومن ثم تصبح عملية صنع القرار أكثر مرونة ومنطقية فى رفض الفرض الصفرى ويكون بذلك أكثر قوة من الإحصاء الكلاسيكى.

مداخل تحليل القوة الإحصائية

توجد عدة مداخل لتناول تحليل القوة الإحصائية أهمها:

1. قبل إجراء الدراسة Prospective or prior power analysis: فى هذا النوع من التحليل يتبنى الباحث مستوى قوة إحصائية اللازمة للوصول إليها ولتكن P= 0.80، ومستوى دلالة إحصائية معينة مثلًا 0.05 = α، وحجم تأثير مفترض فى ضوء تراث الظاهرة أو دراسة استطلاعية. ويكون هدف الباحث تحديد حجم العينة المطلوب لتحقيق قوة إحصائية محددة مسبقا قبل إجراء الدراسة وهذا التحليل خطوة أساسية فى مرحلة تصميم الدراسة وهذا التحليل للقوة يفيد فى تقليل التكلفة والموارد نتيجة زيادة حجم العينة بدون داعى ويوصى Cohen (1992) بأهمية إجراء هذا النوع من التحليل للدارسة قبل إجرائها.

2. بعد إجراء الدراسة Retrospective or Post hoc Power analysis: وهنا يجرى الباحث تحليل القوة بعد الانتهاء من الدراسة للتقييم البعدى لنتائج الدراسة والتعرف على احتمالية الوصول إلى تفسيرات واستنتاجات صحيحة، ويتم تقدير القوة خاصة للنتائج غير الدالة إحصائيًا وذلك لأن عدم الدلالة قد لا يعود لانعدام الدلالة الفعلية فى المجتمع (تأثير حقيقى صغير)، بل إن الاختبار الإحصائى له قوة إحصائية ضعيفة غير قادرة على كشف الفروق الموجودة فى المجتمع ويتم تقدير القوة للاختبار من خلال معرفة مستوى الدلالة الإحصائية α وحجم التأثير وحجم العينة.

3. أثناء الدراسة Statistical Power during study: الاستخدام الثالث لتحليل القوة فى مجال تقيم البرامج أو التجارب عندما يستخدم تصميم القياسات التتبعية،

فالسؤال الذى يطرح نفسه هو: إلى متى تستمر إجراءات البرنامج أو الدراسة لتحقيق مستوى قوة مرضى؟

4. تحليل الحساسية Sensitivity analysis: ينصب الاهتمام فى تحليل الحساسية على تقدير حجم التأثير فى المجتمع بمعلومية α و P وN، ويكون مفيد فى تقييم الدراسات المنشورة وهذا يجيب على تساؤل وهو: ما هو حجم التأثير الذى بإمكان الاختبار تحقيقه فى ظل قوة إحصائية تساوى 0.80=؟ ويتم تقدير الحساسية قبل إجراء الدراسة.

5. تحليل المحكمية Criterion analysis: فى تحليل المحكمية يتم تقدير α كوظيفة للقوة وحجم التأثير وحجم العينة.

تقدير القوة الإحصائية

يتم تقدير القوة الإحصائية بالطرق الآتية:

1. جداول Cohen (1988): قدم Cohen (1988) جداول يتم من خلالها تقدير القوة لمعظم الاختبارات الإحصائية مثل T, χ^2, r, F, Rوغيرها، وذلك بمعلومية مستوى الدلالة الإحصائية وحجم العينة وحجم التأثير.

2. البرامج الإحصائية: ومنهاPower & Precision، G-power، NQuery، Stat، PASS، NCSS، وحدد Thomas & Kerbs (1997) المواصفات المثالية لأى برنامج وهى:

- يغطى اختبارات عديدة بسيطة أو متدرجة، وبارامترية ولابارامترية.
- يسمح بمرونة فى التعامل مع المواقف الجديدة.
- يعطى نتائج دقيقة ويسهل تفسير مخرجاته.
- يقدر القوة وحجم التأثير وحجم العينة أى يغطى معظم تحليلات القوة.
- يقدر معالم المدخلات اللازمة لتقدير القوة مثل حجم التأثير.
- يأخذ فى حسبانه مؤشرات متنوعة لحجم التأثير.

- يتميز بالجودة فى عرض الأشكال أو الرسومات أو الجداول.
- يسهل توثيقه.

وتم إجراء الدراسات السابقة الخاصة بالقوة الإحصائية منذ زمن بعيد **ففى مجال علم النفس** قام Cohen (1962) فى دراسته المشهورة بتقدير القوة الإحصائية للدراسات المنشورة فى مجلة علم النفس الاجتماعى وغير العاديين عام 1960. وتوصل إلى أن متوسط القوة الإحصائية للاختبارات التى لها حجم صغير 0.18، ولحجم تأثير متوسط 0.48، ولحجم تأثير كبير 0.83، ومتوسط القوة العام 0.48، وهذا يعنى أن نصف النتائج تعطى عدم دلالة إحصائية بالرغم من وجود فروق حقيقية ولكن الاختبارات غير قادرة على إظهارها. وبعد 24 عامًا أعاد Sedlmeier & Gigerenzer (1989) دراسة كوهين على 54 دراسة منشورة فى نفس المجلة فى مجلد 1984، وتوصل إلى أن دراستين أشارت إلى القوة الإحصائية ولم تقدر أى منهما القوة وبلغ وسيط القوة لأحجام التأثير الصغيرة والمتوسطة والكبيرة 0.14 ,0.44 ,0.90 على التوالى، أى بعد 24 عامًا فإن وسيط القوة أقل قليلًا عما كان عليه فى عام 1962 وهذا يدل على أنه بمرور الزمن فإن قوة الاختبار تقل أو ثابتة نسبيًا إلى حد ما.

وقدر Rossi (1990) القوة الإحصائية لـ 25 دراسة منشورة فى مجلة علم النفس غير العادين ومجلة علم النفس الاجتماعى والشخصية وتوصل إلى أن متوسط القوة لحجم تأثير صغير 0.26، ولحجم تأثير متوسط 0.71، ولحجم تأثير كبير 0.85.

وفى مجال التربية قام Brewer (1972) بتقدير القوة الإحصائية فى ثلاث مجلات تربوية وتوصل إلى أن وسيط القوة فى المجلة الأمريكية للبحوث التربوية لأحجام التأثير الصغيرة 0.14، للمتوسطة 0.58، وللكبيرة 0.78. ولمجلة البحوث فى التدريس (التعليم) هي 0.22 ,0.71 ,0.87 على التوالى ولمجلة البحث ربع سنوية هى: 0.80, 0.52, 0.14 على التوالى.

وفى مجال التربية العلاجية قدر Woolley (1983) القوة الإحصائية للدراسات المنشورة فى مجلة التربية العلاجية فى المدة من 1980 حتى 1988 ، وتوصل إلى أن

91 % من الدراسات التى لها حجم تأثير صغيرة تمتلك قوة أقل من 0.50، و25% من الدراسات التى لها حجم تأثير متوسط لها قوة أقل من 0.5 ، و 4% من الدراسات التى لها حجم تأثير كبير لها قوة أقل من 0.50، ووسيط القوة لحجم تأثير صغير هو 0.14، ولحجم تأثير متوسط 0.71، وحجم تأثير كبير 0.99.

وفى مجال الإدارة حسب Mazen, Graf, Kellogy & Hemmasi, (1987) متوسط القوة الإحصائية للدراسات المنشورة فى عدة مجلات للإدارة وكانت لحجم تأثير صغير 0.31، ولحجم تأثير متوسط 0.77، ولحجم كبير 0.86.

وقدر Verma & Goodle (1995) متوسط القوة لدراسات فى مجلة الإدارة التطبيقية ومجلة علوم القرار وتوصل إلى أن 60% من الاختبارات لها قوة أقل من 0.80. وكانت الاختبارات التى لها أعلى قوة إحصائية هى الانحدار يليه الارتباط ثم اختبار T المستقلة.

وتوصل Casher & Geiger (2004) فى تقيميه لـ 53 دراسة فى مجال الإدارة إلى أن 4 دراسات (9.3%) ناقشت القوة ومن هذه الدراسات الأربعة ثلاثة دراسات قدرت القوة الإحصائية تراوحت من 0.48 إلى 0.98، وأن متوسط القوة لـ 77 اختبار هى 0.29.

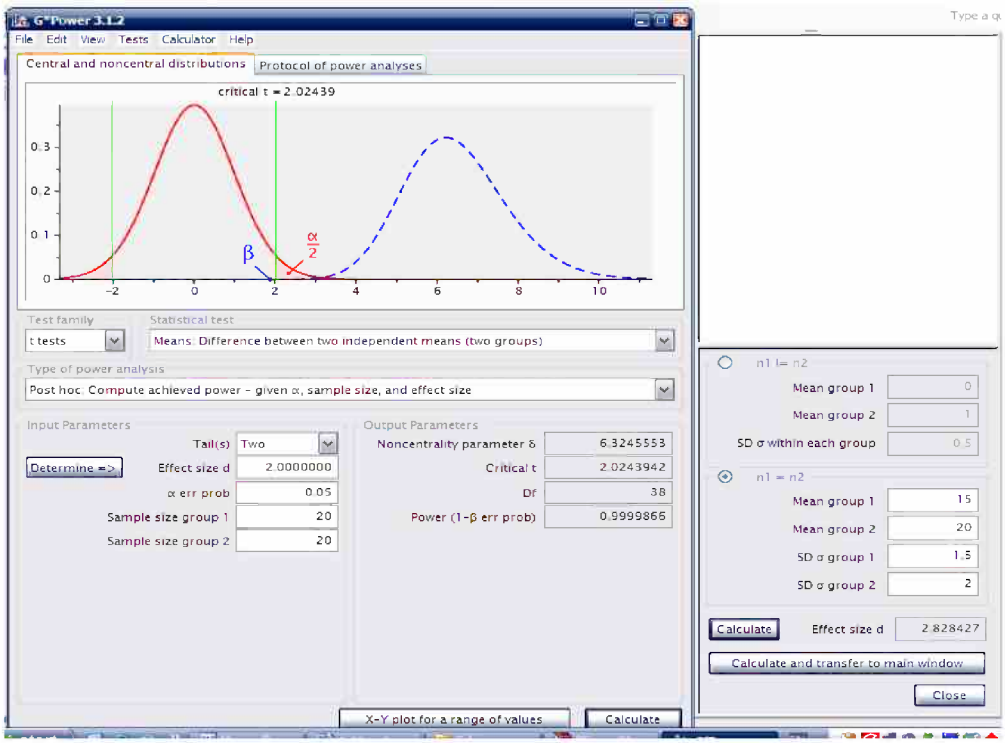
وفى البيئة العربية قام النجار (2006) بتقدير قوة الاختبار فى مجلة الإدارة فى المملكة العربية السعودية وبلغ متوسط القوة لحجم تأثير صغير 0.19، ولحجم تأثير متوسط 0.79، ولحجم تأثير كبير 0.93.

فى الدراسات النفسية التربوية توصل عامر(2014) وسيط القوة الإحصائية للاختبارات الإحصائية 0.84 وهذا يعنى أن الاختبارات الإحصائية أو بمعنى اشمل الدراسات النفسية العربية لديها احتمال 84% للكشف عن الفروق الحقيقية عندما تكون موجودة وأن الخطأ من النوع الثانى هو 0.16 وهى نسبة مقبولة حيث لا يفضل أن تزيد عن 0.20 وعلى ذلك يمكن القول إن الدراسات النفسية تمتع بمستوى قوة إحصائية مرغوب.

ويتضح أن الدراسات السابقة اتبعت منهجية تقدير القوة الإحصائية لكل الاختبارات معًا في ضوء ثلاثة مستويات من حجم التأثير الضعيف والمتوسط والكبير وذلك في ضوء محكات Cohen (1988) لحجم التأثير، ويلاحظ أن الدراسات التى لها حجم التأثير الضعيف والمتوسط لم تصل إلى المستوى المرغوب من القوة وهو 0.80 وذلك لجميع التخصصات.

تقدير حجم التأثير والقوة الإحصائية

لتقدير القوة الإحصائية للاختبار يستلزم تحديد ثلاثة معالم وهى مستوى الدلالة الإحصائية وحجم العينة وحجم التأثير ولكن حجم التأثير ولم يقدر من قبل صاحب الدراسة وعلى هذا قام الباحث الحالى بالاعتماد على برنامج G-Power (Faul, Erdfelder, lang & Buchner, 2007) لاستخدامه فى غرضين هما تقدير حجم التأثير وتقدير قوة الاختبار، وهذا البرنامج يتكون من شاشة بها قوائم ثم مربع أبيض لعرض الرسومات، وبعد ذلك تظهر أيقونة هى Test Family، وتتكون من Exact (تتضمن معامل ارتباط بيرسون والانحدار) وعائلة F-Test (وتتضمن تحليل التباين و التغاير و MANOVA)، وعائلة T-test تتضمن اختبار (ت) والاختبارات اللابارامترية) وعائلة Z. واعتمد الباحث الحالى على عائلة T-test وعائلة Exact. وبعد ذلك تم تحديد نوع الاختبار ونوع تحليل القوة، كما فى الشكل الآتى:



الشكل (1.11): مخرج برنامج G-Power فى حالة اختبار "ت" للعينات المستقلة.

وبرنامج G-power قائم على معادلات Cohen (1988) لتقدير حجم التأثير يأخذ بحدود كوهين لتحديد أما إذا كان حجم تأثير صغير أو متوسط أو كبير وهى فى حالة معامل ارتباط بيرسون والانحدار 0.01 صغير و 0.09 متوسط و 0.25 كبير، وفى حالة اختبار (ت) لعينتين مستقلتين 0.2 صغير و 0.5 متوسط و 0.8 كبير ثم أدخلت قيمتى حجم التأثير والقوة إلى برنامج SPSS فى ملف البيانات.

والفشل فى رفض الفرض الصفرى يترك الباحث فى حيرة فهل يرجع هذا إلى قوة إحصائية غير كافية للاختبار لرفض الفرض الصفرى أم إلى عدم وجود فروق حقيقية للظاهرة ؟ ولكن ماذا لو كانت القوة الإحصائية للدراسة أقل من 0.80 ؟ ربما يرجع هذا إلى أن الفروق الحقيقة موجودة وأن الاختبار ليس لديه القوة الكافية لإظهارها وهذا نتيجة إلى عيوب فى انتقاء الاختبار، والتحقق من مسلماته أو القياسات أو فى تصميم الدراسة أو غيرها من الأمور المنهجية والإحصائية، وقد يرجع انخفاض القوة إلى أن الفروق الحقيقة صغير جدًا فى الواقع الحقيقى وبالتالى فهذا ليس عيبًا فى الدراسة، إنما هو طبيعة الظاهرة تحت الدراسة.

ولكن هل توجد استراتيجيات للتعامل مع القوة المنخفضة؟ فيرى البعض أنه لزيادة القوة الإحصائية يجب زيادة مستوى الدلالة الإحصائية α لأنه من شانه أن يقلل مستوى الخطأ من النوع الثانى (β) ولكن Daniel (1993) يرى أنه حلًا غير مرغوبًا كما يجب معرفة حجم التأثير للظاهرة تحت الدراسة من خلال تقدير حجم التأثير قبل إجراء الدراسة عن طريق مراجعة التراث وأيضا يمكن زيادة القوة من خلال زيادة أحجام العينات لأنها الضمانة الأكيدة لزيادة قوة الاختبار.

ومجال دراسات القوة الإحصائية فى البيئة العربية يكاد يكون محدود جدًا خاصة فى الدراسات النفسية واوصى عامر (2014) عدة توصيات أهمها:

- على الباحثين حساب القوة الإحصائية لكل اختبار إحصائى فى الدراسة وهذا يسمح للباحثين تفهم المخاطر الناتجة من عدم الدلالة الإحصائية للنتائج. فالدراسات مع قوة إحصائية عالية والتى ليس لها دلالة إحصائية تعطى تدعيم قوى للقرار، بينما الدراسات مع قوة منخفضة تعطى تدعيم ضئيل سواء للفرض الصفرى أو البديل.
- على محررى المجلات التأكيد على ضرورة تقدير القوة الإحصائية على الأقل بعد إجراء الدراسة كشرط للنشر فى المجلات باعتباره مؤشر للجودة الإحصائية والمنهجية للدراسة.

حجم التأثير

Effect Size

لا أحد ينكر أهمية اختبارات الفروض الإحصائية الصفرية Null Statistically Hypotheses Testing (NSHT) كجزء ضرورى لتحليل البيانات فى الاستدلال الإحصائى كما أنها تلعب دورًا كبيرًا فى اتخاذ القرارات فهى استراتيجية صناعة القرار الإحصائى ولكنها غير كافية لتفسير بيانات العلوم الاجتماعية (Cohen, 1994) وعلى الرغم من الانتقادات الموجهة إلى اختبارات الدلالة الإحصائية إلا أن الباحثون أصبحوا أسيرين لها دون الأخذ فى الاعتبار مؤشرات أخرى تعبر عن جودة البحث مثل حجم التأثير وقوة الاختبار وانتقد Schmidt & Hunter (1997) اختبارات الفروض

الإحصائية الصفرية لكونها تعوق نمو المعرفة العلمية ولا تقدم إسهامات ايجابية، فالدلالة الإحصائية تعنى وجود التأثير ولكنها لا تخبرنا بشيء عن حجم هذا التأثير، ومن أهم الانتقادات الموجهة إلى اختبارات الفروض الإحصائية الكلاسيكية:

1. تأثرها الواضح بحجم العينة ، فكلما زاد حجم العينة يزداد احتمال الحصول على نتائج دالة إحصائية (Kirk, 1996).

2. لا تعطى القيمة المعنوية أو التطبيقية لنتائج الدراسة، أى أنها تخبرنا عن القرار ولكنها تعجز أن تخبرنا عن معنوية الفروق وحجمها.

واشار Rosnow & Rosenthal (1989) إلى ثلاثة عوامل فى أزمة تحليل البيانات وهى:

- الاعتماد الكلى على الدلالة الإحصائية وهى تعطى قاعدة قرارية ثنائية (قبول – رفض) وهذا يقود إلى خط صارم بين الدلالة وعدم الدلالة الإحصائية، وهذا ما يشار إليه بأمزوجة التصنيفClassification Fallacy. وهذا القرار التصنيفى غير مناسب للوصول إلى استنتاجات من البيانات.
- كثيرًا من الدراسات تجرى باختبارات لها قوة إحصائية منخفضة وهذا يؤدى إلى الفشل فى أن تصل القيمة الاحتماليةP إلى رفض الفرض الصفرى.
- تقدير أهمية النتائج فى ضوء الدلالة الإحصائية فقط دون الأخذ فى الاعتبار مؤشرات أخرى.

وأشار Nickerson (2000) بوجود خطأ فى تفسير الدلالة الإحصائية وهو الاعتقاد السائد بأن القيمة الصغرى لـ P تعنى أن تأثير المعالجة كبير والقيمة الكبيرة لـP تعنى أن تأثير المعالجة صغير، أى أن الدلالة الإحصائية تعنى الدلالة التطبيقية والعملية للدراسة وهذا ليس صحيحًا بمعنى أن القيمة الصغيرة لـ P تعنى أن تأثير المعالجة تحت الدراسة كبير لأنه قد يكون قيمة P لتأثير معالجة ما دالة إحصائيًا عندα=0.01 ولكن ليس لها دلالة عملية بمعنى يكون حجم التأثير صغير.

وعـدم الدلالـة الإحصـائية لنتـائج الدراسـات يعطـى انطباعًـا عنـد المحكمـين ومحـررى المجـلات بضـعف هـذه الدراسـات ولـذلك يجاهـد البـاحثون بشـتى الطـرق للحصـول علـى دلالـة إحصـائية حتـى لـو كانـت تافهـة (الصـياد، 1988). فالدلالـة الإحصـائية تـؤدى إلـى الحصـول علـى سـطحية القـرار وتفتقـد إلـى عمـق القـرار وهـذا مـا أدى إلـى ظهـور المقولـة "The end of statistical significance" ودعـا Schmidt (1996) إلـى هجـر وتـرك اختبـارات الفـروض الصـفرية واحلالهـا بحجـم التـأثير وفتـرات الثقـة فـى صـناعة القـرارات الإحصـائية. وعلـى الـرغم مـن الانتقـادات الموجهـة لاختبـارات الفـروض الإحصـائية إلا أنـه لا يوجـد إطـار بـديل مقتـرح لهـا ونتيجـة لهـذه الانتقـادات أوصـت كثيـرا مـن المؤسسـات والبـاحثين باسـتخدام حجـم التـأثير Effect size كجـزء مكمـل لاختبـارات الفـروض لأنهـا تمـدنا بمعلومـات لا تسـتطيع اختبـارات الفـروض الكلاسـيكية إعطاءهـا (APA, 1994; Kirk, 1996; Thompson 1998). وحتـى أن الجمعيـة النفسـية الأمريكيـة APA أوضـحت أنـه عنـد اسـتخدام NHST فـإن تقـدير حجـم التـأثير وفتـرات الثقـة هـو الحـد الأدنـى المتوقـع فـى تقريـر الدراسـات. وعلـى الـرغم مـن هـذه التوصـيات إلا أن اسـتخدام الدلالـة الإحصـائية اسـتمر مسـيطرًا علـى التحلـيلات الإحصـائية فـى الدراسـات التربوية والنفسية.

الدلالة الإحصائية وحجم التأثير

توجد العديد من الفروق بين الدلالة الإحصائية والدلالة العملية كالتالى:

الجدول (1.11): الفروق بين الدلالة الإحصائية وحجم التأثير.

الدلالة الإحصائية	حجم التأثير
• تجيب عن هل التأثير الملاحظ حقيقى أم يعود للصدفة؟ • المظهر الأولى لاختبارات الفروض الكلاسيكية. • لا تعطى الدلالة التطبيقية للقرار. • القرار فى ضوئها قطبى ثنائى (رفض – قبول). • تتأثر بحجم العينة. • يتم تقديرها مسبقا قبل إجراء البحث أو من خلال البيانات. • لا يمكن إجراء اختبارات الفروض إلا بدونها. • قيم p إذا كانت < 0.001 دليل قاطع ومن 0.001 إلى 0.01 تعطى دليل جوهرى و من 0.01 إلى 0.05 دليل مقبول و 0.05< لا دليل على الدلالة الإحصائية	• تجيب عما حجم هذا التأثير أو قوة العلاقة؟ • المظهر المكمل لاختبارات الفروض الكلاسيكية. • تعطى القيمة التطبيقية والعملية للقرار. • القرار فى ضوئها على متصل كمى. • لا تتأثر بحجم العينة. • يتم تقديرها قبل إجراء الدراسة وأيضا بعد إجرائها. • يمكن إجراء اختبارات الفروض بدونها. • قيم d إذا كانت < 0.2 حجم تأثير ليس له قيمة عملية ومن 0.2 إلى 0.49 حجم تأثير صغير ومن 0.5 إلى 0.79 حجم تأثير متوسط و 0.8 فأكثر حجم تأثير كبير.

وعلى ذلك فإن الاعتماد على قيمة P كمؤشر للدلالة الإحصائية غير كافٍ لصناعة قرار. فلو اعتمدت الدراسة على عينة كبيرة فمن المحتمل الحصول على دلالة إحصائية فى حين لو اعتمدت على عينة صغيرة من الأفراد (قوة غير كافية) فمن المحتمل ان تؤدى قيمة P إلى قرار غير صحيح ، وعلى ذلك لا بد من تضمين مؤشر آخر بجانب

اختبارات الفروض حتى نصل إلى الطمأنينة والثقة فى صناعة القرار. وعلى الرغم من أهمية استخدام حجم التأثير فى اختبارات الفروض إلا أن تطبيقية فى الدراسات والبحوث النفسية والتربوية ما زال محدودًا وقد يعود ذلك إلى غياب تدريس هذا الموضوع فى مقررات الماجستير أو الدكتورة أو نتيجة إلى عدم الاهتمام به فى المجلدات الإحصائية أو عدم احتواء البرامج الإحصائية على مؤشرات عديدة لتقدير حجم التأثير.

وان كان استخدامها محدودًا فى الاختبارات الشائعة مثل اختبارات النموذج البسيط مثل "T" أو "F" أو معاملات الارتباط أو غيرها. فإنها تستخدم بصورة محدودة جدًّا فى الاختبارات المتدرجة مثل تحليل التباين المتدرج MANOVA.

مفهوم حجم التأثير

أخذ مفهوم حجم التأثير Effect size (ES) مسميات مختلفة فى التراث البحثى منها قوة العلاقة Strength of association أو قوة الارتباط Strength of relation أو مقدار التأثير magnitude of effect أو حجم التباين المفسر Explained variance والمناسبة العملية Practical Relevance ويرى Richardson (2011) أن مصطلح تأثير effect مرتبط باستخدام معالجات مختلفة فى البحث التجريبى ولكن المسمى الأكثر انتشارًا فى التراث هو حجم التأثير.

وتعرف بأنها تقدير لدرجة وجود الظاهرة فى المجتمع (Cohen, 1988)، أو تكميم لحجم الفروق أو قوة العلاقة بين متغيرين فأكثر وبصفة عامة حجم النواتج ويعرفها Olejnik & Algina (2003) بأنه مؤشر معيارى وتقدير للمعلم ومستقل عن حجم العينة ويكمم مقدار أو حجم الفروق أو العلاقة بين المتغيرات فى المجتمع. ويعطى حجم التأثير مؤشرًا للدلالة العملية لنتائج الاختبارات الإحصائية أو بصفة عامة لنتائج الدراسة وكلما زادت قيمة حجم التأثير فإن ذلك يعنى تواجد الظاهرة بدرجة كبيرة.

أهمية حجم التأثير

لحجم التأثير أهمية كبيرة فى مجال الدراسات وهى كالآتى:

- طريقة لتكميم فعالية البرنامج أو التجربة على متصل.
- تساعد الباحثون على تفسير النتائج.
- تستخدم قبل إجراء الدراسة فى مرحلة التخطيط لترشدنا فى تقدير حجم العينة.
- تساعد على التخلص من عيوب الدلالة الإحصائية أى أنها مكمل للدلالة الإحصائية عند صناعة القرار (Thompson, 1998).
- مؤشر لمدى تطبيق نتائج الدراسة فى الواقع الحقيقى عكس الدلالة الإحصائية.
- تستخدم لدمج أو مقارنة نتائج الدراسات لمجال معين وذلك لعدم تأثرها بحجم العينة أو ظروف الدراسة وهو ما يطلق عليه بأسلوب ما وراء التحليل أى إنها المادة الخام لأسلوب ما وراء التحليل (الصياد، 1988) .
- تشكل مكون هام فى مجال نشر البحوث فى المجلات العلمية حيث تشترط معظم الدوريات العلمية تقدير حجم التأثير كشرط للنشر (Kirk, 2004).
- تعتبر مكون أساسى فى تحليل أو تقدير قوة الاختبار الإحصائية قبل وبعد إجراء الدراسة (Cohen, 1988; 1994).

وعلى ذلك يتضح الفوائد العديدة من تقدير حجم التأثير وهى إعطاء القيمة التطبيقية للنتائج واستخدامها فى تقدير حجم العينة اللازمة لتحقيق مستوى قوة إحصائية معينة وايضًا الدور الهام الذى تلعبه فى تطبيق أسلوب ما وراء التحليل للدارسات فى مجال معين.

وشدد Rosnow & Rosenthal (1989) على ضرورة تقدير حجم التأثير عند تقدير للنتائج الدالة إحصائيًا وغير الدالة إحصائيًا وهذا عكس ما يعتقده الباحثين من أن تقدير حجم التأثير يكون للنتائج الدالة إحصائيًا فقط.

تصنيف مؤشرات حجم التأثير Effect size Taxonomy

تناول العديد من الباحثين تصنيفات مختلفة لمؤشرات حجم لتأثير أهمها:

1. تصنيف (Kirk, 1996; Richardson, 1996) وفيه قسمت مؤشرات حجم التأثير إلى:

أ- مؤشرات الفروق المعيارية بين المتوسطات: وتتضمن مؤشر كوهين d وجلاس Δ (دلتا) وهيدجز (g) وهو تحدد حجم الفروق ليتضمن متغير تصنيفى مستقل ومتغير تابع متصل (اختبار T المستقلة).

ب- مؤشرات قوة العلاقة أو التباين المفسر مثل مؤشرات R^2 و $\eta 2$ (ايتا) و ω^2 (اوميجا) و ابسيلون و r^2 و φ^2 (فاى) وغيرها. وهى مؤشرات تحدد نسبة التباين المفسر فى المتغير التابع نتيجة التغير الحادث فى المتغير المستقل.

2. صنفها Fren & Monroe (1996) **إلى:**

أ- مؤشرات الفروق بين المتوسطات : تتضمن مؤشرات d و Δ و g.

ب- مؤشرات العلاقة وتتضمن:

- **المؤشرات الارتباطية:** وتتنوع حسب طبيعة البيانات المحللة وأشهرها معامل ارتباط يرسون (r) ومعامل ارتباط سيرمان (ρ) ومعامل ارتباط Point-biserial (r_{pb}) ومعامل ارتباط فاى (φ) ومعامل ارتباط كرامير Crammar (V) وأخيرًا مؤشر η وهى صيغة تعميمة لـ r وغيرها.
- **مؤشرات التباين المفسر:** تستخدم فى تصميمات تحليل التباين مثل η^2 و $\omega 2$ و R^2 و R^2_{adj} واختبار Interclass(Icc) وغيرها.

3. صنفت إلى عائلتين هما:

أ. عائلة d: تتضمن مؤشرات d , Δ , gوغيرها.

ب .عائلة r: وتتضمن مؤشرات r, r_{pb}, φ , $\eta 2$, $\omega 2$.

4. تصنف مقاييس حجم التأثير إلى:

أ. **مقاييس معيارية:** هى مقاييس حرة من وحدات القياس ويتم تقديرها فى ضوء الانحراف المعيارى للبيانات ويمكن لهذه المقاييس مقارنتها ببعض أو على الأقل تحويل أحدهما للآخر ويندرج تحتها كل المؤشرات السابقة.

ب. **مقاييس غير معيارية:** تقدر فى ضوء وحدات خام بدون الأخذ فى الإعتبار الانحراف المعيارى للبيانات وهى غير مستقلة (حرة) عن مستوى القياس مثل الفروق بين المتوسطات أو الفروق فى الوسيط أو الفروق فى النسب المئوية وغيرها.

اولًا: مؤشرات حجم الفروق المعيارية بين المتوسطات: وهى تحدد مقدار الفروق بين مجموعتين وهى أكثر ملائمةً للتصميم التجريبى ذى المجموعتين ومن أهم هذه المؤشرات:

1. مؤشر كوهين d لـ Cohen (1988): أكثر مؤشرات حجم الفروق شيوعا لسهولة تفسيره ومعياريته ويستخدم بكثرة فى دراسات ما وراء التحليل والتصميمات التجريبية وشبة التجريبية ويتحدد بالمعادلة الآتية:

$$d = \frac{\overline{x_1} - \overline{x_2}}{\sigma_{pooled}}$$

$\overline{x_1}$ متوسط المجموعة التجريبية، $\overline{x_2}$ متوسط المجموعة الضابطة و σ_{pooled} الانحراف المعيارى المشترك للمجتمع ويقدر كالآتى:

$$\sigma p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}$$

حيث n_1, n_2 أفراد عينة المجموعة التجريبية وأفراد عينة المجموعة الضابطة على التوالى و s^2_1, s^2_2تباين درجات المجموعة التجريبية والمجموعة الضابطة ويفضل استخدام هذا المؤشر إذا توفر التوزيع الاعتدالى للمتغير التابع (المتصل) وايضا تساوى تباين المجموعتين، وتقدر قيمة d من قيمة اختبار (T) للعينات المستقلة إذا كان $n2=n_1$ كالآتى:

$$d = \frac{2T}{\sqrt{df}}$$

حيث df درجات الحرية وتساوى n1+n2 -2 ، وفى حالة n1≠n2 فإن المعادلة السابقة تصبح كالآتى:

$$d = T\left(\frac{n_1 + n_2}{\sqrt{df(n_1 n_2)}}\right)$$

الصيغتين السابقتين تصلح فى الدراسات التى تذكر قيمة الاختبار T فقط ويمكن تقدير قيمة d فى حالة استخدام اختبار F للمقارنة بين مجموعتين كالآتى:

$$d = \frac{2\sqrt{F}}{\sqrt{df_{error}}}$$

حيث df_{error} هى درجات حرية الخطأ (بين المجموعات) وتساوى N-C حيث N حجم العينة و C عدد المجموعات وتساوى 2. وتوجد إشكالية عند حساب مؤشر d فى حالة المقارنات المتعددة المصاحبة لتحليل التباين (توكى وشيفية وغيرها) حيث غالبًا لا يوجد تجانس بين تباينات المجتمعات عند مقارنة كل مستويين ببعضها البعض (Ferguson, 2009).

2. مؤشر جلاس Δ (دلتا) لـ (Glass (1976): اقترحه Glass (1976) كمؤشر لحجم التأثير مشابهة لمؤشر d وتتحدد قيمته بالمعادلة الآتية:

$$\Delta = \frac{\overline{x_1} - \overline{x_2}}{\sigma_c}$$

حيث σc الانحراف المعيارى لمجتمع المجموعة الضابطة.

ونلاحظ أن Glass استعان بالانحراف المعيارى للمجموعة الضابطة، فى حين اعتمد Cohen على الانحراف المعيارى المشترك للمجموعتين فى المجتمع. ويفترض مؤشر Δ تساوى تباينات المجموعتين الضابطة والتجريبية ($S^2c = S^2e$) وفى حالة عدم تساوى فإن مؤشر دلتا متحيز تحيز موجب أى يعطى تقديرًا عاليًا لحجم التأثير .

ولكن هذا المؤشر أقل تطبيقًا فى الدراسات التى لا تتضمن مجموعة ضابطة (مثل المقارنة بين الذكور والإناث) (Ferguson, 2009). ويمكن استخدام مؤشر Δ فى حالة تصميم المجموعة الواحدة وقياس قبلى وبعدى حيث يستخدم اختبار T المرتبطة ويكون المقام هو الانحراف المعيارى للقياس القبلى (Lipsy & Wilson, 2001).

3. مؤشر هيدجز (g) لـ(Hedges (1981: عدل Hedges (1981) مؤشر دلتا لـ Glass، حيث تخلص من محدداته واقترح إحلال الانحراف المعيارى للمجموعة الضابطة بالانحراف المعيارى المشترك لعينيتى المجموعتين وتتحدد قيمة g من المعادلة الآتية:

$$g = \frac{\overline{x_1} - \overline{x_2}}{S_P}$$

وتتحدد قيمة S_P كالآتى:

$$S_P = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

ويفترض مؤشر g ان توزيع المتغير التابع المتصل اعتدالى التوزيع. ويرى Durlak (2009) أن مؤشرى g , d متحيزان تحيز موجب عندما تكون أحجام العينات صغيرة ونقح Hedges & Olkin (1985) هذا المؤشر للتخلص من التحيز الموجب خاصة فى حالة العنيات الصغيرة كالآتى:

$$g_{adj} = g\left(1 - \frac{3}{4df - 1}\right) = g\left(1 - \frac{3}{4 - 9\mathrm{N}}\right)$$

ويتضح أن الاختلاف فى الصيغ الثلاثة لـ d , Δ , g فى المقام فاعتبرها Cohen الانحراف المعيارى المشترك لمجتمع العينتين (n) و اعتمد Hedges على الانحراف المعيارى المشترك للعينتين (n-1) ، فى حين اعتمد Glass على الانحراف المعيارى للمجموعة الضابطة ويرى Hedges & Olkin (1985) أنه فى حالة العينات الكبيرة فإن المؤشرات الثلاثة يكون لها نفس الأداء ولكن مؤشر g أفضل فى حالة العنيات الصغيرة خاصة مع استخدام الصيغة المنقحة للمؤشر ويتفق معهما Ferguson (2009) فى أن مؤشر g أكثر معيارية من مؤشرى d,Δ ولكن الواضح ان مؤشر g أقل تحيزًا من مؤشر Δ عندما $s^2 \neq s^2_1$.

فالمؤشرات الثلاثة تقيس حجم التأثير (حجم الفروق) عندما يكون المتغير المستقل تصنيفى بمستويين والمتغير التابع متصل وعلى ذلك تستخدم فى حالة اختبار T المستقلة أو اختبار (F) للمقارنة بين مجموعتين ويمكن استخدامهم فى حالة اختبار T المرتبطة. فى ضوء ذلك فهذه المؤشرات انسب للاختبارات التى تعالج التصميمات التجريبية ذو المجموعتين التى تستخدم T وعلى الرغم من ان مؤشر g يفضل عند البعض على مؤشر d.

ثانيًا: مؤشرات قوة العلاقة: وهى المؤشرات البديلة لفروق المتوسطات المعيارية لتقدير حجم التأثير وتنقسم إلى المؤشرات الارتباطية ومؤشرات التباين المفسر.

1. المؤشرات الارتباطية: تقيس هذه المؤشرات قوة العلاقة بين متغيرين فأكثر وهى كالتالى:

- **معامل ارتباط بيرسون** (r): يقيس درجة التباين المشترك أو قوة العلاقة الخطية بين متغيرين متصلين. ويكون أدائه أفضل إذا كان توزيع البيانات اعتداليًا والعلاقة بين المتغيرين خطية. وتقدير حجم التأثير من خلال مؤشر r أكثر معياريةً وأسهل فى تفسيره نتيجة اعتماده على معايير فى تحديد قوة العلاقة ويقع مداه من 1.0 (موجبة) إلى 1.0- (سالبة) عكس مؤشر كوهين " d " فليس له معايير واضحة لتفسيره (Rosenthal et al., 2000) وعلى الرغم من مميزات r إلا أنه يعطى تقديرًا متحيزًا موجبًاPositively bias خاصة فى حالة العينات الصغيرة (Wang & Thompson, 2006).

- **معامل ارتباط سيبرمان** (r_s): هو حالة خاصة من معامل ارتباط بيرسون ويستخدم عندما تكون المتغيرات رتبية، ويستخدم اختبار كندل تاو (T) ومعامل ارتباط جاما لقياس حجم التأثير بين أزواج الرتب (Ferguson,2009).
- **معامل ارتباط** (φ): هو حالة خاصة من معامل ارتباط بيرسون ويستخدم كمقياس لحجم التأثير عندما يكون المتغيرين تصنفين (2 × 2) حيث يستخدم اختبار كا٢ (χ^2) وتتحدد قيمته كالآتى:

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

ويستخدم مؤشر φ بكثرة فى دراسات المجال الطبى ولكن يجب أن نتعامل معه بحذر ويري(1993) Tatsuoka بأنه يمكن التعبير عن حجم التأثير بمربع φ:

$$ES(\varphi)^2 = \frac{\chi^2}{N}$$

وإذا كانت قيمتها 0.01 فأكثر فهى ذات دلالة عملية.

- **مؤشر كرامير** Crammer(V): يستخدم إذا زادت مستويات أحد المتغيرات التصنيفية عن 2 مثل جداول (3×3 أو 2×3) أو غيرها ويتحدد بالآتى:

$$V = \sqrt{\frac{\chi^2}{N(K-1)}}$$

  - K الحد الأدنى من الصفوف أو الأعمدة.

مؤشرات تحليل الانحدار: يمكن استخدام مؤشرات معامل الانحدار المعيارى β لكل متغير أو معامل الارتباط المتعدد (R) لكل معادلة انحدار و قيمة β دائمًا تكون أقل من قيمة مربع معامل الارتباطR^2(Ferguson, 2009) وكذلك يستخدم معامل الارتباط الجزئى Partial r كمؤشر لحجم التأثير إذا تم عزل تأثير المتغير الدخيل وفى هذه الحالة فإن Partial r و β أفضل من مؤشر r للتعبير عن حجم التأثير ويفضل استخدام مربع معامل الارتباط المصحح R^2_{adj} كمؤشر لحجم التأثير.

ويمكن تقدير مؤشر r من خلال مؤشر d وإذا زدات قيمة d تزداد قيمة r . وإذا كانتn_1= n_2 تصبح المعادلة السابقة كالآتى:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

وتقدر قيمة r من قيمة الاختبار الإحصائى T كالآتى:

$$r = \sqrt{\frac{T^2}{T^2 + df}}$$

وتقدر قيمة r من قيمة الاختبار الإحصائى F للمقارنة بين مجموعتين كالتالى:

$$r = \sqrt{\frac{F}{F + df_{error}}}$$

  - df_{error} = N- k
  - k عدد المجموعات أو مستويات المتغير المستقل.

3. مؤشرات نسبة التباين المفسر: هى المؤشرات التى تقيس نسبة التباين المفسر فى المتغير التابع نتيجة المتغير المستقل ومن أهم هذه المؤشرات:

أ. **مؤشر مربع إيتا (η^2):** هذا المؤشر مقرونًا بتصميم تحليل التباين ويقدر من المعادلة الآتية:

$$\eta^2 = \frac{SS_C}{SS_T}$$

- SS_C مجموع المربعات بين المجموعات.
- SS_T مجموع المربعات الكلى.

وهذا المؤشر يعطى تقديرًا لنسبة العلاقة بين متغيرين فى العينة (Richardson, 2011) ويري(2009) Ferguson أن η^2 هو مؤشر لقوة العلاقة بين المتغيرات مثل مؤشرات r, R, R^2, φ, r_{pb}. ولكن مؤشر η^2 مفضل عن مؤشر r^2 عندما تكون العلاقة بين المتغيرات شبه منحنية Curvilinear وتكون قيمته أعلى قليلًا من r^2 (Tatsuoka, 1993).

وأشار Richardson (2011) إلى ان الفرق بين η^2 وr^2 مساوى لمربع معامل الارتباط المتعدد، وفى حالة إذا كان المتغير المستقل بمستوين فإن قيمةη^2 مساوية لمؤشر r_{pb}(Tatsuoka, 1993). ويمكن تقدير مؤشر η^2 من قيمة اختبار F كالآتى:

$$\eta^2 = \frac{(k-1) - F}{(k-1)\big(F + (N-k)\big)}$$

ويري Olejnik & Algina (2003) بأنه لا تتم مقارنة مؤشر η^2 عبر الدراسات المختلفة ولذلك تم تصحيح مؤشر η^2 كالآتى:

$$\eta^2 = \frac{SS_C}{SS_e + SS_T}$$

- SS_e مجموع مربعات داخل المجموعات (الخطأ).

ويمكن تقدير مؤشر η^2لاختبار T كالآتى:

$$\eta^2 = \frac{T^2}{T^2 + N - 2}$$

ويـري (2009) Ferguson أن الفـرق بـين مؤشـرى η^2 وη^2_p ضـئيل جـداص مـع العينـات الكبيـرة وأحيانـاً يخطـى البـاحثون ويعتبـروا ان مؤشـر η^2_pهـو نفسـه مؤشـر η^2 وذلك فى مخرج SPSS.

ب . مؤشـر (f) Cohen: يقـدر نسـبة التبـاين المفسـر فـى حالـة اختبـار تحليـل التبـاين الأحادى ويمكن تقدير قيمة f كالآتى:

$$f = \sqrt{\frac{\omega 2}{1 - \omega 2}}$$

ويمكن تقدير مؤشر f من قيمة اختبار F كالآتى:

$$f = \sqrt{\frac{(k-1)(F-1)}{Nk}}$$

د. مؤشر اوميجا تربيع(ω^2) لـ Hayes (1963): يعتبر أكثر مؤشرات قياس نسبة التباين المفسر انتشارًا وذلك فى التصميمات ذات العامل أو المعالجة الواحدة ANOVA(Fern & Monroe, 1996) ومؤشر ω^2 أكثر تفضيلًا من مؤشر η^2 لتقدير حجم التأثير. وعلى الرغم من أهمية مؤشرى ω^2وابسيلونζ^2 كتصحيح لمؤشر η^2 إلا أن (Kirk (2004 يرى محدودية استخدامهما فى الدراسات ولكن Ferguson (2009) يرى أن مؤشرى ω^2 و ζ^2 أكثر استخدامًا فى تصميمات ANOVA وأن مؤشر ω^2 مثل مؤشر ζ^2 يقدر حجم التأثير فى المجتمع ولذلك فهو مؤشر استدلالى عكس مؤشر η^2 وتتحدد قيمة مؤشر ω^2من المعادلة الآتية:

$$\omega^2 = \frac{SS_C - MS_e}{SS_T - MS_e}$$

- MS_e متوسط مجموع المربعات

بينما عبر عنها Fern & Monroe (1996) بالصيغة الآتية:

$$\omega^2 = \frac{SS_C - (k-1)MS_e}{SS_T + MS_e}$$

ويقدر مؤشر ω^2لاختبار Tكالآتى:

$$\omega^2 = \frac{T^2 - 1}{T^2 + n_1 + n_2 - 2}$$

ويقدر مؤشر ω^2من قيمة الاختبار الإحصائى F للمقارنة بين مجموعتين كالآتى:

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + Nk}$$

ويعتبر مؤشر ω^2 أقل تحيزًا من مؤشرى η^2و ζ^2 و يمكن حسابه للتصميمات البحثية الأكثر تعقيدًا ولكن قيمته تتأثر بمكونات التباين فى هذه التصميمات وتعطى تقديرًا صغيرًا لحجم التأثير Underestimate (Fern & Monroe, 1996) .

ولا يمكن مقارنة مؤشر ω^2 عبر دراسات مختلفة كما هو الحال فى مؤشر η^2 ويوصى باستخدام الصيغة المصححة لمؤشر ω^2وهى كالآتى:

$$\omega^2_{partial} = \frac{SS_C - df_C MS_e}{SS_C + (N - df_c)MS_E}$$

حيث df_c درجات الحرية بين المجموعات وتساوى k-1.

وقيمة ω^2_P هى تقريبًا نفس قيمة ω^2 فى التصميمات الأحادية العامل ولكنهما يختلفا فى حالة التصميمات المعقدة (Fern & Monroe, 1996).

ثالثًا: مؤشرات تقديرات المخاطرة: يستخدم فى التصميم البحثى الذى يتضمن متغيرًا مستقلًا تصنيفى ومتغير تابع تصنيفى ويستخدم بكثرة فى أبحاث المجال الطبى ولتوضيح مؤشرات المخاطرة يمكن عرض الجدول الآتى:

المجموعة	لا تحسن	تحسن
ضابطة	a	b
تجريبية	c	d

- a ، b ، c ، d هى تكرارات.

وتعتمد تقدير مؤشرات المخاطرة على نسبة الأفراد فى كل خلية وتتضمن المؤشرات الآتية:

1. مؤشر نسبة المخاطرة Risk Ratio (RR): وتقدر بالصيغة الآتية:

$$RR = \frac{a(c+d)}{c(a+b)}$$

القيمة + 1.00 تشير إلى أنه لا فروق فى المخاطرة بين المجموعتين وأقل من 1.00 مخاطرة أقل بالمجموعة الضابطة مقارنة بالتجريبية والقيمة 2.00 تعنى أن المجموعة الضابطة لها ضعف المخاطرة مقارنة للمجموعة التجريبية والقيمة بين 1.00 و 2.00 قيمة غير عملية وهذا المؤشر مفضل عن r (Feguson, 2009).

2. مؤشرنسبة الارجحية Odds Ratio (OR): تحدد بالصيغة الآتية:

$$OR = \frac{ad}{bc}$$

القيمة 1.50 (تدل على تأثير صغير) أعلى من 3.00 يعنى حجم تأثير قوى نسبيا Moderately Strong ومدى هذا المؤشر من صفر إلى ∞ ومن عيوب OR هى إذا كان أحد الخلايا تكراراتها صفر فلا يمكن تقديره (Breaugh, 2003; Durlak, 2009) ومؤشر OR أحيانا مفضل على مؤشرات فروق المتوسطات المعيارية (d, g, Δ) وذلك بتحويل المتغير المتصل إلى متقطع عن طريق تقسيمه من خلال قيمة المتوسط (فوق المتوسط – تحته) أو الوسيط (Durlak, 2009) وقيمة OR.

3. مؤشر Risk Difference (RD): وتتحدد بالصيغة الآتية:

$$RD = \frac{a}{(a+b)} - \frac{c}{(c+d)}$$

هـذا المؤشـر سـهل تفسـيره ويعـرض الفـروق الفعليـة فـى المخـاطرة بـين المجمـوعتين فالقيمـة 0.4 تشـير إلـى ٠.٤ فـروق المخـاطرة بـين المجموعـة التجريبيـة والضـابطة (Ferguson, 2009)

رابعًا : مداخل بديلة

1. مؤشر Common language effect Size (CLES): اقترح McGraw & Wong (1992) مؤشر حجم التأثير (CLES) ويعبر عن حجم التأثير فى ضوء احتمالات وهو يعطـى مقدار الفروق معبرا عنها باحتمالات وتعنى ان احتمال درجات فردا فى المجموعة التجريبية تكون أعلى من درجات فردًا آخر فى المجموعة الضابطة. ولتوضيح قيمة هذا المؤشر فى حالة المقارنة بين الذكور والإناث على متغير ما متصل فإذا كان متوسط الذكور 69.7 والانحراف المعيارى 2.8 ومتوسط الإناث 64.3 و S=2.6 فإن قيمة مؤشر CLES=0.92 وهذا يعنى أن احتمال أن الذكور أطول من الإناث بنسبة 92%.

وهـذا المؤشـر يقـى نفسـه مـن عـدم تـوفر الاعتداليـة (Breaugh, 2003) ويقـدر هـذا المؤشر من خلال تقريب اختبار Z كالآتى:

$$z = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{S_e^2 + S_c^2}}$$

ومثال على ذلك:

$$z = \frac{51.476 - 39.454}{\sqrt{(11.007)^2 + (14.628)^2}} = 0.652\,, P(Z0.652) = 0.74$$

وهذا المؤشر يستخدم للمقارنة بين متوسطين ولكن Dunlap (1994) وسع استخدام هذا المؤشر ليناسب معامل ارتباط بيرسون.

2. مؤشر نسبة الكفاءة السببية Causal Efficacy Ratio: لاحظ Prentice & Miller (1992) أن لحجم التأثير وظيفةً مرتبطةً بقوة المعالجة للمتغير التابع وأكد Abelson (1995) ان المعالجة يتم تجاهلها بدرجة كبيرة أثناء تقدير حجم التأثير واقترح Abelson تقدير مؤشر CER ونسبته تعكس حجم التأثير الخام وهو الفرق بين متوسط المجموعتين على

المتغير التابع مقسوما على حجم السببية CauseSize (الفرق بين متوسط المجموعتين على المتغير المستقل).

تفسير حجم التأثير

يوجد القليل من التوجهات لتفسير حجم التأثير ولا توجد معايير واضحة ومتفق عليها لكل مؤشر لتفسيره.

1. معايير Cohen (1988):

الجدول (2.11): معايير Cohen (1988) لتفسير مؤشرات حجم التأثير.

المؤشر	ضعيف	متوسط	كبير
الفروق بين المجموعات d, Δ , g	0. 2	0.5	0. 8 ≥
مؤشر f	0.10	0.25	0.40≥
مؤشر r	0.10	0.30	0.50≥
مؤشرات التباين المفسر , R^2_{adj} , r^2 R^2, ω^2, η^2,	0.01	0.06	0.14≥

2. معايير اقترحها Ferguson (2009):

الجدول (3.11): معايير Ferguson (2009) لتفسير حجم التأثير

المؤشر	صغير	متوسط	كبير
الفروق بين المجموعات d, Δ, g	0.41	1.15	2.70
مؤشر العلاقة (r , R, φ, ρ, β)	0.20	0.50	0.80
التباين المفسر r^2 , R^2 , ω^2, η^2,	0.04	0.25	0.64
تقديرات المخاطرة (RR , OR)	2.0	3.0	4.0

ووضع Lispsey (1998) معايير لمؤشرات الفروق بين المجموعات وهى القيمة 0.15 صغير و 0.45 متوسط 0.90 فأكثر و وضع Hinkle, Wiersma, & Jurs (1979) معايير لمعامل ارتباط حيث0.30 منخفض و 0.70 متوسط و 0.90 فأكثر قوى.

وهذه المعايير أو الإرشادات ما هى الا اجتهادات بحيث لا تطبق بطريقة جامدة (Cohen, 1994)، وأشار Pedhazur & Schmelkin (1991) بأن القيمة العالية لحجم التأثير عند باحث ما فى وقت ما لتخصص ما يختلف عند باحث آخر فى تخصص أخر. ومن الخطأ القول إن حجم التأثير الكبير فى القيمة ذو دلالة عملية وان حجم التأثير الصغير فى القيمة يعتبر تافهة وليس له أى معنى (Prentice & Miller, 1992) ويعتمد تفسير حجم التأثير أيضًا على الظروف التى تتم فيها الدراسة. وتوجد عوامل اخرى عند تفسير حجم التأثير مثل تحليل الكلفة – المنفعة كمحك خارجى للحكم ما إذا كان حجم التأثير ذا دلالة عملية أم لا.

وإن تفسير حجم التأثير للدراسة لا بد أن يتم فى ضوء المرحلة التعليمية للمجتمع المستهدف (روضة – ابتدائى – ثانوى – جامعى)، وفى ضوء المحتوى المقاس فتفسير حجم التأثير لمحتوى الرياضيات مختلف عن تفسير حجم التأثير لمحتوى اللغة العربية، ويفسر حجم التأثير أيضًا على أساس المجموعات الفرعية فمثلًا تفسيره لموضوع ما على الذكور يختلف عن تفسيره عند الإناث وأيضًا بالنسبة للموقع الجغرافى وذلك نظرًا لوجود فروق بين المجتمعات ويفسر أيضًا فى ضوء التراث البحثى المشابهة للظاهرة موضوع الدراسة وهذا يفيد تمامًا مجال دراسات ما وراء التحليل. وأنه يمكن لأحجام التأثير الصغيرة يمكن أن تكون لها معنى وهذا يتوقف على مجال دراسة الظاهرة. ووجود قاعدة وحيدة لتفسير حجم التأثير مثل معايير Cohen وغيرها لا تكون مناسبة لكل المجالات.

قام (1996) kirk بمسح استخدام حجم التأثير فى أربعة مجلات تابعة للرابطة النفسية الأمريكية (APA) فى عام 1995 واتضح أن نسبة المقالات التى قدرت حجم التأثير (سواء حجم تأثير فروق معيارية أو قوة علاقة) تراوحت من 12% لمجلة

علـم الـنفس التجريبـى إلـى 77% لمجلـة علـم الـنفس التطبيقـى وأن 60% مـن أحجـام التـأثير المسـتخدمة هـو مؤشـر R^2 فـى المجـلات الأربعـة. ومسـح Keselman et al., (1998) اسـتخدام حجـم التـأثير لـ 226 دراسـة فـى مجـال التربيـة وعلـم الـنفس فـى 17 مجلـة منشـورة فـى الفتـرة مـن 1994 حتـى 1995 وتبـين أن 16 دراسـة قـدرت حجـم التأثير بنسبة % 7.1. وفحص Fiddler et al. (2005) استخدام حجم التأثير فى مجلـة علـم الـنفس الإرشـادى الاكلينكـى فـى الفتـرة مـن 1993 إلـى 2001 واسـتخدمت مؤشـرات حجـم التـأثير المعياريـة لاختبـارات ANOVA وX^2 وهـى علـى التـوالى 32% و13% و 16%. وفحص Kampense, Dyba, Hannay, & Sjoberg (2007) 78 دراسـة فـى مجـال هندسـة البرمجيـات واشـتملت علـى 429 اختبـار وجـد ان نسـبة توصـيف حجـم التـأثير 29% مـن الدراسـات قـدرت حجـم التـأثير فـى ضـوء مؤشـرات معياريـة وغيـر معياريـة ونسـبة اسـتخدام المؤشـرات المعياريـة % 7.7 وان نسـبة اسـتخدام مؤشـرات حجـم التـأثير لاختبـارات ANOVA 3% ولاختبـار T 16% ولـ X^2 6% وتوصل إلى أن وسيط مؤشر g_{adj} لهذه الدراسات 0.60.

وحلـل Zientek, capraro & Capraro (2008) 217 اختبـارًا إحصـائيًا فـى 174 دراسـة منشـورة قبـل 2005 ووجـد ان 39% مـن الدراسـات قـدرت حجـم التـأثير وأن 94% مـن هـذه الدراسـات اعتمـدت علـى معـاملات الارتبـاط ومعـاملات الانحـدار كمؤشـر لحجـم التـأثير و 7% اسـتخدمت حجـم التـأثير فـى الاختبـارات الفارقـة مثـل مؤشـر d. وحلـل Alhija & Levy (2009) 183 اختبـارًا إحصـائيًا فـى 99 دراسـة فـى مجـلات نفسـية وتربويـة فـى الفتـرة مـن 2003 إلـى 2004 و توصـلا إلـى أن معظـم الدراسـات اعتمـدت علـى مؤشـرات r و R و X^2 كمقيـاس لقـوة العلاقـة (86%) وأن (57%) مـن الدراسات الفارقة قدرت حجم التأثير باستخدام مؤشرات d و μ^2 وμ_P^2 .

وأجرى Sun, Pan, & Wang (2010) مسحًا لـ 1243 دراسـة منشـورة فـى 14 مجلـة علميـة تربويـة ونفسـية فـى الفتـرة مـن 2005 حتـى 2007 ووجـد أن 49% مـن المقـالات قـدرت حجـم التـأثير وفقـط 57% منهـا فسـرت قيمـة حجـم التـأثير و المؤشـر الأكثـر اسـتخداما هـو R. وراجـع McMillan (2011) أربـع مجـلات مـن إصـدارات APA فـى

الفترة من 2008 إلى 2010 متضمنة 417 دراسة وتوصل إلى أن 74% من الدراسات قدرت حجم التأثير وكانت أكثر ثلاثة مؤشرات استخداما هى R^2 و η^2 و d ويوجد استخدام محدود جدًا لمؤشر (g) Hedges و odds ratio(OR) و Cohen (f) و ω^2 وأن 20% من الدراسات تضمنت تفسيرًا لمؤشر حجم التأثير وأشار إلى زيادة استخدام حجم التأثير من 1997 حتى 2010.

وظهر الاهتمام بحجم التأثير نتيجة عدم الرضا عن NHST وزيادة الاهتمام بإجراء دراسات ما وراء التحليل واستخدامه فى تحليلات القوة الإحصائية للدراسات.

ويعتبر حجم التأثير جزءًا مكملًا للدلالة الإحصائية إلا أن توظيفها فى البحث النفسى التربوى العربى محدودًا إلى حد ما وذلك بمقارنته بالتراث السيكولوجى الأجنبى وقد يعود هذا إلى عدة أسباب عدم إدراك الباحثين لأهمية مؤشر حجم التأثير فى إعطاء القيمة التفسيرية والتطبيقية للقرار الإحصائى وأيضا غياب هذا المفهوم فى المجلدات الإحصائية والممارسات البحثية وعدم احتواء البرامج الإحصائية على مؤشرات تقدير حجم التأثير ما عدا القليل منها مثل مؤشر η^2_p و R^2 و R^2_{adj} وعلى الرغم من أهمية تقدير حجم التأثير لأسلوب ما وراء التحليل وتقدير القوة الإحصائية إلا أن استخدامه فى الاختبارات الإحصائية محدود ما عدا أسلوب تحليل الانحدار وقد يكون ذلك نتيجة لأن مخرج SPSS يعطى هذا المؤشر ولا حاجة لحسابه بمعادلات رياضية ويوظف فى عرض جداول النتائج فى الدراسات النفسية.

على الرغم من تأكيد الجمعية النفسية الأمريكية APA بتضمين مؤشر حجم التأثير فى الدراسات النفسية حتى يتم قبول هذه الدراسات للنشر وليس هذا فقط، بل مناجاة الباحثين فى مجال العلوم المختلفة بتقدير مؤشر حجم التأثير لمحاولة التغلب على محددات إستراتيجية اختبارات دلالة الفروض الصفرية والملفت للنظر فى البحث النفسى العربى أن استخدام مؤشرات حجم التأثير لأسلوب تحليل التباين منعدم تمامًا عكس أسلوب تحليل الانحدار وهذا يعكس إشكالية للباحثين فى التعامل مع نتائج دراساتهم.

أما بالنسبة للقيمة التقديرية لحجم التأثير فى البحث النفسى العربى فإنها فى مجملها تقع فى حجم تأثير متوسط وذلك لمؤشر r^2 وهذا يعنى حقيقة أن القيمة العملية

والتطبيقية لنتائج هذه الدراسات محددوه وان توصيات هذه الدراسات خاصة فى الاختبارات التى استخدمت تحليل التباين وتحليل الانحدار واختبار T المستقلة واختبار r غير قابلة للتطبيق فى المجالات النفسية التربوية ويمكن فى هذه الحالة أن لا نقلق ونغضب من عدم الأخذ بتوصيات البحوث النفسية التربوية لأن قيمتها القرارية متوسطة وخاصة إذا أخذنا فى الاعتبار مؤشرات أخرى تلعب دورًا فى تفسير حجم التأثير مثل التكلفة والمنفعة وان كان هذا الاستنتاج للقيمة التقديرية لحجم التأثير فى البحث النفسى التربوى العربى فيه درجة من القسوة والحدة إلا أن تفسير حجم التأثير لا بد أن يتم فى صورة ديناميكية وليس استاتكية ويكون فى ضوء الفئة المستهدفة وأبحاث المجال والمرحلة العمرية وطبيعة دراسة الظاهرة، فقد أشار (Trusty, Thompson, & Petrocelli 2004) أنه يمكن لأحجام تأثير صغيرة أن يكون لها معنى وهذا يتوقف على مجال دراسة الظاهرة.

والنتائج الدالة إحصائيا ليست بالضرورة دالة عمليا وهذا يؤكده Nickerson (2001) بوجود خطأ فى تفسير الدلالة الإحصائية فالقيمة الصغيرة لـ P ليست بالضرورة تعنى أن حجم التأثير المقابل لها دال عمليا وكما هو واضح أن 68.1% من الاختبارات الدالة إحصائيا عند $\alpha = 0.01$ ليست دالة عمليًا.

الإرشادات اللازمة للباحثين لتقدير حجم التأثير

هذه الإرشادات تم الوصول إليها فى ضوء التراث البحثى وتتضمن:

- **تنظيف البيانات قبل التحليل:** وذلك لمعالجة البيانات الغائبة والقيم المتطرفة Outlier والتعرف على أخطاء الإدخال للبيانات.
- **التحقق من مسلمات الاختبار:** وذلك لاختيار النموذج الإحصائى والاختبار الإحصائى وهذه المسلمات مثل الاعتدالية وذلك لأن تقدير حجم التأثير فى حالة الإحصاء البارامترى يعطى تقديرات متحيزة خاصة عندما لم يتحقق مسلمات استخدامها (Ferguson, 2009).
- **عرض الإحصاء الوصفى لبيانات الدراسة:** لأن ذلك يساعد القارئ والباحث على تقدير حجم التأثير خاصة إذا كانت الدراسة لم تتضمن تقدير حجم التأثير أو يريد

الباحث تقدير حجم تأثير آخر ومن أهم مؤشرات الإحصاء الوصفى المتوسطات والتباينات كذلك متوسطات الأخطاء المعيارية فى التصميمات العاملية ANOVA ويوصى فى التصميمات العاملية عرض MSE, df, F لكل تأثير حتى يمكن تقدير مؤشر حجم التأثير وكذلك أكد Sun et al. (2010) بضرورة تحديد حجم العينة للمجموعات والمتوسطات والانحرافات المعيارية فمثلا فى تحليل التباين إذا كانت توجد فروق دالة إحصائية فإنه من خلال المتوسطات يمكن التعرف على أى مجموعة أحدثت هذه الفروق قبل إجراء المقارنات البعدية Post hoc وللطرق اللابارامترية يجب عرض التكرارات والنسب المئوية و المؤشرات الوصفية.

- **تقدير حجم التأثير:** وذلك لتفسير أهمية النتائج أو مقدار تأثير التجربة لأنها ضرورية كجزء مكمل لاختبارات الفروض وذلك لوجود مخاطرة لعمل استنتاجات خاطئة عندها لا تتم تقدير حجم التأثير وذلك لأن النتائج غير الدالة إحصائيًا يحكم عليها خطأ بأنها ليست لها أهمية عملية وهنا يقع الباحث فى خطأ استدلالى وكذلك النتائج الدالة إحصائيًا يُفهم خطأ بأنها دالة عملية.
- **استخدام الخريطة التصنيفية لمؤشرات حجم التأثير** يفضل استخدام مؤشر d لاختبار (T) ويفضل مؤشر ω^2 للتصميمات العاملية ANOVA ومؤشر R^2 لتحليل الانحدار و r^2 لمعامل ارتباط بيرسون. إلا أن Ferguson (2009) يوصى باستخدام أحجام التأثير المصححة عن المؤشرات غير المصححة لأنها أكثر دقة لحجم التأثير فى المجتمع وهو مؤشر استدلالى وعلى ذلك يفضل استخدام g_{adj}و η^2_p و ω^2_p و R^2_{adj} خاصة للعينات الصغيرة.
- يرى Ferguson (2009) أن مؤشرات Partial R ومعاملات الانحدار المعيارية أكثر دقة من مؤشر r للتعبير عن حجم التأثير فى ذلك فى الانحدار المتعدد.
- **تفسير حجم التأثير:** لا بد من مناقشة وتفسير الأهمية التطبيقية أو العملية لحجم التأثير وفى قضية التفسير لا بد أن يوجد فى الاعتبار بانه لا توجد معايير موضوعية موحدة متفق عليها بين كل الباحثين ويرى Synder & Lawrson (1992) أن هذه المعايير جامدة واعتبارية ولا ينصح الأخذ بها بصورة مطلقة للحكم على قيمة حجم التأثير بأنها منخفضة أو مرتفعة وعندما يزيد قيمة حجم

التأثير عن نقطة القطع الموصى بها لا يعنى هذا أن حجم التأثير له دلالة اكلينيكية أو معنوية. وحجم التأثير لها محددات يحددها Sun et al., (2010) بأنها تعتمد على المتوسطات والانحرافات المعيارية للعينة وهذا يختلف من عينة إلى أخرى أى انها دالة وظيفية لحجم العينة ولذلك يجب الحذر عن تفسير حجم التأثير من دراسة إلى أخرى.

- وعند تفسير حجم التأثير لا بد أن يؤخذ فى الاعتبار عدة عوامل منها تأثره بالعينة والخصائص السيكومترية لأدوات القياس كالصدق والثبات. فالعينات الصغيرة وغير العشوائية تعطى تقديرات متحيزة وأيضًا يعتمد حجم التأثير على التصميم البحثى فالتصميمات التى تفشل فى ضبط المتغيرات الخارجية (الدخلية) تعطى أحجام تأثير كبيرة مقارنة بالتصميمات التى تضبط المتغيرات الخارجية خاصة إذا كانت هذه المتغيرات الخارجية مرتبطة ببعضها ارتباطًا عاليًا (Olejnik & Algine, 2003).

وعند التفسير لا بد أن يؤخذ فى الاعتبار طبيعة حجم التأثير فى التخصص فمعايير التفسير لحجم تأثير (صغير – متوسط – كبير) فى علم النفس قد تختلف عن التخصصات الأخرى السلوكية والطبيعية وكذلك يؤخذ فى الاعتبار عند مناقشة الأهمية التطبيقية عوامل الكلفة والمنفعة.

- **التعبير عن حجم التأثير فى ضوء فترات الثقة:** يعطى معلومات إضافية وقرار على متصل لإبراز الأهمية العملية لحجم التأثير، ويوصى باستخدامها كبديل لاتخاذ القرار فى اختبارات الفروض الكلاسيكية.
- **تضمين حجم العينة اللازمة فى مرحلة تصميم الدراسة:** للحصول على حجم مرغوب وذو معنى لحجم التأثير وذلك فى ضوء الدراسات السابقة والهدف من التقدير القبلى لحجم التأثير هو الحصول على قرارات لها قوة إحصائية معينة للاختبارات المستخدمة وهذا يجنبنا الوقوع فى تناقض بين حجم التأثير ومستوى الدلالة الإحصائية.

الفصل الثانى عشر

اختبارات الفروض للتحقق من الاعتدالية

Normality

قبل التطرق إلى التحقق من الاعتدالية لا بد من التطرق إلى:

البيانات الغائبة أو المفقودة Missing data

تعتبر قضية البيانات الغائبة من أهم المشكلات فى تحليل البيانات وتحدث عندما يرفض أو ينسى المفحوص الاستجابة على مفردة أو أكثر وهذا يؤثر على تعميم النتائج، والإجراءات الإحصائية لأى أسلوب تتطلب أن تكون وحدة التحليل تامة أو كاملة البيانات. ويتعامل الباحث دائمًا مع البيانات الكاملة، والمشكلة التى تواجهه الباحث أثناء التحليل الإحصائى هو كيفية معالجة البيانات المفقودة (غياب بيانات على متغير واحد فأكثر لفرد واحد فأكثر) وتظهر البيانات المفقودة نتيجة أخطاء الإدخال، أو مشاكل أثناء جمع البيانات أو الرفض للاستجابة على مفردة ما فأكثر وقضية كيفية تحليل قواعد بيانات بها درجات غائبة هى معقدة.

وتوجد عدة استراتيجيات للتعامل مع البيانات الغائبة منها كما اشار اليها (Kline, 2016):

- **طريقة الحالة المتاحة Available case method**: وهى تحليل البيانات المتاحة وهذا يتضمن نوعين أساسيين هما:

1. **طريقة الحذف List-wise deletion**: الحالات التى لها درجات غائبة على أى متغير تستبعد من كل التحليلات، وهذه الطريقة تتعامل مع الحالات **كاملة البيانات فى التحليل Complete case analysis**، وعلى ذلك يحدث نقص لحجم العينة، وهى من الإجراءات البسيطة والمباشرة ويمثل هذا **بالاتجاه المحافظ**، ويطلق على البيانات فى هذه الحالة **Complete data only**. ومن المتوقع ان تكون الأخطاء المعيارية بعد تطبيق هذه الإستراتيجية أكبر من مثيلاتها على قاعدة البيانات الكلية قبل الحذف، ولكنها تعطى تقديرات دقيقة لمعالم النموذج، ومن أهم مميزاتها هى ان كل التحليلات تكون لنفس العدد من الحالات وعلى ذلك تقدر مصفوفة التغاير (الارتباط) للحالات كاملة البيانات فقط.

2. طريقة الحذف Pair-wise deletion: وتسمى أيضًا تحليل الحالات المتاحة فقط Available case analysis ففى الطريقة السابقة يحذف الفرد الذى لديه بيانات غائبة على أحد المتغيرات من التحليل، ولكن فى **طريقة Pair-wise** يبقى الفرد فى التحليل للاستفادة منه بالبيانات الموجودة له على بقية المتغيرات، وربما يكون هذا حل فعال ولا يسبب فقد الحالة كلها. ولكن إذا كانت البيانات الغائبة على المتغير التابع فإن الحالة كلها تحذف من التحليل، وهذه الطريقة تعطى معالجات إحصائية بأحجام عينات مختلفة من تحليل إلى أخر. بافتراض أن حجم العينة = 300 ، وتتضمن حالات غائبة، إذا 280 حالة بدون درجات غائبة على المتغيرات فأحيانًا يتم تقدير معامل الارتباط ($r_{x1\ x2}$= 0.30, N = 290) وأحيانًا ($r_{x1\ y}$ = 0.85, N = 255)، وأحيانًا (r_{x2y} = 0.43, N = 280) أى تختلف حجم العينة من تحليل لأخر وهذا يقود إلى عدم الاتساق الرياضى. وإذا كان حجم العينة كبيرًا كبرًا كافيًا ويوجد عدد قليل من المفردات بها بيانات غائبة فلا توجد فروق عملية بين الطريقين ولكن إذا وجد عدد كبير من أفراد العينة لديهم بيانات مفقودة على مفردات المقياس ففى هذه الحالة فإن الطريقة المفضلة هى pair wise ، وفيما يلى مثال للبيانات الغائبة:

الجدول (1.12): مثال لمجموعة بيانات غير كاملة.

CNo	X_1	X_2	Y
1	42	8	13
2	34	10	12
3	22	12	-
4	مفقود	14	8
5	24	16	7
6	16	مفقود	10
7	30	مفقود	10

- **الطرق التعويضية الانفرادية Single imputation Methods**: وعلى الرغم من انتشار طريقتى pairwise و listwise إلا أنه توجد بعض الطرق التعويضية للتعامل مع البيانات الغائبة وأهمها:

1. استبدال البيانات الغائبة **بالمتوسط العام لاستجابات الأفراد على المتغير** وعلى ذلك تستخدم العينة كاملة العدد فى التحليل group – mean substitution. فمثلا فى جدول (1.2) نلاحظ أن المتغير X_1 به الحالة 4 ليس لها قيمة على هذا المتغير، وبتقدير المتوسط لـ X_1 يتبين أن $\bar{X}_1 = \frac{148}{7} = 21.14$ بذلك تكون القيمة الغائبة للحالة 4 على المتغير X_1 هى 21.14، ولكن استبدال البيانات الغائبة بالمتوسط يغير أو يشوه خصائص التوزيع للبيانات عن البيانات قبل التعويض بالمتوسط، وذلك لأنها تقلل من التباينات للمتغيرات التى لها بيانات غائبة. ويفضل عدم استخدامها إذا كانت البيانات الغائبة كثيرة وتستخدم إذا كان حجم البيانات الغائبة صغيرًا.

2. التعويض عن طريق حساب الانحدار Regression: تستبدل البيانات الغائبة بالدرجة المنبئة عن طريق حساب الانحدار المتعدد لدرجات المتغير الذى يتضمن بيانات غائبة من خلال المتغيرات الأخرى فى ملف البيانات ففى البيانات الموضحة بالجدول السابق يعتبر X_1 و X_2 منبئين بالمتغير Y:

$$\hat{y} = b_1 X_1 + b_2 X_1 + a.$$

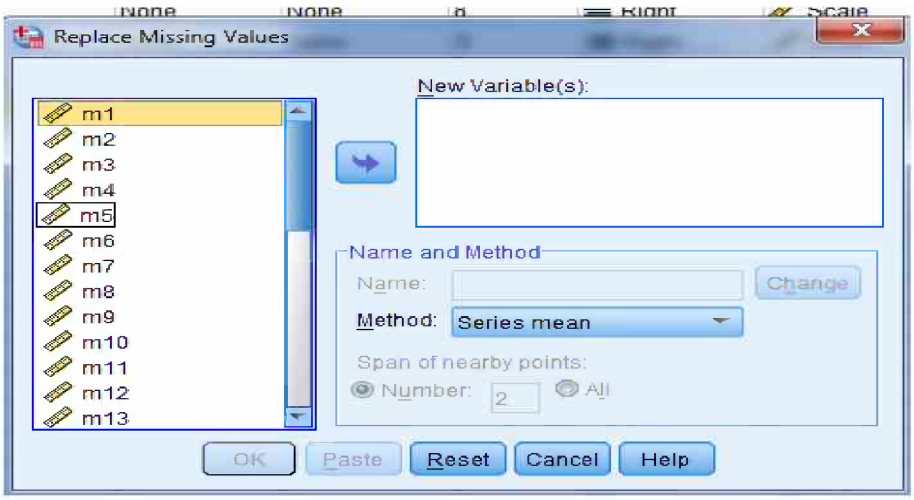
وهكذا بالنسبة ل X_1 تابع والمتغيرات الأخرى مستقلة X1 = b1y + b2x2 +a واستخدام هذه الطريقة يتطلب معلومات أكثر من طريقة التعويض بالمتوسط. ويمكن استخدام الانحدار اللوجستى Logistic regression إذا كانت المتغير التابع تصنيفى. وعمومًا هى أقل مرغوبية مقارنة بطرق التعظيم المتوقع.

وإذا تواجدت البيانات الغائبة لدى عدد معين من الأفراد فإنه يمكن حذفهم من قاعدة البيانات (Tabachnick & Fidell, 2007)، ويفضل إذا استخدمت إحدى الطرق التعويضية للتعامل مع البيانات الغائبة أن تعيد التحليل مرة أخرى مع مصفوفة الارتباطات بدون استخدام الطرق التعويضية بمعنى أن البيانات كاملة. وإذا كانت النتائج متماثلة فتوجد ثقة فى الطريقة المستخدمة لمعالجة البيانات الغائبة ولو حدث اختلاف فالباحث بحاجة إلى معرفة أسباب هذا التناقض ويجب عرض نتائج التحليلين

(قبل وبعد) معالجة البيانات الغائبة. حيث يعوض عن البيانات الغائبة من اقرب قيمة للحالة فى ملف البيانات قبله أو بعده.

استبدال البيانات الغائبة فى SPSS

ويتم إجراء الأمر من خلال: 1. اضغط قائمة **Transform** واختر **replace missing data** تظهر النافذة التالية:



2. انقل المتغيرات المراد معالجتها من أثر البيانات الغائبة إلى مربع **New variables**

3. اضغط على وظيفة **Method** فتظهر قائمة تحتها الطرق التى تقوم بمعالجة البيانات المفقودة أو الغائبة احدد طريقة معالجة البيانات الغائبة وليكن استبدالها بمتوسط المتغير.

4. اضغط **Ok** وبالرجوع إلى ملف البيانات يلاحظ استبدال البيانات المفقودة بمتوسط درجات المتغير.

القيم المتطرفة Outliers

فى بعض الأحيان تتضمن بيانات العينة قيمة متطرفة أو أكثر على متغير واحد وهى outlier Univariate حيث تختلف تمامًا عن بقية البيانات، وهى بدورها تؤثر على قيمة العلاقات بين المتغيرات وهذا ينعكس سلبًا على تقديرات معالم النموذج نتيجة تأثيرها على العلاقات بين المتغيرات وكذلك على مؤشرات الالتواء والتفرطح. وتوجد

القيم المتطرفة نتيجة عدة أسباب أهمها وجود أخطاء إدخال أو ترميز أو نتيجة وجودها بصورة طبيعية. وإذا كانت القيم المتطرفة موجودة بصورة خاطئة تُحذف أما إذا كانت حقيقية فيجب التعامل معها، مثل إدخال عمر شخص 210 عامًا ولكنها خطأ إدخال فعمره الحقيقى 21 عامًا ويطلق على القيمة المتطرفة فى متغير واحد بـ Univariate Outlier. ولا يوجد تعريف واحد للقيمة المتطرفة ولكن القاعدة هى الدرجات التى لها أكثر من ثلاثة أضعاف الانحرافات المعيارية عن المتوسط. ويمكن تشخيصها من خلال فحص التوزيعات التكرارية للدرجات المعيارية Z فإذا كانت lZl >3.0 فإن هذا يشير إلى وجود قيم متطرفة، بينما القيم المتطرفة المتدرجة Multivariate Quilter وهى درجات متطرفة على متغيرين فأكثر وهذا بدوره يزيد من الوقوع فى الخطأ من النوع الأول أو الخطأ من النوع الثانى، وهذا مثال على متغير يتضمن درجة متطرفة (1, 2, 5, 5, 7, 11, 50) وتعتبر 50 قيمة متطرفة ويوجد عدة طرق لتشخيص القيم المتطرفة المتدرجة أهمها:

1. **إحصاء (D) Mahalanobis distance** : حيث يشير إلى المسافة بوحدات الانحراف المعيارى بين مجموعة من الدرجات للحالة الواحدة ومتوسطات كل المتغيرات ، وفى العينات الكبيرة ذات التوزيعات الاعتدالية فإن توزيع مؤشر D^2 يعامل مثل توزيع χ^2. فقيمة D^2 مع القيمة المنخفضة لمستوى الدلالة الإحصائية (P) يؤدى إلى رفض الفرض الصفرى وعلى ذلك يوجد تطرف فى التوزيع، وينصح دائما باستخدام مستوى دلالة إحصائية (α) عند أدنى قيم مثل 0.001 = α وتقوم بعض البرامج مثل SPSS و SAS بطباعة مؤشر D^2.

2. إذا كان الدرجات المعيارية كبيرة جدًا على متغير واحد فأكثر حيث المتغيرات التى لها درجات معيارية تزيد عن 3.29 (اختبار ذو ذيلين 0.001 > P) فيدل هذا على وجود قيم متطرفة.

3. العرض البيانى لبيانات المتغير مثل المدرج التكرارى أو Boxplot.

واحد الطرق للتعامل مع الدرجات المتطرفة هو إجراء تحوير للمتغير الذى يتضمن القيم المتطرفة. حيث تؤدى هذه الاستراتيجية إلى جعل توزيع المتغير أكثر اعتدالية

التحقق من الاعتدالية

اما فيما يتعلق بمداخل التحقق من مسلمة الاعتدالية فمنها ما هو وصفى من خلال:

1. **عـــرض مقـــاييس النزعـــة المركزيـــة:** تســاوى قـيم المتوسـط والوسـيط والمنـوال يفيـد بوجود الاعتدالية.

2. **حســاب مؤشــرى الالتــواء والتفــرطح:** ولتفسـير قيمـة الالتـواء والتفـرطح فـى ضـوء قـيم مطلقـة لهـم أو حـدود قطـع، أشـارت دراسـات المحاكـاة إلـى أنـه إذا كانـت قيمـة الالتـواء SK 3< فإنـه يوجـد التـواء شـديد، بينمـا التفـرطح أكثـر مرونـة فالقيمـة المطلقـة أكبـر مـن 7.0 تشير إلى تفرطح شديد (Chou & Bentler, 1995; West, Finch & Curran 1995). وتوجـد أراء أكثـر مرونـة فيمـا يخـص التفـرطح فـإذا كانـت قيمتـه أكبـر مـن 8.0، ولكـن القاعـدة العامـة هـى إذا كـان قيمـة التفـرطح أكبـر مـن > (10.0) فإنـه توجـد مشـكلة فيمـا يخـص التوزيـع، والقيمـة 20 فـأكثر تشـير إلـى توزيـع يعـانى بشـدة مـن عـدم الاعتداليـة .(Kline, 2016)

ويتم التحقق الاستدلالى من مسلمة الاعتدالية من خلال:

اختبار كولوموجوروف – سميرنوف لعينة واحدة

Kolomogorov–smirnov one sample test

الهـــدف: يسـتخدم للتحقـق مـا إذا كانـت البيانـات الفتريـة أو النسـبية اعتداليـة التوزيـع فـى المجتمـع وهـى مـن فئـة اختبـارات التوزيـع المنـتظم Test of Distribution . Symmetry ، وفــى التوزيــع الاعتــدالى فــإن المنحنــى يشــبه الجــرس، والاعتداليــة شــرط ضـرورى لاسـتخدام الإحصـاء البـارامترى وعلـى الباحـث التأكـد مـن الاعتداليـة وفـى حالـة عدم توافرها فإنه يفضل استخدام الاختبارات اللابارامترية.

اختبارات الفروض لقضية بحثية: جمع الباحث بيانات لمتغير (التحصيل مثلًا) لـ 100 فرد وهى كالآتى:

7	5	1	9	2	3	6	7	2	1
8	3	3	2	3	3	1	3	3	3
3	3	8	3	3	6	1	1	3	5
1	1	1	3	1	3	3	3	6	6
1	1	5	5	3	6	3	3	16	3
3	4	5	5	3	3	1	3	1	1
12	3	6	8	6	3	1	3	3	3
3	1	3	1	2	3	1	5	3	1
3	1	5	1	3	3	2	6	2	3
6	2	3	5	1	3	1	2	1	3

واراد الباحث التحقق ما إذا كان توزيع هذه الدرجات اعتدالية التوزيع فى المجتمع؟

الخطوات البحثية:

1. **سؤال البحث:** هل توزيع الدرجات اعتدالية فى المجتمع؟

2. **فرض البحث:** توزيع الدرجات توزيعًا اعتداليًا.

3. **متغيرات البحث:** المتغير (x) – فترى –تابع فى البحث.

4. **النمـــوذج الإحصـــائى:** نمـوذج المتغيـر الواحـد – إحصـاء لابـارامترى. و**الاختبـــار المناسب** كولموجوروف– سميرنوف (K.S) لعينة واحدة.

اختبارات الفروض الصفرية

1. **الفـــــروض الإحصـــــائية** H0: الـــدرجات اعتداليـــة التوزيـــع فـــى المجتمـــع

HA: الدرجات غير اعتدالية التوزيع

2. **الاختبـــــــــار الإحصـــــــــائى:** كولموجـــــــوروف- ســـــــميرنوف (K.S)

3. **مستوى الدلالة الإحصائية وقاعدة القرار:** لمستوى دلالة إحصائية

$\alpha = 0.05$ وحجـم العينـة $N = 100$ يمكـن البحـث فـى جـداول(كولموجـوروف – سيمرنوف) عن القيمة الحرجة أو الجدولية.

4. **الحسابات:** بإنشاء جدول التوزيع التكرارى الآتى:

x	Frequency	Cum. freq	Observed (xi) (SN(Xi))	Z score	Fo (Xi) القيمة المتنبأة	القيمة المطلقة للفرق $\|(Foxi) - SN(xi)\|$
1	24	24	0.24	- 0.98	0.1635	0.08
2	8	32	0.32	- 0.57	0.2843	0.04
3	41	73	0.73	- 0.17	0.4325	0.30
4	1	74	0.74	0.24	0.5948	0.15
5	9	83	0.83	0.65	0.7422	0.09
6	9	92	0.92	1.05	0.8531	0.07
7	2	94	0.94	1.46	0.9279	0.01
8	3	97	0.97	1.87	0.9693	0.00
9	1	98	0.98	2.27	0.9884	0.01
10	0	98	0.98	2.68	0.9963	0.02
11	0	98	0.98	3.09	0.999	0.02
12	1	99	0.99	3.49	0.9998	0.01

13	0	99	0.99	3.90	0.9999	0.01
14	0	99	0.99	4.31	0.9999	0.01
15	0	99	0.99	4.72	0.9999	0.01
16	1	100	1.00	5.12	0.9999	0.00

وباتباع الخطوات الآتية:

ا. حساب التكرار الملاحظ لكل قيمةFrequency.

ب. حساب التكرار التراكمى التصاعدي Cum. Freq.

ج. حساب قيمة$SN_{(xi)}$ حيث:

$$S_{N(xi)} = \frac{\text{التكرار التراكمى}}{\text{حجم العينة}}$$

- N حجم العينة
- xi القيمة فى التوزيع .

وعليه فان S_N للدرجات2,1 كالآتى:

$$S_{100}(1) = \frac{24}{100} = 0.24$$

$$S_{100}(2) = \frac{32}{100} = 0.32$$

وهكذا لقيمة التكرار التراكمى المقابل لكل قيمة.

د. حساب الدرجة المعيارية Z وهى درجة معيارية متوسطها صفر والانحراف المعيارى لها واحد. ولحساب التوزيع التراكمى الاعتدالى لاختبار K.S لعينة واحدة لا بد من حساب المتوسط:

$$\mu = \frac{\sum X}{N} = \frac{\text{مجموع القيم}}{\text{عددها}} = 3.41$$

والانحراف المعيارى : $\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N}} = 2.458$

وعليه يتم حساب الدرجة المعيارية Z لكل درجة مقاسة حيث :

$$Z = \frac{X-\mu}{\sigma}$$

$$Z_{X1} = \frac{1-3.41}{2.458} = -0.98$$

$$Z_{X2} = \frac{2-3.41}{2.458} = -0.57$$

وتستخدم الدرجات المعيارية Z لايجاد القيمة المتنبأ بها Fo(xi) عن طريق الكشف فى جدول التوزيع الاعتدالى المعيارى (الملاحق)، وبالكشف فى جداول التوزيع الاعتدالى عن Z = - 0.98 فإن P(Fo(xi)) = 0.1635، وبالكشف عن Z = - 0.57 فان:

$$P(Fo(xi)) = 0.2843$$

هـ. حساب القيمة المطلقة للفرق بين $S_N(Xi)$ و $F_o(x)$:

$$D = |Fo(x) - S_N(Xi)|$$

وتكون القيمة المطلقة للدرجة 1 ، 2 هى:

$$D_{X1} = |0.1635 - 0.241| = 0.08$$

$$D_{X2} = |0.2843 - 0.321| = 0.04$$

وهكذا لباقى الدرجات وعليه فإن اقصى قيمة مطلقة هى D = 0.30 (للستة عشر درجة فى الجدول) وهى تعتبر القيمة المحسوبة.

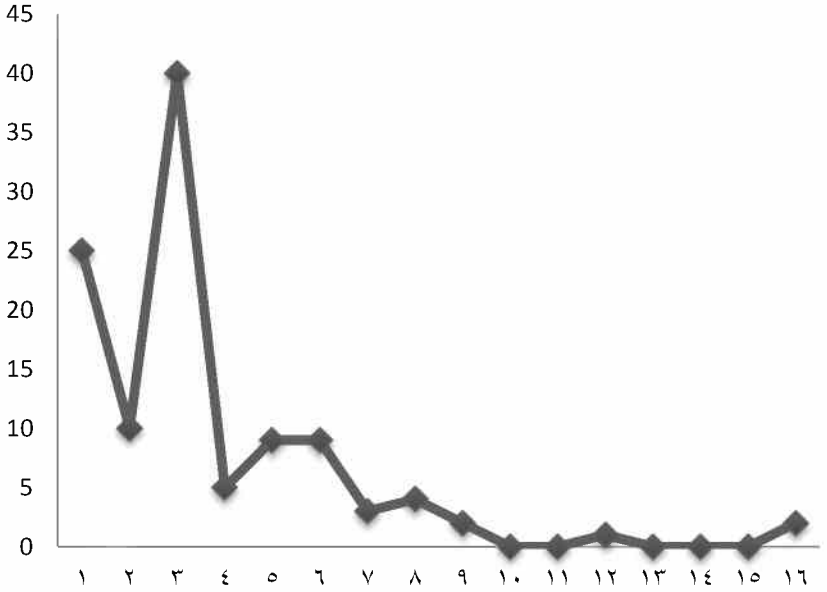
وإذا كان N > 35 فإن تقريب العينات الكبيرة يجب أن يستخدم لتحديد الدلالة الإحصائية فلمستوى دلالة إحصائية $\alpha = 0.05$ ، ولحجم العينة N = 100 يمكن البحث فى جداول(كولوموجوروف – سيمرنوف) عن القيمة الحرجة أو الجدولية فان:

$$= \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{100}} = 0.136$$

(الجدولية) (0.136) ≤ المحسوبة 0.30

وبالتالى فإن القيمة المحسوبة D أكبر من القيمة الحرجة ولذلك نرفض الفرض الصفرى وعليه فإن توزيع درجات المتغير (X) غير اعتدالى ، وعليه فإن استخدام الإحصاء البارامترى غير مرغوب ويفضل استخدام الإحصاء اللابارامترى .

وبالتأكيد على ان التوزيع ليس اعتدالى تم عرض الشكل البيانى الآتى:



والتمثيل البيانى تم فى ضوء عرض الدرجات على المحور الأفقى (السينى) والتكرارات على المحور الرأسى (الصادى) ويظهر من الشكل ان التوزيع ليس اعتدالى إنما هو ملتوى التواء موجب.

كتابة نتائج اختبار *K.S* لعينة واحدة وفقا لـ*APA*

اوضحت نتائج اختبار $K.S$ للتحقق من الاعتدالية للمتغير:

$P < 0.05$ ، $D(100)=0.30$ ووعليه فإنه توجد دلالة إحصائية، بالتالى فإن التوزيع غير اعتدالى.

التحقق من اعتدالية البيانات فى SPSS

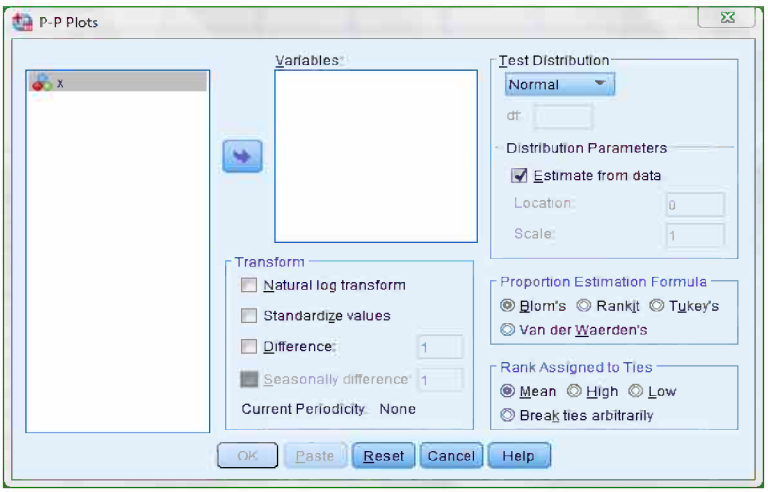
- **التحقق من الاعتدالية من العرض البيانى:**

اولًا: إدخال البيانات: يتم إدخال بيانات متغير X في المثال السابق في البرنامج وذلك من خلال تسميته من خلال الضغط على Variable view ثم كتابة مسمى المتغير تحت عمود Name ثم اضغط على Data view

- **شكل plots probability-Probability(P-P)**: في هذا المنحنى يتم حساب الرتب المناظرة للدرجات ولكل رتبة يتم حساب قيمة Z الفعلية المناظرة وهذه هي القيمة المتوقعة ثم المطابقة بين الدرجات الخام بالدرجات المعيارية المناظرة لرتب الدرجات فإذا كانت الدرجات لها توزيع اعتدالى فإن الدرجات Z الفعلية سوف تكون على خط قطرى مستقيم. والفكرة في هذا المنحنى هو مقارنة نقاط أو احداثيات البيانات بالخط المستقيم القطرى وإذا وقعت الدرجات على القطر فإن بيانات المتغير اعتدالية والابتعاد عن القطر يدل على الابتعاد عن الاعتدالية.

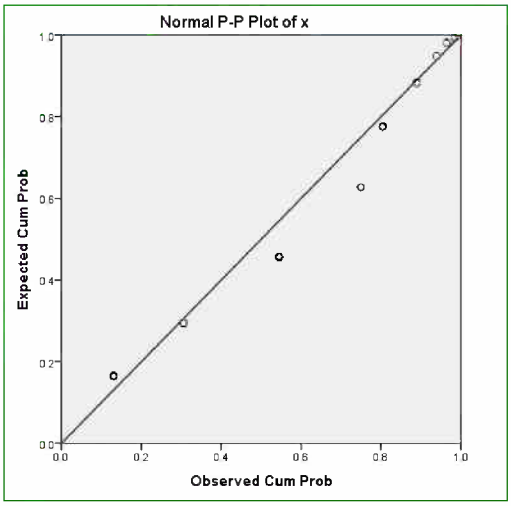
لتنفيذه اتبع الخطوات الآتية:

1 . اضغط Analyze ثم اختار Descriptive Statistics ثم اضغط على p-p.plots تظهر الشاشة الآتية:



2 . انقل متغير X إلى مربع Variables. ثم اضغط على OK.

ثالثا: المخرج:



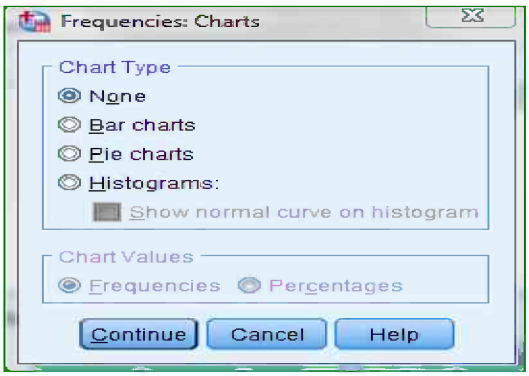
كمـا هـو ملاحـظ أن الإحـداثيات أو النقـاط لا تقـع تمامًـا عـن الخـط القطـرى وعليـه فـإن البيانـات تبتعـد عـن الاعتداليـة ولا يتـوافر فيهـا هـذه المسـلمة، وكمـا هـو ملاحـظ أن القـيم الملاحظـة علـى المحـور السـينى والقيمـة المتوقعـة علـى المحـور الصـادى. ويتضـح تمامًـا أن الإحـداثيات أو نقـاط البيانـات لا تقـع علـى الخـط المسـتقيم وعليـه فـإن البيانـات غيـر اعتدالية.

- المدرج التكرارى **Histogram**: تنفيذ الأمر :

1 . اضغط Analyze ثم اضغط Descriptive Statistics ثم اضغط Frequencies تظهر الشاشة:

2. اضـغط علـى المتغيـر x ثـم انقلـه إلـى مربـع Variables

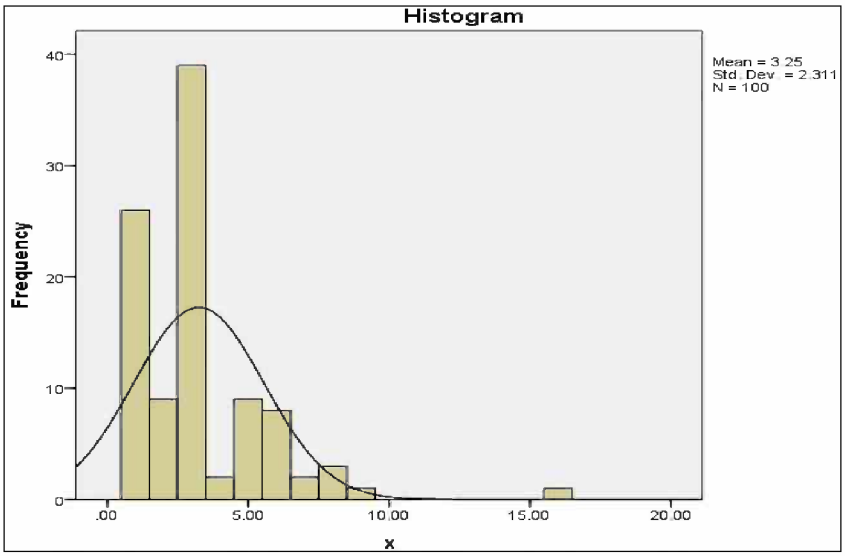
3. اضـــــغط علـــــى اختيـــار Charts علـى اليمـين تظهر الشاشة الآتية:



4. اضغط على Histograms و Show with normal curve

5. اضغط على Continue ثم Ok

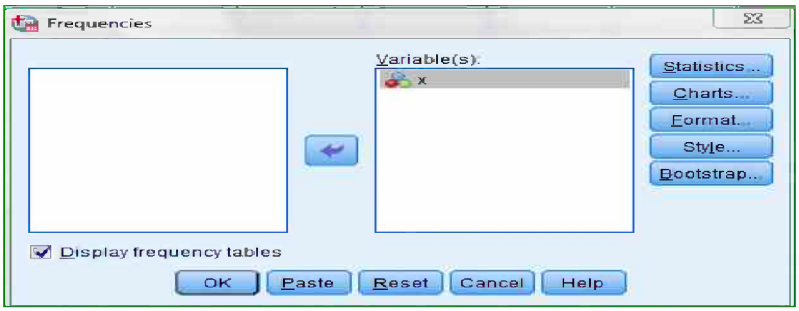
المخرج :



يظهر من المخرج ان المنحنى ملتوى ناحية اليمين بمعنى التواء موجب.

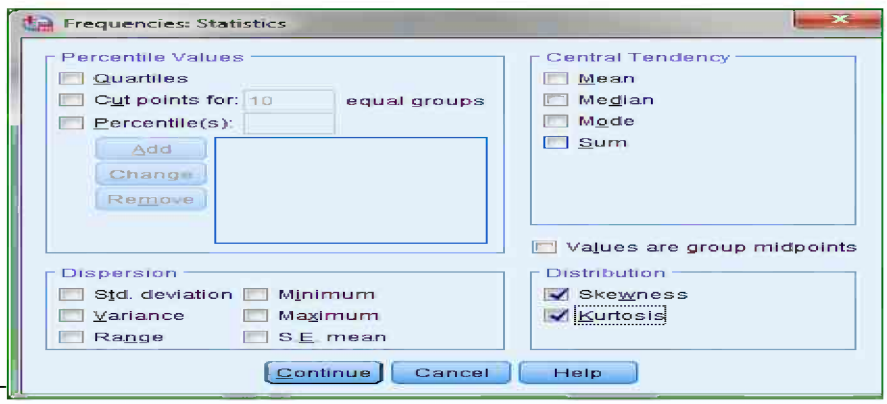
حساب مؤشرى التفرطح والالتواء باعتبارهما من خصائص التوزيع ويمكن تنفيذها فى البرنامج:

Analyze→Descriptive Statistics→Frequencies



2. انقل المتغير x إلى مربع Variables

3. اضغط على Statistics تظهر الشاشة:



4 . اضغط على Skewness ,Kurtosis

5. اضغط Continue ثم OK

المخرج:

Statistics

x

N	Valid	100
	Missing	0
Skewness		2.164
Std. Error of Skewness		.241
Kurtosis		8.447
Std. Error of Kurtosis		.478

يتضح أن قيمة الالتواء = (2.164) أى زادت عن الواحد الصحيح بما يدل على وجود التواء وبما أن قيمته موجبة اذن هو التواء موجب وقيمة التفرطح = (8.447) أى زادت عن الواحد الصحيح بل قيمتها عالية جدًا بما يدل على ان التوزيع ليس اعتدالى بل يوجد تفرطح وعليه فإن البيانات غير اعتدالية التوزيع، ويمكن تحويل قيم الالتواء والتفرطح إلى قيمة Zلاختبار دلالتها الإحصائية لها حيث ان:

$$Z_{skew} = \frac{S}{SE_{skew}}$$

$$\text{Z kurt} = \frac{\text{k - 0}}{\text{SEkur}}$$

حيث S قيمة الالتواء و SEskew الخطأ المعيارى للالتواء و K قيمة التفرطح و SE kurt الخطأ المعيارى للتفرطح على حدة وعليه فإن :

$$\text{Z skew} = \frac{2.164}{0.241} = 8.979$$

$$\text{Z kurt} = \frac{8.447}{0.478} = 17.671$$

لاختبار الدلالة الإحصائية للالتواء والتفرطح يتم مقارنة Zskew و Zkurt بـ 1.96 وهى قيمة Z لاختبار ذو ذيلين عند 0.05 أو مقارنتها بـ 2.56 قيمة Z لاختبار ذو ذيل واحد عند 0.05 وعليه فأن:

$$1.96 < (8.979) \rightarrow Z_{skew}$$

$$1.96 < (17.671) \rightarrow Z_{kurt}$$

واذًا توجد دلالة إحصائية لقيمتى الالتواء والتفرطح وعليه فإن التوزيع غير اعتدالى. ولكن عليك ان تكون حذرًا عند استخدام اختبار Z لأنه يعطى دلالة للعينات الكبيرة و ينصح Field (2014) باستخدام Z فى حالة العينات الصغيرة والمتوسطة (50,100, 150) ولكن إذا زادت حجم العينة عن 200 يفضل عدم الاعتماد على الدلالة الإحصائية لمؤشرى الالتواء والتفرطح والاعتماد على قيمتهما المطلقة.

اختبارات الدلالة الإحصائية للتحقق من الاعتدالية

يوجد اختبارين للتحقق من الدلالة الإحصائية للاعتدالية وهما اختبار كولوموجروف – سميرنوف لعينة واحدة كذلك اختبار شابيرو ويلك Shapiro -wilk ولكن من محدداتهما تأثرهما بأحجام العينات الكبيرة حيث من المتوقع مع حجم عينة كبير نحصل على دلالة إحصائية حتى لو ابتعد توزيع درجات المتغير قليلا عن الاعتدالية.

الفروض الإحصائية: الفرض الصفرى (H0): توزيع درجات المتغير المتصلx اعتدالى .

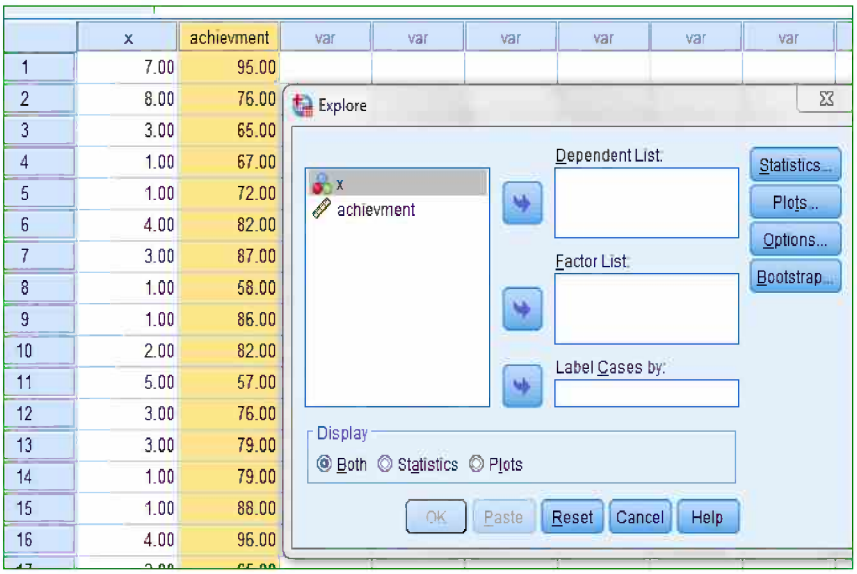
الفرض البديل (HA) : توزيع درجات المتغير غير اعتدالى.

وعدم الدلالة الإحصائية للاختبارين يعنى توافر الاعتدالية.

لتنفيذ الاختبارين فى برنامج الـ SPSS اتبع الآتى:

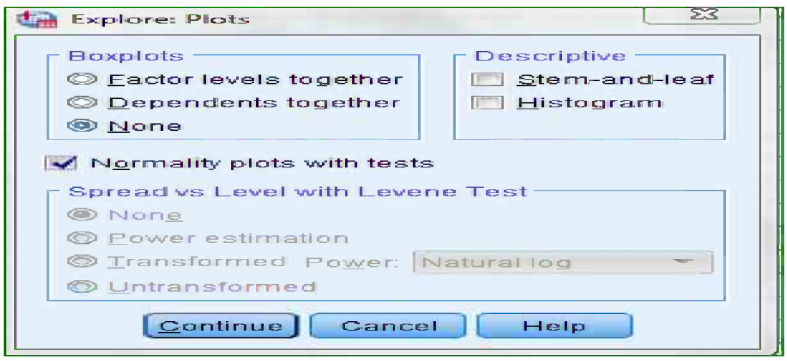
اولًا: إدخال البيانات: 1. اكتب مسمى المتغير X تحت عمود Name ، ثم اضغط على data view

ثانيًا: تنفيذ الأمر: 1 . اضغط على Analyze ثم اضغط DescriptiveStatistics ثم اضغط على Explore يعطى الشاشة الآتية:



2 . انقل المتغير X إلى المربع Dependent list عن طريق الضغط على السهم →

3 . اضغط على اختيار Plots تظهر الشاشة الآتية:



4 . اضغط على اختيار Normality Plots with tests وهذا يعطى رسم Q-Q Plot

5. اضغط على Continue ثم اضغط Ok

ثالثًا: المخرج: يعطى إحصائيات:

Descriptives				
			Statistic	Std. Error
x	Mean		3.2500	.23110
	95% Confidence Interval for Mean	Lower Bound	2.7914	
		Upper Bound	3.7086	
	5% Trimmed Mean		3.0111	
	Median		3.0000	
	Variance		5.341	
	Std. Deviation		2.31104	
	Minimum		1.00	
	Maximum		16.00	
	Range		15.00	
	Interquartile Range		3.00	
	Skewness		2.164	.241
	Kurtosis		8.447	.478

ثم عرض البرنامج:

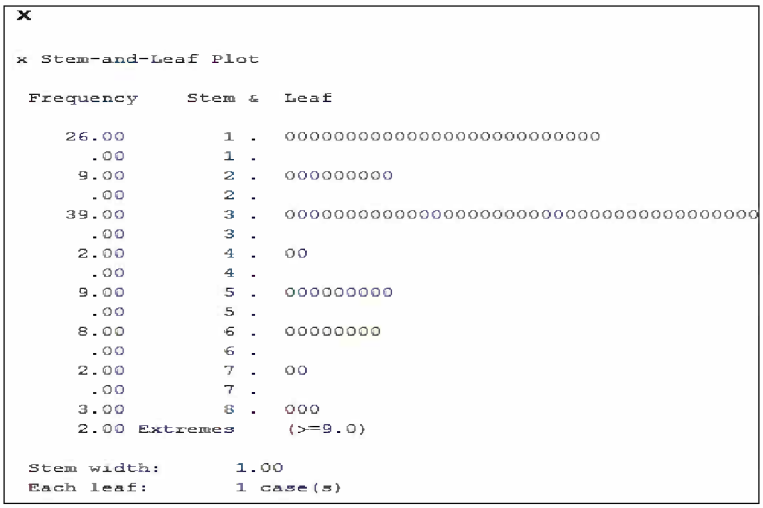
Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
x	.283	100	.000	.788	100	.000
a. Lilliefors Significance Correction						

وبالنسبة لاختبار K.S يتضح أن قيمته : 0.283 = (100) K.S بينما كانت قيمته بالحسابات اليدوية 0.30 تقريبًا وكانت درجات الحرية 100= df واتضح ان قيمة الاختبار دالة إحصائيًا حيث : **α(0.05) > (0.00) P(Sig)** وتم رفض الفرض الصفرى على ذلك توجد دلالة إحصائية وعليه فإن توزيع المتغير X غير اعتدالى.

وهكذا بالنسبة لاختبار Shapiro-wilk حيث إن إحصائية أو قيمة الاختبار 0.788 و df = 100.

وبالنسبة للقرار بما أن : **P(Sig) (0.00) < (0.05)** وعليه يتم رفض الفرض الصفرى وبالتالى توجد دلالة إحصائية وعلى ذك فإن توزيع درجات X غير اعتدالى.

وعرض الشكل الآتى:



```
x

x Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

    26.00        1 .  00000000000000000000000000
      .00        1 .
     9.00        2 .  000000000
      .00        2 .
    39.00        3 .  000000000000000000000000000000000000000000000
      .00        3 .
     2.00        4 .  00
      .00        4 .
     9.00        5 .  000000000
      .00        5 .
     8.00        6 .  00000000
      .00        6 .
     2.00        7 .  00
      .00        7 .
     3.00        8 .  000
     2.00 Extremes    (>=9.0)

 Stem width:      1.00
 Each leaf:       1 case(s)
```

ويتضح وجود التواء موجب.

الفصل الثالث عشر

معامل ارتباط بيرسون الخطى

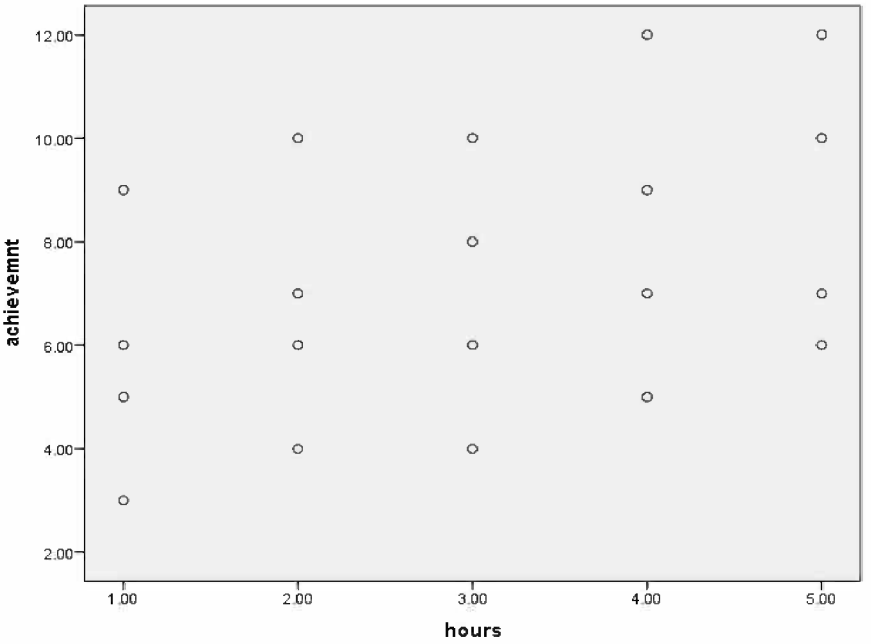
Linear correlation Coefficient Pearson

الارتباط هو أسلوب إحصائى يستخدم لقياس ووصف العلاقة بين متغيرين وقياسات المتغيرين موجودة بالفعل ولا يوجد أى محاولة لضبط أو معالجة المتغيرات، وبيانات هذا الأسلوب تتبع من تصميم البحث الارتباطى، والإحصاء المستخدم لقياس الارتباط يسمى معامل الارتباط coefficient Correlation وهو قياس كمى يستخدم لقياس قوة واتجاه العلاقة الخطية أو الارتباط بين متغيرين أو عاملين وقيمته تتراوح من 1.00 – إلى 1.00-. معامل الارتباط بين متغيرين يسمى correlation Bivariate ويتبع إحصاء المتغيرين statistics Bivariate، وإذا وجدت علاقة Association بين متغيرين يقال انهما مرتبطين وهذا يعنى يوجد بينها تغير Co-vary بمعنى تغير الدرجات على أحد المتغيرين يتبعه تغير على المتغير الآخر بكلمات اخرى المتغيرين ليس مستقلين. ولا بد التميز بين العلاقات والفروق فاذا وجدت فروق بين متوسطات المعالجات فهذا يعنى وجود ارتباطات بين المعالجات و القياسات التابعة، ولكن فى التجربة الباحث يهتم بدراسة متغير مستقل بمستويات ومتغير تابع متصل، وعليه يهتم بدراسة فروق بينما العلاقات يهتم الباحث بمتغيرين (قد يكون احدها مستقل والآخر تابع) وهذا ينطبق على معامل ارتباط بيرسون وكل التحليلات الإحصائية القائمة على الارتباطات هى جزء من النموذج الخطى العام General linear model.

العرض البيانى لشكل الانتشار Scatter plot or diagram

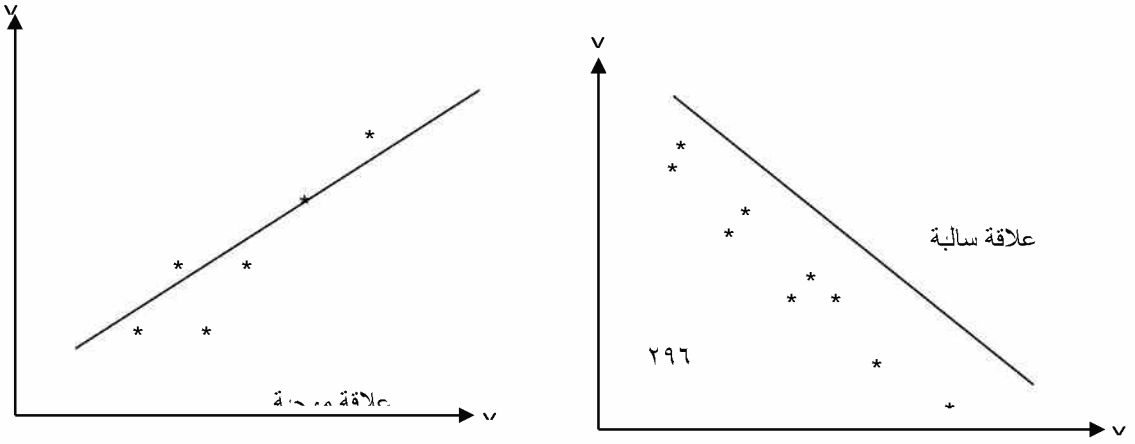
احد الطرق الشائعة لفحص العلاقة بين متغيرين هى عرضها فى ضوء شكل رسم بيانى يسمى شكل الانتشار Scatter plot أو يسمى Scatter diagram وهو عرض بيانى لبيانات متغيرين على محورى (Y,X) ويستخدم لتلخيص العلاقة بين المتغيرين ويوضع المتغير Y على المحور الصادى ويتم تمثيل زوجى القيم بنقطة وتسمى نقاط البيانات أو الاحداثى يطلق عليها Bivariate point ونمط الشكل يصف قيمة واشارة

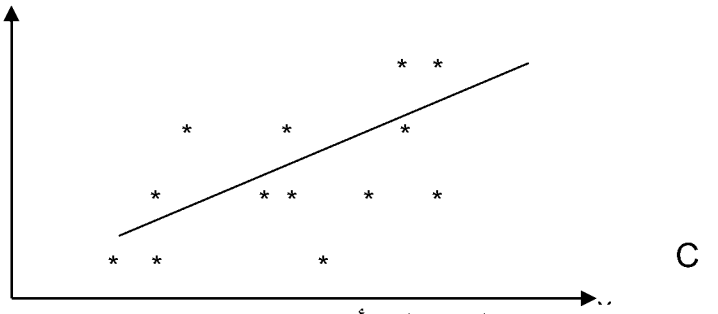
الارتبـاط و فيمـا يلـى شـكل يبـين هـذا فـافترض يوجـد بيانـات لمتغيـرين عـدد السـاعات الـذى يقضيه الطالب فى الاستذكار والتحصيل كالآتى:



B

A





الشكل (1.13) : أشكال انتشار للعلاقة بين Y, X

لاحظ فى شكل (A) ان قيم التحصيلY تزيد كلما زادت قيم الذكاء X ولاحظ أن الاحداثيات فى شكل A , B تتجمع تقريبًا على الخط المستقيم و هذا يعنى أفضل مطابقة للبيانات وهذه الخطوط يطلق عليها خطوط الانحدار Regression lines، وإلى أى درجة النقاط(الاحداثيات) تتجمع حول خط الانحدار هو مرتبط بالارتباط بين Y,X ومدى تمركز النقاط حول خط الانحدار يشير إلى علاقة قوية بين المتغيرين وإذا وقعت النقاط كلها على الخط المستقيم (الانحدار) تمامًا فإنها تكون علاقة ارتباطية تامة سواء كانت موجبة (1.00+) أو سالبة (1.00-) وكلما زاد تشتت وابتعاد النقاط عن الخط المستقيم انخفضت قيمة الارتباط وفى العلوم النفسية والاجتماعية إذا كانت درجة العلاقة 0.81 فاكثر فإنها علاقة قوية أو قوية جدًا(Howell, 2013). وبجانب أهمية مقاييس الارتباط فى تقدير العلاقة بين المتغيرين فإنه مفيد جدًا لإجراء التنبؤ Prediction فاذا وجدت علاقة ارتباطية بين متغيرين فمن المحتمل أن يكون أحدهما سببًا للآخر، ولكن الارتباط بين المتغيرين ليس كافيًا كأساس لثبوت أو وجود السببية Causality ولكنه يعتبر الخطوة الأولى للبرهنة على ان المتغيرين مرتبطين سببيًا. الارتباط والانحدار مرتبطين تمامًا، فالاستفادة من الارتباط هو استخدامه فى التنبؤ فلا انحدار بدون ارتباط، فالارتباط يحدد ما إذا كانت توجد علاقة بين المتغيرين ويحدد حجمها واتجاهها بينما الانحدار يهتم أساسا باستخدام الارتباط للتنبؤ.

والاستخدام الآخر للارتباط هو استخدامه لتقدير معاملات الثبات سواء معامل الاستقرار الناتج من الاختبار واعادتهTest – Retest وكذلك الاختبار مكافئه Test – Equivalence حيث هو مؤشر لمدى اتساق الدرجات عبر مرات التطبيق لنفس الأداة أو الأداة ومكافئها.

حقيقة يعتبر معامل الارتباط على درجة كبيرة من الأهمية فى عالم الإحصاء المتدرج سواء كان فى الانحدار المتعدد أو التحليل العاملى أو تحليل المسار أو نمذجة المعادلة البنائية لأن كل هذه الأساليب قائمة على الارتباط.

ومعامل الارتباط يستخدم لتقدير الصدق Validity فى مظاهره العديدة سواء كان صدق المحكى من خلال معامل الارتباط بين قياسات المقياس الذى قمت بإعداد هو محك اخر(صدق محكى) أو صدق تميزى من خلال معامل الارتباط بين بنائين مختلفين أو صدق تقاربى من خلال معامل الارتباط بين بنائين متشابهين وايضًا الصدق البنائى من خلال التحليل العاملى سواء الاستكشافى أو التوكيدى حيث يعتبر معامل الارتباط هو البداية لهذه الأساليب الإحصائية.

هدف تحليل الارتباط

الهدف من إجراء تحليل الارتباط هو اكتشاف وجود علاقات ذات دلالة معنوية بين المتغيرات ويساعدنا تحليل الارتباط على الآتى:

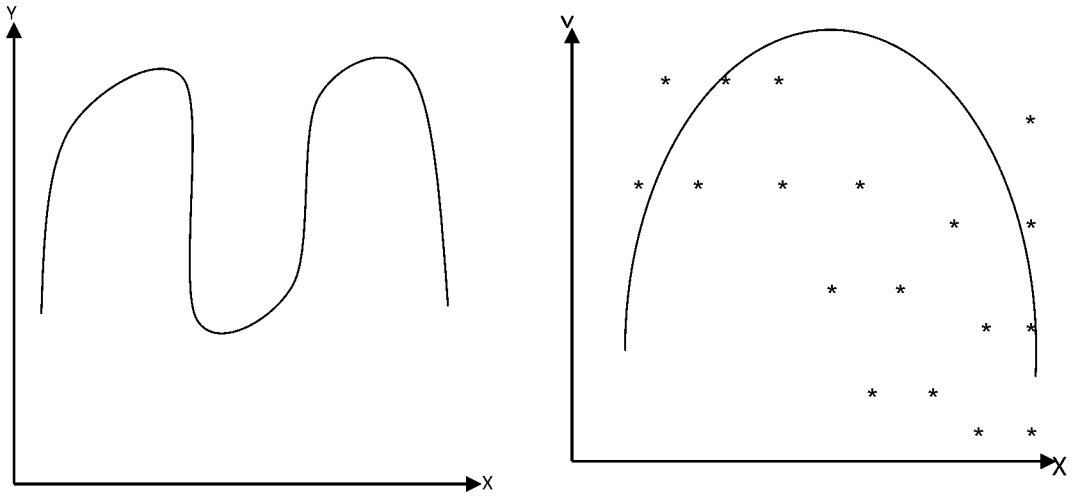
أولًا: اتجاه العلاقة Direction of the relation : فإشارة الارتباط سواء كانت:

- موجبة Positive: المتغيرين يتغيروا فى نفس الاتجاه فكلما زاد قيم المتغير **X** للأفراد فإن قيم المتغير Y تزيد أيضًا وكلما انخفضت قيم **X** تنخفض قيم **Y** (انظر شكلA)، بكلمات اخرى توجد علاقة طردية بين المتغيرين.
- سالبة Negative: يشير إلى ان المتغيرين فى اتجاه معاكس فكلما زادت قيم X تنخفض أو تنتقص قيم Y وهذه علاقة معكوسة Inverse relationship (شكلB 1.2) بكلمات اخرى توجد علاقة عكسية بين المتغيرين.
- علاقة صفرZero: عدم وجود علاقة خطية (عدم وجود خط تتجمع حوله النقاط) بين المتغيرين.

العلاقة التامة وغير التامة

ففى شكل (A 1.13) إذا كانت النقاط فى شكل الانتشار تقع على الخط المستقيم تمامًا فيقال ان العلاقة تامة موجب Perfect positive relationship وقد تكون علاقة تامة سالبة شكل(B 1.2)، وفى العلوم النفسية هذه العلاقة غير متوفرة وإذا توفرت فيقال ان العلاقة بين المتغيرين خطية تامة بمعنى أن تعكس خط مستقيم تتوزع عليه الاحداثيات حيث تكون العلاقة غير تامةImperfect حيث تتوزع بعض النقاط بعيدًا عن الخط المستقيم ويمكن أن تكون علاقة ارتباطية جزئية سواء كانت موجبة أو سالبة.

اما العلاقة غير الخطية Non-linear relationship وهى علاقة منحنية وهى تمثل كالآتى:



الشكل (2.13): شكل الانتشار للعلاقات غير الخطية .

اذا تحققت هذه العلاقات فلا داعى لاستخدام معامل ارتباط بيرسون لأنه يفترض ان تكون العلاقة خطية بقدر الامكان بين المتغيرين Linearity.

ثانيًا: مقدار أو حجم أو اتساق العلاقة: معامل الارتباط يتراوح من 1.0– إلى 1.0 + فإشارة معامل الارتباط تحدد ما إذا كان الارتباط موجبًا أو سالبًا، ولكن قيمته تصف مقدار حجم الارتباط فالقيمة العالية تشير إلى ارتباط قوى، فالقيمة 1.0+ تشير إلى معامل ارتباط موجب تام و1.0– تشير إلى معامل ارتباط سالب تام.

فيما يلى اطار لتحديد حجم قيمة معامل الارتباط كالاتي Dancy & Reidy (2011):

تام	+1	−1
	0.9	−0.9
قوي	**0.8**	−0.8
	0.7	−0.7
	0.6	−0.6
متوسط	**0.5**	−0.5
	0.4	−0.4
	0.3	−0.3
ضعيف	**0.2**	−0.2
	0.1	−0.1
صفر	**0**	0

ووضع Hinkle et al. (1994) معايير لمعامل ارتباط r =0.30 منخفض و 0.70 متوسط و 0.90 فأكثر قوى، بينما وضعها Cohen (1988) القيمة 0.1 ضعيف والقيمة 0.30 متوسط والقيمة 0.50 فاكثر قوى.

معامل ارتباط بيرسون الخطي

الهدف: يقيس درجة أو قوة واتجاه العلاقة الخطية بين متغيرين متصلين(فترى على الأقل) ويطلق عليه Pearson Product – Moment Correlation.

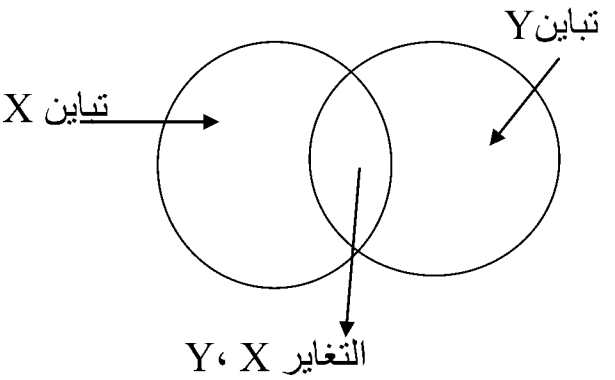
حسابه: توجد صيغ عديدة لحساب معامل ارتباط بيرسون أهمها:

1. **التغاير Covariance:** معامل الارتباط للبيانات قائم على إحصاء التغاير Cov_{xy} و هو يعكس درجة تغاير للمتغيرين معًا و يمكن حسابة كالآتى:

$$Cov_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{N-1}$$

- $\bar{x}$ متوسط X، $\bar{Y}$متوسط قيم Y

- Nحجم العينة.



حيث ان:

$$r = \frac{cov_{xy}}{SD_x SD_y}$$

- SD_x , SD_y الانحراف المعيارى لدرجات X , Y

2. **من الدرجات المعيارية:** تقدر كالآتى:

$$r = \frac{\Sigma Zx Zy}{N-1}$$

- Zx الدرجة المعيارية لـx, Zy الدرجة المعيارية لـ Y
- ΣZx Zy مجموع حاصل درجات الدرجات المعيارية لـ X فى الدرجات المعيارية لـ Y
- N حجم العينة و احيانًا يتم القسمة على N (إحصاء وصفى).

3. **من الدرجات الخام:** يمكن ترجمة معادلة بيرسون من الدرجات المعيارية إلى الصيغة الآتية:

$$r_{xy} = \frac{N\sum XY - N\sum X \sum Y}{\sqrt{(N(\sum X^2 - (\sum X)^2)}(\sqrt{N(\sum Y^2 - (\sum Y)^2)}}$$

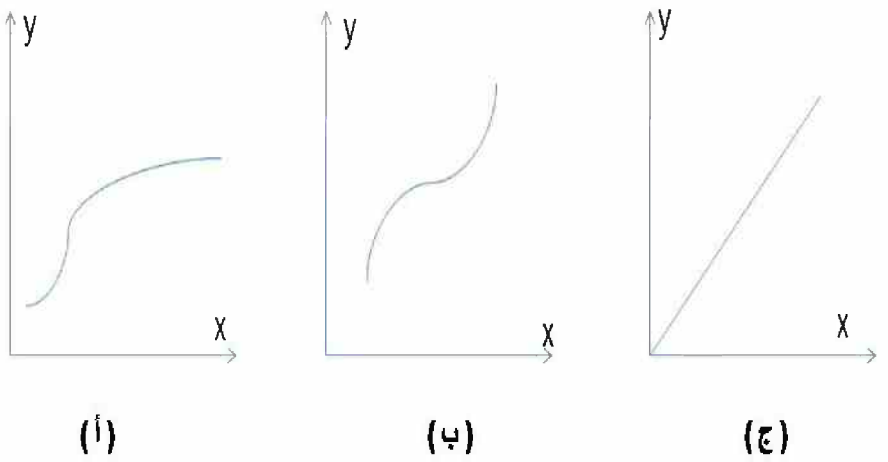
- ΣXY مجموع حاصل ضرب X فى Y.
- ΣX مجموع قيم X، ΣY مجموعة قيم Y.
- ΣX^2 مجموع مربعات X، $(\Sigma X)^2$ مربع مجموع X.

- ΣY^2 مجموع مربعات Y، $(\Sigma Y)^2$ مربع مجموع Y ، N حجم العينة.

العوامل المؤثرة فى حجم معامل الارتباط

يلعب معامل الارتباط دورًا أساسيًا فى أساليب إحصائية متنوعة ولذلك فمن الضرورى فهم العوامل المؤثرة فيه وهى كالآتى:

1. الخطية Linearity: يشير معامل ارتباط بيرسون إلى درجة الارتباط الخطى بين متغيرين، ومدى ابتعاد أحد المتغيرين عن هذه العلاقة الخطية يؤثر على حجم معامل الارتباط، ويتطلب استخدام معامل ارتباط بيرسون توافر العلاقة الخطية بين المتغيرين ويمكن أن تكون العلاقة منحنية خفيفة لا تؤثر فى اتجاه معامل الارتباط بل تؤثر فى قيمته كالآتى:



الشكل (3.13): أمثلة العلاقة الخطية و غير الخطية الخفيفة.

وفى هذا التمثيل بين المتغيرين فإن العلاقة موجبة لأنه كلما زادت x تزداد y، وهذه الأشكال (أ، ب) يطلق عليها **S-shaped**. وعلى ذلك لا بد من التحقق من توافر الخطية من خلال العرض البيانى أو رسم المنحى. وشكل (ج) هو مثال للعلاقة الخطية التامة وتعتبر العلاقة الخطية مؤشر على وجود الاعتدالية.

2. الثبات للقياسات: القياسات فى العلوم الاجتماعية والسلوكية دائمًا غير تامة الثبات ، وذلك لوجود أخطاء قياس واقعة على المتغيرات. ويمكن التعبير عن العلاقة بين الدرجة الحقيقية والمقاسة كالآتى:

$$T=X \pm E$$

- T هى الدرجة الحقيقة، ، X الدرجة المشاهدة (المقاسة)، E خطأ القياس.

وكلما كانت T قريبة جدًا من X فإن أخطاء القياس تكون أقل ما يمكن. ويعرف الثبات مدى تطابق الدرجة المقاسة مع الدرجة الحقيقة فإذا كانت E=0 فإن T=X، ولكن هذا صعب المنال فى قياسات العلوم الإنسانية، وعليه فإن قياساتها تتضمن أخطاء، فإذا كان الثبات = 0.80 فإن هذا يفيد بان 20% من تباين المتغير (الاختبار) يعزو إلى الخطأ. وعليه فإن الثبات المنخفض للقياسات (أخطاء عالية) تؤثر على حجم معامل الارتباط وكذلك معاملات الانحدار ويحدث لمعاملات الارتباط تقلص لقيمته Attenuation (Kenny, 1987).

3. **التجميع الكلى Aggregation:** عندما يتم التعبير عن الدرجة فى ضوء المتوسط أو تجميع لعدد من المفردات فى متغير واحد فإنها تسمى Aggregate data وعمومًا فإن الارتباط المقدر من الدرجة الكلية (التجمعية) أكبر من الارتباط المقدر للدرجات المفردة Individual score، وعلى ذلك فإن تجميع الدرجات المفردة فى درجة كلية يؤدى إلى تضخم Inflation لقيمة معامل الارتباط لأن ثبات مجموع المفردات معًا أكبر من ثبات كل مفردة على حدة.

4. **ارتباط الكل بالجزء Part – Whole Correlation:** الارتباط يتضمن متغيرين ولكن فى بعض الأحيان أحد المتغيرين مشتق من الآخر أو متضمن فى الآخر وعلى ذلك يوجد تباين مشترك كبير بينهما.

الجدول (1.13): ارتباط الكل بالجزء.

المتغير الأول	المتغير الثاني	التحيز فى r
X	**X+ Y**	**موجب**
X	**Y-X**	**سالب**
X+Z	**X+ Y**	**موجب**
X + Z	**Y – X**	**سالب**
Y / X	**W / X**	**موجب**

5. نـوع مسـتوى القيـاس للمتغيـرات: يوجـد أربعـة مسـتويات قيـاس شـائعة وهـى الاسـمى، الرتبـى، الفتـرى، والنسـبيولا ينصـح التعامـل مـع مصـفوفة ارتبـاط أو تغـاير بهـا خلـيط مـن مسـتويات القيـاس.وتتنوع معـاملات الارتبـاط حسـب مسـتوى القيـاس الـذى تتعامـل معـه كالآتى:

الجدول (2.13): أنواع معامل الارتباط حسب مستوى القياس.

م	معامل الارتباط	مستوى القياس
1	**بيرسون (r)**	**كلًا من المتغيرين فترى على الأقل**
2	**سبيرمان – كندل تاو (ρ)**	**كلا من المتغيرين رتبي**
3	**فاى (φ)**	**كلًا من المتغيرين اسمى بمستويين**
4	**Point – bi-serial الثنائى السلسل (r_{pb})**	**احد المتغيرات فترى (التابع) والآخر المستقل تصنيفى بمستوين**
5	**التسلسلى الرتبى (جاما) Gamma Rank bi-serial**	**احد المتغيرين رتبى والآخر اسمي**
6	**معامل كرامير (v)**	**كلا من المتغيرين اسمى (مثل فاى) و يكون احدهما بأكثر من مستويين .**
7	**Poly-serial**	**احد المتغيرين فترى والآخر رتبى أو تصنيفى مع ثلاثة مستويات فأكثر مع بناء تحتى متصل**
8	**الرباعى Tetra- choric**	**كلا من المتغيرين تصنيفى بمستويين (0, 1) مع بناء تحتى متصل وتوزيع اعتدالي**
9	**Poly-choric**	**كلا من المتغيرين ترتيبى مع بناء تحتى متصل أو تصنيفى بأكثر من مستويين مع بناء تحتى متصل**

6. المـدى للمتغيـرات: لا بـد أن تكـون المتغيـرات لهـا مـدى معقـول لأنـه لـو كـان مـدى الـدرجات مقيـد (صـغير) فـإن مقـدار قيمـة الارتبـاط تـتقلص deflation، وذلـك لأن عـدم وجـود تجـانس بـين الـدرجات يـؤدى إلـى تبـاين أق لوبـدوره يـؤثر علـى الارتبـاط ويحـدث ذلـك للمتغيـرات التصـنيفية وأحيانًـا الرتبيـة. وتقلـص المـدى يـؤثر بدرجـة كبيـرة علـى معـاملات الارتبـاط، ولـذلك فإنـه يفضـل اسـتخدام معـاملات الانحـدار كبـديل لمعـاملات الارتبـاط لأن معـاملات الانحـدار لا تتـأثر بـتقلص المـدى وهـذا ينطبـق علـى المتغيـرات التصـنيفية ذات البناء التحتى المتصل.

7. القـيم المتطرفـة Outliers: القيمـة المتطرفـة هـى التـى تبتعـد بدرجـة كبيـرة عـن المتوسـط والوسـيط، وجـود القـيم المتطرفـة فـى البيانـات يـؤدى إلـى تشـوية حجـم الارتبـاط،

ويمكـن أن يكـون معامـل الارتبـاط خـادع Spurious وهـذا يرجـع إلـى سـلوكيات غيـر طبيعية تظهرها العينة بمعنى وجود بيانات متطرفة لاحد أفراد العينة ففى المثال الآتى:

الجدول (3.13) : مجموعات بيانات تتضمن قيمة متطرفة(Kenney, 1987).

الشخص	1	2	3	4	5
X	**2**	**3**	**1**	**4**	**10**
Y	**3**	**2**	**1**	**4**	**1**

فـى الجـدول توجـد قيمـة متطرفـة فـى المتغيـر X وهـى 10، وعلـى ذلـك فـإن قيمـة معامـل ارتبـاط بيرسـون 0.271- = r ولكـن إذا اسـتبدلت القيمـة 10 بالقيمـة 3 مـثلا فـإن (r = 0.572)، ونلاحـظ هـذا التغيـر الـدراماتيكى مـن(X=10)0.27- إلـى (X=3)0.572، وعلـى ذلـك فـإن القـيم المتطرفـة تـؤدى إلـى تقلـص قيمـة معامـل الارتبـاط وكـذلك إشارته أو العكس أحيانًا(Kenney, 1987).

اختبارات الفروض لقضية بحثية (بيانات من: Hinkle et al., 1994)

اراد باحث دراسة العلاقة بين درجات اختبار الاستعدادات والأداء الأكاديمى لعينة مكونة من 15 فردًا وكانت البيانات كالآتى:

الطالب	X الاستعداد	Y التحصيل
1	**595**	**68**
2	**520**	**55**
3	**715**	**65**
4	**405**	**42**
5	**680**	**64**
6	**490**	**45**
7	**564**	**56**
8	**580**	**59**
9	**615**	**56**
10	**435**	**42**

11	**440**	**38**
12	**515**	**50**
13	**380**	**37**
14	**510**	**42**
15	**564**	**53**
	X^{-}= 534.00	**Y^{-}=51.49**
	S_x= 96.53	**S_y=10.11**

الخطوات البحثية:

1. **سؤال البحث:** هل توجد أو ما طبيعة العلاقة الارتباطية بين الاستعداد و الأداء الأكاديمى؟

2. **فرض البحث:** توجد علاقة ارتباطية بين الاستعداد والأداء الأكاديمى.

3. **منهج البحث:** المنهج الارتباطى.

4. **متغيرات البحث:** فى معامل ارتباط بيرسون ليس من الضرورى تحديد أى المتغيرات مستقلة وايهما تابع ولكن فى ضوء الاطار النظرى: الاستعداد (فترى – متصل –مستقل) ، الأداء الأكاديمى (فترى – متصل – تابع).

5. **النموذج الإحصائى:** إحصاء النموذج البسيط أو Bivariate statistics **والأسلوب الإحصائى** معامل ارتباط بيرسون الخطى شرط توافر العلاقة الخطية وهو إحصاء بارامترى.

خطوات اختبارات الفروض الصفرية

على الرغم ان الباحثون يستخدمون معامل ارتباط بيرسون كإحصاء وصفى فى اغلب الاحيان ويفسرونه فى ضوء قيمته ولكن قد تكون قيمة معامل ارتباط بيرسون مرتفعة ولكنها غير دالة إحصائيًا نتيجة صغر حجم العينة، وعلى ذلك لا بد من اختبار الدلالة الإحصائية لمعامل ارتباط بيرسون إذا كان هدف الباحث هو الاستدلال أو التعميم من العينة إلى المجتمع وخطوات اختبارات الفروض الصفرية كالآتى:

1. صياغة الفروض الإحصائية:

- الفرض الصفرى: $H0: \rho = 0$

حيث ρ معامل الارتباط بين المتغيرين فى المجتمع، أى لا توجد علاقة ارتباطية بين المتغيرين.

- الفرض البديل: $H_A : \rho \neq 0$

توجد علاقة ارتباطية بين المتغيرين.

إذا كان فرض الباحث وجود علاقة ارتباطية موجبة تصبح الفروض الإحصائية كالآتى:

$$H0 : \rho \leq 0$$

$$H_A : \rho > 0$$

2. الاختبار المناسب ومسلماته: أسلوب معامل الارتباط بيرسون ومسلماته كالتالى:

- الاعتداليةNormality: توزيع كل متغير على حده اعتداليًا وإذا تحققت الاعتدالية فإن العلاقة بين المتغيرين تكون خطية وإذا لم تتحقق هذه المسلمة فإن العلاقة تكون غير خطية ويمكن التحقق من العلاقة الخطية من خلال عرض شكل الانتشار.
- انتقاء العينة عشوائيًا.
- الاستقلالية Independence: درجة كل فرد على أحد المتغيرات تكون مستقلة عن درجة فرد آخر فى المجموعة التى ينتمى اليها.
- المتغيرين من مستوى القياس الفترى على الأقل أى أن البيانات متصلة.

3. مستوى الدلالة الإحصائية و قاعدة القرار: حدد الباحثα = 0.50، درجات حرية:

$$df= n-2= 15-2=13$$

والاختيار ذو ذيلين ولتقيم الدلالة الإحصائية للارتباط فإنها تقدر خلال:

المدخل الأول: من خلال اختبار T وفى هذه الحالة تكون صيغة اختبار T للارتباط كالآتى:

$$T_r = \frac{Sample\ Statistics(r) - \rho}{\sqrt{\frac{(1-r^2)}{(n-2)}}}$$

$$T_r = \frac{r\sqrt{n-2}}{\sqrt{1-r2}}$$

حيث: $\rho = 0$ وهى المحددة فى الفرض الصفرى، r معامل ارتباط بيرسون المقدر.

ودرجات الحرية لاختبار T هى:

$$df = n-2=15-2 = 13$$

وتكون هذه قيمة T المحسوبة ويتم مقارنتها بـT الجدولية عند $\alpha= 0.05$ ، $df = 13$، ولاختيار ذو ذيلين تكون قيمتها الحرجة(الجدولية)=2.16 . وإذا كانت:

T المحسوبة > T الحرجة نرفض الفرض الصفرىH0.

المدخل الثانى: يمكن اختبار الدلالة الإحصائية لمعامل بيرسون من خلال الكشف فى جدول معامل الارتباط عند دلالة إحصائية 0.05 واختبار ذو ذيلين ودرجات حرية 13 فإن قيمةr الجدولية =0.514 ، وعليه إذا كانت **r المحسوبة > r الجدولية** نرفض H0.

المدخل الثالث: مقارنة قيمة p الاحتمالية(فى مخرج SPSS) بقيمة α حيث إذا كانت: $\mathbf{P \leq \alpha}$ نرفض H0

4. الحسابات: يتم حساب قيمة**r** من الصيغة المعيارية كالآتى:

X	Y	Zx	Zy	ZxZy
595	**68**	**0.63**	**1.64**	1.03
520	**55**	**-0.15**	**0.35**	-0.05
715	**65**	**1.88**	**1.34**	2.52
405	**42**	**-1.34**	**-0.94**	1.26
680	**64**	**1.51**	**1.24**	1.87
490	**45**	**-0.46**	**-0.64**	0.29
565	**56**	**0.32**	**0.45**	0.14
580	**59**	**-0.48**	**0.74**	0.36
615	**56**	**0.84**	**0.45**	0.38
435	**42**	**1.03**	**-0.94**	0.97
440	**38**	**-0.97**	**-1.33**	1.29
515	**50**	**-0.20**	**-0.15**	0.03
380	**37**	**-1.60**	**-1.43**	2.29

510	**42**	**-0.25**	**-0.94**	0.24
505	**53**	**0.32**	**0.15**	0.05
Σ 8010	**772**	**0**	**0**	12.67

X^{-}=53.4y^{-} = 51.47

Sx = 96.53 Sy = 10.11

حيثZx , Zy هى الدرجات المعيارية وتقدر كالآتى:

$$Zx = \frac{X - X^{-}}{Sx}$$

فمثلًا الدرجة المعيارية لـ595:

$$Zx_{595} = \frac{595-534}{96.53} = 0.63$$

$$Zy = \frac{y - y^{-}}{Sy}$$

$$Z_{68} = \frac{68-51.47}{10.11} = 1.64$$

وعليه فإن معامل ارتباط بيرسون يقدر كالآتى:

$$r = \frac{\Sigma ZxZy}{n-1} = \frac{12.67}{14} = 0.90$$

ويمكن حساب معامل ارتباط بيرسون الدرجات الخام:

X	Y	XY	X^2	Y^2
595	**68**	**40460**	**354025**	4624
520	**55**	**28600**	**270400**	3025
715	**65**	**46475**	**511225**	4225
405	**42**	**17010**	**164025**	1764
680	**64**	**43520**	**462400**	4096
490	**45**	**22050**	**240100**	2025
565	**65**	**31640**	**319225**	3136
580	**59**	**34220**	**336400**	3481
615	**56**	**34440**	**378255**	3136
435	**42**	**18270**	**189255**	1764
440	**38**	**16720**	**193600**	1444
515	**50**	**25750**	**265225**	2500

380	**37**	**14060**	**144400**	1369
510	**42**	**21420**	**260100**	1764
565	**53**	**29945**	**319225**	2809
Σ 8010	**772**	**42580**	**4407800**	41162

وتطبق الصيغة الآتية:

$$r_{xy} = \frac{N\sum xy - N\sum x\ \sum y}{\sqrt{(N(\sum x^2 - (\sum x)^2)}(\sqrt{N(\sum y^2 - (\sum y)^2)}}$$

$$= \frac{15(424580) - (8010)(772)}{\sqrt{(15(4407800) - (8010)^2(15(41162)(772))}} = 0.90$$

وكذلك تقدير معامل ارتباط بيرسون من خلال التغاير كالآتى:

$$\text{Cov} = \frac{\Sigma(x-x^-)(y-y^-)}{n-1}$$

وعليه لا بد من حساب مجموع حاصل ضرب انحراف القيمة X عن متوسطها فى انحراف القيمة Y عن متوسطها وبحسابه كان ناتجه:

$$\Sigma(x\text{-}x^-)(y\text{-}y^-) = 12.332$$

$$\text{Cov} = \frac{12.332}{15-1} = 880.86$$

اذا : $$r = \frac{covxy}{sxsy}$$

sx , sy الانحراف المعيارى لـ x ، y إذا :

$$r = \frac{880.66}{(96.53)(10.11)} = 0.90$$

ولحساب قيمة Tمن خلال r:

$$T = \frac{0.90-0}{\sqrt{\frac{(1-(0.9)^2)}{15-2}}} = \frac{0.90}{\sqrt{\frac{0.19}{13}}}$$

5. القرار والتفسير: أ. بمقارنة قيمة T المحسوبة بـT الحرجة حيث: T المحسوبة ( )> T الحرجة (2.16)، إذا يرفض H0 وعليه توجد علاقة ارتباطية بين درجات الاستعداد والأداء الأكاديمى وهى دالة إحصائية عند 0.05.

ب. بمقارنة قيمة r المحسوبة ب r الحرجة: r **المحسوبة (.90) > r الحرجة (.514)** وعليه يرفض H0.

6. حجم التأثير: لحساب نسبة التباين المفسر يتم تربيع معامل الارتباط ويسمى معامل التحديد Coefficient of determination وقيمته تنحصر من 0.0 الى1.0 وإذا فإن حجم التباين المفسر للبيانات:

$$r^2 = (0.90)^2 = 0.81$$

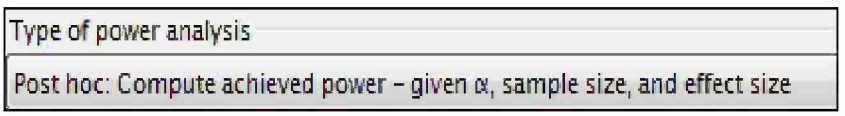
وعليه يمكن استنتاج أن درجات الاستعدادات فسرت 81% من تباين الأداء الأكاديمى وعلى ذلك فإن مؤشر r^2 هو مقياس لقوة العلاقة. ومعامل الاغتراب Coefficient of alienation:

$$=1 - r^2$$

القوة الإحصائية: لحساب القوة الإحصائية باستخدام برنامج**G-Power اتبع الآتى:**

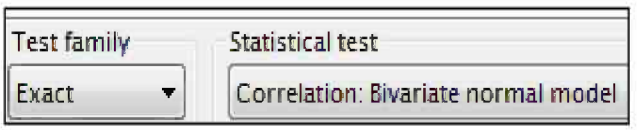
1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. تحت Type of power analysis اختار:

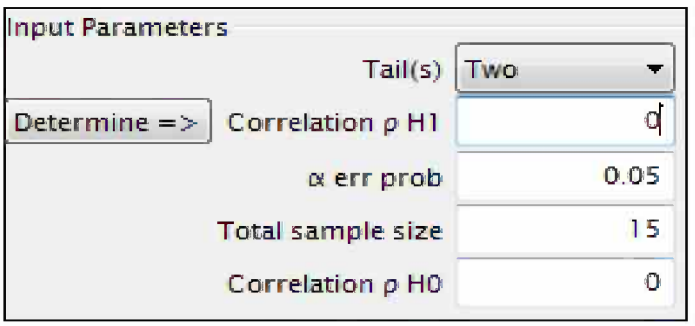


3. تحت Test family اختار Exact

4. تحت Statistcal test اختار:



5. ادخل المعالم الآتية تحت Input parameters:

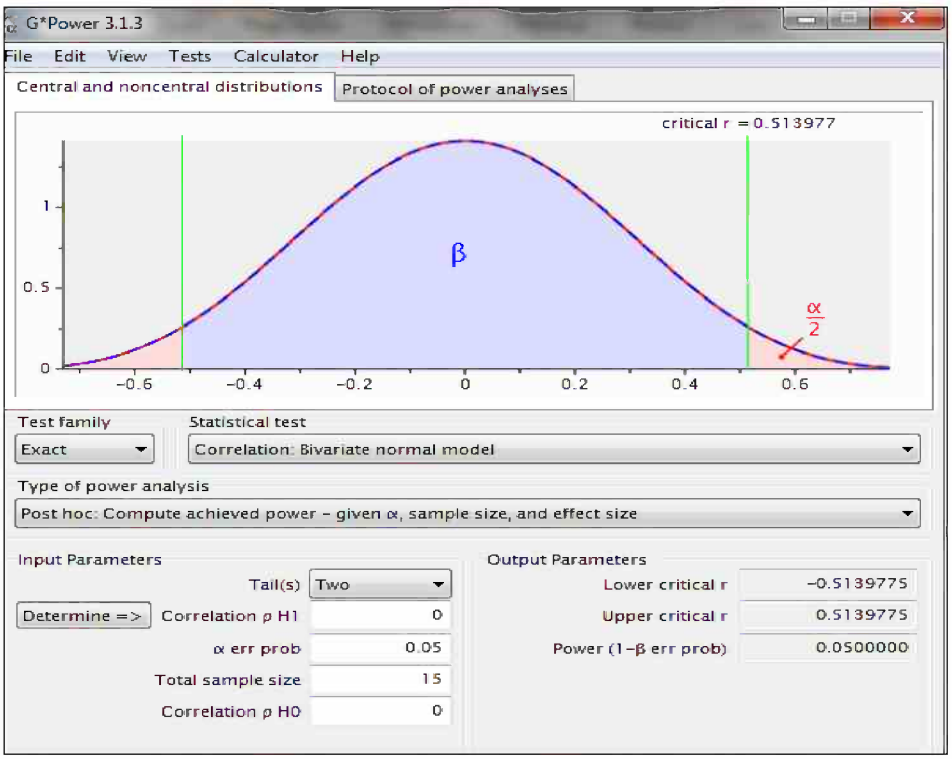


- نوعية الاختبار: ذو ذيلين

- مستوى الدلالة الإحصائية: 0.05

- حجم العينة الإجمالى= 15

6. اضغط Calculated تظهر المخرجات الآتية:



يتضح أن القوة الإحصائية= 0.05 وهذا مستوى ضعيف جدًا وقد يكون هذا نتيجة لعوامل عديدة منها تصميم بحث ردئ وجودة بيانات منخفضة وحجم عينة صغير.

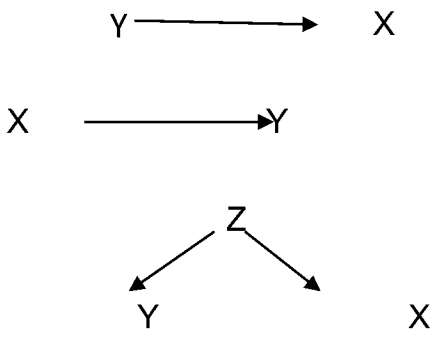
الارتباطية والسببية

اذا وجدت علاقة خطية دالة إحصائيًا بين متغيرين فيمكن افتراض وجود حد أدنى للسببية بينهما ولكن تفسير الارتباط ليس سببيًا لأننا لا نعرف اتجاه السببية بينهما Direction of causality ، ويوجد ثلاثة احتمالات ممكنة لاتجاه السببية بين X , Y هى:

1. أن المتغير x(مفهوم الذات) يسبب المتغير Y (التحصيل).

2. كلما زاد التحصيل (Y) يسبب زيادة فى مفهوم الذات.

3. وجود متغير ثالث مثل المناخ المدرس وهو ما يعرف بمشكلة المتغير الثالث Third variable problem وتوصف أيضًا بالمتغير الوسيط Mediator variable يسبب X , Y ويمكن عرضهم كالآتى:



الشكل (4.12) : ثلاث احتمالات للعلاقة للسببية بين Y,X.

وعمومًا الارتباط لا يعنى السببية:

Correlation Does Not imply causation.

Correlation not proof cause - effect relationship between the two varibles

وادعاء السببية بين خلال الارتباط فيه مخاطرة كبيرة لأن السببية الحقيقية لا يمكن اثباتها لا من خلال تجريب ولكن يمكن القول إن ادعاء السببية من خلال الارتباطات هو الحد الأدنى لادعاء السببية، وهذا الأمر ينطبق على كل الأساليب الإحصائية التى تتحقق من العلاقات السببية من خلال الارتباطات كما فى تحليل الانحدار وتحليل المسار ونموذج المعادلة البنائية.

معامل الارتباط المصحح Adjusted r (radj)

كما نعلم ان معامل الارتباط فى العينة هو تقدير غير متحيز لمعامل الارتباط فى المجتمع rho (ρ) ولكن عندما تكون العينة أو عدد الدرجات أو الملاحظات صغيرة فإنه يعتبر تقدير متحيز لمعامل الارتباط فى المجتمع ولتصحيح معامل الارتباط من صغر حجم العينة قدم Howell (2013) معادلة تصحيحية كالآتى:

$$radj = \sqrt{1 - \frac{(1-r^2)(N-1)}{N-2}}$$

واذا كان حجم العينة كبيرًا فلا نتوقع فروق كبيرة بين r , radj.

فاذا كان $r_{yx} = 0.529$ وحجم العينة $N = 107$ فان:

$$radj = \sqrt{1 - \frac{(1-(0.529)^2\ (106)}{107-2}} = 0.522$$

على ذلك الفروق بسيطة جدًا.

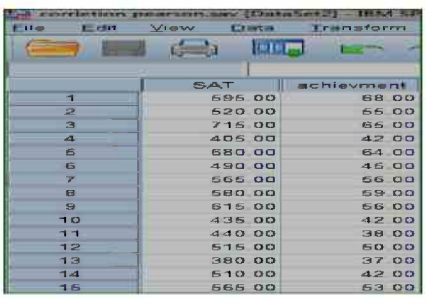
كتابة نتائج معامل الارتباط فى تقرير البحث وفقًا لـ APA

تـم حسـاب معامـل ارتبـاط بيرسـون بـين الأداء الأكـاديمى والاسـتعدادات $r_{(13)}=0.90$, $n=15$, $P<0.01$ ، حيث 13 هـى درجـات الحريـة (2-15). وعلـى ذلـك تـم تحديـد قيمـة معامـل الارتبـاط مقرونًـا بحجـم العينـة ودلالتـه الإحصـائية لاختبـار ذو ذيـل واحـد أو ذو ذيلين، وإذا وجدت متغيرات عديدة يفضل عرضها فى مصفوفة ارتباط كالآتى:

الجدول (4.12) مصفوفة الارتباط بين الدخل والعمر والذكاء.

	الدخل	العمر	الذكاء
الدخل	**1.00**		
العمر	**0.41**	**1.00**	
الذكاء	**027**	**0.02**	**1.00**

$N = 30$، $P^{**}<0.01$ ذو ذيلين، $P^{xx}< 0.05$ ذو ذيلين



تنفيذ معامل ارتباط بيرسون فىSPSS

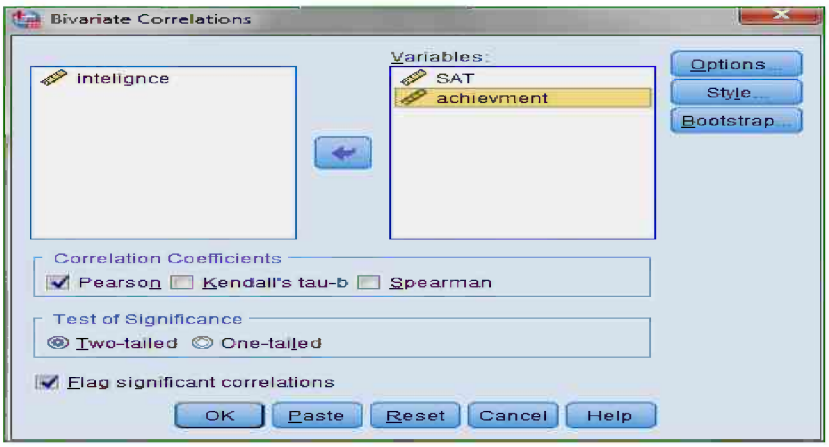
اولًا : إدخال البيانات:1. اضغط Variable view

2. اكتب مسمى المتغيرات تحت عمود Name:الصف الأول: الاستعداد Sat والصف الثانى: الأداء الأكاديمى achievement

3. اضغط data view وادخل المتغيرين فى عمودين. وفيما يلى شاشة البرنامج:

ثانيًا: تنفيذ الأمر: 1.اضغط على Analyze ثم اضغط Correlate ثم اضغط Bivariate

تظهر الشاشة:

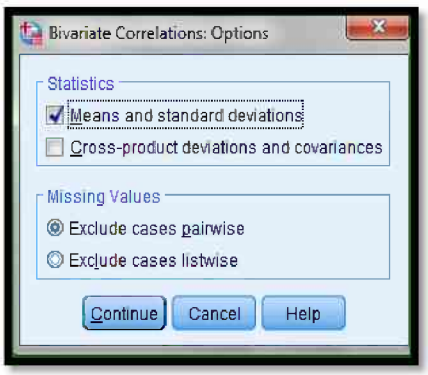


2. اضغط على المتغيرين Achievment , SAT معًا (اضغط ctrl) أو انقل كل متغير على حده ثم اضغط على السهم → لتنقلهما إلى المربع Variables

3. يوجد تحت معاملات الارتباط Correlation Coefficient ثلاث معاملات ارتباط اضغط على Pearson.

4. تحت مربع Test of Significance ثم اضغط على اختيار Two – Tailed وإذا كان فرضك ذو اتجاه واحد علاقة ارتباطية موجبة أو سالبة اضغط على One – tailed

5. تأكد ان الاختيار Flag significance نشطة



6. اضغط على Options يعطى شاشة فرعية كالآتى:

7. اضغط فى مربع Statistics اضغط على اختيار means and standard deviations ليعطى المتوسطات و الانحرافات المعيارية لكل متغير فى التحليل.

8. فى مربع Missingvalues حدد الاستراتيجية التى تتعامل بها مع البيانات المفقودة سواء pairwise أو listwise فاضغط على pairwise

9. اضغط على Continue

10. اضغط على Bootstrap تظهر النافذة الآتية:

11. اضغط Perform bootstrapping :

- حدد عدد العينات التى ترغب فى توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها وليكن 500 لاحظ أنه إذا كان حجم العينة المتاحة كبير يمكن تحديد عدد العينات بـ 1000 مثلًا وهى Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة. ويتم ال
- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01 ويتم الحصول على فترات الثقة المئينية Percentile Confidence intervals من خلال ترتيب فترات الثقة من الاصغر إلى الاكبر فمثلًا فترات الثقة 95% فإنه يعتمد على المئينى percentile 5^{th} و Percentile 95^{th} كنقطتى للفترة. وقد يحدث اختلاف بين فترات الثقة البارامترية التى تقدر من خلال اختبارات الفروض الكلاسيكية الصفرية عن فترات الثقة اللابارامترية المقدرة من خلال البوتستراب وهذا نتيجة لعد الاعتدالية للبيانات.
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيطة.

12. اضغط OK لتنفيذ الامر.

- **ثالثًا المخرج** :اعطى المخرج الآتى: المتوسطات والانحرافات المعيارية:

```
CORRELATIONS
  /VARIABLES=SAT achievment
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.
```

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
SAT	534.0000	96.53275	15
achievment	51.4667	10.10563	15

- مصفوفة الارتباط بين المتغيرين:

Correlations

		SAT	achievment
SAT	Pearson Correlation	1	.903**
	Sig. (2-tailed)		.000
	N	15	15
achievment	Pearson Correlation	.903**	1
	Sig. (2-tailed)	.000	
	N	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

وهى مصفوفة قطرها واحد صحيح هو معامل ارتباط المتغير بنفسه والقيم فوق القطر هى نفسها تحت القطر ولذلك يفضل عرض القيم أسفل القطر وتكونت كل خلية فى المصفوفة من ثلاثة قيم:

- قيمة معامل الارتباط $r = 0.903$
- قيمة p لتحديد الدلالة الإحصائية لمعامل الارتباط و $p = 0.000$ و بما أن $0.00 <$ $(0.01) = \alpha$ ، وعليه فهى علاقة ارتباطية دالة إحصائيًا عند 0.01 وبالتالى فهى دالة عند 0.05.

- حجم العينة هو 15

وعليه توجد علاقة ارتباطية دالة إحصائيًا بين التحصيل والاستعدادات، وبالتالى نرفض H_0 واعطى البرنامج العلامة (**) تعنى دالة عند 0.01 كما اوضحها البرنامج أسفل الجدول وذلك لاختيار ذو ذيلين و أحيانًا يضع (*) يعنى دالة عند مستوى دالة إحصائية 0.05.

ويرى Green & Salkind (2014) إذا تم تقدير معامل الارتباط بين ثلاثة متغيرات فى تحليل واحد فيجب تصحيح مستوى الدلالة الإحصائية باستخدام تصحيح Boneferroni حيث يتطلب قسمة مستوى الدلالة الإحصائية (0.05 مثلًا) على عدد معاملات الارتباطات فاذا وجدت مصفوفة تتضمن خمسة متغيرات وعليه يوجد 10 معاملات الارتباط وبالتالى تصحح مستوى الدلالة الإحصائية كالآتى:

$$\alpha = \frac{0.05}{10} = 0.005$$

وبالتالى يجب مقارنة القيم P بـ0.005 .

القرار فى ضوء تطبيق استراتيجية Bootstrap:

Bootstrap Specifications

Sampling Method	Simple
Number of Samples	1000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

				SAT	achievment
SAT	Pearson Correlation			1	.903**
	Sig. (2-tailed)				.000
	N			15	15
	Bootstrap[b]	Bias		0	.001
		Std. Error		0	.042
		95% Confidence Interval	Lower	1	.811
			Upper	1	.972

اعطى الخطأ المعيارى 0.045 اما فترات الثقة للتحصيل (0.97 ,0.811) وهاتين القيمتين لا يقع بينها قيمة الصفر الموضوعة فى الفرض الصفرى بالتالى نرفض الفرض الصفرى وهذا يعنى وجود علاقة ارتباطية بين التحصيل والاستعدادات.

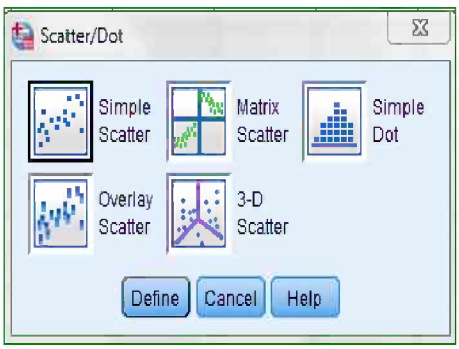
التحليل البديل لمعامل ارتباط بيرسون

إذا لـم تتحقـق مسـلمات بيرسـون مثـل الاعتداليـة والخطيـة فإنـه يمكـن باسـتخدام معامـل ارتبـاط سـبيرمان أو كنـدال تـاو وذلـك بالتعامـل مـع رتـب الـدرجات وبـذلك تصـبح المتغيـرات رتبية.

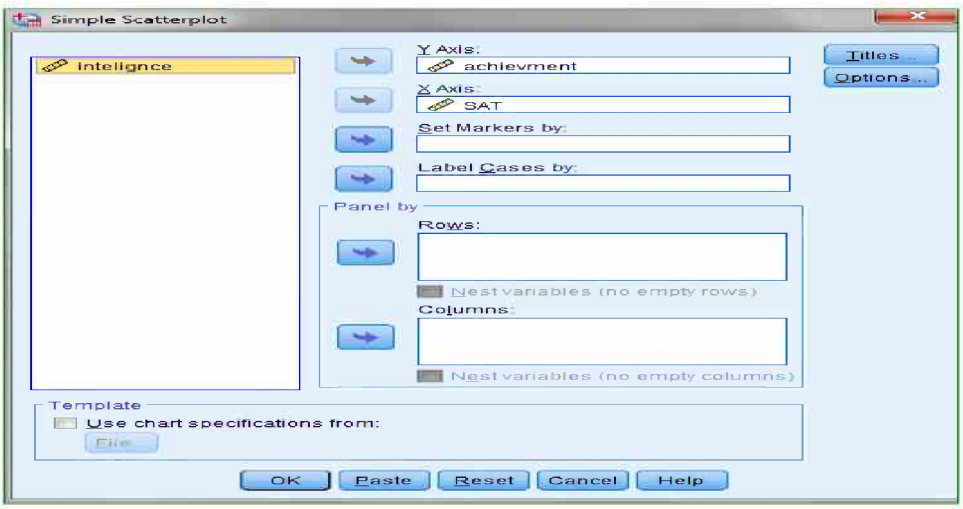
استخدام الرسم البيانى لعرض النتائج

لعرض شكل الانتشار للعلاقة بين المتغيرين اتبع الخطوات الآتية:

1. اضـغط علـى Graphs ثـم اختـار legacy Dialogs ثم اضـغط علـى اختيـار Scatter dot تظهر الشاشة:

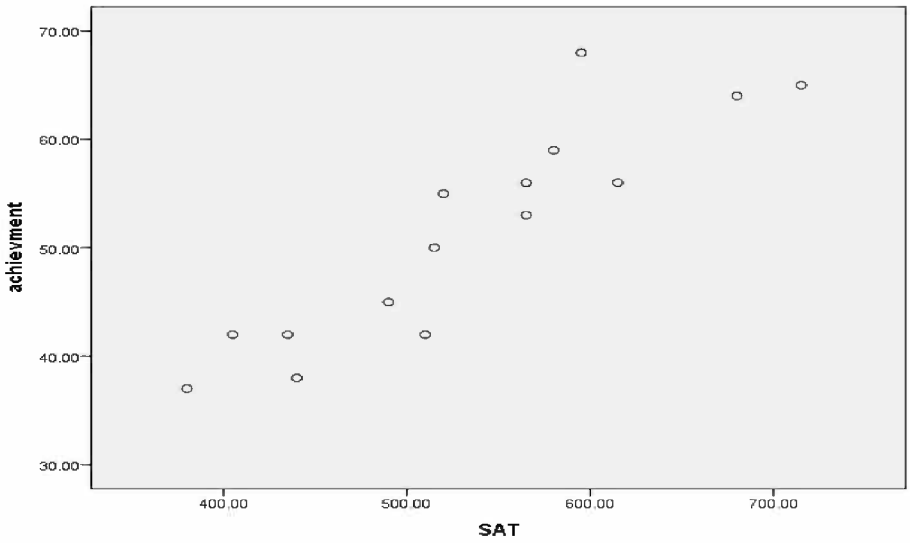


2. اضـغط علـى simple Scatter ثم اضغط على define تظهر الشاشة الآتية:



3. انقل SAT إلى مربع X Axis وachievement إلى مربع Y Axis

4. اضغط OK يظهر المخرج الآتى:



معامل الارتباط الجزئي

Partial correlation coefficient

يهدف إلى تقدير لمعامل الارتباط بين متغيرين بعد ضبط تأثير متغير ثالث يستخدم للكشف عن التضخم أو التقلص الخادع لمعامل ارتباط بين متغيرين نتيجة تدخل متغير ثالث (z) وتقدر كالآتى:

$$r_{xy.z} = \frac{rxy - (rxz)(ryz)}{\sqrt{(1-rxy^2)(1-ryz^2)}}$$

- $r_{xy.z}$ ، معامل الارتباط الجزئى بين y,x
- r_{yx} معامل ارتباط بين y, x
- r_{xz} معامل ارتباط بين z,x ، r_{yz} معامل ارتباط بين z,y

فإذا كان: $r_{xy} = 0.91$, $r_{xz} = 0.85$, $r_{yz} = 0.97$

فان: $r_{xy.z} = \frac{0.91-(0.85)(0.97)}{\sqrt{1-(0.85))^2(1-(097))^2}} = 0.65$

لاحظ أن معامل الارتباط بين y,x= 0.91 ومعامل الارتباط بين y,x و بعد ضبط اثر المتغير الثالث Z هو 0.65 بمعنى حدث تقلص لمعامل الارتباط بدرجة كبيرة، وان التباين المفسر قبل الضبط 82.8% والتباين المفسر بعد الضبط 42.2% ، وهذا التدخل سبب حدوث معامل ارتباط زائف أو خادع و يتم ضبطه إحصائيًا باستخدام

معامل الارتباط الجزئى وتجريبيًا من خلال الضبط التجريبى وهذا يدعونا إلى الحذر عند تفسير معامل الارتباط أو الحذر عند استخدامه للاستدلال السببى فى الأساليب الإحصائية المتدرجة.

ويرى Field (2009) إذا افترضنا ان التحصيل يرتبط سلبيًا مع قلق الامتحان وايجابيًا مع وقت الاستذكار فإن وقت المراجعة للامتحان ترتبط ارتباط سالب مع قلق الامتحان وعلى ذلك فإن الوقت يرتبط مع كلًا من التحصيل والقلق. فاذا افترضنا ان القلق فسر 19.4% من تباين التحصيل والوقت فسر 15.7% من تباين التحصيل و الوقت فسر 50.2% من تباين القلق ، و الوقت و القلق يسهموا معًا فى تفسير 19.4% (المحسوبة على القلق) و على ذلك يهدف معامل الارتباط الجزئى حساب العلاقة بين القلق و التحصيل نقية خالصة مع عزل اثر الوقت على القلق والتحصيل

فالقلق هذا متغير وسيط ينقل أثر الوقت إلى التحصيل وكذلك يؤثر عليه الوقت.

مسلمات معامل الارتباط الجزئى

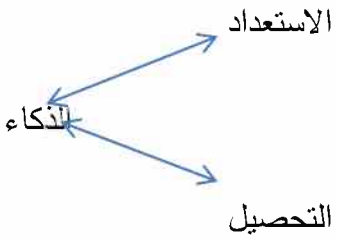
1. توافر الاعتدالية المتدرجة للمتغيرات وإذا تحققت الاعتدالية من اتحاد المتغيرات معًافان هذا يفترض تحقق الاعتدالية لكل متغير حدة.

2. العينة العشوائية واستقلال درجات كل حالة عن درجات الحالات الأخرى.

حجم التأثير لمعامل الارتباط الجزئي

هو نفسه مربع معامل الارتباط الجزئى كما هو الحال مع معامل ارتباط بيرسون.

قضية: فى مثال معامل الارتباط بيرسون السابق نفترض تم إضافة متغير ثالث الذكاء:



وعليه هل توجد علاقة بين الاستعداد والتحصيل بعد عزل اثر الذكاء؟

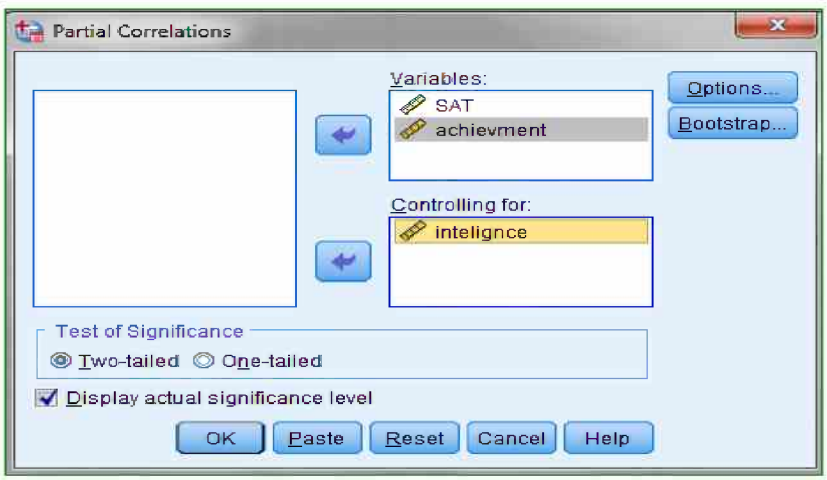
تنفيذ معامل الارتباط الجزئى فىSPSS

اولًا: إدخال البيانات: 1. اضغط Variable view

2. اكتب مسمى المتغيرات تحت عمود Name كالآتى: الصف الأول: الاستعداد SAt، الصف الثانى : التحصيل achievement، الصف الثالث Intelligence

3.اضغط dataview وأدخل المتغيرات فى ثلاثة أعمدة.

ثانيًا: تنفيذ الأمر :1.اضغط على Anlayze ثم Correlate ثم اضغط على Partial تظهر الشاشة:

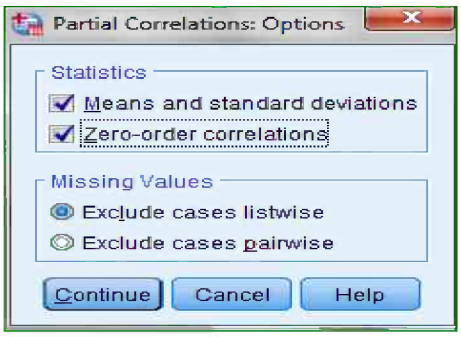


2. انقل متغيرات التحصيل والاستعداد معًا أو كلاهما إلى مربع Variables

3. اضغط على المتغير المراد ضبطه وهو intelligence ثم نقله إلى مربع Controlling for

4. تأكد ان الاختيار Two- tailed إذا كان الفرض ذا اتجاهين أو One- tailed إذا كان الفرض ذا اتجاه واحد، وذا اتجاهين نشطة من غير الضغط عليها لأنها default البرنامج.

5. تأكد أن الاختبار Display actual significance نشطة.



6. اضغط على Options تظهر الشاشة الآتية:

7. اضغط فى مربع Statistics اضغط على اختيار Means and standard deviations ليعطى المتوسطات والانحرافات المعيارية لكل متغير فى التحليل واضغط zero- order correlation وهى معامل الارتباط بين SAT, Achievment

8. اضغط Continue ثم OK

ثالثًا: تفسير المخرج: 1. يعطى المخرج الإحصاء الوصفى لكل المتغيرات كالآتى:

Descriptive Statistics

	Mean	Std. Deviation	N
SAT	534.0000	96.53275	15
achievment	51.4667	10.10563	15
intelignce	118.2000	11.32128	15

2.اعطى البرنامج مصفوفة الارتباط و هى مكونة من جزئين كالآتى:

Correlations

Control Variables			SAT	achievment	intelligence
-none-[a]	SAT	Correlation	1.000	.903	.240
		Significance (2-tailed)	.	.000	.388
		df	0	13	13
	achievment	Correlation	.903	1.000	.188
		Significance (2-tailed)	.000	.	.503
		df	13	0	13
	intelligence	Correlation	.240	.188	1.000
		Significance (2-tailed)	.388	.503	.
		df	13	13	0
intelligence	SAT	Correlation	1.000	.900	
		Significance (2-tailed)	.	.000	
		df	0	12	
	achievment	Correlation	.900	1.000	
		Significance (2-tailed)	.000	.	
		df	12	0	

a. Cells contain zero-order (Pearson) correlations.

أ. **الجزء الأعلى:** يعرض مصفوفة معامل ارتباط بيرسون بدون ضبط بين المتغيرات واتضح ان العلاقة بين الاستعداد والتحصيل هى 0.903 وهى داله عند 0.01.

ب. **الجـزء الأسـفل:** يعـرض معامـل الارتبـاط الجزئـى بـين الاسـتعدادات والتحصـيل بعـد عـزل أثـر الـذكاء علـى كـلٍّ مـن الاسـتعدادات والتحصـيل اتضـح ان قيمتـه 0.900، أى أنــه انخفــض مــن 0.903 إلــى0.900 وأيضًــا معامــل الارتبــاط الجزئــى دال إحصــائيًا عنـد 0.05 وكـذلك عنـد 0.01 وإذا اعتمـد الباحـث علـى معامـل ارتبـاط بيرسـون بـين المتغيـرين فـإن التبـاين المفسـر فـى التحصـيل جـراء الاسـتعدادات تقريبًـا 0.810 بينمـا التبـاين المفسـر للتحصـيل جـراء الاسـتعدادات بعـد اسـتبعاد اثـر الـذكاء هـو 0.815 تقريبًا.

معامل الارتباط شبه الجزئى Part correlation

هـو حـذف تـأثير متغيـرات خارجيـة تـؤثر علـى العلاقـة بـين متغيـرين مـا مـن خـلال حـذف تأثيرهـا علـى أحـد المتغيـرين ولـيس علـى كلاهمـا كمـا فـى حالـة Partial. والصـيغة لتقدير معامل الارتباط بينY و X1 مع حذف تأثير X2 على X1 فقط هى كالآتى:

$$r_y(x_1, x_2) = \frac{r_{yx_1} - r_{yx_1} r_{x_1x_2}}{\sqrt{1 - r_{x_1x_2}}}$$

وهـذه المعادلـة تعطـى التبـاين الكلـى بـين Y و X1 منقحًـا مـن التـداخل المشـترك بـين X1 و X_2 ولـيس مـن تـداخلX1 مـع Y.وعلـى ذلـك فـإن العلاقـة بـينX_1 , Y مـع عـزل تـأثيرX_2 علىX1 فقط هـى (x1, x2) ry بينمـا معامـل الارتبـاط الجزئـى هـو عـزل تـأثير X_2 علـى كـل مـن Y و X1 (Kline, 2016)، وعلـى ذلـك عنـدما تـرتبط المنبئـات ببعضـها فـإن أوزان بيتـا والارتباطـات الجزئيـة هـى البـديل لمعرفـة القـوة التفسـيرية لكـل متغير منبئ مع ضبط المتغيرات الأخرى.

الفصل الرابع عشر

معامل الارتباط الثنائى الاصيل

Point bi-serial Correlation (r_{pb})

معامل ارتباط بيرسون (r) يطبق فى المواقف التى تكون فيها العلاقة بين المتغيرين خطية، أى أن المتغيرين يقاسوا على مقياس متصل ووجود درجة كبيرة من الاعتدالية وكذلك تجانس التباينات. وتوجد مجموعة مختلفة من مقاييس العلاقة أو الارتباط، فعلى سبيل المثال يوجد موقف بحثى فيه متغيرين من مستوى قياس رتبى أو من مستوى قياس اسمى تصنيفى وعلى ذلك توجد معاملات ارتباطات تناسب هذه المواقف غير معامل ارتباط بيرسون. وتوجد بدائل لمعامل ارتباط بيرسون فى هذه المواقف وتقدر هذه البدائل كما لو استخدمنا صيغة معامل ارتباط بيرسون وبعضها بصيغ اخرى.

واحيانًا تقاس المتغيرات فى صورة ثنائية Dichotomy مثل ذكر-أنثى، ناجح-راسب، مجموعة ضابطة-مجموعة تجريبية فهذه المتغيرات لا تقاس رقميًا ولذلك فهى غير متصلة وليس لها بناء تحتى متصل أو توزيع متصل مثل المنحنى الاعتدالى ولو استخدمنا معامل ارتباط بيرسون(r) لقياس العلاقة بين هذه النوعية من المتغيرات فإنه توجد مشاكل عديدة نتيجة عدم توافر مسلمة الاتصالية.

ويستخدم معامل الارتباط الثنائى الأصيل $\mathbf{r_{pb}}$ عندما يوجد متغيران أحدهما تصنيفى ثنائى (اسمى) والآخر متصل (فترى على الأقل) وقد يكون توزيعه اعتدالى أو شبه اعتدالى وعلى ذلك عندما تستخدم صيغة معامل ارتباط بيرسون فى هذا الموقف ويسمى معامل الارتباط $\mathbf{r_{pb}}$ وعلى ذلك فإن $\mathbf{r=r_{pb}}$، وصيغة حساب $\mathbf{r_{pb}}$ مشتقة من معامل ارتباط بيرسون وهى كالآتى:

$$r_{pb}=\left(\frac{\overline{X}_{y1}-\overline{X}_{y2}}{SY}\right)\left(\sqrt{Pq}\right)$$

- $\overline{X}_{y1}$ متوسط درجات المتغير التابع المتصل للتصنيف الأول للمتغير التصنيفى.
- $\overline{X}_{y2}$ متوسط درجات المتغير التابع المتصل للتصنيف الثانى للمتغير التصنيفى.
- SY الانحراف المعيارى للمتغير التابع.
- q، P نسبة الدرجات فى كل مستوى من المتغير التصنيفى.

اختبارات الفروض لقضية بحثية: أراد باحث دراسة العلاقة بين الجنس والاتجاه نحو الإنترنت وكانت الدرجات لـ 12 طالب كالآتى:

الجنس	التحصيل
1	**23**
1	**9**
1	**12**
1	**19**
1	**24**
2	**32**
2	**10**
2	**8**
2	**20**
2	**12**
2	**24**
2	**34**

حيث 1 ذكور ، 2 إناث

الخطوات البحثية

1. سؤال البحث: هل توجد علاقة بين الجنس والتجاه نحو الانترنت؟

2. فرض البحث: توجد علاقة بين الجنس والاتجاه نحو الانترنت.

3. المتغيرات: الجنس: اسمي- مستقل بمستويين، الاتجاه نحو الانترنت: تابع- فتري- متصل.

4. منهج البحث: منهج ارتباطى.

5. النموذج الإحصائى: بسيط- إحصاء بارامترى والاختبار الإحصائى معامل الارتباط الثنائى الاصيل r_{pb}.

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية:

$$H0: \rho = 0$$

$$H_A: \rho \neq 0$$

ρ معامل الارتباط بين المتغيرين فى المجتمع

2. الاختبار الإحصائى: معامل ارتباط بيرسون للمتغير التصنيفى والآخر المتصل وهو معامل الارتباط r_{pb} ويقدر كالآتى:

$$r_{pb} = \left(\frac{\bar{X}_{y1} - \bar{X}_{y2}}{SY}\right)\left(\left(\sqrt{Pq}\right)\right)$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: تبنى الباحث $\alpha = 0.05$ وبالنسبة لقاعدة القرار والدالة الإحصائية لهذ الاختبار تقدر من خلال اختبار T لعينات مستقلة:

$$T^2 = \frac{r_{pb}^2}{\left(1 - r_{pb}^2\right)/df}$$

$$T = \frac{r_{pb}\sqrt{N-2}}{\sqrt{1 - r_{pb}^2}}$$

$$df = N - 2 = 12 - 2 = 10$$

4. الحسابات:

ا- تكويد المتغير التصنيفى حيث الذكور= 1 ، الإناث = 2 وعليه يوجد 5 ذكور و 7 إناث.

وتكون نسبة الذكور P: $P = \frac{5}{12} = 0.42$

ونسبة الإناث q: $q = \frac{7}{12} = 0.58$

لاحظ أن $p + q = 1$

ب- حساب متوسط المتغير التابع Y لكل مستوى فمتوسط درجات الذكور $\bar{X}_{y1}$ كالآتى:

$$\overline{X}_{y1}=\frac{23+19+\cdots\quad+19}{5}=17$$

ومتوسط درجات الإناث: $\overline{X}_{y2}$

$$\overline{X}_{y2}=\frac{32+10+\cdots\quad+34}{7}=20$$

جـ- حساب الانحراف المعيارى للمتغير التابع YSY = 9.056:

د- حساب معامل الارتباط:

$$r_{pb}=\left(\frac{17-20}{9.056}\right)\left(\sqrt{0.42x0.58}\right)=-0.163$$

وهذا معامل ارتباط ضعيف وسالب بمعنى اتجاه العلاقة ليس له معنى لأنه من الصعب القول كلما انخفض الجنس يزداد الاتجاه نحو الإنترنت ولكن يمكن التفسير من خلال النظر إلى قيمة المتوسط للإناث وللذكور والملاحظ أن متوسط الاتجاه نحو الانترنت للإناث أعلى من متوسط الذكور.

5.القرار والتفسير: القرار بتحويل r_{pb} الى قيمة اختبار T المناظرة كالآتى:

$$T^2=\frac{r_{pb}{}^2}{\left(1-r_{pb}{}^2\right)/_{df}}=\frac{0.163}{(1-0.163)/_{10}}=0.272$$

وعليه فالجذر التربيعى هو : $T=\sqrt{0.272}=0.522$

ومن ثم فإن قيمة T المحسوبة = 0.522 وبالكشف فى جداول T لاختبار ذو ذيلين ومستوى دلالة إحصائية 0.05 و df = 10 فإن T الجدولية = 2.228، وعلى ذلك فان: T **المحسوبة < T الجدولية (2.228)**، وعليه نقبل الفرض الصفرى القائل الجنس والتجاه نحو الانترنت غير مرتبطين.

6. **حجم التأثير:** يقدر حجم التأثير من خلال:

- مؤشر إيتا تربيع η^2 كالآتى:

$$\eta^2=\frac{T^2}{T^2+df}=\frac{(0.522)^2}{(0.522)^2+10}=0.0272$$

ومربع قيمة r_{pb} وهى:

$$r^2= (r_{pb})^2=(-0.163)^2=0.0272$$

والمؤشـرين اعطـوا نفـس القيمـة وهـذا يفيـد ان الجـنس فسـر 2.72% مـن تبـاين الاتجـاه نحو الانترنت وهذا حجم تأثير ضعيف جدًا.

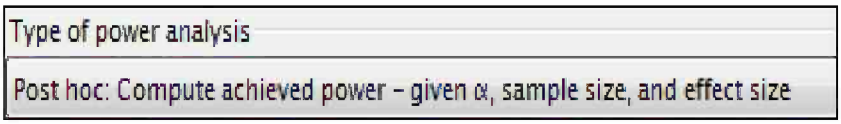
- مؤشر d لـ Cohen(1988) :

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

7. القوة الإحصائية: لحساب القوة الإحصائية باستخدام برنامج **G-Power** اتبع الآتى:

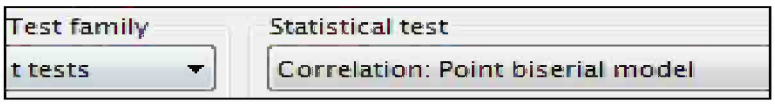
1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختار:

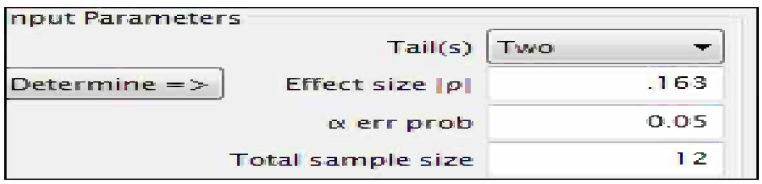


3. أسفلTest family اختار t- tests

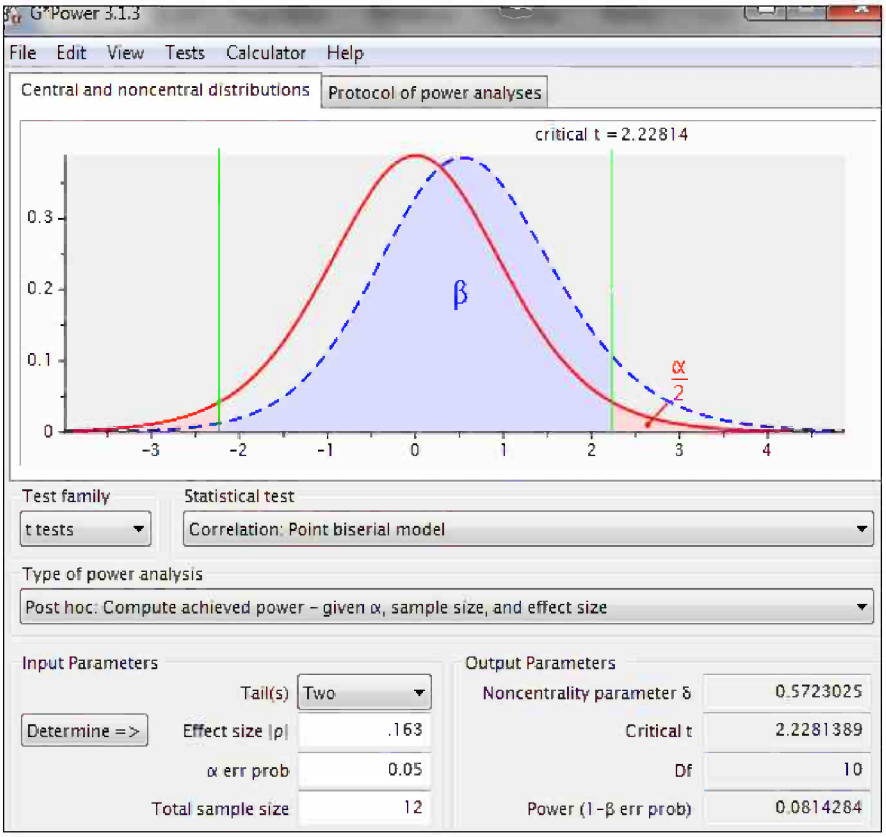
4. تحت Statistcal test اختار:



5. ادخل معالم المدخلات كالآتى:



6. اضغط calculated يظهر المخرج الآتى:



يتضح أن القوة الإحصائية= 0.0814 وهذا مستوى ضعيف جدًا. وهذا انعكس على عدم قدرة الاختبار على رفض الفرض الصفرى ولذلك عدم الدلالة الإحصائية ليست نتيجة لأنها حقيقية فى المجتمع بل قد تكون نتيجة عدم قدرة الاختبار على رفض الفرض الصفرى وهذا بدورة يضخم الخطأ من النوع الثانى (عدم دلالة على مستوى العينة بينما هى موجودة فى المجتمع).

كتابة نتائج r_{pb} فى تقارير البحوث وفقًا لـ APA

تم تقدير معامل الارتباط r_{pb} بين الجنس والاتجاه نحو الانترنت:

$$r_{pb}= -0.163,\ p > 0.05,\ \eta^2,\ r^2=0.0272$$

تنفيذ معامل ارتباط r_{pb} فى برنامج SPSS

أولًا: إدخال البيانات: تم تعريف المتغيرين من خلال الضغط على variable view أسفل الشاشة تظهر الشاشة التى يحدد بها المتغيرات اكتب مسمى المتغيرين تحت

عمـــود Name ولـــيكن attitude و sex ثـــم اضـــغط علـــى DataView وقـــم بإدخـــال البيانات.

ثانيًا تنفيذ الأمر:1. اضغط على Analyze→ Bivariate→Correlate

ونفس خطوات معامل ارتباط بيرسون.

الفصل الخامس عشر

الانحدار البسيط

Simple Regression

تناولنا سابقًا كيف يمكن تقدير العلاقة الارتباطية بين متغيرين وهذه الارتباطات فى غاية الأهمية ولكننا فى هذا الفصل نعتمد على الارتباط للذهاب خطوة ابعد وهو التنبؤ بمتغير تابع من متغير أو متغيرات أخرى مستقلة. وتحليل الانحدار هى مجموعة من الأساليب الإحصائية التى تسمح بتقدير العلاقة بين المتغير التابع Dependent variable (DV) ومجموعة من المتغيرات المستقلة Independent variables (IV)، فمثلًا القدرة القرائية فى المرحلة الابتدائية مرتبطة بعدة متغيرات مستقلة مثل الذكاء والنمو المعرفى والعمر وغيرها. وكما نعلم أن الانحدار والارتباط متداخلين ولكن الانحدار يستخدم غالبًا عندما يكون هدف الباحث هو التنبؤ Prediction فلا تنبؤ بدون ارتباط والارتباط متطلب أساسى للانحدار ولكن الارتباط يستخدم عندما يكون هدف الباحث دراسة العلاقة بين متغيرين.

والانحدار أسلوب إحصائى يستخدم بصورة متزايدة فى كافة التخصصات سواء العلوم النفسية والسلوكية والاجتماعية وكذلك العلوم الطبيعية والطبية وغيرها، ويستخدم للتنبؤ بمتغير ناتج Outcome variable (Y) أو تابع من متغير منبئى Predictor أو مستقل(X) وفى هذه الحالة يسمى انحدار بسيط Simple Regression أو من متغيرات منبئة عديدة وفى هذه الحالة يسمى انحدار متعدد Multiple Regression.

أهداف الانحدار:

يستخدم الانحدار لهدفين رئيسين هما التنبؤ والتفسير ويمكن أن يستخدم مصطلح التفسير Explanation مع الانحدار لتحديد التباين المفسر فى التحصيل (المتغير التابع) من خلال تباين الدافعية (المتغير المستقل) ويمكن أن يستخدم فى التنبؤ Prediction وهو استخدام الدافعية للتنبؤ بدرجة التحصيل أى بناء معادلة تنبؤ.

والفرق بين التفسير والتنبؤ كما يشير keith (2014) هو أن التفسير ضرورى للتنبؤ فلو أنك فسرت ظاهرة فإنك تستطيع التنبؤ بحدوثها وعلى الجانب الآخر فالتنبؤ ليس

ضـروريًا للتفسـير وعمومًـا فـى أبحـاث العلـوم الإنسـانية والسـلوكية يهـتم البـاحثين بتفسـير لماذا تحدث الظاهرة أكثر من التنبؤ بحدوثها.

الانحدار الخطى البسيط

يهـدف إلـى دراسـة تـأثير متغيـر مسـتقل (X) علـى متغيـر تـابع (Y) والباحـث يحـاول أن يفهـم كيـف ان المتغيـر (X) يـؤدى إلـى التغيـر فـى (Y)، بكلمـات أخـرى الهـدف هـو تحديـد كيـف والـى أى درجـة أن التبـاين أو التغيـر فـى سـلوك المتغيـر التـابع يعتمـد علـى المتغيـر المسـتقل. والانحـدار البسـيط أداة إحصـائية تسـمح لنـا بـالتنبؤ بدرجـة فـرد علـى المتغيـر التـابع مـن درجتـه علـى المتغيـر المسـتقل واحـد ومثـل معامـل الارتبـاط لبيرسـون، فـإن الانحـدار البسـيط يسـتخدم مـع بيانـات متغيـرين ويجـب أن تكـون العلاقـة بينهمـا خطية(عامر، 2007).

علـى سـبيل المثـال أراد باحـث تحديـد تـأثير سـاعات المـذاكرة (X) علـى التحصـيل(Y) بكـل تأكيـد الأداء علـى التحصـيل يتـأثر بعوامـل أخـرى غيـر (X) ومـن غيـر المحتمـل أن كـل الأفـراد الـذين يتعرضـون لـنفس مسـتوى التغيـر فـى(X) سـوف يظهـروا نفـس التغيـر المماثـل علـى (Y) لوجـود متغيـرات اخـرى تـؤثر علـى المتغيـر التـابع Y وعنـدما تكـون متوسـطات (Y) للمسـتويات المختلفـة لـ (X) تتغيـر مـن فـرد إلـى آخـر وتقـع علـى خـط مسـتقيم فإنـه يوجـد انحـدار خطـى بسـيط لـ Y علـى X وتكـون معادلـة الانحـدار فـى المجتمع:

$$Yi = \alpha + \beta\ Xi + \varepsilon i$$

- Y درجة الفرد i فى المتغير التابع
- α (ألفا) متوسط المجتمع عندما تكون $x = 0$ أو ثابت Y
- (β)(بيتا) معامل الانحدار فى المجتمع أو ميل خط الانحدار
- εi (إيبسلون) قيمة الخطأ أو الخطأ العشوائى Random Disturbance للفردi
- Xi درجة المتغير المستقل لفرد i

ومعامـل الانحـدار β يشـير إلـى قيمـة تـأثير المتغيـر المسـتقل فـى المتغيـر التـابع وأى وحـدة تغيـر فـى المتغيـر المسـتقل X يقابلهـا تغيـر متوقـع مسـاوى لحجـم β فـى المتغيـر Y وعلـى ذلك درجة أى فرد على المتغير التابع تتكون من جزئين:

1. جزء $\alpha + \beta X$: وهو جزء من الدرجة Y للفرد الذى يتعرض لمستوى معين من X

2. جزء الخطأ العشوائى εi لكل فرد

والمعادلة السابقة يعبر عنها فى العينة ضوء المعالم الآتية:

$$Y = a + b\,x + e$$

حيث a هى تقدير لـ α و b هى مقدر لـ β و e هى مقدر لـ ε

ومعادلة الانحدار فى ضوء المتغيرات المقاسة (الدرجات الخام) يعبر عنها:

$$\hat{Y} = a + b\,X$$

مع عدم وضع القيمة e ولكن يتم وضعها إذا تعاملنا مع الانحدار فى ضوء نمذجة المعادلة البنائية حيث تأخذ أخطاء القياس فى الحسبان.

ويتم تقدير الثوابت كالآتى:

$$b = \frac{\sum XY}{\sum X^2}$$

$$a = \bar{Y} + b\bar{X}$$

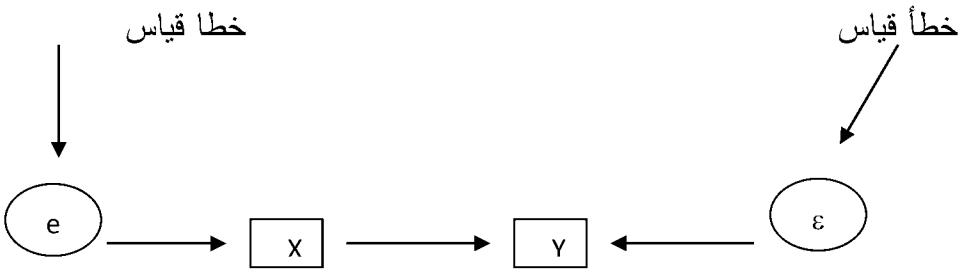
- $\sum xy$ هو مجموع حاصل ضرب Y , X
- $\sum X^2$ مجموع مربع قيم X
- $\bar{X}$متوسط X و $\bar{Y}$ متوسط Y

مسلمات تحليل الانحدار البسيط (Pedhazure, 1997)

لكل أسلوب إحصائى شروط ومسلمات وعدم تحقق هذه المسلمات يؤدى إلى تحيزات خطيرة واستنتاجات خاطئة وتحليل للبيانات ليس له معنى وهذه المسلمات هى:

- القيم المتطرفة: حيث وجودها يعظم اختبارات الدلالة الإحصائية.
- الخطية: العلاقة بين X و Y وهى علاقة خطية بمعنى انحدار Y على X يفترض أن يكون خطى.
- يفترض أن المتغير X يقاس بدون أخطاء قياس.
- يفترض أن المتغير المستقل مثبت variable Fixed فعندما يريد إعادة التجربة مثلًا فإنه تستخدم نفس قيم X ففى المثال السابق يمكن أن يكون لدى الباحث نفس القيم من عدد الساعات لإعادة التجربة.

- المتغير التابع والمستقل اعتدالى التوزيع فى المجتمع وعلى ذلك إذا كان المتغير X و Y اعتدالى التوزيع فمن المتوقع ان تكون العلاقة بينهما خطية. وعدم الاعتدالية تقلل من القوة الإحصائية للاختبار وتكون قيمة P غير صادقة.
- اختيار العينة عشوائيًا من المجتمع ودرجات أى متغير مستقل عن درجات المتغير الآخر.
- وعكس المتغير X فإن المتغير التابع Y هو عشوائى بمعين لكل قيمة لـ X فإنه يوجد مدى محتمل من القيم لـY وتذكر الدرجة Yi يتكون من جزء مثبت (α + BX) و جزء الخطأ عشوائى (ε)



وفيما يلى مسلمات خاصة بالأخطاء القياس وهى:

- متوسط أخطاء كل قياس Yi عبر العديد من التطبيقات أو المرات هى صفر.
- الأخطاء المرتبطة بالقياس أو الملاحظة Yi لا ترتبط مع الأخطاء المرتبطة بأى قياسات أخرى.
- تباين الأخطاء لكل قيم X ثابت بمعنى التباين للأخطاء هو نفسه على كل مستويات X وهذه الخاصية يشار إليها بـ Homoscedas`ticity وسبق التحقق منها فى المثال وعندما تختلف التباين للأخطاء لقيم X تسمى Heteroscedasticity
- لا يفترض أن ترتبط الأخطاء بالمتغير المستقل X
- لاختبارات الدلالة الإحصائية (F ,T) يتطلب أن يكون توزيع الأخطاء اعتدالى.

وهذه المسلمات لا بد من توافرها للحصول على أفضل تقدير غير متحيز لمعامل بالانحدار ولكنها تظهر بوضوح فى الانحدار المتعدد. حقيقة هذه المسلمات فى غاية

الأهمية بالإضافة إلى مسلمات معامل ارتباط بيرسون وهى الاعتدالية والخطية والثبات وغيرها.

اختبارات الفروض لقضية بحثية مأخوذة من: (Pedhazure, 1997)

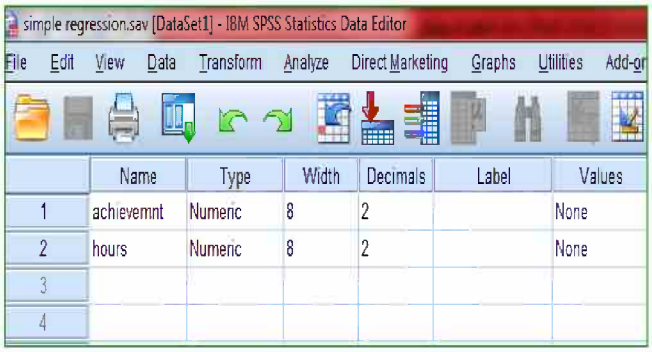
بفرض أن باحث دراسة تأثير عدد الساعات ذالتى يقضيها الفرد على الانترنت (X) على التحصيل فى الرياضيات لـ 20 طالب وكانت البيانات كالآتى:

X	Y
1	3
1	5
1	6
1	9
2	4
2	6
2	7
2	10
3	4
3	6
3	8
3	10
4	5
4	7
4	9
4	12
5	7
5	10
5	12
5	6

وطرح الباحث السؤال الآتى: ما هى المعادلة الخطية التى تسمح بالتنبؤ بالتحصيل من عدد الساعات؟ أو ما الإسهام النسبى لعدد الساعات للتنبؤ بالتحصيل؟

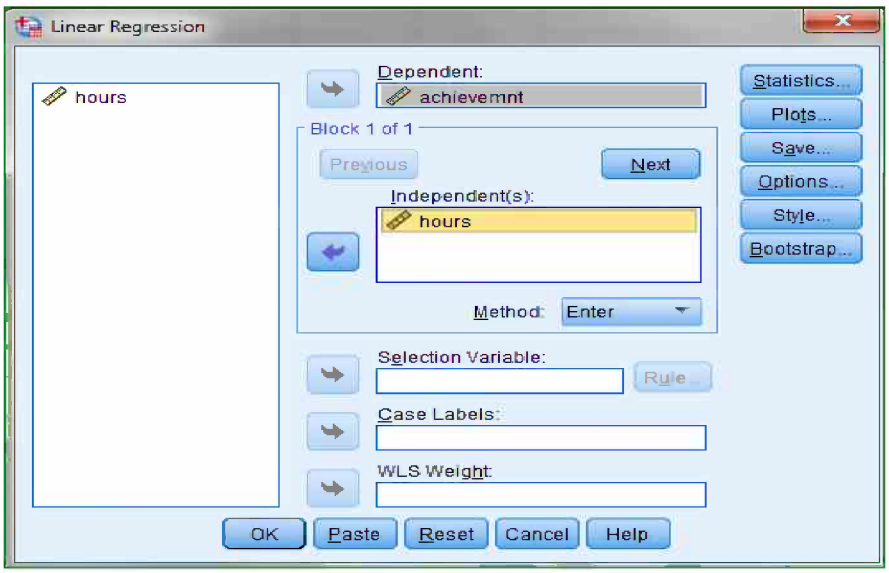
تنفيذ الانحدار البسيط فىSPSS

أولًا: إدخال المتغيرات: 1. بعد فتح البرنامج اضغط علىVariable view تظهر الشاشة الفرعية:

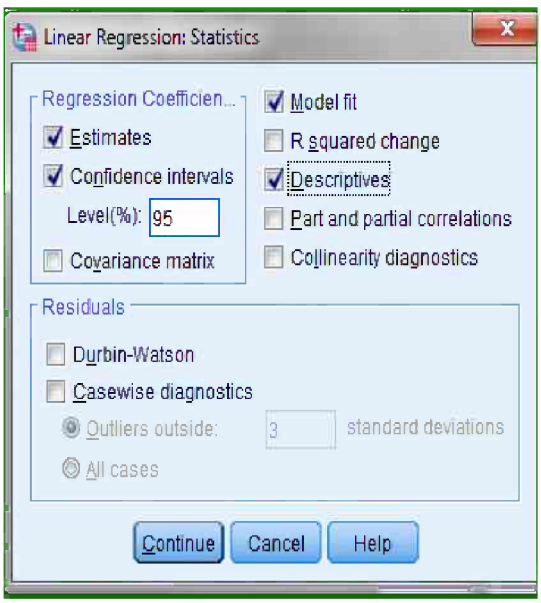


تحت عمود Name اكتب التحصيل achievement وكذلك متغير عدد الساعات hours ثم اضغط Data view لترجع إلى شاشة إدخال البيانات

ثانيًا: تنفيذ الأمر: 1. اضغط على Analyze→Regression→Linear تظهر الشاشة:



2. انقل achievement↳ إلى مربع Dependent ثم انقل hours إلى مربع↳Independent

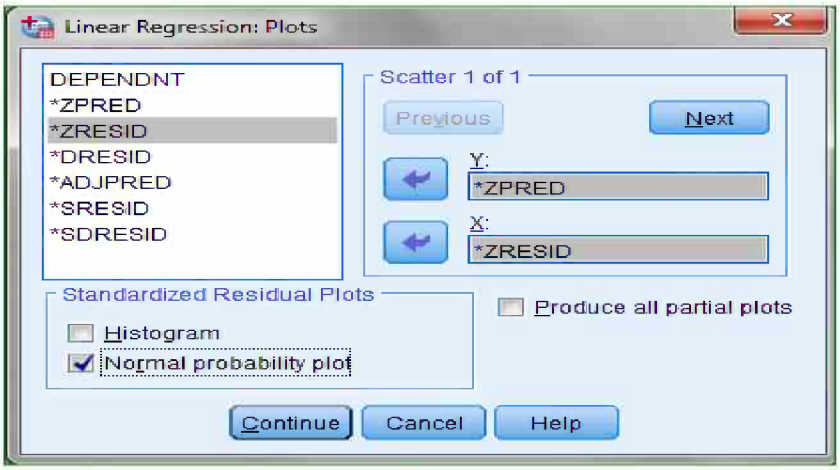


3. اضغط على اختيار Statistics على يمين الشاشة تظهر الشاشة:

اضغط على خيارات Estimates و Confidence intervals و Model fit و Descriptive

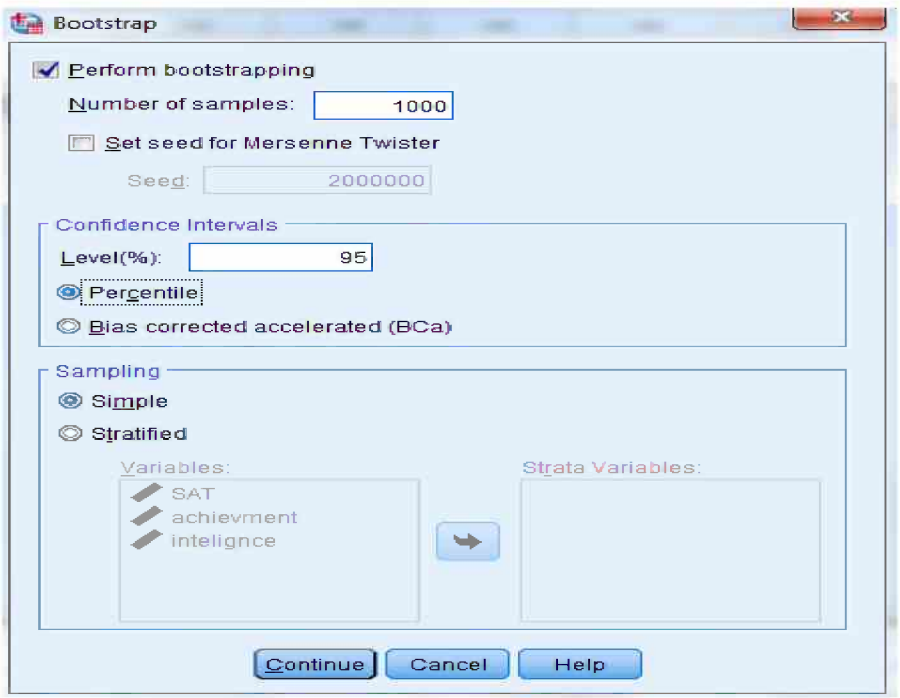
4. اضغط على Continue

5. اضغط على الاختيار Plots تظهر النافذة الآتية:



6. اضغط على ZPRED* وانقله إلى ↳ مربع Y واضغط على ZPRESID وانقله إلى ↳ المربع X، واضغط على Normal probability plot

7. اضغط على Bootstrap تظهر النافذة الآتية:



8. اضغط Perform bootstrapping:

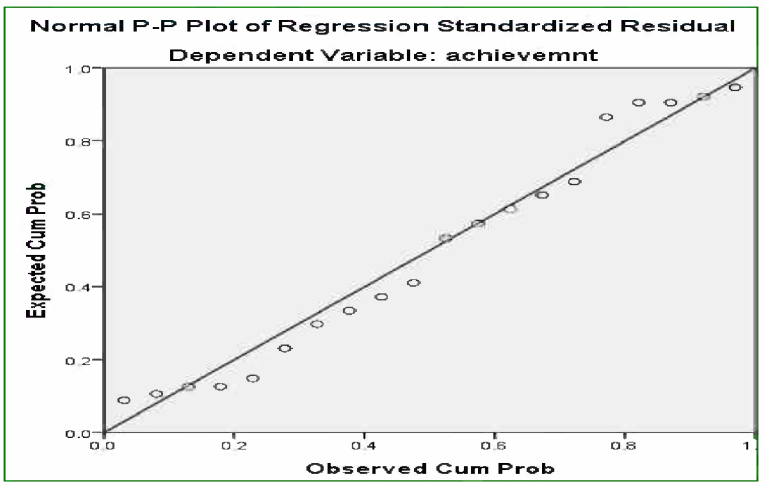
- حدد عدد العينات التي ترغب في توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها وليكن 500 لاحظ أنه إذا كان حجم العينة المتاحة كبير يمكن تحديد عدد العينات بـ 1000 مثلًا وهي Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.

- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01 ويتم الحصول على فترات الثقة المئينية Percentile Confidence intervals من خلال ترتيب فترات الثقة من الاصغر إلى الاكبر فمثلًا فترات الثقة 95% فإنه يعتمد على المئينى percentile 5^{th} و Percentile 95^{th} كنقطتى للفترة. وقد يحدث اختلاف بين فترات الثقة البارامترية التى تقدر من خلال اختبارات الفروض الكلاسيكية الصفرية عن فترات الثقة اللابارامترية المقدرة من خلال البوتستراب وهذا نتيجة لعدم الاعتدالية للبيانات.
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيطة.

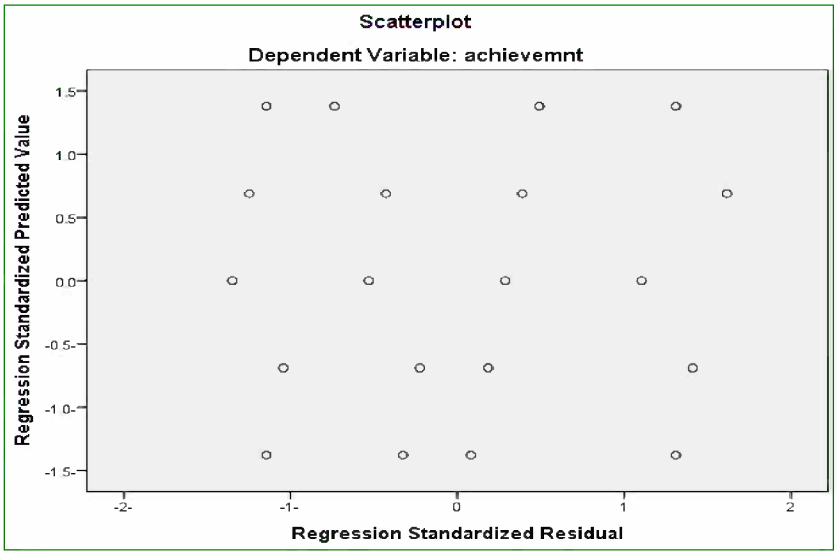
9. اضغط على Continue ثم اضغط OK

ثالثا: المخرج:

- **مسلمات البيانات:** بالنظر فى شكلNormal probability plot أعطى المنحنى الآتى:



لاحظ أن الدوائر الصغيرة (الإحداثيات) فى الشكل تكاد تقترب من الخط المستقيم وهذا يعطى دليل على أن الأخطاء (الفرق بين القيم المتنبأ بها والقيم المقاسة) اعتدالية التوزيع ولكن هذه المسلمة لم تتحقق بدرجة تامة نظرًا لابتعاد العديد من الدوائر الصغيرة عن الخط المستقيم ولذلك يمكن أن نشك فى توافر الاعتدالية للبيانات كذلك اعطى البرنامج الشكل الآتى:



وهـذا الشـكل يعـرض البـواقى المعياريـة والقـيم المتنبـأ بهـا المعياريـة ونلاحـظ أن الـدوائر الصـغيرة تتـوزع أو تنتشـر عشـوائيا وهـذا الشـكل يشـير إلـى تسـاوى التباينـات للمتغيـر التـابع لكـل مسـتوى مـن مسـتويات المتغيـر المسـتقل وهـذا يشـار إليـه Homoscedasity وهـى متطلب أساسى للانحدار بمعنى أن الأخطاء المعيارية تتوزع عشوائيًا.

ولا بـد مـن التحقـق مـن مسـلمة الاعتداليـة للمتغيـرات achievement و hours قبـل إجراء تحليل الانحدار الخطى البسيط (انظر فصل التحقق من الاعتدالية).

ثانيا: تفسير مخرج الانحدار والتنبؤ: أعطى البرنامج إحصاء وصفى للمتغيرات كالآتى:

Descriptive Statistics

	Mean	Std. Deviation	N
achievemnt	7.3000	2.61775	20
hours	3.0000	1.45095	20

ثم أعطى مصفوفة الارتباط بين المتغيرين:

Correlations

		achievemnt	hours
Pearson Correlation	achievemnt	1.000	.416
	hours	.416	1.000
Sig. (1-tailed)	achievemnt	.	.034
	hours	.034	.
N	achievemnt	20	20
	hours	20	20

ويتضح أن 0.416 = r وهو من النوع المتوسط والجزء الثانى يعبر عن الدلالة الإحصائية لاختبار ذو ذيل واحد وهى 0.034 =p وهى دالة إحصائيًا عند 0.05 والجزء الثالث من الجدول يعبر عن حجم العينة وهى 20 فردًا.

ثم أعطى البرنامج معلومات عن الانحدار كالآتى:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	hours[b]	.	Enter

a. Dependent Variable: achievemnt

b. All requested variables entered.

المتغيرات المستقلة الداخلة فى المعادلة Variables Entered وهى hours ولا يوجد متغيرات مستبعدة أو محذوفة من المعادلة والطريقة المستخدمة فى تحليل الانحدار Method هى طريقة Enter ثم عرض أسفل الجدول المتغير التابع Dep . var هي achievement

المطابقة العامة للنموذج Overall fit of the model: عرض البرنامج ملخص النموذج كالآتى :

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.416[a]	.173	.127	2.44609

a. Predictors: (Constant), hours

b. Dependent Variable: achievemnt

- القيمة R وهى معامل الارتباط المتعدد 0.416 وهى قيمة معامل ارتباط بيرسون بين التحصيل وعدد الساعات أو هى معامل الارتباط بين درجات المتغير التابع المقاسة Yودرجات المتغير التابع المتنبأ بها من المعادلة $\hat{Y}$.
- قيمة R^2 وهى مربع قيمة معامل الارتباط 0.173 = (0.416)2 = R^2 وهى تعرض التباين المفسر explainedVariance فى المتغير التابع عن طريق المتغير المستقل وهذا يعنى أن عدد الساعات المذاكرة فسرت 17.3% من تباين التحصيل .

وهنا نتساءل عما إذا كان معامل الارتباط المتعدد R ومربعة R^2 دال إحصائيًا؟ وعليه يوجد فرضان إحصائيان:

الفرض الصفرى: معادلة الانحدار لا تصلح لاستخدامها فى التنبؤ:

H0: R(ρ)= 0او Beta= 0

الفرض البديل: معادلة الانحدار تصلح لاستخدامها فى التنبؤ:

$$HA: R(\rho) \neq 0 \text{ أو } Beta \neq 0$$

حيث Beta معامل الانحدار المعيارى فى المجتمع.

ويستخدم اختبار تحليل التباين F لاختبار الدلالة الإحصائية لمعادلة الانحدار كالآتى:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.500	1	22.500	3.760	.068[b]
	Residual	107.700	18	5.983		
	Total	130.200	19			

a. Dependent Variable: achievemnt

b. Predictors: (Constant), hours

حيث:

$$F = \frac{SS_{regression} / df_{regression}}{SS_{residual} / df_{residual}} = \frac{MS_{regression}}{MS_{rresidual}}$$

- $SS_{regression}$ مجموع مربعات الانحدار وهو مقياس للتباين المفسر فى المتغير التابع عن طريق المتغير المستقل ويقدر كالآتى:

$$SS_{regression} = \frac{(\sum XY)^2}{\sum X^2}$$

- $SS_{residual}$ مجموع مربعات البواقى وهو التباين غير المفسر فى معادلة الانحدار ويقدر كالآتى:

$$\sum Y^2 = SS_{regression} + SS_{residual}$$

اذًا:

$$SS_{residual} = \sum Y^2 - SS_{regression}$$

ودرجات الحرية للانحدار هى:

$$df_{regression} = K=1$$

- K عدد المتغيرات المستقلة.

ودرجات الحرية للبواقى:

$$df_{residual} = N - K - 1 = 20 - 1 - 1 = 18$$

- N حجم العينة.

وتحسب قيمةF لاختبار الدلالة الإحصائية لمعادلة الانحدار كالآتى:

$$F = \frac{22.5/1}{107.7/18} = \frac{22.5}{5.983} = 3.76$$

وهـى نفـس القيمـة فـى المخـرج ولكـن فـى الجـدول السـابق أعطـى قيمـة P=0.068 وهـذا يعنـى أنهـا غيـر دالـة إحصـائيًا، حيـث P(0.068) > 0.05 وعليـه فـإن معادلـة التنبـؤ لا تصـلح للتنبـؤ. وإذا بحثنـا فـى جـدول F بـدرجتى حريـة 1 ، 18 نحصـل علـى F الجدوليـة = 4.41 بينمـا F المحسوبة 3.76، وبمـا ان**F المحسـوبة F> (3.76) الجدوليـة (4.41)** وعليـه نقبـل الفـرض الصـفرى أى معادلـة التنبـؤ لا تصـلح لأغـراض التنبـؤ أو انحدار التحصيل على عدد الساعات لا يختلف عن الصفر.

ويمكن حساب قيمة F من المعادلة الآتية:

$$F = \frac{R^2/K}{(1-R^2)(N-K-1)} = \frac{0.1728/1}{(1-0.1728)(20-1-1)} = 3.76$$

وعلـى ذلـك فـلا داعـى للنظـر إلـى الدلالـة الإحصـائية لمعـالم معادلـة الانحـدار وهـى معامـل الانحدار b والثابت a ولكن نعرضه هنا للتوضيح فقط.

ويمكن حساب التباين المفسر:

$$R^2 = \frac{SS_{reg}}{SS_{total}} = \frac{SS_{model}}{SS_t} = \frac{22.50}{130.2} = 0.17$$

معالم النموذج Model parameter : وهى مكونات الانحدار **أعطى المخرج الآتى:**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.050	1.283		3.937	.001
	hours	.750	.387	.416	1.939	.068

a. Dependent Variable: achievemnt

كما نعلم أن المعادلة كالآتى:

$$\hat{Y} = a + b\,X + e$$

حيث درجة أى فرد على المتغير التابع (التحصيل) هى نتيجة لثابت (a) إضافة إلى معامل (b) وخطأ، فقيم a، b تم عرضها فى العمود الثانى Unstandardized coefficients فى الشكل السابق فى الجزء الأول وهى المعاملات غير المعيارية: a = 5.050 , b = 0.750 ، حيث يستخدم SPSS الثابت b تحت مسمى B والثابت a وهو الدرجة المتنبأ بها على المتغير التابع لشخص درجته صفر على المتغير المستقل، وb هو معامل الانحدار غير المعيارى وعليه فإن معادلة الانحدار غيرالمعيارية هى:

Achievement = 0.750 hours+ 5.050

$$\hat{Y} = 0.750\ X + 5.050$$

وهذا يعنى أنه للتنبؤ بدرجة التحصيل فإننا نضرب الوقت المستغرق فى مذاكرة الرياضيات فى 0.750 ونضيف إليها الثابت 5.05، ولذلك فإذا كانت الوقت المستغرق فى مذاكرة الرياضيات ساعة لطالب فإن درجته المتنبأ بها فى الامتحان:

$$= 0.70 + 5.05 = 5.70$$

حقيقة إن معامل الانحدار يخيرنا بان تغير وحدة واحدة فى المتغير المستقل يقابلها تغير 0.70 وحدة فى المتغير التابع. ولكن السؤال المطروح لماذا نتنبأ بدرجة الطالب فى الرياضيات مثلا ونحن نعرف درجة تحصيله الحقيقة؟ الإجابة هى استخدام المعادلة السابقة لأغراض التنبؤ لمجموعة أخرى من التلاميذ لها نفس الخصائص أو للتنبؤ فى المستقبل لمجموعات أخرى لاحقة.

نتائج تحليل البوتستراب:

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval	
						Lower	Upper
1	(Constant)	5.050	-.002-	1.234	.001	2.664	7.521
	hours	.750	-.004-	.384	.075	.034	1.478

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

لاحظ أن تحليل البوتسراب اعطى فترات الثقة لمعاملات الانحدار حيث فى ضوء الدلالة الإحصائية وفترات الثقة فإن معامل الانحدار غير دال إحصائيًا. لاحظ فروق

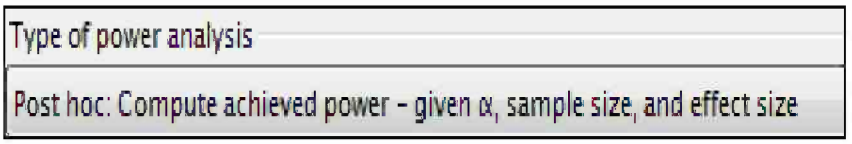
طفيفة بين اختبارات الفروض باستخدام عينة واحدة وباستخدام معاينات البوتستراب نظرًا لصغر حجم العينة لقاعدة البيانات المتولد منه البيانات.

القوة الإحصائية للانحدار البسيط باستخدام برنامج G-power

لحساب القوة الإحصائية باستخدام برنامج **G-Power اتبع الآتى:**

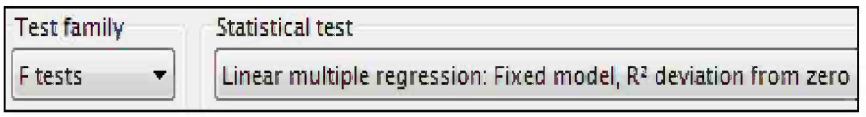
1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختر:



3. أسفل Test family اختر F- tests

4. أسفل Statistcal test اختر:



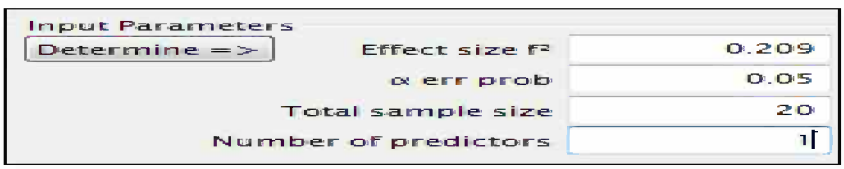
5. أدخل المعالم الآتية تحت Input parameters:

-حجم التأثير f^2 كالآتى:

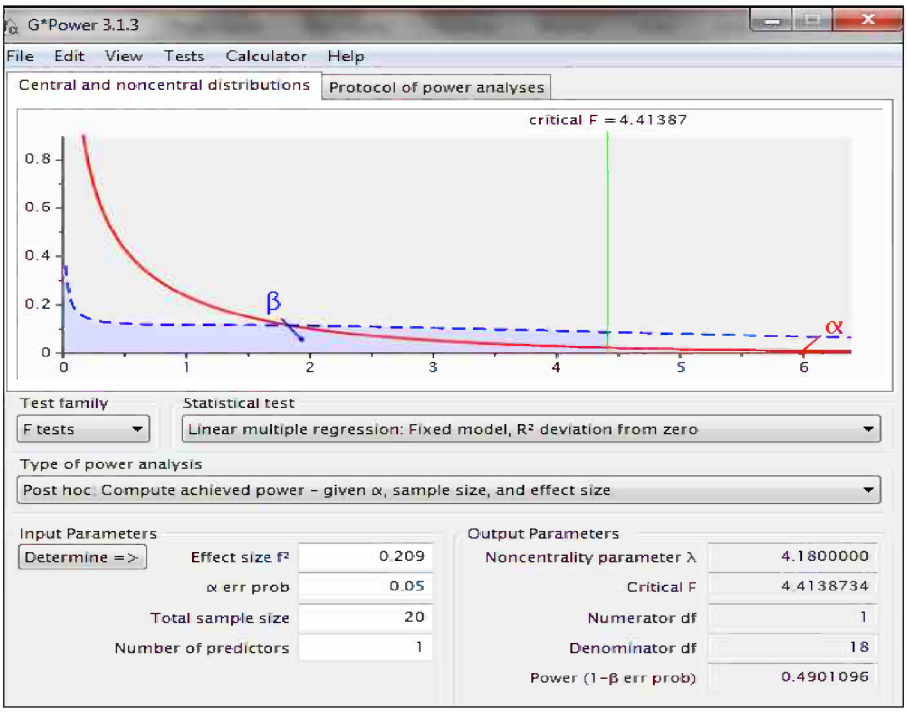
$$f^2 = \frac{R^2}{1-R^2} = \frac{0.1728}{1-0.1728} = \frac{0.1728}{0.827} = 0.209$$

وفى ضوء معاير Cohen(1988) فإن 0.02 حجم تأثير ضعيف، 0.15 متوسط، 0.35فاكثر حجم تأثير كبير.

- مستوى الدلالة الإحصائية: 0.05، حجم العينة الاجمالى= 20، عدد المنبئات=1



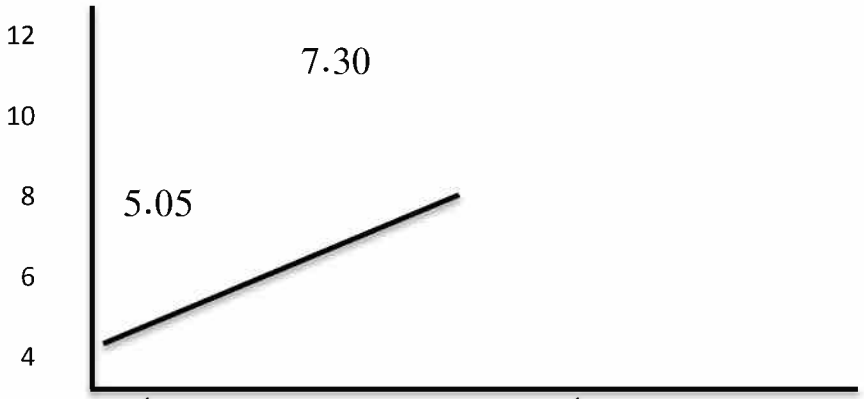
6. اضغط Calculated يظهر المخرج الآتى:



يتضح أن القوة الإحصائية= 0.490 وهذا مستوى قوة منخفض وبدورة يؤدى إلى تضخم الخطأ من النوع الثانى بيتا(β) وقد يرجع إلى صغر حجم العينة.

خط الانحدار Regression line

تستخدم معادلة الانحدار لرسم العلاقة بين الوقت والتحصيل حيث يوضع التحصيل على المحور الصادى (Y) وعدد الساعات على المحور السينى (X) وكما نعلم أن الثابت a=5.050 هى قيمة Y عندما تكون X=0 وهذه النقطة الأولى أما النقطة الثانية فهى يعبر عنها بمتوسط عدد الساعات ($\bar{X}$) = 3.00 ومتوسط التحصيل وهو 7.30



وهذا الميل يمثل الزيادة المتنبأ بها فى Y لكل وحدة فى X بمعنى أن زيادة ساعه واحدة فى المذاكرة يتوقع زيادة التحصيل بـ 0.75 نقطة واحدة، وهذا هو تفسير معامل الانحدار غير المعيارى.

معامل الانحدار المعيارى Standardized Regression Coefficient

بالنظر فى المخرج السابق نلاحظ وجود المعاملات المعيارية (Beta) وراعى أن معامل الانحدار غير المعيارى فسرناه كتغير فى المتغير التابع لكل وحدة تتغير فى المتغير المستقل وفى هذا المثال فإن 0.75 = b وهذا يعنى أن لكل وحدة واحدة فى ساعات المذاكرة تتنبأ بـ 0.75 وحدة فى التحصيل وتفسر بيتا β مثل b ولكن فى ضوء وحدات الانحراف المعيارى 0.416 = β وتعنى أن لكل انحراف معيارى زيادة فى عدد الساعات فإن التحصيل يزيد بـ 0.416 انحراف معيارى وعلى ذلك فإن β هى معامل الانحدار المعيارى عندما يتم التعبير عن قيم المتغير المستقل والتابع بالدرجات المعيارية (Z)

ويمكن تحويل b إلى β كالآتى :

$$\beta = b\,\frac{SD_y}{SD_x}$$

وعليه ففى المثال:

$$\beta = 0.75 \times \frac{1.4509}{2.6177} = 0.416$$

ولاحظ أن معامل الانحدار المعيارى هو قيمة معامل ارتباط بيرسون بين المتغيرين ومعادلة الانحدار المعيارية هى:

$$\acute{Z}\ \text{Achievement} = 0.416\ Z_{\text{hours}}$$

لا تتضمن الثابت a.

ولكن على ماذا يعتمد الباحث معامل الانحدار المعيارى أم اللامعيارى؟ كلاهما مفيد ولكن إذا كانت توجد مقارنة بين معاملات الانحدار للمتغيرات المستقلة بالتابعة فيفضل استخدام معامل الانحدار المعيارى لأنه يسمح بالمقارنة عبر متغيرات بوحدات قياس مختلفة.

عدم تحقق المسلمات

يمكن أن يكون لتحليل الضلاعة Robust analysis فى حالة الابتعاد أو عدم تحقق هذه المسلمات دورًا فى إجراء التحليل ما عدا عدم تحقق مسلمات أخطاء القياس وكذلك أخطاء التخصيص.

أخطاء القياس: أخطاء القياس فى المتغير التابع لا تؤدى إلى تحيز فى معامل الانحدار ولكنها تؤدى إلى زيادة فى الخطأ المعيارى للتقدير وعليه ظهور خلل فى اختبارات الدلالة الإحصائية (T , F) وأخطاء القياس فى المتغير المستقل تؤدى إلى تقدير منخفض Under- estimation لمعامل الانحدار وهذا نتيجة لانخفاض الثبات لقياسات المتغير المستقل حيث:

الدرجة الحقيقة لـ (T) = الدرجة المقاسة ± (X) أخطاء قياس.

وعندما تزيد الأخطاء فإنه ذلك يعنى عدم الحصول على التقدير الحقيقى للصفة أو الخاصة، والثبات يشير إلى دقة القياس وتتراوح قيمته من 0.0 إلى 1.00 حيث القيمة المرتفعة تعنى دقة القياس فمثلًا r_{tt} B = b ، حيث b إحصائية ، B معلم ، r_{tt} الثبات، فإذا كان r_{tt}=1.00 فإن B = b ، ولكن إذا كان r_{tt} = 0.70 فإنه يو جد 30 % تقدير منخفض لـ B وعليه فإن b هى تقدير منخفض لـ B.

وفى البحوث التجريبية يكون ثبات المتغير المستقل عالى لأنه تحت الضبط بينما فى البحوث غير التجريبية فإن ثبات المتغيرات المستقلة يكون من نوع المنخفض إلى المتوسط ويتراوح من 0.50 إلى 0.80 وهذا دائما يؤدى إلى وجود تحيز فى معاملات الانحدار.

أخطاء التخصيص أو التحديد Specification errors

تشير إلى الأخطاء المرتكبة فى تخصيص النموذج المراد اختباره أو عدم تحقق أى من المسلمات المراد توافرها فى النموذج، فأى نموذج انحدار هو انعكاس لبناء نظرى للظاهرة موضع الدراسة وعندما يكون بنائه بعيدًا عن الإطار النظرى بمعنى تضمين متغيرات غير مناسبة أو حذف متغيرات مناسبة أو وجود علاقة منحنية خطية Curvilinear بدلًا من العلاقة الخطية وهذا يلقى بظلاله فى البحوث غير التجريبية أكثر من البحوث التجريبية.

العلاقة بين الانحدار والأساليب الإحصائية الأخرى

تحليل التباين ANOVA يركز على الفروق بين المجموعات بينما الانحدار يركز على التنبؤ بمتغير تابع واحد أو أكثر من العديد من المتغيرات المستقلة والعمليات المستخدمة فى الأسلوبين واحدة ويعتبر اختبارى T, ANOVA تحت مظلة الانحدار بل هما حالات خاصة منه. حيث تعتبر ANOVA هى جزء من النموذج الخطى العام General Linear model (Howell, 2014)

وان المعادلة العامة لهذه الأساليب هى:

$$Y = \mu + B + e$$

ويفترض أن درجة الفرد على المتغير التابع هى مجموع المتوسط العام والتباين نتيجة تأثير المعالجة التجريبية B واضافة أو طرح تأثير الخطأ e

وهذه معادلة الانحدار :

$$Y= a + b\,x + e$$

حيث الدرجة على المتغير التابع هى مجموع ثابت لكل الأفراد (a) بالإضافة إلى تباين (b) يرجع إلى المتغير المستقل X زائدا أو ناقص خطأ عشوائى (e) وهاتين المعادلتين واحد ولهما نفس التفسير ولذلك فإن ANOVA هو جزء من النموذج الخطى العام.

السببية Causality

فى الانحدار يمكن القول إن الدافعية تساعد فى تفسير التحصيل وليس ان الدافعية مسببة للتحصيل وهذا يعنى أننا لا نستنتج (ندعى) السببية من الارتباطات:

Don't infer Causality from correlations

وتذكر الآتى:

1. ليس من المناسب استنتاج السببية من بيانات ارتباطية.

2. ليس من المناسب استنتاج السببية ما لم توجد معالجة تجريبية للمتغير المستقل.

ويمكن القول العبارات الآتية من تحليل الانحدار:

- التدخين يزيد من احتمالية الإصابة بسرطان الرئة عند الإنسان
- سمات الشخصية تؤثر فى النجاح فى الحياة
- طلاق الوالدين يؤثر فى تحصيل الأطفال وسلوكياتهم
- احتمال سمات الشخصية ان تسبب بصورة جزئية النجاح فى الحياة "

حقيقة يمكن عمل استنتاجات سببية من بيانات ارتباطية ولكننا لا نشعر بالثقة والمصداقية كما لو كانت من بيانات تجريبية وعليه فإن البحوث الارتباطية يمكن أن تدعى وجود شيء من السببية على استحياء أى انها سببية احتمالية وليست مطلقة.

الفصل السادس عشر

اختبارى Z و T لاختبار فروض حول متوسط عينة واحدة

قبل عرض اختبار Z لا بد من عرض:

توزيعات المعاينات Sampling distributions

التوزيع العينى لإحصاء ما يغطى كل القيم التى يمكن أن تأخذها الإحصائية واحتمال الحصول على أى قيمة نتيجة للصدفة فقط، ويوجد توزيع عينى مختلف عن الآخر لكل حجم عينة N واحد الطرق لاشتقاق التوزيعات للعينة هى نابعة من اعتبارات الاحتمالات الأساسية. وعلى ذلك يمكن اشتقاق التوزيع العينى لأى إحصاء بواسطة:

- تحديد كل العينات المختلفة المحتملة بأحجام N حيث تتكون من درجات المجتمع.
- حساب الإحصائية (المتوسط مثلًا) لكل عينة من العينات المختلفة.
- حساب احتمال الحصول على كل قيمة للإحصاء لو تم إجراءه بالصدفة.

يوجد عاملان يحددان شكل التوزيع العينى للمتوسط هما شكل الدرجات الخام فى المجتمع وحجم العينة، فاذا كان شكل الدرجات فى المجتمع اعتدالى فإن التوزيع العينى يكون اعتدالى بغض النظر عن حجم العينة وإذا كان التوزيع فى المجتمع غير اعتدالى فإن شكل التوزيع للعينة يعتمد على حجم العينة، وإذا كان حجم العينة كبيرًا فإن شكل التوزيع فى العينة يكون تقريبًا اعتدالى لحجم عينة $N \geq 30$ فإن النظر إلى شكل التوزيع لمجتمع الدرجات الخام غير ضرورى لأن توزيع العينة يكون تقريبًا اعتداليًا، ولذلك فإن معظم المجتمعات فى العلوم السلوكية لا تختلف كثيرًا عن الاعتدالية إذا كانت حجم العينة $N \geq 30$ من هنا تأتى مقولة " يفضل استخدام الإحصاء البارامترى إذا كان حجم العينة كبيرًا وهذا يتحدد لحجم العينة 30 فأكثر.

توجد ثلاثة أساسيات لتحديد التوزيع العينى هى المتوسط والانحراف المعيارى أو التباين وشكله.

هدف اختبارZ

يهدف إلى معرفة معالم الفرض الصفرى فى المجتمع أو اختبار فروض حول متوسط المجتمع أى تحديد ما إذا كانت للمعالجة أثر على أفراد المجتمع علمًا بأن المتوسط والانحراف المعيارى للمجتمع يكون معروف قبل إجراء التجربة، ويهدف اختبارZ أيضًا إلى المقارنة بين بيانات الدراسة الحالية بنتائج دراسة اخرى التى تعتبر بمثابة معلم المجتمع.

اختبارات الفروض لقضية بحثية (فى:Pagano, 2013)

قام باحث بإعداد برنامج لتنمية مهارة القراءة فى مجتمع المرحلة الثانوية فى محافظة أو مدينة ما علمًا بان معالم درجات المجتمع فى مادة القراءة كانت كالآتى : 16= σ ، 75 =μ وقام بتطبيق البرنامج لمدة شهر على عينة عشوائية مكونة من 100 طالب فى المرحلة الثانوية . وتساءل هل درجات القراءة بعد تطبيق البرنامج تزيد عن 75 أو هل متوسط درجات العينة بعد تطبيق البرنامج يختلف اختلافًا جوهريًا عن متوسط المجتمع ؟

الخطوات البحثية:

1. مشكلة البحث: هل متوسط درجات القراءة للعينة بعد تطبيق البرنامج يختلف عن متوسطها فى المجتمع ؟ أو هل للبرنامج أثر على مهارة القراءة ؟

2. فرض البحث: يصاغ فرض البحث اما بصورة موجهة أو غير موجهة ويصاغ كالآتى: توجد فروق بين متوسط درجات القراءة للعينة بعد تطبيق البرنامج عن متوسط درجات القراءة لمجتمع المرحلة الثانوية.

3. منهج البحث: يمكن استخدام اختبار Z للتصميمات الآتية:

ا- منهج شبة تجريبى ويتضمن التصميمات الآتية:

- قياس قبلى وبعدى (المجموعة الواحدة)
- المجموعة الوحدة وقياس بعدى.

ب- منهج وصفى يتضمن جمع البيانات على متغير واحد.

4. النموذج الإحصائى: نموذج المتغير الواحد وإحصاء بارامترى و**الاختبار الإحصائى** هو اختبار Z حول متوسط مجتمع ما والانحراف المعيارى معروف.

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية: بما أن المعلمة هى متوسط المجتمع فإن :

$$H0: \mu = 75$$

$$H_A: \mu \neq 75$$

وإذا كان فرض البحث توجد فروق ولذلك فهو غير موجه، أما إذا كان فرض البحث أن للمعالجة تأثير أو أن متوسط العينة يزيد عن متوسط المجتمع فإن الفرض البديل هو موجه لأعلى أو ذو ذيل واحد لأعلى، بكلمات اخرى أن للتجربة أثر فعال فى تنمية القراءة ويكون الفرض البديل :

$$H_A : \mu > 75$$

أو أن متوسط العينة أقل من متوسط المجتمع و الفرض البديل ذو ذيل واحد لأدنى: $H_A : \mu < 75$

2. الاختبار الإحصائى ومسلماته: الاختبار هو Z حول متوسط مجتمع وتكون صيغة حسابه كالآتى:

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}}$$

- $\overline{x}$ متوسط درجات العينة

- $\sigma_{\bar{x}}$ الانحراف المعيارى للتوزيع العينى للمتوسط وهو يساوى الخطأ المعيارى للمتوسط ولذلك يطلق عليها الخطأ المعيارى.

وحيث إن: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

اذًا:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$$

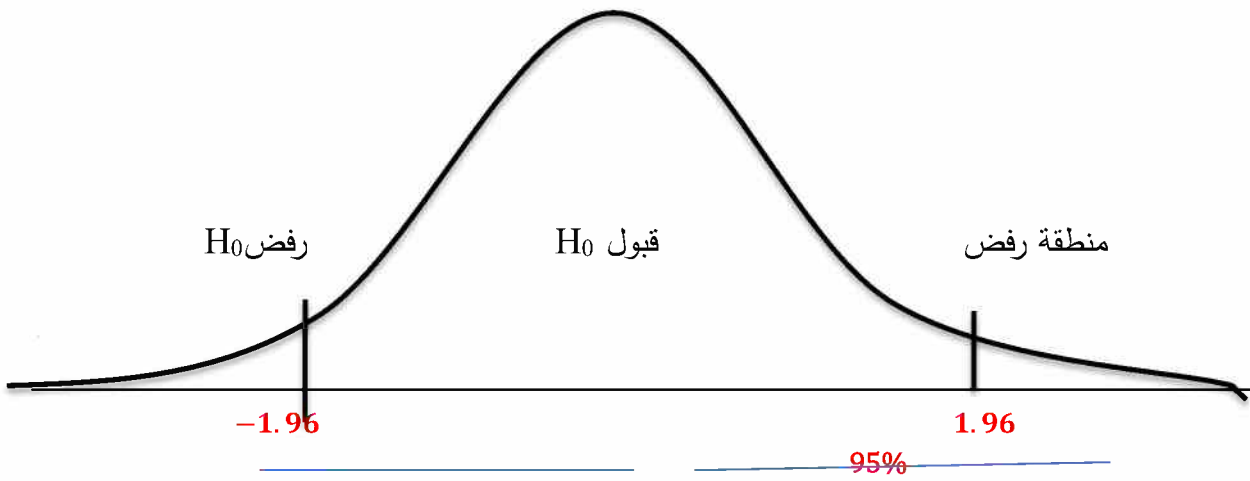
- σ الإنحراف المعيارى للعينة
- N حجم العينة.

ولهذا الاختبار مسلمات هى:

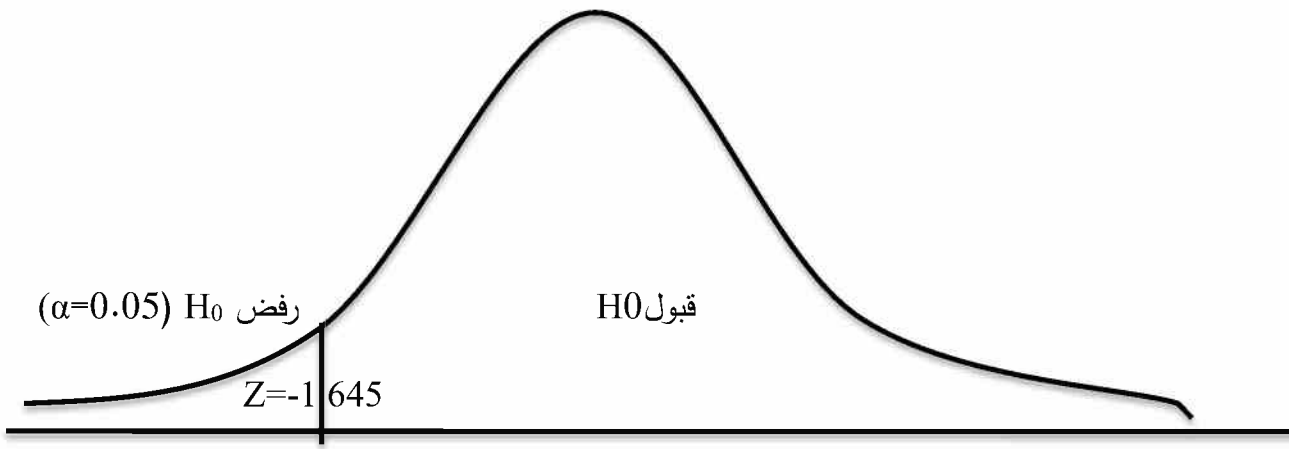
- المعاينة العشوائية.
- مستوى القياس فترى على الأقل.
- التوزيع العينى لبيانات العينة اعتدالى.
- يستخدم لاختبار فروض حول متوسط مجتمع إذا كان متوسط المجتمع والانحراف المعيارى معروفين.
- يفضل استخدامه لعينة $N \geq 30$ وذلك للحصول على التوزيع الإعتدالى.
- استقلالية القياسات بمعنى أن درجات الأفراد مستقلة عن بعضها البعض.

3. مستوى الدلالة الإحصائية وقاعدة القرار: تبنى الباحث مستوى دلالة إحصائية $\alpha = 0.05$ والتوزيع العينى لاختبار Z هو اعتدالى، وهى كما يراها Pagano (2013) هى المنطقة تحت المنحنى التى تتضمن كل القيم لإحصائية للاختبار التى تسمح برفض الفرض الصفرى، ومنطقة الرفض هى منطقة التوزيع العينى التى تعكس تلك القيم لمتوسط العينة التى تكون غير محتملة أو غير صحيحة إذا كان الفرض الصفرى حقيقى والقيم الحرجة للاختبار هى تلك القيم فى التوزيع العينى التى تعكس بداية منطقة رفض H_0.

وبالتالى يتم تحديد منطقة القبول والرفض وذلك بتحديد القيمة الحرجة أو الجدولية لـ Z من خلال: $\alpha = 0.05$ ، اختبار ذو ذيلين وبالكشف فى جدول Z (ملحق) فان: $Z_c.v = \pm 1.96$ ، وعليه فإن منطقة القبول والرفض فى التوزيع الاعتدالى لـ Z كالآتى:



أما إذا كان الاختبار ذا ذيل واحد لأدنى:



وتكون قاعدة القرار كالآتى: إذا كانت **Z (المحسوبة)** $\leq Z$ (الحرجة) نرفض الفرض الصفرى.

4. حســـاب إحصـــائية الاختبـــار: وبعـد حسـاب الآتـى: متوسـط المجتمـع $\mu = 75$ ، متوسط العينة $\bar{x} = 72$، الانحراف المعيارى للعينة $\sigma = 16$

إذًا الخطأ المعيارى:

$$\sigma_{\bar{X}} = \frac{16}{\sqrt{100}} = 1.6$$

إذًا:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{72 - 75}{16} = -1.88$$

5. القـرار والتفسـير: Z المحسـوبة (1.88−) > Z الجدوليـة (1.96)، اعتمـد علـى القيمـة المطلقـة وعليـه نقبـل الفـرض الصـفرى بمعنـى لا توجـد فـروق ذات دلالـة إحصـائية بـين متوسـط العينـة ومتوسـط المجتمـع أو بكلمـات أخـرى: البرنـامج لـيس لـه أثـر فـى تنميـة مهـارة القـراءة لـدى أفـراد عينـة طـلاب المرحلـة الثانويـة. ولكـن إذا تـم تبنـى الفـرض البـديل الموجـه فـإن القـرار يكـون اختبـار ذو ذيـل واحـد: **Z المحســوبة (1.88-) < Z الجدوليــة (1.645)**، وعليــه يــتم رفــض الفــرض الصــفرى بمعنــى أنــه توجــد فــروق ذات دلالــة إحصـائية بـين متوسـط العينـة ومتوسـط المجتمـع لصـالح المتوسـط الأدنـى وهـو المجتمـع وعليه فإن البرنامج له أثر فعال فى زيادة مهارة القراءة.

القرار فى ضوء فترات الثقة

صـناعة القـرار مـن خـلال حسـاب فتـرات الثقـة لاختبـار Z كمـا أوردهـا Aron et al. (2013) لمستوى$\alpha = 0.05$:

حساب الخطأ المعيارى لتوزيع المتوسطات : $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = 1.6$ –

– لفترات الثقة 95% أضف 1.96مضروبًا فى الخطأ فوق وتحت متوسط العينة:

(الحد الأعلى)Upper limit=$\bar{x} + 1.96(\sigma\bar{x})$

(الحد الأدنى) Lower limit = $\bar{x} - 1.96(\sigma\bar{x})$

فى حالة الاختبار ذو ذيلين:

C1 (72+1.96 (1.6), (72-1.96 (1.6))

CI(75.136, 68.864)

وعليه فإن قيمة المتوسط الموضوعة فى الفرض الصفرى (75) تقع بين الحد الأعلى والأدنى ومن ثم يتم قبول الفرض الصفرى وبالتالى تكون استراتيجية القرار كالآتى:

الموقف	القرار	التفسير
الإحصائية فى منطقة الرفض H_0	**رفض الفرض الصفرى**	**توجد فروق دالة إحصائيا**
الإحصائية ليست فى منطقة الرفض	**فشل رفض الفرض الصفرى**	**لا توجد فروق دالة إحصائيا**

6. حجم التأثير: يقدر حجم التأثير لاختبار z كما عرضها Cohen (1988) كالآتى:

$$d = \frac{\text{فروق المتوسطات}}{\text{الإنحراف المعيارى}} = \frac{\mu \text{ معالجة} - \mu \text{ لا معالجة}}{\sigma} = \frac{\bar{x} - \mu \text{ مجتمع}}{\sigma}$$

$$= \frac{72-75}{1.6} = 0.187$$

وتكون حدوده وفقا لـ Cohen(1988) كما يلى:

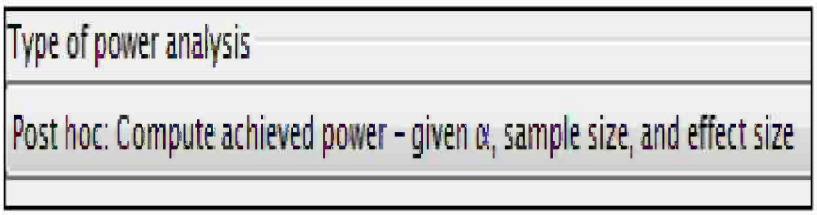
مقدار d	تفسير d
0.2	ضعيف
0.50	متوسط
$d \geq 0.80$	كبير

على الرغم من أن كثير من الباحثين لا يرون ضرورة لحساب حجم التأثير فى حالة النتائج غير الدالة إحصائيًا إلا أنه يفضل استخدامها لتحديد الحجم المطلق لتأثير المعالجة لكى يسترشد بها الباحثين فى دراساتهم السابقة وإجراء دراسات ما وراء التحليل.

7. **القوة الإحصائية:** لتحديد القوة الإحصائية للمثال السابق باستخدام برنامج G-Power اتبع الآتى:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

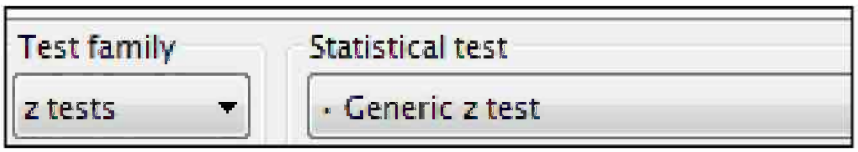
2. تحتType of power analysisاختار:



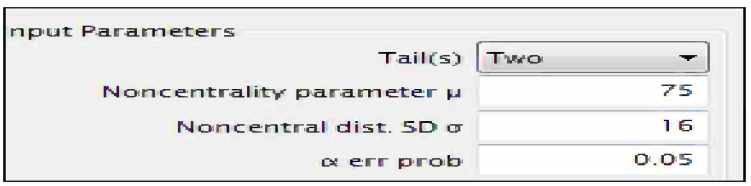
لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

3. تحت Test family اختار Z- tests

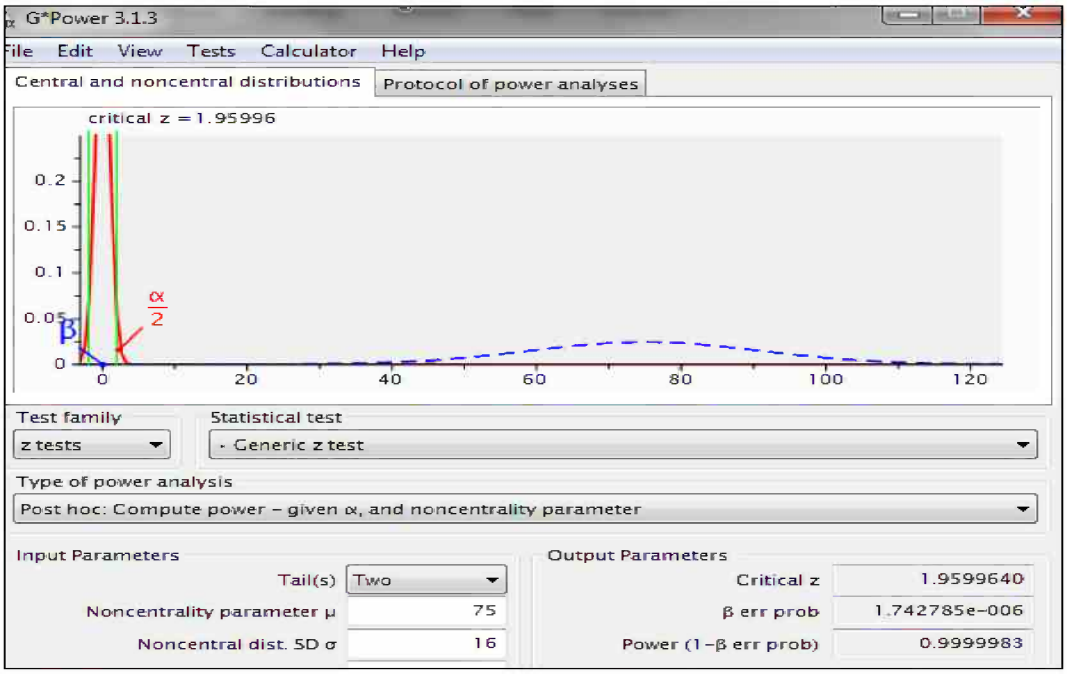
4. تحت Statistical test اختار:



5. ادخل المعالم Input parameters:



6.اضغط Calculated تظهر شاشة الناتج كالآتى:



يتضح أن القوة الإحصائية=0.999 وهو مستوى عالٍ جدًا.

كتابة نتائج Z فى الدراسات المنشورة وفقا لـ APA

اتضح أن البرنامج ليس له أثر دال إحصائيًا على درجات القراءة لطلاب المرحلة الثانوية:P>0.05 , Z=-1.88. أو أتضح وجود فروق بين متوسط درجات الاستعداد لطلاب علم النفس ومتوسط درجات الاستعداد لمجتمع طلاب الجامعة حيث: Z = 9.60 , P < 0.05 ,ES = 0.80

اختبار T لاختبار فروض حول متوسط مجتمع عينة واحدة

Student's T test for single sample

تناولنا اختبار Z وهو مناسب فى المواقف البحثية التى يكون فيها المتوسط والانحراف المعيارى للمجتمع معروف، وهذه النوعية من المواقف نادرة نسبيًا ولكن المواقف البحثية الأكثر شيوعًا فى البحث هى تلك التى يكون متوسط الفرض الصفرى للمجتمع معلوم والانحراف المعيارى غير معلوم، فى هذه الحالة لا يمكن استخدام اختبار Z وبدلا منه نستخدم Tوهو كثير الشبه باختبار Z

هدف اختبار T لعينة واحدة

يستخدم لاختبار فروض حول متوسط مجتمع على متغير ما ومعرفة ما إذا كان متوسط عينة ما $\bar{X}$ يختلف عن متوسط المجتمع μ تمامًا مثل Z ولكن الانحراف المعيارى للمجتمع غير معلوم ويفضل أن يكون حجم العينة 30 فأكثر وتوزيع درجات المجتمع اعتدالية. ويرى Huck (2012) أن الفرق الوحيد بين Z و **T** هو أن اختبار Z يعطى قيمة محسوبة أكبر قليلًا مما يجب أن تكون وقيمة P الاحتمالية أصغر مما يجب أن تكون وهذا التحيز يكون محدودًا إذا كان حجم العينة 30 فأكثر.

معادلات Z و T

Z	T
$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}}$	$\mathbf{T} = \frac{\bar{x} - \mu}{s / \sqrt{N}}$
$= \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$	$= \frac{\bar{x} - \mu}{S_{\bar{x}}}$

حيث S = تقدير لـ σ و $S_{\bar{x}}$ تقدير لـ $\sigma_{\bar{x}}$ من العينة

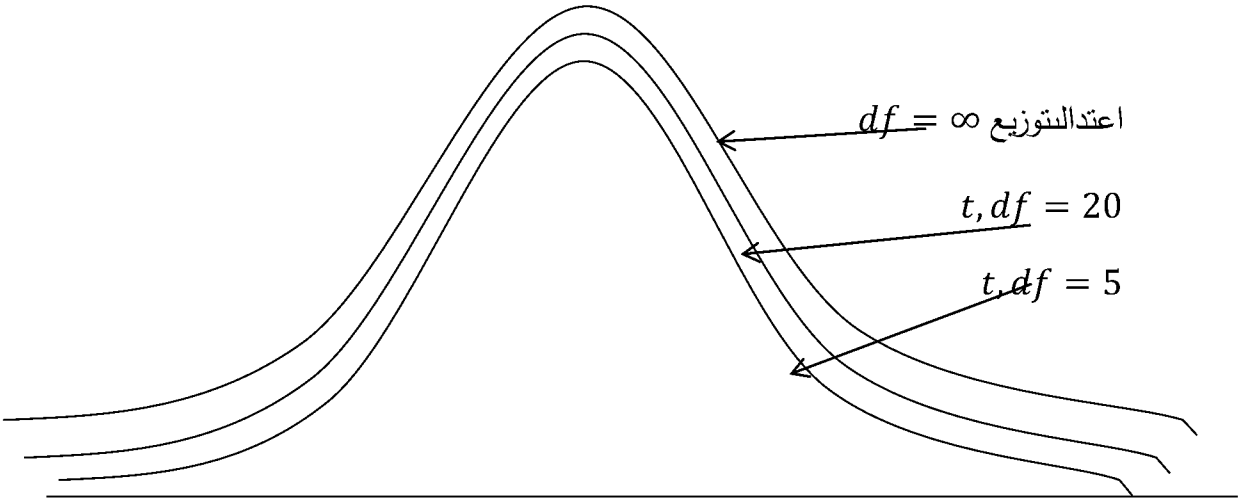
وعلى ذلك فإن الباحث يقدر الانحراف المعيارى للمجتمع σ باستخدام الانحراف المعيارى للعينة، ولكن هل استخدام S للعينة لتقدير σ لمجتمع يؤثر على الاختبار الإحصائى؟ بكل تأكيد خاصة إذا كانت أحجام العينات صغيرة فإن التوزيع العينى للعينة يكو غير مناسب، وتوزيع **T** هو عائلة من التوزيعات المنتظمة أو الاعتدالية التوزيع تتغير كلما تغيرت أحجام العينة (Hinkle et al.,1994) .

تطبيقات T

- اختبار فروض حول نقطة منتصفية Midpoint، فإذا كان مدى المتغير من 0 إلى 10 فإنه يمكن اختبار فروض لمتغير ما حول القيمة 5 وذلك لتحديد ما إذا كان متوسط المجتمع يختلف عن 5.
- اختبار فروض حول قيمة متوسطة لدرجات متغير ما فى ضوء دراسات سابقة.

- اختبار فروض حول متوسط مجتمع ما محدد مسبقا مثل متوسط التحصيل مثلا أو أى متغير آخر فى مجتمع ما معروف.

التوزيع العينى لـ T: يتم الحصول على التوزيع العينى لـ Tمن خلال حساب إحصاء T لعدد كبير من العينات، وكما نعلم أن التوزيع العينى لـ Z يكون اعتدالى خاصة إذا كان حجم العينة أكبر من 30 وبالمثل فإن التوزيع العينى لـ T تقريبًا اعتدالى مثل توزيع Z ويتحدد هذا التوزيع عن طريق درجات الحرية، وتقترب اعتدالية Tمن Z إذا كان حجم العينة كبير، ودرجات الحرية هو مفهوم رياضى حيث عدد درجات الحرية للإحصاء يعتمد على حجم العينة ودرجات الحرية لاختبار T:df=N-1



وكما نرى أن التوزيع يقترب تمامًا من الاعتدالية إذا زادت درجات الحرية ويكون اعتدالى بدرجة تامة إذا كانت df >120 وتوزيع Tهو دالة وظيفية لدرجات الحرية.

اختبارات الفروض لقضية بحثية

أراد باحث معرفة أثر التدريب على استراتيجية ما للتبكير بالمرحلة الزمنية لبداية النطق عند الأطفال حديثى الولادة، علما بأن متوسط المجتمع μ=13 (شهرًا) واختار الباحث عينة عشوائية مكونة من 15 طفلًا وكانت بيانات العينة كالآتى:

Age (X)	x^2	$x - \bar{x}$	$(x - \bar{x})^2$
8	**64**	**-3**	**9**

9	81	-2	4
10	100	-1	1
15	225	4	16
18	324	7	49
17	289	6	36
12	144	1	1
11	121	0	0
7	49	-4	16
8	64	-3	9
10	100	-1	1
11	121	0	0
8	64	-3	9
9	81	-2	4
12	144	1	1
∑ 165	1971	0	156

وأراد معرفة ما إذا كان للبرنامج أثر فى التبكير بالمرحلة الزمنية لنطق الكلام؟

الخطوات البحثية

1. **سؤال البحث:** هل توجد فروق بين متوسط العمر لبداية الكلام لأفراد العينة عن متوسط العمر للمجتمع؟

2. **فرض البحث:** للبرنامج التدريبى أثر على التبكير بالمرحلة الزمنية لنطق الكلمات عند الأطفال حديثى الولادة أو توجد فروق بين متوسط عمر العينة عن متوسط عمر المجتمع (13).

3. **متغيرات البحث:** المرحلة العمرية التى يبدأ عندها النطق: متغير نسبى- متصل- تابع.

4. **النموذج الإحصائى:** نموذج المتغير الواحد (الصفرى) – إحصاء بارامترى، و**الاختبار الإحصائى المناسب** هو T لعينة واحدة.

خطوات اختبارات الفروض الصفرية

1. **الفروض الإحصائية:**

$$H_0 : \mu = 13$$
$$H_A : \mu \neq 13$$

2. الاختبار الإحصائى المناسب ومسلماته: هو اختبار **T** لعينة واحدة ويتحدد بالصيغة:

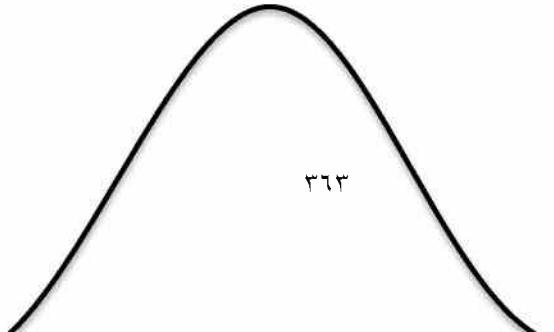
$$\mathbf{T} = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

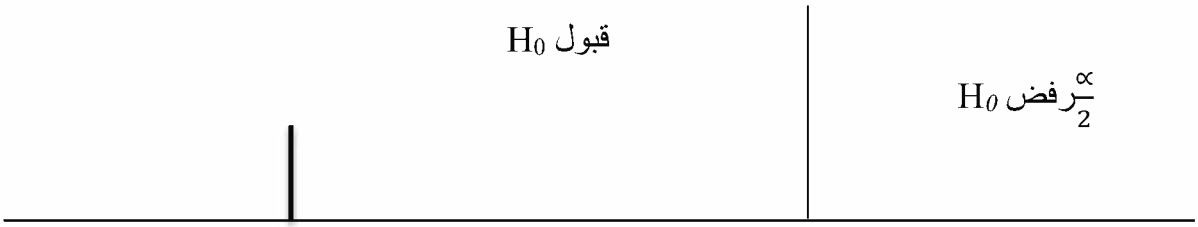
والاختبار ذو ذيلين و df= N-1=15-1=14 ، **ومسلماته:**

- **الاعتداليـــة:** بيانـات العينـة تخضـع للتوزيـع الاعتـدالى وهـذه المسـلمة ضـرورية فـى حالـة استخدام T للعينـات الصـغيرة ولكـن إذا كانـت **30** ≤ **N** فـإن الخطـأ المعيـارى يكـون صغيرًا وهذه المسلمة أقل تأثيرًا على النتائج.
- **المعاينـــة العشـــوائية:** البيانـات تـأتى مـن عينـة عشـوائية مـن المجتمـع حتـى نثـق فـى النتـائج ونعممهـا ومـن غيـر المناسـب اسـتخدام اختبـارات الفـروض مـع عينـات غيـر عشوائية.
- **الاستقلالية:** تفترض أن أى قياسات للأفراد مستقلة عن بعضها البعض.

3.**الدلالـة الإحصـائية وقاعـدة القـرار:** افتـرض الباحـث أن 0.05 =α ويـتم قسـمة مسـتوى الدلالـة الإحصـائية علـى جـانبى المنحنـى$\frac{\alpha}{2}$، ودرجـة الحريـة df=14 ، ولاختبـار ذو ذيلين ، وبالبحث فى جدول T للحصول على قيمة T الحرجة وهى: ± 2.145

وتحديد منطقة القبول والرفض كالآتى:





و إذا كانت T **المحسوبة** $\leq$ T **الحرجة** نرفض الفرض الصفرى ، وإذا كان الاختبار ذو ذيل واحد تكون منطقة الرفض أصغر عندما $\alpha = 0.05, df=14$

4. حساب إحصائية الاختبار: بما ان: $\bar{x} = 11$ و $S = 3.34$ اذاً:

$$\mathbf{T} = \frac{11 - 13}{3.34/\sqrt{15}} = -2.32$$

أشار Pagano (2013) إلى تقدير T من خلال الدرجات الخام كالآتى :

$$\mathbf{T} = \frac{\bar{x} - \mu}{\sqrt{\frac{SS}{N(N-1)}}}$$

حيث إن:

$$SS = \sum (x - \bar{x})^2 = \sum x^2 - \frac{\sum (x)^2}{N}$$

$$= 1971 - \frac{(165)^2}{15} \quad = 156$$

$$\mathbf{T} = \frac{11 - 13}{\sqrt{\frac{156}{15 \times 14}}} = -2.32$$

5.القرار والتفسير: حيث أن: T المحسوبة $|-2.32| <$ **T** الحرجة **2.14** (اعتمد على أن القيمة المطلقة بدون الإشارة السالبة) وعليه يتم رفض الفرض الصفرى لأن القيمة المحسوبة وقعت فى منطقة رفض الفرض الصفرى وبالتالى توجد فروق ذات دلالة إحصائية بين متوسط

أعمار العينة عند النطق بالكلام عن متوسط عمر المجتمع أو أن التجربة أو البرنامج لها أثر فى التبكير بنطق الكلمات عند الأطفال حديثى الولادة.

6. حجم التأثير: ولمعرفة ما إذا كان للمتغير المستقل أثر حقيقى على المتغير التابع، ولتوضيح الأثر الحقيقى أو الفعلى أو الجوهرى لا بد من تحديد مقداره ولذلك لا بد من تقدير حجم التأثير. يتم تقدير حجم التأثير لهذا الاختبار من خلال مؤشرات عديدة أهمها:

أ- مؤشر Cohen (1988):

$$d = \frac{\bar{x} - \mu}{S}$$

- d قيمة حجم التأثير
- S الانحراف المعيارى لبيانات العينة

اذًا:

$$\mathrm{d} = \frac{(11 - 13)}{3.34} = 0.60$$

وفى ضوء حدود ومعايير كوهين 0.2 - 0.0 ضعيف جدًا و0.79 - 0.5 متوسط و $0.80 \leq$ d كبير وعليه فإن حجم التأثير من النوع المتوسط وفقا لمعايير كوهين لتفسير حجم التأثير ويتم تفسيرها أيضًا حسب التخصص والدراسات السابقة.

ب- نسبة التباين المفسر Explained(r^2): الطريقة الثانية لتحديد حجم التأثير لاختبار T لعينة واحدة هى تحديد التباين المفسر للمعالجة على المتغير التابع ويتم ذلك من خلال مؤشرات عائلة r وتسمى مؤشر ايتا تربيع η^2 كالآتى:

$$r^2 = \frac{T^2}{T^2 + df} \text{ او } \eta^2$$

- Tهى القيمة المحسوبة
- df درجات الحرية

وعليه فإن:

$$r^2 = \frac{(-2.32)^2}{(-2.32)^2 + (15-1)} = \frac{5.3824}{19.3824} = 0.277$$

حيث إن التجربة فسرت 27.7 % من تباين العمر الزمنى لبداية الكلام عند الأطفال ويفسر فى ضوء معايير كوهين حجم تأثير كبير (صغير 0.01 و متوسط 0.09 وكبير **0.25** فاعلى). وعرض(2014) Gravetter &Wallnau هذه المعادلة لحساب حجم التأثير تحت مسمى مؤشر r^2 بينما عرضها (2015) Privitera تحت مسمى مؤشر إيتا تربيع ولكن المؤلف يميل إلى صيغة r^2 .

ج- مؤشر أوميجا ω^2 :وتتحدد قيمته كما يلى :

$$\omega^2 = \frac{T^2 - 1}{T^2 + df}$$

$$\omega^2 = \frac{(-2.32)^2 - 1}{(-2.32)^2 + (15-1)} = \frac{4.382}{19.382} = 0.260$$

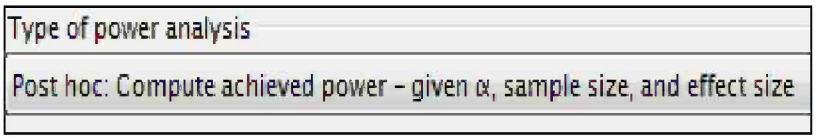
حدد Cohen (1988) معايير للمؤشرات الثلاثة:

المؤشر لحجم	d	r
تافهة	———	———
صغيرة	$d < 0.2$	$r < 0.30$
متوسطة	$0.2 < d < 0.8$	$0.3 < r < 0.5$
كبيرة	$d \geq 0.8$	$r \geq 0.5$

7.القوة الإحصائية: لتقدير القوة الإحصائية للمثال السابق باستخدام برنامج G-Power اتبع الاتى:

1.افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختار:

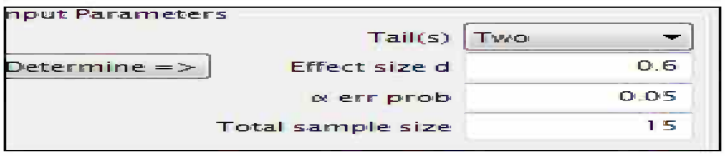


لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

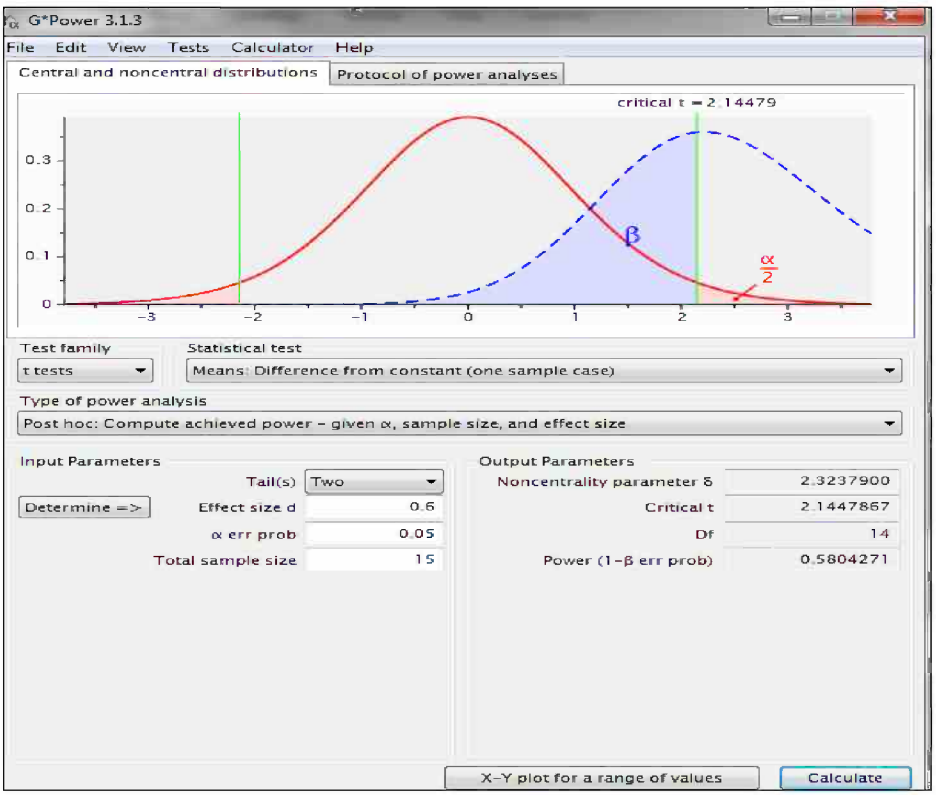
3. أسفل Test family اختار t- tests

4. أسفل Statistical test اختار:

5. ادخل المعالم Input parameters:



6. اضغط على Calculated يعطى شاشة الناتج كالآتى:



يتضح أن القوة الإحصائية= 0.58 وهذا حجم قوة غير مناسب اذ يجب أن يزيد عن 0.80 وقد يرجع هذا إلى صغر حجم العينة.

فترات الثقة لمتوسط المجتمع

كما نعلم اختبارات الفروض الصفرية قائمة على تقدير معالم المجتمع من إحصائيات العينة وهذا المدخل يطلق عليه تقديرات النقطة Point estimates والمدخل الآخر هو فترات الثقة وهو مدى من القيم من المحتمل أن يقع فيه معلمة المجتمع ولتكن المتوسط. فمثلًا فترة الثقة 95% هى فترة يمكن أن تتضمن قيمة الإحصائية فى المجتمع فى هذه الفترة باحتمال 95%.

بناء 95% فترة الثقة: حد الثقة الأدنى: $\text{الادنى CI} = \bar{x} - s_{\bar{x}} T_{c.v}$

- $Tc.v$ القيمة الحرجة أو الجدولية
- $\bar{x}$ متوسط العينة.

حد الثقة الأعلى: $\text{الأعلى CI} = \bar{x} + s_{\bar{x}} T_{c.v}$

بالبحث فى جدول T عند $\alpha = 0.05$ ، $df=15-1=14$، و لـ 95 ولاختبار ذو ذيلين فان:

$$\text{الادنى} = 11 - \frac{3.34}{\sqrt{15}}(\ 2.145)$$

$$\text{الأعلى} = 11 + \frac{3.34}{\sqrt{15}}(\ 2.145\ )$$

إذًا حدّى الثقة هما ( )

بالتالى فإن قيمة متوسط المجتمع (القيمة المفترضة فى الفرض الصفرى) تقع خارج هذه الفترة وعليه نرفض H_0 وكلما كانت فترات الثقة كبيرة كلما زادت الثقة أن يكون متوسط المجتمع يقع فى هذا المدى.

مثال آخر لحساب فترات الثقة

اهتم باحث بتحديد متوسط وزن عمر الأطفال فى سن العام ونصف واختار 15 طفلًا، وقدر متوسط الوزن للعينة 7.2، والانحراف المعيارى 0.48 فإن فترات الثقة 95% لاختبار ذى ذيلين كالآتى:

$$CI\text{الادنى} = \bar{x} - s_{\bar{x}} T_{c.v\,(0.025)}$$

$$s\bar{x} = \frac{0.48}{\sqrt{15}} = 0.124$$

ولتحديد قيمة T الحرجة عند α= 0.025 ، df=15-1=14 إذا Tالحرجة= 2.145 وعليه فإن حدى الثقة كالآتى :

$$\mu\text{الادنى} = 7.2 - 0.124(2.145)$$

$$= 7.2 - 0.266 = 6.93$$

$$\mu \text{ الأعلى} = \bar{x} + s_{\bar{x}} T_{0.025}$$

$$= 7.2 + 0.124(2.195)$$

$$= 7.2 + 0.266 = 7.47$$

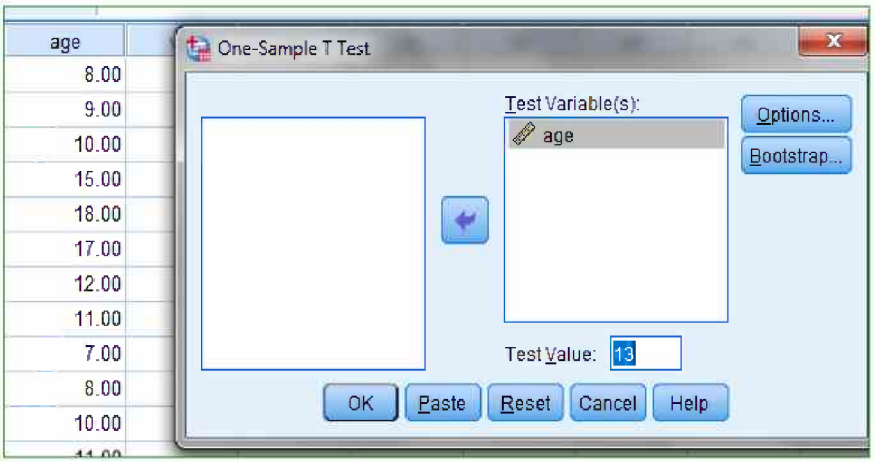
وعليه فإن فترات الثقة 95% هى (7.47- 6.93).

تنفيذ اختبار T لعينة واحدة فى SPSS

اولًا: إدخال البيانات: 1. اضغط على Variable view ثم اكتب اسم المتغير فى عمود Name العمر:age

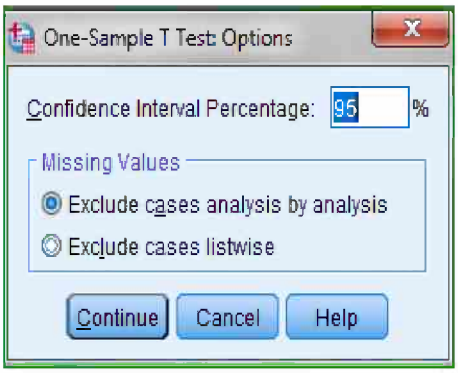
2. اضغط على Data view، ابدأ فى إدخال البيانات

ثانيًا : تنفيذ الأمر:1.اضغط Analyze→Compare Means→One sample T test تظهر الشاشة الآتية:



2. انقل age إلى مربع Test variables

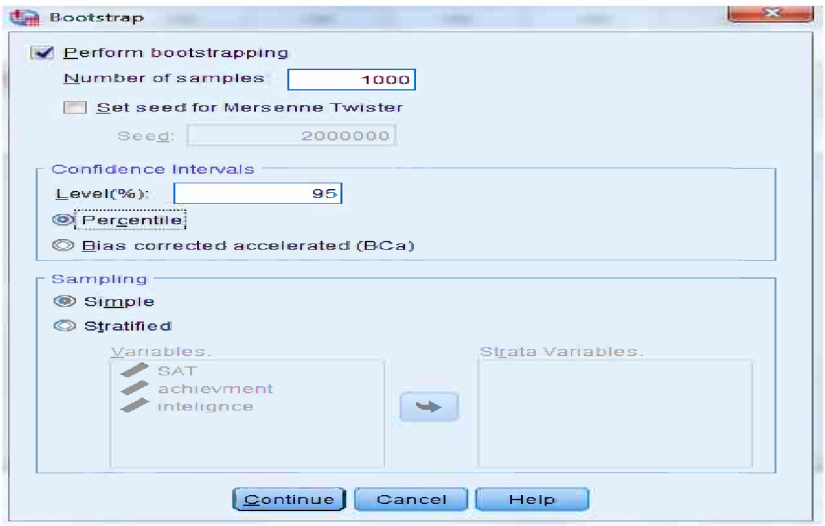
3. فى مربع Test value اكتب القيمة المراد اختبار الفروض حولها وهى 13



4. اضغط على اختيار Options واكتب حدود الثقة سواء 95 أو 99%

5. اضغط Continue

6.اضغط على Bootstrap تظهر النافذة الآتية:



7. اضغط Perform bootstrapping:

- حدد عدد العينات التى ترغب فى توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها وليكن 500 لاحظ أنه إذا كان حجم العينة المتاحة كبير يمكن تحديد عدد العينات بـ 1000 مثلًا وهى Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة. ويتم ال
- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01 ويتم الحصول على فترات الثقة المئينية Percentile Confidence intervals من خلال ترتيب فترات الثقة من الأصغر إلى الأكبر فمثلًا فترات الثقة 95% فإنه يعتمد على المئينى Percentile 5^{th} و Percentile 95^{th} كنقطتى للفترة. وقد يحدث اختلاف بين فترات الثقة البارامترية التى تقدر من خلال اختبارات الفروض الكلاسيكية الصفرية عن فترات الثقة اللابارامترية المقدرة من خلال البوتستراب وهذا نتيجة لعدم الاعتدالية للبيانات.
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيط.

8. اضغط Continue ثم OK

```
T-TEST
  /TESTVAL=13
  /MISSING=ANALYSIS
  /VARIABLES=age
  /CRITERIA=CI(.95).
```

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	15	11.0000	3.33809	.86189

ثالثًا: تفسير المخرج : الجدول الأول كالآتى: المتوسط Mean = 11، والانحراف المعيارى Std. Dev = 3.33 ، و N = 15 . ثم إحصائيات الاختبار كالآتى:

One-Sample Test

	Test Value = 13					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	-2.320-	14	.036	-2.00000-	-3.8486-	-.1514-

حيث قيمة T المحسوبة = −2.32 ، و df=14، والقيمة الاحتمالية(P) Sig. (2.tail) = 0.036 وعليه فان: 0.05(α) > 0.036 (p) بالتالى نرفض H_0 أى أن متوسط أعمار

الأطفال يختلف عن متوسط المجتمع 13. بمعنى للبرنامج أثر فى التبكير فى المدة الزمنية للنطق.

ويمكن صناعة القرار فى ضوء فترات الثقة للفروق وهى:

$$CI_{95} = (-3.848, -1.514)$$

وبما أن فروق المتوسطات -2.00 =Mean difference وعليه فإنها تقع فى مدى حدود الثقة بمعنى 95% من العينات سوف تعطى قيمة فروق تقع فى هذا المدى وهذا يعطينا ثقة عالية فى قيم إحصائيات العينات كتقدير لقيمة المجتمع غير المعروفة وفى الدلالة الإحصائية وعليه توجد دلالة إحصائية وبالتالى نرفض H_0 .

نتائج البوتستراب:

Bootstrap for One-Sample Test

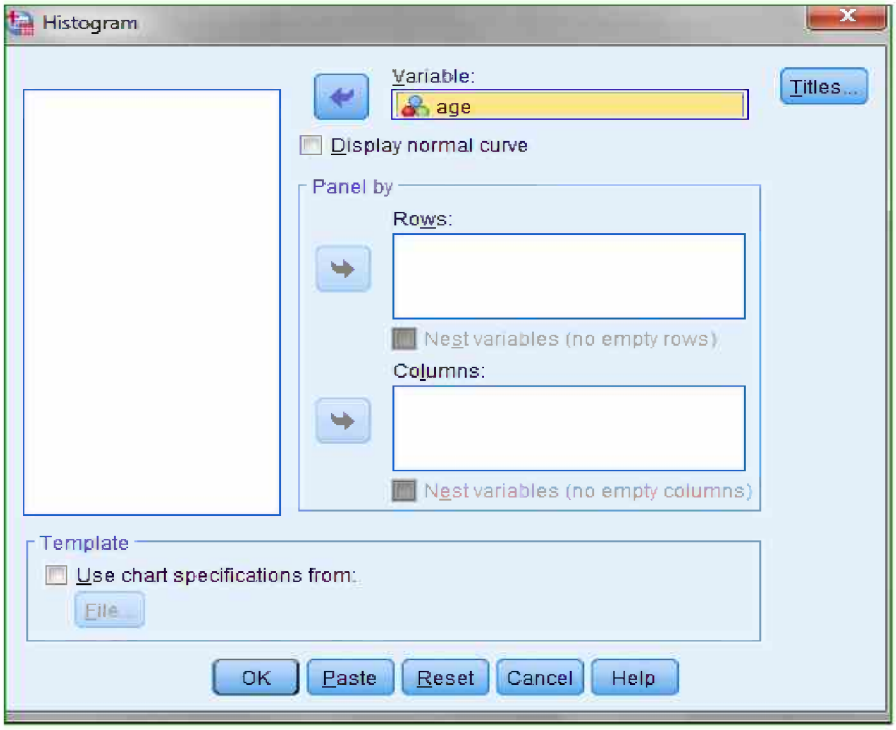
	Mean Difference	Bootstrap[a]			
				95% Confidence Interval	
		Bias	Std. Error	Lower	Upper
age	-2.00000-	-.03447-	.84455	-3.60000-	-.26667-

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

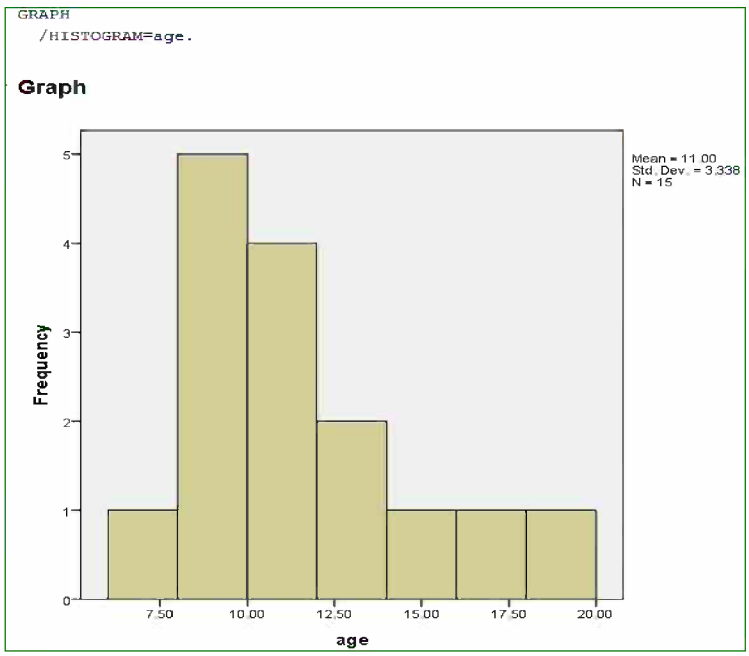
يتضح أن متوسط الفروق هو نفسه كما فى اختبارات الفروض الصفرية ولكن اختلفت فترات الثقة 95% وكانت اضيق مما يدل انها أكثر دقة من التعامل مع العينة العشوائية الواحدة والقرار هو رفض الفرض الصفرى.

عرض النتائج فى ضوء المدرج التكراري

1. اضغط Graphs→Legacy Dialogs →Histogram



2. انقل المتغير age إلى مربع Variable اضغط OK يعطى المدرج كالآتى:



كتابة نتائج T عينة واحدة وفقا لـ APA

اتضح من نتائج اختبار **T** وجود فروق دالة إحصائيا بين متوسط اعمار العينة عند النطق بالكلام ومتوسط اعمار المجتمع:

$$\mathbf{T}(14) = -2.32, p < 0.05, d = 0.60,\ \ \mathrm{CI95}\ (-3.848\ , -1.514)$$

عند استخدام اختبار T لتقيم ما إذا كان عمر الذين يتلقون البرنامج عند النطق يختلف إحصائيًا عند 0.05.، واتضح أن متوسط العينة 11 دال إحصائيًا عند 0.05.

$$\mathbf{T}(14) = -2.32, p < 0.05\ , r^2 =$$

$$\mathbf{T}(14) = -2.32\ , \mathrm{p} = 0.004\ , \mathrm{d} = 0.6$$

حيث قيمة p هى القيمة الاحتمالية لمخرج SPSS

وأن فترات الثقة 95% للمتوسط تتراوح بين ـــــــ و ـــــــــ وأن حجم التأثير d=0.60 وهو من النوع المتوسط والشكل الآتى يوضح توزيع درجات عمر الـ 15 طفل عند النطق.

الفصل السابع عشر

اختبار فروض لإحصائيات أخرى

(معامل الارتباط – النسبة)

فى الفصل السابق تناولت اختبارات فروض حول متوسط ما فى حالة ما إذا كان الانحراف المعيارى σ معلوم لدرجات المجتمع أو فى حالة σ غير معروفة واستخدمنا اختبارىTو z. وأيضًا تناولت تقدير اختبارات الفروض وكذلك فترات الثقة كمدخلين للتحقق من الفروض أو اختبار الحالة فى المجتمع. ولكن فى بعض الاحيان يستخدم معالم مثل معامل الارتباط فى المجتمع (ρ) أو النسبة (P) ونريد اختبارات فروض حولها وبالتالى يهدف الباحث إلى اختبار فروض صفرية كالآتى:

$H_0: P = a$، $H_0: \rho = a$

- a قيمة المعلم فى المجتمع
- ρ معامل الارتباط فى المجتمع
- P نسبة المجتمع (أحيانًا تكتب πpi)

وفى هذا الإطار فإن الصيغة العامة للإحصاء المستخدم:

$$= \frac{\text{إحصائية العينة} - \text{المعلم}}{\text{الخطأ المعيارى للإحصائية}}$$

وعليه فإن الصيغة العامة لفترة الثقة:

$$CI = \text{Statistic} \pm \text{critical value (SE)}$$

- SE هى الخطأ المعياري
- Statistic قيمة إحصائية الاختبار المحسوبة.
- critical value(cv) القيمة الحرجة (القيمة الجدولية).

اختبار فروض حول نسب عينة واحدة

قد لا يكون المتغير فى صورة فترية أو نسبية ولكنه يكون عبارة عن نسبة العينة وليس المتوسط ويكون السؤال: هل نسبة العينة تختلف عن نسبة المجتمع لمتغير ما؟

قضية بحثية (مأخوذة من (Haeley, 2012)

فى مجتمع محافظة ما أراد باحث التحقق من نسبة الأمية فى هذه المحافظة التى تختلف عن نسبة الأمية على مستوى الجمهورية وهى 39% وجمع عينة من 122 فردا واتضح أن نسبة الأمية فى هذه العينة 43%.

الخطوات البحثية

1. **سؤال البحث:** هل توجد فروق جوهرية بين نسبة الأمية فى محافظة ما عن نسبتها على مستوى الوطن؟

2. **فرض البحث:** توجد فروق بين النسبة على مستوى المحافظة بنظيرتها على مستوى الوطن.

3. **متغيرات البحث:** نسبة الأمية عبارة عن نسب.

4. **منهج البحث:** وصفى.

5. **النموذج الإحصائى:** نموذج المتغير الواحد وإحصاء بارامترى.

6. **الاختبار المناسب:** Z للعينات الكبيرة، 30 $\leq$ N أما إذا كانت العينة صغيرة فيفضل استخدام الاختبار اللابارامترى Binomail test .

خطوات اختبارات الفروض الصفرية

1.الفروض الإحصائية:

$$H_0 : P \text{ أو } \pi = 0.39$$

$$H_A: P \text{ أو } \pi \neq 0.39$$

على ذلك فإن الفرض البديل غير موجه ذو ذيلين.

2. الاختبار المناسب وشروطه: اختبارz ، **حيث:**

$$z = \frac{P_s - P_u}{\sqrt{\text{Sp}}}$$

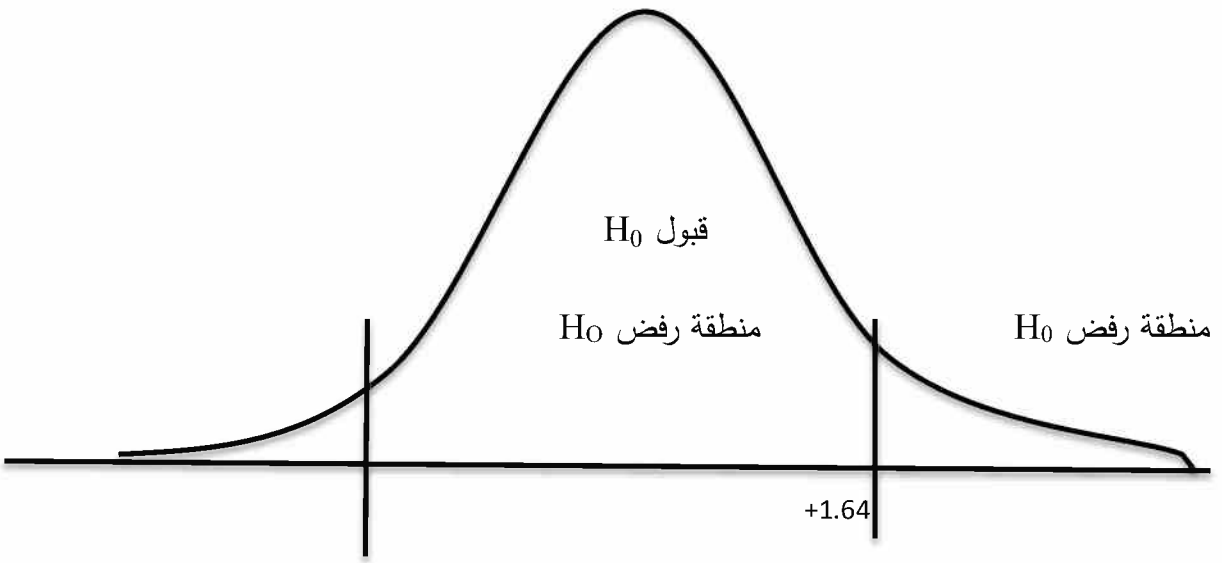
- P_s نسبة العينة، P_u نسبة المجتمع أو الفرض الصفرى.
- N حجم العينة
- Sp الخطأ المعيارى وتتحدد بالاتى:

$$S_p = \sqrt{\frac{pq}{n}}$$

- P هى نسبة العينة
- q=(1-P) نسبة المجتمع التى ليس لديه السمة(نسبة الامية).

ومسلماته هى نفس مسلمات اختبار Z.

3. الدلالة الإحصائية وقاعدة القرار: اختبر الباحث الفرض عند α=0.10 ، واختبار ذو ذيلين وبالبحث فى جداول Z يتضح أن: الحرجة Z = ± ١.٦٥



4. الحسابات: حيث: P = 0.39 ,q = 1-0.39 = 0.61, N = 122

$$z = \frac{0.43 - 0.39}{\sqrt{\frac{0.39\ X\ 0.61}{122}}} = \frac{0.04}{0.044} = 0.91$$

5. القرار والتفسير: Z المحسوبة (0.91 Z >) الحرجة (1.65)، وهو أن نسبة الأمية فى المحافظة لا تختلف عن نسبتها فى عموم البلاد.

فترتى الثقة:

$$CI\ 90 = P \pm Z\ crit\ (Sp) =$$
$$0.43 \pm\ (0.044)(1.65)$$

مثـال آخـر: قـام مركـز الوحـدة المحليـة بمسـح لتحديـد الخـدمات التـى يحتاجهـا اهـالى قريـة مـا تتضـمن 120000 مـواطن اختيـرت عينـة مـن 400 مـواطن وافتـرض فريـق البحـث أن 30% مـن سـكان المنطقـة يحتـاجون دار حضـانة للأطفـال تحـت سـن السادسـة وأشـار 88 مـن أفـراد العينـة أنهـم بحاجـة إلـى دار حضـانة، وأرادوا التحقـق مـا إذا كانـت توجـد فـروق بين نسبة المجتمع ونسبة العينة فى مدى احتياجهم لدار حضانة.

خطوات اختبارات الفروض

1. الفروض الإحصائية: $\boldsymbol{H_0 : P = 30\% = 0.30}$

$$\boldsymbol{H_A : P \neq 0.30}$$

2. الاختبار الإحصائى المناسب ومسلماته: اختبار Z وصيغته كالآتى:

$$z = \frac{P_s - P_u}{Sp}$$

$$S_p = \sqrt{\frac{pq}{\pi}}$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: حيث $\alpha = 0.01$ ، واختبار ذو ذيلين إذا: $Z\ c.v = \pm 1.96$، حيث إذا كانت:

(المحسوبة) $Z\ c.v < Z$ نرفض الفرض الصفرى.

4. الحسابات: $P = \frac{88}{400} = 0.22$

$$S_p = \sqrt{\frac{(0.22)(0.78)}{400}} = 0.02$$

$$Z = \frac{0.22 - 0.30}{0.02} = -4.00$$

5. القرار والتفسير: بما ان Z المحسوبة (-4.00 **$<$ Z** الحرجة (1.96)، **إذا** نرفض الفرض الصفرى وعليه توجد فروق ذات دلالة إحصائية بين نسبة المجتمع ونسبة العينة.

القرار فى ضوء فترتى الثقة

بما أن 0.05 =α إذا فترات الثقة تكون 95% ($1-0.05$) ويتم بناءها حول نسبة العينة وعليه:

$$CI95 = P \mp 1.96\ (Sp)$$

$$= 0.22 \mp 1.96\ (0.02)$$

$$= 0.22 \mp 0.04$$

$$= (0.18 \quad , \quad 0.26)$$

الأعلى , الأدنى

وعليه فإن قيمة Z المحسوبة أو القيمة المتضمنة فى الفرض الصفرى (0.30) تقع خارج هذه الفترة وعليه نرفض الفرض الصفرى وبالتأمل فى بيانات العينة نجد أن نسبة العينة أقل من نسبة المجتمع وهذا يعنى أن 95% ثقة للفترة من 0.18 وحتى 0.26 تتضمن نسبة المجتمع.

اختبار فروض حول معامل الارتباط فى المجتمع

تحوير فيشر Fisher's z Transformation

يهدف اختبار تحوير أو تعديل فيشر Z إلى اختبار فروض حول معامل ارتباط بين متغيرين فى المجتمع ومعرفة ماذا كانت الارتباط فى العينة يختلف اختلافًا جوهريًا عن معامل الارتباط فى المجتمع.

اختبارات الفروض لقضيه بحثية (Hinkle et al., 1994)

أراد مجموعة من علماء الاجتماع دراسة العلاقة بين دخل الأسرة واتجاهاتهم نحو تربية الأطفال ووجدوا أن التراث البحثى توصل فى معظمه أن العلاقة تدور حول 0.65 وأراد اختبار ما إذا كانت العلاقة بين المتغيرين لدى أفراد العينة المكونة من 30 عائلة تختلف عن العلاقة الموجودة بين المتغيرين فى الدراسات السابقة.

الخطوات البحثية

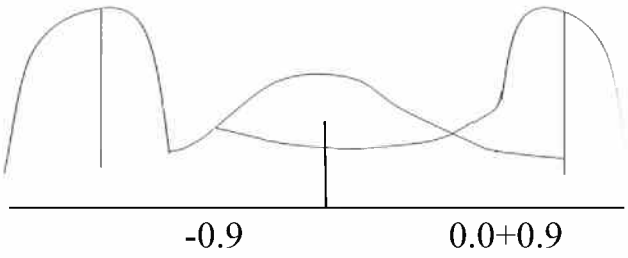
1. **سؤال البحث:** هل توجد تختلف العلاقة بين دخل الاسرة واتجاهاتهم نحو تربية الأطفال لدى العينة وعن العلاقة بين المتغيرين فى المجتمع 0.65 ؟

2. **فرض البحث:** توجد فروق فى العلاقة بين دخل الأسرة واتجاهاتهم لدى العينة عن العلاقة بينهما فى المجتمع. أو تختلف العلاقة فى العينة عن 0.65.

3. **متغيرات البحث:** يوجد متغيرين هما الدخل: نسبى – متصل، والاتجاه: فترى – متصل

4. **منهج البحث:** منهج ارتباطى.

5. **النموذج الإحصائى:** هو البسيط والإحصاء بارا مترى ولأن حسابات معامل الارتباط فى طبيعته وصفية. ولكن اختبارات الفروض قائمة حول معلم معين لعينة واحدة، فالإحصاء هو لعينة الواحدة حول معلم واحد فهو إحصاء استدلالى لمتغير واحد.



الشكل (1.17): التوزيع العينى لمعامل الارتباط عندما (.90 -, 0.0, .90)=r

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية: $H_0: \rho = 0.65$

$H_A: \rho \neq 0.65$

والاختبار ذو ذيلين.

2. الاختبار الإحصائى المناسب و مسلماته: يقدر من الصيغة الآتية:

$$\frac{Z_r - Z_\rho}{S_{Zr}}$$

- Z_ρ قيمة تحوير Z لمعامل الارتباط فى المجتمع
- Z_r قيمة تحوير Z لمعامل الارتباط فى العينة
- S_{Zr} الخطا المعيارى لـ Z

ولتحوير معامل الارتباط إلى Z كالآتى (تحوير فيشر Z):

$$Zr = (\frac{1}{2}) \log e \left(\frac{1+r}{1-r} \right)$$

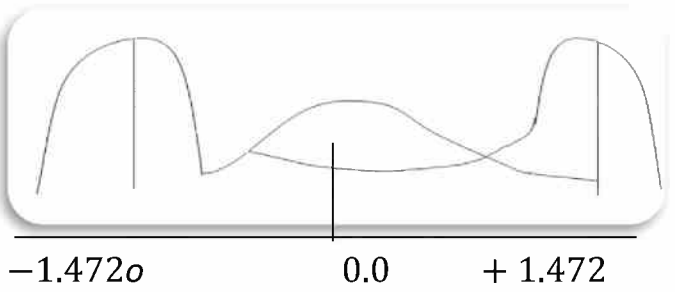
3. **مستوى الدلالة الإحصائية والتوزيع العينى وقاعدة القرار:** اختبر الباحث هذا الفرض عند α =.10 والتوزيع العينى لمعامل الارتباط يقدر مثل التوزيع العينى للمتوسط وهو التوزيع العينى لمعاملات الارتباط المحسوبة من عينات من المجتمع ويتغير التوزيع العينى ووفقًا لقيمة معامل الارتباط واشارته وكلما زاد ومعامل الارتباط فإن التوزيع العينى يكون ملتوى ناحية اليمين للمعاملات السالبة وناحية اليسار للمعاملات الموجبة وعندما يكون قريب من الصفر فإن التوزيع العينى يكون

تقريبًا اعتدالى، وتتحسن الاعتدالية بزيادة حجم العينة ولان التوزيع العينى يكون ملتوى كلما زادت قيمة معامل الارتباط فإنه من غير المناسب استخدام التوزيع الاعتدالى لاختبارات الفروض حول ρ إلا إذا كانت ρ =0 وهذا مما أدى بـ Fisher بإجراء تعديل Z

وحيث إن الخطأ المعيارى لـ Z:

$$S_{zr} = \sqrt{\frac{1}{n-3}}$$

ولذلك فإن التوزيع العينى لمعاملات الارتباط المحورة أو المعدلة كالآتى:



الشكل(2.17) : التوزيع العينى لمعامل الارتباط المعدل Zr لـ ρ- 0.95,0.0,+0.90

ولتحديد قيمة Zr الجدولية فإنه يتم حساب معامل الارتباط لبيانات العينة وبفرض أن قيمته 0.61 =r ، ويجب تحويل قيمة ρ = .65 (فى المجتمع) و r = .61(فى العينة) إلى تحوير Z لاختبار فروض حولها. بالبحث فى جدول (انظر ملحق جدول تحوير فيشر) فإن تحوير قيمة ρ عن طريق قراءة العمود r حتى نصل إلى 65. فإن قيمة فيشر المقابلة: Zρ = 0 .775

وكذلك بالنسبة لـ 0.61 فتكون قيمة فيشر المقابلة لها: Z_r= 0.759

وعليه ρ =0.65 ⟶ Z ρ = 0.775

r= .61 ⟶ Z_r= 0.759

وبتقدير الخطا المعيارى:

$$S_{zr} = \sqrt{\frac{1}{n-3}}$$

$$= \sqrt{\frac{1}{30-3}} = 0.192$$

وعلى ذلك فإنه لـα =0.10 واختبار ذو ذيلين فإن: Z = ±1.645 (الجدولية)

وإذا كانت Z المحسوبة ≤ Zالحرجة نرفض Ho.

4. الحسابات:

$$Z=\frac{Z_r - Z_\rho}{S_{zr}}$$

$$= \frac{0.759\text{-}0.775}{0.192} = \text{-}\ 0.344$$

5. **القرار والتفسير:** **Z المحسوبة $|-0.344|$ < Z الحرجة $|1.645|$** ، وعليه نقبل الفرض الصفرى ويفشل الباحث فى رفض H0.

بنـــاء فتـــرات الثقـــة: عنـد رفـض H0 أو قبولـه فمـن الأفضـل بنـاء فتـرات الثقـة لأنهـا تكون أكثر معلوماتية وبما أن α =0.10 فان:

$$=1\text{-}0.1= 0.90$$

وذلك يتم بناء 90% فترة الثقة:

$$CI_{90}\ (\text{الأعلى}) = Z_r + 1.645\ (S_{zr})$$

$$= 0.759 + 10645\ (0.192) = 1.025$$

$$CI_{90}\ (\text{الادنى}) = 0.759\text{-}\ 1.645\ (0.192) = 0.393$$

وبالبحث فى جدول Z فيشر عن قيم معامل الارتباط المناظرة لقيمتى فيشر 1.025و 0.393 وهما كالآتى:

$$CI_{90} = (0.374, 0.773)$$

وعليه فإن الباحث فشل فى رفض H_0 حيث لا فروق بين معامل الارتباط فى العينة عن معامل الارتباط فى المجتمع.

وفيما يلى معلومات عن اختبارات الفروض لعينة واحدة كما طرحها Hinkle et al. (1994) :

الفرض الصفرى	المعلم	الإحصائية	SE	الاختبار	df	التوزيع العينى
$H_0 : \mu = 0$	μ	$\bar{x}$	$s\bar{x} = \frac{s}{\sqrt{n}}$	$\frac{\bar{x} - \mu}{s\bar{x}}$	N-1	توزيع T
$H_0 : \mu = 0$	μ	$\bar{x}$	$\sigma\bar{x} = \frac{\sigma}{\sqrt{N}}$	$\frac{\bar{x} - \mu}{\sigma\bar{x}}$	———	Z
$H_0: \rho = q$	$z\rho$	r	$s_{zr}\sqrt{\frac{1}{N-3}}$	$\frac{zr - zp}{S_{zr}}$	———	اعتدالى
$H_0: \rho = 0$	ρ	r	--	$T = r\sqrt{\frac{N-2}{1-r^2}}$	N-2	T
$H_0: P=q$	Pu	Ps	$sp = \sqrt{\frac{pq}{N}}$	$T = \sqrt{\frac{N-2}{S_p}}$	———	اعتدالى

$\bar{x}$: المتوسط

وعرض Hinkle et al. (1994) ملخص (فترات الثقة) ولتقدير الفترة للمعالم لعينة واحدة كالآتى:

المعلم المقدر	الإحصائية	فترة الثقة
μ	$\bar{x}$	$\bar{x} \mp (T_{c.v})(\frac{s}{\sqrt{n}})$
Z	zr	$zr \mp z_{c.v}(\sqrt{\frac{1}{N-3}})$
P	P	$P \mp z_{c.v}(\sqrt{\frac{pq}{N}})$

حيث c.v القيمة الحرجة (Critical value).

الفصل الثامن عشر

اختبار T المرتبطة أو المعتمدة

Correlated or Dependent samples T test

فى الفصول السابقة تناولت اختبارات الفروض وتقدير فترات الثقة حول متوسط مجتمع من خلال عينة واحدة لمتغير واحد لاختبارى Z أو T وكذلك اختبارات فروض حول النسبة وحول معاملات الارتباط فى المجتمع وفى هذا الفصل نتناول اختبارات الفروض وتقدير معالم عينتين Two samples حيث يتعامل الباحثون فى العلوم النفسية والسلوكية مع مواقف بحثية تتضمن عينيتين مستقلتين وهذه النوعية من المواقف البحثية تتولد من تصميمات بحثية مختلفة سواء كانت تجريبية أو شبة تجريبية أو سببية مقارنة وغيرها، وعلى ذلك فهذه التصميمات تولد بيانات لمجموعتين ويكون هدف الباحث المقارنة بين درجات أو بيانات المجموعتين على متغير تابع كالتحصيل مثلًا. وغالبًا تتم المقارنة بين المجموعتين فى متوسط الدرجات على المتغير التابع فعلى سبيل المثال نريد اختبار فرض حول الفروق بين الذكور والإناث فى التحصيل أو فى الابتكارية أو ربما يوجد موقف بحث تجريبى تأخذ المجموعة التجريبية معالجة اما المجموعة الأخرى الضابطة لا تأخذ معالجة أو تأخذ معالجة تقليدية.

أو نريد اختبار ما إذا كانت توجد فروق بين متوسط أداء المجموعتين فى القياسات البعدية للمتغير التابع ولو كانت المجموعتين بينهما ارتباط كائن بمعنى نفس المجموعة قبلى وبعدى أو المجموعة الأولى تماثل أو تناظر أو هى نفسها المجموعة الثانية فهذه النوعية من البيانات تتبع من عينات مترابطة Correlated samples أو هذه النوعية من البيانات تتولد من تصميم القياسات المتكررة Repeated measures design وهذه العينات المترابطة توجد فى صورة ثلاث حالات:

الحالة الأولى: توجد مجموعة وحيدة من الأفراد ويتم تطبيق عليها قياسات تتبعية مثل بيانات التصميم التجريبى قياس قبلى وقياس بعدى Pre and post design وبالتالى يوجد ارتباط منطقى فى البيانات لأن كل فرد يأخذ درجة فى الاختبار القبلى

ويأخذ درجة أيضًا فى الاختبار البعدى وعلى ذلك نفس الاشخاص تطبق عليهم قياسات متكررة .

الحالة الثانية: المزاوجة Matched – pairs design: بمعنى أن كل فرد فى المجموعة الأولى يناظر فرد آخر له نفس الخصائص فى المجموعة الثانية ويمكن أن تحدث المزاوجة فى حدود وخصائص معينة مثل نسبة الذكاء IQ أو الطول والحالة الاقتصادية وغيرها ومتغير المزاوجة نسبة الذكاء IQ ليس متغير تابع ويتعقد الموقف البحثى إذا كان الباحث مهتم بجمع بيانات لأكثر من متغيرين، وتحدث المزاوجة فى التصميمات التجريبية عند التكافؤ بين التجريبية والضابطة.

الحالة الثالثة: عندما توجد توأم بيولوجية خاصة المتماثلة حيث يوضع فرد فى مجموعة ما والفرد الآخر فى المجموعة الأخرى وهنا يوجد تماثل جينى.

ويكون التصميم البحثى كالآتى:

قياسات قبلية	قياسات بعدية
درجات المتغير التابع للأفراد	درجات المتغير التابع لنفس الأفراد
معالجة 1	معالجة 2
n_1	n_1
n_2	n_2
n_3	n_3

وعلى ذلك فإن التصميم البحثى يعطى مجموعتين من البيانات تأتى من نفس المجموعة من الأفراد وهذا التصميم يسمى تصميم القياسات المتكررة أو تصميم داخل الأخرى Within – subject design ويشار للعينات المرتبطة بأزواج العينات Matched samples , paired samples أو العينات المعتمدة within samples أو Dependent samples . ولكن إذا كان درجات المجموعتين المراد مقارنتهما لا تقع فى أى من الحالات الثلاثة (قبلى – بعدى) أو (الازدواج المتماثلة) أو (التوائم) فإن البيانات تكون لمجموعتين مستقلتين ويطلق عليها عينات مستقلة Independent

samples بمعنى أن أفراد المجموعة الأولى لا ترتبط بأفراد المجموعة الثانية بل يوجد استقلالية هو الاكثر شيوعًا واستخداما فى الدراسات البحثية وهذه النوعية من البيانات تتولد من تصميم بحثى مستقل أو تصميم بين الأفراد Between subject design ويكون التصميم البحثى كالآتى:

المتغير المستقل

ذكر أو تجريبية	انثى أو ضابطة
درجات المتغير التابع	درجات المتغير التابع

ويطلق البعض على هذه النوعية من العينات بالعينات غير المتزاوجة Unpaired samples أو عينات غير متماثلة Unmatched samples أو بالعينات غير المرتبطة Uncorrelated samples.

حقيقة فى الفصول السابقة تناولنا تصميم العينة الواحدة قياسات بعدية ويتم تحليلها باستخدام z ,T ، والموقف البحثى الذى يتضمن تجربة بمعالجتين أكثر ميزة من تصميم العينة الواحدة وذلك لأن تصميم تجربة العينة الواحدة يتطلب معرفة معلم واحد للمجتمع (μ) على الأقل ولكن فى معظم المواقف العملية هذا غير متاح ولذلك فإن تصميم المعالجتين أو المجموعتين لا يتطلب تحديد معلم المجتمع عند اختبار الفروض. وفى تصميم المجموعتين أو المعالجتين يكون التفسير أكثر دقة وعليه سنعرض فى الجزء التالى:

اختبار T المرتبطة أو المعتمدة

الهدف: يهدف إلى تقيم ما إذا كان متوسط الفروق بين متغيرين للعينات المرتبطة يختلف عن الصفر فى المجتمع أو ما إذا كانت توجد فروق بين القياسات القبلية والقياسات البعدية لمتغير تابع ما ويستخدم كما سبق عرضه فى تصميمات القياسات المتكررة حيث يحصل كل فرد على درجات قبل المعالجة وبعد المعالجة وميزة هذا

التصـميم هـو ان نفـس الأفـراد يتعرضـوا لـنفس المعالجـات. حيـث إن تصـميم القياسـات المتكررة يمكن أن يكون كالآتى:

- تصميم المجموعة الواحدة وقياس قبلى وبعد:

$O_1 \quad X \quad O_2$

- تصميم المجموعة الواحد التى تتعرض لأكثر من معالجة:

$O_1 \quad X_1 \quad O_2 \quad X_2 \quad O_3 \quad X_3 \quad O_4$

مقارنة بين T لعينة واحدة و T لمجموعات مرتبطة:

نلاحـظ أن اختبـار T لعينـات مرتبطـة قـائم علـى متوسـطات فـروق الـدرجات difference scores ولـيس متوسـطات الـدرجات الخـام كمـا هـو الحـال فـى T لعينـه واحـدة. وفيمـا يلـى مقارنـة بـين معـادلات حسـاب T لعينـة واحـدة T لعينـات مرتبطـة، فاللعينة الواحدة:

$$T = \frac{\overline{x} - \mu}{(S_{\bar{x}})}$$

$$T = \frac{\overline{x} - \mu}{\sqrt{\frac{SS}{N(N-1)}}}$$

$$SS = \sum x^2 - \frac{(\sum x)^2}{N}$$

T المرتبطة:

$$T = \frac{\overline{D} - \mu_D}{S_D / \sqrt{N}}$$

$$SS_D = \sum D^2 - \frac{(\sum D)^2}{N}$$

- D الفرق بين الدرجات، $\overline{D}$ متوسط الفروق بين الدرجات
- μ_D متوسط الفروق بين الدرجات فى المجتمع

- S_D الانحراف المعيارى لفروق الدرجات فى العينة.
- Nعدد فروق الدرجات (حجم العينة).

اختبـارات الفـروض لقضـيه بحثيـة (Pagano, 2013): اجـرى باحـث فـى مجـال علـم الـنفس والتربيـة دراسـة لاختبـار مـا إذا كـان إشـراف المعلـم واتبـاع تعليماتـه يـؤثر علـى الوقـت الـذى يقضـيه الطالـب فـى القـراءة الجيـدة لموضـوع مـا فقـام بقيـاس قبلـى للوقـت المسـتغرق لقطعـة قـراءة وثـم قـدم تعليمـات وقواعـد للقـراءة الجيـدة ثـم قـاس بعـدى وحسـب الوقـت المسـتغرق فـى القـراءة القبليـة والبعديـة وكانـت النتـائج كـالآتى (وقـت القـراءة بالثانية):

الفرد	قبلى	بعدى	D	*D*
1	220	210	220-210=10	100
2	245	220	245-220=25	625
3	215	195	20	400
4	260	265	-5	25
5	300	275	25	625
6	280	290	-10	100
7	250	220	30	900
8	310	285	25	625
∑			120	3400

وأراد اختبـار مـا إذا كانـت توجـد فـروق فـى وقـت القـراءة لموضـوع مـا فـى حضـور وغياب إرشادات المعلم عند مستوى دلاله إحصائية 0.05؟

خطوات اختبارات الفروض: لاحظ أن هذه النوعية من الدراسات تقع تحت تصميم المجموعة الواحدة وقياس قبلى وبعدى.

1.الفروض الإحصائية:

$$H_0 : \mu_D(\mu_1 - \mu_2) = 0$$
$$H_A : \mu_D(\mu_1 - \mu_2) \neq 0$$

توجد فروق فى متوسطات الوقت المستغرق فى القراءة فى حضور وفى غياب تعليمات المعلم والاختبار ذو ذيلين.

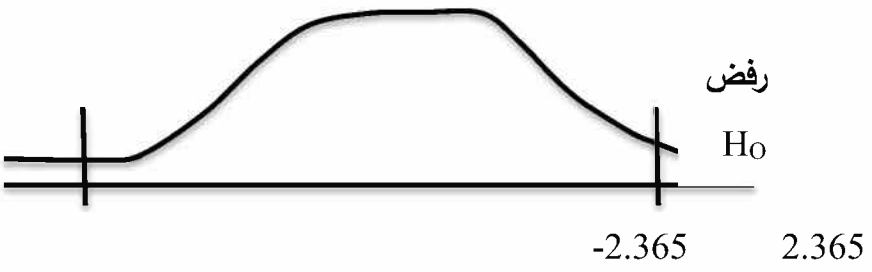
2. الاختبار المناسب ومسلماته:

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}}$$

$$df = N - 1 = 8 - 1 = 7$$

3. الدلالة الإحصائية وقاعدة القرار: إذا كانت $\alpha = 0.05$ ولاختبار ذو ذيلين و $df = 7,$

بالبحث فى جداول T فإن قيمة T الحرجة = ± 2.365



وإذا كانت T المحسوبة $\leq$ T الجدولية $|2.361|$ نرفض H_0

4.الحسابات:

$$\bar{D}(M_D) = \frac{\sum D}{N} = \frac{120}{8} = 15$$

$$S_{\bar{D}} = \sqrt{\frac{SS_D}{n-1}}$$

حيث: $$SS_D = \sum D^2 - \frac{(\sum D)^2}{N}$$

$$SS_D = 3400 - \frac{(120)^2}{8} = 1600$$

$$S_D = \sqrt{\frac{1600}{8-1}} = \sqrt{228.57} = 15.12$$

وعليه فإن الخطأ المعيارى لفروق الدرجات هى:

$$S_{\bar{D}} = 15.12/\sqrt{8} = 5.35$$

وبالتالى :

$$T = \frac{\bar{D}-\mu_D}{S_{\bar{D}}} = \frac{D^{-}-0}{\sqrt{SS_D/N)}} = 15/(5.35)=2.80$$

5. القـــرار والتفســـير: حيـــث**T المحســـوبة** (2.80) > **T** الحرجة **2.365**) ، وعليـــه نـرفض H_0 وبالتـالى توجـد فـروق ذات دلالـة إحصـائية بـين زمـن القـراءة فـى حضـور المعلـم وإتبـاع تعليماتـه وزمـن القـراءة فـى عـدم وجـود المعلـم، ولمعرفـة اتجاهـه فهـى لصـالح المتوسط الاكبر.

6. حجم التأثير: يُقدر حجم التأثير خلال المؤشرات الآتية:

- مؤشر Cohen (1988):

$$d = \frac{\bar{D}}{S_D} = \frac{15}{15.12} = 0.99$$

وعلى ذلك فإن الوقت المستغرق فى القراءة يساوى0.99 وحدة انحراف معيارى أكبر من الوقت المستغرق فى عدم وجود تعليمات وهذا حجم تأثير كبير.

- مؤشر مربع إيتا η^2 أو التباين المفسر:

$$\eta^2 = \frac{T^2}{T^2 + df} = \frac{(2.804)^2}{(2.804)^2 + 7} = 0.53(53\%)$$

ويعبر عنها بالصيغة الآتية:

$$r^2 = \frac{T^2}{T^2 + df}$$

إذًا إرشـادات المعلـم أو البرنـامج فسـرت 53% مـن تبـاين الوقـت المسـتغرق فـى القـراءة، بمعنى أن البرنامج فسر 53% من تباين الوقت.

- مؤشر اوميجا ω^2:

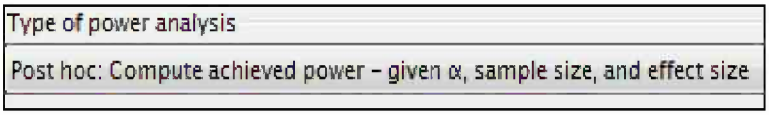
$$\omega^2 = \frac{T^2 - 1}{(T^2 + df)} = \frac{(2.804)^2 - 1}{(2.804)^2 + 7} = 0.46$$

إذًا 46% من تباين الوقت يرجع إلى التعليمات.

7.القـوة الإحصـائية: لتقـدير القـوة الإحصـائية للمثـال السـابق باسـتخدام برنـامج G-Power اتبع الآتى:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

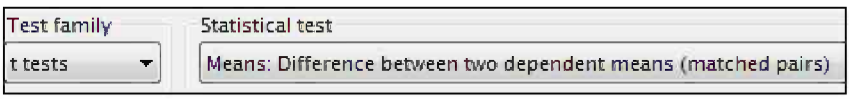
2. تحت Type of power analysis اختر:



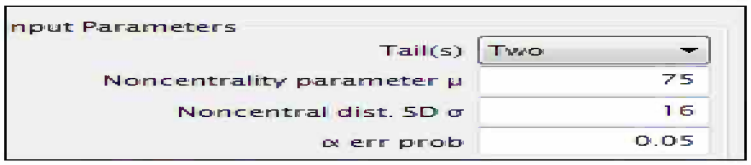
لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

3. تحت Test family اختر t– tests

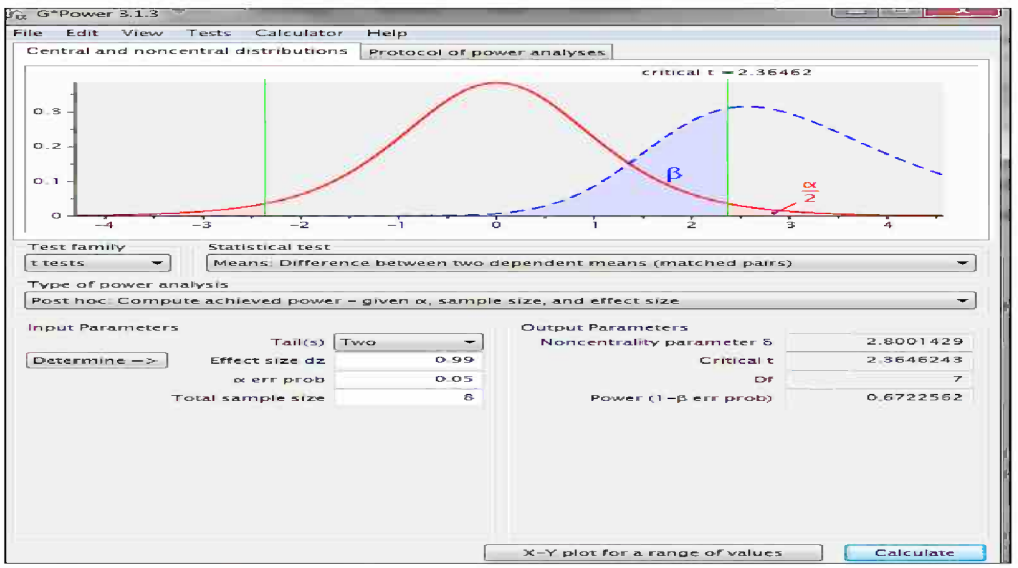
4. تحت Statistical test اختر:



5. أدخل المعالم Input parameters:



6. اضغط Calculated تظهر شاشة الناتج كالآتى:



يتضح أن القوة الإحصائية =0.67 وهو مستوى قوة غير مرغوب وقد يرجع إلى صغر حجم العينة.

القرار فى ضوء فترات الثقة

بما ان $\alpha = 0.05$ إذا فترات الثقة لـ $95\%(1-0.05)$:

$$CI_{95\%} = \overline{D} \pm T_{c.v}(S_D-) = 15 \pm 2.365(5.35)$$

$$(2.360, 27.6394)$$

وعليه فإن القيمة المتضمنة فى الفرض الصفرى 0.0 وقعت خارج مدى فترات الثقة بالتالى نرفض H_0.

كتابة نتائج اختبار T المرتبطة وفقا لـAPA

تم إجراء الاختبار T المرتبطة القيم بما إذا كانت توجد فروق فى الاوزان قبل وبعد البرنامج وأوضحت النتائج ان متوسط فروق الدرجات $\overline{D} = 7.26$ ومتوسط درجات الاختبار القبلى $5.02 = S_D$, $\bar{x}=83.23$ ومتوسط درجات الاختبار البعدى $\bar{x} = 90.49, S_D=8.40$ وحيث $p < 0.05$:$T(16)= 4.18$ هى فروق دالة إحصائيًا وان فترات الثقة لمتوسطات فروق الدرجات كانت $cI_{95}=(3.57,10.95)$

$$\ldots\ldots\ldots, T(7) = 2.804, p < 0.05\ (d = 0.99)$$

$$\ldots\ldots\ldots, T_{(7)} = 2.804, p < 0.05(\omega^2 = 0.46), r^2 = 0.53$$

وعموما مؤشر ω^2 أكثر دقة من مؤشر μ^2

العوامل التى تؤثر على اختبار T المرتبطة

توجد العديد من العوامل التى تؤثر على نواتج اختبارات الفروض لاختبار T وهى:

1. **حجم فروق الدرجات:** فالفروق الكبيرة بين الدرجات تزيد من احتمالية رفض Ho و كذلك يزيد من حجم التأثير.

2. **تشتت أو اختلاف الدرجات وحجم العينة:** يؤثروا على حجم الخطأ المعيارى فى مقام معادلة T، فالخطأ المعيارى يتناسب عكسيًا مع حجم العينة، فالعينات الكبيرة تقود إلى خطأ معيارى صغير، وهذا مرتبط بتباين العينة حيث التباينات الكبيرة تقود إلى خطأ معيارى أكبر وكذلك فإن العينات الكبيرة ينتج عنها قيمة كبيرة لـ T وهذا يزيد

من احتمال رفض H_0. والتباينات لها تأثير كبير على حجم التأثير، فالتباينات العالية ينتج عنها قيم منخفضة لحجم وعلى العكس حجم العينة ليس له تأثير على مؤشر d ولكن تأثيرها محدود على مؤشر η^2 (Pagano, 2013) .

اختبار T المرتبطة والاختبار اللابارامترى الإشارة Sign

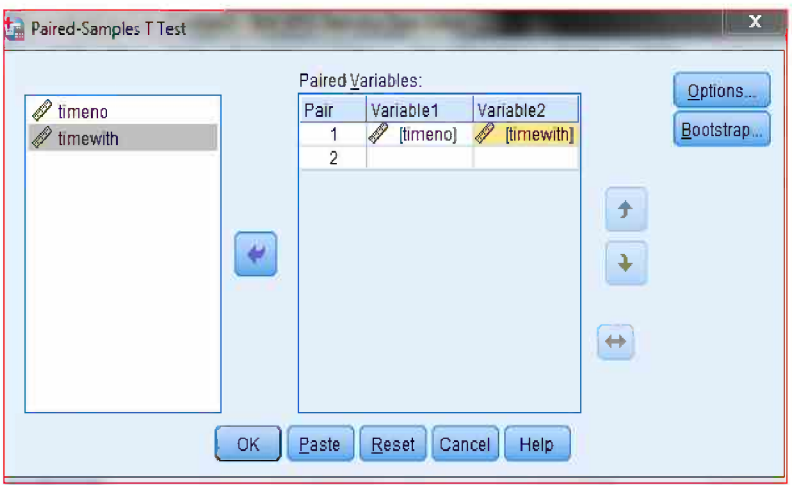
من الممكن إجراء اختبارات الفروض للقضيتين السابقتين التى تم تناولهما باستخدام اختبار T من خلال الاختبار اللابارامترى الإشارة Sign. ولكن اختبار T المرتبطة أكثر قوة إحصائيًا من اختبار الإشارة حيث تقريبًا تكون النتائج نفسها إذا توافرت الاعتدالية.

تنفيذ اختبار T المرتبطة فى SPSS

اولًا: إدخال البيانات: 1.اضغط على Variable view ثم اكتب مسمى المتغيرات فى عمود Name كالآتى: الوقت المستغرق للقراءة مع تعليمات المعلم Timewith ، وقت بدون تعليمات المعلم: Timeno

2. اضغط على Data view يوجد المتغيرين فى ملف البيانات فى عمودين ابدأ فى إدخال البيانات.

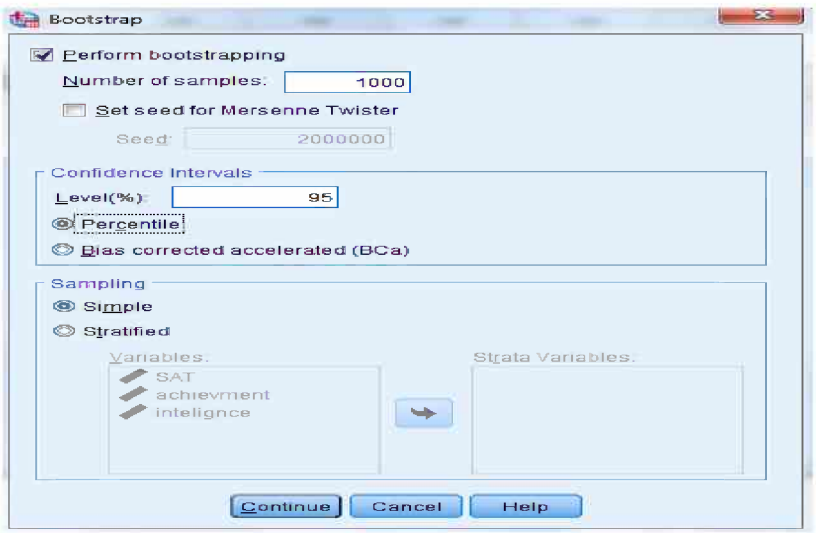
ثانيًا : تنفيذ الامر:1. اضغط Analyze → Compare Means → Paired samples T test تظهر الشاشة الآتية:



2. انقل Timewith إلى مربع Variable 1 ويمكن الضغط على المتغيرين معًا من خلال الضغط على مفتاح ctr l ونقلهما إلى مربع Variable

3. انقل Timeno إلى مربع Variable 2

4. اضغط على Bootstrap تظهر النافذة الآتية:



5. اضغط Perform bootstrapping :

- حدد عدد العينات التى ترغب فى توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 1000 مثلًا وهى Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيطة.

6. اضغط Contiue ثم OK

4. اضغط OK

```
T-TEST PAIRS=timeno WITH timewith (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.
```

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	timeno	260.0000	8	34.74396	12.28385
	timewith	245.0000	8	37.60699	13.29608

ثالثًا: المخرج : اعطى وصف إحصائى للزمن مع وجود تعليمات وبدون:

- متوسط زمن القراءة بدون تعليمات Mean = 260 ، الانحراف المعياري=34.74
- متوسط الزمن مع وجود تعليمات Mean = 245 وكذلك الانحراف المعياري
- الخطأ المعيارى: $SE = \frac{S}{N} = \frac{34.74}{\sqrt{8}} = 12.28$

واعطى معامل ارتباط بيرسون بين زوج المتغيرات وهو الزمن مع وجود تعليمات والزمن بدون تعليمات لأن التصميم قياسات متكررة:

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	timeno & timewith	8	.916	.001

$r = 0.916 \ , \ P = 0.001 < 0.05$

وهو معامل ارتباط كبير ودال إحصائيًا. ثم اعطى البرنامج أهم جدول Paired samples test:

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	timeno - timewith	15.00000	15.11858	5.34522	2.36055	27.63945	2.806	7	.026

والفرق بين المتوسطين:

$$= 260.00 - 245.0 = 15$$

- الانحراف المعيارى للفروق بين الدرجات: $SD = 15.11$
- والخطأ المعيارى للمتوسط: $SD = \frac{15.11}{\sqrt{8}} = 5.35$ وهذه مقام قانون حساب T
- قيمة T المحسوبة = 2.806

$$= \frac{\text{Mean } (\bar{D})}{S_{\bar{D}}} = \frac{15}{5.34} = 2.806$$

- Df درجات الحرية : N-1=7، القيمة الاحتمالية (2 tailed) Sig. = 0.026

وعليه يوجد 2.6 % فرصة فى ان قيمة T تكون أكبر من القيمة المفترضة فى الفرض الصفرى، وعلى ذلك 0.05 > 0.026 وبالتالى نرفض H_0 بمعنى توجد فروق ذات دلالة إحصائية عند 0.05 بين متوسط الزمن المستغرق فى القراءة مع تعليمات المعلم و متوسط الزمن المستغرق فى القراءة بدون تعليمات المعلم، والدلالة تكون فى صالح المتوسط الأعلى حيث الوقت المستغرق بدون تعليمات المعلم.

- أخيرًا 95% فترة الثقة للفروق بين المتوسطات Mean difference فتخيل لو اخذنا 100 عينة من المجتمع وقدرنا فروق الدرجات وحسبنا متوسط الفروق D وحسبنا فترة الثقة لهذا المتوسط فهو يخبرنا بالحدود الممكنة التى يمكن أن يقع فيها هذا المتوسط وهى:(2.360, 27.6394)CI95، وعليه فإن قيمة متوسط الفرق تقع فى هذا المدى وأن القيمة المفترضة فى الفرض الصفرى = 0.0 لا تقع فى هذا المدى وعليه نرفض H_0.

نتائج البوتستراب:

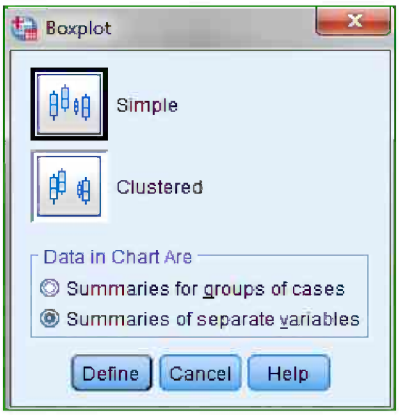
Bootstrap for Paired Samples Test

			Bootstrap[a]				
						95% Confidence Interval	
		Mean	Bias	Std. Error	Sig. (2-tailed)	Lower	Upper
Pair 1	timeno - timewith	15.00000	.15390[b]	5.17296[b]	.087[b]	3.75000[b]	25.00000[b]

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

b. Based on 999 samples

الملاحظ أن فترات الثقة هى المتولدة من 1000 عينة وهى لا تتضمن القيمة المتولدة فى الفرض الصفرى وعليه توجد فروق دالة إحصائيًا وفترات الثقة اضيف من فترات الثقة للعينة الوحيدة مما يدل على دقة التقديرات.

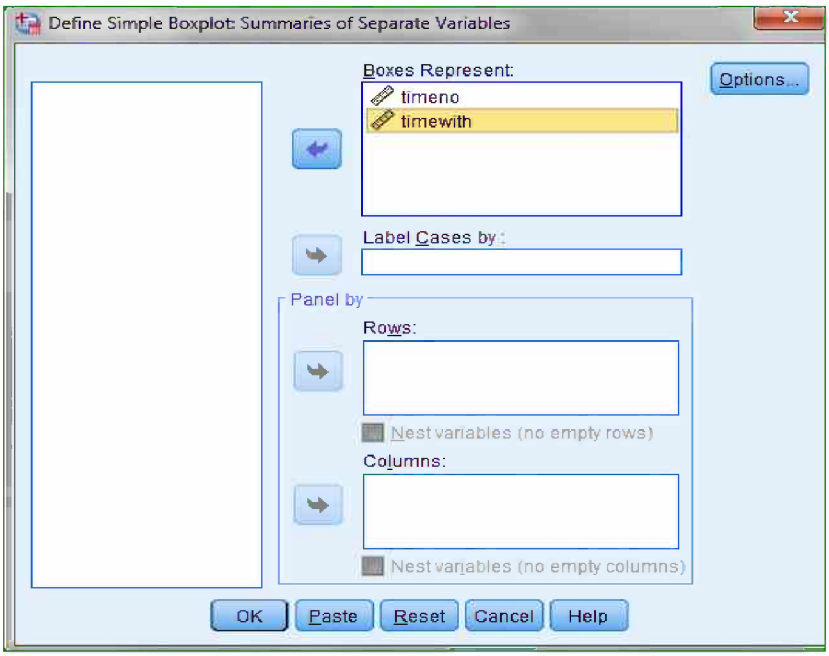


عرض النتائج بيانيًا فى ضوء Boxplot كالآتى:

1. Graphs→Legacy Dialogs →Boxplot يعطى الشاشة الآتية:

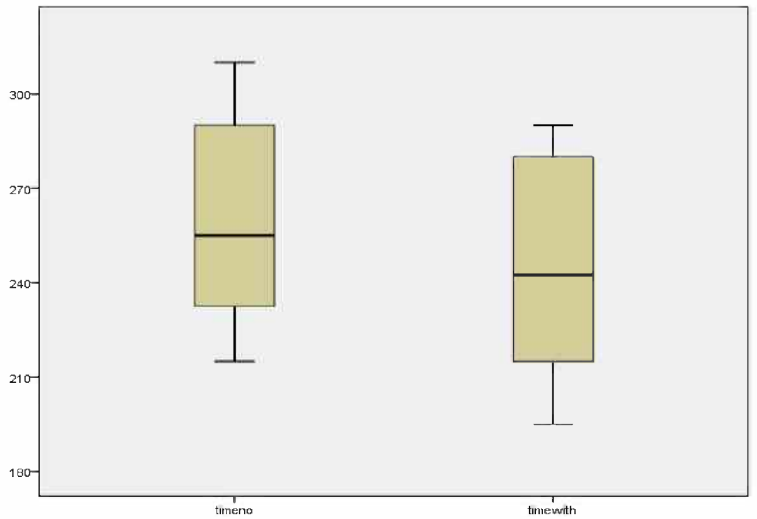
2. اضغط Simple، ثم اضغط Summaries of Separate Variables

3. اضغط Define:



5 . انقل المتغيرين إلى مربع Boxes Represent

6 . اضغط Ok يعطى الشكل الآتى:



وكما هو واضح من الشكل وسيط زمن القراءة مع تعليمات المعلم أقل من وسيط الزمن بدون تعليمات المعلم (الخط وسط الصندوق) كما هو واضح أن الإرباعى الأدنى والأعلى لزمن القراءة بدون تعليمات أكبر من نظيرتها مع تعليمات المعلم.

الفصل التاسع عشر

اختبار T المستقلة

Independent-Samples T test

يعتبر اختبار T أشهر اختبارات الفروق المستخدمة لتقيم ما إذا كانت توجد فروق بين متوسط مجتمع مجموعتين مستقلتين، وكل فرد فى هذا الاختبار لا بد أن يكون له درجات على المتغيرين أحدهما متغير تصنيفى اسمى مستقل مثل الجنس أو المجموعة التجريبية والمجموعة الضابطة والآخر تابع متصل فترى على الأقل مثل التحصيل والابتكارية وغيرها.

تطبيقات اختبار T المستقلة

- الدراسات التجريبية
- الدراسات شبه التجريبية
- الدراسات السببية المقارنة

اختبار T المستقلة: تباين المجتمع الأول = تباين المجتمع الثانى $\sigma_2^2 = \sigma_1^2$

اختبارات الفروض لقضية بحثية Hinkle et al.(1994)

اهتم معلم بدراسة الفروق فى تحصيل الرياضيات لطلاب الصف الثالث الابتدائى الذين يتعلموا تحت بيئتين مختلفتين، فتلقت العينة الأولى بيئة عالية التنظيم والتخطيط من حيث التدريس وفى حين تلقت العينة الثانية بيئة أقل تنظيمًا وتخطيطًا من حيث معينات التدريس، وبعد ستة أسابيع من التدريس الحصول على الدرجات التحصيلية للعينتين، واراد المعلم اختبار ما إذا كانت توجد فروق فى تحصيل الرياضيات بين البيئتين؟

الخطوات البحثية

1. **سؤال البحث:** هل يختلف تحصيل تلاميذ العينة الأكثر تنظمًا فى التدريس عن تحصيل العينة الأخرى الأقل تنظيمًا التى تناولت تدريسيًا فى بيئة أقل تنظيمًا؟ أو هل للبيئة التدريسية أثر فى تحصيل التلاميذ؟

2. **متغيرات البحث:** البيئة التدريسية: مستقل– اسمي– تصنيفى، تحصيل الرياضيات: تابع – فترى – متصل.

3. **فرض البحث:** توجد فروق فى تحصيل الرياضيات بين تلاميذ البيئة الاكثر تنظيمًا وتلاميذ البيئة الأقل تنظيمًا.

4. **تصميم البحث:** باستخدام اختبار T فى التصميمات البحثية التجريبية وشبه التجريبية والسببية المقارنة، وفى المثال السابق فإن الدراسة من النوع شبه التجريبى.

5. **النموذج الإحصائى:** بسيط حيث يوجد متغيرين وعليه فإحصاء النموذج البسيط Bivariate statistics هو إحصاء بارامترى و**الاختبار المناسب:** اختبار T المستقلة.

خطوات اختيارات الفروض الصفرية

1. **الفروض الإحصائية:** $H_0: \mu_1 = \mu_2 (\mu_1 - \mu_2 = 0)$

$H_A: \mu_1 \neq \mu_2 (\mu_1 - \mu_2 \neq 0)$

- μ_1 متوسط مجتمع العينة الأولي
- μ_2 متوسط مجتمع العينة الثانية

وهو اختبار ذوذيلين.

2. **الاختبار الإحصائى ومسلماته:** الاختبار الإحصائى T المستقلة ومسلماته كالآتى:

- الأعتداليةNormality: يفترض أن توزيع المتغير التابع إعتدالى التوزيع وهذه المسلمة على جانب كبير من الأهمية خاصة للعينات الصغيرة لأن الخطأ المعيارى يكون كبير بينما للعينات الكبيرة $N > 30$ يكون الخطأ المعيارى صغير.
- **المعاينة العشوائية** : يفترض اختيار العينات عشوائيًا.
- الإستقلالية Independence of observation: أفراد المجموعة الأولى مستقلة عن أفراد المجموعة الثانية وهذا يعنى ان قياسات أى فرد ما فى المجموعة مستقلًا

عـن قياسـات فـرد آخـر فـى نفـس المجموعـة وإذا لـم توجـد اسـتقلالية فـإن قيمـة P الاحتمالية لا يمكن الوثوق بها.

- **تساوى أو تجانس التباينات** Equal variances or homogeneity of variance :

تباين مجتمع المجموعة (العينة) الأولى مساويًا لتباين مجتمع العينة الثانية${\sigma_2}^2 = {\sigma_1}^2$، ويرى Privitera (2015) ان هذه المسلمة مرضية إذا كانت نسبة تباين المجتمع الأكبر إلى تباين المجتمع الأصغر أقل من 2.

والملاحـظ أن الاسـتقلالية والعشـوائية هـى نـواحى متعلقـة بمنهجيـة الدراسـة ويجـب علـى البـاحثين التحقـق مـن مسـلمات الاعتداليـة وتجـانس التباينـات بعـد جمـع البيانـات قبـل إجـراء اختبـار T، وأحيانًـا يكتـب فـى تقـارير البحـوث تـم التحقـق مـن الاعتداليـة للبيانـات باسـتخدام اختبـار كولمـوجروف– سـيمرنوف لـدرجات المتغيـر التـابع حيـث إذا كانـت قيمـة الاختبـار غيـر الدالـة إحصـائيًا فـإن التوزيـع اعتـدالى. وللتحقـق مـن تجـانس التباينـات يـتم إجـراء اختبـار Leven's F و إذا كانـت(P) $<$ (α)0.05، وعلـى ذلـك نقبـل الفـروض الصـفرى وبالتالى يوجد تساوى التباينات.

- **المتغير التابع:** يكون من مستوى القياس الفترى أو النسبى.

تكون صيغتها كالآتى:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

- $\bar{x}_1$ ، $\bar{x}_2$متوسط العينة الأولى والثانية على التوالى.
- ${s_1}^2$ ، ${s_2}^2$تباين المجموعة الأولى والثانية على التوالى.
- n_1 ، n_2حجم العينة الأولى والثانية على التوالى.

التباين المشترك Pooled variance

المعادلـة السـابقة تعـرض الخطـأ المعيـارى $(S\bar{x}_1 - \bar{x}_2)$ لإحصـاءTللعينـات المسـتقلة وهـذه الصـيغة تكـون صـحيحة فقـط فـى المواقـف التـى تكـون فيهـا أحجـام العينتـين متسـاوية $n_1=n_2$ ولكـن فـى المواقـف التـى تكـون فيهـا أحجـام العينـات مختلفـة فـإن الصـيغة

السابقة تكون متحيزة ولذلك تكون غير مناسبة والتحيز يأتى من حقيقة أن المعادلة السابقة تعالج العينتين كما لو كان تباينهما متساوى $s_1{}^2=s_2{}^2$، وعندما تكون أحجام العينات مختلفة فإن التباينات غير متساوية تمامًا ولا يجب معالجتها على أساس إنها متساوية. وكما نعلم أن تقدير معالم المجتمع من العينات الكبيرة أفضل من العينات الصغيرة، وعليه فإن التباين المتحصل يعكس تقدير دقيق لـ σأفضل من العينات الصغيرة و أحد طرق تصحيح هذا التحيز فى الخطأ المعيارى هو دمج تباين العينتين فى قيمة واحدة ويسمى التباين المدمج أو المشترك ويتم الحصول عليه بأخذ متوسطات أو دمج تباين العينتين باستخدام درجات الحرية كالآتى:

$$S_p^2 = \frac{(n_1 - 1)s_1{}^2 - (n_2 - 1)s_2{}^2}{n_1 - n_2 - 2}$$

والبسط يعكس مجموع التباينات وكلًا منهما تم وزنه عن طريق درجات الحرية وعليه فإن اختبار T كالآتى:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

اذا:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1{}^2+(n_2-1)s_2{}^2}{n_1+n_2-2}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

وعمومًا عندما يكون أحجام العينات متساوية فإنه لا توجد فروق بين استخدام التباين كلًا على حدة واستخدام التباين المدمج فى الصيغة السابقة حيث يكون القرار واحد ولكن إذا كان أحجام العينات غير متساويه فإنه التباين المشترك يحدث فروق قليلة. ولاختبار T درجات حرية:

$$\text{df}=(n_{1_}\ 1) + (n_2 - 1) = n_1 + n_2 - 2$$

$$= 9 + 8 - 2 = 15$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: افترض الباحث $\alpha = 0.05$ ولاختبار ذوذيلين و $df = 15$ وبالبحث فى جدول Tفإن قيمة T الجدولية= 2.13

4. الحسابات:

X_1	X_2
35	52
51	87
66	76
42	62
37	81
46	71
60	55
55	67
53	
n_1=9	n_2=8
$\sum x1$=445	$\sum x2$=551
$\overline{x_1}$=49.44	$\overline{x_2}$=68.88
$\sum {x_1}^2$=22865	$\sum {x_2}^2$=39009
$S1^2$=10778	${S_2}^2$=151.27
S_1=10.38	S_2=12.30

إذًا:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{49.44 - 68.88}{\sqrt{\frac{(9-1)107.78+(8-1)151.27}{8+9-2}\left(\frac{1}{8}+\frac{1}{9}\right)}}$$

$$T = \frac{-19.44}{5.50} = -3.35$$

5. القـرار والتفسـير: حيـث:|3.53-| >Tالحرجـة 2.13، وعليـه يُـرفض H0 وبالتـالى توجـد فـروق دالـة إحصـائيًا فـى تحصـيل الرياضـيات بـين تلاميـذ العينـة الـذين تلقـوا تدريسًـا فى بيئة أكثر تنظيمًا وتلاميذ العينة ذو التدريس فى البيئة الأقل تنظيمًا.

القرار فى ضوء فترتى الثقة: يتم بناء فترات الثقة لـ CI_{95}:

$$CI_{95}=(\bar{x}_1-\bar{x}_2) \pm T_{c.v}(S_{\bar{x}_1-\bar{x}_2})$$

$$= -19.44 \pm 2.13\ X\ 5.50$$

$$=(-31.81, 7.7)$$

وعليه فإن قيمة صفر الموضوعة فى الفرض الإحصائى الصفرى تقع خارج مدى فترة الثقة وبالتالى نرفض H_0

6. حجم التأثير: يتم حساب حجم التأثير لاختبار T المستقلة من المؤشرات الآتية:

أ- مؤشر كوهين d وتكون صيغته كالآتى:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_p^2}$$

حيث إن S_p^2 التباين المدمج:

$$S_p^2 = \frac{(n_1-1){s_1}^2 - (n_2-1){s_2}^2}{n_1 - n_2 - 2}$$

$$= \frac{(9-1)107.78 + (8-1)151.27}{8+9-2}$$

$$\frac{862.24 + 1058.89}{15}$$

$$= \frac{1921.13}{15} = 128.07$$

$$S_p = \sqrt{128.07} = 11.31$$

$$d = \frac{19.44}{11.31} = 1.75$$

وهو حجم تأثير كبير فى ضوء معايير Cohen.

ب- مؤشر مربع إيتا r^2 أو η^2 أو التباين المفسر:

$$\eta^2 = \frac{T^2}{T^2 + df} = \frac{T^2}{T^2 + n1 + n2 - 2}$$

$$= \frac{(3.53)^2}{(3.53)^2 + 8 + 9 - 2} = \frac{12.4609}{27.4609} = 0.4537$$

اذًا البيئة التدريسية فسرت 45.37 % من تباين التحصيل وهذا حجم تأثير كبير.

جـ- مؤشر اوميجا ω^2 :

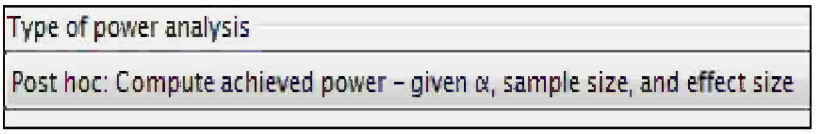
$$\omega^2 = \frac{T^2 - 1}{T^2 + df}$$

$$= \frac{(3.53)^2 - 1}{(3.53)^2 + 8 + 9 - 2} = \frac{11.4609}{27.4609} = 0.417$$

7. القوة الإحصائية: لتحديد القوة الإحصائية للمثال السابق باستخدام برنامج G-Power اتبع الاتى:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

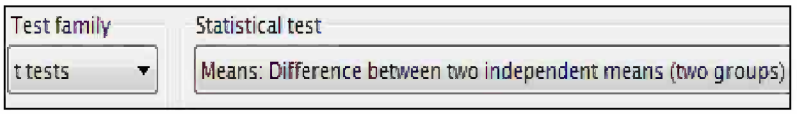
2. تحت Type of power analysis اختار:



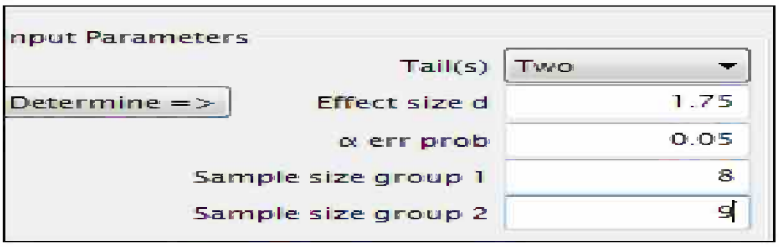
لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

3. تحت Test family اختار T- tests

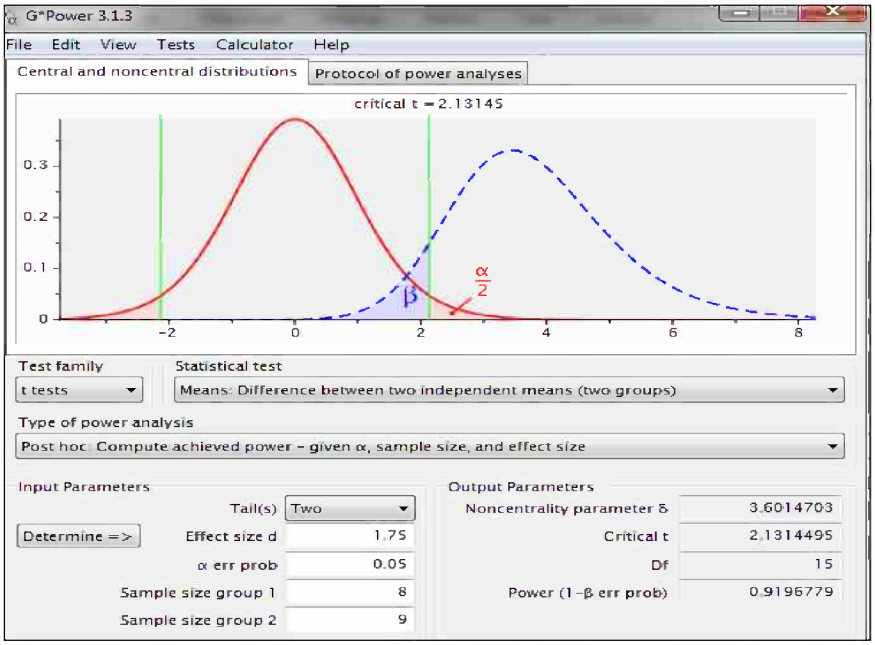
4. تحت Statistical test اختار:



5. ادخل المعالم Input parameters:



6. اضغط Calculated تظهر شاشة الناتج كالآتى:



ويتضح أن القوة الإحصائية= 0.919 وهذا مستوى قوة مرغوبة وقد يرجع هذا إلى جودة التصميم البحثى وجودة البيانات.

عرض نتائج اختبار T المستقلة وفقًا لـ APA

لتقييم الفروق بين التلاميذ الذين يتعرضون لطريقة تدريس ما و أقرانهم الذين يتعرضون لطريقة تدريس آخرى فى التحصيل الرياضى أوضحت نتائج اختبار T وجود فروق ذات دلالة إحصائية حيث: $T(15) = 3.53,\ p < 0.05,\ (d = 1.75)$، و أن متوسط تحصيل الطلاب فى طريقة التدريس الأول $(\bar{x}_1 = 49.44\ ,\ SD = 10.38)$ ومتوسط تحصيلهم فى طريقة التدريس الثانية $(\bar{x}_2 = 68.88, SD = 12.30)$ وأن فترات الثقة CI_{95} $(-31.18\ ,\ 7.7)$ وأتضح أن طريقة التدريس فسرت 45.3% من تباين التحصيل. أو متوسط التحصيل فى الرياضيات

للمجموعة الأولى $\bar{x}_1 = 49.44$ أو للمجموعة الثانية (68.88) اتضح وجود فروق بينهما ذات دلالة إحصائية حيث $T(15)=3.53, p< 0.05$, $d=1.75$, $CI_{95}(-31.18, 7.7)$

اختبار T المستقلة فى حالة عدم تجانس التباينات $\sigma_1^2 \neq \sigma_2^2$

فى الأمثلة السابقة استخدمنا التوزيع العينى الاعتدالى T تحت مسلمة تجانس التباين وعدم تحقق هذه المسلمة تكون النتائج غير دقيقة عكس تحقق فرضيه $n_1 = n_2$، فلو لم تتحقق هذه الافتراضية فإنها غير ذات أهمية فى تأثيرها على نتائج الاختبار. وعلى ذلك فإن عدم تحقق مسلمة التجانس للتباينات يتطلب إجراء معين وهذا الإجراء هو تعديل فى حساب الخطأ المعيارى للفروق $(\sigma_{\overline{x_1}-\overline{x_2}})$ وكذلك لدرجات الحرية. وقبل مناقشة هذا التعديل فلا بد من التعرض لاختبار التحقق من مسلمة التجانس للتباين.

اختبار التحقق من مسلمة تجانس التباين

للتحقق من هذه المسلمة فلا بد من إتباع خطوات اختبارات الفروض:

1. الفروض الإحصائية:

$H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$

- σ_1^2 تباين درجات مجتمع العينة الأولى.
- σ_2^2 تباين درجات مجتمع العينة الثانية.

وإذا تم رفض الفرض الصفرى فإن مسلمة التجانس لم تتحقق ولكن قبول H_0 يعنى تجانس التباينات وفى حالة رفض H_0 فلا بد من وجود تعديل أو توفيق adjustment للخطأ المعيارى ودرجات الحرية.

2. **الاختبار الإحصائى المناسب ومسلماته:** مسلماته مثل مسلمات T المستقلة والاختبار المناسب هو F للتجانس ويطلق عليه فى برنامج SPSS Leven's test وهى:

$$F=\frac{S_1^2}{S_2^2}$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: توزيع F تقريبًا اعتدالى ولهذا الاختبار درجتى حرية، كل درجة مرتبطة بتقديرات التباين للمجموعتين وبفرض أن لدينا بيانات لاختبار T حيث:n_1=31 ، n_2=41 و , s_1^2=105.96 ، s_2^2=36.42 فإن:

$$df_1 = n_1 - 1 = 31 - 1 = 30$$

$$df_2 = n_2 - 2 = 41 - 1 = 40$$

وبالبحث فى جدول F بدرجتى حرية 30 (البسط) ، 40(المقام) و α=0.02 فإن القيمة الحرجة أو الجدولية: $F_{c.v} = 2.20$

4. الحسابات:

$$F = = \frac{105.96}{63.42} = 2.91 \quad \frac{S_1^2}{S_2^2}$$

حتى لو اعتبرنا التباين الأصغر فى البسط و التباين الأكبر فى المقام فإن:

$$F = \frac{36.42}{105.96} = 0.34$$

وبالبحث فى جدول Fحيث درجات حرية البسط (numerator) 40 و درجة حرية المقام 30 denominator فإن F الحرجة=2.30.

5. القرار : F المحسوبة(2.91) >F الحرجة(2.20)، وعليه يتم رفضH0 حيث لا يوجد تجانس. وعند إجراء اختبار F للتجانس يوضع التباين الأعلى فى البسط والتباين الأقل فى المقام وذلك إذا كان الاختبار ذا ذيلين (Hinkle et al.,1994)

تقدير الخطأ المعيارى ($\sigma_{\overline{x_1}-\overline{x_2}}$) للفروق بين المتوسطات عند $\sigma_1^2 \neq \sigma_2^2$

عندما لا تتحقق مسلمة تجانس التباينات فإنه الإجراء البديل لاختبار الفرض الصفرى H0:$\mu_1 = \mu_2$ هو:

- تقدير الخطأ المعيارى للفرق بين المتوسطات بطريقة مختلفة ويصبح صيغة الخطأ المعيارى :

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- تعديل حساب درجات الحرية ويتم تقدير درجات الحرية من خلال تعديل Welch– Satterhwaite من الصيغة الآتية :

$$df=\frac{(\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2})}{\frac{\frac{{s_1}^2}{n_1}}{n_{1-1}}+\frac{\frac{{s_2}^2}{n_2}}{n_{2-1}}}$$

مثال : أراد باحث التحقق من الفرض الإحصائى الآتى:

$H_A: \mu_1 \neq \mu_2$ ،$H0: \mu_1 = \mu_2$

ولا يوجد تجانس فى التباينات وكانت نتائج العينتين كالآتى:

الثانية	الأولي	
19.60	27.20	$\bar{x}$
31	16	n
62.60	145.90	S^2
7.09	12.07	S

التحقق من مسلمة التجانس من خلال اختبار F كالآتى:

$$F = \frac{{s_1}^2}{{s_2}^2} = \frac{145.90}{62.60} = 2.33$$

وبالكشف فى جدول F ملحق( ) بدرجتى حرية:

$df_1 = n_1 - 1 = 16 - 1 = 15$(البسط)

$df_2 = n_2 - 1 = 31 - 1 = 30$(المقام)

و عند $\alpha = 0.10$ *يتضح أن:* $F_{c.v} = 2.01$ *وعليه نرفض* H0 *وبالتالى لا يوجد تجانس بين* ${\sigma_1}^2$ ، ${\sigma_2}^2$ *وعليه لحساب اختبار* T *يجب إجراء تعديل للخطأ المعيارى و لدرجات الحرية كالآتى:*

$$= \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$$T = \frac{27.2 - 19.6}{\sqrt{\frac{145.90}{16} + \frac{62.60}{31}}}$$

$$= \frac{7.6}{3.34} = 2.28$$

وإجراء تعديل على درجات الحرية كالآتى :

$$df = \frac{\left(\frac{145.90}{16} + \frac{62.60}{31}\right)^2}{\left(\frac{145.90}{16}\right)^2 / 15 + \left(\frac{62.60}{31}\right)^2 / 30}$$

$$= \frac{(9.12 + 2.02)^2}{(9.12)^2 / 15 + (2.02)^2 / 30}$$

$$= \frac{124.10}{5.54 + 0.14} = 21.85$$

ولا تقدر df كما فى الصيغة الآتية:

$$df = n_1 + n_2 - 2 = 6 + 31 - 2 = 45$$

وبالتالى لاختبار ذى ذيلين عند0.10= α و درجات حرية تقريبًا22 *فإن* 1.717=$T_{c.v}$ *وعليه* Tالمحسوبة(2.28) >الحرجة (1.717) وبالتالى نرفضH0

المناعة أو الضلاعة ضد عدم تحقق مسلماتT

يوصف اختبار T بأنه ضليع أو منيع Robust حيث توجد عدة مسلمات ضرورية لاختبار T مثل الاعتدالية وتجانس التباين وبدرجة أقل مسلمة تساوى أحجام العينات ولكنه يعتبر أقل تأثرًا بعدم تحقق مسلماته بدرجة متوسطة مثل الاعتدالية أو تجانس التباينات، وكما نعلم أن عدم تجانس التباين يتم التغلب عليه بإجراء و تعديل فى صيغة $\sigma_{\overline{x1}-\overline{x2}}$ وdf، ويرى Howell (2013) أن الاستنتاج العام من الدراسات أنه إذا تساوت أحجام العينات ولم تتحقق مسلمة التجانس للتباينات فإنه يوجد تأثير صغير جدًا على قيمة Tو تكون فى حدود 0.02± عند α =0.05 ولكن إذا ابتعد التوزيع عن الاعتدالية (ملتوى بشدة) فإنه تحدث مشاكل خطيرة خاصة إذا وجد عدم تجانس للتباين ولكن وجود التجانس تبدو الخطورة أقل.

ويقترح Howell (2013) عندما تكون أحجام العينات غير متساوية بدرجة كبيرة فيجب تطبيق مدخل Welch–Satterhwaite وفيها تستخدم صيغة درجات الحرية المعدلة، ويطرح تساؤل هو هل من الضرورى التحقق من مسلمة التجانس للتباينات؟ وفى هذا الشأن يقترح Cai & Hayes (2007) إن إجراء هذا الاختبار للتحقق من مسلمة التجانس غير ضرورى فالإجراء التقليدى هو إذا وجد تجانس فإنه يستخدم تقدير التباين المدمج Pooled variance وإذا وجد عدم تجانس يعتمد على التباينات المنفصلة كلًا على حدة، و يروا من الأفضل استخدام التباينات كلًا على حدة حتى لو وجد تجانس لأن نسب الخطأ تبقى قريبة من $\alpha=0.05$> وعليه فإن استخدام Tفى حالة عدم تحقق التجانس لا يؤثر على القيمة القرارية للاختبار ويقال أن الاختبار ضليع إذا كان أقل حساسية بدرجة نسبية لعدم تحقق شروط أو مسلمات استخدامه، وعدم حساسية اختبار T لعدم الاعتدالية وعدم تجانس التباينات يعتمد على حجم العينة فاذا كان= n_1 n_2 وحجم كل عينة أكبر من 30 فإن اختبار Tيستخدم بدرجة مناسبة ومعقولة بدون خطأ يستحق الوقوف عنده إذا كان درجة عدم تحقق مسلماته بدرجة متوسطة. ولكن إذا كانت الدرجات بها التواء شديد وعدم تجانس للتباينات بدرجة كبيرة يفضل استخدام البديل اللابارامترى.

ويرى Huck (2012) أنه إذا لم تتحقق مسلمة الاعتدالية أو التجانس يوجد ثلاث بدائل متاحة هى:

- استخدام صيغة معينة لتعويض عدم تحقق الاعتدالية أو عدم التجانس كما سبق ذكره من إجراء تعديل للخطأ المعيارى ودرجات الحرية.
- تعديل أو تحوير الدرجات الخام عن طريقة تحوير البيانات Data transformation لتقليل حدة عدم الاعتدالية أو عدم التجانس.
- استخدام اختبارات لا بارامترية مثل مان – ويتنى لا تتضمن مسلمات حول المجتمع.

ويمكن أن تكتب فى تقارير البحوث عبارات مثل:

عند المقارنة بين مجموعتين تم إجراء اختبار Tفى حالة ما إذا كشف اختبار F للتحقق من التجانس (عدم دلالة إحصائية) وإذا وجدت دلالة إحصائية فإن Wellch'sT-test يستخدم بدلا من StudentT-test.

او تم تحوير البيانات باستخدام الجزر التربيعى لمحاولة الوصول إلى الاعتدالية. أو لم يعتمد الباحث على اختبار T لعدم تحقق الاعتدالية وعدم تجانس التباينات ولذلك استخدم الاختبار اللابارامترى المقابل مان–ويتنى حيث إنه الأنسب.

يوجد ثلاث معادلات لحساب T وهى T لعينة واحدة:

$$\mathbf{T} = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

وT لعينة مرتبطة:

$$\mathrm{T} = \frac{\bar{D} - \mu_D}{S_D/\sqrt{N}}$$

وT لعينات مستقلة:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

ومن الواضح أن القيمة العالية لـT المحسوبة تعنى احتمال كبير لرفض H_0 بمعنى قوة عالية للاختبار الإحصائى المستخدم، ومن الواضح أن التأثير الحقيقى العالى للمعالجة (المتغير المستقل) يعنى أن قيم $\bar{x}_1, \bar{x}_1 - \bar{x}_2$ و $\bar{D}$ عالية وعليه فإن قيمة بسط المعادلات الثلاثة تكون عالية، ومن الأفضل استخدام مستويات متغير مستقل حتى يسمح للمجرب أن يعظم قيمة التأثير للكشف عن الفروق الحقيقية فى المجتمع ومقام هذه المعادلات الثلاثة دالة وظيفيه لحجم العينة فكلما زاد حجم العينة فإن قيمة المقام تنقص وعلى ذلك فإن التأثير الحقيقى للتجربة (زيادة البسط) مع حجم عينة كبير (نقصان المقام) يعطى قيمة عالية لقيمة الاختبار Tوعليه توجد احتمالية عالية لرفض

H_0 وعلى ذلك يكون الاختبار قوى إحصائيًا و على ذلك فإن زيادة حجم العينة يزيد القوة الإحصائية لاختبار T.

والمقام يتغير كوظيفة لتباين العينة ففى Tلعينة واحدة فإن الانحراف المعيارىS مقياس للتشتت و S_D فى تجارب المجموعات المرتبطة ، S فى تجارب المجموعات المستقلة، فكلما زاد التباين فإن المقام يزيد وبدوره تنقص قيمة Tولذلك فالعينات عالية التباين تقل من القوة الإحصائية لـ T ولذلك فمن المرغوب تقليل تباين العينات كلما أمكن وجعلها متجانسة بقدر الامكان، وأحد الطرق لتقليل التباين هو إجراءات الضبط فى التجربة وعليه ففى تجارب القياسات المرتبطة من المتوقع ان يكون التباين أقل من تباين التجارب المستقلة ولذلك فإن اختبار T المرتبطة أكثر قوة إحصائيًا من Tالمستقلة.

على الرغم من أن الاختبار الإحصائى لمقارنة متوسطين T إلا إنه يمكن إجرائه باستخدام اختبار Z أو F ولكن اختبارZ يعطى نتائج متحيزة قليلًا حيث يمكن أن يؤدى إلى زيادة الخطأ من النوع الأول أكبر من مستوى الدلالة الإحصائية الذى يتبناه الباحث وهذا التحيز يزيد مع صغر حجم العينة، فى المقابل فإن اختبار F غير متحيز ويكون الاستنتاج الخاص بـ H_0 فى حالة استخدام F مماثل لاستنتاج اختبار T وبالتالى لا توجد إشكالية فى استخدام T وF للمقارنة بين مجموعتين، وفى ضوء هذه الحقيقة فإن بعض الباحثون يفضلون استخدام اختبار F عند المقارنة بين متوسطى عينتين حتى أنه فى بعض التقارير البحثية يستخدم F وتكتب النتائج باستخدامه للمقارنة بين مفهوم الذات للذكور($\bar{x}_1$=21.1,S_D=31.1) و مفهوم الذات للإناث ($\bar{x}_2$=29.9, S_D=42.3) فنجد أن $F_{(1,596)}$= 8.51, p< 0.05، حيث إن $F_{(1,596)}$ هما درجتى الحرية ويكون :K-1 df_1= حيث K عدد المجموعات أو العينات: $df_1 = 2 - 1 =$ 1، درجة الحرية الثانية هى df_c=N-K حيث Nحجم العينة الكلى (ذكور و إناث): df_2=598-2=596

وفى بعض الأحيان يقوم الباحثون بعرض جدول تحليل التباين لـ F.

المقارنة بين متغيرات تابعة متعددة

أحيانًا يتم جمع بيانات لعينة واحدة أو عينتين على أكثر من متغير تابع ثم يقوم الباحث باختبار فرض صفرى لكل متغير تابع على حدة على الرغم أنه من الأفضل استخدام T^2هو تلنج أو تحليل التباين المتدرج MANOVA. ولكن الباحث يستخدم اختبار T أوF لكل متغير تابع على حدة **ويكتب فى تقرير البحث الآتى:** وبإجراء اختبارT للفروق بين منخفضى ومرتفع التحصيل فى تقدير الذات و سمات الشخصية تبين بالنسبة لتقدير الذات $T=5.28$, $p < 0.001$ وبالنسبة للانبساطية $T_{(89)}=7.8$ $p<0.001$. أو **يكتب فى تقرير البحث:** بإجراء اختبار F للفروق بين القياس البعدى للمجموعة التجريبية والضابطة فى إبعاد سمات الشخصية هى للبعد الأول $F_{(1,642)}=32.474,p<0.001$ و بالنسبة للبعد الثانى $F(1,642)=(4.453),P<0.0001$ ، وفى هذه الحالة يجب استخدام تصحيح Bonferroni Adjustment لأنه عندما يختبر الباحث فروض صفرية عديدة وكل فرض يناظر متغير تابع فمن المتوقع تضخم الخطأ من النوع الأول حيث يحدث له زيادة عن القيمة التى افترضها الباحث فى بداية اختبار الفروض بطريقة غير مباشرة، وأحد الطرق للتعامل مع هذه المشكلة والأكثر استخدامًا هو تطبيق أسلوب تصحيح بونفيرونى وسوف يشار إليه أثناء تحليل التباين.

تقدير حجم التأثير وتحليلات القوة لاختبارT

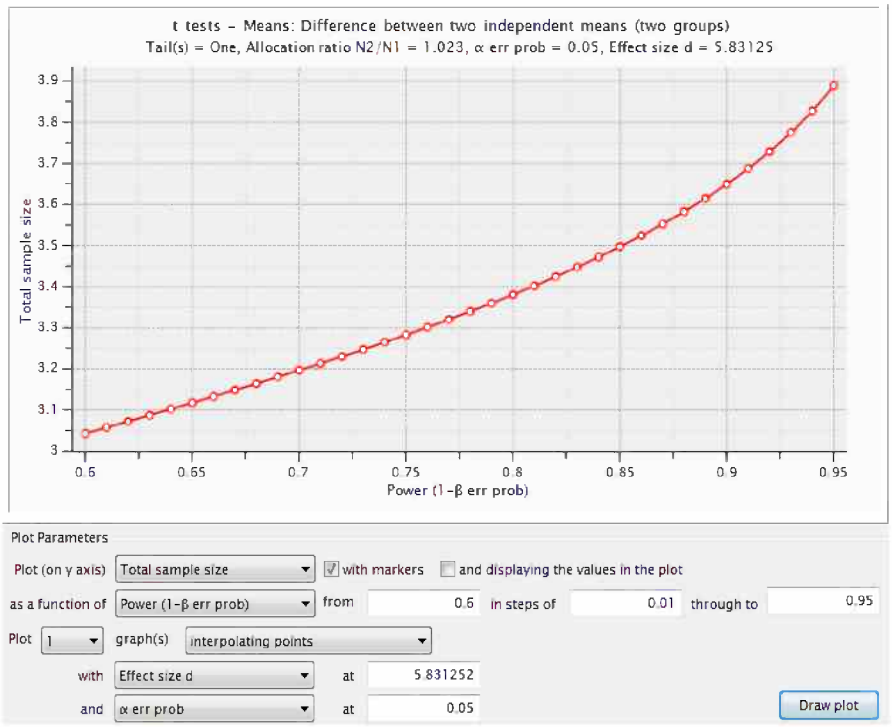
كما سبق وأوضحنا عند التعامل مع متوسط أو متوسطين باستخدام اختبارات الفروض فإن الباحث يجب أن يميز بين الدلالة الإحصائية والدلالة العملية ومن أهم مؤشرات حجم التأثير شيوعًا المصاحبة لاختبارT هى مؤشر كوهين (d) أو مؤشر μ^2 للمقارنة بين متوسطين أو مؤشر Partial eta squared (μ^2p) المصاحب لاختبار F للمقارنة بين متوسطين ومؤشر (μ^2p) هو مماثل لمؤشر r^2 لتقدير التباين المفسر. و فيما يلى أهم مؤشرات حجم التأثير المصاحب لاختبارات الفروض أو متوسطين بإستخدام T ، F كما حددها Huck (2012) :

المؤشر	صغير	متوسط	كبير
كوهين d	0.20	0.50	0.80
إيتا μ	0.10	0.24	0.37
إيتا تربيعية η^2	0.4	0.06	0.14
أوميجا ω^2	0.01	0.06	0.14
$\mu^2 p$	0.01	0.06	0.14
Cohen f	0.10	0.25	0.40

حيث μ^2p هى اختصارًا Squared Partial Eta

والمؤشرات الأكثر استخدامًا فى تقارير البحوث هو مؤشر d,μ^2p.

والشكل التالى يوضح الرسم البيانى لمنحنى العلاقة بين قوة الاختبار (محور السينات)، وحجم العينة الكلى:



تعقيب: فى نهاية هذا الفصل لا بد من عرض الآتى:

- النتائج غير الدالة إحصائيًا لا تعنى أن H_0 حقيقى.
- الفرض الصفرى لا يجب اعتباره حقيقى لأن الاختبار فشل فى رفضه، فعند المقارنة بين المجموعات فى المتوسطات القبلية فالباحث يأمل ألا يرفض H_0 لأنه يريد أن

تكـون للمجموعـات الضـابطة والتجريبيـة نفـس الـدرجات فـى بدايـة الدراسـة ولكـن يوجـد العديـد مـن الأسـباب لأن نعتبـر مـن الخطـورة أن الاعتقـاد بـأن H_0 حقيقـى إذا لـم يـرفض. وتوجـد عـدم منطقيـة لقبـول H_0 لأمـر يتعلـق بجـودة البيانـات وذلـك لحـدوث نـوع مـن الـتقلص Attenuation حيـث لهـا ثبـات مـنخفض، فالثبـات المـنخفض يقـود إلـى الفشـل فـى رفضH_0ولـذلك فـإن قبـول H_0 نتيجـة عـدم الثبـات للبيانـات. والاعتبـار الآخـر لعـدم رفـض H_0 هـو القـوة الإحصـائية وهـذا متعلـق بحجـم العينـة، فالفشـل فـى الحصـول علـى دلالـة إحصـائية قـد يكـون بسـبب حجـم العينـة فربمـا تكـون صـغيرة جـدًا فربمـا يختلـف متوسـط مجتمـع المجموعـة الضـابطة$\mu_{control}$ عـن متوسـط مجتمـع المجموعـة التجريبيـةμ_{exper} إختلافًـا كبيـرًا ولكـن الدراسـة ينقصـها الحساسـية (القـوة) للكشف عن هذه الفروق.

*و أخيرًا قدم*Aron et al. (2013) *مراجعة للأنواع الثلاثة لـ* T:

الملمح	*عينة واحدة*	*عينتين مرتبطتين*	*عينتين مستقلتين*
تباين المجتمع معلوم	لا	لا	لا
متوسط المجتمع معلوم	*نعم*	لا	لا
عدد درجات كل فرد	1	2	1
فروق الدرجات(D)	لا	*نعم*	لا
التوزيع	*توزيع* T	*توزيع* T	*توزيع*T
درجات الحرية	df=n-1	df=n-1	$df= n_1+n_2-2$
المعادلة	$(\bar{x} - \mu)/S_{\bar{x}}$	$\frac{D}{S_{\bar{D}}}$	$\frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1-\bar{x}_2}}$
إحصاء العينة	$\bar{X}$	$\bar{D}$	$\bar{X}_1$-$\bar{X}_2$

تنفيذ اختبار T المستقلة فى SPSS: مثال Hinkle et al.(1994)

أولا إدخال البيانات:1. اضغط Variable View

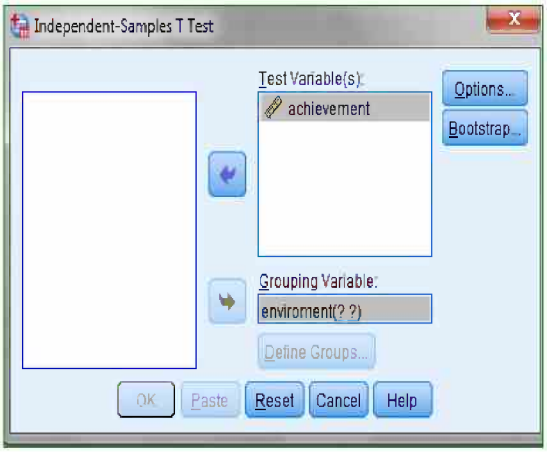
2. اكتب مسمى المتغيرات عمود Name وهما:

المتغير المستقل: البيئة environment: بيئة أقل تنظيما وتخطيطا (1)، بيئة عالية التنظيم والتخطيط (2). والمتغير التابع: التحصيل achievement

3. اضغط Data View تظهر شاشة البيانات

ثانيًا: تنفيذ الأمر:

1. اضغط Analyze→Compare Means →Independent-Samples تظهر الشاشة الآتية:



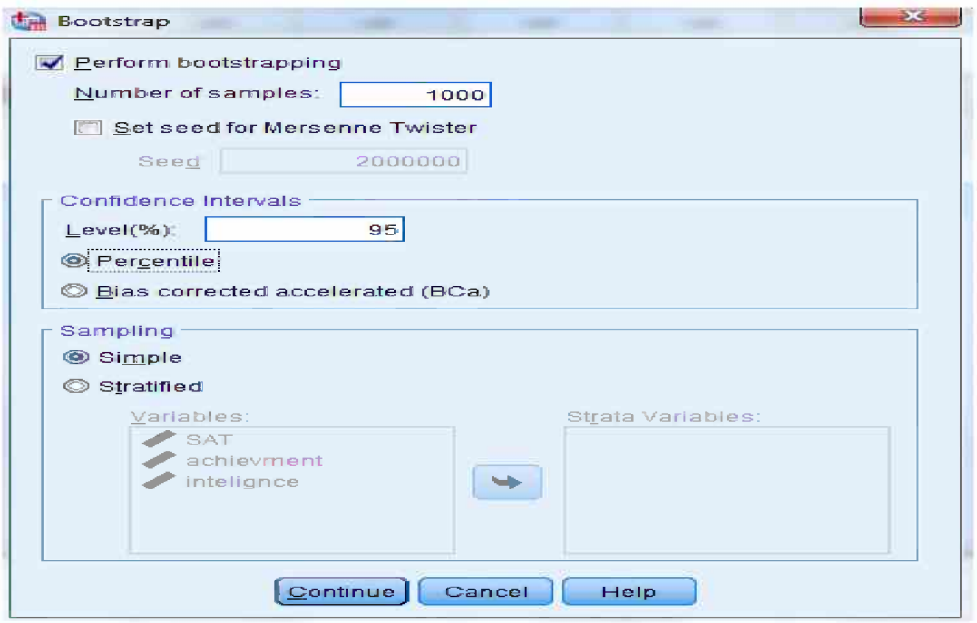
2. اضغط achievement وانقله إلى مربعTest Variables

3. اضغط environment وانقله إلى مربعGrouping Variable

4. اضغط DefineGroups تظهر الشاشة الآتية:

أكتب الكود 1 أمام Group 1 وأكتب الكود 2 أمام Group 2

5. اضغط على Bootstrap تظهر النافذة الآتية:



6. اضغط Perform bootstrapping :

- حدد عدد العينات التى ترغب فى توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 1000 مثلًا وهى Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيطة.

7. اضغط Continue ثم اضغط Ok

ثالثًا: تفسير المخرج: الجدول الأول: الإحصائيات الوصفية:

Group Statistics

	enviroment	N	Mean	Std. Deviation	Std. Error Mean
achievement	1.00	9	49.4444	10.38161	3.46054
	2.00	8	68.8750	12.29910	4.34839

حيث أعطى المتوسط Mean والانحراف المعيارى Std. Deviation والخطأ المعيارى للمتوسط Std. Error Mean: $SE = \frac{10.38}{\sqrt{9}} = 3.46$ والواضح أن عدد أفراد المجموعة الأولى 9 والثانية 8 وأن متوسط تحصيل العينة الأولى أكبر من متوسط تحصيل العينة الثانية.

الجدول الثانى يتضمن إحصائيات اختبار Tالجزء الأول:

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
achievement	Equal variances assumed	.275	.608	-3.533-	15	.003
	Equal variances not assumed			-3.496-	13.823	.004

الجزء الثانى:

t-test for Equality of Means			
Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		Lower	Upper
-19.43056-	5.49904	-31.15149-	-7.70962-
-19.43056-	5.55732	-31.36416-	-7.49695-

فالجزء الأول يتضمن اختبار Levene s Test للتحقق من مسلمة اختبار T المستقلة وهى تساوى أو تجانس تباينات العينتين:

$$H_0:\sigma_1^2=\sigma_2^2$$

$$H_A:\sigma_{1\neq}^2\sigma_2^2$$

ويتضح أن قيمة F=0.275 و 0.05 < Sig (p)=0.608 وهى غير دالة إحصائيًا وعليه تتحقق مسلمة استخدام اختبار T المستقلة وهى تجانس التباينات . وعليه تعتمد على نتائج اختبار T فى حالة Equal Variances Assumed وإذا رفض الباحث الفرض الصفرى فعليه نعتمد على نتائج اختبار T فى حالة Equal Variances Not Assumed أعطى المخرج:

- قيمة T المحسوبة =3.533-
- Df =8+9-2=15
- قيمة p الاحتمالية هى 0.003 =Sig (2tailed)، وبما أن: 0.05 > 0.003، وعليه يُرفض H_0 وبالتالى يوجد فرق دال إحصائيًا عند 0.05 فى التحصيل بين بيئتى التعلم. وإذا تبنى الباحث اختبار ذو ذيل واحد وعليه: P=0.003/2=0.0015 ، وعليه رفض H_0 لاختبار ذى ذيل واحد أسهل من رفضه لاختبار ذى ذيلين وعلى ذلك فإن الاختبار الموجه أكثر قوة إحصائية من الاختبار غير الموجه.

والمتوسطات التى نحصل عليها من العينة هى تقديرات النقطة point estimates ولكن بدلًا من هذا فإن التقديرات تكون عباره عن مدى وهى أكثر واقعية من النقطة الوحيدة وهو للفترة التى لها حد أدنى وحد أعلى وتسمى حدود الثقة وهى تسمح بمعرفة كيف نثق فى ان متوسط المجتمع يكون داخل فترة معينة. وإذا لم تتضمن حدود الثقة القيمة المفترضة فى الفرض الصفرى وهى فى هذه الحالة صفر فإن

الفـرض الصـفرى يـرفض، بمـا أن القيمـة المتضـمنة فـى الفـرض الصـفرى تقـع خـارج منطقة رفض H0 وعليه فإن حدود الثقة:

$$CI_{95}=(-31.151,-7.709)$$

وعليه لا تتضمن القيمة صفر بالتالى يرفض H0. ويمكن تغيير حدود الثقة إلى 99% من خلال اختيار Option فى امر T.

نتائج البوتستراب:

Bootstrap for Independent Samples Test

		Mean Difference	Bootstrap[a]				
			Bias	Std. Error	Sig (2-tailed)	95% Confidence Interval	
						Lower	Upper
achievement	Equal variances assumed	-19.43056-	-.10651-	5.58973	008	-31.14925-	-8.88376-
	Equal variances not assumed	-19.43056-	-.10651-	5.58973	011	-31.14925-	-8.88376-

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

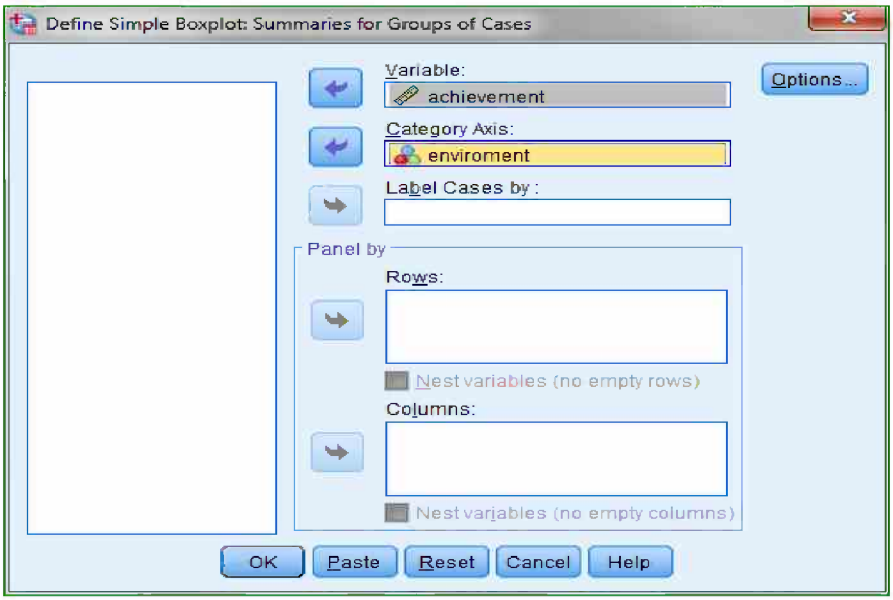
عليك المقارنة بين نتائج البوتستراب ونتائج اختبارات الفروض الصفرية لبيانات العينة المجمعة من الأفراد، لاحظ لا اختلاف جوهرية بين النتائج وذلك لصغر حجم العينة.

عرض النتائج بيانيًا

يمكن عرض النتائج من خلال Histogramأو Boxplot كالآتى:

1. اضغط Boxplot ← Legacy Dialogs ← Graphs تظهر الشاشة الآتية:

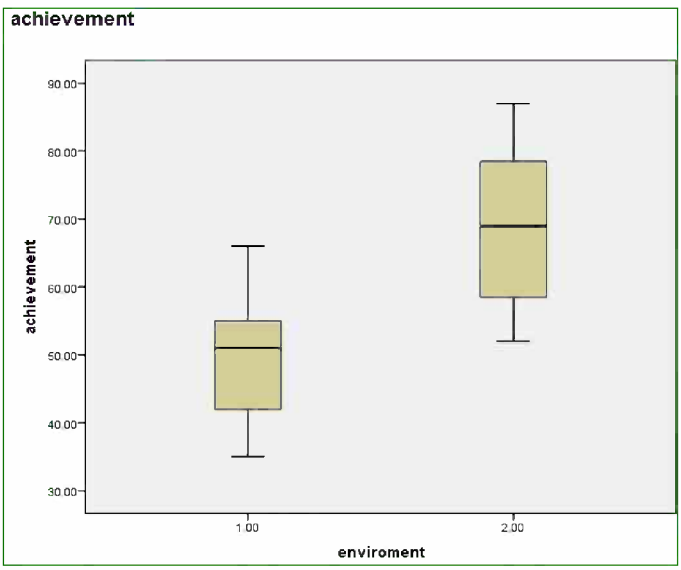
2. اضغط Simple ثم Define:



3. انقل متغير Achievement إلى مربع Variable

4. انقل متغير Environment إلى مربع CategoryAix ثم اضغط OK

المخرج:



الملاحظ أن وسيط العينة 2 أكبر من وسيط العينة 1(الخط داخل المستطيل أو الصندوق) وهذا يدل على أنه أكثر تحصيلًا.

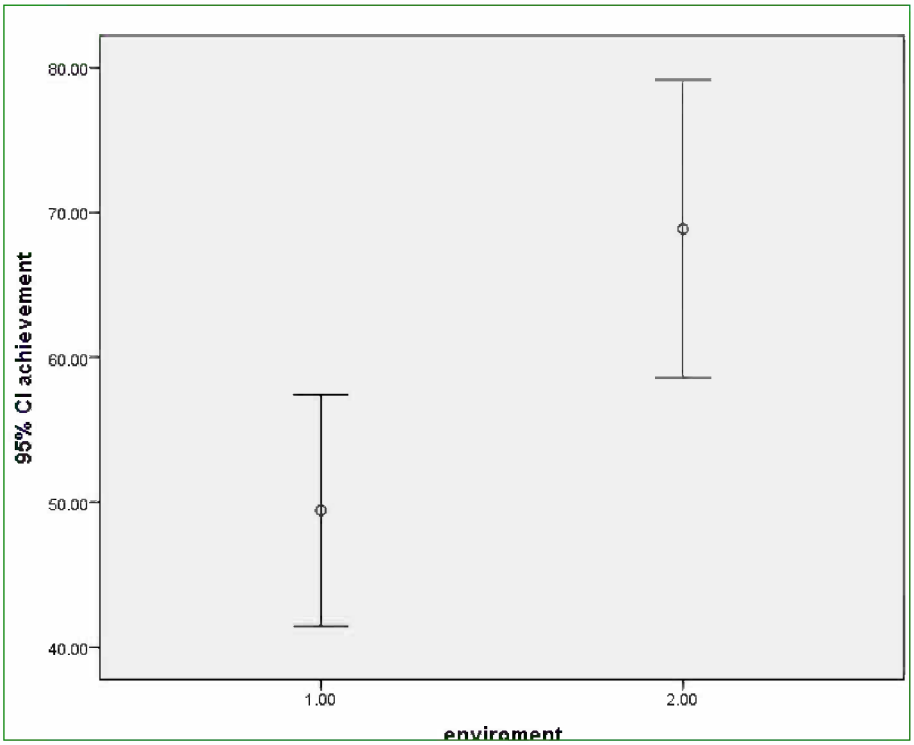
ويمكن عرض Error Bar **كالآتى:**

Graphs→ Legacy Dialogs→ ErrorBar

2. اضغط Simple ثم Define

3. انقل achievement الى مربع Variable

4. انقل Environment إلى CategoryAix ثم اضغط Ok



حيث الدائرة O تعكس المتوسط .

الفصل العشرون

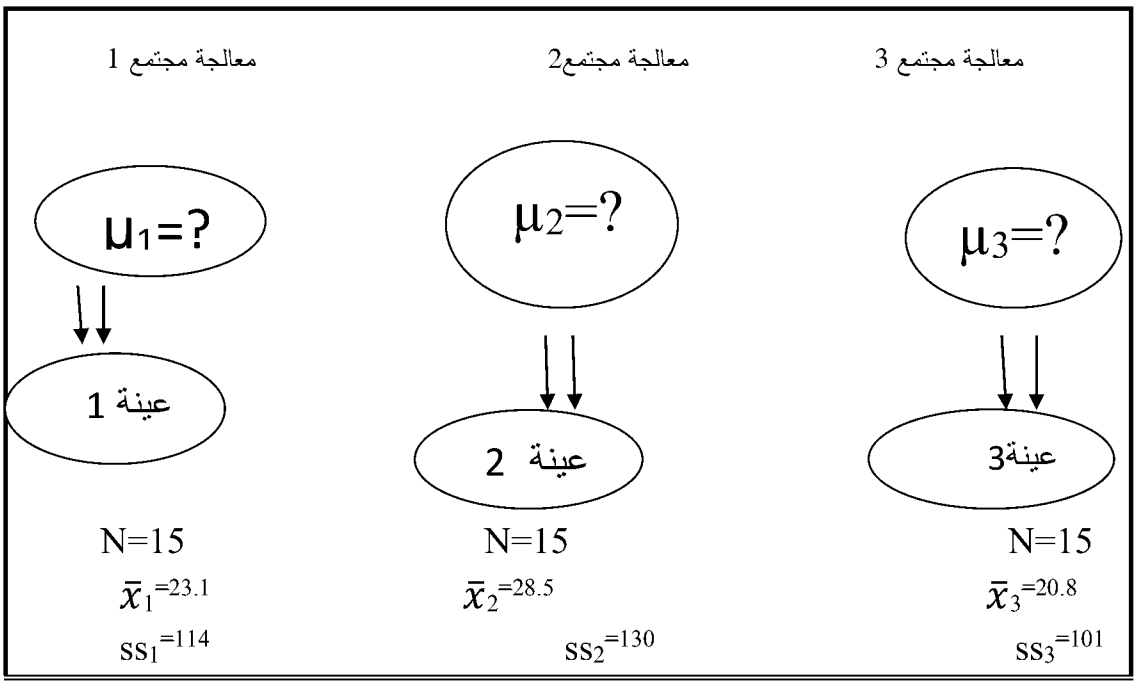
تحليل التباين احادى الاتجاه

One way analysis of variance (ANOVA)

فى الفصول السابقة تناولت اختبارات إحصائية استدلالية لاختبارات فروض حول متوسط مجتمع واحد أو متوسطين مثل اختبارات z وT، وفى هذا الفصل سوف تتسع المناقشة عندما يكون هدف الباحث إجراء اختبارات فروض تتعامل مع ثلاث متوسطات فاكثر فالأسلوب الإحصائى المستخدم فى هذه النوعية من الدراسات يطلق عليه تحليل التباين Analysis of variance أو اختصارًا ANOVA و احيانًا يسميه البعض تحليل التباين احادى التصنيف Analysis of variance one way classification. وتحليل التباين هو أسلوب إحصائى استدلالى يستخدم على نطاق واسع فى كثير من التخصصات خاصة فى البحوث التربوية والنفسية والاجتماعية، وفى هذا الفصل يكون التركيز على الصيغة البسيطة لتحليل التباين الذى يتعامل مع متغير مستقل واحد تصنيفى متعدد المستويات (المجموعات) ومتغير تابع فترى على الأقل ثم نتعرض بعد ذلك إلى تحليل التباين المتعدد والمتعدد المتدرج.

وتحليل التباين هو إجراء لاختبارات الفروض لتقويم فروق بين المتوسطات ولمعالجتين أو لمجتمعين فأكثر، ويتضح أن اختيار T يتشابه مع اختبار F فى ان كلاهما يؤدوا نفس الوظيفة وهى اختبار فروق بين متوسطات فى المجتمع ولكن اختبار T محدود فقط بمعالجتين أو بمجموعتين وبذلك يتعامل مع التصميم الأكثر شيوعًا وهو تصميم المجموعتين. وتحليل التباين يسمح بدراسة الظاهرة الإنسانية بصورة أكثر شمولية وكلية من اختبار T حيث يمكن تضمين أكثر من متغير مستقل (تحليل التباين المتعدد) وأيضًا يمكن تضمين متغير مستقل فأكثر وأكثر من متغير تابع فى نفس التحليل تحليل التباين المتدرج (MANOVA) وكذلك يظهر تأثير التفاعلات بين مستويات المتغيرات المقاسة المستقلة، ولذلك فإن ANOVA يجعل الباحثين أكثر مرونة فى تصميم تجاربهم وذلك بالتعامل مع أكثر من معالجتين أو مجموعتين مثل تصميم سولمون ذى

الأربـع مجموعـات. وفيمـا يلـى شـكل يوضـح الموقـف الامثـل لاسـتخدام ANOVA احـادى الاتجاه:



الشكل (1.20): الشكل الامثل لاستخدام ANOVA.

فـى هـذا الشـكل نلاحـظ وجـود ثـلاث عينـات يمثلـوا ثـلاث مجتمعـات والهـدف مـن التحليـل هـو تحديـد مـا إذا كـان يوجـد فـروق بـين العينـات الثلاثـة لأن هـذا يعطـى دليـل كـافى لاستنتاج فروق فى متوسطات المجتمعات الثلاثة وعليه يتم استنتاج تفسيرين:

- عدم وجود فروق بين متوسطات المجتمعات الثلاثة أو المعالجات الثلاثة.
- وجود فروق فى متوسطات المجتمعات أو المعالجات الثلاثة.

لاحـظ أن هـذين التفسـيرين ينـاظروا الفـرض الصـفرى (H_0) والفـرض البـديل (H_A) فـى اختبارات الفروض.

ويسـتخدم هـذا الاختبـار مـع التصـميمات التجريبيـة بـين المجموعـات أو تصـميم بـين المجموعـات المسـتقلة أحاديـة الاتجـاه أو العامـل One way between subject design حيـث إن المتغيـر المسـتقل متعـدد المسـتويات ويطلـق عليـه مسـتويات العامـل و يشـاع اسـتخدام مفهـوم عامـل ليشـير إلـى المتغيـر المسـتقل أو متغيـر المعالجـة فـى

التصـــميمات التجريبيـــة أو شـــبه التجريبيـــة، بينمـــا شـــروط المعالجـــات Treatment conditions تشير إلى مستويات العامل كما هو موضح:

الجدول (1.20): مثال لتصميم لبيانات لمجموعات مستقلة.

عامل (طريقة التدريس)		
معالجة 1	معالجة 2	معالجة 3
5	4	6
6	3	5
7	2	7
8	4	8
$\mathbf{n_{1=4}}$	$n_{2=4}$	$n_{1=4}$

علـى ذلـك فـإن الهـدف مـن اسـتخدام ANOVA هـو عمـل اسـتنتاجات حـول مـا إذا كانـت يوجـد فـروق بـين متوسـطات المجتمعـات بغـض النظـر عـن عـدد العينـات أو المجموعـات المتضـمنة ويصـبح السـؤال الرئيسـى هـو هـل متوسـطات المجتمعـات المختلفـة متسـاوية أو غيـر متسـاوية؟ وكمـا سـبق الإشـارة اثنـاء الحـديث عـن اختبـار T يمكـن اسـتخدام تحليـل التبـاين لاختبـار دلالـه الفـروق بـين متوسـطين أيضًـا كبـديل لاختبـار T ولـذلك نعـود فنقـول أنه يمكن استخدام اختبار F للتحقق من الدلالة الإحصائية بين متوسطين فاكثر.

الفرق بين تحليل التباين البسيط وانواع اخرى من ANOVA

فـى هـذا الفصـل نعـرض تحليـل التبـاين البسـيط ويشـار إليـه تحليـل التبـاين احـادى الاتجـاه One way ANOVA. ولكـن يوجـد انـواع أكثـر تعقيـدًا لـ ANOVA كمـا سـوف نتعـرض موضـع آخـر مثـل تحليـل التبـاين ذو اتجـاهين Two-way ANOVA وتحليـل التبـاين للقياسـات المتكـررةRepeated measures ANOVA وتحليـل التبـاين المتـدرجMultivariate ANOVA علـى الـرغم انهـم جميعـا يركـزوا حـول المتوسـطات لكنهم يختلفوا فى ثلاثة مظاهر:

- عدد المتغيرات المستقلة
- عدد المتغيرات التابعة

- قياسات العينات سواء كانت مستقلة أو مرتبطة

ففى تحليل التباين الاحادى يوجد متغير مستقل واحد تصنيفى ومتغير تابع واحد كمى فترى أو نسبى متصل وعينات ذو قياسات مستقلة ويشاع فى التراث بـ Simple ANOVA أو F-test، وتتم المقارنة بين مستويات أو مجموعات المتغير المستقل فى المتغير التابع الكمى المتصل وبالتالى فإن المتغير المستقل هو كيفى تصنيفى متعدد المستويات اثنان فأكثر ويطلق عليه البعض تحليل التباين احادى العامل، حيث يشير العامل والمتغير المستقل نفس المعنى ويفضل العامل فى حالة التصميمات التجريبية خاصة المعقدة.

اما ANOVA Two-way يعنى أن الاختبار يتعامل مع متغيرين مستقلين تصنيفين وتابع واحد متصل بينما ANOVA Three-way يعنى ان الاختبار يتعامل مع ثلاث متغيرات مستقلة تصنيفية وهكذا، لاحظ أن المتغيرات المستقلة كيفية تصنيفية اسمية متعددة المستويات بينما يوجد متغير تابع وحيد متصل واحيانًا يطلق عليه البعض تحليل التباين المتعدد Multiple ANOVAحيث المتغيرات المستقلة اثنين فأكثر.

وعلى ذلك فإنه يمكن استخدام ANOVA لتقويم فروق المتوسطات فى دراسة بحثية تتضمن أكثر من عامل. فعلى سبيل المثال اراد باحث المقارنة بين أسلوبين مختلفين للعلاج ثم فحص فعاليتهما الآنية واستمرار فاعليتهما عبر الزمن، فالبحث يتضمن مجموعتين مختلفتين أحدهما تلقت أحد طرق العلاج بينما الأخرى تلقت الطريقة الثانية وقام بقياس كل مجموعات عبر فترات زمنية مختلفة والتصميم كالآتى:

- العامل الأول: طرق العلاج (قياسات مستقلة)
- العامل الثانى: الوقت حيث تم قياس كل مجموعة عبر ثلاث فترات زمنية (قياسات متكررة). على ذلك فالدراسة تتضمن عاملين ويطلق عليها تصميم ثنائى العامل أو تصميمات عاملية وتسمى التصميمات المختلطة Mixed design

		الوقت		
			بعد المعالجة مباشرة	بعد 6 شهور
العلاج	الأولي	درجات	درجات	درجات
	الثانية	درجات	درجات	درجات

تصميم بحثى يتضمن عاملين.

اما تحليل التباين المتدرج MANOVA فهو يتعامل مع أكثر من متغير تابع متصل من مستوى فترى على الأقل ومتغير مستقل تصنيفى واحد فأكثر. اما إذا كانت القياسات على نفس المجموعة كما فى تصميم داخل الأفراد أو القياسات المتكررة فيستخدم تحليل التباين للقياسات المرتبطة A within- groups ANOVA ، بينما إذا كانت القياسات لمجموعات مستقلة يستخدم تحليل التباين الاحادى للمجموعات المستقلة Between-groups ANOVA

مسلمات تحليل التباين

لتحليل التباين عدة مسلمات أهمها:

- **تجانس التباين Homogeneity variance**: فالمسلمة الأساسية لتحليل التباين هو تساوى تباينات مجتمعات العينات اى:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \text{----} = \sigma_n^2$$

واحيانًا يطلق عليها **Homoscedasticity** وعدم توافر هذه المسلمة بمعنى عدم التجانسHeterogeneity of variance يؤدى إلى تضخم قيمة التباين فى بسط اختبار F وهذا يزيد من احتمالية ارتكاب الخطأ من النوع الأول (رفض خاطئ للفرض الصفرى) ويتم التحقق منها باستخدام Leven's test أو اختبار بارتليت Bartlett's test، واذا لم تتحقق هذه المسلمة فمن الأفضل استخدام إحصاءWelsh وإجراء اختبارات المقارنات البعديةPost-hoc tests يكون محل شك.

- **الاعتداليـــة:** يفتـرض ان تكـون توزيـع بيانـات المتغيـر التـابع اعتداليـة فـى مجتمعاتهـا، وهـذه المسـلمة علـى درجـة كبيـرة مـن الأهميـة خاصـة للعينـات الصـغيرة. والابتعـاد بدرجـة متوسـطة عـن الاعتداليـة ليسـت مشـكلة خطيـرة، ودائمًـا مـع العينـات المتوسـطة والكبيـرة تتـوفر الاعتداليـة. ويـرى Green & Salkind (2014) فـى بعـض المجتمعـات غيـر الاعتداليـة التوزيـع فلحجـم عينـة 15 لكـل مجموعـة كافيـة لإعطـاء تقـديرات دقيقـة لـ P. والقـوة الإحصـائية لاختبـار F تـنقص كلمـا كـان توزيعـات المجتمعـات غيـر اعتداليـة (ملتويـة – متفرطحـة)،ودائمًا يـتم التحقـق منهـا باسـتخدام اختبـار كولموجـوروف–سيميرنوف أو اختبار شابيرو–ويلك Shapiro-Wilk.
- **الاســتقلالية:** يفتـرض ان كـل فـرد مسـتقل بذاتـه عـن الآخـرين واختيـار أى فـرد فـى نفـس المجموعـة مسـتقل عـن أى فـرد آخـر فـى ذات المجموعـة، كمـا أن اختيـار أفـراد أى مجموعة لا يؤثر على اختيار فراد المجموعات الأخرى.
- **المعاينة العشوائيةRandomness** : يفترض ان يتم اختيار أفراد كل مجموعة عشوائيًا من مجتمعها ذلك حتى نزداد طمأنينة فى تعميم النتائج. ولكن كثيرًا من الباحثين لا يعطون انتباه لهذه المسلمات عند استخدام ANOVA.

وإذا لم تتحقق مسلمتى الاعتدالية وتجانس التباينات فإن الباحث امامه اربعة بدائل Huck (2012):

- حدد واستبعد القيم المتطرفة.
- حور أو عدل البيانات لتقليل عدم الاعتدالية.
- استخدم حالات خاصة لـ ANOVA وخاصة تعديل درجات الحرية df لـ Welch لـ ANOVA.
- استخدم بديل آخر لـ ANOVA لا يتطلب هذه المسلمات الصارمة مثل الاختبار اللابارامترى المقابل (كروسكال والاس).

مصادر التباين فى ANOVA

يستخدم تحليل التباين لتحليل تباين مجموعتين أو مستويين فاكثر للعامل وكل مستوى من العامل أو المتغير المستقل هو مجموعة وإذا كان متوسطات المجموعات متساوية فإنه من المحتمل ان يكون التباين بين المجموعات متساوى والفروق الكبيرة بين متوسطات المجموعات يؤدى إلى تباينات كبيرة بين متوسطات المجموعات، والفرض الصفرى لـ F للمقارنة بين متوسطات عديدة:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \text{----} = \mu_k$$

- μ_1 متوسط مجتمع العينة الأولى، μ_2 متوسط مجتمع العينة الثانية، μ_3 متوسط مجتمع العينة الثالثة، μ_k متوسط مجتمع العينة k.

بينما الفرض البديل دائمًا غير موجهه ويصاغ إذا حدث أن أحد متوسطات المجموعات أو أكثر تختلف تأثيرها على المتغير التابع عن متوسطات المجموعات الأخرى فيمكن أن يصاغ كالآتى:

$$H_A: \mu_1 \neq \mu_2 \neq \mu_3 \text{ ----} \neq \mu_k \quad ,$$

$$H_A: \mu_1 \neq \mu_2 = \mu_3 = \text{----} = \mu_k$$

توجد فروق على الأقل لمتوسط واحد عن بقية المتوسطات للمجتمعات، وعليه فإن للتجربة أو للمعالجة تأثير على المتغير التابع (تذكر أن الفرض البديل هو فرض البحث).

وكذلك يفترض اختبار F ان التباينات متساوية:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \text{----} = \sigma_k^2$$

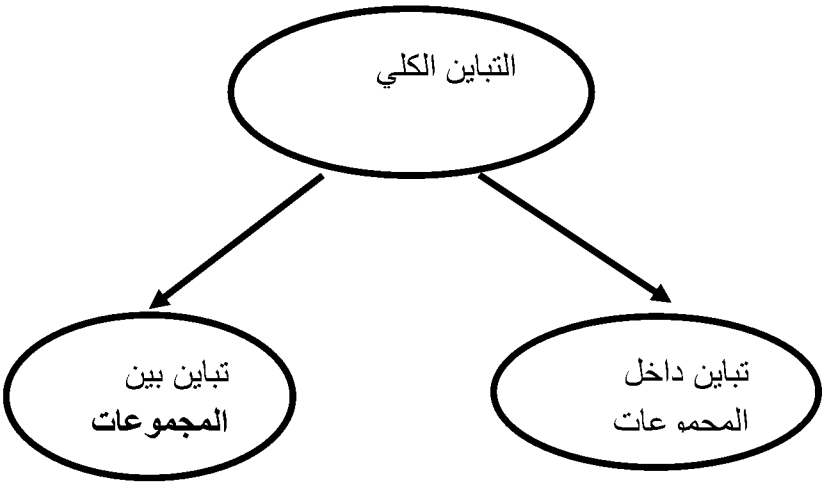
- σ_1^2 تباين مجتمع المجموعة الأولى.
- σ_2^2 تباين مجتمع المجموعة الثانية.
- σ_k^2 تباين مجتمع المجموعة K.

وان تحليل التباين يجزئ التباين الكلى Total variability(SS_{total}) للبيانات إلى مصدرين أساسين:

1. التباين أو الاختلاف داخل كل مجموعة ويسمى مجموع المربعات داخل المجموعات أو الخطأ Within-groups Sum of Squares (SS_{within})

2. التباين بين المجموعات ويطلق عليه مجموع المربعات بين المجموعات Between-groups Sum of Squares ($SS_{Between}$)

ويمكن عرض ذلك بالشكل الآتى:



الفروق نتيجة:

-عوامل غير منتظمة(الخطأ)

- عوامل عشوائية (فروق فردية)

الفروق نتيجة:

- تأثير المعالجة

تقدير التباين داخل المجموعات MS_{within}

أحيانًا يشار اليه بتباين الخطأ أو البواقى MS_e، وفى تحليل التباين فإن التباين داخل المجموعات:

$$MS_W = \frac{SS_1 + SS_2 + SS_3 + \cdots . + SS_k}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) \ldots . + (n_k - 1)}$$

- SS تباين درجات المجموعة (مجموع المربعات).
- K عدد المجموعات.

ويمكن تبسيط المعادلة السابقة كالآتى:

$$MS_W = \frac{SS_1 + SS_2 + SS_3 + \cdots . + SS_k}{N - K}$$

حيث N حجم العينة الكلى فى كل المجموعات:

$$N = n_1 + n_2 + n_3 + n_4 + \ldots . + n_k$$

وبسط المعادلة السابقة يرمز له بالرمز SS_W كما أن N-Kهى درجات الحرية الأولى (داخل المجموعات أو الخطأ أو البواقى) فى تحليل التباين وعليه:

$$MS_W = \frac{SS_w}{df_w}$$

حيث df_w درجات الحرية داخل المجموعات:

$$df_w = N-K = df_1 + df_2 + df_3 + \ldots . + df_k$$

$$df_1 = n_1 - 1 \ , \ df_2 = n_2 - 1$$

و df_w يطلق عليها df_{error} أو درجات الحرية للخطأ.

حيث مجموع المربعات داخل المجموعات:

$$SS_w = SS_1 + SS_2 + SS_3 + \cdots + SS_k$$

وتقدر مجموع المربعات داخل المجموعات من الدرجات الخام كالآتى:

$$SS_w = \sum X^2 - \left(\frac{(\sum X_1)^2}{n_1} + + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} + \cdots . + \frac{(\sum X_k)^2}{n_k}\right)$$

- $\sum x^2$ (T^2) مجموع مربعات كل الدرجات.
- $(\sum X_1)^2$ مربع مجموع درجات المجموعة الأولى.
- $(\sum X_2)^2$ مربع مجموع درجات المجموعة الثانية وهكذا.
- $\sum x^2$ إلى ويشار بـ T^2

تقدير التباين بين المجموعات $MS_{Between}$

التباين الثانى فى تحليل التباين هو قائم على الاختلاف بين المجموعات وعليه فالمعادلة العامة من الدرجات الخام كالآتى:

$$SS_B = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \cdots + \frac{(\Sigma X_K)^2}{n_3} - \frac{(\Sigma X)^2}{N}$$

- $(\Sigma X_1)^2$ مربع مجموع درجات المجموعه الأول.
- $(\Sigma X_2)^2$ مربع مجموع درجات المجموعه الثانية.
- $(\Sigma X)^2$ مربع مجموع درجات كل المجموعات
- ويشار إلى ΣX بـ T .

ويطلق على درجات الحرية بين المجموعات بدرجات الحرية المعالجة:

$$df_{between} = df_B = K - 1$$

التباين الكلى SS_{total}

يقدر التباين الكلى من الصيغة الآتية:

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

- ΣX^2 مجموع مربعات كل درجات العينات المتضمنة فى التحليل.
- $(\Sigma X)^2$ مربع مجموع كل درجات العينات.

النسبة F (The F Ratio)

كما سبق توضيحه كلما زادت قيمة التباين بين المجموعات MS_B دل ذلك على تأثير كبير للمتغير المستقل أو المعالجة، وان مسلمه تحليل التباين هو ان المتغير المستقل يؤثر فقط على التباين بين المعالجات وليس تباين كل مجموعة ولذلك فإن التباين داخل المجموعات لا يتغير مع تأثير المتغير المستقل، وتقدر قيمة F المحسوبة من الصيغة الآتية:

$$F = \frac{MS_B}{MS_W}$$

حيث:

$$MS_B = \frac{SS_B}{(K-1)}$$

$$MS_w = \frac{SS_w}{(N - K)}$$

لذلك فإن القيمة الكبيرة لـF تعنى تأثير حقيقى كبير للمتغير المستقل.

ويكون تلخيص جدول F كالتالى :

مصدر التباين	مجموع المربعات SS	df	متوسط مجموع المربعات MSS	قيمه F
بين المجموعات	SS_B	$K-1$	$MSS_B = \frac{SS_B}{K-1}$	$F = \frac{MSS_B}{MSS_W}$
داخل المجموعات	SS_W	$N-K$	$MSS_W = \frac{SS_W}{N-k}$	
الكلى	SS_T	$N-1$		

المشاكل الناتجة من استخدام اختبار T للمقارنة بين كل مستويين من المستويات المتعددة للمتغير المستقل

لو استخدم الباحث اختبار T للمقارنة بين كل مستويين من مستويات المتغير المستقل المتعددة عند مستوى ألفا 0.05 فإنه توجد مشكله وهى ان احتمال ارتكاب الخطأ من النوع الأول يتضخم عما افترضه الباحث، حيث بإجراء سلسلة من اختبارات T بدلًا من اختبار F فإنه يحدث تضخم للخطأ من النوع الأول بطريقه غير مباشرة وهذا التضخم للخطأ من النوع الأول يتحدد من المعادلة الآتية:

$$= 1 - (1 - \alpha)^{k}$$

حيث k عدد اختبارات T المستقلة.

فاذا كان لدينا متغير مستقل بخمسه مستويات وهذا يستلزم إجراء 10 اختبارات T المستقلة بدلًا من اختبار واحد لـF فإن درجه تضخم $\alpha = 0.05$ تكون كالآتى :

$$= 1 - (1 - 0.05)^{10} \quad = 0.40$$

واحتمال ارتكاب الخطأ من النوع الأول للمقارنات بين المجموعات الخمسة ليس 0.05 كما هو مفترض انما 0.4 . وعليه فعلى الباحث التميز بين Test wise Level alpha وهى المخاطرة لعمل الخطأ من النوع الأول للاختبارات الانفرادية لـ T فعلى سبيل المثال فى تجربه ما تضمنت ثلاثة معالجات بالتالى فهى تتطلب ثلاثة اختبارات لـT لمقارنة كل الفروق بين المتوسطات:

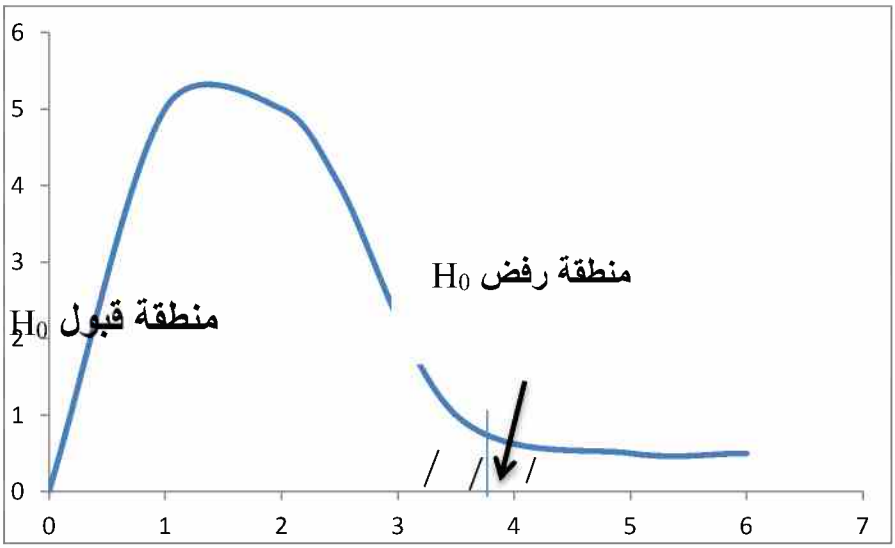
الاختبار الأول: مقارنة المعالجة الأولى مع الثانية، الاختبار الثانى: مقارنة المعالجة الأولى مع الثالثة، الاختبار الثالث: مقارنة المعالجة الثانية مع الثالثة. وعليه فإن ارتكاب الخطأ من النوع الأول يزيد:

$$= 1 - (1 - 0.05)^3$$

ولذلك فإن اختبار F يستخدم لتجنب هذه الاشكالية وهى تضخم الخطأ من النوع الأول.

التوزيع العينى لاختبار ANOVA

يعتبر التوزيع العينى لـ F غير اعتدالى حيث إنه يكون ملتويًا ناحية اليمين.



ويتم البحث فى جداول F فى ضوء درجتى حرية بين المجموعات تكون فى صف جدول F، df = K - 1(البسط) ودرجات حريه داخل المجموعات عمود جدول F، df = N - K(المقام) وبمستوى الدلالة الإحصائية المحددة 0.05 ويتم وضع القيمة الجدولية أو الحرجة على المنحنى لتحديد منطقتى القبول والرفض لـ H_0 . وتوجد

خاصيتين لهذا التوزيع هما قيمه F دائما موجبة، ولان البسط والمقام لـF يقيسوا نفس التباين للفرض الصفرى الحقيقى فإن قيمه F تكون قريبة من الواحد الصحيح وعلى ذلك فالتوزيع F فى المنطقة الموجبة (0 → 1,2,.....) وناحية اليمين.

اختبارات الفروض لقضيه بحثية(Pagano,2013)

قام باحث بقياس الضغوط النفسية لخمسه عشر طالب فى المرحلة الثانوية وقسمهم إلى ثلاث مجموعات، المجموعة الأولى تكونت عشوائيًا من 5 أفراد وقاس الضغوط النفسية لديهم بعد العودة من اجازة الصيف، والمجموعة الثانية 5 أفراد وقاس الضغوط فى منتصف الترم، والمجموعة الثالثة 5 أفراد وقاس الضغوط قبل امتحانات نهاية الترم واراد اختبار ما إذا كانت توجد فروق بين المجموعات الثلاثة فى الضغوط عند 0.05 ؟

الخطوات البحثية

1. **سؤال البحث: هل** توجد فروق بين درجات المجموعات الثلاثة فى الضغوط ؟
2. **فرض البحث:** توجد فروق فى درجات الضغوط بين المجموعات الثلاثة أو فى المواقف الثلاثة.
3. **منهج البحث:** يستخدم تحليل التباين الاحادى الاتجاه لتحليل بيانات متولدة من نوعية دراسات مختلفة مثل: الدراسات التجريبية، والدراسات شبه التجريبية، فى القضية السابقة فإن الدراسة تقع ضمن المنهج شبه التجريبى.
4. **المتغيرات فى تحليل التباين احادى الاتجاه:** يوجد متغيرين احداهما مستقل بمستويين فأكثر ويوجد متغير تابع واحد. توقيت التطبيق (بعد الإجازة، منتصف الترم، نهاية الترم) وهو متغير مستقل – اسمى – بثلاث مستويات والضغوط: تابع – فترى – متصل.
5. **الإحصاء المستخدم:** إحصاء النموذج البسيط البارامترى و**الاختبار المناسب** تحليل التباين الاحادى.

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية: $H_0: \mu_1 = \mu_2 = \mu_3$

μ_1 متوسط مجتمع العينة أو المجموعة الأولى (العودة من الإجازة)، μ_2 متوسط مجتمع العينة الثانية، μ_3 متوسط مجتمع العينة الثالثة.

$$H_A: \mu_1 \neq \mu_2 = \mu_3$$

$$H_A: \mu_1 \neq \mu_2 \neq \mu_3 \text{ او}$$

توجد فروق على الأقل بين متوسط أحد العينات الأولى والعينات الأخرى أو بين متوسطات العينات الثلاثة.

2. الاختبار الإحصائى ومسلماته: الاختبار هو F:

$$F = \frac{MS_B}{MS_{W(error)}}$$

فى مجلدات يشار إلى داخل المجموعات بالخطأ وكأن تقول مجموع مربعات الخطأ أو درجه حرية الخطأ وهذا يكون فى ضوء التعبير عن الدرجة المقاسة فى ضوء النموذج البنائى Structural Model حيث يتعامل مع الدرجة ليست تامة الثبات بل تتضمن جزء من الخطأ.

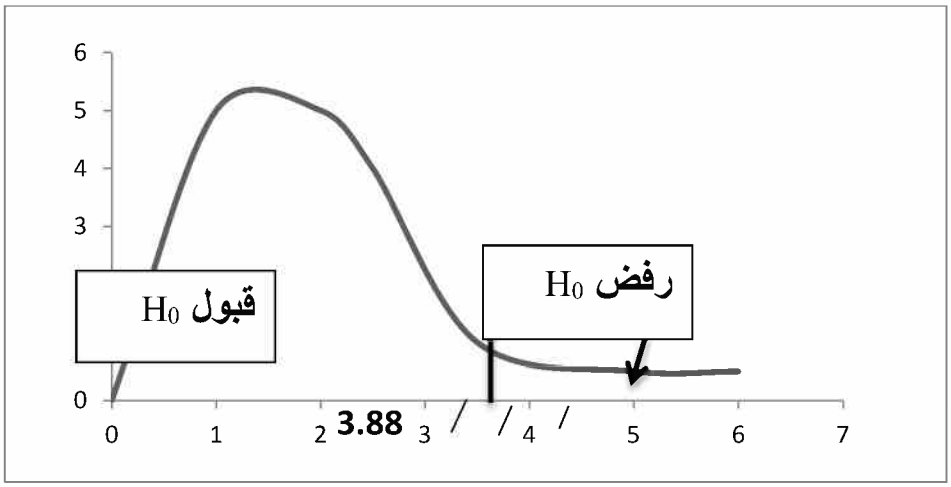
ودرجات الحرية: $df_B = K - 1 = 3 - 1 = 2$

$$df_W = N - K = 15 - 3 = 12$$

$$df_{total} = N - 1 = 15 - 1 = 14$$

3. مستوى الدلالة الإحصائية وقاعده القرار: لمستوى دلاله إحصائية 0.01 ولدرجتى حريه (البسط – الصف) $df_B = 2$ و(المقام – العمود) $df_W = 12$ وبالبحث فى جداول F (ملحق) فان:

$F_{C.V}$ الحرجة =3.88 وعليه فإن منطقتى الرفض لـH_0



وعليه إذا كانت: **F المحسوبة> F الحرجة(3.88)**، نرفض H_O

- **الحسابات:**

المجموعة 1		المجموعة الثانية 2		المجموعة الثالثة	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
2	**4**	**10**	**100**	**10**	100
3	**9**	**8**	**64**	**13**	169
7	**49**	**7**	**49**	**14**	196
2	**4**	**5**	**25**	**13**	169
6	**36**	**10**	**100**	**15**	225
Σ20	**102**	**40**	**338**	**65**	859
$n_1 = 5$		**$n_2 = 5$**		$n_3 = 5$	
$\bar{x}_1 = 4$		**$\bar{x}_2 = 8$**		$\bar{x}_3 = 13$	

- **مجموع كل الدرجات لكل المشاركين فى المجموعات:**

$$\Sigma x = T = 2+3+7+2+6+10+8+ \ldots\ldots +13+15= 125$$

حيث T مجموع درجات لأفراد فى المجموعات الثلاثة .

- **مجموع مربع كل الدرجات لكل المشاركين فى المجموعات:**

$$\Sigma X^2 = T^2 = 4+9+49+4+36+100+ \ldots\ldots +169+225 = 1299$$

حيث T^2مجموع مربعات كل الدرجات للعينه الكلية .

$$\bar{X} = \bar{T}1 = \frac{T1}{N} = \frac{\text{مجموع كل الدرجات}}{\text{حجم العينه الكلى}}$$

$$= \frac{125}{15} = 8.333$$

- **حساب مجموع المربعات بين المجموعات SS_B:**

$$SS_B = [\ \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3}] - \frac{(T)^2}{N}$$

$$= \left[\frac{(20)^2}{5} + \frac{(40)^2}{5} + \frac{(65)^2}{5}\right] - \frac{(125)^2}{15}$$

$$=203.333$$

- **حساب مجموع مربعات داخل المجموعات (الخطأ) SS_W:**

$$SS_W = (T)^2 - [\ \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3}]$$

$$= 1299 - \left[\frac{(20)^2}{5} + \frac{(40)^2}{5} + \frac{(65)^2}{5}\right] = 54$$

- **حساب مجموع المربعات الكلى:**

$$SS_{Total} = SS_W + SS_B$$

او يمكن تقديره من الصيغة الآتية:

$$SS_{Total} = T^2 - \frac{(T)^2}{N}$$

$$= 1299 - \frac{(125)^2}{15} = 257.333$$

وعليه: $SS_{Total} = 257.333$

$$257.333 = 54 + 203.333$$

وتُحسب SS_B كالآتى: $SS_B = SS_T - SS_W$

بينما $SS_W = SS_T - SS_B$

- **حساب التباين بين المجموعات:**

$$\mathrm{MS}_B = \frac{\mathrm{SS}_B}{\mathrm{df}_B} = \frac{203.333}{2} = 101.667$$

- **حساب التباين داخل المجموعات:**

$$MS_W = \frac{\mathrm{SS}_W}{\mathrm{df}_W} = \frac{54}{12} = 4.50$$

- **حساب نسبة F :**

$$F = \frac{MS_B}{MS_W} = \frac{101.667}{4.50} = 22.59$$

5. القرار والتفسير: القيمة المحسوبة لـ**F (22.59) > القيمة الحرجة لـF 3.88**، وعليه يرفض الباحث الفرض الصفرى وبالتالى توجد فروق ذات دلالة إحصائية فى متوسط درجات الضغوط النفسية بين المجموعات الثلاثة عند 0.05.

6. حجم التأثير: يتم حساب حجم التأثير فى تحليل التباين من عدة مؤشرات كالآتى:

ا. **مؤشر ايتا η^2** أو **r^2** : ويقدر من الصيغة الآتية:

$$\eta^2 = \frac{\mathrm{SS}_B}{\mathrm{SS}_T}$$

ويشار إلى هذا المؤشر بنسبه الارتباط :

$$\mathrm{r}^2 = \eta^2 = \frac{203.333}{257.333} = 0.790$$

وهذا يعنى ان المتغير المستقل فسر 79.0 % من تباين درجات الضغوط.

ويمكن تقدير η^2 من الصيغة الآتية:

$$\eta^2 = \frac{1}{1 + \frac{\mathrm{df}_W}{\mathrm{F} \times \mathrm{df}_B}}$$

$$= \frac{1}{1 + \frac{12}{22.59 \times 2}} = 0.79$$

واشار Howell (2013) إلى أنه توجد صيغة لـη^2 بحيث تكون مقياس لحجم التأثير وليس كنسبة تباين مفسر وتسمى Percent Reduced in Error (RRE)

ب– مؤشر اوميجا تربيعω^2 : البديل الأفضل ل η^2 لتقدير حجم التأثير أو تقدير حجم التأثير التجريبى هو مؤشر ω^2 . ويوجد صيغتين لـω^2 اعتمادًا على ماذا كان المتغير المستقل مثبت Fixed أو عشوائى Random . والمتغير العشوائى هو نادر الاستخدام فى التصميمات التجريبية احادية الاتجاه ولكنه فى غاية الأهمية فى التصميمات العاملية. ولنموذج تحليل التباين المثبت Fixed Model analysis of variance يقدر حجم التأثير التجريبى كالآتى:

$$\omega^2 = \frac{SS_B - (K-1)MS_W}{SS_{tot} + MS_W}$$

لاحظ أن مؤشر ω^2 يتميز عن مؤشر η^2 فى الآتى:

- يصحح مؤشر η^2 بتضيمن متوسط مربعات الخطأ داخل المجموعات فى المعادلة.
- يصحح عدد المجموعات بإضافة درجات الحرية بين المجموعات df_B فى بسط المعادلة.

وعلى ذلك يقدر للمثال السابق كالآتى:

$$\omega^2 = \frac{203.333 - (3-1)4.500}{257.333 + 4.500} = 0.742$$

وعلى ذلك فإن مواقف الضغوط فسرت 74.2 % من تباين درجات الضغوط وهذا حجم تأثير كبير.

وحدود قيم η^2 و ω^2 هى:

النسبة أو القيمة	التفسير
0.01 – 0.05	صغير
0.06 – 0.13	متوسط
≥ 0.14	كبير

لاحظ أن قيمة η^2 (0.792) أكبر قليلًا من قيمهω^2 (0.742) لأن مؤشر ω^2 يعطى تقدير أكثر دقة ولذلك يوصى باستخدامه عن مؤشر η^2 على الرغم ان الفروق بينهما ضئيلة جدًا.

ج – مؤشر d : هذا المؤشر طرحه Steiger (2004) ويسمى Root – Mean square Standardized effect (RMSSE) ويقدر كالآتى :

$$d = RMSSE = \frac{(\mu_j - \mu_G)}{\sigma}$$

- μj متوسط مجتمع كل مجموعة.
- μG متوسط متوسطات مجتمعات المجموعات Grand Mean.
- σ الانحراف المعيارى .

وهذه الصيغة فى المجتمع كالآتى :

$$d = \sqrt{\left(\frac{1}{K-1}\right) \Sigma \left(\frac{(\mu_j - \mu_G)}{\sigma}\right)^2}$$

وصيغتها فى العينات:

$$d = \sqrt{\left(\frac{1}{K-1}\right) \left(\frac{\Sigma (X_j - \bar{X}_G)^2}{MS_W}\right)}$$

وفى المثال السابق يوجد ثلاثة متوسطات هى 4 , 8 , 13 بالتالى فإن متوسطهم هو:

$$= \frac{13+4+8}{3} = \frac{25}{3} = 8.33$$

ويتم تقدير البسط كالآتى:

$$\Sigma(X_j - \bar{X}_G) = \Sigma (13 - 8.33)^2 + (8 - 8.33)^2 + (4 - 8.33)^2$$

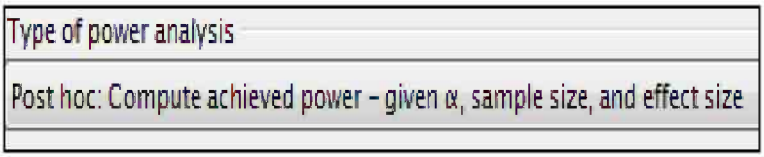
وعليه فإن:

$$d = \sqrt{\frac{1}{3-1} \left(\frac{(1.77)^2 + (0.33)^2 + (-3.77)^2}{4.50}\right)} =$$

7. تقدير القوة الإحصائية لـ ANOVA باستخدام برنامج G-Power

لتحديد القوة الإحصائية للمثال السابق باستخدام برنامج G-Power اتبع الاتى:

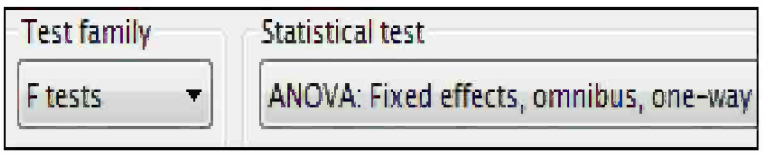
1. افتح البرنامج تظهر الشاشة الافتتاحية.
2. أسفلType of power analysisاختار :



لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

3. أسفلTest family اختر F- test

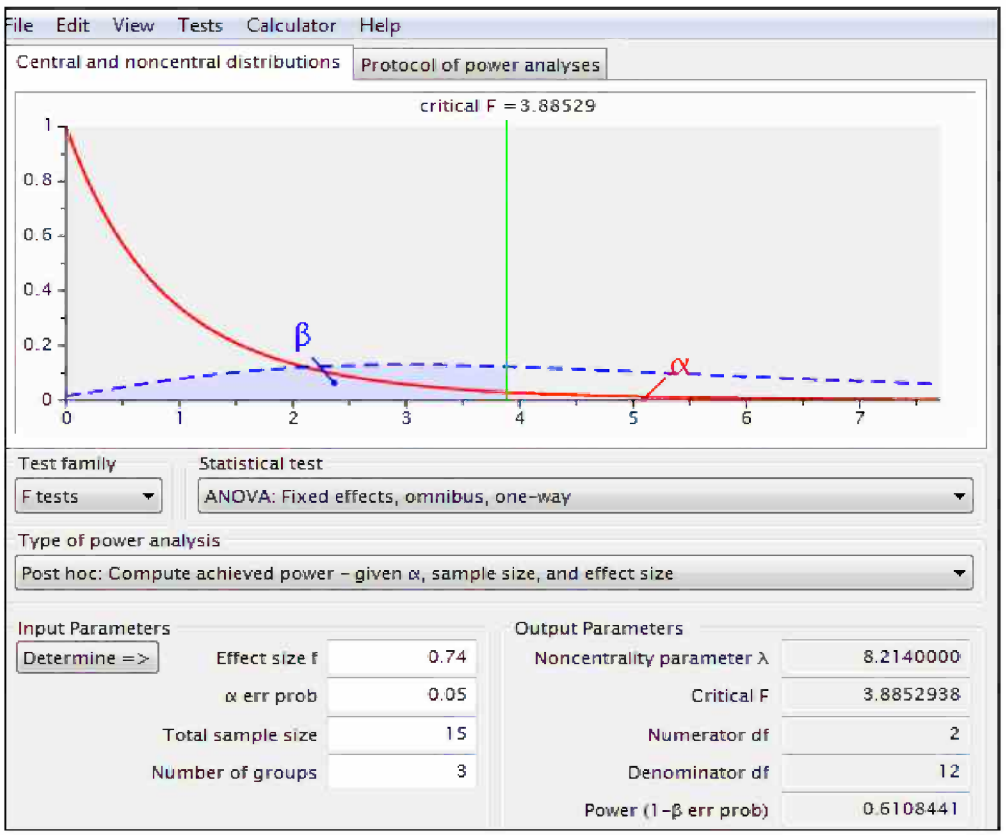
4. أسفلStatistical test اختر :



5. أدخل المعالم Input parameters:

- حجم التأثير المقدر= 0.74 (حجم تأثير كبير)، مستوى الدلالة الإحصائية، $\alpha = 0.05$ ، عدد المجموعات K=3 ، حجم العينة= 15

6.اضغط Calculated تظهر شاشة الناتج كالآتى:



لاحظ تحت output parameters قيمة F الحرجة = 3.88 ودرجات الحرية البسط(بين المجموعات) Numerator df = 2 بينما درجات الحرية المقام(داخل المجموعات) والقوة الإحصائية للاختبار = 0.611 وهذا مستوى قوة غير مناسب وعلى ذلك فإن عدم رفض الفرض الصفرى يكون نتيجة عدم قدرة الاختبار على رفضة لانخفاض القوة الإحصائية.

ولكن فى المدخلات البرنامج اعتمدنا على مؤشر ω^2 ولكن البرنامج يعتمد على مؤشر f وبحساب هذا المؤشر كالآتى:

$$f = \sqrt{\frac{(k-1)(F-1)}{Nk}}$$

$$f = \sqrt{\frac{(3-1)(22.59-1)}{15 \times 3}} = \sqrt{\frac{2 \times 21.59}{45}} = 0.97$$

وبتعديل حجم التأثير فى المدخلات من 0.74 إلى 0.97 فإن المخرجات كالآتى:

Input Parameters			Output Parameters	
Determine =>	Effect size f	0.97	Noncentrality parameter λ	14.1135000
	α err prob	0.05	Critical F	3.8852938
	Total sample size	15	Numerator df	2
	Number of groups	3	Denominator df	12
			Power (1-β err prob)	0.8477721

وعلى ذلك فإن القوة الإحصائية = 0.8477 وهى مستوى مناسب وجيد وانعكس ذلك على رفض الفرض الصفرى.

القوة الإحصائية لتحليل التباين

القوة الإحصائية لتحليل التباين تتأثر بنفس المتغيرات أو العوامل التى تؤثر على اختبار T وهى:

- تتغير القوة بتغير حجم العينة N وتزيد القوة تزيد.
- ترتبط القوة طرديًا مباشره مع حجم التأثير الحقيقى للمتغير المستقل كلما كان الحجم الحقيقى كبير تزيد القوة.
- القوة ترتبط عكسيًا مع تباين العينة، التباين المرتفع للعينة يؤدى إلى قوة المنخفضة للكشف عن الفروق الحقيقية.

القوة وحجم العينة: كلما زاد عدد الأفراد فى التجربة يؤدى إلى زياده قيمه F المحسوبة، حيث بزيادةN فإن مقام معادله F (MS_W) يقل وعليه تزيد قيمه F وزيادة N فى بسط F له تأثير ايجابى على زيادة قيمة (MS_B) وهذا بدوره يؤدى إلى زياده F وعليه تزيد احتماليه رفض H0 وبالتالى زياده القوة الإحصائية للاختبار.

القوة والتأثير الحقيقى للمتغير المستقل (حجم التأثير) :التأثير الحقيقى الكبير للمتغير المستقل يعطى قيم كبيره لـ$(\overline{X}_1 - \overline{X}_G)$و$(\overline{X}_2 - \overline{X}_G)$و$(\overline{X}_3 - \overline{X}_G)$ وعليه يؤدى إلى زيادة قيمه MS_B (البسط) وبدوره يزيد من قيمه F المحسوبة وهذا يؤدى إلى احتمال كبير لرفض H0 بالتالى تزيد القوة للاختبار.

القوه وتباين العينة: SS_1 هو مجموع مربعات المجموعة الأولى، SS_2 هو مجموع مربعات المجموعة الثانية، SS_3 هو مجموع مربعات المجموعة الثالثة وهم مؤشرات لقياس التباين داخل كل مجموعة وعليه فإن زيادةSS_1 و SS_2 و SS_3 يؤدى إلى زيادة المقام لـF وهو متوسط مجموع المربعات داخل المجموعات (الخطأ) وعليه فإن زيادة MS_W يؤدى إلى نقصان F المحسوبة وعلى ذلك فإن زيادة التباين داخل المجموعات يؤدى إلى نقصان القوة .

تحديد حجم العينة لـ ANOVA

لتوضيح تحديد حجم العينة لاختبار **ANOVA** فإنه يستلزم تحديد المعطيات الآتية:

1. **مستوى الدلالة الإحصائية (α):** يوصى دائمًاً ان تكون نسبة الخطأ من النوع الأول إلى الخطأ من النوع الثانى 4:1 وعليه فإن وضع α= 0.05 فإن β= 0.20

2. **تحديد مستوى القوة الإحصائية المرغوب:** $P = 1 - 0.20 = 0.80$ وهذا المستوى يوصى به (1988) Cohen كأفضل مستوى مرغوب ويفضل الا يقل عن 0.80.

3. **تحديد أدنى حجم تأثير متوقع:** يحدد فى ضوء الدراسات السابقة، أو ماذا يتوقع الباحث من تأثير حقيقى للمتغير المستقل بفرض أن الدراسات السابقة لحجم تأثير ثلاثة طرق للمعالجة على متغير تابع هى ES=0.74. ومن ثم إذا اعتبرنا:

$$P = 0.80 \, , ES = 0.74 \, , \alpha = 0.05$$ وعدد المجموعات K=3

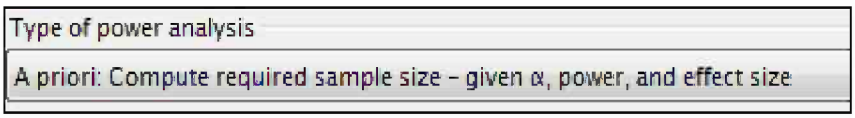
فانه يتم تحديد حجم العينة فى ضوء مدخلين:

1. **مدخل جداول Cohen (1988):** فإن حجم العينة المتطلب لتحقيق هذا هو 40 لكل مجموعة وعليه فإن حجم العينة الاجمالى لتنفيذ التجربة: $4 = X\ 40 = 160$

2. المدخل الثانى هو استخدام التحليل القبلى للقوة الإحصائية Prior power باستخدام برنامج G.Power ولتحديد حجم العينة المتطلب لدراسة الفروق بين المجموعات الثلاثة فى الضغوط النفسية (المثال السابق):

1. افتح البرنامج تظهر الشاشة الافتتاحية:

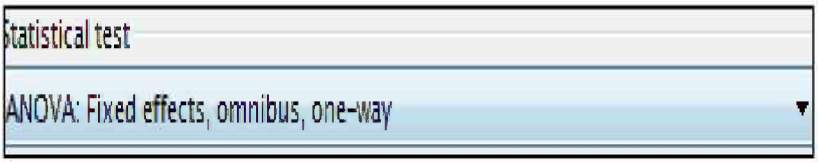
2. أسفل Type of power analysis اختار



لاحظ اعتمدنا على تحليل القوة القبلى Aprior وليس مدخل Post hoc اثناء تقدير القوة الإحصائية للاختبار.

3. تحت Test family اختر F- test

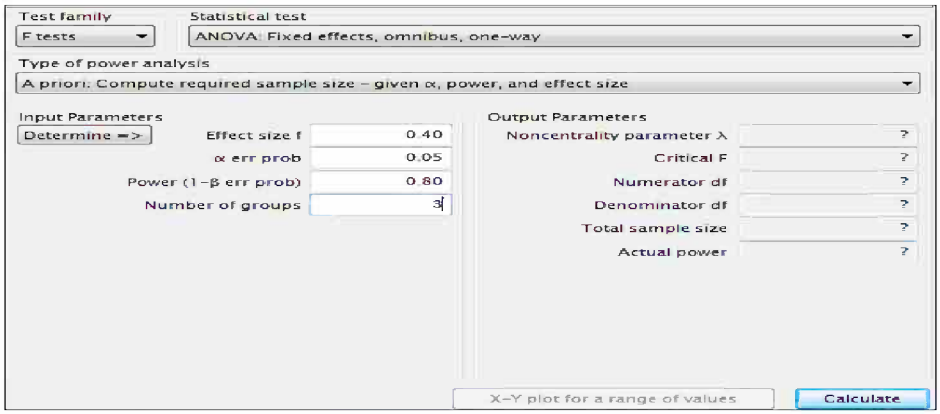
4. تحت Statistical tests اختر:

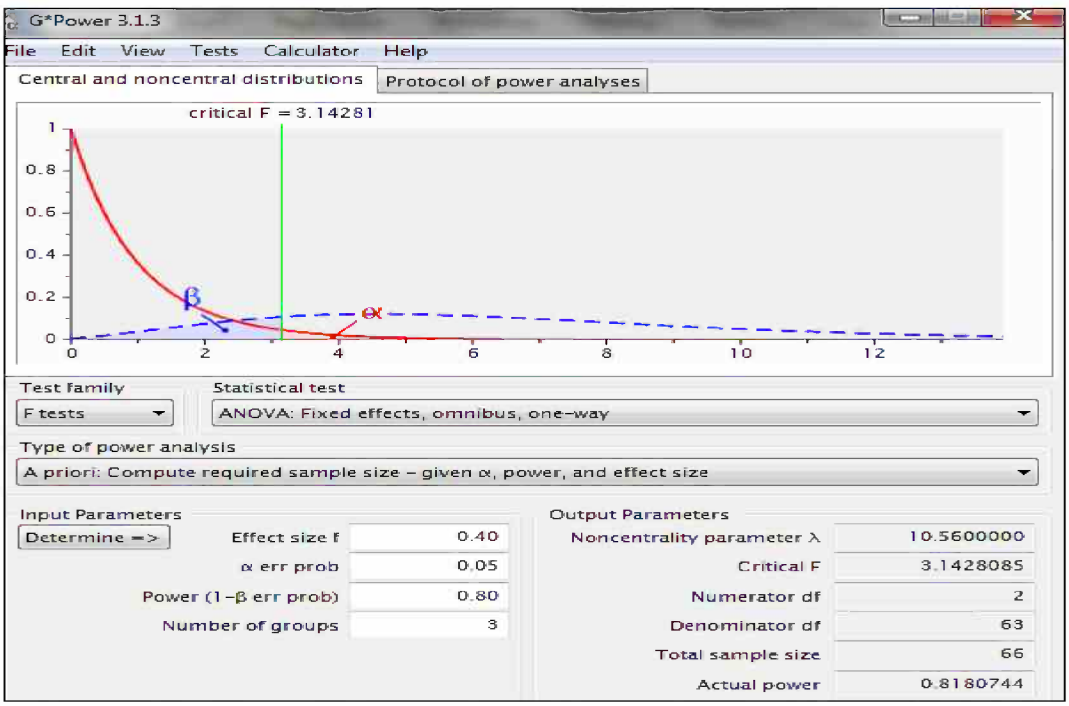


5. ادخل المعالم Input parameters:

- حجم التأثير وفقا لمؤشر f لكوهين التى يتبناها البرنامج= 0.40(حجم تأثير كبير)،
  القوة الإحصائية= 0.80(الحد الأدنى)
- مستوى الدلالة الإحصائية $\alpha = 0.05$
- عدد المجموعات K=3

6.اضغط Calculated تظهر شاشة الناتج كالآتى:





لاحظ أن حجم العينة الاجمالى 66 مقسمة إلى ثلاثة مجموعات وعلى ذلك فإن كل مجموعة تتضمن 22 فردا وتكون قيمة F=3.14 Critical والقوة الفعلية=0.81. وعليك ان تقارن بين حجم العينة فى مدخل Cohen ومدخل G.Power

العلاقة بين اختبار T المستقلة و ANOVA

عندما يقيم الباحث الفروق بين متوسطى مجتمعين اثنين فقط لعينات مستقلة فإنه يستخدم اختبار T المستقلة وكذلك يمكن استخدام تحليل التباين الاحادى **ANOVA**، ولا يوجد فروق بينهما حيث يؤدوا إلى نفس الاستنتاج بشان قبول أو رفض Ho . والعلاقة بين F و T:

$$F = T^2$$

وحيث إن اختبار T يقارن المسافة Distance بين متوسطى العينة (البسط) والمسافة المحسوبة للخطأ المعيارى (المقام) $T = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}}}$ ، فى حين ان اختبار F يقارن بين تباين العينتين والتباين هو مقياس لمربع المسافة.

ويوجد العديد من الاعتبارات عند مقارنه إحصاء T بالنسبة الي F هى :

1. تختبر نفس الفروض عند وجود معالجتين فقط وهى:

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2$$

2. درجات الحرية لـTو df لمقام اختبار F (df_w) واحدة فاذا كانت معالجتين لكل معالجة ستة أفراد فإن df لـ T:

$$= n_1 + n_2 - 2 = 12 - 2 = 10$$

و df لمقام F (داخل المجموعات) هى:

$$df_w = N - K = 12 - 2 = 10$$

ودرجتى الحرية لـF هى df =(1,10)

3. يختلف توزيع T عن توزيع F فاذا كانت df لـT (18) و df لـF

(1,18) فـإن قيمـه F الحرجـة عنـد 0.05 وقيمـة T الحرجـة عنـد 0.05 كـالآتى :

$F = (2.10)^2 = (T\text{الحرجة})^2 = (4.41)$الحرجة

مثال لبيانات تقدر باستخدام T وباستخدام ANOVA

قضية (1994) Hinkle et al.:

اهتم باحث بإجراء مقارنة بين متوسط عدد المكالمات التى يتسلمها طلاب الجامعة المقيمين فى المدينة الجامعية ومقارنتها بعدد المكالمات لطلاب الجامعة المقيمين فى منازلهم وتم اختيار عينه عشوائية من المجتمعين العينة الأولى 10 طلاب والعينة الثانية10 طلاب عند $\alpha =$ 0.05

خطوات اختبارات الفروض باستخدام T, F: كما نعلم أنه يوجد متغيرين أحداهما مستقل وهو مكان الإقامة بمستويين والمتغير التابع نسبى وهو عدد المكالمات والفروض الإحصائية لـ**T** وايضًا لـ F:

$$H_0 : \mu_1 = \mu_2 \ , \ H_A : \mu_1 \neq \mu_2$$

وبحساب اختبار **T** للبيانات:

	طلاب المدينة	منازلهم
	2	4
	6	6
	3	4
	2	2
	4	12
	0	16
	2	4
	15	10
	1	0
	0	5
ΣX	**35**	63
ΣX^2	**299**	613
$\bar{X}$	**3.5**	6.3
S^2	**19.61**	29.01

$\Sigma T = 98$, $T^2 = 912$

$$\mathbf{T} = \frac{3.5 - 6.30}{\sqrt{\frac{(10-1)(19.61)+(10-1)(24.01)}{10+10-2}\left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{-2,80}{\sqrt{4.36}} = -1.34$$

ولاختبار ذو ذيلين و $\alpha = 0.05$ فان:

$$\mathbf{T}(\text{الحرجة})(18) = \pm\ 2.101$$

وبما ان: $1.34 < 2.10$ وبالتالى يفشل الاختبار فى رفض$H0$ بحيث لا توجد فروق.

وبحساب F كالآتى:

$$SS_B = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} - \frac{(\Sigma T)^2}{N}$$

$$= \frac{(35)^2}{10} + \frac{(63)^2}{10} - \frac{(98)^2}{20}$$

$$= 519.40 - 480.20 = 39.20$$

$$SS_w = \Sigma T^2 - \left[\frac{(\Sigma X_1)^2}{10} + \frac{(\Sigma X_2)^2}{10}\right]$$

$$= 912 - \left[\frac{(35)^2}{10} + \frac{(63)^2}{10}\right]$$

$$= 912 - 51.40 = 392.40$$

$$SS_T = \Sigma T^2 - \frac{(\Sigma T)^2}{N}$$

$$= 912 - \frac{(98)^2}{10} = 912 - 480.20 = 431.80$$

وعليه فإن:

$$MS_B = \frac{39.20}{1} = 39.20$$

$$MS_w = \frac{392.60}{18} = 21.81$$

وبالتالى:

$$F = \frac{39.20}{21.81} = 1.80$$

وبالبحث فى جدول F بدرجة حرية المقام 8 ودرجة حرية البسط 1 ومستوى دلالة إحصائية 0.05فان $F_{c.v}$ =4.41 الحرجة وعليه فإن:

F المحسوبة(1.80) < $F_{c.v}$ الحرجة(4.41)

اذًا نقبل الفرض H_0 بعدم وجود فروق فى عدد المكالمات بين الطلاب ساكنى المدن الجامعية والطلاب ساكنى منازلهم.

إجراء تصحيح بونيفرونى فى حالة إجراء F مرات عديدة فى نفس الدراسة

يمكن لباحث جمع بيانات لمتغيرين تابعين أو أكثر فى نفس الدراسة ويوجد على الأقل ثلاث مجموعات فيقوم الباحث بحساب اختبار F لكل متغير تابع على حدة فى الدراسة وهذا هو المألوف فى البحوث ويستخدم مستوى دلاله إحصائية واحدة لكل المقارنات للمتغيرات التابعة ولكن فى هذه الحالة يحدث تضخم للخطأ من النوع الأول خاصة إذا وجد ارتباطات بين المتغيرات التابعة وربما يصل الباحث إلى رفض غير صحيح (خاطئ) على الأقل لأحد الفروض الصفرية ولكن توجد فنيات عديدة للتعامل مع مخاطر تضخم الخطأ من النوع الأول وهو إجراء تصحيح بونيفرونى Bonferroni adjustment حيث يتم قسمة مستوى ألفا على عدد مرات إجراء ANOVA فى الدراسة فاذا تم إجراء تحليل التباين ثلاث مرات فإن قيمة ألفا المعدلة هى:$0.016=\frac{0.05}{3}=$

وإذا تم إجراء تحليل التباين مرتين فإن ألفا المعدلة هى:

$$=\frac{0.05}{02}=0.025$$

ويأخذ الباحث قراره فى ضوء α المعدلة وكن على علم لا بد من إجراء هذا خاصة إذا كان الباحث يريد حساب تحليل التباين لأبعاد مفهوم ما مثل الابتكارية (الاصالة–المرونة–الطلاقة) أو الدافعية وغيرها من المفاهيم متعددة الابعاد.

كتابة تحليل التباين فى تقرير البحث وفقا لـAPA

فى كتابة نتائج ANOVA فى تقرير البحث فلا بد من عرض متوسطات المجموعات والانحرافات المعيارية كتابة فى فقرة أو فى جدول. على الرغم ان هذه الإحصائيات لا تدخل فى حساب نسبه F ولكنها ضرورية. وفيما يلى تقرير الدراسة: " المتوسطات والانحرافات المعيارية تم عرضهما فى جدول ( ). اشارت نتائج تحليل التباين إلى

وجود فروق دالة إحصائيًا بين التوقيتات الثلاثة فى الضغوط: $F\ (2,12)\ =\ 22.59,\ P < 0.05, \eta^2 = 0.792$

او $F\ (2,12) = 22.59, P < 0.05, \omega^2 = 0.742$

والباحثون يفضلون كتابة ANOVA فى تقرير البحث كالآتى:

مصدر التباين	مجموع المربعاتSS	درجات الحرية df	متوسطات المربعات MSS	نسبة Fالمحسوبة
بين المجموعات	**203.33**	**2**	**101.667**	**22.59**
داخل	**54.00**	**12**	**4.500**	**-**
الكلى	**257.333**	**14**	**-**	**-**

* داله إحصائيًا عند $\alpha = 0.05$ ، $Fc.v = 3.88$ ويتم رفض H_0 .

وإذا تم عرض الجدول بالصيغة السابقة يمكن أن يتبعها الفقرة الآتية: كما هو واضح فى الجدول السابق وجود فروق فى الضغوط:

$$F\ (2,12\ ) = 22.59\ ،\ ،\ P = 0.00\ , \omega^2 = 0.742$$

تحذيرات

- **النتائج الدالة وغير الدالة إحصائيًا:** عندما يحصل الباحث على نتائج دالة إحصائيًا من تحليل التباين الاحادى فهذا يعنى رفض H_0 والشىء الوحيد للنتائج الدالة هو ان الاختلاف بين مجموعة من المتوسطات هو أكبر مما توقعنا لو ان متوسطات كل المجتمعات متماثلة أو متساوية، ولكن يريد الباحث ان يكون أكثر تفصيلًا أو خصوصية عن ما إذا كانت $\mu1$ أكبر من $\mu2$ أو أكبر من $\mu3$ أو أى زوج من المتوسطات سبب هذه الدلالة ولذلك فنحن يجب أن نذهب ابعد من الدلالة الإحصائية لـ ANOVA و لذلك سنتطرق فى الفصل القادم للاختبارات اللاحقة أو البعدية Post hoc or Follow up tests
- اعلم أهم شئ فى تحليل التباين وفى كل الأساليب الإحصائية الاستدلالية هو ثبات البيانات فنقص الثبات يجعل من الصعب رفضH_0 ونقص الصدق يجعل النتائج

غير مضمون تعميمها وهدف تحليل التباين هو التعميم على مستوى المجتمع من خلال متوسطات العينة.

- عدم رفض الفرض الصفرى لا يعنى ان متوسطات كل متوسطات المجتمع متساوية.
- يجب إجراء تصحيح بونيفرونى عند إجراء تحليلات عديدة لـANOVA فى نفس الدراسة خاصة لأبعاد المفاهيم لأنه يضبط الخطأ من النوع الأول.
- يجب الا تجرى تحليل تباين معيوب، فعليك التحقق من مسلماته وخاصة الاعتدالية وتجانس التباينات.
- إجراء التحليل القبلى للقوة كمدخل لتحديد حجم العينة حتى يدخل الباحث فى دراسته على ارضية قوية.

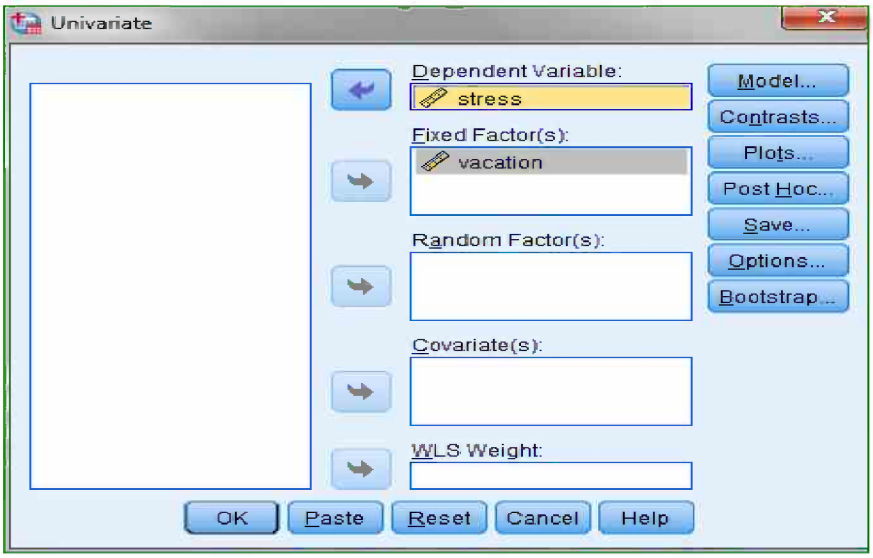
تنفيذ تحليل التباين ANOVA فىSPSS: مثال Pagano(2013)

أولا: إدخال البيانات: 1. أضغط Variable view، فى عمود Name حدد مسمى المتغيرات وهى كالآتى: **المتغير المستقل:** الأجازة Vacation ثم تكوده إلى ثلاث مستويات: (بعد العودة مباشرة من الأجازة (أول الترم)=1، منتصف الترم=2، نهاية الترم=3)، **المتغير التابع:** الضغوط النفسية Stress.

3. فى عمود Values حدد تعريف 1,2,3.

4. أضغط Data view تظهر شاشة البيانات بها عمودين. ابدأ فى إدخال المتغيرات.

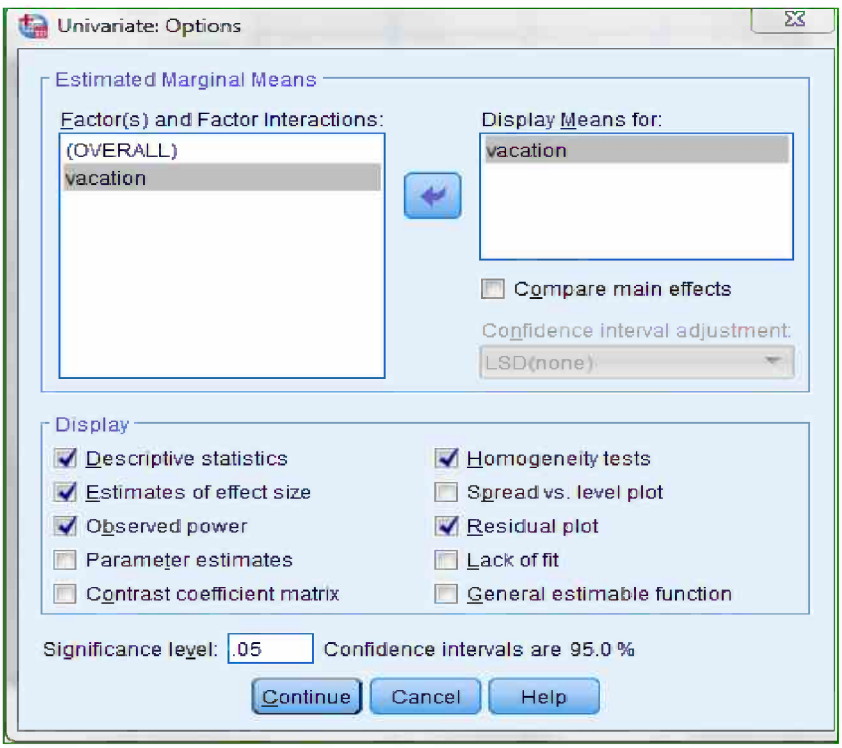
ثانيًا: تنفيذ الأمر: الطريقة الأولى: Analyze → General Linear Model →Univariate→تظهر الشاشة الآتية:



2. انقل المتغير التابع stress إلى مربع Dependent variable

3. انقل المتغير المستقل vacation إلى مربع fixed factor

4. اضغط اختيار Options على يمين الشاشة تظهر الشاشة الآتية:

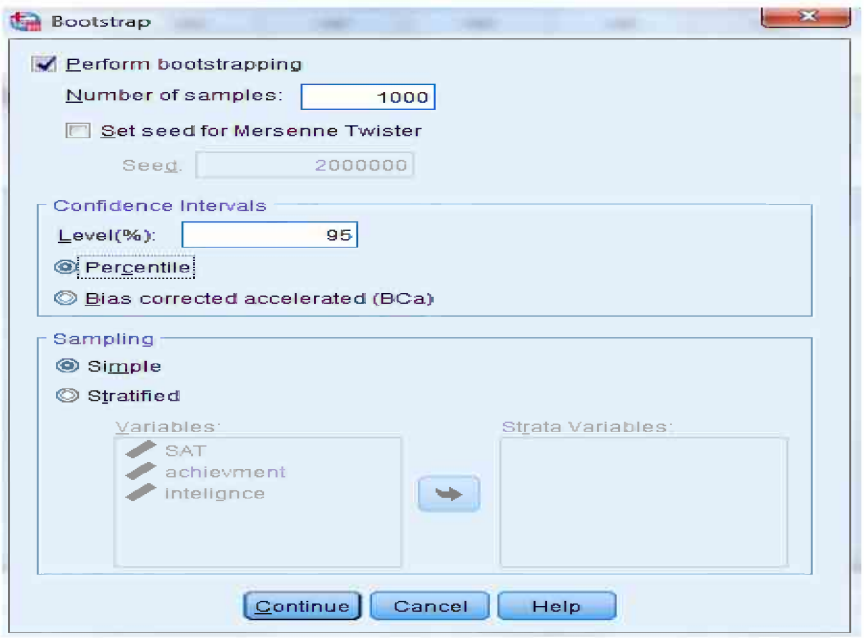


5. اضغط vacation فى مربع Factor(s)and factor Inter وانقله إلى مربع Display Means

6. اضغط على Estimate of effect size, Descriptive statistics,

Residual plot, Observed power, Homogeneity tests,

7. اضغط على Bootstrap تظهر النافذة الآتية:



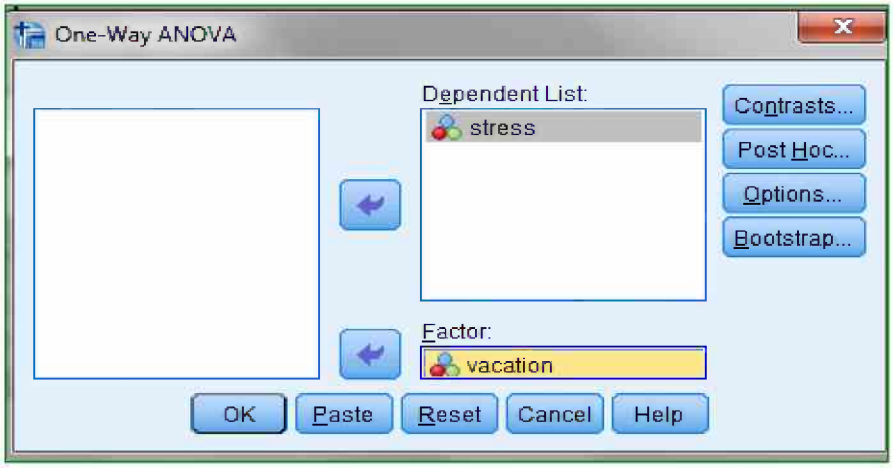
8. اضغط Perform bootstrapping :

- حدد عدد العينات التى ترغب فى توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 200 مثلًا، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التى ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهى بسيطة.

9.اضغط Continue ثم اضغط Ok

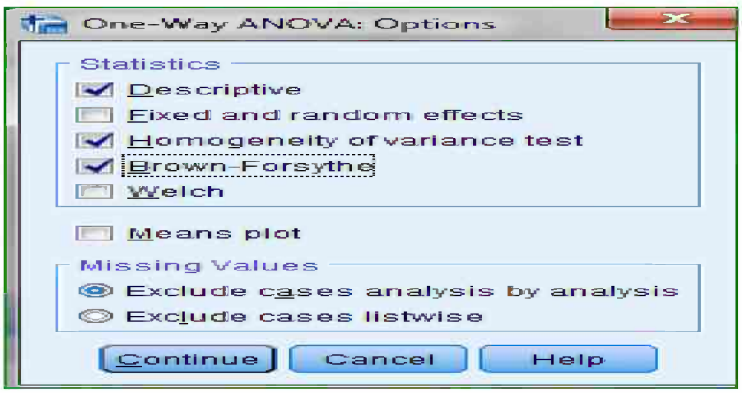
الطريقة الثانية:1.اضغطAnalyze→ Compare means →One way ANOVA

تظهر الشاشة الآتية:



2. انقل Stress إلى مربع Dependent list وانقل Vacation إلى مربع Factor

3. اضغط اختيار Options يعطى الشاشة الآتية:



- اضغط Descriptive, Homogeneity of variancetest, واختبارى Brown – forsythe , welch

4. اضغط Continue ثم اضغط OK

Between-Subjects Factors

		N
vacation	1.00	5
	2.00	5
	3.00	5

ثالثا: تفسير المخرج:

- أعطى **الجدول الأول:**

عدد أفراد كل عينة، حيث أفراد عينة بداية الترم=5 ومنتصف الترم=5 ونهاية الترم =5 أفراد

الجدول الثانى: الإحصاء الوصفى لدرجات كل عينة كالآتى:

Descriptive Statistics

Dependent Variable: stress

vacation	Mean	Std. Deviation	N
1.00	4.0000	2.34521	5
2.00	8.0000	2.12132	5
3.00	13.0000	1.87083	5
Total	8.3333	4.28730	15

ومن الواضح أن متوسط Mean درجات الضغوط للعينة نهاية الترم 13 أكبر من المتوسطات يليه متوسط عينة منتصف الترم.

وكذلك أعطى الانحراف المعيارى Std. Deviation لدرجات كل عينة.

الجدول الثالث: اختبار Levin's لاختبار تجانس البيانات.

Levene's Test of Equality of Error Variances[a]

Dependent Variable: stress

F	df1	df2	Sig.
.686	2	12	.522

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

والوضح أن قيمة F=0.686 ، $df_1=(c-1)=2, df_2=(15-3)=2$,

والدلالة الإحصائية sig=0.522 وبما أن $p > 0.05(\alpha)$ (0.522)، وعليه نقبل H_0

وبالتالى يوجد تساوى أو تجانس لتباينات العينات الثلاثة.

الجدول الرابع: اختبارات تأثيرات بين الأفراد:

Tests of Between-Subjects Effects

Dependent Variable: stress

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	203.333[a]	2	101.667	22.593	.000	.790	45.185	1.000
Intercept	1041.667	1	1041.667	231.481	.000	.951	231.481	1.000
vacation	203.333	2	101.667	22.593	.000	.790	45.185	1.000
Error	54.000	12	4.500					
Total	1299.000	15						
Corrected Total	257.333	14						

a. R Squared = .790 (Adjusted R Squared = .755)

b. Computed using alpha = .05

والواضح فى هذا الجدول أن قيمة F المقابلة لـ Vacation هى22.59 $F=(\frac{101.667}{4.5})$ ،

و Error هى مجموع المربعات داخل المجموعات أو الخطأ أو البواقى بما أن Sig =0.000

وعليه 0.000 < 0.05، وبالتالى نرفض H_0 وعليه توجد فروق ذات دلالة إحصائية فى الضغوط بين المجموعات الثلاثة.

وأتضح أن مؤشر حجم التأثير (مربع إيتا) الجزئية هى: $\eta_p^2=0.790$ وهى حجم تأثير من النوع القوى بمعنى أن الأجازة فسرت 79 %من تباين تقدير الذات.

ولاحظ أسفل هذا الجدول أعطى : R square= 0.790, R^2_{adj} = 0.755، على اعتبار ان تحليل التباين هو جزء من تحليل الانحدار وأن قيمة R^2هى تقريبًا نفسها قيمة η_p^2 وهذا هو مؤشر لمطابقة النموذج Model fit.

وأعطى البرنامج الجدول الآتى:

Estimated Marginal Means

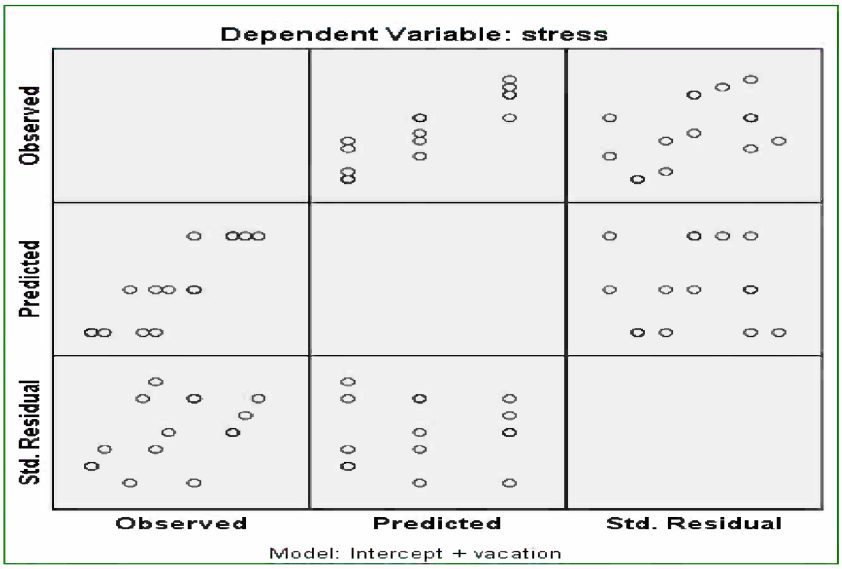
vacation

Dependent Variable: stress

vacation	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	4.000	.949	1.933	6.067
2.00	8.000	.949	5.933	10.067
3.00	13.000	.949	10.933	15.067

وهى متوسطات كل عينة والخطأ المعيارى للمتوسط.

ثم أعطى البرنامج scatter plot للمتغير التابع stress لكلٍّ من الدرجات المقاسة والدرجات المتنبأ بها وقيم البواقى المعيارية:



والمهم فى هذا العرض البيانى هو الصف الأوسط (الثانى) حيث تختبر مسلمة التوزيع العشوائى لأخطاء القياس Homoscedastiaty.

نتائج البوتستراب:

Bootstrap for Multiple Comparisons

Dependent Variable: stress

	(I) vacation	(J) vacation	Mean Difference (I-J)	Bootstrap[a]			
				Bias	Std. Error	95% Confidence Interval	
						Lower	Upper
Tukey HSD	1.00	2.00	-4.00000-	-.07893-[b]	1.35987[b]	-6.85046-[b]	-1.18715-[b]
		3.00	-9.00000-	.04024[b]	1.14749[b]	-11.20948-[b]	-6.38662-[b]
	2.00	1.00	4.00000	.07893[b]	1.35987[b]	1.18715[b]	6.85046[b]
		3.00	-5.00000-	.11917[b]	1.28958[b]	-7.63871-[b]	-2.45718-[b]
	3.00	1.00	9.00000	-.04024-[b]	1.14749[b]	6.38662[b]	11.20948[b]
		2.00	5.00000	-.11917-[b]	1.28958[b]	2.45718[b]	7.63871[b]
Sidak	1.00	2.00	-4.00000-	-.07893-[b]	1.35987[b]	-6.85046-[b]	-1.18715-[b]
		3.00	-9.00000-	.04024[b]	1.14749[b]	-11.20948-[b]	-6.38662-[b]
	2.00	1.00	4.00000	.07893[b]	1.35987[b]	1.18715[b]	6.85046[b]
		3.00	-5.00000-	.11917[b]	1.28958[b]	-7.63871-[b]	-2.45718-[b]
	3.00	1.00	9.00000	-.04024-[b]	1.14749[b]	6.38662[b]	11.20948[b]
		2.00	5.00000	-.11917-[b]	1.28958[b]	2.45718[b]	7.63871[b]

a. Unless otherwise noted, bootstrap results are based on 200 bootstrap samples

b. Based on 188 samples

لاحظ أن نتائج البوتسراب اعطت للمقارنة المتعددة واعطى فترات الثقة لمقارنة كل زوج على حدة.

أما بالنسبة لمخرج الطريقة الثانية:

فالمخرج أكثر بساطة حيث عرض اختبار Levin's Test

وكذلك جدول ANOVA كالآتى:

ANOVA

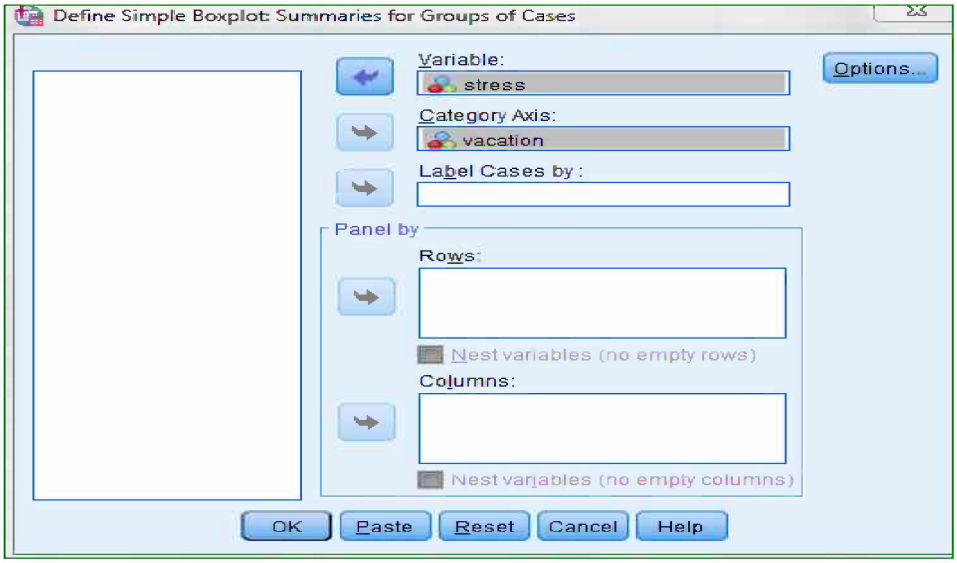
stress

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	203.333	2	101.667	22.593	.000
Within Groups	54.000	12	4.500		
Total	257.333	14			

عرض نتائج ANOVA بيانيًا

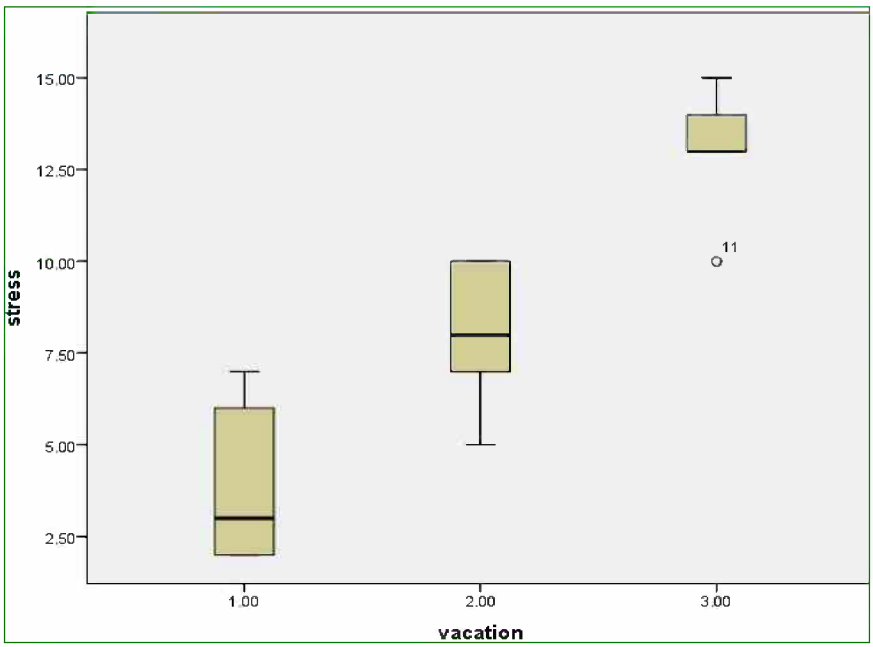
1. اضغط Graphics → Legacy Dialogs → Boxplot

2. اضغط Simple ثم Define تظهر الشاشة الآتية:



3. انقل stress إلى مربع Variable

4. انقل vacation إلى مربع CategoryAxis ثم اضغط Ok



نلاحظ أن توزيع الضغوط فى الفترات الثلاثة مختلف، حيث أعلاهم هو فى نهاية الترم يليه منتصف الترم ثم فى بداية الترم. وتعتبر قيمة الحالة 11 فى توزيع نهاية الترم 3 قيمة متطرفة وهى القيمة 10 ولكنها غير ذلك فى بداية ومنتصف الترم.

الفصل الحادى والعشرون

المقارنات البعدية المتعددة بين متوسطات المعالجات أو التحليل البعدي

Multiple Comparisons Among treatment means

or

Post hoc analysis

فى تحليل التباين الاحادى قيمة F الدالة إحصائيًا تشير إلى ان كل المعالجات أو مستويات المتغير المستقل ليس لها نفس التأثير على المتغير التابع، وفى تحليل التباين يتم اختيار الفرض الصفرى $\mu1=\mu2=\mu3$. بينما تكون قيمة الخطأ فى النوع الأول كما حددها الباحث مسبقًا ولتكن 0.05 أو 0.01 أو غيرها وهذا يسمى احيانًا بـ Omnibus test، ففى مثال الفصل السابق قيمة F المتحصل عليها جراء المقارنة بين ثلاثة مواقف لا تعنى ان مستوى الضغوط واحد فى هذه المواقف وبعد رفض H_0 فإننا نواجهه مشكلة فى تحديد أى زوج من المتوسطات المسئول عن احداث الدلالة الإحصائية ولتحديد أى متوسطات تختلف عن بعضها البعض إحصائيًا أو المسئولة عن احداث دلالة اختبار F تستخدم اختبارات المقارنات المتعددة البعدية أو المخططة وعلى ذلك فإنها تهدف إلى تحديد أى المجموعات أو المعالجات تختلف عن بعضها، أى ان قيمة اختبار F تخبرنا فقط عما إذا كانت الدلالة موجودة ولكنها لا تخبرنا عن أى المتوسطات أو المجموعات بالضبط هى التى احدثت الدلالة الإحصائية. ويمكن تسميتها المقارنات المخططة والمقارنات البعدية Post hoc and planned Comparisons.

فافترض أن باحثًا قارن بين ثلاثة عينات أو معالجات وافترض أن متوسطات العينات هى$\overline{X}_3=10$,$\overline{X}_2=5$, $\overline{X}_1=3$وعلى ذلك يوجد ثلاثة فروق بين المتوسطات هى: فرق نقطتين بين$\overline{X}_2$, $\overline{X}_1$ ، فرق خمس نقاط $\overline{X}_3$,$\overline{X}_1$ ، وفرق سبعة نقاط بين $\overline{X}_3$, $\overline{X}_1$ ، فلو استخدمANOVA لتقيم هذه الفروق فإن الدلالة الإحصائية تشير إلى وجود على الأقل أحد فروق المتوسطات كبيرة بدرجة كافية للحصول على دلالة إحصائية، وفى المثال السابق فإن فرق سبع نقاط هو أكبر الفروق الثلاثة ولذلك فإنه

يشير إلى دلالة إحصائية بين المعالجة الأولى والمعالجة الثالثة ($\mu1\neq\mu2$) ولكن ماذا عن فرق الخمس نقاط هل إنه كافى لإحداث دلالة إحصائية؟، وماذا عن الفرق بين $\overline{X}_1$, $\overline{X}_2$ هل هو دال إحصائيًا أيضًا؟ وعلى ذلك فإن الهدف من الاختبارات المقارنات المتعددة الاجابة على هذه التساؤلات. وعلى ذلك فإن الاختبارات البعدية Post hoc tests هى اختبارات فروض اضافيه يتم إجرائها بعد ANOVA أو تحليل التباين لتحديد بالضبط أى فروق بين المتوسطات دال وايهما غير دال إحصائيًا.

وعلى ذلك فإن الاختبارات البعدية في ANOVA يتم إجرائها فى حالة:

1. رفض الفرض الصفرى H_0 حيث متوسطات كل المجتمعات غير متساوية.
2. عندما تتم المقارنة بين ثلاثة مجموعات أو معالجات فأكثر.

وكما نعلم ان نسبة الخطأ لكل مقارنة بين متوسطين Error Rate Per Comparison (PC) وهى الاحتمال للوقوع فى الخطأ من النوع الأول فمثلًا عند إجراء اختبار T للمقارنة بين مجموعتين ونرفض الفرض H_0 حيث T المحسوبة تزيد عن T الحرجة فإن الباحث يعمل فى نسبةPC وهى 0.05. وكثيرًا من إجراءات المقارنة المتعددة تحاول مباشرة ضبط FW وإذا وجدت مقارنات عديدة فإن عائله الخطأ هى:

$$\alpha_E=1-(1-\alpha)^c$$

وإذا كانت المقارنات مرتبطة فإن نسبة الخطأ لكل مقارنه لا تتغير ولكن عائلة الخطأ تتأثر.

وعلى ذلك فإن α_E كالآتى:

$$(FW)\ \alpha_E= C\alpha$$

- C عدد المقارنات بين المجموعات

وعلى ذلك فإن عند إجراء مقارنات عديدة بين المتوسطات لاختبار F يحدث تضخم للخطأ من النوع الأول α وعن طريق قسمهαE على C يعطى القيم الفعلية للخطأ من النوع الأول عند المقارنة بين كل زوج من المتوسطات فان:

$$\alpha=\frac{\alpha E}{C}$$

فاذا تمت المقارنة بين خمسة متوسطات (αE=0.05 , C=10) فإن المقارنة بين كل متوسطين فى ضوء مستوى دلالة إحصائية:

$$\alpha = \frac{0.05}{10} = 0.005$$

وهذا الإجراء يتحكم فى عائلة نسبة الخطأ FW أو αE ولكن هذا الإجراء يطلق عليه شديد التحفظ حيث من الممكن أن تكون نتائج المقارنات بين المتوسطات غير دالة فى حين ان اختبار F دال إحصائيًا.

وقد صنف Hinkle et al. (1994) المقارنات المتعددة إلى:

اولًا: اختبار المقارنات المتعددة أو البعدية اللاحقة لـF الدالة وهذه الفئة تتضمن:

1. المقارنات المزدوجة Pairwise contrasts: لاحظ استخدام مصطلح pairwise تعنى مقارنة متوسطات كل مجموعتين فى نفس الوقت حيث عند مقارنة A , B , C فإن Pairwise comparison بين المجموعات الثلاثة تضمن مقارنة Aفى مقابل B وA فى مقابلC ومقارنة C فى مقابل B ولأربع مجموعات توجد سته مقارنات، عدد أزواج المقارنات يتحدد بالآتى:

$$= \frac{4X3X2}{4} = \frac{24}{4} = 6$$

ومصطلح Non-pairwise or complex فيه يتم مقارنة متوسط ما فى مقابل باقى المتوسطات، ففى المجموعات A, B, C تتم مقارنة المجموعة A مع متوسط B, C معًا.

وهذه الفئة من المقارنات المزدوجة تتضمن:

أ- اختبارات فى حاله تساوى أحجام العينات فى المجموعات وتتضمن اختبارات توكى Tukey ونيومان-كيولز Newman-keuls.

ب- فى حالة عدم تساوى أحجام العينات فى المجموعات وتتضمن اختبار توكي-كرامير Tukey-kramer.

2. المقارنات المعقدة: تتضمن اختبار شيفيه Scheffe

ثانيًا: المقارنات المخططه Planned Contrast: وتشمل المقارنات العمودية Orthogonal Contrasts. فى كثير من المقالات تظهر مصطلح Contrast وهى مرادفة لمصطلح Comparison ومصطلح Post hoc contrast مرادف Follow-up comparison.

وقدم Privitera (2014) تصنيف للمقارنات المتعددة فى ضوء مدى تحفظها فى رفض الفرض الصفرى وهى كالآتى:

اختبار شيفيه	الأكثر تحفظًا
إجراء Bonferroni	(الأقل قوة)
Tukey's HSD Test	↓
Newman-Keuls	الأكثر ليبرالية
Fishr's LDS test	(الأكبر قوة إحصائية)

HDS: honestly significance difference

لاختبار ANOVA الذى يتعامل مع تصميمات بين الأفراد يوجد اختبارين أكثر تحفظًا هما شيفيه وبونفيرونى وهى أفضل اختبارين للتحكم فى تضخم ألفا ويشير Privitera (2014) إلى ان إجراء بونفيرونى هو الإجراء الوحيد الأكثر تحفظًا عندما يكون عدد أزواج المقارنات أكثر من ثلاثة.

ويكون الفرض الصفرى فى المقارنات البعدية كالآتى :

$$H_0: \mu_A = \mu_B$$

$$H_0: \mu_A = \mu_C \text{ او}$$

$$H_0: \mu_B = \mu_C \text{ او}$$

وإذا تضمن البحث أربع مجموعات بالتالى يوجد ستة فروض صفرية.

المقارنات البعدية فى مقابل المقارنات القبلية أو المخططة

المقارنات القبلية يتم اختيارها قبل جمع البيانات بينما البعدية يتم تحديدها بعد جمع البيانات فالمقارنات القبلية يطلق عليها المقارنات المخططة حيث يتم التخطيط لها فى مقدمة أو فى إجراءات البحث قبل جمع البيانات فى ضوء النظرية أو الدراسات السابقة وهذه المقارنات يمكن أن تكون موجهة أو غير موجهة فى ضوء التوقع بينما A posterior or post hoc ليست مخطط له قبل إجراء الدراسة وليست قائم على نظرية أو بحث مسبقًا (دراسات سابقة) وذلك لأن الباحث ليس لديه خلفية مسبقة عن أى المجموعات التى سوف تحدث الفروق بين المجموعات ولذلك يقوم الباحث بعمل كل المقارنات الممكنة بين مستويات المتغير المستقل وتحديد أى مجموعات احدثت الفروق.

والمقارنات المخطط لها مسبقًا لا تصحح تضخم الخطأ من النوع الأول نتيجة المقارنات المتعددة التى تحدث مع المقارنات البعدية وهذا التصحيح للمقارنات المتعددة غير المخطط لها(البعدية) يجعل من الصعوبة رفض H_0، وعليه فإن رفض الفرض الصفرى فى المقارنات المتعددة البعدية أصعب من إجراء المقارنات المخطط لها مسبقًا وعلى ذلك فإن المقارنات القبلية لها قوة إحصائية أعلى من المقارنات البعدية العديدة بين كل الازواج الممكنة بين المجموعات.

وبتطبيق ذلك على تجربة قام بها باحث للمقارنة بين اربعة انواع مختلفة من العلاج للاكتئاب A, B, C, D، فالباحث عنده توقع مسبقًا عن مقارنة بين طريقتين A, C فهذه مقارنة مخطط لها. ولكن إذا اراد الباحث إجراء مقارنات بين كل زوج من انواع العلاج ستة مقارنات ففى هذه الحالة يعتمد الباحث على مقارنات بعدية ولكن الاجابة التقليدية أو الإجراء التقليدى هو إجراء مقارنات بين كل الازواج الممكنة.

وحقيقه يوجد جدل حول ما إذا كان الباحث يقوم بإجراء المقارنات المتعددة فى حاله ما إذا كان قيمة F غير دالة إحصائيًا فكانت الاجابة القديمة بدون وجود دلالة إحصائية فلا داعى أو من غير المناسب إجراء المقارنات المتعددة ولكن Howell (2013) يرى أن هذه النصيحة خاطئة حيث إن المنطق وراء المقارنات المتعددة لا يتطلب دلالة إحصائية لـ F لان اختبارات الفروض لـ F وللمقارنات المتعددة مختلفة تمامًا مع مستويات مختلفة من القوة الإحصائية حيث إن قيمه F توزع الفروق بين المجموعات عبر عدد من درجات الحرية للمجموعات وهذا له تأثير على قيمة F فى الموقف الذى

يوجد فيه متوسطات مجموعات عديدة متساوية ببعضها البعض ، واختبارات الفروض ودلالتها الإحصائية للمقارنات المتعددة صممت دلالتها الإحصائية بدون النظر أو الاعتبار لقيمةF، لاحظ أن التحليل المبدئى لـ ANOVA يطلق على قيمة F بـ Omnibus F-test، بل ان (1987) Wilcox أشار إلى أنه لا يوجد منطقية لتطبيق اختبار F على الإطلاق وبدلًا من ذلك فإنه ذهب مباشرة إلى إجراء المقارنات المتعددة بدون حساب F ويشير (2013) Howell إلى ان هذا هو رأى الكثير من الباحثين وعلى ذلك فلا بد من مقارنة المجموعات وبغض النظر عن دلالة F .

فى عرض اختبارات المقارنات المتعددة سوف اعتمد على مثالين هما المثال الذى سبق عرضه فى تحليل التباين مأخوذ من (2013) Pagano ومثال آخر يتم عرضه فى سياق الحديث مأخوذ من (2013) Howell

قضية بحثية(2013) Howell (البيانات فقط):

وفيما يلى خمسة طرق للتنشئة الاسرية وتم اختيار عينة عشوائية فى كل مجموعة وبعد تطبيق طرق التدريس كانت درجات التحصيل كالآتى:

	1	2	3	4	5
	3	**2**	**14**	**29**	24
	5	**12**	**6**	**20**	26
	1	**13**	**12**	**36**	40
	8	**6**	**4**	**21**	32
	1	**10**	**19**	**25**	20
	1	**7**	**9**	**18**	33
	4	**11**	**9**	**20**	27
	9	**19**	**21**	**17**	30
$\bar{x}$	**4.00**	**10.0**	**11**	**24.0**	29.0
S	**3.16**	**5.13**	**6.72**	**6.37**	6.16

وتم حساب تحليل التباين لدراسة الفروق بين أساليب التنشئة الخمسة وكانت النتائج كالآتى:

مصدر التباين	SS	df	MS	F
بين المجموعات	**349760**	**4**	**87440**	**27.33**
داخل المجموعات(الخطأ)	**1120.00**	**35**	**32.00**	

$P<0.05$

وعلى ذلك توجد فروق دالة إحصائيًا بين المجموعات الخمسة.

حجم التأثير: يقدر حجم التأثير من خلال مؤشرى:

$$\eta^2 = \frac{SS_B}{SS_{total}} = \frac{3497.60}{4617.60} = 0.76$$

وهذا يعنى ان طريقة التنشئة فسرت76%من تباين التحصيل.

والتقدير غير المتحيز لحجم التأثير هو ω^2 كالآتى:

$$\omega^2 = \frac{SS_{between} - (k-1)MS_{within}}{SS_{total} + MS_{within}}$$

$$\omega^2 = \frac{3497.60 - 4(32)}{4617.60 + 32} = \frac{3369.6}{4649} = 0.72$$

المقارنات القبلية

والمقارنات القبلية تسمى احيانًاContrasts ويتم التخطيط لها قبل جمع البيانات وتوجد العديد من هذه النوعية من المقارنات أهمها:

أولًا : اختبارات T المتعددة: حيث يتم إجراء اختبار **T** بين كل زوج من المجموعات وفى إجراء اختبار **T** لا بد من توافر مسلمة التجانس للتباينات حيث يتم دمج التباينات المفردة أو المشتركة Pooled التى تدخل فى حسابT مع MSerror التى تدخل فى حساب تحليل التباين وتقيم اختبار **T** فى ضوء درجات الحرية الخطأ أو داخل المجموعات dfw (N-K) ولكن Howell (2013) يرى الاعتماد على درجات الحرية المحسوبة من اختبار **T**(n1+n2-2) وعندما يحدث عدم تجانس للتباين ولكن أحجام العينات متساوية فلا تستخدم MSw ولكن تستخدم التباينات المفردة وتقيم **T** فى ضوء df = (N-2) وعندما يحدث عدم تجانس للتباينات واحجام عينات غير متساوية فإنه يستخدم التباينات المفردة وتصحيح درجات الحرية فى ضوء مدخل satterthewaite -Welch . وكما نعلم أن إجراء العديد من

المقارنات المتعددة يضخم قيمة الخطأ من النوع الأول، فمثلًا مع إجراء مقارنتين أو اختبار **T** مرتين فإن قيمتها تتضخم إلى 0.02 عندما تكون α=0.01.

وفى المثال السابق فإن الباحث يتوقع ان تكون المقارنات بين المجموعة الأولى والثالثة وكذلك المقارنة بين المجموعة الثانية والخامسة هى السبب الرئيسى فى إحداث الفروق وعليه فإن صيغة**T** تكون كالآتى:

$$T = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_w}{n} + \frac{MS_w}{n}}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{2MS_w}{n}}}$$

وفى حالة المقارنة بين المجموعة الأولى والمجموعة الثالثة فان:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2MS_w}{n}}} = \frac{4 - 11}{\sqrt{\frac{2(32)}{8}}} = \frac{-7}{\sqrt{8}} = -2.47$$

والمقارنة بين المجموعة الثانية والخامسة:

$$T = \frac{29 - 10}{\sqrt{\frac{2(32)}{8}}} = \frac{19}{\sqrt{8}} = 6.772$$

وبالبحث فى جداول **T** بـ α=0.05 و ، dfw=N-K=40-5=35 (درجات الحرية داخل المجموعات)، يتضح أن **T** الحرجة =2.03 وعليه يرفض H_0 فى المقارنتين ويمكن القول إن الفرق بين المجموعتين الأولى والثانية، والفرق بين المجموعتين الثانية والخامسة هى التى احدثت هذه الدلالة الإحصائية لاختبار F وان كانت الفروق بين المجموعة الثانية و الخامسة هى بالذات كان لها دورًا كبيرًا فى احداث هذه الفروق.

وبعد ان تناولنا أحد طرق المقارنات القبلية نتطرق إلى طرق اخرى مثل المقارنات العموديةOrthogonal-contrasts واختبار Dunn-sidak test و Bonferroni و Dunn's test من خلال مخرج برنامج SPSS نظر لتعقيدات حسابية.

تصحيح بونيفرونى أو اختبار دان Bonferroni (Dunn's test)

عند المقارنة بين متوسطات عديدة فاحد طرق التحكم فى عائلة نسبة الخطأ هو استخدام مستوى الدلالة الإحصائية α أكثر تحفظًا، فاستخدام 0.01 أفضل من 0.05 وأحد الطرق الأساسية لضبط تضخم ألفا من خلال ضبط نسبة الخطأ المقارن بين كل مجموعتين وتم اقتراحه بواسطة Dunn ويعرف تحت مسمى Bonferroni T، فعند عمل ثلاثة مقارنات وكل مقارنة باحتمال α=0.5 الخطأ من النوع الأول، فإن احتمال ان لا يزيد الخطأ من النوع الأول عن 0.15 = 0.05 x 3.

فاذا كانت Cهى عدد المقارنات و α احتمال الخطأ من النوع الأول لكل مقارنة بين مجموعتين فان:

$$Fw \leq C\alpha$$

واذا تم وضع $\alpha = \frac{\alpha}{c}$ لكل مقارنة فان:

$$Fw \leq C\alpha = C\ \frac{\alpha}{C}$$

و إذا كانت α للمقارنة = 0.05 فعند إجراء ثلاثة مقارنات فإن α لكل مقارنة هى:

$$\alpha = \frac{0.05}{3} = 0.0167$$

والقضية هنا ان جداول T لا تعطى قيم جدولية عند مستوى ألفا 0.017 ويمكن تنفيذ ذلك من خلال برنامج الكمبيوتر حيث يعطى قيمة P. ويمكن تقدير اختبار بونيفرونى للمقارنات الزوجية من المعادلة الآتية:

$$T' = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_W}{n} + \frac{MS_W}{n}}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{2MS_W}{n}}}$$

حيث 'T عند $'(\frac{\alpha}{C}, df)$

وهذا الاختبار يصلح فى حالة عدد محدود من المقارنات المخطط لها ويعطى برنامج SPSS وSAS اختبار بونيفرونى مع اختبارات المقارنات البعدية ولذلك فإن استخدامه غير مناسب فى هذه الحالة لأنه يقع ضمن المقارنات القبلية المخطط لها

مسبقًا. وبتطبيق اختبار بونيفرونى على المثال السابق للمقارنة بين المجموعة الخامسة والثالثة كالآتى:

$$= \frac{29 - 10}{\sqrt{\frac{2(32)}{8}}} = \frac{197}{\sqrt{8}} = 6.72$$

وبين المجموعة الثالثة والأولى:

$$= \frac{11 - 4}{\sqrt{\frac{2(32)}{8}}} = \frac{7}{\sqrt{8}} = 2.47$$

وبين المجموعة الثانية والثالثة:

$$= \frac{10 - 11}{\sqrt{\frac{2(32)}{8}}} = \frac{-1}{\sqrt{8}} = -0.35$$

وفى حالة إجراء أربع مقارنات فإن:

$$Fw = 3 \times 0.05 = 0.15$$

ويتم تقدير **T'** عند:

$$\alpha = \frac{0.05}{4} = 0.0125$$

$$\alpha = \frac{0.05}{3} = 0.0167$$

ولاختبار ذو ذيلين فإنه يتم البحث عن **T** الجدولية بـ $0.0167 = \alpha$ و $dfw = 40$ $35 = 5 - 40$ وعليه فان:

T الحرجة $\left(\frac{\alpha}{2}(0.00625), 35\right) = \pm 2.052$

ومن ثم يرفض H_0 فى حالة المقارنة بين المجموعة الخامسة والثالثة بينما يتم قبول الفرض H_0 فى حاله المقارنة بين المجموعة الثانية والثالثة والمقارنة بين المجموعة

الثالثة والأولى. وقدر اختبار بونيفرونى عند مستوى دلالة إحصائية $\frac{\alpha}{c}$ ولكن إذا تم تقدير مستوى الدلالة الإحصائية عند:

$$\alpha = 1 - (1 - \alpha)^{\frac{1}{c}}$$

فان هذا يسمى اختبار Dunn-sidak test وهى نفس حساب اختبار بوينفرونى وهذا متاح فى برنامج SPSS وبمقارنة القوة بين اختبار بونيفرونى واختبار دان- سيداك فإن القوة الإحصائية تميل قليلًا لصالح دان- سيداك.

إجراءات بونيفرونى متعددة المراحل Multistage Bonferonni procedures

كما سبق واوضحنا ان اختبار بونيفيرونى يستخدم عند اختبارات فروض متعدد لضبط نسبة الخطأ (Fw) وهذه الإجراءات لها ميزة فى وضع إطار محدد لخطأ Fw عندα عند إجراء الفروض الصفرية المحتملة بين ازواج المتوسطات وذلك مثل اختبار توكى ولكنه أقل تحفظًا من اختبار توكى عندما يكون الاهتمام قائم على مجموعه فرعية من المقارنات.

واختبار بونيفرونى قائم على اختبار اربعة مقارنات للمتوسطات بقيمة:

$$\frac{\alpha}{4} = 0.0125$$

والاختبار المتعدد المراحل يتبع نفس المبدأ ولكن الاختلاف هو ان يبدأ بأعلى فرق (أو معامل ارتباط)، ثم ينتقل إلى الفرق الأقل الذى يليه وتستمر حتى أول فرق غير دال إحصائيًا.

كتابة نتائج اختبارات المقارنات القبلية فى البحث وفقًا لـ APA

اتضح وجود تأثير دال إحصائيًا بين المجموعات الخمسة، حيث اتضح وجود فروق دالة إحصائيًا بين المجموعة الخامسة ($\bar{x}$ = 29) مقارنة بالمجموعة الثالثة ($\bar{x}$= 10) حيث:

$$T(35) = 6.72 \text{ , } tc.v(\frac{\alpha}{3} = .0167) = \pm 2.52$$

ثانيًا: المقارنات البعدية: عندما لا تكون المقارنات مخطط لها مسبقًا فإننا نستخدم اختبار المقارنات البعدية حيث يقوم الباحث بإجراء مقارنات بين كل الازواج الممكنة بدون وجود خلفية أو أساس نظرى وهذه المقارنات غير

مخطط لها قبل إجراء التجربة أو البحث وكما نعلم أنه يوصى باستخدام المقارنات القبلية و اختبار بونيفرونى فى حالة وجود عدد محدود من المقارنات المحدد مسبقًا. ويرى Howell (2013) أن الاستخدام الشائع للمقارنات البعدية يفقدها قوتها عن طريق طرح أسئلة فيما يخص المقارنات بين مجموعات الباحث ليس بحاجة اليها وليس لديه تفسير نظرى أو منطقى لها ولكن يرى أن هذه المقارنات لها ميزة فى التجارب التى تتضمن فروض كثيرة وكذلك فروض يمكن أن يفترضها أو يصيغها الباحث بعد جمع البيانات. وفى هذه الحالة توجد العديد من أساليب المقارنات المتعددة البعدية المتاحة أهمها:

1. إجراء الفروق الأقل دلالة لفيشر Fisher's Least Significance Difference (LSD): يعتبر أحد أقدم الطرق للمقارنات البعدية ويعرف بـ Fisher's protected T ، ويشكو الباحثين من هذا الإجراء عندما يوجد الكثير من المتوسطات ولكن عندما يوجد متوسطات محدودة للمقارنة (فقط ثلاثة) فإن هذا الإجراء مفيد وفى غاية من الأهمية ويتم اجراؤه من خلال كل زوج من المقارنات بين المتوسطات باستخدام اختبار T والفرق الوحيد بين LSD و اختباراتT المتعددة القبلية الذى تناولناه فى المقارنات القبلية هو ان LSD يتطلب دلالة إحصائية لقيمة F فى تحليل التباين وفى إجراء LSD فإن عائلة الخطأ تساوى α. ويعتبر من أكثر طرق المقارنات البعدية ليبرالية ويمتاز بالقوة بالكشف عن التأثير أو فروق المتوسطات بين مجموعتين ولا يتطلب تساوى حجم العينات فى كل المجموعات ويتم حسابه كالآتى:

$$Fisher\ (LSD) = T_{c.v}\sqrt{MS_w\left[\frac{1}{n_1}+\frac{1}{n_1}\right]}$$

وقيمة Tc.v الحرجة من خلال درجات الحرية dfw مثلًا:

$$LSD_{(3,5)} = 2.052\sqrt{32\left[\frac{1}{8}+\frac{1}{8}\right]}$$

وبمقارنة T الحرجة يتم رفض H_0 ، ولكن البعض يوصى بعدم استخدام هذا الاختبار لأنه لا يقى نفسه من تضخم الخطأ من النوع الأول ولكن إذا كانت التجربة تتضمن ثلاثة مجموعات أو متوسطات فإن هذا الإجراء ممتاز لأن قيمة Fw تبقى ثابته عند α المفترضة، وهو أكثر اختبارات المقارنات البعدية ليبرالية وفى نهاية المتصل يعتبر شيفيه أكثر تحفظًا .

إحصاء Studentized Range Statistic (Q)

نعلم ان التوزيع الاحتمالى للمقارنات البعدية قائم على هذا الإحصاء ويتحدد هذا التوزيع بالاتى:

$$Q_r = \frac{\bar{X}_L - \bar{X}_u}{\sqrt{\frac{MS_w}{n}}}$$

$\bar{X}_L - \bar{X}_u$ فرق متوسطات المعالجات الأكبر والأصغر و k عدد المعالجات.

وإحصاءQ مشابهًا تمامًا لصيغة T فى المقارنات القبلية.

$$T=\frac{\bar{X}_i-\bar{X}_j}{\sqrt{\frac{2MS_W}{n}}}$$

الفرق الوحيد هو ان صيغةQ تتضمن$\sqrt{2}$ فى المقام أى ان:

$$Q = \frac{T}{\sqrt{2}}$$

وعلى ذلك ففى توزيعQ يختلف عدد المجموعات بينما عدد الأفراد فى المجموعات لا يتغير ولإحصاءQ توزيع عينى وقيمته الحرجة تعتمد على عدد المجموعات (K) ودرجات الحرية المرتبطة بتقدير متوسطات مجموع المربعات داخل المجموعات(الخطأ) MSw. " وعليه فإن اختبارات المقارنات البعدية مثل توكى وتوكي–كرامير ونيومان–كيولز تستخدم توزيع Q كتوزيعات عينية لضبط نسب تضخم الخطأ من النوع الأول.

ولحَساب اختبار Q لمتوسطات المجموعات الخمسة السابقة : $\overline{X}_1, \overline{X}_2$, $\overline{X}_3, \overline{X}_4, \overline{X}_5$ على التوالى: 29 ,24, 11, 10, 4

و كما نعلم ان: $MS_w = 32$, $df_w=40-5=35$, $n=8$، وان أعلى متوسط $\overline{X}_5=29$ وان أدنى متوسط $\overline{X}_1=4$ ، k عدد المجموعات =5 وعليه:

$$Q_5 = \frac{\bar{X}_1 - \bar{X}_5}{\sqrt{\frac{MS_w}{n}}} = \frac{29-4}{\sqrt{\frac{32}{8}}} = \frac{25}{\sqrt{4}} = 12.5$$

لاحظ أن k لم تتضمن فى الحسابات وبالبحث فى جدولQبـ $df = 35$,$k = 5$ يتضح ان:

$Q_{(0.05)}$ (5,35) = 4.07الحرجة

وحيث إن : **Q المحسوبة(4.07) < Q الحرجة(12.5)**، وعليه نرفض H_0 أى يوجد فرق دالة إحصائيًا بين المتوسط الأعلى والمتوسط الادنى. وهذا الاختبار هو المستخدم فى برنامج الكمبيوتر لحساب المقارنات المتعددة البعدية.

2. اختبار توكى Tukey's test: معظم اختبارات المقارنات البعدية قائمة على عمل John Tukey ويسمى الاختبار باسمه. ويمكن تسميته Tukey's HSD (Honestly Significance Differences) أو يطلق عليه اختبار فرق الدلالة الكلية Wholly Significance Difference (WSD) ويستخدم إحصاء Q للمقارنات ما عدا انQ_{HSD} لتوكى هو القيمة العظمى لـQ_r. بكلمات أخرى لو وجد خمس متوسطات فكل الفروق يتم اختبارها ويتم تثبيت عائلة الخطأ عند مستوى α لكل الفروض الصفرية الممكنة، واختبار توكى مفضل كاختبار بين أزواج المتوسطات الممكنة لأنه يضبط α ويحد تضخمها ولو طبقنا اختبار توكى على المتوسطات الخمسة السابقة وكما نعلم ان :

$$Q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_w}{n}}}$$

و $df_w = 40 - 5 = 35$ و $k = 5$ فإن القيمة الحرجة لـQ $= 4.07$ $= Q_{c.v}$ وعليه:

$$\bar{X}_i - \bar{X}_j = Q\sqrt{\frac{MS_w}{n}}$$

$$= 4.07\sqrt{\frac{32}{8}} = 8.14$$

فـــان كـــان الفـــروق بـــين أى متوســـطين أكبـــر مـــن 8.14 فـــإن قيمـــة Q دالـــة إحصائيًا ومن ثم:

$$\bar{X}_1 - \bar{X}_4 = 4 - 24 = -20$$

$$\bar{X}_1 - \bar{X}_5 = 4 - 29 = -25$$

$$\bar{X}_2 - \bar{X}_4 = 10 - 24 = -14$$

$$\bar{X}_2 - \bar{X}_5 = 10 - 29 = -19$$

$$\bar{X}_3 - \bar{X}_4 = 11 - 24 = -13$$

$$\bar{X}_3 - \bar{X}_5 = 11 - 29 = -18$$

وعليه فإن كل الفروق بين ازواج المتوسطات دالة إحصائيًا لان:

$$|\bar{X}_i - \bar{X}_j|(8.14) > \text{الحرجة} Q$$

وباقى فروق المتوسطات غير دالة إحصائيًا، ويمكن حساب اختبار Q بطريقة أخرى:

$$Q_r = \frac{\bar{X}_1 - \bar{X}_5}{\sqrt{\frac{MS_w}{n}}} = \frac{4 - 29}{\sqrt{\frac{32}{8}}} = 12.5$$

وعليه فإن Q_r المحسوبة $12.5 >$ Q الحرجة 8.14

وهكذا لبقية ازواج المتوسطات وعليه فان:

$$\bar{X}_1 = \bar{X}_2 = \bar{X}_3 \neq \bar{X}_4 = \bar{X}_5$$

والاشارة تساوى تشير إلى عدم رفض H_0. ويمكن ترجمتها الى:

$$(\overline{X}_1 = \overline{X}_2 \ , \overline{X}_1 = \overline{X}_3 , \overline{X}_2 = \overline{X}_3 , \overline{X}_4 = \overline{X}_5)$$

$$\overline{X}_1 \neq \overline{X}_4 \ , \overline{X}_1 \neq \overline{X}_5 , \overline{X}_2 \neq \overline{X}_4 , \overline{X}_2 \neq \overline{X}_5 , \overline{X}_3 \neq \overline{X}_4 , \overline{X}_3 \neq \overline{X}_5$$ بينما:

وعليه نقبل الفرض البديل.

صيغة Tukey في حالة $n_1 \neq n_2$ و $\sigma_1^2 \neq \sigma_2^2$

وفى حالة عدم تساوى العينات تصبح الصيغة كالآتى:

$$\sqrt{\frac{MS_W}{n}} = \sqrt{\frac{\frac{MS_W}{n_i} + \frac{MS_W}{n_j}}{2}}$$

ويعتبر تساوى أحجام العينات هى الطريقة المستخدمة default فى برنامج SPSS وتصبح درجات الحريه كالآتى:

$$df = \frac{\left(\frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{S_j^2}{n_j}\right)^2}{n_j - 1}}$$

وهذا التصحيح سبق الإشارة إليه فى الحديث عن اختبار T فى حاله عدم تجانس التباينات

كتابة نتائج اختبار توكى فى تقرير البحوث وفقا APA

تم إجراء اختبار توكى البعدى لمقارنة ازواج المتوسطات واتضح ان الفروق بين قيم المتوسطات دالة إحصائيًا.

3. طريقة نيومان-كيولز Newman-keuls method: هذه الطريقة مشابهة لطريقة توكى ماعدا انها تعتمد على الفروق بين المتوسطات المرتبة. ودار حولها جدلًا كبيرًا الا انها مازالت تستخدم فى برنامجSPSS حيث تتغير القيمة الحرجة لـ Q فى ضوء المدى لمجموعة المتوسطات فعندما يتم ترتيب المتوسطات للمجموعات من

الأدنى إلى الأعلى فاذا حدثت مقارنة بين المجموعة الأولى والثالثة فإن عدد المجموعات K=3 ، وليس K=5 كما هو الحال فى اختبار توكى ويبدو ان ميزه اختبار نيومان-كيولز أنه اختبار قوى إحصائيًا ولكن من عيوبه أنه يفتقد بعض الضبط لـ α.

فعند المقارنة بين $\overline{X}_5$, $\overline{X}_1$ فإن k = 5 ، وعند المقارنة بين $\overline{X}_3$, $\overline{X}_1$ فإن k = 3، وعند المقارنة بين $\overline{X}_3$, $\overline{X}_2$ فإن k = 2.

وقبل حساب اختبار نيومان- كيولز لا بد من ترتيب المتوسطات من الأدنى إلى الأعلى ثم حساب قيمة α حسب قيمة k وعليه فإن قيمة Q الحرجة تختلف من مقارنة لأخرى عكس اختبار توكى الذى يعتبر فيه قيمة Q الحرجة واحدة لكل المقارنات، وحسابات توكى هى نفسها حسابات نيومان- كيولز ماعدا الاختلاف فى قيمة Q الحرجة وكلًا منهما يستخدم ومناسب للمقارنات المزدوجةPairwise وإذا كانت نسبة F دالة إحصائيًا فى تحليل التباين وعندما تكون أحجام العينات متساوية.

وعليه فإنه توجد اختلافات فى نتائج توكى عن نيومان-كيولز فيما يتعلق بالدلالة الإحصائية للفرق بين متوسطين حيث توجد فروق دالة إحصائيًا باستخدام نيومان-كيولز مقارنة بطريقة توكى ولهذا السبب فإن طريقة نيومان-كيولز أكثر قوة إحصائيًا من توكى ولكن هذه الميزة تقابل بمحدد لنيومان-كيولز وهو ان نسبة الخطأ من النوع الأول يحدث لها تضخم عن مستوى ألفا المفترض مسبقًا(α) ولكن هذا التضخم لـ α يكون أقل من تضخمها فى حالة استخدام المقارنات المتعددة باستخدام T ولكن باستخدام طريقة توكى تظل قيمة α ثابتة عند المستوى المفترض مسبقًا.

4. طريقة Ryan: هذا الإجراء للمقارنات المتعددة البعدية يعرف بـ REGWQ test وهو تعديل لطريقة Tukey. ويجعل قيمة FEW عند مستوى α المفترض ويفضل Howell (2013) هذا الاختبار على توكى ولكنه نادر الاستخدام.

5. اختبار بنجامينى – هوتشبيرج Benjamini – Hochberg Test: كل اختبارات المقارنات البعدية السابقة تركز على ضبط نسبة عائلة الأخطاء(FM) ، ولكن Benjamini & Hochberg (2000) طور اختبارات ليست قائمة على FM

واستخدم ما يسمى نسبة اكتشاف الخطأ False discovery rate(FDR) بدلًا من FM وتسألوا عن ماهى النسبة المئوية للنتائج الدالة إحصائيًا ويمكن بعد ذلك ايجاد النتائج الدالة إحصائيًا بالخطأ. فافترض توجد تسع مقارنات ووجدنا اربع تأثيرات دالة إحصائيًا واحد هذه التأثيرات الدالة إحصائيًا هى حقيقة خطأ من النوع الأول (رفض H_0 وهو فى المجتمع صحيح) ويعرف FDR بالآتى:

$$FDR= \frac{\text{عدد المقارنات الدالة وخطأ}}{\text{عدد المقانات الدالة وهى حقيقة}} = \frac{1}{4} = 0.25$$

للمزيد عن هذا الاختبار (انظر: Howell, 2013) وهذا الاختبار متوفر فى برنامج SPSS.

المقارنات المتعددة البعدية للعينات غير المتساوية

كما سبق وأشرنا أنه يمكن استخدام اختبار توكى للعينات غير المتساوية فى المجموعات بعد تصحيح df و MSw فى العرض الآتى تقدم طريقة للعينات غير المتساوية وهى:

1. طريقة توكى-كرامير (TK) Tukey-Kramer: كما سبق واوضحنا ان طريقتى توكى ونيومان-كيولز تستخدم بفاعلية إذا كان أحجام المجموعات متساوية ولكن إذا كانت أحجام العينات غير متساوية فإنه يوجد تعديل لطريقة توكى و تسمى توكى- كرامير وهى تتضمن تعديل فى مقام إحصاء Q كالآتى:

$$Q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_W\left(\frac{\frac{1}{ni}+\frac{1}{nj}}{2}\right)}}$$

مثال: افترض أن باحث درس تقدير الذات بين أربع مجموعات أو عينات وكانت نتائج ANOVA كالآتى:

المجموعة	الأولى	الثانية	الثالثة	الرابعة
$\bar{X}$	21.86	23	17.29	16.80
N	7	9	7	10

وملخص جدول ANOVA كالآتى:

مصدر التباين	SS	df	MS	F.Ratio	Fc.v
بين المجموعات	**255.45**	**3**	**85.15**		
داخل	**235.89**	**29**	**8.13**	**10.47**	**2.93**
الكل	**491.34**	**32**			

وعليه رفض الباحثH_0: $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

ولإجراء المقارنات المتعددة كالآتى:

بين المجموعة الأولى والثانية:

$$Q_{1-2} = \frac{21.86 - 23}{\sqrt{8.13(\frac{\frac{1}{7}+\frac{1}{9}}{2})}} = -1.12^*$$

بين المجموعة الأولى والثالثة:

$$Q_{1-3} = \frac{21.86 - 17.29}{\sqrt{8.13(\frac{\frac{1}{7}+\frac{1}{7}}{2})}} = 4.24^*$$

بين المجموعة الأولى والرابعة:

$$Q_{1-4} = \frac{21.86 - 16.80}{\sqrt{8.13(\frac{\frac{1}{7}+\frac{1}{10}}{2})}} =$$

بين المجموعة الثانية والثالثة:

$$Q_{2-3} = \frac{23.0 - 17.29}{\sqrt{8.13(\frac{\frac{1}{9}+\frac{1}{7}}{2})}} = 5.62^*$$

بين المجموعة الثانية والرابعة:

$$Q_{2-4} = \frac{23 - 16.18}{\sqrt{8.13(\frac{\frac{1}{9}+\frac{1}{10}}{2})}} = 6.69^*$$

المجموعة الثالثة والمجموعة الرابعة:

$$Q_{1-2} = \frac{17.29 - 16.80}{\sqrt{8.13(\frac{\frac{1}{7}+\frac{1}{10}}{2})}} = 0.49$$

ولحساب القيمة الحرجة لـ Q بدرجات حرية: $k= 4, df = 33 - 4 = 29$،

إذًا Q الحرجة= 3.86

وعليه فإن الفروق بين المجموعتين الأولى والثالثة، والمجموعة الأولى والرابعة، والمجموعة الثانية والثالثة، وبين الثانية والرابعة هى فروق دالة إحصائيًا، وعليه فإن المجموعة الأولى تختلف عن الثالثة والرابعة وكذلك متوسط المجموعة الثانية يختلف عن متوسطات المجموعة الثالثة والمجموعة الرابعة.

2. اختبار Dunnett's test: صمم هذا الاختبار للمقارنة بين أحد المعالجات (عادة الضابطة) مع أى من المتوسطات أو المعالجات الأخرى وهو أكثر قوة فى تحقيق ذلك الغرض، أى أنه يهتم بالمقارنة بين أى متوسط لمجموعة معينة فى مقابل أى من متوسطات المجموعات الباقية ويحدث إذا اهتم الباحث فى التصميم الذى يتضمن معالجات عديدة ومجموعة ضابطة واحدة، فيقارن متوسط كل معالجة بالمجموعة الضابطة. ومع هذا الاختبار يوجد القليل من الفروض الصفرية حيث مع ثلاث مجموعات (2 تجريبية وواحد ضابطة):

$$H_0: \mu_1 = \mu_2$$
$$H_0: \mu_1 = \mu_3$$

ولا يهتم بدراسة الفروق بين μ_2 و μ_3

3.**اختبار Tamhan's post hoc test** : ويشار اليه بـ Tamhan's T2 test حيث يقوم بإجراء مقارنات بين كل زوج من المتوسطات فى حالة عدم تساوى التباينات بين المجموعات .

الاختبارات البعدية للمقارنات المعقدة: طريقة شيفيه Post hoc test for complex comparison: Scheffe method

فـى كثيـر مـن المواقـف يهـتم الباحـث باختبـار فـروض أكثـر تعقيـدًا مقارنـة بالمقارنـات البسـيطة بـين كـل زوج مـن المتوسـطات ويعتبـر اختبـار شـيفيه مـن أقـدم الاختبـارات وقـائم علـى إجـراء Tukey ويسـمح بالمقارنـة بـين أى متوسـط وبـاقى متوسـطات كـل الازواج معـا عنـد α. ولـيس مقارنـات مزدوجـة (كـل متوسـطين معًـا). ولا ينصـح شـيفيه باسـتخدام لكـل زوج علـى حـدة وعلـى الـرغم عـن ذلـك فـإن برنـامج SPSS لا يأخـذ بهـذه النصـيحة. وفـى اختبـار شـيفيه فـإن كـل فـرض يصـاغ فـى ضـوء اتحـاد خطـى للمعـاملات والمتوسـطات وهـذا الاتحاد يشار اليه بـContrast. وشكل الفرض الصفرى يكون كالآتى:

$$H_0: \sum_{k=1}^{k} C_K \mu_X = 0$$

حيث $\sum_{k=1}^{k} C_K = 0$

ويضاف المتوسط مضروبًا فى المعاملات وأن مجموع المعاملات = صفر. ففى المثال السابق فاذا كان الباحث مهتم بتحديد ما إذا كان يوجد فروق بين المجموعة الثانية عن بقية المجموعات فى تقدير الذات فان:

$$H_0 : \mu_2 = \frac{\mu_1 + \mu_3 + \mu_4}{3}$$

حيث 3 فى المقام هى عدد المتوسطات فى البسط. ويمكن أن يصاغ الفرض الصفرى فى ضوء المقارنات:

$$3\ \mu_2 = \mu_1 + \mu_3 + \mu_4$$

بالقسمة على 3:

$$H_0: \mu_2 = \frac{\mu_1}{3} + \frac{\mu_3}{3} + \frac{\mu_4}{3}$$

$$H_0: \frac{1}{3}\mu_1 + (-1)\ \mu_2 + \frac{1}{3}\mu_3 + \frac{1}{3}\mu_4 = 0$$

لاحظ أن مجموع المعاملات يساوى صفر:

$$(\frac{1}{3} - 1 + \frac{1}{3} + \frac{1}{3}) = 0$$

بالضرب x 3:

$$H_0: (1)\ \mu_1 + (-3)\ \mu_2 + (1)\ \mu_3 + 1\ (\mu_4) = 0$$

وعندما تكون أحجام العينات غير متساوية من الضرورى تصحيح هذه المعاملات وهى C_4 , C_3 , C_1 ويتم تقديرها كالآتى:

$$C_1 = \frac{7}{7+7+10} = 0.24$$

$$C_2 = -1\left(\frac{9}{9}\right) = -1$$

$$C_3 = \frac{7}{7+7+10} = 0.29$$

$$C_4 = \frac{10}{7+7+10} = 0.42$$

وتكون صيغه طريقه شيفيه:

$$F = \frac{\sum(C_K \bar{X}_K)^2}{MS_w \left[\frac{{C_K}^2}{nK}\right]}$$

والتوزيع العينى هو اختبار F وتتحدد القيمة الحرجة :

القيمة الحرجة Critical value = Fc.v × (K – 1)

وعليه فإن القيمة الحرجة ل F عند $\alpha = 0.05$ ، $dF_1 = 4 - 1 = 3$ (البسط)، $dF_2 =$ (المقام) $33 - 4 = 24$ هى : $F_{c.v} = 2.93$

والقيمة الحرجة المستخدمة فى شيفية:

$$Fc.v \times (K - 1) = 2.93 * 3 = 8.79$$

وعليه فإن قيمة شيفيه:

$$F = \frac{(C_1\bar{X}_1 + C_2\bar{X}_2 + C_3\bar{X}_3 + C_1\bar{X}_4)^2}{MS_w \left(\frac{(C_1)^2}{n_1} + \frac{(C_2)^2}{n_2} + \frac{(C_3)^2}{n_3} + \frac{(C_4)^2}{n_4}\right)}$$

$$F = \frac{(0.29\,(2.86) + (-1)(23.0) + 0.29\,(17.29) + 0.42\,(16.80))^2}{8.13\left(\frac{(0.29)^2}{7} + \frac{(-1)^2}{9} + \frac{(0.29)^2}{7} + \frac{(0.42)^2}{10}\right)}$$

$$= \frac{21.07}{1.24} = 16.96$$

بالتالى توجد فروق بين متوسط مجتمع المجموعة الثانية وبقيه متوسطات مجتمعات المجموعات الأخرى وبالمثل يمكن إجراء المقارنات القبلية باستخدام المقارنات المتعامدة المخططة Planned or orthogonal Contrasts وتحليل الاتجاه Trend analysis للمزيد انظر (1994) Hinkel et al. .

أى من اختبارات المقارنات المتعددة مفضله؟

اختيار إجراء المقارنات المتعددة الانسب لموقف بحثى معين عملية ليست سهلة حيث يوجد العديد من الاختبارات ولكنها تختلف عن بعضها البعض.

- إذا استخدمت مقارنات قبلية ومخطط لها قبل جمع البيانات واردت إجراء مقارنة واحده فيقترح Howell (2013) إجراء اختبار T مع تصحيح MSw , df فى حاله عدم وجود تجانس للتباين وإذا كانت لديك مقارنة معقدة بمعنى $\bar{X}_1$ مع $\bar{X}_2$, $\bar{X}_3$ معًا استخدم المقارنة الخطية Linear Contrast (متاح فى SPSS) .
- واذا كان لديك مقارنات قبليه عديد (قبل جمع البيانات) وليس بالضرورة بين كل زوج فاستخدام إجراء Multistage Bonferroni
- اذا كان لديك عدد كبير من المجموعات وترغب فى عمل مقارنات عديدة (بعد جمع البيانات) سواء كنت مهتم أو غير مهتم بكل المقارنات المحتملة بين ازواج المتوسطات فيصبح باستخدام Tukey أو إجراء REGWQ مفضل. ويشير Howell (2013) انا لا اعتقد فى هذا الموقف ان أفضل اختبار هو شيفيه ولكنه هو اختبار شائع الاستخدام، ويفضل استخدام اختبار بنجامين- هوتسبيرج لأنه أكثر قوه عند التعامل مع نسب خطا مختلفة.

وبعد ان استعراض اختبارات المقارنات المتعددة سوف يطبق هذا على اختبار تحليل التباين التى سبق عرضة فى فصل تحليل التباين الأحادى الاتجاه وهى المقارنة بين مستوى الضغوط بين ثلاثة مواقف:

مصدر التباين	SS	dF	MS	F	Fc.v
بين المجموعات	203.33	2	101.667		
داخل المجموعات	54	12	4.500	22.59	3.88
الكلى	257.333	14			

	المجموعة الأولى	المجموعة الثانية	المجموعة الثالثة
$\bar{x}$	4	8	13
N	5	5	5

اولًا: المقارنات القبلية (المحددة قبل جمع البيانات):

1.إجراء اختبار T المتعددة.

أ – مقارنه المجموعة الأولى والثالثة:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2MS_w}{n}}}$$

$$T = \frac{4 - 13}{\sqrt{\frac{2\times1.5}{5}}} = -6.71$$

ب –المجموعة الثانية والثالثة:

$$T = \frac{8 - 13}{\sqrt{\frac{2\times4.5}{5}}} = -3.74$$

حيث قرر الباحث قبل إجراء التجربة إجراء المقارنات فى حالة الحصول على دلاله إحصائية لـF وبالبحث فى جدول Tبـ $\alpha = 0.05,$ ، $df_w = N - K=15-3=12$ واختبار ذو ذيل واحد حيث افترض الباحث ان:

$H_A : \mu_1 < \mu_2$

$H_A : \mu_2 < \mu_3$

فان :(T) الحرجة = 1.78

$1.78 < T\ 2,3\ ,\ T\ 1,2$

وعليه توجد فروق لصالح المجموعة الثالثة (اعتمد على القيمة المطلقة).

ثانيًا: المقارنات البعدية:

1.**اختبار توكى (HSD):** الخطوة الأولى: حساب قيمه Q وهى كالآتى:

ا. المجموعة الأولى والثانية:

$$Q = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{MS_W}{n}}} = \frac{8-4}{\sqrt{\frac{4.50}{5}}} = 4.21$$

ب . مقارنة المجموعة الأولى والثالثة:

$$Q = \frac{13-4}{\sqrt{\frac{4.50}{5}}} = 9.48$$

ج . مقارنة المجموعة الثانية والثالثة:

$$Q = \frac{13-8}{\sqrt{\frac{4.50}{5}}} = 5.27$$

د. حساب قيمة Q الحرجة : من خلال جدول Q (ملحق) حيث $K = 3$ (عدد المجموعات)، $df_w = 15 - 3 = 12$، $\alpha = 0.05$، فان:

$$Qc.v = 3.77 \text{ (الحرجة)}$$

وعليه فإن قيمه Q المحسوبة للمقارنة الثلاثة $Q <$ Q (الحرجة)

$$3.77 < 5.27 , 9.48 , 4.21$$

بالتالى فإن $\mu_1 \neq \mu_2 \neq \mu_3$ حيث تختلف المواقف الثلاثة فى الضغوط النفسية.

2. اختبار شيفيه: وفيه يتم المقارنة بين المجموعة الأولى مع المجموعة الثانية والثالثة معًا وكذلك المجموعة الثانية مع المجموعة الأولى والثالثة معًا، ويرى(2013) Panago ان اختبار شيفيه يثبت قيمه α لكل المقارنات البعدية المحتملة ولذلك فإنه من أكثر اختبارات المقارنات البعدية امانًا ضد عدم ارتكاب أو تضخم الخطأ من النوع

الأول ورغم عن ذلك يستخدم لإجراء مقارنات كل زوج من المتوسطات وليس كل متوسط مع بقيه المتوسطات مجتمعة معًا كما اشرنا سابقا انظر مخرج SPSS.

مقارنه بين المقارنات المخططة (القبلية) واختبار توكى (HSD) واختبار شيفيه

فى المثال السابق تم حساب المقارنات المتعددة باستخدام ثلاثة طرق كالتالى:

مقارنات مخططة			مقارنات بعدية			
T المتعددة			توكى		شيفيه	
مقارنات المجموعات	T	P	Q	P	F	P
2*1			4.21	0.029	4.44	0.036
3*1	-6.71	0.00001	9.49	0.001006	22.50	0.00009
3*2	-3.73	0.001	5.27	0.008	6.49	0.010

الجدول يعرض قيمه الإحصاء المحسوبة لكل مقارنه وكذلك قيمه P (مخرج SPSS) حيث يتم رفض H_0 إذا كان $P \leq \alpha$، ويمكن أن نلاحظ أن قيمه P لاختبار توكى تتراوح من ستة إلى ثمانية مرات أعلى من القيمة الاحتمالية لـ P فى المقارنات المخططة T، وكذلك القيمة الاحتمالية P لاختيار شيفيه تتراوح أو أكبر1.2 إلى 1.5 مرة منP لاختبار توكى لنفس المقارنات وهذا يدل على أن المقارنات القبلية (المخطط لها) أكثر قوة إحصائيًا من المقارنات البعدية (شيفيه – توكى). وفى المقارنات البعدية فإن اختبار توكى أكثر قو من شيفيه، ولكن المقارنات القبلية لا تصحح (تضبط) احتمال زيادة الوقوع فى الخطأ من النوع الأول ولكنها لا بد أن تكون قائمة على منطقية ومعنى من قبل الباحث الذى يقوم بالتجربة بمعنى أن يحدد مسبقًا فى فروضه أى المجموعات التى ربما تسهم فى حدوث الدلالة الإحصائية لتحليل التباين.

والاختيار بين شيفيه وتوكى يعتمد على فلسفه البحث، فإذا أراد الباحث الحصول على اختبار أكثر قوة إحصائيًا لرفض (H_0) بمعنى الحصول على دلالة إحصائية بين المقارنات مع ضبط معقول للخطأ من النوع الأول فإن اختبار Tukey مفضل عن

Scheffe(Pagano, 2013) . ويرى Gravetter& Wallnau (2014) بأنه يمكن الحصول على دلالة إحصائية بين ازواج المتوسطات بواسطة استخدام توكى ولكن لا نحصل عليها باستخدام شيفيه، ولكن شيفيه أكثر أمانًا لأن لديه وقاية كبيرة من تضخم الخطأ من النوع الأول. ولكن إذا اراد الباحث الذهاب ابعد من المقارنات المزدوجة الثنائية بين المتوسطات Post hoc pair-wise comparisons بمعنى المقارنة بين المجموعة الواحدة وبقية المجموعات معا فإنه يفضل استخدام شيفيه.

ومن الأفضل استخدام طريقة توكى _ كرامير فى حاله عدم تساوى العينات ولكن هذا نادر الاستخدام فى التراث. كما يفضل استخدام طريقة نيومان – كيولز لأنها أكثر قوة إحصائيًا من طريقة توكى.

وفيما يلى ملخص لطرق المقارنات المتعددة:

الإجراء	H_0	استخدامه	الإحصائية	التوزيع التحتي
توكى	$M_i = M_k$	**مقارنات زوجية مع تساوى n**	$Q = \frac{\bar{X}_i - \bar{X}_k}{\sqrt{\frac{MS_w}{n}}}$	**توزيع Q مع df (r, N –K)**
نيومان- كرامير	$M_i = M_k$	**ازدواج المقارنات مع تساوى n**	$Q = \frac{\bar{X}_i - \bar{X}_k}{\sqrt{\frac{MS_w}{n}}}$	**توزيع Q مع df(r,N –K)**
توكى – كرامير	$M_i = M_k$	**ازواج المقارنات مع عدم تساوى n**	$Q = \frac{\bar{X}_i - \bar{X}_k}{\sqrt{MS_w \frac{1/n1 + 1/nk}{2}}}$	**توزيع Q مع df(r, N–K)**
شيفيه	$\sum C_k\mu_k = 0$	**مقارنات زوجية مع عدم تساوى n ومقارنات معقده**	$F = \frac{(\sum Ck\bar{X}k)^2}{MS_w \sum(\frac{C_k^2}{n_k})}$	**توزيع F مع df (K - 1 , N-K)**
المقارنات المتعامدة المخطط لها	$\sum C_k\mu_k = 0$	**مقارنات متعامدة**	$F = \frac{(\sum Ck\bar{X}k)^2}{MS_w \sum(\frac{C_k^2}{n_k})}$	**توزيع F مع df (1 , N-K)**

K عدد المجموعات، r عدد المقارنات، N حجم العينه الكلى، n حجم العينة فى لمجموعة الواحدة .

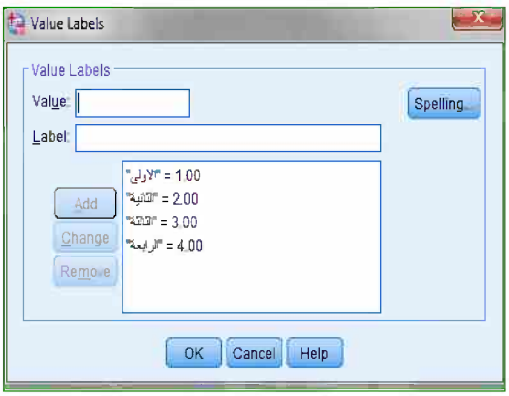
ويبدو ان إجراءات المقارنات البعدية مثل فيشر واختبار Duncan ونيومان – كيولز هى أقل استخدامًا فى الدراسات، فى حين أن اختبار توكى واختبار شيفيه هى أكثر استخدامًا واحيانًا يستخدم الباحث مقارنات بعدية ويطلقوا عليها مقارنات مخططة أو العكس صحيح.

تنفيذ اختبار F مقرونًا بالمقارنات المتعددة البعدية فى برنامج SPSS

أولًا: إخال البيانات: 1.اضغط Variable view وحدد مسمى المتغيرات فى عمود Name كالآتى: **المتغير المستقل: طريقة** التنشئة Method ويأخذ الكود الآتى (طريقة التنشئة الأولى =1، الثانية =2، الثالثة=3، الرابعة=4 ،الخامسة=5)، **المتغير التابع** التحصيل:achievement

2. فى عمود Values أكتب الكود المقابل للمتغيرMethod كالآتى:

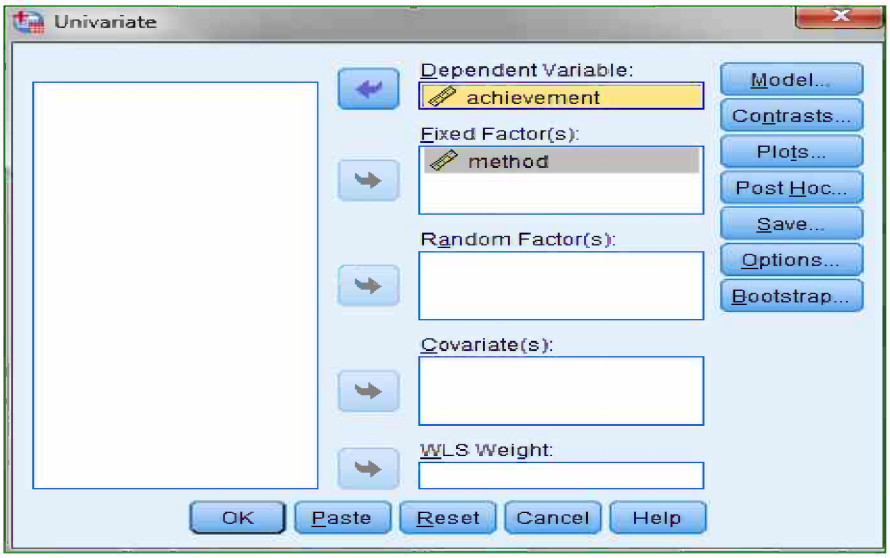
ا- اضغط على الخانة تظهر الشاشة الآتية:



ب – أكتب الكود1 أمام Value ثم أكتب مسماة أمام Label ثم اضغط Add ثم أكتب 2 ومسماه وأضغط Add وهكذا.

ثانيًا: تنفيذ F مقرونًا بالمقارنات البعدية:

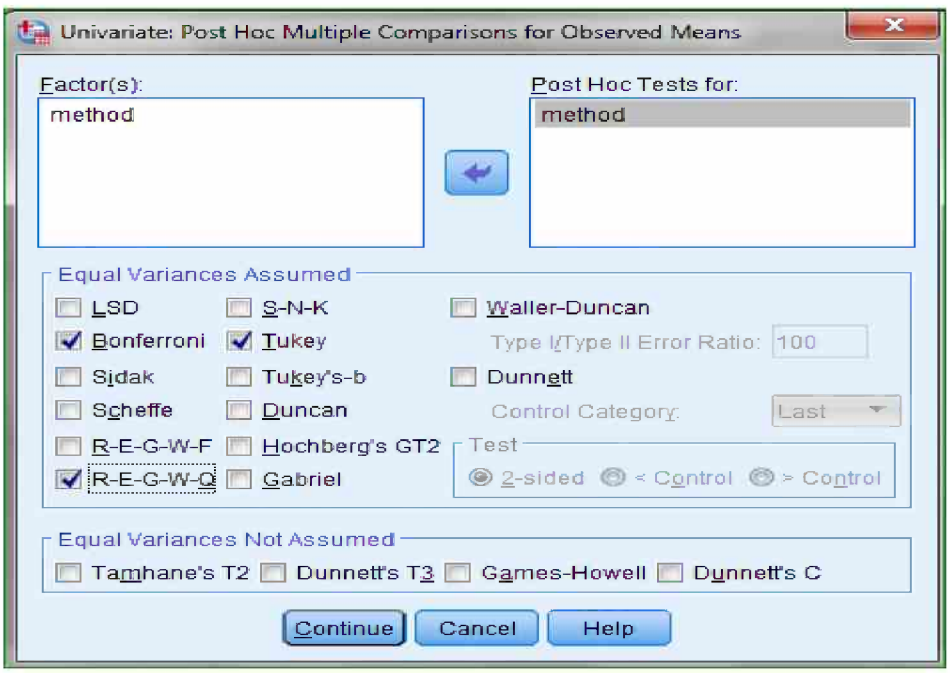
1. اتبع خطوات تنفيذ اختبار Fمن خلال Analyze→General Linear Model→ Univariate



2. اضغط على Option, اختار Homogeneity of variance test

تنفيذ Post hoc

3. اضغط اختيار Post Hoc يعطى الشاشة الآتية:



4. انقل متغير method إلى مربع Post hoc test

5. اضغط Bonferroni، Tukey، R.E.G.W.Q

6. اضغط Continue ثم اضغط Ok

أما فى حالة تنفيذ الأمر من خلال: Analyze → Compare Means → One –Way ANOVA واتبع باقى الخطوات.

ويوجد العديد من طرق المقارنات المتعددة البعدية كما سبق عرضها ويفضل فى حالة تحقق تساوى البيانات استخدام Tukey, R-E-G-W-Q وكذلك يفضل Bonferroni لأنها تتحكم فى الخطأ من النوع الأول ولا يفضل استخدام kauls - Newman لأنها لا تضبط تضخم الخطأ من النوع الأول وكذلك لا يفضل استخدام least-significance difference((LSD) لأنها لا تتحكم فى الخطأ من النوع الأول أيضًا وهو مكافئ لإجراء اختبار T لكل زوج من المستويات .

ويفضل استخدام GT2,Gabriel Hochbergs و Tukey-Kramer عندما تكون أحجام عينات المجموعات غير متساوية. وعندما لا تتحقق مسلمة التجانس للتباينات توجد عدة خيارات وأهمها T3 Dunnetts، وعليه ففى المقارنات المتعددة عند تحقق مسلمة تجانس التباينات وتساوى العينات عبر المستويات فإن المفضل هو REGWQ, Tukey، وإذا كان أحجام العينات مختلفة عبر المستويات بدرجة خفيفة يفضل استخدام Gabriel ولكن إذا وجد تفاوت كبير فى أحجام العينات عبر المستويات يفضل استخدام GT2 Hochbergs هذا مع تساوى التباينات وكذلك يستخدم طريقة Tukey-Kramerوطريقة Newman kauls

ثالثًا: المخرج: اعطى جدول التحقق من مسلمة تجانس التباينات كالآتى:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: achievement

F	df1	df2	Sig.
1.149	4	35	.350

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + method

وبما أن $P=0.350 > 0.05$ وعليه نقبل H_0.

ثم أعطى جدول ANOVA كالآتى:

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3480.900[a]	4	870.225	27.418	.000
Intercept	9703.225	1	9703.225	305.717	.000
method	3480.900	4	870.225	27.418	.000
Error	1110.875	35	31.739		
Total	14295.000	40			
Corrected Total	4591.775	39			

a. R Squared = .758 (Adjusted R Squared = .730)

حيث أظهرت النتائج أن$P=0.000$, $F=27.418$ ، وعلية يتم رفض H_0 بمعنى وجود فروق ذات دلالة إحصائية بين متوسطات التحصيل بين المجموعات الخمسة ولمعرفة أى مجموعة أحدثت هذه الدلالة تم إجراء المقارنات البعدية وكان الجدول كالآتى:

Multiple Comparisons

Dependent Variable: achievement

	(I) method	(J) method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	الاولى	الثانية	-6.0000-	2.81688	.231	-14.0987-	2.0987
		الثالثة	-7.0000-	2.81688	.117	-15.0987-	1.0987
		الرابعة	-19.8750-*	2.81688	.000	-27.9737-	-11.7763-
		5.00	-25.0000-*	2.81688	.000	-33.0987-	-16.9013-
	الثانية	الاولى	6.0000	2.81688	.231	-2.0987-	14.0987
		الثالثة	-1.0000-	2.81688	.996	-9.0987-	7.0987
		الرابعة	-13.8750-*	2.81688	.000	-21.9737-	-5.7763-
		5.00	-19.0000-*	2.81688	.000	-27.0987-	-10.9013-
	الثالثة	الاولى	7.0000	2.81688	.117	-1.0987-	15.0987
		الثانية	1.0000	2.81688	.996	-7.0987-	9.0987
		الرابعة	-12.8750-*	2.81688	.001	-20.9737-	-4.7763-
		5.00	-18.0000-*	2.81688	.000	-26.0987-	-9.9013-
	الرابعة	الاولى	19.8750*	2.81688	.000	11.7763	27.9737
		الثانية	13.8750*	2.81688	.000	5.7763	21.9737
		الثالثة	12.8750*	2.81688	.001	4.7763	20.9737
		5.00	-5.1250-	2.81688	.379	-13.2237-	2.9737
	5.00	الاولى	25.0000*	2.81688	.000	16.9013	33.0987
		الثانية	19.0000*	2.81688	.000	10.9013	27.0987
		الثالثة	18.0000*	2.81688	.000	9.9013	26.0987
		الرابعة	5.1250	2.81688	.379	-2.9737-	13.2237

فبالنسبة لطريقة (HSD) Tukey نلاحظ أنه لا توجد فروق بين متوسط المجموعة الأولى ومتوسط الثانية (لا دلالة إحصائية)P=0.231، والأولى مع الثالثة لا دلالة P=0.117، الأولى مع الرابعة توجد دلالة P=0.000، والأولى مع الخامسة توجد دلالة P=0.000، وكذلك المجموعة الثانية مع الرابعة، والثانية مع الخامسة أحدثت دلالة، وكذلك متوسط المجموعة الثالثة مع الرابعة، والثالثة مع الخامسة وهكذا.

كما أعطى نتائج Bonferroni وهى نفس الاستنتاجات التى توصلنا إليها من خلال اختبار توكى كالآتى:

Bonferroni	الاولى	الثانية	-6.0000-	2.81688	.403	-14.4395-	2.4395
		الثالثة	-7.0000-	2.81688	.179	-15.4395-	1.4395
		الرابعة	-19.8750-*	2.81688	.000	-28.3145-	-11.4355-
		5.00	-25.0000-*	2.81688	.000	-33.4395-	-16.5605-
	الثانية	الاولى	6.0000	2.81688	.403	-2.4395-	14.4395
		الثالثة	-1.0000-	2.81688	1.000	-9.4395-	7.4395
		الرابعة	-13.8750-*	2.81688	.000	-22.3145-	-5.4355-
		5.00	-19.0000-*	2.81688	.000	-27.4395-	-10.5605-
	الثالثة	الاولى	7.0000	2.81688	.179	-1.4395-	15.4395
		الثانية	1.0000	2.81688	1.000	-7.4395-	9.4395
		الرابعة	-12.8750-*	2.81688	.001	-21.3145-	-4.4355-
		5.00	-18.0000-*	2.81688	.000	-26.4395-	-9.5605-
	الرابعة	الاولى	19.8750*	2.81688	.000	11.4355	28.3145
		الثانية	13.8750*	2.81688	.000	5.4355	22.3145
		الثالثة	12.8750*	2.81688	.001	4.4355	21.3145
		5.00	-5.1250-	2.81688	.774	-13.5645-	3.3145
	5.00	الاولى	25.0000*	2.81688	.000	16.5605	33.4395
		الثانية	19.0000*	2.81688	.000	10.5605	27.4395
		الثالثة	18.0000*	2.81688	.000	9.5605	26.4395
		الرابعة	5.1250	2.81688	.774	-3.3145-	13.5645

Based on observed means.
The error term is Mean Square(Error) = 31.739.

*. The mean difference is significant at the .05 level.

ثم أعطى البرنامج Homogenous Subset :

Homogeneous Subsets

achievement

	method	N	Subset 1	Subset 2
Tukey HSD[a,b]	الأولى	8	4.0000	
	الثانية	8	10.0000	
	الثالثة	8	11.0000	
	الرابعة	8		23.8750
	5.00	8		29.0000
	Sig.		.117	.379
Ryan-Einot-Gabriel-Welsch Range[b,c]	الأولى	8	4.0000	
	الثانية	8	10.0000	
	الثالثة	8	11.0000	
	الرابعة	8		23.8750
	5.00	8		29.0000
	Sig.		.075	.182

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 31.739.

a. Uses Harmonic Mean Sample Size = 8.000.

b. Alpha = .05.

c. Critical values are not monotonic for these data. Substitutions have been made to ensure monotonicity. Type I error is therefore smaller.

وهذا الجدول يتضمن الطريقة بكودها ثم حجم العينة فى كل طريقة ثم متوسطات المجموعات حيث إن متوسط درجات تحصيل طريقة التنشئة الأولى =4، بينما متوسط درجات تحصيل عينة طريقة التنشئة الخامسة=29. ويتضمن أحد الأعمدة متوسطات ثلاث مجموعات هى الأولى والثانية والثالثة ولا يوجد بين أى زوج منها دلالة إحصائية، والعمود الثانى متوسطات المجموعة الرابعة والخامسة لا يوجد دلالة بينها ولكن بين أى متوسط من الأولى أو الثانية أو الثالثة مع متوسطات الرابعة والخامسة يوجد دلالة إحصائية.

وتكتب النتائج كالآتى: أشارت نتائج تحليل التباين الأحادى بوجود فروق دالة إحصائيا بين متوسطات تحصيل المجموعات الخمسة : F=27.418 , P=0.000 η_p^2 =0.758, 0.05 > ، وبإجراء اختبارات المقارنات المتعددة بإستخدام توكى أشارت النتائج إلى وجود فروق بين متوسطات المجموعة الثانية والخامسة و d=2.37 , p=0.000 وفروق بين متوسطات المجموعة الخامسة والأولى p= 0.000 وبين المجموعة الخامسة والثالثة، والخامسة والرابعة p=0.000.

ويمكن إجراء الاختبارات البعدية باستخدام شيفيه حيث يسمح بمقارنة أى متوسط مع باقى الأزواج. ويمكن تنفيذه باتباع حل الخطوات السابقة وفى Post hoc

اضغط scheffe يظهر المخرج الآتى:

Multiple Comparisons

Dependent Variable: achievement

Scheffe

(I) method	(J) method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
الاولى	الثانية	-6.0000-	2.81688	.356	-15.1563-	3.1563
	الثالثة	-7.0000-	2.81688	.211	-16.1563-	2.1563
	الرابعة	-19.8750-*	2.81688	.000	-29.0313-	-10.7187-
	5.00	-25.0000-*	2.81688	.000	-34.1563-	-15.8437-
الثانية	الاولى	6.0000	2.81688	.356	-3.1563-	15.1563
	الثالثة	-1.0000-	2.81688	.998	-10.1563-	8.1563
	الرابعة	-13.8750-*	2.81688	.001	-23.0313-	-4.7187-
	5.00	-19.0000-*	2.81688	.000	-28.1563-	-9.8437-
الثالثة	الاولى	7.0000	2.81688	.211	-2.1563-	16.1563
	الثانية	1.0000	2.81688	.998	-8.1563-	10.1563
	الرابعة	-12.8750-*	2.81688	.002	-22.0313-	-3.7187-
	5.00	-18.0000-*	2.81688	.000	-27.1563-	-8.8437-
الرابعة	الاولى	19.8750*	2.81688	.000	10.7187	29.0313
	الثانية	13.8750*	2.81688	.001	4.7187	23.0313
	الثالثة	12.8750*	2.81688	.002	3.7187	22.0313
	5.00	-5.1250-	2.81688	.517	-14.2813-	4.0313
5.00	الاولى	25.0000*	2.81688	.000	15.8437	34.1563
	الثانية	19.0000*	2.81688	.000	9.8437	28.1563
	الثالثة	18.0000*	2.81688	.000	8.8437	27.1563
	الرابعة	5.1250	2.81688	.517	-4.0313-	14.2813

Based on observed means.
The error term is Mean Square(Error) = 31.739.

*. The mean difference is significant at the .05 level.

ويتضح أن هذا المخرج يتشابه تمامًا مع مخرج Tukey وأن برنامج SPSS لا يتعامل مع شيفيه كما يجب أن يكون ولكنه يتعامل على أساس مقارنة كل زوج على حدة.

إجراء المقارنات القبلية المخطط لها مثل Dunncan ,Sidak

ويتم تنفيذها بإتباع الخطوات السابقة فى Post hoc اضغط على أى من هذه الاختبارات، وفيما يلى المخرج:

Homogeneous Subsets

achievement

Duncan[a,b]

method	N	Subset		
		1	2	3
الاولى	8	4.0000		
الثانية	8		10.0000	
الثالثة	8		11.0000	
الرابعة	8			23.8750
5.00	8			29.0000
Sig.		1.000	.725	.077

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 31.739.

a. Uses Harmonic Mean Sample Size = 8.000.

b. Alpha = .05.

وهو مثل مخرج شيفيه وتوكى ولكن هنا الباحث حدد مسبقًا الدلالة الإحصائية لصالح أى مجموعة لأنها مخطط لها.

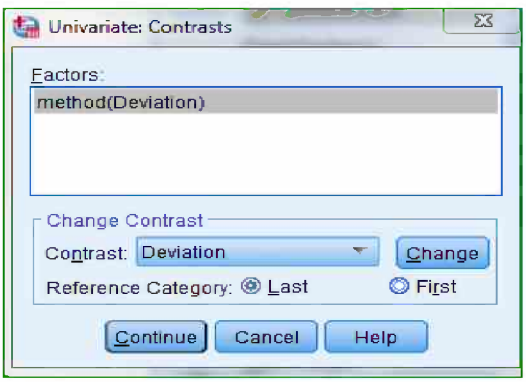
إجراء المقارنات Contrast

هى من نوع اختبارات المقارنات القبلية المخطط لها والفرق بين المقارنات القبلية والمقارنات البعدية هو الفرق بين الاختبارات ذوات ذيل واحد وذوات ذيلين، فالاختبارات القبلية لدينا فروض عن طبيعة اتجاه الدلالة لصالح أى مجموعة. ففى المثال السابق للباحث فكرة مسبقة عن طبيعة الفرق بين المجموعة الخامسة والأولى وهى اتجاه الفرق لصالح المجموعة الخامسة. وعند المقارنة بين المجموعة التجريبية والضابطة فالباحث لديه فكرة مسبقة عن اتجاه الدلالة لصالح المجموعة التجريبية. وتعتبر الفرق بين المجموعة الخامسة والأولى هى مقارنة أولى 1 Contrast فمثلًا تخيل تصميم تجريبى بثلاث مجموعات تجريبية ومجموعة ضابطة فالمقارنة الأولى تتضمن مقارنة المجموعات التجريبية الثلاثة على حدة بالمجموعة الضابطة وعلية توجد فروق بين المجموعات التجريبية الثلاثة والمجموعة الضابطة، والمقارنة الثانية تتضمن مقارنة المجموعتين التجريبيتين E_1، E_2 بالمجموعة التجريبية E_3 بينما المقارنة الثالثة تتضمن مقارنة E_1 بـ E_2 .

لتنفيذ المقارنات فى SPSS اتبع كل الخطوات السابقة:

1. اتبع خطوات تنفيذ اختبار F من خلال Analyze → General Linear Model

2. اضغط اختيار Contrast تظهر الشاشة الآتية:



3. يمكن اختيار مقارنات مختلفة عديدة فى المربع امام Contrast اضغط على السهم تظهر عدة بدائل وهى:

Deviation : مقارنة متوسط كل مستوى أو مجموعة مع المتوسط العام لباقى المستويات Grand Mean

Simple: مقارنة متوسط كل مستوى أو مجموعة بمتوسط مستوى محدد مسبقًا كالمجموعة الضابطة.

Difference: مقارنة متوسط كل مستوى (ماعدا الأول) بمتوسط المستويات السابقة على حدة.

Helmert: مقارنة متوسط كل مستوى (ما عدا الأخير) بمتوسط المستويات اللاحقة على حده.

Repeated: مقارنة متوسط كل مستوى بالذى يليه مباشرة فمثلًا (1, 2) و(2, 3) (3, 4) و(4 ,5).

Polynomial : مقارنة التأثير الخطى أو تأثير Quadratic ، وهذا الاختيار فى غاية الأهمية لمعرفة الاتجاه الخطى Linear Trend للمتوسطات.

حيث Quadratic تعنى أنه يوجد تغير وحيد فى اتجاه الخطية حيث تغير المتوسط للمجموعة فى ضوء المستويات الباقية.

4. اختر Deviation ثم اضغط Change واضغط Continu ثم اضغط Ok

المخرج: كما هو واضح فإن البرنامج أعطى مقارنة بين متوسط المجموعة الأولى مع متوسطات المجموعات الأربعة الأخرى معا حيث: Contrast Estimate = - 11.575، Sig= 0.000، وبما أن 0.05> 0.000، توجد فروق ذات دلالة إحصائية بين متوسط المجموعة الأولى ومتوسط باقى المجموعات معًا.

وهكذا لمتوسط المجموعة الثانية وباقى المتوسطات معًا حيث0.05 > Sig = 0.004.

method Deviation Contrast[a]			Dependent Variable achievement
Level 1 vs. Mean	Contrast Estimate		-11.575-
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-11.575-
	Std. Error		1.782
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	-15.192-
		Upper Bound	-7.958-
Level 2 vs. Mean	Contrast Estimate		-5.575-
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-5.575-
	Std. Error		1.782
	Sig.		.004
	95% Confidence Interval for Difference	Lower Bound	-9.192-
		Upper Bound	-1.958-
Level 3 vs. Mean	Contrast Estimate		-4.575-
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-4.575-
	Std. Error		1.782
	Sig.		.015
	95% Confidence Interval for Difference	Lower Bound	-8.192-
		Upper Bound	-.958-
Level 4 vs. Mean	Contrast Estimate		8.300
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		8.300
	Std. Error		1.782
	Sig.		.000
	95% Confidence Interval	Lower Bound	4.683

وإذا كان الباحث يهدف مقارنة بين متوسط كل مجموعة بمتوسط محدد مسبقًا مثل متوسط المجموعة الضابطة. وإذا اعتبرنا أن المجموعة الأولى الضابطة فيتم اختيار بديل Simple وتتبع الخطوات الآتية:

1. اضغط Contrast.

2. فى مربع Change Contrast اضغط على السهم واختر بديل Simple.

3. أمام Reference Category اضغط First باعتبار أن المجموعة الضابطة لها كود1

4. اضغط على Continue ثم اضغط على Ok.

method Simple Contrast[a]			Dependent Variable: achievement
Level 2 vs. Level 1	Contrast Estimate		6.000
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		6.000
	Std. Error		2.817
	Sig.		.040
	95% Confidence Interval for Difference	Lower Bound	.281
		Upper Bound	11.719
Level 3 vs. Level 1	Contrast Estimate		7.000
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		7.000
	Std. Error		2.817
	Sig.		.018
	95% Confidence Interval for Difference	Lower Bound	1.281
		Upper Bound	12.719
Level 4 vs. Level 1	Contrast Estimate		19.875
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		19.875
	Std. Error		2.817
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	14.156
		Upper Bound	25.594
Level 5 vs. Level 1	Contrast Estimate		25.000
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		25.000
	Std. Error		2.817
	Sig.		.000
	95% Confidence Interval	Lower Bound	19.281

لاحظ تمت المقارنة بين المجموعة الأول وكل متوسطات باقى المجموعات على حدة، حيث توجد فروق بين متوسطات تحصيل المجموعة الأولى والثانية0.04 = Sig وكذلك بين الأولى والثالثة 0.018 = Sig وهكذا. وإذا كان الباحث يهدف إلى المقارنة بين كل متوسط المجموعة بمتوسط المجموعة التى تسبقها مباشرة فاتبع الخطوات الآتية:

1. فى المربع الذى امام Contrast اضغط على السهم واختارDiffirence ثم اضغط على Change

2.اضغط على Continue ثم اضغط على OK.

method Difference Contrast			Dependent Variable: achievement
Level 2 vs. Level 1	Contrast Estimate		6.000
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		6.000
	Std. Error		2.817
	Sig.		.040
	95% Confidence Interval for Difference	Lower Bound	.281
		Upper Bound	11.719
Level 3 vs. Previous	Contrast Estimate		4.000
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		4.000
	Std. Error		2.439
	Sig.		.110
	95% Confidence Interval for Difference	Lower Bound	-.952-
		Upper Bound	8.952
Level 4 vs. Previous	Contrast Estimate		15.542
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		15.542
	Std. Error		2.300
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	10.872
		Upper Bound	20.211
Level 5 vs. Previous	Contrast Estimate		16.781
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		16.781
	Std. Error		2.227
	Sig.		.000
	95% Confidence Interval	Lower Bound	12.260

الفصل الثانى والعشرون

ما وراء التحليل

Meta-Analysis

ماهية ما وراء التحليل

فى عام 1952 اثار Hans Eysenck جدلًا واسعًا فى مجال علم النفس الاكلنيكى وذلك بان برامج العلاج النفسى ليس لها جدوى أو فائدة على المرضى، وفى منتصف عام 1970 وجدت المئات من الدراسات فى العلاج النفسى بعضها اعطى نتائج موجبة وبعضها نتائج سالبة وبعضها نتائج صفرية، والمراجعة الكيفية لهذه الدراسات لم تستطيع حل هذا الجدل القائم بمدى جدوى برامج العلاج النفسى. وللرد على ادعاء Eysneck قام Glass(1977) & Smith بمراجعة كمية إحصائية معيارية وذلك بحساب متوسط الفروق ابين متوسطات لمجموعات التجريبية والضابطة لـ 375 دراسة فى مجال العلاج النفسى وسما هذه الطريقة الإحصائية لمراجعة الدراسات السابقة كميًا بما وراء التحليل Meta- analysis واستنتجا ان برامج العلاج النفسى فعالة، ولكن Eysneck (1978) لم يقتنع بهذه الطريقة وحاول ان يقلل من شان هذه الطريقة بقوله انها ممارسة الغباء الذائد عن الحد An exercise in mega-silliness وبالرغم من انتقاد Eysneck والآخرين لهذا الأسلوب إلا أنها أصبحت أكثر انتشارًا وقبولًا كطريقة لتلخيص نتائج الدراسات الإمبريقية داخل العلوم الاجتماعية والسلوكية والطبية.

ومنذ ابتدع علماء النفس أسلوب ما وراء التحليل منذ 1970 اصبح أكثر استخدامًا فى كافة التخصصات وخاصة الطبية وحدثت تطورات كبيرة فى هذه المنهجية سواء بالكتب، الدراسات، والنماذج التحليلية، ومن التطورات الحديثة تنفيذ ما وراء التحليل باستخدام برامج إحصائية متخصصة مثل CMA و MetaXL وكذلك فى البرامج الإحصائية العامة مثل SPSS, SAS, Stata والبرنامج الإحصائى المتقدم R .

مفهوم ما وراء التحليل

يعتبر أسلوب ما وراء التحليل مدخل منهجى وإحصائى شامل للوصول إلى الاستنتاجات من التراث البحثى وأحد الطرق لإحداث الدمج أو التوليف المنظم للتراث الإمبريقى (الدراسات الأولية) لظاهرة ما بهدف التكامل بين النواتج للوصول إلى تعميمات أو حل التناقضات بين الدراسات. ويُعرف ما وراء التحليل بانه التحليل الإحصائى لمجموعة كبيرة من نتائج الدراسات المفردة بهدف احداث تكامل لنواتجها (Glass, 1976)، وما وراء التحليل طريقة لإجراء التحليل الإحصائى لنواتج البحوث لكثير من الدراسات المفردة المستقلة لموضوع محدد (Borenstein et al., 2009; Hedges & Olkin, 1985) ، أو التوليف الكمى لبيانات من دراسات متعددة (Leandro, 2005)، أو مجموعة من الطرائق الإحصائية لتوليف النتائج الكمية من دراسات متعددة لإعطاء ملخص للنتائج الإمبريقية لموضوع محدد (Littlell, Corcoran, & Pillai, 2008)، أو مدخل قوى وفعال لتلخيص ومقارنة نتائج التراث البحثى لظاهرة ما (Card, 2012).

ويطلق على أسلوب ما وراء التحليل بالتوليف الكمى Quantitative synthesis ويستخدم احيانًا كمرادف للتوليف البحثى، ويتضمن أسلوب ما وراء التحليل مجموعة من الأساليب الإحصائية المستخدمة فى المراجعة المنظمة للدراسات السابقة لتقليل التحيز كلما امكن. والإجراءات الإحصائية المستخدمة فى أسلوب ما وراء التحليل يمكن تطبيقها لأى مجموعة بيانات. ويهدف ما وراء التحليل إلى الوصول إلى استنتاج وحيد وشامل لنتائج دراسات مختلفة لنفس الموضوع باستخدام نفس الطرق، ويساعد فى فهم نتائج تجارب التدخل (المنهج التجريبى وشبة التجريبى)، ولا يساعد ما وراء التحليل فى الحصول المعرفة التراكمية فقط بل يعطى مؤشر واتجاه واضح عن ماهى الفجوات أو الموضوعات فى التراث البحثى التى هى بحاجة لمزيد من الدراسة والبحث فى المستقبل.

الحاجة إلى ما وراء التحليل فى العلوم السلوكية والاجتماعية

الملاحظ تراكم كم هائل من الدراسات فى تراث علم النفس والعلوم الاجتماعية، فالحاجة اليوم ليس لمزيد من الدراسات الإمبريقية التى تتناول ظاهرة ما بل لإعطاء نظرة تقييمية

شاملة للكم الهائل من البيانات المتراكمة على مدى أكثر من خمسين عامًا بمعنى اعادة تحليل هذا التراث البحثى المتصارع، والاتجاه إلى التكامل بين النتائج المتعارضة لمحاولة الوصول إلى معرفة عامة وشاملة تتوحد فيها هذه النتائج المتضاربة فى محاولة الوصول إلى بناء نظرية متكاملة.

فالدراسات ذات أحجام العينات الصغيرة Small-sample studies فى البحث النفسى التى تحاول الاجابة على نفس السؤل ينتج عنها نتائج متضاربة وهذه الدراسات تعتمد على اختبارات الدلالة الإحصائية وهذا يجعل نتائج الدراسات أكثر تعارضًا لكون القرار ثنائى(نعم-لا)، ولذلك فإن ما وراء التحليل يهدف إلى حدوث تكامل عبر هذه الدراسات للكشف عن النمط العام للعلاقات التى تكمن فى التراث البحثى وهذا هو الأساس لبناء نظرية جديدة أو تطويرها. ما وراء التحليل يصحح النتائج (التأثيرات- العلاقات) المتعارضة من خطأ المعاينة وأخطاء القياس واى عوامل اخرى تؤدى إلى ظهور النتائج المتعارضة. وهدف أى علم هو انتاج المعرفة المتراكمة التى تتضمن العلاقات والنظريات التى تستطيع تفسير الظواهر المرتبطة به. فعلى سبيل المثال النظريات التى تفسر كيفية تطور ونمو سمات الشخصية عند الاطفال والراشدين عبر الزمن وكيف لهذه السمات ان تؤثر فى حياتهم، أو النظرية التى تفسر العوامل المساهمة فى الرضا الوظيفى أو النظرية التى تفسر العوامل المساهمة فى والتحصيل وغيرها.

وبدون تحديد واضح لطبيعة العلاقات بين متغيرات الظاهرة فلا نستطيع بناء نظرية متماسكة لتفسيرها، وكذلك الحصول نتائج متعارضة مثل الحصول على علاقات قوية موجبة وعلاقات قوية سالبة ولا علاقات فإن هذا بدوره لا يساعد على بناء نظرية قوية متماسكة تستطيع تفسير جوانب الظاهرة المختلفة. والمهمة الأساسية فى العلوم الاجتماعية والسلوكية هو بناء النظرية، والنظرية الجيدة هى التى لها القدرة على تفسير طبيعة العمليات والتفاعلات التى تحدث داخل الظاهرة وتفسرها بصورة سهلة وبسيطة. والنظرية عبارة عن تفسيرات نظرية وعلاقات بين متغيرات الظاهرة وتنتج من تراكم من الاطروحات النظرية مقرونة بالتحقق الإمبريقى من الدراسات.

ومن خلال المقارنة الكمية لنواتج الدراسات المتنوعة فإن ما وراء التحليل يكتشف معرفة جديدة غير واضحة فى الدراسات المفردة واحيانًا يجيب على أسئلة لا يمكن حلها من الدراسات المفردة، فالدراسة المفردة غير قادرة على الحكم على مقارنة فعالية Effectiveness برنامج تدريبى سواء للأفراد مرتفعى ومنخفضى القدرة العقلية، ومن خلال مقارنة متوسط إحصاء حجم التأثير d لـ Cohen عبر المجموعات المختلفة للدراسات فإن ما وراء التحليل يجيب عن هذا السؤال، وكذلك يمكن دراسة المتغيرات المتفاعلة Moderator variables التى احيانًا يصعب دراستها من خلال الدراسات المفردة عن طريق ما وراء التحليل.

وفى الواقع لا توجد دراسة تامة بدون اخطاء، فكل الدراسات تتضمن اخطاء قياس فى كل قياساتها حيث من الصعب القول انها تامة الصدق البنائى خاصة فى العلوم السلوكية والنفسية حيث تتداخل وتتشابك فيها المتغيرات ببعضها البعض بالتالى فإن المفهوم فى حد ذاته ليس خالصًا نقيًا(Hunter & Schmidt, 2004)، وبالإضافة إلى خطأ القياس فإن الدراسات تعانى من خطأ المعاينة حيث تتضمن الدراسة مجموعة صغيرة من المجتمع وهذا بدوره يؤدى إلى نتائج متعارضة ويبدو هذا واضحًا فى العلوم السلوكية والطبيعية، وتناول Hedges (1987) باستخدام طرق ما وراء التحليل لدراسة الاختلاف لنتائج الدراسات فى 13 مجال فى الفيزياء الطبيعية و13 مجال بحثى فى علم النفس اظهرت النتائج وجود اختلافات كبيرة عبر دراسات الفيزياء كما هو الحال فى علم النفس، بمعنى أن الاختلاف فى نتائج الدراسات لا يختلف من تخصص إلى آخر بغض النظر عن طبيعة المفاهيم المراد قياسها.

وما وراء التحليل ليس أسلوب لإعطاء رؤية شاملة عن التراث البحثى المتراكم ولكنه يمدنا باتجاهات واضحة عن طبيعة الدراسات التى نحن بحاجة اليها مستقبلًا بمعنى ما نوعية الدراسات الأولية التى نحتاجها للتغلب اوسد فجوة بحثية أو التغلب على بعض المحددات التى يكشف عنها أسلوب ما وراء التحليل. ويبدو تأثير ما وراء التحليل واضح فى المجلات النفسية الامريكية التى تتبع APA خاصة Psychological Bulletin فقد زاد استخدامه فى الدراسات المنشورة فى الفترة الاخيرة ويؤكد معظم

محـررى المجـلات علـى توظيـف هـذا الأسـلوب فـى دراسـاتهم كمـدخل لمراجعـة الدراسـات السـابقة لأن هـذا بـدوره يحسـن جـودة الاسـتنتاجات مـن التـراث البحثـى ممـا يـنعكس إيجابيًـا على بناء نظرية متماسكة وأكثر قدرة على تفسير الظاهرة.

ويـرى (1997) Hunt ان مـا وراء التحليـل أكثـر تـأثيرًا فـى ابحـاث المجـال الطبـى مـن العلـوم النفسـية والاجتماعيـة، ويبـدو أن فعاليـة مـا وراء التحليـل أكثـر فـى العلـوم الطبيـة نظـرًا لطبيعـة الابحـاث التجريبيـة التـى تعتمـد علـى عـدد محـدود مـن الأفـراد قـد يصـل إلـى فـرد واحـد مـن ثـم فـإن مـا وراء التحليـل يسـعى إلـى احـداث تكامـل بـين هـذه الدراسـات للوصول إلى استنتاج عام.

فـاذا كـان باحـث مهـتم بدراسـة اثـر المتغيـر المسـتقل X علـى المتغيـر التـابع Y وبـدأ فـى جمـع التـراث المتـاح فـى هـذا المجـال، فالباحـث يصـنف كـل نتـائج الدراسـات فـى ثلاثـة تصـنيفات، الأول يتضـمن الدراسـات التـى اثبتـت وجـود علاقـة موجبـة ودالـة إحصـائيًا، والثـانى يتضـمن تلـك الدراسـات التـى توصـلت إلـى علاقـة سـالبة ودالـة إحصـائيًا، والثالـث يتضـمن تلـك الدراسـات التـى توصـلت إلـى علاقـة غيـر دالـة إحصـائيًا، وعليـه فالباحـث غيـر قـادر للوصـول إلـى اسـتنتاج مـن هـذه الدراسـات المنفصـلة، ويـرى أن هـذا مـدعاة لإجراء دراسته وبعد إجراء الدراسة ووصل إلى وجود اثر سالب ويفسر نتائج دراسته بان هـذا يتفـق مـع دراسـات ويتعـارض مـع دراسـات اخـرى ، ويصـدر توصـية بـإجراء مزيـد مـن الدراسـات للوصـول إلـى تأكيـد إلـى هـذه النتيجـة ويظـل البحـث فـى هـذا المجـال علـى هـذه الـوتيرة وهكـذا دواليـك وهـذا لا يحـدث تطـور فـى العلـم، وعليـه فـإن الدراسـات المفـردة الأوليـة نادرًا ما تحل التناقضات فى العلوم الاجتماعية والنفسية (Ellis, 2010).

ويـرى (2004) Hunter & Schmidt ان الحقيقـة المؤلمـة ان معظـم التـراث البحثـى اظهـر نتـائج متعارضـة فبعضـها وجـد علاقـات دالـة إحصـائيًا والـبعض غيـر دالـة إحصـائيًا. ووجـدت فـى التـراث النفسـى والاجتمـاعى تقريبًـا 50-50% (Cohen, 1962; Sedlmeier & Gigerenzer, 1989) بمعنـى أن نصـف النتـائج دالـة إحصـائيًا ونصـفها غيـر دالـة إحصـائيًا، وهـذا هـو الحـال الطبيعـى فـى معظـم العلـوم السـلوكية والاجتماعيـة، وعليـه فمـن الصـعب بنـاء نظريـات ومعرفـة متراكمـة متماسـكة، مـن ثـم فـإن

مـا وراء التحليـل يلعـب دوًا كبيـرًا فـى عمليـة البحـث فـى العلـوم النفسـية والاجتماعيـة. المراجعـة المنظمـة الكيفيـة السـردية ومـا وراء التحليـل همـا طـريقتين متمـايزتين لتوليـف الدراسـات السـابقة ويسـتخدمان لتحديـد وتحليـل وتوليـف الادلـة الكميـة مـن الدراسـات السابقة.

ولا يمكـن اسـتخدامه لدراسـات تتضـمن موضـوعات مختلفـة فـى نفـس مـا وراء التحليـل مثـل دراسـات تتنـاول بـرامج لعـلاج الاكتئـاب ودراسـات تتنـاول الفـروق بـين الجنسـين فـى الوحـدة النفسية، فيشـار إلـى هـذا بإشـكالية "**التفـاح والبرتقـال**" فـى مـا وراء التحليـل بمعـن اسـتخدامه لتوليـف دراسـات لا تتعامـل مـع نفـس المفـاهيم أو العلاقـات ولكـن هـذا مـا اتبعـه Smith & Glass (1977) حيـث دمـج نـواتج دراسـات مـن معالجـات مختلفـة مثـل العـلاج النفسـى المعرفـى، العـلاج السـلوكى المعرفـى، العـلاج الجشـتالتى وغيرهـا فـى دراسـة واحـدة للحصـول علـى الفعاليـة العامـة لهـذه البـرامج. وبالمثـل لـيس مـن الأفضـل تضـمين نـواتج دراسـات مـن تصـميمات بحثيـة مختلفـة التـى تتضـمن أسـاليب إحصـائية مختلفـة حتـى لـو تعاملـت مـع نفـس الموضـوع، فعلـى سـبيل المثـال لا تتضـمن دراسـات لتـأثير معالجـات علـى الاكتئـاب باسـتخدام تصـميم المجموعـة الضـابطة والتجريبيـة مـع دراسـات مـن منهجيـة بحثية اخرى تتعامل مع الاكتئاب كدراسات الملاحظة أو العلاقة.

وعليـه فـإن نتـائج مـن تصـميمات بحثيـة مختلفـة بمعالجـات إحصـائية مختلفـة مـن الصـعب دمجهـا فـى نفـس مـا وراء التحليـل ولكـن مـن الأفضـل إجـراء مـا وراء تحليـل منفصـل للنـواتج التجريبية وللنواتج دراسات العلاقة.

ما وراء التحليل فى مقابل التحليل الأولى أو الثانوى

قبـل تنـاول أسـلوب مـا وراء التحليـل لا بـد مـن التفريـق بـين مصـطلحات مـا وراء التحليـل و التحليـل الأولـى Primary analysis والتحليـل الثـانوى Secondary analysis . ويـرى Glass (1976) ان **التحليـل الأولـى** يحـدث عنـدما يقـوم باحـث بجمـع البيانـات مـن العينـة أو الشـركات أو أى وحـدة معاينـة خـلال تطبيـق ادوات القيـاس ثـم تحليلهـا للإجابـة

علـى أسـئلة بحثـه أو دراسـته وهـذا مـا يحـدث فـى معظـم ابحـاث الماجسـتير والـدكتوراه ومـا بعد الدكتورة.

امـا **التحليـل الثـانوى** يشـير إلـى اعـادة تحليـل بيانـات تـم جمعهـا مسـبقًا للإجابـة عـن أسـئلة بحثيـة مختلفـة أو الاجابـة عـن علـى نفـس الأسـئلة بطريقـة منهجيـة اخـرى مثـل اسـتخدام أسـاليب إحصـائية اخـرى لـم تتـاح عنـد إجـراء الدراسـة الأوليـة. ودراسـات التحليـل الثـانوى يمكـن أن يجريهـا نفـس الباحـث أو بـاحثين اخـرين مختلفـين إذا كـان باسـتطاعتهم الحصـول على البيانات الخام من صاحب الدراسة الأولية.

لاحـظ أن التحليـل الأولـى والثـانوى يتطلـب الحصـول علـى البيانـات الخـام الاصـلية للدراسـة. فـى المقابـل فـإن **مـا وراء التحليـل** يتطلـب التحليـل الإحصـائى لنتـائج (ليسـت البيانـات الخـام) أكثـر مـن دراسـة (ليسـت لدراسـة واحـدة). فـى هـذا يتضـح مظهـرين لمـا وراء التحليل:

اولًا: يتضـمن نتـائج الدراسـات كوحـدة التحليـل خاصـة ان النتـائج تكـون فـى صـورة أحجـام **التـأثير**، والحصـول علـى أحجـام التـأثير لا يتطلـب الوصـول للـدرجات الخـام حيـث يمكـن حسـابها مـن البيانـات المطبوعـة فـى الدراسـات الأوليـة أو الثانويـة المنشـورة مثـل المتوسطات والتباينات واحجام العينات ودرجات الحرية.

ثانيًـا: يعتبـر تحليـل لنتـائج دراسـات متعـددة وعليـه فـإن الدراسـة المفـردة هـى وحـدة التحليـل ويتراوح حجم الدراسات من دراستين كحد أدنى إلى مئات الدراسات فأكثر إذا كان متاح.

يهـدف مـا وراء التحليـل إلـى الوصـول إلـى اسـتنتاج مـن عينـة مـن الدراسـات فـى المقابـل فـإن التحلـيلات الأوليـة والثانويـة تهـدف للوصـول إلـى اسـتنتاج مـن عينـة مـن الأفـراد، وبالتـالى يعتبـر مـا وراء التحليـل شـكل مـن أشـكال مراجعـة التـراث البحثـى. وعليـه فـإن مـا وراء التحليـل لـيس دراسـة اوليـة أو ثانويـة لحـل قضـية معينـة أو للإجابـة عـن سـؤال بحثـى محـدد انمـا فقـط هـو أسـلوب إحصـائى يسـعى لتكامـل النتـائج الدراسـات المختلفـة لـنفس الموضـوع وباسـتطاعته ضـبط خطـأ المعاينـة واى اخطـأ اخـرى غيـر مرغوبـة لإعطـاء أساس متكامل وشامل للاستنتاجات.

حقيقة من الصعب القول بوجود دراسة متقنة تمامًا Perfect study فكل الدراسات تعانى من بعض المحددات المنهجية وتتضمن اخطأ القياس، فأى دراسة ليست تامة صدق البناء Construct validity بالإضافة لعوامل أخرى تشوه نواتج الدراسة مثل خطأ المعاينة وعليه لا يمكن الوصول إلى استنتاجات علمية من دراسة مفردة وحيدة أو مجموعة فرعية من الدراسات.

خطوات ما وراء التحليل أو توليف البحوث Research Synthesis

لتوضيح أهمية عملية توليف البحوث التى تتضمن أحد صورها ما وراء التحليل فمن الضرورى اتباع الخطوات العلمية لعملية التوليف أو الدمج وهى تتضمن خمسة مراحل مناظرة لتلك المتبعة فى البحوث الأولية كالتالى (Card, 2012; Cooper, 1998; Cooper & Hedges, 2009; Kline, 2013; Little et al., 1998):

- **صياغة مشكلة البحث** Problem formulation: المرحلة الأولى فى ما وراء التحليل هو تحديد المشكلة أو القضية المراد دراستها، أو ما يرغب الباحث فى الاجابة عنه وتحديد المتغيرات أو الابنية والمجتمع المتضمن فى دراسة ما وراء التحليل المزمع انجازها. فاذا كان هدف الباحث الوصول إلى استنتاج عن السببية Causality فالباحث يعتمد على الدراسات التجريبية وربما الدراسات الطولية ودراسات تحليل المسار، ويمكن أن يستخدم التراث تعريفات إجرائية لنفس المفهوم أو ربما يستخدم الباحث مصطلحات متشابهة لوصف مفاهيم مختلفة ولذلك فمن الضرورى تحديد مصطلحات وعينة الدراسة (طلاب مرحلة تعليمية أو عينة صعوبات تعلم وغيرها) جيدًا قبل البدء فى بحث ومراجعة الادبيات البحثية وكل ذلك فى تحديد الدراسات المتجانسة المناسبة لإجراء دراسات ما وراء التحليل.
- **البحث عن عينة الدراسات المرتبطة بموضوع البحث (الدراسات الأولية)** literature Searching أو Sampling: وفى هذا الشأن يبدأ الباحث فى الحصول عن الدراسات الاكثر ارتباطًا ومن المناسب الحصول على عينة من الدراسات ممثلة تمثيلًا جيدًا للتراث البحثى للظاهرة موضع الدراسة من مصادر متعددة سواء مقالات، كتب، تقارير لهيئات، وغيرها. ولو ان الدراسات غير ممثلة للتراث

البحثى عندئذ تكون الاستنتاجات متحيزة لواقع الظاهرة واحد التهديدات جراء ذلك هو **تحيز النشر** Publication bias ويحدث هذا للدراسات التى تفشل فى الحصول على تأثيرات دالة إحصائيًا فهذه الدراسات أقل احتمالية للنشر فى الدوريات العلمية ومن الصعب الحصول عليها. وللتغلب على هذه الاشكالية يحاول الباحث الحصول على الدراسات غير المنشورة مثل دراسات الماجستير والدكتوراه ودراسات المؤتمرات وتقيم ماذا كانت توجد فروق فى التأثيرات بين الدراسات المنشورة والدراسات غير المنشورة. ويوجد تهديد آخر هو ان المراجعين للتراث يعتمدون على الدراسات التى تعتمد على اللغة التى يعرفونها ويستبعدون التراث المنشور فى بلاد أخرى وثقافات أخرى، وعلى الرغم أنه من غير العملى ان يعرف الباحث كل اللغات فى مجتمعات اخرى ولكن على الباحث ان يكون مدركًا لمحددات ذلك على استنتاجات الدراسة ويجب أن يتميز بالشفافية Transparency أن يذكر الأساليب التى استخدامها فى البحث عن الادبيات البحثية. ويرى Kline (2013) لا يوجد حد أدنى مطلق للدراسات المتضمنة فى دراسة ما وراء التحليل الا أنه يبدو لى ان 20 دراسة مختلفة على الأقل هو العدد المعقول لإجراء ما وراء التحليل والفشل فى الحصول على عدد كافى من الدراسات يمثل فجوة معرفية.

- **تقويم جودة الدراسات السابقة لتحديد مدى تضمينها Evaluation Quality of Studies:** وهذه المرحلة تتضمن قراءة التراث والوصول إلى استنتاجات عن طبيعة العلاقات أو التأثيرات بين المتغيرات. وتوضيح معايير الاستبعاد Exclusion أو التضمين للدراسات التى تتضمن طبيعة الابنية أو العينات غير المناسبة أو المجتمع المناسب لطبيعة الدراسة التى بصدد إجرائها أو التصميم البحثى أو طبيعة النواتج، والمعلومات المتحصل عليها فى المرحلة تفيد فى اعادة صياغة المشكلة واستبعاد الدراسات التى لم يتوفر فيها معايير التضمينInclusion.
- **التكويد وحساب حجم التأثير** Coding and Compute Effect Size: يوجد العديد من المداخل لكيفية الوصول إلى استنتاجات عن الدراسة تتراوح من المرجعة الكيفية إلى ما وراء التحليل. وباستخدام ما وراء التحليل لا بد من التكويد Coding البناء لخصائص أو عوامل الدراسة وحساب أحجام التأثير وكذلك تحديد وزن كل

دراسة فى ضوء حجم العينة وتباين الخطأ المعيارى وكذلك يمكن تصحيح أحجام التأثير من المحددات المنهجية مثل ضيق المدى أو الثنائية التصنيفية الاصطناعية التى تسبب تقلص لأحجام التأثير.

- **تحليل البيانات Data analysis**: بعد حساب حجم التأثير تبدأ عملية التوليف الكمى لأحجام التأثير المختلفة باستخدام أى من نماذج ما وراء التحليل سواء المثبتة أو العشوائية أو المختلطة وإجراء **التحليلات المتفاعلة** للتعرف على مدى اسهام خصائص الدراسات فى تفسير الاختلاف بين أحجام التأثير.
- **وصف وتفسير وعرض النتائج Describe, Interpret ,and Report the results**: يتم عرض النتائج والرسومات والجداول ثم مناقشتها وتفسيرها وعرض التوصيات والتطبيقات العملية أو الدراسات المقترحة. وفى هذه المرحلة لا بد أن يتميز الباحث بالشفافية ويبتعد عن التحيز الشخصى خاصة فى حالة استخدام المراجعة الكيفية.

يرى Little et al. (2008) ان ما وراء التحليل يتضمن مهمة أو الكثير من المهام الآتية:

- تقدير متوسط التأثير العام.
- تقدير التباينات أو الاختلافات Heterogeneity فى أحجام التأثير عبر الدراسات.
- تقدير التأثيرات لمجموعات فرعية من المجتمع موضع اهتمام الدراسة.
- اكتشاف التأثيرات المحتملة للمتفاعلات (خصائص الدراسات) على اختلاف أحجام التأثير.
- تقدير أثر تحيز النشر وتحيز العينات الصغيرة على أداء ما وراء التحليل.

وتوجد بعض المفاهيم الخاطئة الخاصة بما وراء التحليل منها:

السائد	الحقيقة
نشأ فى ابحاث المجال الطبى	نشأ فى العلوم الاجتماعية والسلوكية واصبح أكثر استخدامًا فى العلوم الطبية.
مناسب للدراسات التجريبية	مناسب لمعظم طرائق البحث مثل البحث الارتباطى والسببى المقارن.
يجب أن يتضمن دراسات من تصميمات تجريبية تتضمن مجموعة ضابطة.	يمكن أن يتضمن دراسات من تصميمات عينات غير عشوائية(منهج شبه تجريبى) مثل تصميم قياس قبلى- بعدى.
يتطلب دراسات كثيرة.	يمكن استخدامه لدراستين أو ثلاثة فقط.
يتطلب دراسات ذات عينات كبيرة.	يمكن استخدامه لدراسات ذات أحجام عينات صغيرة وتستخدم اختبارات لتصحيح من تحيز العينة الصغيرة.
يتغلب على مشاكل الجودة(الصدق) للدراسات الاصلية	ما وراء التحليل لا يحسن من جودة الدراسات الاصلية.

مميزات ما وراء التحليل

يمتلك ما وراء التحليل مميزات ومصادر قوة عديدة كالآتى (Borenstein et al., 2009; leandro, 2005; Littlel et al., 2008; Lisepy & Wilson, 2002)

- **الموضوعية:** ما وراء التحليل يتغلب على محددات التلخيص الكيفى التقليدى للدراسات السابقة ويمدنا بطريقة فعالة لتلخيص نتائج كم هائل من الدراسات والكشف عن علاقات لا يمكن ادراكها من الدراسات المفردة ويمدنا بشفافية غير متاحة فى المراجعة السردية التقليدية Narrative review لنتائج الدراسات وهذا بدوره يقلل التحيزات والاجتهادات الشخصية. ويعتبر ما وراء التحليل طريقة منظمة وموضوعية لمعالجة معلومات من عدد كبير من نواتج الدراسات.
- **التغلب على محددات اختبارات الدلالة الإحصائية:** أحد طرائق مراجعة التراث طريقة العد Vote-counting المعتمدة على عد النتائج الدالة والنتائج غير الدالة وهى تعانى من محدد أساسى وهو الاعتماد على اختبارات الدلالة الإحصائية وهى تتأثر بحجم العينة فدراسات العينات الكبيرة تعطى دلالة للعلاقات أو التأثيرات الضعيفة،

فالدلالة الإحصائية لا تخبرنا عن قوة أو حجم العلقة أو التأثير ولذلك فإن طريقة العد ربما تخبرنا عن دلالة إحصائية موجبة أو سالبة ولكنها حقيقة ليس لها دلالة عملية. كما ان دراسات العينات الصغيرة غالبًا تعانى من قوة إحصائية منخفضة للكشف عن التأثيرات الدالة إحصائيًا بمعنى تعطى نتائج غير دالة إحصائيًا بينما هى فى طبيعتها دالة عمليًا أو اكلنيكيًا. وباستخدام ما وراء التحليل يمدنا بتقديرات عامة أكثر دقة نتيجة دمج الدراسات ذات أحجام العينات الصغيرة والكبيرة فى دراسة واحدة وهذا من شانه زيادة القوة الإحصائية للكشف عن التأثيرات الدالة. وبذلك يتغلب أسلوب ما وراء التحليل على محددات اختبارات الدلالة الإحصائية.

وعلى ذلك فإن ما وراء التحليل تميز عن اختبارات الدلالة الإحصائية بانة يتعامل مع مؤشر حجم التأثير من كل دراسة وليس الدلالة الإحصائية أو قيمة P لتحديد ما إذا كان حجم التأثير متسق عبر الدراسات وكذلك تضمين كل أحجام التأثير لكل الدراسات فى توليف إحصائى واحد وهذا فى غاية الأهمية لحساب حجم تأثير تلخيصى، على ذلك فإن ما وراء التحليل يذهب أبعد من لدلالة الإحصائية ويسمح ما وراء التحليل بتقدير مدى الاختلاف بين التأثيرات ويميز بين الاختلاف الحقيقى والاختلاف الزائف.

- **وراء التحليل إجراء مفيد فى بناء معرفة علمية صادقة** قائم على عملية منظمة لتلخيص النواتج البحثية تشمل معاير لتحديد مجتمع نتائج الدراسات وكذلك استراتيجيات بحثية منظمة لتحديد واسترجاع الدراسات وتكويد خصائص الدراسة والنواتج وتحليل البيانات للوصول إلى استنتاجات صادقة.
  - **قدرة ما وراء التحليل للوصول إلى علاقات وتأثيرات لا يمكن الوصول الها من الطرق المختلفة لمراجعة التراث.** فالطرق الكيفية أو السردية لنتائج الدراسات لا تخبرنا عن الفروق بين الدراسات. ومن خلال التكويد المنظم (أحد مراحل ما وراء التحليل) لخصائص الدراسات يمكن المقارنة بين الدراسات من حيث طبيعة المعالجة وتصميم البحث وإجراءات القياس وبتقدير حجم التأثير لكل دراسة ودمجها

معًا عبر الدراسات المختلفة يعطى تقديرات لحجم التأثير أكثر قوة إحصائيًا من الدراسات المفردة.

- **الحصول على تقديرات معالم أفضل Parameter Estimates**: نتيجة لخطأ المعاينة يمكن لأى دراسة ان تنتج تقدير غير دقيق للتأثير خاصة فى المجتمعات الكبيرة، فالمقاييس متعددة المفردات أكثر ثباتًا من المقياس ذو المفردة الواحدة، وبالمثل فتوليف النتائج من دراسات عديدة يعطى تقديرات أكثر دقة ومناعة لمعالم المجتمع من الدراسة المفردة.
- **إجراء التحليل المتفاعل Moderator analysis** : المراجعة السردية لا يمكن أن تكشف عن مدى تأثير نوعية المشاركون أو المعالجة أو تصميم الدراسة على نواتج الدراسة، ويستطيع ما وراء التحليل الكشف عن تأثير هذه العوامل أو الخصائص المسئولة عن الاختلاف بين النتائج أو أحجام التأثير مثل خصائص العينة(عمرها، جنسها)، نوعية المعالجة، مكان إجراء الدراسة، ونوعية القياسات(ملاحظة، تقرير ذاتى)، جودة تصميم البحث، مصدر تمويل الدراسة(خاص، حكومى)، تاريخ النشر، الخلفية المهنية للباحث (Kline, 2013). ولو ان التحليل المتفاعل كشف عن عدم تأثير هذه العوامل على التأثير العام عندئذ فإن نواتج ما وراء التحليل تشير إلى ان النتائج لديها مناعة أو صمود وقابلة لتعميمها إلى مواقف اخرى، لو ان الدراسات غير متسقة بين المؤسسات أو المواقف فلا بد من فحص وكشف مصادر الاختلاف. وتحليل التفاعل يمكن تحديد البرامج أو المعالجات أكثر نجاحًا. ولذلك فإن **تحليل المتغير المتفاعل** يهدف إلى تقدير ما إذا خصائص الدراسة تفسر الاختلاف بين أحجام التأثير.
- **تقليل الخطأ والتحيز Minimizing error and bias** : يمكن لما وراء التحليل تقليل خطأ المعاينة والتحيز من خلال توليف عدد كبير من الدراسات لموضوع ما.

ويرى Littlel et al. (2008) أن من أهم مميزات أسلوب ما وراء التحليل هو التغلب على محددات المراجعة الكيفية السردية للدراسات، وكذلك يمدنا بطريقة فعالة لتلخيص نتائج كم هائل من الدراسات ويكشف عن علاقات غير مدركة فى الدراسات المفردة

وتمدنا بشفافية فى مراجعة الدراسات كذلك يعطى معلومات أكثر دقة عن النتائج المختلطة عن طريق التلخيص الكمى للتأثيرات.

وكذلك يرى leandro (2005) ان ما وراء التحليل يسهم فى التالى:

- زيادة القوة الإحصائية للتحليلات أو المقارنات.
- يعطى تقدير أكثر دقة لتأثيرات المعالجات.
- دمج أو توليف أو تقيم نتائج الدراسات المتعارضة.
- الاجابة عن أسئلة بحثية جديدة.
- تحليل مجموعات فرعية من لعينات مختارة من دراسات مختلفة.
- تحليل الاتجاهات العامة فى دراسة موضوع ما.
- لفت الانتباه إلى المجالات البحثية التى بحاجة إلى دراسة وبحث.

الانتقادات الموجهة لأسلوب ما وراء التحليل

منذ نشأة أسلوب ما وراء التحليل على يد Smith & Glass (1977) واخرون يلاقى العديد من الانتقادات أو المحددات معظمها مرتبط بالتطبيقات غير المنطقية وسوء استخدام الأسلوب (Littell et al., 2008). ومن أهم الانتقادات الموجهه لهذا الأسلوب (Borenstien et al., 2009; Card, 2012; Littell et al., 2008; Lisepy & Wilson, 2002; Rosenthal & DiMatteo, 2001)، كالتالى:

- يحتاج إلى **جهد وخبرة عالية لفهمه وتنفيذه** من جمع الدراسات وتكويدها تقدير أحجام التأثير المناسبة وتحليلها إحصائيًا وتفسيرها وذلك مقارنة بالمراجعة الكيفية السردية، وعليه يتطلب جهد إحصائى عالى لتنفيذه والالمام بفنياته ولكن كأى أسلوب يبدو صعب عندما تقرأه أول مرة، ولكن معظم تطبيقات ما وراء التحليل لا تتطلب تحليلات معقدة، وأيضًا كتابة تقرير دراسة ما وراء التحليل تحتاج إلى مهارة عالية مقارنة بتقرير الدراسات الأولية

- بـدون اتبـاع القواعـد المتبعـة لأسـلوب مـا وراء التحليـل مـن جمـع الدراسـات المرتبطـة وتكويدها وتحليلها تصبح عبارة عن **تدريب إحصائى** لا معنى له.
- **التحديـد غيـر المناسـب للمشـكلة:** عنـد إجـراء مـا وراء التحليـل لا بـد تحديـد نوعيـة الدراسـات أو المعالجـات المتضـمنة فـى التحليـل وغيـاب المعنـى المنطقـى مـن تضـمين الدراسـات المختلفـة التـى تتضـمن معالجـات وعينـات وتصـميمات بحثيـة ونـواتج مختلفـة يعطـى نتـائج مـن الصـعب تفسـيرها ويعتبـر هـذا مـن الانتقـادات الكلاسـيكية لـما وراء التحليل.
- يشـبه الـبعض خلـط دراسـات مختلفـة التصـميمات والعينـات كـالخلط بـين **التفـاح والبرتقـال** حيـث تتبـاين الدراسـات فـى الجـودة المنهجيـة والمقـاييس والعينـات وغيرهـا مـن المظـاهر المنهجيـة، ولكـن محلـل أو مسـتخدم الأسـلوب يعتبـر ان هـذا الخلطـة مناسـبة عنـدما يكـون الـبعض مهـتم بعمـل سـلطة فواكـه، فمـن غيـر المناسـب ان يكـون الباحـث مهـتم بالتفـاح فقـط أو يريـد ان يميـز التفـاح عـن البرتقـال ولا بـد مـن وجـود منطقيـة نظريـة قويـة لإجـراء مـا وراء التحليـل لإعطـاء نتـائج ذات معنـى ويمكـن التغلـب علـى ذلـك بدراسة العوامل المسئولة عن تفسير التباين بين الدراسات المختلفة.
- **تقـدير غيـر مناسـب لجـودة الدراسـة:** مـا وراء التحليـل لدراسـات منخفضـة الجـودة لهـا ثبـات مـنخفض وتصـميم بحثـى ضـعيف ومشـاكل إجرائيـة ينـتج عنهـا نتـائج غيـر صـادقة. ويطلـق علـى هـذه المشـكلة " زبالـة دخلـت، زبالـة خرجـت". وللتغلـب علـى هـذه المشـكلة لا بـد مـن وضـع معـايير لانتقـاء الدراسـات فـى ضـوء تصـميم البحـث الجيـد ونوعيـة المعالجـات التـى تتنـاول المتغيـرات المسـتقلة لتضـمين الدراسـات فـى التحليـل ويسـعى مـا وراء التحليـل مـا إذا كـان التنـوع المنهجـى يـؤثر علـى النتـائج مـن خـلال تحليل المتفاعلات.
- **انخفـاض الصـدق الخـارجى** External validity : كـأى نـوع مـن الدراسـات فـإن نتـائج مـا وراء التحليـل يمكـن تعميمهـا فقـط إلـى المجتمعـات والمعالجـات والمواقـف التـى تشـبه المواقـف التـى تمـت فـى التجربـة، فالمعالجـات التجريبيـة التـى تـتم فـى ضـوابط منهجيـة صـارمة وعينـات متجانسـة والتركيـز علـى متغيـر مسـتقل وحيـد فـى الغالـب وتـدريب عـالى الجـودة للمعـالج أو المتـدرب، فالنتـائج المتحصـل عليهـا مـن هـذه النوعيـة

من الدراسات من الصعب تعميمها فى المواقف الطبيعة سواء فى المدرسة أو الجامعة أو المستشفى أو غيرها. وعليه فنتائج ما وراء التحليل لهذه الدراسات تعانى من نفس المحدد وهو انخفاض القرة التعميمية فى المواقف الطبيعية.

- **اعتماده على أحجام التأثير** Effect sizes: لدمج الدراسات كميًا فمن الضرورى الحصول على قياسات تكون قابلة للمقارنة عبر الدراسات، ولذلك يحصل الباحث على مجموعة من أحجام التأثير لتحديد قوة أو حجم العلاقات بين المتغيرات. وفى ما وراء التحليل يحسب حجم التأثير لكل دراسة على حدة ثم تُوزن بحجم العينة ومتوسطها لتمدنا بالأثر العام Overall effect . وفى هذا الشأن ينتقد البعض ما وراء التحليل باعتماده على حجم التأثير فقط وعلى الرغم أنه يمدنا بمؤشر محكم لمتوسط التأثير عبر الدراسات ولكنه ربما يكون غير مفهوم ومتداول لمعظم الأفراد. ولكن هذا المحدد يمكن التغلب عليه بوضع قياسات لهذا المؤشر لكى تكون مفهومة لدى الاكلينيكيين وصانعى القرار. ولكن تتوافر بعض الفنيات للتحقق من مسلمة التجانس Homogeneity لتحديد ما إذا كانت أحجام التأثير من الدراسات المختلفة متجانسة من عدمه.
- **بعض فنياته الجديدة لم تصل إلى درجة مصداقية عالية:** على سبيل المثال الفنيات المستخدمة لتصحيح تحيز النشر مازالت محل جدل ونقاش بين المتخصصين.
- **يتجاهل المعلومات الكيفية البناءة:** على الرغم أن ما وراء التحليل لا يمكن استخدامه لتوليف البيانات الكيفية ولكن المعلومات الكيفية على درجة كبيرة من الأهمية حيث من الممكن اقتراح قضايا ومعلومات ذات مغزى واضاءات اثناء التجريب. ولكن أحد المداخل لما وراء التحليل توليف الادلة الأفضل Best evidence synthesis الذى يحاول ان يدمج بين فنيات المراجعة الكيفية والكمية. فالبحث الكيفى على درجة كبيرة من الأهمية حيث يمدنا برؤى وبمعلومات ثرية تسهم فى فهم الظاهرة مقارنة بالبحث الكمى الإمبريقى. والتقدم العلمى يكون محددًا لو اعتمد فقط على البيانات الكمية المتحصل عليها من طرائق البحث الكمية.
- **المعاينة غير المناسبة وطرق جمع البيانات:** لا يعطى ما وراء التحليل انتباه كافى لمصادر الخطأ المحتمل والتحيز فى معاينة (عينة الدراسات) ومعايير تقيم الدراسات

المدمجة، وعدم الانتباه لثبات وصدق البيانات المستخلصة للتحليل من الدراسات. ويعتمد تحليلات ما وراء التحليل فى معظمها على الدراسات المنشورة على الرغم من عدم وجود فرق بين نتائج التحليل بين الدراسات المنشورة والدراسات غير المنشورة (De Smidt & Gorey, 1997).

- **تحيز تقرير النتائج Outcome reporting bias** : دائمًا يتم نشر النتائج الدالة إحصائيًا مقارنة بالدراسات ذات النتائج غير الدالة إحصائيًا(الصفرية) وهذا يعرف بتحيز تقرير النتائج عليه فإن الباحث يعتمد على حساب حجم التأثير للنتائج الدالة إحصائيًا التى تكون متاحة للباحث وهذا بدوره يؤدى إلى استنتاجات مربكة ومتحيزة لأحجام التأثير، وهذا مرتبط **بتحيز النشر** حيث الدراسات الدالة نتائجها إحصائيًا أكثر مقبولية للنشر فى الدوريات العلمية مقارنة بنظيرتها غير الدالة إحصائيًا بالتالى لا نستطيع الوصول إلى استنتاجات صادقة لأنه يعتمد على الدراسات المنشورة الدالة إحصائيًا.
- **التحيز فى انتقاء الدراسات التى تعتمد على لغة الباحث** بينما الدراسات فى البلدان الأخرى بلغات اخرى تكون غير متاحة للباحث وكذلك مدى الإتاحية Availability حيث بعض الدراسات تحتاج إلى تكلفة عالية للحصول عليها وهذا غير متاح للباحث. فالباحث العربى لديه إتاحية ومرغوبية للحصول على الدراسات المكتوبة باللغة الانجليزية لأنه أكثر الفة بها مقارنة بالدراسات الفرنسية والالمانية والصينية وغيرها، وهذا بدوره يقود إلى استنتاجات خاطئة عن النظرية أو الظاهرة موضع الدراسة وإذا اعتمد التحليل على مجموعة متحيزة من الدراسات الدالة إحصائيًا ويتجاهل النتائج غير الدالة إحصائيًا بالتالى يعطى نتائج متحيزة، ولكن المؤيدون يرون ان الأسلوب يتضمن العديد من الفنيات لتشخيص هذه المشكلة وهى تحيز الدراسات ومحاولة معالجتها. وهذه المشكلة ليست قاصرة فقط على ما وراء التحليل انما أيضًا قائمة فى المراجعات الكيفية للدراسات السابقة.
- **مصداقية استخلاص البيانات** Data extraction: كثيرًا من الباحثين متأكدون من قدراتهم العالية لاستخلاص وتقويم المعلومات من الدراسات ولكن بدون وجود أكثر من مقدر أو مكود وحساب ثبات أو اتفاق المقدرين المستقلين Inter-rater

agreement فإن القارئ لا يستطيع التأكد من مدى مصداقية البيانات المستخلصة من الدراسات. ووجد Gotzsche, Hrogjartsson, Maric, & Trendal (2007) إلى وجود اخطأ فى استخلاص بيانات 10 دراسات (37%) من 27 دراسة لما وراء التحليل، وهذه بدوره لا يعطى نتائج دقيقة بدون تجنب اخطأ إدخال البيانات .

- يلخص المعالجات أو تأثيرات الدراسات بمعنى يحاول ان يقلل النتائج فى قيمة واحدة ويتجاهل التنوع بين الدراسات حيث إذا حدث خلل أو عيب فى استخلاص هذه القيمة فله عواقب خطيرة على صناعة القرار. ولكن المؤيدون يرون أن الهدف من الأسلوب هو توليف أحجام التأثير وليس اعطاء ملخص لها ولو كانت هذه التأثيرات متجانسة فإن التحليل يظهر تأثير لديه مناعة عبر الدراسات المتضمنة فى التحليل ولكن لو ان الاختلافات جوهرية بين أحجام التأثير فإن الاهتمام يتحول من ملخص لحجم التأثير إلى تباين التأثير.
- تتعامل ما وراء التحليل مع التأثيرات الرئيسية للمتغيرات ولا تستطيع أن تتعامل مع التفاعلات بين المتغيرات فى تأثيرها على المتغير التابع انما يمكن أن يتم ذلك من خلال التحليلات المتفاعلة.

وفيما يلى استعراض لخطوات ما وراء التحليل:

الخطوة الأولى: تحديد المشكلة أو موضوع البحث Topic or problem specification: مثل أى طريق أو منهجية من طرق البحث فإن الخطوة الأولى لإجراء ما وراء التحليل تبدأ بتحديد الموضوع المراد دراسته أو السؤال المطلوب الاجابة عليه وهذا ضرورى لتحديد اهداف دراسة ما وراء التحليل وهذا التحديد للموضوع يعتبر كمرشد لاختيار الدراسات السابقة وتكويد المعلومات من هذه الدراسات وتحليل نتائج البيانات. وتحديد المشكلة يتطلب تعريف وتحديد المتغيرات وطبيعة العلاقة بينهما وكيفية اختبارها إمبريقيًا، والمتغيرات فى العلوم الاجتماعية يجب تحديدها بطريقتين الأولى التعريف المفاهيمى أو النظرى وهى تحدد صفات المتغير وخصائصه، فمثلًا التعريف النظرى للذكاء هو القدرة على التعلم، الاستدلال، الفهم، وأى نشاط عقلى آخر،

فـالتعريف النظـرى شـامل وعـريض بينمـا التعريـف الإجرائـى أكثـر تحديـدًا ويكـون قابـل للقياس ووصف لخصائص للأحداث الملاحظة لتحديد مدى وجود الظاهرة أو المفهوم.

ويـرى Lipsey & Wilson (2001) يجـب أن تكـون صـياغة المشـكلة واضـحة ومباشـرة واكثـر تفصـيلًا مثـل: هـل البـرامج القائمـة علـى الوسـائط المتعـددة فعالـة فـى تحسـين التحصـيل؟، هـل البـرامج القائمـة علـى اسـتراتيجيات الـتعلم المـنظم ذاتيًـا لهـا تـأثيرات فعالـة فـى تنميـة تقـدير الـذات أو أى نـواتج اخـرى مثـل التحصـيل أو الاتجـاه أو التفكير الناقد وغيرها من النواتج الانفعالية والمعرفية والسلوكية؟

لاحـظ أن صـياغة المشـكلة وتحديـدها يسـاعد فـى تحديـد طبيعـة المتغيـرات المسـتقلة والتابعـة، فـالمتغير المسـتقل موضـع الاهتمـام والدراسـة هـو خصـائص البرنـامج بينمـا المتغيـرات التابعـة تقـدير الـذات، التحصـيل، الاتجـاه، التفكيـر الناقـد. وهـذا التحديـد يعتبـر بمثابـة المرشـد لتحديـد الكلمـات المفتاحيـة وبـدورها المسـاعدة فـى البحـث عـن التـراث البحثى والدراسات السابقة.

ودائمًا تصاغ عناوين ما وراء التحليل لدراسة تأثيرات المعالجات كالآتى:

أثر (المعالجة أو البرنامج) لـ (المشكلة أو المتغير الناتج) فى (مجتمع)

مثال: العلاج المعرفى السلوكى لتخفيف اضطرابات القلق للأطفال والبالغين

أثر (البرنامج **أو التدخل)** مقارنة **بـ (الشرط أو المجموعة الضابطة) لـ (المشكلة أو الناتج) فى(مجتمع)**

مثـال: أثـر العـلاج السـلوكى مقارنـة بـالعلاج المعرفـى فـى تخفيـف اعـراض الاكتئـاب لـدى نزلاء السجون.

تحديد أهداف وأسئلة أو موضوع البحث المراد الاجابة فى ما وراء التحليل

وتتضمن عدة خطوات أهمها:

الموضوع والمجال Topic and scope

يرى Littell et al. (2008) أن الخطوة الأولى لإجراء توليف البحوث هى تحديد الأسئلة الجوهرية التى سوف تكون بمثابة مرشد للباحث لجمع التراث البحثى ومن الضرورى تحديد نطاق ومجال والمسلمات الأساسية لإجراء دراسات ما وراء التحليل.

والأسئلة والفروض الجيدة تأتى من مصادر عديدة منها الممارسة البحثية، الخبرة، والمراجعة المستمرة لتراث ظاهرة ما. والسؤال البحثى يجب أن يكون قابلًا للبحث، محدد بوضوح، ممتع، جديد، اخلاقى، ومناسب. فمثلًا هل البرنامج فعال؟ فهذا ليس سؤال محدد لأن التجربة أو البرنامج ليس فعالًا لكل الفئات بمعنى له شروط وظروف معينة، ولكن الصيغة الأفضل للأسئلة عن تأثير برنامج أو تدخل ما هو: هل المعالجة X1 أكثر فعالية من المعالجة X2 على المتغير التابع Y لمجتمع ما؟ وهذا من شأنه تحديد ما هى المعلومات التى يحاول الباحث الحصول عليها ومنها تحديد المجتمع والمعالجات والمواقف والنواتج، كذلك تحديد المكونات المستبعدة من الدراسات السابقة.

فمثلًا أحد المراجعات يهدف إلى تقدير أثر البرامج المدرسية على مدى واسع من النواتج مثل كفاءة الذات، العلاقات بين الأقران، التحصيل الأكاديمى، وهكذا. ومراجعة ما وراء تحليل اخرى تركز على أثر هذه البرامج المدرسية لدى التلاميذ منخفضى الدخل، ومراجعة ثالثة تركز على أثر البرامج المدرسية على مدى محدود من النواتج أو على ناتج واحد مثل التحصيل الأكاديمى، بالتالى هذه الأسئلة تعكس أطر مختلفة للبحث فى التراث.

وكلما كان الباحث أكثر وضوحًا فى تحديد الموضوع المراد دراسته فى ما وراء التحليل كلما كانت الدراسات المجمعة من التراث البحثى اكثرًا تجانسًا ولكن عدم التحديد الدقيق للموضوع المراد دراسته يؤدى إلى اتساع دائرة البحث فى التراث ويحصل على دراسات غير متجانسة، وعليه تظهر مشكلة التفاح والبرتقال فى ما وراء التحليل ولكن البعض يرى أن الهدف العام هو البحث فى سلة الفواكه ومحاولة ايجاد العلاقات بينهم.

ومن الأفضل أن تكون القضية المراد بحثها محددة المفاهيم والأسباب والعواقب وأن تكون مستندة لنظرية أو نموذج لأن هذا يساعد الباحث فى صياغة العلاقات المفترضة بين المعالجات والنواتج على المدى القصير وعلى المدى الطويل.

معايير الصلاحية أو تضمين واستبعاد الدراسات Eligibility or Inclusion and Exclusion Criteria:

بعد تحديد الموضوع المراد دراسته باستخدام ما وراء التحليل واتضحت الفكرة عن طبيعة النواتج البحثية المناسبة للموضوع، يبدأ الباحث فى تحديد الدراسات البحثية المراد دمجها فى ما وراء التحليل ومن أهم مظاهر هذا التحديد هو مجتمع الدراسات البحثية المناسبة. ومن الضرورى وضع معايير Criteria لتحديد الدراسات المتضمنة أو المستبعدة من التحليل، وهذه المعايير فى ضوء النظرية أو النموذج القائم عليه موضوع البحث التى سبق تحديدها فى الخطوة السابقة وهى مجموعة من المحكات عن ملامح أو خصائص الدراسات المراد تضمينها أو استبعادها، ويجب تحديدها قبل البدء بالبحث فى التراث البحثى للموضوع.

وتحديد هذه المعايير يفيد فى الآتى (Card, 2012; Littell et al., 2008):

- تمدنا بمرشد موضوعى وصادق للدراسات المراد تضمينها أو استبعادها وبدورها تقلل التحيز الشخصى Subjectivity الذى يحدث لو لم تتوفر هذه المعايير، وكذلك الابتعاد عن تحيز الانتقاء للدراسات كلما أمكن. وإذا لم تحدد هذه المعايير مسبقًا فإن تضمين الدراسات يعتمد على التفضيل الشخصى والإتاحية للدراسات وعوامل اخرى وهذا بدوره يبتعد عن الموضوعية.

- تحدد طبيعة المجتمع المراد دراسته، فاذا كان الموضوع منصب على مرحلة التعليم الثانوى فيتم استبعاد كل الدراسات دون ذلك.

- الالتزام بالشفافية Transparency التى تعتبر من أهم خصائص دراسات ما وراء التحليل مما يساعد باحثين آخرين على إعادة الدراسة مرة أخرى Replication، فافترض أنك حصلت على 100 دراسة لإجراء ما وراء التحليل ولكن بعد تطبيق معايير التضمين والاستبعاد قررت استبقاء 60 دراسة ولو استخدم باحث آخر هذه المعاير سوف يصل إلى نفس الـ 60 دراسة.

ويحدد Lipsey & Wilson (2001) معايير صلاحية الدراسات أو معاير استبعاد أو تضمين الدراسات في ما وراء التحليل بالآتى: الملامح المميزة لأهلية الدراسة، المستجيبين في الدراسة، المتغيرات الأساسية، تصميمات البحث، المدى اللغوى والثقافى، زمن الدراسة، نوع النشر. ويحددها Littell et al. (2008) بالاتى: التصميمات البحثية، المجتمعات، التدخلات أو البرامج، المقارنات، والنواتج المقاسة المراد تضمينها أو استبعادها. بينما يحددها Card (2012) بالاتى: تعريفات المفاهيم موضع اهتمام الدراسة، خصائص العينة، تصميم الدراسة، وقت انجاز الدراسة، نوع النشر، ومعلومات عن حجم التأثير.

وفيما يلى عرض لأهم هذه المعايير:

- **الملامح المميزة أو التدخلات** Distinguishing Features or Interventions: ما هى الملامح المميزة للموضوع المراد تحليله؟ فلو كان سؤال ما وراء التحليل: ما فعالية برنامج أو تدخل ما؟ فإن الملمح الأساسى طبيعة التدخل أو البرنامج وعليه لا بد من تحديد خصائص البرنامج أو المعالجة المستخدمة وتعريفها والاسس النظرية التى تقوم عليها والفنيات والانشطة المستخدمة ومؤهلات القائمين على تطبيقه ومدته، وبعض دراسات ما وراء تستبعد البرامج التى تقل جلساتها عن 8 أو 12 (Littell et al., 2008). ولو تعامل ما وراء التحليل مع مقارنة مجموعات (مثل الفروق بين الجنسين) فإن المعيار تحديد نوعية المجموعات، لو أن موضوع ما وراء التحليل العلاقة بين بنائين (X, Y) فإن المعيار تحديد طبيعة الابنية وتعريفها إجرائيًا، وعليه لا بد من تحديد التعريفات الإجرائية Operational definition للمفاهيم المتضمنة فى التحليل. ويرى بعض الباحثين والخبراء أنه يمكن إجراء ما وراء التحليل بدون التعمق فى هذا المعيار ولكن له مخاطر لأن التراث البحثى خاصة فى العلوم النفسية والاجتماعية يتضمن مفاهيم لها مسميات مختلفة ولكن لها نفس المعنى أو العكس مثل مفهوم الابتكار والابداع أو التأجيل والتسويف والتلكؤ الأكاديمى مسميات مختلفة لنفس المعنى.

- **خصائص العينة أو المشاركين** Sample characteristics or participants: تعتبر خصائص المشاركين أو المستجيبين المشتق منهم البيانات فى الدراسة من أهم معايير تضمين أو استبعاد الدراسة من التحليل، وفى الدراسات الأولية المشاركين هم الأفراد. فاذا كان موضوع ما وراء منصب على مجتمع الاعاقة السمعية فلا بد أن تكون الدراسات تتضمن عينات الاعاقة السمعية، وكذلك إذا كانت موضوع ما وراء التحليل فعالية برنامج لمجتمع المرحلة الجامعية فلا بد أن تتضمن الدراسات عينة طلاب المرحلة الجامعية، ويمكن تضمين بعض الخصائص الديموجرافية مثل الجنس أو الموقع أو الحالة الاقتصادية الاجتماعية. فمثلًا دراسة ما وراء التحليل للذكور فقط أم للإناث فقط. ويمكن أن تكون دراسة ما وراء التحليل قاصرة على بيئة معينة مثل البيئة المصرية أو السعودية أو غيرها وعليه لا بد أن تقتصر الباحث على الدراسات فى البيئة المصرية أو السعودية أو أى بيئة اخرى.
- **المتغيرات المفتاحية أو النواتج** Key Variables or Outcomes : ماهى المتغيرات الأساسية المتضمنة فى الموضوع أو المشكلة المراد دراستها؟ ففى الدراسات التجريبية تكون النواتج (المتغيرات التابعة) محور الاهتمام، وفى دراسات مقارنة المجموعات تكون المتغيرات المراد إجراء المقارنات عليها محور اهتمام ما وراء التحليل، فى الدراسات الارتباطية يكون المتغيرين X, Y محك أساسى لانتقاء الدراسات. ويوجد اعتبار هام لتضمين قياسات النواتج وهو مصداقيها وثباتها. وايضًا توقيت قياس النواتج فالاعتماد على القياسات البعدية لا يخبرنا عن مدى حدث استمرارية للتحسن ولذلك يهتم بالقياسات البعدية Follow up measures.
- **تصميم البحث** Research Method : ماهى الملامح المنهجية والتصميمات البحثية التى تؤهل الدراسة لأن تكون فى ما وراء التحليل؟ المحك الأساسى الذى يحدد منهج أو تصميم البحث هو شكل النواتج البحثية. ففى دراسات التدخل لا بد من استخدام دراسات تجريبية يتم توزيع أفراد العينة عشوائيًا على المجموعة التجريبية والمجموعة الضابطة وان لم يكن توزيع عشوائى يصبح منهج شبه تجريبى وكذلك يعتمد المحلل على الدراسات التجريبية ذات تصميم المجموعتين وقياس قبلى بعدى ويستبعد أى دراسات اعتمدت على تصميم المجموعة الواحدة ذات القياس

القبلى والبعدى ويمكن أن يستبعد الدراسات التى لم يتم التوزيع فيها عشوائى على المجموعة التجريبية والمجموعة الضابطة أو يمكن الاعتماد عليها لو استخدمت قياسات قبلية للمتغير التابع أو المتغيرات الاكثر ارتباطًا به. ويمكن الاعتماد على الدراسات ذات الجودة المنهجية العالية (توزيع عشوائى للمجموعات) أو الدراسات شبة التجريبية ذات القياسات القبلية للمجموعة الضابطة والتجريبية للمتغيرات الاكثر ارتباطًا لتحديد التماثل أو التكافؤ بين المجموعتين على الرغم انها تمدنا بتأثيرات عالية للبرامج والصدق الداخلى لها منخفض. ويرى Shadish & Myers (2004) بضرورة إجراء تحليل منفصل للدراسات التجريبية ذات المجموعات العشوائية وللدراسات شبه التجريبية (توزيع غير عشوائى) واختبار الفرق بينهما باستخدام تحليل التفاعل فلو اعطت أحجام تأثير متجانسة فيمكن دمج نتائج التصميمين فى ما وراء تحليل واحد. توصل Shadish & Myers (2004) إلى ان الدراسات التى اعتمدت على تصميم المجموعات غير العشوائية (شبه التجريبية) تعطى أحجام تأثير أعلى من نظيرتها(التجريبية) ذات التوزيع العشوائى وهذا تحيز موجب. وكذلك يحدد الباحث معيار التضمين للدراسات التى حددت مبررات لكيفية التغلب على مهددات الصدق الداخلى Internal validity للاستنتاج السببية بمعنى اثبات ان التحسن الحادث يتم ارجاعه إلى البرنامج فقط وليس إلى عوامل دخيلة فى التجربة. تصميم الدراسة وجودة التنفيذ تلعب دورًا كبيرًا فى اختلاف النتائج وتقود إلى استنتاجات خاطئة.

ودائمًا يتم استبعاد دراسات المجموعة الواحدة قياس قبلى وبعدى لأنها معرضة بدرجة شديدة لانخفاض الصدق الداخلى لها بمعنى تكون عرضة للمتغيرات الدخيلة حيث لا يمكن ضبط تأثير هذه المتغيرات وعليه لا يمكن ارجاع التحسن إلى المعالجة بل إلى عوامل اخرى ليست فى الحسبان(Littell et al., 2008)، ولكن منهجيًا يمكن إجراء دراسات ما وراء التحليل لدراسات المجموعة الواحدة ومقارنتها بدراسات ما وراء التحليل لتصميمات المجموعتين لنفس الموضوع.

وعلى ذلك لا بد أن يحدد الباحث نوعية المنهج البحثى المستخدم فى الدراسات المتضمنة فى تحليل ما وراء فهل يعتمد على دراسات تجريبية فقط؟ أم دراسات شبة تجريبية فقط؟ أن كانت شبه تجريبية هل تصميم المجموعة الضابطة والتجريبية وقياس قبلى وبعدى أم تصميم المجموعة الواحدة وقياس قبلى وبعدى؟

ويجب على الباحث أن يتبع الخطوات الآتية:

- ضع سقف أدنى من التصميمات المقبولة التى تناولت الموضوع.
- قدر مدى الاختلاف بين الدراسات المتضمنة فى ضوء تصميم البحث وخصائص التنفيذ التى ربما تزيد من مخاطر التحيز.
- استخدم تحليل التفاعل لتقدير تأثيرات التصميمات المختلفة.

- **المدى اللغوى والثقافى** Cultural and Linguistic Range: لا بد من الاعتماد على الدراسات فى البيئة الثقافية للباحث بقدر الامكان نظرًا لاختلاف المفاهيم من ثقافة إلى اخرى. فمثلًا معنى مفاهيم مثل السرقة أو الاحتيال أو السعادة فى البيئة العربية مختلف عن معناها فى البيئات الأخرى، وهذا بدوره يؤدى إلى دمج دراسات مختلفة من حيث التعريف الإجرائى للمتغيرات. ولكن إذا تأكد الباحث ان المفهوم له نفس المعنى عبر الثقافات المختلفة فلا مانع من التضمين ولكن من الأفضل إجراء مقارنة بين الدراسات عبر البيئات الثقافية المختلفة من خلال تحليل المتفاعل.

ولا بد الاخذ فى الاعتبار معيار لغة النشر وهذا مرتبط بقضية تحيز النشر، فهل يعتمد الباحث على الدراسات باللغة العربية فقط أم على الدراسات العربية والمنشورة باللغة الإنجليزية، هذا يتوقف على الهدف مما راء التحليل هل تحديد فعالية البرنامج فى البيئة المصرية فقط أم فى البيئة العربية أم الوصول إلى بناء أو تطوير نظرية بالتالى على أكبر قدر من الدراسات فى كل اللغات.

- **إطار الوقت** Time Frame: من المهم وضع حدود على المدة الزمنية التى تغطيها عملية البحث فى التراث البحثى فمن غير المعقول ان يتناول الباحث دراسة الموضوع منذ مثلًا 1970 فهذا يمثل صعوبة كبيرة على الباحث من حيث الوقت

والتكلفة. يمكن للباحث على الدراسات فى المدة الزمنية من مثلًا 2000 إلى 2010 أو الدراسات المنشورة بعد عام 2010. ولكن يمكن إجراء ما وراء التحليل التراكمى Cumulative meta-analysis لاختبار ما إذا كان تأثير المعالجة تختلف عبر الزمن وهذا يتطلب تضمين الدراسات التى اجريت خلال كل الفترات الزمنية وايضًا يمكن استخدام مدخل التحليل المتفاعل Moderator analysis لدراسة تأثير الزمن على فعالية معالجة ما عبر فترات زمنية مختلفة.

- **نوع النشر** Publication Type: ماهى طبيعة الدراسات التى يرغب الباحث تضمينها فى تحليل ما وراء؟ فمثلًا يرغب الباحث فى تضمين الدراسات المنشورة فى مجلات علمية محكمة والدراسات غير المنشورة مثل ابحاث الماجستير والدكتوراه وابحاث المؤتمرات وكذلك التقارير الحكومية وغيرها. فالاعتماد على الدراسات المنشورة فقط تعتبر اشكالية كبيرة نتيجة الاحتمالية العالية لوجود تحيز النشر جراء ان الابحاث ذات النواتج الدالة تجد طريقها للنشر عكس الابحاث ذات النواتج غير الدالة التى فى احياناً كثيرة يتم رفضها للنشر. ويرى McLeod & Weiz (2004) ان الدراسات غير المنشورة أكثر جودة من نظيراتها غير المنشورة، وعليه فإن تحيز النشر يمثل اشكالية خطيرة لمصداقية المراجعة للتراث البحثى.

وكلما كان الباحث أكثر مرونة فى تبنى معاير منهجية فإن هذا يؤدى إلى نتائج غير دقيقة ولكن وضع معايير منهجية صارمة يؤدى أيضًا إلى الحصول على عينة صغيرة من الدراسات ويمكن استبعاد دراسات مفيدة من التحليل وهذا من شانه يؤدى إلى الحصول على عينة من الدراسات غير ممثلة تمثيلًا جيدًا لمجتمع الدراسات وبالتالى يحد من القدرة التعميمية للنتائج. واجريت العديد من دراسات ما وراء التحليل لمعرفة مدى وجود فروق بين دراسات ما وراء التحليل ذات المعايير المنهجية الصارمة بنظيرتها ذات المعايير المنهجية المرنة ولم تتضح وجود فروق(Shadish, 1995) .

ويرى Lipsey & Wilson (2001) ان الجودة المنهجية فى تراث البحث الاجتماعى والسلوكى فقيرة، فالتقارير غامضة عن توضيح القضايا والتفاصيل المنهجية والإجرائية، فباحث ما وراء التحليل يعانى من ان الدراسات لا تمده بالمعلومات الكافية عن المعايير

المنهجيـة التـى يتبناهـا. والجـودة المنهجيـة ليسـت شـئ بسـيط انمـا تتضـمن عوامـل عديـدة العينـة، العشـوائية، الصـدق، الثبـات، التصـميم، التكـافؤ، وغيرهـا. فالمعـايير المفضـلة لدراسـة لموضـوع مـا فقـد تختلـف عـن دراسـة موضـوع آخـر ولكـن المعـايير التـى لا خـلاف عليها هى قضايا ثبات البيانات والتصميم التجريبى المستخدم.

وبعـض البـاحثون وضـعوا اطـر وقـوائم لتقـدير الجـودة المنهجيـة ولكـن لا يوجـد اتفـاق بـين البـاحثين حـول مكونـات هـذه الاطـر وهـى تتضـمن الخصـائص المنهجيـة الجيـدة التـى يجـب أن تتسـم بهـا الدراسـات. وتوجـد علاقـة تبادليـة بـين قيـود معـايير التضـمين للدراسـات المتضـمنة فـى مـا وراء التحليـل والملامـح المنهجيـة للدراسـات، فكلمـا زاد تنـوع واتسـاع هـذه المعـايير كلمـا حصـل الباحـث علـى مـدى واسـع لدراسـات ذات ملامـح منهجيـة وإجرائية متنوعة.

عمومًـا معـايير التضـمين والاسـتبعاد المرنـة والصـارمة لهـا مميـزات أهمهـا (Card, 2012):

- تضـمين أكبـر عـدد مـن الدراسـات الممكنـة واسـتبعاد عـدد محـدود بينمـا المعـايير الصـارمة تسـتبعد عـدد كبيـر مـن الدراسـات وتضـمن عـدد قليـل. والاختيـارين يعكسـوا نقطتـى النهايـة علـى متصـل. والمعيـار الاهـم للتضـمين هـو مجتمـع الدراسـات التـى بصـدد الوصـول إلـى اسـتنتاجات بشـأنه، وكـذلك معيـار الجـودة المنهجيـة فهـل الباحـث علـى اسـتعداد لتضـمين دراسـات تعـانى مـن خلـل منهجـى مـا وهـذا مـن شـأنه يزيـد مـن انتقـاد أسـلوب مـا وراء التحليـل بمعنـى " زبالـة دخلـت التحليـل، زبالـة خرجـت مـن التحليل" ولكن المعايير الصارمة تجعل استنتاجات النتائج لمجتمع محدود.
- المعـايير المرنـة تجعـل الدراسـات أكثـر تنوعًـا فـى ملامحهـا المنهجيـة بينمـا المعـايير الصـارمة ينـتج عنهـا عـدد محـدود مـن الدراسـات ومتشـابهة فـى ملامحهـا وهـذا يـؤدى إلى قوة إحصائية غير مناسبة لتقويم متوسط حجم التأثير.
- **معلومـات حجـم التـأثير** Effect Size Information: فـى النهايـة مـن المعـايير الهامـة تـوافر البيانـات الكافيـة لحسـاب حجـم التـأثير، وفـى معظـم الدراسـات تتـوافر هـذه

المعلومات مثل المتوسطات والانحرافات المعيارية وحجم العينة. وإذا لم تتوافر هذه المعلومات فيمكن التواصل مع المؤلف لطلب هذه المعلومات.

ولا بد من اعداد بروتكول أو إطار أو خطة مكتوبة يتضمن ماهية المعايير التى يعتمد عليها باحث ما وراء التحليل حتى يسمح لمراجع آخر يكود النواتج البحثية المتطلبة لإنجاز دراسة ما وراء التحليل.

ويتضمن البروتكول الاتى:

- غلاف يتضمن عنوان دراسة ما وراء التحليل، تفاصيل الاقتباس، والتواصل مع المؤلفين.
- مقدمة (الخلفية واهداف الدراسة).
- المنهجية وتتضمن معايير التضمين والاستبعاد، طرق البحث (قواعد البيانات، الكلمات المفتاحية، استراتيجيات البحث)، استخلاص البيانات (شيت التكويد، تقدير ثبات المقدرين)، تحليل البيانات والتوليف (كيفية تنفيذ ما وراء التحليل، تقدير تحيز النشر، التحليل المتفاعل، تحليل الحساسية)، التفسير، الجداول والاشكال، والمراجع.

تحديد نواتج البحث المتطلبة لإجراء دراسة ما وراء التحليل

يرى Lipsey & Wilson (2001) أن نواتج الدراسات الكمية تأخذ أشكالًا عديدة مثل:

- الفروق بين متوسطات المجموعات (الفروق بين الذكور والإناث فى الدافعية).
- العلاقات بين المتغيرات (العلاقة بين القلق والتحصيل).
- نسب لملاحظات أو ظواهر معينة فى المجتمع (نسبة مدمنى الفيس بوك أو نسبة انتشار الجريمة فى المجتمع).

تتعدد النواتج المراد تجميعها من الدراسات أهمها:

- **حجم التأثير:** تجمع نتائج الدراسات لاستخدامها فى ما وراء التحليل فى صورة حجم التأثير ولمقارنة النتائج بصورة لها معنى لا بد أن يستخدم نفس حجم التأثير لتكويد كل النتائج حتى تكون قابلة لاستخدامها فى ما وراء التحليل، وعليه فلا بد من تحديد شكل النواتج البحثية المناسبة لما وراء التحليل وتعرض فى ضوء إحصاء حجم التأثير العام وإذا وجد صور مختلفة لأحجام التأثير فلا بد من تصنيف النتائج وإجراء ما وراء التحليل منفصل لكل حجم تأثير وكذلك يمكن إجراء تحويلات حجم التأثير من صورة إلى اخرى مثل تحويل مؤشر d إلى مؤشر معامل الارتباط r.

على ذلك فالمكون العام لنوعى التحليل سواء دمج أو مقارنة هو أحجام التأثيرات حيث تعكس ماهية أسلوب ما وراء التحليل. وتتنوع مؤشرات حجم التأثير الاكثر استخدامًا فى ما وراء التحليل منها مؤشرات فروق المتوسطات المعيارية، معامل الارتباط، مؤشر نسبة أو احتمال الترجيح Odds-ratio وغيرها، وعلى الرغم من المؤشرات المختلفة لأحجام التأثير الا انها فى معظمها تعكس العلاقة بين متغيرين، والعلاقة المتحصل عليها من الدراسات الارتباطية تعكس درجة التلازم أو التغاير بين متغيرين ويمكن أن تكون مقياس تقريبى للعلاقات السببية Causal relations ولكن العلاقات من الدراسات التجريبية فهى مقياس حقيقى للسببية على الرغم من مهددات الصدق الداخلى فى التصميمات التجريبية.

واى من انواع العلاقات هى مقياس لحجم التأثير المستخدم فى ما وراء التحليل، ويوجد العديد من مؤشرات حجم التأثير متضمنة المتوسطات، النسب، ثبات المقاييس، درجات التغير الطولى (الدراسات الطولية) ولكنها أقل استخدامًا فى ما وراء التحليل. واحجام التأثير المشتقة من دراسات ارتباطية، طولية، شبة تجريبية، تجريبية تعكس مدى من الأسئلة البحثية يمكن الاجابة عليها باستخدام ما وراء التحليل (Cooper, 2009; Lipsey & Wilson, 2001).

- **إحصائيات النزعة المركزية** Central tendency statistics: نواتج البحث تصف خصائص المتغيرات المقاسة لعينة وحيدة من المستجيبين وتوزيع قيم المتغيرات تلخص فى ضوء مؤشرات النزعة المركزية مثل المتوسط، الوسيط، المنوال، والنسب.

وإذا تـوافرت هـذه المؤشـرات لنـواتج الدراسـات فـيمكن إجـراء مـا وراء التحليـل للدراسـات. لاحـظ أن هـذه المؤشـرات تكـون لدراسـات متنوعـة تناولـت نفـس المتغيـر وغالبًـا تسـتخدم نفـس المقيـاس كمتوسـط الـذكاء كمـا قـيس فـى اختبـار سـتانفورد- بينيـه. ويمكـن حسـاب مؤشـرات النزعـة المركزيـة لدراسـات اسـتخدمت مقـاييس مختلفـة لـنفس المفهـوم شـرط تحديد التعريف الإجرائى للمتغير عبر كل الدراسات.

- **مقارنـات المجموعـات Group contrasts:** هـذا النـوع مـن نـواتج البحـوث يتضـمن متغيـر واحـد فـأكثر تـم قياسـه لمجمـوعتين فـأكثر مـن الأفـراد ثـم تـتم المقارنـة بـين المجموعـات ويـتم عـرض المتوسـطات أو النسـب لكـل مجموعـة. وفـى هـذا الإطـار يوجـد شكلين لمقارنة المجموعات وهى موضع اهتمام ما وراء التحليل:

- **البحـوث التجريبيـة أو المحـاولات الاكلينيكيـة Experimental or clinical trials**: حيـث تـتم المقارنـة بـين مجموعـة تعرضـت لمعالجـة مـا تجريبيـة (مجموعـة تجريبيـة) بمجموعـة لـم تتلقـى معالجـة أو معالجـة تقليديـة (مجموعـة ضـابطة) حيـث تكون المقارنة على قيم المتغير الناتج (التابع) وتفسر على انها تأثير المعالجة.

- **بحـوث الفـروق بـين المجموعـات**Group differences research: تـتم المقارنـة فـى ضـوء خصـائص المجموعـة مثـل المقارنـة بـين الـذكور والإنـاث فـى التحصـيل أو أى متغيـرات اخـرى تابعـة أو المقارنـة بـين طـلاب ذى صـعوبات الـتعلم ومنخفضـى التحصـيل بـأقرانهم العـاديين فـى القـدرة القرائيـة، وهـذه النوعيـة مـن الدراسـات تنبـع مـن دراسات **المنـهج السببى المقارن** Causal comparative method.

وهـذه النوعيـة مـن الدراسـات سـواء التجريبـى أو مقارنـة المجموعـات هـى شـائعة فـى التـراث البحثـى لمـا وراء التحليـل وخاصـة البحـوث التجريبيـة التـى تهـدف إلـى تقـدير فعاليـة المعالجـة أو التـدخل (Lipsey & Wilson, 2001; Rosenthal, 1994). وعنـدما تـتم المقارنـة بـين ثـلاث مجموعـات فـأكثر فيـتم دراسـة كـل مجمـوعتين معًـا علـى حـدة نظـرًا لأن الفنيـات المسـتخدمة للمجموعـات الثلاثـة معًـا غيـر متاحـة (Lipsey & Wilson, 2001).

- **العلاقة بين المتغيرات** Association between variables: هذا النوع من نواتج البحوث يعكس التغاير عبر الأفراد لمتغيرين لتحديد درجة العلاقة بينهما، كالعلاقة بين ما وراء التعلم وما وراء الذاكرة، ويعتبر معامل الارتباط مؤشر لحجم التأثير أو العلاقة بين الحالة الاقتصادية الاجتماعية للأسرة وتحصيل الرياضيات فى المرحلة الثانوية، وتعتبر مؤشرات كأى تربيع أو معامل فأى أو لمدا lambda أو نسبة Odds-ratio مؤشرات حجم التأثير، وفى هذه الفئة من البحث يوجد تصنيفين عريضين من اهتمامات ما وراء التحليل هما:

- **بحوث القياس** Measurement research: وهذه النوعية من بحوث العلاقة تكون منصبة على تقدير خصائص ادوات القياس مثل حساب معامل الارتباط بين درجات الاختبار واعادته Test-retest كمؤشر لمعامل الثبات، أو استخدام معامل الارتباط لتقدير الصدق التنبئى Predictive validity بين درجات اختبار القدرات والتحصيل اللاحق فى الجامعة ويستخدم للتحقق من صدق اختبار الاستعدادات أو القدرات.

- **بحوث الفروق الفردية** Individual differences research: وهذه النوعية من بحوث الارتباط هى الاكثر استخدامًا وهى تهتم بدراسة التغاير بين سمات مختارة للأفراد كالعلاقة بين ادمان الانترنت والوحدة النفسية، أو العلاقة بين التنشئة الاسرية والعنف المدرسى، أو العلاقة بين وقت المذاكرة درجات الطلاب فى الرياضيات وغيرها.

وتعتبر دراسات ما وراء التحليل لأبحاث العلاقات أو الارتباطات شائعة الاستخدام فى التراث وذلك لتوافر العديد من مؤشرات أحجام التأثير المناسبة (Hunter & Schmidt, 2004; Rosenthal, 1994)، وتعتبر قيمة معامل الارتباط مؤشرًا لحجم التأثير.

الخطوة الثالثة: تحديد التراث البحثى Identifying Research Literature : بعد التحديد الجيد لمعايير التضمين والاستبعاد للدراسات المناسبة لما وراء التحليل التى حددت طبيعة مجتمع الدراسات، فإن الخطوة التالية هى البحث

عن الدراسات الإمبريقية التى تتوافق مع اطار المعاينة ومن المحتمل ان يتضمن مجتمع الدراسات عينة ممثلة بدرجة كافية، بالإضافة إلى ان المحلل الإحصائى لما وراء غالبًا يرغب فى تقسيم الدراسات إلى تصنيفات مختلفة مع توافر عدد كافى من الدراسات فى كل تصنيف ليسمح بإجراء التحليل والمقارنة لذلك فإن الباحث يحاول تحديد واسترجاع كل دراسة فى المجتمع وليس عينة من هذا مجتمع الدراسات.

ما وراء التحليل هى دراسة ملاحظة مسحية Observational Study قائمة على نتائج الدراسات المتاحة للعلاقة بين المتغيرين فى التراث البحثى، وعليه فالباحث مهتم فى المقام الأول بعض قائمة من بعض أشكال نواتج الدراسات التى تمثل إحصائيات أحجام التأثير التى يعتمد عليها محلل ما وراء التحليل.

إجراء عملية البحث

تتضمن عملية البحث عن الدراسات عدة خطوات كما حددها Reed & Baxter (2009) كالتالى:

- **تحديد استراتيجية البحث:** وضع خطة لعملية البحث فى التراث البحثى تتضمن السؤال البحثى أو الموضوع المراد البحث فيه وتوصيفات هذا الموضوع من حيث عينته ومنهجه ومقاييسه وغيرها، وماهى قواعد البيانات المراد البحث فيها، وما الفترة الزمنية المراد البحث فيها.
- **خطوات عملية البحث:** وتتضمن الخطوات الآتية:

- **تحديد الموضوع:** متضمنًا المتغيرات، العينات، والمنهجية.
- **تحديد المصطلحات** أو **الكلمات المفتاحية:** من الأفضل تضمين مترادفات المفهوم أو المصطلح للحصول على أكبر عدد من الدراسات.
- **بروفيل البحث:** ما طبيعة الربط المنطقى بين المفاهيم مثل وضع AND أو علامة (+) وتستخدم للحصول فقط على الدراسات التى تتضمن المفهومين فقط، اما OR تستخدم للبحث فى مفهومين بنفس المعنى بهدف توسيع دائرة البحث مثل Program or intervention أو High school or secondary school، أما الاختيار NOT يستخدم لاستبعاد عينة معينة مثل.

- **البحث الرئيسى:** يتضمن تحديد قواعد البيانات الرئيسية الالكترونية والحصول على المصادر أو الدراسات الرئيسية.
- **تقويم الدراسات:** وهذه خطوة تتضمن تقويم الدراسات من حيث ارتباطها بالموضوع وهذه عملية صعبة وتحتاج جهد ووقت.

فمثلًا عنوان دراسة ما وراء التحليل: أثر برنامج مستند على عادات العقل لتنمية التفكير المنظومى فإن الكلمات المفتاحية: عادات العقل والتفكير المنظومى واحيانًا كلمة برنامج.

وإذا كنت بصدد البحث فى قواعد البيانات الاجنبية فراعى الكلمات المترادفة مثل مفهوم المراهقة adolescence وقد تستخدم بمعنى Teen وعليه يتم البحث كالآتى (adolescence* or Teen*) وكلمة البرنامج يمكن أن نبحث عنها بالمترادفات الآتية:

(Program* or service* or* treatment* or intervention* or prevent*).

ولذلك يجب استشارة المتخصصين لتحديد مترادفات المصطلح. فالبحث فى قاعدة البيانات لعنوان "برنامج لخفض الاكتئاب لدى المراهقين" فإن الكلمات المفتاحية كالآتى:

(Adolescence* or teen) And depression* and (Program* or service* or treatment* or intervention* or prevent*)

وهذا البحث أكثر تحديدًا لأنه يمدنا بالدراسات التى تتضمن المراهقة والاكتئاب والبرنامج معًا.

او "Adolescence, depression, and intervention"

بالتالى يسترجع الدراسات التى تضمنت هذه الكلمات معًا.

فكلمة * أو or ؟ تستخدم للبحث باستخدام كلمات مختلفة لها نفس المعنى فى التراث البحثى.

وكلمة "و" and تستخدم للبحث للعلاقات بين مفهومين مثل عادات العقل ودافعية الإنجاز، يمدنا بالدراسات التى تتضمن العلاقة بينهما.

معظم برامج البحث تتضمن كلمة لا "Not" أى استبعاد الدراسات التى تضمن كلمة أو مصطلح معين. ولتحديد الدراسات المستبعدة يمكن استخدام كلمة Not. ولمحدودية مجال عملية البحث يمكن وضع العينة موضع اهتمام الدراسة ككلمة مفتاحية.

فيما يلى أمثلة لكلمات مفتاحية:

- **نوع العينة:** المراهقون، المراهقة، الذكور، الإناث، أطفال الأحداث، طلاب المرحلة الثانوية أو المدارس العليا، تلاميذ المرحلة الابتدائية، المعاقون سمعيًا..... إلخ.
- **نوع التدخل Type of intervention أو البرنامج:** برنامج تدريبى، برنامج علاجى، برنامج إرشادى، برنامج تحدى، ..... إلخ
- **نوع البحث Type of Research:** تقويم، تأثيرات، فعالية، تجربة، تجريبى مجموعة ضابطة، ارتباطى،..... إلخ

استراتيجية البحث الخلفى Backward Searches

بعد جمع وتراكم مجموعة كبيرة من الدراسات لتطبيق معاير التضمين الممكنة فى ما وراء التحليل يبدأ الباحث فى عملية التكويد لهذه الدراسات ولكن لا بد أن يقرأ محتوى الدراسات كاملة خاصة المنهجية المستخدمة والنتائج التى تتضمن معظم المعلومات المراد تكويدها وكذلك يمكن الاستشهاد ببعض الدراسات المذكورة فى الدراسات المجمعة والمناسبة للموضوع المراد درسته وايضًا من الضرورى قراءة دراسات ما وراء التحليل السابقة التى اجريت فى نفس الموضوع ان وجدت أو موضوعات قريبة لموضوع بحثك. وعملية البحث فى الدراسات المناسبة التى وردت فى الدراسات المجمعة تسمى البحث الخلفى وهذا المدخل مفيد فى تحديد الدراسات القديمة ولكن يوجد عيب فى هذا المدخل هو احتمال التحيز حيث إن الدراسات التى تعطى نتائج مرغوبة (دالة إحصائيًا) أكثر احتمالية لتضمينها فى الدراسة

المجمعة مقارنة بالدراسات غير المرغوبة (النتائج التى لا تعطى دلالة إحصائية). وعلى الرغم من قضية التحيز إلا أن استراتيجية البحث الخلفى تمدنا بمصدر جيد للبحث خاصة انها احيانًا تمدنا بدراسات تفشل مصادر البحث الأخرى امدادنا بها.

استراتيجية البحث الامامى Forward Searches

بينما البحث الخلفى يحاول الوصول للدراسات المذكورة فى الدراسات التى جمعتها فإن البحث الامامى محاولة ايجاد الدراسات التى استشهدت بالدراسات التى جمعتها ويمكن إجراء ذلك من خلال قواعد بيانات متخصصة مثل مؤشر استشهاد العلوم الاجتماعية Social Science Citation Index أو قاعدة بيانات علم النفس PsycINFO فهى تمتلك هذه الخاصية. ولإجراء البحث الامامى ادخل معلومات الدراسة التى تعتبر مناسبة لما وراء التحليل والتى تمتلكها فى قاعة البيانات وسوف يبين لك الدراسات الأخرى التى استشهدت بها لأن هذه الدراسات يمكن أن تساعد الباحث فى الحصول على دراسات جديدة ويمكن إجراء ذلك فى أى قاعدة بيانات شرط إدخال الكلمات المفتاحية التى استخدمتها الدراسة. ويمكن إجراء البحث الامامى لكل الدراسات المتضمنة فى التحليل بعد تطبيق معايير التضمين. ويرى Card (2012) أن هذه الاستراتيجية البحثية غالبًا تكون محدودة فى دقتها.

التواصل مع الباحثين والخبراء فى المجال Communication with Researchers and Experts

المدخل الاخير للحصول على الدراسات هو استشارة الخبراء أو الباحثين فى مجال دراسة الموضوع المراد إجراء ما وراء التحليل فيه، ويمكن استشارة الخبير فى مدى مناسبة معايير التضمين أو الاستبعاد وقائمة الدراسات التى حددتها للتحليل وسؤاله عما كنا بحاجة لتضمين دراسات إضافية، ويفضل سؤال الخبراء الذين لهم اسهامات ورؤى عن الموضوع وذلك لإثراء دراسة ما وراء التحليل، ويمكن التواصل مع الخبراء من خلال الإيميل أو البريد المدون فى الدراسات المنشورة لهم. وللمزيد عن اليات البحث فى التراث البحثى والحصول على الدراسات السابقة يفضل الرجوع إلى Cooper (1998).

مصداقية البحث: هل إجراءات البحث كافية؟

بغض النظر عن طرق البحث المختلفة فى التراث التى اعتمد عليها الباحث، فالسؤال الذى من الضرورى التأكد منه هل إجراءات البحث فى التراث مناسبة وكافية؟ يمكن التأكد من مناسبة مدى كفاية البحث فى التراث من خلال الطرق الآتية Card (2012):

الأولى: هل عينة الدراسات التى حصلنا عليها ممثلة تمثيلًا جيدًا لمجتمع الدراسات أم متحيزة؟ ويمكن التأكد من ذلك من خلال أحد إجراءات ما وراء التحليل للتأكد من إشكالية تحيز النشر.

الثانى: هل عينة الدراسات المتحصل عليها كافية لإعطاء قوة إحصائية مناسبة لتقيم الفروض التى بصدد التحقق منها أو هل كافية لإعطاء مدى ضيق من فترات الثقة لأحجام التأثير؟

الثالث: هل وجد الباحث كل الدراسات المتاحة فى المجال أم فقدت دراسات كان يجب تضمينها فى ما وراء التحليل؟

التساؤلان الأول والثانى يؤثران على جودة الاستنتاجات الإمبريقية لدراسة ما وراء التحليل. اما التساؤل الثالث أقل أهمية فيما يخص الاستنتاج مما وراء التحليل. وللإجابة على هذه التساؤلات هو بذل المزيد من الجهد للحصول على أكبر قدر من الدراسات الممثلة لمجتمع الدراسات كلما أمكن.

ويقدم Card (2012) عدة اقتراحات للتأكد من مصداقية عملية البحث:

- إجراء البحث فى التراث البحثى باستخدام عدة طرائق للبحث مثل قواعد البيانات المختلفة والمجلدات البحثية والتواصل مع الباحثين والمؤتمرات غيرها.
- عمل قائمة من الدراسات المتحصل عليها من استراتيجيات البحث المختلفة وعرضها على بعض الباحثين المتخصصين فى موضوع الدراسة لفحصها لمعرفة مدى مناسبتها فى ضوء معايير التضمين والاستبعاد، ولو أقروا بمدى

كفايتها ابدأ فى إجراء ما وراء التحليل ولو أشاروا إلى غياب بعض الدراسات فيجب اعادة استراتيجيات البحث مرة اخرى.

- خذ قائمة الدراسات التى استبقيت عليها وابدأ استراتيجيات البحث الخلفى والأمامى ومن الأفضل البدء بالبحث الخلفى لأنه يستغرق وقتًا أقل، ويمكن الحصول على بعض الدراسات الاضافية، وإذا لم تحصل على المزيد من الدراسات فابدأ بإجراء استراتيجية البحث الأمامى فى الدراسات التى أبقيت بعد تحقيقها معايير التضمين والاستبعاد على ان تبدًا بالدراسات القديمة اولًا إلى ان تصل إلى نقطة عندها لا تمدنا هذه الاستراتيجية بدراسات جديدة عندئذ اوقف إجراء هذه الاستراتيجية.

بعد ذلك ابدأ فى ترميز الدراسات ولكن يمكن إجراء استراتيجية البحث الخلفى اثناء القراءة المتعمقة فى الدراسات وتساعد فى الحصول على المزيد من الدراسات الاضافية وإذا شعرت أنك لا تحصل على مزيد من الدراسات بنهاية عملية التكويد فإنه يعطى درجة من الثقة ان عملية البحث كانت تقريبًا شاملة وممثلة للتراث. وعلى الرغم من هذه الثقة فى شمولية عملية البحث فى الادبيات البحثية فإن Card (2012) ينصح بخطوتين اضافتين:

- اجلس فى المكتبة أو على الانترنت وحاول ان تطلع على جدول محتويات المجلات الاكثر ارتباطًا بالموضوع فى السنوات الخمسة الاخيرة مثلًا وإذا لم تحصل على مزيدًا من الدراسات فإن هذا يزيد الثقة فى شمولية ومصداقية عملية البحث وإذا وجدت المزيد من الدراسات فإنك بحاجة للمزيد من البحث للحصول على الدراسات الغائبة والمناسبة لطبيعة الموضوع ويمكن استخدام كلمات مفتاحية أو مصطلحات مختلفة غير التى استخدمتها من قبل.
- بعد الطمأنينة بان البحث فى التراث كافى أرسل قائمة الدراسات التى حصلت عليها من التراث مرة اخرى إلى الخبراء فى التخصص ليقيموا مدى كفاية ومناسبة قائمة الدراسات لما وراء التحليل وعليه ذلك فالخبراء لهم دور فعال فى عمل البحث.

بناء قاعة بيانات للدراسات المتطلبة لما وراء التحليل

بعد جمع الدراسات السابقة فإن الخطوة التالية هى كيفية تنظيم عرض الدراسات، فعملية البحث والحصول على الدراسات السابقة تستغرق وقتًا وجهد إلا أن عملية التنظيم تيسر على الباحث إدراك العلاقات بين هذه الدراسات. والتنظيم الجيد يتطلب وضع إطار تنظيمى محدد يتضمن مكونات أساسية مثل **اولًا:** معايير التضمين والاستبعاد التى تستخدمها لتقيم الدراسات المجمعة وفيها تحدد الدراسات المستبعدة فى ضوء كل معيار وعلى ذلك توضح للقارئ مبررات استبعاد دراسة ما مما يوفر الشفافية. **ثانيًا:** يجب أن توجد قائمة محددة لطرائق البحث فى التراث البحثى مثل PsycINFO و ERIC مثلًا حتى تتيح للآخرين اعادة عملية البحث. ويمكن الاستعانة بالدراسات النظرية لأنها مفيدة فى تحديد دراسات اضافية وكذلك فى تفسير نتائج ما وراء التحليل.

وعملية البحث تتكون من جزئين:

- ايجاد المراجع المنشورة التى تلبى معايير التضمين.
- الحصول على نسخ من هذه الدراسات لقراءتها وتكويدها فى ضوء محكات التضمين أو الاستبعاد.

والمبادئ الأساسية للمعاينة تشير إلى ان العينات المتاحة Convenience samples من غير المحتمل ان تكون ممثلة تمثيلا جيدًا لمجتمع الدراسات الكبيرة، فالبحث فى قاعدة البيانات PsycINFO ليس ضمانًا للحصول على قدر كافى من الدراسات المتطلبة لما وراء التحليل، وتزداد هذه المشكلة فى واقع البحث فى البيئة العربية لعدم توافر قواعد البيانات الكافية المتضمنة للتراث البحثى خاصة فى العلوم السلوكية والاجتماعية. وينصح Littell et al. (2008) بالحصول على عينة ممثلة لمجتمع الدراسات وتجنب اشكالية تحيز النشر أو أى مصادر تحيز اخرى مرتبطة.

ولا مانع من الحصول على عينة طبقية من الدراسات للموضوع حيث تفيد فى دراسات مجموعات فرعية، فمثلًا أراد باحث دراسة فعالية برنامج تدخلى لتخفيف العسر القرائى فلا مانع من جمع دراسات لبرامج فى المرحلة الابتدائية والمرحلة الاعدادية.

ولتنظيم جمع الدراسات السابقة المرتبطة يفضل بعض الباحثون وضع خرائط منظمة Systematic maps للتراث البحثى لإعطاء نظرة لعدد وانواع الدراسات التى تناولت دراسة الموضوع، وتساعد هذه الخرائط أو الاطر فى تحديد وتصنيف الدراسات فى ضوء مجتمعها، الموقع، تصميم البحث وغيرها من مظاهر البحث المختلفة.

الخطوة الثالثة: تكويد خصائص الدراسات Coding Study Characteristics: إجراء ما وراء التحليل يتطلب تقدير حجم التأثير ثم إجراء إحصاء استدلالى حول متوسط حجم التأثير وهذا يتطلب إعداد برتوكول تكويد Coding Protocol الذى يحدد المعلومات المطلوب استخلاصها من الدراسات. وعملية تكويد البيانات من الدراسات الأولية غالبًا تكون عملية صعبة ومتعبة وتحتاج وقت كبير وهى من العمليات الحرجة والبناءة فى دراسة ما وراء التحليل (Hunter & Schmidt, 2004)، وعن طريق عملية التكويد يستخلص المكود بيانات ومعلومات وخصائص عن الدراسات يمكن أن تفيد فى إجراء التحليل المتفاعل Moderator analysis بمعنى إجراء تحليلات فى ضوء خصائص الدراسة مثل تحليل للذكور وتحليل للإناث وغيرها من المتغيرات التصنيفية وهذا يفيد فى التنبؤ بالتباين أو الاختلاف فى أحجام التأثير عبر الدراسات المختلفة، ومن خلال تكويد خصائص الدراسات فمن المهم وصف مكونات الدراسة لمعرفة مدى مناسبتها لإجراء ما وراء التحليل وكذلك معرفة الاستنتاج العام الذى توصلت اليه عينة الدراسات وكذلك تحديد مدى وجود فجوات فى التراث البحثى للظاهرة مثل استخدام نوعية من المقاييس دون غيرها أو التركيز على نوع معين من العينات دون غيرها، وعليه فإن تكويد خصائص الدراسات يفيد فى إجراء التحليل المتفاعل أو وصف عينة الدراسات، والمكود يقرأ تقرير الدراسة ثم يملئ بروتكول التكويد بالمعلومات المناسبة من الدراسة.

وتكويد دراسات ما وراء التحليل مثل تكويد الدراسات المسحية فمن المهم اعداد استبيان التفريغ بعناية وتدريب المكودين على ملئ الاستبيان وتقدير صدق وثبات البيانات. وفى هذا الجزء التالى نعرض لكيفية اعداد بروتكول التكويد والإجراءات المستخدمة لاستخدامه لتكويد الدراسات وتدريب المكويدين.

بناء برتوكول التكويد Developing A coding Protocol

فـى ضـوء دراسـات مـا وراء التحليـل لا توجـد صـيغة موحـدة لعمليـة تكويـد الدراسـات بـل تختلـف مـن باحـث إلـى آخـر وحسـب اهـداف الدراسـة ويتضـمن بروتكـول التكويـد خصـائص الدراسـات المـراد إجـراء مـا وراء التحليـل لهـا وهـذا يتوقـف عـل طبيعـة المعلومـات فـى الموضوع أو الظاهرة المراد دراستها.

ويحـدد Lipsey & Wilson (2001) الهـدف مـن التكويـد هـو بنـاء قاعـدة بيانـات للتحليـل الإحصـائى ومـن الأفضـل اسـتخدام مفـردات ذات النهايـة المغلقـة كلمـا امكـن ويحـاول الباحـث ان يحـدد مسـبقًا الاختيـارات المحتملـة لكـل مفـردة فـى اسـتبيان التكويـد حيـث تسـهل علـى المكـود ملـئ مفـردات الاسـتبيان أو القائمـة مثـل مفـردة حجـم العينـة، ولكـن توجـد مواقـف لا يسـتطيع الباحـث تحديـد بـدائل مفـردات الاسـتبيان وعليـه فهـو مضـطر إلـى بنـاء مفـردات ذات نهايـة مفتوحـة وفـى هـذه الحالـة يلجـأ الباحـث إلـى مجموعـة مـن التصـنيفات الفرعيـة أو الابعـاد ولكـن هـذا يسـتغرق وقـت وجهـد مـن الباحـث أو المكـود. كلمـا كـان المكـود مـدربًا كلمـا زادت قدرتـه علـى تجميـع المفـردات فـى اطـر أو تصـنيفات متقاربـة. ويفضـل ان يكـون بروتكـول التكويـد بسـيط وقصـير كلمـا أمكـن ومـنظم فـى اجـزاء واضـحة، ومـن المهـم أيضًـا اعطـاء تعريـف كامـل وواضـح لكـل مفـردة وإرشـادات لكيفيـة تكويـدها لمسـاعدة المكـود للـتحكم فـى الحـالات الغامضـة وغيـر المعتـادة. وفـى الممارسـة العمليـة بنـاء بروتكـول تكويـد يـدوى Coding manual حيـث يتـرجم مـاهى المعلومـات المـراد الحصـول عليهـا مـن كـل مفـردة وكـذلك اعطـاء توصـيف للبـدائل المختلـف للإجابـة واعطـاء إرشـادات للحـالات التـى بهـا غمـوض. وكتـاب التكويـد اليـدوى يكـون بمثابـة الـدليل المرشد للمكود عندما يواجهه صعوبات.

ويحدد Card (2012) اعتبارين يمكن تطبيقهما لمراجعات ما وراء التحليل هما:

- **أسـئلة البحـث موضـع الاهتمـام:** مثـل إجـراء الدراسـات الأوليـة مـن تحديـد المتغيـرات فـإن التخطـيط لمـا وراء التحليـل يتطلـب تحديـد السـؤال البحثـى الـذى علـى أسـاس يـتم تكويـد خصـائص الدراسـة أو توصـيفات الدراسـة Study

descriptors مثل(الطريقة أو المنهج، القياسات، العينات، المتغيرات المستقلة والتابعة، نوعية المعالجة، الإجراءات وغيرها)، فلو ان سؤال البحث منصبًا على متوسط أحجام التأثير عبر الدراسات(دمج الدراسات) فالباحث ليس بحاجة سوى إلى معلومات مثل حجم التأثير وحجم العينة أو أى بيانات تستخدم فى حساب حجم التأثير، وفى هذه الحالة لا نحتاج الا إلى عدد محدود من خصائص الدراسات. وفى المقابل إذا كان هدف الباحث مقارنة الدراسات لتحديد ما إذا كانت دراسات بملامح معينة تعطى أحجام تأثيرات أكبر من أحجام تأثيرات لدراسات بملامح مختلفة وبالتالى فالباحث بحاجة إلى معلومات كثيرة لخصائص الدراسات لتكويدها.

ويوجد نوعية من الأسئلة البحثية تتضمن متغيرات متفاعلة، فبعض محللى ما وراء التحليل يرغبون فى التنبؤ بعدم التجانس فى أحجام التأثير من خصائص الدراسات المكودة، وعليه فإن الباحث بحاجة إلى تكويد عدد كبير من الدراسات خاصة التى يتوافر لها معيارين **الأول:** خصائص الدراسات تكون متماثلة فى معظمها وتجنب غياب البيانات خاصة للمتغير المراد التنبؤ به، **والثانى:** تنوع خصائص الدراسات خاصة عبر على الأقل بعض الدراسات وهذا التنوع ضرورى عبر الدراسات لمعرفة مدى التلازم أو التغاير لأحجام التأثيرات، وعليه يتم إدخال خصائص الدراسات المكودة فى نموذج انحدارى تنبؤى لاكتشاف العلاقات بين هذه الخصائص والتباين فى أحجام التأثيرات.

- **الاعتبار الخاص بمظاهر الدراسة:** خصائص الدراسة تخبرنا بتفاصيل عن مجال البحث، وعمومًا يجب أن يوضع فى الاعتبار أربعة أنواع أساسية لإجراء ما وراء التحليل فى العلوم الاجتماعية: خصائص العينة، عملية القياس، التصميم البحثى، والمصدر. ويمكن عرضها كالآتى (Card, 2012; Valentine, 2009):

المظهر العام	المظاهر الخاصة	امثلة
العينة	إجراءات المعاينة	نوعية المجتمع، العينة الممثلة، العينة العشوائية، البلد.....الخ.
	الملامح الديموجرافية	الجنس، العمر، الحالة الاقتصادية الاجتماعية، الذكاء، الموقع....الخ
خصائص القياس	مصدر البيانات	تقرير ذاتى(الاستبيان- المقاييس)، ملاحظة، مقابلة وغيرها.
	عملية القياس	ملاحظة منظمة أم غير منظمة، تطبيق فردى أم جماعى، مقيد بوقت أو غير مقيد
	قياسات معينة مستخدمة	الصورة كاملة أم مختصرة، مقياس مترجم أم من البيئة أو الثقافة.
خصائص التصميم	نوع التصميم البحثى	المجموعة الواحدة وقياس قبلى وبعدى، تصميم المجموعتين قياس قبلى وبعدى.
	خصائص مميزة للتصميم	نوع المعالجة التى تلقتها المجموعة الضابطة، الفترة الزمنية بين القياسات الطولية.
المصدر	النشر	دراسة منشورة أو غير منشورة، جودة النشر
	عام النشر	السنة، تاريخ جمع البيانات.
	التمويل	ممولة أم غير ممولة، جهة التمويل.
	خصائص الباحثين	الجامعة، الجنس، الجنسية.

جودة الدراسة	الصدق الداخلى	معاينة عشوائية، ضبط المتغيرات الدخيلة، التزييف، الانسحاب.
	الصدق الخارجى	العينات من المجتمع الممثلة له، التجريب يناظر الواقع.
	الصدق البنائى	ثبات القياسات (معايير التضمين)، خصائص القياسات.

وفيما يلى عرض لاهم خصائص الدراسات (Card, 2012; Littell et al. 2008; Lipsey& Wilson, 2002; Wilson, 2009):

1. **كود الدراسة:** يتم وضع كود الدراسة رقميًا أو حرفيًا واحيانًا يتم تحديد مؤلف الدراسة وسنة النشر.

2. **المفاهيم Constructs:** دائمًا تركز الدراسات التجريبية على متغير تابع واحد فهى تسعى إلى تنمية مهارات التواصل أو جودة العلاقات الاجتماعية أو تقدير الذات أو مفهوم الذات وغيرها من المتغيرات الناتجة، اما الدراسات الارتباطية تركز على العلاقة بين متغيرين مثل العلاقة بين سلوك الجريمة والحالة الاقتصادية الاجتماعية أو العلاقة بين ما وراء التعلم والتحصيل، وعليه يقرر محلل ما وراء التحليل ماهى المتغيرات موضع اهتمام دراسته لتكويدها. فمثلًا ما وراء التحليل لدراسات برامج القراءة العلاجية فهى تنصب على متغير ناتج وحيد هو القدرة القرائية على الرغم لو ان بعض الدراسات تبنت مفاهيم اخرى مرادفة مثل الاتجاه نحو القراءة يمكن تضمينها.

3. **خصائص العينة Sample:** تتضمن هذه الخصائص إجراءات المعاينة والخصائص الديموجرافية، **فإجراءات المعاينة** المفضل تكويدها تتضمن: **المجتمع** الذى اشتقت منه سواء طلاب جامعة، مرحلة ثانوية، مدمنين المخدرات، نزلاء المستشفيات النفسية، نزلاء السجون بقضايا الارهاب والتطرف وغيرها، **نوعية العينة**

سواء عشوائية بسيطة أم طبقية، عينة متاحة لموقع عمل الباحث، مقصودة عمدية، **بلد العينة** مثل مصر، السعودية، المغرب...وغيرها. **والخصائص الديموجرافية** مثل الجنس، مكان المعيشة (ريف – بدو – حضر)، الديانة، الحالة الاقتصادية الاجتماعية، العمر، واى توصيف آخر للعينة مثل الذكاء. و**حجم العينة** يعتبر من أهم البيانات فى دراسة ما وراء التحليل حيث يساعد فى حساب الخطأ المعيارى وكذلك تحديد العينات الفرعية كلما أمكن. والباحث ليس بصدد تكويد كل خصائص العينة مالم يكن لها دور فى ما وراء التحليل.

وتقدر أحجام التأثير ما وراء التحليل من علاقات أو تأثيرات محسوبة من قياسات مطبقة على عينة من المستجيبين أو المشاركين وأحيانًا يتم عرض النتائج فى ضوء مجموعات فرعية فمثلا تعرض العلاقة بين التحصيل والتنشئة الأسرية لعينة الريف ولعينة المدينة على حدة إذا كان هذا موضع اهتمام دراسة ما وراء التحليل حيث تقدر أحجام التأثير للذكور وللإناث كلًا على حدة مثلًا لدراسة الفروق بينهما لتحديد مدى استجابة المجموعتين للمعالجة. وفى إعداد بروتكول التكويد من الضرورى تحديد ما إذا كانت أحجام التأثير تقدر للعينة الاجمالية أم تقدر لعينات فرعية.

4. خصائص القياس Measuresement: من الضرورى التميز بين المفاهيم الإجرائية أو القياسات Operational definitions فى الدراسة عن المفاهيم النظرية Theoritical contrusts لأنه توجد مقاييس مختلفة تقيس نفس المفهوم، فمثلًا فى دراسة لتقدير المشاكل السلوكية للأطفال يوجد قياسات مختلفة مثل تقديرات الاخصائى النفسى وتقديرات الوالدين والتقديرات الذاتية. فى العلوم الاجتماعية والنفسية توجد مداخل عديدة لعملية القياس وقياسات عديدة متخصصة للمتغيرات المتضمنة فى ما وراء التحليل، ولذلك من الضرورى تكويد خصائص القياسات للمتغير أو المتغيرين فى ما وراء التحليل. ومن المظاهر الهامة التى من الضرورى تكويدها هو **مصدر البيانات المقاسة** مثل تقرير ذاتى، تقرير من مصدر آخر مثل الوالدين أو المعلم أو ملاحظة الباحث أو غيرها، **ملامح عملية القياس** مثل ملاحظة مفتوحة أم مغلقة، مقابلة فردية أم جماعية، اختبار سرعة محدد بوقت

أم اختبار قوة غير محدد بوقت. ويمكن فى بعض الحالات تثبيت المقياس المستخدم فى الدراسة كإجراء ما راء التحليل لأثر العلاج المعرفى على الاكتئاب كما قيست بقائمة بيك Beck depression inventory ويمكن استبعاد الدراسات التى اعتمدت على مقاييس ذات جودة منخفضة.

وتوجد اشكالية فيما يخص المواقف التى تستخدم أكثر من مقياس لقياس نفس المفهوم فى نفس الدراسة ولكن المدخل الأفضل هو ترميز كل مقياس على حدة ولكن هذا يؤدى إلى عدم تماثل عبر الدراسات، والأفضل الاعتماد على الدراسات التى اعتمدت على المقياس الاكثر استخدامًا، واحيانًا يكود محلل ما وراء التحليل نتائج الدراسات التى اعتمدت على مقاييس متعددة لنفس المفهوم ثم يقدر متوسط حجم التأثير. ومن المهم تضمين الخصاص السيكومترية مثل معاملات الصدق والثبات كمعايير للتضمين فإذا كان معامل الثبات 0.8 فيُكتب فى ملف الإدخال بنفس القيمة ويفضل تحديد نوعية طريقة حساب الثبات ونوعية معامل الثبات هل ألفا كرونباخ أم معامل ارتباط بيرسون ومن الطبيعى ان يعتمد الباحث على دراسات عالية الجودة لا يقل عن 0.70.

5. وقت القياس Time of measurement: يمكن قياس المفهوم على نفس المجموعة أو مجموعات فرعية من الأفراد فى أكثر من وقت. ففى دراسات المعالجات يوجد قياسات قبلية ثم بعدية ثم قياس تتبعى. فى الدراسات الطولية تجمع البيانات من نفس الأفراد على نفس المقياس فى فترات زمنية متتالية، وعليه يقرر الباحث ما إذا كان يقدر ويكود أحجام تأثير فى كل اوقات القياسات (قبلى- بعدى- تتبعى) ولكن فى الدراسات التجريبية يكود الباحث حجم التأثير للنتائج الناتج فى القياس البعدى.

6. خصائص التصميم Design Characteristics: لا بد من تكويد خصائص التصميم مثل طبيعة التصميم المستخدم فى الدراسة سواء تصميم المجموعة الواحدة وقياس قبلى وبعدى أو تصميم مقارنة المجموعات شبة التجريبية أو تصميمات التجريبية العشوائية (الحقيقية)، وايضًا تكويد طبيعة المجموعة الضابطة كونها تلقت

معالجة تقليدية أو لم تتلقى معالجة على الاطلاق. أو الدراسات الارتباطية المحددة بالتصميمات عبر عرضية Cross-sectional designs.

7. **خصائص المصدر Source characteristics**: توجد بعض المعلومات الخاصة بمصدر النشر مثل **مسمى المجلة** أو **طبيعة النشر** دراسة منشورة أو غير منشورة وذلك لتقيم تحيز النشر، و**سنة النشر** فى المجلة أو المؤتمر أو بحث دكتوراه وذلك لدراسة التحليل المتفاعل لمعرفة اتجاهات أحجام التأثير عبر فترات زمنية مختلفة، **ومصدر تمويل** الدراسة لأن هذا من أكبر العوامل التى تؤدى احيانًا إلى تحيز النتائج، و**خصائص الباحث** مثل الجنس أو الثقافة أو المؤسسة التى ينتمى اليها لدراسة تحيز الباحثين بمعنى مدى اختلاف أحجام التأثير للدراسات التى أجراها الذكور عن الدراسات التى اجريت عن طريق الإناث.

تكويد حجم التأثير Effect Size Coding

اى ناتج إحصائى فى أى دراسة منتقاة فى ما وراء التحليل يجب أن تعرض فى ضوء قيمة حجم التأثير المصاحب، وتوجد العديد من المعادلات لحساب حجم التأثير وتتنوع حسب الاختبار الإحصائى المستخدم والتكويد لحجم التأثير تتضمن حساب قيمته لكل ناتج بحثى أو ناتج الاختبار الإحصائى للمتغير موضع الاهتمام باستخدام صيغ ومعادلات إحصائية. وتوجد العديد من البرامج الإحصائية الحاسوبية المستخدمة لتقدير حجم التأثير مثل G-Power, SAS, STATA , SPSS وغيرها من البرامج. وطور Shadish, Robinson, & Lu (1999) برنامج حاسوبى Effect Size Calculator ومتاح فى الموقع(http:/ www.assess.com/) . والمعلومات المراد تكويدها المرتبطة بحجم التأثير تقع فى عدة تصنيفات متداخلة هى:

- المعلومات التى تصف **طبيعة المتغيرات** المراد تقدير حجم التأثير لها مثل المفهوم المراد قياسه وتعريفه الإجرائى والطبيعة الإحصائية للمتغير سواء تصنيفى أو رتب أو متصل.
- **توقيت قياس المتغير** مثل قياسات قبلية وبعدية أو قياسات تتبعية.

- حساب حجم التأثير **للعينة الاجمالية أم لعينات فرعية** مثل حسابه لعينة الإناث فقط ولا يمكن تضمين حجم التأثير للعينة الاجمالية وللعينات الفرعية فى تحليل واحد لانهما معتمدين إحصائيًا ولكن يمكن تضمين أحجام الاثر للذكور وللإناث فى تحليل واحد لانهما مستقلين.
- عملية التكويد تتطلب **بيانات إحصائية** عن حجم الاثر مثل المتوسطات، النسب، أحجام العينة، التباينات، الانحرافات المعيارية، الارتباطات القبلية والبعدية وقيمة P الاحتمالية.... الخ. وكذلك يتضمن بروتكول التكويد معلومات إحصائية متطلبة لعمل تصحيحات أو تعديلات لحجم الاثر من عدم الثبات، التصنيفية، والمدى المقيد للمتغيرات. وتوجد معلومات إحصائية اضافية مثل الدلالة الإحصائية للنتائج، الاختبار الإحصائى المستخدم، طبيعة البيانات الغائبة وفيما يلى ملخص للمعلومات الإحصائية Lipsey & Wilson, (2001):

(المتغير/ المفهوم)، (التوقيت)، (العينات الفرعية)، (حجم العينة)، (المتوسطات أو النسب)، (الانحرافات المعيارية أو التباينات)، (إجراء حساب حجم التأثير)، (ثبات المتغيرات)، (فترات الثقة لحجم التأثير)، (تصنيفية المتغيرات)، (المدى الضيق للمتغيرات)، (الاختبار الإحصائى المستخدم ودلالته الإحصائية)، (رقم الصفحة فى تقرير البحث المقدر فيها حجم التأثير).

تكويد جودة الدراسة Coding Study quality

ينصح البعض ان يتضمن ما وراء التحليل تكويد جودة الدراسة ويمكن إدخال عامل الجودة كمتغير فى التحليل باستخدام التحليل المتفاعل لأحجام الأثر ويرى Card (2012) ان هذا التضمين لعامل الجودة يمثل اشكالية وذلك لأنه ذات طبيعة متعددة الابعاد وليس ذات بعد واحد كما يعتقد البعض لأن الجودة تتضمن مظاهر عديدة خاصة فى ضوء صدق البحث سواء داخلى أو خارجى أو بنائى Construct أو صدق الاستنتاج الإحصائىStatistical Conclusion Validity وداخل كل نوع من

انـواع الصـدق الاربعـة توجـد مظـاهر عديـدة Shadish, Cook, & Campbell (2002):

- **الصـدق الـداخلى**Internal validity : يشـير إلـى أى درجـة ان تصـميم الدراسـة أو المتغيـر المسـتقل يسـمح باسـتنتاج السـببية، بمعنـى أن التغيـر الحـادث فـى المتغيـر التـابع يرجـع ويرجـع فقـط إلـى المعالجـة ولـيس إلـى متغيـرات اخـرى دخيلـة. يعتبـر تصـميم الدراسـة مـن أكثـر الامـور تـأثيرًا فـى الصـدق الـداخلى فالدراسـات التجريبيـة (المعاينـة العشـوائية) أكثر صدقًا داخليًا من الدراسات شبه التجريبية (لا عشوائية).

- **الصـدق الخـارجى** External validity: يشـير إلـى أى درجـة نـواتج الدراسـة قابلـة للتعمـيم لمجتمـع الدراسـة وكـذلك لعينـات مختلفـة لـنفس المجتمـع وشـروط اخـرى غيـر الشـروط التـى اجريـت فيهـا التجربـة ولطـرق قيـاس اخـرى لـنفس المفهـوم موضـع الدراسـة. والدراسـات الاكثـر صـدقًا خارجيًـا هـى التـى اعتمـدت علـى عينـات عشـوائية مـن مجتمـع الدراسـة وهـذا الشـرط يكـاد يكـون نـادرًا فـى الدراسـات النفسـية والاجتماعيـة. مـن الضـرورى فـى دراسـات مـا وراء التحليـل تكويـد مـاذا كانـت الدراسـات المجمعـة حققـت هـذا المعيـار مـن عدمـه، ولكـن إجـراء مـا وراء التحليـل لدراسـات مفـردة محدوديـة الصـدق الخـارجى يـؤدى إلى استنتاجات عالية الصدق الخارجى وهذا من أهم مميزات ما وراء التحليل.

- **صـدق المفهـوم** Construct validity: يشـير إلـى أى درجـة المقـاييس المسـتخدمة فـى الدراسـة تقـيس البنـاء النظـرى الـذى ينـوى الباحـث قياسـه. ولكـن هـذا المصـطلح يشـير فـى هـذا الصـدد إلـى خصـائص القيـاس وهـى الصـدق والثبـات للقياسـات ولـذلك يجـب تكويـد ثبـات المقـاييس للقياسـات المتضـمنة للمتغيـرين فـى التحليـل لإجـراء تصـحيحات أحجـام الأثـر، وكـذلك يمكـن تضـمين قياسـات الصـدق لأن مـن معـايير التضـمين ان يكـون التعريـف الإجرائـى للمفهـوم واضـح ويمكـن تصـحيح الصـدق غيـر الـدقيق مـن خـلال أسـاليب مـا وراء التحليـل، خاصـة حينمـا يقـدر مـن خـلال معامـل الارتبـاط (معامـل الصـدق) بـين المقيـاس المسـتخدم فـى الدراسـة ومقيـاس معيـارى متفـق عليـه لـنفس المفهـوم ويطلق عليه الصدق المحكى.

- **صدق الاستنتاج الإحصائى:** يشير إلى صدق الاستنتاجات الإحصائية فيما يخص العلاقة بين المسبب المفترض والتأثير أو الناتج من التجربة.

وجودة الدراسة بناء تحتى ومتعدد المظاهر من تصميم الدراسة وانتقاء العينة وطبيعة المعالجة الإحصائية ويمكن اخذها فى الاعتبار كمتغير وسيط فى ما وراء التحليل.

ويقترح Card (2012) توصيف المظاهر المختلفة لجودة الدراسة المهمة فى المجال أو التخصص واعتبارها متغيرات متفاعلة لأحجام الاثر بين الدراسات.

ماذا عن الدراسات ضعيفة المنهجية؟

كثير من الخبراء يستبعد الدراسات ضعيفة المنهجية من دراسة ما وراء التحليل ولكن Hunter & Schmidt (2004) يرى أن هذا غير مناسب ولا يفضل استبعادها لأن عدم المناسبة المنهجية Methodological inadequacies لا يمكن اختبارها إمبريقيًا ولا توجد دراسة بدون محددات وقصورات منهجية على الرغم ان هذه المحددات لا تسبب دائمًا تحيز فى النتائج.

الاختلافات المنهجية بين الدراسات تسبب اختلاف النتائج عبر الدراسات نتيجة على الأقل لخطأ المعاينة فى الدراسات ويمكن اختبار عدم المناسبة المنهجية من خلال اعتبارها متغير متفاعل ويمكن دراسة اثرها فى دراسة ما وراء التحليل وإذا لم توجد فروق بين الدراسات عالية المنهجية والدراسات ضعيفة المنهجية على ذلك لا توجد اشكالية عدم المناسبة المنهجية، وعليه يجب تضمين الدراسات الضعيفة المنهجية والجيدة المنهجية ثم دراسة الفروق بينهما ويكون القرار فى النهاية فى ضوء الدراسات عالية الجودة وإذا لم يوجد فروق بينهما(الضعيفة – الجيدة) يكون التحليل على كل الدارسات، وعليه فلا بد من تقيم المناسبة المنهجية إمبريقيًا ولا يتم الحكم عليها فى ضوء التقييمات الذاتية الشخصية حيث توصل Cooper (1998) ان معامل الاتفاق بين المقدرين للحكم على المناسبة المنهجية على أفضل الاحوال 0.50.

على ذلك لا بد من تضمين سؤال هام فى دراسة ما وراء التحليل لدراسة الفروق بين الدراسات المناسبة والدراسات غير المناسبة لتحديد مدى تضمين الدراسات الضعيفة منهجيًا فى دراسة ما وراء التحليل.

توصيفات الدراسة Study Descriptors

كل الخصائص المكودة المتحصل عليها من الدراسة يطلق عليها مواصفات الدراسة وطرح Lipsey & Wilson (2001) إطار شامل لمفردات بروتكول التكويد، ولكن من الأفضل لبناء هذا البروتكول إجراء مسح لعينة من الدراسات المجمعة لتحديد طبيعة المعلومات المراد تكويدها، وهذا الإطار يتضمن مجموعة من التصنيفات لتحديد مواصفات الدراسة كالآتى:

- **المظاهر الأساسية للدراسة** Substantive aspects: مثل طبيعة المشاركين فى العينة، المعالجات المستخدمة، والواقع الثقافى. وهذه المتغيرات تسمح باختبار ما إذا كان يوجد اختلاف فى أحجام الاثر عبر المعالجات المختلفة أو عبر الثقافات المختلفة أو عبر عينات مختلفة.

- **الطريقة والإجراءات** Method and procedures: وهذا النوعية من المعلومات المنهجية تسهم فى دراسة العلاقة بين التنوع المنهجى ونتائج الدراسة حيث يكشف عن التحيز فى النتائج فى ضوء الممارسات المنهجية المختلفة، وفى اغلب الاحوال الاختلاف فى نتائج الدراسات يرجع إلى الفروق المنهجية بين الدراسات، وعليه إذا لم يتم تكويد المتغيرات المنهجية فإن الباحث يصل إلى استنتاجات غير صحيحة.

- **توصيفات المصدر** Source descriptors: ومتغيرات المصدر تتعلق بالإطار العام للدراسة، عملية النشر، خصائص الباحث أو المؤلفين، تاريخ النشر، وغيرها. هذا يفيد مثلًا فى ما إذا يوجد اختلاف لأحجام الاثر عبر التخصصات المختلفة (علم النفس فى مقابل الصحة النفسية).

ويمكن تلخيص توصيفات الدراسة كالآتى:

القضايا الأساسية

مصدر العينة

- الخصائص الديموجرافية (المستوى الاقتصادى الاجتماعى، العمر، الجنس، المستوى التعليمى، مكان المعيشة وغيرها).
- الخصائص الشخصية (القدرات المعرفية، السمات الشخصية).
- الخصائص التشخصية أو ملامح خاصة (نزيل سجن، صعوبة تعلم، مصاب بمرض معين، معاق سمعيًا وغيرها).
- المتغير المستقل (التدخل أو المعالجة أو البرنامج).
- النظرية القائمة عليها الدراسة أو البرنامج.
- خصائص المنظمة أو المؤسسة (عمرها، حجمها، الهيكل الإدارى).
- خصائص القائم بتطبيق البرنامج ومؤهلاته.

الطريقة والإجراءات

- إجراءات المعاينة (عشوائية- عمدية).
- تصميم المسح (البريد، التليفون، المقابلة، دراسة طولية، دراسة عبر عرضية، دراسة مستقبلية، دراسة أرشيفية، دراسة تاريخية).
- التآكل أو الانسحاب للعينة من التجربة.
- مهددات الصدق الداخلى (العوامل الدخيلة فى التصميم).
- القوة الإحصائية.
- جودة القياسات (الصدق والثبات).
- كيفية تحليل البيانات.
- طبيعة المجموعة الضابطة (لم تتلقَ معالجة أو تلقت معالجة بديلة).
- طريقة توزع أفراد العينة على مجموعات التصميم.
- دور المجرب أو الباحث فى المعالجة أو البرنامج.

توصيفات المصدر

شكل النشر (مجلة، كتاب، مجلد، رسالة دكتوراه، تقرير حكومى أو مؤسسى)، تاريخ النشر، بلد النشر واللغة، تمويل الدراسة والجهة المانحة، خصائص الباحث (الجنس، التخصص، والدرجة العلمية).

ومن الضرورى تكويد هذه المتغيرات وتختلف عملية التكويد حسب الهدف من دراسة ما وراء التحليل وطبيعة الدراسات المتضمنة فى التحليل.

تقويم عملية التكويد Evaluatig Coding Process

بعد الانتهاء من تحديد خصائص الدراسة المراد تكويدها فإن الخطوة التالية اقرأ بعناية الدراسات لتكويد المعلومات منها وبطبيعة الحال المعلومات المكودة فى ضوء خصائص الدراسة وحجم الاثر ولا بد من التأكد من عملية الحكم التى يطبقها المكود عندما يتفاعل مع الدراسة لأن هذه العملية تختلف من مكود إلى مكود آخر، ومن دراسة إلى دراسة أخرى، وهذا بالطبع يؤثر فى تحليل بيانات ما وراء التحليل ويوجد مبدآن أو قضيتان على درجة كبيرة من الأهمية مرتبطان ببروتكول التكويد يحددهم (Card, 2012; Lipsey & Wilson, 2002) بالآتى:

- الشفافية والاعادة للتكويد Transparency & Replicability of Coding:

عند كتابة التقرير النهائى لدراسة ما وراء التحليل يجب اعطاء تفاصيل كافية عن عملية التكويد حتى يعرف القارئ كيفية عمل قرارات التكويد(الشفافية)، وكذلك إذا ما طبق باحث آخر نفس قواعد التضمين هل يصل إلى نفس الاستنتاج أو القرارات التكويدية؟ ولتحقيق هذين المبدأين فمن الضرورى اعطاء وصف كامل لكيفية تكويد كل دراسة كميًا، فبعض خصائص الدراسة تكود بطريقة مباشرة ولا تتطلب جهد من المكود ودائمًا لا تختلف من مكود إلى آخر مثل متوسط عمر العينة، ولكن عندما تكون خصائص الدراسة أقل وضوحًا حيث تتطلب احكام تكويد عالية الاستدلال فمن الضرورى كتابة تقرير كامل عن عملية التكويد للتأكد من الشفافية والاعادة وهذا بدوره يسهم فى تقدير ثبات المكودين، فعلى سبيل المثال لتكويد طبيعة القياسات المستخدمة وتصميمات الدراسات فيجب كتابة قيم مختلفة لهذا الكود التصنيفى.

- **ثبات التكويد Reliability of coding:** لتقويم عملية الاعادة لنظام التكويد فمن الضرورى تقدير ثبات المكود الذى بدوره يأخذ بعدين هما:

- **اتساقه عبر ذاته** Intra-coder reliability : أى اتساق نفس المكود من مناسبة إلى اخرى بمعنى أن يكود الباحث مجموعة فرعية من الدراسات مرة تلو الأخرى بوجود فترة زمنية بينهما يمكن أن تكون كحد أدنى خمسة عشر يومًا أو بأقل تقدير اسبوع ويقارن بين النتائج ولكن الباحث يكون على دراية بتكويد الدراسات فى المرة الأولى وعليه يحدث تضخم لقيمة الثبات واحد الاسباب لتقدير ثبات المكود عبر الزمن التحقق من معرفة مدى وجود تغير دراماتيكى اثناء عملية التكويد عبر الزمن. وفى دراسات ما وراء التحليل الصغيرة ربما يوجد مكود واحد لكل الدراسات وإجراء اتساق بين المكودين يصبح أقل أهمية ولكن كلٌّ من النوعين ضروريان لإجراء ما وراء التحليل.

- **الاتساق بين مكودين مختلفين** Inter-coder : حيث يكود باحثان أو مكودان نفس العينة من الدراسات المجمعة بصورة مستقلة ويفضل ان تكون عينة الدراسات من 20 إلى 50 دراسة للحصول على ثبات دقيق نسبيًا ولكن كثير من دراسات ما وراء التحليل لا تتضمن هذا العدد الكبير من الدراسات وفى هذه الحالة يجب استخدام كل الدراسات التى استوفت معايير التضمين والاستبعاد ويجب عدم معرفة بان مكود آخر يقوم بعملية التكويد لأن المعرفة تزيد من عملية الحذر واليقظة فى التكويد ويعطى تقدير عاليOverestimate للثبات الحقيقى.

وباستخدام مدخلى تقدير ثبات المكودين فمن المفيد تقدير الثبات لكل خصائص دراسة تم تكويدها على حدة وايضًا تقدير ثبات كل متغير أو خاصية عبر كل الدراسات.

وتوجد العديد من المؤشرات لتقدير ثبات لمكودين من أهمها:

- **نسبة الاتفاق(AR)** Agreement rate: يقيس نسبة الدراسات لدى المكودين أو نفس المكود عبر مناسبتين قدروا نفس الكود التصنيفى. ويقدر كالآتى:

$$AR = \frac{\text{عدد الاتفاقات}}{\text{عدد الدراسات}}$$

- عدد الاتفاقات: عدد الدراسات التى اخذت نفس الكود للخاصية.

- عدد الدراسات: عدد الدراسات التى كودها المكودين على خاصية الدراسة.

فمثلا فى دراسة ما وراء التحليل تضمنت 20 دراسة وتضمن نظام التكويد خاصية نوعية العينة سواء عشوائية أو غير عشوائية وعرضت على مكودان فاتفق تقدير المحكمان على نوعية العينة لـ 15 دراسة بينما اختلفت الاستجابة لـ 5 دراسات وعليه فإن معامل الاتفاق:

$$AR = \frac{15}{20} \times 100 = 75\%$$

ويرى البعض ان الحد الأدنى لقبول الثبات هى 70% . وهو بسيط فى حسابه ويعتبر مؤشرًا لثبات التكويد ولكن توجد بعض المحددات هى قضية التحيز فيمكن أن يكون أحد المكودين متحيز لبديل معين وعليه يوجد احتمال لحدوث حكم أو تقدير بالصدفة أو الخطأ.

- **المعامل كابا لكوهين** Cohen Kappa(K): هذا المعامل يعتبر مؤشر لتقدير ثبات المقدرين Inter-rater reliability للأكواد التصنيفية (نعم- لا أو مناسب- غير مناسب) ويعالج قضية التحيز التى يعانى منها معامل الاتفاق

ويقدر المعامل Kappa من المعادلة الآتية:

$$\mathbf{K=} \frac{\sum fo - \sum fe}{N - \sum fe}$$

حيث: $\sum fo$ مجموع التكرارات الملاحظة فى الخلايا القطرية، $\sum fe$ مجموع التكرارات المتوقعة فى الخلايا القطرية، N حجم العينة أو عدد الدراسات.

ويتم حسابه من جدول اقتران ثنائى مثل حسابات كا تربيع. وهذا الإحصاء مفيد كمؤشر لثبات التكويد للخصائص الدراسة والاستجابة على مستويات اسمية تصنيفية. ويرى Card (2012) إذا كانت الاستجابة على مقياس ترتيبى تدريج خماسى مثلًا فيمكن استخدام مؤشر كابا الموزون، والمحدد الرئيسى لهذا المعامل هو يتطلب عدد كبير من الدراسات تتراوح من 40 إلى 50 دراسة للحصول على تقدير دقيق للثبات.

- **معامل ارتباط بيرسون(r) Pearson Correlation**: يكون مفيد إذا كانت بعض خصائص الدراسة تكود على مقياس رتبى متعدد التصنيفات. وإذا كان التكويد للمفردة عبر كل الدراسات على مقياس رتبى متعدد عبر مناسبتين متتاليتين لنفس

المكـود يمكـن اسـتخدام أحـد الاختبـارات لتحديـد الفـروق بـين القياسـات التكويديـة باستخدام اختبار T المرتبطة.

بناء وتنظيم نظام أو بروتكول التكويد

بعـد تحديـد الخصـائص المـراد تكويـدها، فـالخطوة التاليـة التخطـيط لتكويـد الدراسـات، والمرشـد لعمـل التخطـيط هـو بروتكـول التكويـد الـذى يتضـمن وسـيط التكويـد الشـيت يـدوى أو الكمبيـوتر الـذى يفـرغ فيـه المعلومـات المكـودة بالإضـافة إلـى كتـاب التكويـد الـذى يتضـمن تعليمـات عمليـة التكويـد. وفـى النهايـة نحصـل علـى قاعـدة بيانـات قابلـة للتحليـل فى برامج ما وراء التحليل.

التكويد اليدوى Manual Coding

كتـاب التكويـد اليـدوى هـو مجموعـة مفصـلة مـن التعليمـات تصـف نوعيـة المعلومـات المطلـوب استخلاصـها مـن البحـوث والدراسـات فـى صـورة كميـة لإدخالهـا البرنـامج الكمبيـوترى لمـا وراء التحليـل، وهـذا الكتـاب التكويـدى يخـدم ثلاثـة اهـداف **الأول** يمـد المكـود بمرشـد لنقـل المعلومـات أو خصـائص الدراسـة إلـى وسـيط التكويـد (الشـكل الـورقى) وهـذه المعلومـات فـى كتـاب التكويـد تسـتخدم فـى المواقـف النموذجيـة كـذلك فـى المواقـف التـى تتضـمن اشـكاليات معينـة. **الثـانى** يهـدف إلـى توحيـد المعلومـات المتحصـل عليهـا عبـر المكـودين بالتـالى أى حـدوث اتسـاق عـن طريـق تعليمـات واضـحة وشـفافة لمـا يقـوم بـه كـل مكـود. **الثالـث** يعتبـر مفتـاح التكويـد اليـدوى بمثابـة وثيقـة لعمليـة التكويـد التـى تعتبـر كالمرشد لقاعة بيانات ما وراء التحليل.

ويعـد مفتـاح التكويـد اليـدوى الأولـى قبـل تكويـد الدراسـات واى معلومـات يمكـن أن تضـاف الية اثناء تكويد الدراسات وعليه فهو قابل للتعديل بالحذف أو الاضافة.

وفيما يلى مثال التكويد اليدوى:

مسمى الدراسة: تكتب مسماها حسب APA

1- رقم الدراسة: يحدد رقم محدد لكل دراسة.

2- ما نوع النشر للدراسة؟

-كتاب – مجلة – دكتوراه أو ماجستير – مؤتمر – تقرير حكومى

3- ما سنة النشر؟.....

توصيفات العينة

4- متوسط عمر العينة:....

5- جنس العينة:

- ذكر – أنثى – خليط

6- المرحلة الدراسية أو نوعية العينة:

– رياض اطفال - ابتدائى – اعدادى - ثانوى

– جامعى – دراسات عليا – لم يحدد

وتختلف بدائل هذه الفقرة حسب موضوع البحث فاذا كان البحث مهتم بمدمن المخدرات فتكون تصنيفات العينة كالتالى:

- حشيش – افيون – ترامدول - بانجو – تامول - لم يحدد

7- طريقة انتقاء العينة:

–عشوائية بسيطة – عشوائية طبقية - عمدية – لم يحدد

8-حجم العينة: .....

9- بيئة العينة:

– ريفية – حضرية - بدوية – ساحلية

توصيفات تصميم البحث

10- نوعية المنهج المستخدم:

– تجريبى – شبة تجريبى - سببى مقارن - ارتباطى - لم يحدد

11- وحدة المعاينة:

– طلاب – فصل دراسى – نزلاء سجون – اطفال شوارع – لم يحدد

12- توزيع أفراد العينة على مجموعات التصميم:

– عشوائية بسيطة – عشوائية طبقية – غير عشوائية - لم يحدد

13- هل تحقق من التكافؤ فى الاختبارات القبلية؟:

- نعم – لا

14 - نوعية التصميم التجريبى:

– المجموعة الواحدة قبلى بعدى

– المجموعتين وقياس قبلى بعدى وعشوائية

- المجموعتين وقياس قبلى بعدى وعشوائية

- المجموعتين وقياس بعدى

– المجموعات الثلاثة وقياس قبلى وبعدى

– تصميم سولمون ذو الأربع مجموعات

15– نوعية المعالجة أو التدخل أو البرنامج:

- إرشادى – علاجى – تدريبى

فاذا كان البرنامج لتخفيف الاكتئاب:

- معرفى – سلوكى – سلوكى معرفى – اسرى

16– **مدة البرنامج بالأسبوع:** .....

17– **النظرية القائم عليها البرنامج:** .....

18– **طبيعة المعالجة التى تلقتها المجموعة التجريبية:**

-معالجة تجريبية اخرى – معالجة تقليدية – لم تتلقى أى معالجة – لم يحدد

تكويد حجم التأثير

لتكويد حجم التأثير لكل دراسة يجب تكويد المفردات الآتية:

1- كود أو رقم الدراسة.

2– عدد حجم التأثير فى الدراسة الواحدة.

توصيفات المتغير التابع

3– نوع حجم التأثير:

– للقياسات القبلية – القياسات البعدية – القياسات التتبعية

4- المتغير أو المتغيرات الناتج:.....

بيانات حجم التأثير

5- نوعية البيانات لحساب حجم التأثير:

– متوسطات وانحرافات المعيارية – قيمة T - قيمة F

– قيمة كأى تربيع – التكرارات والنسب للمتغيرات التصنيفية

6– رقم الصفحة الذى قدرت فيه حجم التأثير فى البحث:.....

حجم العينة:

المجموعة الضابطة:.....

المجموعة التجريبية:.....

المتوسطات والانحرافات المعيارية

متوسط المجموعة التجريبية:.....

متوسط المجموعة الضابطة:.....

الانحراف المعيارى للمجموعة التجريبية.....

الانحراف المعيارى للمجموعة الضابطة.....

النسب والتكرارات

عدد أفراد المجموعة التجريبية اللذين حققوا نجاح أو تحسن

عدد أفراد المجموعة الضابطة اللذين حققوا نجاح أو تحسن

نسبة أفراد المجموعة التجريبية اللذين حققوا تحسن

نسبة أفراد المجموعة الضابطة اللذين حققوا تحسن

اختبارات الدلالة

- قيمة اختبار T
- قيمة اختبار F(df=1)
- قيمة كا تربيع

حساب حجم التأثير

- قيمة حجم التأثير:
- قيمة حجم التأثير: كبيرة – متوسطة – ضعيفة.

قاعدة البيانات لما وراء التحليل

ناتج عملية التكويد يجب أن تترجم إلى ملف بيانات الكترونى لإجراء ما وراء التحليل كالآتى:

كود الدراسة	المؤلفين	تاريخ النشر	حجم العينة	العمر	الجنس	المرحلة	انتقاء العينة
01	عامر	2003	30	17.5	1	4	1
02	الهادى	2006	40	13	3	3	2
03	القطراوى	2009	50	16.8	2	4	3

فالمتغيرات أو الأعمدة تمثل خصائص الدراسة المكودة أما الصفوف تمثل الدراسات، وقاعدة البيانات تتضمن كل خصائص الدراسة من العينة والمنهج والطريقة وحجم التأثير وعليه فإن هذا الجدول يتضمن عدد كبير من الأعمدة.

الخطوة الرابعة: حساب وتكويد حجم التأثير Computing and Coding Effect Size: الخطوة المهمة فى ما وراء التحليل تكويد أو قياس النواتج البحثية على مقياس كمى ومن أهم النواتج البحثية موضع اهتمام الدراسة العلاقات بين بنائين أو متغيرين. ويقوم أسلوب ما وراء التحليل على إحصاء حجم الأثر حيث يعتبر المادة الخام ومن أهم المعلومات المستخلصة من الدراسات المتضمنة فى التحليل. ويشير Borenstein et al. (2012) إلى أن مصطلحى تأثيرات المعالجة Treatment effects واحجام التأثير Effect sizes تستخدم بمعانى مختلفة بين الباحثين، فيستخدم مصطلح حجم التأثير فى الطب أو المعالجات الطبية ليشير إلى تأثير المعالجة ويتضمن إحصائيات مثل نسب المخاطرة Risk ratios أو فروق المخاطرة Risk difference، بينما فى العلوم النفسية والاجتماعية يستخدم مصطلح حجم التأثير ليشير إلى فروق المتوسطات المعيارية أو الارتباطات حيث انهما أكثر استخدامًا فى دراسات ما وراء التحليل فى الدراسات النفسية والاجتماعية.

وفيما يلى عرض لثلاثة انواع رئيسية لأحجام التأثير:

مؤشرات الفروق بين المتوسطات

فروق المتوسطات غير المعيارية Ustandardized Mean Difference

عندما تحدث مقارنة بين مجموعتين(ضابطة – تجريبية، ذكور – إناث) على متغير متصل ما (التفكير الابداعى – الاكتئاب) فإن إحصاء حجم التأثير يمكن التعبير عنه بالفروق بين متوسطى المجموعتين ويسمى إحصاء حجم التأثير غير المعيارى أو الخام Unstandardized effect size، ولكن يفترض ان تستخدم الدراسات نفس المقياس فمثلًا هدفت الدراسات إلى دراسة فعالية برامج الالعاب التعليمية فى تنمية الذكاء، واعتمدت الدراسات على مقياس ستانفورد بينيه فى هذه الحالة يستخدم الفروق الخام بين المتوسطات مؤشر لحجم التأثير بفاعلية أو يستخدم للقياسات فى المجال الطبى التى لها مقياس موحد ومعروف مثل ضغط الدم أو يستخدم لمقاييس واسعة الانتشار مثل مقياس Beck للاكتئاب أو تقدير الذات لـ Rosenberg وغيرها.

ويقدر حجم التأثير كالآتى (Lipsey & Wilson, 2001; Borenstien etal., 2009):

$$ES_{um}(D) = \overline{x_1} - \overline{x_2} = M_1 - M_2$$

- $\overline{x_1}$ ، M_1 متوسط المجموعة الأولى
- $\overline{x_2}$ ، M_2 متوسط المجموعة الثانية

ولتقدير الخطأ المعيارى لفروق المتوسطات غير المعيارية لاستخدامها فى تصحيح أو موازنة حجم التأثير كالآتى (Card, 2012; Lipsey & Wilson, 2001):

$$SE_{um} = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \approx \frac{2S_p}{\sqrt{N_{total(equal\ sample\ size)}}}$$

- S_p الانحراف المعيارى المدمج اوالمشترك Pooled للمجموعتين.
- n_1، n_2 حجمى المجموعتين الأولى والثانية.
- N_{total} حجم العينة الكلى عند تساوى حجم العينة فى كل مجموعة.

ويقدر الانحراف المعيارى المدمج للمجموعتين كالآتى:

$$S_{pooled} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2}}$$

- S_1^2 تباين قيم المجموعة الأولى.
- S_2^2 تباين المجموعة الثانية.

مثـال: افتـرض أن دراسـة تجريبيـة هـدفت إلـى المقارنـة بـين المجموعـة التجريبيـة والمجموعـة الضـابطة فـى العـدوان وكـان متوسـط المجموعـة الأولـى($\overline{x_1}$)= 103 ومتوسـط المجموعـة الثانيـة ($\overline{x_2}$) = 100 والانحـراف المعيـارى للعينتـين S1=5.5، S2=4.5 وحجـم العينـة للعينتين متساوى $n_1 = n_2 = 50$ وعليه فإن حجم التأثير:

$$SE_{um} = 103 - 100 = 3$$

والانحراف المعيارى المدمج كالآتى:

$$S_{pooled} = \sqrt{\frac{(50-1) \times 5.5^2 + (50-1) \times 4.5^2}{50+50-2}} = 5.0249$$

والخطأ المعيارى لفروق المتوسطات:

$$SE_{um} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 5.0249 \sqrt{\frac{1}{50} + \frac{1}{50}} = 1.0050$$

الإجـراءات الحسـابية: تتطلـب حسـاب حجـم التـأثير مـن خـلال حسـاب الفـرق بـين المتوسـطين مباشـرة وتـوافر الانحـراف المعيـارى للمجمـوعتين واحجـام عينـة. ولكـن بفـرض ان بعـض الدراسـات لـم تـذكر أحـد هـذه المؤشـرات خاصـة الانحـراف المعيـارى فـى هـذه الحالة يمكن حساب الانحراف المعيارى من قيمة الاختبار مباشرة مثل T و F.

مؤشرات فروق المتوسطات المعيارية Standardized Mean Difference(انظر الفصل العاشر)

مؤشرات قوة العلاقة (انظر الفصل القوة الإحصائية وحجم التأثير)

هـى المؤشـرات البديلـة لفـروق المتوسـطات المعياريـة لتقـدير حجـم التـأثير وتنقسـم إلـى المؤشرات الارتباطية ومؤشرات التباين المفسر.

المؤشـــرات الارتباطيـــة: تقـيس هـذه المؤشـرات قـوة العلاقـة بـين متغيـرين فـأكثر وهـى كالتالى:

معامـــل ارتبـــاط بيرســـون (r): يقـيس درجـة التبـاين المشـترك أو قـوة العلاقـة الخطيـة بـين متغيرين متصلين. وتقدر قيمته كالآتى:

$$r_{xy} = \frac{N\sum XY - N\sum X \sum Y}{\sqrt{(N(\sum X^2 - (\sum X)^2)}(\sqrt{N(\sum Y^2 - (\sum Y)^2)}}$$

- ΣXY مجموع حاصل ضرب X فى Y.
- ΣX مجموع قيم X، ΣY مجموعة قيم Y.
- ΣX^2 مجموع مربعات X، $(\Sigma X)^2$ مربع مجموع X.
- ΣY^2 مجموع مربعات Y، $(\Sigma Y)^2$ مربع مجموع Y ، N حجم العينة.

ويكون أدائه أفضل إذا كان توزيع البيانات اعتداليًا والعلاقة بين المتغيرين خطية.

والارتبـاط الموجـب يعنـى أن الأفـراد مرتفعـى الـدرجات علـى المتغيـر X لهـم درجـات مرتفعـة علـى المتغيـر Y بينمـا الارتبـاط السـالب الأفـراد مرتفعـى الـدرجات علـى المتغيـر X لهم درجات منخفضة على المتغير Y .

وتعتبـر قيمـة معامـل الارتبـاط مؤشـرًا لحجـم التـأثير، وهـو مؤشـر أكثـر معياريـة وأسـهل فـى تفسـيره نتيجـة اعتمـاده علـى معـايير فـى تحديـد قـوة العلاقـة ويقـع مـداه مـن 1.0 (موجـب) إلـى 1.0- (سـالب) عكـس مؤشـر كـوهين " d " فلـيس لـه معـايير واضـحة لتفسـيره (Rosenthal et al., 2000)، وتفسـير قيمـة r كمـا وضـحها Cohen (1988) كمؤشر لحجم التأثير إذا كانت $r = \pm 0.10$ حجم تأثير ضعيف، $r = \pm$ 0.30 حجم تأثير متوسط، $r \geq \pm 0.5$ حجم تأثير كبير، ولكن عليك الحذر من ان هـذه المعـايير ليسـت ثابتـة فـى كـل التخصصـات أو المجـالات. ويفسـر r مـن خـلال مربـع قيمتـه (r^2) ويطلـق عليـه **معامـــل التحديـــد** Coefficient of determination، وعلـى ذلـك يكـون مربـع معامـل الارتبـاط ضـعيفًا إذا كانـت $r^2 = 0.01$، متوسـط $r^2 = 0.09$، وكبيـر $r^2 \geq 0.25$. فمثـلًا إذا كـان r =.50 فـإن مربـع معامـل الارتبـاط $r^2 = .25$ بمعنـى أن أحـد المتغيـرين فسـر 25% مـن تبـاين المتغيـر الثانى.

وتباين الخطأ أو الخطأ المعيارى لمعامل الارتباط تقريبًا Borenstien et al., (2009):

$$V_r = \frac{(1-r^2)^2}{N-1}$$

حيث N حجم العينة، r معامل الارتباط.

ومعظم تحليلات ما وراء التحليل لا تجرى على معامل الارتباط ولذلك يحول معامل الارتباط قبل دمج أحجام التأثير عبر الدراسات المختلفة إلى تحوير أو مقياس فيشر transformation(Z_r) Fisher (ليس اختبار الدلالة الإحصائية Z) ويجرى التحليل على القيم المحورة، ويقدر Z_r كالتالى (Card, 2012; Lipsey & Wilson, 2001):

$$(ES_{Z_r})Z_r = \frac{1}{2}Ln(\frac{1+r}{1-r})$$

حيث Ln أو log_e اللوغاريتم الطبيعى.

وعليه فإن الدراسات التى تعتمد على معامل الارتباط كمؤشر لحجم التأثير فى الدراسات الارتباطية فى ما وراء التحليل يجب تحويله إلى الدرجة أو تحوير Fisher ثم تحويل القيم النهائية لفيشر إلى معامل الارتباط لتفسير النتائج.

ويحول معامل ارتباط بيرسون إلى تحوير فيشر فى ما وراء التحليل لأن التوزيع العينى لمعامل الارتباط فى المجتمع يكون ملتوى ناحية اليمين للمعاملات السالبة وناحية اليسار للمعاملات الموجبة وعندما يكون قريب من الصفر فإن التوزيع العينى يكون تقريبًا اعتدالى وتتحسن الاعتدالية بزيادة حجم العينة ولان التوزيع العينى يكون ملتوى كلما زادت قيمة معامل الارتباط فإنه من غير المناسب استخدام التوزيع الملتوى لاختبارات الفروض حول معامل الارتباط فى المجتمع ρ إلا إذا كانت 0= ρ وهذا أدى بـ Fisher لإجراء تعديل فيشر Z لأنه توزيع منتظم Symmetric.(Lipsey & Wilson, 2001, Hedges &Olkin, 1985)

وتباين تحوير Z_r:

$$V_z = \frac{1}{N-3}$$

حيث N حجم العينة.

والخطأ المعيارى لـ Z_r:

$$SE_{Z_r} = \sqrt{V_z}$$

$$= \sqrt{\frac{1}{N-3}}$$

وكلما زادت حجم العينة فإن مقام المعادلة السابقة يزيد وبالتالى تنقص قيمة الخطأ المعيارى. ويجب إدخال الخطأ المعيارى لتحوير فيشر إلى قاعدة البيانات اثناء عملية التكويد لاستخدامه فى عملية موازنة أحجام التأثير، وعليه عند استخدام تحوير فيشر لا نستخدم تباين معامل الارتباط انما نعتمد على تباين الدرجة Z_r. ولان الدرجة Z_r أقل قابلية للتفسير لعدم امتلاكها معايير واضحة مثل معايير معامل الارتباط فتحول نتائج الدرجة فيشر Z_r إلى معامل الارتباط من خلال التالى (Hedges & Olkin, 1985):

$$r = \frac{e^{2Z_r} - 1}{e^{2Z_r} + 1}$$

حيث r معامل ارتباط بيرسون، Z_r تحوير فيشر المقابل إلى r.

فعلى سبيل المثال فى دراسة ما قدرت معامل الارتباط بـ 0.50 مع حجم عينة 100 بالتالى فإن تحوير فيشر Z_r كالآتى:

$$Z_r = 0.5 \times log_e\left(\frac{1+0.5}{1+0.5}\right) = 0.5493$$

وتباين الدرجةZ_r :

$$V_z = \frac{1}{100-3} = 0.0103$$

والخطأ المعيارى للدرجة Z_r:

$$SE_z = \sqrt{0.0103} = 0.1015$$

ولتحويل الدرجةZ_r إلى معامل الارتباط:

$$r = \frac{e^{(2\times 0.5493)} - 1}{e^{(2\times 0.5493)} + 1} = 0.50$$

مثـــال: أراد باحـث دراســة العلاقــة بـين ادمــان الانترنــت والوحــدة النفسـية بالتــالى جمــع الدراسـات الـت تناولـت هـذه العلاقـة مـن التـراث فـى البيئـة العربيـة وهنـا يعتبـر معامـل الارتبـاط مؤشـر لحجـم التـأثير ثـم يحـول إلـى تحـوير Fisher وبعـد حسـاب المتوسـط العام لتحوير Z يحول مرة اخرى إلى معامل الارتباط.

حساب قيمة معامل الارتباط من قيم الاختبارات

بعـض الدراسـات الأوليـة المنشـورة تتضـمن نتـائج الإحصـائيات الاسـتدلالية دون حسـاب حجـم التـأثير ويفضـل فـى كثيـر مـن تحلـيلات مـا وراء التحليـل التعبيـر عـن معظـم الإحصائيات فى ضوء معامل الارتباط r وهذا الموقف يحدث لعدة اسباب منها:

1. تتضـمن الدراسـات الدلالـة الإحصـائية لمعامـل الارتبـاط بـدون قيمـة معامـل الارتبـاط وهذا الموقف نادر الحدوث.

2. بعـض البـاحثين يميلـون إلـى تقسـيم المتغيـر المتصـل إلـى تصـنيفين مثـل مرتفـع ومـنخفض التحصـيل أو مرتفعـى ومنخفضـى الدافعيـة ثـم المقارنـة بينهمـا علـى المتغيـر التـابع باسـتخدام اختبـار T أو F، ولكـن هـذا تقسـيم اصـطناعى وبـدوره يـؤدى إلـى تقلـص Attenuation قيمـة حجـم التـأثير (Hunter & Schmidt, 2004)، ولـذلك مـن المفيـد حسـاب قيمـة معامـل الارتبـاط بـل وتصـحيحه مـن التقسـيم الاصـطناعى للمتغيـرات المتصلة.

3. بعـض الدراسـات تُقسـم أو تصـنف كـلًا مـن المتغيـرين المتضـمنين فـى معامـل الارتبـاط إلـى تصـنيفين ثـم تحلـل المتغيـرين مـن خـلال جـداول الاقتـران باسـتخدام إحصـاء χ^2 وفـى هذة الحالة فالباحث بحاجة إلى تصحيح قيمة معامل الارتباط.

وتقدر قيمة r كمؤشر لحجم التأثير من قيمة اختبار **T** المستقلة كالتالى:

$$r = \sqrt{\frac{T^2}{T^2 + df}}$$

- T قيمة الاختبار المحسوبة
- df درجات الحرية: df=N-2=n1+ n2 -2

وتقدر قيمة r من اختبار F لعينتين مستقلتين و $n_1 \neq n_2$ كالتالى:

$$r = \sqrt{\frac{F}{F + df}}$$

- df درجات الحرية للمقام أو الخطأ N-C= ، df حجم العينة، C عدد المجموعات ويساوى 2.

ويقدر قيمة r من اختبار T المرتبطة كالتالى:

$$r = \sqrt{\frac{T^2}{T^2 + df}}$$

- df درجات الحرية (df =N-1)

ويقدر قيمة r من اختبار F لعينتين مرتبطتين كالتالى:

$$r = \sqrt{\frac{F_{\text{المرتبطة}}}{F_{\text{المرتبطة}} + df}}$$

والبعض ضد عدم دمج نتائج اختبارات العينات المستقلة مع نتائج اختبارات العينات المرتبطة، وينصح Lipsey & Wilson (2001) بإجراء ما وراء تحليل منفصل لنتائج الاختبارات المستقلة (تصميم بين المجموعات أو الأفراد) عن نتائج الاختبارات المرتبطة (تصميم داخل المجموعات أو الأفراد)، ولكن Card (2012) يرى لإجراء فصل لا بد من إجراء تحليل التفاعل باعتبار نوعية المنهجية متغير تصنيفى (عينات مستقلة فى مقابل عينات مرتبطة) وإذا اتضح فروق فى أحجام التأثير يجرى ما وراء التحليل منفصل.

وتقدر r من إحصاء χ^2 كالتالى:

$$r = \sqrt{\frac{\chi^2}{N}}$$

- N حجم العينة

حساب معامل الارتباط r من الإحصائيات الوصفية

بعض الدراسات تعرض الإحصائيات الوصفية مثل المتوسطات والانحرافات المعيارية واحجام العينات للمتغير تابع للمجموعتين ويمكن استخدامها لحساب r من خلال حساب قيمة فروق المتوسطات المعيارية ثم تحويلها إلى r من خلال حساب مؤشر d أو g ثم تحويلهما إلى r من المعادلة التالية (Card, 2012):

$$r = \sqrt{\frac{g^2 n1 n2}{g^2 n1 n2 + (n1 + n2)df}}$$

بعض الدراسات البحثية تعرض النتائج فى ضوء جدول اقترانى لمتغيرين تصنيفين 2×2 وفى هذا الموقف يتم حساب قيمة معامل الارتباط فأى ويفسر على أنه معامل ارتباط بيرسون r.

وفيما عرض للحدود المتفق عليها لمؤشرات حجم التأثير:

الجدول (1.22): حدود حجم التأثير لمؤشرات بعض الاختبارات.

الاختبار	المؤشر	صغير	متوسط	كبير
T المستقلة	d, Δ, g	.20	.50	.80
الارتباط	(r^2) r	(.01).10	(.09).30	(.25).50
جدول الاقتران	W,V, φ,c	.10	.30	.50
ANOVA	f	.10	.25	.40
	ω^2, μ^2	.01	.06	.14

الانحدار المتعدد	f^2	.02	.15	.35
	R^2	.02	.13	.26

ما وراء التحليل لثبات الاتساق الداخلى

يهدف مقياس الاتساق الداخلى أو الثبات الداخلى للمقياس التى تحديد درجة التناسق بين مفردات المقياس، وتوجد العديد من مقاييس الاتساق الداخلى ولكن اكثرهم انتشارًا المعامل ألفا كرونباخ ويقدر فى ضوء عدد المفردات فى المقياس j ومتوسط الارتباطات بين هذه المفردات($\bar{r}$) كالآتى:

$$\alpha = \frac{j\,\bar{r}}{1 + (j-1)\,\bar{r}}$$

متى يتم إجراء ما وراء التحليل لمقاييس الاتساق الداخلى؟ يرى Card (2012) موقفين لاستخدام الاتساق الداخلى فى ما وراء التحليل هما:

الأول: عند تصحيح الثبات فى الدراسات من نقص الثبات Unreliability .

الثانى: عندما يهدف الباحث إلى الوصول إلى متوسط المعامل ألفا كرونباخ لدراسات مختلفة استخدمت نفس المقياس وهذا فى غاية الأهمية خاصة للمقاييس المعروفة فى دراسة ظاهرة معينة مثل مقياس كاتل للشخصية أو اختبار ستانفورد بينية أو العوامل الخمسة الكبرى للشخصية وغيرها لتقيم استخدام هذه المقاييس فى دراسة الظاهرة. وتوجد طرق مطروحة فى التراث لإجراء ما وراء التحليل لمعامل ألفا كرونباخ منها ما قدمه Rodriguez & Maeda (2006) حيث تم تحوير المعامل ألفا لمؤشر لحجم التأثير كالتالى:

$$ES_\alpha = \sqrt[3]{1-\alpha}$$

- α معامل الثبات فى الدراسة.

والخطأ المعيارى لهذا التحوير لمعامل ألفا كالآتى:

$$SE_{ES_\alpha} = \sqrt{\frac{18j(N-1)(1-\alpha)^{\frac{2}{3}}}{(j-1)(9N-11)^2}}$$

- j عدد المفردات فى المقياس فى الدراسة.
- N حجم العينة.
- α معامل الثبات فى الدراسة.

وبعد حساب متوسط معامل الاتساق المحور ألفا يحو إلى معامل ألفا مرة اخرى حتى يسهل تفسيره كالآتى:

$$\alpha = 1 - ES_\alpha^3$$

- ES_α معامل ألفا المحور فى ما وراء التحليل.

واخيـرًا توجـد مجموعـة مـن الاشـكاليات أو التحـديات عـن إجـراء مـا وراء التحليـل لهـذه النوعيـة مـن أحجـام التـأثير التـى عرضـناها فـى هـذا الفصـل يحـددها (Card, 2012; Lipsey & Wilson, 2001):

- عـدم تـوافر البيانـات المتـوفرة لتقـدير حجـم التـأثير، فمـثلًا فـى الدراسـات الطوليـة لا تتـوفر للباحـث معامـل الارتبـاط بـين مرتـى القيـاس للمتغيـر عكـس معامـل الاتسـاق الـداخلى الـذى تتـوفر البيانـات المطلوبـة لحسـاب حجـم التـأثير. وعـدم تـوفر البيانـات اللازمة لحساب حجم التأثير يتطلب الاتصال المباشر بمؤلفى هذه الدراسات.

- التحـدى الثـانى هـو عـدم الاتسـاق فـى الطـرق التحليليـة وتقريـر أحجـام التـأثير المناسـب، فمـثلًا فـى تحليـل الانحـدار عنـد حسـاب حجـم التـأثير حيـث تتضـمن الدراسـات متغيـرات منبئـات أو تغـايرات متنوعـة غيـر المتغيـرين موضـعى الاهتمـام وتتضـمن الدراسـات طـرق تحليليـة متنوعـة فبعضـها اعتمـد علـى طريقـة التحليـل Enter وبعضـها اعتمـد على Stepwise.

ايضًـا عنـد إجـراء مـا وراء التحليـل علـى التحليـل العـاملى الاستكشـافى حيـث يعتمـد علـى تشـبعات المفـردات بالعوامـل كمؤشـر لحجـم التـأثير فيوجـد اخـتلاف واضـح فـى الطـرق التحليليـة فـى ضـوء طـرق التقـدير أو الاسـتخلاص (المكونـات الأساسـية- طـرق التحليـل العـاملى) وطـرق التـدوير (مائـل- متعامـد) وكيفيـة تحديـد عـدد العوامـل، وبكـل تأكيـد

يمكن أن تختلف النتائج بين هذه الطرق التحليلية المختلفة وهذا يمثل تحدى كبير امام الباحث عند دمج هذه النوعية من الدراسات.

ولهذا فإن استخدام هذه المؤشرات فى تراث ما وراء التحليل قليل ويفضل الاعتماد على مؤشرات التقليدية مثل r ,g ,OR، وقبل إجراء هذه النوعية من الدراسات يجب التأكد من توافر المعلومات أو البيانات اللازمة لحساب حجم التأثير.

الخطوة الخامسة: دمج أحجام التأثير Combining Effect Sizes: بعد جمع الدراسات وحساب أحجام التأثير ثم إجراء التصحيحات يبدأ الباحث فى دمج هذه أحجام التأثيرات فى ما وراء التحليل لحساب حجم التأثير العام عبر الدراسات المختلفة للوصول إلى الاستنتاج العام عن الظاهرة أو المتغيرات موضع الدراسة. وعملية الدمج تكون سهلة لو ان الدراسات اتبعت نفس المنهجية والعينات والقياسات والتصميمات ونفس المفاهيم النظرية. ومن الأسئلة الرئيسية التى يحاول ما وراء التحليل الاجابة عليها **الأول:** ما هو حجم التأثير النموذجى للعلاقة بين المتغيرين موضع الدراسة من عينة من الدراسات التى تناولت المتغيرين؟ **الثانى:** هل يوجد اختلافات بين أحجام التأثير المتضمنة فى التحليل بمعنى انها أحجام تأثير متجانسة؟ والاجابة على السؤال الثانى فى غاية الأهمية لتحديد طبيعة نوع النموذج المستخدم لحساب حجم التأثير العام وايضًا ضرورية لتحديد مصادر التنوع والاختلاف بين أحجام التأثير من الدراسات المختلفة من خلال استراتيجية التحليل المتفاعل Moderator analysis.

يتضمن ما وراء التحليل تصنيفين من المتغيرات (Lipsey & Wilson, 2001):

1. أحجام التأثير للمتغيرات التابعة موضع الاهتمام (التحليل الرئيسى).

2. المتغيرات الوصفية أو خصائص الدراسات التى ولدت أحجام التأثير والتى تؤسس للمتغيرات المستقلة فى التحليل (التحليل الثانوى).

وتحليل البيانات يتضمن وصف توزيعات المجموعة المنتقاة لأحجام التأثير من حساب المتوسطات والتباينات لأحجام التأثير ثم بعد ذلك بحث العلاقة بين أحجام التأثير

والمتغيرات الوصفية أو خصائص الدراسات من خلال استخدام تحليل التباين ANOVA أو ما وراء الانحدار.

اشكالية تحليل ما وراء التحليل

يعتبر الفرد أو الشخص هو وحدة التحليل فى العلوم السلوكية فى الدراسات البحثية باختلاف منهجياتها المختلفة بينما وحدة التحليل فى ما وراء التحليل هى الدراسة البحثية وهنا توجد اشكاليات عديدة أهمها:

1. الدراسات فى ما وراء التحليل يمكن أن تنتج أكثر من حجم تأثير حيث يمكن للدراسة الواحدة ان تتضمن علاقة بين المتغيرين للذكور وللإناث أو علاقة بين متغيرين على فترات زمنية مختلفة ولكن الاعتماد على أكثر من حجم تأثير لنفس الدراسة يسبب الاعتمادية الإحصائية Statistically dependence وهذا اختلال لمسلمة الاستقلالية اللازمة لتحليل البيانات وهذا يسبب خطأ جوهرى للتحليل الإحصائى الاستدلالى لأحجام التأثير. والاعتماد على أكثر من حجم تأثير للدراسة يسبب زيادة فى حجم عينة أحجام التأثير المعتمدة وهذا يشوه الخطأ المعيارى.

2. الدراسات المجمعة لحساب أحجام التأثير تتضمن أحجام عينات مختلفة من الأفراد وبما أن حساب أحجام التأثير يعتمد على المتوسطات والانحرافات المعيارية و الارتباطات، وعليه فإن الخصائص الإحصائية لأحجام التأثير تعتمد فى المقام الأول على حجم العينة، فحجم التأثير القائم على حجم عينة كبير يتضمن خطأ معاينة صغير وبالتالى يكون تقديره أكثر دقة وثبات من ذلك حجم التأثير القائم على حجم عينة صغير، وبالتالى إذا تم إجراء التحليل الإحصائى لما وراء التحليل بدون الاخذ فى الاعتبار حجم التأثير كأننا عالجنا كل الدراسات كما لو كانت لها نفس حجم العينة ويكون لهما نفس الاسهام لنتائج ما وراء التحليل، ولكن أحجام التأثير المقدرة من أحجام عينات كبيرة يجب أن تلعب دورًا أكبر من نظيرتها المقدرة من أحجام عينات صغيرة لأنها أقل خطأ معاينة، وعليه لا يجب معالجة أحجام التأثير بنفس الدرجة من الأهمية فى التحليل الإحصائى ولكن يجب أن تعطى وزن Weight أكبر لأحجام التأثير من

الدراسات ذات العينات الكبيرة فى الحسابات الإحصائية من تلك التى تعتمد على أحجام عينات صغيرة ولذلك لا بد من إجراء التحليل الموزون Weighted analysis.

ويمر تحليل البيانات فى ما راء التحليل بعدة خطوات:

اولًا: إجراء تصحيحات لأحجام التأثير من المحددات المنهجية للدراسات (الفصل السابع).

ثانيًا: حساب متوسط حجم التأثير وتحليل توزيع أحجام التأثير: وتعتبر هذه من أهم الخطوات لإجراء ما وراء التحليل وتتضمن الخطوات الآتية:

1. إنشاء قاعدة بيانات لمجموعة من أحجام التأثير المستقلة.
2. حساب المتوسط الموزون Weighted mean.
3. حساب فترات الثقة والدلالة الإحصائية حول المتوسط الموزون.
4. اختبار التجانس Homogeneity بين أحجام التأثير المختلفة.

وإذا لم يوجد تجانس بين أحجام التأثيرات بمعنى اختلافات بينهم فمن الضرورى اعادة التحليل من الخطوة 2 حتى 4 باستخدام نموذج إحصائى مختلف إلى جانب تحليلات اضافية.

بناء قاعدة بيانات لمجموعة مستقلة من أحجام التأثير Creating an Independence Set of Effect Size

كما سبق وأشرنا أنه يجب أن تكون أحجام التأثيرات مستقله عن بعضها البعض بمعنى لا يجب تضمين أكثر من حجم تأثير من نفس الدراسة لنفس العينة من الأفراد، ويرى البعض ان عدم تضمين أحجام التأثير للعينات الفرعية فى الدراسة (مثلًا ذكور – إناث)، وكذلك أحجام تأثير من دراسات مختلفة لنفس الفريق البحثى (Wolf, 1990). ولكن دراسات ما وراء التحليل يتم جمعها بناءً على المفهوم المتضمن فى الدراسة

ويمكن أن تتضمن الدراسة أكثر من مفهوم وعليه يمكن الاعتماد على أكثر من حجم تأثير في الدراسة التي تتناول مفاهيم مختلفة وهذا نادرًا ما يحدث لأنه لا يتناسب مع منهجية ما وراء التحليل التي تركز على متغيرين، ولكن يمكن إجراء ما وراء التحليل لأحجام تأثير لدراسات تتضمن أكثر من متغيرين مثل دراسات الانحدار المتعدد وتحليل التباين المتدرج وهذا يتم من خلال ما وراء التحليل المتدرج Multivariate meta-analysis. ولو تضمنت الدراسة أكثر من حجم تأثير للمفهوم بمعنى اعتمدت على تعريفات إجرائية مختلفة فلا يجب تضمينها جميعها في نفس التحليل كما لو كانت بيانات مستقلة. إذا تضمنت الدراسة أحجام تأثير متعددة فيمكن اختصارها بإحدى الطريقتين Lipsey & Wilson (2001):

الأولى: حساب متوسطهما والتعامل مع المتوسط كحجم تأثير واحد.

الثانى: الاكتفاء بتضمين حجم تأثير واحد واستبعاد بقية أحجام التأثير الأخرى والاختيار يتم في ضوء مدى مناسبة التعريف الإجرائي في الدراسة مع التعريفات الإجرائية في كل الدراسات الأخرى وجودة عملية القياس وإذا تساوت هذه المعايير فيتم الاختيار عشوائى.

حساب متوسط حجم التأثير The mean effect size

قبل حساب متوسط حجم التأثير فإن اعطاء وزن متساوى لكل حجم تأثير فيه شئ من التجاهل لحقيقة ان الدراسات ليست بنفس القدر من الجودة وان كلها يمكن أن تسهم بنفس القدر من الاسهام في حساب متوسط حجم التأثير. في الحقيقة توجد دراسات أكثر جودة ودقة من حيث الخصائص المنهجية مثل إجراءات الضبط وحجم العينة الكافى من نظيراتها الأقل جودة من حيث أقل ثبات وحجم عينة صغير بالتالى فمن غير المقبول ان تتساوى مع الدراسات عالية الجودة في حساب حجم التأثير العام. من هنا تأتى أهمية عملية الموازنة Weighting للدراسات المتضمنة في التحليل بما يتناسب مع دقتها وجودتها وحجم العينة.

مثـال(Card, 2012): افتـرض أن دراسـة تتضـمن عينـة 10 أفـراد وتوصـلت إلـى ان العلاقـة بـين X، Y تسـاوى 0.20 (تحـوير فيشـر= 0.203)، والدراسـة الثانيـة اعتمـدت علـى عينـة 10000 فـردًا و معامـل الارتبـاط بـين X، Y 0.30 (تحـوير فيشـر = 0.310)، ولكـن قبـل حسـاب متوسـط العلاقـة بـين المتغيـرين فـى الدراسـتين فمـن المهـم التأمـل فـى دقـة تقـديرات أحجـام التـأثير للدراسـتين، فالدراسـة الأولـى اعتمـدت علـى عينـة 10 أفراد والخطأ المعيارى لتحوير فيشر Z_r لها كالآتى:

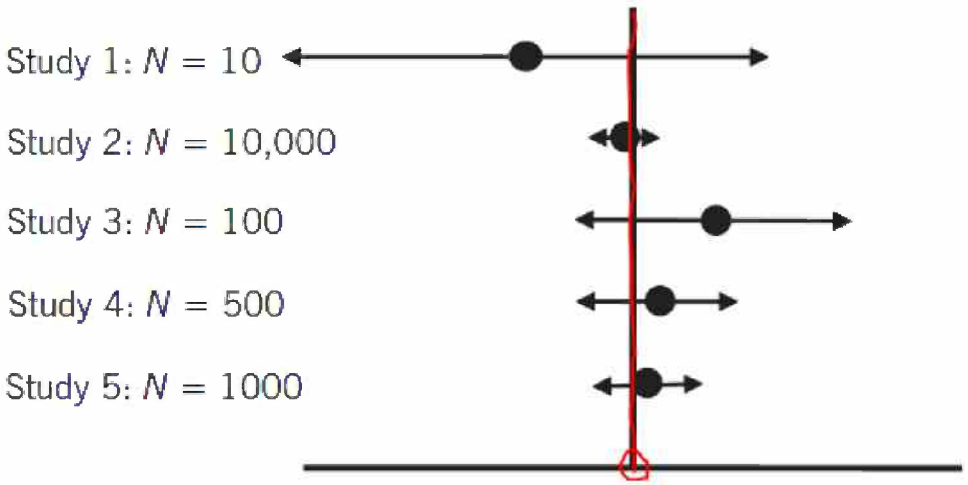
$$SE_{zr} = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{10-3}} = 0.378$$

فى حين ان الخطأ المعيارى لـZ_r للدراسة الثانية:

$$SE_{zr} = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{10000-3}} = 0.010$$

ولاحـظ أن الخطـأ المعيـارى فـى حالـة اسـتخدام عينـات كبيـرة قليـل مقارنـة بنظيـره فـى حالـة عينـات صـغير، وعليـه فـإن أحجـام التـأثير المتولـدة مـن عينـات كبيـرة أكثـر دقـة وتقتـرب مـن حجـم التـأثير الحقيقـى الفعلـى فـى المجتمـع. والخطـأ المعيـارى يتناسـب عكسـيًا مـع حجـم العينة.

والشكل التالى يبين مفهوم الدقة لأحجام التأثير:



الشكل (1.22): عرض لمفهوم الدقة لأحجام التأثير.

الشكل يعرض خمسة دراسات متفاوتة حجم العينات وتختلف فيما بينها فى الدقة لأحجام التأثير فى المجتمع. فالخط المستقيم فى المنتصف(الأحمر) يمثل حجم التأثير الحقيقى فى المجتمع.

فالدراسة الأولى أنتجت حجم تأثير الممثل بالدائرة على يسار الخط الراسى وهى تعتبر أقل من حجم التأثير الحقيقى ولها خطأ معيارى كبير لأن حجم العينة 10 صغير نسبيًا وهذا يؤدى إلى فترة ثقة كبيرة أو واسعة (الخط الأفقى)، وعليه لو اعتمد الباحث على هذه الدراسة فإنها تعطى تقدير منخفض لحجم التأثير الحقيقى فى المجتمع.

اما الدراسة الثانية (N=10000) تتضمن حجم عينة كبير جدًا فإن حجم التأثير فى هذه الدراسة (الدائرة السوداء) يكون قريب جدًا من حجم التأثير الحقيقى فى المجتمع ومدى فترات الثقة يكون ضيقًا لأن هذه الدراسة لها خطأ معيارى صغير جدًا. بالتالى تمتلك دقة عالية فى تقدير حجم التأثير الحقيقى فى المجتمع.

وعلى ذلك يجب اعطاء وزن وأهمية للدراسة الثانية أكثر من الدراسة الأولى فى التحليل الإحصائى لما وراء التحليل للوصول إلى حجم التأثير الحقيقى فى المجتمع.

ولكن الباحث لا يمكن الحصول على حجم التأثير الحقيقى فى المجتمع لمعرفة مدى اقتراب أو ابتعاد أحجام التأثير فى الدراسات من حجم التأثير الحقيقى ولكن الهدف الأسمى من إجراء ما وراء التحليل هو محاولة الوصول أفضل تقدير لحجم التأثير، وعلى ذلك فإن الهدف من عملية الموازنة فى ما وراء التحليل اعطاء وزن أكبر لبعض الدراسات التى لها اخطاء معيارية صغيرة ودقة صغيرة (أحجام عينات كبيرة) مقارنة بنظيرتها التى لها اخطاء معيارية كبيرة ودقة كبيرة (أحجام عينات صغيرة).

وتتم عملية الموازنة من خلال إجراء إحصائى منطقى ومقبول وهو وزن أحجام التأثير عن طريق مقلوب أو معكوس تباينات الأخطاء المعيارية. فاذا كانت وزن الدراسة i هو w_i ، والخطا المعيارى لحجم التأثير فى الدراسة SE_i ، فإن وزن الدراسة هو مقلوب أو معكوس تباين الخطأ المعيارى كالآتى:

$$w_i = \frac{1}{SE_i^2}$$

- SE_i الخطا المعيارى لحجم التأثير المقدر للدراسة i

مثال: قاعدة بيانات لدراسة ما وراء التحليل لـ 22 دراسة للعلاقة بين العدوان والرفض (Card, 2012)

الجدول (1.22): بيانات دراسة ما وراء التحليل لـ 22 دراسة لـ Card et al. (2008)

Sample size (N)	Effect size (r)	Transformed ES (Z_r)	Standard error (SE)	Weight (w)	wES ($w * Z_r$)	wES^2 ($w * Z_r^2$)
228	.525	.583	.0693	208.12	121.27	70.66
491	.198	.201	.0579	297.81	59.74	11.98
65	.311	.322	.1325	56.95	18.34	5.90
458	.554	.624	.0484	427.07	266.55	166.37
929	.161	.162	.0346	835.96	135.72	22.03
904	.336	.349	.0347	831.95	290.69	101.57
74	.396	.419	.1223	66.89	28.02	11.73
151	.617	.721	.0855	136.66	98.48	70.97
150	.557	.628	.0845	139.90	87.90	55.23
590	.575	.655	.0426	550.97	361.12	236.69
180	.039	.039	.0831	144.75	5.70	0.22
139	.358	.375	.0892	125.71	47.14	17.68
132	.049	.049	.0922	117.70	5.81	0.29
60	.000	.000	.1386	52.05	0.00	0.00
839	.326	.339	.0381	689.52	233.68	79.19
262	–.048	–.048	.0642	242.93	–11.73	0.57
266	.454	.489	.0636	246.94	120.85	59.15
209	.253	.258	.0715	195.52	50.54	13.06
314	.160	.162	.0589	288.73	46.71	7.56
881	.477	.519	.0351	810.71	420.41	218.01
517	.469	.509	.0455	482.45	245.37	124.79
228	.572	.651	.0702	202.92	132.06	85.94
				7,152.21	2,764.36	1,359.60

وهذا الجدول يتضمن:

- حجم العينة لكل دراسة Sample size.
- معامل ارتباط بيرسون وهو مؤشر لحجم التأثير Effect Size.
- تحوير فيشر لكل معامل ارتباط Z_r.

- الخطأ المعيارى لمعامل التحوير SE_{zr} لفيشر فمثلًا الدراسة الأولى حجم العينة 228 وعليه فإن الخطأ المعيارى لتحوير فيشر كالتالى:

$$SE_{zr} = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{228-3}} = 0.0693$$

وهكذا لكل دراسة.

- الوزن لكل حجم التأثير w_i وهى كالآتى:

$$w_{study1} = \frac{1}{SE_i^2} = \frac{1}{(0.0693)^2} = \frac{1}{0.00480249} = 208.2253$$

وهكذا لكل دراسة.

- حاصل ضرب وزن الدراسة فى حجم التأثير الدراسة $w_i * z_r$ فالدراسة الأولى:

$$w_1 * z_{r1} = 208.2253 \times 0.583 = 121.3953$$

- حاصل ضرب مربع حجم التأثير للدراسة فى وزن $w_i * z_r^2$ فالدراسة الأولى:

$$w_1 * z_{r1}^2 = 208.2253 \times 0.583^2 = 70.7735$$

وهذه الدراسات تم تصحيحها من محددين أساسين فى هذه الدراسات وهى التصنيفية الثنائية الاصطناعية وعدم الثبات إذا كان مناسب وفى ضوء البيانات المتاحة من الدراسات.

حساب المتوسط الموزون لأحجام التأثير Weighted Mean Effect Size

تتعدد مؤشرات النزعة المركزية فمنها **المنوال** Mode وهو القيمة الاكثر حدوثًا أو شيوعًا ولا يفضل على الاطلاق استخدامه فى تحليلات ما وراء التحليل لعرض القيمة النموذجية المنتصفية لأحجام التأثير. أما **الوسيط** Median وهو القيمة المنتصفية بعد ترتيب القيم تنازليًا أو تصاعديًا من السهل تحديدها لأنها تعتمد على قيمة واحدة فى التوزيع إذا كان عدد القيم فردى أو قيمتين إذا كان عدد القيم زوجى. فالوسيط فى بيانات الجدول السابق لأحجام التأثير r = 0.35 ومن أهم مميزات الوسيط عدم تأثره بالالتواء لأحجام التأثير لأنه يستبعدها من طرفى التوزيع وهذا عكس المتوسط الاكثر تأثرًا بالالتواء. ولكن نظرًا لاعتماد الوسيط على قيمة أو قيمتين فلا يأخذ فى اعتباره كل

أحجام التأثير الموزونة للدراسات ولذلك يعتبر الوسيط مؤشر مكملًا للمتوسط. ومقياس النزعة المركزية الاكثر استخدامًا في دراسات ما وراء التحليل **المتوسط** Mean وهو مجموع القيم مقسومًا على عددها وهو يتضمن كل أحجام التأثير الموزونة.

ويقدر المتوسط الموزون لحجم التأثير للدراسات المختلفة من الاوزان w_i واحجام التأثير ES_i من كل دراسة i كالآتى (Card, 2012; Lipsey & Wilson, 2002):

$$\overline{ES}(M_{ES}) = \frac{\sum(w_i ES_i)}{\sum w_i}$$

- w_i وزن كل دراسة i .
- ES_i حجم التأثير لكل دراسة i.
- $\overline{ES}$ او M_{ES} متوسط حجم التأثير.

وعليه فإن متوسط حجم التأثير للدراسات هو مجموع حاصل ضرب حجم التأثير لكل دراسة في وزنها مقسومًا على مجموع اوزان الدراسات. لو كان وزن كل الدراسات مساويًا للواحد الصحيح أو كان وزن كل الدراسات متساوى فإن المتوسط يكون مجموع أحجام التأثير مقسومًا على عددها.

والمعادلة السابقة تطبق على أى نوع من أحجام التأثير سواء معامل الارتباط وتحويره Z_r حيث يقدر المتوسط الموزون لتحوير فيشر ثم يحول مرة اخرى إلى معامل الارتباط، ومؤشرات فروق المتوسطات المعيارية مثل g, d ومؤشر Odds Ratio.

مثال: فى الجدول السابق المتضمن 22 دراسة يتم حساب المتوسط الموزون كالآتى:

$$\overline{Z_r} = \frac{\sum(w_i Z_{ri})}{\sum w_i}$$

$$= \frac{\sum(.583 \times 208.12) + .201 \times 297.81 + \cdots + .509 \times 482.45 + .651 \times 202.92}{\sum(208.12 + 297.81 + \cdots + 482.45 + 202.92)}$$

$$= \frac{2764.36}{7152.21} = 0.387$$

وعليه فإن معامل ارتباط بيرسون المقابل لقيمة متوسط معامل التحوير فيشر $\overline{Z_r}$ كالتالى:

$$r = \frac{e^{(2\times 0.387)} - 1}{e^{(2\times 0.387)} + 1} = 0.368$$

وعليه يوجد معامل ارتباط موجب من الدرجة المتوسطة بين العدوان ورفض الاقران.

اختبارات الدلالة الإحصائية وفترات الثقة لمتوسط حجم التأثير

حساب متوسط حجم التأثير ليس كافى انما لا بد من اختبار الدلالة الإحصائية وتقدير فترات الثقة حول هذا المتوسط التى تمثل مدى من أحجام التأثير فى المجتمع من المحتمل ان تقع فيه قيمة المتوسط المحسوبة من الدراسة. على سبيل المثال95% فترات الثقة من 0.05 إلى *0.45* حول حجم التأثير تشير إلى ان احتمال 95% ان متوسط حجم التأثير فى المجتمع يقع بين هاتين القيمتين، وإذا لم تتضمن فترات الثقة القيمة صفر بالتالى فإن متوسط حجم التأثير دال إحصائيًا عند مستوى دلالة إحصائية *0.05* لـ *95%* فترة ثقة.

ولا تتم اختبارات الفروض الصفرية أو الدلالة الإحصائية بدون حساب الخطأ المعيارى لمتوسط حجم التأثير ويقدر كالآتى(Hedges & Olkin, 1985):

$$SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w_i}}$$

- w_i وزن كل دراسة

مثال: فى الجدول السابق يقدر الخطأ المعيارى كالآتى:

$$SE_{\overline{Z}_r} = \sqrt{\frac{1}{\sum w_i}} = \sqrt{\frac{1}{7152.21}} = 0.118$$

لاحظ أن الدراسات لو تضمنت أحجام عينات كبيرة بالتالى فإن مجموع w_i يكون كبير بالتالى فالخطأ المعيارى يكون صغير والعكس صحيح.

وبعد حساب الخطأ المعيارى يعتمد عليه الباحث فى إجراء الاستدلال إحصائى وحساب فترات الثقة من خلال اختبار ولد Wald أو اختبار Z (لا تخلط بين اختبار Z للدلالة

الإحصائية وتحوير فيشر Z_r) من خلال قسمة قيمة المعلم (المتوسط الموزون) مقسومًا على الخطأ المعيارى كالآتى:

$$Z = \frac{\overline{ES}}{SE_{\overline{ES}}}$$

- $\overline{ES}$ متوسط حجم التأثير لكل الدراسات.
- $SE_{\overline{ES}}$ الخطأ المعيارى لمتوسط حجم التأثير.

وهذا الاختبار يتم تقييمه وفقًا للتوزيع الاعتدالى المعيارى لاختبار Z وقيمتها الجدولية أو الحرجة عند مستوى دلالة إحصائية α = 0.05 هى 1.96 (p ≤ 0.05) ولمستوى دلالة إحصائية 0.01 هى 2.58 (p ≤ 0.01)، وعليه لو كانت قيمة Z المحسوبة أكبر من القيمة الجدولية أو الحرجة بالتالى توجد دلالة إحصائية بمعنى أن العلاقة أو الفروق بين المتغيرين دالة إحصائيًا.

ويكون الفرض الصفرى والبديل فى هذه الحالة:

$$H_0: \overline{ES} = 0$$

$$H_A: \overline{ES} \neq 0$$

- $\overline{ES}$ متوسط حجم التأثير فى المجتمع.

وتقدر فترات الثقة من الخطأ المعيارى كالآتى (Hedges & Olkin, 1985):

$$\overline{ES}_{lower} = \overline{ES} - Z_{(1-\alpha)}SE_{\overline{ES}}$$

$$\overline{ES}_{upper} = \overline{ES} + Z_{(1-\alpha)}SE_{\overline{ES}}$$

- $\overline{ES}_{lower}$ حد الثقة الادنى.
- $\overline{ES}_{upper}$ حد الثقة الأعلى.
- $\overline{ES}$ متوسط حجم التأثير.
- $Z_{(1-\alpha)}$ القيمة الجدولية عند 0.05 أو 0.01 لاختبار ذو ذيلين.
- $SE_{\overline{ES}}$ الخطأ المعيارى لمتوسط حجم التأثير.

مثال: بالرجوع إلى جدول البيانات فإن قيمة اختبار Z كالآتى:

$$Z = \frac{\overline{ES}}{SE_{\overline{ES}}} = \frac{0.387}{0.118} = 32.70, \mathrm{P} < .001$$

بالتالى فإن متوسط حجم التأثير دال إحصائيًا بمعنى أكبر من الصفر (القيمة فى الفرض الصفرى) لأن القيمة المحسوبة 32.70 أكبر من القيمة الجدولية 1.96، وعليه يوجد معامل ارتباط موجب من النوع المتوسط ودال إحصائيًا بين العدوان ورفض الاقران.

ولحساب 95% فترات الثقة كالآتى:

$$\overline{ES}_{lower} = 0.387 - 1.96 \times 0.0118 = 0.363$$

$$\overline{ES}_{upper} = 0.387 + 1.96 \times 0.0118 = 0.410$$

ثم تحول قيمتى فترتى الثقة لتحوير فيشر إلى معامل الارتباط المقابل لتصبح الحد الأدنى 0.348 = r وهكذا للحد الأعلى 0.388 = r، بالتالى فترتى الثقة95% 0.348 0.388 ,)، وعليه فإن القيمة صفر المفترضة فى الفرض الصفرى لا تقع فى مدى فترات الثقة وعليه توجد دلالة إحصائية.

وتكتب نتائج ما وراء التحليل فى الدراسة كالآتى: اتضح من نتائج ما وراء التحليل لـ 22 دراسة تناولت العلاقة بين العدوان والرفض ان متوسط معامل الارتباط 0.368 وهى دالة إحصائيًا حيث انها أكبر من الصفر (01. > P)، وان 95% فترات الثقة حول متوسط معامل الارتباط تراوحت من 0.348 إلى 0.388

اختبار التجانس بين أحجام التأثير Homogeneity among Effect Sizes

السؤال المهم المطروح هو ما إذا كانت أحجام التأثير المختلفة التى قدر منها المتوسط العام تقدر نفس حجم التأثير فى المجتمع؟ وهذا السؤال يهتم بمدى توافر التجانس بين أحجام التأثيرات المختلفة. فى التوزيع المتجانس فإن تشتت أو اختلاف أحجام التأثير حول متوسطها لا يكون أكبر مما توقعناه من خطأ المعاينة المرتبط بعينات الأفراد فى الدراسات الأولية التى قدر منها أحجام التأثير، وعليه فإن التوزيع المتماثل لأحجام التأثير للدراسات لا يختلف عن متوسط حجم التأثير الحقيقى فى المجتمع.

الاختبار الإحصائى الذى يرفض الفرض الصفرى للتجانس يشير إلى ان الاختلاف أو عدم التجانس بين أحجام التأثير أكبر مما توقعناه من خطأ المعاينة، وعليه فإن حجم التأثير لا يقدر متوسط حجم التأثير العام فى المجتمع، بكلمات اخرى وجود فروق أو انحرافات بين أحجام التأثير فإنها لا ترجع فقط إلى خطأ المعاينة بل ترجع إلى مصادر اخرى ربما تكون مرتبطة بخصائص الدراسات المختلفة فى هذه الحالة لا يوجد تجانس ولكن تباين واختلافات.

انظر إلى الشكل (1.22) تلاحظ أن كل الدراسات لها حجم تأثير واحد متضمنًا فى الخط الراسى الذى يعرض حجم التأثير العام فى المجتمع فى هذا الموقف يوجد تجانس وان أى ابتعاد لحجم التأثير فى الدراسة(الدائرة) يكون متوقع نتيجة خطأ المعاينة وليس لأى مصدر آخر وفى هذا الموقف يوجد تجانس. ولكن وجود اختلافات أكبر من خطأ المعاينة فى هذه الحالة لا يصلح الخط المستقيم فى الشكل (1.22) لعرض أحجام التأثير انما توجد عدة بدائل للتعامل مع هذا التوزيع غير المتجانس نتناولها لاحقًا.

وتقيم الدلالة الإحصائية للتجانس لأحجام التأثير من خلال إحصاء Q ويسمى اختبار Hedges Q test للتجانس وهذا الاختبار يعرض مدى الاختلاف فى أحجام التأثير بين الدراسات المختلفة المتضمنة فى ما وراء التحليل ويقدر من كالتالى (Card, 2012; Hedges & Olkin, 1985; Lipsey & Wilson, 2001):

$$Q = \sum w_i \, (ES_i - \overline{ES})^2 = \sum (w_i ES_i^2 - \frac{(\sum (w_i ES_i)^2}{\sum w_i}$$

- w_i وزن الدراسة i.
- ES_i حجم التأثير للدراسةi .
- $\overline{ES}$ متوسط حجم التأثير الموزون العام لكل الدراسات.

درجات الحرية لاختبار Q:

$$df = k-1$$

- K عدد الدراسات فى ما وراء التحليل.

ويعتبر مربع انحراف حجم التأثير لكل دراسة عن متوسط حجم التأثير لكل الدراسات (الجزء الأول من المعادلة) هو أفضل تقدير لحجم التأثير فى المجتمع وهو الخط الراسى فى الشكل (1.22).

ولكن كيف يمكن استنتاج وجود تجانس من عدمه؟ يعامل إحصاء Q فى ضوء توزيع كاى تربيع χ^2 برجات حرية k-1 وعلى ذلك يجب البحث فى جدول χ^2 لاستخراج قيمتها الجدولية وإذا كانت:

χ^2 الجدولية ≥ قيمة Q المحسوبة

نرفض الفرض الصفرى بمعنى عدم وجود تجانس أو وجود اختلاف بين أحجام التأثير أكثر مما توقعناه نتيجة خطأ المعاينة.

مثال: فى الجدول (1.22) تم حساب:

$$\sum w_i = 7152.21$$

$$\sum w_i ES_i^2 = 1359.60$$

$$\sum w_i ES_i = 2764.36$$

وبالتعويض فى المعادلة:

$$Q = \sum (w_i ES_i^2 - \frac{(\sum (w_i ES_i)^2}{\sum w_i} = 1359.60 - \frac{(2764.36)^2}{7152.21} = 291.17$$

حيث درجات الحرية: $df = k - 1 = 22 - 1 = 21$ وبالبحث فى جدول χ^2 الاتى:

الجدول(2.22): القيم الحرجة لـ χ^2 لدرجات الحرية ومستوى الدلالة الإحصائية.

df	p = .1	p = .05	p = .01	p = .005	p = .001	df	p = .1	p = .05	p = .01	p = .005	p = .001
1	2.706	3.841	6.635	7.879	10.828	45	57.505	61.656	69.957	73.166	80.077
2	4.605	5.991	9.210	10.597	13.816	50	63.167	67.505	76.154	79.490	86.661
3	6.251	7.815	11.345	12.838	16.266	55	68.796	73.311	82.292	85.749	93.168
4	7.779	9.488	13.277	14.860	18.467	60	74.397	79.082	88.379	91.952	99.607
5	9.236	11.070	15.086	16.750	20.515	65	79.973	84.821	94.422	98.105	105.988
6	10.645	12.592	16.812	18.548	22.458	70	85.527	90.531	100.425	104.215	112.317
7	12.017	14.067	18.475	20.278	24.322	75	91.061	96.217	106.393	110.286	118.599
8	13.362	15.507	20.090	21.955	26.124	80	96.578	101.879	112.329	116.321	124.839
9	14.684	16.919	21.666	23.589	27.877	85	102.079	107.522	118.236	122.325	131.041
10	15.987	18.307	23.209	25.188	29.588	90	107.565	113.145	124.116	128.299	137.208
11	17.275	19.675	24.725	26.757	31.264	95	113.038	118.752	129.973	134.247	143.344
12	18.549	21.026	26.217	28.300	32.909	100	118.498	124.342	135.807	140.169	149.449
13	19.812	22.362	27.688	29.819	34.528	110	129.385	135.480	147.414	151.948	161.581
14	21.064	23.685	29.141	31.319	36.123	120	140.233	146.567	158.950	163.648	173.617
15	22.307	24.996	30.578	32.801	37.697	130	151.045	157.610	170.423	175.278	185.571
16	23.542	26.296	32.000	34.267	39.252	140	161.827	168.613	181.840	186.847	197.451
17	24.769	27.587	33.409	35.718	40.790	150	172.581	179.581	193.208	198.360	209.265
18	25.989	28.869	34.805	37.156	42.312	160	183.311	190.516	204.530	209.824	221.019
19	27.204	30.144	36.191	38.582	43.820	170	194.017	201.423	215.812	221.242	232.719
20	28.412	31.410	37.566	39.997	45.315	180	204.704	212.304	227.056	232.620	244.370
21	29.615	32.671	38.932	41.401	46.797	190	215.371	223.160	238.266	243.959	255.976
22	30.813	33.924	40.289	42.796	48.268	200	226.021	233.994	249.445	255.264	267.541
23	32.007	35.172	41.638	44.181	49.728	210	236.655	244.808	260.595	266.537	279.066
24	33.196	36.415	42.980	45.559	51.179	220	247.274	255.602	271.717	277.779	290.556
25	34.382	37.652	44.314	46.928	52.620	230	257.879	266.378	282.814	288.994	302.012
26	35.563	38.885	45.642	48.290	54.052	240	268.471	277.138	293.888	300.182	313.437
27	36.741	40.113	46.963	49.645	55.476	250	279.050	287.882	304.940	311.346	324.832
28	37.916	41.337	48.278	50.993	56.892	300	331.789	341.395	359.906	366.844	381.425
29	39.087	42.557	49.588	52.336	58.301	350	384.306	394.626	414.474	421.900	437.488
30	40.256	43.773	50.892	53.672	59.703	400	436.649	447.632	468.724	476.606	493.132
35	46.059	49.802	57.342	60.275	66.619	450	488.849	500.456	522.717	531.026	548.432
40	51.805	55.758	63.691	66.766	73.402	500	540.930	553.127	576.493	585.207	603.446

وبالبحث فى الجدول السابق بدرجات حرية 21 وبمستوى دلالة إحصائية مثلًا 0.01 فإن ^{2} الجدولية= 38.932 اذًا:

^{2} الجدولية(38.932) < Q المحسوبة(291.71) وعليه يتم رفض الفرض الصفرى بمعنى وجود اختلاف بين أحجام التأثير أى عدم تجانس بينهم بمعنى أن أحجام التأثير كلها لا تقدر حجم التأثير الحقيقى فى المجتمع وعلى ذلك فإن أحجام التأثير اختلفت عن حجم التأثير الحقيقى اختلافًا أكبر مما توقعناه نتيجة خطأ المعاينة. ويشير Card (2012) إلى تحذيرين عند تفسير نتائج Q وهما:

1. هذا الاختبار يمدنا بمعلومات عن احتمال وجود أو عدم وجود تجانس فقط لأحجام التأثير ولا يخبرنا عن مقدار Magnitude هذا الاختلاف أو عدم التجانس Heterogeneity .

2. يجب الاخذ فى الاعتبار القوة الإحصائية لهذ الاختبار ونقص هذه القوة للاختبار فيجب أن يكون الباحث على حذر فى تفسير النتائج غير الدالة إحصائيًا كدليل للتجانس (الفرض الصفرى).

بدائل اختبار التجانس بين أحجام التأثير

مدخل Hunter & Schmidt

افترض مدخل بديل لاختبار عد التجانس بين أحجام التأثير لا يعتمد على اختبارات الدلالة. هذا المخل يجزئ تباين حجم التأثير المقاس من الدراسة إلى مكونين هما جزء يعزى إلى خطأ المعاينة لعينات الأفراد فى الدراسة والثانى يعزى إلى الفروق بين الدراسات. يفترض ان يكون التوزيع متجانس لو ان 75% أو أكثر من الاختلافات بين أحجام التأثير ترجع إلى خطأ المعاينة وهذه قاعدة افتراضية متعارف عليها وليست نقطة قطع صارمة.

مؤشر I^2

على الرغم ان إحصاء Q والدلالة الإحصائية المرتبطة به من المؤشرات المفيدة لتحديد مدى وجود تجانس من عدمه بين أحجام التأثير فى ما وراء التحليل الا أنه لا يشير إلى كيفية ان يكون هذا الاختلاف بين أحجام التأثير. ومن المؤشرات المفيدة لتحديد التجانس فى ما وراء التحليل مؤشر I^2 ويفسر كنسبة مئوية للاختلاف بين أحجام التأثير الموجود بين الدراسات بالنسبة إلى الاختلاف الكلى بين أحجام التأثير (Higgins & Thompson, 2002; Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006)، ويتم تقديره كالآتى:

$$I^2 = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{Q - (k-1)}{Q} \times 100 \ , when\ Q > (k-1)$$

- τ^2 الاختلاف المقدر بين الدراسات.
- σ^2 الاختلاف داخل الدراسات.
- Q الإحصاء المستخدم لحساب الدلالة الإحصائية للتجانس.

- k عدد الدراسات.

والجزء الايمن من المعادلة يعتمد على إحصاء Q وعدد الدراسات k وقيمة I^2 تكون تقريبًا عند الصفر عندما تكون Q أقل من المتوقع تحت الفرض الصفرى للتجانس. والمقام مكون من Q وهو عبارة عن التباين الكلى بين أحجام التأثير بينما البسط يتكون من الاختلاف الكلى بين أحجام التأثر مطروحًا منه الاختلاف المتوقع جراء خطأ المعاينة.

مؤشر I^2 هو النسبة بين الاختلاف بين الدراسات(الاختلاف الكلى مطروحًا من الاختلاف نتيجة خطأ المعاينة) بالنسبة إلى الاختلاف الكلى بين الدراسات ويعبر عنه فى ضوء نسبة مئوية على مقياس من 0.0% إلى 100% وعلى ذلك فإن هذا المؤشر يعبر عن مقدار الاختلاف بين أحجام التأثير فى ما وراء التحليل ويستخدم لمقارنة الاختلافات عبر الدراسات المختلفة لما وراء التحليل، وهذا المؤشر أقل استخدامًا فى دراسات ما وراء التحليل كما ليس له معايير متفق عليها لتحديد مقدار الاختلاف سواء صغير أو متوسط أو كبير، ولكن Huedo-Medin et al. (2006) اقترحوا المعايير الآتية:

$I^2 \approx 0\%$ تجانس بين أحجام التأثير.

$I^2 \approx 25\%$ قدر صغير من الاختلاف بين أحجام التأثير.

$I^2 \approx 50\%$ قدر متوسط من الاختلاف.

$I^2 \approx 75\%$ قدر كبير من الاختلاف.

مثال: يقدر مؤشر I^2 لبيانات لـ 22 دراسة كالآتى:

$$I^2 == \frac{Q-(k-1)}{Q} \times 100 = \frac{291.17-(22-1)}{291.17} = 92.8\%$$

وهذا يعبر عن اختلافات كبيرة جدًا بين أحجام التأثير.

عدم الاستقلالية بين أحجام التأثير

فـى هـذا الفصـل عرضـنا لأهميـة الاسـتقلالية لأحجـام التـأثير ومـن المفضـل الاعتمـاد علـى حجـم تـأثير واحـد مـن الدراسـة لأن حسـاب أكثـر مـن حجـم تـأثير مـن نفـس الدراسـة يحطـم مسـلمة الاسـتقلالية. فـى مـا وراء التحليـل يجـب الاعتمـاد علـى العشـوائية بقـدر الامكـان فـى عينـة الدراسـات المتضـمنة فـى مـا وراء التحليـل نظـرًا لاسـتخدام اختبـارات الفـروض الإحصائية حول المتوسط الموزون لأحجام التأثير.

احجام التأثير المتعددة من القياسات المتعددة

أحـد مصـادر أحجـام التـأثير فـى الدراسـة ان الدراسـة يمكـن أن تتضـمن مقـاييس مختلفـة لـنفس المفهـوم أو البنـاء فـى الدراسـة، فمـثلًا فـى الدراسـات الـ 22 اعتمـد أحـد المـؤلفين علـى العلاقـة بـين العـدوان والـرفض مـن الاقـران مـن وجهـة نظـر زملائـه(r =.556) وايضًـا مـن وجهـة نظـر المعلمـين(r =.338)، واحـد الدراسـات اعتمـدت علـى دراسـة هـذه العلاقـة علـى فتـرات زمنيـة مختلفـة مـثلًا علـى مسـتوى تلاميـذ الصـف الثـانى وعلـى مسـتوى تلاميـذ الصف الرابع.

فى هذه المواقف يرى Card (2012) يوجد اختيارين للحصول على حجم تأثير وحيد:

1. اختيـار حجـم التـأثير الـذى يقـع فـى اهتمـام الدراسـة التـى يقـوم بهـا الباحـث وهـذا فـى ضـوء معـايير التضـمين والاسـتبعاد التـى اعتمـد عليهـا الباحـث ففـى الدراسـة الأولـى يضـع الباحـث معيـار تقـدير العلاقـة فـى ضـوء تقـديرات الاقـران ولـيس تقـدير المعلـم. ومـن الضـرورى وضـع المعيـار لصـناعة القـرار قبـل إجـراء مـا وراء التحليـل ولـيس فـى ضـوء حجم التأثير.

2. وهـو الاكثـر اسـتخدامًا فـى الدراسـات وهـو اخـذ متوسـط أحجـام التـأثير المتعـددة فـى نفـس الدراسـة لـنفس المتغيـرين موضـع الاهتمـام واسـتخدام متوسـط أحجـام التـأثير كمؤشـر وحيـد لحجم التأثير المتولد من الدراسة.

احجام التأثير المتعددة من مجموعات فرعية من المشاركين فى الدراسة

المصدر الآخر لأحجام التأثير المتعددة من نفس الدراسة هو أحجام التأثير من المجموعات الفرعية من الدراسة (ذكور – إناث، المجموعات التجريبية المتعددة) الأمر يتوقف على معايير التضمين أو الاستبعاد هل تهتم الدراسة بالعينة الاجمالية أم العينات الفرعية، فإذا كان الاهتمام بالعينات الفرعية بالتالى يجب حساب متوسط أحجام التأثير من هذه المجموعات الفرعية ويجد اعتبارين فى هذه الحالة هما:

1. عند حساب متوسط المجموعات الفرعية فيجب الاعتماد على مجموع العينات للمجموعتين عند حساب الخطأ المعيارى فى الدراسة.

2. يجب حساب المتوسط الموزون لأن العينات الفرعية الكبيرة تسهم بقدر أكبر فى المتوسط من العينات الصغيرة.

أحجام التأثير من التقارير العديدة لنفس الدراسة

المصدر الثالث لعدم الاستقلالية لأحجام التأثير فى ما وراء التحليل هو ان نفس الدراسة يعاد نشرها أكثر من مرة فقد تكون رسالة دكتوراه أو ماجستير ثم يعاد نشرها فى كتاب أو دراسة فى مجلة علمية متخصصة، وقد تكون دراسة معروضة فى مؤتمر ثم يعاد نشرها فى مجلة متخصصة أو تختلف أحجام التأثير فى نفس الدراسة نتيجة لأن الباحث قد يعتمد على مقياس فى الدراسة الأولى مختلف عن مقياس نفس الدراسة فى مكان النشر الآخر، فى هذه الحالة يجب اخذ متوسط أحجام التأثير لنفس الدراسة ذات المقاييس المختلفة ومكان النشر المختلف وهذ نادر الحدوث فى دراسات ما وراء التحليل.

نماذج تحليل ما وراء التحليل Analysis Models for Meta-analysis

فى العرض السابق تناولنا مدخل لكيفية حساب متوسط حجم التأثير واختبار الدلالة الإحصائية وحساب فترات الثقة حول المتوسط وهذا المدخل يفترض التجانس بين أحجام التأثير بمعنى أن الاختلافات بين أحجام التأثير المقدرة من الدراسات وحجم التأثير الحقيقى فى المجتمع ترجع إلى فقط خطأ المعاينة، وهذا المدخل فى تحليل ما

وراء التحليل يسمى نموذج التأثيرات المثبتة Fixed-effects model ويسميه Hedges (2009) بالنموذج الاستدلالى الشرطى Conditional Inference Model ولا يفضل استخدام هذا النموذج إذا ثبت ان الفروق أو الاختلافات بين أحجام التأثير ترجع إلى المتغيرات المتفاعلة Moderator variables أو خصائص الدراسات بالإضافة إلى خطا المعاينة. وإذا وجدت اختلافات بين أحجام التأثير أكبر مما نتوقعه من خطا المعاينة ونتيجة لخصائص الدراسات فإن نموذج تحليل ما وراء التحليل يسمى نموذج التأثيرات العشوائية Random effects mode ويسمى النموذج غير الشرطى Unconditional Model.

وعرضنا اختبار Q لفحص مسلمة التجانس بين أحجام التأثير المقاسة من الدراسات المختلفة وإذا تحققت يستخدم نموذج التأثيرات المثبتة ولكن الدلالة الإحصائية تعنى رفض الفرض الصفرى بمعنى اختلافات بين أحجام التأثير أكبر مما توقعناه جراء اخطأ المعاينة فى الدراسات، وهذا الاختلاف يلقى بتحدى كبير على استخدام النموذج المثبت لأنه عند حساب المتوسط الموزون فى الفصل السابق حدث موازنة لكل حجم تأثير فى ضوء الخطأ المعيارى الذى يأخذ فى اعتباره اخطاء المعاينة فقط وهذا كافى لإجراء التحليلات الإحصائية. ولكن فى ظل عدم التجانس بين أحجام التأثير نتيجة مصدر آخر للاختلافات غير خطأ المعاينة بالتالى إجراء موازنة فى ضوء الخطأ المعيارى غير كافى فلا بد من إجراء موازنة لمصادر الاختلافات الأخرى بين أحجام التأثير للدراسات، وعليه فإن نموذج التأثيرات المثبتة غير ملائم لتحليل توزيع أحجام التأثيرات غير المتجانسة. كما ان تحليلات ما وراء التحليل تتعامل مع عينة صغيرة نسبيًا من أحجام التأثيرات أو الدراسات تتراوح فى معظمها من 15 دراسة إلى 25 دراسة وهذا من شانه ان تكون القوة الإحصائية لاختبار Q غير كافية لرفض الفرض الصفرى حتى لو وجدت مصادر اختلاف بين الدراسات غير خطأ المعاينة وعليه فاذا لم تتحقق مسلمة نموذج التأثيرات المثبتة فإن الباحث امامه ثلاث اختيارات حددها Lipsey & Wilson (2001) بالتالى:

الأول: افترض أن الاختلافات بين أحجام التأثير ترجع إلى مصادر اختلاف عشوائية اخرى غير خطأ المعاينة لعينات الأفراد فى الدراسات وهذه المصادر لا يمكن تحديدها. وفى هذه الحالة يتبنى الباحث **نموذج التأثيرات العشوائية** Random effects model حيث يفترض ان التباين أو الاختلاف بين أحجام التأثير يرجع إلى مكون أو خطأ عشوائى آخر غير خطأ المعاينة ويجب تضمين هذا الخطأ إلى النموذج الإحصائى لتحليل أحجام التأثيرات، وهذا الخطأ الاضافى يتصرف مثل خطأ المعاينة حيث يؤدى إلى ابتعاد أو انحراف حجم التأثير المقاس عن حجم التأثير الحقيقى فى المجتمع وعلية فإن النموذج العشوائى يتعامل مع مصدرين للخطأ هما:

- خطا معاينة على مستوى الدراسات Study Level حيث إن عينة الدراسات غير ممثلة تمثيلًا جيدًا لمجتمع دراسات الظاهرة التى تتم عليها دراسات ما وراء التحليل.
- خطأ معاينة على مستوى الأفراد أو الاشخاص Subject level حيث إن عينة الأفراد فى كل دراسة غير ممثلة تمثيلًا جيدًا للمجتمع.

ولذلك لا بد من إجراء تصحيح للمتوسط مرتين وليس مرة واحدة كما عرضنا فى الفصل السابق (من خطأ معاينة الأفراد) ولا بد من تصحيح كل حجم تأثير من خطأ المعاينة للأفراد وخطأ المعاينة للدراسات، وهذا يتطلب إجراءات إحصائية مختلفة عما عرضناه فى تحليل نموذج التأثيرات المثبتة فى الفصل السابق متمثلة فى نموذج التأثيرات العشوائية.

الثانى: استخدم **نموذج التأثيرات المثبتة** وافترض مسلمة هو ان الاختلاف غير خطأ معاينة الأفراد يكون منتظم Systematic مشتق من فروق غير محددة بين الدراسات، وهذا يتطلب تحليل الاختلاف بين أحجام التأثير فى ضوء خصائص الدراسات التى تولدت منها أحجام التأثير وفى هذه الحالة يفترض الباحث ان وجود اختلافات فى توزيع أحجام التأثيرات نجمت عن مصادر غير اخطاء المعاينة للأفراد ويكون هذا الاختلاف غير عشوائى والاختلافات الكبيرة بين أحجام التأثير ترجع إلى المتغيرات المتفاعلة مثل الخصائص المنهجية(الضبط – المعالجات المختلفة– التصميمات

البحثيـة- الإجـراءات- حجـم العينـة- نوعيـة العينـة، وغيرهـا) وكـذلك خصـائص العينـة وهـذه العوامـل تجعـل حجـم التـأثير مـنخفض أو كبيـر فـى الدراسـات. وفـى هـذه الحالـة إجـراء موازنـة Weighting لحجـم التـأثير فـى ضـوء خطـأ المعاينـة للأفـراد يكـون مقبـول ومناسـب. والإجـراءات الإحصـائية لدراسـة الاخـتلاف بـين أحجـام التـأثير للدراسـات المختلفة يسمى **نمـوذج التـأثيرات المثبتـة: تجزئـة تبـاين أحجـام التـأثير** Fixed effects model: Partitioning effect size variance

الثالـث: افتـرض أن التبـاين بـين أحجـام التـأثير يرجـع إلـى مصـادر اخـرى غيـر خطـأ المعاينـة هـى مكونـة مـن مصـدرين احـدهما عوامـل منتظمـة يمكـن تحديـدها (خصـائص الدراسـات) والآخـر مصـادر عشـوائية، و لإجـراءات الإحصـائية لتحليـل مـا وراء التحليـل تسمى **نمـوذج التـأثيرات المختلطـة** Mixed effects model حيث إن الاختلاف بين أحجـام التـأثير لا يرجـع إلـى خطـأ المعاينـة فقـط بـل إلـى مكـونين اخـرين احـدهما منتظم(خصـائص الدراسـات) والآخـر عشوائى(خصـائص معاينـة الدراسـات مـن مجتمـع الدراسـات التـى تناولـت العلاقـة بـين المتغيـرين). وهـذا النمـوذج يتشـابه مـع نمـوذج التـأثيرات المثبتـة: تجزئـة تبـاين أحجـام التـأثير فـى تحليـل خصـائص الدراسـات كمتغيـرات متفاعلـة ولكنـه يختلـف عنـه فـى أنـه يسـمح بتضـمين المكـون العشـوائى فـى التحليـل كتبـاين بـواقى ليصـبح فـى النمـوذج بعـد تقـدير المكـون المنـتظم، بكلمـات اخـرى يفتـرض ان تبـاين أحجـام التـأثير غيـر خطـأ المعاينـة يرجـع إلـى مكـونين احـدهما منـتظم حيـث يعـرض العلاقـة بـين خصـائص الدراسـات وحجـم التـأثير، والآخـر يمثـل الخطـأ العشـوائى نتيجـة خطـأ معاينـة للدراسـات الممثلـة لمجتمـع الدراسـات التـى تناولـت المتغيـرين ويجـب تضـمينه فـى الوظيفـة الموازنيـة Weighting function عنـد تحليـل حجـم التـأثير حيـث يعـاد حسـاب المتوسـط المـوزون وفتـرات الثقـة بعـد اسـتخدام اوزان مختلفـة ولـيس وزن واحـد كمـا فـى النموذج المثبت.

نموذج التأثيرات المثبتة Fixed effects model

مصـطلح التـأثيرات المثبتـة يشـير إلـى حقيقـة ان مسـتويات المعالجـة فـى التجربـة تعتبـر مثبتـة ومصـدر الاخـتلاف هـى عينـات أو مجموعـات الأفـراد والمتغيـر التـابع. ويفتـرض

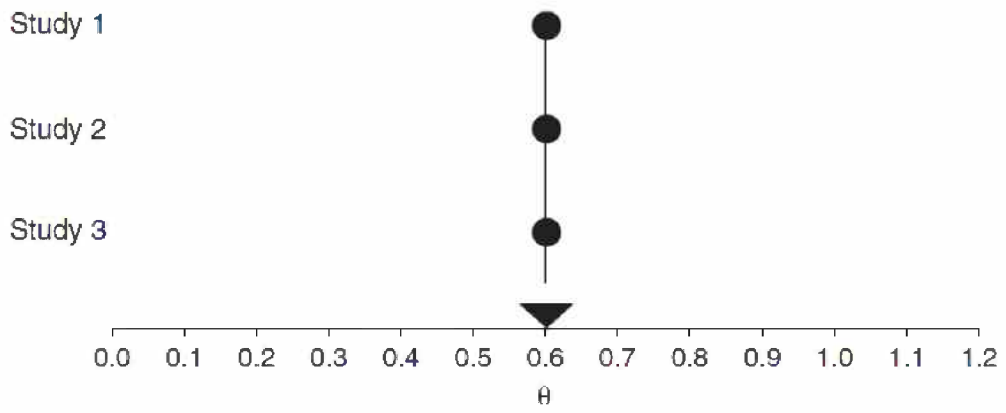
هـذا النمـوذج ان حجـم التـأثير المقـدر مـن الدراسـة ينـاظر حجـم التـأثير الحقيقـى فـى المجتمـع مـع وجـود خطـأ عشـوائى نتيجـة عوامـل الصـدفة المرتبطـة بخطـأ المعاينـة فـى الدراسـة النـاتج عـن الاختيـار غيـر العشـوائى للعينـة مـن مجتمـع الدراسـة وايضًـا نتيجـة ان العينـة غيـر ممثلـة تمثـيلًا جيـدًا لمجتمـع العينـة وكـذلك لصـغر حجـم العينـة. ولأسـباب نظريـة فربمـا لا يعتقـد الباحـث ان هـذه الافتراضـات مناسـبة أو متـوفرة لتوزيـع أحجـام التـأثير. فعلـى سـبيل المثـال يعتقـد بوجـود فـروق عشـوائية بـين الدراسـات ترجـع إلـى اختلافـات فـى الإجـراءات والمواقـع المختلفـة التـى اجريـت فيهـا الدراسـات وغيرهـا مـن العوامـل التـى تكـون ابعـد مـن خطـأ المعاينـة فـى الدراسـات، وربمـا يكـون الاختلافـات ترجـع إلـى عينـة الدراسـات فـى مـا وراء التحليـل حيـث لا تكـون ممثلـة لمجتمـع الدراسـات تمثـيلًا جيـدًا وهـذا يـؤدى إلـى خطـا معاينـة خـاص بالدراسـات بالإضـافة إلـى خطـأ المعاينـة الخـاص بعينـات الأفـراد فـى الدراسـة. ويـرى Borenstein et al. (2009) أن النمـوذج المثبـت فـى مـا وراء التحليـل يفتـرض ان أحجـام التـأثير للدراسـات تقـيس نفـس حجـم التـأثير الحقيقـى فـى المجتمـع بمعنـى أن كـل العوامـل التـى تـؤثر فـى حجـم التـأثير واحـدة فـى كـل الدراسـات ولـذلك فـإن حجـم التـأثير هـو نفسـه (مثبـت) عبـر كـل الدراسـات والاخـتلاف عـن حجـم التأثير الحقيقى يرجع إلى خطأ المعاينة.

وعلـى ذلـك فـإن نمـوذج التـأثيرات المثبتـة يفتـرض ان كـل الدراسـات فـى مـا وراء التحليـل تـأتى مـن نفـس المجتمـع وهـذا يعنـى ان حجـم التـأثير الحقيقـى يكـون هـو نفسـه عبـر كـل الدراسـات وفـى هـذه الحالـة يفتـرض وجـود تجـانس بـين أحجـام التـأثير المختلفـة وان مصـدر الاخـتلاف بـين حجـم التـأثيرات يرجـع إلـى مكونـات مـن داخـل الدراسـات مثـل خطـأ المعاينـة أو خطـأ القيـاس (Hegger, 2006). ويهـتم هـذا النمـوذج بالتبـاين داخـل الدراسـة نتيجـة اختلاف المعاينات ولا يهتم بالتباين بين الدراسات

ويفترض هذا النموذج (Kline, 2013):

- وجود مجتمع واحد متجانس من الدراسات مع حجم تأثير حقيقى وحيد.
- اخـتلاف حجـم التـأثير للدراسـة عـن حجـم التـأثير الحقيقـى فـى المجتمـع يرجـع إلى التباين داخل الدراسة بمعنى إلى خطأ المعاينة.

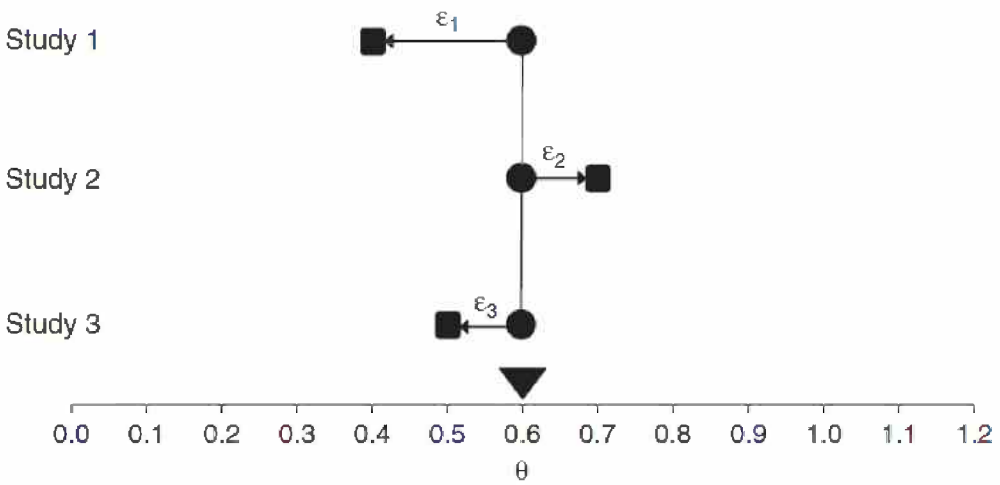
ويمكن توضيح ذلك بالشكل الاتى:



الشكل (2.22): نموذج التأثيرات المثبتة: التأثيرات الحقيقية.

فاذا افترضنا ان حجم التأثير العام أو الحقيقى فى المجتمع هو 0.60 (كما معروض فى المستطيل على المحور الافقى فى الأسفل) ويرمز له بالرمز θ وحجم التأثير الحقيقى لكل دراسة معروض على الخط الراسى(دائرة). ويفترض نموذج التأثيرات المثبت ان حجم التأثير الحقيقى يجب أن يكون 0.60 وهذا معروض فى الدوائر التى تقع مباشرة على الخط المستقيم، ولكن هذا صعب حدوثة فى الواقع التطبيقى الإمبريقى للدراسات لأن حجم التأثير يختلف من دراسة لأخرى نتيجة خطأ المعاينة فى كل دراسة فلو كان هذا الخطأ مساويًا للصفر فى الدراسة فإن حجم التأثير لها يكون مساويًا لحجم التأثير فى الحقيقى فى المجتمع.

ولكن فى الممارسة العملية فإن حجم العينة فى كل دراسة يكون محدد بحجم معين ولذلك فيوجد قدرًا من خطأ المعاينة يجعل حجم التأثير فى الدراسة ليس هو نفسه حجم التأثير الحقيقى وقد تزداد قيمته أو تنقص. ويمكن توضيح ذلك بالشكل الاتى:



الشكل (3.22): نموذج التأثيرات المثبتة فى ظل تأثير أخطاء المعاينة للدراسات.

لاحظ فى هذا الشكل ان حجم التأثير الحقيقى لكل دراسة ما زال 0.60 (الدوائر) بينما حجم التأثير المقاس(المربعات) ■ اختلف عن حجم التأثير الحقيقى من دراسة إلى اخرى.

ففى الدراسة الأولى خطأ المعاينة (ε_1)= -0.20 وعليه فإن حجم التأثير المقاس:

$$ES_1 = 0.60 - 0.20 = 0.40$$

وفى الدراسة الثانية خطأ المعاينة(ε_2)= 0.10 وعليه فإن حجم التأثير المقاس:

$$ES_2 = 0.60 + 0.10 = 0.70$$

وفى الدراسة الثالثة خطأ المعاينة(ε_3)= -0.10 وعليه فإن حجم التأثير المقاس:

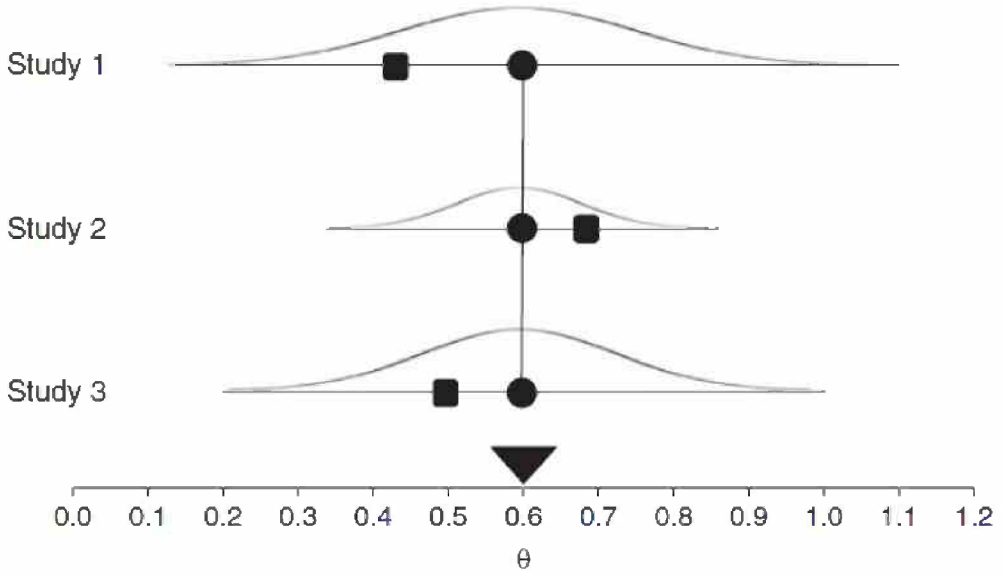
$$ES_3 = 0.60 - 0.10 = 0.50$$

بصفة عامة فإن حجم التأثير المقاس فى كل الدراسة ES_i كالآتى:

$$ES_i = \theta + \varepsilon_i$$

- θ حجم التأثير العام الحقيقى فى المجتمع.
- ε_i ابيسلون خطأ المعاينة للدراسة i.

لاحظ أن الدراسة الأولى ان حجم التأثير نقص بمقدار وحدتين وفى الدراسة الثانية زاد بمقدار وحدة وفى الدراسة الثالثة نقص بمقدار وحدة وعليه فإن الخطأ فى أى دراسة عشوائى ويمكن تقدير التوزيع العينى لهذه الأخطاء كالآتى:



الشكل (4.22): التوزيع العينى لأخطاء المعاينة فى النموذج المثبت.

وفى هذا الشكل نلاحظ أن توزيع اجحام التأثير المقاسة حول حجم التأثير الحقيقى فى ضوء المنحنى الطبيعى، ففى الدراسة الأولى حيث حجم عينة صغير فمن المتوقع ان يقع حجم التأثير المقاس على جانبى حجم التأثير الحقيقى فى مدى من 0.20 إلى 1.00، فى المقابل فى الدراسة الثانية حيث حجم العينة كبير نسبيًا فمن المتوقع ان يقع حجم التأثير المقاس فى مدى ضيق من 0.40 إلى 0.80.

نموذج التأثيرات العشوائيةRandom Effects Model

فى نموذج التأثيرات المثبتة افترضنا ان حجم التأثير الحقيقى نفسه فى كل الدراسات ولكن هذه المسلمة من الصعب تحقيقها. فلمجموعة من الدراسات المراد دمجها فى ما وراء التحليل فإنها ليست متشابهة أو متماثلة تمامًا ان تعطى نفس حجم التأثير للعلاقة بين المتغيرين، فهذه الدراسات لها منهجيات وإجراءات وتصميمات واحجام عينات مختلفة بل مقاييس مختلفة، على ذلك فإن الدراسات مختلفة فى خصائصها وقيمة حجم التأثير مختلف من دراسة إلى اخرى.

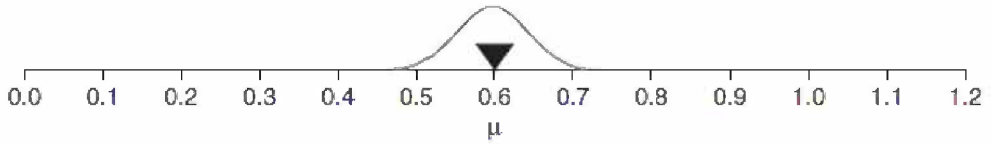
فمثلًا الدراسات التى تتناول أثر العلاج السلوكى المعرفى لتخفيف الاكتئاب فإن حجم التأثير يكون قوى فى الدراسات التى تتضمن عينة شديدة الاكتئاب مقارنة بدراسة اخرى تضمنت عينة متوسطى الاكتئاب كذلك يكون حجم التأثير كبير فى الدراسات التى تضمنت عدد من الجلسات أكبر من نظيرتها أقل جلسات، وعلى ذلك فإن الدراسات المختلفة تنتج أحجام تأثير مختلفة. وفى الدراسات التى تتضمن برامج التدخل التربوى النفسى فإن مقدار حجم التأثير يعتمد على المصادر المتاحة سواء المستوى الاقتصادى الاجتماعى للعينة أو حجم الفصل أو عمر العينة أو مستوى ذكاء الأفراد أو عوامل اخرى تختلف من دراسة إلى أخرى، وعلى ذلك ففى البرامج النفسية التربوية الاجتماعية يوجد كثيرًا من التغيرات أو المتغيرات الدخيلة التى تؤدى إلى الاختلافات بين أحجام التأثير للدراسات عبر الدراسات المختلفة التى تتناول نفس المتغيرات لنفس الظاهرة.

ويرى Borenstein et al. (2009) أن أحد الطرق للتعامل مع هذه الاختلافات عبر الدراسات هو استخدام نموذج التأثيرات العشوائية حيث يتم افتراض ان التأثيرات الحقيقية تتوزع اعتدالى وليس اجبارها كلها على خط راسى كما فى النموذج المثبت بالتالى لا يوجد حجم تأثير حقيقى لكل الدراسات انما يختلف من دراسة إلى اخرى.

ويفترض هذا النموذج Kline (2013):

- يوجد توزيع لمجتمع أحجام التأثير بمعنى وجود حجم تأثير حقيقى لكل دراسة وليس حجم تأثير حقيقى واحد لكل الدراسات.
- يوجد مصدرين لتباين الخطأ أحدهما تباين داخل الدراسات Within-studies variance نتيجة خطأ المعاينة والآخر التباين بين الدراسات Between-studies variance الذى يختص بتوزيع كل مجتمعات أحجام التأثير حول متوسط حجم التأثير العام للمجتمعات. وهذا التباين فى مجمله من مصادر عشوائية لا يمكن تحديدها.

ويمكن توضيح ذلك بالشكل التالى:



الشكل (5.22): توزيع أحجام التأثير الحقيقية فى نموذج التأثيرات العشوائية.

وفى الشكل السابق فإن متوسط كل أحجام التأثير الحقيقية 0.60 ولكن أحجام التأثير للدراسات تتوزع اعتداليًا حول هذا المتوسط وطول أو مسافة المنحنى تقترح ان معظم أحجام التأثير الحقيقية تقع فى المدى من 0.50 إلى 0.70.

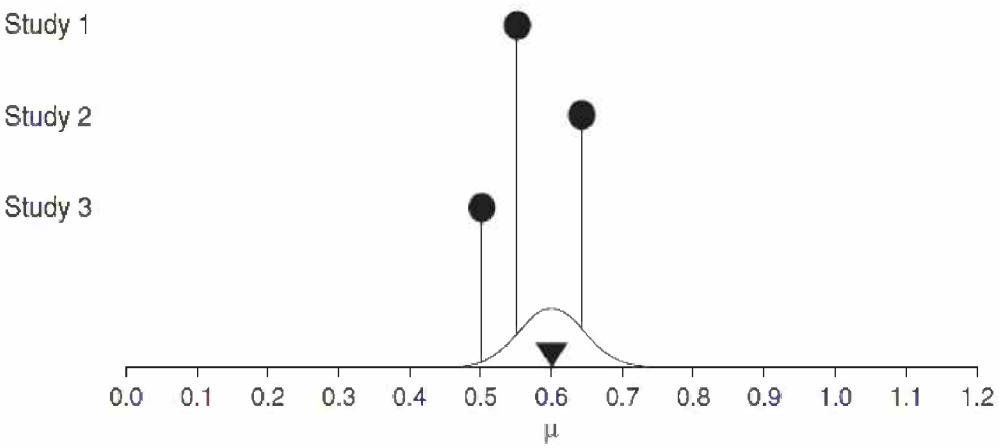
مثال: افترض أن دراسة ما وراء التحليل لثلاثة دراسات وان أحجام التأثير الحقيقية لهذه الدراسات هى:

$$\theta_1 = 0.50$$

$$\theta_2 = 0.55$$

$$\theta_3 = 0.65$$

ولو ان حجم اعينات فى الدراسات الثلاثة كبيرًا كبرًا كافيًا فإن خطأ المعاينة يكون صفر ويكون حجم التأثير المقدر هو نفسه حجم التأثير الحقيقى لهذه الدراسة ويمكن التعبير عن ذلك بالشكل الآتى (Borenstien et al., 2009):



الشكل (6.22): التأثيرات الحقيقية فى نموذج التأثيرات العشوائية.

فـى هـذا الشـكل لاحـظ أن حجـم التـأثير الحقيقـى العـام فـى المجتمـع (μ) يسـاوى 0.60(علـى الخـط الافقـى) بينمـا حجـم التـأثير الحقيقـى للدراسـة الأولـى 0.50 بالتـالى حـدث ابتعـاد لحجـم التـأثير للدراسـة الأولـى عـن حجـم التـأثير الحقيقـى العـام فـى المجتمـع بحـوالى 0.10 وهـذا يمثـل نـوع مـن الخطـأ لاحـظ أن هـذا الخطـأ لا يرجـع إلـى خطـأ المعاينة لأننا افترضنا ان خطأ المعاينة مساوى للصفر.

ولكـن حجـم العينـة فـى كـل دراسـة لـيس ممـثلًا للمجتمـع تمثـيلًا كافيًـا وعليـه فـإن خطـأ المعاينـة لا يسـاوى صـفر. فـاذا كـان حجـم التـأثير الحقيقـى للدراسـة θ_i فـإن حجـم التـأثير المقـاس فـى الدراسـة ES_i يكـون أكبـر أو أصـغر مـن حجـم التـأثير الحقيقـىθ_i نتجـة خطـأ المعاينـة. ففـى الشـكل (6.22) الدراسـة الثالثـة لهـا حجـم تـأثير حقيقـى 0.50 وخطـأ المعاينة -0.10 بالتالى حجم التأثير المقدر للدراسة 0.40.

وفـى الشـكل (6.22) يتضـح أن المسـافة بـين المتوسـط العـام الحقيقـى فـى المجتمـع (μ) وحجم التأثير المقاس فى الدراسة يتكون من جزئين هما:

– المسـافة بـين المتوسـط العـام لحجـم التـأثير الحقيقـى(μ) وحجـم التـأثير الحقيقـى للدراسـة θ_i ويطلق على هذا الخطأ أو الفرق زيتا ζ_i.

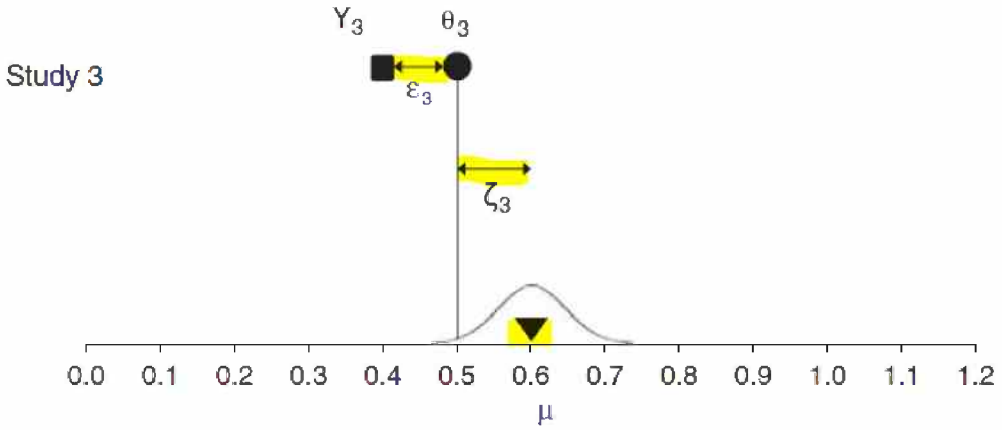
– المسـافة بـين حجـم التـأثير الحقيقـى للدراسـة θ_i وحجـم التـأثير المقـاس للدراسـةES_i وهذا الفرق نتجة خطأ المعاينة فى الدراسة ويطلق عليه الخطأ ابيسلون ε_i.

وعلى فإن حجم التأثير المقاس فى الدراسة كالتالى:

$$ES_i = \mu + \zeta_i + \varepsilon_i$$

- ES_i حجم التأثير المقاس فى الدراسة i.
- μ متوسط توزيع أحجام التأثير فى المجتمع.
- ζ_i (زيتـا) انحـراف أو ابتعـاد ثابـت لحجـم التـأثير المقـاس عـن متوسـط حجـم التـأثير الحقيقى.
- ε_i(ابيسلون) انحراف شرطى لحجم التأثير الحقيقى للدراسة نتيجة خطأ المعاينة.

ويمكن توضيح ذلك بالشكل الآتى (Borenstien et al., 2009):



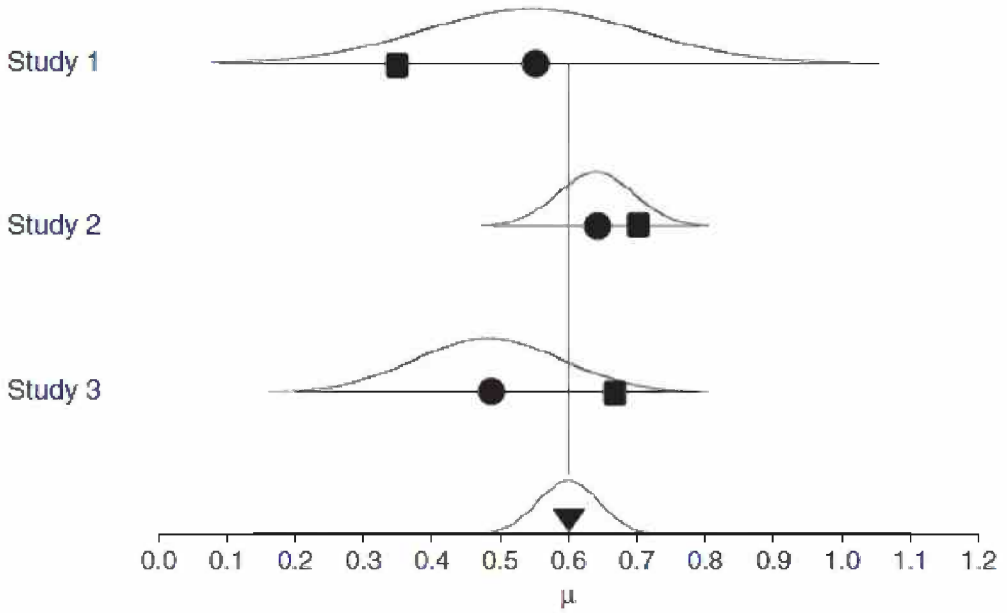
الشكل (7.22): حجم التأثير الحقيقى والمقاس فى نموذج التأثيرات العشوائية.

فالدراسة الثالثة متوسط حجم التأثير الحقيقى لمجتمع الدراسات (μ)=0.60 ، بينما حجم التأثير الحقيقى فى الدراسة θ_3= 0.50 ،وعليه فإن المسافة بين متوسط حجم التأثير الحقيقى للدراسات وحجم التأثير الحقيقى للدراسة 0.10- وهذا الخطأ نشأ نتيجة خطأ المعاينة لعينة الدراسات فى ما وراء التحليل.

والفرق بين حجم التأثير الحقيقى للدراسة θ_3 وحجم التأثير المقاس $ES_3(Y_3)$ يساوى 0.10- وهذا يرجع إلى خطأ المعاينة للأفراد داخل الدراسات وعليه فإن الفرق بين متوسط حجم التأثير الحقيقى للدراسات وحجم التأثير المقاس فى الدراسة يساوى 0.20- وعليه فان:

$$ES_3(Y_3) = 0.60 - 0.10(\zeta_3) - 0.10(\varepsilon_3) = 0.40$$

والمسافة من متوسط حجم التأثير الحقيقى للدراسات μ(المستطيل) إلى حجم التأثير الحقيقى للدراسة θ_i (الدوائر) يعتمد على الانحراف المعيارى لتوزيع التأثيرات الحقيقية عبر الدراسات ويطلق عليه تاو τ اوتباينها τ^2 ويكون توزيعها اعتدالى كالتالى:



الشكل (8.22): توزيع أحجام التأثير الحقيقية للدراسات فى نموذج التأثيرات العشوائية.

لاحظ أن أحجام التأثير الحقيقية للدراسات توزعت عشوائيًا حول متوسط التأثيرات الحقيقية فى نموذج التأثيرات العشوائية بينما فى نموذج التأثيرات المثبتة تم اجبار كل أحجام التأثير ان تقع على الخط الراسى.

الفروق بين نموذج التأثيرات المثبتة والعشوائية فى ما وراء التحليل

توجد العديد من الفروق المفاهيمية، التحليلية عند حساب المتوسط الموزون ودلالتها الإحصائية، والتفسيرية بين النموذجين كالآتى (Card, 2012; Kisamore & Brannick, 2008; Mengersen et al., 2013):

الفروق المفاهيمية النظرية Conceptual Differences

فى النموذج المثبت الفرق بين حجم التأثير الوحيد الحقيقى فى المجتمع θ وحجم التأثير المقاس فى الدراسة نتيجة خطأ المعاينة ε_3، ويتلامس حجم التأثير الحقيقى للدراسات مع حجم التأثير الحقيقى فى المجتمع ،اى ان متوسط حجم التأثير ثابت لكل الدراسات.

بينما فى النموذج العشوائى حجم التأثير الحقيقى للدراسة θ لا يتلاقى أو يتلامس مع متوسط حجم التأثير الحقيقى العام للدراسات فى المجتمع μ بل يتوزع حوله بانحراف معيارى τ، وعليه فإن متوسط حجم التأثير العام للدراسات μ بانحراف معيارى تاو τ وليس حجم تأثير وحيد فى المجتمع θ كما فى النموذج المثبت، ويفترض ان كل حجم تأثير مقاس يختلف عن متوسط مجتمع أحجام التأثير فى ضوء خطأ المعاينة على مستوى الأفراد بالإضافة إلى قيمة اخرى للتباين من مصادر اخرى نتيجة الاختلاف بين الدراسات ويفترض انها تتوزع عشوائيًا . وعليه فإن تباين حجم التأثير المقاس يتكون من مكونين (Lipsey & Wilson, 2001):

- تباين مرتبط بخطأ المعاينة للأفراد على مستوى الدراسة.
- تباين مرتبط بالتأثيرات العشوائية الأخرى نتيجة الاختلاف بين عينة الدراسات.

ويمكن التعبير عن التباين المرتبط بحجم التأثير المقاس كالآتى:

$$S_T^2 = S_\theta^2 + S_i^2$$

- S_θ^2 التباين نتيجة مصادر عشوائية أو نتيجة الاختلاف بين الدراسات.
- S_i^2 التباين المرتبط بخطأ المعاينة للأفراد.

الفروق التحليلية أو الحسابية Analytic differences

حجم التأثير المقاس فى الدراسة فى النموذج المثبت هى حاصل جمع:

- حجم التأثير الحقيقى الوحيد فى المجتمع θ.
- الخطأ ε نتجة خطأ المعاينة فى الدراسات.

بينما حجم التأثير المقاس فى النموذج العشوائى حاصل جمع:

- μ متوسط توزيع أحجام التأثير فى المجتمع.
- ζ_i (زيتا) انحراف أو ابتعاد ثابت لحجم التأثير المقاس عن متوسط حجم التأثير الحقيقى نتيجة خطأ معاينة الدراسات فى ما وراء التحليل.
- ε_i(ابيسلون) انحراف شرطى لحجم التأثير الحقيقى للدراسة نتيجة خطأ المعاينة الأفراد فى الدراسات.

الفروق التفسيرية Interpretive Differences

مـا وراء التحليـل باسـتخدام النمـوذج المثبـت يهـدف إلـى الوصـول إلـى اسـتنتاجات شـرطية Conditional Inferences حـول مجموعـة محـددة مـن الدراسـات المتضـمنة فـى التحليـل فقـط ولـيس علـى مجتمـع الدراسـات التـى تناولـت المتغيـرين بمعنـى " ان متوسـط حجـم التـأثير لـ 15 دراسـة....." ولا يعمـم علـى مجتمـع الدراسـات التـى تناولـت المتغيرين ويمكن القول إن نتائج نموذج التأثيرات المثبتة وصفية.

بينمـا مـا وراء التحليـل باسـتخدام نمـوذج التـأثيرات العشـوائية يهـدف إلـى الوصـول لاسـتنتاجات غيـر مشـروطة Unconditional Inferences ليسـت قاصـرة علـى عينـة الدراسـات فـى التحليـل بـل إلـى مجتمـع الدراسـات المحتملـة التـى تناولـت دراسـة العلاقـة بـين المتغيـرين بمعنـى أن " تـراث الدراسـات للعلاقـة بـين المتغيـرين X, Y توصـل إلـى ان حجـم التـأثير أو العلاقـة بـين المتغيـرين....."، ويمكـن القـول إن نتـائج نمـوذج التـأثيرات العشـوائية اسـتدلالية. ولـذلك فـإن نمـوذج التـأثيرات العشـوائية أكثـر مناسـبة فـى الدراسـات النفسـية التربويـة الاجتماعيـة نتيجـة تـراكم كـم هائـل مـن الدراسـات للوصـول إلـى اسـتنتاجات أكثر تعميمية وليست قاصرة على عينة الدراسات فى التحليل.

وعليـه فـإن نمـوذج التـأثيرات العشـوائية لا يفتـرض ان الدراسـة مشـتقة مـن نفـس المجتمـع كمـا هـو الحـال فـى نمـوذج التـأثيرات المثبتـة بـل انهـا تمثـل مجتمـع فريـد فـى حـد ذاتهـا، وفـى هـذه الحالـة يوجـد مصـدرين للاختلافـات أحـدهما مـن الدراسـة كمـا هـو الحـال فـى نمـوذج التـأثيرات المثبتـة والآخـر مـن الاختلافـات بـين الدراسـات ولـذلك يسـمى بنمـوذج عـدم التجانس (Egger, 2006, Mengersen et al., 2013)

العمليات الحسابية لنموذج التأثيرات العشوائية

تناولنـا فـى الفصـل الثـامن تحليـل نمـوذج التـأثيرات المثبتـة مـن حسـاب المتوسـط المـوزون لأحجـام التـأثير فـى ضـوء الخطـأ المعيـارى، وحسـاب اختبـار التجـانس وفتـرات الثقـة والدلالـة الإحصـائية باسـتخدام اختبـار Z. امـا عـن العمليـات الحسـابية لنمـوذج التـأثيرات العشـوائية فإنهـا تختلـف عـن النمـوذج المثبـت نظـرًا لوجـود مكـون آخـر للتبـاين يتمثـل فـى

المكـون العشـوائى ζ_i، وعليـه فـلا بـد مـن إجـراء موازنـة لكـل حجـم تـأثير فـى ضـوء تعـديلين احـدهما خـاص بالخطـأ المعيـارى المـرتبط بخطـأ المعاينـة للأفـراد والآخـر مـرتبط بـالمكون العشـوائى الـذى ربمـا يرجـع فـى جـزء منـه إلـى خطـأ معاينـة الدراسـات المتضـمنة فـى مـا وراء التحليل.

ويـرى Card (2012) أن مـا وراء التحليـل للنمـوذج العشـوائى يتضـمن أربـع خطـوات عامة هى كالآتى:

اولًا: تقدير التجانس أو عدمه بين أحجام التأثير للدراسات من خلال اختبار Q.

ثانيًا: تقدير التباين بين أحجام التأثير فى المجتمع.

ثالثـا: اسـتخدام التبـاين لأحجـام التـأثير فـى المجتمـع للحصـول علـى اوزان التـأثيرات العشوائية Random-effects weights لأحجام التأثير المقاسة فى الدراسات.

رابعًـا: اسـتخدام اوزان التـأثيرات العشـوائية لتقـدير متوسـط التـأثيرات العشـوائية لأحجـام التـأثير والأخطـاء المعياريـة المرتبطـة بهـا لاسـتخدامها فـى حسـاب الدلالـة الإحصـائية وفترات الثقة.

1. تقـدير التجـانس مـن عدمـه بـين أحجـام التـأثيرات: يقدر التجانس من عدمه من خلال اختبارQ كالآتى:

$$Q = \sum w_i \, (ES_i - \overline{ES})^2 = \sum (w_i ES_i^2 - \frac{(\sum (w_i ES_i)^2}{\sum w_i}$$

- w_i وزن الدراسة i.
- ES_i حجم التأثير للدراسة i .
- $\overline{ES}$ متوسط حجم التأثير الموزون العام لكل الدراسات.

درجات الحرية لاختبار Q:

$$df = k-1$$

- K عدد الدراسات فى ما وراء التحليل.

وفى العرض السابق تم تقدير اختبار Q= 291.17 وبمقارنتها بقيمة إحصاء كأى تربيع الجدولية اتضح ان قيمتها أكبر من القيمة الجدولية لكاى تربيع، وعليه رفض الفرض الصفرى بالتالى توجد دلالة إحصائية بمعنى لا يوجد تجانس أو الاختلاف بين أحجام التأثير أكبر مما هو متوقع نتيجة خطأ المعاينة بين أحجام التأثير لـ 22 دراسة فى ما وراء التحليل. وهذا الاستنتاج يفرض على الباحث استخدام نموذج التأثيرات العشوائية وليس النموذج المثبت لأن النموذج المثبت يتطلب تجانس بين أحجام التأثير المتضمنة فى دراسة ما وراء التحليل.

2. تقدير تباين مجتمع أحجام التأثير Population Variability : وعدم التجانس بين أحجام التأثير لا يرجع فقط إلى خطأ المعاينة بل أيضًا إلى مكون عشوائى يسبب انحرافات حقيقية عن المتوسط العام لحجم التأثير الحقيقى لمجتمعات حجم التأثير ويتم تقدير تباين أحجام التأثير الحقيقية للدراسات من خلال معلم τ^2 وهو التباين أحجام التأثير الحقيقية للدراسات ولكن الباحث لا يعرف أحجام التأثير الحقيقة ولكن يمكن تقدير تباينها كالآتى (Borenstien et al., 2009; Card, 2012):

$$\tau^2(v_\theta) = \frac{Q - (K-1)}{(\sum w_i) - \frac{(\sum w_i^2)}{(\sum w_i)}}$$

- Q إحصاء التجانس.
- K عدد الدراسات أو أحجام التأثير.
- w_i وزن كل دراسة من خطأ المعاينة.

ولو ان قيمة هذه المعادلة بالسالب فإن تباين مجتمع أحجام التأثير صفر وهذا يحدث لو ان توزيع أحجام التأثير متجانس وهذا يتطابق مع نموذج التأثيرات المثبتة بالتالى شروط النموذج العشوائى لم تتحقق.

ولتقدير تباين مجتمع أحجام التأثير الحقيقية للدراسات الـ 22 فى الجدول الآتى (انظر: Card, 2012):

الجدول (3.2): بيانات الـ 22 دراسة وحساباتها فى النموذج العشوائى.

Sample size (N)	Age (years)	Effect size (Z_r)	Standard error (SE)	Weight (w)	wES	wES^2	w^2	w^*	w^*ES
228	9.2	.583	0.0693	208.12	121.27	70.66	43315.0	21.94	11.51
491	9.4	.201	0.0579	297.81	59.74	11.98	88693.4	22.66	4.49
65	4.5	.322	0.1325	56.95	18.34	5.90	3242.9	17.14	5.34
458	8.0	.624	0.0484	427.07	266.55	166.37	182388.9	23.20	12.85
929	14.7	.162	0.0346	835.96	135.72	22.03	698829.8	23.83	3.84
904	7.5	.349	0.0347	831.95	290.69	101.57	692141.3	23.83	8.00
74	6.0	.419	0.1223	66.89	28.02	11.73	4474.6	17.95	7.11
151	9.5	.721	0.0855	136.66	98.48	70.97	18675.3	20.80	12.84
150	16.0	.628	0.0845	139.90	87.90	55.23	19570.9	20.87	11.62
590	9.0	.655	0.0426	550.97	361.12	236.69	303565.1	23.48	13.51
180	4.8	.039	0.0831	144.75	5.70	0.22	20953.4	20.98	0.83
139	3.6	.375	0.0892	125.71	47.14	17.68	15802.7	20.52	7.35
132	4.1	.049	0.0922	117.70	5.81	0.29	13853.1	20.30	1.00
60	4.6	.000	0.1386	52.05	0.00	0.00	2709.6	16.67	0.00
839	14.5	.339	0.0381	689.52	233.68	79.19	475442.6	23.69	7.73
262	8.7	–.048	0.0642	242.93	–11.73	0.57	59016.0	22.28	–1.07
266	9.5	.489	0.0636	246.94	120.85	59.15	60977.2	22.31	10.12
209	15.5	.258	0.0715	195.52	50.54	13.06	38226.3	21.79	5.51
314	9.0	.162	0.0589	288.73	46.71	7.56	83367.0	22.61	3.63
881	8.0	.519	0.0351	810.71	420.41	218.01	657255.5	23.81	11.35
517	8.0	.509	0.0455	482.45	245.37	124.79	232756.5	23.34	10.94
228	9.0	.651	0.0702	202.92	132.06	85.94	41174.7	21.88	12.52
				7152.21	2764.36	1359.60	3756432	475.89	161.02

وبتقدير قيمة تباين مجتمع أحجام التأثير الحقيقية كالآتى:

$$\tau^2 = \frac{Q-(K-1)}{(\sum w_i) - \frac{(\sum w_i^2)}{(\sum w_i)}} = \frac{291.17-(22-1)}{7152.21 - \frac{3756432}{7152.21}} = \frac{270.17}{6627} = 0.0408$$

3. حساب اوزان التأثيرات العشوائية: بعد تقدير تباين مجتمع أحجام التأثير فإن الخطوة التالية حساب اوزان التأثيرات العشوائية لكل دراسة. المنطق من تقدير الاوزان مماثل لما تم حسابه من موازنة أو تصحيح حجم التأثير فى نموذج التأثيرات المثبتة من عدم الدقة لحجم التأثير من الخطأ المعيارى نتيجة خطأ المعاينة للحصول على حجم

تـأثير أكثـر دقـة وفـى هـذا التصـحيح فـإن حجـم التـأثير المصـحح أو الـدقيق هـو دالـة وظيفية نتيجة حجم التأثير المقاس وانحراف المعاينة فى كل دراسة.

وفـى النمـوذج العشـوائى فـإن حجـم التـأثير يكـون وظيفـة لحجـم التـأثير المقـاس بالإضـافة إلـى مكـونين مـن عـدم الدقـة Imprecision ولـيس مكـون واحـد كمـا فـى نمـوذج التـأثيرات المثبتة هما:

- ζ_i (ابسـاى) انحـراف أو ابتعـاد ثابـت لحجـم التـأثير المقـاس عـن متوسـط حجـم التـأثير الحقيقـى نتيجـة خطـأ معاينـة الدراسـات فـى مـا وراء التحليـل، ويقـدر مـن خلال تباين مجتمع أحجام التأثيرات τ^2.
- ε_i(زيتـا) انحـراف شـرطى لحجـم التـأثير الحقيقـى للدراسـة نتيجـة خطـأ المعاينـة الأفراد فى الدراسات، ويقدر من خلال الخطأ المعيارى SE.

وعلـى ذلـك فـإن اوزان التـأثيرات العشـوائية تتكـون مـن تبـاين مجتمـع أحجـام التـأثير بين الدراسات نتيجة عوامل مختلفة والخطأ المعيارى الخاص بكل دراسة.

وتقدر اوزان التأثيرات العشوائية كالآتى:

$$w_i^* = \frac{1}{\tau^2 + SE_i^2}$$

- w_i^* اوزان التأثيرات العشوائية.
- τ^2 تباين مجتمع أحجام التأثيرات.
- SE_i الخطأ المعيارى المرتبط بحجم التأثير لكل دراسة.

وعلـى ذلـك يـتم تصـحيح حجـم التـأثير لكـل دراسـة مـن خـلال اوزان التـأثيرات العشـوائية. بـالرجوع إلـى جـدول البيانـات فـى جـدول(3.**22**) تـم حسـاب w_i^* لكـل دراسـة ثـم حسـاب حاصل ضرب وزن التأثير العشوائى w_i^* × حجم التأثير ES_i لكل دراسة.

وفـى الجـدول (3.**22**) بالنسـبة للدراسـة الأولـى فـإن التصـحيح فـى ضـوء نمـوذج التـأثيرات المثبتة كالآتى:

$$w_1 = \frac{1}{SE^2} = \frac{1}{(0.0693)^2} = 208.12$$

ولكن فى النموذج العشوائى تم حساب وزن جديد كدالة لكلًا من تباين مجتمع أحجام التأثير 0.0408 والخطأ المعيارى لكل دراسة SE=0.0693 وعليه يصبح تصحيح حجم التأثير للدراسة كالآتى:

$$w_i^* = \frac{1}{\tau^2 + SE_i^2} = \frac{1}{0.0408 + 0.0693^2} = 21.94$$

لاحظ أن قيمة تصحيح أو وزن حجم التأثير للدراسة الأولى فى نموذج التأثيرات المثبتة هى 208.12 بينما قيمة التصحيح فى ضوء نموذج التأثيرات العشوائية 21.94.

وبالنظر إلى اوزان التأثيرات العشوائية لـ 22 دراسة (العمود الثانى يمينًا) فى الجدول تلاحظ الاتى:

- قيمة هذه الاوزان w_i^* أصغر بكثير من اوزان التأثيرات المثبتة w_i وهذا يعنى ان مجموعها عبر الدراسات يكون أصغر من نظيراتها فى النموذج المثبت وهذا يعنى ان الخطأ المعيارى لمتوسط حجم التأثير يكون كبير.
- على الرغم أن كل الدراسات لها تقريبًا نفس الترتيب النسبى فى ضوء نموذج التأثيرات العشوائى أو المثبتة (الدراسة التى لها أعلى وزن على أحدهما تكون لها أعلى وزن على النموذج الآخر) إلا أن التناقض فى الاوزان عبر الدراسات أقل بالنسبة للنموذج العشوائى مقارنة بالنموذج المثبت.

الخطوة الرابعة: حساب متوسطات التأثيرات العشوائية والاستدلال حولها: الخطوة الاخيرة فى تحليل نموذج التأثيرات العشوائية تقدير متوسط حجم التأثير ثم إجراء اختبارات الدلالة الإحصائية وحساب فترات الثقة حول هذا المتوسط. ويقدر متوسط حجم التأثير تمامًا مثل متوسط حجم التأثير فى النموذج المثبت مع استبدال الوزن w_i بالوزن w_i^* وتصبح المعادلة كالتالى(Card, 2012; Lipsey & Wilson, 2001):

$$\overline{ES} = \frac{\sum(w_i^* ES_i)}{\sum w_i^*}$$

وبتطبيق هذه المعادلة على الـ 22 حجم التأثير فى الجدول:

$$\overline{Z_r} = \frac{\sum(w_i^* ES_i)}{\sum w_i^*} = \frac{161.02}{475.89} = 0.338$$

لاحظ أنه تم تحوير حجم التأثير معامل الارتباط r إلى تحوير فيشر Z_i ثم قدر متوسط تحويرات فيشر لكل الدراسات $\overline{Z_r}$ وقيمتة 0.338 ثم تعاد هذه القيمة مرة اخرى إلى متوسط معامل الارتباط وهى $\bar{r}$ =0.326 . انظر نموذج التأثيرات المثبتة حيث إن قيمة $\overline{Z_r}$ =0.387 بينما $\bar{r}$ =0.387. أى ان متوسط حجم التأثير فى النموذج العشوائى أقل من متوسط حجم التأثير فى النموذج المثبت وهذا يرجع إلى أنه تم تصحيح حجم التأثير فى النموذج العشوائى من نوعين من الخطأ بينما فى النموذج المثبت تم تصحيحها من خطأ واحد فقط وهو خطأ المعاينة.

ويقدر الخطأ المعيارى لمتوسط حجم التأثير كالآتى (Card, 2012; Lipsey & Wilson, 2001):

$$SE_{\overline{Z_r}} = \sqrt{\frac{1}{\sum(w^*)}} = \sqrt{\frac{1}{475.89}} = 0.0458$$

لاحظ أن الخطأ المعيارى فى النموذج العشوائى أكبر من الخطأ المعيارى فى النموذج المثبت 0.0118(انظر الفصل الثامن).

وتقدر الدلالة الإحصائية من خلال اختبار Z (للدلالة وليس تحوير فيشر):

$$Z = \frac{\overline{Z_r}}{SE_{\overline{Z_r}}} = \frac{0.338}{0.0458} = 7.38, P < 0.01$$

ويتضح أن قيمة Z أكبر من 1.96 (قيمة Z الجدولية عند مستوى دلالة إحصائية 0.05 لاختبار ذو ذيلين) أو أكبر من 2.58(قيمة Z الجدولية عند مستوى دلالة إحصائية 0.01 لاختبار ذو ذيلين)، على ذلك يرفض الفرض الصفرى أى توجد علاقة ارتباطية موجبة ودالة إحصائيًا بين العدوان ورفض الاقران.

لاحظ أن قيمة Z للنموذج العشوائى أقل بكثير من نظيرتها للنموذج المثبت 32.70(انظر الفصل الثامن)، وهذا يرجع إلى ان الخطأ المعيارى(المقام) فى النموذج العشوائى (0.0458) أكبر من الخطأ المعيارى فى النموذج المثبت (0.0118) وهذا ادى إلى انخفاض قيمة Z فى النموذج العشوائى عن نظيراتها فى النموذج المثبت.

وتقدر فترات الثقة من الخطأ المعيارى كالتالى(Hedges & Olkin, 1985):

$$\overline{ES}_{lower} = \overline{ES} - Z_{(1-\alpha)}SE_{\overline{ES}}$$

$$\overline{ES}_{upper} = \overline{ES} + Z_{(1-\alpha)}SE_{\overline{ES}}$$

- $\overline{ES}_{lower}$ حد الثقة الادنى.
- $\overline{ES}$ متوسط حجم التأثير.
- $Z_{(1-\alpha)}$ القيمة الجدولية عند 0.05 أو 0.01 لاختبار ذو ذيلين.
- $SE_{\overline{ES}}$ الخطأ المعيارى لمتوسط حجم التأثير.

ولكن قبل حساب فترات الثقة تحول قيمة متوسط تحوير فيشر إلى متوسط معامل الارتباط. ويقدر فترات الثقة 95% CI كالآتى:

$$\bar{r}_{lower} = \bar{r} - Z_{(1-\alpha)}SE_{\bar{r}} = 0.326 - 1.96 \times .0458 = .2363$$

$$\bar{r}_{lower} = \bar{r} + Z_{(1-\alpha)}SE_{\bar{r}} = 0.326 + 1.96 \times .0458 = .349$$

المقارنة بين نموذج التأثيرات المثبتة ونموذج التأثيرات العشوائية

قدم Borenstien et al.(2009) مقارنة جيدة ومستفيضة بين النموذجين ففى النموذج المثبت نفترض ان حجم التأثير الحقيقى هو نفسة لكل الدراسات وان متوسط حجم التأثير المقدر من الدراسات هو تقدير لحجم التأثير الحقيقى فى المجتمع وان السبب الوحيد للاختلاف بين أحجام التأثير هو خطأ المعاينة(عينات الدراسات)، فى حين فى النموذج العشوائى نفترض ان حجم التأثير الحقيقى يختلف من دراسة إلى اخرى وان الدراسات المتضمنة فى التحليل هى عينة عشوائية لاحجام التأثير للظاهرة موضع الدراسة وان ملخص حجم التأثير هو تقدير لمتوسط أحجام التأثير. وفى النموذج المثبت مصدر الخطأ أو عدم التاكد هو نتيجة خطأ المعاينة بينما فى نموذج التأثيرات

العشوائية يوجد مصدرين للاخطأ هما خطأ المعاينة بالاضافة إلى خطأ ناتج عن الفروق بين الدراسات، وعليه فإن الأخطاء المعيارية وفترات الثقة وتباين أحجام التأثير فى النموذج العشوائى أكبر أو اوسع من نظيراتها للنموذج المثبت.

اما الفرض الصفرى فى النموذج المثبت هو ان حجم التأثير فى كل دراسة يساوى صفر بينما فى النموذج العشوائى متوسط حجم التأثير لكل الدراسات يساوى صفر وبالتاكيد هذا ينعكس عل عملية تعميم نتائج النموذجين.

ويفضل استخدام نموذج التأثيرات المثبتة إذا تحقق الشرطين:

- الدراسات المتضمنة فى التحليل متماثلة أو متجانسة.
- الهدف من التحليل حساب حجم التأثير العام للدراسات المتماثلة دون تعميم النتائج لمجتمعات غير مجتمع عينة الدراسات.

فلو ارادت شركة ادوية التحقق من فاعلية عقار ما واستخدمت مئات المرضى من خلال المجموعة التجريبية والضابطة ولان الشركة تجرى الدراسة على 100 مريض فى نفس الوقت فالشركة تجرى عشرة دراسات فى نفس الوقت على العينة وفى هذة الحالة فإن الدراسات العشرة متماثلة لأن العينة لها خصائص متشابهة ونفس المعالجة(العقار) ونفس القياسات ونفس الفريق البحثى، وعليه فإن هذة النوعية من الدراسات متماثلة أو متجانسة (الشرط الأول) وتقوم على نفس المجتمع فإن الهدف من دراسة ما وراء التحليل دراسة مدى فعالية العقار على نفس العينة وتعمم النتائج على نفس المجتمع وليس مجتمعات مرضى اخرى(الشرط الثانى). ولكن يبدو هذا التجانس فى دراسات العلوم الطبية ولكنه بعيد المنال فى دراسات العلوم النفسية والاجتماعية.

وعلى النقيض من المثال السابق حيث يجمع الباحث بيانات أو دراسات من سلسلة من الدراسات فى التراث البحثى لظاهرة ما لباحثين مستقلين بإجراءات وعينات ومنهجيات وببرامج مختلفة بمعنى أن الدراسات ليست متجانسة أو متكافئة ولذلك لا يمكن افتراض وجود حجم تأثير عام.

واذا كان عدد ت الدراسات صغير جدًا فى ما وراء التحليل فإن تقدير التباين بين الدراسات τ^2 يكون غير دقيق وعلى الرغم صلاحية استخدام نموذج التأثيرات العشوائية ولكن المعلومات المتطلبة لتنفيذة تكون ناقصة وفى هذة الحالة امام الباحث عدة خيارات (Borenstein et al., 2009):

1. عرض أحجام التأثير بصورة منفصلة ولا تعرض ملخص التأثير بالتالى لايمكن الوصول إلى استنتاجات حول حجم التأثير وفترات الثقة المرتبطة به، وهنا يلجأ الباحث إلى عرض الدراسات فى ضوء طريقة العد كطريقة لمراجعة الدراسات السابقة وربما يصل إلى استنتاج خاطئ.

2. إجراء ما وراء التحليل باستخدام نموذج التأثيرات المثبتة حيث يعطى تحليل وصفى للدراسات المتضمنة فى التحليل.

ومعظم الباحثون فى العلوم الاجتماعية يتبنوا مدخل Hedges & Olkin (1985) و Rosonthal & Rubin (1982) للتاثيرات المثبتة لسهولة فهمة وبساطتة ولكن يؤدى إلى تقدير متحيز لحجم التأثير الحقيقى لأن البيانات الحقيقية لاحجام التأثير لا تقيس نفس حجم التأثير انما تتنوع باختلاف مجتمع عينات الدراسة وخصائص الدراسات أو ما يعرف بالمتغيرات المتفاعلة (Hagger, 2006)، واثبتت دراسات المحاكاة ان باستخدام نموذج التأثيرات المثبتة لاحجام التأثير تولدت من مجتمعات مختلفة يؤدى إلى تضخم الخطأ من النوع الأول قبول وجود حجم تأثير مفترض بينما فى الواقع لا يوجد تأثير حقيقى وعلى ذلك يجب تبنى نموذج التأثيرات العشوائية لـ Hunter & Shmidt (1990, 2004) أو Field, Hedges & Vevea (1998) (2003).

راجع Hunter & Shmidt (2000) طرق ما وراء التحليل المستخدمة فى الدراسات فى مجلة Psychological Bulletin ووجدوا ان الـ 21 دراسة استخدمت كلها نموذج التأثيرات المثبتة متبوعة بالتحليلات المتفاعلة ولم تستخدم أى من الدراسات نموج التأثيرات العشوائية وحذروا من ان استخدام نموذج التأثيرات المثبت لأنه يضخم نسبة

الخطأ الأول بنسبة أعلى من 11% لدراسات ما وراء التحليل التى تتضمن حجم عينة 25 دراسة، وأعلى من 28% للدراسات التى تتضمن حجم عينة أكثر من 100 دراسة وفى هذا الشأن يتفق مع Field (2003).

عرض مثال تطبيقى (فى: Ellis, 2010)

اهتم باحث بإجراء ما وراء التحليل للعلاقة بين ادمان الانترنت والتحصيل وجمع ثلاثة دراسات والبيانات كالآتى:

الجدول (4.22): بيانات ثلاثة دراسات للعلاقة بين ادمان الفيس بوك والتحصيل.

الباحث	r	P	N	ثبات المتغير التابع
A (2009)	-0.48	0.02	80	0.70
B (2014)	-0.58	0.001	112	0.92
C (2017)	0.05	0.33	32	-

- r معامل ارتباط بيرسون.
- p القيمة الاحتمالية للخطأ من النوع الأول.
- N حجم العينة.

إجراء ما وراء التحليل فى ضوء مدخل Hunter & Schmidt (1990, 2004)

حساب متوسط حجم التأثير الموزون

الطريقة الاسهل والاسواء لحساب متوسط حجم التأثير كالآتى:

$$M_{ES}(\bar{r}) = \frac{-0.48 \pm 0.58 + 0.05}{3} = -0.337$$

لاحظ أن هذا المتوسط أصغر من حجم التأثير أو معامل الارتباط للدراستين A, B وهذا يعنى ان المتوسط متحيز ويتضح أن حجم التأثير للدراسة C مختلف عن الدراستين B, A.

ولكن المدخل الاكثر دقة ومنطقية وهو اعطاء وزن أو أهمية للدراسة بما يتناسب مع حجم العينة فمن غير المنطقى ان دراسة تتضمن عينة 112 ان تتساوى مع دراسة عينتها 32 وعلى ذلك فلا بد من حساب متوسط حجم التأثير الموزون من خلال ضرب كل حجم تأثير فى حجم عينها مقسومًا على حجم العينة الكلى كالآتى:

$$\bar{r} = \frac{\sum n_i r_i}{\sum n_i}$$

وعليه فإن متوسط حجم التأثير الموزون كالتالى:

$$= \frac{(80 \times -0.48) + (112 \times -0.58) + (32 \times 0.05)}{80 + 112 + 32}$$

$$\overline{r} = \frac{(-38.4) + (-65.0) + (1.6)}{224} = -0.454$$

الملاحظ أن المتوسط الموزون 0.454- أكبر فى قيمته المطلقة من المتوسط غير الموزون 0.337- وهو قريب من قيمة حجمى التأثير للدراستين B,A بمعنى أن المتوسط الموزون أكثر دقة فى التعبير عن حجم التأثير الفعلى للدراسات.

تصحيح حجم التأثير من اخطاء القياس أو الثبات المنخفض

للحصول على أحجام تأثير أكثر دقة يجب تصحيحها من عدم الثبات للمقاييس فالمقياس المستخدم فى قياس المتغير التابع فى الدراسة A أقل من ثبات المقياس المستخدم فى الدراسة B وهذا يؤدى إلى انكماش حجم التأثير عن قيمته الحقيقية نتجة اخطاء القياس ولتصحيح حجم التأثير من خطأ القياس من خلال المعادلة الآتية:

$$ES_{adjusted} = \frac{r}{\sqrt{\alpha}}$$

- α المعامل ألفا كرونباخ.

وعليه فإن حجم التأثير المصحح للدراسة A كالتالى:

$$ES_A(r)_{adjusted} = \frac{-0.48}{\sqrt{0.70}} = -0.574$$

وحجم التأثير المصحح للدراسة B:

$$ES_B(r)_{adjusted} = \frac{-0.58}{\sqrt{0.92}} = -0.605$$

امـا الدراسـة الثالثـة لـم يـذكر الباحـث معامـل الثبـات ألفـا كرونبـاخ ولـذلك يقـدر علـى متوسط الثبات فى الدراستين **B, A**:

$$= \frac{0.70 + 0.92}{2} = 0.81$$

وحجم التأثير المصحح للدراسة الثالثة **C** كالتالى:

$$ES_C(r)_{adjusted} = \frac{0.05}{\sqrt{0.81}} = 0.056$$

ويمكن إجراء التصحيح فى ضوء عدم الثبات المتغيرين فى الدراسة كالآتى:

$$ES(r)_{adjusted} = \frac{r}{\sqrt{\alpha_1 \alpha_2}}$$

وعليه فإن متوسط حجم التأثير الموزون المصحح من خطأ القياس أو نقص الثبات:

$$= \frac{(80 \times -0.574) + (112 \times -0.605) + (32 \times 0.056}{80 + 112 + 32}$$

$$= \frac{(-45.92) + (-67.76) + (1.79)}{224}$$

$$= -0.50$$

لاحـظ أن متوسـط حجـم التـأثير المـوزون تغيـرت قيمتـة مـن 0.337 إلـى 0.454 نتجـة تصـحيحة مـن خطـأ المعاينـة والـى 0.50 نتجـة تصـحيحة مـن خطـأ المعاينـة وخطـأ القياس، وبالتالى نستطيع ان نثق فى النتيجة الثالثة لأنها أقل تحيز.

حساب الدلالة الإحصائية لمتوسط حجم التأثير

لحساب الدلالة الإحصائية لمتوسط حجم التأثير (متوسط معامل الارتباط) يوجد مدخلين هما:

المدخل الأول: تحوير معامل الارتباط إلى تحوير فيشر Z_r وتحديد الدلالة الإحصائية ولكن Hunter & Schmidt (1990, 2004) اجروا ما وراء الحليل على قيم معامل الارباط مباشرة دون تحويره إلى تحوير فيشر وفى هذه الحالة لا بد من حساب الخطأ المعيارى المرتبطة بمتوسط حجم التأثير والخطأ المعيارى لوصف انتشار التوزيع العينى لأحجام التأثير فى المجتمع بمعنى أنه نوع خاص من الانحراف المعيارى وفى هذا المثال فإن التوزيع العينى يتكون من ثلاثة أحجام تأثير، والبعض يفضل مصطلح التباين كتعبير عن الخطأ المعيارى ولكن المصطلح الاكثر استخدامًا هو الجذر التربيعى للتباين. وتباين عينة معاملات الارتباطات v_r تقدر من مجموع حاصل ضرب مربع الفرق بين كل حجم التأثير للدراسة ومتوسط حجم التأثير مضروبًا فى حجم العينة ثم يقسم الناتج على حجم العينة الكلى كالآتى Ellis (2010):

$$v_r = \frac{\sum n_i(r_i - \bar{r})^2}{\sum n_i}$$

$$= \frac{(80 \times (-.574 - -.50)^2 + (112 \times (-.605 - -.500)^2 + (32 \times (.056 - -.500)^2}{80 + 112 + 32}$$

$$v_r = \frac{0.400 + 1.232 + 9.888}{224} = 0.51$$

ويتم تقدير الخطأ المعيارى كالآتى:

$$SE_{\bar{r}} = \sqrt{\frac{v_r}{K}} = \sqrt{\frac{0.051}{3}} = 0.130$$

وتقدر الدلالة الإحصائية فى ضوء اختبار تحويل معامل الارتباط إلى الدرجة المعيارية Z كالآتى:

$$Z = \frac{\bar{r}}{SE_{\bar{r}}}$$

$$= \frac{0.500}{0.130} = 3.85$$

وقيمة Z الجدولية لمستوى دلالة إحصائية 0.05 لاختبار ذو ذيلين 1.96 ولمستوى 0.01 هى 2.58 ، وبما أن قيمة Z المحسوبة أكبر من قيمتها الجدولية، وعليه يرفض الفرض الصفرى بالتالى فإن النتائج دالة إحصائيًا.

المدخل الثانى: تقدير فترات الثقة: تقدر 95% فترات الثقة من المعادلات الآتية:

$$CI_{95\%\ lower} = \bar{r} - Z\alpha_{/2} SE_{\bar{r}} = -.500 - (1.96 \times .13) = -.755$$

$$CI_{95\ upper} = \bar{r} + Z\alpha_{/2} SE_{\bar{r}} = -.500 + (1.96 \times .13) = -.245$$

وبما أن قيمة المعلم فى الفرض الصفرى هى صفر، وبما أن القيمة صفر لا تقع بين حدى الثقة الأدنى والأعلى، وعليه فإن النتائج دالة إحصائيًا.

زيادة الخطأ المعيارى يجعل فترات الثقة أكثر اتساعًا مما يجب أن تكون عليه.

اختبار الاختلاف أو عدم التجانس بين توزيع تقديرات حجم التأثير

اتساع فترات الثقة يشير إلى احتمال ان توزيع أحجام التأثير غير متجانسة وهذا يعنى ان أحجام التأثير لا تتمركز حول متوسط وحيد لاحجام التأثير فى المجتمع وتتوزع حول متوسطات عديدة لاحجام التأثير وتقدر اختبار عدم التجانس من خلال اختبار Q كالآتى:

$$Q = \sum (n_i - 1)\,(r_i - \bar{r})^2$$

$$= ((80-1) \times (-.574 - -.50)^2) + ((112-1) \times (-.650 - -.500)^2) + ((32-1) + (.056 - -.500)^2)$$

$$= (79 \times .005) + (111 \times .011) + (31 \times .309)$$

$$= 0.395 + 1.221 + 9.579$$

$$= 11.195$$

وتختبر الدلالة الإحصائية من خلال اختبار كاى تربيع بدرجات حرية K-1 حيث K عدد أحجام التأثير، وبما أن قيمة اختبار كاى تربيع لدرجات حرية 2 ومستوى دلالة إحصائية 0.05 هى 5.991 ، بالتالى فإن قيمة Q المحسوبة أكبر من 5.991، وعليه نرفض الفرض الصفرى بوجود اختلاف بين أحجام التأثير، وهذا يتطلب اما إجراء تحليلات المتفاعلة لمعرفة أى من خصائص الدراسات التى احدثت هذه الدلالة أو تحليل ما وراء التحليل فى ضوء نموذج التأثيرات العشوائية. ولكن لبيانا هذا المثال لا يمكن إجراء تحليل التفاعل نظرًا لقلة الدراسات وعدم اتاحة الفرصة لعمل مجموعات فرعية من الدراسات أو أحجام التأثير.

تفسير النتائج

يتضمن التفسير الاجابة عن الأسئلة الآتية: ما دلالة قيمة متوسط حجم التأثير فى ضوء معايير كوهين؟ ماهى الفئة المستفيدة من هذه النتجة وكيف تستفيد منها؟ ماذا اضافت هذه النتيجة إلى التراث البحثى؟ هل النتائج دعمت أو تناقضت ما هو موجود فى التراث البحثى؟ ماهى محددات الدراسة ان وجدت؟

إجراء ما وراء التحليل لتوليف r باستخدام مدخل Hedges وزملائه

(Hedges, 1981,1992; Hedges & Olkin,1985; Hedges & Vevea, 1998)

نموذج التأثيرات المثبتة: باستخدام مدخل Hedges et al لتوليف أحجام التأثير يتم فى ضوء نموذج التأثيرات المثبتة كالآتى:

الأول: يتم تحويل أحجام التأثير الخام r إلى تحوير فيشر Z_r كالآتى:

$$Z_r = \frac{1}{2} Ln \frac{1+r}{1-r}$$

ويتم تقدير تباين الخطأ المعيارى كالآتى:

$$v_i = \frac{1}{n_i - 3}$$

ووزن الدراسة w_i مقلوب تباين الخطأ المعيارى كالآتى:

$$w_i = n_i - 3$$

الجدول (5.22): بيانات ما وراء التحليل فى ضوء البيانات الخام ونموذج التأثيرات المثبت والتأثيرات العشوائية.

Raw study data			Fixed-effects sums			Random-effects sums		
n	r	z	w	wz	wz^2	w^2	w^*	w^*z
80	−.48	−.523	77	−40.27	21.06	5,929	11.35	−5.94
112	−.58	−.662	109	−72.21	47.84	11,881	11.86	−7.86
32	.05	.050	29	1.45	0.07	841	9.12	0.46
			215	−111.03	68.97	18,651	32.34	−13.34

لحساب متوسط حجم التأثير الموزون لـ Z_r كالآتى:

$$\bar{Z}_r = \frac{\sum w_i z_i}{\sum w_i} = \frac{\sum(-0.48 \times 77) + (-0.58 \times 109) + (0.05 \times 29)}{77 + 109 + 29}$$
$$= \frac{-111.03}{215} = -0.52$$

وتباين المتوسط الموزون:

$$v_{\cdot} = \frac{1}{215} = 0.005$$

والخطأ المعيارى للمتوسط كالآتى:

$$SE_{\bar{Z}} = \sqrt{0.005} = 0.071$$

وحساب اختبار Z (ليس تحوير فيشر) للدلالة الإحصائية كالآتى:

$$Z = \frac{\bar{Z}_r}{SE_{\bar{Z}}} = \frac{0.52}{0.071} = 7.32$$

ووفترات الثقة كالآتى:

$$CI_{95\%\ lower} = \bar{Z} - Z\alpha_{/2} SE_{\bar{Z}} = -.52 - (1.96 \times .071) = -.66$$

$$CI_{95\,upper} = \bar{Z} + Z\alpha_{/2} SE_{\bar{Z}} = -.52 + (1.96 \times .071) = -.38$$

ويتم رفض الفرض الصفرى بمعنى وجود علاقة ارتباطية سالبة ودالة إحصائيًا بين التحصيل وادمان الانترنت.

واختبار الاختلاف بين أحجام التأثير كالآتى:

$$Q = \sum w_i (Z_i - \bar{Z})^2 = wz^2 - \frac{(wz)^2}{w}$$

$$= 68.97 - \frac{(-111.03)^2}{215} = 11.63$$

يتضح وجود دلالة إحصائية بمعنى اختلاف بين أحجام التأثير وعلى ذلك فالمدخل الأفضل لتحليل بيانات هو:

نموذج التأثيرات العشوائية

لإجراء نموذج التأثيرات العشوائية لا بد من تقدير التباين داخل الدراسات v_i (نتجة خطأ المعاينة) والتباين بين الدراسات τ^2 وتقدر قيمة τ^2 كالآتى:

$$\tau^2(v_\theta) = \frac{Q - (K-1)}{(\sum w_i) - \frac{(\sum w_i^2)}{(\sum w_i)}}$$

$$= \frac{11.63 - (3-1)}{215 - \frac{18651}{215}} = \frac{9.63}{128.25} = 0.075$$

حساب الاوزان أو مقلوب تباين الخطأ المعيارى كالآتى:

$$w_i^* = \frac{1}{v_i^*}$$

حيث v_i^* كالآتى:

$$v_i^* = (v_i + 0.075)$$

$$= \left(\text{خطأ المعاينة للعينات} \right) 0.005 + \quad (\text{خطأ معاينة الدراسات}) 0.075$$

$$= 0.08$$

ويتم حساب المتوسط الموزون كالآتى:

$$\overline{z}^{*}_{r} = \frac{\sum w_i^* z_i}{\sum w_i^*} = \frac{-13.34}{32.34} = -0.41$$

وتباين المتوسط الموزون كالآتى:

$$v_i^* = \frac{1}{\sum w_i^*} = \frac{1}{32.34} = 0.31$$

والخطأ المعيارى للمتوسط الموزون لمعامل الارتباط المحور فى نموذج التأثيرات العشوائية:

$$SE_{\overline{z}^{*}_{r}} = \sqrt{0.031 =} 0.176$$

اختبار Z للدلالة الإحصائية كالآتى:

$$Z = \frac{\overline{z}^{*}_{r}}{SE_{\overline{z}^{*}_{r}}} = \frac{0.41}{0.176} = 2.33$$

لاحظ أن الاختبار دال إحصائيًا عند 0.05 وليس دال إحصائيًا عند مستوى دلالة إحصائية 0.01 وهذا يختلف عن نتائج النموذج المثبت.

وفترات الثقة:

$$CI_{95\%\ lower} = \overline{Z}_r^* - Z\alpha_{/2} SE_{\overline{Z}_r^*} = -.41 - (1.96 \times 0.176) = -.66$$

$$CI_{95\ upper} = \overline{Z}_r^* + Z\alpha_{/2} SE_{\overline{Z}_r^*} = -.41 + (1.96 \times 0.176) = -.38$$

ولتفسير النتائج يجب تحويل تحوير فيشر إلى معامل الارتباط كالآتى:

$$\overline{r} = \frac{e^{2\overline{Z}} - 1}{e^{2\overline{Z}} + 1}$$

بالتعويض بقيمة $\overline{z}$ فى النموذج المثبت كالآتى:

$$r = \frac{e^{2\bar{Z}} - 1}{e^{2\bar{Z}} + 1} = \frac{e^{2\times -0.52} - 1}{e^{2\times -0.52} + 1} = -0.48$$

بالتعويض بقيمة $\overline{z}^*_r$ فى النموذج العشوائى:

$$r = \frac{e^{2\bar{Z}} - 1}{e^{2\bar{Z}} + 1} = \frac{e^{2\times -0.41} - 1}{e^{2\times -0.41} + 1} = -0.39$$

وعليه فإن نتائج ما وراء التحليل فى ضوء معامل الارتباط كالآتى:

النموذج المثبت: $-0.48 (CI_{95} : -0.57, -0.36)$

النموذج العشوائى: $-0.39 (CI_{95} : -0.64, \ -0.07)$

مقارنة بين المداخل

تم تحليل نتائج الدراسات الثلاثة باستخدام مدخل Hunter & Schmidt (2004) ومدخل Hedges et al ويوجد فروق مختلفة بين المدخلين وفيا يلى معادلات المدخلين:

الجدول (6.22): المعادلات البديلة المستخدمة فى المدخلين (Ellis, 2010).

		Hunter and Schmidt		Hedges et al.	
Output	Generic	*d*	*r*	*d*	*r*
Weighted mean ES	$\overline{ES} = \frac{\sum w_i ES_i}{\sum w_i}$	$\bar{d} = \frac{\sum w_i d_i}{\sum d_i}$	$\bar{r} = \frac{\sum n_i r_i}{\sum n_i}$	$\bar{d} = \frac{\sum w_i d_i}{\sum w_i}$	$\bar{z} = \frac{\sum w_i z_i}{\sum w_i}$
Variance of sample ESs	$v_i = \frac{1}{\sum w_i}$	$v_{wd} = \frac{\sum w_i (d_i - \bar{d})^2}{\sum w_i}$	$v_{wr} = \frac{\sum n_i (r_i - \bar{r})^2}{\sum n_i}$	$v_i = \frac{1}{\sum w_i}$	$v_i = \frac{1}{\sum w_i}$
z score	$z = \frac{\lvert\overline{ES}\rvert}{SE_{\overline{ES}}}$	$z = \frac{\lvert\bar{d}\rvert}{SE_{\bar{d}}}$	$z = \frac{\lvert\bar{r}\rvert}{SE_{\bar{r}}}$	$z = \frac{\lvert\bar{d}\rvert}{SE_{\bar{d}}}$	$z = \frac{\lvert\bar{z}\rvert}{SE_{\bar{r}}}$
Homogeneity statistic	$Q = \sum w_i (ES_i - \overline{ES})^2$	-	$\chi^2_{k-1} = \sum \frac{(n_i - 1)(r_i - \bar{r})^2}{(1 - \bar{r}^2)^2}$	$Q = \sum w_i (d_i - \bar{d})^2$	$Q = \sum w_i (z_i - \bar{z})^2$

- ES حجم التأثير.

- $\overline{ES}$ متوسط حجم التأثير (r, d, zr).
- K عدد الدراسات أو أحجام التأثير.
- Wi وزن الدراسة.
- SE الخطأ المعيارى.

مصدر المعادلات: (Hedges & Vevea, 1998; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001).

والجدول يعرض الفروق بين المدخلين فى أربع معادلات أساسية هى متوسط حجم التأثير الموزون، التباين الخطأ المعيارى، اختبار Z للدلالة، واختبار تقدير الاختلاف بين أحجام التأثير. فالصيغة العامة أو المعيارية Generic تم عرضها فى العمود الأول اما الأعمدة الباقية توضح صيغ حسابات مداخل Hedges & Vevea (1998) و Hunter & Schmidt (2004). والفروق بين هذه المداخل كالتالى:

الأول: معامل الارتباط أو حجم التأثير بين المدخلين حيث حول Hedges معاملات الارتباط إلى تحويرات فيشر Z_r فى حين اعتمد Hunter & Schmidt على معاملات الارتباطات مباشرة دون تحوير. فالارتباطات المحورة تصحح متوسط معامل الارتباط من التحيز السالب الصغير ولكن فى نفس الوقت تسهم بقدر صغير من التحيز الموجب للنتائج. والاختيار بين تحليل معامل الارتباط 0.454−(خطأ المعاينة) او0.50- (خطأ المعاينة وخطأ القياس) وتحليل تحويرات معامل الارتباط (المثبت 0.48− والعشوائى0.39−) وهذا تحيز سالب يعتمد على مدى تفضيل الباحث للتحيزات الموجبة والسالبة المحدودة للنتائج.

الثانى: التباين فى البيانات الذى بدورة يؤثر على الاوزان التى تعطى للتقديرات وتستخدم فى حساب تباين المعاينة والاخطاء المعيارية و فترات الثقة. فعند دمج معاملات الارتباط فى ضوء مدخل H.S يكون قائم على حجم العينة لكل دراسة بينما الاوزان فى ضوء مدخل Hedges فى ضوء N-3 ، ولكن الشئ المفيد فى مدخل H.S هو إجراء تصحيحات لاحجام التأثير فى ضوء محددات الدراسة مثل نقص الثبات والصدق والتصنيفية الثنائية بينما فى مدخل Hedges للنموذج العشوائى يدخل تصحيح

آخر للتباين بين أحجام التأثير نتيجة عينة الدراسات وهذا يفسر الاختلاف فى النتائج كالآتى:

مدخل Hunter & Schmidt المصحح فى ضوء خطأ المعاينة فقط:

$$-0.454(CI_{95}: -0.693\ to -0.215)$$

مدخل Hunter & Schmidt المصحح فى ضوء تصحيح من نقص الثبات وخطأ المعاينة:

$$-0.500\ (CI_{95}: -0.755\ \ to -0.245)$$

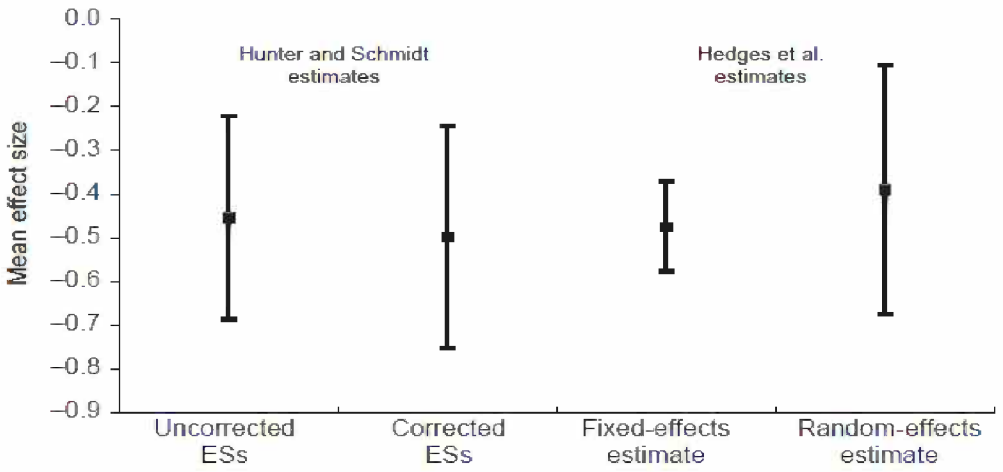
مدخل Hedges et al (التأثيرات المثبتة):

$$-0.478\ (CI_{95}: -0.572\ \ to -0.365)$$

مدخل Hedges et al (التأثيرات العشوائية):

$$-0.391\ (CI_{95}: -0.639\ \ to -0.068)$$

ويمكن عرض المتوسطات للمداخل الاربعة فى الشكل الاتى(2010)Ellis:



الشكل (9.22): عرض متوسط أحجام التأثير فى المداخل الاربعة.

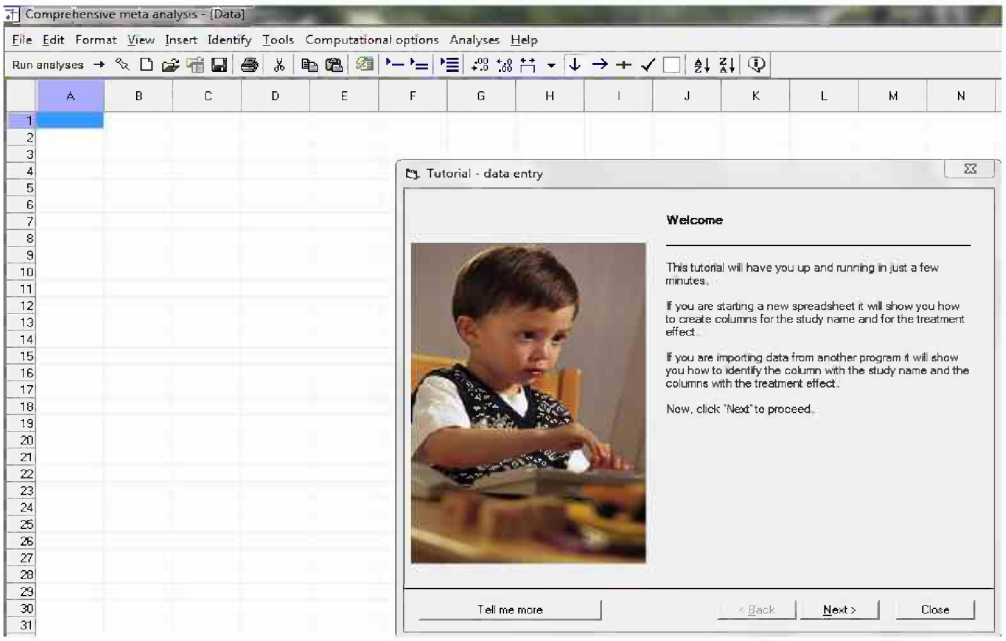
ويلاحظ الاختلاف الواضح بين المتوسطات فى المداخل الاربعة وتساءل Ellis (2010) ماهى النتائج الاكثر دقة؟ هذا السؤال نال اهتمام كثير من الباحثين امثال(Field, 2005; Hall & Brannick,2002; Schulze, 2004) ، ففى ضوء دراسة المحاكاة لـ Field (2005) توصل إلى ان طرق Hedges وزملائة تنتج

تقديرات دقيقة لفترات الثقة بينما طريقة Hunter & Schmidt تعطى تقديرات دقيقة لتقديرات المتوسط ويعطى مدى ضيق لفترات الثقة عما هو يجب أن يكون بمعنى أنه يستبعد حجم تأثير حقيقى من مدى حجم التأثير المقدر من خلاله وهذا يتفق مع نتائج المثال السابق عرضه حيث اوسع مدى لحجم التأثير من خلال طريق Hedges باستخدام نموذج التأثيرات العشوائية حيث أكبر بنسبة 12% من فترات الثقة من خلال طريقتى Hunter & Schmidt. واوصى Field (2005) بان نضع معظم جهودنا فى تحليلات ما وراء التحليل فى ضوء مدخل Hunter & Schmidt وفى نهاية هذه المناقشة نؤكد على ما أشار إليه Schulze (2004) بان بعض المداخل أفضل من الأخرى لمهام مختلفة ولكن الطريقة الوحيدة الأفضل هى مجموعة من الإجراءات يجب توليفها. وقام Hall & Brannick (2002) بمقارنة نموذج التأثيرات العشوائية فى ضوء مدخلى Hedges & Vevea (1998) و Hunter & Schmidt (1990) خصوصًا فيما يخص بمدى جدوى تحوير معامل الارتباط إلى إحصاء فيشر التى اعتمد عليها Hedges، واستخدموا طريقة المحاكاة لمونت كارلو توصلوا إلى فروق طفيفة فى متوسط حجم التأثير وبعض الفروق الملاحظة فى فترات الثقة لصالح مدخل Hunter & Schmidt

ما وراء التحليل باستخدام برنامج Comprehensive Meta-Analysis(CMA) Version 3

تم تطوير هذا البرنامج على يد مجموعة من الباحثين هم Borenstein, :Hedges, Higgins, Rothstein ومجموعة كبيرة من الهيئة الاستشارية امثال Harris Cooper وEgger Matthias و Mark Lipsey وغيرهم. ويتم تنزيل البرنامج على جهاز الكمبيوتر على سطح المكتب أو فى قائمة كل البرامج.

افتح البرنامج من خلال Start ثم All Programs ثم اضغط Comprehensive Meta-Analysis تظهر الشاشة الآتية:

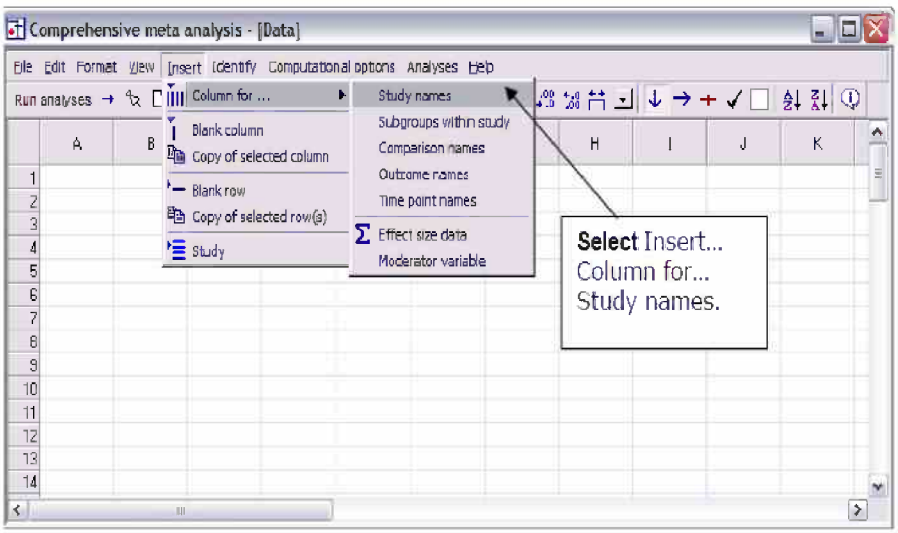


- اضغط Close.

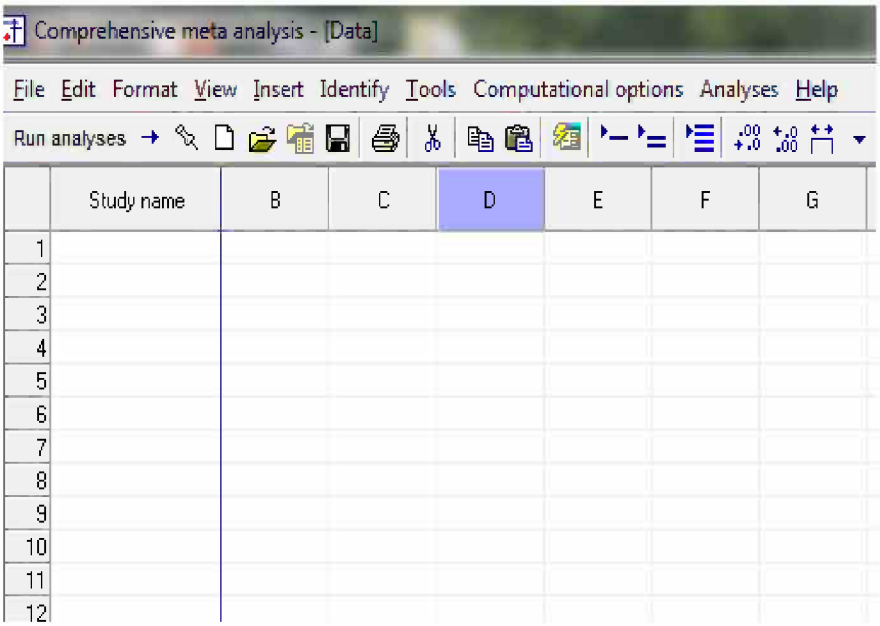
الجزء الأول: إدخال البيانات الأساسية والتحليل

لإدخال البيانات **إنشاء عمود لأسماء الدراسات** كالتالى:

Insert……Colum for…… Study names

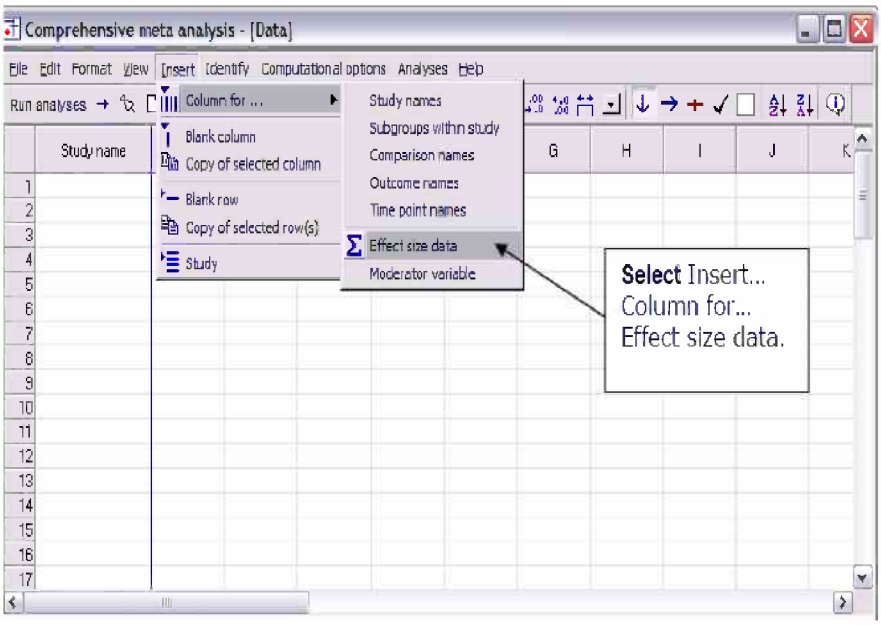


يقوم البرنامج بحشر أسماء الدراسات أو مؤلف الدراسات كالتالى:

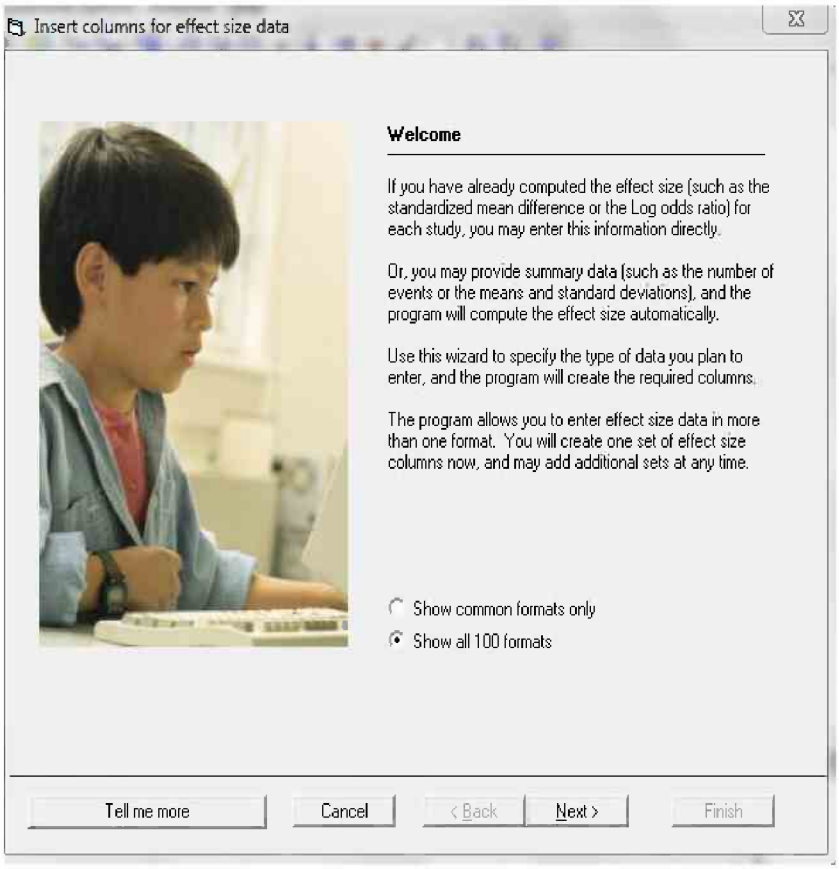


إنشاء أعمدة لبيانات حجم التأثير كالتالى:

Insert……Colum for…… Effect Size data



تظهر الشاشة الآتية:



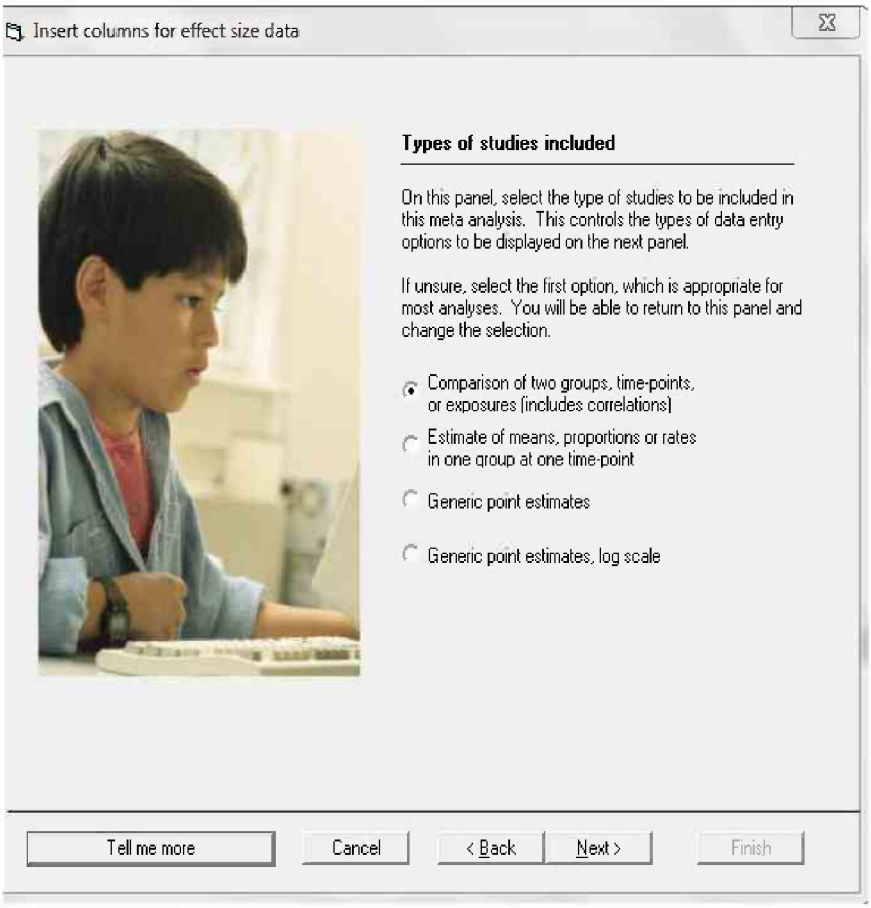
هذه الشاشة تمدنا بخيارات لإدخال بيانات حجم التأثير كالآتى:

الأول: لو قدر الباحث حجم التأثير مباشرة مثل مؤشرات فروق المتوسطات المعيارية أو لوغاريتم Odds ratio لكل دراسة فادخلها مباشرة فى عمود حجم التأثير.

الثانى: إذا ادخلت البيانات التلخيصية مثل عدد الاحداث أو الحالات أو المتوسطات والانحرافات المعيارية للمجموعات فإن البرنامج يقدر أحجام التأثير مباشرة من تلك البيانات.

من الشاشة السابقة حدد نوعية طبيعة إدخال البيانات المراد ادخالها ثم يقوم البرنامج بإنشاء الأعمدة المتطلبة لذلك. ويسمح البرنامج بإدخال البيانات بأشكال عديدة.

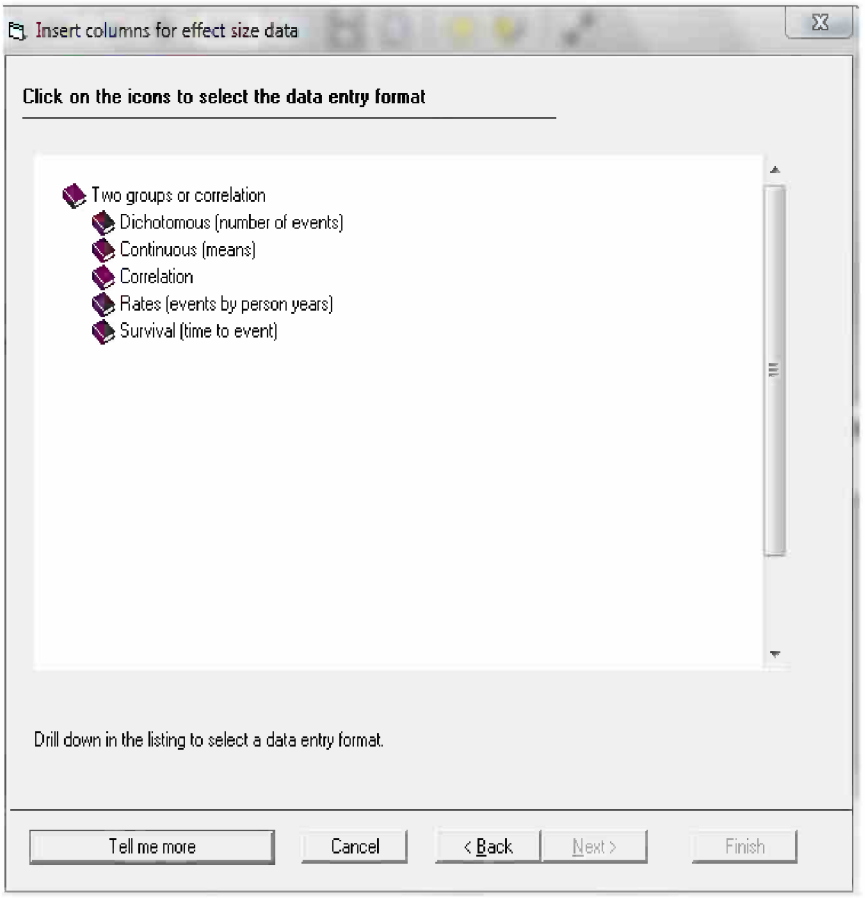
- اضغط على اختيار Next تظهر الشاشة الآتية:



هذه الشاشة تسمح للباحث اختيار نوعية بيانات حجم التأثير كالتالى:

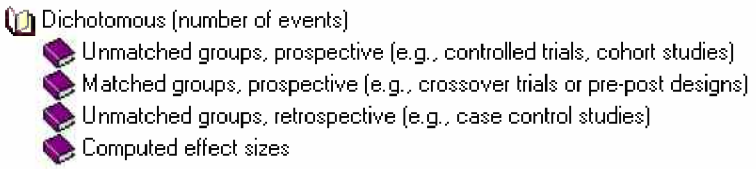
- المقارنة بين مجموعتين أو التدخلات وتتضمن العلاقات.
- تقدير المتوسطات، النسب، والتقديرات لمجموعة واحدة.
- التقديرات اللوغاريتمية.

- اضغط على البديل الأول المتضمن المقارنة بين مجموعتين أو العلاقات ثم اضغط Next تظهر الشاشة الآتية:



يظهر عدة بدائل منها:

- البيانات التصنيفية الثنائية Dichotomous وتتضم عدة خيارات فرعية أهمها:

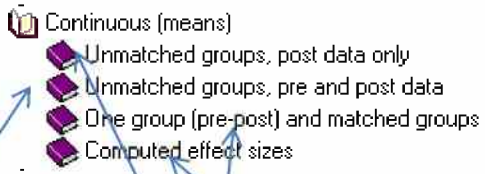


– المجموعات غير المرتبطة Unmatched groups وتحتها تتضمن الاحداث وحجم العينة فى كل مجموعة، وعدم الأحداث والعينة فى كل مجموعة، والاحداث أو عدم الاحداث فى كل مجموعة، واختبار كاى تربيع وحجم العينة.

– المجموعات المترابطة Matched groups

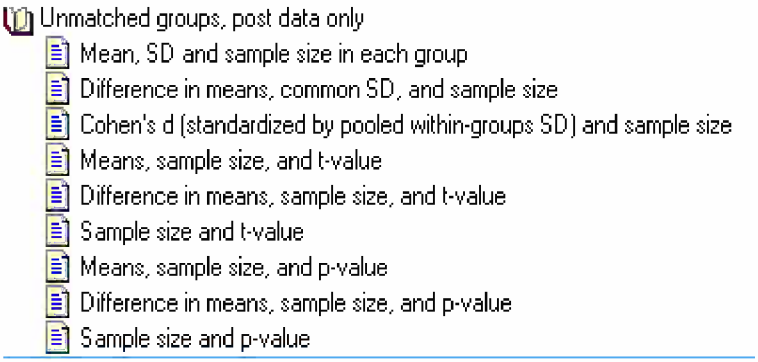
– أحجام التأثير المحسوبة.

- النواتج المتصلة Continous: وتتضمن:

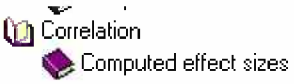


وتتضمن المجموعات المستقلة للبيانات البعدية فقط، بيانات المجموعة غير المتزاوجة وقياس قبلى وبعدى، وبيانات المجموعة الواحدة قبلى بعدى والمجموعات المتزاوجة، وحجم التأثير.

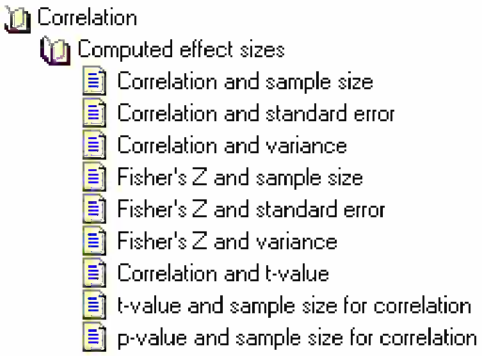
وإذا ضغط الباحث على اختيار المجموعات المستقلة وقياس بعدى فقط تظهر الشاشة الخيارات الآتية:



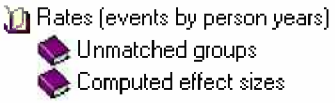
وعلى الباحث ان يختار شكل البيانات المراد ادخالها سواء المتوسط والانحراف المعيارى SD وحجم العينة فى كل مجموعة، أو مؤشر d لـ Cohen وحجم العينة أو حجم العينة وقيمة P الاحتمالية وغيرها.

- معامل الارتباط: 

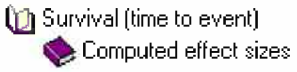
: قيم معامل الارتباط مباشرة. بالضغط على Computed effect size تظهر اختيارات إدخال معامل الارتباط كالآتى:



- التقديرات:

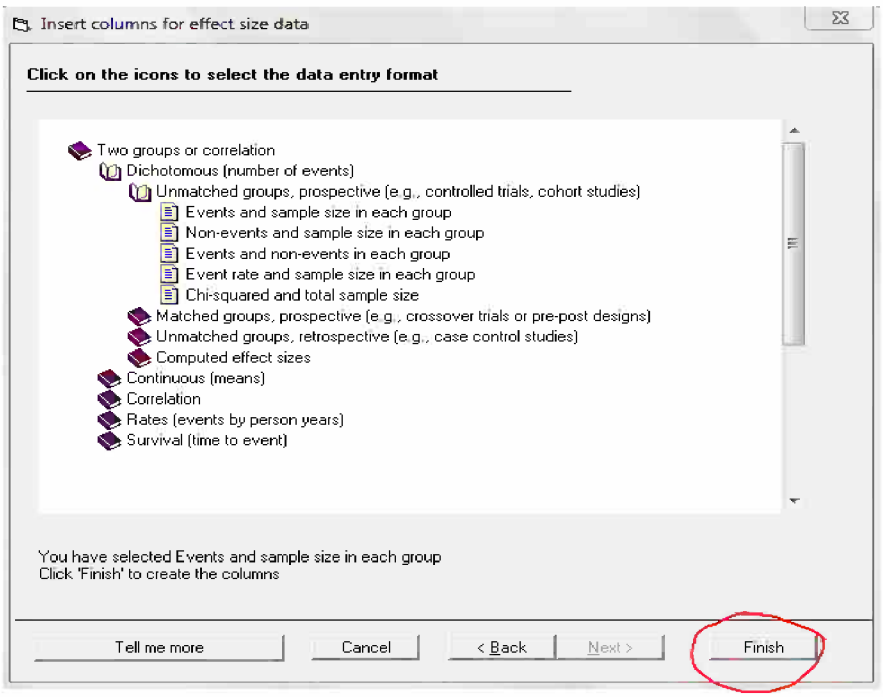


- البيانات الطولية:

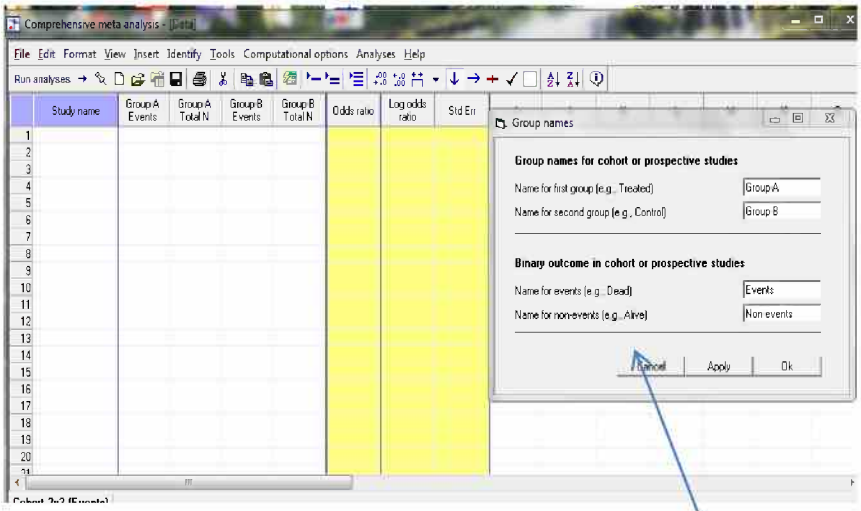


فاذا اهتم الباحث بدراسة ما وراء التحليل لدراسات تجريبية حيث المتغير المستقل والتابع تصنيفى ثنائى كما فى الدراسات الطبية وفى هذة الحالة فإن مؤشر حجم التأثير Odds ratio(OR) فعلى الباحث اتباع الاتى:

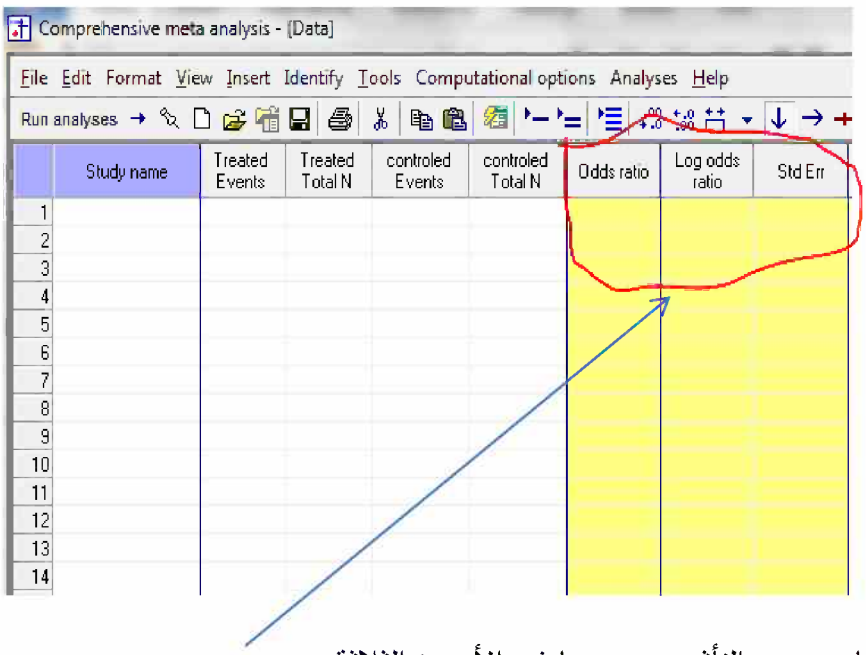
- اضغط Dichotomous ثم unmatched groups ثم اختار Events and sample size in each group تظهر الشاشة الآتية:



- اضغط Finish تظهر الشاشة الآتية المتضمنة الأعمدة لإدخال البيانات:



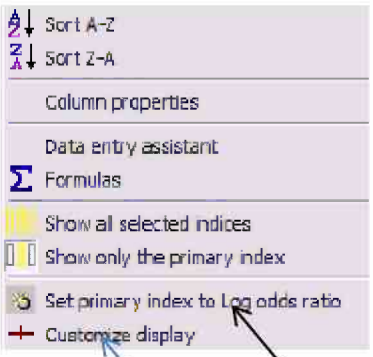
وفى المربع الفرعى يمكن تغير مسمى المجموعات فبدلًا من Group A يمكن تسميتها بالمجموعة المعالجة Treated والمجموعة Group B يمكن تسميتها بالمجموعة الضابطة Control ثم اضغط Apply تظهر الشاشة الآتية:



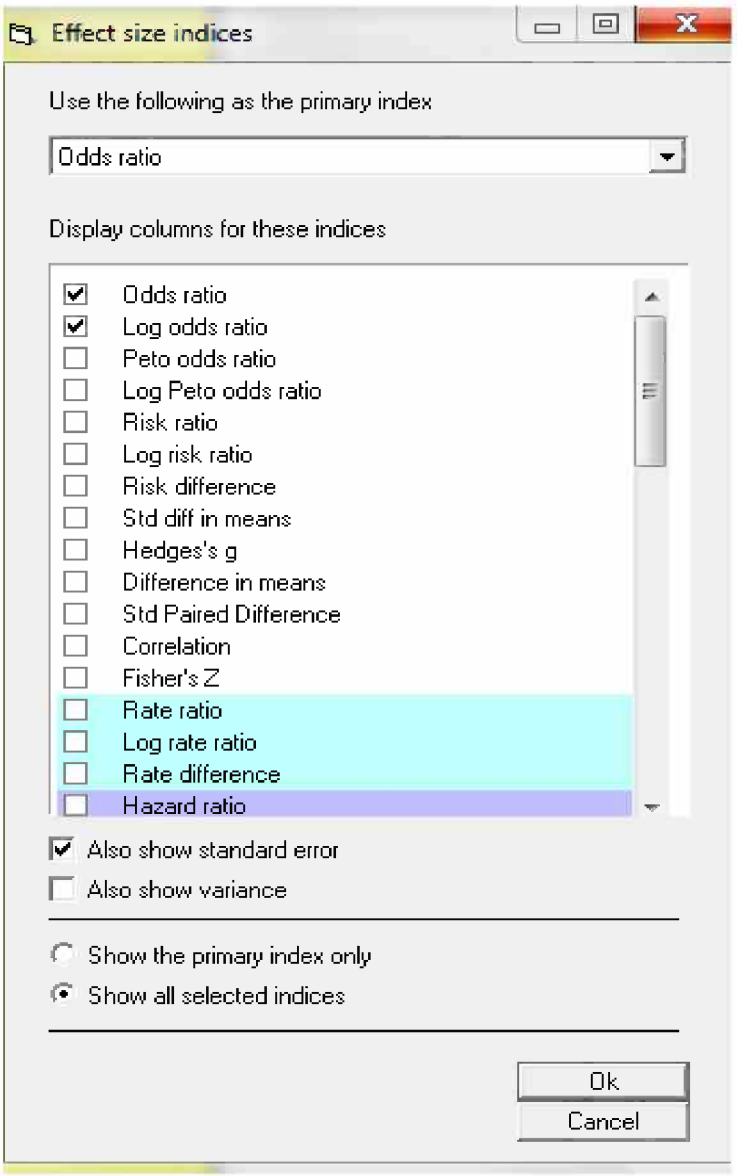
يحسب البرنامج حجم التأثير ويعرضها فى الأعمدة الثلاثة.

ويمكن استعراض مؤشرات اخرى لحجم التأثير من خلال:

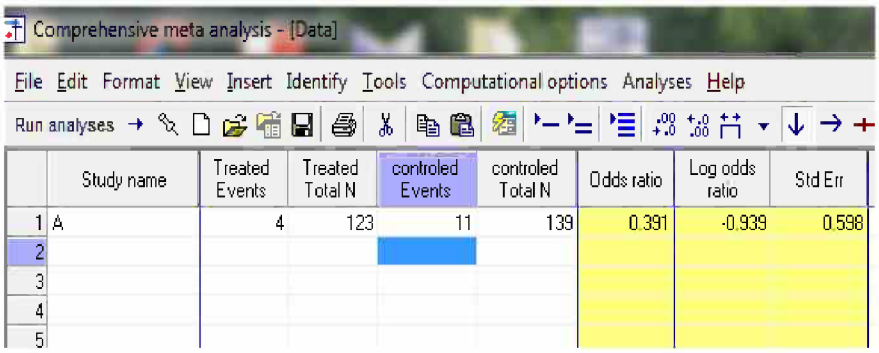
- اضغط كليك يمين على عمود Odds ratio تظهر قائمة كالآتى:



- اضغط set primary index أو Customized Display تظهر المؤشرات الآتية:

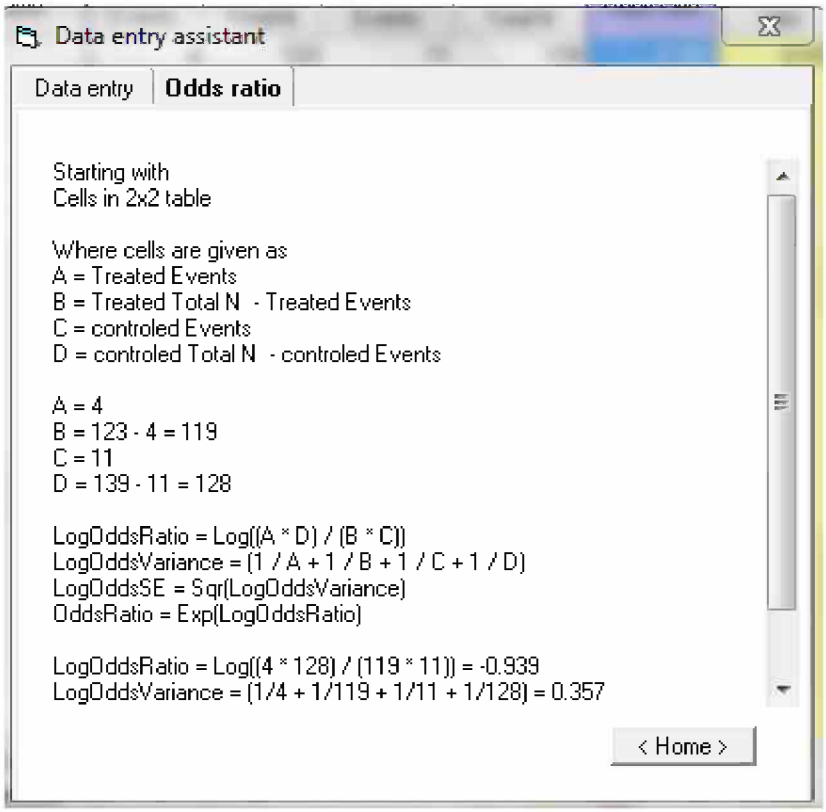


- ابدأ بادخل البيانات كما فى الشكل الآتى واضغط على خانة Odds ratio يقوم بحساب مؤشرات حجم التأثير ويعرضها فى الأعمدة الصفراء الثلاثة:

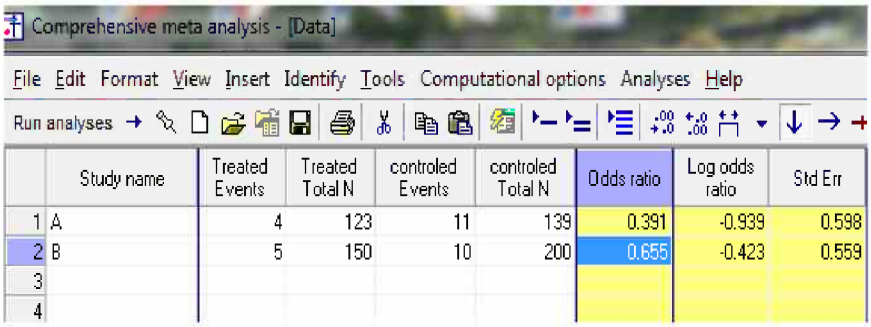


	Study name	Treated Events	Treated Total N	controled Events	controled Total N	Odds ratio	Log odds ratio	Std Err
1	A	4	123	11	139	0.391	-0.939	0.598
2								
3								
4								
5								

ولاظهار معادلات حساب مؤشرات حجم التأثير اضغط مرتين بالموس على خانة أو المربع 0.391 فى عمود Odds ratio تظهر الشاشة الفرعية الآتية:



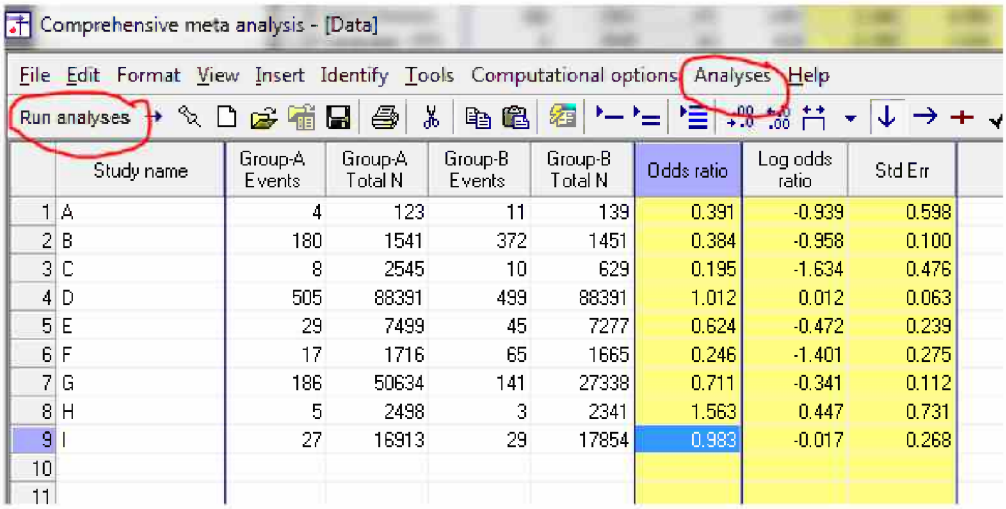
ثم يقوم الباحث بإدخال بيانات دراسة جديدة ويقوم بالضغط على مربع إحصاء Odds ratio لحساب مؤشرات حجم التأثير كالآتى:



	Study name	Treated Events	Treated Total N	controled Events	controled Total N	Odds ratio	Log odds ratio	Std Err
1	A	4	123	11	139	0.391	-0.939	0.598
2	B	5	150	10	200	0.655	-0.423	0.559
3								
4								

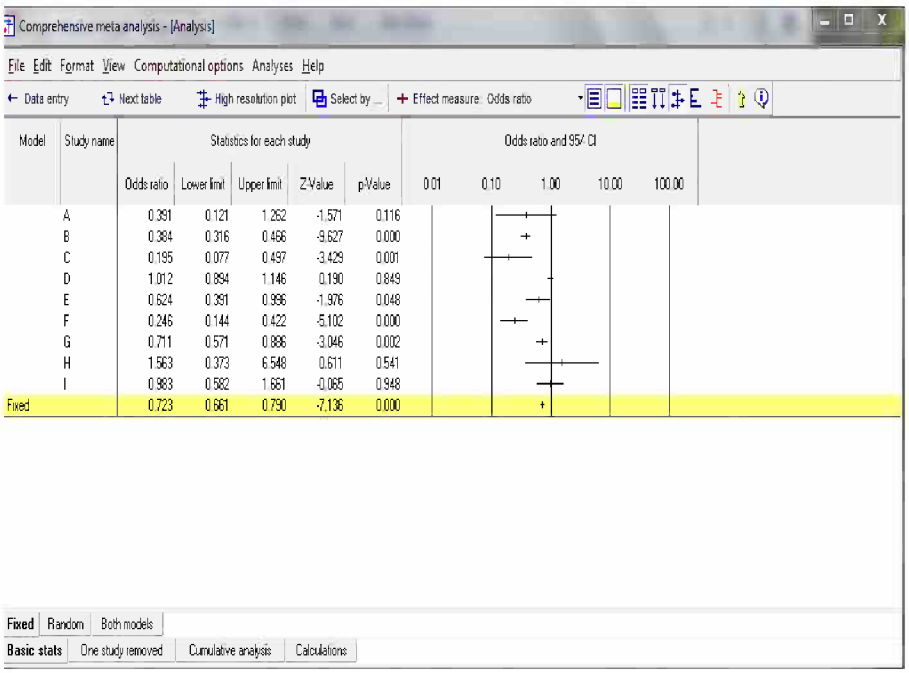
إجراء التحليل

ويقوم الباحث بإدخال باقى بيانات دراسة ما وراء التحليل وإجراء التحليل كالآتى:



	Study name	Group-A Events	Group-A Total N	Group-B Events	Group-B Total N	Odds ratio	Log odds ratio	Std Err
1	A	4	123	11	139	0.391	-0.939	0.598
2	B	180	1541	372	1451	0.384	-0.958	0.100
3	C	8	2545	10	629	0.195	-1.634	0.476
4	D	505	88391	499	88391	1.012	0.012	0.063
5	E	29	7499	45	7277	0.624	-0.472	0.239
6	F	17	1716	65	1665	0.246	-1.401	0.275
7	G	186	50634	141	27338	0.711	-0.341	0.112
8	H	5	2498	3	2341	1.563	0.447	0.731
9	I	27	16913	29	17854	0.983	-0.017	0.268
10								
11								

- اضغط على Run analyses أو اضغط على قائمة Analysis لتنفيذ التحليل كالآتى:

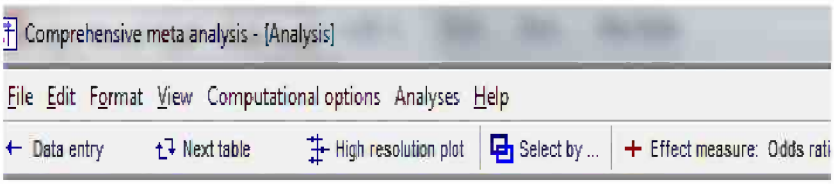


لاحظ أنه عرض الإحصائيات الخاصة بكل دراسة تحت الأعمدة المسمى Statistics for each study متضمنة مؤشر حجم التأثير Odds ratio و 95% فترات الثقة بينما الصف الاخير يبين الإحصاء التلخيص (متوسط حجم التأثير) حيث باستخدام النموذج المثبت فإن متوسط Odds ratio= 0.723 بفترات ثقة 0.66- إلى 0.709 وقيمة اختبار Z للدلالة الإحصائية لمتوسط حجم التأثير 7.00- وهى أكبر من 1.96 اى ان العلاقة بين المتغيرين دالة إحصائيًا بمعنى للمعالجة اثر فى احداث حسن.

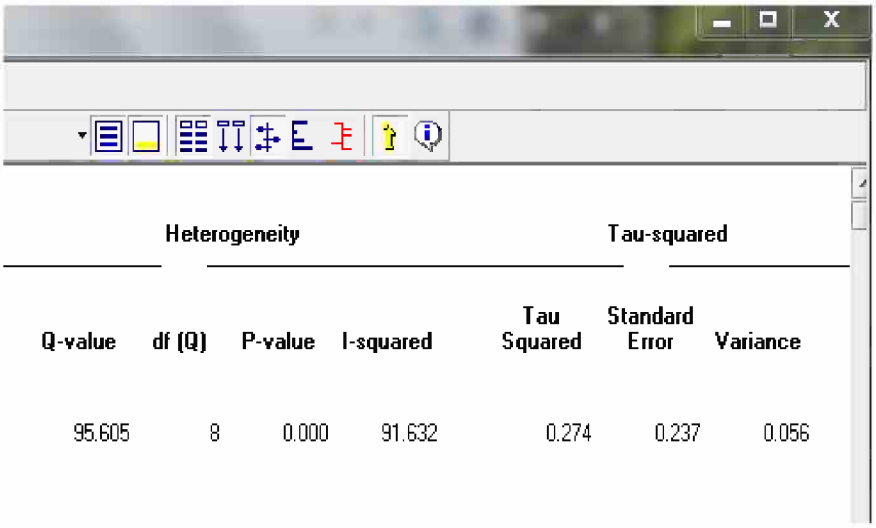
وعلى الناحية اليمنى عرض شكل Forest وهذا الشكل يوضح الحد الأعلى والأدنى لفترات الثقة ويعطى انطباع بمدى ابتعاد الدراسات عن بعضها البعض. وهذة الشاشة تتضمن عدة ايقونات منها:

عرض ملخص للنتائج

- اضغط Next table تظهر الجدول الاتى:



Model		Effect size and 95% interval			Test of null (2-Tail)	
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value
Fixed	9	0.723	0.661	0.790	-7.136	0.000
Random	9	0.556	0.373	0.828	-2.887	0.004

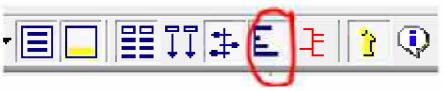


Heterogeneity				Tau-squared		
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance
95.605	8	0.000	91.632	0.274	0.237	0.056

فـى هـذا الجـدول تـم عـرض اختبـار عـدم التجـانس بـين الدراسـات Q وكانـت قيمتـة 95.60 ويـتم اختبـار دلالتهـا فـى ضـوء اختبـار كـاى تربيـع لـدرجات حريـة (8=1-9) ومسـتوى دلالـة 0.50 وهـى دالـة إحصـائيًا حيـث P=0.000 وعليـه يوجـد اخـتلاف بـين أحجـام التـأثير بالتـالى فـإن اسـتخدام النمـوذج المثبـت غيـر مناسـب والأفضـل الاعتمـاد على نتائج نموذج التأثيرات العشوائية.

عرض اوزان الدراسات

يمكن عرض اوزان الدراسات من خلال الضغط على ايقونة:



تعرض إحصائيات الدراسات المفردة.

تعرض الإحصائيات الملخصة.

تعرض الإحصائيات الأساسية المخفية.

تعرض الاعداد غير المرئية.

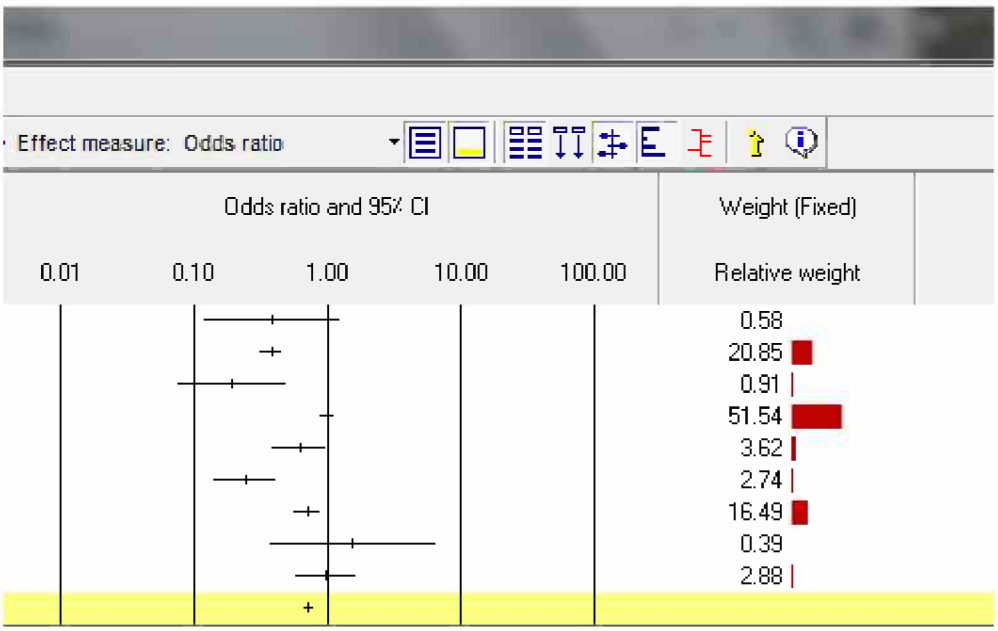
تعرض شكل Forest.

تعرض اوزان الدراسات.

تعرض البواقى.

تعرض تفاصيل.

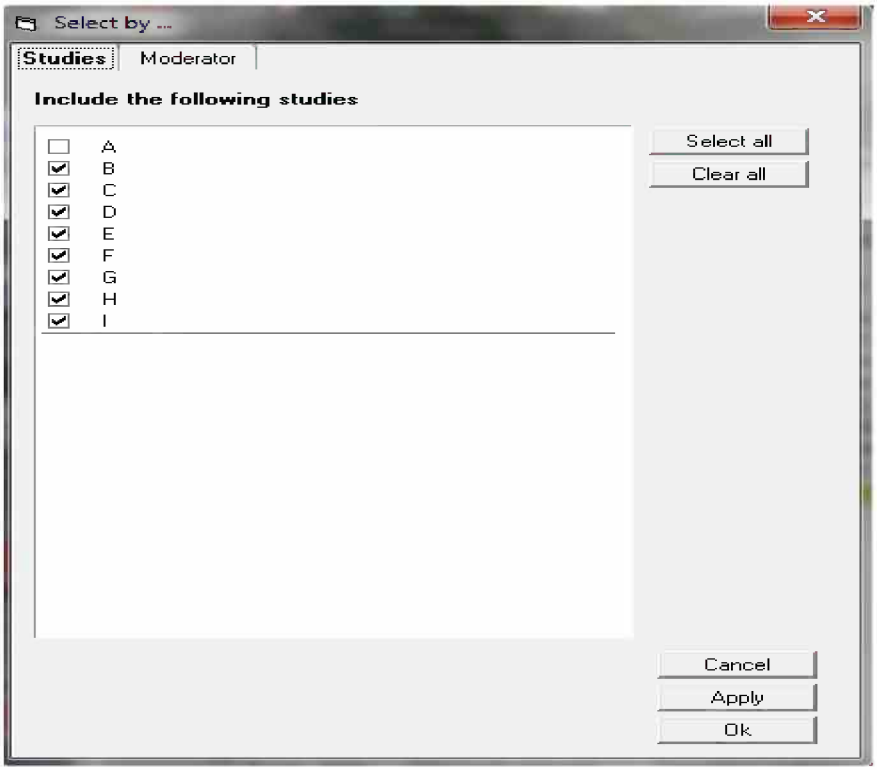
وبالضغط على تعطى الاوزان النسبية للدراسات كالآتى:



عرض نتائج ما وراء التحليل مع حذف إحدى الدراسات (تحليل الحساسية)

- اضغط على Select case:

يعطى الجدول الآتى:

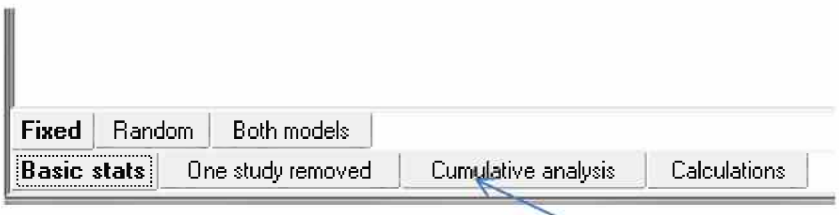


ثم اضغط على الدراسةA يستبعدها من التحليل.

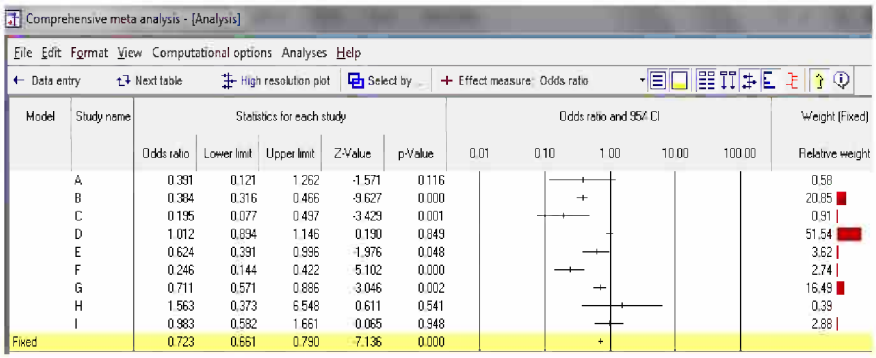
التحليل التراكمى Cumulative Meta-analysis

هو الية لعرض سلسلة منفصلة من تحليلات ما وراء التحليل فى جدول أو شكل واحد حيث تدمج الدراسة الأولى مع الثانية فى تحليل ثم الدراسات الثلاثة الأولى فى تحليل وهكذا، وهذا الدمج يتم فى ضوء بعض العوامل مثل التسلسل الزمنى أو الجودة المنهجية وهذا يبين كيفية تغير متوسط حجم التأثير عبر فترات زمنية ويمكن اعتبار هذا التحليل أداة تربوية للحكم على مدى تطور الممارسات البحثية التربوية عبر فترات زمنية متعاقبة (Borenstein et al.,2009).

من خلال الشريط الأسفل يوجد:



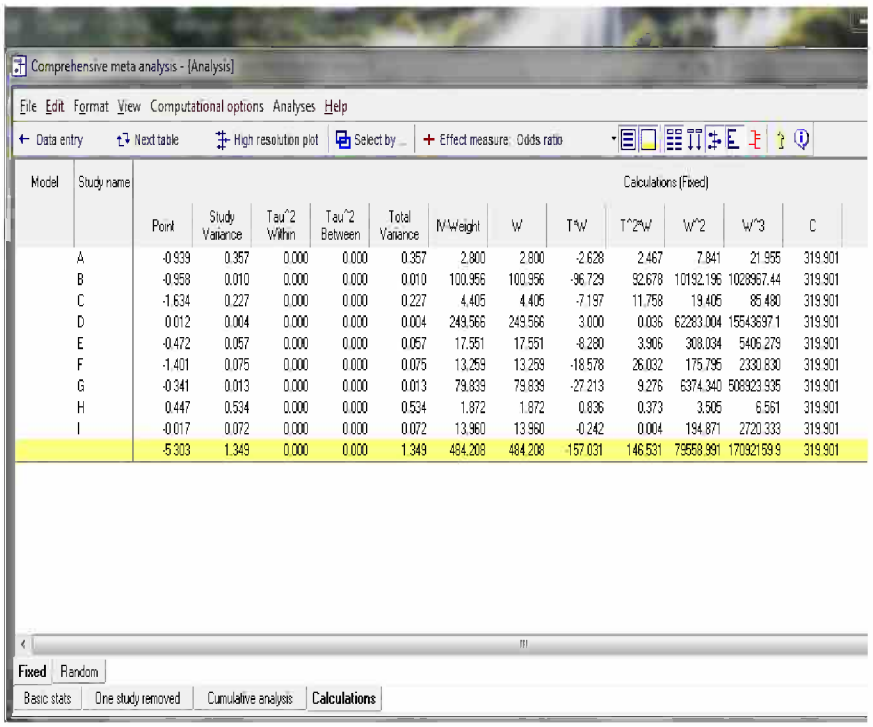
اضغط عليه يعطى النتائج الآتية:

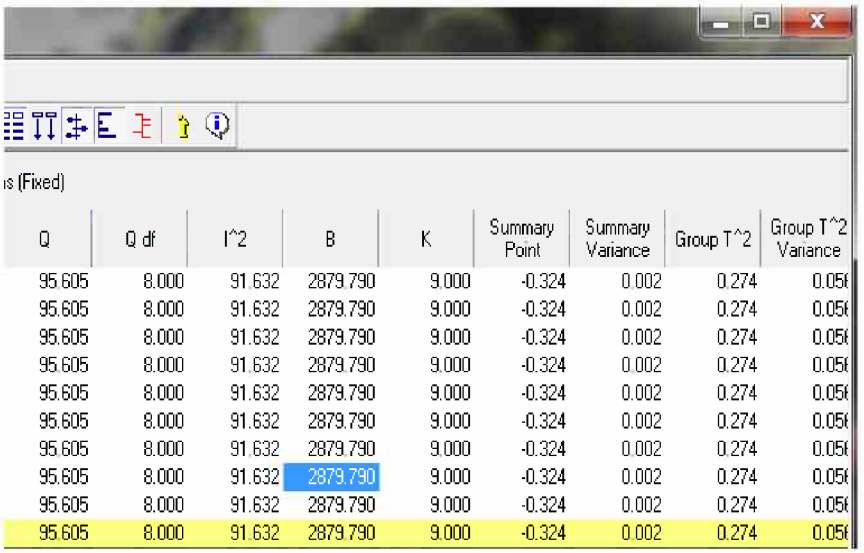


يقوم البرنامج بعرض النتائج متراكمة حيث الصف الثانى يعرض نتائج الدراستين الأولتين السابقيتين والصف الثالث يعرض نتائج الدراسات الثلاثة السابقة وهكذا.

عرض الحسابات التفصيلية

- اضغط علىCalculations أسفل الشاشة يعطى الحسابات التفصيلية الآتية:

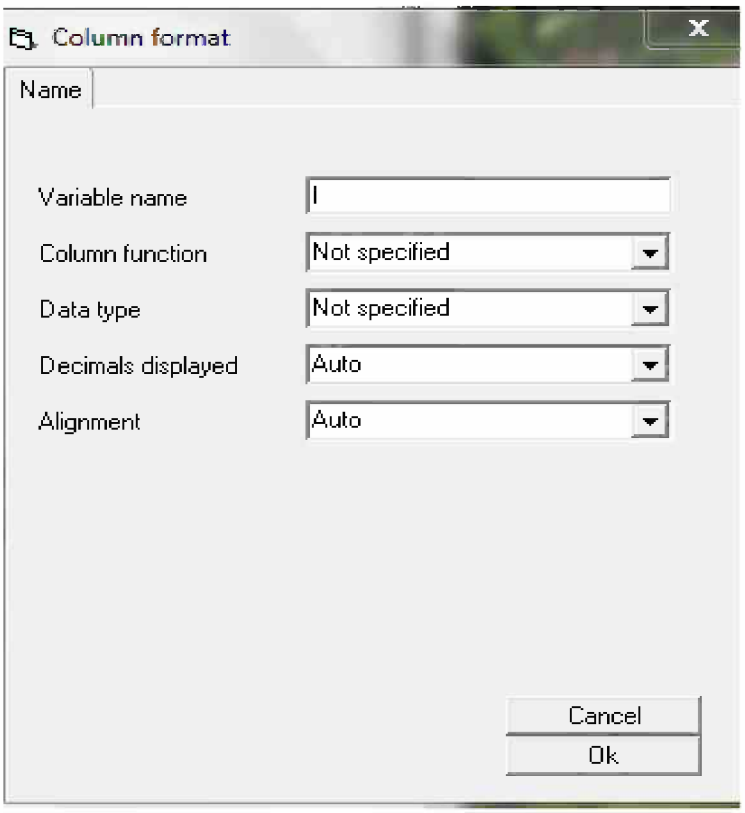




إجراء التحليلات المتفاعلة

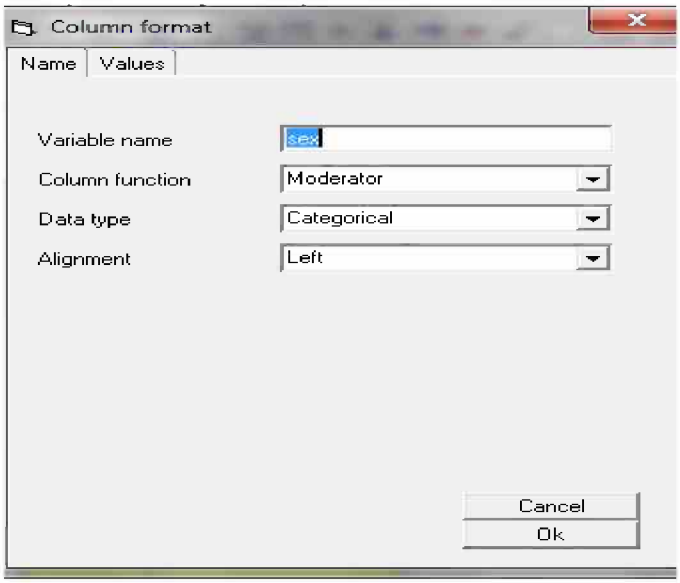
لإجراء التحليلات المتفاعلة باستخدام متغير تصنيفى (خاصية الدراسة) وليكن جنس العينة لا بد من إنشاء عمود المتغير المتفاعل كالتالى:

- اضغط مرتين متتاليتن بالموس على أى عمود لايتضمن بيانات تظهر الشاشة التالية:



● امـام function Column اضـغط علـى السـهم لأسـفل واختـار Moderator امـام data type اختار Categorical.

●اكتب مسمى المتغير التصنيفى وليكن الجنس Sex.



●اضغط OK يظهر مسمى المتغير فى شاشة البيانات.

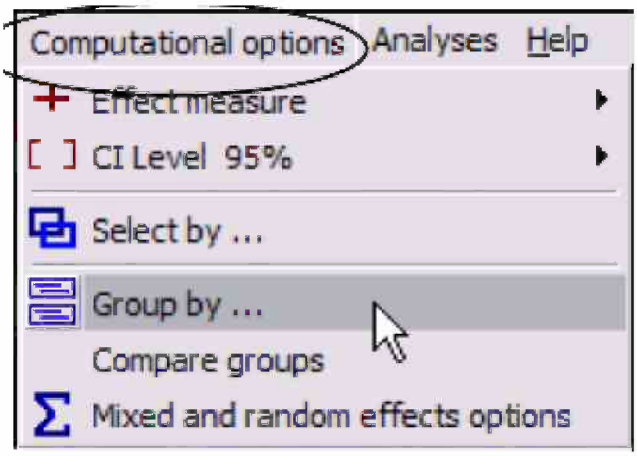
●ابدأ فى إدخال بيانات المتغير المتفاعل كما يلى:

File Edit Format View Insert Identify Tools Computational options Analyses Help

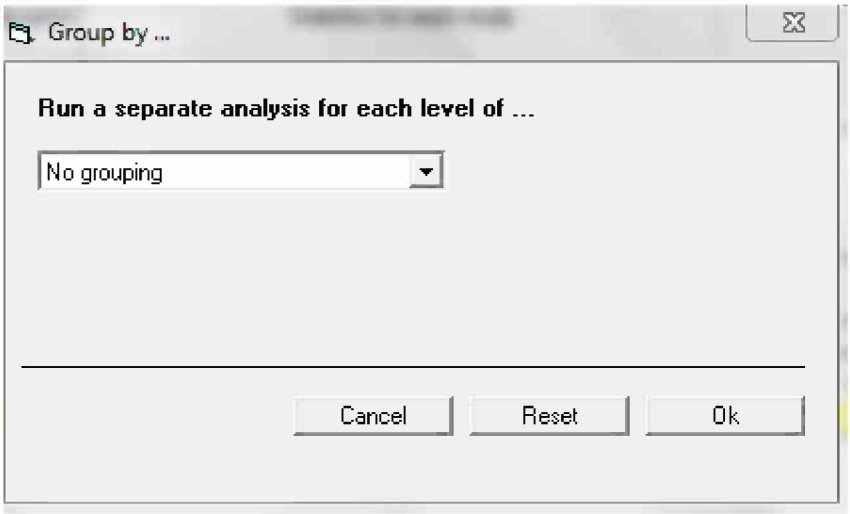
Run analyses

	Study name	Group-A Events	Group-A Total N	Group-B Events	Group-B Total N	Odds ratio	Log odds ratio	Std Err	sex
1	A	4	123	11	139	0.391	-0.939	0.598	M
2	B	180	1541	372	1451	0.384	-0.958	0.100	F
3	C	8	2545	10	629	0.195	-1.634	0.476	M
4	D	505	88391	499	88391	1.012	0.012	0.063	F
5	E	29	7499	45	7277	0.624	-0.472	0.239	M
6	F	17	1716	65	1665	0.246	-1.401	0.275	F
7	G	186	50634	141	27338	0.711	-0.341	0.112	M
8	H	5	2498	3	2341	1.563	0.447	0.731	F
9	I	27	16913	29	17854	0.983	-0.017	0.268	M
10									

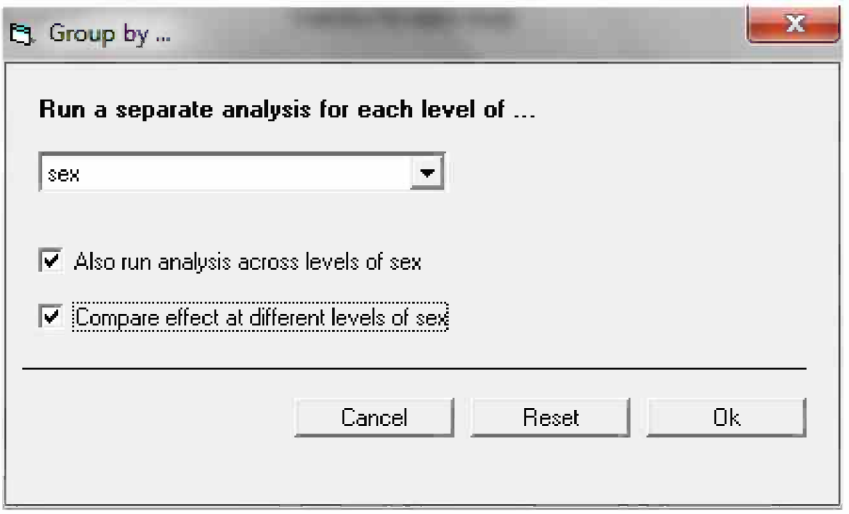
– اضـغط Run analyses ويظهـر المخـرج اضـغط علـى قائمـة Computational options تظهر الشاشة الآتية:



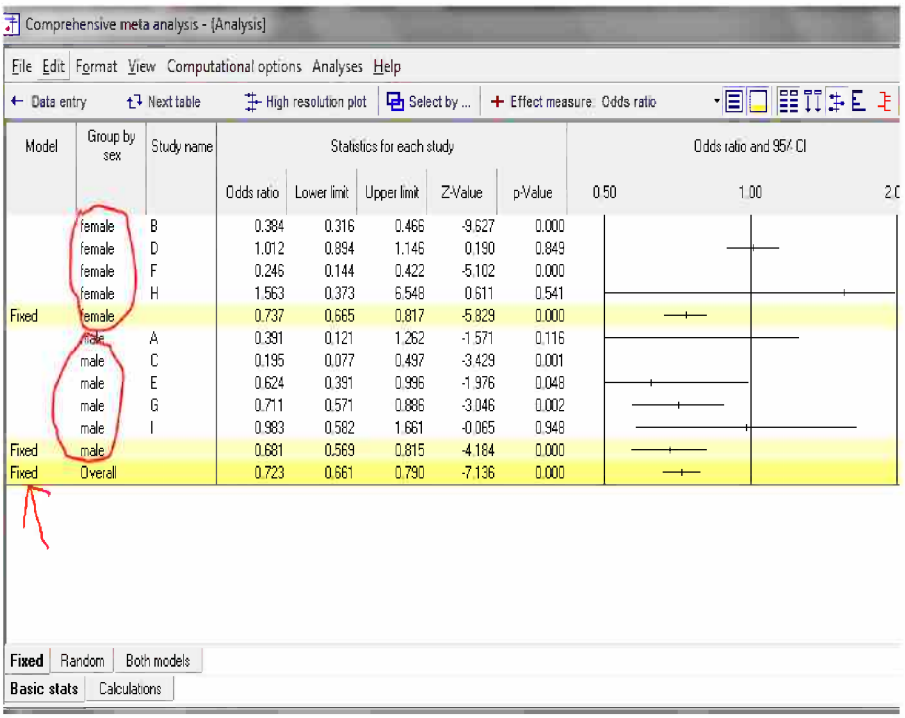
- اضغط على Group by يعطى المربع التى:



- اضغط على السهم الأسفل فى امام grouping No واختار المتغير المتفاعل Sex ونشط الايقونتين فى المربع وهذا يعنى إجراء التحليل لكل مستوى من مستويات المتغير التصنيفى وكذلك المقارنة بين المستويات المختلفة ويعطى الاختلاف بين وداخل المجموعات. وتكون المربع كالآتى:



- اضغط OK يظهر المخرج الاتى:



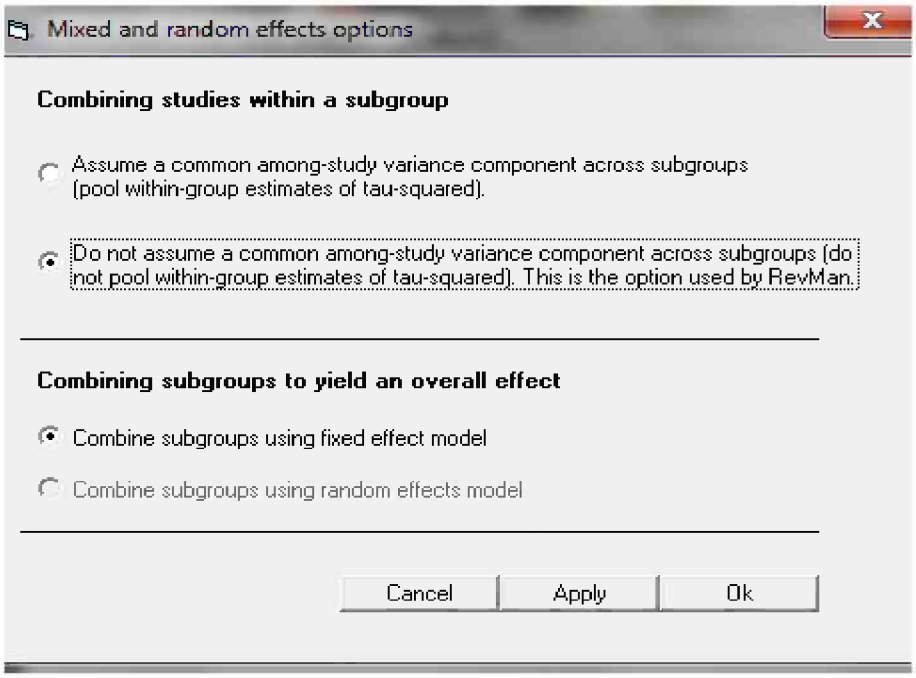
لاحظ أن المخرج اعطى إحصائيات كل مجموعة على حدة المظللة بالدائرة بالاضافة إلى إحصائيات المجموعتين معًا (المشار إليه بالسهم).

اختيار نموذج الحسابات أو التحليل

لاختيار نموذج التحليل:

- **اضغط** Computational options واختار Mixed and random effects

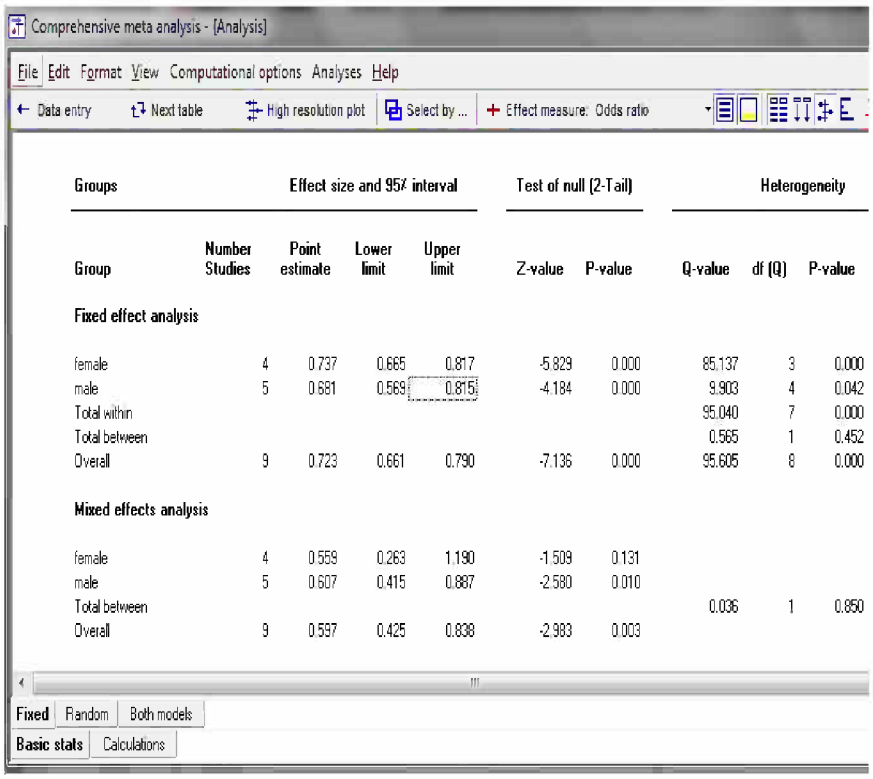
تظهر الشاشة الآتية:



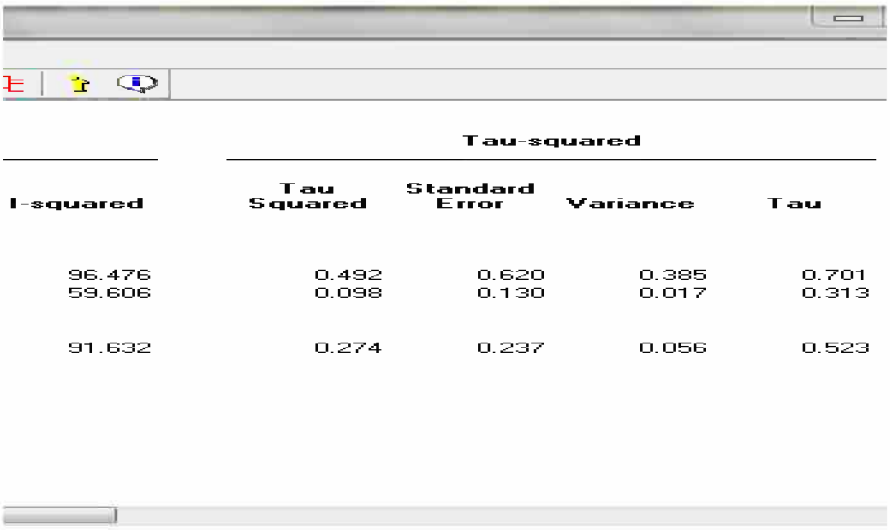
- اضغط Ok .

عرض الإحصائيات الإضافية لتحليل المجموعة

- اضغط على اختيار Next table يظهر الإحصائيات الإضافية لكل مجموعة بالاضافة إلى قيم التباين بين المجموعات وداخل المجموعات كالآتى:



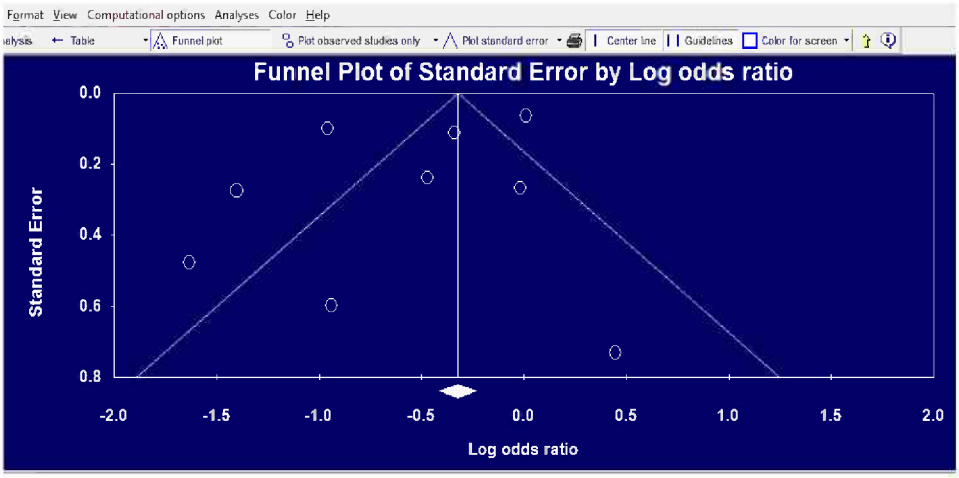
Groups		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity		
Group	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value
Fixed effect analysis									
female	4	0.737	0.665	0.817	-5.829	0.000	85.137	3	0.000
male	5	0.681	0.569	0.815	-4.184	0.000	9.903	4	0.042
Total within							95.040	7	0.000
Total between							0.565	1	0.452
Overall	9	0.723	0.661	0.790	-7.136	0.000	95.605	8	0.000
Mixed effects analysis									
female	4	0.559	0.263	1.190	-1.509	0.131			
male	5	0.607	0.415	0.887	-2.580	0.010			
Total between							0.036	1	0.850
Overall	9	0.597	0.425	0.838	-2.983	0.003			



I-squared	Tau-squared			
	Tau Squared	Standard Error	Variance	Tau
96.476	0.492	0.620	0.385	0.701
59.606	0.098	0.130	0.017	0.313
91.632	0.274	0.237	0.056	0.523

إجراء شكل القمع لتشخيص تحيز النشر

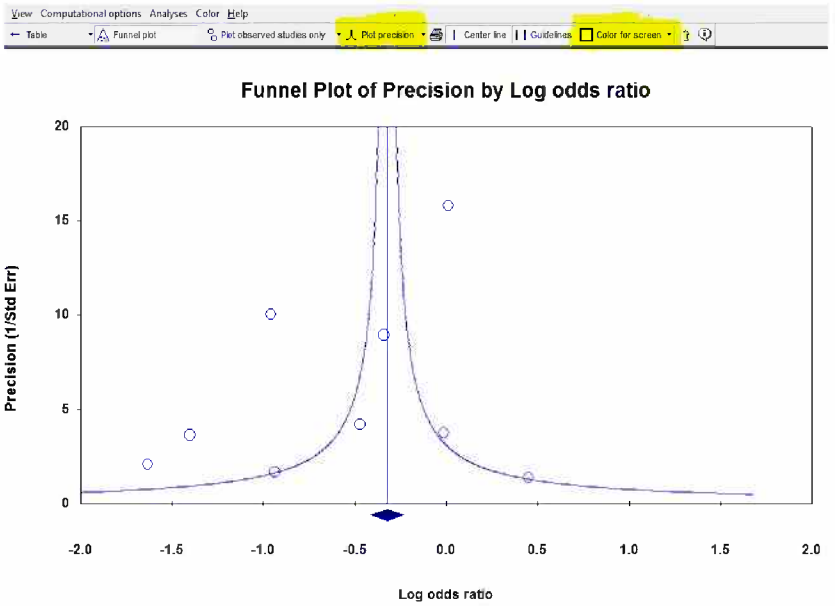
- اضغط على قائمة Analysis واختار Publication bias يعطى الشكل الآتى:



وتم عرض شكل القمع فى ضوء الخطأ المعيارى و لوغاريتم Odd ratio ويتضح من الشكل وجود تحيز نشر نتجة وقوع احداثيات عديدة خارج القمع وايضًا ان أحجام التأثير لا تتوزع اعتداليًا حول القمع. وإذا وقعت معظم الدراسات أسفل الشكل على جانب واحد من الشكل أكثر من جانب آخر فهذا يدل على عدم انتظامية شكل القمع بالتالى وجود تحيز للدراسات.

ويمكن عرض الشكل فى ضوء الدقة كالتالى:

-الضغط على Plots standard error أعلى الشكل واختيار Plot precision يظهر الشكل الاتى:



ويمكن تغير لون الشكل من خلال الايقونة المظللة بالأصفر Color of .screen

وتوجد طرق لتقدير التحيز النشر باستخدام إجراء احذف واملئ Trim and Fill وفكرته قائمة على شكل القمع ولتنفيذه:

- اضغط على قائمة View فى الشكل السابق واختار Trim and Fill يظهر المخرج الاتى:

Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0.72303	0.66141	0.79039	0.55612	0.37341	0.82823	95.60464
Adjusted values	0	0.72303	0.66141	0.79039	0.55612	0.37341	0.82823	95.60464

Look for missing studies where?

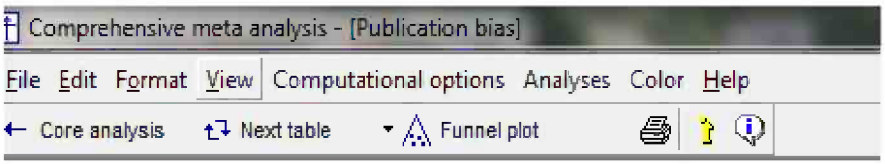
- ○ Not specified
- ◉ To left of mean
- ○ To right of mean

وبالضغط على الايقونة أعلى الشاشة يقدم تقرير وتفسير للنتائج. ولايوجد تأثير للدراسة الغائبة على تغير الشكل واعطى متوسط حجم التأثير وفترات الثقة لنموذج التأثيرات المثبتة والعشوائية.

معامل الارتباط الرتبى لتشخيص التحيز

وهو أحد طرق الاستدلال الإحصائى لتشخيص التحيز ويقدر من خلال معامل ارتباط كندال تاو b بين أحجام التأثير والاخطاء المعيارية المشتقة من أحجام العينات يقدر كالتالى:

- ضغط على قائمة View ثم اختار test correlation Rank يظهر المخرج التالى:



Begg and Mazumdar rank correlation

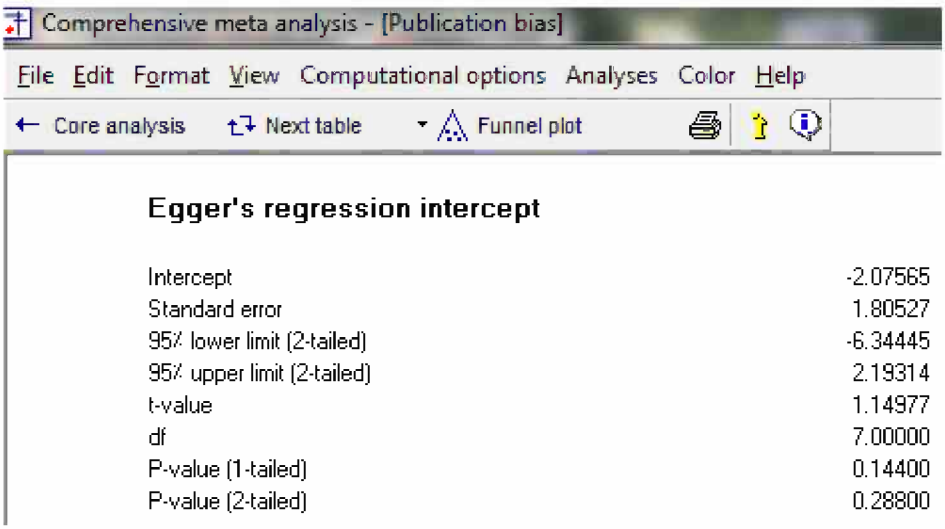
Kendall's S statistic (P-Q)	-2.00000
Kendall's tau without continuity correction	
Tau	-0.05556
z-value for tau	0.20851
P-value (1-tailed)	0.41741
P-value (2-tailed)	0.83483
Kendall's tau with continuity correction	
Tau	-0.02778
z-value for tau	0.10426
P-value (1-tailed)	0.45848
P-value (2-tailed)	0.91697

ويتضح عدم الدلالة الإحصائية حيث P=0.458 وعليه لا يوجد تحيز للنشر وهذا يتعارض مع تحليل شكل القمع وذلك لأن اختبار كندل ليس الاختبار الصادق فى هذة الحالة نظرًا لصغر حجم عينة الدراسات ونقص القوة الإحصائية للاختبار ولا يمكن الاعتماد علية فقط كدليل للحكم على تحيز النشر.

مدخل Egger Regression

هذا المدخل يعتمد على التنبؤ باحجام التأثير من خلال مؤشر الدقة (مقلوب الخطأ المعيارى) ويتم تنفيذة كالتالى:

- اضغط على قائمة View ثم اختار Regression test يظهر المخرج الاتى:



ويتم الحكم على وجود تحيز من خلال Intercept(B0) وقيمتة -2.07 بفترات ثقة (-6.344, 2.1931) وقيمة P الاحتمالية 0.14400 وهى غير دالة إحصائيًا وعليه لا يوجد تحيز للنشر انظر قيمة P لمعامل الارتباط الرتبى 0.458.

توصيات وتقويم دراسة ما وراء التحليل

وقدم Pigott (2012) عدة توصيات لزيادة جودة دراسة ما وراء التحليل كالتالى:

- **الرؤية الواضحة لمشكلة البحث والأسلوب:** يتطلب أسلوب ما وراء التحليل فهم عميق لطبيعة الأسئلة المراد الاجابة عليها لأن هذا الأسلوب يتسم بالحداثة إلى حدًا ما فى تطبيقه فى الدراسات، وكذلك فهم لطبيعة الأسلوب المستخدم ومواصفات الدراسات المتضمنة وحجم عينة الدراسات المناسب للتحليل واختيار نوعية النموذج المستخدم وكيفية تفسير الفروق بين الدراسات وكيفية تفسير النتائج ومدى الاستفادة منها فى تحسين الممارسات البحثية واستخدامها لصانعى السياسات المختلفة.

- **وجود خطة مسبقة لدراسة ما وراء التحليل:** وجود تصور مسبق عن طبيعة العلاقة بين المتغيرين وتحديد أهم خصائص الدراسات المراد تكويدها واهم العوامل التى يمكن أن تسبب الاختلافات بين الدراسات لتضمينها فى التحليل لإجراء التحليلات المتفاعلة.
- **الدقة فى تفسير نتائج ما وراء التحليل:** الفهم الواضح للمشكلة ووضع الخطة المناسبة للتحليل يقود إلى الوصول إلى نتائج تتسم بالموثوقية، وأثناء التفسير راعى أنه يجب تحويل حجم التأثير اللوغاريتمى إلى حجم تأثير قابل للتفسير. ويجب أن يكون التفسير مرتبط بالنظرية سواء بتأيدها أو رفضها أو محاولة إدخال تحسينات أو تعديلات عليها كذلك ان وجدت اختلافات بين أحجام التأثير فيجب البحث عن العوامل التى احدثت هذة الاختلافات من خلال إجراء ANOVA أو ما وراء الانحدار.

تجنب اشكاليات عند كتابة عرض نتائج ما وراء التحليل

قدم Card (2012) عشرة اشكاليات شائعة عند عرض نتائج ما وراء التحليل وفيما يلى عرض هذه الاشكاليات والمقترحات لمحاولة التغلب عليها:

1. **عدم الربط بين المنطقية المفاهيمية وتحليلات البيانات:** أحد أهم المشاكل عند كتابة تقرير نتائج ما وراء التحليل عدم الربط بين الخلفية النظرى للمراجعة أو للدراسة والتحليلات والنتائج المعروضة. أى تحليل يتم إجرائه يكون لسبب وهذا السبب يكون معروضًا فى مقدمة الدراسة، وبعض الباحثين يعرضوا احيانًا بعض التحليلات الاستكشافية ولكن يجب التركيز فى المقام الأول على التحليلات التى تجيب على الأسئلة البحثية أو تخدم اهداف الدراسة.

2. **عدم اعطاء عرض وتفصيلات كافية عن المنهجية:** فى دراسات ما وراء التحليل يجب عرض وصف كافى عن استرتجيات البحث التى اعتمد عليها الباحث فى الحصول على الدراسات، معايير التضمين أو الاستبعاد، إجراءات التكويد لخصائص الدراسات، كيفية حساب حجم التأثير، واستراتجية التحليل المستخدمة. ويجب عرض

الدراسة قبل نشرها على متخصصين فى ما وراء التحليل لمراجعة النسخة الأولية من الدراسة.

3. **عدم كتابة تفاصيل النتائج:** عند عرض نتائج ما راء التحليل يجب عرض المعلومات الخاصة بالدراسات المفردة والإحصائيات التلخيصية مثل متوسط حجم التأثير واختبار الاختلاف بين أحجام التأثير ودلالته الإحصائية وتحليل المتفاعلات ان وجدت واختبار تشخيص التحيز وعرض الأشكال البيانية.

4. **عرض التعقيدات الفنية للأسلوب للاجابة عن سؤال البحث:** يعتبر أسلوب ما وراء التحليل من الأساليب التى تتسم بالتعقيد إلى حدًا ما ولكن بقدر الامكان استخدام الفنيات البسيطة كلما كان ممكن لأن استخدام الفنيات المعقدة يمكن أن يقع الباحث فى اخطاء. ومن الصعب فهم هذه الفنيات المعقدة ولا مانع من استخدام الفنيات المعقدة إذا كانت ضرورية للاجابة عن الأسئلة البحثية المعقدة. ولتجنب عدم فهم تقرير بحث ما وراء التحليل اعرض البحث على زملائك لترى مدى فهمهم للبحث وإذا وجدت صعوبة أو استنتاج غير دقيق فيجب اخذ ذلك فى الاعتبار فى الصورة النهائية للنشر.

5. **نسيان لماذا تجرى دراسة ما راء التحليل:** وهذا واضح عند كتابة جزء المناقشة فينسى الباحث الوصول إلى استنتاج واضح عن النتائج ما وراء التحليل للاجابة عن أسئلة البحث ومدى اسهام الدراسة فى زيادة الفهم للظاهرة المراد دراستها وكذلك اصدار توصيات احيانًا لا ترتبط بمشكلة البحث.

6. **الفشل فى تحديد محددات عينة الدراسات:** لاتعتقد ان الدراسات المتضمنة فى ما وراء التحليل تمثل مجتمع الدراسات تمثيلًا دقيقًا، فتوجد محددات فى الحصول على الدراسات وكذلك بعد استبعاد العديد من الدراسات فى ضوء معايير التضمين والاستبعاد فاصبحت عينة صغيرة نسبيًا فى ضوء مجتمع الدراسات للمتغيرين المراد دراستهما ويجب عرض نتائج تقيم تحيز النشر وتحديد خصائص هذه الدراسات يحد من القدرة التعميمية لنتائج الدراسة.

7. **عدم عرض الملامح الوصفية للدراسات:** الفشل فى اعطاء المعلومات الأساسية والوصفية عن الدراسات المتضمنة فى ما وراء التحليل وهذه المعلومات أو خصائص الدراسات على درجة كبيرة من الأهمية سواء فى استخدام بعض الخصائص فى إجراء التحليلات المتفاعلة أو للوصول إلى استنتاجات من الدراسة ولذلك من الضرورى عرض هذه الخصائص فى بداية جزء النتائج.

8. **استخدم نموذج التأثيرات المثبتة فى التحليل فى حالة وجود تجانس بين أحجام التأثير:** يفترض نموذج التأثيرات المثبتة ان الاختلاف بين أحجام التأثير يرجع إلى خطأ المعاينة فقط، وإذا لم يتحقق التجانس بين أحجام التأثير من الأفضل استخدام نموذج التأثيرات العشوائية أو نموذج التأثيرات المثبتة مع تحليل التفاعلات أو النموذج المختلط. ويتم تقيم عدم التجانس بين أحجام التأثير من خلال اختبار Q وإذا كانت له دلالة إحصائية فيجب استخدام نموذج التأثيرات العشوائية.

9. **عدم عرض محددات تحليلات ما وراء التحليل المتفاعلة Meta-analysis moderator analyses:** تعتبر التحليلات المتفاعلة من التحليلات الأساسية فى ما وراء التحليل وان نتائج هذة التحليلات مرتبطة بخصائص الدراسات ويجب عدم تفسير نتائجها بصورة سببية لأن طبيعة المتغيرات المتضمنة فى هذه التحليلات غير تجريبية.

10. **لاتعتقد وجود طريق صحيح Right way لإجراء وكتابة ما وراء التحليل** على الرغم من كل التوصيات الملموسة فى هذا العرض الا انها توصيات مجردة ومطلقة ولكن من الضرورى اتباع الصيغ المتبعة المحددة من قبل الجمعيات العلمية مثل الجمعية النفسية الامريكية APA وكذلك المجلات العلمية فى هذا الشأن.

وعلى الرغم من مميزات ما وراء التحليل كأسلوب لتحليل وتلخيص النواتج الكمية للتراث البحثى الخاص بموضوع ما ولكنه كأى أسلوب يشوبه العديد من المحددات والمشاكل ويمكن عرض البعض منها المرتبط بتفسير النتائج للممارسة ولصانعى السياسة وللبحث. وبعض هذه المحددات كالتالى Lipsey &Wilson (2001):

المناسبة المنهجية للأساس البحثى Methodological adequacy

جودة نتائج ما وراء التحليل تتوقف على طبيعة الدراسات المتضمنة فى التحليل. فاذا كانت الدراسات المتضمنة فى التحليل فقيرة المنهجية فلا نتوقع ان نصل إلى نتائج مفيدة وصادقة بالتالى تحد من قدرة الباحث على تعميم النتائج.

الدراسات المتضمنة فى دراسة ما وراء التحليل متنوعة فى الجودة المنهجية فبعضها جيد وبعضها يتضمن محددات منهجية وعلى ذلك فلا بد على الباحث ان ان يختبر إلى مدى تختلف النتائج باختلاف الجودة المنهجية للدراسات وهل نفس النتائج المتحصل عليها من كل الدراسات هى نفسها للدراسات عالية الجودة وللدراسات منخفضة الجودة ولتحقيق ذلك يمكن تصنيف الدراسات إلى دراسات جيدة ودراسات فقيرة المنهجية وإجراء التحليل المتفاعل. ويمكن تضمين بعض الخصائص المنهجية مثل عمر العينة لإجراء تحليل الانحدار الموزون لمعرفة مدى اسهامها فى التنبؤ بحجم التأثير وهذه النوعية من التحليلات للمظاهر المختلف للمنهجية لتحديد تاثيرها على حجم التأثير يطلق عليها تحليلات الحساسية Sensitivity Analyses.

تداخل الملامح الجوهرية والمنهجية

يمكن تؤدى نتائج الدراسات الضعيفة منهجيًا إلى نتائج غير موثوق بها. ويقوم الباحث بالمقارنة بين أحجام التأثير فى ضوء متغيرات تصنيفية مختلفة مثل المقارنة بين أحجام التأثير لمعالجات مختلفة كالمقارنة بين العلاج السلوكى المعرفى فى مقابل أسلوب الاسترخاء فى أحجام التأثير أو المقارنة بين الذكور والإناث أو المرحلة التعليمية أو الفئة العمرية، وهذه التحليلات تهدف إلى دراسة العلاقة بين هذة المتغيرات واحجام التأثير.

أهمية التباين

بالتاكيد يوجد فروق فردية بين الأفراد فى الدراسات الأولية كذلك يوجد نفس الفروق بين أحجام التأثير المتضمنة فى ما وراء التحليل وليس من الحكمة لباحث ما وراء التحليل اعطاء اهتمامه لقيم متوسط حجم التأثير بدون الوضع فى الاعتبار تباين أحجام التأثير. وفى ما وراء التحليل يتم تحديد مجتمع الدراسات التى تقيس نفس المتغيرين أو

نفس المفاهيم ومن الطبيعى ان يوجد اختلاف بين أحجام التأثير للدراسات المختلفة لاسباب عديدة، وفى الوضع الطبيعى إذا كان توزيع أحجام التأثير متجانس بمعنى أنه من المتوقع ان لا تختلف أحجام التأثير أكثر مما هو متوقع نتجة خطأ المعاينة وفى هذه يستخدم الباحث نموذج التأثيرات المثبتة.

وعندما يوجد اختلاف بين أحجام التأثير فليس من الحكمة ان يتم دمج كل أحجام التأثير لحساب المتوسط ومن الضرورى تحديد لماذا اختلفت أحجام التأثير من خلال تحديد خصائص الدراسات المنهجية والجوهرية واى ملامح اخرى ويتم تكويدها للمساعدة فى تحديد العوامل التى تسبب هذا الاختلاف فربما يكتشف الباحث ان المفاهيم الإجرائية المختلفة لنفس المفهوم، المعالجات المختلفة، العينات المختلفة وغيرها من العوامل يمكن أن تسبب الاختلاف فى أحجام التأثير من ثم يصبح تحديد مصادر الاختلاف بين أحجام التأثير هدف اسمى من حساب متوسط حجم التأثير. ويعتبر نموذج التأثيرات العشوائية من أفضل الخيارات للتعامل مع توزيع غير متجانس لاحجام التأثير.

الفجوات البحثية والقدرة التعميمه Research Gaps and Generalizability

من النواتج القيمة لما وراء التحليل هو اعطاء وصف ملخص لطبيعة حال الدراسات فى تراث الظاهرة فمن الضرورى فحص التوزيع التكرارى لتوصيفات أو خصائص الدراسات المنهجية أو الجوهرية مثل نوعية التصميم والقياسات المستخدمة والعينات والمفاهيم الإجرائية والمعالجات غيرها، وهذه الخصائص تكشف عن القصورات والمحددات فى التراث البحثى للظاهرة وإذا لم تمثل هذه الخصائص فى الدراسات المتضمنة فى ما وراء التحليل فإن تعميم النتائج يكون موضع شك وتساؤل. واحد التضمينات هو ان التأثيرات فى ما وراء التحليل الخاصة بمدى كفاءة معالجة معينة لفئة معينة ليس من الضرورى أن تكون ممثلة للتاثيرات التى تحدث فى الممارسة الاكلنيكية فى المواقف غير البحثية.

واحد محددات تعميم النتائج ان الدراسات تتم فى مواقف معملية وليس من الضرورى ان تعميم نتائجها فى الممارسات الطبيعة فى المدرسة أو المستشفيات أو العيادات هذة تختلف عن طبيعة الابحاث التجريبية الحقيقية المعملية الصارمة.

تطبيقات ما وراء التحليل للممارسة ولصانعى السياسات

إجراء دراسة ما وراء التحليل بصورة متقنة وجيدة يمكن أن يكون لها اسهامات فعالة فى الممارسة العملية وصناعة السياسات والقرارات، فالدراسات المفردة تختلف فيما بينها فى النتائج والجودة ولذلك فإن ما وراء التحليل يعطى إطار أو رؤية شاملة للدراسات المفردة. وهذا التجميع يقلل خطأ المعاينة المسئول عن تضارب نتائج الدراسات. يرى Hunter & Schmidt (2004) أن ما وراء التحليل يمدنا بإجراء امبريقى كمى لتحديد النواقص المنهجية إذا تحقق التى:

الأول: ان يجمع الباحث مجموعة من الدراسات الشاملة التى تمثل مجتمع الدراسات تمثيلًا جيدًا.

الثانى: تحديد ماهية العيوب المنهجية فى الدراسات.

الثالث: تطبيق ما وراء التحليل لكل الدراسات الجيدة والضعيفة منهجيًا وليس كما هو سائد فى دراسات ما وراء التحليل استبعاد الدراسات الضعيفة منهجيًا فى ضوء معايير الاستبعاد التى يضعها الباحث فتعامل دراسات ما وراء التحليل مع الجانب المضئ ويستبعد الجانب المظلم وعليه ما الفائدة إذا كان الهدف الأسمى مما وراء التحليل تقيم التراث البحثى ومحاولة تحديد المحددات المنهجية واخبار الباحثين فى المجال بأهمية تلاشى هذه المحددات.

الرابع: إذا وجدت اختلافات بين الدراسات فى التحليل فإن هذا الاختلاف ربما أو لا يفسر عن طريق دراسات ما وراء التحليل المنفصلة الضعيفة منهجيًا والجيدة منهجيًا. ومن الأفضل تحديد هذة المشكلات المنهجية وقياسها وتصحيحها أفضل من استبعادها وهذا المدخل إلى تبناه Hunter & Schmidit (2004) فى كتابهما: **طرق ما وراء التحليل: تصحيح الخطأ والتحيز فى نتائج البحوث.**

قضايا فى ما وراء التحليل

عرض Borenstien et al. (2009) العديد من القضايا الخلافية والجوهرية اثناء إجراء دراسة ما وراء التحليل كالتالى:

هل توجد منطقية لخلط دراسات بمنهجيات وعينات ومقاييس مختلفة فى تحليل واحد؟

شبهه البعض تحليل ما وراء التحليل كالخلط بين التفاح والبرتقال، ولكن Rosenthal من المطورين الاوائل لأسلوب ما وراء التحليل قدم لهم اجابة مفيدة وهو أنه يمكن جعل هذا الخلط له معنى إذا كنت مهتم بعمل سلطة فواكهة فهذ أكثر فائدة من اكل البرتقال فقط أو التفاح فقط.

الهدف من ما وراء التحليل ليس فقط تجميع بيانات من مجموعة متماثلة من الدراسات وتوليفها لتقدير حجم التأثير العام، ولكن من الممكن الذهاب ابعد من ذلك من خلال تحديد مدى تشتت أو ابتعاد أحجام التأثير عن المتوسط العام. فاذا كان الهدف تحديد اثر أو فعالية التدخلات التدريسية على التحصيل فهل من الأفضل تضمين الدراسات التى تتضمن الدراسات التى تتضمن قياس المهارات اللغوية والمهارات الرياضية؟ إذا كان الهدف تقدير الاثر على الأداء بصفة عامة فالاجابة نعم ولكن إذا كان الهدف تقدير الاثر على المهارة اللغوية فقط فالاجابة لا. وهل من الأفضل تضمين الدراسات فى المرحلة الاعدادية والثانوية معًا؟ الاجابة تتوقف على طبيعة وهدف البحث.

ايضًا هل من المنطقى تضمين دراسات تجريبية مع دراسات ارتباطية أو ملاحظة فى نفس التحليل؟ وماذا عن الدراسات شبه تجريبية؟ وهل من المقبول تضمين دراسات من تصميم مجموعات مستقلة مع دراسات من تصميم مجموعات مرتبطة أو متزاوجة؟ الاجابة على هذة الأسئلة تتوقف عل طبيعة السؤال البحثى.

هل يجب أن تكون الدراسات المتضمنة فى التحليل متشابهة؟ من وجهة النظر الإحصائية لا يوجد قيود على على التماثل أو التشابهة للدراسات المتضمنة فى التحليل سواء تشابه فى ضوء العينات أو التدخلات أو المنهجيات، يكون التحليل مفيد وله معنى إذا تناولنا هذه التنوعات فى التحليل فهذا مفيد فى تفسير الظاهرة موضع الدراسة

ودراسة مدى استقرار وتماسل النظرية. ومدى اتساع معايير التضمين ومحدوديتها يتوقف على طبيعة الدراسة فمثلًا فى الدراسات الدوائية والعلاج بادوية كلما كانت الدراسات المتضمن فى التحليل محكمة ومتجانسة ومعايير التضمين محددة تمامًا لفئة معينة فإن هذا يعطى تقديرات دقيقة. ولكن فى معظم تحليلات ما وراء التحليل معايير التضمين تكون واسعة وليست مقيدة وهذا يعطى قدرًا كبيرًا من التنوع بين الدراسات وهذا امر مرغوب فى ما وراء التحليل (Borestein et al., 2009).

ودراسة ما وراء التحليل الجيدة هى التى تتضمن تنوع فى متغيراتها المتفاعلة التى تفسر الاختلاف فى أحجام التأثير وكذلك يمكن تفسير الاختلاف بين أحجام التأثير بالاعتماد على نموذج التأثيرات العشوائية، وايضًا بتضمين المتغيرات التصنيفية مثل الجنس اوالمرحلة التعليمية فهذا يفيد فى إجراء ما وراء تحليلات منفصلة فى الدراسة الواحدة وهذا يشار اليه ما وراء التحليل للمجموعات الفرعية Subgroup meta-analysis ولو كان التأثير متماثل عبر المجموعات المختلفة عندئذ يقال ان للتاثير مناعة أو ضلاعة Robust وهذا لن يحدث بتبنى معايير تضمين محدودة، بالتالى إذا كان الهدف تقدير متوسط حجم التأثير فلا مانع من معايير محدودة ضيقة ولكن إذا كان الهدف الكشف عن التنوع والاختلاف بين أحجام التأثير فلا مفر من اتساع معايير التضمين.

هل من الممكن تضمين دراسات من تصميمات بحثية مختلفة؟ النوعية المناسبة من الدراسات المتضمنة فى ما وراء التحليل يتوقف على نوعية السؤال البحثى الذى يطرحه الباحث فى دراسته، فمثلًا ما وراء التحليل الذى يتضمن تقيم فعالية برنامج أو تدخل معين سيعتمد على الدراسات التجريبية وايضًا يعتمد على نوعية التجارب هل التصميمات التى تتضمن مجموعتين واى انتقاء للمجموعتين للعينة عشوائية أو مقصودة، أم تصميم المجموعة الواحدة اويعتمد الباحث على الدراسات السببية الارتباطية كل هذا يتوقف على الهدف من الدراسة.

ويرى البعض تضمين كل الدراسات التجريبية سواء كانت حقيقية أو شبة تجريبية وكذلك تضمين كل التصميمات ثم دراسة الفروق فى أحجام التأثير بين التصميمات

المختلفة من خلال تحليل المتفاعل. بينما يرى البعض ان ما وراء التحليل الذى يهتم بدراسة تاثيرات التدخلات تكون مقصورة فقط على الدراسات التجريبية الحقيقية التى تتمتع بصدق داخلى مرتفع وبكل تاكيد يترتب على ذلك استبعاد الدراسات الملاحظة الارتباطية التى تتناول العلاقة بين المتغيرين على مجموعة واحدة بدون معالجة، ولكن Borenstein et al.(2009) يرى أن الأفضلية فى الدراسات المتضمنة ليس بطبيعة التصميم ولكن بمدى جودة الدراسة فى اعطاء تقدير غير متحيز لحجم التأثير ولكن يجب تحليل الدراسات التجريبية منفصلة عن دراسات الارتباطية وربما يتم دمجهما فى تحليل واحد لو لم يختلفوا فى حجم التأثير ويتناولوا نفس السؤال.

حتى فى دمج دراسات المقارنة بين متوسطات المجموعة التجريبية فى مقابل المجموعة الضابطة يستخدم المجموعات المستقلة أم المجموعات المرتبطة، وفى هذا الشان لا يوجد مانع أو أى مشكلة فنية فى دمج البيانات من تصميم المجموعات المستقلة مع تصميم المجموعات المرتبطة.

هل تدمج دراسات باستخدام تحليلات مختلفة أو عرضت النتائج بصورة مختلفة؟ يتعامل دراسة ما وراء التحليل مع نتائج مختلفة لنفس المتغيرين. فبعض الدراسات تقيس المتغير التابع فى ضوء المتوسط وبعضها فى ضوء اعداد مثل قياس الفترة الزمنية فى ضوء عدد الايام أو الاسابيع وهذا شائع فى تخصصات مثل الزراعة والطب اونسبة الأفراد أو النباتات التى حدث لها تحسن، حتى فى نفس الدراسة يتم عرض النتائج فى صور مختلفة نتيجة لاستخدام مداخل مختلفة للتحليل البيانات. ومهما اختلفت طرق عرض النتائج لنفس المتغيرين عبر الدراسات المختلفة فإن العبرة هى بكيفية حساب حجم التأثير حيث يمكن تحويل الصيغ المختلفة لحجم التأثير إلى مؤشر موحد مثل معامل الارتباط r.

كم عدد الدراسات اللازمة لإجراء ما وراء التحليل؟ السؤال الشائع المطروح فى ما وراء التحليل هو كم عدد الدراسات اللازمة لإجراء ما وراء التحليل؟ يرى Borenstein et al.(2009) باستخدام نموذج التأثيرات المثبتة لإجراء ما وراء التحليل يمكن الاعتماد على دراستين لأن الحصول على ملخص لحجم التأثير لدراستين أفضل واكثر دقة من

دراسة واحدة، المهم ان اهتمام الباحث ليس دراسة الاختلافات بين أحجام التأثير لأن هذا يهتم بشئ آخر ابعد من خطأ المعاينة، وملخص حجم التأثير يتم توصيفة فى فترات الثقة التى تصف عدم التاكد لتقدير حجم التأثير.

ولكن باستخدام نموذج التأثيرات العشوائية حيث يكون الاختلافات بين أحجام التأثير ضرورة قصوى ومصدر آخر للخطأ فإن عدد الدراسات الصغيرة يجعل تقدير الخطأ غير صادق أو ثابت لأن تقدير التباين بين الدراسات τ^2 يعتبر مصدر آخر للخطأ وليس فقط خطأ المعاينة كما فى نموذج التأثيرات المثبتة من ثم عددالدراسات الصغيرة يسبب تقدير تحيز للخطأ المعيارى المرتبط بمتوسط حجم التأثير ويسبب تضخم أو تقدير موجب متحيز لمتوسط حجم التأثير ويعطى تقديرات غير دقيقة لفترات الثقة. ولكن Pigott (2012) يرى أن الاجابة الكاملة والحقيقية لهذا السؤال تتوقف على القوة الإحصائية للاختبارات المتضمنة فى ما وراء التحليل سوا متوسط حجم التأثير أو الاختلاف أو تحليل المتفاعل فى ضوء نموذج التأثيرات المثبتة ونموذج التأثيرات العشوائية، وعليه فإن القاعدة فى عينة الدراسات فى ما وراء التحليل هى "الاكثر أفضل" وهى قاعدة فى الدراسات الأولية لأن تضمين عدد قليل من الدراسات فى التحليل فمن الأفضل عدم تلخيصها إحصائيًا ومن الأفضل مراجعتها كيفيًا ولكن هذا لا يعنى عدم إجراء ما وراء التحليل بل إجرائه وفى هذة الحالة يجب عدم الوثوق فى النتائج وتعميمها.

حقيقة فى ظل تراكم دراسات ما وراء التحليل ظهر الجيل الثانى من ما وراء التحليل وهو ما وراء التحليل من الرتبة الثانية Second-Order Meta-analysis وهى إجراء ما وراء التحليل لدراسات ما وراء التحليل نفسها بالتالى فإن عينة الدراسات ليست دراسات اولية انما دراسات ما وراء التحليل ولكن المشكلة التحليلية لإجراء هذا النوع من الدراسات هى خطا المعاينة (للمزيد يمكن الرجوع إلى Hunter & Schmidt, 2004).

معاير الجودة فى دراسة ما وراء التحليل

قدم Cooper (2009) قائمة تتضمن عدد من الخطوات الإرشادية أو المعايير التى يمكن أن تساعد على جودة الدراسة وتقيمها ومدى مقبوليتها للنشر وهى كالآتى:

تحديد المشكلة

- هل توجد منطقية ومبررات لدراسة المشكلة؟
- هل التعريفات النظرية واضحة للمتغيرات الأساسية؟
- هل متغيرات الدراسات الأولية مناسبة لتمثل مفاهيم دراسة ما وراء التحليل؟
- هل يوجد عدد كافى من الدراسات الأولية تناولت متغيرات دراسة ما وراء؟
- هل التصميمات البحثية واضح ومناسبة للاجابة عن الأسئلة؟
- هل قواعد أو معايير التضمين أو الاستبعاد واضحة؟

جمع البيانات

- هل توجد استرتجيات بحثية عديدة وقواعد بيانات للبحث عن الدراسات المناسبة؟
- هل استخدم الباحث مصطلحات مترادفة للمتغيرات الرئيسية اثناء البحث فى قواعد البيانات؟
- هل استخلاص البيانات من الدراسات تميز بالموضوعية والثبات؟
- هل استخلص الباحث معلومات تكويدية ملائمة من الدراسات المتعددة؟
- هل الطرق المستخدمة فى استخلاص البيانات مرضية، موثقة، واضحة، وقابلة للاعادة من قبل الآخرين؟
- هل تعامل مع البيانات الغائبة (الدراسات الغائبة) بطرق مناسبة؟

حساب وتوليف أحجام التأثير

- هل استخدم مؤشر حجم التأثير المناسب؟
- هل قدر متوسط حجم التأثير؟
- هل قدر فترات الثقة؟
- هل قدر الدلالة الإحصائية لمتوسط حجم التأثير؟

- هل اختار النموذج الإحصائى المناسب؟
- هل حسب المتوسط الموزون فى ضوء مقلوب تباين الخطأ المعيارى؟
- هل حسب درجة الاختلاف بين أحجام التأثير؟
- هل أجرى تحليل التفاعلات إذا حدث اختلاف بين أحجام التأثير؟

اختبار ضلاعة النتائج

- هل حسب تحليل الحساسية واستخدمه فى تفسير النتائج؟
- هل تناول قضية البيانات الغائبة من خلال اختبارات تحيز النشر؟

عرض النتائج

- هل عرض النتائج فى ضوء جداول وأشكال بشكل جيد؟
- هل الاستخلاصات من الدراسة واضحة وكاملة ومرضية؟

تفسير النتائج

- هل أصدر تعميمات ومحددات الدراسة؟
- هل تفسير النتائج مرتبطة بما انتجته الدراسة من نتائج ومرتبط بالقيمة العملية بمجال الدراسة؟
- هل حدد نواحى معينة بحاجة إلى دراسات اخرى؟

ومن خلال العرض السابق لفصول الكتاب يمكن لمعد هذا الكتاب من طرح قائمة لكيفية قراءة وتقيم جودة دراسة ما وراء التحليل كالآتى:

- هل العنوان مصاغ بطريق توحى انها دراسة ما وراء التحليل؟
- هل الأسئلة مصاغة بطريقة مناسبة؟
- هل استراتجية البحث واضحة وشاملة (قواعد البيانات أو أى مصادر اخرى)؟
- هل معايير التضمين والاستبعاد واضحة وكافية؟
- هل قيم الجودة المنهجية للدراسات بقدر الامكان؟
- هل بروتكول التكويد شامل وواضح؟

- هل تحقق من موضوعية عملية التكويد وثباتها؟
- هل عينة الدراسات ممثلة لمجتمع الدراسات فى التراث البحثى؟
- هل اختار حجم التأثير المناسب لطبيعة المتغيرات؟
- هل اعتمد على أكثر من حجم تأثير من نفس الدراسة؟
- هل قدر حجم التأثير للعينات الفرعية فى الدراسة؟
- هل أجرى تصحيح لحجم التأثير من المحددات المنهجية كلما أمكن؟
- هل قدر الدلالة الإحصائية
- هل قدر فترات الثقة حول متوسط حجم التأثير؟
- هل حدد البرنامج المستخدم لإجراء التحليل ومبررات ذلك؟
- هل حدد مدخل التحليل المستخدم كطريقة Hedges أو Hunter & Schmidt ؟
- هل حدد نموذج التحليل المستخدم سوا مثبت اوعشوائى أو مختلط ومبررات اختياره؟
- هل قدر اختبار الاختلاف بين الدراسات؟
- وإذا وجد اختلاف هل أجرى التحليل المتفاعل لتحديد سبب الاختلاف؟
- هل تحقق من تحيز النشر اوالبيانات الغائبة؟
- هل أجرى تحليل الحساسية فى الدراسة؟
- هل نتائج الدراسة منطقية وصحيحة؟
- هل عرض النتائج فى جداول واضحة؟
- هل عرض النتائج فى رسومات بيانية خاصة شكل Forest plot؟
- هل اعطى تفسيرات واضحة للنتائج؟
- هل طرح الاستفادة والأهمية من نتائج الدراسة؟
- هل عرض الصعوبات ومحددات الدراسة؟

المراجع

الصياد، عبد العاطي أحمد. (1985). النماذج الاحصائية في البحث التربوي والنفسي العربي بين ما هو قائم وما يجب ان يكون. *رسالة الخليج العربي، 16* ، 211 – 251.

الصياد، عبد العاطى أحمد. (1988). الدلالة العملية وحجم العينة المصاحبة للدلالة الإحصائية لاختبار (ت) فى البحث التربوى والنفس العربى. بحوث مؤتمر البحث التربوى الواقع والمستقبل، *رابطة التربية الحديثة بالقاهرة، 2*، 197-233.

الصياد، عبد العاطى احمد. (1994). الإحصاء التربوى. مذكرة غير منشورة كلية التربية، جامعة قناة السويس.

الصياد، عبد العاطى أحمد وعامر، عبد الناصر السيد. (2013). نحو معيار موحد لتقدير حجم التأثير لاختبار ت لعينتين مستقلتين فى البحث التربوى النفسى العربى. *المجلة المصرية للدراسات النفسية، 23*، 80، 5- 21.

النجار، عبد الله بن عمر. (2006). دراسة تحليلية لقوة الاختبار الإحصائى فى البحوث الإدارية المنشورة. *المجلة العلمية لجامعة الملك فيصل (العلوم الإنسانية والإدارية)*، *7*، 261- 293.

عامر، عبد الناصر السيد. (2007). حجم العينة فى تحليل الإنحدار المتعدد. *المجلة المصرية للدراسات النفسية، 17*.

عامر، عبد الناصر السيد. (2008). الدقة التنبؤية لدرجة اختبار القبول وتحصيل اللغة الفرنسية فى الثانوية العامة للتنبؤ بتحصيل الجامعة لشعبة اللغة الفرنسية. *مجلة كلية التربية بنها، 18*، 39-49.

عـامر، عبـد الناصـر السـيد.(2012). النمـاذج والإختبـارات الإحصـائية المسـتخدمة فـي البحـث النفسـي المصـري والعربـي. *المجلـة المصـرية للدراسـات النفسـية، 22*، 74، 355- 371.

عـامر، عبـد الناصـر السـيد (2012). اسـتخدام وتقـدير حجـم التـأثير فـى الدراسـات النفسـية التربويـة فـى المجـلات العربيـة (1994 -2011). *مجلـة جامعـة عـين شـمش للقياس والتقويم، 2* ، 1-40 .

عـامر، عبـد الناصـر السـيد. (2014). تقـدير القـوة الإحصـائية للدراسـات النفسـية فـى المجـلات العربيـة (2000-2012). *مجلـة دراسـات تربويـة ونفسـية، جامعـة الزقازيق، 84*،213-244.

عـامر، عبـد الناصـر السـيد. (2018). *مـا وراء التحليـل للعلـوم السـلوكية والنفسـية: الاسـس النظرية والمنـهجية والتطبيقات*. الرياض: مكتبة العبيكان الرقمية.

Alhija, F. N., & levey, A. (2009). Effect Size reporting practices in published articles. *Educational and Psychological Measurement, 69*, 245- 265.

American psychological Association (APA). (1994). *Publication manual of the American Psychological Association* (5th. Ed.). Washington, DC: Author.

American psychological Association (APA). (2001). *Publication manual of the American Psychological Association* (5th Ed.). Washington, DC: Author.

American psychological Association (APA). (2010). *Publication manual of the American Psychological Association* (6th Ed.). Washington, DC: Author.

Amstrong, J. S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting, 23*, 321-327.

Aron, A., Coups, E. J., & Aron, E. N. (2013). *Statistics for psychology (6th.ed).* Boston: Pearson Education, Inc.

Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics*, *35*, 73- 80.

Baroudi, J. J., & Orlikowski, W. J. (1989). The problem of statistical power in MIS research, *13*, 87-106 .

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*, 60-83.

Borenstein, M. (1994). The case for confidence intervals in controlled clinical trails. Controlled Clinical Trials, *15*, 411-428.

Borenstein, M. (2009). Effect size for continuous data. IN H. Cooper, L. V. Hedge, & J. C. Valentine(Eds.), *The handbook of research synthesis and meta analysis* (2nd.ed) (pp. 222-235). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex: John Wiley & Sons, Ltd.

Borman, G. D. & Grigg, J. A. (2009). Visual and narrative interpretation. In H. Cooper, L. V. Hedges, & J. C. Valentine(Eds.), *The handbook of research synthesis and meta-analysis (2nd ed., pp. 497-519).* New York: Russell Sage.

Breaugh, J. A. (2003). Effect Size Estimation: factors to consider and Mistakes to Avoid. *Journal of Management*, *29*, 79-N97.

Brewer, J. (1972). On the power of statistical tests in the American Educational research journal. *American & Educational Research Journal*, *9*, 391 – 401.

Cai, L., & Hayes, A. F. (2008). A new test of linear hypotheses in OlS regression under heteroscedasticity of unknown form. *Journal of Educational and Behavioral Statistics*, *33*, 21-40

Capraro, R. M., & Capraro, M. M. (2002). Treatments of effect size and statistical significances testes in textbooks. *Educational and Psychological Measurement, 62*, 771 – 782.

Card, N. A. (2012). *Applied meta-analysis for social sciences research.* New York: Guilford Press, Inc.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, inter-correlations, to maladjustment. *Child Development, 79*, 1185-1229.

Casher, H. C., & Geiger, S. W. (2004). Statistical power and the testing of null hypotheses: a review of contemporary management Research and Recommendation for future studies. *Organizational Research Methods, 7*, 151 – 167.

Chase, J., & Chase, B. (1976). Statistical power analysis of applied psychological research. *Journal of Applied Psychology, 61*, 234 – 237.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. *In. R. H. Hoyle (Eds.), Structural equation modeling : concepts, issues, and applications* (PP. 37- 59). Thousand Oaks, CA: Sage

Coe, R. (2004). Issues arising from the use of effect size in analyzing and Re porting research. In. I. Schrage & K. Elliot (Eds.,) But what does it mean? The use of effect Size in educational research. (PP. 80- 100). Institute of Education: university of London.

Cohen, J. (1962). The statistical power of abnormal–social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145- 153.

Cohen, J. (1969). *Statistical power analysis for the Behavioral sciences.* London: Acadmic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed).* Hillsdale: Lawrence ERLBaum Associates, Inc.

Cohen, J. (1990). Things I have learned (So far). *American Psychologist, 45*, 1304 – 1312.

Cohen, J. (1992). A power premier. *Psychological Bulliten*, 112, 155- 159.

Cohen, J. (1994). The Earth is round (P < 0.05). *American Psychologist, 49*, 997 – 1003.

Comery, A. L., & Lee, H. B. (1992). *A First course in factor analysis (2nd .ed).* Hillsadale, NJ: Erbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field ettings.* Boston: Houghton Mifflin.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews.* Thousand Oaks, CA: Sage.

Cooper, H. & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in society, 1*, 104-126.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87*, 442-449.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on centeral and noncentral distribution. *Educational and Psychological Measurement, 61*, 530-572.

Dancy, C. P., & Reidy, J. (2011). *Statistics without maths for psychology (5th.ed).* Harlow: Pearson, Prentice Hall.

Daniel, T. D. (1993, November). A statistical power analysis of the Quantitative techniques used in the "journal research in music Education 1987 – 1991". *Paper presented at the Annual Meeting of the mid – south Educational Research Association. New Orleans, L.A.*

Dattalo, P. (2008). *Detemining sample size: Balancing power, precision, and practicality.* Oxford: University Press.

De Smidt, G. A., & Gorey, K. M. (1997). Unpublished social research: Systematic replication of a recent meta-analysis of published intervention effectiveness research. *Social Work Research, 21*, 58-62.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th.ed*).* Los Angeles: SAGE Publications, Inc.

Diamanatopoulos, A., & Schlegelmilch, B. B. (2003). *Taking the Fear out of data anslysis.* London: The Dryden Press.

Dulark, J. A. (2009). How to select, calculate, and interpret Effect Size. *Journal of Pediatric Psychology, 3*, 917- 928.

Dunlap, W. P. (1994). Generalizing the common language effect size indicators to bivariate normal correlations. *Psychological Bulletin, 116*, 509- 611.

Dunn, D.S. (2001). *Statistics and data analysis for behavioral sciences.* New York: MaGraw-Hill Higher education, Inc.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412.

Efron, B. (1982). *Th jackknife, the bootstrap and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* New York, NY: Cambridge University Press.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*, 319-324.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, 33*, 517.

Fan, X. (2001). Statistical significance and effect size in education research: two side of coin. *Journal Educational Research*, 94, 275-383.

Faul, F., ErdFelder, E., Lang, A. G., & Buchner, A. A. (2007). G - power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39* , 175- 219 .

Felt, L. S. (1965). The approximate sampling distribution of kuder-richardson coefficient twenty. *Psychometrika*, 30, 357-370.

Felt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement*(3rd ed., pp. 105-146). Washington, DC: American Counncil on Education, Macmillan.

Ferguson, C. J. (2009). An effect Size primer: A Guide for clinicians and Researchers. Professional Psychology: *Research and Practice, 2*, 1- 7.

Fern, E. F., & Monroe, K. B. (1996). Effect Size Estimates: Issues and problems in interpretation. *Journal of Consumer Research, 23*, 89- 105.

Fidler, F., Cumming, G., Thomason, N., Pannuzzo D., Smith, J., Fyffe, P., Edmonds, H., Harrington, C., & Schmitt, R. (2005). Toward improved statistical reporting in the journal of consulting and clinical psychology. *American psychological Association, 73*, 136-143.

Field, A. (2009). *Discovering statistics using SPSS (3rd.ed).* Sage Publications.Ltd.

Field, A. (2013). *Discovering statistics using SPSS (4th.ed).* Sage Publications.Ltd.

Field, A. P. (2003). The problems using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77-96.

Field, A. P. (2005)." Is meta-analysis of correlation coefficients accurate when population correlations vary?" *Psychological Methods*, *10*, 444-467.

Finch, W. H. & Frencch, B. F. (2019). *Educational and psychological measurement.* New York: Routledge.

Frohlich, M., Emrich, E., Pieter, A., & Stark, R. (2009). Out come Effect Sizes in Sport sciences. *International Journal of Sports Sciences*, *3*, 175-179.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation*, 17, 1-13.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, *5*, 3-8.

Glass, McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Good, P. (1999). Resampling Methods: Apractical guide to data analysis. Boston: Birkhauser.

Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences (8th.ed).* Belmont: Wadsworth, Cengage Learning.

Green, S.B. (1991). How many subjects does it take to do regression analysis? Multivariate Behavioral Research, *27*, 499-510.

Green, S. B., & Salkind, N. J. (2014). *Using SPSS for windows and macintosh: Analysing and understanding data (7th.ed).* Boston: Pearson Education, Inc.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*, 265-275.

Haase, R. F. (1974). Power analysis of research in counselor education. *Counselor Education and Supervision, 14*, 124–132.

Hagger, M. S. (2006). Meta-analysis in sport and exercise research: Review, recent, development, and recommendations. *European Journal of Sport, 6*, 103-115.

Hall, J. A., & Brannick, M. T. (2002). Comparison of two random effects methods of met-analysis. *Journal of Applied Psychology, 87*, 377-389.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behavioral Research, 34*, 489-499.

Harris, R. J. (1985). *A primer of multivariate statistics.* New York: Academic.

Hayes, W. L. (1963). *Statistics for psychology.* New York: Holt, Rinehart & Winston.

Hayes, W. L. (1988). *Statistics* (4th Ed.). New York: Holt, Rinehart, & Winston.

Healey, J. F. (2012). Statistics: A tool for social sciences. Belmont: Wadsworth, Cengage Learning.

Hedges, L. V. (1981). Distribution Theory of Glass's estimator of effect Size and related estimators. *Journal of Educational Behavioral Statistics, 6*, 107- 128.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*, 119-137.

Hedges, L. V. (2009). Statistical considerations. IN H. Cooper, L. V. Hedge, & J. C. Valentine(Eds.), *The handbook of*

research synthesis and meta analysis(2nd.ed) (pp. 37-47). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V. & Piggot, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6,* 203-217.

Hentschke, H. & Stuttgen, M. C. (2011). Computation of measures of effect Size for neuroscience data Sets. *European Journal of Neuroscience, 34,* 1887- 1894.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558.

Hill, C. J., Bloom, H. S., Black, A. R., & lipsey, M. W. (2007). Empirical Benchmarks for Interpreting Effect Sizes in Research. MDRC working papers on Research Methodology, Available from: http://www. Mdrc.Org

Hinkle, D. E., &Wiersma, W., & Jurs, S.G.(1994). *Applied statistics for the behavioral sciences (3rd.ed).* Boston: Houghton.

Houser, J. (2007). How many are enough? Statistical power analysis and sample size estimation in clinical research. *Journal of Clinical Research Best Practice, 3,* 1- 5.

Howell, D. C. (2013). *Statistics methods for psychology (8th.ed).*Belmont: Wadsworth, Cengage Learning.

Howell, D. C. (2014). *Fundemental statistics for the behavioral sciences (9th.ed).* Belmont: Wadsworth, Cengage Learning.

Howitt, D., & Cramer, D. (2011). *Introduction to statistics in psychology (5th.ed).* Harlow: Pearson Education Limited.

Huck, S. W.(2012).*Reading statistics and research (6th.ed).* Boston: Pearson.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistics or I^2 index? *Psychological Methods, 11,* 193-206.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, 8, 3-7.

Hunter, J. E., & Schmidt, F. L. (1977). A critical analysis of the statistical and ethical implications of various definitions of tests fairness. *Psychological Bulletin, 83*, 1053-1071.

Hunter, J.E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge in psychology. *International Journal of Selection and Assessment, 8,* 275-292.

Hunter, J.E., & Schmidt, F. L. (2004). *Meta-analysis: Correcting error and bias in research findings (2nd Ed).* Newbury Park, CA: Sage Publication, Inc.

Hunter, J.E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Huston, H. L. (1993). November Meaningfulness, statistical significance, Effect size, and power analysis: A General Discussion with Implication for MANOVA. Paper presented at the Annual meeting of the Mid–South Educational Research Association: New Orleans, LA.

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Jornal of Consumer Psychology*, 13, 478-487.

Isaac, C. S., & Michael,W. B. (2009). *Handbook in research and evaluation.* San Diego, CA: Edlts

Iversson, A., Anderson, M. B., Johnson, V., & Lindwall, M. (2013). To adjust or not adjust: Nonparametric effect size,

Confidence intervals, and Real – world meaning. *Psychology of Sport and Exercise, 14*, 97– 102.

Kampense, V. B., Dyba, T., Hannay, J. E., & Sjoberg, D. L. (2007). A Systematic Review of effect Size in Software engineering experiments. *Information and Software Technology, 49*, 1073- 1086.

Kenny, D. A. (1987). *Statistical for social and behavioral sciences* .Canada: Little Brown Company limited.

Keselman , H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselnon, J. C., & Levin. J. R. (1998). Statistical practices of educational Researchers: on analysis of their ANOVA, MANOVA, and ANCOVA analysis. *Review of Educational Research, 68*, 350- 386.

Kirk, R. E. (1996). Practical Significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746- 759.

Kirk, R. E. (2001). Promoting good Statistical practice: Some Suggestions. *Educational and Psychological Measurement, 61*, 213 – 218.

Kirk, R. E. (2008). *Statistics: An introduction (5th.ed).*Belmont: Thomson Wadsworth.

Kisamore, J. L., & Brannick, M. T. (2008). An illustration of the consequances of meta-analysis model choice. *Organizational Research Methods, 11*, 35-53.

Kline, R. B. (2013): *Beyond significance testing: Statistics reform in the behavioral science (2nd Ed).* Washington: American Psychological Association.

Kline, R. K. (2016). *Principles and practice of structural equation modeling (4th.ed).* New York: Guilford publications, Inc.

Kotrlik, J. W., & William, H. A. (2003). The incorporation of effect Size in information technology, learning, and

performance Research. *Information Technology, Learning, and Performance Journal*, *21*, 1- 7.

Ledesma, R. D., Macbeth, G., & Dekohan, N. G. (2009). Computing Measures with vista: The visual statistics system. Tutorials, in *Quantitative Methods for Psychology*, *5*, 25-34.

Levine , T . & Hullett, C. (2002). Eta squared, partial Eta squared, and misreporting of effect Size in communication research . Human Communication Research, *28*, 612- 625.

Lipsey, M. W. (1998). Design sensitivity: statistical power for Applied experimental research. In L. Beckman & D. J. Rog (Eds.), *Handbook of Applied Social Research Methods* (PP. 39- 68). Thousand Oaks, CA: sage.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta–analysis. Thousand.* Oaks, CA: Sage.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic review and meta-analysis.* OXFORD: University Press.

Mazen, A. M., Graf, L. A., Kellogy, C. E., & Hemmasi, M. (1987). statistical power in contemporary management research. The Academy of Management Journal , *30* , 369 – 380 .

McGrath, R. E., & Meyer, G. J. (2006). When effect Size disagree: the case of r and d. *Psychological Methods*, *11*, 386 – 401.

McGraw, k. O., & Wong, S. P. (1992). A Common language effect size Statistic. *Psychological Bulletin*, *111*, 361- 365.

McLean, J. E., & Ernest, J. M. (1998). The role of Statistical Significance testing in educational research. *Research in the Schools*, *5*, 15- 22.

Mcleod, B. D., & Weis, J. R. (2004). Using dissertation to exmine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology*, *72*, 235-251.

McMillan, J. H. (2011). Reporting and discussing Effect Size: Still the road les traveled. *Practical Assessment Research & Evaluation, 16*, 1- 11.

Mengersen, K., Schmid, C. H., Jennions, & Gurevitch, J. (2013). Statistical models and approaches to inference. *In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), Handbook of meta-analysis in ecology and evolution(pp. 89-107).* Oxford: Princeton University Press.

Miller, M. K. (2014). *Nonparametric statistics for social and behavioral sciences.* Boca Raton: Tylor & Francis Group, LLC

More, J. G. & Hudson, W.W. (1995). The problem of sample size estimation: Confidence intervals. Social Work Research, *19*, 121-127.

Nickerson, R. S. (2000). Null hypothesis significance testing: A Review of an old and continuing controversy. *Psychological Methods, 5*, 241– 301.

Nolan, S. A., & Henzien, T. E. (2012). *Statistics for the behavioral sciences(2nd.ed).* New York: Worth Publishers.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*(3rd ed.). New York: McGraw-Hill.

Olejnik, S., & Algina, J. (2003). Generalized Eta and omega squared statistics: Measures of effect Size for some common Research design. *Psychological Methods*, 434 – 444 .

Olejnik. S., & Algina, J. (2000). Measures of effect Size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241 – 286.

Ormee, J. G. & Hudson, W. W. (1995). The problem of sample size estimation: Confidence intervals. *Social Work Research, 19* , 121-127

Pagano, R. P. (2013). *Understanding statistics in the behavioral sciences(10th.ed).* Belmont: Wadsworth, Cengage Learning.

Pedhazur, W. J., & Schmelkin, L. p. (1991). *Measurement, design, and analysis: an integrated approach.* Hillsdale, NJ: Erlbaum.

Pigott, T. D. (2012). *Advances in meta-analysis.* New York: Springer Science + Business Media, LLC.

Pigott, T. D., & Wu, M. J. (2008). Methodological issues in meta-analysis standard deviations: Comments on Bond and Depaulo (2008). *Psychological Bulletin, 134*, 498-500.

Prentice, D. A., & Miller, D. T. (1992). When small effect are impressive. *Psychological Bulletin, 112*, 160- 164.

Privitera, G. J. (2015). *Statistics for the behavioral sciemces(2nd.ed).* Canda: Sage Publications, Inc.

Reed, J. G. & Baxter, P. M. (2009). Using reference database. *IN H. Cooper, L. V. Hedge, & J. C. Valentine(Eds.), The handbook of research synthesis and meta analysis(2nd.ed) (pp. 73-101).* New York: Russell Sage Foundation.

Reynolds, C. R., & Livingston, R. B. (2014). *Mastering modern psychological testing: Theory and methods.* Harlow: Pearson Education Limited.

Richardson, J. T. E. (1996). Measures of effect Size. *Behavior Research Methods, Instruments, and Computers, 28*, 12- 22.

Richardson. J. T. E. (2011). Eta squared and partial Eta squared as measures of effect Size in educational Research. *Educational Research Review, 6*, 135- 147.

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*, 306-322,.

Rosenthal, R. (1991). *Meta-analysis procedures for social research(2nd ed).* Newbury Park, CA: Sage.

Rosenthal, R. (1994). Statistically describing and combining studies. *In H. Cooper & L. V. Hedges (Eds.),The handbook*

of research synthesis(pp.231-244). New York: Russell Sage Foundation.

Rosenthal, R. (1994). Statistically describing and combining studies. *In H. Cooper & L. V. Hedges (Eds.),The handbook of research synthesis(pp.231-244).* New York: Russell Sage Foundation.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B (2000). *Contrasts and effect sizes in behavioral Research: Acorretational approach.* New York: Cambridge university press.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical Procedures and The Justification of Knowledge in Psychological science. *American Psychologist, 44,* 1276 – 1286.

Rosnow, R.., & Rosenthal, R. (2003). Effect Size for experimenting psychologist. *Canadian Journal of Experimental Psychology, 57,* 221-237.

Rossi, J. S. (1990). Statistical power of psychological research: what have we gained in 20 years?. *Journal of Consulting and Clinical Psychology,* 58, 646- 655.

Sawyer, A. G., & Balla, A. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Review, XVIII,* 275 – 290.

Schmidt, F. (1996). Statistical significance Testing and cumulative Knowledge in Psychology: Implications for the Training of researchers. *Psychological Methods, 1,* 115- 129.

Schmidt, F. L., & Hunter, J. E. (1997). *Eight common but false objections to the discontinuation of Significance testing in the analysis of research data* In *L. L. Harlow, S. A. Mulaik & J. H . Steiger (Eds.,), what if there were no significance tests?* (PP. 37-64). Mahwah, NJ: Erlbaum.

Sedlmeier, p., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?. *psychological Bulletin, 105,* 309- 316.

Shadish, W .R. (1995). The Logic of generalization: Five principles common to experiments and ethnographies. American Journal of Community Psychology, *23*, 419-247.

Shadish, W. R., & Cook, T. D. (2002). *Expermintal and quasi-experimental designs for generalized causal inference.* Boston: Hougheton Mifflin Company.

Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program and manual for effect size calculation.* St. Paul, MN: Assessment Systems Cooperation.

Shadish, W., & Myers, D. (2004). Campbell Collaboration research design policy brief. Retrieved from http//www.campbellcollabartion.org/MG/ResDesPolicyBrief.pdf

Sink, G. A., & Mvududu, N. H. (2010). Statistical power, sampling, and effect size: three keys to research relevancy. *Counseling Outcome Research and Evaluation, 1*, 1– 18.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach alpha. *Psychometrika*, 74, 107-120.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect Size estimates. *Journal of Experimental Education, 16*, 334- 349.

Steiger. J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in analysis of variance and contrast analysis. *Psychological Methods, 9,* 164-182.

Stout, D. E., & Ruble, T. L. (1995). Assessing the practical Significance of empirical Results in accounting education Research: The use of Effect Size information. *Journal of Accounting Education, 13*, 281- 298.

Sun, S., pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic Journals in education and psychology. *Journal of Educational Psychology, 102*, 989- 1004.

Tabachnick, B. G., & Fidel, L. S. (2007). *Using multivariate statistics* (4th Ed.). Boston: Allgn & Bacon.

Tatsuoka, M. (1993). *Effect Size. In G. Keren & C. lewis (Eds.),* A *Handbook for Data analyses in the behavioral sciences: Methodological Issues.* New Jersey: Lawrence Erlbaum Associates, publishers.

Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of American, 78,* 128 – 139.

Thompson, B. (1998). Statistical Significance and effect Size Reporting: portrait of a possible future. *Research in the Schools, 5,* 33 -38.

Trusty, J., Thompson, B., & petrocelli , J. V. (2004). Practical guide for reporting effect Size in Quantative Research in the Journal of Counseling & Development. *Journal of Counseling & Development, 82,* 106- 110.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect Sizes. *Journal of Counseling and Psychology, 51,* 473-481.

Verma, J. P. (2004). *Data analysis in management with spss software.* New Delhi: Spriger

Verma, R., & Goodale, J. C. (1995). Statistical power in operations management research. *Journal of Operations Management, 13,* 139 – 152.

Wang, L. W. (2010). Retrospective statistical power: fallacies and Recommendation. *New Born & Infant Nursing Reviews, 10,* 55- 59.

Wang, Z., & Thompson, B. (2007). In The Pearson r^2 biased and if so, what on the best correction formula. *Journal of Experimental Education, 75,* 109- 125.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation modeling with non normal variables: problems and remedies. *In R. H. Hoyle (Eds.), Structural equation modeling: concepts, issues, and applications (PP. 65- 75).* Thousand Oaks, CA: Sage.

Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology, 38,* 29-60.

Wilcox, R. R. (2009). *Basic statistics: Undrestanding conventional methods and modern insights.* Oxford: Oxford University Press.

Wilson, D. V. (2009). Systematic coding. *IN H. Cooper, L. V. Hedge, & J. C. Valentine (Eds.), The handbook of research synthesis and meta analysis(2nd.ed) (pp. 160-175).* New York: Russell Sage Foundation.

Wilson, S. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from Meta–analysis. *Psychological Methods, 6,* 413 – 429.

Wilson-vanvoorhis, C. R., & Morgan, B. L. (2007). Understanding Power and Rule of thumb for determining sample size. *Tutorial in Quantitative Methods for psychology, 3,* 43 – 50.

Wilsson Van Voorhis, C.R. & Morgan, B. L. (2007). Understanding powr and rules of thumb for determining sample sizes. *Tutorials in Quantative Methods for Psychology,* 3, 43-50.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills: Sage Publications.

Wolf, I. M. (1990). Methodological observations on bias. *In K. W. Wachter & M. L. Straf (Eds.). The future of meta-analysis (pp. 139-151).*New York: Russell: Sage.

Woolley, T. W. (1983). A comprehensive power analytic investigation of research in medical education. *Journal of Medical Research, 58,* 710 – 715.

Zientek, L. R, Capraro, M. M., & Capraro, R. M. (2008). Reporting practices in quantative teacher education research: On look at the evidence cited in AERA panel report. *Educational Researcher, 37,* 208- 216.

Zimmerman, D. W., & Williams, R. H. (1989). Statistical power analysis and reliability of measurement. *Journal of General Psychology, 116,* 359- 369.