

سلسلة الاحصاء والبحث للعلوم السلوكية والنفسية

باستخدام حزم البرامج الاحصائية

الجزء الاول

ما بعد اختبارات الدلالة الاحصائية

فترات الثقة، وحجم التأثير، 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 الثاني: 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 الاعتيادي

الفصل التاسع: اختبارات الفروض الصفرية وفترات الثقة

الفصل العاشر: حجم التأثير والقوة الإحصائية

الفصل الحادي العاشر: العينات

الفصل الثاني عشر: اختبارات الفروض للتحقق من الاعتدالية

الفصل الثالث عشر: معامل ارتباط بيرسون الخطى ومعامل الارتباط الجزئي

الفصل الرابع عشر: معامل الارتباط الثنائي 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 أحادي الاتجاه أو البسيط بين المجموعات

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

الفصل الثاني والعشرون: ما وراء التحليل

المراجع

مقدمة السلسلة

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

في هذه السلسلة تحول الاهتمام من اختبارات الفروض الصفرية التي تعاني الكثير من الانتقادات والمحددات منها ثنائية القرار في ضوء $P \leq 0.05$ حيث تؤدي إلى دلالة أو عدم دلالة إحصائية بمعنى هل يوجد تأثير من عدمه؟، وهذا لا يؤدي إلى تقدم العلم وتطوره، وانتقلت إلى إصلاح الإحصاء Statistics reform من خلال إعطاء مدي من القيم لقبول أو رفض الفرض الصفرية من خلال فترات الثقة، وكذلك الانتقال من وجود أو عدم وجود تأثير إلى مقدار وحجم هذا التأثير من خلال المكمل الأساسي حجم التأثير، وكذلك الانتقال من الدلالة الإحصائية للدراسة الواحدة إلى استخدام ما وراء التحليل للإجابة عن الأسئلة البحثية من خلال إعطاء مؤشر تلخيصي للدراسات في التراث التي تناولت العلاقة بين متغيرين في الدراسة، وعليه استخدام فترات الثقة وحجم التأثير كمكملين لاختبارات الفروض الصفرية بجانب مؤشر هام للدقة الإحصائية وهو القوة الإحصائية؛ 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 المجالات النفسية والتربوية وفي كل التخصصات.

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

وفي مقابل القطبية في صناعة القرار الإحصائي 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 في الفرض الصفرى. وكذلك تم عرض ما وراء التحليل باعتبارها أحد الطرق البديلة لاختبارات الدلالة الإحصائية للإجابة على أسئلة الدراسة. وبعد تناول الاختبار أو الأسلوب الإحصائي وحساباته اليدوية، عرضت كيفية إجرائه في البرنامج الإحصائي Version SPSS 23 بكل خطواته من إدخال البيانات وخطوات تنفيذه في البرنامج وكيفية تفسير النتائج وحساب القوة الإحصائية من خلال برنامج G-Power.

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

وتضمنت هذه السلسلة ثلاثة أجزاء كالآتي:

الجزء الأول: ما بعد اختبارات الدلالة الإحصائية : فترات الثقة، وحجم التأثير، والقوة

الإحصائية وما وراء التحليل

"SPSS, G-Power, & CMA"

الجزء الثاني: تحليل البيانات اللابارامترية في ضوء حجم التأثير "SPSS"

الجزء الثالث: تحليل البيانات المتعددة والمتدرجة في 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 (أكثر من متغير مستقل ومتغير تابع وحيد)، وللنموذج المتدرج Multivariate (متغير مستقل واحد فأكثر وأكثر من متغير تابع) (الصياد، 1985) والاختبارات اللابارامترية التي يتناولها الباحث في العلوم السلوكية والنفسية والتربوية بل معظم التخصصات الأكاديمية واعتمدت في إعداد هذه السلسلة على كتب ومراجع إحصائية متخصصة وكذلك على خبرتي التدريسية لمقررات القياس والإحصاء في كليتي لمدة تزيد على خمسة وعشرون عامًا وخبرتي البحثية في مجال علم النفس والقياس والإحصاء، وكذلك على الاستشارات واستفسارات الباحثين أثناء تحليل بياناتهم.

وتمتاز هذه السلسلة بالآتي:

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

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

واتمنى أن اكون قدمت عملاً مفيداً للطلاب والباحثين فى كافة التخصصات ليساعدهم على تصميم بحوثهم وتحليل بياناتهم بكل موثوقية وبقليل من الخوف والرغبة من التعامل مع البيانات وبعيداً 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 اطل 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 الإلكتروني 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 بادرارة 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 والاجتماعي والتربوي حيث توظف الدراسات والبحوث هذه الاختبارات بدرجة كبيرة ودائماً يفضل الباحث هذه النوعية من الاختبارات لأنها أكثر الفة بها، كما انها تناسب كثيراً من البيانات، ولكن يوجد محدد رئيسي في استخدامها وهو أن الباحثون لا يتحققون من مسلمات هذه الاختبارات خاصة العشوائية والاعتدالية. وهذا ما حاولت تجنبه في عرض اختبارات الفروض لهذه الاختبارات. وتناول هذا الجزء اختبارات نموذج المتغير الواحد واختبارات النموذج البسيط حيث انها تمثل 73.1% من الاختبارات الإحصائية المستخدمة في البحث النفسي والتربوي العربي (عامر، 2012).

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

وتناولت مقاييس الإحصاء الوصفي مثل مقاييس النزعة المركزية والتشتت وكيفية عرض البيانات بيانياً باعتبارها الخطوة الأولى للتعامل مع البيانات فهي تهدف إلى عرض وتنظيم وتلخيص البيانات بيانياً وعددياً لمحاولة فهمها بصورة أكثر وضوحاً وعرضت لكيفية حساب المتوسط المذهب 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 في برنامج G-Power. واعتمدت في هذا الجزء على مداخل حديثة لصناعة القرار اكدت عليها معظم المؤسسات البحثية والمجلات العلمية ومنها الاعتماد على حجم التأثير لما له من أهمية في 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، أيضاً نتيجة للانتقادات الموجهة للطريقة التقليدية الكيفية السردية فى مراجعة الدراسات السابقة حث تودى 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 الإحصائي المناسب لفروض دراسته حتى نصل إلى قرارات تتسم بالموثوقية والدقة الإحصائية.
- واتمنى أن اكون قدمت عملاً مفيداً للطلاب والباحثين فى كافة التخصصات مجال الإحصاء يساعدهم على تصميم بحوثهم وتحليل بياناتهم بكل موثوقية وبقليل من الخوف وبعيداً 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 موفق والمستعان

أ.د. عبد الناصر السيد عامر

الفصل الأول

مفاهيم أساسية

Basic Concepts

منذ اهتمامى بعمل الإحصائي وتحليل البيانات فإن زملائي 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 الإحصائي الأمثل يبدو مفقوداً 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 الإحصائي.

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 أسلوباً تحليلياً تطبيقياً 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) بأنها مجموعة من الإجراءات والقواعد التى تستخدم لاختصار أحجام كبيرة من البيانات إلى نسب معينة وتسمح لنا بعمل استنتاجات. واستخدما 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 وتوفير 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 الاستدلالي **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 الوصفي: تستخدم الإحصاء الوصفي عندما يكون الهدف وصف مجموعة من البيانات فقط وتستخدم لتصنيف وتلخيص البيانات فى صورة مبسطة كتقدير المتوسط أو الانحراف المعياري للبيانات، فعلى سبيل المثال المدرس فى الفصل يبنى اختبار تحصيلي لقياس مهارات اللغة العربية ثم يطبقه ويصححه ويعطى درجات للطلاب بعد ذلك يقوم بعرض بتمثيل الدرجات بيانياً من خلال الرسومات وتقدير متوسط التحصيل وتحديد نسب النجاح والرسوب فى الفصل. والإحصاء الوصفي مجموعة من الإجراءات تهدف إلى عرض ووصف وتلخيص البيانات بهدف تيسير فهمها والوصول لانطباعات مبدئية عنها بالتالى فإن الإحصاء الوصفي 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 الاستكشافية (**Exploratory data analysis (EDA)**، ومن أهم الأساليب الإحصائية التى تقع فى الإحصاء الوصفى مقاييس النزعة المركزية ومقاييس التشتت والتوزيعات التكرارية والرسومات البيانية وغيرها.

الإحصاء الاستدلالي: بعد وصف البيانات وأصبحنا على وعى بماذا تقول على مستوى وصفى سطحى نهتم بعد ذلك بالإحصاء الاستدلالي 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، فمثلاً فى تصميم التجارب نحن ندرك أنه من المستحيل تطبيق التجربة على المجتمع ككل ولظروف خاصة بعملية الضبط والتكلفة والوقت يتم سحب عينات من المجتمع الأصلي.

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 تعبيراً صادقاً وحقيقياً للمجتمع إذا كانت العينة ممثلة تمثيلاً جيداً للمجتمع أو تخمينات غير حقيقية إذا كانت العينة غير ممثلة للمجتمع، وبالتالي يتم استنتاج معالم المجتمع من خلال ما نعرفه من إحصائيات العينة. فالإحصاء الاستدلالي هو مجموعة من الطرق لعمل استدلالات عن خصائص المجتمع من خلال معرفة خصائص العينة.

ومن الخطأ القول إن الإحصاء الوصفى والإحصاء الاستدلالي تصنيفين متميزين ومختلفين عن بعضها البعض فكلاهما مكمل للآخر بمعنى أن الإحصاء الوصفى هى ضرورة للإحصاء الاستدلالي، فالإحصاء الاستدلالي هى وصفية أيضاً لأن البيانات التى تقدر على مستوى العينة هى وصفية للعينة ولكن الإحصاء الاستدلالي تمتلك وظيفة اضافية إلى جانب الوصف وهى الوصول إلى تعميمات من العينة للمجتمع، ودائماً الإحصاء الوصفى يتناول متغير واحد على حدة بينما الإحصاء الاستدلالي يتناول دراسة متغير واحد فاكثراً وعليه لإحصاء استدلالي 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 لاستدلالي فى القرن العشرين خاصة من 1890 حتى 1950 على يد الإخوة Karl & Egon Pearson وفisher و Jerzy Neyman 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 Princeton من خلال العالم John Tucky بالولايات المتحدة الأمريكية 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 التجريبي 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 إحصائي 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 المصادقية، أو يتعامل مع بيانات منخفضة الثبات. وكما نعلم أن العمل الإحصائي يعتمد على بيانات واهم شئ يتوفر فى هذه البيانات جودتها بمعنى ثباتها يكون مرتفع والتعامل مع بيانات ثباتها أقل من 0.70 دائماً يؤدي إلى نتائج مشوهة وذلك لأن القياسات فى مجملها كانت لأخطاء وليس للظاهرة الحقيقية المراد قياسها.

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

الفصل الثانى

الإحصاء وعملية البحث (خطوات البحث)

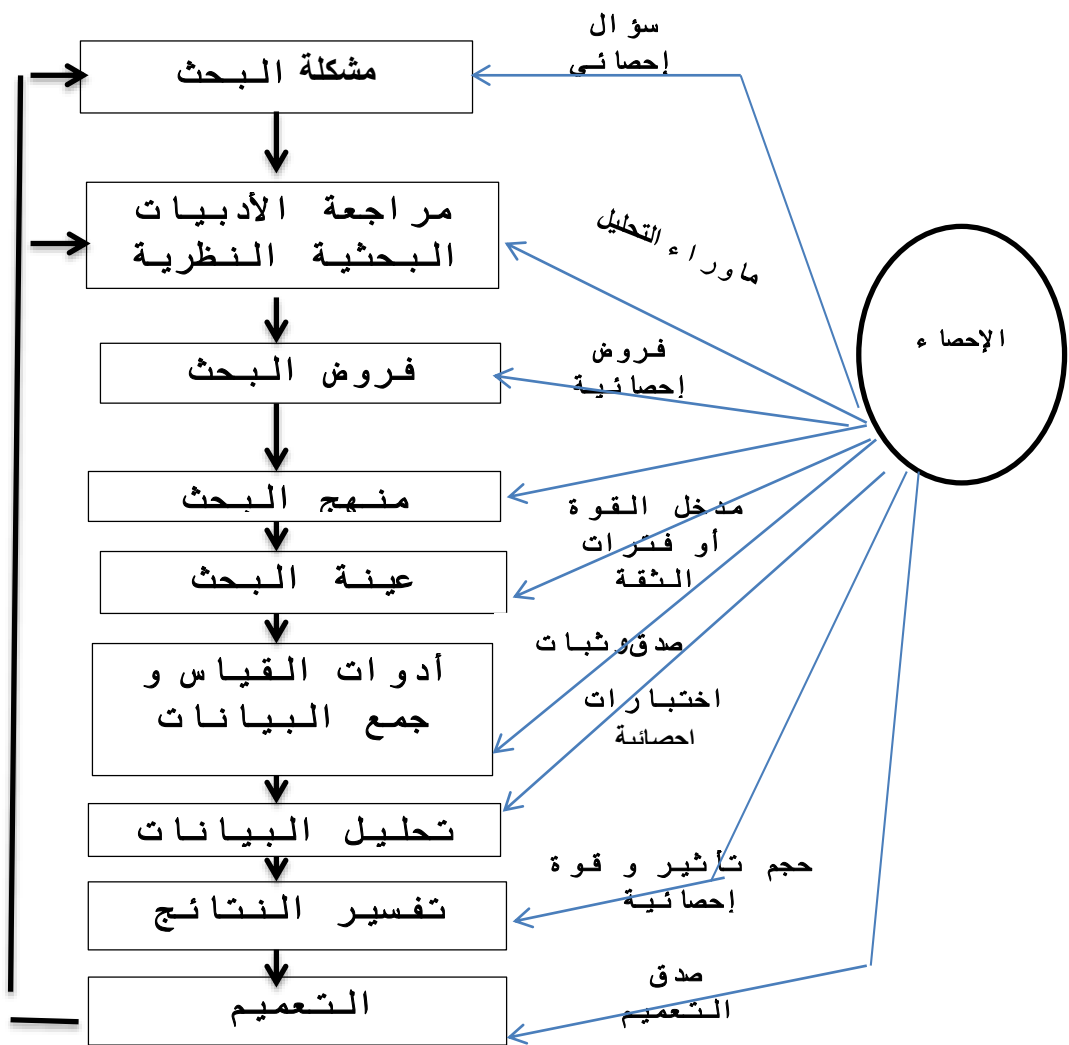
Statistics and Research Process

يختلف الباحثون حول مدى إسهام الإحصاء فى خطوات أو عمليات البحث، فيرى البعض أن دورها يقتصر على التحليل الإحصائي 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، فمثلاً 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 فسأله "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**، مثلاً 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 التي تناولت هذه المتغيرات من خلال تحديد الكلمات المفتاحية **Key words** والبحث في قواعد البيانات المتخصصة، والباحث يكون لديه كمًا هائلًا من الدراسات والبحوث، وفي 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 البعدي، مستوى الدلالة الإحصائية، حجم التأثير، القوة الإحصائية، و... إلخ.

والسائد في 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 هي المراجعة الكمية **Quantitative review** أو ما يطلق عليها البعض ما وراء التحليل **Meta-analysis** وهي استراتيجية منهجية لمراجعة الدراسات السابقة كميًا والوصول إلى متوسط حجم التأثير أو الارتباط بين متغيرين ما وهذا الاستخدام غائب في مراجعة الدراسات السابقة في البيئة العربية، وحساب هذا الأسلوب يعتمد في المقام الأول على الإحصاء مثل مؤشرات حجم التأثير.

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

- توجد فروق بين الذكور والإناث في الابتكارية.
- توجد فروق بين الذكور والإناث في 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 الصفريّة 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 التجريبي **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 التجريبي 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. ولكي ندرس السببية من خلال تجربة فلا بد أن تتضمن إجراءات صارمة منها:

- المعالجة لمتغير تجريبي مراد دراسته Manipulation.
- الضبط لكل المتغيرات الدخيلة التي يمكن أن تؤثر في المتغير التابع.
- العشوائية Randomization في اختيار عينة البحث وكذلك عشوائية في توزيع المشاركين على مجموعات التجربة بطريقة تحدث درجة كبيرة من التكافؤ Equivalence.

- المقارنة بين درجات المعالجات المختلفة وهي المجموعات التجريبية (التي تتلقى المعالجة) والمجموعة الضابطة (لا تتلقى المعالجة أو تتلقى مستوى آخر تقليدي من المعالجة)، فالفرق بين نواتج المعالجات دليل على أن المعالجة (المتغير المستقل) المسبب للتغيرات في درجات المتغير التابع.

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

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

عليه التصميم التجريبي للقياسات المستقلة Independent measures experimental Design وفيه تتم المقارنة بين المجموعات المختلفة وتحديد الفروق بينهما وفيما يلي 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 العشوائية، X المعالجة، O₁ قياس بعدى للمجموعة التجريبية، O₂ قياس بعدى للمجموعة الضابطة.

- تصميم المجموعتين التجريبيتين (المعالجتين) وقياس بعدى:

R XA O₁

R XB O₂

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

R XA O₁

R XB O₂

R O₃

- تصميم المجموعة الضابطة وقياس قبلي وبعدى:

R O₁ X O₂

R O₃ O₄

- تصميم المعالجات البديلة مع قياس قبلي:

R O1 XA O2

R O3 XB O4

- تصميم المعالجات المتعددة مع المجموعة الضابطة وقياس قبلي :

RO₁ XA O₂

R O₃ XB O₄

R O₅ O₆

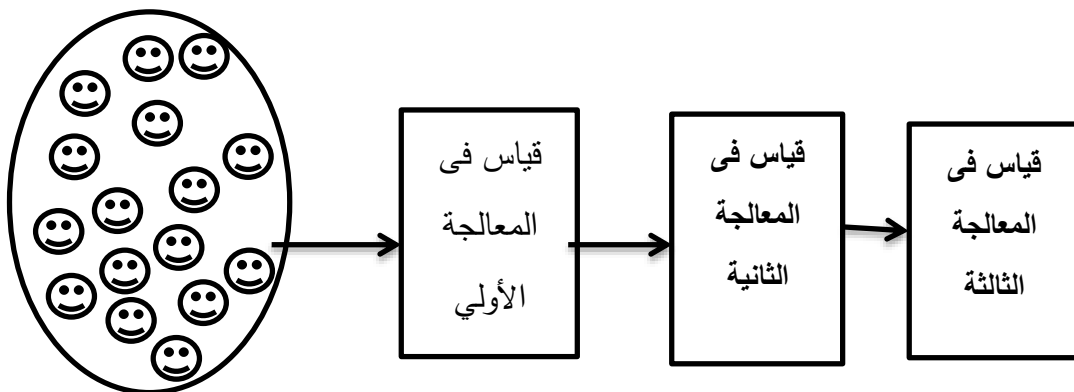
ويمكن تعميمه لأكثر من ثلاث مجموعات:

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 شيوعاً في البحوث النفسية والتربوية والسلوكية هو تصميم المجموعة الضابطة والتجريبية وقياس قبلي وبعدي، وكذلك تصميم المعالجات المتعددة مع المجموعة الضابطة وقياس قبلي. واستخدام الإحصاء في هذه النوعية من التصميمات يستخدم لدراسة الفروق بين القياسات البعيدة للمجموعات. ومن أهم الأساليب الإحصائية المستخدمة اختبار T للعينات المستقلة وكذلك تحليل التباين أحادي الاتجاه (ANOVA) مصحوباً باختبارات المقارنات المتعددة البعيدة Post-hoc لتحديد أي المجموعات التي أحدثت الفروق مثل اختبارات شيفيه، وتوكي ونيومان-كولز وغيرها. وكذلك يمكن استخدام اختبارات الإحصاء اللابارامترى في 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: وفي هذا التصميم تستخدم مجموعة واحدة من الأفراد ويطبق عليها كل معالجات أو مستويات المتغير المستقل ثم قياس أو ملاحظة أداء كل فرد في كل المعالجات ويطلق عليه تصميم القياسات المتكررة لأنه يحدث إعادة للقياسات على نفس الأفراد ولكن لظروف معالجات مختلفة وهذا التصميم سائد أيضاً في الدراسات غير التجريبية مثل الدراسات الطولية التي تتناول التغيرات للمتغيرات عبر فترة زمنية طويلة ويمكن تمثيله كالتالي:

المعالجة 3 المعالجة 2 المعالجة 1 نفس الأفراد



شكل (2.2): بنية التصميم التجريبي 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 فريدمان Friedman أو اختبار T المرتبطة والمقابل اللابارامترى مان - ويتنى أو اختبار الإشارة.

المنهج شبه التجريبي Quasi experimental 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 التجريبي 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 جداً نظراً لأن الظاهرة النفسية ذات متغيرات متعددة ومعقدة ومتفاعلة ولذلك فإن هذه التصميمات تعاني من نقص الصدق الداخلى Internal validity وهو الادعاء بأن التغير فى المتغير التابع يرجع ويرجع فقط إلى المتغير التجريبي وبالتالي فالتجارب التى تجرى فى واقعنا التربوى فى المدارس يغلب عليها التصميم شبه التجريبي ويطلق عليها تجارب الميدان أو الحقل Field experiments وعلى الرغم من ذلك فإن التصميمات شبه التجريبية لها درجة صدق خارجى External validity عالية 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:

$$X \quad O_1$$

ب. 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:

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

هـ. تصميم المعالجة المتكررة:

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

• تصميمات شبه تجريبية مع وجود مجموعة ضابطة بدون قياسات قبلية:

أ- تصميم القياس البعدى فقط مع مجموعات غير متكافئة :

$$N_R \quad X \quad O_1$$

$$N_R \quad O_2$$

N_R تعنى لاعشوائية Non-randomization

ب- تصميم القياس البعدى مع قياس قبلى لعينة مستقلة:

$$N_R \quad \text{----} \quad O_1 \quad X \quad O_2$$

$$N_R \quad \text{----} \quad O_3 \quad O_4$$

• تصميمات شبه تجريبية تستخدم مجموعات مقارنة وقياس قبلى - بعدى:

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

$$N_R O_1 \quad X \quad O_2$$

$$N_R \quad O_3 \quad O_4$$

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

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

$$N_R O_4 O_5 \quad O_6$$

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

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

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

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

المجموعة الضابطة المعالجة المعكوسة:

$$N_R \frac{O_1 \quad X \quad O_2}{O_3 \quad X \quad O_4}$$

● تصميم السلاسل الزمنية: يتضمن:

أ- تصميم السلاسل الزمنية البسيطة:

$$0_1 0_2 0_3 \quad 0_4 \quad 0_5 X \quad 0_6 \quad 0_7 \quad 0_8 \quad 0_9 \quad 0_{10}$$

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

$$\begin{array}{cccccccccccccccc} 0_1 & 0_2 & 0_3 & 0_4 & 0_5 & X & 0_6 & 0_7 & 0_8 & & & & & & & & \\ & & 0_9 & 0_{10} & 0_{11} & 0_{12} & 0_{13} & & 0_{14} & 0_{15} & 0_{16} & & & & & & \end{array}$$

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 هي الأكثر استخدامًا في علم النفس والتربية ولكن أحيانًا يستخدم الباحثون تصميمات معقدة Complex designs حيث يتناول معالجة متغيرين مستقلين فأكثر في نفس التجربة وهذه النوعية يطلق عليها التصميمات العاملية Factorial Designs لأن المتغيرات المستقلة يطلق عليها عوامل ويوجد فيها نوعين من التأثيرات هي التأثيرات الرئيسية Main effects وتأثيرات التفاعل Interactions effects.

Correlational method or design المنهج الارتباطي

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 التجريبي. فالبحث الارتباطى هو أحد الطرق للوصف فى صورة كمية موضحاً 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 سببي فالارتباطية لا تعنى بالضرورة سببية Correlation doesn't mean Causality ولكن وجود علاقة ارتباطية يمكن أن تكون بداية أساسية لدراسات تنبئية ودراسات 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 التجريبي 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 التجريبي حيث فى

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

- الفروق بين الذكور والإناث في العدوان.
 - الفروق بين المدخنين وغير المدخنين في الإصابة بالسرطان.
- ويلجأ الباحث إلى البحث السببي المقارن لدراسة ظاهرة ما عندما يتعذر دراستها تجريبيًا لاعتبارات أخلاقية.
- وفيما يلي ملخص التصميمات البحثية والاختبارات الإحصائية المصاحبة لها:

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

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

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

5. العينة أو المشاركون Sample: يتم انتقاء العينة بصورة عشوائية من مجتمع الدراسة، والمسلمة الأساسية فى الإحصاء الاستدلالي هي العشوائية. والإحصاء ليست مقطوعة الصلة بهذا الجانب فى البحث لأن العشوائية تتم فى ضوء قوانين الاحتمالات، وتلعب الإحصاء دورًا كبيرًا فى اختيار العينة فى ضوء مدخل القوة الإحصائية باستخدام التحليل القبلى للقوة الإحصائية لتحديد حجم العينة المناسب وكذلك فى ضوء مدخل فترات الثقة (للمزيد انظر العينات).

6. إعداد أدوات القياس وجمع البيانات: الإجابة المنظمة لأسئلة البحث تتطلب قياسات صادقة وفى هذه الخطوة يسهم علم القياس النفسى بإسهامات فعالة حيث يقدم أسس ومبادئ إعداد أدوات القياس مثل المقاييس والاختبارات والاستبيانات ومن شروط أداة القياس الجيدة هي الصدق والثبات. فالصدق يعطى أدلة على مصداقية بناء الأداة، أما الثبات يعطى موثوقية للبيانات المراد تحليلها ومعرفة أخطاء القياس Measurement errors. فى هذه الخطوة تلعب الإحصاء دورًا كبيرًا بالغ الأهمية وتقدم معادلات لحساب أدلة الصدق ومؤشرات الثبات.

7. المعالجة الإحصائية وتحليل البيانات: تتم المعالجة الإحصائية فى ضوء فروض البحث وكذلك الفروض الإحصائية وطبيعة المتغيرات المقاسة (متصلة - منفصلة)، وحجم العينة وأهداف التحليل (فروق - علاقة - تنبؤ)، وطبيعة توزيع البيانات (اعتدالية - ملتوية) كل هذه العوامل تحدد نوعية الإحصاء المستخدمة سواء كانت إحصاء بارامترى Parametric statistics أو إحصاء لبارامترى 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 يأتي دور الباحث في 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 مصحوباً 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 لكي 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.

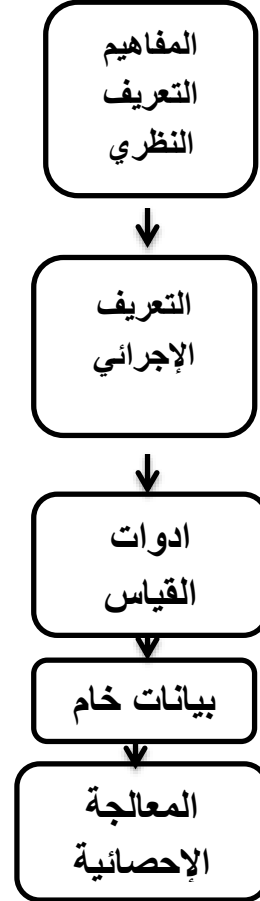
وفي ضوء ذلك تؤدي الإحصاء دوراً أساسياً في كل خطوات وعمليات البحث المختلفة ابتداءً من صياغة المشكلة وانتهاءً بالتعميم وهذا عكس ما يعتقد البعض بأن الإحصاء هي وسيلة تستخدم فقط في تحليل البيانات.

الفصل الثالث

المتغيرات والبيانات

Variables and Data

تتعامل الإحصاء مع البيانات Data وهذه البيانات تأتي من مقاييس أو استبيانات أو اختبارات بهدف قياس متغيرات ما. وفيما يلي 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 ويسمى 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:

أولاً: مستوى القياس الذى تتبع منه: وفى هذا الإطار تصنف المتغيرات حسب مستوى القياس Level of measurement أو مقياس البيانات Data scale هرمياً من الأقل رقيماً (الأسمى) إلى الأكثر رقيماً (النسبى):

1. مستوى القياس الاسمى Nominal scale: فى هذا المستوى البيانات عبارة عن تصنيفات مثل الديانة، الجنس، الوظيفة، التخصص، وغيرها. فمثلاً 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 والأرقام 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 الأكثر 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) وبالتالي يتضمن هذا المستوى التصنيف ثم الترتيب من الرتبة الأعلى إلى الأقل أو العكس. ويرى (2014) Miller 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 تفضيلاً، بينما (1) أقل الأطعمة تفضيلاً.

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 المعياري 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 الرتببة كأنها بيانات فترية باستخدام الإحصاء البارامترى.

3. المقياس الفترى Interval scale: يتضمن التعبير الكمي ويستخدم الأرقام

للتعبير عن الصفات أو الخصائص، والفترية الحقيقية تتضمن توزيع متصل وعلى ذلك فإن الفترات بين الأرقام لها معنى، فدرجة الحرارة (45) هي أقل خمس درجات من درجة الحرارة (50)، ودرجة العدوانية تتراوح من (20) إلى (80)، والدرجة (60) أكثر بعشر وحدات من الدرجة (50) ولكن (40) ليس هي ضعف (20). البيانات على مقياس ليكرت الخماسي أو السباعي أو التساعي يمكن التعامل معها على أساس بيانات فترية. والدرجة على متصل السمة لها معنى مختلف فالدرجة التحصيلية بين 9,10 تختلف عن الدرجة التحصيلية بين 1,2.

والدرجات على المقياس الفترى تعالج الفروق في الدرجات كأنها مسافات متساوية بين هذه الدرجات، ولا يوجد صفر كنقطة بداية، والصفر ليس نقطة حقيقية على المقياس بل هي افتراضية ويسمى صفر نسبي أو اعتباري Relative zero حيث انعدام الصفة على المستوى المقاس لا يعنى انعدامها على المستوى الحقيقى. والعلاقة بين الدرجة المقاسة X والدرجة الحقيقية T :

$$T = X + E$$

• E خطأ القياس

وإذا كانت $X = 0$ فإن:

$$T = E$$

وبيانات العلوم النفسية والاجتماعية تحلل على أساس أنها بيانات فترية ويعتبر استخدام إجراءات الإحصاء البارامترى لتمثيل البيانات الفترية محل تساؤل لأسباب عديدة أهمها ان اعتدالية البيانات الفترية هي محل جدل. عمومًا متغيرات العلوم النفسية والاجتماعية التى تقاس بطريق غير مباشر بصورة كمية تقع تحت هذا المستوى. وعلى ذلك فإن القياسات الفترية تسعى إلى التصنيف والترتيب من الأقل إلى الأكبر وكذلك توضح النسب بين المجموعات المتكافئة من الأرقام ويمكن استخدام إحصائيات مثل المتوسط والانحراف المعياري وهذه ضرورية للإحصاء البارامترى مثل اختبارات T و ANOVA و r حيث تفترض اعتدالية البيانات. وعلى الرغم ان البعض ينادى بضرورة

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

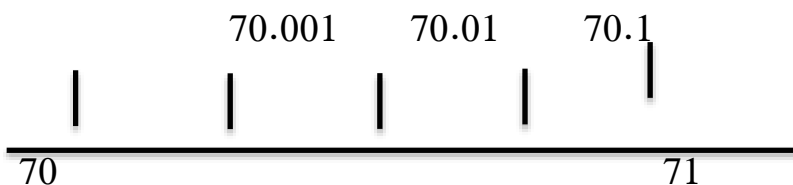
4. المقياس النسبي Ratio scale: يستخدم البيانات الكمية ولها فترات متساوية بين الارقام ويتوفر الصفر الحقيقى حيث انعدام الصفة على المستوى المقاس يفيد بانعدامها على المستوى الحقيقى، وايضا فالصفر يعنى لا شئ فالدخل والطول والمسافة والوقت هما متغيرات نسبية، فالفرق بين الوزن 40 و 50 هو نفسه الفرق بين 60 و 70. ويمكن القول إن الاستجابة A هي 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 كالاتى:

$$T = X$$

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 المعياري كمعالم للمجتمع وعليه فإن الإحصاء البارامترى المناسب لبيانات هذا المستوى.

ثانياً: تصنيف المتغيرات حسب اتصالياتها الى:

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



وإذا كانت اداة قياس تتميز بالحساسية الشديدة فإن عدد الأوزان الواقعة بين القيمتين 70, 71 لا نهائى، والمتغيرات المتصلة غير قابلة للعد نظرياً، ودائماً المتغيرات الفترية والنسبية هى متغيرات متصلة. وكأمثلة على المتغيرات المتصلة: العمر، الوزن، الطول، المسافة، والأمن النفسى. ولكن عدد أفراد الأسرة ليس متغيراً 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، فمثلاً 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 أيضاً **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 التجريبي أو التصميم البحثى كالاتى:

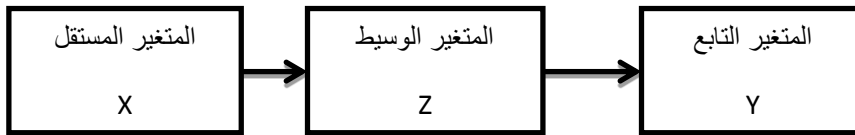
1. المتغير المستقل Independent variable: هو المتغير الخاضع لسيطرة الباحث أو سيطرة المجرب، وهو خاصية أو سمة تؤثر فى الناتج أو التأثيرات ، ويمكن للمتغير المستقل ان يؤثر فى المتغير التابع من خلال متغير وسيط أو ناقل بالتالى يكون التأثير غير مباشر، وتسمى المتغيرات المستقلة متغيرات منبئة، معالجات، ومتغيرات سابقة، فمثلاً 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 زمنياً عن المتغير (Y) ويسمى المتغير المستقل أحياناً 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. **ومتغيرات**

المعالجة هي نوع من المتغيرات المستقلة وتستخدم في البحوث التجريبية والمعالجات التربوية، فيعطى المعالجة لمجموعة بينما يستبعدا من المجموعة الأخرى. ويصبح السؤال ما إذا كانت المجموعة التى تلقت المعالجة تتفوق أو تختلف على المجموعة التى لم تتلق المعالجة. بالتالى فإن هذا المتغير تصنيفى يمكن أن يكون بمستويين فأكثر. وفى التراث البحثى تسمى هذه المتغيرات بالمعالجات أو المستويات أو ظروف المعالجة. والمتغيرات المستقلة فى التربية وعلم النفس كثيرة مثل برنامج علاجى أو تدريبى، مهارات ما وراء المعرفة، أثر عادات العقل، أثر عادات الاستذكار، عدد ساعات الاستذكار، العلاج السلوكى، العلاج العقلانى، وغيرها.

2. المتغير التابع Dependent variable: هو خاصية أو سمة تعتمد على أو تتأثر بالمتغير المستقل ويكتب فى تقارير البحوث تحت مسمى الناتج، التأثير، المحك، ويمكن للدراسة الواحدة ان تتضمن أكثر من متغير تابع ويمكن أن يكون قياسات كمية متصلة أو قياسات كيفية منفصلة. هو المتغير الناتج ويعتمد على سلوك المستجيبين ويتغير حدوثه نتيجة تغير المتغير المستقل أى أنه النتيجة، فالنجاح الأكاديمى فى الجامعة هو المتغير التابع وهو المتغير الذى يأتى بعد المتغير المستقل أو يحدث معه فى نفس اللحظة، ونواتج الدراسة (البيانات) هى متغيرات ناتجة تابعة. وامثلة للمتغيرات التابعة فى التربية وعلم النفس التحصيل، تنمية الابتكارية، تنمية الاتجاه، خفض القلق، تنمية الذكاء الاجتماعى، المهارات الإدارية، المناخ الأسرى، خفض الاكتئاب.

3. المتغير الضابطة Control variable : هى متغيرات مستقلة ولكن يضبط اثرها اثناء التجربة، فمثلا اراد باحث دراسة اثر التماسك الاسرى على الاكتئاب، فلدراسة هذا فإنه توجد متغيرات اخرى تؤثر على الاكتئاب مثل المستوى الاقتصادى للأسرة والوحدة النفسية وهى متغيرات تؤثر على النواتج فلا بد أن يضبطها الباحث ويستبعد اثرها عند تصميمه للتجربة من خلال قياسها. ومن امثلة هذه النوعية من المتغيرات الحالة الاقتصادية الاجتماعية، النوع، الجنس، والذكاء. ويمكن ضبط هذه المتغيرات إحصائيًا وإذا كانت متغيرات متصلة يطلق عليها تغايرات Covariates من خلال أساليب إحصائية مثل تحليل التغاير Analysis of Covariance

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



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

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

- ما هو النواتج في دراستي التي تحتاج لتفسير؟ (متغيرات تابعة).
- ما هي المتغيرات أو العوامل التي تؤثر في الناتج؟ (متغيرات مستقلة).
- ما هي المتغيرات التي احتاج لقياسها والتي من المحتمل أن تشكل عوامل أساسية في تأثيرها على النواتج وليس متغيرات أخرى؟ (المتغيرات الضابطة).
- ما هي المتغيرات التي ربما تؤثر على نواتج الدراسة ولا يمكن قياسها؟ (المتغيرات الدخيلة).

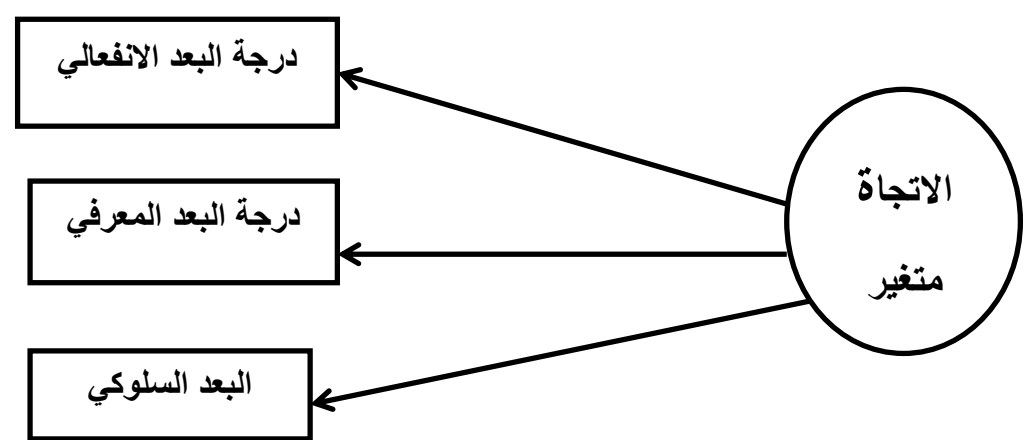
6. المتغيرات المتفاعلة **Moderatig variables**: هذه المتغيرات تستحق الانتباه وغالباً ما تستخدم في التصميمات والتجارب التربوية، هذه المتغيرات هي جديدة وتولد

نتيجة التفاعل مع متغيرات أخرى فمثلاً 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 المقاسة Observed 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 وأحياناً 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 (2003) Diamantopoulos & Schlegelmilch 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، التمر، 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 كالاتى:

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

م	وحدات التحليل	الجنس	العمر	دخل الالاب (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 التابعة 2 < 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 وفي هذا تسمى استجابات متعددة Multiple responses.

أنواع البيانات

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

أولاً: فى ضوء معناها: وتنقسم الى:

- أ- **حقائق Facts:** وهى عبارته عن خصائص أو مواقف موجودة منذ زمن بعيد مثل العمر والجنس والدخل والديانة وغيرها ويطلق عليها فى الدراسات بالبيانات الأساسية أو البيانات الديموجرافية Demographic data.
- ب- **معارف:** وهى معارف الفرد عن قضية ما أو مخترع ما أو صاحب انجاز رياضى ما وغيرها.

ج- **اتجاهات وآراء 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 مستقبلاً، ماذا يرغب الطالب ان يتخصص فى الجامعة؟

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

• **البيانات الأولية:** تجمع لأغراض محددة وواضحة عند الباحث، فمثلاً أراد 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 الملاحظة Observation methods باستخدام أجهزة فنية أو تسجيل الملاحظات أو غيرها.

• **البيانات الثانوية:** هى البيانات التى لم تجمع مباشرة للدراسة ولكنها تجمع لأغراض أخرى من مصادر أخرى مثل الإحصائيات المطبوعة Published 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 العلمية أو خدمات قواعد البيانات Database services التى تمدنا بمعلومات عبر الكمبيوتر من مختلف دول العالم مثل قاعده البيانات ERIC أو Academic research elliet حيث تساعد فى نقل مجموعات البيانات إلكترونياً من موقع إلى آخر عبر الدول.

ثالثًا: فى ضوء بعد الوقت: وهذا مرتبط بفترة زمنية سواء كانت فى فترة واحدة أو عدد من المراحل أو الفترات الزمنية. والبيانات فى هذا 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

الثانوية العامة منذ عام 2000 حتى الآن، وهذا يعرف بتصميم القياسات المتكررة Repeated measures design ومثل تتبع تحصيل التلاميذ في منهج العلوم عبر شهور السنة وهى تتبع نفس المتغير عبر فترات زمنية مختلفة ومما يلى مثال توضيحي:

الجدول (2.3): مثال لبيانات مستعرضة ولبينات طولية.

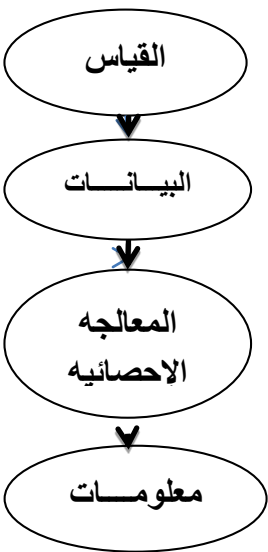
البيانات			المتغيرات
C	B	A	
10		10	تحصيل العلوم فى اكتوبر
11		20	تحصيل العربى اكتوبر
12		30	تحصيل الرياضيات اكتوبر
15	15		تحصيل العلوم مارس
10	25		تحصيل العربى مارس
7	20		تحصيل الرياضيات مارس

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 زمنية 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

إلى قرارات حول ظاهره معينة، ومعالجة البيانات الخام تكون باستخدام أساليب وتقنيات الإحصاء الوصفي والاستدلالي. ويمكن عرض العلاقة بين البيانات والمعلومات كآتي:



الشكل (3.3): العلاقة بين القياس والبيانات والمعلومات.

الفصل الرابع

أساسيات التعامل مع برنامج SPSS

يعتبر برنامج SPSS أحد أهم البرامج الشائعة الاستخدام في تحليل بيانات العلوم الاجتماعية والنفسية والسلوكية بالإضافة إلى برامج أخرى مثل SAS, Minitab, R, S-PLUS, STATA EXCEL وغيرها، إلا أنه أكثرهم يسرًا وسهولة في استخدامه. ويهتم البرنامج بحساب مؤشرات الإحصاء الوصفي من مقاييس النزعة المركزية ومقاييس التشتت والعلاقة والرسومات البيانية وكذلك كل الأساليب الإحصائية الاستدلالية البسيطة والمتدرجة بالإضافة إلى دور البرنامج في 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 في سهولة التعامل معه ووجود وظائف وخصائص لإجراء معالجة البيانات خصوصًا في اختبارات دلالات الفروق، إضافة إلى تفوق البرنامج في وجود طرق لمعالجة البيانات الغائبة بمجموعة من الطرق سواء بالحذف أو التعويض.

الكمبيوتر والإحصاء

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

برامج ممتازة: مثل برنامج R وهو متاح مجانًا ومتوفر على الموقع: <http://cran.R-project.org> حتى يتضمن كل الطرق الحديثة التي طورت في السنوات الأخيرة وكذلك الأساليب الكلاسيكية. وبرنامج S-Plus وهذا البرنامج مشابهة في جودته لبرنامج R.

برامج جيدة جدًا: مثل برنامج SAS.

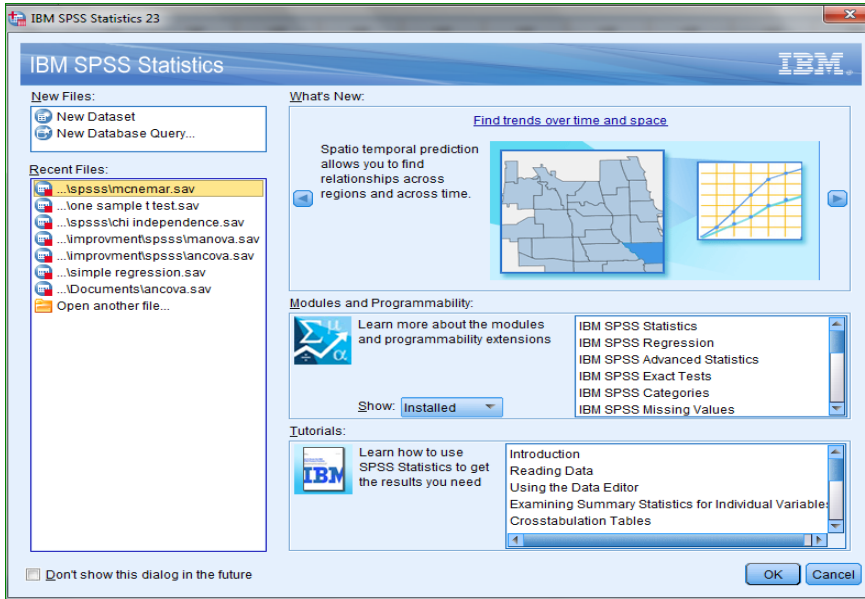
برامج جيدة: مثل برنامج Minitab

برامج غير مرضية: مثل برنامج SPSS وهو أكثر البرامج شهرة واستخدامًا وكذلك

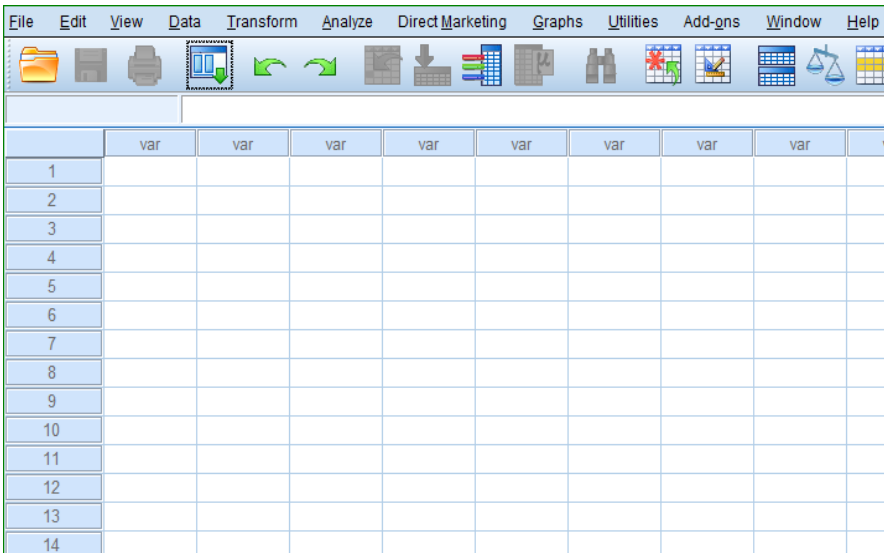
برنامج Excel.

فتح البرنامج

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



2. اضغط OK أو Cancel تظهر الشاشة الافتتاحية للبرنامج كالتالي:



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

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

قوائم البرنامج

أولاً: قائمة File: اضغط على File تظهر الأوامر الآتية:

New: إنشاء ملف بيانات ويمكن أيضاً إنشاء ملف أوامر Syntax.

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 في موضع آخر.

Exit: الخروج من البرنامج.

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

ثالثاً: قائمة View : تستخدم لإظهار أو إخفاء Status Bar أو شريط القوائم أو إظهار fonts أو Grid lines .

رابعاً: 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.

خامساً: قائمة التحويل Transform: تستخدم في تحويل قيم متغير إلى قيم أخرى وفي إنشاء متغير جديد وحساب متغير جديد من خلال امر Compute وإعادة ترميز قيم متغير والتعامل مع البيانات الغائبة.

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

سابعاً: قائمة The Direct Marketing: قائمة متخصصة جداً وتسمح لمستخدمى البرنامج تصميم قوائم لأهداف خاصة ولن تستخدم في هذا الكتاب.

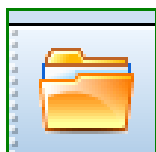
سابعًا: قائمة الرسومات **Graphs menu**: وتستخدم لرسم المدرج والخط والمنحنى و **Bar** وغيرها من الرسومات والأشكال البيانية.

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.

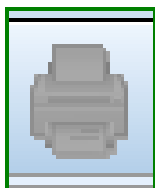
شريط **Tool Bar**: يتضمن عدة أيقونات كالاتى:



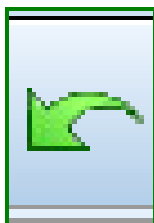
- فتح ملف بيانات سبق حفظه Open data document:



- حفظ بيانات جديدة Save this document:



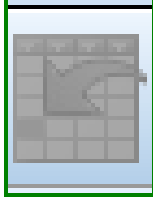
- طبع ملف Print :



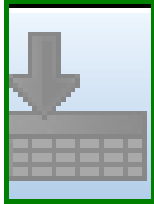
- العودة إلى تغير سابق Undo a user action:



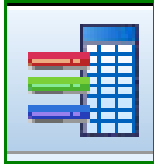
- العودة إلى تغيير لاحق Redo a user action:



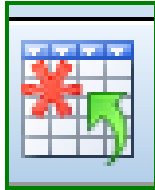
- الذهاب إلى حالة معينة برقمها Go to case :



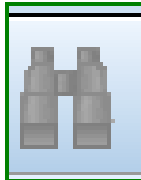
- الذهاب إلى متغير ما Go to variable:



- إعطاء معلومات عن متغير Variables:



- إجراء إحصاء وصفى Run descriptive statistics:



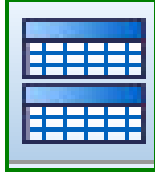
- إيجاد قيمة معينة Find:



- حشر حالات في ملف البيانات Insert cases:



- حشر متغير جديد في ملف البيانات Insert variable:



- تجزئة ملف البيانات Split file:



- وزن حالات Weight cases :



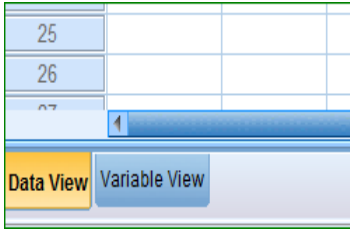
- اختيار حالات بمحكات معينة Select cases:



- إظهار كل المتغيرات Show all variables:

تسمية المتغيرات Naming Variable

اتبع الخطوات الآتية:



1. في الجزء الأسفل على يسار الشاشة الافتتاحية اضغط

على Variable view تظهر الشاشة الآتية:

	Name	Type	Width	Decimals	Label	Values
1	sex	Numeric	8	2		None
2	age	Numeric	8	2		None
3	achievement	Numeric	8	2		None
4						
5						
6						

حيث توجد مسميات عديدة وهي Width

Name , Type , ... إلخ، وكل صف

يعكس متغير، والأعمدة تمثل خصائص

المتغيرات.

2. تحت عمود Name في الخلية

الأولى (الصف الأول - العمود الأول)

اكتب مسمى المتغير وليكن الجنس واكتبه Sex أو أي مسمى آخر، ثم اضغط Enter

أو السهم، ثم اكتب مسمى المتغير الثاني وليكن العمر Age في الخلية الثانية ثم

اضغط Enter أو السهم، ثم اكتب مسمى المتغير الثالث وهكذا.

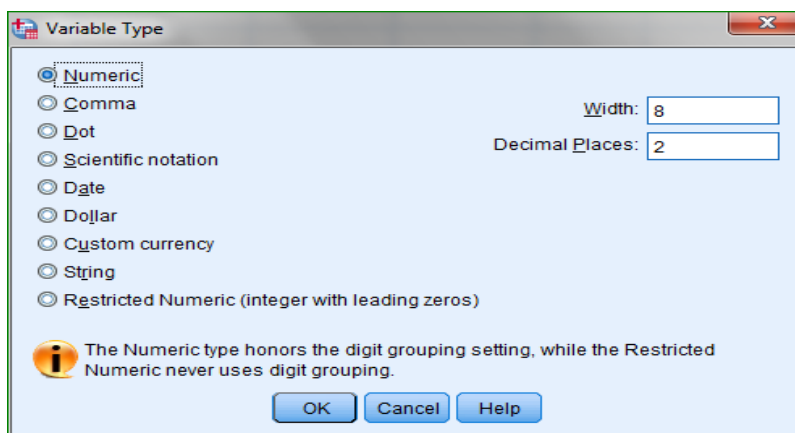


ويجب مراعاة الآتي:

- أ - مسمى المتغير لا يزيد عن 32 حرف في النسخ الحديثة بينما في نسخة 12 لا يزيد عن 16 حرف، أما في نسخ 11 ، 10 لا يزيد عن ثمانية أحرف.
- ب - لا يتضمن الاسم مسافة.

خصائص المتغيرات

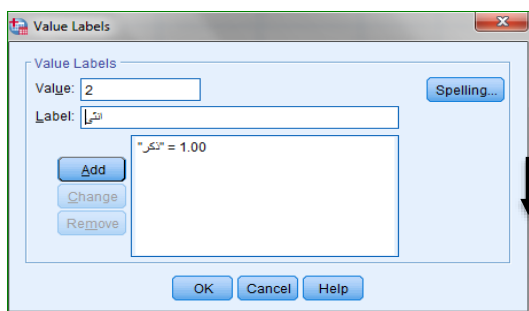
- **Type** : يحدد نوع البيانات المدخلة للمتغير سواء أرقام أو حروف أو رمز علمي أو عملة مالية أو تاريخ أو مقطع **text** واضغط على العلامة [...] لتحديد نوع المتغير تظهر الشاشة الآتية:



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

المتغيرات التصنيفية، فمثلاً لتحديد توصيف كود الجنس اضغط على العلامة الزرقاء تظهر الشاشة الآتية:

- اكتب 1 في 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 ترتيبي 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.

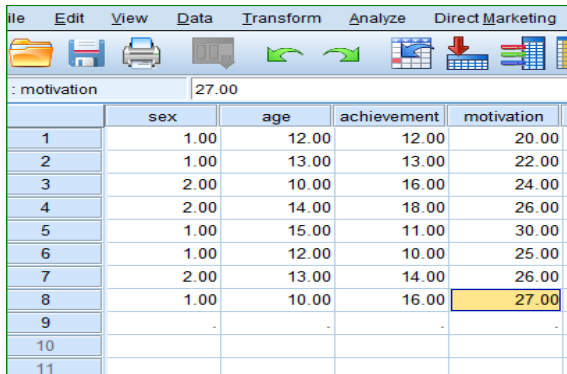
وأهم هذه الأعمدة هى **Measure** , **Value** , **Type** حيث تستخدم بصفة دائمة، وعليه تم

إدخال أربعة متغيرات هى **Sex** والتحصيل **Achievement** والدافعية **Motivation** كما فى

الشاشة الموضحة أسفـل.

إدخال البيانات

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



	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				

عمود متغير.

ويمكن إدخال بيانات المتغير (العمود)

فاضغط بالماوس على الخلية الأولى ثم

اكتب قيمة المتغير وليكن 1 ثم اضغط

Enter ينتقل إلى الحالة الثانية (الصف

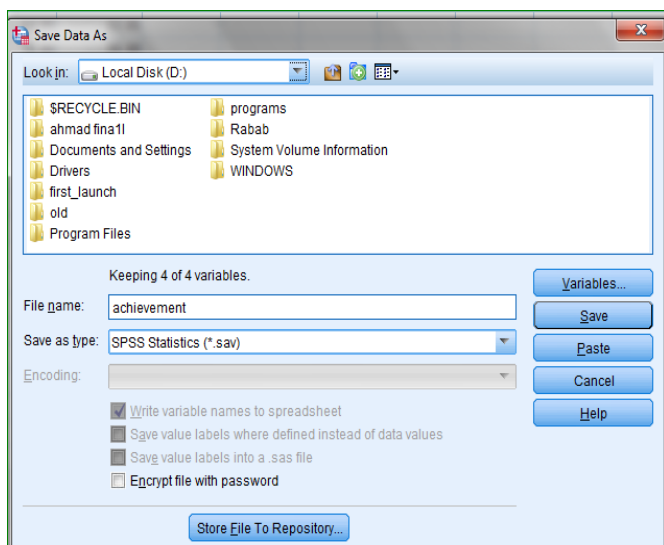
الثانى) اكتب 1 وهكذا، أما إذا تم

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 في مستطيل
File name وحدد موضع الحفظ
 ثم اضغط **Save** وإذا أردت تغيير
 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** — **Open** — **File**

2. إذا ظهر لك مسمى الملف في البرنامج اضغط عليه مرتين بالماوس أو اضغط عليه بالماوس
 ثم اضغط **Ok**.

حشر حالة في ملف البيانات

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

2	1.00	13.00	13.00	22.00
3				
4	2.00	10.00	16.00	24.00

1. اضغط **Edit** ثم اضغط **Insert cases**

2. اضغط بالماوس على الحالة 3 تظهر كالاتى:

حشر متغير في ملف البيانات

1. اختار موضع المتغير المراد إدخاله أو حشره ودائماً اختار الخلية على اليمين، فمثلاً يرغب
 الباحث في حشر متغير الذكاء **Intelligence** بعد متغير **Age** فأضغط بالماوس على متغير
Achievement.

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

2. اضغط قائمة **Edit** ثم اضغط **Insert**

variables تظهر الشاشة:

حذف حالة

لحذف حالة في ملف البيانات:

1. اضغط على الصف الخاص بالحالة.

2. اضغط **Edit** ثم اضغط **Clear** أو **Del** في لوحة المفاتيح أو اضغط على يمين الماوس

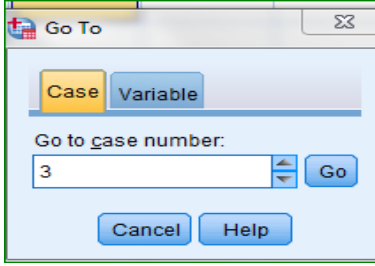
ثم اضغط **Clear**.

حذف متغير

لحذف متغير من ملف البيانات:

1. اضغط على العمود الخاص بالمتغير.
2. اضغط قائمة **Edit** ثم اضغط **Clear** أو مفتاح **Del** في لوحة المفاتيح.

للذهاب إلى حالة بعينها



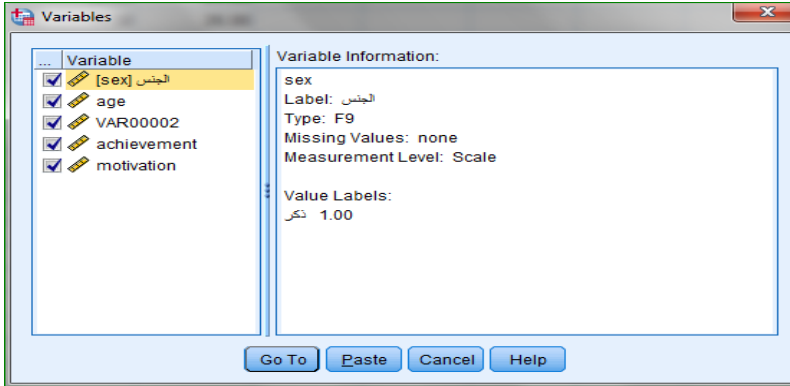
1. اضغط على قائمة **Edit**
2. اضغط **Go to case** تظهر الشاشة:
3. اكتب رقم الحالة في مربع **Go to case number** لتكن 500

4. اضغط **Go**

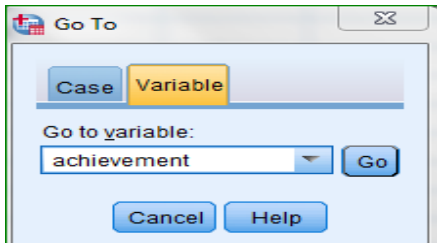
إعطاء نظرة عن المتغيرات المتضمنة في ملف البيانات

1. اضغط. **Utilities**

2. اضغط **variables**



البحث عن متغير ما في ملف البيانات



1. من قائمة **Edit** اضغط **Go to variable**:
2. اكتب مسمى المتغير في مربع **Go to variable**
3. اضغط **Go**

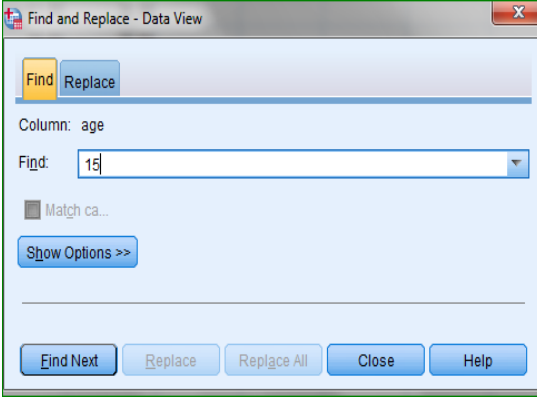
الذهاب إلى المتغير في قاعدة البيانات وهذا مهم في قواعد البيانات الكبيرة.

البحث عن قيمة معينة

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

2. اضغط قائمة Edit

3. اضغط Find تظهر الشاشة:



4. اكتب القيمة 15 مثلاً التي تمثل العمر

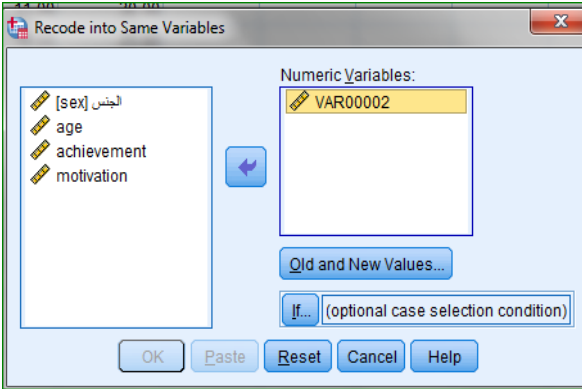
5. اضغط Find Next يقوم البرنامج بتحديد القيمة للمتغير.

إعادة ترميز البيانات

أحياناً عند إدخال بيانات مفردات المقاييس فإن بعض العبارات تكون موجب الصياغة وبعض العبارات سالبة الصياغة ويريد الباحث إعادة ترميز تصحيح العبارات السلبية فمثلاً القيمة 5 تصبح 1 ، القيمة 1 تصبح 5 وهكذا ولتنفيذ ذلك اتبع الآتي:

1. اضغط قائمة Transform

2. اضغط Recode into Same Variables



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 مثلاً ثم في New

1 Valu اكتب القيمة الجديدة

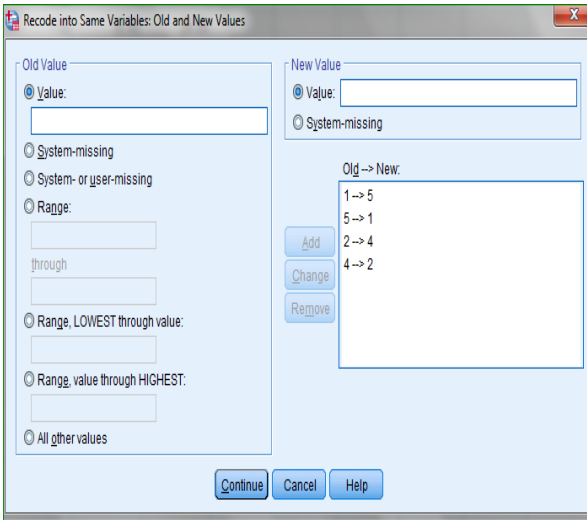
6. اضغط Add وهكذا لباقي القيم

المراد تكويدها

7. اضغط Continue ثم اضغط Ok

إنشاء متغير جديد

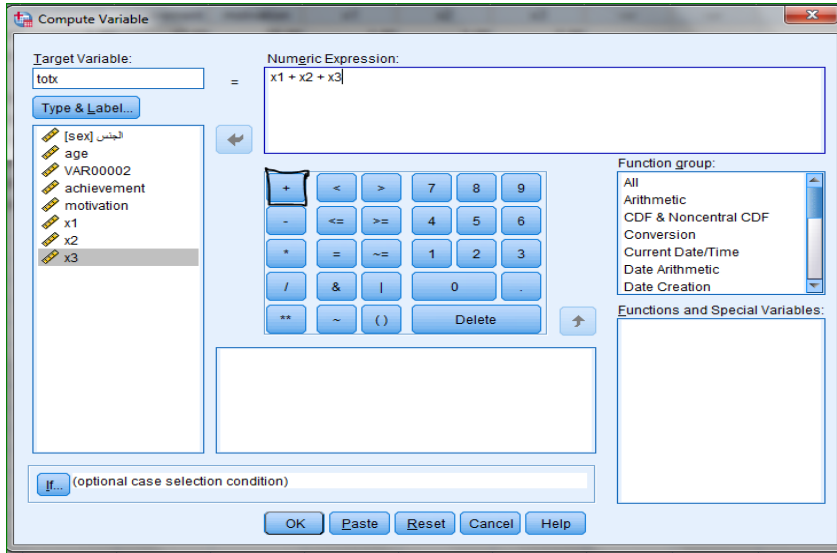
يقوم الباحث بإدخال درجات مفردات



المقاييس أو الاختبارات في ملف البيانات ولكن المقياس يتكون من إبعاد فرعية ويرغب في 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 مثلاً **totx**

4. اضغط على المتغير **X1** مثلاً وانقله بالسهم إلى مربع **Numeric Expression** ثم اضغط علامة (+) (بالموس).

5. اضغط على المتغير **X2** في مربع **Type & Lable** وانقله إلى مربع **Numeric Expression** ثم اضغط بالموس 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 في ملف البيانات تحت

مسمى **totx**

حفظ الرسم البياني

1. اضغط **File** ثم **save**

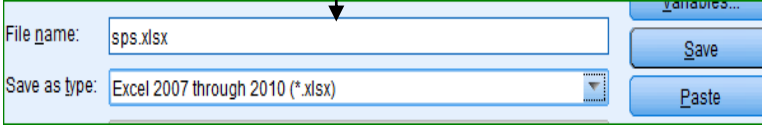
2. اكتب مسمى الملف **line** مثلاً في مربع **File Name** لاحظ أن امتداد الملف **SPV**.

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 في **Excel** وهو **.xlsx**.



3. اضغط **Save** على ذلك يتم حفظ الملف البيانات كملف بيانات في **Excel** في برنامج

Excel، ويمكن حفظ ملف البيانات **SPSS** في أى برنامج آخر مثل:

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

استيراد ملف من برنامج آخر في برنامج **SPSS**

1. اضغط **File** ثم اختر **Open**

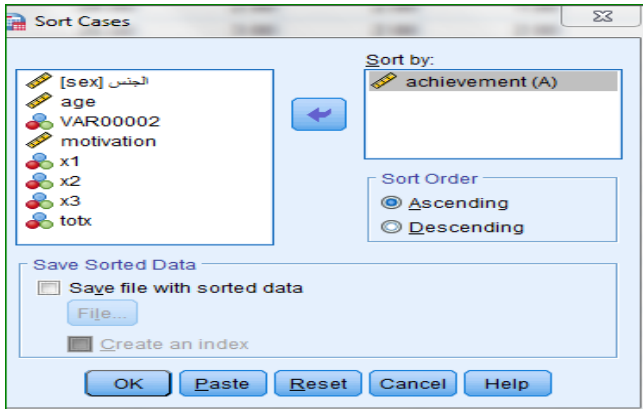
2. حدد ملف البيانات في **Excel** واختر امتداد **Save** من الملف وهي ملف **SPSS**.

ترتيب البيانات Sorting Data

حيث يرتب قيم المتغيرات تنازلي أو تصاعدي في ملف البيانات كالاتي:

1. اضغط قائمة **Data**

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



3. انقل المتغير المراد ترتيب قيمه

وليكن achievement.

4. حدد نوع الترتيب تصاعدي

Ascending أو تنازلي

Descending

5. اضغط **Ok**.

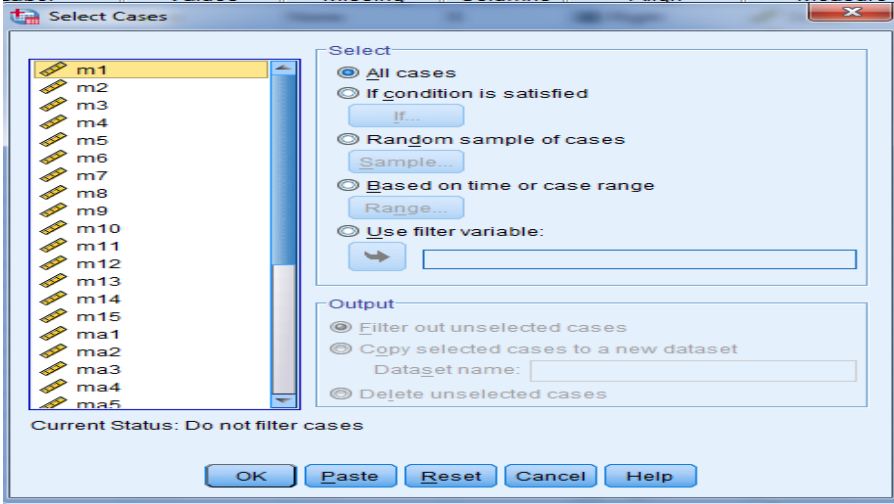
لاحظ الترتيب يتم لكل المتغيرات

في ملف البيانات.

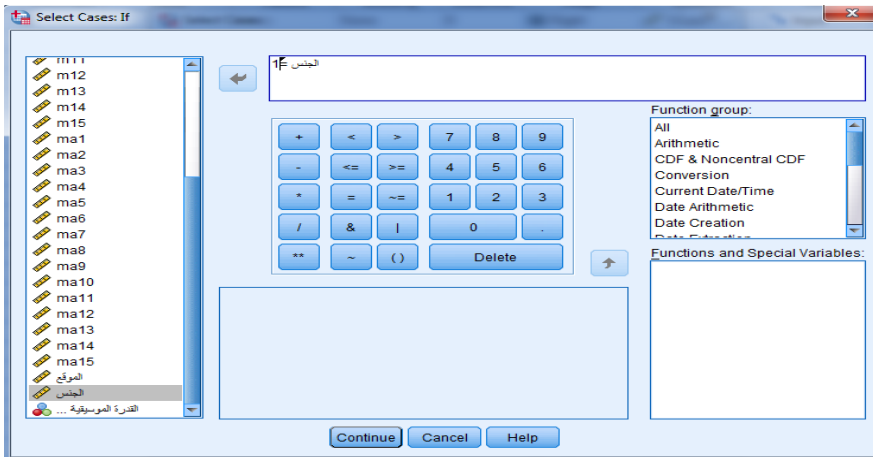
اختيار حالات Select cases

يتم اختيار عينة من الحالات موضع الدراسة كاختيار طلاب من شعبة اللغة الفرنسية مثلاً لاختبار فرض إحصائي عليهم، أو اختيار طلاب من عينة الذكور من العينة الكلية، ويتم إجراء الأمر كالتالي:

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



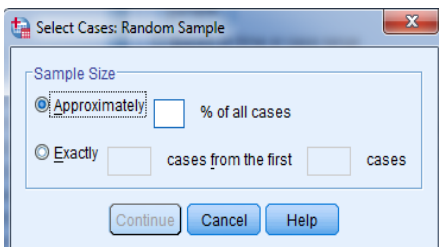
2. اضغط اختيار **If condition is satisfied** فيقوم باختيار أحد تصنيفات أفراد العينة وليكن الجنس = 1 (الذكور).



3. اضغط على **Continue** ثم **OK**.

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

أما اختيار **Random sample of**



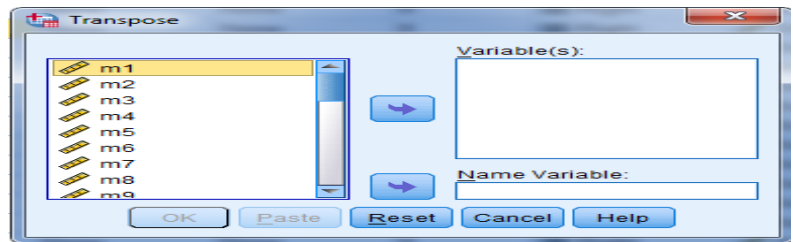
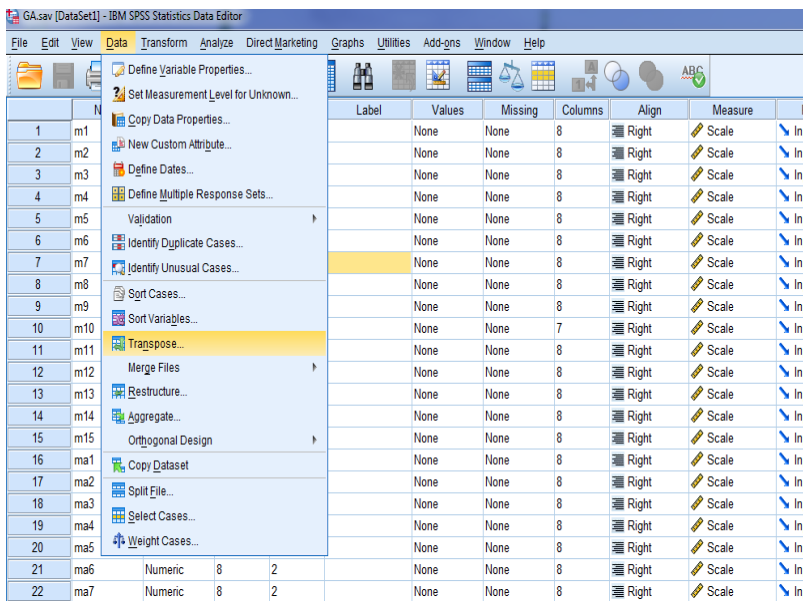
cases فيستخدم بغرض اختيار عينة عشوائية من ملف البيانات بغرض انجاز البحوث والضغط على **Random sample cases** تظهر النافذة التالية:

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

التحويل Transpose

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

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



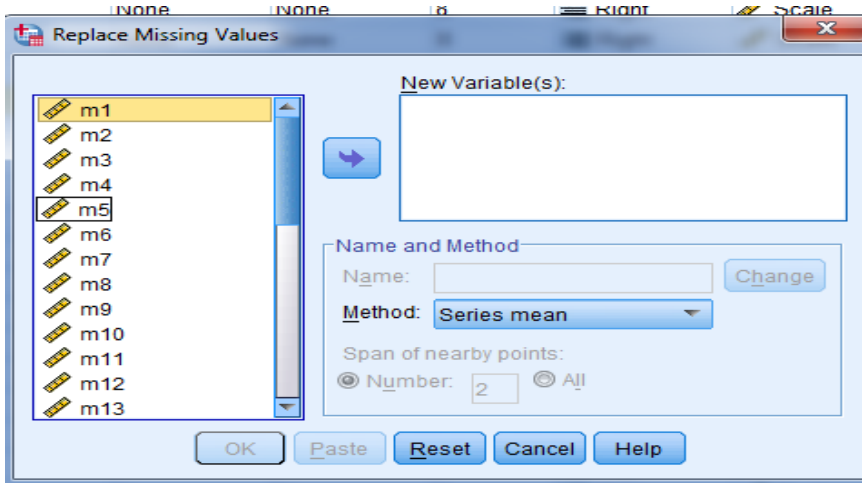
2. انقل المتغيرات المراد تدويرها إلى مربع **Variables**

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

استبدال البيانات الغائبة

ويتم إجراء الأمر من خلال:

1. اضغط قائمة Transform واختار **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 التي 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.

الفصل الرابع

الثبات

Reliability

بعد جمع البيانات وادخلها لا بد من تقدير جودة البيانات من خلال تقدير الثبات، ولكي يكون المقياس ثابت فإن الدرجات التي يمدنا بها تمثل أو تعكس الحالة الحقيقية للمتغير المراد تقديره وأي تغير في الدرجة المقاسة لا بد أن يقابلها تغيير فعلى حقيقى فى البناء التحتى للسمة المقاسة. والمقياس تام الثبات Perfect reliability يجب أن يكون انعكاساً تاماً 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 أولاً إلى ماذا نقيس، فالقياس دائماً منصّباً حول مفهوم أو بناء أو ظاهرة، والمفهوم تعبيرات مجردة ونتعرف عليه من خلال ملاحظات وفى البحث العلمى نتكلم أيضاً 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 أخرى فمثلاً 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 أحداث ملاحظة أو مقاسة عن طريق تحديد ماذا على الباحث عمله لقياس المفهوم.

والباحث دائماً يبحث عن إجابات واضحة عن أسئلته المراد دراستها ويريد أن يتأكد من مظاهر وخطوات بحثه صادقة وثابته وكلاً 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 وتتأظر الدرجة الحقيقة لهذه الخاصية. بكلمات أخرى نحن نفترض تتأظر تام بين الدرجة المقاسة أو الملاحظة **Observed(X)** والدرجة الحقيقة **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 نسبياً 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 الحقيقة 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 الحقيقة 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 إدارة 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

تنقص درجات الآخرين، وهذه الزيادة والنقصان يحدث بينهما توازن ويمكن أن يمحوا أحدهما الآخر.

مفهوم الثبات

يعرف بالثبات حيث إن الدرجة القياسية تعكس تمامًا الدرجة الحقيقية ($T = X$) أى أن ثبات القياسات يعنى أن الدرجة الحقيقية تساوى أو تتأطر تمامًا الدرجة المقاسة وفى هذه الحالة يقال ان الاداة تامة الثبات. أو إلى أى مدى أن القياسات حرة خالية من كلاً 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، إذاً:

$$X = T + S$$

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

المحك الأساسى لتقويم جودة أدوات القياس هو الثبات ويقال أن أداة القياس تتميز بالثبات لو أعطت نتائج متماثلة (أو قريبة للتماثل) عندما تطبق عدة مرات على نفس الأفراد تحت نفس الظروف، على سبيل المثال لو طبقنا اختبار ذكاء على مجموعة من الأفراد اليوم وطبقنا نفس الاختبار على نفس الأفراد تحت نفس الظروف الأسبوع القادم فنحن سوف نحصل على درجات متقاربة أو هى نفسها فى التطبيقين، بالتالى فإن الثبات هو استقرار أو اتساق القياسات المتولدة من أداة قياس محددة عبر الزمن على نفس الأفراد تحت نفس الظروف، وعلى ذلك فالثبات هو اتساق أو استقرار القياس ويعنى الاتساق للدرجات المتحصل عليها من تطبيق إلى آخر. فإذا كان الاختبار ثابت فنحن نتوقع أن التلاميذ الذين يحصلون على درجات عالية فى التطبيق الأول سوف يحصلون على درجات عالية فى التطبيق القادم، ويمكن للدرجات المتحصل عليها من 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 والباقي 10% يعكس تباين الخطأ.

وعموماً 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 منخفضاً، وعندما يكون صغيراً يكون الثبات عالياً.

طرق تقدير الثبات:

معاملات الثبات يمكن تصنيفها فى ثلاثة تصنيفات عريضة كالتالى:

- المعاملات المشتقة من ادارة نفس الاختبار فى مناسبات مختلفة (ثبات الاختبار واعدته).

- 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 وإعادة 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 فأكثر 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 اتساقاً وثباتاً ولكن متغير الحالة المزاجية لا يتميز بالاستقرار والثبات لفترة طويلة. وثبات الاختبار وإعادة غير مناسب وغير واقعى فى الأغراض التربوية.

واستقرار الدرجات عبر فترة زمنية شهر أو شهرين تعتبر دليلاً كافياً لثبات الاختبار وإعادة ولو كانت الفترة الزمنية بين مرتى إدارة 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 يؤدي إلى زيادة الثبات بصورة اصطناعية. ويمكن تجنب ذلك من خلال جعل الفترة الزمنية بين مرتى التطبيق طويلة مثلاً 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 زمنياً. 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 الأقل استقراراً مثل سمات الشخصية والحالة المزاجية.

الصيغ المتكافئة أو المتوازية Equivalent forms method or parallel

تستخدم عند تطبيق صورتين متكافئتين تماماً 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 فى الصيغتين فإن معامل الارتباط هو معامل الثبات ويطلق عليه معامل الاستقرار (فترة زمنية) والتكافؤ (صيغ مكافئة) 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 دليلاً قوياً 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 نستعير عنها بالعلاقة القوية بين المفردات وعلى ذلك فالمقياس المتسق داخلياً لا بد أن تكون العلاقات بين مفرداته قوية، العلاقات القوية بين المفردات Inter-item correlations تقتضى أن كل المفردات تقيس نفس الشيء بالتالى علاقات قوية بين المفردات والمتغير الكامن، بالتالى المقياس أحادى البعد Unidimensional للمقياس متعدد الأبعاد (كل بعد كأنه اختبار) يتكون من مفردات ترتبط ارتباطاً عالياً مع بعضها البعض (DeVellis, 2017).

للمفاهيم المعقدة مثل الذكاء أو الشخصية تقاس باستخدام اختبار أو استبيان حيث يتكون من مفردات عديدة والفكرة هي أن المفردة بحد ذاتها أو السؤال غير كافٍ لإعطاء صورة كاملة أو قياس كامل للبناء والمثال الشائع هو استخدام امتحانات عديدة لقياس الأداء في مقرر ما، والقياس النهائي لكل فرد يتحدد عن طريق اضافة أو متوسط استجابات الأفراد عبر مجموعة كبيرة من المفردات والمسلمة الأساسية في 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 الشئ 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 عشوائياً والطريقة الأكثر استخداماً 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 معاً فسوف يكون الاختبار أكثر ثباتاً 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 كالاتى:

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

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

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 متساوياً.

• 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 كالاتى:

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

- عدد الأسئلة (المفردات) K .

- متوسط الدرجة الكلية $\bar{X}_T$.

- تباين الدرجة الكلية S_T^2 .

فإذا كان $K = 50$ ، $\bar{X} = 40$ ، $s^2 = 16$ 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 ثابت منخفض.

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

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

- K عدد المفردات

- p_i معامل الصعوبة (عدد الاستجابات الصحيحة على الفردة مقسومًا على عدد

الأفراد الذين حاولوا الاستجابة على الفردة).

- 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 & Livingston, 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 $(1-p)$.

فمثلاً معامل الصعوبة للمفردة الأولى Item 1 كالتالي:

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

• n عدد الطلاب الذين اجابوا اجابة صحيحة على المفردة.

• N عدد الطلاب الذين حاولوا الاجابة على المفردة.

ومعامل السهولة:

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

إذا حصل ضرب معامل الصعوبة في 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 في الصعوبة:

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

وهكذا لكل المفردات بالتالي مجموع حاصل ضرب معامل الصعوبة في السهولة لكل المفردات

كالتالي:

$$\begin{aligned}\sum p_1 q_1 &= (0.8333 \times 0.1667) + (0.5 \times 0.50) + (0.6667 \times 0.333) \\ &\quad + (0.8333 \times 0.1667) + (0.6667 \times 0.3333) \\ &= 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222 = 0.972\end{aligned}$$

اما تباین الدرجة الكلية يقدر كالتالى:

- حساب المتوسط:

$$\bar{X}_T = \frac{4 + 5 + 2 + 1 + 5 + 4}{6} = \frac{21}{6} = 3.5$$

- حساب التباين كالتالى:

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

وبالتطبيق فى معادلة كودر-ريتشاردسون 20 :

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

ويقدر الثبات باستخدام معادلة كودر-ريتشاردسون 20 كالتالى:

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

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

- **المعامل ألفا Alpha Coefficient** : ابتدعه (1951) Cronbach ويرمز له بالرمز α وهو أحد الطرق الأكثر انتشارًا لتقدير معامل الاتساق الداخلي، وتعتبر صيغ كودر- ريتشاردسون 20 و 21 حالة خاصة من ألفا للمتغير اللاتيني ~٤٩ فئة لطريقة ألفا عندما تكون الأسئلة موضوعية تصحيحها (1، 0) وكما نعلم ان التباين 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 معينين).
 ويقوم معامل ألفا بتجزئة التباين الكلى بين مجموعة من المفردات إلى مكونين أحدهما التباين الحقيقى والآخر تباين الخطأ وقيمة ألفا تعبر عن التباين الفعلى الكلى بين المفردات التى تعزى إلى المتغير الكامن أو البناء التحتى للسمة وليس إلى شئ اخر .
 وتحدد قيمة المعامل ألفا كرونباخ بالصيغة الآتية:

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

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

وهذا المقياس للثبات يمكن التعبير عنه فى ضوء الارتباطات حيث هو عبارة عن متوسط الارتباطات الداخلية بين كل المفردات ببعضها البعض أو بكلمات أخرى هو يمدنا بمتوسط كل معاملات الثبات باستخدام التجزئة النصفية الممكنة. ولذلك فإن قيمة المعامل ألفا قائمة على متوسط كل معاملات الارتباط الداخلية بين المفردات وعدد المفردات فى المقياس وهذا يعنى أنه كلما زادت عدد المفردات فى المقياس ارتفعت 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 في المقياس.

وتتميز صيغة α كرونباخ بالتعقيد في حساباتها ولذلك فهي متوفرة في 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 Polytomously .

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

$$\begin{aligned}\sum SD_i^2 &= SD_1^2 + SD_2^2 + SD_3^2 + SD_4^2 + SD_5^2 \\ &= 0.8056 + 0.3333 + 1.4722 + 1.5556 + 1.4722 \\ &= 5.6389\end{aligned}$$

وتباين الدرجة الكلية:

$$SD_T^2 = 18.81$$

إذاً المعامل ألفا كالتالي:

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

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

ويرى (Sijtsma 2009) أن المعامل ألفا ليس الطريقة الأمثل لتقدير الثبات لأنه يمثل الحد الأدنى للثبات الفعلي لمجموعة من المفردات ويشير إلى أن فكرة الاتساق الداخلي غامضة ومبسطة عن الاتساق الداخلي للمقياس، فالاتساق يدور حول البناء العاملي لمجموعة من المفردات واستخدام ألفا لهذا الغرض تبسيط كبير، كما أن ألفا صمم لبيانات متصلة وحسابه يعتمد على معامل ارتباط بيرسون بالتالي 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 حد ما، غير موافقو غير موافق على الإطلاق) وفي هذه المقاييس لا نستطيع افتراض تساوي الفترات بمعنى أن الفرق بين استجابة موافق بشدة وموافق ليس هو نفس الفرق بين غير موافق بشدة وغير موافق، بالتالي عدم تساوي الفترات يقود إلى تقدير غير دقيق للمعامل ألفا واقتروا المعامل ألفا الرتبي Ordinal alpha حيث لا يفترض القياسية الفترية ومدخلهم استبدل معامل ارتباط بيرسون بمعامل الارتباط الرباعي Tetrachoric (بين المفردات تصحيحها 1, 0) أو معامل ارتباط Polychoric (اختيارات استجابات أو تصنيفات رتبية) ولأنهما يفترضوا أن البيانات ذات بناء تحتى متصل. واستخدام ألفا بصورتها الأصلية لبيانات رتبية يعطى تقلص لقيمة معامل الثبات بالتالي يؤدي إلى حدوث تحيز مقارنة باستخدام ألفا الرتبية (Gadermann et al., 2012).

ولاحظ (Dunn, Baguley, & Brunsden 2014) أن الصعوبات الرئيسية مع استخدام المعامل ألفا كمقياس للاتساق الداخلي تتلخص في الآتي:

- اعتماده على مسلمات من الصعب تحقيقها وهي 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 وهذا من الصعب تحقيقه إذ تختلف المؤشرات أو المفردات من حيث جودتها من مفردة إلى أخرى وكذلك يمكن عدم تحقق الاحادية البعد للمقياس خاصة إذا كان المقياس يتضمن مجموعة من المفاهيم المتميزة مثل الشخصية أو الاتجاهات وغيرها من المفاهيم بشكل تام. فمثلاً مقياس القلق فإنه يتضمن بعض المفردات التي تبدو أنها تتداخل مع الاكتئاب وهذا لا يؤدي إلى أن المفردات تستخلص نفس القدر من تباين المتغير الكامن وهذه أهم مسلمات النظرية الكلاسيكية لأن أحادية البعد للمفردات تفترض أن 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 α if item deleted والمدخل العام المستخدم لتحديد تأثير حذف المفردة من المقياس على الثبات الإجمالي لا يعكس هذا التأثير على ثبات مجتمع للمفردات وذلك للوصول إلى أفضل مجموعة من المفردات الممتثلة لبناء التحتى. فيمكن أن تزداد ألفا لعدة أسباب وليس نتيجة زيادة في الدرجة الحقيقية وهذا الحذف قائم على أساس أن كل المفردات لها نفس القدر المتساوي من تباينات الأخطاء ولكن هذا لا يمكن تحقيقه في تطبيقات ألفا ولكن أخطاء القياس الواقعة على المفردات مختلفة عن بعضها البعض، بالتالي فإن حذف المفردة ربما لأنها لا تمتلك تباين من الدرجة الحقيقية ولكنها أقل خطأ تباين.

■ عرض تقدير المعامل ألفا في قيمة تقديرية واحدة بدون فترات الثقة حول هذا التقدير فإنه من الصعب الثقة في هذا التقدير وهذا ما دعا كثير من الباحثين إلى إجراء اختبارات الفروض حول المعامل ألفا.

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 في حدود خطأ محددة مسبقاً وليكن 95% ثقة حيث يقع معامل الثبات الحقيقي بين حدى الثقة الأعلى

والأدنى، فمثلاً لو أن الثبات 0.85 له 95% حدى الثقة 0.80 و 0.90 فإن هذا يشير إلى أن احتمال 95% إلى أن قيمة الثبات الحقيقى تقع داخل فترة الثقة هذه واحتمال أن 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، فمثلاً 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 أن الاختلافات بين متوسط العينات يحدث بطريقة عشوائية لأن كل العينات اشتقت من نفس المجتمع ويكون أن يكون توزيع المتوسطات توزيعاً اعتدالياً 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 جوهري وهو بدلاً من أخذ 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 مثل SPSS ، R ، وغيرها. بل يمكن اختبارات فروض حول قيمة ألفا يحددها مسبقاً فى ضوء الدراسات السابقة.

المعامل أوميغا (ω) Omega Coefficient

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

الثبت بأنه 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 اوميجا أكثر تفضيلاً من المعامل ألفا ولكنه أقل استخداماً 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) :

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

- 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 كالاتى:

$$\begin{aligned} \omega &= \frac{(\sum_{i=1}^I a_i)^2}{\sigma_X^2} \\ &= \frac{.80 + 0.91 + 0.85 + 0.94 + 0.89 + .92 + 0.90 + 0.88 + 0.97 + 0.91}{10} = \frac{8.97}{10} \\ &= 0.897 \end{aligned}$$

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

ألفا الطبقيّة أو لمقياس مكون من أبعاد فرعية Stratified Alpha

وفى حالة وجود مقياس مكون من مقاييس فرعية (أبعاد) مثل مقاييس الذكاء أو الشخصية فيوجد مدخل معيارى لتقدير معامل ألفا للمقياس ككل واقترح Felt & Brennan (1989) الصيغة الآتية لتقدير ألفا لمقياس مكون من أبعاد 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 أربعة أبعاد. وتباينات الأبعاد الفرعية ومعامل ألفا لها كالتالى:

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

وعليه فإن معامل ثبات للمقياس ككل كالتالى:

$$\alpha_{stratified} = 1 - \frac{(2.3 \times 0.79) + (1.8 \times 0.82) + (2.5 \times 0.85) + (1.9 \times 0.74)}{35} =$$

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

ولكن إذا وجدت فروق جوهرية كبيرة بين تشبعات المفردات بالعوامل فإن المعامل ألفا الطبقي 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 ألفا

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 الالم هو كيفية الحصول على معامل الثبات على مستوى المجتمع أو بكلمات اخرى إجراء استدلال إحصائى لمعامل الثبات على مستوى المجتمع. فمثلاً 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 معامل الثبات ألفا ولنفا لم تكن تماماً القيمة 0.75 بل اختلفت عنها بالزيادة أو النقصان. ولو تم تقدير معامل الثبات من العديد من العينات العشوائية من نفس المجتمع ونفس 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 بناءً على ألفا المحورة

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

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

• $\ln$ تحويل لوغاريتمى.

وتباين هذا التقدير Z كالتالى:

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

• N حجم العينة الكلى.

ويقدر الحد الأدنى والأعلى لفترات الثقة لـ Z باستخدام تحويل 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 تحويل Z للمعامل ألفا كالتالى:

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

وعليه فإن 95% فترات الثقة، الحد الأدنى كالتالى:

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

والحد الأعلى كالتالى:

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

و 95% 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:

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

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 إعادة المعاينة بالاستبدال (توليد عينات جديدة من قاعدة البيانات الأصلية) ويستخدم فى اغراض عديدة أهمها الحصول على الخطأ المعياري وفترات الثقة لإحصائيات عديدة خاصة عندما لا تتحقق المسلمات التحتية للطرق الإحصائية وأهمها توفر الاعتدالية (Wilcox, 2012) وتوجد ثلاثة مداخل لاستخدام منهجية البوتستراپ للحصول على فترات الثقة للمعامل ألفا واحد هذه الطرق تطبق كالتالى (Finch & French, 2019):

- حساب قيمة المعامل ألفا لبيانات العينة.
- اختيار عشوائياً مجموعة من الأفراد من العينة المقدر منها معامل الثبات الفأ، مع تكرار هذه العملية للحصول على عينات جديدة ويمكن أن تصل إلى 10000

- حساب المعامل ألفا لكل عينة مولدة باستخدام هذه الاستراتيجية.
- حساب التوزيع العيني للإحصائي للمعامل ألفا المتولدة من العينات.
- حساب فترات الثقة للمعامل الفأ.

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

فى كل الأحوال فإن المعامل ألفا أكثر استخدامًا على الرغم من محدوداته ولكنه الحد الأدنى لمعامل الثبات فمع الأخذ فى الحسبان فى بناء واختيار المفردات التى يجب أن تكون أحادية البعد يكون استخدامه مفيد فى تقدير معامل الثبات للأبنية النفسية والاجتماعية ولكن يمكن أن تكون الفروق بين المعامل ألفا والمعامل أوميغا لا تؤثر فى القيمة القرارية لاستخدام أداة القياس فيمكن أن تكون ألفا 0.80 والمعامل أوميغا 0.82 فإن ثبات الاختبار جيد فى الحالتين، ولكن المعامل أوميغا أكثر دقة من المعامل ألفا (DeVellis, 2017).

• **ثبات المقدرين Inter-rater reliability:** ويطلق عليه ثبات الملاحظين inter-judge reliability أو ثبات المحكمين observer reliability يتضمن تقدير مقدار الاتساق بين باحثين أو ملاحظين أو مصححين أو مقدرين الذين يلاحظوا ويسجلوا سلوكيات الأفراد. وفى بعض المواقف البحثية فإن المقدرين يلاحظوا السلوكيات ويضعوا تقديرات أو أحكام. فمثلاً إذا طلب من باحثين أو مقدرين تقدير سلوك العدوانية للأطفال فى المدرسة، فكل محكم أو مقدر يضع تقديرات أو أحكام عن العدوانية ولكن استخدام احكام ملاحظ واحد تعتبر غير ثابتة وعلى ذلك فإن ثبات المقدرين يشير إلى أى درجة يوجد اتفاق أو اتساق بين تقديرات الملاحظين. فإذا أعطى مقدر للسلوك العدوانى 15 وأعطى الآخر 18 فإن الفرق بينهما يعكس خطأ القياس.

وتوجد طريقتين لتقدير ثبات المقدرين فإذا كان المقدرين يسجلوا ما إذا كان سلوك الأطفال (متوفرة - غير متوفرة) مثل متعاون، متنافس، عدوانى فإن المؤشر المعروف هو معامل كبا Cohen's kappa (k) ويستخدم لقياس الاتفاق بين المحكمين وتتراوح قيمته من 0 إلى 1.00. وإذا كان المقدرين يقدرون السلوك بصورة كمية على مقياس تدرجى أو تصحيح اختبارات بحيث يعطى درجة كلية فيتم تقدير معامل الارتباط بين درجات المصححين أو الملاحظين عبر كل الأفراد أو المستجيبين ويجب أن يزيد الارتباط عن 0.70 بينما يرى البعض أن يكون الارتباط المرغوب بين المصححين على الأقل 0.80.

اختيار معامل الثبات المناسب

لو صمم الاختبار لتطبيقه في أكثر من فترة زمنية لنفس الأفراد مثل تطبيقه في القياس القبلي والبعدي 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.

إذا صمم الاختبار لتطبيقه أو ادارته مرة واحدة فقط فإن تقدير معامل الاتساق الداخلي اختيار مفضل شرط وجود تجانس بين فقرات المقياس، فطريقة التجزئة النصفية تقدر خطأ التباين الناتج عن معاينة المحتوى في 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 بالتالي يتطلبوا تجانس مفردات الاختبار. فإذا كان لدينا اختبار لقياس مهارة الجمع يتضمن 20 مفردة بالتالي فإن المعامل ألفا يعطى تقدير جيد لمعامل الثبات، ولكن لدينا اختبار حساب يتضمن مهارة الجمع والطرح معاً فإن محتوى الاختبار غير متجانس بالتالي فإن ألفا و KR-20 يعطوا تقدير منخفض للثبات ولكن الاختيار المفضل في هذه الحالة التجزئة النصفية (Reynolds & Livingston, 2014) لأن الهدف من التجزئة النصفية المقارنة بين صيغتين متكافئتين. فمثلاً يقيس الاكتئاب فمن المتوقع ان يكون المحتوى متجانس بالتالي 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):

- لو ان نتائج الاختبار تستخدم لصناعة قرارات هامة مثل الشهادات أو الانتقاء أو التسكين فمن المنطقي ان يكون معامل الثبات 0.90 أو حتى 0.95.
- لاختبارات الشخصية والتحصيل يفضل ان يكون الثبات 0.80 فاعلى.
- لاختبارات التحصيل التى يعدها مدرس الفصل والاختبارات التحصيلية المستخدمة فى المسح والتشخيص يجب أن يكون معامل الثبات على الاقل 0.70.
- البعض يقترح ان معامل الثبات 0.60 يكون مقبول لأبحاث المجموعات وتقويمات الأداء والمقاييس الإسقاطية ولكن إشكالية هذه القيمة تعنى ان 40% من التباين المقاس للسمة أو المفهوم يرجع إلى الخطأ العشوائى.

العوامل المؤثرة فى معامل الثبات

يتأثر ثبات المقاييس بعدة عوامل أهمها:

- زيادة عدد المفردات (استخدام قياسات متعددة) فى المقياس، فزيادة قيمة الثبات ترتبط بزيادة عدد مفردات المقياس. ولكن هذه المسلمة ليست صحيحة فى كل الاحوال ويجب تهذيب العبارة الآتية "زيادة عدد مفردات المقياس يترتب عليها زيادة ثبات المقياس" إلى العبارة الآتية "المفردات عالية الجودة السيكمترية من يؤدى إلى ثبات عالٍ بغض النظر عن زيادة عدد المفردات.
- معامل الصعوبة والتمييز: فالمفردات السهلة جداً أو الصعبة جداً تؤدى إلى ثبات منخفض وينطبق هذا على معامل التمييز فالمفردات غير المميزة بين منخفضى ومرتفعى القدرة أو السمة تؤدى إلى ثبات منخفض.
- تباينات العينة Sampling variation: يكون الثبات منخفض كلما كان أفراد العينة أكثر تجانساً بغض النظر عن السمة المقاسة، لاحظ أن مقام طرق الثبات سواء ألفا أو كودر - ريتشاردسون 20 وغيرها يعتمد على تباين الدرجات وكلما زاد التباين زاد معامل الثبات المقدر.

مثال: مقياس مكون من 5 مفردات وكانت تباينات المفردات كالتالى (Finch & French, 2019)

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

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 وبقي 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 المجموعة 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 يؤدي إلى زيادة ثبات الأداة. وحاول بقدر الإمكان تتبه 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 وأيضاً 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 يؤدي إلى زيادة قدرة المفردة على إظهار مدى أوسع من الفروق الفردية وهذا بدوره يزيد من قيمة الثبات.

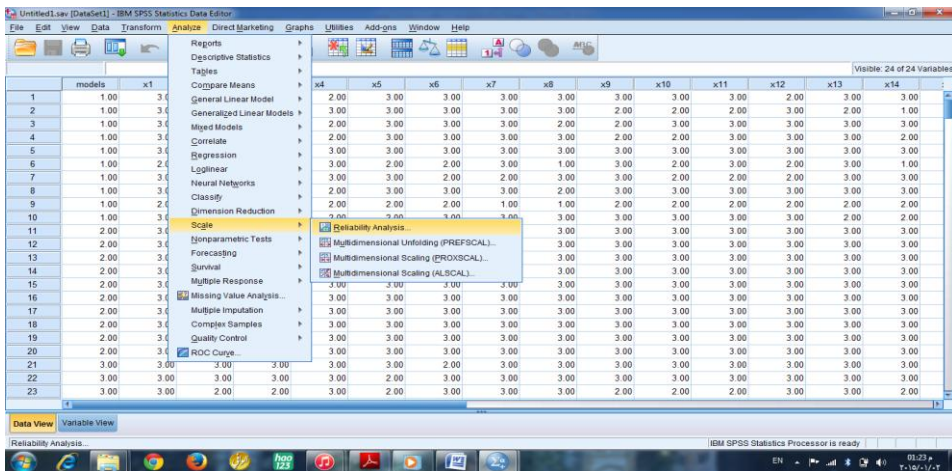
بقدر الإمكان كلما ذلك ممكن استخدم مقاييس موجودة بالفعل ونتميز بالصدق والثبات بدلاً من بناء مقاييس جديدة لأن الصدق والثبات لهذه المقاييس تم تقديره.

إجراء تحليل الثبات في SPSS:

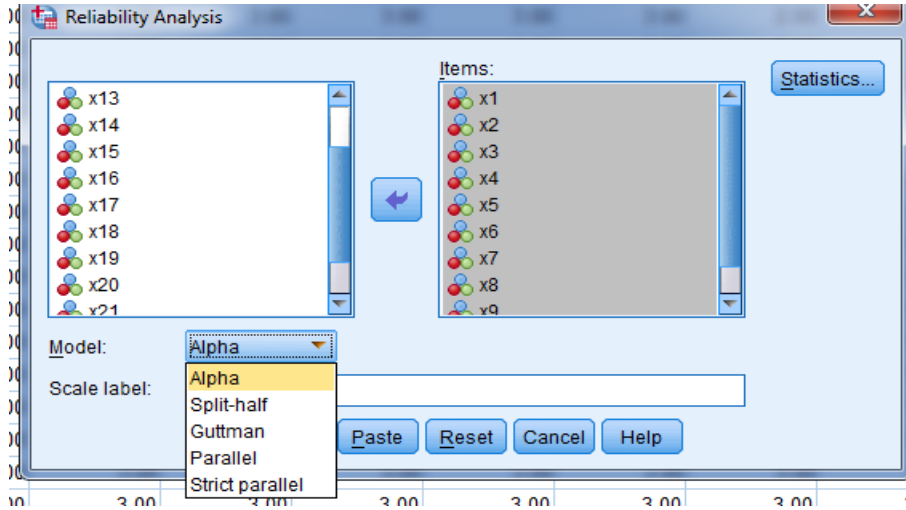
ويستخدم هذا الأمر لتحليل ثبات الاختبارات التحصيلية والنفسية والمعرفية، بمختلف أنواع الاستجابة المستخدمة فيها المتصلة Scale، أو الأسمية Nominal، أو الثنائية Binary أو الرتب في ضوء مقياس ليكرت.

ويتم تنفيذ الأمر من خلال:

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

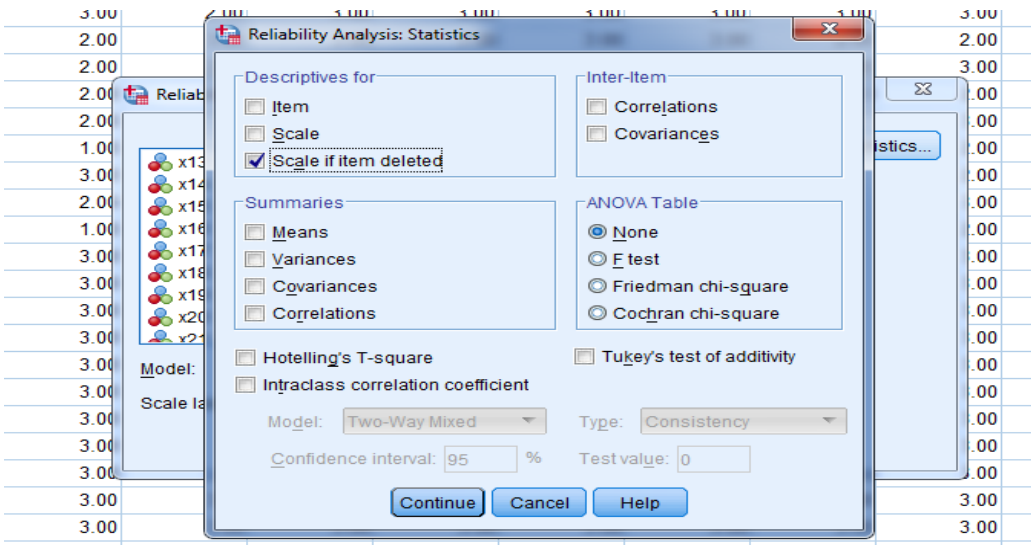


وعندئذ تظهر شاشة الآتية:



والتي تحتوى مربعين أحدهما به المتغيرات المتضمنة فى مصفوفة البيانات الخام، والأخرى مستطيل يتم نقل المتغيرات المراد تحليلها ويسمى Items، وتحتوى النافذة على مربع يسمى Model يتم اختيار طريقة تحليل الثبات اختر طريقة ألفا كرونباخ 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.

- وللتعرّف على خصائص المفردات مثل معاملات الارتباط والتغاير بين كل مفردة من مفردات الاختبار تيم اختيار correlations لتقدير معاملات الارتباط أو covariances لتقدير التغاير.

-اضغط Continue ثم Ok

-المخرج:

Reliability
[DataSet1] E:\رهنوي\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

ويوضح الجدول الأول

عدد الحالات الفعلية التي

استجابت على مفردات

المقياس ككل وهي Valid

"309" ونسببتهم "99.7%"

من إجمالي العينة الكلية،

أما الأفراد الذين لم يعطوا استجابات على مفردات المقياس Excluded وعددهم فرد واحد بنسبة "0.3%" من إجمالي العينة الكلية.

ويوضح الجدول الثانى قيمة معامل ثبات ألفا كرونباخ للمقياس ككل

Cronbach's Alpha وبلغت قيمته "0.808"، بينما بلغ عدد المفردات الكلى للمقياس

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 في حالة استبعاد كل مفردة على حدة وكلما زاد المتوسط فإن هذا يعنى أن المفردة لا يمكن استبعادها وأنها تمثل البنية الصحيحة في ضوء إطار نظري منطقي استندت وصيغت في ضوءه المفردات.
- Scale variance if item deleted تباين المقياس في حالة استبعاد كل مفردة على حدة وكلما قل التباين فإن هذا يعنى أن استبعاد المفردة يؤدي إلى انخفاض تباين الأفراد على مفردات الاختبار.
- Corrected item total correlation معامل الارتباط المصحح للمفردة على حدة ويجب استبعاد المفردات التي يقل معامل ارتباطها المصحح عن القيمة "0.3".
- Alpha if item deleted معامل الثبات للمقياس بعد استبعاد كل مفردة على حدة ويجب استبعاد المفردات التي يزيد معامل الثبات المقياس بعد حذفها عن معامل الثبات الكلي "0.818".

الفصل السادس

التوزيعات التكرارية وتمثيلها بيانياً

Frequencies distributions

فى هذا الفصل نعرض نوعين من إجراءات وصف وتلخيص البيانات وهى التوزيعات التكرارية وتمثيلها بيانياً، 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 (2013) Pagano 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 جدول منظمة لمجموعة من الأرقام المفردة. ويعرفها (2008) kirk بأنه 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 مثلاً.

ج - 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

التكرارى المجمع (الفئات) Grouped. وعليه فإن التوزيع التكرارى هو الجدولة أو التوبيب لمجموعة من الدرجات بحيث تعرض القيم وعدد مرات تكرارها فى التوزيع.

التوزيع التكرارى غير المجمع للمتغيرات الكمية

يتم فيه وضع قيم المتغير X فى عمود وتعرض فيه القيم من الاكبر إلى الاصغر 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. تحديد اكبر قيمة فى التوزيع $10=X$ واصغر قيمة $4=X$ وعلى ذلك فالعمود الأول يكون فيه قيم X مرتبة من 10 إلى 4.

2. رصد تكرار كل درجة فى العمود الثانى فمثلاً 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

والملاحظ فى التوزيع أنه يوجد درجتين لهما أكبر تكرار فى التوزيع وهما 8 ، 9 بينما الدرجة 4 حصل عليها طالب واحد فقط والدرجة 5 لم يحصل عليها أى طالب، ومجموع التكرارات هى تعبر عن عدد الأفراد أو عدد الدرجات فى التوزيع بالتالى:

$$\sum F = N$$

ومجموع درجات X يعبر عنها كالتالى:

$$\sum X = \sum XF = 10 \times 2 + 9 \times 5 + 8 \times 7 + 7 \times 3 + 6 \times 2 + 5 \times 0 + 4 \times 1 = 146$$

مثال: فيما يلى درجات عشرة 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 كالاتى:

$$P = F / N$$

حيث F تكرار الدرجة، N العدد الكلى من الأفراد.
ولدرجة 4 فإن النسبة كالاتى:

$$P = 2/10 = 0.2$$

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

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

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 أساسيين 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 بيانياً.

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 (96-41) 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 بدلاً 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 مثلاً 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 الأقل وهذه القاعدة يؤكد عليها (Gravetter & Wallnau, 2014; Pagano, 2013).

ويرى Kirk (2008) إلى أنه يجب أن تكون عدد الفئات من 10 إلى 20 فئة ما لم يوجد عدد محدود من القيم فى التوزيع فمثلاً توزيع به 25 درجة فقط فلا داعى لتمثيله بعشر فئات ويمكن الاختصار 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:

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

3. كل الفئات يكون لها نفس المسافة أو الطول.

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

5. ابدأ الجدول التكرارى بالفئة التى تتضمن الدرجات العظمى بمعنى اجعلها الفئة الأولى من أعلى وعلى الرغم أن Kirk (2008) 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 كثيرا 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 تحصيلي 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: $\text{Range} = 99 - 46 = 53$

2. حدد طول الفئة وليكن عرض الدرجات فى 10 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 التي تتضمن أقل قيمة (الفئة الدنيا) أسفل التوزيع.

لتحديد الحد الأدنى للفئة الدنيا يوجد متطلبين:

أ- يجب أن تتضمن أقل قيمة في التوزيع.

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

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

وعلى ذلك يكون جدول الفئات كالتالي:

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
			$\Sigma F = N = 70$

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

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

وهكذا لباقي الفئات. لاحظ في هذا الجدول تم عرض الحدود الحقيقية Real limits للفئة وذلك لأن المتغيرات المتصلة تمتلك عدد لانهاى 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 الحقيقية 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 ويعمل الحدود الحقيقية 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 كبيراً. فمثلاً فى الفئة 85-89 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.

الفترة الفئوية	F	التكرار النسبي	التكرار التراكمي	التراكمي 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		

فالجداول 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:

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

فمثلاً الفئة (45-49) تكرارها 1 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 = 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.

أما التكرار التراكمي المئوي لكل فئة يتم حسابه عن طريق قسمة التكرار التراكمي للفئة إلى العدد الكلي للأفراد مضروباً في 100 %

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

• $\text{cum } F$ التكرار التراكمي

وللفئة 60-64 في التكرار التراكمي المئوي كالاتي :

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

ويمكن إضافة عمود آخر يمثل التكرار النسبي المئوي. والتكرار التراكمي والتكرار المئوي مفيد في حساب المئينيات الرتب المئينية.

التمثيل البياني للبيانات

العرض البياني لبيانات كيفية (اسمية أو رتبية)

أولاً: الوصف والتمثيل البياني 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 يلي بيانات لمتغير كيفي وهو عدد الطلاب في تخصصات كلية الطب:

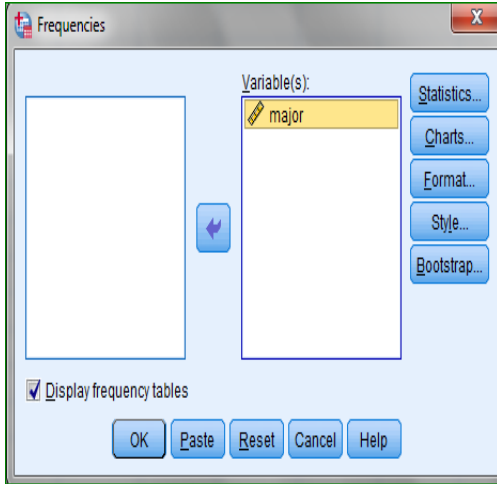
التخصص	F	Prop F	% F
باطنه	92	0.42	42
جراحة	33	0.15	15
أطفال	85	0.38	38
جلدية	11	0.05	5
المجموع	221	1.00	100

لعرض التوزيع التكرارى في SPSS:

أولاً: إدخال البيانات: 1. اضغط Variable view

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

3. اضغط على data view ثم قم بإدخال البيانات في عمود واحد.



ثانيًا: تنفيذ الأمر: 1. اضغط على Analyze ثم اختار Descriptives ثم اضغط على Frequencies يعطى الشاشة:
2. انقل المتغير major من خلال الضغط على السهم $\Rightarrow$ إلى مربع Variable، ثم اضغط على OK.

ثالثًا: المخرج: أعطى البرنامج

إحصائيات Statistics 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 Valid = 221
- بينما لا توجد بيانات غائبة Missing
- العمود الأول هو القيم (الكود) وهى أربعة تخصصات كما تم ترميزها، فمثلاً الكود (1) تخصص باطنة يوجد أمامه التكرار (العمود الثانى) Frequency = 92 فرداً
- النسبة المئوية Percent :

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

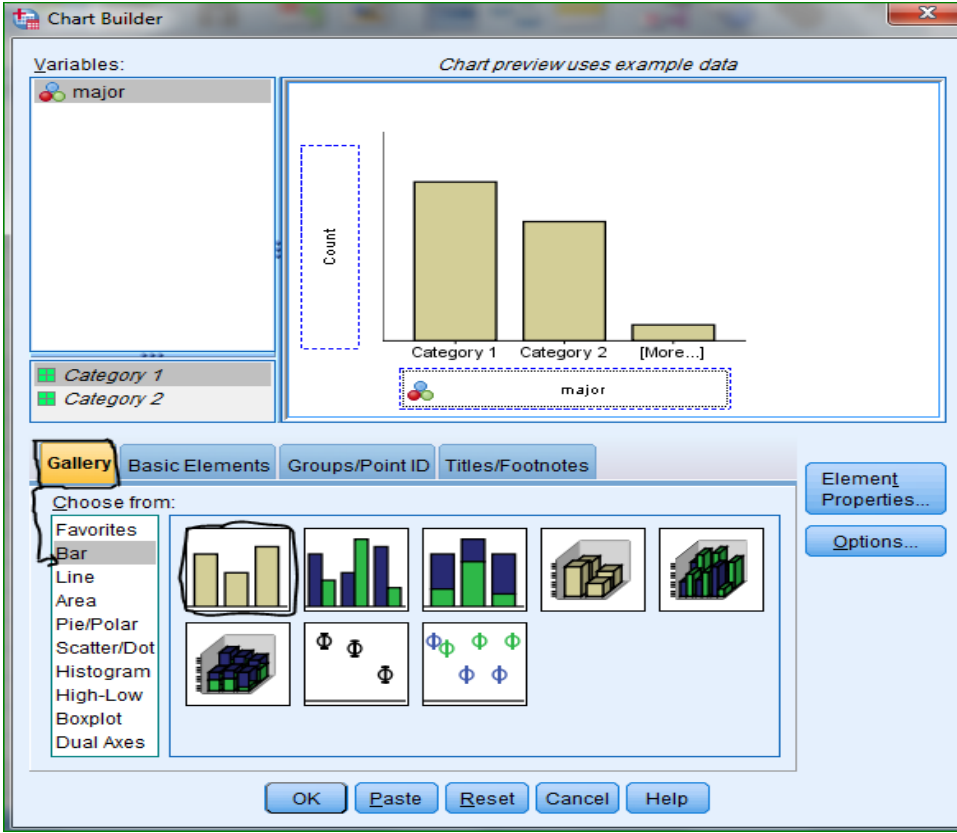
إذا 41.6% من طلبة الطب فى تخصص باطنة و 38.5 % تخصص جراحة.

- Valid percent النسبة المئوية الحقيقة أو الصادقة وقيمتها مثل 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.

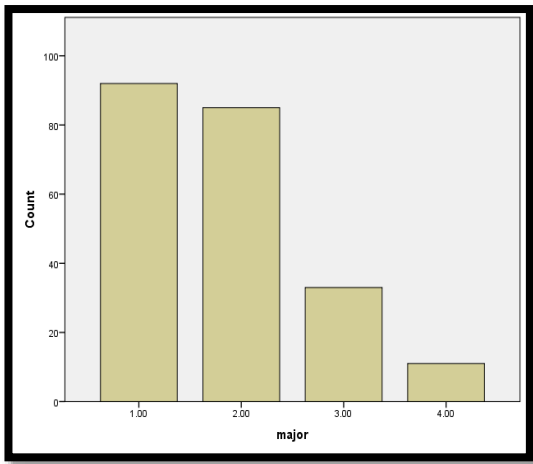
التمثيل البيانى فى ضوء Bar

1. اضغط على Graphs واختر Chart Builder 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 أسفل Choose from ثم اضغط مرتين متتاليتين 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 هي الجلدية

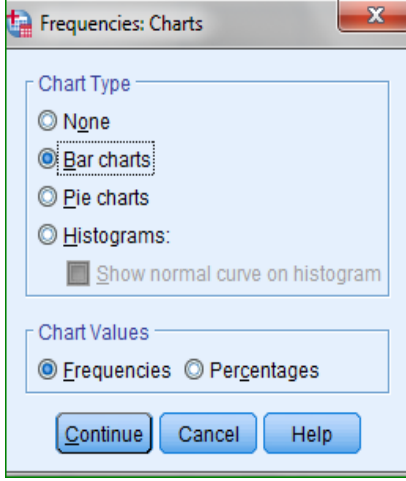
4.

ويمكن تنفيذ شكل Bar كالاتي:

1. اضغط على Analyze ثم اختر Descriptive statistics 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 Percentages, Frequencies، اضغط على

Frequencies ويمكن أن تختار النسبة المئوية.

6. اضغط على Continue لترجع إلى شاشة الأمر

الرئيسية ثم اضغط OK.

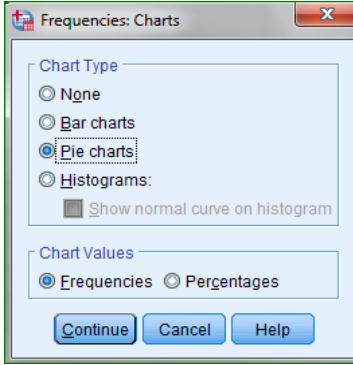
المخرج: سبق عرضه.

شكل Pie الدائرى

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

2. انقل المتغير major الى مربع Variables

3. اضغط على اختيار Charts تظهر الشاشة التالية:



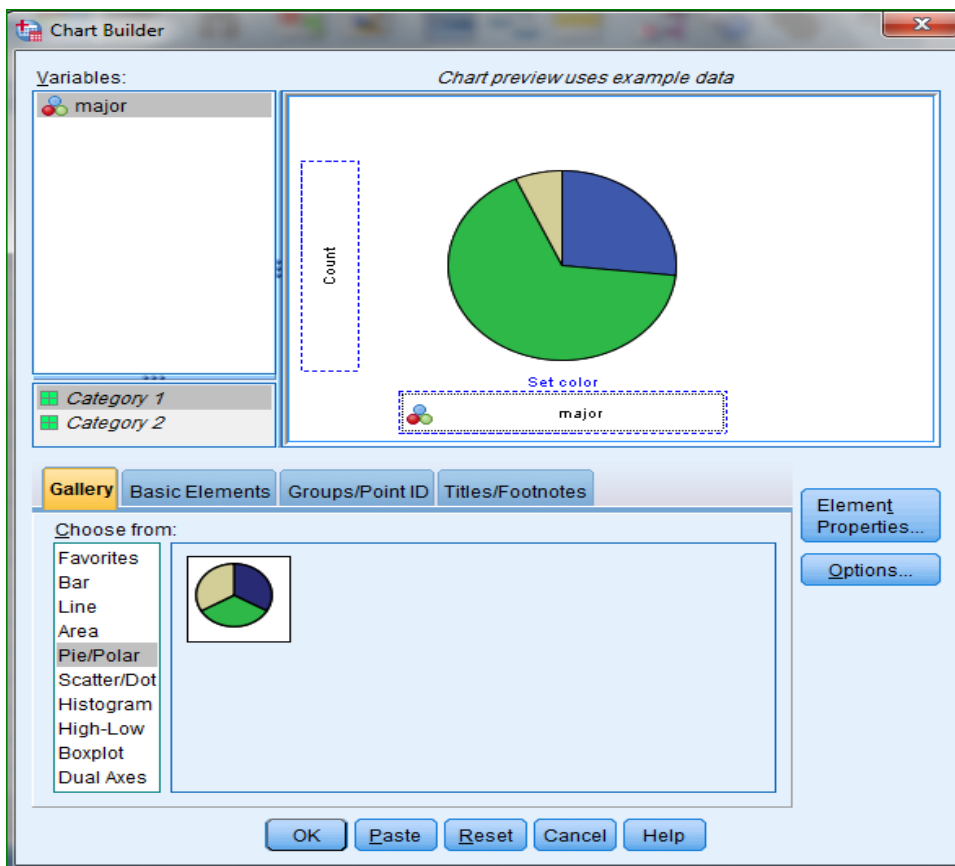
4. اختار Pie charts يمكن أن تختار عرض شكل

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 اختيارين OK، Define variable اضغط على OK يعطى الشاشة التالية:



3. في وسط الشاشة يسارًا اختار Gallery في كل الاحوال هي نشطة في البرنامج

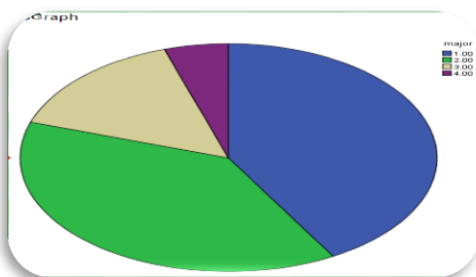
إذا لم تختارها ثم اضغط على Pie / polar

4. اضغط على الشكل Pie في يمين الشاشة تظهر في الجزء الأعلى من الشاشة.

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

6. اضغط على OK يعطى شكل Pie

كالآتي:

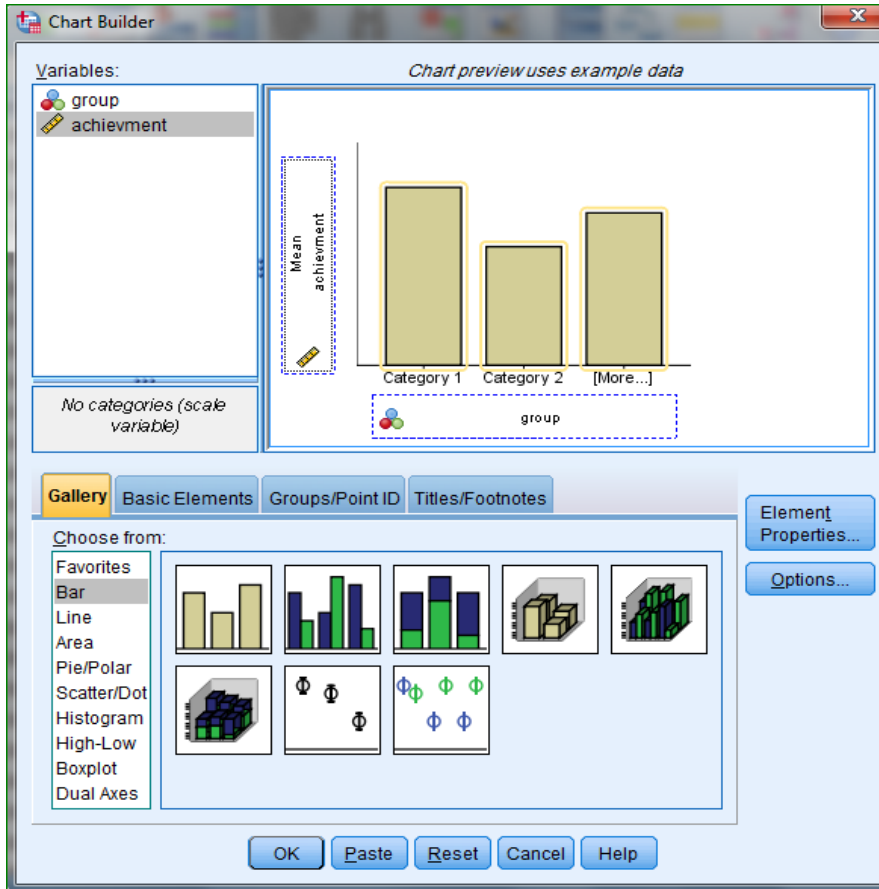


عرض شكل Bar للتعبير عن متوسطات مجموعات مستقلة

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

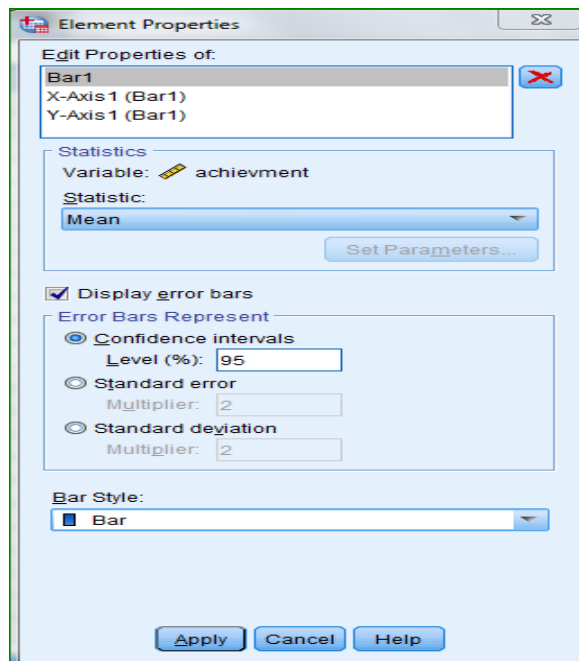
1. المجموعة group: ثلاثة مجموعات وتم توكيدهم 3,2,1 (مثل بيانات ANOVA أو حتى اختبار T المستقلة).

2. المتغير التابع achievement: التحصيل للمجموعات الثلاثة وللتعبير عن متوسطات المجموعات الثلاثة في شكل Bar اتبع الخطوات الآتية:
1. اضغط على قائمة Graphs ثم اختر Chart Builder
2. تظهر شاشة بها اختارين احدهما OK اضغط على OK
3. اضغط على Gallery وتحتها عدة اختيارات اضغط على Bar تظهر عدة أشكال أو رسومات بيانية:



4. اضغط مرتين متتاليتين على أول شكل Simple Bar يظهر في الجزء الأعلى من الشاشة رسم بياني عبارة عن إحداثيات السيني X Axis وإحداثي Y Axis فالإحداثي X يوضع عليه المتغير المستقل (group) والإحداثي Y يوضع عليه المتغير التابع Y (achievement)

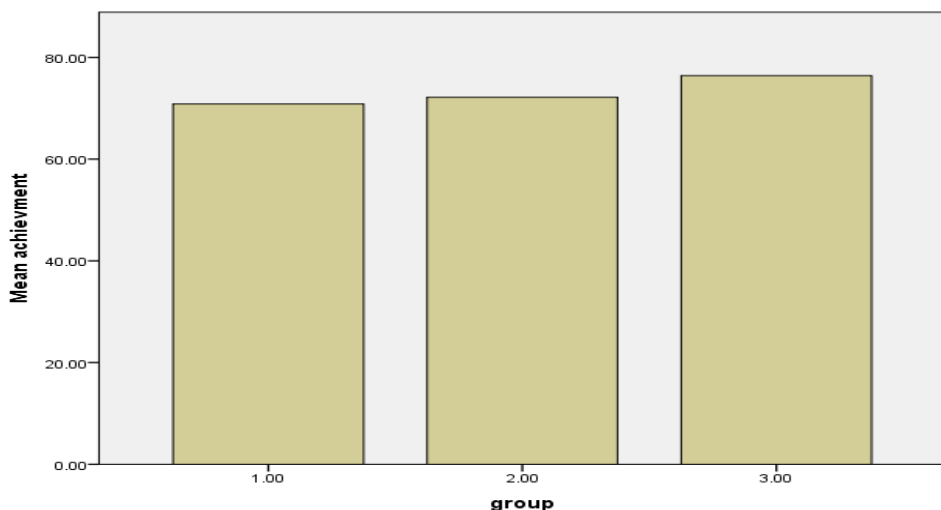
5. على يمين الشاشة يوجد اختيار Element properties اضغط عليه يعطى نافذة فرعية كالتالى:



وهذه الشاشة أو المربع الحوارى يتضمن ثلاثة ملامح رئيسية هى:
 - Statistics تحتها عدة بدائل اضغط على السهم يعطى اختيارات اضغط على Mean.
 - اضغط على Display error bars وهذا لأخطاء تظهر 95% فترة الثقة ويمكن تغييرها 90 مثلاً 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 على Apply أسفل الشاشة الفرعية.

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



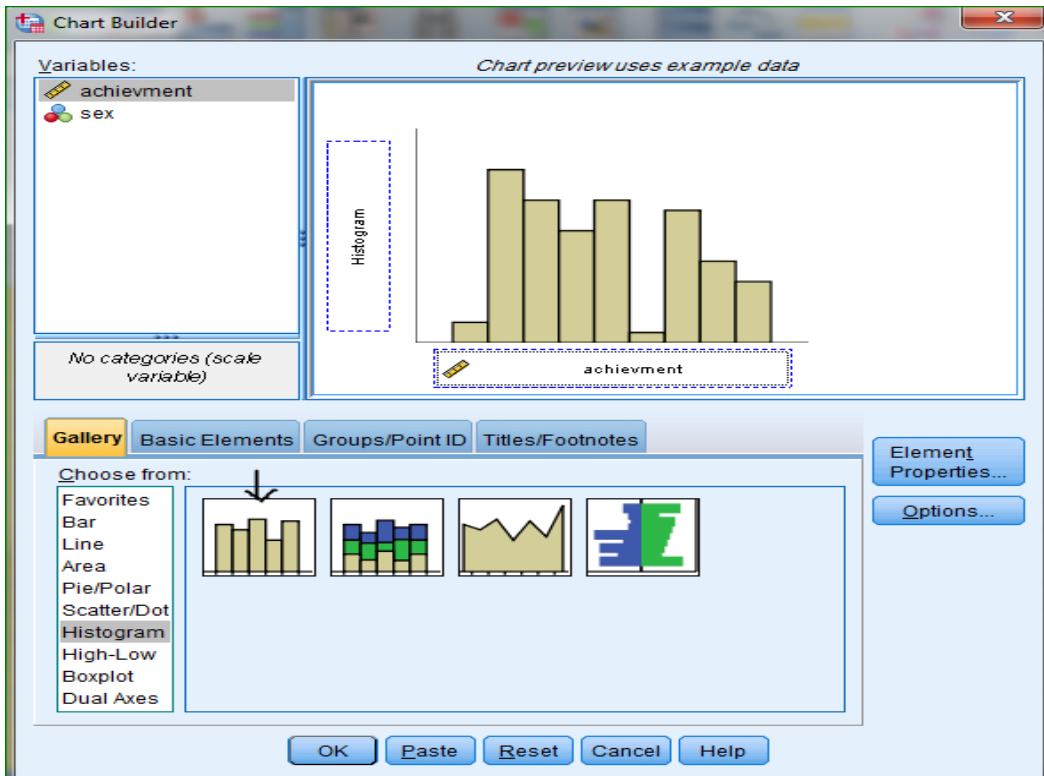
العرض البيانى للمتغيرات المتصلة (الفترة والنسبية)

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

فئة ولعرض المدرج التكرارى لمتغير واحد لبيانات التحصيل فى SPSS اتبع الخطوات الآتية(لمتغير التحصيل) :

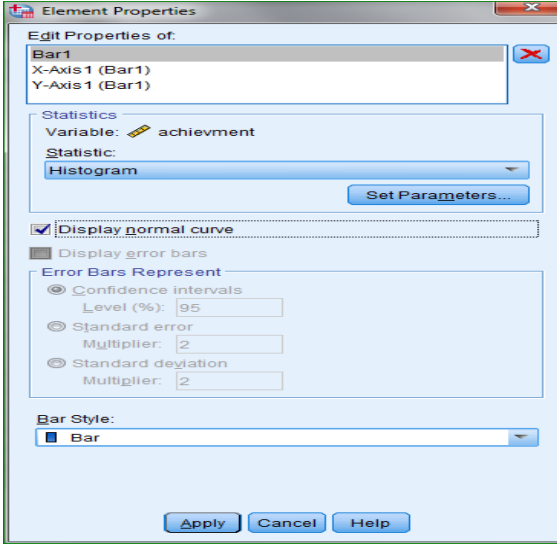
1. اضغط على قائمة Graphs ثم اختار Chart Builder تظهر رسالة فى مربع اضغط على OK

2. اضغط على Gallery ثم اضغط على Histogram يظهر لك مجموعة من الرسومات على يمين المربع اضغط على الشكل الأول وهو يعبر عن Simple Histogram تظهر الشاشة الآتية:



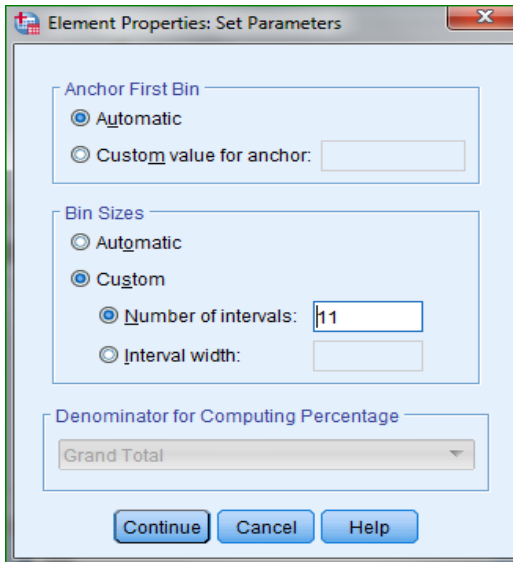
3. فى الجزء الأعلى من الشاشة Chart preview uses example data يظهر رسم المدرج التكرارى اسحب متغير Achievement فى مربع variables وضعة أسفل المدرج على محور X على يمين الشاشة.

4. اضغط على Element properties على يمين الشاشة:



5. اضغط على السهم في مربع Statistics يعطى عدة بدائل اختار Histogram وهى تعتبر default

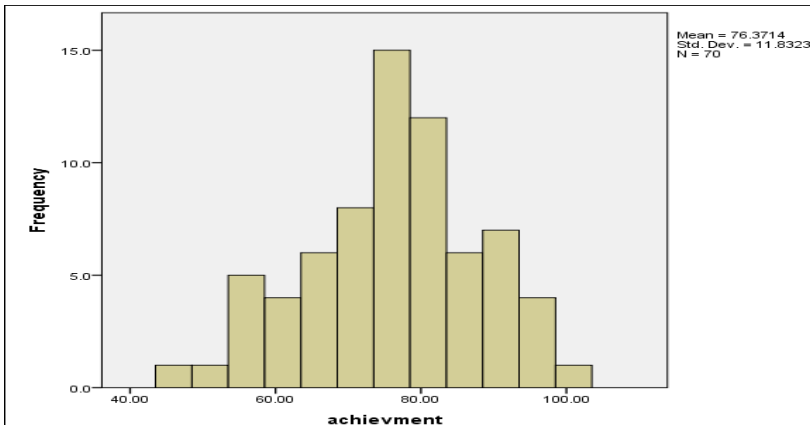
6. اضغط على Set parameters تظهر النافذة الآتية:



7. فى مربع Bin sizes حيث يمكن أن نترك للبرنامج حرية فى عمل الفئات أو يمكن تحديدها وذلك بالضغط على اختيار Custom ومنها يتم تحديد عدد الفئات مثلا 11 فئة كما فى المثال أو يمكن تحديد طول الفئة بدلا من عدد الفئات وفى هذه الحالة اضغط على Custom ثم حدد طول الفئة فى 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 الآتى:

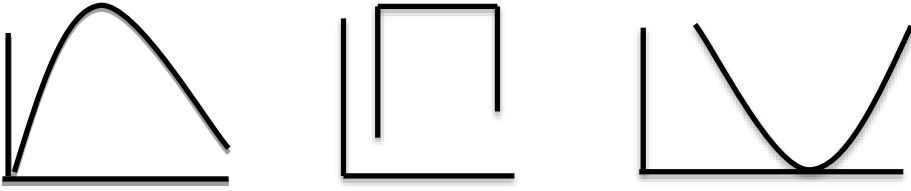


حيث إن غالبية الدرجات فى وسط المنحنى، وفى يمين المدرج عرض متوسط الدرجات
 $Mean = 76.37$ والانحراف المعياري $Std. DV = 11.832$ وعدد الطلاب $= 70$
ويمكن تنفيذ المدرج بدون تحديد عدد فئات ونترك الحرية للبرنامج لتحديد عدد فئاتها ولتنفيذها
اتبع كل الخطوات السابقة ولكن فى نافذة **Element properties** اضغط على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

أشكال المنحنيات

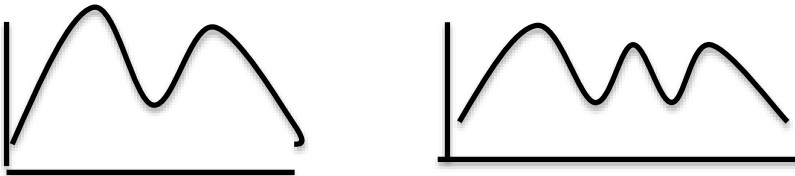
تأخذ التوزيعات التكرارية أشكال عديدة مختلفة وتصنف إلى:

1. التوزيعات المنتظمة أو المتماثلة **Symmetrical distribution**:



شكل U

مستطيل اعتدالى



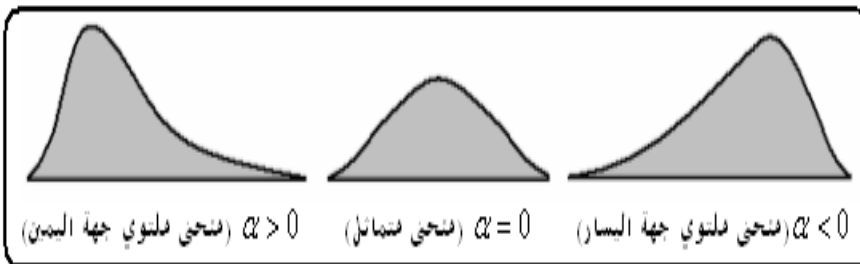
Bimode

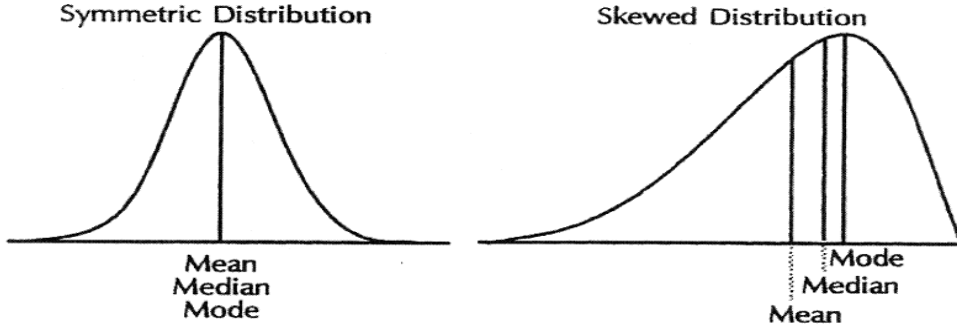
Multimodal

الشكل (1.6): امثلة لتوزيعات منتظمة.

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

2. التوزيعات الملتوية: فالتوزيع المنتظم عندما تقع الدرجات تقريباً بالتساوى على جانبي 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:

$$\text{المتوسط} < \text{الوسيط} < \text{المنوال}$$

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

$$\text{المتوسط} > \text{الوسيط} > \text{المنوال}$$

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:

$$\text{المتوسط} = \text{الوسيط} = \text{المنوال}$$

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

$$\text{المنوال} - \text{المتوسط} \cong 3(\text{الوسيط} - \text{المتوسط})$$

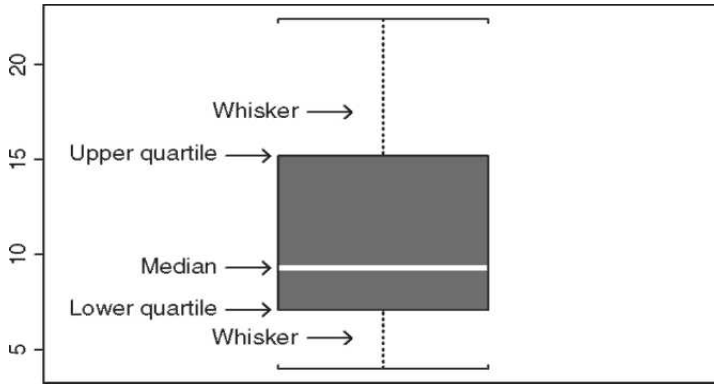
3. **التوزيع المتفرطح Kurtosis**: التوزيع المتفرطح هو يتشابه مع المنحنى الاعتدالى ولكن قمة المنحنى أكثر تفرطحًا أو تدببًا والتفرطح تأتى من الكلمة اللاتينية Kyros وهى تعنى منحنى.

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 مدببًا، وعندما يتركز عدد أكبر على طرفى

المنحنى ويقل بالقرب من المنتصف يكون المنحنى مفطحاً أو منبسّطاً.

شكل Boxplot

المدرج التكرارى لا يعطى توصيف دقيق عن مدى وجود قيم متطرفة فى التوزيع من عدمه ولذلك يتم عرض شكل Boxplot. فقام John Tuckey بتطوير عرض بياني 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:



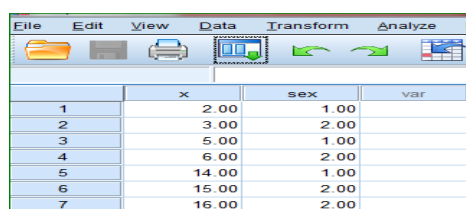
وإذا كان خط الوسيط داخل الصندوق قريباً من الإرباعي الأعلى دل ذلك على ان التوزيع ذو التواء سالب، وإذا كان قريب من الإرباعي الأدنى دل ذلك على ان التوزيع ذو التواء موجب، والخط السميك داخل الصندوق هو قيمة الوسيط أو المئينى 50% بمعنى القيمة التى يقع فوقها % 50 من القيم وتحتها 50% من القيم والخطوط الراسية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

وتعرف القيمة المتطرفة Outlier هى أكبر من 1.5 طول الصندوق بعيداً 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	الإرباعي الأولى = 6
5	14	
6	15	
7	16	
8	16	الوسيط = 16
9	16	
10	21	
11	22	
12	23	الإرباعي الأعلى = 23
13	24	
14	33	
15	45	

تنفيذ الشكل البياني Boxplot للمتغير X في SPSS



	x	sex	var
1	2.00	1.00	
2	3.00	2.00	
3	5.00	1.00	
4	6.00	2.00	
5	14.00	1.00	
6	15.00	2.00	
7	16.00	2.00	

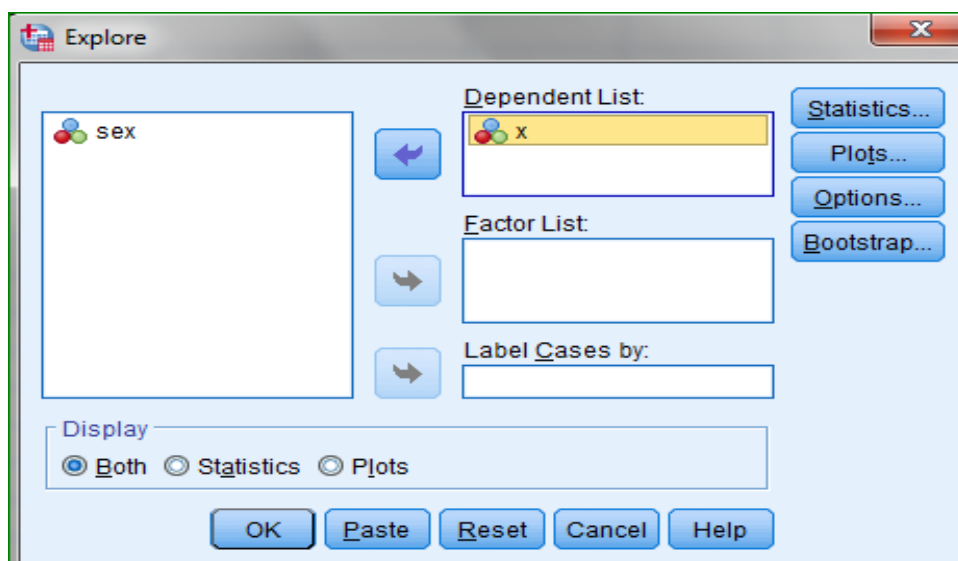
اولاً: إدخال البيانات: 1. اضغط variable view وفي

عمود name اكتب مسمى المتغير x

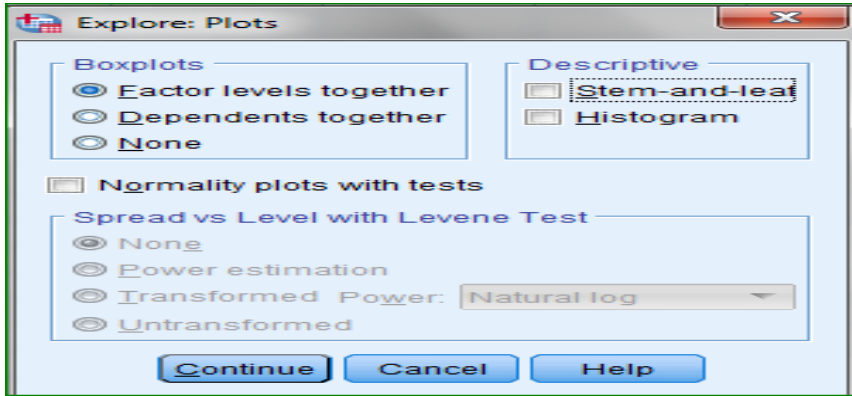
2. اضغط data view ابداء في إدخال البيانات

ثانياً: تنفيذ الأمر:

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

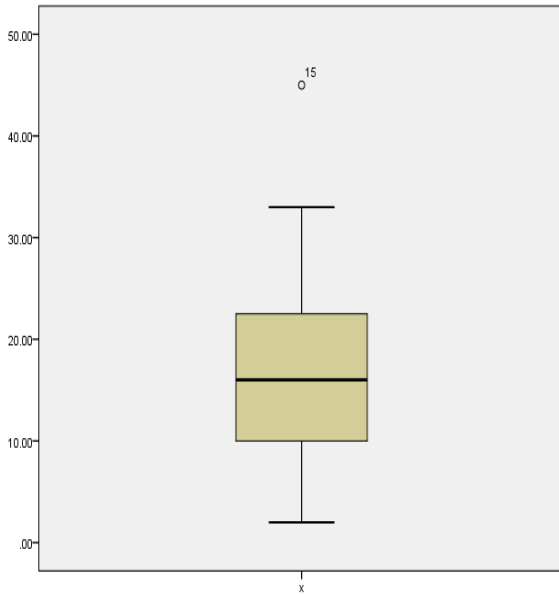


2. انقل المتغير X إلى مربع Dependentlist اضغط على اختيار Plots تظهر النافذة الآتية:



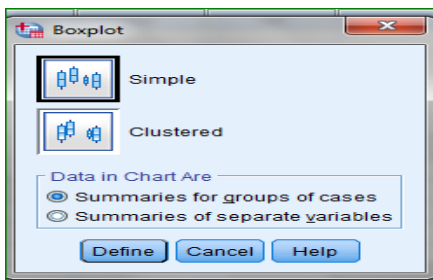
3. في مربع Boxplot اضغط على اختيار Factors levels.

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



لاحظ أن الخط في وسط المستطيل يمثل الوسيط بينما الإرباعي الأعلى يناظر 23 والإرباعي الأدنى 6 وان القيمة المناظر للخط السميك (33)T هي قيمة كبيرة في التوزيع ولكنها ليست متطرفة وأعطى البرنامج العلامة (O) تعبر عن قيمة متطرفة هي تناظر قيمة الحالة (15) وهي 45 وإذا أعطى البرنامج العلامة (*) فهي تعبر عن قيمة متطرفة جداً أي تعادل 3 مرات المدى الربيعي.

ولتنفيذ Boxplot لدرجات متغير كمي عبر مستويات متغير كيفي



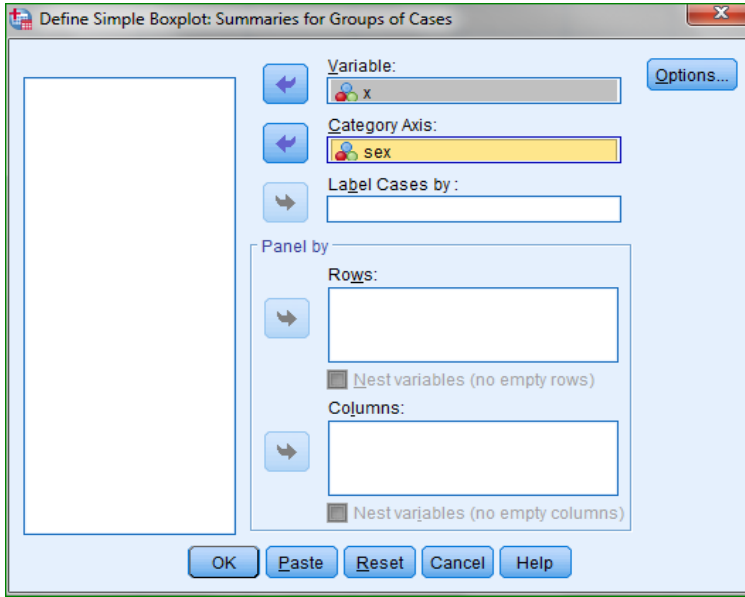
1. اضغط على قائمة Graphs ثم اضغط legacy

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:

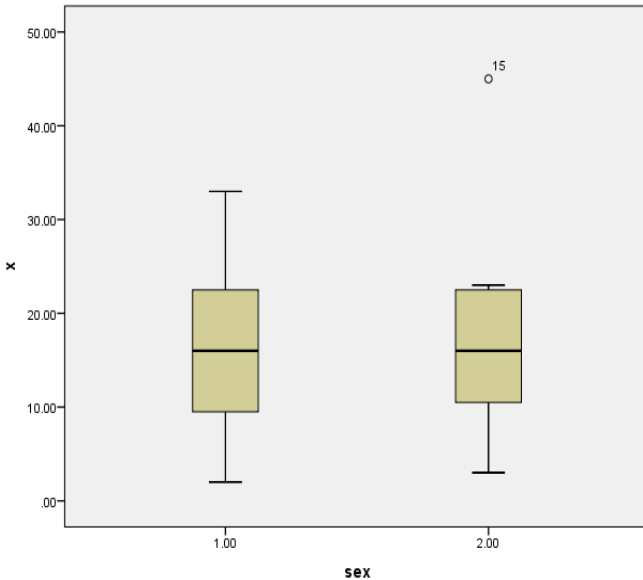


4. اضغط على متغير X وانقله إلى مربع Variable

5. اضغط على Sex ثم اضغط على وانقله إلى مربع Category Axis ثم اضغط

OK

ثالثاً: 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 بصري لتوزيع المتغير في 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

كما سبق ان المدرج التكرارى يعرض البيانات فى ضوء قيم مجمعة فى فئات، وهذا العرض يفقد كثيرًا من المعلومات الخاصة بالدرجات المفردة داخل الفئات وعلى الجانب الآخر فإن التوزيع التكرارى للدرجات المفردة (لا يستخدم الفئات) يبعث على الملل إذا كان حجم البيانات كبيرة، والمدخل البديل لتجنب عيوب المدرج التكرارى والتوزيع التكرارى هو عرض شكل الورقة أو الجذع والساق وهو مدخل عام لتحليل البيانات ويعرف (EDA) Exploratory Data Analysis وهو عرض بياني آخر لتمثيل البيانات وهذا النوع من العرض مفيد خاصة لقواعد البيانات الصغيرة لأن كل درجة يتم تدوينها فى الشكل البياني وبناء العرض البياني للساق والورقة يتطلب تجميع القيم فى Stems (تجمعات) وبعد ذلك يتم عرضها عبر كل صف ولتوضيح عرض هذا الشكل فيما يلى مجموعة من البيانات كالاتى:

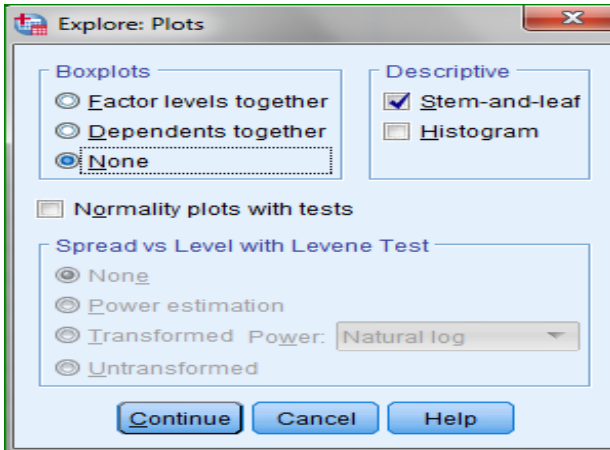
12, 33, 24, 42, 11, 14 , 23, 32, 46, 10, 16, 26, 29, 15, 30, 47, 52, 44, 19

ولتنفيذ ذلك خلال SPSS اتبع الخطوات الآتية:

أولاً: إدخال البيانات: 1. اضغط view variable وفى عمود Name اكتب اسم المتغير X.

2. اضغط data view ابداء فى إدخال البيانات.

ثانياً: تنفيذ الامر: 1. اضغط Analyze ثم Descriptive Statistics اضغط Explore تظهر الشاشة (كما فى Boxplot).



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

3. فى مربع Descriptive اضغط

على Stem and leaf

4. اضغط Continue ثم اضغط OK

ثالثاً: 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 (.) هي Stem بينما القيم بعد النقطة هي Leaf وهذا يعنى ان القيم التى تبدأ بالواحد هي 11,12,14، 10، وهكذا وتكرارها 7 قيم، بينما القيم التى تبدأ ب 2 هي 3, 4, 6, 9 وهكذا.

ولبناء شكل الورقة والساق توضع القيمة 12

مع الساق 1 وتوضع القيمة 50 مع 5، ويتم وضع اول صف بالقيم التى تبدأ ب 1 وهي (19 , 16 , 15 , 14 , 12 , 11 , 10) واربع قيم تبدأ مع 2 وهي (29 , 26 , 24 , 23) و ثلاث قيم تبدأ ب 3 وهي (33 , 32 , 30) 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 هي 10 بينما أى رقم فى leaf يمثل الحالة أو الحالات. ولكن ماذا لو كان Stem width = 1 فإن الصف الأول قيمة تكون كالاتى ، 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 وهي تكون الساق Stem أو المحور الرأسى فمثلا القيم 10 ، 11 ، 12 فإن وحدات الكسر Unitdigit هي 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 بيانياً لإعطاء صورة عن شكل توزيع الدرجات سواء كان منحى اعتدالى أو ملتوى أو متفرطح ولكن هذا ليس كافياً 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 أحدهما 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 Central tendency ويشار إليها أحياناً بمقاييس الموقع Measures of 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 أن تخبرنا بعض الشيء عن أين يقع مركز التوزيع أو تحاول أن 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 أيضاً بالمتوسط الحسابى Arithmetic average ويحسب عن طريق مجموع كل الدرجات فى التوزيع مقسوماً على عدد هذه الدرجات ويرمز بالمتوسط فى المجتمع بالرمز μ وتتطابق ميو وتسمى معلم Parameter ومتوسط العينة بالرمز $\bar{X}$ وتتطابق X

بارا أو M وتسمى إحصائية Statistic ولكن معظم المجلدات الإحصائية تستخدم $\bar{X}$ ليعبر عن متوسط العينة بينما فى الدراسات المنشورة دائماً يستخدم الرمز M ويقدر كالاتى :

$$M = \bar{X} = \frac{\sum X_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum x_n}{n}$$

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

بينما متوسط المجتمع يقدر كالاتى :

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

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

هذا المؤشر أو المقياس يحدد متوسط مجموعة درجات لعينات مختلفة من حيث العدد أو عينات غير متساوية الأفراد بكل تأكيد لنفس المتغير بالتالى متغير واحد يقاس على عينتين فأكثر غير متساوية الحجم وتكون صيغته كالاتى:

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

$$\overline{X_w} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3 + \dots + n_i \bar{x}_i}{n_1 + n_2 + n_3 + \dots + n_i}$$

$$\frac{A}{\pi} = r^2$$

• $\bar{X}$ متوسط كل عينة.

• n حجم كل عينة.

مثال: التحصيل في ثلاثة عينات مختلفة كالآتي:

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

لاحظ أن حجم العينة في كل تصنيف مختلف وإذا تم حساب المتوسط العادي للعينات الثلاثة يكون كالآتي:

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

فان المتوسط العام للمتوسطات الثلاثة معاً 59 وهذا غير صحيح لأنه لم يأخذ في الاعتبار عدد الدرجات التي أسهمت في حساب كل متوسط على حدة وعليه فإن المتوسط الموزون كالآتي:

أ - يقدر المجموع الموزون كالآتي:

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

ب- قسمة المجموع الموزون على أحجام العينات كالآتي:

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

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 الحقيقي وهذا يجعل المتوسط الموزون أكثر دقة لحساب متوسط عينات غير متساوية الأحجام وعلى ذلك فالمتوسط الموزون يستخدم في حالة وجود عينة طبقية.

مثال آخر: أجرى أستاذ امتحان في علم النفس لنفس لشعبتين وكان متوسط الشعبة الأولى 90 وعدد الطلاب 20، ومتوسط الشعبة الثانية 70 وعدد الطلاب 40 وعليه فإن المتوسط العام للشعبتين معا كالآتي:

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

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

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

خصائص المتوسط

- المتوسط حساس لكل قيمة في التوزيع: عند حساب المتوسط فإننا نضيف كل القيم في المعادلة ولذلك فأى تغير فى أى درجة يسبب تغير فى قيمة المتوسط وهذه الخاصية غير متوفرة فى حساب المنوال والوسيط.
- مجموع انحرافات القيم عن متوسطها يساوى صفرا ويرمز لها

$$\sum (X - \bar{X}) = 0 \text{ ويمكن توضيح ذلك بالمثال الآتى:}$$

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:

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

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

وهذا يرجع إلى حقيقة أن قيمة المتوسط هى نقطة التوازن Balance point فى التوزيع.

- المتوسط شديد الحساسية للقيم شديدة التطرف Exterm values أو Outlier وهذا من حقيقة هى أن المتوسط يعتبر النقطة أو القيمة المحورية أو التى يتمركز حولها القيم فإذا كان لدينا الشكل الآتى:

X	2	4	6	8	10

$$\bar{X} = 6.00$$

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

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

وعليه يحدث تغير دراماتيكي 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 في التوزيع تعطى قيمة أقل من مجموع انحرافات القيم عن أى قيمة أخرى في التوزيع $\sum (X - \bar{X})^2$ 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 $\bar{X} = 5$ كما هو واضح ان مجموع مربعات انحرافات القيم عن المتوسط $\sum (X - \bar{X})^2$ هو أقل قيمة 20 مقارنة بمجموع مربعات انحرافات القيم عن أى قيمة أخرى في التوزيع. وهذه الخاصية على درجة كبيرة من الأهمية في الإحصاء الاستدلالي بمعنى تحت كل الظروف فإن المتوسط هو أقل عرضة أو تأثراً لتباينات درجات العينة وهذه الخاصية أيضاً موجودة لمؤشر الوسيط والمنوال ولكن المتوسط هو أقل مؤشرات النزعة المركزية تأثراً بالاختلافات في درجات العينة بكل تأكيد ما عدا وجود قيم شديدة التطرف في التوزيع وهذا ما يجعل المتوسط ملك الإحصاء الاستدلالي لأنه أكثر استقراراً بالتغيرات التي تحدث في قياسات الظاهرة. ولكن تذكر أن المتوسط يستخدم للمتغيرات الكمية التي يعبر عنها رقمياً وأيضاً لحسابه تستخدم كل الدرجات في التوزيع ولذلك فهو يعطى صورة جيدة للقيم في التوزيع وأيضاً من أهم مميزات المتوسط أنه يدخل في حساب مؤشرات إحصائية أخرى مثل الانحراف المعياري والتباين ولذلك فإنه

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

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

هو متوسط يقدر من البيانات حيث يستبعد نسبة مئوية من نهايتي التوزيع، فعلى سبيل المثال إذا كان 100 ملاحظة أو درجة ونريد حساب 10% المتوسط المذهب بمعنى نستبعد 10% من درجات طرفي 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 المذهب وهو التغلب على محددات المتوسط وهو التعامل مع التوزيعات الملتوية وقد ايد ذلك بشدة (2009) Wilcox وكذلك يدعم Howell (2014) ويجب على الباحثين التعامل مع هذا المؤشر لأنه حديث في التراث البحثي.

مثال: احسب المتوسط المذهب 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 و 43 .

3. حساب المتوسط:

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

ويقدر المتوسط المذهب 20% كالتالي:

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

وبالتالى يتم استبعاد أعلى قيمتين وأدنى قيمتين وعليه يكون عدد القيم المتبقية هي 6 كالاتى:

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 المذهب 20% فإن الصيغة هي :

$$0.2n = 0.2 \times 99 = 19.8 = 1.9\% \text{ عدد القيم المحذوفة.}$$

Awinsorized Mean

عندما يتم حساب المتوسط المذهب 10% تم استبعاد 10% من القيم الدنيا والعليا من الدرجات ولكن عند حساب $X_{Awin}10\%$ يتم حذف القيمتين 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 كالاتى:

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 كالاتى:

11, 14, 17, 21, 25, 26

وإذا استُبدلت أدنى قيمتين ب 11 وأقصى قيمتين ب 26 ويصبح التوزيع كالاتى :

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

ويكون المتوسط كالاتى:

$$20\% X_w = 18.8$$

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

وهذا المتوسطات المذهبة تقترب من فلسفة Bootstrap.

الوسيط Median

هو المقياس الثانى للنزعة المركزية والهدف من الوسيط هو تحديد النقطة الوسيطة أو المنتصفة فى التوزيع Midpoint ، ويعرفه (Gravtter & Wallanu 2014) ،
بانه إذا كانت درجات التوزيع موضوعة فى ترتيب من الاصغر إلى الاكبر فهو النقطة المنتصفية فى التوزيع، أو بكلمات اخرى هو النقطة على المقياس التى تحتها 50% من درجات التوزيع وفوقها 50% من الدرجات، ويعرف أيضاً بالنقطة التى تقسم الدرجات إلى مجموعتين متساوين الحجم أو التكرار أو العدد وحساب الوسيط يعتمد على عدد الدرجات.

ويرى Pagano (2013) بان الوسيط هو المئينى خمسين P50 ويرمز له بالرمز (Md) ورمزه واحد سواء كان للعينة أو للمجتمع.

حساب الوسيط

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

1. ترتيب الدرجات تصاعدي أو تنازلي

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 تصاعدي أو تنازلي:

$$1, 2, 4, \boxed{5}, 7, 8, 9$$

2. تحديد رتبة الوسيط:

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

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

ثانياً: 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 تصاعدي أو تنازلي:

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 فمثلاً الوسيط للآتى:

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. فمثلاً 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.

- فى مواقف بحثية لا تأخذ فيها المتغيرات درجات كان يتعامل مع رتب مثلاً وفى هذه الحالة لا يمكن استخدام المتوسط ويستخدم الوسيط كبديل جيد بمعنى القياسات تكون نابعة من مستوى القياس الرتبى ويكون من الأفضل إجراء اختبارات فروض حول الوسيط وليس المتوسط.
- فى توزيعات النهايات المفتوحة Open – ended distribution مثل التوزيع الآتى:

X	F
5 فأكثر	3
4	2
3	2
2	3
1	6

وعليه فلا يستطيع الباحث حساب المتوسط ولكن يمكن حساب الوسيط لأنه لا تدخل النهاية المفتوحة 5 فأكثر فى حساباته وحيث يكون الوسيط = 1.5.

2. الوسيط أكثر عرضة للاختلافات أو تباينات العينة من المتوسط ولكنه أقل عرضة لاختلافات العينة من المنوال.

المنوال (MO) Mode

هو القيمة الأكثر حدوثاً أو تكراراً أو شيوعاً فى التوزيع وعلى العكس فإن المتوسط وأحياناً الوسيط قد يكونوا قيم ليست فى التوزيع وميزة المنوال أنه يظهر القيمة التى يحصل عليها أكبر عدد من الأفراد وهذا المؤشر يتعلق أكثر بالمتاجر أو حملات البيع للأزياء أو غيرها حيث يركز الأفراد على الزى الأكثر استخداماً ويعتبر أسهل مقاييس النزعة المركزية حساباً.

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

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

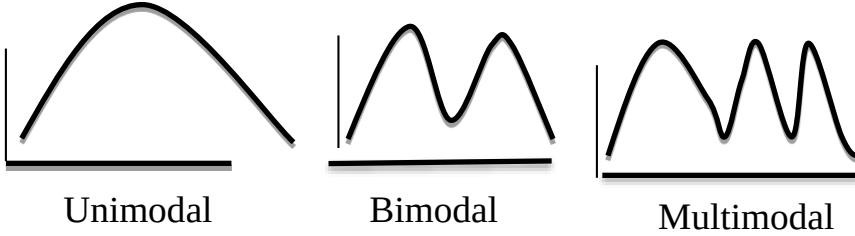
ويقال إن التوزيع فى هذه الحالة ذو منوال واحد Unimodal ويمكن للتوزيع أن يكون له أكثر من منوال مثل التوزيع الآتى:

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 محدوداً لأنه لا يتميز بالاستقرار من عينة إلى أخرى وغالباً ما يوجد أكثر من منوال للتوزيع.

وكذلك يوجد توزيعات متعددة المناويل Multimodal



مقاييس النزعة المركزية والمنحنى المنتظم

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

$$\text{المتوسط} = \text{الوسيط} = \text{المنوال}$$

وهذا يتحقق فى المنحنى الاعتدالى، إذا كانت قيم المتوسط والوسيط والمنوال مختلفة فإن التوزيع غير منتظم. وإذا كان التوزيع ملتوياً فإن المتوسط لا يساوى الوسيط لأن المتوسط يتأثر بالدرجات المتطرفة والمتوسط يكون قريباً 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 وأيضاً فترية ونسبية بعد تحويلها إلى رتب وتوزيع ذات النهايات المفتوحة - وجود قيم متطرفة فى التوزيع.

المنوال: متغيرات من مستوى قياس اسمى ومتغيرات منفصلة يستخدم 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 استخدام المتوسط فقط

لأنه يعطى صورة صادقة للتوزيع.

وفيما يلي ملخص للمواقف التي يستخدم فيها مؤشرات النزعة المركزية:

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

كتابة مقاييس النزعة المركزية في تقارير البحوث وفقا لـ APA

تستخدم مقاييس النزعة المركزية بصورة شائعة في 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 في صورة جداول وكذلك عرضها على الرسومات.

في تقرير النتائج فإن كثير من المجالات يستخدم إرشادات الجمعية النفسية الأمريكية APA كما هو معروض في 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 كالتالي:

التجريبية	الضابطة
20	15
25	17

إجراء مؤشرات النزعة المركزية في برنامج SPSS

مثال: فيما يلي الوقت الذي استغرقه 20 طالبًا لكتابة عشرة استخدامات للكمبيوتر في اختبار الابتكارية :

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

لإجراء مقاييس النزعة المركزية اتبع الآتي:

أولاً: إدخال البيانات: 1. بعد فتح البرنامج اضغط على view Variable أسفل

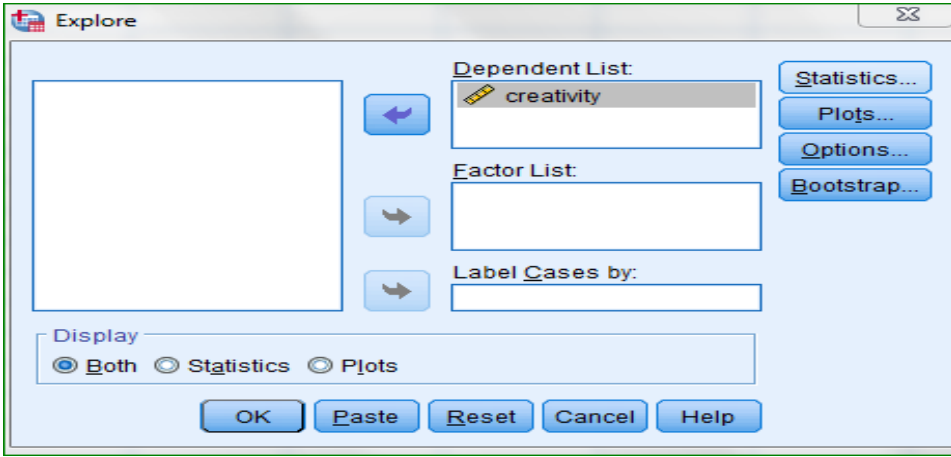
الشاشة على اليمين ثم اكتب اسم المتغير Creativity تحت عمود Name

2. اضغط على أيقونة Data view تظهر شاشة البيانات حيث يشغل متغير

Creativity عمود واحد ثم ابدأ في إدخال البيانات.

ثانياً: تنفيذ الأمر: 1- اضغط على Analyze ثم Descriptive Statistics ثم اضغط على

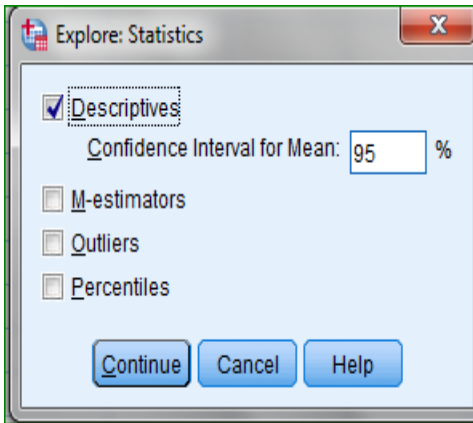
Explore تظهر الشاشة الآتية :



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

3. اضغط على اختيار Statistics تظهر

الشاشة:



4. اضغط على Descriptive

5. اضغط على Continue لتعود إلى

الشاشة الأمر الرئيسية ثم اضغط على

OK

ثالثاً: 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%

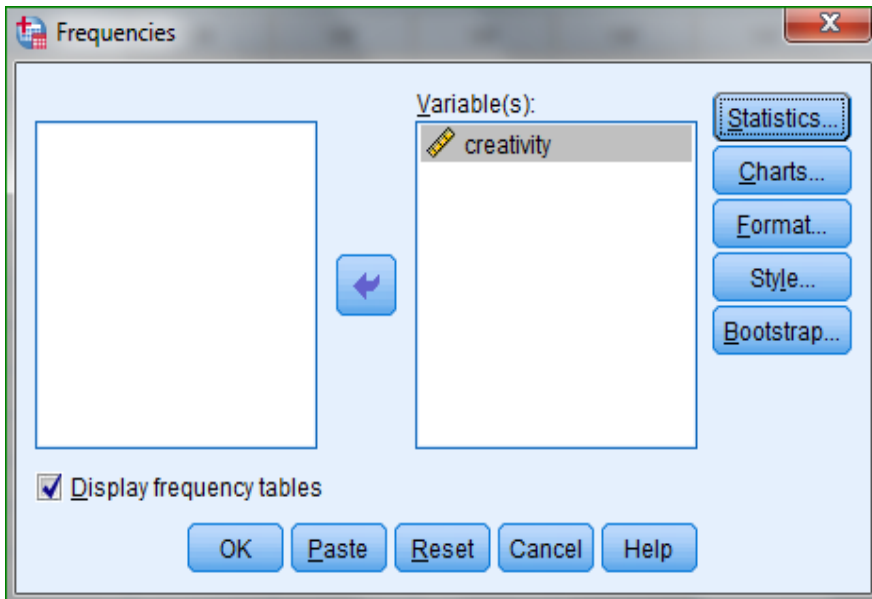
Descriptives			
		Statistic	Std. Error
creativity	Mean	80.0000	7.47346
	95% Confidence Interval for Mean	Lower Bound	64.3579
		Upper Bound	95.6421
	5% Trimmed Mean	78.9444	
	Median	80.0000	
	Variance	1117.053	
	Std. Deviation	33.42234	
	Minimum	36.00	
	Maximum	143.00	
	Range	107.00	
	Interquartile Range	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 المهدب 5% Trimmed Mean وهذا معناه أن البرنامج استبعد 5% من القيم العليا و 5% من القيم الدنيا وأصبحت قيمته 78.949 ونلجأ إلى هذا المؤشر عندما يوجد قيم متطرفة فى التوزيع، كما اعطى البرنامج مؤشر 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 7 وعندئذ يقال أن التوزيع متفرطح.

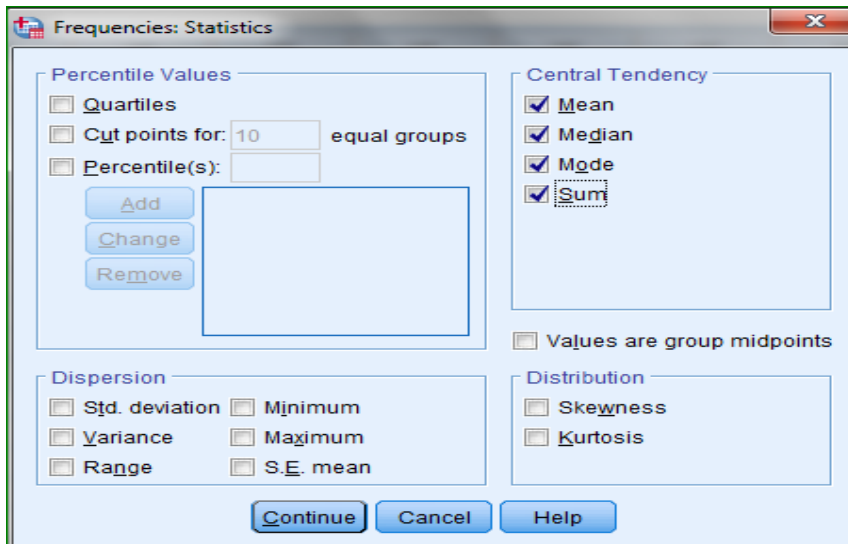
ولكن المخرج السابق لم يعطى المنوال Mode.

ويمكن تنفيذ مقاييس النزعة المركزية:

1. اضغط علي Analyze ثم Descriptive Statistics ثم اضغط على Frequencies تظهر الشاشة:



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



3. على يمين الشاشة في مربع Central Tendency اضغط علي Mean , Sum , Median , Mode اضغط على Continue ثم اضغط OK. ويمكن عدم تنشيط Display frequency table في حالة عدم الحاجة إليها وذلك في 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 : عدد القيم الفعلية فى

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	20	100.0	100.0	
	l				

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

العمود الثانى: Frequency وهو تكرار كل قيمة أو درجة فى التوزيع حيث الدرجة 36 حصل عليها طالب واحد بالتالى تكرارها 1 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 Percent Valid : عمود Percent يكون مقامه N للطلاب الحاضرين والغائبين اما فى Valid percent 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 $10 = 5.0 +$ 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

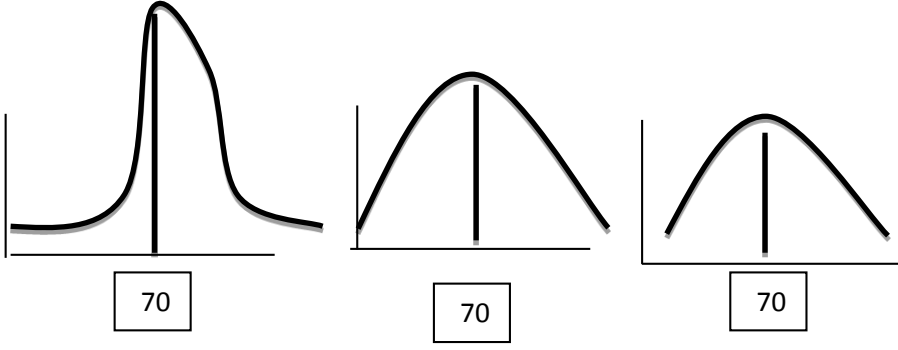
فى الفصل السابق تناولنا مقاييس عديدة مرتبطة بمركز التوزيع ومهما يكن فإن القيمة المنتصفية للتوزيع سواء المتوسط أو الوسيط أو المنوال لا تعطى صورة كاملة عن التوزيع، وعليه فنحن بحاجة إلى مقاييس إضافية لتوضح إلى أى درجة بيانات أو قياسات الأفراد تتباعد أو تحرف عن هذه القيمة المتوسطة، فالمتوسط ربما يعكس الموقع 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 مثل r , T , F .

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

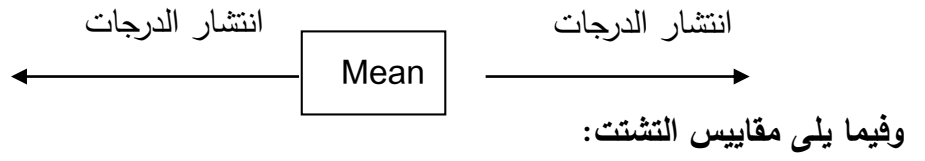
ومصطلح Variability له نفس المعنى فى الإحصاء، فنحن نطلق على مفهوم ما بالمتغير وهذا يعنى أنه له حالات مختلفة وليس حالة واحدة وهكذا بالنسبة لدرجات التوزيع فإذا كانت كلها لها نفس القيمة بالتالى لا يوجد اختلاف أو تنوع (8 ، 8 ، 8)، وإذا وجدت فروق صغيرة بين الدرجات بالتالى فالاختلاف يكون صغير (8 ، 8 ، 7 ، 9)، وإذا وجدت فروق كبيرة فإن الاختلاف يكون كبير، وعلى ذلك فإن الاختلاف يعطى مؤشر كمى للفروق بين الدرجات ويصف مدى انتشار أو تجمع الدرجات وقيمه تتراوح من 0 إلى ∞ +

وفيما يلي 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 في ضوء مسافة Distance على المحور الأفقى أو محور X



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:

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

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

ويمكن حساب المدى من الصيغة الآتية:

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

والصيغة الأولى هي التي تستخدم أثناء عرض الكتاب بينما الصيغة الثانية هي المستخدمة في برامج الكمبيوتر.

مثال: 2 , 3 , 5 , 8 , 10

$$R = 10 - 2 = 8$$

ولكن إذا اعتمدنا على صيغة المدى من الحدود الحقيقة فإن:

$$R = 10.5 - 1.5 = 9$$

مثال:

18, 12, 28, 15, 20

$$R = 28 - 18 = 16$$

ويفضل (Gravetter & Wallanw (2014) 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 السيني X أو الخط الأفقي الذي 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 تأثيره بالقيم المتطرفة.

الانحراف المعياري (SD) Standard deviation

يعتبر الانحراف المعياري أكثر مقاييس التشتت استخداماً وأهمية حيث يعتبر بجانب المتوسط الملوك العظماء فى الإحصاء الوصفى والعمود الفقرى للإحصاء الاستدلالي البارامترى، والانحراف المعياري يستخدم المتوسط للتوزيع كنقطة مرجعية يقيس الاختلاف أو التشتت اخذاً 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 المعياري لبيانات العينة بالرمز S أو SD وليانات المجتمع بالرمز σ

ويقدر الانحراف المعياري في 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 - μ	(X - μ) 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 إشارة الدرجة الانحرافية موجبة بالتالي 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 $(\frac{\sum(x - \bar{X})}{N})$ يساوي صفر.

وحيث إن:

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

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

للتباين أو الاختلاف لأنه دائما يساوى صفر وهذا نتيجة للقيم الموجبة والسالبة حيث تلغى أحدهما الأخرى ولذلك فإن الحل هو تربيع كل درجة انحرافية كما هو واضح فى العمود الثالث.

3. حساب الانحراف المعياري من الصيغة:

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

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

ومعادلة حساب الانحراف المعياري من العينة S ويستخدم لتقدير الانحراف المعياري فى المجتمع ولكن صيغة الانحراف المعياري للعينة يتم القسمة على $n - 1$ بدلا من n لاعطاء تقدير دقيق وغير متحيز لـ σ وتصبح المعادلة:

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

وهذه الصيغة المستخدمة فى إجراءات الإحصاء الاستدلالي ولكن إذا كان الانحراف المعياري لبيانات العينة للوصف فقط فالمقام يكون n لاحظ استخدمنا N للمجتمع 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 المعياري من الدرجات الخام بصيغة أخرى اعتمادًا على الدرجات الخام كالآتي:

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

• 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 يلي المثال التالي:

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

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

2. الانحراف المعياري حساس لكل درجة في التوزيع، فإذا كانت الدرجات قريبة من المتوسط فإن الانحراف المعياري يكون صغيراً والعكس إذا كانت الدرجات تبتعد بعيداً عن المتوسط فإن الانحراف المعياري يكون كبيراً. والتوزيع ذو انحراف معياري كبير يكشف عن مدى واسع من الفروق الفردية بين الأفراد.

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

4. الانحراف المعياري دائماً يأخذ قيم موجبة $S(SD) > 0$ والقيم السالبة لا معنى لها. والخصائص السابقة تجعل الانحراف المعياري مقياس جيد للتعبير عن تشتت الدرجات مقارنة بالمدى.

التباين Variance

التباين هو مربع الانحراف المعياري 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 وايضاً لاستخدامه فى الإحصاء الاستدلالي.

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 الاستدلالي.

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 المعياري:

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

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

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

الإرباعيات Quartiles

الإرباعيات Quartiles تقسم التوزيع إلى أربعة أجزاء متساوية، والمئينات Percentiles تقسم التوزيع إلى 100 جزء متساوي، يوجد أربعة إرباعيات هي الإرباع الأول Q1 وهو يقابل المئين P25 ، والإرباع الثاني Q2 وهو يقابل المئين P50 ، والإرباع الثالث Q3 وهو يقابل المئين P75 ، والإرباع الرابع Q4 وهو يقابل المئين P100 ، وعلى ذلك فإن الإرباعيات تقسم التوزيع إلى أربعة أجزاء وكل جزء يتضمن 25% من البيانات ويطلق على الإرباع الأول بالإرباع الأدنى Lower quartile والإرباع الثاني بالوسيط والإرباع الثالث يسمى

الإرباع الأعلى Upper quartile

ولتقسيم البيانات إلى أربعة أجزاء اتبع الآتي:

1. تحديد الوسيط لكل البيانات Q2.

2. تحديد الوسيط للدرجات تحت Q2 فهذا الإرباع الأدنى.

3. تحديد الوسيط للدرجات فوق Q2 فهذا يكون الإرباع الأعلى.

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

لحساب الإرباعيات اتبع الآتي:

1. ترتيب القيم تصاعدي:

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

2. حساب رتبتى الوسيط:

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

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

إذا القيمتين المقابلتين لرتبتى الوسيط هي 85 و 87 إذا:

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

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):

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

إذا كانت القيمة المقابلة للرتبة 4 هى 76 إذا قيمة الإرباعى الأدنى $Q_1 = 76$
وهى القيم التى يقع تحتها 25 % وفوقها 75 % من القيم .

4. للقيم فوق الوسيط Q_2 (86) وهى:

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

حساب رتبة الوسيط :

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

إذا القيمة التى ترتيبها 4 هى 92 إذا هى قيمة الإرباعى الأعلى وعليه $Q_3 = 92$
وهى القيمة التى يقع تحتها 75% من القيم وفوقها 25% من القيم.

نصف المدى الربيعي Q

هو نصف المسافة بين قيمة الإرباعى الأدنى Q_1 و قيمة الإرباعى الأعلى Q_3

$$Q_3 - Q_1$$

$$\frac{2}{2}$$

ويطلق عليه الانحراف الربيعى Quartile Deviation والقيمة الصغيرة تشير إلى
تشنت 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 التشنت 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.
ودائمًا يوضع نصف المدى الربيعى بجانب الوسيط لاعطاء وصف كامل عن التوزيع

$Q \pm Md$ و يعطى قيمتين أو نقطتين يقع بينهما 50% من القيم إذا كان التوزيعات منتظمة واستخدام Q و Md فى الإحصاء الاستدلالي محدود.

المئينات (P%) Percentile

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 و P_{60} هى القيمة أكثر يقع تحتها 60% من القيم. ولحساب المئينات فيما يلى المثال الآتى لدرجات 70 طالباً فى امتحان الإحصاء (Pango, 2013)

Class	F	Cum F	Cum %
95-99	4	70	$70/70 \times 100 = 100$
90-94	6	66	49.29
85-89	7	60	85.71
80-89	10	53	75.71
75-79 (74.5 – 79.5)	16	54	$61.43 \rightarrow P_{50}$
70-74	9	27	38.57
65-69 (64.5 – 69.5)	7	18	$25.71 \rightarrow P_{20}$
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:

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

- $P\%$ هي النقطة المئوية
- X_L الحد الحقيقي الأدنى للفئة التي تقع فيها رتبة المئيني
- P_R رتبة المئيني.
- i طول الفئة
- f_b التكرار التراكمي تحت الفئة التي تتضمن رتبة المئيني
- f_i التكرار المقابل للفئة التي تتضمن رتبة المئيني
- N حجم العينة

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

1. حساب رتبة المئيني:

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

إذاً 35 هي رتبة المئيني P_{50} .

2. حدد الحد الحقيقي الأدنى للفئة التي تقع فيها رتبة المئيني 50 (في ضوء عمود Cum F) إذا فإن 35 تقع في التكرار التراكمي 43 والفئة المقابلة لها 75-79 وحددها الحقيقي الأدنى (74.5)

3. طول الفئة: $i = 79.5 - 74.5 = 5$

4. f_b هي التكرار التراكمي تحت الفئة التي تتضمن رتبة المئيني = 27

5- f_i التكرار المقابل للفئة التي تتضمن رتبة المئيني هي 16

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

$$= 74.5 + 2.5 = 77$$

إذاً القيمة 77 هي المئيني 50 بمعنى يقع تحتها 50% من القيم.

ولحساب المئيني 20 كالآتي:

1. حساب رتبة المئينى 20 كالاتى:

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

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

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) وهى 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 الأول K ولكن أفضل طريقة للمقارنة بين الانحرافات المعيارية على اختبارات مختلفة لها متوسطات مختلفة هو معامل الاختلاف ويقدر كالاتى:

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

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 تقريباً 23% 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 (IQV) The Index of Qualitative Variation

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):

$$I_{qv} = \frac{K(100^2 - \sum P^2)}{100^2(K - 1)}$$

- K عدد التصنيفات والمستويات للمتغير
- $\sum P^2$ 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	انجليزي	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
				$\sum P^2 = 5018.3$

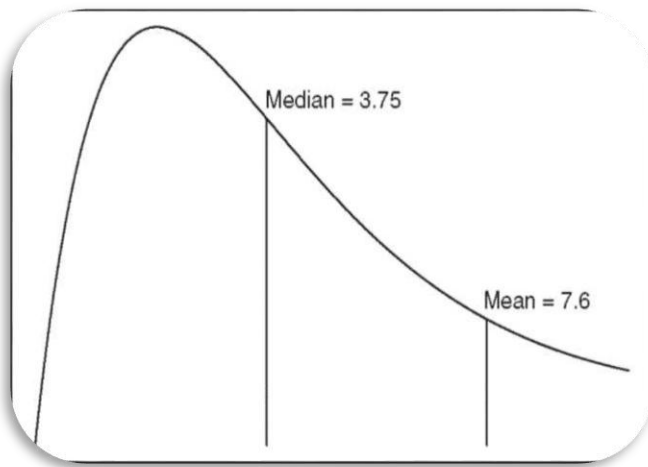
وعليه:

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

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 تقريباً 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 كالاتي:



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 المعياري ولو كان التوزيع منتظم Symmetrical 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
$\sum X = 36$	$\sum(X - \bar{X}) = 0$	$\sum(X - \bar{X})^2 = 22$	$\sum(X - \bar{X})^3 = 24$	$\sum(X - \bar{X})^4 = 18$

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

إذا الانحراف المعياري:

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

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

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

يقدر التقترح من الصيغة الآتية:

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

ويقدر للبيانات السابقة كالاتى :

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

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

أولاً: إدخال البيانات بعد فتح الشاشة الرئيسية

1. اضغط على Variable view أسفل الشاشة على اليمين

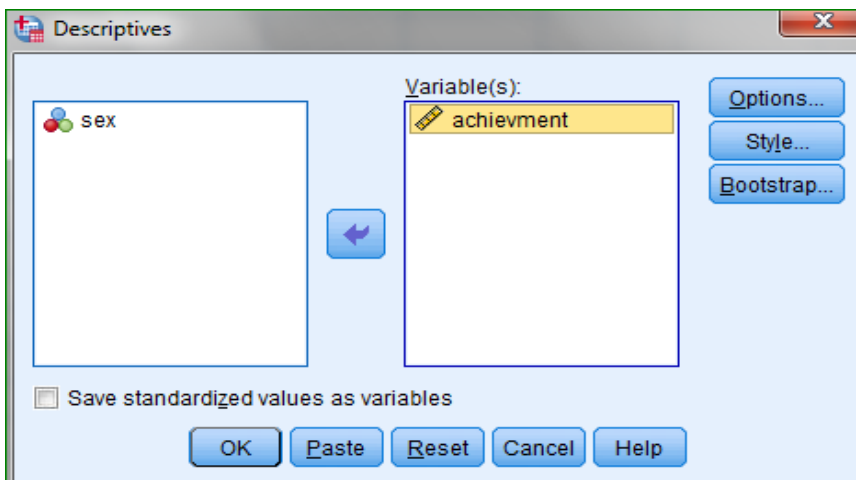
2. اكتب اسم المتغير فى عمود Name : achievement

3. اضغط على أيقونة Data view، قم بإدخال البيانات فى عمود واحد

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

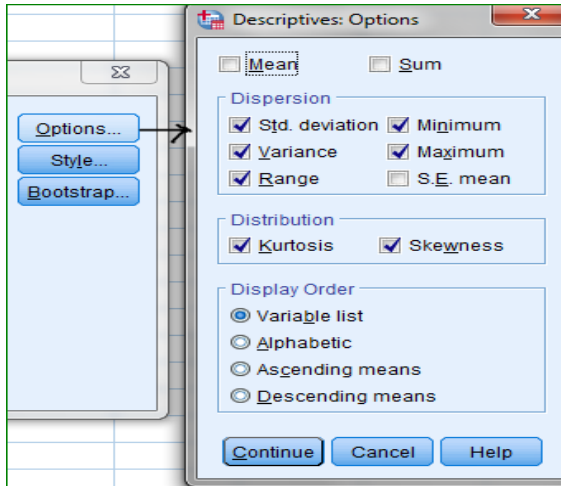
1. اضغط على Analyze ثم Descriptive Statistics ثم اضغط على

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 مؤشر التفرطح

والالتواء فى مربع Distribution

6. اضغط على Continue ثم OK

ثالثاً: المخرج:

```
DESCRIPTIVES VARIABLES=achievement
  /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	Statistic
achievement	70	53.00	46.00	99.00	11.83237	140.005	-.300	.289
Valid N (listwise)	70							

كما هو واضح أعطى الإحصائيات:

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

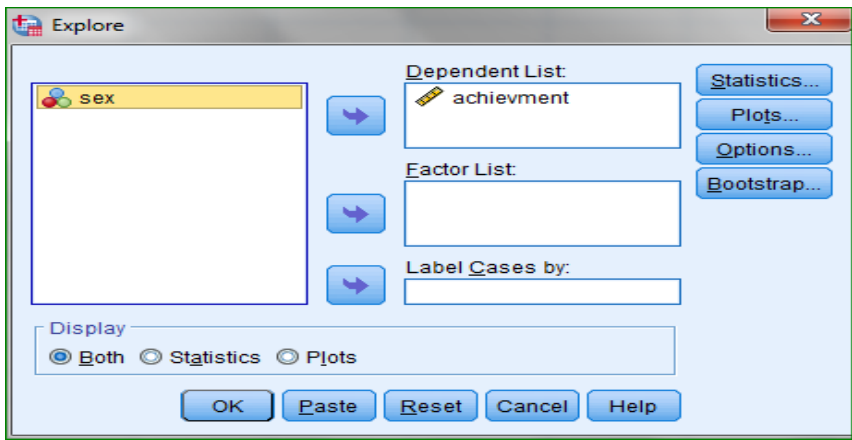
- الانحراف المعياري والتباين

- وقيمة $Skewness = -0.300$, أى يوجد درجة ضعيفة من

الالتواء السالب ولكن فى المجل المنحنى اعتدالى إلى حد 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. اضغط على Analyze ثم Statistics Descriptive ثم اضغط على Explore يعطى الشاشة الآتية:



2. انقل متغير achievement إلى مربع
Dependent list

3. اضغط على اختيار Statistics تظهر الشاشة
الآتية:

4. اضغط على اختيار Descriptive و
Quitlers و Percentiles

5. اضغط على Continue ثم OK
ثالثا: المخرج:

```
EXAMINE VARIABLES=achievement
/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
achievement	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 المهذب Trimmed Mean مع استبعاد 5% من الدرجات الدنيا وكذلك القصوى = 76.62 . ونلاحظ أنه لم يتغير كثيرًا، والوسيط Median = 77 لأنه لا يوجد قيم متطرفة.

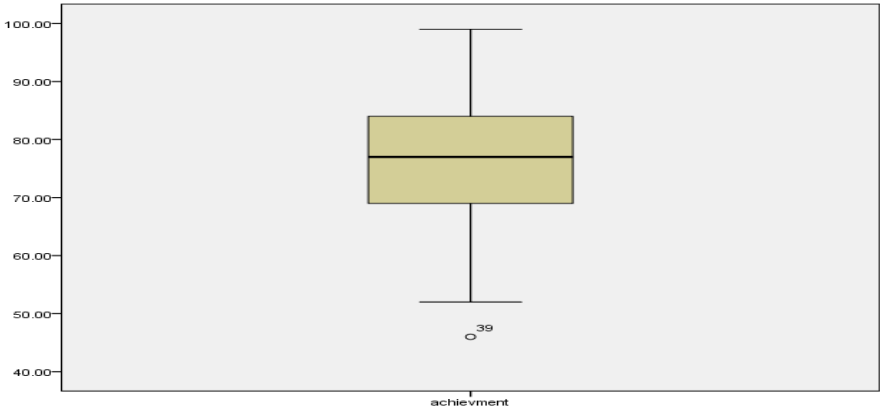
والتباين Variance = 140.005، الانحراف المعياري Std. Deviation = 11.83، المدى Range = 53، نصف المدى الربيعي Interquartile Range = 15.75. لاحظ أن قيم الانحراف المعياري ونصف المدى الربيعي متقاربة إلى حد ما ولكن المدى قيمته كبيرة جدًا 53 لأنه يعتمد على القيم المتطرفة، والالتواء = -0.300، التفطح = -0.284،

- أعطى البرنامج المئينات كالآتي:

Percentiles		Percentiles						
		5	10	25	50	75	90	95
Weighted Average	achievement	55.1000	58.2000	68.7500	77.0000	84.5000	93.0000	95.4500
(Definition 1)								
Tukey's Hinges	achievement			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 الثانى (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 هي قيمة متطرفة كما هو واضح من العلامة (0).

The normal Curve and المنحنى الاعتدالى والدرجات المعيارية Standardized Scores

افتراض 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 ضوءها 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 يتضمن 1000 درجة و توجد به الدرجة 132، فيمكن تحديد كم نسبة مئوية من الدرجات (%) أقل أو أكبر من الدرجة 132 كما سبق توضيحه من الرتبة الميئية 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 والميئيات.

واحد المشاكل لعمل مقارنات لها معنى هو قياس متغيرات مختلفة على وحدات قياس مختلفة، فعلى سبيل المثال يقاس الطول بالسنتيمتر 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 في ضوء المتوسط والانحراف المعياري.

وإذا وقعت الدرجة المعيارية فوق المتوسط فإنها تكون موجبة، وإذا وقعت تحت المتوسط تكون سالبة. وتقدر الدرجة المعيارية في 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 المعياري
4. حساب الدرجة المعيارية لكل درجة من المعادلة السابقة
5. التعبير عن المسافة في ضوء الانحراف المعياري 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 المعياري:

X	X ²
1	1
4	16
5	25
7	49
8	64
$\sum X = 25$	$\sum X^2 = 155$

$$SS = 155 - \frac{(25)^2}{5} = 30 \quad \text{وبما ان:}$$

$$S = \sqrt{\frac{30}{4}} = 2.7386 \quad \text{إذا:}$$

وعليه:

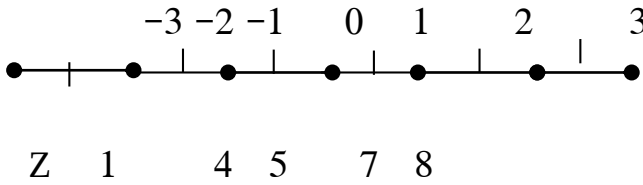
X	Z
1	

$$Z = \frac{X - \bar{X}}{\sigma} = \frac{1 - 5}{2.7366} = -1.46$$

4	-0.37
5	0.00
7	0.73
8	1.10

لا تتسبب إشارة Z حيث تغييرها من السالب إلى الموجب يحدث فروق كثيرة.

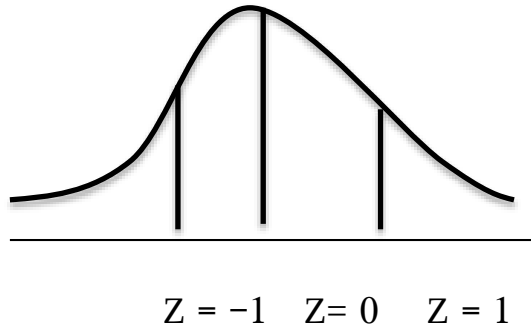
الدرجة Z كمقياس يمكن تمثيل الدرجة المعيارية على مقياس كالتالي:



وعلى ذلك الدرجة 8 تبعد 1.10 وحدة معيارية أعلى المتوسط بينما الدرجة 4 تبعد 0.37 وحدة معيارية تحت المتوسط.

ويوجد ملحقين أساسيين للدرجة المعيارية Z:

الأول: أن توزيع Z دائماً متوسطة - صفر وإذا كانت الدرجة في التوزيع هي المتوسط فإنها تتناظر الدرجة المعيارية صفر بمعنى ابتعادها هو صفر وحدة انحراف معيارى.
الثاني: توزيع Z دائماً له انحراف معيارى مساوياً للواحد الصحيح.



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 تبعد 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 $S = 5$ والمتوسط $\bar{X} = 8$ فإن:

$$X = 1 \times 5 + 8 = 13$$

وفى ضوء ما سبق فإن:

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 دائماً $= 1$
 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 إلى Z باستخدام المتوسط والانحراف المعيارى للتفاح ثم تحول Z إلى المئىنى لتقول أن 85 % من كل التفاح المتبقى يقع فوق هذه التفاحة.

استخدام Z لعمل مقارنات

تخيل أنك 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 في مقررات مختلفة.

للدرجة Z فوائد أهمها:

1. الدرجة Z تعطى إحساس أو فهم عام لموقع الدرجة في علاقتها بمتوسط الدرجات و الانحراف المعياري في التوزيع.
 2. تسمح الدرجة المعيارية Z بمقارنة الدرجات من توزيعات مختلفة.
 3. يمكن تحويل الدرجات Z إلى المئينات Percentiles .
- ولا نشكل توزيع Z اعتدالي فلا بد أن نتعرض للتوزيع الاعتدالي.
- تنفيذ إحصاء Z في برنامج SPSS (للمثال السابق)**

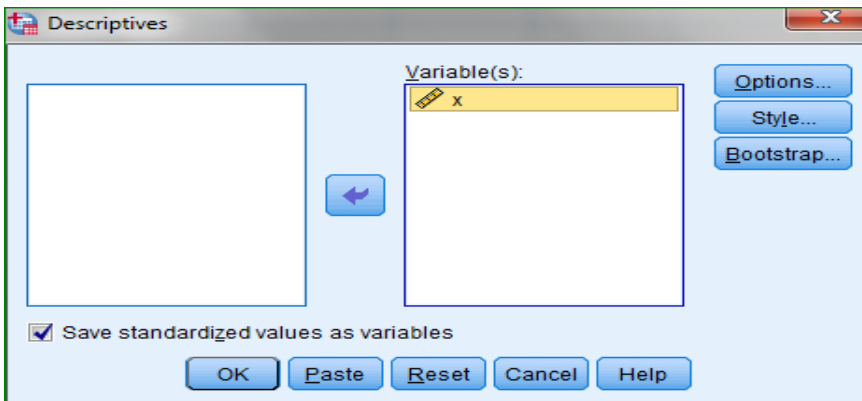
أولاً: إدخال البيانات: 1. اضغط علي Variable view في أسفل الشاشة على.

2. اكتب مسمى المتغير x في عمود Name

3. اضغط علي Data view

ثانياً: تنفيذ الامر: 1. اضغط على analyze ثم اختر Descriptive statistics ثم اضغط

على Descriptive تظهر الشاشة الآتية:



2. انقل المتغير X إلى مربع Variables (اضغط على السهم $\rightarrow$)

3. اضغط على اختيار Save Standardized values as variables

4. اضغط على OK

ثالثاً: تفسير المخرج: يعطى الدرجات المعيارية ملف البيانات حيث يقوم البرنامج بإنشاء عمود تحت مسمى ZX لكل درجة خام وبالرجوع إلى ملف البيانات من خلال الضغط على قائمة Window ثم اضغط على الملف المتضمن البيانات .

File	Edit	View	Data	Transform	Analyze
5 : x		8.00			
	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					

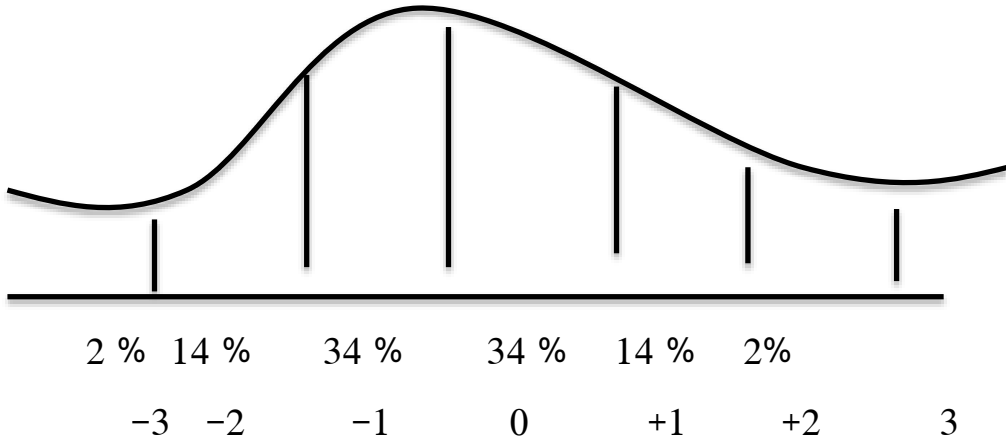
التوزيع الاعتدالي

هو توزيع نظري لدرجات للمجتمع النسبة ويشبه الجرس ومنتظم Symmetric وأحادي المنوال Unimodal ولذلك فمن السهل معرفة بعض الشيء من النسبة المئوية لأي منطقة تحت المنحنى.

ولإنشاء المنحنى عليك تذكر ما عرضناه في التوزيعات التكرارية وتمثيلها بالمدرج التكراري ولذلك فنحن نستطيع تحديد النسبة المئوية للمنحنى الاعتدالي التي تقع تحت أو فوق أي خط رأسى في هذا المنحنى. والإحصائيون حددوا نسب معينة تقع داخل أي منطقة محددة في المنحنى الاعتدالي وعليك ان تضع في اعتبارك ان المنحنى الاعتدالي هو منتظم بمعنى 50% من القيم تقع تحت المتوسط و 50 % من القيم تقع فوق المتوسط واحد جانبي المنحنى هو مرآة تماماً للأخرى للجانب الآخر.

والشكل الآتي يوضح المنحنى الاعتدالي بالنسب التي تقع داخله (Aron et al.,

) 2013

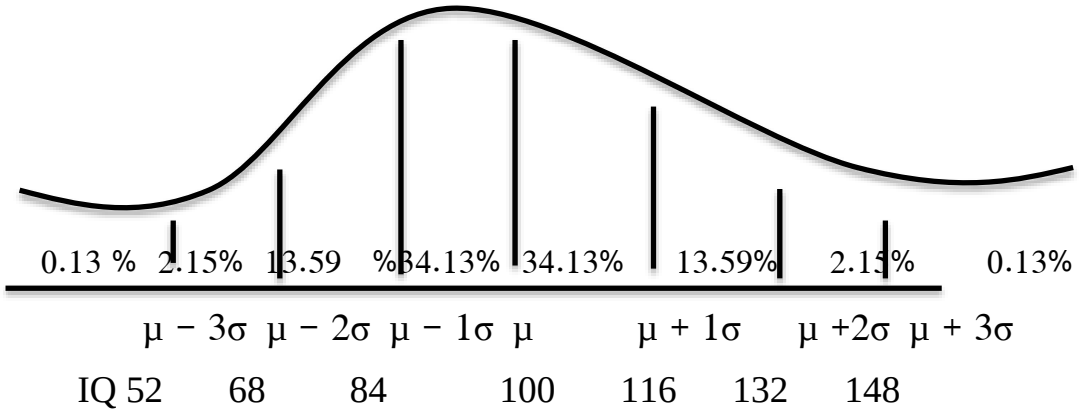


الشكل (1.7): المنحنى الاعتدالى مع النسب المئوية التقريبية للدرجات بين المتوسط ووحدة الانحراف المعيارى تحت وفوق المتوسط

ففى الشكل السابق تم تقسيم المنحنى الاعتدالى إلى نصفين متساوين تقريبا 34% من الدرجات تقع بين المتوسط (0) والدرجة المعيارية (-1) و 14% 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 (34 + 34 + 14 + 14) وتقريبا 100% من الدرجات (34 + 34 + 14 + 14 + 2 + 2) تقع فى نطاق الدرجة المعيارية -3 إلى +3 وهذه تعتبر إرشادات لتحديد النسبة المئوية المرتبطة ب 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 و 115 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

$$\mu + 2\sigma = 100 + 32 = 132$$

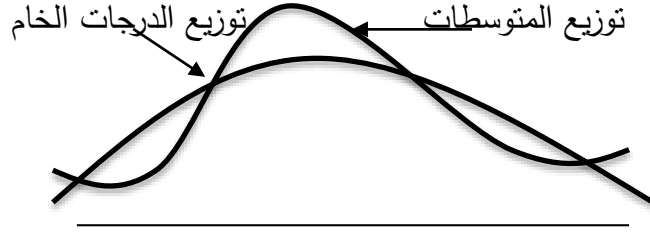
و 2.26 % من الدرجات تقع بين الدرجات 132 و 148

$$\mu + 3\sigma = 100 + 48 = 148$$

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 بيانياً للحصول على المنحنى الذى يمثلها ويعتبر توزيع المتوسطات أكثر ثباتاً واستقراراً من الدرجات حيث له انحراف معيارى أقل من الانحراف المعيارى للدرجات الخام.



ولاحظ أن توزيع المتوسطات ضيقة وليس لها انتشار عالٍ كما في حالة توزيع الدرجات الخام وأن الخطأ المعياري هو مسمى الانحراف المعياري لتوزيعات المتوسطات وهو الانحراف المعياري للمجتمع مقسوماً على الجذر التربيعي لحجم العينة:

$$\sigma_{\mu(\bar{X})} = \frac{\sigma}{\sqrt{N}}$$

وإذا كان الانحراف المعياري لتوزيع مجموعة من الدرجات 5 لعينة 10 أفراد فإن الخطأ المعياري: $= \frac{10}{\sqrt{5}} = 1.58$

ولاحظ الفرق أن الانحراف المعياري للدرجات = 5 بينما للمتوسطات 1.58 ولذلك فإن تشتت المتوسطات أصغر من تشتت الدرجات الخام ولذلك يتم تقدير Z للمتوسطات كالتالي:

$$Z = \frac{X - \bar{X}}{\sigma_{\bar{X}}}$$

حيث:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

وهذه هي الصيغة التي تستخدم في الفصول السابقة عند تناول اختبارات Z، T.

الفصل التاسع

اختبارات الفروض ومداخل صناعة القرار الإحصائي

Hypothesis testing & Statistical Decision Making Approaches

تسمح الإحصاء الاستدلالي للباحثين بجمع البيانات من العينة للوصول إلى استنتاجات أو نتائج أو معلومات عن السلوك في المجتمع الذي بدوره يكون كبيراً بدرجة لا نستطيع دراسة كل أفراده، فمثلاً متوسط العينة يكون تقدير غير متحيز للمجتمع إذا تم اختيار العينة بصورة عشوائية من المجتمع ويقال عندئذ أن قيمة المتوسط في العينة $\bar{X}$ تعكس تقريباً قيمة المتوسط في المجتمع (μ)، ولذلك عندما يستخدم الباحث الإحصاء الاستدلالي عليه أن الاجابة عن أربعة أسئلة أساسية هي:

- ما هو المجتمع المراد دراسته؟
- كيف تم اشتقاق أو سحب العينة من المجتمع؟
- ما هي خصائص العينة التي 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 المنطقي أن يقوم باحث بملاحظة كل فرد في المجتمع ولذلك يقوم بجمع بيانات العينة واستخدامها للإجابة عن الأسئلة البحثية التي تدور حول الحالة في المجتمع ولذلك فإن اختبارات الفروض هي إجراء إحصائي يسمح للباحثين باستخدام بيانات العينات للوصول إلى استدلالات أو استنتاجات عن المجتمع موضع الاهتمام (Gravetter & Wallnau, 2014). أو بكلمات أخرى هي الإجراء

الإحصائي للوصول إلى صناعة قرار لمشكلة ما يتناولها الباحث، أو هي الطريقة المستخدمة لتقييم بيانات العينة لنعلم أكثر عن خصائص المجتمع. ويعرفها Privitera (2015) بأنها الطريقة المستخدمة لاختبار ادعاء أو فرض عن المعالم في المجتمع باستخدام بيانات العينة.

والفرض هو تفسير مفترض أو تخمين ذكي أو توقع لظاهرة ما ويختبر باستخدام إجراءات الطريقة العلمية وغالبًا يتم صياغته عن الحالة في المجتمع، وعلى الرغم أن تفاصيل اختبار الفرض تتغير من موقف بحثي 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 الإحصائي هي:

- يصيغ الباحث فرض في ضوء دراسات سابقة أو استدلال منطقي.
- يضع الباحث خصائص وشروط عن العينة (عمرها، نوعها).
- يحسب متوسط الدرجات أو أي 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 في متن الدراسة البحثية ويعتمد على نوع الدراسة ونوعية البيانات، وتفاصيل اختبار الفرض تتغير من موقف بحثي إلى آخر.

المدخل العام لاختبارات الفروض الصفرية

يوجد جدل كبير حول الخطوات المراد إتباعها للتحقق من الفروض خلال اختبارات الفروض فيراها البعض أربعة خطوات (Hinkle ; Gravetter & Wallnau, 2014) وبينما يفترضها البعض ستة خطوات (Howell, et al., 1994; Privitera, 2015)

(Huck, 2012; Nolan & Heinzen, 2012; 2013) ، وسواء أربعة خطوات أو ستة فيوجد اختلاف بينهما فى ماهية هذه الخطوات، فمثلاً (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 إحصائياً باستخدام تصميم إحصائى أمثل.

وإذا كانت الخصائص المراد دراستها فى المجتمع مقاسة كمياً 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 معلمي Non- Parametric Hypothesis وهذه الفروض المعلمية وغير المعلمية تعرف بالفرض الإحصائى.

وتصاغ الفروض الإحصائية فى صورة فرضين أحدهما صفرى أو العدمى (H_0) Null Hypothesis والآخر بديل (H_A) Alternative Hypothesis وكليهما مكملين لبعضهما البعض. وفى الحقيقة الفرض الصفرى هو الوسيلة لاختبار فرض البحث إحصائياً ولذلك فإن رفض الفرض الصفرى يسمح للباحثين بقبول فرض البحث أو الفرض البديل، وعليه فإن الإحصاء الاستدلالي 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 صفرىاً لأنه فى كل الأحوال سواء صاغه أو لم يصيغه فإنه يتحقق منه إحصائياً 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 H_0 حيث H تشير إلى Hypothesis بينما (0) Zero تشير إلى فرض التأثير صفرى Zero effect Hypothesis ، ويفترض أن يكون الفرض الصفرى حقيقى. ويتم صياغة الفرض الصفرى لرفضه، وإذا فشل الباحث فى رفض H_0 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 $\mu=100$: H_0 ، حيث يوضع الرمز H_0 أولاً متبوعاً 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 $H_0: \rho=0$ أى لا يوجد علاقة أو ارتباط بين المتغيرين. وللباحث الحرية فى اختيار أى قيمة يريد اختبارها فى المجتمع مثلاً:

$$H_0: \mu = 118, \mu = 100, \mu = 0.0$$

وليس شرطاً أن يكون المعلم مساوياً للصفر وإنما مساوياً لقيمة معينة يفترضها الباحث. والقيود الإحصائية على القيمة العددية التى تظهر فى H_0 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 مثلاً:

$$H_0: \rho = +1.30$$

$$H_0: \rho = 1.00$$

فهذا غير مقبول لأن قيمة معامل الارتباط في المجتمع لا تزيد عن 1.30 أو لا تصل إلى 1.00 خاصة في العلوم النفسية والاجتماعية. فإذا صاغ باحث الفرض الصفرى كالتى: $H_0: \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)، فهذا يعنى أن متوسط المجتمعين متساوٍ ولا فرق بينهما وهذا يكون مفاده أن الفرق بينهما صفرًا.

وعلى الرغم أن الباحثون لهم الحرية في أى قيمة تختبر حولها الفروض إلا أن القيمة صفر تبقى هي الأكثر استخدامًا في 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 كالتى:

$$H_0: \mu_1 - \mu_2 = 5$$

والفرض الصفرى قائم حول اختبار المعلم في المجتمع من خلال بيانات العينة. ونادرًا ما يشير الباحثون أو يتحدثون عن الفرض الصفرى في تقرير بحوثهم.

- **الفرض البديل Hypothesis Alternative:** 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 H_A أو H_1 حيث H هي اختصار لـ Hypothesis و A هي اختصار لـ Alternative، وعلى ذلك ففي أى دراسة بحثية يبدأ الباحث في صياغة فرض البحث لاختبار معلم المجتمع ويقوم بصياغة الفرض الصفرى للتحقق من فرض البحث إحصائيًا، وعلى ذلك فالفرض البديل هو أن الادعاء بأن الفروق في النتائج بين المعالجات أو الظروف أو المجموعات ترجع إلى المتغير المستقل، فإذا كان الفرض الصفرى يتعامل مع قيمة محتملة مثلًا لمعامل ارتباط

بيرسون في المجتمع: $p=0.00$ ، فإن الفرض البديل يجب أن يتعامل مع نفس القيمة المفترضة في المجتمع:

$$H_A: \rho > , < , \neq$$

وإذا كان الفرض الصفري يتعامل مع فروق بين متوسطات مجتمعين:

$$H_0: \mu_1 = \mu_2$$

فإن الفرض البديل: $H_A: \mu_1 \neq \mu_2, \mu_1 > \mu_2, \mu_1 < \mu_2$

وعلى ذلك فإن H_0 ، H_A متماثلين في أنهما يتعاملوا مع نفس المجتمع ونفس الاهتمام الإحصائي (المعلم) ونفس المتغيرات، الاختلاف الوحيد بينهما هو القيمة المفترضة في المجتمع المتضمنة في H_A تختلف عن تحديدها في H_0 فإذا كان: $\rho = 0$: H_0 ، فإن $\rho \neq 0$: H_A ، ودائمًا اختبارات الفروض تنتهي بقرار إما رفض أو الفشل في رفض الفرض الصفري. فإذا رفض H_0 ب التالى فإن H_A يعكس الحالة في المجتمع، بكلمات أخرى فإن H_0 , H_A وجهين متعارضين بمعنى 50% قبول H_0 و 50% قبول H_A .

والقضية في صياغة الفرض H_A هي هل يوضع $0.1 \neq$ 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 الآتي:

أجرى باحث دراسة لمقارنة الذكور والإناث في الذكاء وكان اهتمامه هو المتوسط وعلية فإن الفرض الصفري : $\mu = \mu$ إناث ذكور: H_0

وإذا كان فرض البحث لا يوجد فروق بين الذكور والإناث في الذكاء فإن الفرض البديل: $\mu \neq \mu$ ذكور: H_A ، وعندئذ يقال ان الفرض البديل بانه غير موجه. ولكن إذا كان فرض البحث توجد فروق بين الذكور والإناث في الذكاء لصالح الذكور أو توجد فروق بين الذكور والإناث في الذكاء لصالح الإناث عندئذ يقال أن الفرض البديل موجه حيث:

$$H_A: \mu_1 > \mu_2$$

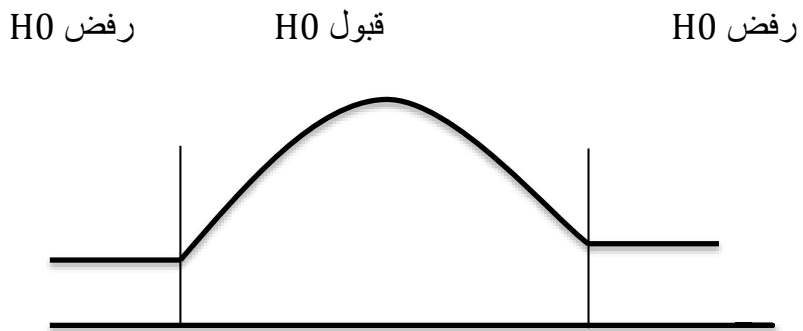
$$H_A: \mu_1 < \mu_2$$

أو

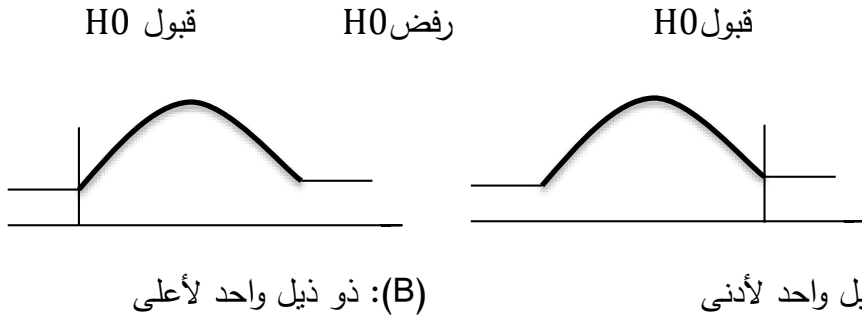
وعملية تحديد طبيعة اتجاهية الفرض البديل سواء كان موجهة أو غير موجهة على غاية كبيرة من الأهمية فى إجراء اختبارات الفروض، ولا يمكن صناعة قرار حول H_0 ما لم يتم تحديد خاصية الاتجاهية سواء موجهة أو غير موجهة. ويرى (Huck 2012) أنه فى معظم الدراسات الإمبريقية توضع الفروض البديلة غير موجهة، ويمكن عرض فروض غير موجهة وموجهة كالتالى:

فرض صفرى	فرض بديل غير موجهة	فرض بديل موجهة
$H_0: \rho = 0.20$	$H_A: \rho \neq 0.20$	$H_A: \rho > 0.20$
$H_0: \mu_1 = \mu_2$	$H_A: \mu_1 \neq \mu_2$	$H_A: \mu_1 > \mu_2$

ويضع الباحثون H_A بصورة غير موجهة لأنهم غير متأكدين ما إذا كانت المعلمة فى الفرض الصفرى كبيرة أو صغيرة أو لصالح أى مجتمع، ويمكن صياغته بصورة موجهة إذا كان هناك تدعيم من الدراسات السابقة أو النظرية. ويرى (Huck 2012) من غير المناسب للباحثين فحص البيانات أولاً وتحديد فى أى اتجاه تكون المعلمة فى الفرض البديل وعلى ذلك فالقرار فى تحديد اتجاهية الفرض البديل يجب أن يكون مستقل تماماً عن أى حديث عن البيانات أو فحصها. والفرض البديل غير الموجهة يطلق عليه الباحثين اختبار ذو ذيلين Two tailed test لأن قاعدة القرار تكون على جانبى (ذيلى) التوزيع العينى الاعتنالى.



بينما الفرض البديل الموجه يسمى اختبار ذو ذيل واحد One tailed test وقد يكون اتجاهية المعلمة لأعلى (أكبر من <) $\mu_1 > \mu_2$ $H_A:$ عندئذ يطلق عليه اختبار ذو ذيل واحد لأعلى Upper one tailed test، وقد تكون اتجاهية المعلمة لأدنى (أصغر من >) $\mu_1 < \mu_2$ $H_A:$ عندئذ يطلق عليه اختبار ذو ذيل واحد لأدنى Lower one tailed test.



على ذلك فالفرض غير الموجه يكون محدد بالعلامة ($\neq$). 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 الصفري وأحيانًا يؤجلها البعض إلى بعد مرحلة جمع البيانات وتلخيصها، واختبارات الفروض لا تسمح مطلقًا بوجود ذاتية للباحث لتؤثر على الوصول إلى القرار مما يؤدي إلى وجود تحيز فيما يخص التحقق من الفرض الصفري. وعليه فإن الهدف تحديد أو الإعتماد على نقطة قطع لتحديد ما إذا كانت بيانات العينة متسقة أو متعارضة مع الفرض الصفري. ومستوى الدلالة الإحصائية يعتبر بمثابة نقطة القطع الموضوعية لتحديد ما هو القرار الذى نتوصل إليه فيما يخص الفرض الصفري، وهذا التحديد لا بد أن يتم قبل جمع البيانات وهذه نقطة القطع يتم فى ضوءها 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 0.10).

(10% ، 0.05 (5%) ، 0.01 (1%) والقيمة الأكثر استخدامًا وشيوعًا في البحث النفسى والسلوكى هي 0.05، ويرى (Gravetter & Wallnau (2014 أن مستوى ألفا level Alpha 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 $\alpha = 0.05$ 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 فى ذيل المنحنى أو التوزيع تعرف بأنها نواتج لا تتسق مع الفرض الصفرى ومن غير المحتمل حدوثها لو أن الفرض الصفرى حقيقى. وعندما تقع متوسط بيانات عينة البحث فى منطقة الرفض فإننا نستدل على أن البيانات غير متسقة مع الفرض الصفرى وبالتالي نرفضه.

ويستخدم كل الباحثون مصطلح مستوى الدلالة ويستخدم أيضًا مستوى ألفا، بينما يذكر البعض $p = 0.05$ وأحيانًا يشير البعض إلى أن رفض الفرض الصفرى إذا كانت $p < 0.05(\alpha)$ وهكذا بالنسبة لـ $\alpha = 0.01$ 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 فى قاعدة القرار وذلك فى الحصول على القيمة الحرجة Critical value للاختبار (القيمة الجدولية) من خلال الكشف فى الجداول لكل اختبار محدد.

وبغض النظر عن القرار حول H_0 بعد مقارنة p بـ α أو مقارنة القيمة الحرجة بالقيمة المحسوبة للإحصاء المستخدم فمن المحتمل أن يصل الباحث إلى قرار خاطئ، فاختبار فرض صفرى فى ضوء فرض بديل باستخدام بيانات العينة يصاحبه نوعان من الخطأ قد يكون بسبب عدم انتقاء العينة عشوائيًا أو عدم توافر مسلمات الاختبار الإحصائية 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 (Type I error (α):** من المحتمل أن بيانات العينة تقودنا إلى رفض الفرض الصفري (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 الصفري 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 الصفري 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 فى الأدبيات البحثية المرتبطة بظاهرة ما. ويمكن للباحثين أن ينشؤا نظريات ويعدوا تجارب أخرى فى ضوء هذه النتائج الخاطئة وبدوره يترتب عليه إهدار للوقت والموارد.

• **الخطأ من النوع الثانى (β) Type II error:** عندما يرفض الباحث الفرض

الصفري فهذه مخاطرة الخطأ من النوع الأول، بالمثل عندما ي فشل الباحث فى رفض الفرض الصفري (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 الصفري 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 H_0 على مستوى العينة بينما هو خطأ على مستوى المجتمع. والخطأ من النوع الثانى يحدث عندما يكون متوسط العينة ليس فى نطاق منطقة الرفض لـ H_0 على الرغم أن للمعالجة أثر فى العينة، ويحدث هذا غالباً عندما يكون تأثير المعالجة أو المتغير المستقل صغير نسبياً 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 كالتالى:

الحالة الحقيقية (المجتمع H_0) ؟

العينة	المجتمع H_0 حقيقى لا أثر للتجربة	H_0 غير حقيقى أثر للتجربة
أقبل H_0	لا أخطاء No errors	خطأ من النوع الثانى (بيتا β)
أرفض H_0	خطأ من النوع الأول (ألفا α)	لا أخطاء No errors

واختيار $\alpha = 0.05$ فإنه الفرصة لرفض الفرض الصفري الحقيقى يساوى 5 مرات من 100 مرة القرار فيهم صحيح. وإذا كانت $\alpha = 0.01$ 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 100 الحقيقى 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 آلاف 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 آخرين 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			
حكم القاضى	برئ	قرار صحيح ($1-\alpha$)	قرار خاطئ (β)
	مذنب	قرار خاطئ α	قرار صحيح ($1-\beta$)
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 نادراً 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 إحصائية α متحفظة Conservative مثل 0.05 ، 0.01 ، وأحياناً 0.001 لتقليل مخاطر الوصول إلى تقرير بحث خاطئ. ومستوى ألفا 0.000001 يعنى تقريباً لا مخاطرة لارتكاب الخطأ من النوع الأول.

ولكن ماذا لو أن باحث تبنى مستوى ألفا أكبر من 0.05، مثلاً جعلها 0.10 وهذا يعطى احتمالية عالية لرفض H_0 . 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 الدراسات التى ترفض H_0 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 (Huck 2012) 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 خطورة 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 دوراً هاماً 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 دوراً فى احتمالية الوقوع فى الخطأ من النوع الثانى؟ والثانى: لو رفض الباحث H_0 فهل مستوى ألفا يشير إلى أن احتمالية H_0 حقيقية؟ 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 H_0 فليس من المناسب أن يعود الباحث مرة أخرى ويعيد النظر فى مستوى α ويفسر مستوى ألفا على أنه مؤشر الاحتمالية أن H_0 صحيح، ومستوى ألفا فى الدراسة يشير فقط إلى أن 5 مرات من 100 مرة H_0 يرفض إذا كان H_0 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 التجريبي 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 ، $\alpha=0.05$ 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 تغييرها بعد تحليل البيانات ويفضل أن يكون مستوى الدلالة الإحصائية واحد لكل أسئلة الدراسة.

ثالثاً: اختيار الاختبار الإحصائى المناسب: الاختبار الإحصائى هو تكتيك 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 التي تلعب دوراً في اختيار الاختبار الإحصائي:

1. **الهدف من التحليل أو نوع الفرض المراد اختبارها:** في استخدام الإحصاء توجد عدة أهداف للتحليل وهي:

• **وصف واكتشاف إحصائيات البيانات Descriptive focus:** وفيها يهدف الباحث إلى إعطاء صورة ملخصة لبيانات العينة وكذلك تمثيل البيانات بيانياً وهذه النوعية من الإحصاء الوصفى تتضمن النسبة المئوية (20% مدمن الفيس بوك مثلاً) أو مقاييس النزعة المركزية ومقاييس التشتت والرسومات البيانية وكل هذه الأساليب تندرج تحت الإحصاء الوصفى.

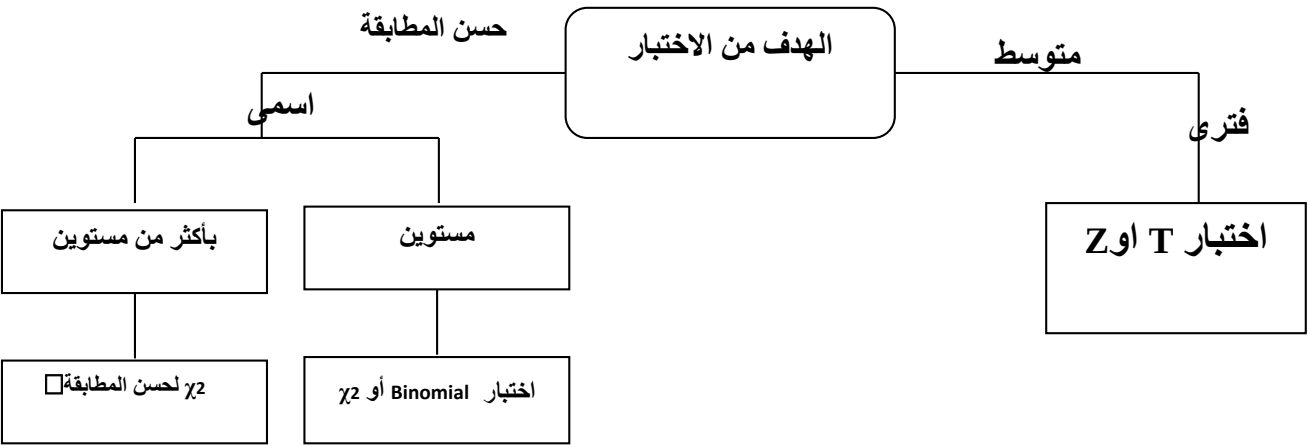
• **عندما يكون هدف الباحث الاستدلال من بيانات العينة للحالة أو الوضع فى المجتمع فإن هدف الباحث هو التقدير Estimation focus** وعلى ذلك فلا بد للباحث التركيز على اختبارات الإحصاء الاستدلالي أو اختبارات الدلالة الإحصائية للتحقق من الفروض أو الأسئلة الآتية:

• **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، واختبار F المستقلة أو المرتبطة وغيرها، والاختبارات اللابارامترية مثل مان - ويتنى، اختبار كروسكال - والاس، فريدمان، ويلكوكسون وغيرها.

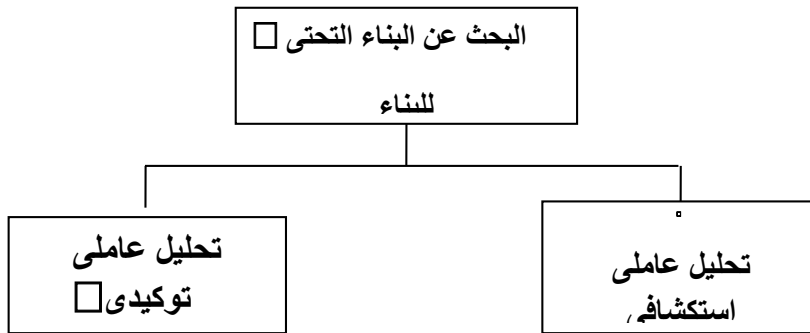
• **العلاقة بين متغيرين فأكثر:** يهدف الباحث إلى دراسة للعلاقة بين متغيرين وتوجد اختبارات عديدة مثل معامل ارتباط سبيرمان ومعامل ارتباط بيرسون ومعامل ارتباط كندال وغيرها.

وفيما يلي مخطط يوضح أهم الاختبارات العلاقة:

- دراسة مدى حسن المطابقة بين بيانات العينة والمجتمع: بمعنى اختبار فروض حول متغير واحد.



- البحث عن البناء التحتي للمتغيرات: وإذا كان هدف البحث العثور عن البناء الداخلي لمفهوم ما وذلك لمجموعة فرعية من المفردات المكونة للأبعاد.



- 2 . خصائص التوزيع **Distributional charecteristics**: فالمسئلة الأساسية في استخدام الاختبارات الإحصائية البارامترية هي أن بيانات العينة تكون مسحوبة من مجتمع ذات توزيع اعتدالي، وهذه المسئلة ضرورية لاختبار فروض حول متوسطات تباينات أو نسب أو ارتباطات ، وكذلك من أهم المسلمات الواجب توافرها هي تساوى التباينات (تجانسها) بمعنى أن تباينات المجتمعات متساوية، بالتالي فاختبارات فروض المتوسطات (ANOVA , T) بين مجموعتين فأكثر لا بد أن يتوفر لها هذه المسلمات وهذه المعالم مثل المتوسط والتباين يستخدم لها اختبارات بارامترية Parametric tests بينما في حالة عدم توافر مسلمتي 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 كبيراً .

3. مستوى القياس للمتغيرات: فمستويات القياس عالية الرتبة (الفترة - النسبة) تحتاج إلى إحصاء أكثر تعقيداً وهي تختبر فروض حول معالم حيث لها متوسط وانحراف المعياري 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 الاستدلالي يكون أقل موثوقية وذلك لأن الاختيار العشوائى للعينة يقلل من أخطاء المعاينة. اما فيما يخص حجم العينة فإن بعض الإجراءات الإحصائية خاصة البارامترية لا تؤدي بكفاءة ما لم يكن حجم العينة كبيراً ويفضل البعض أن يزيد عن 30 فرداً، فمع العينات الكبيرة يفضل استخدام الإحصاء البارامترى ولكن مع أحجام العينات الصغيرة يفضل استخدام الإحصاء اللبارامترى خاصة إذا لم تتوفر 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، فمثلاً إذا كانت القياسات على نفس الأفراد عبر الزمن (قياس قبلى - قياس بعدى) فانه 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

المستقلة Independent Measure Design وهو أن الأفراد المختلفين يقاس عليهم نفس الخاصية وهذه النوعية تتبع 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 أيضاً **Univariate analysis**. أما إذا كان التحليل يتضمن أكثر من متغير تابع حيث النموذج الإحصائي إما 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 تحديداً أيهما تابع وأيهما مستقل ويسمى النموذج فى هذه الحالة بالنموذج المتدرج **Multivariate Model** مثل التحليل العائلى الاستكشافى فإن الإحصاء المستخدم فى هذه الحالات الإحصاء المتدرج **Multivariate statistics**.

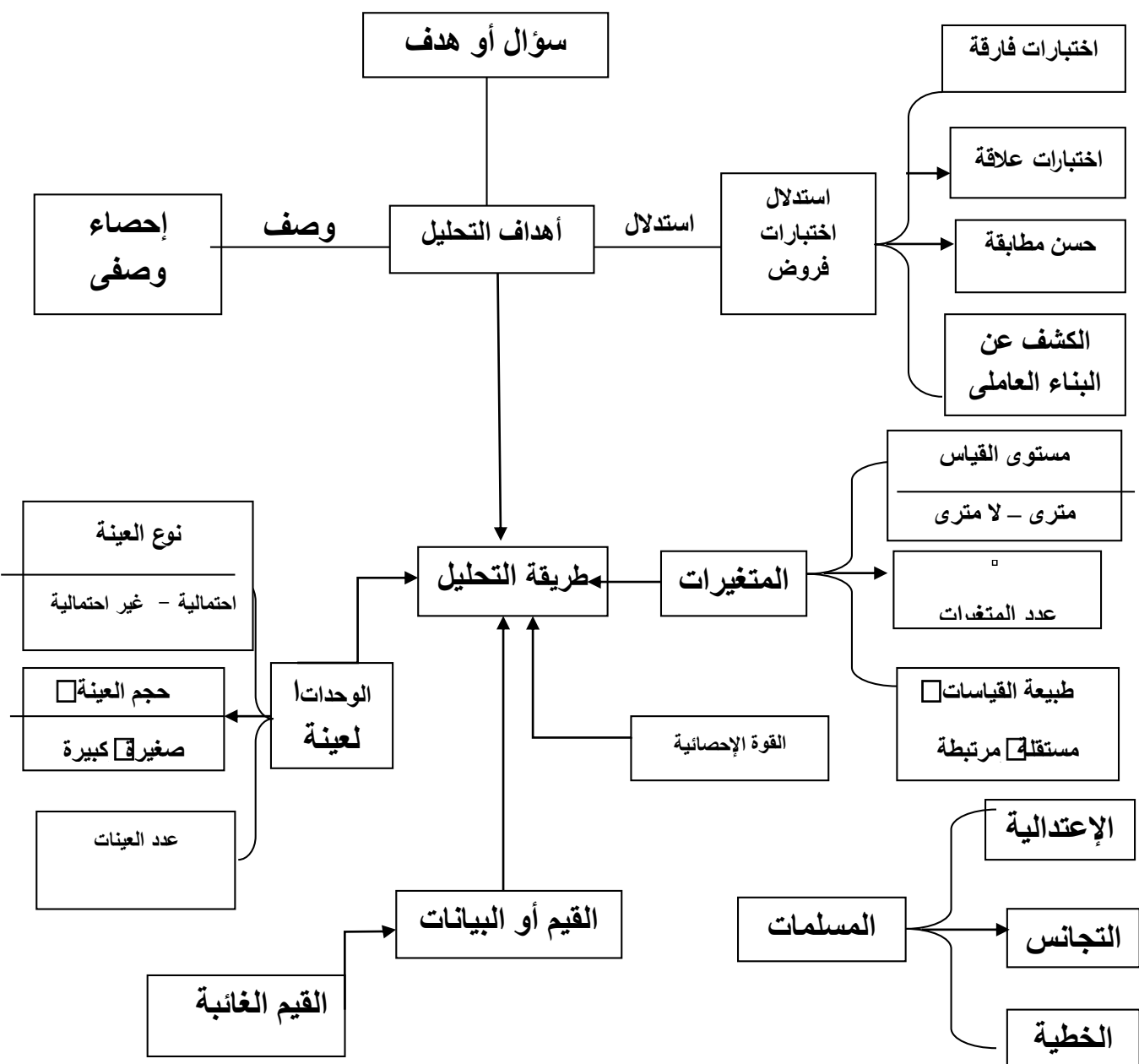
7. طبيعة البيانات المتحصل عليها من المفحوصين: القضية الأساسية هى ان المفردات التى لا يستجيب عليها المفحوصين **Non-Response Item** أو البيانات المفقودة **Missing Data** على متغير أو أكثر فى متغيرات التحليل، والأثر المباشر للبيانات الغائبة هو تقليل حجم العينة وهذا بدوره يؤثر على نوعية الإحصائى المناسب، حيث يطبق الباحث فى دراسته على 50 فرداً 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 الصفري عندما يجب أن يرفض أو قدرة الاختبار فى الحصول على

دلالة إحصائية وهي موجودة بالفعل في مجتمع الدراسة وقوة الاختبار هي متلازمة مع الخطأ من النوع الثاني β حيث: $P = 1 - \beta$

والاختبارات البارامترية أكثر قوة إحصائية من الاختبارات اللابارامترية حيث إن الاختبار البارامترى يكون أكثر قدرة على رفض الفرض الصفري وإن قوة الاختبار الإحصائي تزيد بزيادة حجم العينة، وفي حالة أحجام العينات الصغيرة فإن الاختبار البارامترى واللابارامترى لهما نفس القوة وإذا توافرت مسلمات الاختبار الإحصائي البارامترى وخاصة الاعتدالية فإنه يقال إن الاختبار له صدق إحصائي عالٍ High Statistical Validity هي استخدام الاختبار مع توفر أفضل شروط استخدامه. ولكن هل عدم توفر مسلمات الاعتدالية والتجانس للتباينات لا يستخدم الاختبار البارامترى؟ حقيقة معظم الاختبارات البارامترية تتميز بالمناعة أو الضلاعة النسبية Relative robust أى أنه غير حساس نسبياً للابتعاد الضئيل أو المتوسط Mild-to-Moderate عن مسلماته خاصة مع أحجام العينات الكبيرة وهذه المناعة تكون محددة بكل اختبار بمعنى كل الاختبارات البارامترية ليست لها نفس الدرجة من المناعة ولكن إذا لم تتحقق هذه المسلمات بدرجة كبيرة (التواء شديد - تفرطح شديد) فمن الأفضل عدم الاعتماد على الاختبار البارامترى ويفضل استخدام نظيره اللابارامترى.

في ضوء ما سبق في هذه الخطوة من اختبارات الفروض واستناداً على فرض أو سؤال البحث وتحديد النموذج الإحصائي المناسب، فإن الباحث يستطيع تحديد نوع الإحصاء (بارامترى - لا بارامترى) وبدوره يحدد الأسلوب أو الاختبار الإحصائي المناسب. يمكن تلخيص أهم العوامل المؤثرة في اختيار الأسلوب الإحصائي في الشكل الآتي:



الشكل (3.9): مخطط لأهم العوامل المؤثرة في اختيار الأسلوب الإحصائي

Diamantopoulos & schlegelmilch (2003)

وبعد تحديد الاختبار فإن يجب أن تحدد قيمة درجات الحرية له.

درجات الحرية Degrees of freedom: أى اختبار إحصائي 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

كمية المعلومات التي تكون حرة عند حساب الإحصائية المستخدمة فإذا كان تقدير التباين σ^2 في المجتمع يقدر من خلال S^2 في العينة من خلال عينة عشوائية لعدد n من الدرجات وعليه فإن درجات الحرية تقدر من n وفي هذه الحالة $(n-1)$ وتختلف درجات الحرية من اختبار لآخر، ففي حالة اختبار T المستقلة فإن:

$$df = n_1 + n_2 - 2$$

في حالة χ^2 لاستقلالية متغيرين فإن:

$$df = (r-1)(c-1)$$

• r عدد الصفوف

• C عدد الأعمدة.

لاحظ في هذه النقطة بعد تحديد الاختبار الإحصائي ودرجات الحرية لا بد للباحث أن يتحقق من مسلمة الاختبار الإحصاء المستخدم سواء كانت الاعتدالية، الاستقلالية، العشوائية، والتجانس وغيرها قبل تحليل البيانات.

وبفرض ان باحث اراد اختبار فرضية: ما إذا كان متوسط التحصيل في مجتمع المدرسة الثانوية في مادة الرياضيات يزيد عن 0 ، فإن الاختبار المناسب Z ولا بد من التحقق من مسلمة استخدامه وهي:

• **المعاينة العشوائية Random sampling**: أن يتم اختيار مفردات العينة ولتكن

200 من المجتمع وبصورة عشوائية حتى نثق في تعميم النتائج.

• **القياسات المستقلة**: العينة تتكون من قياسات مستقلة حيث كل فرد في العينة

مستقل عن الآخر، فدرجة أحد الأفراد لا تتأثر بدرجة فرد آخر وهذا يتحقق في العينات العشوائية.

• **التوزيع الاعتدالي Normal distribution**: لتقييم فروض اختبار Z لا بد من

استخدام جداول التوزيع الاعتدالي لتحديد منطقة الرفض، وعليه فإن بيانات العينة لا بد أن تكون اعتدالية التوزيع.

رابعًا: **قاعدة القرار ومنطقة رفض H_0** : أي قيمة يتم حسابها من الاختبار هي

مقياس معبر لنتائج الاختبار وهي تعكس بيانات العينة ويطلق عليها بالإحصائية أو القيمة المحسوبة Calculated value بينما التوزيع الاحتمالي للإحصاء المستخدم

يشير إلى كل قيم الاختبار عندما يكون H_0 حقيقى (فى المجتمع) . وللحصول على التوزيع الاحتمالى فى المجتمع من خلال المتوسط (μ) والتباين S^2 وهو توزيع تم افتراضه لدرجات كل أفراد المجتمع وعملياً لا يمكن الوصول اليه ولكن تم افتراضه نظرياً عن طريق اخذ عينات متعددة من المجتمع وهو غالباً يكون اعتدالى وهذا التوزيع الاحتمالى للاختبار يسمح بتقدير خطأ المعاينة. وكثير من الاختبارات الإحصائية فإن التوزيع الاحتمالى يأخذ أشكالاً معروفة (اعتدالى مثلاً) 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 المعياري للمجتمع

• N حجم العينة

لاحظ أنه يجب أن يكون فى المقام الانحراف المعياري للمجتمع ولكن يستخدم بدلا منه S وذلك للعينات الكبيرة. وإحصاء Z يتبع التوزيع الاعتدالى المعيارى ووضع الإحصائيون جداول إحصائية لاستخدامها فى الممارسة العملية.

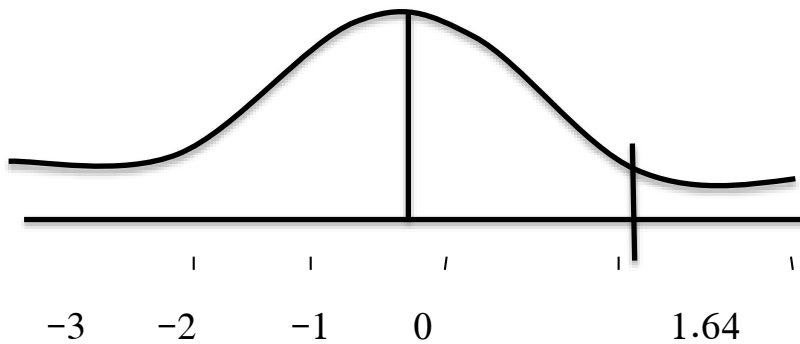
وعلىنا فى هذه الحالة تحديد مدى قيم Z التى تقودنا لرفض الفرض الصفرى أو بكلمات اخرى تحديد منطقة الرفض Rejection region او منطقة الدلالة $\text{Significance region}$ وكل القيم الأخرى التى تشير إلى قبول الفرض الصفرى وتعرف بانها منطقة قبول Acceptance region . والقيمة Z التى تفصل بين منطقة القبول ومنطقة الرفض 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 وبالتالي نرفض H_0 وإذا وقعت داخل منطقه

القبول يفشل الباحث فى رفض H_0 . ولكن كيف يمكن تحديد منطقة الرفض ؟ فى هذه النقطة تلعب مستوى الدلالة الإحصائية دورا هاما التى سبق تحديدها فى الخطوة الثانية، فإذا كان $\alpha = 0.05$ فى هذه الحالة الباحث يقسم التوزيع للاختبار إلى منطقة رفض تغطى 5% من القيم والمنطقة الأخرى تغطى 95% من القيم (منطقة قبول H_0) فإذا كان الفرض الإحصائى:

$$H_0: \mu \leq 3, \mu = 3,$$

$$H_A: \mu > 3$$

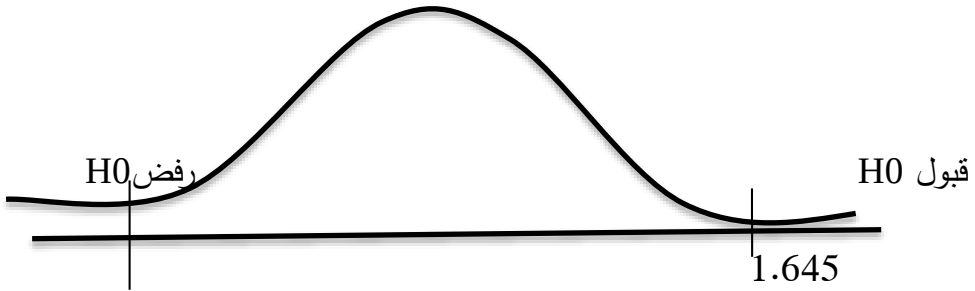
وبالكشف فى جدول Z لاختبار ذو ذيل واحد لأعلى (اليمين) وبمستوى دلالة إحصائية $\alpha = 0.05$ فإن Z الحرجة (الجدولية) = 1.645. وعليه فإن 5% من القيم تزيد عن 1.645 و 95% من القيم تقع أقل من 1.645 وعليه فاحتمال ان إحصائية الاختبار المحسوبة تزيد عن 1.645 عندما يكون الفرض الصفرى حقيقى هى 0.05 (مستوى الدلالة الإحصائية α) وعليه يمكن صياغة قاعدة القرار كالاتى:



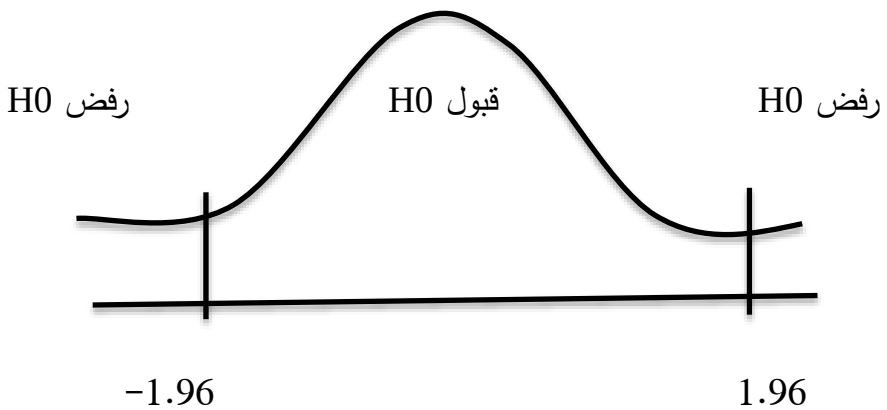
وإذا كانت القيمة المحسوبة للاختبار $\leq$ القيمة الحرجة للاختبار، فإننا نرفض H_0 بمعنى الحصول على دلالة إحصائية والعكس قبول H_0 أى عدم دلالة إحصائية، وهذه القاعدة لمعظم الاختبارات الإحصائية (T, F, r) ولكن هذه القاعدة تكون العكس حيث إذا كانت القيمة الجدولية < القيمة المحسوبة يتم رفض الفرض الصفرى للقليل 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، فإذا ثبت عدم

دالتها عند 0.01 فيغيرها الباحث إلى 0.05. على أساس يتم رفض H_0 عندها. وإذا كانت الفروض الإحصائية: $H_0: \mu = 3$, $H_A: \mu < 3$

فإن الاختبار ذو ذيل واحد لأدنى (اليسار) وبالكشف في جداول z فإن $z = 1.64$ ، وعليه فمنطقة الرفض والقبول H_0 كالآتي :



وهي مثل الاختبار ذو ذيل واحد لأعلى ولكن الفرق هو أن القيم الكبيرة لإحصائية الاختبار المحسوبة في منطقة الرفض سوف تكون أصغر وليس أكبر. وإذا كان الفروض الإحصائية: $H_0: \mu = 3$, $H_A: \mu \neq 3$ ، فإن الفرض البديل غير موجه أو ذو ذيلين وبالكشف في الجدول z للاختبار ذو ذيلين $\alpha = 0.05$ فإنها تكون $z = \pm 1.96$ الدرجة



لذلك توجد منطقتين للرفض كلاً منهما تشغل 0.025 (25%) منطقة رفض تحت المنحنى. ولاحظ توجد قيمتين حرجيتين هي أكبر من 1.96 و - 1.96 . وعلى ذلك فإن تحديد الفرض البديل يلعب دوراً كبيراً في تحديد منطقتي القبول والرفض H_0 ، وكما هو واضح أن رفض الفرض الصفري في الاختبار ذو ذيل واحد أسهل من رفضه لاختبار ذو ذيلين ولذلك فإن الاختبار ذو ذيل واحد أقوى إحصائياً من الاختبار ذو

ذيلين لأن الاختبار ذو ذيل واحد يرفض H_0 عندما يكون الفرق بين العينة والمجتمع صغير نسبياً، والاختبار ذو ذيلين يتطلب ان يكون الفرق كبيراً نسبياً وإذا كان الاختبار ذو ذيلين يعطى دلالة إحصائية فإن نظيرة ذو ذيل واحد يعطى دلالة إحصائية أيضاً ولكن إذا كان الاختبار ذو ذيل واحد يعطى دلالة إحصائية فإن الاختبار ذو ذيلين ليس بالضرورة ان يعطى دلالة 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 1.96 وعليه فهى دالة إحصائياً ولذلك يتم رفض الفرض الصفري" ولكن فى تقارير البحوث يتم عرض القيمة المحسوبة للاختبار فقط وعليه فإن الطريقة الثانية لتقييم 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 الصفري وهى مشابهة لمستوى الدلالة الإحصائية.

على ذلك فإن برامج الكمبيوتر يعطى قيمة p مصاحبة للقيمة المحسوبة للاختبار ويتم القرار فى ضوء مقارنة قيمة p بقيمة مستوى الدلالة الإحصائية α فإذا كانت:

$$P \leq \alpha$$

يرفض الفرض الصفري بمعنى وجود دلالة إحصائية وإذا كانت:

$$P > \alpha$$

نقبل الفرض الصفري.

وهذا المدخل هو الشائع الاستخدام فى تقارير البحوث العلمية.

خامساً: 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. ففي المثال الحالي 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: في نهاية عملية اختبارات الفروض امام الباحث بديلين فيما يخص H_0 اما رفضه (وجود دلالة إحصائية) أو الفشل في رفضه (عدم وجود دلالة إحصائية). ويشير الباحث إلى وجود فروق ذات دلالة جوهريّة أو عدم وجود فروق حقيقية أو النتائج غير دالة إحصائيًا كلها مصطلحات يستخدمها الباحث. ويتم صناعة القرار مما يخص H_0 في ضوء مدخلين:

• مقارنة القيمة المحسوبة للاختبار بنظيرتها الحرجة: القيمة المحسوبة (2.94) < القيمة الحرجة (1.645) وعليه يتم رفض H_0 بمعنى أن متوسط التحصيل يزيد أو أكثر من 3 إذا كان الفرض البديل $\mu > 3$.

• حساب القيمة الاحتمالية للاختبار الإحصائي وهي قيمة p عندما يتم تحليل بيانات العينة باستخدام الكمبيوتر لتقييم الفرض الصفري تستخدم قيمة p الاحتمالية والقيمة الصغيرة لـ p تساعد على رفض H_0 . في هذا المثال فإن قيمة p المرتبطة باختبار z هي 0.0016:

$$0.0016 \leq 0.05$$

وبناء على ذلك يتم رفض H_0 . وتكتب في تقرير البحث كالاتي:

$Z = 2.94, P.0.05$ على ذلك فكلما دال تعنى كبير أو هام أو يستحق الاهتمام ولكن في اطار اختبارات الفروض فإنها لها معنيين مختلفين فتشير إلى أنه من غير المحتمل ان يحدث إذا كان H_0 حقيقى، أو بيانات العينة غير متسقة مع ما توقعنا في الفرض الصفري ولكن هذا لا يعنى أن النتائج في غاية الأهمية أو هامة Important أو وجود فروق حقيقية بين بيانات العينة و H_0 بمعنى فروق كبيرة .large

والنتائج الدالة تكون نتيجة:

- جودة أسئلة البحث بمتغيراته التي 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 $H_0: p=0$ ولكن بعد جمع البيانات وتحليلها تم رفض الفرض الصفري 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 0.0009. 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 تبدأ المحكمة بجمع الادلة لمعرفة حقيقة ما كان المتهم ارتكب "الجريمة". لاحظ أن البحث والمحكمة يحاولوا رفض الفرض الصفري (المتهم برئ).
- إذا توافرت ادله كافية فالباحث يرفض الفرض الصفري ويصل إلى نتيجة وهو ان للمعالجة أثر، وعليه إذا توفرت ادله كافية للقاضى فإنه يرفض H_0 (المتهم برئ) ويصل إلى استنتاج ان المتهم مذنب فى ارتكاب الجريمة.
- اذا لم تتوفر ادله كافية فالباحث يفشل فى رفض H_0 ويصل إلى استنتاج ان المعالجة ليس لها اثر وبالمثل اذا لم تتوفر 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 H_0 فان الباحث يرجع مرة اخرى إلى بيانات بحوثه لإجراء مهمة اضافية هو الذهاب ابعد من صناعة قرار حول H_0 فى سبيل ايضاح بعض الشئ 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 دلالة إحصائية بدون وجود للدلالة العملية أو الأهمية التطبيقية Practical significance وذلك لأن أحجام العينات الكبيرة تزيد من احتمالية رفض H_0 على الرغم ان بيانات العينة تبتعد عن H_0 بمقدار ضئيل جدًا وعليه 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 الصفري 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:

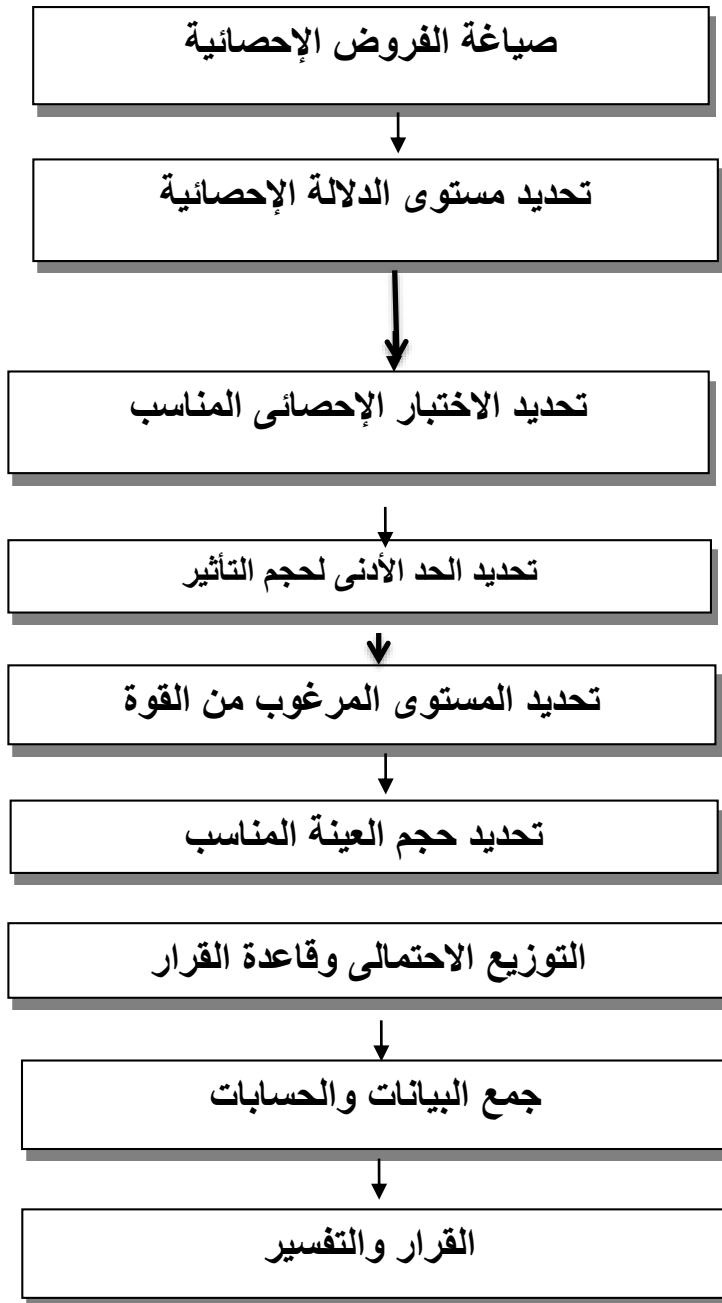
الأولى: عندما يقارن الباحث إحصائية بيانات العينة بالمعلمة فى الفرض الصفري H_0 فإن ذلك يتم فى ضوء القيمة الاحتمالية p ولكنها لا تعتبر هى الفيصل فى كل شئ ولكن لا بد من التركيز على إلى أى درجة بيانات العينة تبتعد أو تتحرف عن H_0 ، فعلى سبيل المثال إذا تناولت دراسة العلاقة بين متغيرين وكان الفرض الصفري $H_0: \rho = 0$ ،فالباحث يحاول النظر إلى كم درجة قيمة r فى العينة تبتعد

عن 0.0، وفى دراسة الفروق بين مجموعتين فإن الفرض الصفري $H_0: \mu_1 = \mu_2$ والباحث يفحص بيانات العينة لمعرفة إلى أى مدى الفرق بين متوسطى X_1 و X_2 يختلف عن الصفر فى الفرض الصفري.

الثانية: الباحث يقيم الفروق المقاسة بين ملخص أو إحصائية بيانات العينة وما يفترضه فى H_0 فان تحليل حجم التأثير يتضمن الحكم ما إذا كان حجم الفروق المقاسة $(\bar{X}_1 - \bar{X}_2)$ فى العينة صغيراً Tiny أو متوسط Medium أو كبير Gigantic or Large 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 لديه 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 الأخيرة فى مدخل التسع خطوات هو مماثل للخطوات الستة السابقة، وهنا يتم التركيز على الخطوات من الرابعة حتى السادسة، ومن الضروري التعرف على الفروق بين مدخل السبعة خطوات ومدخل التسعة خطوات وخاصة فيما يتعلق بحجم التأثير. ففي مدخل السبع خطوات تم حساب حجم التأثير بعد الوصول إلى قرار وذلك بالرجوع مرة أخرى 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 مسبقاً 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 الذكاء IQ، ويكون اهتمامه الإحصائى هو دراسة متوسط المجتمع حيث يكون: $H_0: \mu = 100$ ، فى هذه الدراسة المفترضة باعتبار أن الدرجة العظمى $150 = IQ$ ، $H_A: \mu > 100$ ، فى هذه الدراسة المفترضة باعتبار أن الدرجة العظمى $150 = IQ$ ، 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 H_0 (100) قبل إجراء الدراسة.

ودائماً 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 غالباً 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 الصفري مساوياً لـ 100 حيث يحدد الباحث خط واضح وفاصل بين النواتج عديمة الأهمية والنواتج التى لها أهمية على متصل المتغير التابع فى الدراسة.

وكثيراً 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 أيضاً 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 الشأن فان 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" للوصول 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 لرفض H_0 وكما نعلم ان القوة الإحصائية هى مقدرة الاختبار الإحصائي 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 الصفري والكشف عن الفروق الحقيقية ولاحظ أن مفهومى القوة الإحصائية و α مرتبطين بالأخطاء الاستدلالية التى يمكن أن تحدث، فآلفا 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 فدائما فى تقارير البحوث يجب كتابة مثلاً "لاختبار ذو ذيلين وبقوة إحصائية حددت مسبقا بـ 80. عند مستوى دلالة إحصائية 0.05. 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 عن H_0 ضئيل جدًا، حيث يتم رفض H_0 وهو فى الحقيقة صحيح "تضخم الخطأ 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 هي مرحلة أساسية في تصميم الدراسة هذا يفيد في تقليل التكلفة والموارد حين يزيد الباحث حجم العينة بدون داعى ويوصى (1992) Cohen بأهمية إجراء هذا النوع من تحليل القوة قبل إجراء الدراسة، ويعتبر هذا أحد المداخل المستخدمة لتحديد حجم العينة ويشير الباحث إلى العبارة الآتية في الدراسات "أوضح التحليل القبلى للقوة أنه لتحقيق مستوى قوة 0.80 مع $P < 0.05$ وحجم تأثير متوسط 0.30 مع ثلاثة متغيرات منبئة ان حجم العينة المطلوب 41 فرداً مثلاً. ومستوى قوه 0.80 يشير إلى ان احتمالية ارتكاب الخطأ من النوع الثانى منخفضة وهى 0.20.

واحد مميزات مدخل التسع خطوات هي 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 كبيراً لا داعى له وفى هذا فإن الاختبار يصل إلى رفض الفرض الصفري حتى لو كانت الفروق أو العلاقات ضعيفة جداً.

وإذا لم يعطى اعتبار لمفهومي حجم التأثير والقوة الإحصائية فإن الفشل فى رفض H_0 يعزى لانحرافات تافهه عن H_0 وقد يعزى إلى عدم حساسية الاختبار للكشف عن الفروق الحقيقة نتيجة لصغر الحجم العينة أو ان رفض H_0 نتيجة فروق ضئيلة ويعطى الاختبار دلالة الإحصائية لوجود لها على مستوى حقيقة الظاهرة لأن حجم العينة كبيراً. وأحياناً يذكر الباحثون العبارة الآتية "المحدد الأكبر فى هذه الدراسة صغر حجم العينة وان صغر حجم العينة يقلل من قوة الاختبار الإحصائى ولذلك فإن النواتج الإحصائية تكون معرضة للخطأ من النوع الثانى"، وعلى الرغم ان قيمة التأثير صغيرة جداً ولكنها داله إحصائية فى ضوء حجم العينة الكبيرة (2000) مثلاً وهذا التأثير غير دال عملياً. 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 وقتاً فى تصميم الدراسة لتحديد حجم العينة ولكن هذا يعطى الدراسة قدرًا كبيراً 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 تقريباً عند 5% ولكن هذا صعب المنال ولكن الشئ الوحيد الضامن لهذا هو هجر أو الابتعاد عن استخدام اختبارات الدلالة الإحصائية والاعتماد على استخدام فترة الثقة Confidence intervals كبديل لاختبارات الفروض الصفريّة. وتوجد عدة انتقادات للدلالة الإحصائية أهمها:

• **حساسيتها لحجم العينة Sensitivity to sample size** : عندما يكون حجم العينة صغيراً فإن التأثيرات القوية لا تكون دالة إحصائياً (ارتكاب الخطأ من النوع الثانى)، فى المقابل عندما تكون حجم العينة كبيراً فإن التأثيرات الصغيرة تكون دالة إحصائياً (ارتكاب الخطأ من النوع الأول)، فمثلاً إذا كان معامل الارتباط $r = 0.40$ (20) غير دال إحصائياً عند $P \geq 0.05$ لاختبار ذو ذيلين، بينما $r = 0.07$ (1000) دال إحصائياً. وعلى الجانب الآخر إذا كانت $r = 0.25$ 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 إضاعة عن برهان احتمالى للتناقض بين الفرض الصفرى والفرض البديل.

• **نسب الخطأ والقوة Power and error rate**: الانتقاد الأساسى لاختبارات الفروض الصفريّة النسبة العالية للخطأ من النوع الثانى غير المقبولة Schmidt,

(1996)، وهذا يؤدي إلى نقص القوة الإحصائية المناسبة (Cohen, 1962). والملاحظ في التراث البحثي في التربية وعلم النفس أن متوسط القوة في معظم التراث البحث في علم النفس تراوح من 0.40 إلى 0.60 بل أحياناً 0.20 في بعض التخصصات (Hunter, 1997 , Schmidt, 1996; Sedlmeier & Gigerenzer, 1989). وتوصل Maxwell (2004) إلى أن وسيط القوة في الدراسات غير التجريبية فقط حوالي 0.50. وهذا مفاده أن فرصة 50% لرفض صحيح للفرض الصفري في ضوء بيانات العينة و 50% رفض خاطئ للفرض الصفري. ويشير Kline (2013) في هذه الحالة من الأفضل عدم إجراء البحث وجمع البيانات والاعتماد على مبدأ ملك أو كتابة للعملة لتحديد رفض أو قبول الفرض الصفري.

وتوصل عامر (2014) إلى أن وسيط القوة للدراسات النفسية التربوية العربية ذات حجم التأثير الصغير 0.28 وحجم التأثير المتوسط 0.90، وحجم تأثير كبير 0.99 و في المجمل فإن متوسط القوة الإحصائية 0.72. وفي المجمل في التراث الأجنبي والعربي 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 البحثي لظاهرة ما، فإذا وجد تأثير حقيقي للظاهرة موضع الدراسة ولكن القوة الإحصائية 0.50 فإن هذا يعني أن نصف الدراسات تعطى نتائج دالة بينما النصف الآخر لا يعطى دلالة إحصائية، ولو نشرت هذه الدراسات فإن نصفها يعطى نتائج إيجابية والنصف الآخر نتائج سلبية وقد عبر Simmons, Nelson, & Simonsohn (2011) عن ذلك بعبارة "علم النفس الموجب والخطأ".

وعليه فالاعتماد التقليدي على اختبارات الدلالة الإحصائية في تفسير الدراسات يقود إلى تفسيرات خاطئة عما تعنيه نتائج الدراسات. ويرى Schmidt (1996) أن استخدام المدخل التقليدي في تحليل البيانات يجعل من غير المحتمل الوصول إلى استنتاجات صحيحة في معظم التخصصات. ويشير Cohen (1990) إلى أن العلوم الطبيعية

مثل الفيزياء والكيمياء لا تستخدم اختبارات الدلالة الإحصائية في تفسير بياناتها. ويعتقد الباحثون أنه إذا كانت النتيجة دالة عند 0.05 فإن احتمالية تكرار هذه النتيجة في الدراسات اللاحقة هي $1 - 0.05 = 0.95$. وهذا الاعتقاد خاطئ تمامًا. ولكن احتمال الاعادة فإن القوة الإحصائية للدراسة تقريبًا أقل من 0.95.

الاستنتاجات الخاطئة بعد صناعة القرار عن الفرض الصفري

توجد استنتاجات خاطئة بعد رفض أو الفشل في رفض الفرض الصفري وهي كالتالي (Kline, 2013):

• **خدوة الاحتمالية في 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 الصفري) ونتيجة تكون للتأثيرات الحقيقية الواقعية (رفض الفرض الصفري) وهذه الاعتقادات خاطئة. ولكن توجد عوامل اخرى تلعب دورًا في تفسير وحساب قيمة P في إطار التصميم البحثي للدراسة وعملية القياس وأساليب التحليل الإحصائي. وكثيرًا من الباحثين يعتقدون ان $0.05 < P$ لـ $\alpha = 0.05$ تعنى ان احتمال القرار اللازم لرفض الفرض الصفري هو ان الخطأ من النوع الأول أقل من 0.05.

• **خدوة الاعادة Replicability:** يوجد اعتقاد خاطئ ان النتائج الدالة إحصائيًا

هي قابلة للإعادة إذا اجريت الدراسة على عينات مماثلة اعتقادًا ان $0.05 < P$ 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%.

• **خدوة الحجم Magnitude Fallacy:** الاعتقاد بان القيمة المنخفضة لـ P

تشير إلى تأثير كبير أو قوة المعالجة كبيرة حيث يصف الباحث النتائج التي $P < \alpha$

بانها دالة إحصائية بالتالى فإن النتائج الدالة عند 0.01 أكثر حجماً من النتائج الدالة عند 0.05 وهذا الاستنتاج غير مضمون لأن هذه الدلالة تتأثر بحجم التأثير وحجم العينة لأنه يمكن الحصول على قيم صغيرة جداً لـ P لأحجام العينات الكبيرة بغض النظر عن حجم الاثر أو قوة العلاقة.

- **خدوة المغوبة Meaningfulness fallacy Fallacy:** فى ضوء خدوة

المعنى يعتقد الباحث ان رفض الفرض الصفري يؤكد الفرض البديل وهذا يعكس خطأين. الأول رفض H_0 فى الدراسة لا يعنى تحقق ان الفرض البديل، الثانى حتى لو ان الفرض البديل الإحصائى H_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 الصفري لا يوجد امام الباحث الكثير ليفعله سوى اصدار التوصية.

- **خدوة الصفر وخدوة التكافؤ Zero Fallacy and Equivalence**

Fallacy: يوجد اعتقاد خاطئ بان الفشل فى رفض الفرض الصفري يعنى ان حجم التأثير فى المجتمع يساوى صفر. بينما خدوة التكافؤ تحدث عند الفشل فى رفض الفرض الصفري $H_0: \mu_1 = \mu_2$ فإنه يفسر على أساس ان المجتمعين متكافئين ولكن هذا اعتقاد خاطئ لأنه يمكن أن يحدث اختلاف بينهما فى التباين أو شكل التوزيع.

- **أخدوة الجودة وأخدوة النجاح Quality Fallacy and Success**

Fallacy: الاعتقاد بأن الحصول على الدلالة الإحصائية مؤشر لجودة التصميم التجريبي أو البحثي ويشير أيضاً 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)، وأحياناً الفشل فى رفض الفرض الصفري 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.

- **خدوة الفشل Failure Fallacy:** يوجد اعتقاد خاطئ بان نقص الدلالة

الإحصائية مؤشر لفشل الدراسة فى تحقيق اهدافها ولكن استخدام الطرق التحليلية

غير المناسبة أو القوة الإحصائية المنخفضة تسبب الخطأ من النوع الثانى ويعزو الباحث الفشل فى رفض الفرض الصفري إلى تصميم الدراسة غير الجيد.

• **أخدوة التنقيح Fallacy Reification**: الاعتقاد الخاطئ بان الفشل فى

الحصول على نفس النتائج من دراسات الأخرى لنفس المتغيرات هو فشل فى الوصول لنفس القرار عبر دراسات أخرى، فالنتيجة لا تعتبر قابلة للتكرار لو ان H_0 رفض فى الدراسة الأولى وليس كذلك فى الدراسة الثانية، وهذه الاشكالية تتجاهل حجم العينة، حجم التأثير، والقوة عبر الدراسات المختلفة. ومها اختلفت نتائج الدراسات حول H_0 فإن هذا فى حد ذاته دليل للإعادة Replication .

• **أخدوة الموضوعية Fallacy Objectivity**: الاعتقاد بان اختبارات الدلالة هى

طريقة موضوعية لاختبارات الفروض وكل طرق الاستدلال الأخرى ذاتية هى اضحوة غير مقبولة، فهى ليست الإطار الوحيد فقط لاختبار الفروض بل يوجد التقدير البيزانى Bayesian estimation وهى بديل جيد لاختبارات الدلالة وكذلك فترت الثقة.

• **أخدوة الثائية Sanctification Fallacy**: وتشير إلى اشكالية التفكير

التصنيفى الثائى عن متصل قيم P . فماذا لو $\alpha = 0.05$. وكانت $P = 0.049$ فى مقابل $P = 0.051$ عملياً متمائلين فى ضوء نواتج الاختبار، فالباحث يولى اهتمام للأولى (دالة) ويتجاهل الثانية (غير دالة)، وهذا التغير فى درجة الثقة نتيجة تغير على متصل قيمة P وعالية القيم الكبيرة لـ P تناظر تغيرات صغيرة وغير دالة فى البناء التحتى للمتغيرات. لكن قيم P الأقل من 0.05 مثل 0.048 أو 0.049. فهى توصف بانها غير دالة (Kline, 2013).

• **أخدوة المناعة Robustness Fallacy**: الاختبارات الإحصائية المعملية

البارامترية ليست لديها مناعة أو ضلاعة ضد القيم المتطرفة أو عدم تحقق مسلمات خصائص التوزيع مثل الاعتدالية خاصة للعينات الصغيرة، ولكن كثير من الباحثين يعتقدون العكس والباحثين لا يعطوا مؤشرات على مدى توافر مسلمات الاختبار 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 وأهم جانب آخر من الدقة الإحصائية وهو القوة الإحصائية على الرغم من تأكيد الجمعية النفسية الأمريكية 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 بأنها نوعا من العبث أو عدم التعلم المتدرب عليه Trained incapacity الذى يعوق الباحثين من فهم نتائج دراساتهم، والبعض الآخر يصفها بأنها 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 في المجالات العلمية وتمت مناقشة فكرة الاستبعاد في أعداد أو إصدارات خاصة من مجلات مثل Journal of Psychological and Experimental Education (Thompson, 1993) و Science (Shrout, 1997) وغيرها وكان السؤال الرئيسي: "ماذا يكون الوضع بدون استخدام اختبارات الدلالة الإحصائية؟"، وقدّم (Armstrong 2007) النصيحة الآتية:

الكتابة لكتاب أو تقارير بحوث فلا تذكر اختبارات الدلالة الإحصائية. والكتابة لمجلات فلا بد للبحث عن طرق لتقليل الأضرار المحتملة جراء اختبارات الدلالة، ويجب استبعاد كلمة دلالة لأن النتائج التي ترفض الفرض الصفري ليست بالضرورة دالة وتلك التي تفشل في رفض الفرض الصفري ليست بالضرورة غير دالة.

ونادى بعض الباحثين بهجر أو ترك الدلالة الإحصائية في الدراسات والبحوث أمثال (Schmidt, 1996; Schmidt & Hunter, 1997)، حيث أشاروا إلى أن الدلالة الإحصائية أعاقَت تقدم المعرفة العلمية ولا تقدم إسهامات إيجابية للعلم وهي خطأ رهيب واستراتيجية علمية فقيرة وأحدى الأشياء السيئة التي حدثت في تاريخ علم النفس. وهي استخدام سيئ لفترة طويلة (Cohen, 1994)، بل تقود إلى صناعة قرار بحثي فقير وضعيف.

وقدّمت الجمعية النفسية الأمريكية في إصدارها الخامس والسادس (APA 2001, 2010) عدة توصيات فيما يخص اختبارات الدلالة الإحصائية أهمها:

- اعرض الإحصاء الوصفي مثل المتوسطات والتباينات وحجم العينة لكل مجموعة لأن هذه المعلومات ضرورية لإجراء دراسات لاحقة ثانوية مثل ما وراء التحليل.

- اعرض أحجام التأثير كلما أمكن وعدم عرضه يعتبر قصور في الدراسة.

- من الضروري استخدام فترات الثقة كاستراتيجية لصناعة القرار.

وشدد (Rosnow & Rosenthal 1989) على ضرورة تقدير حجم التأثير عند تقدير للنتائج الدالة إحصائياً وغير الدالة إحصائياً وهذا عكس ما يعتقد الباحثين

من أن تقدير حجم التأثير يكون للنتائج الدالة إحصائياً فقط . وأشار Cohen (1990) إلى أن الناتج الأولي للبحث هي 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 أخرى هي أزمة القياس Measurement crisis فى العلوم السلوكية منذ عهود كثيرة حيث تفتقر البرامج الدراسية للطلاب فى المراحل الجامعية الأولى إلى المقررات السيكمترية 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 بأن 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 وأشار إلى مصطلح الإحصاء الجديدة New statistics وهى ليست إحصاء جديدة على الإطلاق 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 ولأن هذا المؤشر هو إحصائية للعينة فهو تقدير تقريبي 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 الاعداد 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.

والفكير ما وراء التحليل 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 APA (2009) على استخدام فترات الثقة باعتبارها أفضل الاستراتيجيات فى تقرير البحث وأوصت 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 إحصائياً 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 H_0 ومفهومي 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 والحرية للاختبار هى قيم عددية حرة المتريية (بدون وحدات قياس) وليس لها معنى مرتبط بمستويات القياس المرتبطة بالبيانات وهذه تعتبر ميزة للباحثين لاستخدام قيم حرة القياس ولكنها تعطى فهم محدود لكيفية رفض أو عدم رفض H_0 ، وميزة حدود الثقة انها تعطينا هذا الفهم ويتم تقدير حدود الثقة من بيانات أو

إحصاء العينة فبدلاً من حساب القيمة المحسوبة للاختبار فإن الباحث يحسب الفترة في وجود مستوى دلالة إحصائية ألفا وهذه الفترة يعبر عنها بفترة 95% لألفا 0.05 أو فترة 99% لمستوى ألفا 0.01 فبدلاً من التركيز على القيمة الحرجة فإن الباحثين يتجهون إلى الفرض الصفري ويقارن فترة الثقة بالقيمة المتضمنة في الفرض الصفري ($H_0: \mu = 0$) أو ($H_0: \mu = 5$) وهكذا. وقاعدة القرار في الخطوة النهائية هو لو أن القيمة المفترضة في الفرض الصفري (0,5,10) وقعت خارج فترة الثقة فانه يرفض الفرض الصفري H_0 وإذا وقعت داخل فترة الثقة يتم قبول H_0 .

وبشير (2013) Aron et al. إلى أن فترات الثقة (CI) هي مدى من الدرجات بين "الحد الأعلى والحد الأدنى" ومن المحتمل أن تتضمن متوسط المجتمع أو أي مدى من متوسطات المجتمعات الممكنة التي بدورها توجد احتمالية عالية أن تتضمن متوسط العينة. وحدى الثقة هو القيمة الدنيا والعليا لفترات الثقة ويتم عرضها في تقرير البحث كالتالي:

$$(r = 0.27, 95\% \text{ CI } (0.11, 0.42), p = 0.0012)$$

وفي هذا فإن 95% فترة الثقة تتسع من 0.11 إلى 0.42 ولأن الفرض الصفري يساوى صفر $H_0: \mu = 0$ وهذا غير متضمن داخل الفترة وعليه يتم رفض H_0 ، وإذا كان الباحث يختبر فروض حول متوسطات :

$$H_0: \mu_1 - \mu_2 = 0$$

وكانت النتائج: (Mean difference:13 , 95%CI, (-9 , 36) وعلى ذلك القيمة صفر تقع داخل مدى حدى الثقة وعليه يقبل الباحث الفرض الصفري وهذا يعنى أن أعلى فرق بين المجموعات أو المتوسطات هو 36 أو فى فرق فى الاتجاه المعاكس هو -9 وهذا المدى من القيم لفرق المتوسطات يحدث لو أن الدراسة تم اعاذتها 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 أنها الإحصاء

المستخلص من العينة وتستخدم لتقدير معلم المجتمع وهى تقوم عليها اختبارات الفروض ذو الست والتسع خطوات.

ولكن بدلاً من ذلك يمكن أن يستعويض 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 95% Confidence level هو الأكثر استخداماً ويشير إلى ان 95% ان المتوسط يقع بين ذيلى المنحنى. وفترات الثقة تعطى معلومات مثل المعلومات التى تعطىها قيمة p بل وأكثر، 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:

$$CI = \text{Statistic} \pm \text{critical value} * (SE)$$

• Statistic هى أى إحصائية محسوبة سواء متوسط أو تباين وغيرها

• SE الخطأ المعياري للإحصائية.

• Critical value هى القيمة الحرجة المرتبطة بإحصائية الاختبار عند α ، df.

وفى اطار هذا الكتاب نتعرض لكيفية حساب فترات الثقة كثيراً ونعتمد على قيمتها فى برنامج SPSS وعموماً تعتبر فترات الثقة اختبارات للدلالة الإحصائية للفروض الصفرية غير المدخل الكلاسيكى القائم على تقدير النقاط. ويمكن بناء فترات الثقة بدون وجود فرض صفري ولكن لا يمكن إجراء اختبارات الفروض الإحصائية الصفرية بدون الفرض الصفري. ومن النادر جداً 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 القيمة 0.0 (صفر) تقع داخل الفترة فإن التقدير غير دال إحصائياً لمستوى $(1-\alpha)$. وتقدر حدود الثقة باستخدام برامج مثل SPSS, SAS, Minitab. والمؤيدين لاستراتيجية فترات الثقة

يؤكدون على انها تمدنا بمعلومات عن دقة التقديرات وحجم مسافة فترات الثقة تشير إلى الدقة (درجة الخطأ العشوائي 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 واستخداماً من قيمة القوة الإحصائية. وعلى الرغم ان CI حول الفروق تكون اداة لفحص حجم الفروق ويوجد دعم قوى لاستخدام CI اما مكماً أو بديلاً لاختبارات الفروض وفقاً لـ (Borenstein (1994 يوجد ستة خطوات لبناء فترات الثقة كالتى:

1. اختيار مستوى الثقة Confidence level ويشير إلى الاحتمالية إلى ان معالم المجتمع الحقيقى تكون داخل مدى محدود من CI ودائماً يعبر عنها بنسبة مئوية مثل 95% مستوى ثقة و تفترض ان الاحتمالية بان يكون معلم المجتمع الحقيقى داخل فترات الثقة 0.95 .
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 لـ CI 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

وفقاً 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 أخرى مثل استراتيجية Jackknife، واستخدمها (1979) Efron 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 البوتستراب 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 (1993) Efron & Tibshirani المحاولة الأولى لعرض هذه المنهجية وتطبيقاتها إلى مجتمع العلماء والباحثين الاجتماعيين. ونحن لا نعطي 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 البوتستراب مشابه 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 متاح في 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 مقيدتين للنظرية الإحصائية التقليدية وهما:

- مسلمة ان البيانات تتوزع وفقاً للمنحنى الجرسى أو الاعتدالى.
- الحاجة للتركيز على القياسات الإحصائية التى لها خصائص نظرية يمكن تحليلها نظرياً أولاً. وفى هذه الاستراتيجية فإن البيانات الاصلية تستخدم لتوليد عينات عشوائية باستخدام طريقة الاستبدال لإعطاء قواعد بيانات جديدة (Goodd, 1999). وهى أحد الأساليب اللابرامترية 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 البوتسراب يكون تقريبي باستخدام طريقة توليد البيانات لمونت كارلو وهذا التقريب يجعل البوتسراب 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 البوتسراب هى عينة عشوائية بسيطة.

وفيا يلى مثال لتوليد عينة عشوائية بسيطة، بفرض لدينا مجتمع مكونة من ستة أفراد وهى ابجدياً 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}

$$\{A, B, C, E\} .2$$

$$\{A, B, C, F\} .3$$

$$\{A, B, D, E\} .4$$

$$\{A, B, D, F\} .5$$

$$\{A, B, E, F\} .6$$

$$\{A, C, D, E\} .7$$

$$\{A, C, D, F\} .8$$

$$\{A, C, E, F\} .9$$

$$\{A, D, E, F\} .10$$

$$\{B, C, D, E\} .11$$

$$\{B, C, D, F\} .12$$

$$\{B, C, E, F\} .13$$

$$\{B, D, E, F\} .14$$

$$\{C, D, E, F\} .15$$

بالتالى لإتمام عملية توليد البيانات من قاعدة البيانات الأساسية لا بد من استخدام مولد البيانات باستخدام الكمبيوتر. ويفرض تم تحديد ستة عينات مكونة من أربعة أفراد وتم حساب معلم المجتمع من هذه العينات الستة وليكن المتوسط مثلاً.

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 نسبياً فإن عدد العينات العشوائية المتولدة المحتملة من قاعدة البيانات الأولية يكون كبير ايضاً.

عموماً نحن نختار عينة عشوائية لحجم n مع الاستبدال Replacement من المجتمع N ، وفى تطبيقات البوتستراپ فإن المجتمع للمعاينة البوتستراپية لا يكون المجتمع الفعلى موضع الاهتمام ولكن معاينة عشوائية مفترضة من المجتمع. والحقيقة ان متوسط العينات البوتستراپية يكون تقدير غير متحيز لمتوسط المجتمع (Chernick, 2008).

وتتضمن خطوات:

أولاً: بناء مجتمع افتراضى ويسمى **Apseudo population** تماماً كمرأة أو يتماثل مع شكل العينة المتحصل عليها.

ثانياً: 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.

ثالثاً: إجراء التحليل باستخدام الأسلوب الإحصائى لكل عينة مسحوبة.

رابعاً: 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) :

- يحتلظ الأمر لدى مستخدمى اختبارات الفروض حيث النتائج الدالة إحصائياً 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 $P > 0.00000001$ وعلى ذلك فإن عرض النتائج الدالة لا يكون غريب لدى الآخرين فاذا كانت الفروض تافهة فإن الدلالة الإحصائية لا معنى لها.

- النتائج الدالة إحصائياً 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 جداً لا يستحق الوقوف عنده عليك ان تتذكر جيداً ان قوة الاختبار الإحصائى تزيد كلما زاد حجم العينة ولذلك حتى إذا كان الفرق ضئيل جداً ومع أحجام عينات كبيرة فإن الاختبار ينتج دلالة إحصائية.

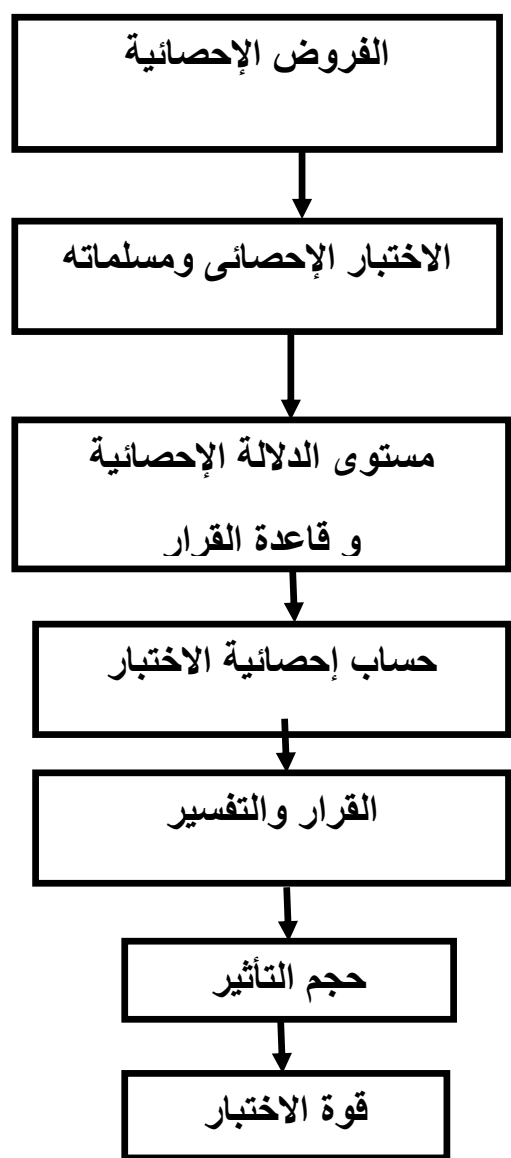
- لا تتجاهل النتائج غير الدالة إحصائياً فالفشل فى رفض الفرض الصفري يبقى على درجة كبيرة من الإثارة، فالفشل فى رفض الفرض الصفري يثير مناقشات وحوارات تكون مدعاة للباحثين لإجراء المزيد من الدراسات خاصة إذا كان هذا يتعارض مع نظرية متماسكة ومتفق على صلاحيتها ومصادقيتها. وعلى ذلك فالنتائج غير الدالة ليست شيئاً سيئاً، 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 H_0 يكون من اهتمام الباحث وعليه فإن الحصول على H_0 ليس أمراً سيئاً طول الوقت.

- لا تحاول استغلال امكانيات برامج الكمبيوتر 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 تخصيصاً "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.

وفى هذا النموذج المكون من سبعة خطوات اعتبرت فيه اولى خطواته هى الفروض الإحصائية ثم الاختبار الإحصائى ومسلماته وتم وضع مستوى الدلالة الإحصائية مع قاعدة القرار، كما تم اعطاء أهمية كبيرة لمفهوم الدقة أو الجودة الإحصائية للاختبار وتم تقسيمها لمكونين هما حجم التأثير وقوة الاختبار الإحصائى وعكس نموذج التسع خطوات عند (2012) Huck الذى استخدم فيه حجم التأثير والقوة الإحصائية فى وسط السلسلة أى فى تصميم الدراسة أو البحث للحصول على حجم العينة غير أننى اعتبرت العكس ان قوة الإحصائية وحجم التأثير فى نهاية السلسلة أو مخرجات اختبارات الفروض وذلك للحكم على الجودة الإحصائية لعملية اختبارات الفرض.

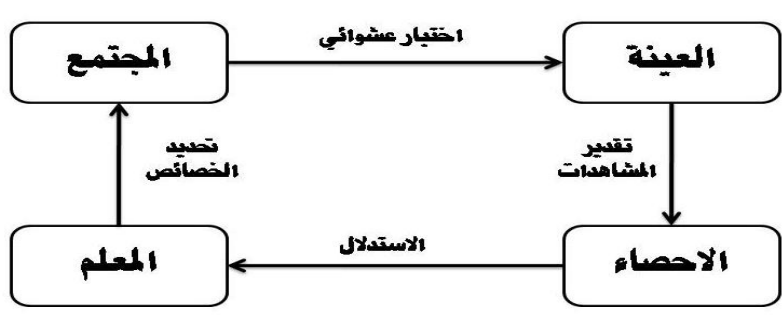
الفصل العاشر

العينات

Samples

يعتبر تحديد حجم العينة على درجة كبيرة من الأهمية وهى من الخطوات الصعبة فى التخطيط للدراسة. وتصميم المعاينة Sampling design وممارسات انتقاء حجم العينة تم مناقشتها فى التراث البحثى بصورة مستفيضه. فالمجتمع هو مجموعة من الأفراد أو العناصر تشترك فى خصائص عامة. فمثلاً 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 اصابوا 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 نظرياً هو مجموعات أو تجمعات من العناصر أو الأفراد أو الأحداث، وتكلفة دراسة المجتمع ككل عملية مستحيلة ومكلفة وتستغرق وقتاً طويلاً واموالاً 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 بآليات 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 في ظل حالات عدم التأكد (الاحتمال) ولفظ إحصائية Statistic يشير إلى المؤشر الإحصائي المحسوب من العينة بينما المؤشر المحسوب من المجتمع يسمى معلم، وعندما يستخدم الإحصاء المحسوب من العينة لتقدير معلم المجتمع فإنه يسمى مقدر Estimator وفيما يلي العلاقة بين العينة والمجتمع في 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 الابتدائي في محافظة ما وعدد التلاميذ في هذا الصف 10000 ، وعندما يكون عدد عناصر المجتمع غير محددة فإنه يسمى مجتمع غير محدد Definite population وليس من الضروري ان تقدير إحصائيات العينة تساوى قيم معالم المجتمع وهذا لا يكون مدعاة للدهشة وذلك لأن إحصاء العينة يأتي 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 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 (2003) Thompson ان العينة في بحوث العلوم الاجتماعية لا تكون ممثلة تمثيلاً جيداً للمجتمع. ولان المعاينة هي عملية 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 هي تلك التي تستخدم نظريات وقوانين الاحتمالات والإحصاء في تعيينها، بمعنى أن العينة الاحتمالية تلك النوع من العينات التي يكون فيها اختيار أى مفردة من مفردات المجتمع عشوائيًا، والعشوائية تعنى ان أى مفردة من مفردات المجتمع لها نفس درجة الاحتمال في 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 فهي سميت بالعينة المقصودة أو العمدية والدراسات التي تعتمد على اختيار العينة عمدًا تعاني من القصور من حيث تعميم نتائجها.

أولاً: العينة العشوائية البسيطة Simple Random Sampe : هي أحد أنواع العينات الاحتمالية. وأساس اختبار مفردات هذه العينة ويمكن في فرض مؤداه أن جميع مفردات المجتمع لها نفس فرص الظهور في 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 استقلالاً إحصائياً. ولاختيار العينة العشوائية البسيطة تقوم بعمل الآتى:

- ترقيم مفردات المجتمع.
- تحديد حجم العينة المناسب.
- اختيار العينة ذات الحجم المحدد عشوائيًا بالآتى :

- طريقة الجداول الإحصائية العشوائية.

- طريقة البطاقات أو الكيس.

- حزم البرامج الإحصائية مثل برنامج SPSS (الحزمة الإحصائية للعلوم 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 في كلية التربية بالجامعة. وبعد الاطلاع على قوائم هؤلاء الطلبة اتضح أن عددهم 500 طالب وطالبة وحدد الباحث عينة حجمها 100 طالب وطالبة أي أن عينة الدراسة 100 ثم قام الباحث بتوزيع أفراد المجتمع من 001 إلى 500 ثم كتب 500 ورقم وعلى كل ورقة اسم كل طالب ووضعها في كيس، ثم سحب أول ورقة فكانت رقم 350 المعروفة أنها لطالب ما من ثم أصبح احتمال كل طالب أن يكون ضمن أفراد العينة $1/500$ ومن هنا فالطالب رقم 350 تم اختياره باحتمال $1/500$ 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 احتمالي ثابت حيث فرص الظهور لكل طالب أن يكون ضمن أفراد العينة متساوية. ولكن إذا لم يعيد الباحث الورقة رقم 350 إلى الكيس مرة أخرى فيصبح أفراد المجتمع 499 وتكون فرص الطالب التالي في العينة $1/499$ من هنا تغير الكسر الاحتمال، في هذه الحالة عينة احتمالية بسيطة بكسر احتمالي متغير. ومن الأفضل أن يعتمد الباحث على عينة احتمالية بكسر احتمالي ثابت.

حقيقة في طرح المثال السابق للتوضيح فقط ويبدو أن طريقة الكيس تبدو غير عملية للمجتمعات الكبيرة ومن الأفضل الاعتماد على الجداول العشوائية أو على الكمبيوتر باستخدام حزمة البرامج الإحصائية SPSS باستخدام أمر 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 متميزة 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 طبقياً عشوائياً 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 الكلى 300 وأنه يوجد ثلاث طبقات متميزة فى المجتمع فإننا سنختار من كل طبقة (اختيارياً عشوائياً 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 نادراً 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 $(n_1) = 2000$

- طبقة طلاب الحضر $(n_2) = 750$

- طبقة طلاب البدو $(n_3) = 250$

لاحظ أن حجم المجتمع: $n_1 + n_2 + n_3 = 3000$

والسؤال الذى يواجهه 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 $n_3 : n_2 : n_1$

هذه النسبة هي 2000 : 750 : 250 أو بالتقريب تصل إلى 8 : 3 : 1

$$1 + 3 + 8 = 12$$

إذاً 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 عشوائياً 67 من طبقة طلاب الريف و 25 من طبقة طلاب المدينة و 8 طلاب من طبقة طلاب البدو.

ثالثاً: العينة العشوائية المنتظمة Systematic Random: نختار العينة فى صورة

عدد متساوٍ بعد بداية محددة عشوائياً وتحديد هذا العدد المتساوى يتم فى ضوء نسبة N/n حيث N عدد مفردات المجتمع و n عدد مفردات العينة المراد انتقاؤها. فمثلاً فى المثال السابق نجد أن المجتمع 3000 وحجم العينة 100 من هنا فإنه يتم اختيار العينة فى ضوء تحديد الفترة المتساوية $30 = 3000/100$ ، ونأخذ بداية فى المدى من واحد (١) إلى (30) وليكن 10 ومن هنا بعد العدد 10 يتم انتقاء المفحوص رقم $40=30+10$ من هنا يتضح العينة للمفردات التى تأخذ الأرقام الآتية: 70 , 40 , 10 .. , 130 , 100، أى يتم تحديد عدد بصورة عشوائية كنقطة بداية ثم بعد فترة محددة متساوية يتم اختيار المفردات الأخرى.

رابعاً: العينة التجميعية Cluster sample: تستخدم لانتقاء عينه عشوائية من

مجتمع كبير جداً أو مجتمع له تنوعات جغرافية مختلفة والهدف العام للعينة التجميعية هو تقليل التكلفة عن طريق زياده كفاءة المعاينة. يكون المجتمع عبارة عن مجموعات

هرمية أو تجمعات من المجموعات وهذه المجموعات تختلف من حيث الحجم والطبيعة، فعلى سبيل المثال مجتمع تلاميذ المدرسة يتم تجمعهم في عدد من الفصول أو عدد من المدارس ويتم انتقاء العينة من الطلاب من هذا المجتمع عن طريق اختيار تجمعات من الطلاب أو الفصول أو المدارس و ليس كما يحدث في العينة البسيطة العشوائية حيث يتم انتقاء مدرسة ما.

القضية مع هذا النوع من المعاينة (على الرغم ان يكون للتجمع نفس الفرصة ان يتم انتقاؤه) هو ان المفردات أو العناصر داخل التجمعات الكبيرة تقل فرصتها ان تكون ضمن العينة النهائية.

وفيما يلي مثال لاختيار عينة عشوائية تجمعية:

المدارس		الأولى		الثانية		الثالثة	
الفصل	فصل 1	فصل 2	فصل 3	فصل 4	فصل 5	فصل 6	
	4	4	4	4	4	4	
التلاميذ							

وجود تحيز Bias فى العينة، وهذه العينة يتم اختيارها مثلاً من مؤسسة تربية 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 فى الدراسة أو معرفتهم لموضوع الدراسة.

3. **العينة Quota sample**: هى الصيغة أو الصورة غير الاحتمالية للعينة الطبقية حيث يتم تمثيل العينة فى ضوء طبقات المجتمع التى يتكون منها ويشار اليها بالعينة الممثلة تمثيلاً جيداً لطبقات المجتمع حيث يتم انتقاء مفردات كل طبقة فى ضوء حجمها ولا يوجد ضمان لفحص دقة التقديرات من كل طبقة.

4. **عينة كرة الثلج أو المتدرجة Snowball sample** تستخدم لتحديد العينة عندما يصعب تحديد المتطوعين الأساسيين للدراسة فمثلاً على سبيل المثال إذا كان من الصعب تحديد العدد المناسب من الاشخاص الصم فإن الشخص الصم المشترك أو الذى تم انتقاؤه يساعدنا فى تحديد اشخاص صم اخرين مستعدين للمشاركة فى الدراسة أى يتم تحديد أفراد العينة من خلال معرفتهم ببعضهم البعض.

مداخل تحديد حجم العينة

العينة هى جزء فقط من المجتمع والتعميم من العينة إلى المجتمع غالباً يتضمن اخطاء وعموماً يوجد نوعين من الالخطاء التى تحدث فى عملية التعميم من العينة إلى المجتمع هما (Dattalo 2008):

- خطأ المعاينة أو الخطأ العشوائى ويكون ناتج عن حظ السحب luck of the draw وهو وجود عناصر كثيرة من نوع واحد وعناصر غير كافية لنوع آخر فكلما زاد حجم العينة فإن الخطأ المنتظم يقل وإذا كان المجتمع متجانس خطأ المعاينة يكون صغيراً.
- 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 خطأ 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 الأدنى.
- تحيز عدم الاستجابة Nonresponsive: إلى أى درجة أن المستجيبين وعدم المستجيبين يختلفون على المتغيرات موضوع الاهتمام وحدوث استقطاب Extrapolation للمستجيبين 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 بأي تقديرات للمعالم، فمعالم عند حدود ثقة 99% تحتاج إلى عينات أكبر من معالم عند حدود ثقة 95%.

وتوجد مداخل عديدة لتحديد حجم العينة وهي:

- مدخل تحليل القوة الإحصائية
- مدخل القواعد المتعارف عليها
- فترات الثقة
- مدخل طرق كميوتيرية معقدة

ويتم عرض هذه المداخل موضعاً نواحي 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 لسهولة استخدامه ومجانية الحصول عليه من خلال الموقع:

[http : WWW.psych.uniduesseldorf.de/aap/projects/gpower/](http://WWW.psych.uniduesseldorf.de/aap/projects/gpower/)

مدخل تحليل القوة الإحصائية

اشار (1979) Cook & Campbell إلى ان صدق الاستنتاج الإحصائي Statistical conclusion validity للدراسة يتحدد بعدة عوامل أهمها تصميم البحث، مستوى الدلالة الإحصائية (الخطأ من النوع الأول)، تباين المجتمع في المتغير المحك (التابع)، حجم الفروق بين القيم الحقيقية للمعلم في المجتمع والقيمة المحددة عن طريق الفرض الصفري (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 ونقل اخطاء المعاينة.

وتعتبر القوة الإحصائية دالة وظيفة للعوامل الآتية:

- مستوى الدلالة الإحصائية (α) أو الخطأ من النوع الأول
- الخطأ من النوع الثاني (β)
- مقدار حجم التأثير المفترض

- الانحراف المعياري أو التباين لحجم التأثير المفترض
- حجم العينة:

والمنطق وراء تحليل القوة هو ان الدراسة لو اعتمدت على حجم عينة كبير كافٍ فإنه توجد احتمالية مسبقة للحصول على تأثير أو فروق إذا كانت موجودة بالفعل وإذا كان حجم العينة صغيرًا فإن الدراسة تفشل في الكشف عن الفروق الموجودة بالفعل في المجتمع وعلى ذلك فالدراسة مضيعة للوقت والجهد والموارد.

ويفيد تحليل القوة الإحصائية في انتقاء حجم العينة في ضوء اسس منطقية بدلا من انتقائها في ضوء القواعد المتعارف عليها Rules of thumb التي ليس لها اسس علمية موضوعية، وفي تحليل القوة قبل إجراء الدراسة لا بد من تحديد حجم التأثير وقيمتي α ، β ، تحديد اتجاه الاختبار، وعلى ذلك يمكن تقدير حجم العينة المناسب.

وفي هذا التحليل يرغب الباحث في تحديد حجم العينة في 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 هي كيفية تقدير حجم التأثير وفي هذا الشأن يوجد ثلاثة مداخل لتقدير حجم التأثير هي:

1. إجراء دراسة استطلاعية ($n=50$) مثلا وتقدير حجم التأثير للاختبار المستخدم.
2. مراجعة الدراسات السابقة وتقدير متوسط حجم التأثير للدراسات التي تناولت نفس متغيرات الدراسة ويمكن إجراء ما وراء التحليل لعدد محدود يتراوح من 5 إلى 10 دراسات للوصول إلى حجم التأثير في الدراسات السابقة.
3. تبنى معايير حجم التأثير للاختبار المستخدم، وهذه المعايير وضعها Cohen (1988)، فمثلاً في حالة 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 يلي مثال لتحديد حجم العينة في حالة استخدام اختبار T باستخدام تحليل القوة القبلي 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 مستقلة ذو ذيلين

2. مستوى الدلالة الإحصائية $\alpha=0.05$

3. القوة المناسبة = 0.8

4. تبني حجم تأثير متوسط = 0.50

وفيما يلي الشاشة الافتتاحية للبرنامج حين يتم تحديد Test family وتتضمن اختبارات عديدة وهي كما هو موضح :

G*Power 3.1.3

File Edit View Tests Calculator Help

Central and noncentral distributions Protocol of power analyses

Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters

Determine =>

Tail(s): Two

Effect size d: 0.5

α err prob: 0.05

Power (1 - β err prob): 0.8

Allocation ratio N2/N1: 1

Output Parameters

Noncentrality parameter δ : ?

Critical t: ?

Df: ?

Sample size group 1: ?

Sample size group 2: ?

Total sample size: ?

Actual power: ?

X-Y plot for a range of values

Calculate

- اختار التحليل القبلي Apriori تحت مربع Type of power analysis حيث يهدف إلى تقدير حجم العينة المناسب.

- اختار من تحت عائلة الاختبارات اختبار t-test

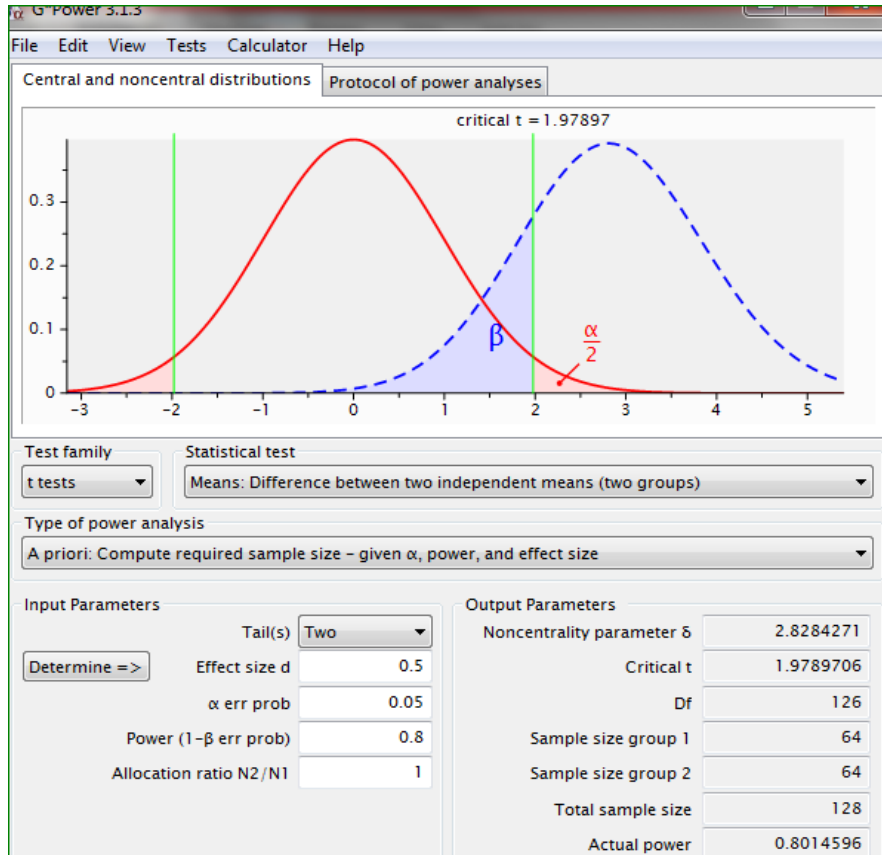
- من تحت Statistical test اختار نوع الاختبار المراد تحدد حجم العينة له وفي هذه الحالة

Means: Difference between two independent means (two groups)

هو :

- ويتضمن البرنامج معالم الإدخال Input parameters قم بإدخالها وهي: تحديد طبيعة اتجاهية الاختبار ذو ذيل واحد أو ذو ذيلين ، حجم التأثير ، قيمة α ، قيمة القوة P ، وتحديد نسبة حجم العينة الأولى N_1 الى الثانية وغالبًا تكون النسبة $N_1:N_2$ وهي 1:1 وتم إدخال المعالم السابقة في البرنامج.

- اضغط Calculate تظهر الشاشة الآتية :



تم اعطاء قيمة T الحرجة = 1.97 و $Df = 126$ وحجم العينة الكلي = 128 مقسمة إلى 64 فردًا في المجموعة الضابطة و 64 فردًا في المجموعة التجريبية وذلك لتحقيق قوة فعلية 0.8014

مدخل القواعد المتعارف عليها Rules of thumbs

قدم (2007) Wilson Vanvoorhis & Morgan إرشادات عملية وإطار لتحديد حجم العينة في ضوء قواعد متعارف عليها ولا بد أن تتضمن الدراسة حجم عينة مناسب لهدفها. ويجب أن تكون كبيرة كبرًا مناسبًا وذلك في محاولة الحصول على دلالة إحصائية. فحجم عينة غير مناسب هو مضيعة للموارد والوقت وحجم عينة كبير غير ضروري وذلك لأنها تحمل

الباحث عبء شديد. غالبًا يتم تجاهل تحديد حجم العينة في ضوء مدخل القوة الإحصائية، ولكن لا بد على الأقل الانتباه إلى القواعد المتعارف عليها، وفيما يلي جدول يوضح ذلك:

الجدول (1.10): حجم العينة في ضوء القواعد المتعارف عليها WilsonVanvoorhis & Morgan (2007)

الاختبار	حجم العينة المناسب
T ,ANOVA	تتضمن الخلية 30 فردا لتحقيق $P=0.8$ وإذا قلت ليس أقل من 7 .
الارتباط والانحدار	50 فردًا على الأقل
χ^2	على الأقل 20. ولا يقل حجم الخلية عن 5
التحليل العاملي	300 فاكتر جيد

وفيما يلي توضيح لذلك:

أولاً: حجم العينة لاختبارات الفروق: حجم الخلية لاختبارات الفروق مثل اختبار T المستقلة و T المرتبطة وتحليل التباين الأحادي ANOVA وتحليل التباين المتعدد وتحليل التباين المتدرج MANOVA ولتحقيق حجم تأثير متوسط إلى كبير يجب أن يكون 30 فردًا في كل خلية لأن هذا يقود إلى قوة إحصائية 80%. وفي بعض الظروف وللبعض الأسباب فإن أدنى حد لحجم العينة 7 أفراد وهذا يؤدي إلى الحصول على قوة 50%، و 14 فردًا لكل خلية ولحجم تأثير 0.50 يؤدي إلى قوة تقريبًا 80%. ويجب الإشارة هنا إلى أن الخلية تمثل المجموعة، في هذه الحالة فإن اختبار T يتضمن خليتين و F يتضمن خليتين فأكثر. ويجب التأكيد على عدة أمور منها:

1. المقارنة بين مجموعات قليلة (2 مثلاً) 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 الضروري توافر حجم عينة كبيراً في كل خلية (Tabachnick & Fidell, 2007).

ثانياً: حجم العينة اللازمة لدراسة العلاقات والانحدار: يوجد العديد من الصيغ المعقدة في هذا الشأن وعموماً يجب استخدام حجم عينة لا تقل عن 50 فرداً للارتباط وأيضاً للانحدار ولكن يجب أن يزيد العدد في حالة زيادة عدد المتغيرات التابعة. وقدم (1991) Green اطارا شاملاً 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 (اختبار T) تستخدم الصيغة الآتية:

$$N > 104 + m$$

وقدم (1985) Harris صيغة تستخدم لتحديد عدد أفراد العينة في تحليل الانحدار لعدد منبئات 5 فأقل فاقترح حجم العينة 50 على الأقل وهو ان العدد الكلى من أفراد العينة يساوى عدد المنبئات بالإضافة الى 50 فرداً:

$$N > m + 50$$

فمثلاً إذا كان تحليل انحدار يتضمن ستة 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 بالنفي 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 تؤدي إلى وفرة في البيانات أكثر من اللازم خاصة إذا كانت عدد المتغيرات المنبئة محدود.

ثالثًا. حجم العينة لأسلوب χ^2 : يستخدم لاختبار الاستقلالية للمتغيرات التصنيفية وفي هذا الأسلوب فإن أفراد العينة يوجد بينها استقلالية والتكرارات المتوقعة الصغيرة في 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 يزداد عن 0.06 (Howell, 2013). والقاعدة الصارمة يجب أن لا يكون التكرار المتوقع في الخلية أقل من 5.

رابعًا. حجم العينة للتحليل العاملي: القاعدة الجيدة للعينة في التحليل العاملي 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.

وأشار (Guadagnoli & Velicer (1988) أن الحلول العاملية مع وجود تشبعات عاليه

للمتغيرات (أكبر من 0.80) لا تتطلب استخدام عينات كبيرة.

مدخل فترات الثقة

الاستراتيجية الأخرى لتحديد حجم العينة هي على أساس الحدود المرغوبة من فترات الثقة (Confidence interval (CI وهو مدى من القيم حول ما إذا كان معلم المجتمع (المتوسط) من المحتمل أن يقع فيه. فعلى سبيل المثال إذا كانت القيمة 0.0 (صفر) تقع داخل الفترة فإن التقدير غير دال إحصائيًا لمستوى $(1-\alpha)$. وتقدر حدود الثقة باستخدام برامج مثل SPSS, SAS, Minitab. والمؤيدين لاستراتيجية فترات الثقة يؤكدون على أنها تمدنا بمعلومات عن دقة التقديرات وحجم مسافة فترات الثقة تشير إلى الدقة (درجة الخطأ العشوائي 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 عند 99% يزيد من دقة حدود الثقة (يمتلك فرص كبرى 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 إحصائي صفرى.

وبعد الدراسة فإن CI يقدم معلومات عن الدقة وتكون أكثر فائدة واستخداماً من قيمة القوة الإحصائية. وعلى الرغم ان CI حول الفروق تكون أداة لفحص حجم الفروق ويوجد دعم قوى لاستخدام CI اما مكملاً أو بديلاً لاختبارات الفروض ويمكن أن تستخدم لتقدير حجم العينة وقارن (1995) Orme & Hudson بين مدخل القوة الإحصائية ومدخل فترات الثقة لتقدير حجم العينة وتوصلا إلى ان استخدام مدخل CI يمدنا بحجم عينة تؤدي 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 محدوداً 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 أيضاً استراتيجيات اعادة المعاينة Resampling strategies وهى متوفرة منذ 1950. ومصطلح الاستراتيجيات المعقدة كمبيوترياً هى تصف مجموعة من المداخل متضمنة طرق Bootstrap ومونت كارلو Monte Carlo methos. وتعتمد استراتيجيات المعاينة العشوائية والمحاكاة الكمبيوترية للحصول على حلول تقاربية للمشكلات الرياضية أو الفيزيائية.

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 وفقاً 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 أخرى فإن تجربة مونت كارلو هي 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 العشوائى

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 عملياته سهلة نسبياً 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 دائماً 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 أيضاً 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 وبالتالي على الخطأ من النوع الأول والقوة الاحصائية وفترات الثقة.

نواحي القوة لاستراتيجيات البرامج الكمبيوترية المعقدة

يرى المؤيدين لهذه الاستراتيجيات كبديل حيوى للإجراءات الاستدلالية الكلاسيكية تتميز بالآتى:

1. الإحصاء الاستدلالي 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 الاستدلالي 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 نفترض الاختبارات التقليدية تساوى التباينات.

2. يشكك بعض الباحثين حول فوائد هذه الاستراتيجيات عند استخدام عينات غير عشوائية وغير ممثلة 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 الشيء من لا شيء " فنفس البيانات تستخدم مره تلو الأخرى.

4. البعض يتساءل عن مدى دقة هذه استراتيجيات فى بعض المواقف فإن هذه الاستراتيجيات أقل دقة من الأساليب البارامترية التقليدية.

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 تمثيلاً جيداً 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 قيوداً 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 يؤدي إلى الوصول إلى استنتاجات متحيزة.

وقدم (1995) Isaac & Michael 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 المنهجي عالى.

- عندما تحتاج إلى الدخول في عمق الظاهرة مثل الدراسات الكيفية من خلال المقابلات.

الفصل الحادى عاشر

القوة الإحصائية وحجم التأثير

Statistical Power & Effect size

تعتبر الإحصاء العمود الفقري لصناعة القرارات والوصول إلى المعلومات والمعارف فى العلوم النفسية والاجتماعية وغيرها. وتقوم نظرية الاستدلال الإحصائي على إستراتيجية اختبارات الفروض التي تقوم على صياغة الفروض الإحصائية وتحديد الاختبار الإحصائي المناسب ومستوى الدلالة الإحصائية ثم صناعة القرار فى ضوء قيمة p حيث إذا كانت $p > 0.05$ نقبل عدم الدلالة الإحصائية (قبول الفرض الصفري H_0) وإذا كانت $p \leq 0.05$ نقبل الدلالة الإحصائية (رفض الفرض الصفري H_0).

ويعتبر (1988) Cohen القوة الإحصائية المؤشر الكمي للحساسية الإحصائية للبحث أو للتجربة، واعتبرها (1981) Sawyer & Ball مظهرًا أساسيًا لصدق الاستنتاج الإحصائي 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 الإحصائي 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 هي التي تصل إلى وجود تأثير نبحث عنه وهو موجود بالفعل في مجتمع الدراسة بأقل قدر من التكلفة والوقت.

مفهوم القوة الإحصائية

يعرف (Pagano 2013) القوة الإحصائية للتجربة بأنها مقياس لحساسية التجربة "Sensitivity" للكشف عن التأثير الحقيقي للمتغير المستقل فنحن نقصد به ذلك التأثير الذي ينتج تغير في المتغير التابع، ولو أن المتغير المستقل لا يؤدي إلى تغير في المتغير التابع فنحن نقول أن المتغير المستقل ليس له تأثير حقيقي. والحصول على تأثير حقيقي يعنى رفض H_0 و لذلك تعرف القوة بأنها قدرة الاختبار الإحصائي على رفض الفرض الصفري إذا كانت للتجربة (المتغير المستقل) لها تأثير حقيقي بالفعل. وعليه فإن الاختبار القوي إحصائياً هو ذلك الاختبار القادر على الحصول على نتائج دالة إحصائياً (رفض H_0) وهي موجودة بالفعل في مجتمع الظاهرة. ويعرف (Howel 2013) القوة بأنها الاحتمالية الصحيحة لرفض H_0 وهو حقيقي مرفوض على مستوى المجتمع (الفرض البديل هو الحقيقي) و لذلك فإن:

$$P=1- \beta$$

والتجربة القوية هي التي تمتلك الفرصة الأفضل لرفض H_0 • وقيمة القوة الإحصائية تتراوح من الصفر حتى الواحد الصحيح ويقال إن التجربة القوية هي التي لها مستوى قوة 0.80 فأكثر. ويرى (Pagano 2013) أن هذا يكون أكثر مرغوبة في البحوث والدراسات ولكن هذا نادراً في العلوم السلوكية والقيم من 0.40 حتى 0.60 هي الأكثر شيوعاً ومن المهم تحديد القوة للتجربة في ضوء:

- **عند تصميم التجربة:** لتحديد حجم العينة ويطلق عليه التحليل القبلي للقوة و التصميم.
- **بعد تحليل النتائج وتفسيرها:** وهذا مفيد في معرفة الدقة الإحصائية للاختبار المستخدم و يعرف هذا بالتحليل البعدي Post hoc procedure ، أى بعد إجراء الاختبار.

وصناعة القرار في نظرية اختبارات الفروض الكلاسيكية تتم بمقارنة قيمة p بمستوى الدلالة الإحصائية (α) وعلى ذلك يتم قبول الفرض الصفري (H_0) أو رفضه (قبول الفرض الإحصائي البديل H_A). وقبول H_0 لا يعنى عدم وجود فروق بالفعل ولكننا لم نمتلك الدليل والوسائل للرفض، وعلى ذلك فإن استخدام الإحصاء الاستدلالي يؤدي إلى الوقوع في أخطاء قرارية على ذلك يوجد خطأين هما: الخطأ من النوع الأول (α) ويتم التحكم في هذا الخطأ قبل إجراء الدراسة من قبل الباحث من خلال تبني مستوى دلالة إحصائية معينة مثل 0.05 أو 0.01 و 0.001 وغيرها. والخطأ من النوع الثاني (β)، وفي الواقع يصعب التحكم في هذا الخطأ قبل إجراء الدراسة ولا يستطيع الباحث التحكم فيه أو السيطرة عليه.

وترتبط القوة الإحصائية بالخطأ من النوع الثاني (β) وتعرف باحتمال تجنب حدوث (β) أو الوصول إلى التأثير الحقيقي باحتمال (β - 1). على ذلك فالقوة الإحصائية مقياس للدرجة التي تكون من خلالها الدراسة قادرة على إظهار الفروق أو العلاقات الحقيقية الموجودة في مجتمع الدراسة (Cohen, 1988)، وتعرف بأنها احتمالية رفض الفرض الصفري المرفوض في المجتمع، بمعنى تجنب حدوث الخطأ من النوع الثاني (β)، أو قدرة الاختبار على الحصول على دلالة إحصائية عندما يكون الفرض الصفري غير حقيقي. فالقوة الإحصائية $0.80 = P$ تعنى وجود احتمال 80% للحصول على دلالة إحصائية وهي موجودة بالفعل بالتالي فإن فرصة ارتكاب (β) تساوى 0.2.

وعلى ذلك فإن دراسات تحليل القوة الإحصائية تجيب على تساؤل ما هي 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 وواقعي والقوة أقل من 0.80 تعنى ثقة منخفضة في مصداقية القرار وعلى ذلك يكون أقصى خطأ من النوع الثاني $\beta = 0.20$ ، ويرى (Verma & Goodale 1995) أن الخطأ α أشد خطورة مقارنة

بخطأ (β) ولتحقيق قوة معقولة لا بد أن تكون نسبة ألفا إلى بيتا هي (4 إلى 1) أى أقصى ألفا يساوي 0.05 وأقصى بيتا يساوي 0.20.

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 كالاتى:

1. حجم العينة (N): كلما زاد حجم العينة قلت أخطاء المعاينة بالتالى تزيد قوة الاختبار وزيادة حجم العينة يؤدي إلى أخطاء أقل وثبات أعلى واحتمالية عالية لرفض الفرض الصفري والحصول على دلالة إحصائية ويشير (1981) Saywer & Ball 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 ، فبدلاً 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 يؤدي 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 الصفري 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 البارامترية 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 في القرار، ومن ثم تصبح عملية صنع القرار أكثر مرونة ومنطقية في رفض الفرض الصفري ويكون بذلك أكثر قوة من الإحصاء الكلاسيكي.

مداخل تحليل القوة الإحصائية

توجد عدة مداخل لتناول تحليل القوة الإحصائية أهمها:

1. قبل إجراء الدراسة Prospective or prior power analysis: في هذا النوع

من التحليل يتبنى الباحث مستوى قوة إحصائية اللازمة للوصول إليها ولتكن $P=0.80$ ، ومستوى دلالة إحصائية معينة مثلاً $0.05 = \alpha$ ، وحجم تأثير مفترض في ضوء تراث الظاهرة أو دراسة استطلاعية. ويكون هدف الباحث تحديد حجم العينة المطلوب لتحقيق قوة إحصائية محددة مسبقاً قبل إجراء الدراسة وهذا التحليل خطوة أساسية في مرحلة تصميم الدراسة وهذا التحليل للقوة يفيد في تقليل التكلفة والموارد نتيجة زيادة حجم العينة بدون داعي ويوصى (Cohen 1992) بأهمية إجراء هذا النوع من التحليل للدراسة قبل إجرائها.

2. بعد إجراء الدراسة Retrospective or Post hoc Power analysis: وهنا

يجرى الباحث تحليل القوة بعد الانتهاء من الدراسة للتقييم البعدي لنتائج الدراسة والتعرف على احتمالية الوصول إلى تفسيرات واستنتاجات صحيحة، ويتم تقدير القوة خاصة للنتائج غير الدالة إحصائياً وذلك لأن عدم الدلالة قد لا يعود لانعدام الدلالة الفعلية في المجتمع (تأثير حقيقى صغير)، بل إن الاختبار الإحصائي له قوة إحصائية ضعيفة غير قادرة على كشف الفروق الموجودة في 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 في مجال تقييم 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 و N ، 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 (1988) Cohen: قدم Cohen (1988) جداول يتم من خلالها تقدير القوة لمعظم الاختبارات الإحصائية مثل T ، χ^2 ، r ، F ، R وغيرها، وذلك بمعلومية مستوى الدلالة الإحصائية وحجم العينة وحجم التأثير.

2. البرامج الإحصائية: ومنها Power & Precision، G-power، NQuery، Stat، PASS، NCSS، وحدد (1997) Thomas & Kerbs الموصفات 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 حسابه 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 (1962) Cohen 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 (1989) Sedlmeier & Gigerenzer 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 التوالي، أى بعد 24 عامًا فإن وسيط القوة أقل قليلاً عما كان عليه فى عام 1962 وهذا يدل على أنه بمرور الزمن فإن قوة الاختبار تقل أو ثابتة نسبياً إلى حد ما.

وقدر (1990) Rossi القوة الإحصائية لـ 25 دراسة منشورة فى مجلة علم النفس غير العاديين 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 (1972) Brewer بتقدير القوة الإحصائية فى ثلاث مجالات 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 التوالي ولمجلة البحث ربع سنوية هي: 0.80، 0.14، 0.52 على التوالي.

وفى مجال التربية العلاجية قدر (1983) Woolley القوة الإحصائية للدراسات المنشورة فى مجلة التربية العلاجية فى المدة من 1980 حتى 1988، وتوصل إلى أن

91 % من الدراسات التي لها حجم تأثير صغيرة تمتلك قوة أقل من 0.50، و 25% من الدراسات التي لها حجم تأثير متوسط لها قوة أقل من 0.5 ، و 4% من الدراسات التي لها حجم تأثير كبير لها قوة أقل من 0.50، ووسط 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 التي لها أعلى قوة إحصائية هى الانحدار يليه الارتباط ثم اختبار T المستقلة.

وتوصل (Casher & Geiger 2004) فى تقييمه 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.

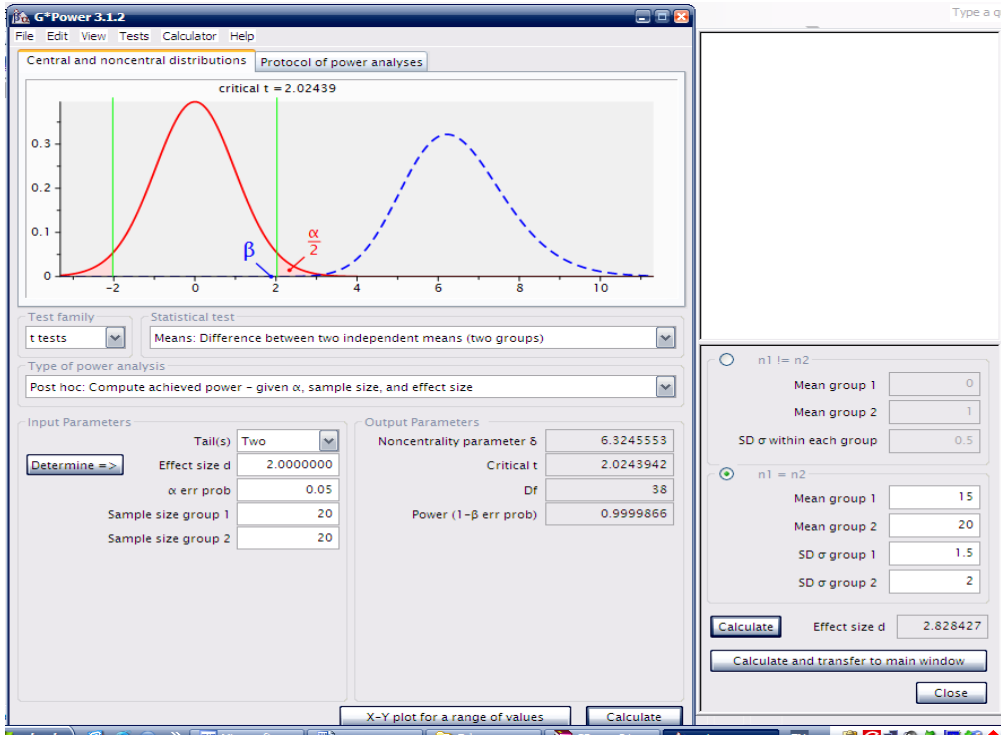
وفى البيئة العربية قام النجار (2006) بتقدير قوة الاختبار فى مجلة الإدارة فى المملكة العربية السعودية وبلغ متوسط القوة لحجم تأثير صغير 0.19، ولحجم تأثير متوسط 0.79، ولحجم تأثير كبير 0.93.

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 معاً في ضوء ثلاثة مستويات من حجم التأثير الضعيف والمتوسط والكبير وذلك في ضوء محكات (1988) Cohen لحجم التأثير، ويلاحظ أن الدراسات التي 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 وهي مستوى الدلالة الإحصائية وحجم العينة وحجم التأثير ولكن حجم التأثير ولم يقدر من قبل صاحب الدراسة وعلى هذا قام الباحث الحالي بالاعتماد على برنامج (Faul, G-Power Erdfelder, lang & Buchner, 2007) لاستخدامه في غرضين هما تقدير حجم التأثير وتقدير قوة الاختبار، وهذا البرنامج يتكون من شاشة بها قوائم ثم مربع أبيض لعرض الرسومات، وبعد ذلك تظهر أيقونة هي 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 الحالي على عائلة T-test وعائلة Exact. وبعد ذلك تم تحديد نوع الاختبار ونوع تحليل القوة، كما في الشكل الآتي:



الشكل (1.11): مخرج برنامج G-Power في 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 وهي في حالة معامل ارتباط بيرسون والانحدار 0.01 صغير و 0.09 متوسط و 0.25 كبير، وفي حالة اختبار (ت) لعينتين مستقلتين 0.2 صغير و 0.5 متوسط و 0.8 كبير ثم أدخلت قيمتي حجم التأثير والقوة إلى برنامج SPSS في ملف البيانات.

والفشل في رفض الفرض الصفري يترك الباحث في حيرة فهل يرجع هذا إلى قوة إحصائية غير كافية للاختبار لرفض الفرض الصفري أم إلى عدم وجود فروق حقيقية للظاهرة؟ ولكن ماذا لو كانت القوة الإحصائية للدراسة أقل من 0.80؟ ربما يرجع هذا إلى أن الفروق الحقيقية موجودة وأن الاختبار ليس لديه القوة الكافية لإظهارها وهذا نتيجة إلى عيوب في انتقاء الاختبار، والتحقق من مسلماته أو القياسات أو في تصميم الدراسة أو غيرها من الأمور المنهجية والإحصائية، وقد يرجع انخفاض القوة إلى أن الفروق الحقيقة صغيرة جدًا في الواقع الحقيقي وبالتالي فهذا ليس عيبًا في 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 شأنه أن يقلل مستوى الخطأ من النوع الثانى (β) ولكن Daniel (1993) يرى أنه حلاً غير مرغوباً كما يجب معرفة حجم التأثير للظاهرة تحت الدراسة من خلال تقدير حجم التأثير قبل إجراء الدراسة عن طريق مراجعة التراث وأيضاً يمكن زيادة القوة من خلال زيادة أحجام العينات لأنها الضمانة الأكيدة لزيادة قوة الاختبار.

ومجال دراسات القوة الإحصائية فى البيئة العربية يكاد يكون محدود جداً 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 مؤشراً 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 دوراً كبيراً 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 إيجابية، فالدلالة الإحصائية تعنى وجود التأثير ولكنها لا تخبرنا بشيء عن حجم هذا التأثير، ومن أهم الانتقادات الموجهة إلى اختبارات الفروض الإحصائية الكلاسيكية:

1. تأثيرها 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.

وإشار (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 التصنيف Classification Fallacy. وهذا القرار التصنيفى غير مناسب للوصول إلى استنتاجات من البيانات.
- كثيرًا من الدراسات تجرى باختبارات لها قوة إحصائية منخفضة وهذا يؤدي إلى الفشل فى أن تصل القيمة الاحتمالية P 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 لـ P 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 عند $\alpha=0.01$ ولكن ليس لها دلالة عملية بمعنى يكون حجم التأثير صغير .

وعدم الدلالة الإحصائية لنتائج الدراسات يعطى انطباعاً عند المحكمين ومحررى المجالات بضعف هذه الدراسات ولذلك يجاهد الباحثون بشتى الطرق للحصول على دلالة إحصائية حتى لو كانت تافهة (الصيد، 1988). فالدلالة الإحصائية تؤدي إلى الحصول على سطحية القرار وتفتقد إلى عمق القرار وهذا ما أدى إلى ظهور المقولة "The end of statistical significance" ودعا Schmidt (1996) إلى هجر وترك اختبارات الفروض الصفرية وإحلالها بحجم التأثير وفترات الثقة فى صناعة القرارات الإحصائية. وعلى الرغم من الانتقادات الموجهة لاختبارات الفروض الإحصائية إلا أنه لا يوجد إطار بديل مقترح لها ونتيجة لهذه الانتقادات أوصت كثيراً من المؤسسات والباحثين باستخدام حجم التأثير Effect size كجزء مكمل لاختبارات الفروض لأنها تمدنا بمعلومات لا تستطيع اختبارات الفروض الكلاسيكية إعطاؤها (APA, 1994; Kirk, 1996; Thompson 1998). وحتى أن الجمعية النفسية الأمريكية APA أوضحت أنه عند استخدام NHST فإن تقدير حجم التأثير وفترات الثقة هو الحد الأدنى المتوقع فى تقرير الدراسات. وعلى الرغم من هذه التوصيات إلا أن استخدام الدلالة الإحصائية استمر مسيطراً على التحليلات الإحصائية فى الدراسات 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 ضوءها 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 جوهري 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 ضوءها على متصل كمي. • لا تتأثر بحجم العينة. • يتم تقديرها قبل إجراء الدراسة وأيضاً 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 تؤدي 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 محدوداً وقد يعود ذلك إلى غياب تدريس هذا الموضوع فى مقررات الماجستير أو الدكتوراة أو نتيجة إلى عدم الاهتمام به فى المجلات الإحصائية أو عدم احتواء البرامج الإحصائية على مؤشرات عديدة لتقدير حجم التأثير.

وان كان استخدامهما محدوداً فى الاختبارات الشائعة مثل اختبارات النموذج البسيط مثل "T" أو "F" أو معاملات الارتباط أو غيرها. فإنها تستخدم بصورة محدودة جداً فى الاختبارات المتدرجة مثل تحليل التباين المتدرج MANOVA.

مفهوم حجم التأثير

أخذ مفهوم حجم التأثير (ES) Effect size 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 التجريبي ولكن المسمى الأكثر انتشاراً 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 مؤشراً 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 كالاتى:

- طريقة لتكثيف فعالية البرنامج أو التجربة على متصل.
- تساعد الباحثون على تفسير النتائج.
- تستخدم قبل إجراء الدراسة في مرحلة التخطيط لترشدنا في 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 (الصيد، 1988) .
- تشكل مكون هام فى مجال نشر البحوث فى المجالات 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 وايضاً 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 إحصائياً وغير الدالة إحصائياً وهذا عكس ما يعتقد الباحثين من أن تقدير حجم التأثير يكون للنتائج الدالة إحصائياً 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 η^2 (ايتا) و ω^2 (اوميغا) 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:** وتتوزع حسب طبيعة البيانات المحللة وأشهرها معامل ارتباط بيرسون
(r) ومعامل ارتباط سيرمان (ρ) ومعامل ارتباط Point-biserial (r_{pb}) ومعامل ارتباط فاى
(ϕ) ومعامل ارتباط كرامير Crammer (V) وأخيراً مؤشر η وهى صيغة تعميمية لـ r
وغيرها.

• **مؤشرات التباين المفسر:** تستخدم فى تصميمات تحليل التباين مثل η^2 و ω^2 و R^2 و
 R^2_{adj} واختبار Interclass (Icc) وغيرها.

3. صنف إلى عائلتين هما:

أ. عائلة d: تتضمن مؤشرات d , Δ , g وغيرها.
ب. عائلة r: وتتضمن مؤشرات r , r_{pb} , ϕ , η^2 , ω^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 الاعتبار الانحراف
المعيارى للبيانات وهى غير مستقلة (حرة) عن مستوى القياس مثل الفروق بين المتوسطات
أو الفروق فى الوسيط أو الفروق فى النسب المئوية وغيرها.

أولاً: مؤشرات حجم الفروق المعيارية بين المتوسطات: وهى تحدد مقدار الفروق بين مجموعتين وهى أكثر ملائمةً للتصميم التجريبي ذى المجموعتين ومن أهم هذه المؤشرات:

1. مؤشر كوهين d لـ **Cohen (1988)**: أكثر مؤشرات حجم الفروق شيوعاً لسهولة تفسيره ومعياريته ويستخدم بكثرة فى دراسات ما وراء التحليل والتصميمات التجريبية وشبه التجريبية ويتحدد بالمعادلة الآتية:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{pooled}}$$

$\bar{x}_1$ متوسط المجموعة التجريبية، $\bar{x}_2$ متوسط المجموعة الضابطة و σ_{pooled} الانحراف المعيارى المشترك للمجتمع ويقدر كالاتى:

$$\sigma_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}$$

حيث n_1, n_2 أفراد عينة المجموعة التجريبية وأفراد عينة المجموعة الضابطة على التوالى و S_1^2, S_2^2 تباين 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 $n_2 = n_1$ كالاتى:

$$d = \frac{2T}{\sqrt{df}}$$

حيث df درجات الحرية وتساوى $n_1 + n_2 - 2$ ، وفى حالة $n_1 \neq n_2$ فإن المعادلة السابقة تصبح كالاتى:

$$d = T \left(\frac{n_1 + n_2}{\sqrt{df(n_1 n_2)}} \right)$$

الصيغتين السابقتين تصلح فى الدراسات التى تذكر قيمة الاختبار T فقط ويمكن تقدير قيمة d فى حالة استخدام اختبار F للمقارنة بين مجموعتين كالاتى:

$$d = \frac{2\sqrt{F}}{\sqrt{df_{error}}}$$

حيث df_{error} هي درجات حرية الخطأ (بين المجموعات) وتساوى $N-C$ حيث N حجم العينة و C عدد المجموعات وتساوى 2. وتوجد إشكالية عند حساب مؤشر d في حالة المقارنات المتعددة المصاحبة لتحليل التباين (توكي 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 وتحدد قيمته بالمعادلة الآتية:

$$\Delta = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_c}$$

حيث σ_c الانحراف المعياري لمجتمع المجموعة الضابطة.

ونلاحظ أن Glass استعان بالانحراف المعياري للمجموعة الضابطة، في حين اعتمد Cohen على الانحراف المعياري المشترك للمجموعتين في المجتمع. ويفترض مؤشر Δ تساوى تباينات المجموعتين الضابطة والتجريبية ($S^2_c = S^2_e$) وفي حالة عدم تساوى فإن مؤشر دلتا متحيز تحيز موجب أى يعطى تقديرًا عاليًا لحجم التأثير .

ولكن هذا المؤشر أقل تطبيقًا في الدراسات التي لا تتضمن مجموعة ضابطة (مثل المقارنة بين الذكور والإناث) (Ferguson, 2009). ويمكن استخدام مؤشر Δ في حالة تصميم المجموعة الواحدة وقياس قبلي وبعدي حيث يستخدم اختبار T المرتبطة ويكون المقام هو الانحراف المعياري للقياس القبلي (Lipsy & Wilson, 2001).

3. **مؤشر هيدجز (g) لـ (Hedges (1981:** عدل Hedges (1981) مؤشر دلتا لـ Glass، حيث تخلص من محدداته واقترح إحلال الانحراف المعياري للمجموعة الضابطة بالانحراف المعياري المشترك لعينيتي المجموعتين وتحدد قيمة g من المعادلة الآتية:

$$g = \frac{\bar{x}_1 - \bar{x}_2}{S_p}$$

وتحدد قيمة S_p كالآتى:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

وبفترض مؤشر g ان توزيع المتغير التابع المتصل اعتدالى التوزيع. ويرى (Durlak 2009) أن مؤشرى d , g متحيزان تحيز موجب عندما تكون أحجام العينات صغيرة ونقح Hedges & Olkin (1985) هذا المؤشر للتخلص من التحيز الموجب خاصة فى حالة العينات الصغيرة كالآتى:

$$g_{adj} = g \left(1 - \frac{3}{4df - 1} \right) = g \left(1 - \frac{3}{4 - 9N} \right)$$

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 العينات 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 تحيزاً من مؤشر Δ عندما $s_1^2 \neq s_2^2$.

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 موجبًا Positively bias خاصة فى حالة العينات الصغيرة (Wang & Thompson, 2006).

- **معامل ارتباط سييرمان (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 تصنيفين (2×2) حيث يستخدم اختبار كا² (χ^2) وتحدد قيمته كالتالى:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

ويستخدم مؤشر ϕ بكثرة فى دراسات المجال الطبى ولكن يجب أن نتعامل معه بحذر ويرى (Tatsuoka (1993 بأنه يمكن التعبير عن حجم التأثير بمربع ϕ :

$$ES(\phi)^2 = \frac{\chi^2}{N}$$

وإذا كانت قيمتها 0.01 فأكثر فهى ذات دلالة عملية.

- مؤشر كرامير (V)Crammer: يستخدم إذا زادت مستويات أحد المتغيرات التصنيفية عن 2 مثل جداول (3×3) أو (2×3) أو غيرها ويتحدد بالآتي:

$$V = \sqrt{\frac{\chi^2}{N(K-1)}}$$

- K الحد الأدنى من الصفوف أو الأعمدة.

مؤشرات تحليل الانحدار: يمكن استخدام مؤشرات معامل الانحدار المعياري β لكل متغير أو معامل الارتباط المتعدد (R) لكل معادلة انحدار و قيمة β دائماً تكون أقل من قيمة مربع معامل الارتباط R^2 (Ferguson, 2009) وكذلك يستخدم معامل الارتباط الجزئي Partial r كمؤشر لحجم التأثير إذا تم عزل تأثير المتغير الدخيل وفي 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 زادت قيمة d تزداد قيمة r . وإذا كانت $n_2 = n_1$ تصبح المعادلة السابقة كالآتي:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

وتقدر قيمة r من قيمة الاختبار الإحصائي T كالآتي:

$$r = \sqrt{\frac{T^2}{T^2 + df}}$$

وتقدر قيمة r من قيمة الاختبار الإحصائي F للمقارنة بين مجموعتين كالآتي:

$$r = \sqrt{\frac{F}{F + df_{error}}}$$

$$df_{error} = N - k$$

- k عدد المجموعات أو مستويات المتغير المستقل.

3. مؤشرات نسبة التباين المفسر: هي المؤشرات التي تقيس نسبة التباين المفسر في 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 في العينة (Richardson, 2011 ويرى (2009) Ferguson أن η^2 هو مؤشر لقوة العلاقة بين المتغيرات مثل مؤشرات 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 (2011) Richardson إلى أن الفرق بين η^2 و r^2 مساوى لمربع معامل الارتباط المتعدد، وفي حالة إذا كان المتغير المستقل بمستويين فإن قيمة η^2 مساوية لمؤشر r_{pb} (Tatsuoka, 1993). ويمكن تقدير مؤشر η^2 من قيمة اختبار F كالآتى:

$$\eta^2 = \frac{(k - 1) - F}{(k - 1)(F + (N - k))}$$

ويرى (2003) Olejnik & Algina 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^2 لاختبار T كالآتى:

$$\eta^2 = \frac{T^2}{T^2 + N - 2}$$

ويري (2009) Ferguson أن الفرق بين مؤشر η^2 و η^2_p ضئيل جداس مع العينات الكبيرة وأحياناً يخطئ الباحثون ويعتبروا ان مؤشر η^2_p هو نفسه مؤشر η^2 وذلك فى مخرج SPSS.

ب. مؤشر (f) Cohen: يقدر نسبة التباين المفسر فى حالة اختبار تحليل التباين الأحادى ويمكن تقدير قيمة f كالاتى:

$$f = \sqrt{\frac{\omega^2}{1 - \omega^2}}$$

ويمكن تقدير مؤشر f من قيمة اختبار F كالاتى:

$$f = \sqrt{\frac{(k - 1)(F - 1)}{Nk}}$$

د. مؤشر اوميغا تربيع (ω^2) لـ (1963) Hayes: يعتبر أكثر مؤشرات قياس نسبة التباين المفسر انتشاراً وذلك فى التصميمات ذات العامل أو المعالجة الواحدة & ANOVA (Monroe, 1996) ومؤشر ω^2 أكثر تفضيلاً من مؤشر η^2 لتقدير حجم التأثير. وعلى الرغم من أهمية مؤشر ω^2 وابسيلون ϵ^2 كتصحيح لمؤشر η^2 إلا أن (Kirk 2004) يرى محدودية استخدامهما فى الدراسات ولكن (Ferguson 2009) يرى أن مؤشر ω^2 و ϵ^2 أكثر استخداماً فى تصميمات ANOVA وأن مؤشر ω^2 مثل مؤشر ϵ^2 يقدر حجم التأثير فى المجتمع ولذلك فهو مؤشر استدلالى عكس مؤشر η^2 وتحدد قيمة مؤشر ω^2 من 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^2 لاختبار T كالاتى:

$$\omega^2 = \frac{T^2 - 1}{T^2 + n_1 + n_2 - 2}$$

ويقدر مؤشر ω^2 من قيمة الاختبار الإحصائي F للمقارنة بين مجموعتين كالاتى:

$$\omega^2 = \frac{(k - 1)(F - 1)}{(k - 1)(F - 1) + Nk}$$

ويعتبر مؤشر ω^2 أقل تحيزاً من مؤشري η^2 و ζ^2 ويمكن حسابه للتصميمات البحثية الأكثر تعقيداً ولكن قيمته تتأثر بمكونات التباين فى هذه التصميمات وتعطى تقديراً صغيراً 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^2 وهى كالاتى:

$$\omega_{partial}^2 = \frac{SS_C - df_c MS_e}{SS_C + (N - df_c) MS_e}$$

حيث df_c درجات الحرية بين المجموعات وتساوى $k-1$.

وقيمة ω_p^2 هى تقريباً نفس قيمة ω^2 فى التصميمات الأحادية العامل ولكنهما يختلفا فى حالة التصميمات المعقدة (Fern & Monroe, 1996).

ثالثاً: مؤشرات تقديرات المخاطرة: يستخدم فى التصميم البحثى الذى يتضمن متغيراً مستقلاً 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 هي تكرارات.

وتعتمد تقدير مؤشرات المخاطرة على نسبة الأفراد في 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 في 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 (Feguson, 2009) r.

2. مؤشنسبة 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 هي إذا كان أحد الخلايا تكراراتها صفر فلا يمكن تقديره (Breaugh, 2003; Durlak, 2009) ومؤشر OR أحيانا مفضل على مؤشرات فروق المتوسطات المعيارية (d, g) 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:

$$RD = \frac{a}{(a + b)} - \frac{c}{(c + d)}$$

هذا المؤشر سهل تفسيره ويعرض الفروق الفعلية فى المخاطرة بين المجموعتين فالقيمة 0.4 تشير إلى ٠.٤ فروق المخاطرة بين المجموعة التجريبية والضابطة (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 يعطى 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 المعياري 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%.

وهذا المؤشر يقى نفسه من عدم توفر الاعتدالية (Breaugh, 2003) ويقدر هذا المؤشر من خلال تقريب اختبار Z كالتى:

$$z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{S_e^2 + S_c^2}}$$

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 وظيفة مرتبطة 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 مقسوماً 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	$0. 8 \geq$
مؤشر f	0.10	0.25	$0.40 \geq$
مؤشر r	0.10	0.30	$0.50 \geq$
مؤشرات التباين المفسر, $R^2_{adj}, r^2, R^2, \omega^2, \eta^2,$	0.01	0.06	$0.14 \geq$

2. معايير اقترحهاFerguson (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 $(r, R, \phi, \rho, \beta)$	0.20	0.50	0.80
التباين المفسر, $r^2, R^2, \omega^2, \eta^2,$	0.04	0.25	0.64
تقديرات المخاطرة (RR, OR)	2.0	3.0	4.0

ووضع (1998) Lippsey معايير لمؤشرات الفروق بين المجموعات وهى القيمة 0.15 صغير و 0.45 متوسط و 0.90 فأكثر و وضع Hinkle, Wiersma, & (1979) Jurs معايير لمعامل ارتباط حيث 0.30 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 (1991) Pedhazur & Schmelkin بأن القيمة العالية لحجم التأثير عند باحث ما فى وقت ما لتخصص ما يختلف عند باحث آخر فى تخصص آخر. 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 أيضاً 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 - ابتدائي - ثانوي - جامعي)، وفى ضوء المحتوى المقاس فتفسير حجم التأثير لمحتوى الرياضيات مختلف عن تفسير حجم التأثير لمحتوى اللغة العربية، ويفسر حجم التأثير أيضاً على أساس المجموعات الفرعية فمثلاً تفسيره لموضوع ما على الذكور يختلف عن تفسيره عند الإناث وأيضاً بالنسبة للموقع الجغرافى وذلك نظراً لوجود فروق بين المجتمعات ويفسر أيضاً فى ضوء التراث البحثى المشابهة للظاهرة موضوع الدراسة وهذا يفيد تماماً 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

علم النفس التجريبي إلى 77% لمجلة علم النفس التطبيقي وأن 60% من أحجام التأثير المستخدمة هو مؤشر R^2 في المجلات الأربعة. ومسح Keselman et al., (1998) استخدام حجم التأثير لـ 226 دراسة في مجال التربية وعلم النفس في 17 مجلة منشورة في الفترة من 1994 حتى 1995 وتبين أن 16 دراسة قدرت حجم التأثير بنسبة 7.1%. وفحص Fiddler et al. (2005) استخدام حجم التأثير في مجلة علم النفس الإرشادي الأكلينيكي في الفترة من 1993 إلى 2001 واستخدمت مؤشرات حجم التأثير المعيارية لاختبارات ANOVA و X^2 وهي على التوالي 32% و 13% و 16%. وفحص Kampense, Dyba, Hannay, & Sjoberg (2007) دراسة في مجال هندسة البرمجيات واشتملت على 429 اختبار وجد أن نسبة توصيف حجم التأثير 29% من الدراسات قدرت حجم التأثير في ضوء مؤشرات معيارية وغير معيارية ونسبة استخدام المؤشرات المعيارية 7.7% وأن نسبة استخدام مؤشرات حجم التأثير لاختبارات ANOVA 3% ولاختبار T 16% ولـ X^2 26% وتوصل إلى أن وسيط مؤشر g_{adj} لهذه الدراسات 0.60.

وحلل (2008) Zientek, capraro & Capraro 217 اختبارًا إحصائيًا في 174 دراسة منشورة قبل 2005 ووجد أن 39% من الدراسات قدرت حجم التأثير وأن 94% من هذه الدراسات اعتمدت على معاملات الارتباط ومعاملات الانحدار كمؤشر لحجم التأثير و 7% استخدمت حجم التأثير في الاختبارات الفارقة مثل مؤشر d. وحلل (2009) Alhija & Levy 183 اختبارًا إحصائيًا في 99 دراسة في مجلات نفسية وتربوية في الفترة من 2003 إلى 2004 و توصلوا إلى أن معظم الدراسات اعتمدت على مؤشرات r و R و X^2 كمقياس لقوة العلاقة (86%) وأن (57%) من الدراسات الفارقة قدرت حجم التأثير باستخدام مؤشرات d و μ^2 و μ_p^2 .

وأجرى (2010) Sun, Pan, & Wang مسحًا لـ 1243 دراسة منشورة في 14 مجلة علمية تربوية ونفسية في الفترة من 2005 حتى 2007 ووجد أن 49% من المقالات قدرت حجم التأثير فقط 57% منها فسرت قيمة حجم التأثير و المؤشر الأكثر استخدامًا هو R. وراجع (2011) McMillan أربع مجلات من إصدارات APA في

الفترة من 2008 إلى 2010 متضمنة 417 دراسة وتوصل إلى أن 74% من الدراسات قدرت حجم التأثير وكانت أكثر ثلاثة مؤشرات استخداما هي R^2 و η^2 و d ويوجد استخدام محدود جدًا لمؤشر Hedges (g) 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 في تحليلات القوة الإحصائية للدراسات.

ويعتبر حجم التأثير جزءًا مكملًا للدلالة الإحصائية إلا أن توظيفها في 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 وأيضاً غياب هذا المفهوم فى المجلات 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 محدودات 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 محدوده 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 دوراً 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 استاتيكية 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 إحصائياً ليست بالضرورة دالة عملياً وهذا يؤكد Nickerson (2001) بوجود خطأ فى تفسير الدلالة الإحصائية فالقيمة الصغيرة لـ P ليست بالضرورة تعنى أن حجم التأثير المقابل لها دال عملياً وكما هو واضح أن 68.1% من الاختبارات الدالة إحصائياً عند $\alpha = 0.01$ ليست دالة عملياً.

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 F , df , MSE 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 إحصائياً يحكم عليها خطأ بأنها ليست لها أهمية عملية وهنا يقع الباحث فى خطأ استدلالى وكذلك النتائج الدالة إحصائياً يفهم 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 R^2_{adj} و ω^2_p و $p\eta^2$ 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 السيكمومترية 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.

- **تضمن 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 دائماً 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 ففي 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 آخر. بافتراض أن حجم العينة = 300 ، وتتضمن حالات غائبة، إذا 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 لآخر وهذا يقود إلى عدم الاتساق الرياضى. وإذا كان حجم العينة كبيرًا كافيًا ويوجد عدد قليل من المفردات بها بيانات غائبة فلا توجد فروق عملية بين الطريقتين ولكن إذا وجد عدد كبير من أفراد العينة لديهم بيانات مفقودة على مفردات المقياس ففي هذه الحالة فإن الطريقة المفضلة هى pair wise ، وفيما يلى مثال للبيانات الغائبة:

الجدول (1.12): مثال لمجموعة بيانات غير كاملة.

Y	X ₂	X ₁	CN ₀
13	8	42	1
12	10	34	2
-	12	22	3
8	14	مفقود	4
7	16	24	5
10	مفقود	16	6
10	مفقود	30	7

- **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 $X_1 = \frac{148}{7} = 21.14$ 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 صغيراً.

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 فى البيانات الموضحة بالجدول السابق يعتبر X_1 و X_2 منبئين بالمتغير Y:

$$\hat{y} = b_1 X_1 + b_2 X_2 + a$$

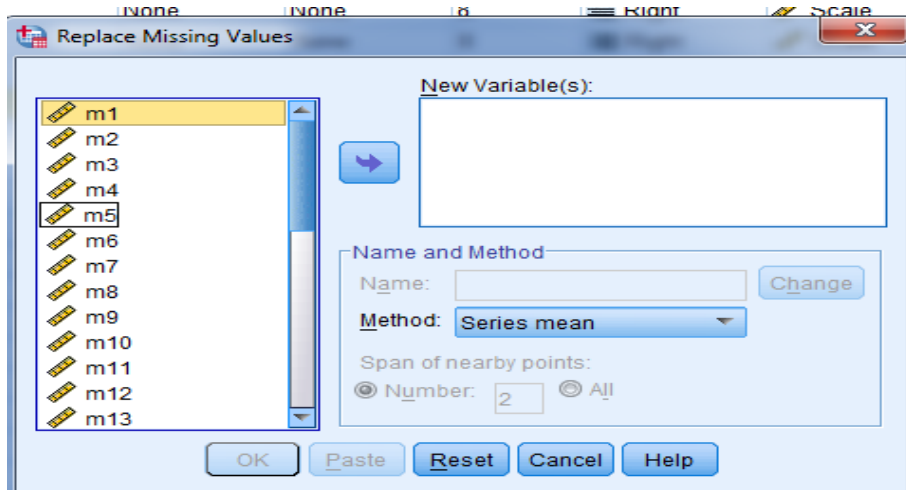
وهكذا بالنسبة لـ X_1 تابع والمتغيرات الأخرى مستقلة $X_1 = b_1 y + b_2 x_2 + a$ واستخدام هذه الطريقة يتطلب معلومات أكثر من طريقة التعويض بالمتوسط. ويمكن استخدام الانحدار اللوجستى Logistic regression إذا كانت المتغير التابع تصنيفى. وعموماً هى أقل مرغوبة مقارنة بطرق التعظيم المتوقع.

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 في ملف البيانات قبله أو بعده.

استبدال البيانات الغائبة في 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 التي 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

في بعض الأحيان تتضمن بيانات العينة قيمة متطرفة أو أكثر على متغير واحد وهي Univariate outlier حيث تختلف تمامًا عن بقية البيانات، وهي بدورها تؤثر على قيمة العلاقات بين المتغيرات وهذا ينعكس سلبًا على تقديرات معالم النموذج نتيجة تأثيرها على العلاقات بين المتغيرات وكذلك على مؤشرات الالتواء والتقرطح. 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 الحقيقي 21 عامًا ويطلق على القيمة المتطرفة في متغير واحد بـ Univariate Outlier. ولا يوجد تعريف واحد للقيمة المتطرفة ولكن القاعدة هي الدرجات التي لها أكثر من ثلاثة أضعاف الانحرافات المعيارية عن المتوسط. ويمكن تشخيصها من خلال فحص التوزيعات التكرارية للدرجات المعيارية Z فإذا كانت $|Z| < 3.0$ فإن هذا يشير إلى وجود قيم متطرفة، بينما القيم المتطرفة المتدرجة Multivariate Quilter وهي درجات متطرفة على متغيرين فأكثر وهذا بدوره يزيد من الوقوع في الخطأ من النوع الأول أو الخطأ من النوع الثاني، 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 المعياري بين مجموعة من الدرجات للحالة الواحدة ومتوسطات كل المتغيرات ، وفي 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) يؤدي إلى رفض الفرض الصفري وعلى ذلك يوجد تطرف في التوزيع، وينصح دائماً باستخدام مستوى دلالة إحصائية (α) عند أدنى قيم مثل $\alpha = 0.001$ وتقوم بعض البرامج مثل SPSS و SAS بطباعة مؤشر D^2 .

2. إذا كان الدرجات المعيارية كبيرة جداً على متغير واحد فأكثر حيث المتغيرات التي لها درجات معيارية تزيد عن 3.29 (اختبار ذو ذيلين $0.001 < P$) فيدل هذا على وجود قيم متطرفة.

3. العرض البياني لبيانات المتغير مثل المدرج التكراري أو Boxplot.

واحد الطرق للتعامل مع الدرجات المتطرفة هو إجراء تحويل للمتغير الذي يتضمن القيم المتطرفة. حيث تؤدي هذه الاستراتيجية إلى جعل توزيع المتغير أكثر اعتدالية

التحقق من الاعتدالية

اما فيما يتعلق بمدخل التحقق من مسلمة الاعتدالية فمنها ما هو وصفى من خلال:

1. عرض مقاييس النزعة المركزية: تساوى قيم المتوسط والوسيط والمنوال يفيد بوجود الاعتدالية.

2. حساب مؤشرى الالتواء والتفرطح: ولتفسير قيمة الالتواء والتفرطح فى ضوء قيم مطلقة لهم أو حدود قطع، أشارت دراسات المحاكاة إلى أنه إذا كانت قيمة الالتواء $SK > 3$ فإنه يوجد التواء شديد، بينما التفرطح أكثر مرونة فالقيمة المطلقة أكبر من 7.0 تشير إلى تفرطح شديد (Chou & Bentler, 1995; West, Finch & Curran 1995). وتوجد آراء أكثر مرونة فيما يخص التفرطح فإذا كانت قيمته أكبر من 8.0، ولكن القاعدة العامة هى إذا كان قيمة التفرطح أكبر من $< (10.0)$ فإنه توجد مشكلة فيما يخص التوزيع، والقيمة 20 فأكثر تشير إلى توزيع يعانى بشدة من عدم الاعتدالية (Kline, 2016).

ويتم التحقق الاستدلالي من مسلمة الاعتدالية من خلال:

اختبار كولوموجوروف – سميرنوف لعينة واحدة

Kolomogorov-smirnov one sample test

الهدف: يستخدم للتحقق ما إذا كانت البيانات الفترية أو النسبية اعتدالية التوزيع فى المجتمع وهى من فئة اختبارات التوزيع المنتظم Test of Distribution . Symmetry ، وفى التوزيع الاعتدالى فإن المنحنى يشبه الجرس، والاعتدالية شرط ضرورى لاستخدام الإحصاء البارامترى وعلى الباحث التأكد من الاعتدالية وفى حالة عدم توافرها فإنه يفضل استخدام الاختبارات اللابارامترية.

اختبارات الفروض لقضية بحثية: جمع الباحث بيانات لمتغير (التحصيل مثلاً) لـ 100 فرد وهى كالاتى:

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 توزيعاً اعتدالياً.
3. متغيرات البحث: المتغير (x) - فترة -تابع فى البحث.
4. النموذج الإحصائي: نموذج المتغير الواحد - إحصاء لابارامترى. والاختبار المناسب كولموجوروف- سميرنوف (K.S) لعينة واحدة.

اختبارات الفروض الصفرية

1. الفروض الإحصائية H_0 : الدرجات اعتدالية التوزيع فى المجتمع
 H_A : الدرجات غير اعتدالية التوزيع

2. الاختبار الإحصائي: كولوموجوروف - سميرونوف (K.S)

3. مستوى الدلالة الإحصائية وقاعدة القرار: لمستوى دلالة إحصائية

$\alpha = 0.05$ وحجم العينة $N = 100$ يمكن البحث فى جداول (كولوموجوروف - سميرونوف) 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 $ F_o(x_i) - S_N(x_i) $
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:

أ. حساب التكرار الملاحظ لكل قيمة Frequency.

ب. حساب التكرار التراكمي التصاعدي Cum. Freq.

ج. حساب قيمة $SN_{(xi)}$ حيث:

$$S_{N(xi)} = \frac{\text{التكرار التراكمي}}{\text{حجم العينة}}$$

• N حجم العينة

• xi القيمة في التوزيع .

وعليه فإن S_N للدرجات 2,1 كالاتى:

$$S_{100}(1) = \frac{24}{100} = 0.24$$

$$S_{100}(2) = \frac{32}{100} = 0.32$$

وهكذا لقيمة التكرار التراكمي المقابل لكل قيمة.

د. حساب الدرجة المعيارية Z وهى درجة معيارية متوسطها صفر والانحراف

المعيارى لها واحد. ولحساب التوزيع التراكمي الاعتنالى لاختبار K.S لعينة واحدة لا

بد من حساب المتوسط:

$$\mu = \frac{\sum X}{N} = \frac{\text{مجموع القيم}}{\text{عددها}} = 3.41$$

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{N}} = 2.458 \quad : \text{ والانحراف المعياري}$$

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 $F_o(x_i)$ عن طريق الكشف في جدول التوزيع الاعتدالي المعياري (الملاحق)، وبالكشف في جداول التوزيع الاعتدالي عن $Z = -0.98$ فإن $P(F_o(x_i)) = 0.1635$ وبالكشف عن $Z = -0.57$ فإن:

$$P(F_o(x_i)) = 0.2843$$

هـ. حساب القيمة المطلقة للفروق بين $F_o(x)$ و $S_N(X_i)$:

$$D = |F_o(x) - S_N(X_i)|$$

وتكون القيمة المطلقة للدرجة 1 ، 2 هي:

$$D_{x1} = |0.1635 - 0.241| = 0.08$$

$$D_{x2} = |0.2843 - 0.321| = 0.04$$

وهكذا لباقي الدرجات وعليه فإن اقصى قيمة مطلقة هي $D = 0.30$ (للسته عشر درجة في الجدول) وهي تعتبر القيمة المحسوبة.

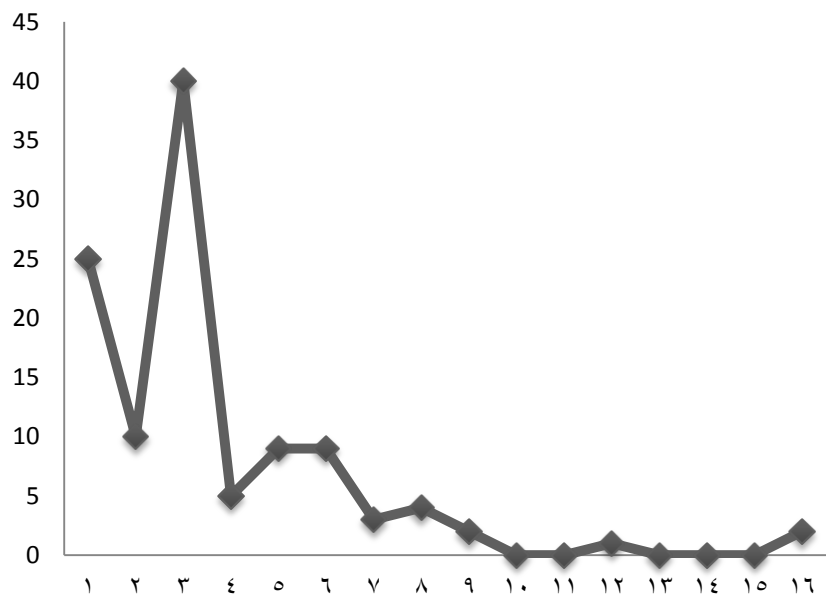
وإذا كان $N > 35$ فإن تقريب العينات الكبيرة يجب أن يستخدم لتحديد الدلالة الإحصائية فلمستوى دلالة إحصائية $\alpha = 0.05$ ، ولحجم العينة $N = 100$ يمكن البحث في جداول (كولوموجوروف - سيمرنوف) عن القيمة الحرجة أو الجدولية فإن:

$$= \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{100}} = 0.136$$

$$0.30 \geq (0.136) \text{ (الجدولية)}$$

وبالتالى فإن القيمة المحسوبة D أكبر من القيمة الحرجة ولذلك نرفض الفرض الصفري 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 البياني الآتى:



والتمثيل البياني 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 وفقاً لـ APA

أوضحت نتائج اختبار $K.S$ للتحقق من الاعتدالية للمتغير:

$D(100)=0.30, P < 0.05$ وعليه فإنه توجد دلالة إحصائية، بالتالى فإن التوزيع غير اعتدالى.

التحقق من اعتدالية البيانات فى SPSS

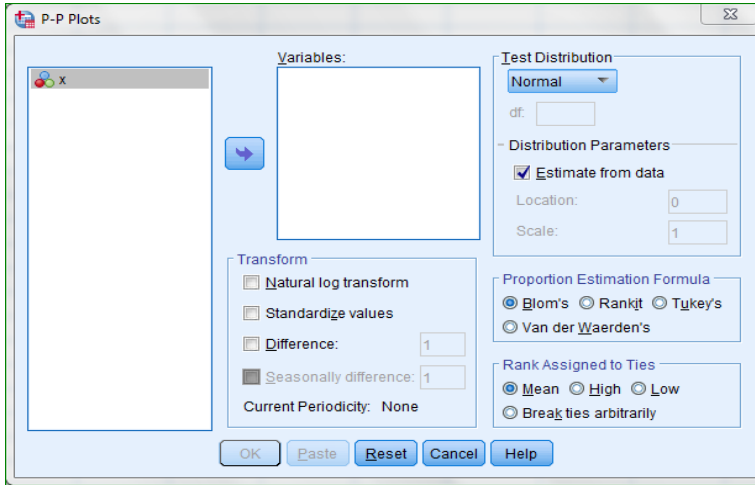
- التحقق من الاعتدالية من العرض البياني:

أولاً: 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 اعتدالي فإن الدرجات Z الفعلية سوف تكون على خط قطري مستقيم. والفكرة في هذا المنحنى هو مقارنة نقاط أو أحداثيات البيانات بالخط المستقيم القطري وإذا وقعت الدرجات على القطر فإن بيانات المتغير اعتدالية والابتعاد عن القطر يدل على الابتعاد عن الاعتدالية.

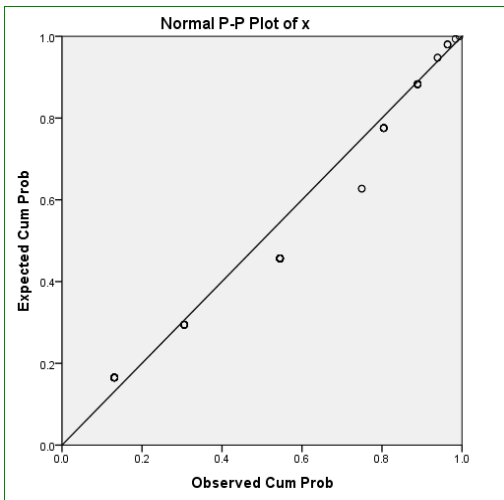
لتنفيذه اتبع الخطوات الآتية:

1 . اضغط Analyze ثم اختر Descriptive Statistics ثم اضغط على p-p.plots تظهر الشاشة الآتية:



2 . انقل متغير X إلى مربع Variables. ثم اضغط على OK.

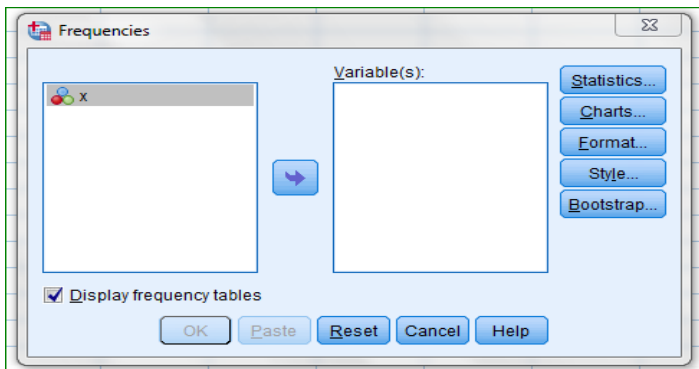
ثالثاً: المخرج:



كما هو ملاحظ أن الإحداثيات أو النقاط لا تقع تمامًا عن الخط القطري وعليه فإن البيانات تبتعد عن الاعتدالية ولا يتوافر فيها هذه المسلمة، وكما هو ملاحظ أن القيم الملاحظة على المحور السيني والقيمة المتوقعة على المحور الصادي. ويتضح تمامًا أن الإحداثيات أو نقاط البيانات لا تقع على الخط المستقيم وعليه فإن البيانات غير اعتدالية.

• المدرج التكرارى Histogram: تنفيذ الأمر:

1. اضغط Analyze ثم اضغط Descriptive Statistics ثم اضغط Frequencies



تظهر الشاشة:

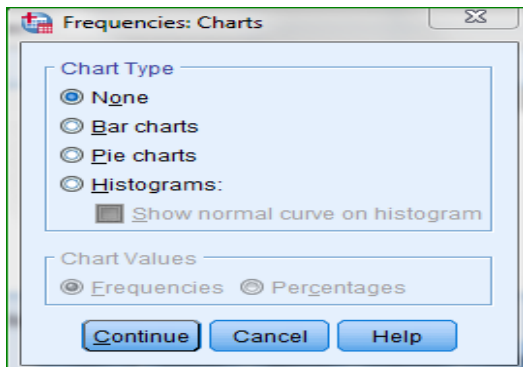
2. اضغط على المتغير x ثم

انقله إلى مربع Variables

3. اضغط على

اختيار Charts على اليمين

تظهر الشاشة الآتية:

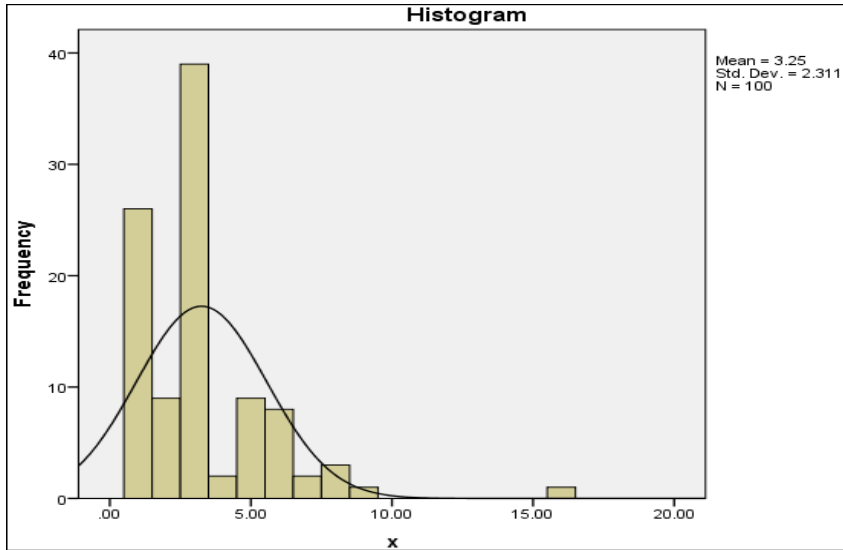


4. اضغط على Histograms و Show with

normal curve

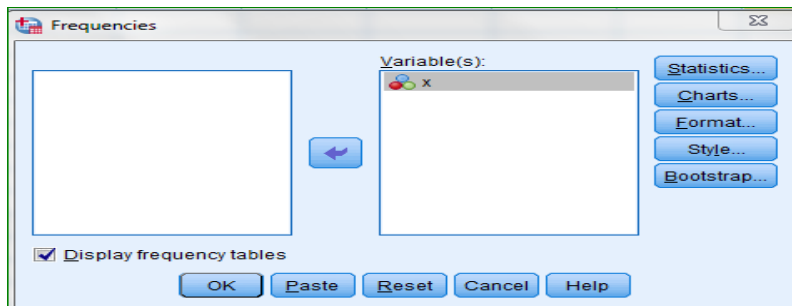
5. اضغط على Continue ثم Ok

المخرج :



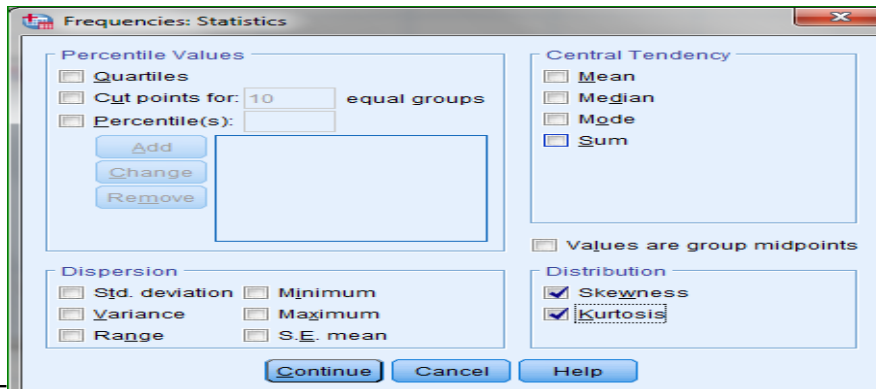
يظهر من المخرج ان المنحنى ملتوى ناحية اليمين بمعنى التواء موجب.
حساب مؤشري التفرطح والتواء باعتبارهما من خصائص التوزيع ويمكن تنفيذها في
البرنامج:

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

Statistics

المخرج:

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 جداً بما يدل على ان التوزيع ليس اعتدالى بل يوجد تفرطح وعليه فإن البيانات غير اعتدالية التوزيع،

ويمكن تحويل قيم الالتواء والتفرطح إلى قيمة Z لاختبار دلالتها الإحصائية لها حيث ان:

$$Z_{skew} = \frac{S}{SE_{skew}}$$

$$k - 0$$

$$Z_{kurt} = \frac{k - 0}{SE_{kur}}$$

حيث S قيمة الالتواء و SE_{skew} الخطأ المعياري للالتواء و K قيمة التفرطح و SE kurt الخطأ المعياري للتفرطح على حدة وعليه فإن :

$$Z_{skew} = \frac{2.164}{0.241} = 8.979$$

$$Z_{kurt} = \frac{8.447}{0.478} = 17.671$$

لاختبار الدلالة الإحصائية للالتواء والتفرطح يتم مقارنة Z_{skew} و Z_{kurt} 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:

$$1.96 < (8.979) \rightarrow Z_{skew}$$

$$1.96 < (17.671) \rightarrow Z_{kurt}$$

وإذا 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 (2014) Field 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 - سميرونوف 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.

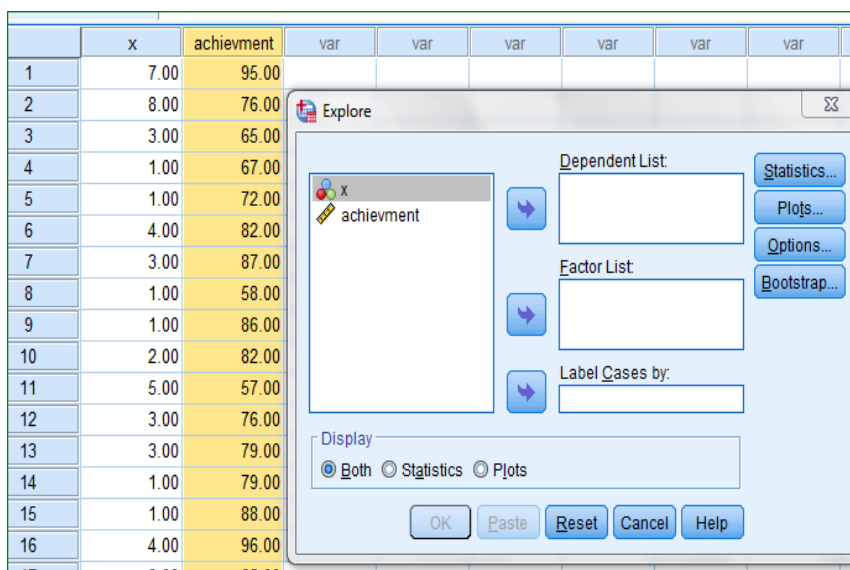
الفروض الإحصائية: الفرض الصفري (H_0): توزيع درجات المتغير المتصل x اعتدالى .
الفرض البديل (H_A) : توزيع درجات المتغير غير اعتدالى.

وعدم الدلالة الإحصائية للاختبارين يعنى توافر الاعتدالية.

لتنفيذ الاختبارين فى برنامج الـ SPSS اتبع الآتى:

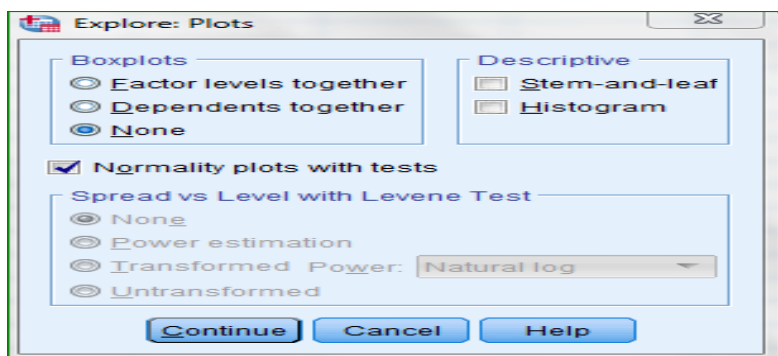
أولاً: إدخال البيانات: 1. اكتب مسمى المتغير X تحت عمود Name ، ثم اضغط على data view

ثانياً: 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:



على اختيار

4. اضغط

Normality Plots with tests وهذا يعطى رسم Q-Q Plot

5. اضغط على Continue ثم اضغط Ok

ثالثاً: 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 df = 100 واتضح ان قيمة الاختبار دالة إحصائيًا حيث : $\alpha(0.05) < P(Sig) (0.00)$ وتم رفض الفرض الصفري على ذلك توجد دلالة إحصائية وعليه فإن توزيع المتغير X غير اعتدالي. وهكذا بالنسبة لاختبار Shapiro-wilk حيث إن إحصائية أو قيمة الاختبار 0.788 و df = 100.

وبالنسبة للقرار بما أن : $(0.05) < (0.00) P(Sig)$ وعليه يتم رفض الفرض الصفري وبالتالي توجد دلالة إحصائية وعلى ذلك فإن توزيع درجات X غير اعتدالي.

وعرض الشكل الآتي:

ويتضح وجود التواء موجب.

الفصل الثالث عشر

معامل ارتباط بيرسون الخطي

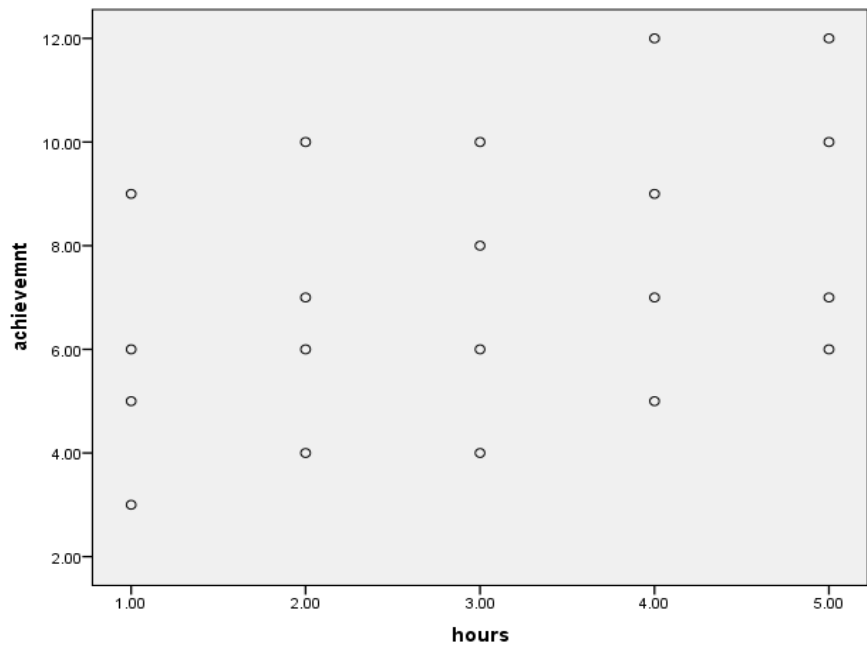
Linear correlation Coefficient Pearson

الارتباط هو أسلوب إحصائي يستخدم لقياس ووصف العلاقة بين متغيرين وقياسات المتغيرين موجودة بالفعل ولا يوجد أي محاولة لضبط أو معالجة المتغيرات، وبيانات هذا الأسلوب تتبع من تصميم البحث الارتباطي، والإحصاء المستخدم لقياس الارتباط يسمى معامل الارتباط $\text{Correlation coefficient}$ وهو قياس كمي 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 $\text{Bivariate correlation}$ ويتبع إحصاء المتغيرين $\text{Bivariate statistics}$ ، وإذا وجدت علاقة Association بين متغيرين يقال انهما مرتبطتين وهذا يعنى يوجد بينها تغير Co-vary بمعنى تغير الدرجات على أحد المتغيرين يتبعه تغير على المتغير الآخر بكلمات أخرى المتغيرين ليس مستقلين. ولا بد التمييز 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 هي جزء من النموذج الخطي العام $\text{General linear model}$.

العرض البياني لشكل الانتشار $\text{Scatter plot or diagram}$

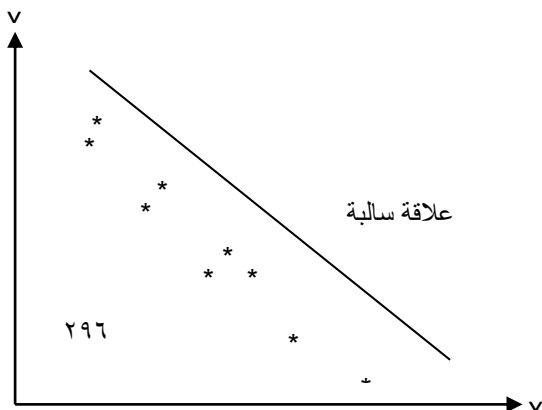
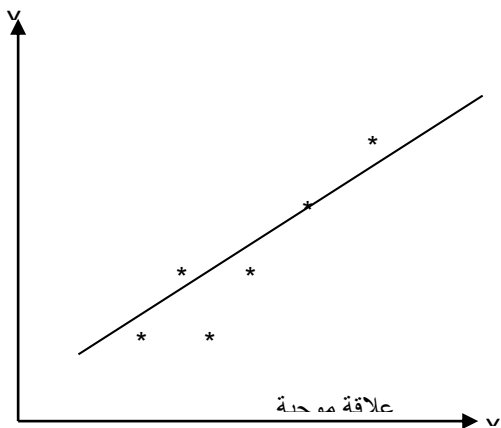
أحد الطرق الشائعة لفحص العلاقة بين متغيرين هي عرضها في ضوء شكل رسم بياني يسمى شكل الانتشار Scatter plot أو يسمى Scatter diagram وهو عرض بياني لبيانات متغيرين على محوري (Y, X) ويستخدم لتلخيص العلاقة بين المتغيرين ويوضع المتغير Y على المحور الصادي ويتم تمثيل زوجي القيم بنقطة وتسمى نقاط البيانات أو الاحداثي يطلق عليها Bivariate point ونمط الشكل يصف قيمة وإشارة

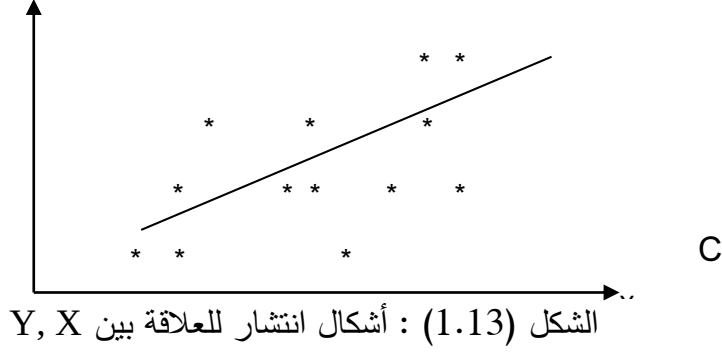
الارتباط و فيما يلى شكل يبين هذا فافترض يوجد بيانات لمتغيرين عدد الساعات الذى يقضيه الطالب فى الاستذكار والتحصيل كالآتى:



B

A





الشكل (1.13) : أشكال انتشار للعلاقة بين Y, X

لاحظ في شكل (A) ان قيم التحصيل Y تزيد كلما زادت قيم الذكاء X ولاحظ أن الاحداثيات في شكل A ، B تتجمع تقريباً 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 تركز 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 وفي العلوم النفسية والاجتماعية إذا كانت درجة العلاقة 0.81 فأكثر فإنها علاقة قوية أو قوية جداً (Howell, 2013). وبجانب أهمية مقاييس الارتباط في تقدير العلاقة بين المتغيرين فإنه مفيد جداً لإجراء التنبؤ Prediction فاذا وجدت علاقة ارتباطية بين متغيرين فمن المحتمل أن يكون أحدهما سبباً للآخر، ولكن الارتباط بين المتغيرين ليس كافياً كأساس لثبوت أو وجود السببية Causality ولكنه يعتبر الخطوة الأولى للبرهنة على ان المتغيرين مرتبطين سببياً. الارتباط والانحدار مرتبطين تمامًا، فالاستفادة من الارتباط هو استخدامه في التنبؤ فلا انحدار بدون ارتباط، فالارتباط يحدد ما إذا كانت توجد علاقة بين المتغيرين ويحدد حجمها واتجاهها بينما الانحدار يهتم أساساً باستخدام الارتباط للتنبؤ .

والاستخدام الآخر للارتباط هو استخدامه لتقدير معاملات الثبات سواء معامل الاستقرار الناتج من الاختبار واعادته Test - 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 تقارى 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:

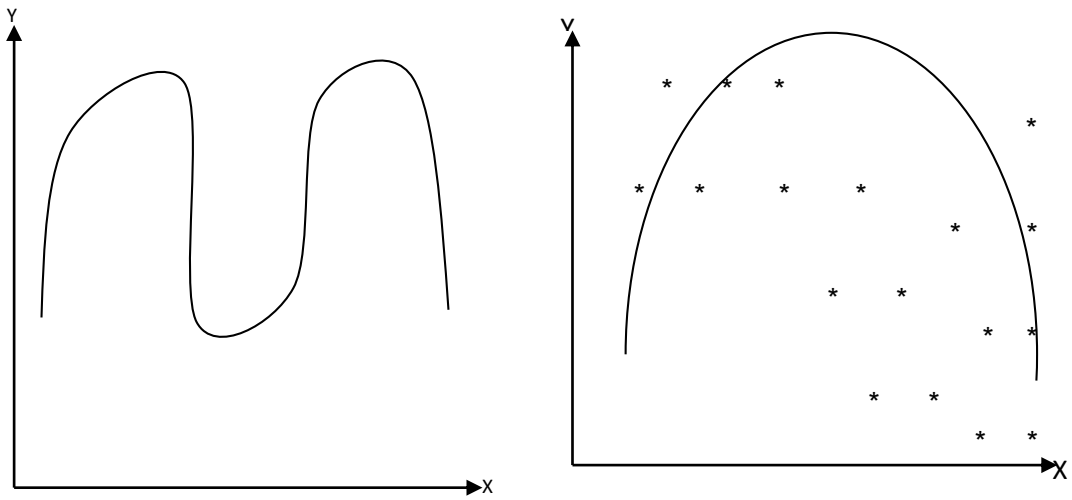
أولاً: اتجاه العلاقة **Direction of the relation** : فإشارة الارتباط سواء كانت:

- موجبة Positive: المتغيرين يتغيروا فى نفس الاتجاه فكلما زاد قيم المتغير X للأفراد فإن قيم المتغير Y تزيد أيضاً وكلما انخفضت قيم X تنخفض قيم Y (انظر شكل A)، بكلمات اخرى توجد علاقة طردية بين المتغيرين.
- سالبة Negative: يشير إلى ان المتغيرين فى اتجاه معاكس فكلما زادت قيم X تنخفض أو تنقص قيم Y وهذه علاقة معكوسة Inverse relationship (شكل B 1.2) بكلمات اخرى توجد علاقة عكسية بين المتغيرين.
- علاقة صفر Zero: 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 تامة Imperfect 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 كالاتى:



الشكل (2.13): شكل الانتشار للعلاقات غير الخطية .

إذا 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 و-1.0 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 (1994) Hinkle et al. معايير لمعامل ارتباط $r = 0.30$ منخفض و 0.70 متوسط و 0.90 فأكثر قوى، بينما وضعها (1988) Cohen القيمة 0.1 ضعيف والقيمة 0.30 متوسط والقيمة 0.50 فأكثر قوى.

معامل ارتباط بيرسون الخطي

الهدف: يقيس درجة أو قوة واتجاه العلاقة الخطية بين متغيرين متصلين (فتري على الأقل) ويطلق عليه Pearson Product – Moment Correlation.

حسابه: توجد صيغ عديدة لحساب معامل ارتباط بيرسون أهمها:

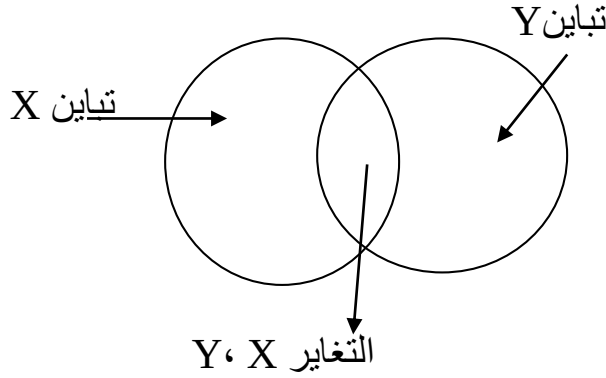
1. التباين **Covariance**: معامل الارتباط للبيانات قائم على إحصاء التباين Cov_{xy} و

هو يعكس درجة تباين للمتغيرين معاً و يمكن حسابة كالآتي:

$$Cov_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N - 1}$$

• $\bar{x}$ متوسط X، $\bar{y}$ متوسط قيم Y

• N حجم العينة.



حيث ان:

$$r = \frac{cov_{xy}}{SD_x SD_y}$$

• SD_x , SD_y الانحراف المعياري لدرجات X , Y

2. من الدرجات المعيارية: تقدر كالتى:

$$r = \frac{\sum Z_x Z_y}{N-1}$$

• Z_x الدرجة المعيارية لـ Z_y الدرجة المعيارية لـ Y

• $\sum Z_x Z_y$ مجموع حاصل درجات الدرجات المعيارية لـ X فى الدرجات المعيارية لـ Y

• N حجم العينة و احياناً يتم القسمة على N (إحصاء وصفى).

3. من الدرجات الخام: يمكن ترجمة معادلة بيرسون من الدرجات المعيارية إلى الصيغة الآتية:

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N(\sum X^2 - (\sum X)^2))(\sum Y^2 - (\sum Y)^2)}}$$

• $\sum XY$ مجموع حاصل ضرب X فى Y.

• $\sum X$ مجموع قيم X ، $\sum Y$ مجموعة قيم Y.

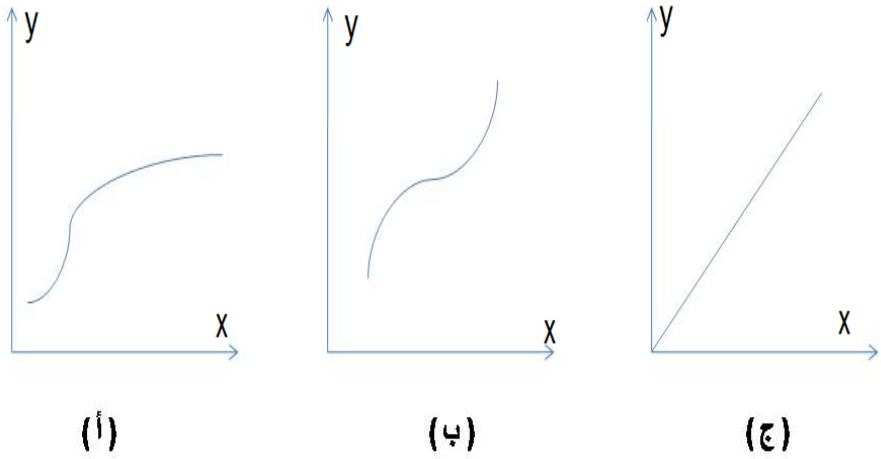
• $\sum X^2$ مجموع مربعات X ، $(\sum X)^2$ مربع مجموع X.

- 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 كالتالى:

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 كالتالى:



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 دائماً غير تامة الثبات ، وذلك

لوجود أخطاء قياس واقعة على المتغيرات. ويمكن التعبير عن العلاقة بين الدرجة الحقيقية والمقاسة كالتالى:

$$T = X \pm E$$

• T هي الدرجة الحقيقية، X الدرجة المشاهدة (المقاسة)، E خطأ القياس.

وكلما كانت T قريبة جداً من X فإن أخطاء القياس تكون أقل ما يمكن. ويعرف الثبات مدى تطابق الدرجة المقاسة مع الدرجة الحقيقية فإذا كانت $E=0$ فإن $T=X$ ، ولكن هذا صعب المنال في قياسات العلوم الإنسانية، وعليه فإن قياساتها تتضمن أخطاء، فإذا كان الثبات $= 0.80$ فإن هذا يفيد بأن 20% من تباين المتغير (الاختبار) يعزى إلى الخطأ. وعليه فإن الثبات المنخفض للقياسات (أخطاء عالية) تؤثر على حجم معامل الارتباط وكذلك معاملات الانحدار ويحدث لمعاملات الارتباط تقلص لقيمتها Attenuation (Kenny, 1987).

3. التجميع الكلى Aggregation: عندما يتم التعبير عن الدرجة في ضوء المتوسط أو تجميع لعدد من المفردات في متغير واحد فإنها تسمى Aggregate data وعموماً فإن الارتباط المقدر من الدرجة الكلية (التجمعية) أكبر من الارتباط المقدر للدرجات المفردة Individual score، وعلى ذلك فإن تجميع الدرجات المفردة في درجة كلية يؤدي إلى تضخم Inflation لقيمة معامل الارتباط لأن ثبات مجموع المفردات معاً أكبر من ثبات كل مفردة على حدة.

4. ارتباط الكل بالجزء Part – Whole Correlation: الارتباط يتضمن متغيرين ولكن في بعض الأحيان أحد المتغيرين مشتق من الآخر أو متضمن في الآخر وعلى ذلك يوجد تباين مشترك كبير بينهما.

الجدول (1.13): ارتباط الكل بالجزء.

المتغير الأول	المتغير الثاني	التحيز في 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. **نوع مستوى القياس للمتغيرات:** يوجد أربعة مستويات قياس شائعة وهى الاسمى، الرتبى، الفترى، والنسبىولا ينصح التعامل مع مصفوفة ارتباط أو تغاير بها خليط من مستويات القياس. وتتنوع معاملات الارتباط حسب مستوى القياس الذى تتعامل معه 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 رتبى
3	فاى (φ)	كلًا من المتغيرين اسمى بمستويين
4	Point – bi-serial	احد المتغيرات فترى (التابع) والآخر المستقل تصنيفى
	الثنائى السلسل (r _{pb})	بمستويين
5	التسلسلى الرتبى (جاما)	احد المتغيرين رتبى والآخر اسمى
	Gamma Rank bi-serial	
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	Tetra- choric	كلا من المتغيرين تصنيفى بمستويين (0, 1) مع بناء تحتى متصل وتوزيع اعتدالى
9	Poly-choric	كلا من المتغيرين ترتيبي مع بناء تحتى متصل أو تصنيفى بأكثر من مستويين مع بناء تحتى متصل

6. **المدى للمتغيرات:** لا بد أن تكون المتغيرات لها مدى معقول لأنه لو كان مدى الدرجات مقيد (صغير) فإن مقدار قيمة الارتباط تنقلص deflation، وذلك لأن عدم وجود تجانس بين الدرجات يؤدي إلى تباين أق لوبدوره يؤثر على الارتباط ويحدث ذلك للمتغيرات التصنيفية وأحيانًا الرتبية. وتقلص المدى يؤثر بدرجة كبيرة على معاملات الارتباط، ولذلك فإنه يفضل استخدام معاملات الانحدار كبديل لمعاملات الارتباط لأن معاملات الانحدار لا تتأثر بتقلص المدى وهذا ينطبق على المتغيرات التصنيفية ذات البناء التحتى المتصل.

7. **القيم المتطرفة Outliers:** القيمة المتطرفة هى التى تبتعد بدرجة كبيرة عن المتوسط والوسيط، وجود القيم المتطرفة فى البيانات يؤدي إلى تشوية حجم الارتباط،

ويمكن أن يكون معامل الارتباط خادع Spurious وهذا يرجع إلى سلوكيات غير طبيعية تظهرها العينة بمعنى وجود بيانات متطرفة لحد أفراد العينة ففي المثال الآتي:

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

في الجدول توجد قيمة متطرفة في المتغير X وهي 10، وعلى ذلك فإن قيمة معامل ارتباط بيرسون $r = -0.271$ ولكن إذا استبدلت القيمة 10 بالقيمة 3 مثلاً فإن $(r = 0.572)$ ، ونلاحظ هذا التغير الدراماتيكي من $(X=10)$ إلى $(X=0.572)$ ، وعلى ذلك فإن القيم المتطرفة تؤدي إلى تقلص قيمة معامل الارتباط وكذلك إشارته أو العكس أحياناً (Kenney, 1987).

اختبارات الفروض لقضية بحثية (بيانات من: Hinkle et al., 1994)

اراد باحث دراسة العلاقة بين درجات اختبار الاستعدادات والأداء الأكاديمي لعينة مكونة من 15 فرداً وكانت البيانات كالاتي:

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

38	440	11
50	515	12
37	380	13
42	510	14
53	564	15
$\bar{Y}=51.49$	$\bar{X}= 534.00$	
$S_y=10.11$	$S_x= 96.53$	

الخطوات البحثية:

1. سؤال البحث: هل توجد أو ما طبيعة العلاقة الارتباطية بين الاستعداد و الأداء الأكاديمي؟
2. فرض البحث: توجد علاقة ارتباطية بين الاستعداد والأداء الأكاديمي.
3. منهج البحث: المنهج الارتباطي.
4. متغيرات البحث: في معامل ارتباط بيرسون ليس من الضروري تحديد أى المتغيرات مستقلة وإيهما تابع ولكن في ضوء الاطار النظرى: الاستعداد (فترى – متصل – مستقل) ، الأداء الأكاديمي (فترى – متصل – تابع).
5. النموذج الإحصائي: إحصاء النموذج البسيط أو Bivariate statistics والأسلوب الإحصائي 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 كالتالى:

1. صياغة الفروض الإحصائية:

- الفرض الصفرى: $H_0: \rho = 0$

حيث ρ معامل الارتباط بين المتغيرين في 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 كالآتي:

$$H_0 : \rho \leq 0$$

$$H_A : \rho > 0$$

2. الاختبار المناسب ومسلماته: أسلوب معامل الارتباط بيرسون ومسلماته كالآتي:

• الاعتدالية Normality: 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 في المجموعة التي ينتمي إليها.

• المتغيرين من مستوى القياس الفترى على الأقل أى أن البيانات متصلة.

3. مستوى الدلالة الإحصائية وقاعدة القرار: حدد الباحث $\alpha = 0.50$ ، درجات حرية:

$$df = n - 2 = 15 - 2 = 13$$

والاختيار ذو ذيلين ولتقييم الدلالة الإحصائية للارتباط فإنها تقدر خلال:

المدخل الأول: من خلال اختبار T وفي هذه الحالة تكون صيغة اختبار T للارتباط كالآتي:

$$T_r = \frac{\text{Sample Statistics}(r) - \rho}{\sqrt{\frac{(1-r^2)}{(n-2)}}}$$

$$T_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

حيث: $\rho = 0$ وهى المحددة فى الفرض الصفرى، r معامل ارتباط بيرسون المقدر.

و درجات الحرية لاختبار T هى:

$$df = n-2=15-2 = 13$$

وتكون هذه قيمة T المحسوبة ويتم مقارنتها بـ T الجدولية عند $\alpha = 0.05$ ، $df = 13$ ، ولاختيار ذو ذيلين تكون قيمتها الحرجة (الجدولية) $= 2.16$. وإذا كانت:

T المحسوبة $< T$ الحرجة نرفض الفرض الصفرى H_0 .

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 قيمة r الجدولية $= 0.514$ ، وعليه إذا كانت r المحسوبة $< r$ الجدولية نرفض H_0 .

المدخل الثالث: مقارنة قيمة p الاحتمالية (فى مخرج SPSS) بقيمة α حيث إذا كانت:

$P \leq \alpha$ نرفض H_0

4. الحسابات: يتم حساب قيمة r من الصيغة المعيارية كالتى:

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

$$X^- = 53.4, y^- = 51.47$$

$$S_x = 96.53$$

$$S_y = 10.11$$

حيث Z_x, Z_y هي الدرجات المعيارية وتقدر كالتالي:

$$Z_x = \frac{X - X^-}{S_x}$$

فمثلاً الدرجة المعيارية لـ 595:

$$Z_{X595} = \frac{595 - 534}{96.53} = 0.63$$

$$Z_y = \frac{y - y^-}{S_y}$$

$$Z_{68} = \frac{68 - 51.47}{10.11} = 1.64$$

وعليه فإن معامل ارتباط بيرسون يقدر كالتالي:

$$r = \frac{\Sigma Z_x Z_y}{n-1} = \frac{12.67}{14} = 0.90$$

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:

$$r_{xy} = \frac{N \sum xy - N \sum x \sum y}{\sqrt{(N(\sum x^2 - (\sum x)^2))(\sqrt{N(\sum y^2 - (\sum y)^2))}}$$

$$= \frac{15(424580) - (8010)(772)}{\sqrt{(15(4407800) - (8010)^2)(15(41162) - (772)^2)}} = 0.90$$

وكذلك تقدير معامل ارتباط بيرسون من خلال التباين كالتالي:

$$\text{Cov} = \frac{\sum(x-x^-)(y-y^-)}{n-1}$$

وعليه لا بد من حساب مجموع حاصل ضرب انحراف القيمة X عن متوسطها في انحراف القيمة Y عن متوسطها وبحسابه كان ناتجه:

$$\sum(x-x^-)(y-y^-) = 12.332$$

$$\text{Cov} = \frac{12.332}{15-1} = 880.86$$

$$r = \frac{\text{cov}xy}{sxsy} \quad \text{إذا :}$$

sy , sx الانحراف المعياري لـ x ، y إذا :

$$r = \frac{880.66}{(96.53)(10.11)} = 0.90$$

ولحساب قيمة T من 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 بـ T الحرجة حيث: T المحسوبة () < T الحرجة

(2.16)، إذا يرفض H0 وعليه توجد علاقة ارتباطية بين درجات الاستعداد والأداء الأكاديمي

وهي دالة إحصائية عند 0.05.

ب. بمقارنة قيمة r المحسوبة ب r الحرجة: r المحسوبة ($.90$) r الحرجة ($.514$) وعليه يرفض H_0 .

6. **حجم التأثير:** لحساب نسبة التباين المفسر يتم تربيع معامل الارتباط ويسمى معامل التحديد Coefficient of determination وقيمته تتحصر من 0.0 الى 1.0 وإذا فإن حجم التباين المفسر للبيانات:

$$r^2 = (0.90)^2 = 0.81$$

وعليه يمكن استنتاج أن درجات الاستعدادات فسرت 81% من تباين الأداء الأكاديمي 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 الآتي:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. تحت Type of power analysis اختيار:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

3. تحت Test family اختيار Exact

Test family	Statistical test
Exact	Correlation: Bivariate normal model

4. تحت Statistical test اختيار:

5. ادخل المعالم الآتية تحت Input parameters:

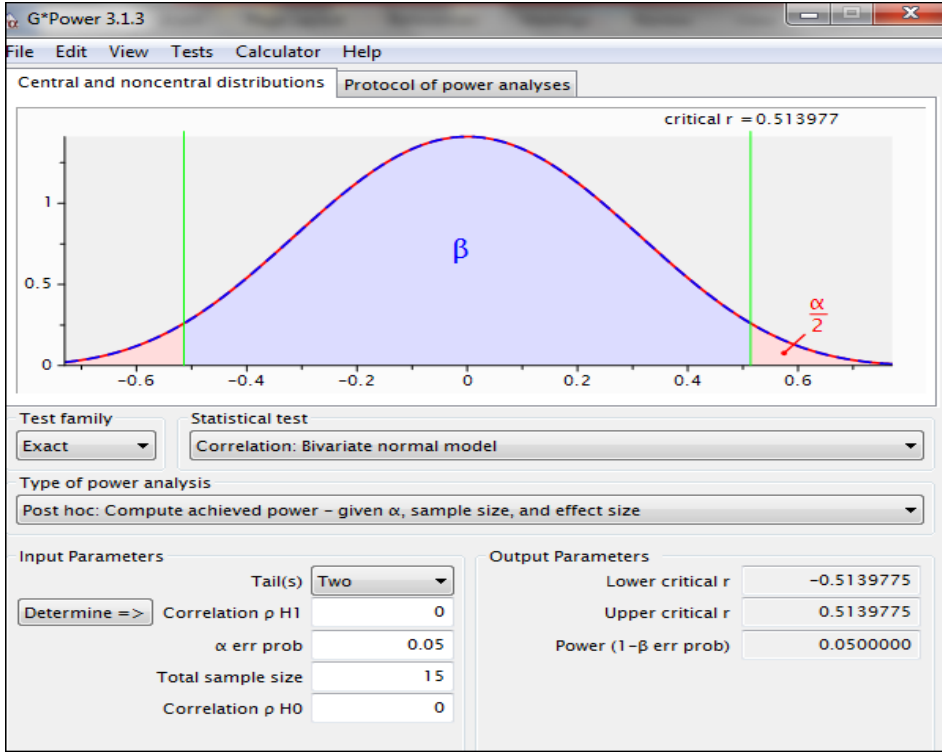
Input Parameters	
Tail(s)	Two
Determine => Correlation p H1	d
α err prob	0.05
Total sample size	15
Correlation p H0	0

- نوعية الاختبار: ذو ذيلين

- مستوى الدلالة الإحصائية: 0.05

- حجم العينة الإجمالي = 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 رديء وجودة بيانات منخفضة وحجم عينة صغير.

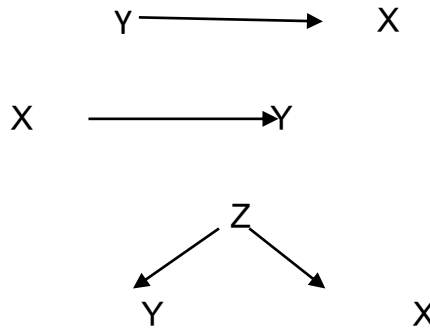
الارتباطية والسببية

إذا وجدت علاقة خطية دالة إحصائية بين متغيرين فيمكن افتراض وجود حد أدنى للسببية بينهما ولكن تفسير الارتباط ليس سببياً لأننا لا نعرف اتجاه السببية بينهما Direction of causality ، ويوجد ثلاثة احتمالات ممكنة لاتجاه السببية بين X , Y هي:

1. أن المتغير x (مفهوم الذات) يسبب المتغير Y (التحصيل).

2. كلما زاد التحصيل (Y) يسبب زيادة في 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 كالآتي:



الشكل (4.12) : ثلاث احتمالات للعلاقة السببية بين Y,X.

وعموماً الارتباط لا يعنى السببية:

Correlation Does Not imply causation.

Correlation not proof cause - effect relationship between the two variables

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 التي تتحقق من العلاقات السببية من خلال الارتباطات كما في تحليل الانحدار وتحليل المسار ونموذج المعادلة البنائية.

معامل الارتباط المصحح Adjusted r (radj)

كما نعلم ان معامل الارتباط في العينة هو تقدير غير متحيز لمعامل الارتباط في المجتمع rho (p) ولكن عندما تكون العينة أو عدد الدرجات أو الملاحظات صغيرة فإنه يعتبر تقدير متحيز لمعامل الارتباط في المجتمع ولتصحيح معامل الارتباط من صغر حجم العينة قدم Howell (2013) معادلة تصحيحية كالآتي:

$$r_{adj} = \sqrt{1 - \frac{(1-r^2)(N-1)}{N-2}}$$

وإذا كان حجم العينة كبيراً فلا نتوقع فروق كبيرة بين r , r_{adj} .

فإذا كان $r_{yx} = 0.529$ وحجم العينة $N = 107$ فان:

$$r_{adj} = \sqrt{1 - \frac{(1-(0.529)^2)(106)}{107-2}} = 0.522$$

على ذلك الفروق بسيطة جداً.

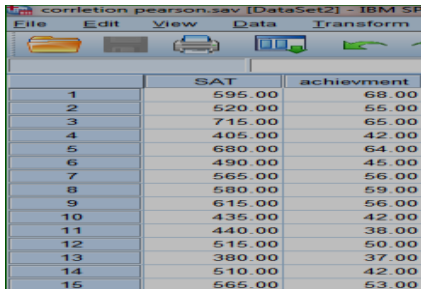
كتابة نتائج معامل الارتباط في تقرير البحث وفقاً لـ APA

تم حساب معامل ارتباط بيرسون بين الأداء الأكاديمي والاستعدادات $r(13)=0.90$, $P < 0.01$, $n=15$, حيث 13 هي درجات الحرية (2-15). وعلى ذلك تم تحديد قيمة معامل الارتباط مقروناً بحجم العينة ودلالاته الإحصائية لاختبار ذو ذيل واحد أو ذو ذيلين، وإذا وجدت متغيرات عديدة يفضل عرضها في مصفوفة ارتباط كالتالي:

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



	SAT	achievement
1	595.00	68.00
2	520.00	55.00
3	715.00	65.00
4	405.00	42.00
5	680.00	64.00
6	490.00	45.00
7	565.00	56.00
8	580.00	59.00
9	615.00	56.00
10	435.00	42.00
11	440.00	38.00
12	515.00	50.00
13	380.00	37.00
14	510.00	42.00
15	565.00	53.00

تنفيذ معامل ارتباط بيرسون في SPSS

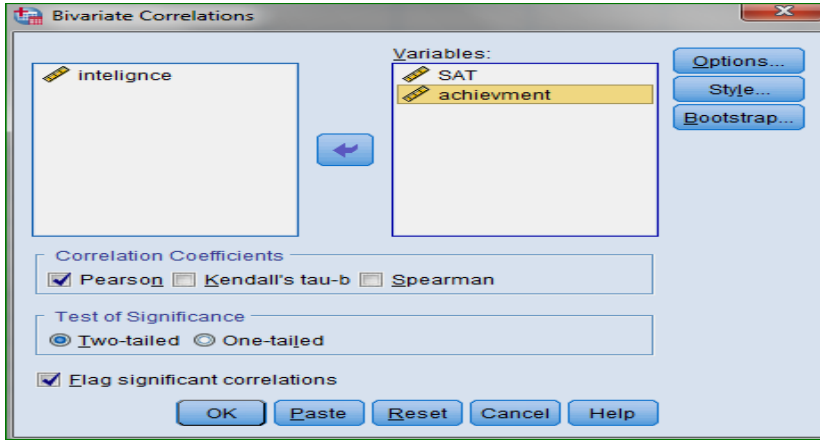
أولاً : إدخال البيانات: 1. اضغط Variable view

2. اكتب مسمى المتغيرات تحت عمود Name:الصف

الأول: الاستعداد Sat والصف الثاني: الأداء الأكاديمي achievement

3. اضغط data view وادخل المتغيرين في عمودين. وفيما يلي شاشة البرنامج:

ثانيًا: تنفيذ الأمر: 1. اضغط على Analyze ثم اضغط Correlate ثم اضغط Bivariate تظهر الشاشة:



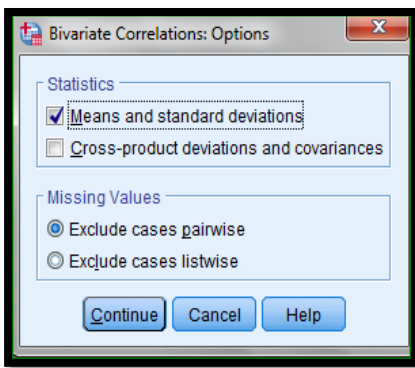
2. اضغط على المتغيرين SAT , Achievement معًا (اضغط ctrl) أو انقل كل متغير على حده ثم اضغط على السهم → لتنتقلهما إلى المربع Variables

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 كالآتي:



7. اضغط فى مربع Statistics اضغط على

اختيار means and standard deviations

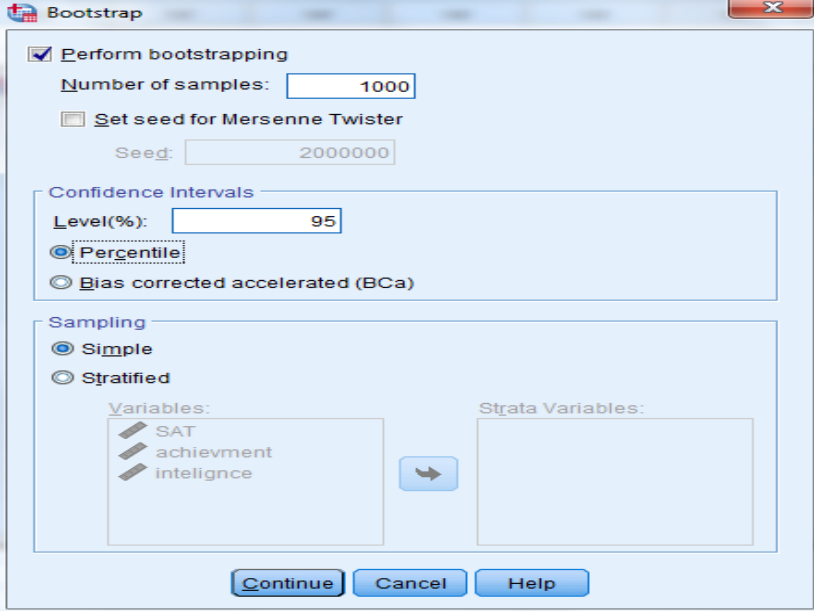
ليعطى المتوسطات و الانحرافات المعيارية لكل متغير فى التحليل.

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 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 مثلاً وهي Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة. ويتم ال
- حدد فترات الثقة التي 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 فمثلاً فترات الثقة 95% فإنه يعتمد على المئينى percentile 5th و Percentile 95th كنقطتى للفترة. وقد يحدث اختلاف بين فترات الثقة البارامترية التي تقدر من خلال اختبارات الفروض الكلاسيكية الصفرية عن فترات الثقة اللابارامترية المقدرة من خلال البوتستراپ وهذا نتيجة لعد الاعتدالية للبيانات.
- حدد نوعية العينة بسيطة أو طبقية Stratified وهي بسيطة.

12. اضغط OK لتنفيذ الامر.

• ثالثاً 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

وهى مصفوفة قطرها واحد

		SAT	achievment
SAT	Pearson	1	.903**
	Correlation		
	Sig. (2-tailed)		
	N		
achievment	Pearson	.903**	1
	Correlation		
	Sig. (2-tailed)		
	N		

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:

**. Correlation is significant at the 0.01 level (2-tailed).

- قيمة معامل الارتباط $r = 0.903$

- قيمة p لتحديد الدلالة الإحصائية لمعامل الارتباط و $p = 0.000$ و بما أن $0.00 < p$

$\alpha = (0.01)$ ، وعليه فهى علاقة ارتباطية دالة إحصائياً عند 0.01 وبالتالي فهى دالة

عند 0.05 .

- حجم العينة هو 15

وعليه توجد علاقة ارتباطية دالة إحصائيًا بين التحصيل والاستعدادات، وبالتالي 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 مثلاً) على عدد معاملات الارتباطات فاذا وجدت مصفوفة تتضمن خمسة متغيرات وعليه يوجد 10 معاملات الارتباط وبالتالي تصحح مستوى الدلالة الإحصائية كالاتى:

$$\alpha = \frac{0.05}{10} = 0.005$$

وبالتالى يجب مقارنة القيم P بـ 0.005 .

القرار فى ضوء تطبيق استراتيجيه Bootstrap:

Bootstrap Specifications

Sampling Method	Simple
Number of Samples	1000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

		SAT	achievement
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

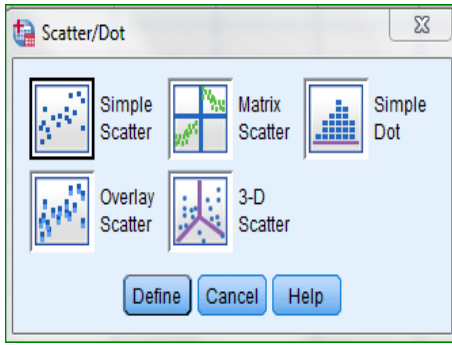
اعطى الخطأ المعياري 0.045 اما فترات الثقة للتحصيل (0.811, 0.97) وهاتين القيمتين لا يقع بينها قيمة الصفر الموضوعة فى الفرض الصفري بالتالى نرفض الفرض الصفري وهذا يعنى وجود علاقة ارتباطية بين التحصيل والاستعدادات.

التحليل البديل لمعامل ارتباط بيرسون

إذا لم تتحقق مسلمات بيرسون مثل الاعتدالية والخطية فإنه يمكن باستخدام معامل ارتباط سبيرمان أو كندال تاو وذلك بالتعامل مع رتب الدرجات وبذلك تصبح المتغيرات رتبية.

استخدام الرسم البياني لعرض النتائج

لعرض شكل الانتشار للعلاقة بين المتغيرين اتبع الخطوات الآتية:



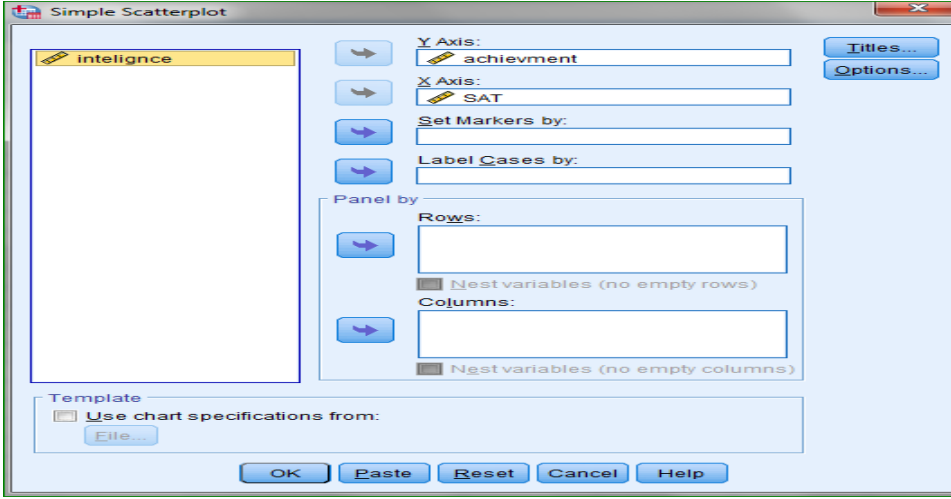
1. اضغط على Graphs ثم اختار legacy

Dialogs ثم اضغط على اختيار Scatter

dot تظهر الشاشة:

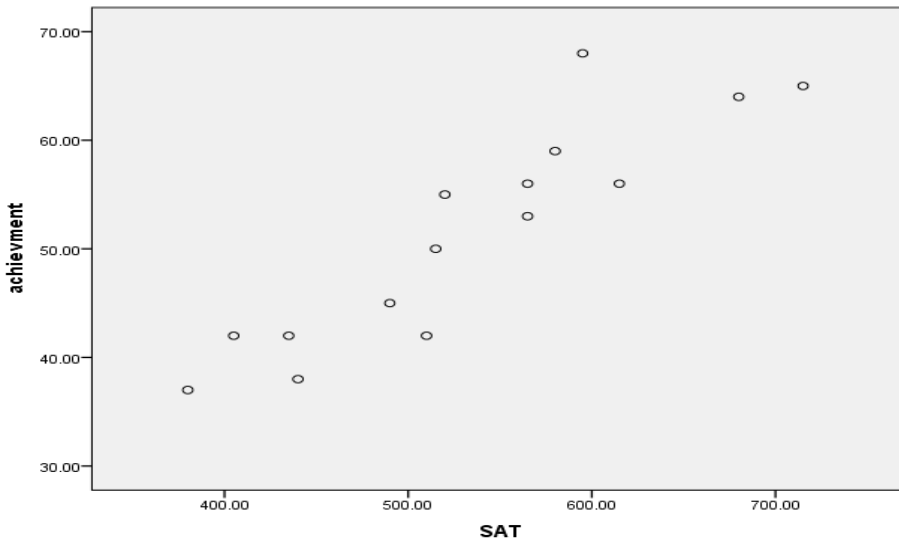
2. اضغط على simple Scatter ثم اضغط

على define تظهر الشاشة الآتية:



3. انقل SAT إلى مربع X Axis و achievement إلى مربع Y Axis

4. اضغط OK يظهر المخرج الآتي:



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 كالاتي:

$$r_{xy.z} = \frac{r_{xy} - (r_{xz})(r_{yz})}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

- $r_{xy.z}$ ، معامل الارتباط الجزئي 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 : $r_{xy} = 0.91$, $r_{xz} = 0.85$, $r_{yz} = 0.97$

$$r_{xy.z} = \frac{0.91 - (0.85)(0.97)}{\sqrt{1 - (0.85)^2(1 - (0.97)^2)}} = 0.65 \text{ فان}$$

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 إحصائياً باستخدام

معامل الارتباط الجزئي وتجريبياً من خلال الضبط التجريبي وهذا يدعونا إلى الحذر عند تفسير معامل الارتباط أو الحذر عند استخدامه للاستدلال السببي في الأساليب الإحصائية المتدرجة.

ويرى (2009) Field إذا افترضنا ان التحصيل يرتبط سلبياً مع قلق الامتحان وإيجابياً مع وقت الاستذكار فإن وقت المراجعة للامتحان ترتبط ارتباطاً سالباً مع قلق الامتحان وعلى ذلك فإن الوقت يرتبط مع كلاً من التحصيل والقلق. فإذا افترضنا ان القلق فسر 19.4% من تباين التحصيل والوقت فسر 15.7% من تباين التحصيل والوقت فسر 50.2% من تباين القلق ، و الوقت و القلق يسهموا معاً في تفسير 19.4% (المحسوبة على القلق) و على ذلك يهدف معامل الارتباط الجزئي 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 الجزئي

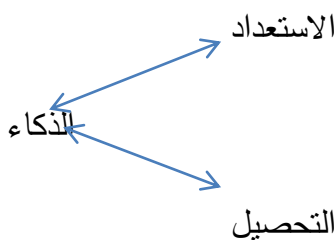
1. توافر الاعتدالية المتدرجة للمتغيرات وإذا تحققت الاعتدالية من اتحاد المتغيرات معاً فإن هذا يفترض تحقق الاعتدالية لكل متغير حدة.

2. العينة العشوائية واستقلال درجات كل حالة عن درجات الحالات الأخرى.

حجم التأثير لمعامل الارتباط الجزئي

هو نفسه مربع معامل الارتباط الجزئي كما هو الحال مع معامل ارتباط بيرسون.

قضية: في 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 الجزئي في SPSS

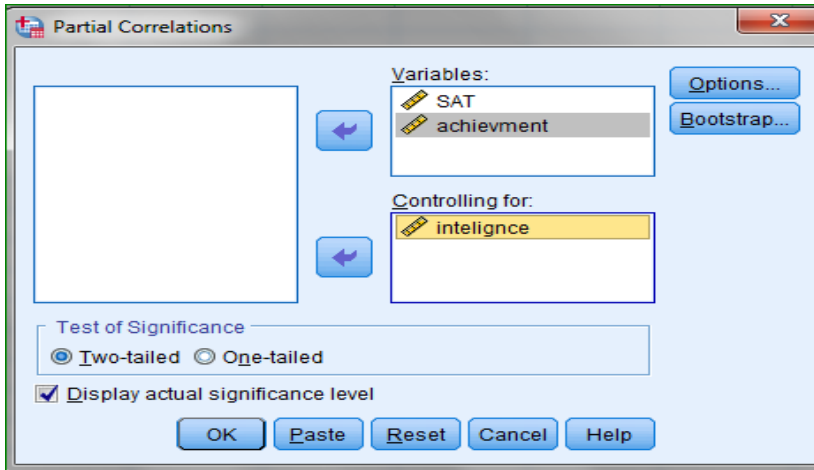
اولاً: إدخال البيانات: 1. اضغط Variable view

2. اكتب مسمى المتغيرات تحت عمود Name كالآتي: الصف الأول: الاستعداد

SAT، الصف الثاني : التحصيل achievement، الصف الثالث Intelligence

3. اضغط dataview وأدخل المتغيرات في ثلاثة أعمدة.

ثانياً: تنفيذ الأمر :1. اضغط على Analyze ثم Correlate ثم اضغط على Partial تظهر الشاشة:

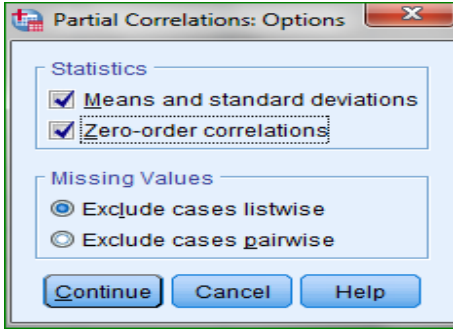


2. انقل متغيرات التحصيل والاستعداد معاً أو كلاهما إلى مربع Variables

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

ثالثاً: تفسير المخرج: 1. يعطى المخرج الإحصاء الوصفى لكل المتغيرات كالتى:

Descriptive Statistics			
	Mean	Std. Deviation	N
SAT	534.0000	96.53275	15
achievment	51.4667	10.10563	15
inteligence	118.2000	11.32128	15

2. اعطى البرنامج مصفوفة الارتباط و هى مكونة من جزئين كالتى:

Correlations					
Control Variables			SAT	achievment	inteligence
-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
	inteligence	Correlation	.240	.188	1.000
		Significance (2-tailed)	.388	.503	.
		df	13	13	0
inteligence	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. الجزء الأسفل: يعرض معامل الارتباط الجزئي بين الاستعدادات والتحصيل بعد عزل أثر الذكاء على كلٍّ من الاستعدادات والتحصيل اتضح ان قيمته 0.900، أى أنه انخفض من 0.903 إلى 0.900 وأيضًا معامل الارتباط الجزئي دال إحصائيًا عند 0.05 وكذلك عند 0.01 وإذا اعتمد الباحث على معامل ارتباط بيرسون بين المتغيرين فإن التباين المفسر فى التحصيل جراء الاستعدادات تقريبًا 0.810 بينما التباين المفسر للتحصيل جراء الاستعدادات بعد استبعاد اثر الذكاء هو 0.815 تقريبًا.

معامل الارتباط شبه الجزئي Part correlation

هو حذف تأثير متغيرات خارجية تؤثر على العلاقة بين متغيرين ما من خلال حذف تأثيرها على أحد المتغيرين وليس على كلاهما كما فى حالة Partial. والصيغة لتقدير معامل الارتباط بين Y و X1 مع حذف تأثير X2 على X1 فقط هى كالتالى:

$$r_y(x_1, x_2) = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{1 - r_{x_1x_2}^2}}$$

وهذه المعادلة تعطى التباين الكلى بين Y و X1 منقحًا من التداخل المشترك بين X1 و X2 وليس من تداخل X1 مع Y. وعلى ذلك فإن العلاقة بين X1 و Y مع عزل تأثير X2 على X1 فقط هى $r_y(x_1, x_2)$ بينما معامل الارتباط الجزئي هو عزل تأثير X2 على كل من Y و X1 (Kline, 2016)، وعلى ذلك عندما ترتبط المنبئات ببعضها فإن أوزان بيتا والارتباطات الجزئية هى البديل لمعرفة القوة التفسيرية لكل متغير منبئ مع ضبط المتغيرات الأخرى.

الفصل الرابع عشر

معامل الارتباط الثنائي 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 أخرى.

وأحياناً تقاس المتغيرات فى صورة ثنائية Dichotomy مثل ذكر-أنثى، ناجح-راسب، مجموعة ضابطة-مجموعة تجريبية فهذه المتغيرات لا تقاس رقمياً 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 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 r_{pb} وعلى ذلك فإن $r = r_{pb}$ ، وصيغة حساب r_{pb} مشتقة من معامل ارتباط بيرسون وهى كالآتى:

$$r_{pb} = \left(\frac{\bar{X}_{y1} - \bar{X}_{y2}}{SY} \right) (\sqrt{Pq})$$

- $\bar{X}_{y1}$ متوسط درجات المتغير التابع المتصل للتصنيف الأول للمتغير التصنيفى.
- $\bar{X}_{y2}$ متوسط درجات المتغير التابع المتصل للتصنيف الثانى للمتغير التصنيفى.
- SY الانحراف المعيارى للمتغير التابع.
- P, q 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 كالتالي:

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 والاتجاه نحو الانترنت؟
2. فرض البحث: توجد علاقة بين الجنس والاتجاه نحو الانترنت.
3. المتغيرات: الجنس: اسمي - مستقل بمستويين، الاتجاه نحو الانترنت: تابع- فترتي - متصل.
4. منهج البحث: منهج ارتباطي.
5. النموذج الإحصائي: بسيط - إحصاء بارامترى والاختبار الإحصائي معامل الارتباط الثنائي الاصيل r_{pb} .

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية:

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

ρ معامل الارتباط بين المتغيرين في المجتمع

2. الاختبار الإحصائي: معامل ارتباط بيرسون للمتغير التصنيفي والآخر متصل وهو معامل الارتباط r_{pb} ويقدر كالتالي:

$$r_{pb} = \left(\frac{\bar{X}_{y1} - \bar{X}_{y2}}{S_Y} \right) \left((\sqrt{Pq}) \right)$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: تبني الباحث $\alpha = 0.05$ والنسبة لقاعدة القرار والدالة الإحصائية لهذا الاختبار تقدر من خلال اختبار T لعينات مستقلة:

$$T^2 = \frac{r_{pb}^2}{(1 - r_{pb}^2) / df}$$

$$T = \frac{r_{pb} \sqrt{N-2}}{\sqrt{1 - r_{pb}^2}}$$

$$df = N - 2 = 12 - 2 = 10$$

4. الحسابات:

أ- تكويد المتغير التصنيفي حيث الذكور = 1 ، الإناث = 2 وعليه يوجد 5 ذكور و 7 إناث.

$$P = \frac{5}{12} = 0.42 \quad \text{وتكون نسبة الذكور } P$$

$$q = \frac{7}{12} = 0.58 \quad \text{ونسبة الإناث } q$$

$$p + q = 1 \quad \text{لاحظ أن}$$

ب- حساب متوسط المتغير التابع Y لكل مستوى فمتوسط درجات الذكور $\bar{X}_{y1}$ كالتالي:

$$\bar{X}_{y1} = \frac{23+19+\dots+19}{5} = 17$$

$\bar{X}_{y2}$: ومتوسط درجات الإناث

$$\bar{X}_{y2} = \frac{32+10+\dots+34}{7} = 20$$

ج- حساب الانحراف المعياري للمتغير التابع $YSY = 9.056$:

د- حساب معامل الارتباط:

$$r_{pb} = \left(\frac{17-20}{9.056} \right) (\sqrt{0.42 \times 0.58}) = -0.163$$

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 الإنترنت للإناث أعلى من متوسط الذكور.

5. القرار والتفسير: القرار بتحويل r_{pb} إلى قيمة اختبار T المناظرة كالاتى:

$$T^2 = \frac{r_{pb}^2}{(1-r_{pb}^2)/df} = \frac{0.163}{(1-0.163)/10} = 0.272$$

$$T = \sqrt{0.272} = 0.522 \quad \text{وعليه فالجذر التربيعى هو :}$$

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 فإن: T المحسوبة $> T$ الجدولية (2.228)، وعليه نقبل الفرض الصفري القائل الجنس والاتجاه نحو الإنترنت غير مرتبطين.

6. حجم التأثير: يقدر حجم التأثير من خلال:

• مؤشر إيتا تربيع η^2 كالاتى:

$$\eta^2 = \frac{T^2}{T^2 + df} = \frac{(0.522)^2}{(0.522)^2 + 10} = 0.0272$$

ومربع قيمة r_{pb} وهى:

$$r^2 = (r_{pb})^2 = (-0.163)^2 = 0.0272$$

والمؤشرين اعطوا نفس القيمة وهذا يفيد ان الجنس فسر 2.72% من تباين الاتجاه 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 الآتي:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختار:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

3. أسفل Test family اختار t- tests

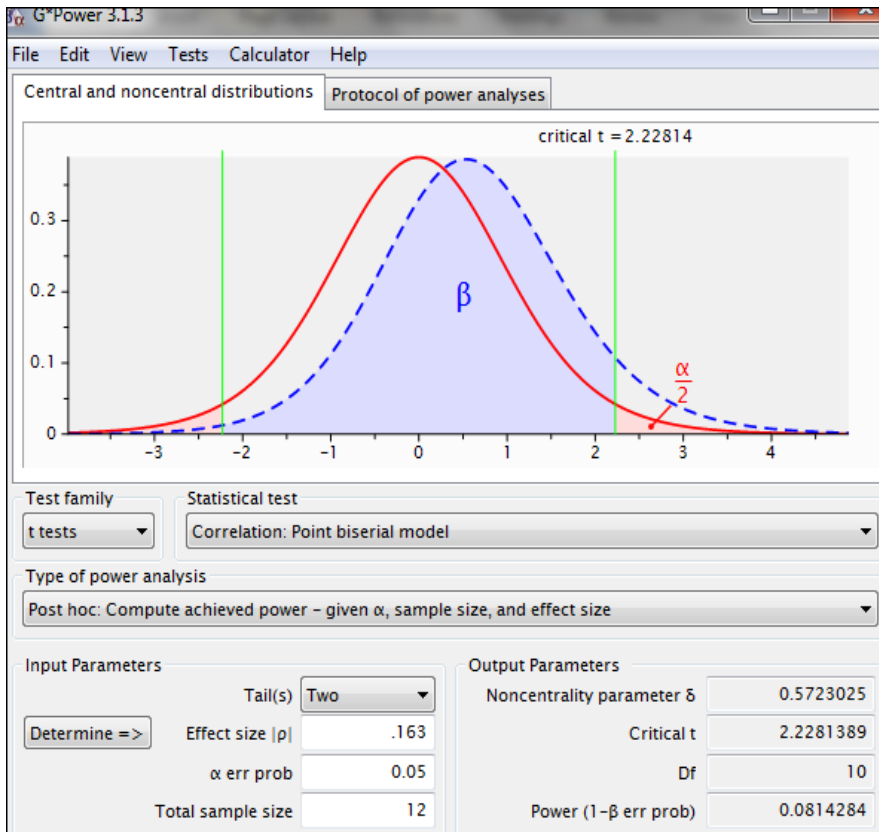
4. تحت Statistical test اختار:

Test family	Statistical test
t tests	Correlation: Point biserial model

5. ادخل معالم المدخلات كالآتي:

Input Parameters	
Tail(s)	Two
Determine => Effect size p	.163
α err prob	0.05
Total sample size	12

6. اضغط calculated يظهر المخرج الآتي:



يتضح أن القوة الإحصائية = 0.0814 وهذا مستوى ضعيف جداً. وهذا انعكس على عدم قدرة الاختبار على رفض الفرض الصفري ولذلك عدم الدلالة الإحصائية ليست نتيجة لأنها حقيقية في المجتمع بل قد تكون نتيجة عدم قدرة الاختبار على رفض الفرض الصفري وهذا بدوره يضخم الخطأ من النوع الثاني (عدم دلالة على مستوى العينة بينما هي موجودة في المجتمع).

كتابة نتائج r_{pb} في تقارير البحوث وفقاً 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} في برنامج SPSS

أولاً: إدخال البيانات: تم تعريف المتغيرين من خلال الضغط على variable view أسفل الشاشة تظهر الشاشة التي يحدد بها المتغيرات اكتب مسمى المتغيرين تحت

عمود Name وليكن attitude و sex ثم اضغط على DataView وقم بإدخال البيانات.

ثانيًا تنفيذ الأمر: 1. اضغط على Analyze→ Bivariate→Correlate

ونفس خطوات معامل ارتباط بيرسون.

الفصل الخامس عشر

الانحدار البسيط

Simple Regression

تناولنا سابقاً كيف يمكن تقدير العلاقة الارتباطية بين متغيرين وهذه الارتباطات فى غاية الأهمية ولكننا فى هذا الفصل نعلم 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)، فمثلاً القدرة القرائية فى المرحلة الابتدائية مرتبطة بعدة متغيرات مستقلة مثل الذكاء والنمو المعرفى والعمر وغيرها. وكما نعلم أن الانحدار والارتباط متداخلين ولكن الانحدار يستخدم غالباً 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 رئيسيين 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

ضروريًا للتفسير وعمومًا في أبحاث العلوم الإنسانية والسلوكية يهتم الباحثين بتفسير لماذا تحدث الظاهرة أكثر من التنبؤ بحدوثها.

الانحدار الخطي البسيط

يهدف إلى دراسة تأثير متغير مستقل (X) على متغير تابع (Y) والباحث يحاول أن يفهم كيف ان المتغير (X) يؤدي إلى التغير في (Y)، بكلمات أخرى الهدف هو تحديد كيف وإلى أى درجة أن التباين أو التغير في سلوك المتغير التابع يعتمد على المتغير المستقل. والانحدار البسيط أداة إحصائية تسمح لنا بالتنبؤ بدرجة فرد على المتغير التابع من درجته على المتغير المستقل واحد ومثل معامل الارتباط لبيرسون، فإن الانحدار البسيط يستخدم مع بيانات متغيرين ويجب أن تكون العلاقة بينهما خطية(عامر، 2007).

على سبيل المثال أراد باحث تحديد تأثير ساعات المذاكرة (X) على التحصيل (Y) بكل تأكيد الأداء على التحصيل يتأثر بعوامل أخرى غير (X) ومن غير المحتمل أن كل الأفراد الذين يتعرضون لنفس مستوى التغير في (X) سوف يظهروا نفس التغير المماثل على (Y) لوجود متغيرات أخرى تؤثر على المتغير التابع Y وعندما تكون متوسطات (Y) للمستويات المختلفة لـ (X) تتغير من فرد إلى آخر وتقع على خط مستقيم فإنه يوجد انحدار خطي بسيط لـ Y على X وتكون معادلة الانحدار في المجتمع:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Y درجة الفرد i في المتغير التابع
- α (ألفا) متوسط المجتمع عندما تكون $x = 0$ أو ثابت Y
- β (بيتا) معامل الانحدار في المجتمع أو ميل خط الانحدار
- ε_i (إيبسلون) قيمة الخطأ أو الخطأ العشوائي Random Disturbance للفرد i
- X_i درجة المتغير المستقل لفرد i

ومعامل الانحدار β يشير إلى قيمة تأثير المتغير المستقل في المتغير التابع وأي وحدة تغير في المتغير المستقل X يقابلها تغير متوقع مساوي لحجم β في المتغير Y وعلى ذلك درجة أي فرد على المتغير التابع تتكون من جزئين:

1. جزء $X + \alpha$: وهو جزء من الدرجة Y للفرد الذى يتعرض لمستوى معين من X

2. جزء الخطأ العشوائى ε_i 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 مقدار لـ β و e هى مقدار 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 كالتالى:

$$b = \frac{\sum XY}{\sum X^2}$$

$$a = \bar{Y} + b\bar{X}$$

• $\sum xy$ 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 يؤدي 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.

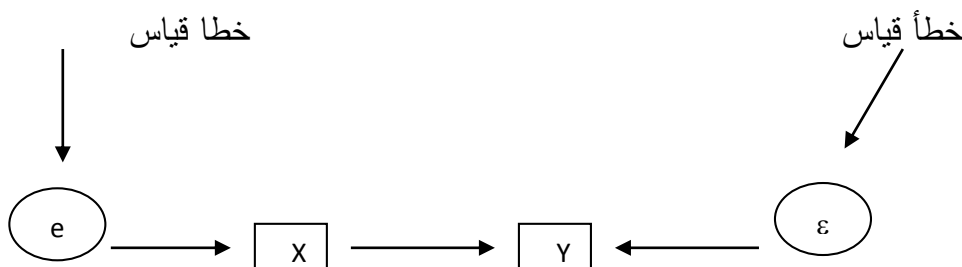
• يفترض أن المتغير X يقاس بدون أخطاء قياس.

• يفترض أن المتغير المستقل مثبت Fixed variable فعندما يريد إعادة التجربة

مثلاً فإنه تستخدم نفس قيم X فى 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 عشوائياً 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 لـ Y وتذكر الدرجة Y_i يتكون من جزء مثبت $\alpha +$ (ε) و جزء الخطأ عشوائى BX



وفيما يلى مسلمات خاصة بالأخطاء القياس وهى:

- متوسط أخطاء كل قياس Y_i عبر العديد من التطبيقات أو المرات هى صفر.
 - الأخطاء المرتبطة بالقياس أو الملاحظة Y_i 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 Homoscedasticity 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 فى SPSS

أولاً: إدخال المتغيرات: 1. بعد فتح البرنامج اضغط على Variable view تظهر الشاشة الفرعية:

	Name	Type	Width	Decimals	Label	Values
1	achievemnt	Numeric	8	2		None
2	hours	Numeric	8	2		None
3						
4						

تحت عمود Name اكتب
التحصيل achievement وكذلك
متغير عدد الساعات hours ثم
اضغط Data view لترجع إلى
شاشة إدخال البيانات

ثانياً: تنفيذ الأمر: 1. اضغط على Analyze→Regression→Linear تظهر الشاشة:

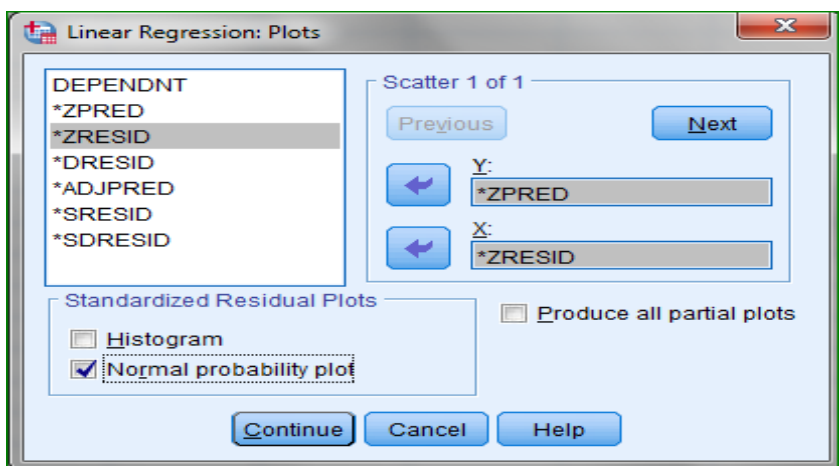
2. انقل achievement إلى مربع Dependent ثم انقل hours إلى
مربع Independent

3. اضغط على اختيار Statistics على
يمين الشاشة تظهر الشاشة:

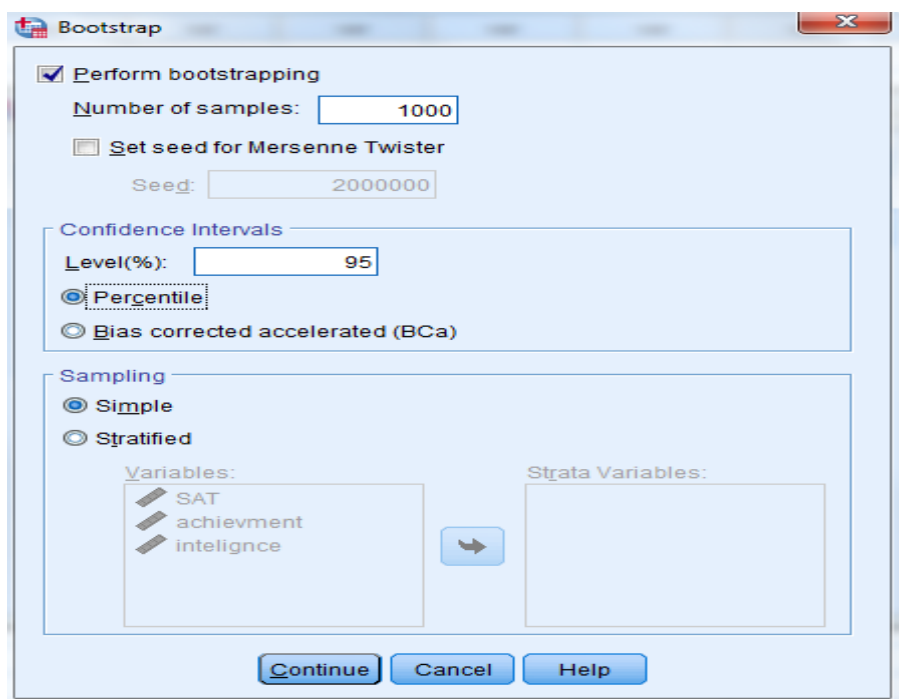
اضغط على خيارات Estimates و
Model fit و Confidence intervals و
Descriptive

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 مربع X، واضغط على Normal probability plot
7. اضغط على Bootstrap تظهر النافذة الآتية:



8. اضغط Perform bootstrapping:

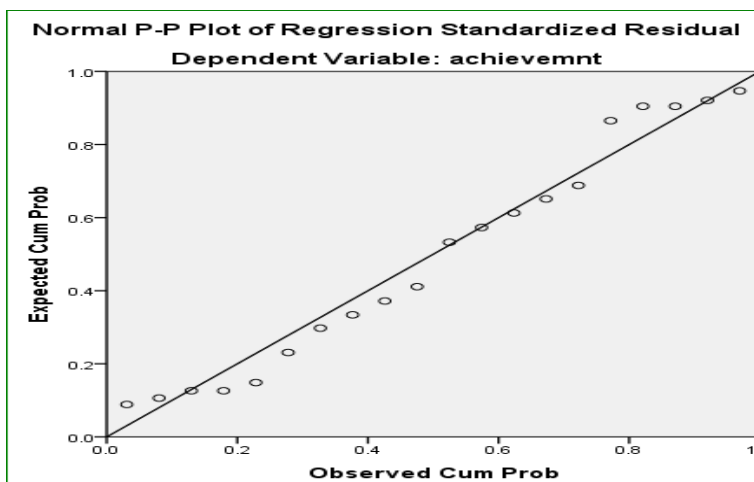
- حدد عدد العينات التي ترغب في توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها وليكن 500 لاحظ أنه إذا كان حجم العينة المتاحة كبير يمكن تحديد عدد العينات بـ 1000 مثلاً وهي Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.

- حدد فترات الثقة التي 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 فمثلاً فترات الثقة 95% فإنه يعتمد على المئينى 5th percentile و 95th Percentile كنقطتي للفترة. وقد يحدث اختلاف بين فترات الثقة البارامترية التي 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.

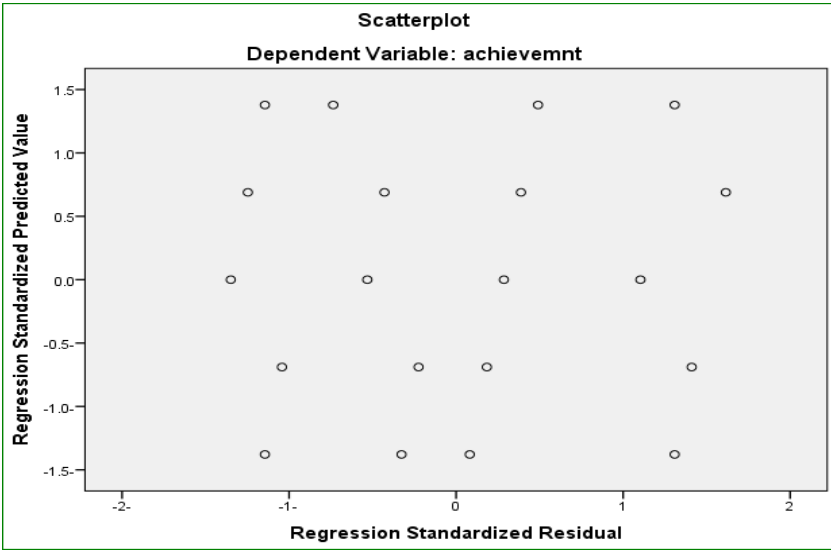
9. اضغط على Continue ثم اضغط OK

ثالثاً: المخرج:

- مسلمات البيانات: بالنظر فى شكل Normal 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 نظراً لابتعاد العديد من الدوائر الصغيرة عن الخط المستقيم ولذلك يمكن أن نشك فى توافر الاعتدالية للبيانات كذلك اعطى البرنامج الشكل الآتى:



وهذا الشكل يعرض البواقي المعيارية والقيم المتنبأ بها المعيارية ونلاحظ أن الدوائر الصغيرة تتوزع أو تنتشر عشوائيا وهذا الشكل يشير إلى تساوى التباينات للمتغير التابع لكل مستوى من مستويات المتغير المستقل وهذا يشار إليه Homoscedasity وهى مطلب أساسى للانحدار بمعنى أن الأخطاء المعيارية تتوزع عشوائياً. ولا بد من التحقق من مسلمة الاعتدالية للمتغيرات hours و achievement قبل إجراء تحليل الانحدار الخطى البسيط (انظر فصل التحقق من الاعتدالية).

ثانيا: تفسير مخرج الانحدار والتنبؤ: أعطى البرنامج إحصاء وصفى للمتغيرات كالاتى:

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 $r = 0.416$ وهو من النوع المتوسط والجزء الثانى يعبر عن الدلالة الإحصائية لاختبار ذو ذيل واحد وهى $p = 0.034$ وهى دالة إحصائياً عند 0.05 والجزء الثالث من الجدول يعبر عن حجم العينة وهى 20 فرداً.

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 achievement هي

المطابقة العامة للنموذج Overall fit of the model: عرض البرنامج ملخص النموذج كالآتى :

Model Summary ^ط				
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 Y ودرجات المتغير التابع المتنبأ بها من المعادلة $\hat{Y}$.

• قيمة R^2 وهى مربع قيمة معامل الارتباط $R^2 = (0.416)^2 = 0.173$ 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 إحصائياً؟ وعليه يوجد فرضان إحصائيان:

الفرض الصفري: معادلة الانحدار لا تصلح لاستخدامها فى التنبؤ:

$$H_0: R(\rho) = 0 \text{ أو } \text{Beta} = 0$$

الفرض البديل: معادلة الانحدار تصلح لاستخدامها في التنبؤ:

$$H_A: R(\rho) \neq 0 \text{ أو } \beta \neq 0$$

حيث β معامل الانحدار المعياري في المجتمع.

ويستخدم اختبار تحليل التباين F لاختبار الدلالة الإحصائية لمعادلة الانحدار كالاتي:

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 في المتغير التابع عن طريق المتغير المستقل ويقدر كالاتي:

$$SS_{regression} = \frac{(\sum XY)^2}{\sum X^2}$$

- $SS_{residual}$ مجموع مربعات البواقي وهو التباين غير المفسر في معادلة الانحدار ويقدر كالاتي:

$$\sum Y^2 = SS_{regression} + SS_{residual}$$

إذا:

$$SS_{residual} = \sum Y^2 - SS_{regression}$$

و درجات الحرية للانحدار هي:

$$df_{regression} = K=1$$

- K عدد المتغيرات المستقلة.

و درجات الحرية للبواقي:

$$df_{residual} = N - K - 1 = 20 - 1 - 1 = 18$$

• N حجم العينة.

وتحسب قيمة F لاختبار الدلالة الإحصائية لمعادلة الانحدار كالاتى:

$$F = \frac{22.5/1}{107.7/18} = \frac{22.5}{5.983} = 3.76$$

وهى نفس القيمة فى المخرج ولكن فى الجدول السابق أعطى قيمة $P=0.068$ وهذا يعنى أنها غير دالة إحصائياً، حيث $P(0.068) < 0.05$ وعليه فإن معادلة التنبؤ لا تصلح للتنبؤ. وإذا بحثنا فى جدول F بدرجتي حرية 1 ، 18 نحصل على F الجدولية = 4.41 بينما F المحسوبة 3.76 ، وبما أن $F > (3.76)$ الجدولية (4.41) وعليه نقبل الفرض الصفري أى معادلة التنبؤ لا تصلح لأغراض التنبؤ أو انحدار التحصيل على عدد الساعات لا يختلف عن الصفر.

ويمكن حساب قيمة F من المعادلة الآتية:

$$F = \frac{R^2/K}{(1 - R^2)(N - K - 1)} = \frac{0.1728/1}{(1 - 0.1728)(20 - 1 - 1)} = 3.76$$

وعلى ذلك فلا داعى للنظر إلى الدلالة الإحصائية لمعامل معادلة الانحدار وهى معامل 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	Sig.
		B	Std. Error	Beta	
1	(Constant)	5.050	1.283		.001
	hours	.750	.387	.416	.068

a. Dependent Variable: achievemnt

كما نعلم أن المعادلة كالاتى:

$$\hat{Y} = a + b X + e$$

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 المنتبأ بها على المتغير التابع لشخص درجته صفر على المتغير المستقل، و b هو معامل الانحدار غير المعيارى وعليه فإن معادلة الانحدار غيرالمعيارية هى:

$$\text{Achievement} = 0.750 \text{ hours} + 5.050$$

$$\hat{Y} = 0.750 X + 5.050$$

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 المنتبأ بها فى الامتحان:

$$= 0.70 + 5.05 = 5.70$$

حقيقة إن معامل الانحدار يخبرنا 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 البوتستراب اعطى فترات الثقة لمعاملات الانحدار حيث فى ضوء الدلالة الإحصائية وفترات الثقة فإن معامل الانحدار غير دال إحصائياً. لاحظ فروق

طفيفة بين اختبارات الفروض باستخدام عينة واحدة وباستخدام معاينات البوتستراب نظراً لصغر حجم العينة لقاعدة البيانات المتولد منه البيانات.

القوة الإحصائية للانحدار البسيط باستخدام برنامج G-power
لحساب القوة الإحصائية باستخدام برنامج **G-Power** اتبع الآتي:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختر:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

3. أسفل Test family اختر F- tests

4. أسفل Statistical test اختر:

Test family	Statistical test
F tests	Linear multiple regression: Fixed model, R^2 deviation from zero

5. أدخل المعالم الآتية تحت Input parameters:

-حجم التأثير f^2 كالآتي:

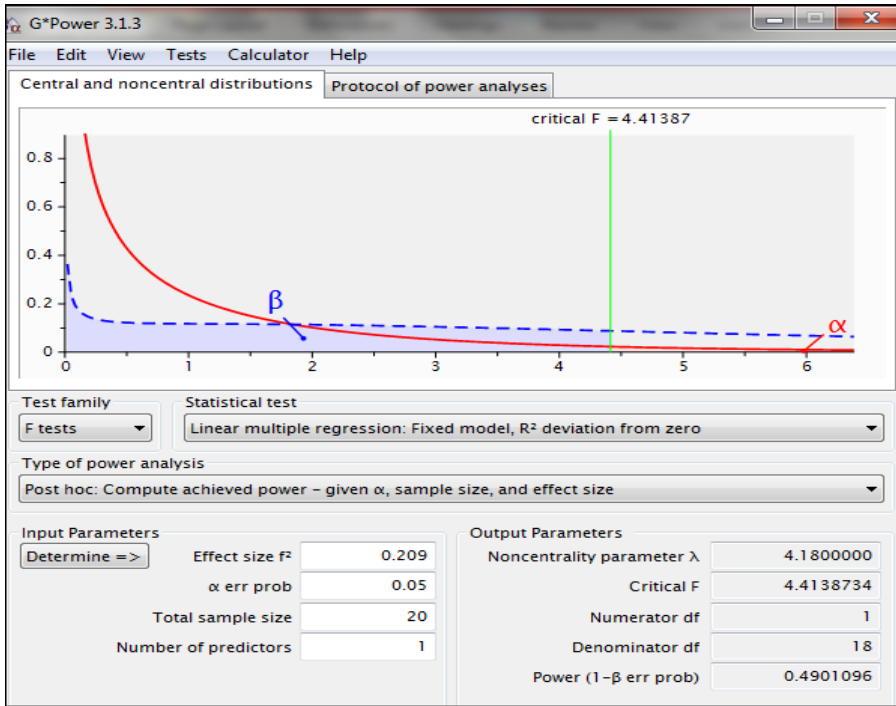
$$f^2 = \frac{R^2}{1 - R^2} = \frac{0.1728}{1 - 0.1728} = \frac{0.1728}{0.827} = 0.209$$

وفي ضوء معايير Cohen(1988) فإن 0.02 حجم تأثير ضعيف، 0.15 متوسط، 0.35 فأكثر حجم تأثير كبير.

- مستوى الدلالة الإحصائية: 0.05، حجم العينة الاجمالي = 20، عدد المنبئات = 1

Input Parameters	
Determine =>	Effect size f^2 0.209
	α err prob 0.05
	Total sample size 20
	Number of predictors 1

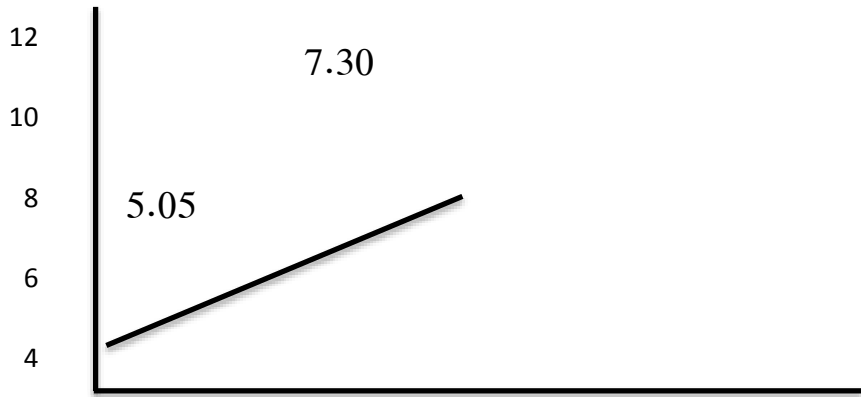
6. اضغط Calculated يظهر المخرج الآتي:



يتضح أن القوة الإحصائية = 0.490 وهذا مستوى قوة منخفض وبدورة يؤدي إلى تضخم الخطأ من النوع الثاني بيتا (β) وقد يرجع إلى صغر حجم العينة.

خط الانحدار Regression line

تستخدم معادلة الانحدار لرسم العلاقة بين الوقت والتحصيل على المحور الصادي (Y) وعدد الساعات على المحور السيني (X) وكما نعلم أن الثابت $a=5.050$ هي قيمة Y عندما تكون $X=0$ وهذه النقطة الأولى أما النقطة الثانية فهي يعبر عنها بمتوسط عدد الساعات $(\bar{X}) = 3.00$ ومتوسط التحصيل وهو 7.30



وهذا الميل يمثل الزيادة المتنبأ بها في Y لكل وحدة في X بمعنى أن زيادة ساعه واحدة في المذاكرة يتوقع زيادة التحصيل بـ 0.75 نقطة واحدة، وهذا هو تفسير معامل الانحدار غير المعياري.

معامل الانحدار المعياري Standardized Regression Coefficient

بالنظر في المخرج السابق نلاحظ وجود المعاملات المعيارية (Beta) وراعى أن معامل الانحدار غير المعياري فسرناه كتغير في المتغير التابع لكل وحدة تتغير في المتغير المستقل وفي هذا المثال فإن $b = 0.75$ وهذا يعنى أن لكل وحدة واحدة في ساعات المذاكرة تتنبأ بـ 0.75 وحدة في التحصيل وتفسر بيتا β مثل b ولكن في ضوء وحدات الانحراف المعياري $\beta = 0.416$ وتعنى أن لكل انحراف معياري زيادة في عدد الساعات فإن التحصيل يزيد بـ 0.416 انحراف معياري وعلى ذلك فإن β هي معامل الانحدار المعياري عندما يتم التعبير عن قيم المتغير المستقل والتابع بالدرجات المعيارية (Z)

ويمكن تحويل b إلى β كالتالى :

$$\beta = b \frac{SD_y}{SD_x}$$

وعليه ففى المثال:

$$\beta = 0.75 \times \frac{1.4509}{2.6177} = 0.416$$

ونلاحظ أن معامل الانحدار المعياري هو قيمة معامل ارتباط بيرسون بين المتغيرين ومعادلة الانحدار المعيارية هي:

$$\hat{Z} \text{ Achievement} = 0.416 Z_{\text{hours}}$$

لا تتضمن الثابت a.

ولكن على ماذا يعتمد الباحث معامل الانحدار المعياري أم اللامعيارى؟ كلاهما مفيد ولكن إذا كانت توجد مقارنة بين معاملات الانحدار للمتغيرات المستقلة بالتابعة فيفضل استخدام معامل الانحدار المعياري لأنه يسمح بالمقارنة عبر متغيرات بوحدات قياس مختلفة.

عدم تحقق المسلمات

يمكن أن يكون لتحليل الضلالة Robust analysis فى حالة الابتعاد أو عدم تحقق هذه المسلمات دوراً فى إجراء التحليل ما عدا عدم تحقق مسلمات أخطاء القياس وكذلك أخطاء التخصيص.

أخطاء القياس: أخطاء القياس فى المتغير التابع لا تؤدي إلى تحيز فى معامل الانحدار ولكنها تؤدي إلى زيادة فى الخطأ المعياري للتقدير وعليه ظهور خلل فى اختبارات الدلالة الإحصائية (F , T) وأخطاء القياس فى المتغير المستقل تؤدي إلى تقدير منخفض Under- estimation لمعامل الانحدار وهذا نتيجة لانخفاض الثبات لقياسات المتغير المستقل حيث:

$$\text{الدرجة الحقيقية لـ } (T) = \text{الدرجة المقاسة } \pm (X) \text{ أخطاء قياس.}$$

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 فمثلاً $r_{tt} = B$ ، 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 يوجد 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 دائماً يؤدي إلى وجود تحيز فى معاملات الانحدار.

أخطاء التخصيص أو التحديد Specification errors

تشير إلى الأخطاء المرتكبة في تخصيص النموذج المراد اختباره أو عدم تحقق أى من المسلمات المراد توافرها في النموذج، فأى نموذج انحدار هو انعكاس لبناء نظرى للظاهرة موضع الدراسة وعندما يكون بنائه بعيداً عن الإطار النظري بمعنى تضمين متغيرات غير مناسبة أو حذف متغيرات مناسبة أو وجود علاقة منحنية خطية Curvilinear بدلاً من العلاقة الخطية وهذا يلقي بظلاله في 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 في الأسلوبين واحدة ويعتبر اختبارى T, ANOVA تحت مظلة الانحدار بل هما حالات خاصة منه. حيث تعتبر ANOVA هي جزء من النموذج الخطى العام (General Linear model) (Howell, 2014)

وان المعادلة العامة لهذه الأساليب هي:

$$Y = \mu + B + e$$

ويفترض أن درجة الفرد على المتغير التابع هي مجموع المتوسط العام والتباين نتيجة تأثير

المعالجة التجريبية B وإضافة أو طرح تأثير الخطأ e

وهذه معادلة الانحدار :

$$Y = a + b x + e$$

حيث الدرجة على المتغير التابع هي مجموع ثابت لكل الأفراد (a) بالإضافة إلى تباين (b) يرجع إلى المتغير المستقل X زائداً أو ناقص خطأ عشوائى (e) وهاتين المعادلتين واحد ولهما نفس التفسير ولذلك فإن ANOVA هو جزء من النموذج الخطى العام.

السببية Causality

في الانحدار يمكن القول إن الدافعية تساعد في 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 في النجاح في الحياة
- طلاق الوالدين يؤثر في تحصيل الأطفال وسلوكياتهم
- احتمال سمات الشخصية ان تسبب بصورة جزئية النجاح في الحياة "

حقيقة يمكن عمل استنتاجات سببية من بيانات ارتباطية ولكننا لا نشعر بالثقة والمصادقية 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 العيى لإحصاء ما يغطى كل القيم التى يمكن أن تأخذها الإحصائية واحتمال الحصول على أى قيمة نتيجة للصدفة فقط، ويوجد توزيع عيى مختلف عن الآخر لكل حجم عينة N واحد الطرق لاشتقاق التوزيعات للعينة هى نابعة من اعتبارات الاحتمالات الأساسية. وعلى ذلك يمكن اشتقاق التوزيع العيى لأى إحصاء بواسطة:

- تحديد كل العينات المختلفة المحتملة بأحجام N حيث تتكون من درجات المجتمع.
- حساب الإحصائية (المتوسط مثلاً) لكل عينة من العينات المختلفة.
- حساب احتمال الحصول على كل قيمة للإحصاء لو تم إجراءه بالصدفة.

يوجد عاملان يحددان شكل التوزيع العيى للمتوسط هما شكل الدرجات الخام فى المجتمع وحجم العينة، فإذا كان شكل الدرجات فى المجتمع اعتدالى فإن التوزيع العيى 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 كبيراً فإن شكل التوزيع فى العينة يكون تقريباً اعتدالى لحجم عينة $N \geq 30$ فإن النظر إلى شكل التوزيع لمجتمع الدرجات الخام غير ضرورى لأن توزيع العينة يكون تقريباً اعتدالياً، ولذلك فإن معظم المجتمعات فى العلوم السلوكية لا تختلف كثيراً عن الاعتدالية إذا كانت حجم العينة $N \geq 30$ من هنا تأتى مقولة "يفضل استخدام الإحصاء البارامترى إذا كان حجم العينة كبيراً وهذا يتحدد لحجم العينة 30 فأكثر.

توجد ثلاثة أساسيات لتحديد التوزيع العيني هي المتوسط والانحراف المعياري أو التباين وشكله.

هدف اختبار Z

يهدف إلى معرفة معالم الفرض الصفري في المجتمع أو اختبار فروض حول متوسط المجتمع أي تحديد ما إذا كانت للمعالجة أثر على أفراد المجتمع علمًا بأن المتوسط والانحراف المعياري للمجتمع يكون معروف قبل إجراء التجربة، ويهدف اختبار Z أيضًا إلى المقارنة بين بيانات الدراسة الحالية بنتائج دراسة أخرى التي تعتبر بمثابة معلم المجتمع.

اختبارات الفروض لقضية بحثية (في: Pagano, 2013)

قام باحث بإعداد برنامج لتنمية مهارة القراءة في مجتمع المرحلة الثانوية في محافظة أو مدينة ما علمًا بأن معالم درجات المجتمع في مادة القراءة كانت كالآتي : $\sigma = 16$ ، $\mu = 75$ وقام بتطبيق البرنامج لمدة شهر على عينة عشوائية مكونة من 100 طالب في المرحلة الثانوية . وتسأل هل درجات القراءة بعد تطبيق البرنامج تزيد عن 75 أو هل متوسط درجات العينة بعد تطبيق البرنامج يختلف اختلافاً جوهرياً عن متوسط المجتمع ؟

الخطوات البحثية:

1. مشكلة البحث: هل متوسط درجات القراءة للعينة بعد تطبيق البرنامج يختلف عن متوسطها في المجتمع ؟ أو هل للبرنامج أثر على مهارة القراءة ؟
2. فرض البحث: يصاغ فرض البحث اما بصورة موجهة أو غير موجهة ويصاغ كالآتي: 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:

أ- منهج شبة تجريبي ويتضمن التصميمات الآتية:

• قياس قبلى وبعدى (المجموعة الواحدة)

• المجموعة الوحدة وقياس بعدى.

ب- منهج وصفى يتضمن جمع البيانات على متغير واحد.

4. النموذج الإحصائي: نموذج المتغير الواحد وإحصاء بارامترى والاختبار الإحصائي هو

اختبار Z حول متوسط مجتمع ما والانحراف المعياري معروف.

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية: بما أن المعلمة هي متوسط المجتمع فإن :

$$H_0: \mu = 75$$

$$H_A: \mu \neq 75$$

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 أخرى 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 الإحصائي ومسلماته: الاختبار هو Z حول متوسط مجتمع وتكون صيغة حسابه

كالآتى:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

• $\bar{X}$ متوسط درجات العينة

• $\sigma_{\bar{x}}$ الانحراف المعياري للتوزيع العيني للمتوسط وهو يساوى الخطأ المعياري للمتوسط ولذلك يطلق عليها الخطأ المعياري.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \quad \text{وحيث إن:}$$

أذا:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$$

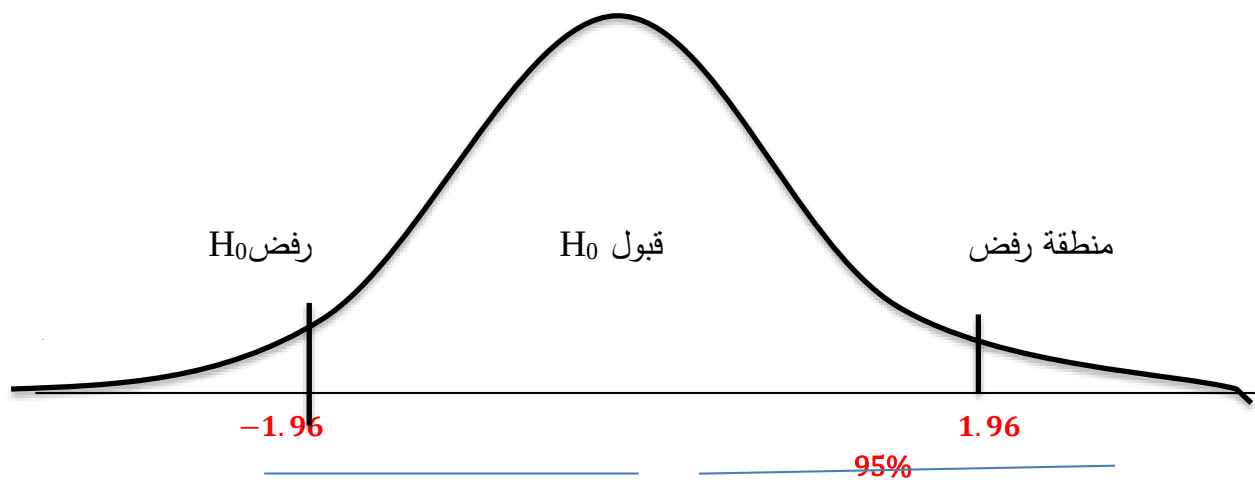
- σ الانحراف المعياري للعينة
- N حجم العينة.

ولهذا الاختبار مسلمات هي:

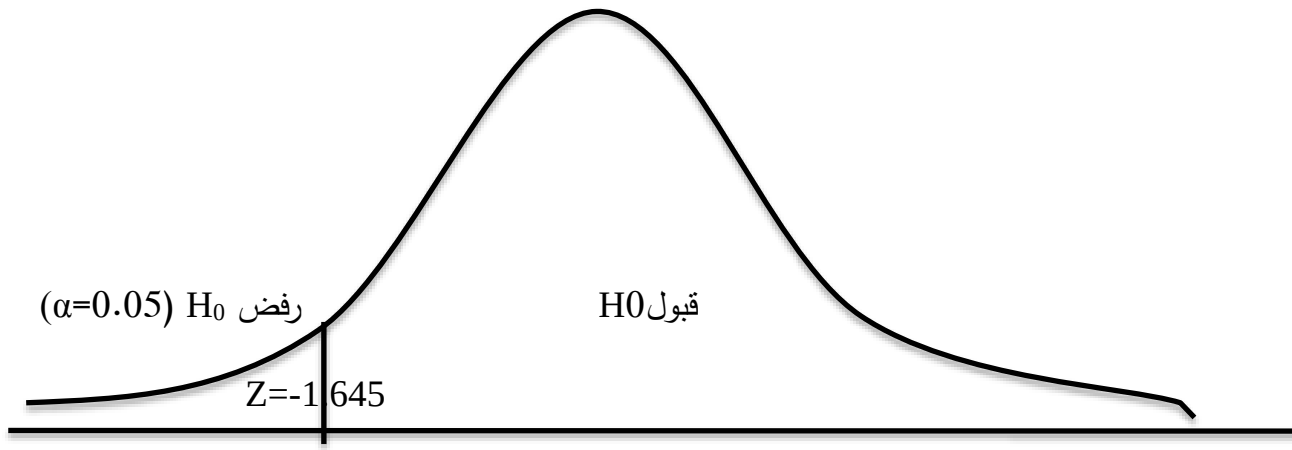
- المعاينة العشوائية.
- مستوى القياس فترى على الأقل.
- التوزيع العيني لبيانات العينة اعتدالى.
- يستخدم لاختبار فروض حول متوسط مجتمع إذا كان متوسط المجتمع والانحراف المعياري معروفين.
- يفضل استخدامه لعينة $N \geq 30$ وذلك للحصول على التوزيع الإعتدالى.
- استقلالية القياسات بمعنى أن درجات الأفراد مستقلة عن بعضها البعض.

3. مستوى الدلالة الإحصائية وقاعدة القرار: تبني الباحث مستوى دلالة إحصائية $\alpha = 0.05$ والتوزيع العيني لاختبار Z هو اعتدالى، وهى كما يراها Pagano (2013) هى المنطقة تحت المنحنى التى تتضمن كل القيم لإحصائية للاختبار التى تسمح برفض الفرض الصفري، ومنطقة الرفض هى منطقة التوزيع العيني التى تعكس تلك القيم لمتوسط العينة التى تكون غير محتملة أو غير صحيحة إذا كان الفرض الصفري حقيقى والقيم الحرجة للاختبار هى تلك القيم فى التوزيع العيني 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 الاعتنالى لـ Z كالآتى:



أما إذا كان الاختبار ذا ذيل واحد لأدنى:



وتكون قاعدة القرار كالآتى: إذا كانت Z (المحسوبة) $Z \leq$ (الحرجة) نرفض الفرض الصفرى.

4. حساب إحصائية الاختبار: وبعد حساب الآتى: متوسط المجتمع $\mu = 75$ ،
متوسط العينة $\bar{x} = 72$ ، الانحراف المعياري للعينة $\sigma = 16$

إذا الخطأ المعياري:

$$\sigma_{\bar{x}} = \frac{16}{\sqrt{100}} = 1.6$$

إذا:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{72 - 75}{1.6} = -1.88$$

5. القرار والتفسير: Z المحسوبة (-1.88) > Z الجدولية (1.96)، اعتمد على القيمة المطلقة وعليه نقبل الفرض الصفري بمعنى لا توجد فروق ذات دلالة إحصائية بين متوسط العينة ومتوسط المجتمع أو بكلمات أخرى: البرنامج ليس له أثر فى تنمية مهارة القراءة لدى أفراد عينة طلاب المرحلة الثانوية. ولكن إذا تم تبني الفرض البديل الموجه فإن القرار يكون اختبار ذو ذيل واحد: Z المحسوبة (-1.88) < Z الجدولية (1.645)، وعليه يتم رفض الفرض الصفري بمعنى أنه توجد فروق ذات دلالة إحصائية بين متوسط العينة ومتوسط المجتمع لصالح المتوسط الأدنى وهو المجتمع وعليه فإن البرنامج له أثر فعال فى زيادة مهارة القراءة.

القرار فى ضوء فترات الثقة

صناعة القرار من خلال حساب فترات الثقة لاختبار Z كما أوردها Aron et al. (2013) لمستوى $\alpha = 0.05$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = 1.6 \quad \text{- حساب الخطأ المعياري لتوزيع المتوسطات :}$$

- فترات الثقة 95% أضف 1.96 مضروباً فى الخطأ فوق وتحت متوسط العينة:

$$\text{Upper limit} = \bar{x} + 1.96(\sigma_{\bar{x}}) \quad (\text{الحد الأعلى})$$

$$\text{Lower limit} = \bar{x} - 1.96(\sigma_{\bar{x}}) \quad (\text{الحد الأدنى})$$

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 الموضوعية في الفرض الصفري (75) تقع بين الحد الأعلى والأدنى ومن ثم يتم قبول الفرض الصفري وبالتالي تكون استراتيجية القرار كالآتي:

الموقف	القرار	التفسير
الإحصائية في منطقة الرفض H_0	رفض الفرض الصفري	توجد فروق دالة إحصائية
الإحصائية ليست في منطقة الرفض	فشل رفض الفرض الصفري	لا توجد فروق دالة إحصائية

6. حجم التأثير: يقدر حجم التأثير لاختبار Z كما عرضها Cohen (1988) كالآتي:

$$d = \frac{\text{مجتمع } \bar{x} - \mu}{\sigma} = \frac{\text{لا معالجة } \mu - \text{معالجة } \mu}{\sigma} = \frac{\text{فروق المتوسطات}}{\text{الانحراف المعياري}}$$

$$= \frac{72-75}{1.6} = 0.187$$

وتكون حدوده وفقاً لـ Cohen (1988) كما يلي:

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 في حالة النتائج غير الدالة إحصائياً إلا أنه يفضل استخدامها لتحديد الحجم المطلق لتأثير المعالجة لكي يسترشد بها الباحثين في دراساتهم السابقة وإجراء دراسات ما وراء التحليل.

7. القوة الإحصائية: لتحديد القوة الإحصائية للمثال السابق باستخدام برنامج G-Power اتبع

الآتي:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. تحت Type of power analysis اختيار:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

لاحظ أننا اعتمدنا على التحليل البعدي Post hoc

3. تحت Test family اختيار Z- tests

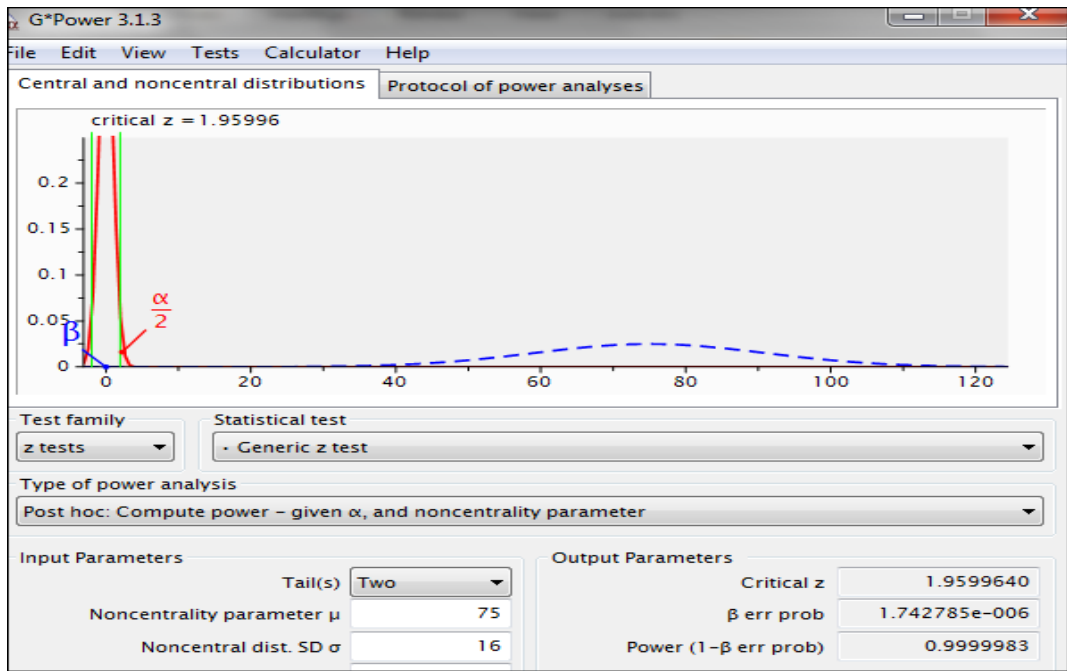
4. تحت Statistical test اختيار:

Test family	Statistical test
z tests	• Generic z test

5. ادخل المعالم Input parameters:

Input Parameters	
Tail(s)	Two
Noncentrality parameter μ	75
Noncentral dist. SD σ	16
α err prob	0.05

6. اضغط Calculated تظهر شاشة الناتج كالاتي:



يتضح أن القوة الإحصائية = 0.999 وهو مستوى عالٍ جداً.

كتابة نتائج Z في الدراسات المنشورة وفقاً لـ APA

اتضح أن البرنامج ليس له أثر دال إحصائياً 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 في المواقف البحثية التي يكون فيها المتوسط والانحراف المعياري للمجتمع معروف، وهذه النوعية من المواقف نادرة نسبياً ولكن المواقف البحثية الأكثر شيوعاً في البحث هي تلك التي يكون متوسط الفرض الصفرى للمجتمع معلوم والانحراف المعياري غير معلوم، في هذه الحالة لا يمكن استخدام اختبار Z وبدلاً منه نستخدم T وهو 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 المعياري 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 قليلاً مما يجب أن تكون وقيمة P الاحتمالية أصغر مما يجب أن تكون وهذا التحيز يكون محدوداً إذا كان حجم العينة 30 فأكثر.

معادلات T و Z

T	Z
$T = \frac{\bar{x} - \mu}{s / \sqrt{N}}$ $= \frac{\bar{x} - \mu}{S_{\bar{x}}}$	$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}}$ $= \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

حيث $S =$ تقدير لـ σ و $S_{\bar{x}}$ تقدير لـ $\sigma_{\bar{x}}$ من العينة

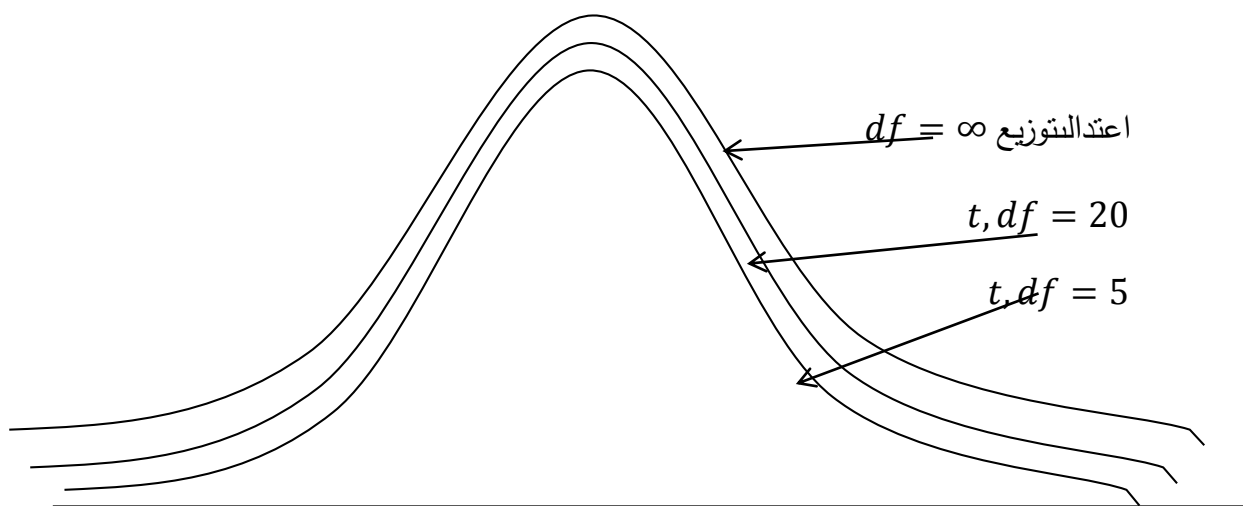
وعلى ذلك فإن الباحث يقدر الانحراف المعياري للمجتمع σ باستخدام الانحراف المعياري للعينة، ولكن هل استخدام S للعينة لتقدير σ لمجتمع يؤثر على الاختبار الإحصائي؟ بكل تأكيد خاصة إذا كانت أحجام العينات صغيرة فإن التوزيع العيني للعينة يكو غير مناسب، وتوزيع T هو عائلة من التوزيعات المنتظمة أو الاعتدالية التوزيع 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 في ضوء دراسات سابقة.

- اختبار فروض حول متوسط مجتمع ما محدد مسبقاً مثل متوسط التحصيل مثلاً أو أى متغير آخر فى مجتمع ما معروف.

التوزيع العينى لـ T: يتم الحصول على التوزيع العينى لـ T من 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 تقريباً اعتدالى مثل توزيع Z ويتحدد هذا التوزيع عن طريق درجات الحرية، وتقترب اعتدالية T من Z إذا كان حجم العينة كبير، ودرجات الحرية هو مفهوم رياضى حيث عدد درجات الحرية للإحصاء يعتمد على حجم العينة ودرجات الحرية لاختبار $df=N-1:T$



وكما نرى أن التوزيع يقترب تماماً من الاعتدالية إذا زادت درجات الحرية ويكون اعتدالى بدرجة تامة إذا كانت $df > 120$ وتوزيع T هو 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، علماً بأن متوسط المجتمع $\mu=13$ (شهرًا) واختار الباحث عينة عشوائية مكونة من 15 طفلاً وكانت بيانات العينة كالاتى:

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
Σ 165	1971	0	156

وأراد معرفة ما إذا كان للبرنامج أثر في التفكير 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 التدريبي أثر على التفكير بالمرحلة الزمنية لنطق الكلمات عند الأطفال حديثي الولادة أو توجد فروق بين متوسط عمر العينة عن متوسط عمر المجتمع (13).

3. متغيرات البحث: المرحلة العمرية التي يبدأ عندها النطق: متغير نسبي - متصل - تابع.

4. النموذج الإحصائي: نموذج المتغير الواحد (الصفري) - إحصاء بارامترى، والاختبار الإحصائي المناسب هو T لعينة واحدة.

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية:

$$H_0 : \mu = 13$$

$$H_A : \mu \neq 13$$

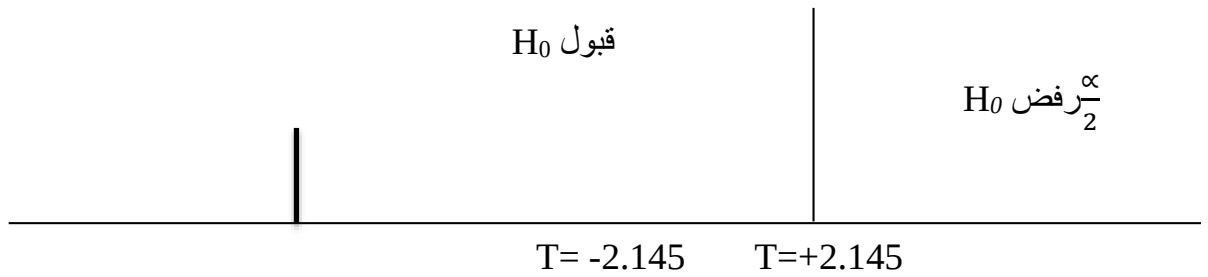
2. الاختبار الإحصائي المناسب ومسلماته: هو اختبار T لعينة واحدة ويتحدد بالصيغة:

$$T = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

والاختبار ذو ذيلين و $df = N - 1 = 15 - 1 = 14$ ، ومسلماته:

- **الاعتدالية:** بيانات العينة تخضع للتوزيع الاعتدالي وهذه المسلمة ضرورية في حالة استخدام T للعينات الصغيرة ولكن إذا كانت $N \geq 30$ فإن الخطأ المعياري يكون صغيراً وهذه المسلمة أقل تأثيراً على النتائج.
- **المعاينة العشوائية:** البيانات تأتي من عينة عشوائية من المجتمع حتى نثق في النتائج ونعممها ومن غير المناسب استخدام اختبارات الفروض مع عينات غير عشوائية.
- **الاستقلالية:** نفترض أن أى قياسات للأفراد مستقلة عن بعضها البعض.

3. **الدلالة الإحصائية وقاعدة القرار:** افترض الباحث أن $\alpha = 0.05$ ويتم قسمة مستوى الدلالة الإحصائية على جانبي المنحنى $\frac{\alpha}{2}$ ، ودرجة الحرية $df = 14$ ، واختبار ذو ذيلين ، وبالبحت في 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 الصفري ، وإذا كان الاختبار ذو ذيل واحد تكون منطقة الرفض أصغر عند $\alpha = 0.05$, $df=14$

4. حساب إحصائية الاختبار: بما ان: $\bar{x} = 11$ و $S = 3.34$ إذا:

$$T = \frac{11 - 13}{3.34 / \sqrt{15}} = -2.32$$

أشار (Pagano (2013 إلى تقدير T من خلال الدرجات الخام كالآتي :

$$T = \frac{\bar{x} - \mu}{\sqrt{\frac{SS}{N(N-1)}}}$$

حيث إن:

$$SS = \sum (x - \bar{x})^2 = \sum x^2 - \frac{\sum (x)^2}{N}$$

$$= 1971 - \frac{(165)^2}{15} = 156$$

$$T = \frac{11 - 13}{\sqrt{\frac{156}{15 \times 14}}} = -2.32$$

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 الصفري لأن القيمة المحسوبة

وقعت في منطقة رفض الفرض الصفري وبالتالي توجد فروق ذات دلالة إحصائية بين متوسط

أعمار العينة عند النطق بالكلام عن متوسط عمر المجتمع أو أن التجربة أو البرنامج لها أثر في التذكير بنطق الكلمات عند الأطفال حديثي 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 (1988) Cohen:**

$$d = \frac{\bar{x} - \mu}{S}$$

- d قيمة حجم التأثير
- S الانحراف المعياري لبيانات العينة

إذا:

$$d = \frac{(11 - 13)}{3.34} = 0.60$$

وفى ضوء حدود ومعايير كوهين 0.2 - 0.0 ضعيف جداً و 0.79 - 0.5 متوسط و 0.80 $d \geq$ كبير وعليه فإن حجم التأثير من النوع المتوسط وفقاً لمعايير كوهين لتفسير حجم التأثير ويتم تفسيرها أيضاً 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 F وتسمى مؤشر إيتا تربيع η^2 كالاتى:

$$r^2 \text{ أو } \eta^2 = \frac{T^2}{T^2 + df}$$

• T هي 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 الزمني لبداية الكلام عند الأطفال ويفسر في ضوء معايير كوهين حجم تأثير كبير (صغير 0.01 و متوسط 0.09 وكبير 0.25 فاعلي). 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 أوميغا ω^2 : وتحدد قيمته كما يلي :

$$\omega^2 = \frac{T^2 - 1}{T^2 + df}$$

$$\omega^2 = \frac{(-2.32)^2 - 1}{(-2.32)^2 + (15 - 1)} = \frac{4.382}{19.382} = 0.260$$

حدد (1988) Cohen 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 الآتي:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. أسفل Type of power analysis اختيار:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

لاحظ أننا اعتمدنا على التحليل البعدي Post hoc

3. أسفل Test family اختبار t- tests

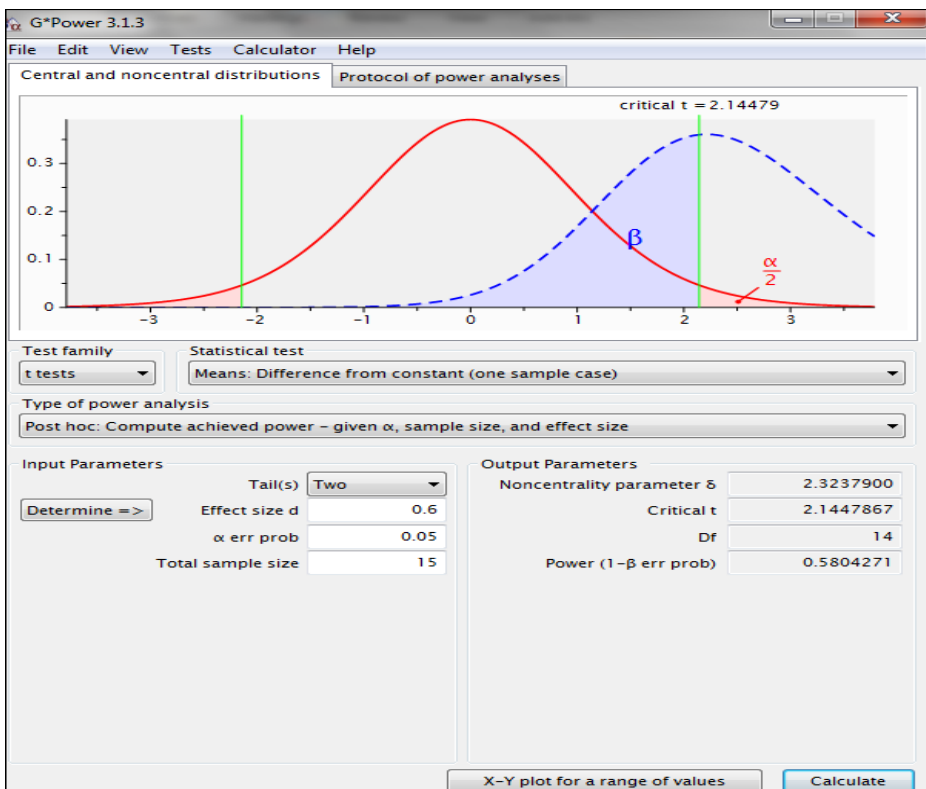
4. أسفل Statistical test اختيار:

Test family	Statistical test
t tests	Means: Difference from constant (one sample case)

5. ادخل المعالم Input parameters:

Input Parameters	
Determine =>	Tail(s) Two
Effect size d	0.6
α err prob	0.05
Total sample size	15

6. اضغط على Calculated يعطى شاشة الناتج كالآتي:



يتضح أن القوة الإحصائية = 0.58 وهذا حجم قوة غير مناسب إذ 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. فمثلاً فترة الثقة 95% هي فترة يمكن أن تتضمن قيمة الإحصائية في المجتمع في هذه الفترة باحتمال 95%.

بناء 95% فترة الثقة: حد الثقة الأدنى: $CI_{\text{الادنى}} = \bar{X} - s_{\bar{X}} T_{c.v}$

• $T_{c.v}$ القيمة الحرجة أو الجدولية

• $\bar{X}$ متوسط العينة.

حد الثقة الأعلى: $CI_{\text{الأعلى}} = \bar{X} + s_{\bar{X}} T_{c.v}$

بالبحث في جدول T عند $\alpha = 0.05$ ، $df = 15 - 1 = 14$ ، و 95% واختبار ذو ذيلين فإن:

$$CI_{\text{الادنى}} = 11 - \frac{3.34}{\sqrt{15}} (2.145)$$

$$CI_{\text{الأعلى}} = 11 + \frac{3.34}{\sqrt{15}} (2.145)$$

إذًا حدّي الثقة هما ()

بالتالى فإن قيمة متوسط المجتمع (القيمة المفترضة في الفرض الصفرى) تقع خارج هذه الفترة وعليه نرفض H_0 وكلما كانت فترات الثقة كبيرة كلما زادت الثقة أن يكون متوسط المجتمع يقع في هذا المدى.

مثال آخر لحساب فترات الثقة

اهتم باحث بتحديد متوسط وزن عمر الأطفال فى سن العام ونصف واختار 15 طفلاً،
وقدر متوسط الوزن للعينة 7.2، والانحراف المعياري 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 $\alpha = 0.025$ ، $df = 15 - 1 = 14$ إذا T الحرجة = 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% هي (6.93 - 7.47).

تنفيذ اختبار T لعينة واحدة فى SPSS

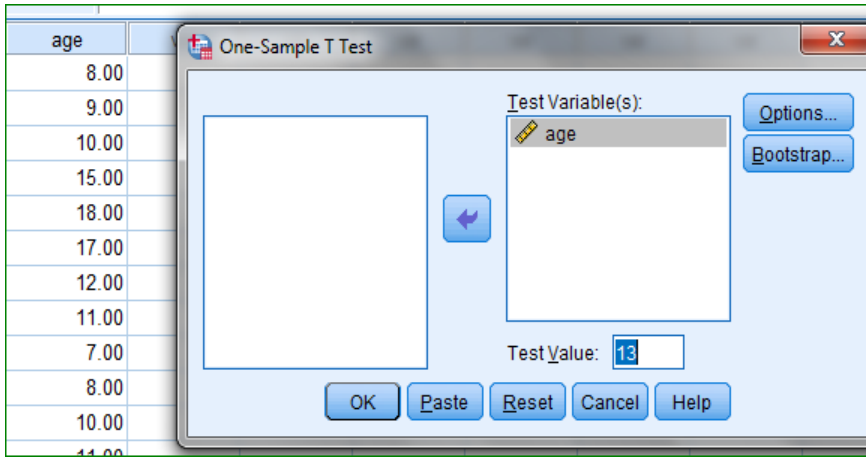
أولاً: إدخال البيانات: 1. اضغط على Variable view ثم اكتب اسم المتغير فى عمود

Name العمر: age

2. اضغط على Data view، ابدأ فى إدخال البيانات

ثانياً : تنفيذ الأمر: 1. اضغط Analyze→Compare Means→One sample T test

تظهر الشاشة الآتية:



2. انقل age إلى مربع Test variables

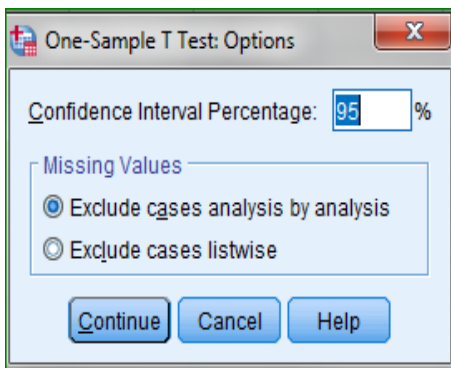
3. في مربع Test value اكتب القيمة المراد

اختبار الفروض حولها وهي 13

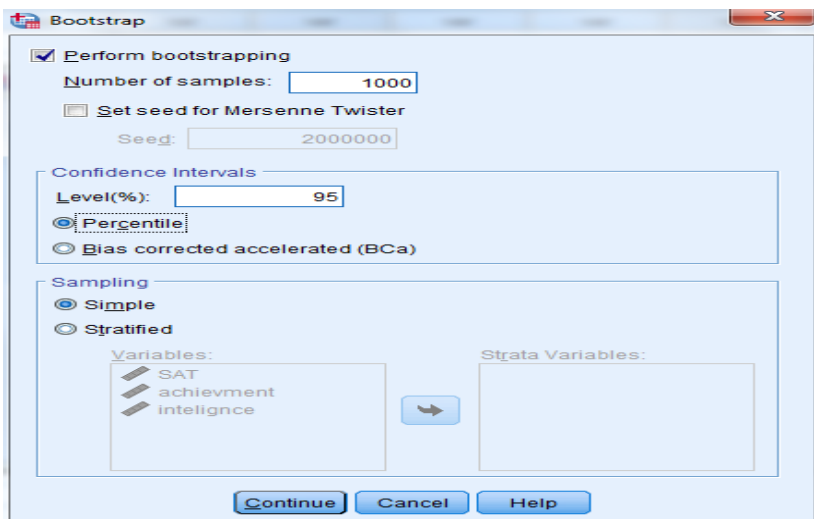
4. اضغط على اختيار Options واكتب حدود

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 مثلاً 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 فمثلاً فترات الثقة 95% فإنه يعتمد على المئينى Percentile 5th و Percentile 95th 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.

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

8. اضغط Continue ثم OK

ثالثاً: تفسير المخرج : الجدول الأول

كالآتى: المتوسط Mean = 11، والانحراف المعياري 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 (Sig. (2.tail) (P) = 0.036 وعليه فان: $0.05(\alpha) < 0.036(p)$ بالتالى نرفض H_0 أى أن متوسط أعمار

الأطفال يختلف عن متوسط المجتمع 13. بمعنى للبرنامج أثر في التكبير في المدة الزمنية للنطق.

ويمكن صناعة القرار في ضوء فترات الثقة للفروق وهي:

$$CI_{95} = (- 3.848 , - 1.514)$$

وبما أن فروق المتوسطات $\text{Mean difference} = -2.00$ وعليه فإنها تقع في مدى حدود الثقة بمعنى 95% من العينات سوف تعطى قيمة فروق تقع في هذا المدى وهذا يعطينا ثقة عالية في قيم إحصائيات العينات كتقدير لقيمة المجتمع غير المعروفة وفي الدلالة الإحصائية وعليه توجد دلالة إحصائية وبالتالي نرفض H_0 .

نتائج البوتستراب:

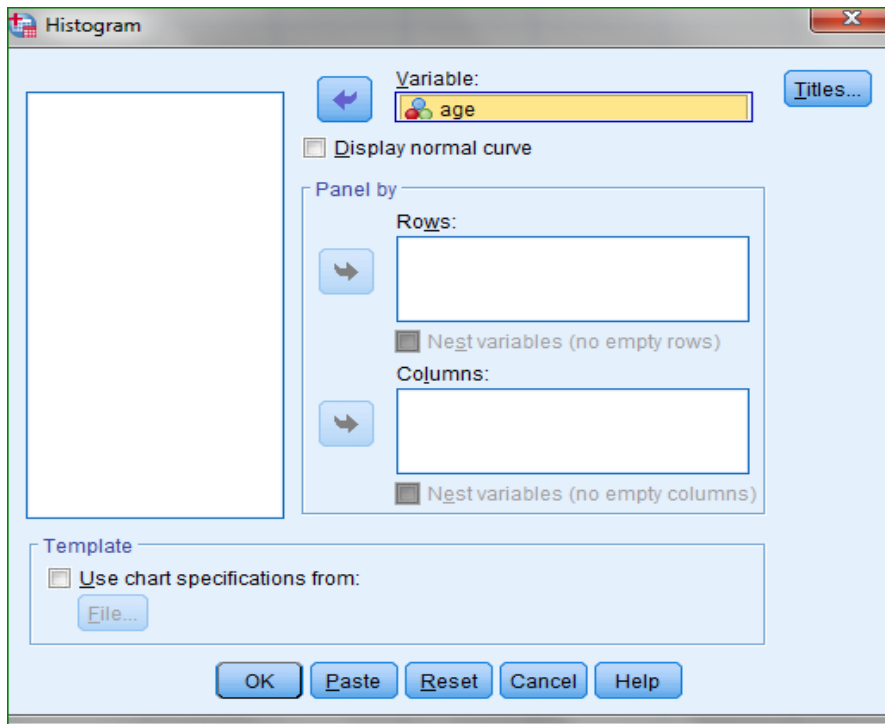
Bootstrap for One-Sample Test					
	Mean Difference	Bootstrap ^a			
		Bias	Std. Error	95% Confidence Interval	
				Lower	Upper
age	-2.00000-	-.03447-	.84455	-3.60000-	-.26667-

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

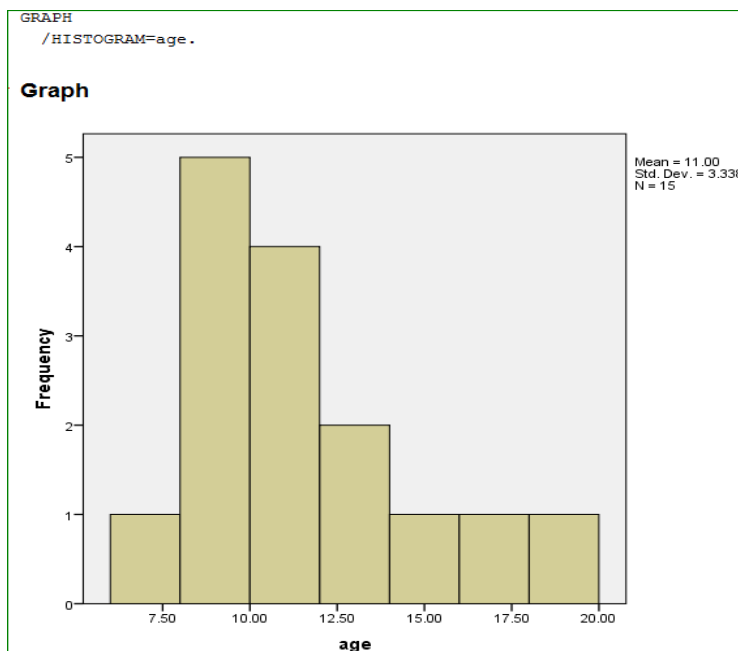
يتضح أن متوسط الفروق هو نفسه كما في اختبارات الفروض الصفرية ولكن اختلفت فترات الثقة 95% وكانت اضيق مما يدل انها أكثر دقة من التعامل مع العينة العشوائية الواحدة والقرار هو رفض الفرض الصفرى.

عرض النتائج في ضوء المدرج التكراري

1. اضغط Graphs→Legacy Dialogs →Histogram



2. انقل المتغير age إلى مربع Variable اضغط OK يعطى المدرج كالاتى:



كتابة نتائج T عينة واحدة وفقا لـ APA

اتضح من نتائج اختبار T وجود فروق دالة إحصائية بين متوسط اعمار العينة عند النطق بالكلام ومتوسط اعمار المجتمع:

$$T(14) = -2.32, p < 0.05, d = 0.60, CI95 (- 3.848, - 1.514)$$

عند استخدام اختبار T لتقييم ما إذا كان عمر الذين يتلقون البرنامج عند النطق يختلف إحصائيًا عند 0.05،، واتضح أن متوسط العينة 11 دال إحصائيًا عند 0.05.

$$T(14) = -2.32, p < 0.05, r^2 =$$

$$T(14) = -2.32, p = 0.004, d = 0.6$$

حيث قيمة p هي القيمة الاحتمالية لمخرج SPSS

وأن فترات الثقة 95% للمتوسط تتراوح بين — و — وأن حجم التأثير $d=0.60$ وهو من النوع المتوسط والشكل الآتي 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 المعياري σ معلوم لدرجات المجتمع أو فى حالة σ غير معروفة واستخدمنا اختبارى T و 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 وبالتالي 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 π)

وفى هذا الإطار فإن الصيغة العامة للإحصاء المستخدم:

$$= \frac{\text{إحصائية العينة} - \text{المعلم}}{\text{الخطأ المعياري للإحصائية}}$$

وعليه فإن الصيغة العامة لفترة الثقة:

$$CI = \text{Statistic} \pm \text{critical value (SE)}$$

- SE هى الخطأ المعياري
- Statistic قيمة إحصائية الاختبار المحسوبة.
- $\text{critical value (cv)}$ 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، $N \geq 30$ أما إذا كانت العينة صغيرة فيفضل استخدام الاختبار اللابارامترى Binomial 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: اختبار Z ، حيث:

$$Z = \frac{P_s - P_u}{\sqrt{S_p}}$$

- P_s نسبة العينة، P_u نسبة المجتمع أو الفرض الصفري.
- N حجم العينة
- S_p الخطأ المعياري وتتحدد بالاتي:

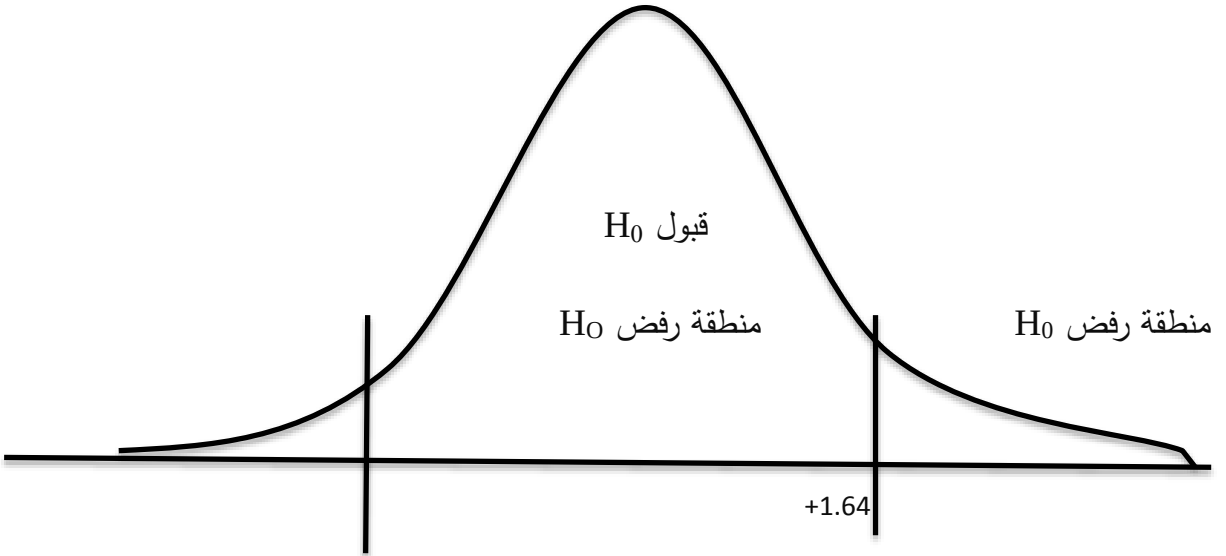
$$S_p = \sqrt{\frac{pq}{n}}$$

- P هي نسبة العينة
- $q = (1-P)$ نسبة المجتمع التي ليس لديه السمة (نسبة الامية).

ومسلماته هي نفس مسلمات اختبار Z .

3. الدلالة الإحصائية وقاعدة القرار: اختبر الباحث الفرض عند $\alpha = 0.10$ ، واختبار ذو ذيلين

وبالبحث في جداول Z يتضح أن: $Z = \pm 1.65$ الحرجة



4. الحسابات: حيث: $N = 122$, $q = 1 - 0.39 = 0.61$, $P = 0.39$

$$z = \frac{0.43 - 0.39}{\sqrt{\frac{0.39 \times 0.61}{122}}} = \frac{0.04}{0.044} = 0.91$$

5. القرار والتفسير: Z المحسوبة ($Z > 0.91$) الحرجة (1.65)، وهو أن نسبة الأمية في المحافظة لا تختلف عن نسبتها في عموم البلاد.

فترتي الثقة:

$$CI\ 90 = P \pm Z\ crit\ (Sp) = \\ 0.43 \pm (0.044)(1.65)$$

مثال آخر: قام مركز الوحدة المحلية بمسح لتحديد الخدمات التي يحتاجها اهالى قرية ما تتضمن 120000 مواطن اختيرت عينة من 400 مواطن وافترض فريق البحث أن 30% من سكان المنطقة يحتاجون دار حضانة للأطفال تحت سن السادسة وأشار 88 من أفراد العينة أنهم بحاجة إلى دار حضانة، وأرادوا التحقق ما إذا كانت توجد فروق بين نسبة المجتمع ونسبة العينة في مدى احتياجهم لدار حضانة.

خطوات اختبارات الفروض

1. الفروض الإحصائية: $H_0 : P = 30\% = 0.30$

$$H_A: P \neq 0.30$$

2. الاختبار الإحصائي المناسب ومسلماته: اختبار Z وصيغته كالاتى:

$$z = \frac{P_s - P_u}{Sp}$$

$$S_p = \sqrt{\frac{pq}{n}}$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: حيث $\alpha = 0.01$ ، واختبار ذو ذيلين إذا: $Z_{c.v} = \pm 1.96$ ، حيث إذا كانت:

(المحسوبة) $Z < Z_{c.v}$ نرفض الفرض الصفري.

$$P = \frac{88}{400} = 0.22 \quad 4. \text{ الحسابات:}$$

$$S_p = \sqrt{\frac{(0.22)(0.78)}{400}} = 0.02$$

$$Z = \frac{0.22 - 0.30}{0.02} = -4.00$$

5. القرار والتفسير: بما ان Z المحسوبة (-4.00) $< Z$ الحرجة (1.96)، إذا نرفض الفرض الصفري وعليه توجد فروق ذات دلالة إحصائية بين نسبة المجتمع ونسبة العينة.

القرار فى ضوء فترتى الثقة

بما أن $\alpha = 0.05$ إذا فترات الثقة تكون 95% ($1-0.05$) ويتم بناءها حول نسبة العينة وعليه:

$$\begin{aligned} CI_{95} &= P \pm 1.96 (S_p) \\ &= 0.22 \pm 1.96 (0.02) \\ &= 0.22 \pm 0.04 \\ &= (0.18, 0.26) \end{aligned}$$

الأعلى , الأدنى

وعليه فإن قيمة Z المحسوبة أو القيمة المتضمنة فى الفرض الصفري (0.30) تقع خارج هذه الفترة وعليه نرفض الفرض الصفري 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

تحويل فيشر Fisher's z Transformation

يهدف اختبار تحويل أو تعديل فيشر Z إلى اختبار فروض حول معامل ارتباط بين متغيرين فى المجتمع ومعرفة ماذا كانت الارتباط فى العينة يختلف اختلافاً جوهرياً عن معامل الارتباط فى المجتمع.

اختبارات الفروض لقضية بحثية (Hinkle et al., 1994)

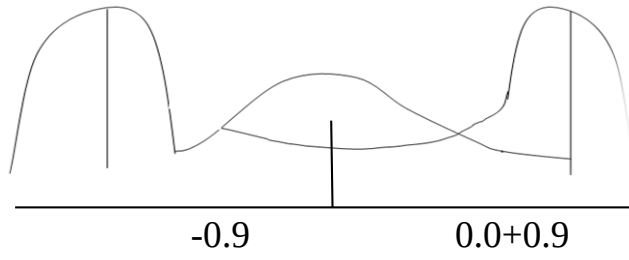
أراد مجموعة من علماء الاجتماع دراسة العلاقة بين دخل الأسرة واتجاهاتهم نحو تربية الأطفال ووجدوا أن التراث البحثى توصل فى معظمه أن العلاقة تدور حول 0.65 وأراد اختبار ما إذا كانت العلاقة بين المتغيرين لدى أفراد العينة المكونة من 30 عائلة تختلف عن العلاقة الموجودة بين المتغيرين فى الدراسات السابقة.

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: نسبي - متصل، والاتجاه: فترى - متصل

4. منهج البحث: منهج ارتباطى.

5. النموذج الإحصائى: هو البسيط والإحصاء بارامترى 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 العيني لمعامل الارتباط عندما $r = (-0.9, 0.0, 0.9)$

خطوات اختبارات الفروض الصفرية

1. الفروض الإحصائية: $H_0: \rho = 0.65$

$H_A: \rho \neq 0.65$

والاختبار ذو ذيلين.

2. الاختبار الإحصائي المناسب و مسلماته: يقدر من الصيغة الآتية:

$$\frac{Z_r - Z_{\rho}}{S_{Z_r}}$$

- Z_{ρ} قيمة تحويل Z لمعامل الارتباط في المجتمع
- Z_r قيمة تحويل Z لمعامل الارتباط في العينة
- S_{Z_r} الخطأ المعياري لـ Z

ولتحويل معامل الارتباط إلى Z كالاتي (تحويل فيشر Z):

$$Z_r = \left(\frac{1}{2}\right) \log e \left(\frac{1+r}{1-r} \right)$$

3. مستوى الدلالة الإحصائية والتوزيع العيني وقاعدة القرار: اختبر الباحث هذا

الفرض عند $\alpha = 0.10$ والتوزيع العيني لمعامل الارتباط يقدر مثل التوزيع العيني للمتوسط وهو التوزيع العيني لمعاملات الارتباط المحسوبة من عينات من المجتمع ويتغير التوزيع العيني ووفقاً لقيمة معامل الارتباط وإشارته وكلما زاد ومعامل الارتباط فإن التوزيع العيني يكون ملتوي ناحية اليمين للمعاملات السالبة وناحية اليسار للمعاملات الموجبة وعندما يكون قريب من الصفر فإن التوزيع العيني يكون

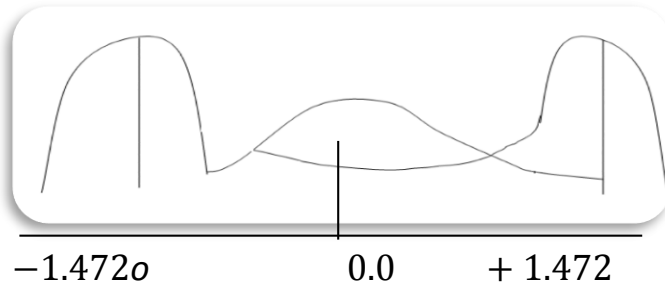
تقريباً اعتدالى، وتحسن الاعتدالية بزيادة حجم العينة ولان التوزيع العينى يكون ملتوى كلما زادت قيمة معامل الارتباط فإنه من غير المناسب استخدام التوزيع الاعتدالى لاختبارات الفروض حول ρ إلا إذا كانت $\rho = 0$ وهذا مما أدى بـ

Fisher بإجراء تعديل Z

وحيث إن الخطأ المعياري لـ Z:

$$S_{Zr} = \sqrt{\frac{1}{n-3}}$$

ولذلك فإن التوزيع العينى لمعاملات الارتباط المحورة أو المعدلة كالاتى:



الشكل (2.17) : التوزيع العينى لمعامل الارتباط المعدل Zr لـ $\rho = -0.95, 0.0, +0.90$

ولتحديد قيمة Zr الجدولية فإنه يتم حساب معامل الارتباط لبيانات العينة وبفرض أن قيمته $r = 0.61$ ، ويجب تحويل قيمة $\rho = 0.65$ (فى المجتمع) و $r = 0.61$ (فى العينة) إلى تحويل Z لاختبار فروض حولها. بالبحث فى جدول (انظر ملحق جدول تحويل فيشر) فإن تحويل قيمة ρ عن طريق قراءة العمود r حتى نصل إلى 0.65. فإن قيمة فيشر المقابلة: $Z\rho = 0.775$

وكذلك بالنسبة لـ 0.61 فتكون قيمة فيشر المقابلة لها: $Zr = 0.759$

وعليه $\rho = 0.65 \longrightarrow Z\rho = 0.775$

$r = 0.61 \longrightarrow Zr = 0.759$

وبتقدير الخطأ المعياري:

$$S_{Zr} = \sqrt{\frac{1}{n-3}}$$

$$= \sqrt{\frac{1}{30-3}} = 0.192$$

وعلى ذلك فإنه لـ $\alpha=0.10$ واختبار ذو ذيلين فإن: $Z = \pm 1.645$ (الجدولية)

وإذا كانت Z المحسوبة $Z \leq$ الحرجة نرفض H_0 .

4. الحسابات:

$$Z = \frac{Z_r - Z_p}{S_{Zr}}$$

$$= \frac{0.759-0.775}{0.192} = - 0.344$$

5. القرار والتفسير: Z المحسوبة $|-0.344| < |1.645|$ الحرجة ، وعليه نقبل الفرض الصفري ويفشل الباحث في رفض H_0 .

بناء فترات الثقة: عند رفض H_0 أو قبوله فمن الأفضل بناء فترات الثقة لأنها تكون أكثر معلوماتية وبما أن $\alpha=0.10$ فإن:

$$= 1-0.1 = 0.90$$

وذلك يتم بناء 90% فترة الثقة:

$$CI_{90} \text{ (الأعلى) } = Z_r + 1.645 (S_{Zr})$$

$$= 0.759 + 1.645 (0.192) = 1.025$$

$$CI_{90} \text{ (الادنى) } = 0.759 - 1.645 (0.192) = 0.393$$

وبالبحث في جدول Z فيشر عن قيم معامل الارتباط المناظرة لقيمتي فيشر 1.025 و 0.393 وهما كالآتي:

$$CI_{90} = (0.374, 0.773)$$

وعليه فإن الباحث فشل في رفض H_0 حيث لا فروق بين معامل الارتباط في العينة عن معامل الارتباط في المجتمع.

وفيما يلي معلومات عن اختبارات الفروض لعينة واحدة كما طرحها Hinkle et al. (1994):

الفرض الصفري	المعلم الإحصائية	SE	الاختبار	df	التوزيع العينى
$H_0: \mu = 0$	μ	$s\bar{x} = \frac{s}{\sqrt{n}}$	$\frac{\bar{x} - \mu}{s\bar{x}}$	N-1	T توزيع
$H_0: \mu = 0$	μ	$\sigma\bar{x} = \frac{\sigma}{\sqrt{N}}$	$\frac{\bar{x} - \mu}{\sigma\bar{x}}$	_____	Z
$H_0: \rho = q$	$z\rho$	$s_{zr} \sqrt{\frac{1}{N-3}}$	$\frac{zr - zp}{S_{zr}}$	_____	اعتدالى
$H_0: \rho = 0$	ρ	--	$T = r \sqrt{\frac{N-2}{1-r^2}}$	N-2	T
$H_0: P=q$	Pu	$sp = \sqrt{\frac{pq}{N}}$	$T = \sqrt{\frac{N-2}{S_p}}$	_____	اعتدالى

$\bar{x}$: المتوسط

وعرض (1994) Hinkle et al. ملخص (فترات الثقة) ولتقدير الفترة للمعالم لعينة واحدة كالآتى:

المعلم المقدر	الإحصائية	فترة الثقة
μ	$\bar{x}$	$\bar{x} \mp (T_{c.v}) \left(\frac{s}{\sqrt{n}} \right)$
Z	zr	$zr \mp z_{c.v} \left(\sqrt{\frac{1}{N-3}} \right)$
P	P	$P \mp z_{c.v} \left(\sqrt{\frac{pq}{N}} \right)$

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 شبه 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 مثلاً. وغالباً 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 أما 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 قبلية وبعدي 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 قبلية وقياس بعدى Pre and post design وبالتالي يوجد ارتباط منطقى فى البيانات لأن كل فرد يأخذ درجة فى الاختبار القبلى

ويأخذ درجة أيضاً 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 كالاتى:

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 الأكثر شيوعاً واستخداماً فى الدراسات البحثية وهذه النوعية من البيانات تتولد من تصميم بحثى مستقل أو تصميم بين الأفراد Between subject design ويكون التصميم البحثى كالاتى:

المتغير المستقل

ذكر أو تجريبية	انثى أو ضابطة
درجات	درجات
المتغير التابع	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 تقييم 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

التصميم هو ان نفس الأفراد يتعرضوا لنفس المعالجات. حيث إن تصميم القياسات المتكررة يمكن أن يكون كالاتى:

- 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:

نلاحظ أن اختبار T لعينات مرتبطة قائم على متوسطات فروق الدرجات difference scores وليس متوسطات الدرجات الخام كما هو الحال فى T لعينه واحدة. وفيما يلى مقارنة بين معادلات حساب T لعينة واحدة T لعينات مرتبطة، فاللعينة الواحدة:

$$T = \frac{\bar{x} - \mu}{(S_{\bar{x}})}$$

$$T = \frac{\bar{x} - \mu}{\sqrt{\frac{SS}{N(N-1)}}}$$

$$SS = \sum x^2 - \frac{(\sum x)^2}{N}$$

T المرتبطة:

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}}$$

$$SS_D = \sum D^2 - \frac{(\sum D)^2}{N}$$

- D الفرق بين الدرجات، $\bar{D}$ متوسط الفروق بين الدرجات
- μ_D متوسط الفروق بين الدرجات فى المجتمع

• S_D الانحراف المعياري لفروق الدرجات فى العينة.

• N عدد فروق الدرجات (حجم العينة).

اختبارات الفروض لقضيه بحثية (Pagano, 2013): اجرى باحث فى مجال علم النفس والتربية دراسة لاختبار ما إذا كان إشراف المعلم واتباع تعليماته يؤثر على الوقت الذى يقضيه الطالب فى القراءة الجيدة لموضوع ما فقام بقياس قبلى للوقت المستغرق لقطعة قراءة و ثم قدم تعليمات وقواعد للقراءة الجيدة ثم قاس بعدى وحسب الوقت المستغرق فى القراءة القبالية والبعدية وكانت النتائج كالاتى (وقت القراءة 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
Σ			120	3400

وأراد اختبار ما إذا كانت توجد فروق فى وقت القراءة لموضوع ما فى حضور

وغياب إرشادات المعلم عند مستوى دلالة 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 في متوسطات الوقت المستغرق في القراءة في حضور وفي غياب تعليمات المعلم والاختبار ذو ذيلين.

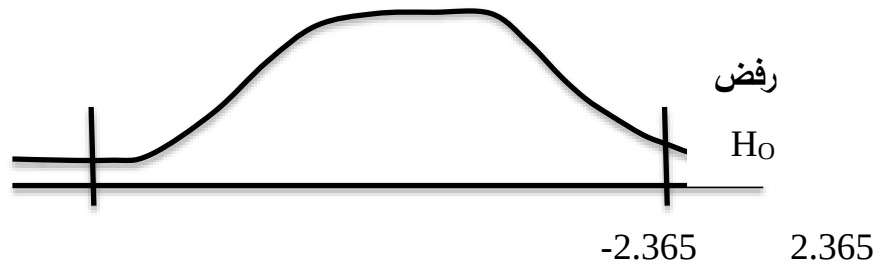
2. الاختبار المناسب ومسلماته:

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}}$$

$$df = N - 1 = 8 - 1 = 7$$

3. الدلالة الإحصائية وقاعدة القرار: إذا كانت $\alpha = 0.05$ واختبار ذو ذيلين و $df = 7$

بالبحث في جداول T فإن قيمة T الحرجة = ± 2.365



وإذا كانت T المحسوبة $T \leq$ الجدولية $|2.361|$ نرفض H_0

4. الحسابات:

$$\bar{D}(M_D) = \frac{\sum D}{N} = \frac{120}{8} = 15$$

$$S_{\bar{D}} = \sqrt{\frac{SS_D}{n-1}}$$

$$SS_D = \sum D^2 - \frac{(\sum D)^2}{N} \quad \text{حيث:}$$

$$SS_D = 3400 - \frac{(120)^2}{8} = 1600$$

$$S_D = \sqrt{\frac{1600}{8-1}} = \sqrt{228.57} = 15.12$$

وعليه فإن الخطأ المعياري لفروق الدرجات هي:

$$S_{\bar{D}} = 15.12 / \sqrt{8} = 5.35$$

وبالتالى :

$$T = \frac{\bar{D} - \mu_D}{S_{\bar{D}}} = \frac{D^- - 0}{\sqrt{SS_D/N}} = 15/(5.35) = 2.80$$

5. القرار والتفسير: حيث T المحسوبة (2.80) $T <$ الحرجة (2.365) ، وعليه نرفض H_0 وبالتالي توجد فروق ذات دلالة إحصائية بين زمن القراءة فى حضور المعلم وإتباع تعليماته وزمن القراءة فى عدم وجود المعلم، ولمعرفة اتجاهه فهى لصالح 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 يساوى 0.99 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:

$$r^2 = \frac{T^2}{T^2 + df}$$

إذاً إرشادات المعلم أو البرنامج فسرت 53% من تباين الوقت المستغرق فى القراءة، بمعنى أن البرنامج فسر 53% من تباين الوقت.

• مؤشر اوميغا ω^2 :

$$\omega^2 = \frac{T^2 - 1}{(T^2 + df)} = \frac{(2.804)^2 - 1}{(2.804)^2 + 7} = 0.46$$

إذاً 46% من تباين الوقت يرجع إلى التعليمات.

7. القوة الإحصائية: لتقدير القوة الإحصائية للمثال السابق باستخدام برنامج G-

Power اتبع الآتى:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. تحت Type of power analysis اختر:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

لاحظ أننا اعتمدنا على التحليل البعدي Post hoc

3. تحت Test family اختر t- tests

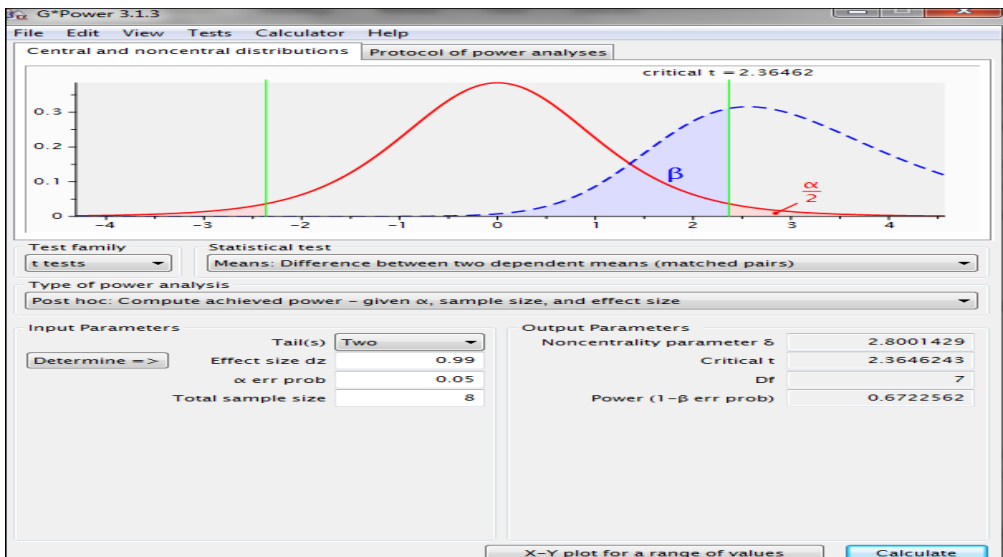
4. تحت Statistical test اختر:

Test family	Statistical test
t tests	Means: Difference between two dependent means (matched pairs)

5. أدخل المعالم Input parameters:

Input Parameters	
Tail(s)	Two
Noncentrality parameter μ	75
Noncentral dist. SD σ	16
α err prob	0.05

6. اضغط Calculated تظهر شاشة الناتج كالآتي:



يتضح أن القوة الإحصائية = 0.67 وهو مستوى قوة غير مرغوب وقد يرجع إلى صغر حجم العينة.

القرار في ضوء فترات الثقة

بما ان $\alpha = 0.05$ إذا فترات الثقة لـ $95\% (1 - 0.05)$:

$$CI_{95\%} = \bar{D} \pm T_{c.v}(S_D) = 15 \pm 2.365(5.35) \\ (2.360, 27.6394)$$

وعليه فإن القيمة المتضمنة في الفرض الصفري 0.0 وقعت خارج مدى فترات الثقة بالتالي نرفض H_0 .

كتابة نتائج اختبار T المرتبطة وفقاً لـ APA

تم إجراء الاختبار T المرتبطة القيم بما إذا كانت توجد فروق في الأوزان قبل وبعد البرنامج وأوضح النتائج ان متوسط فروق الدرجات $\bar{D} = 7.26$ ومتوسط درجات الاختبار القبلي $S_D = 5.02$ ، $\bar{x} = 83.23$ ومتوسط درجات الاختبار البعدي $\bar{x} = 90.49$ ، $S_D = 8.40$ ، وحيث $p < 0.05$ ، $T(16) = 4.18$ هي فروق دالة إحصائياً وان فترات الثقة لمتوسطات فروق الدرجات كانت $CI_{95} = (3.57, 10.95)$

$$....., T(7) = 2.804, p < 0.05 (d = 0.99)$$

$$... .., T_{(7)} = 2.804, p < 0.05 (\omega^2 = 0.46), r^2 = 0.53$$

وعموماً مؤشر ω^2 أكثر دقة من مؤشر μ^2

العوامل التي تؤثر على اختبار T المرتبطة

توجد العديد من العوامل التي تؤثر على نواتج اختبارات الفروض لاختبار T وهي:

1. **حجم فروق الدرجات:** فالفرق الكبيرة بين الدرجات تزيد من احتمالية رفض H_0 و كذلك يزيد من حجم التأثير.

2. **تشقت أو اختلاف الدرجات وحجم العينة:** يؤثران على حجم الخطأ المعياري في مقام معادلة T، فالخطأ المعياري يتناسب عكسياً مع حجم العينة، فالعينات الكبيرة تقود إلى خطأ معياري صغير، وهذا مرتبط بتباين العينة حيث التباينات الكبيرة تقود إلى خطأ معياري 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

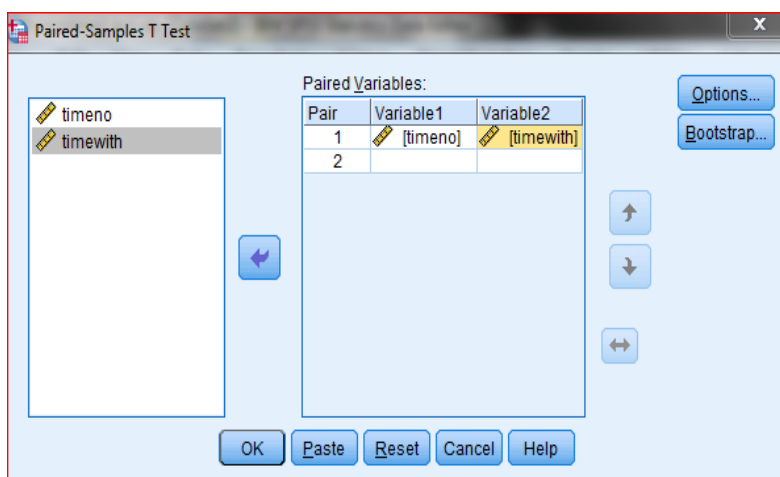
من الممكن إجراء اختبارات الفروض للقضيتين السابقتين التي تم تناولهما باستخدام اختبار T من خلال الاختبار اللابارامترى الإشارة Sign. ولكن اختبار T المرتبطة أكثر قوة إحصائياً من اختبار الإشارة حيث تقريباً تكون النتائج نفسها إذا توافرت الاعتدالية.

تنفيذ اختبار T المرتبطة في SPSS

أولاً: إدخال البيانات: 1. اضغط على Variable view ثم اكتب مسمى المتغيرات في عمود Name كالتالي: الوقت المستغرق للقراءة مع تعليمات المعلم Timewith ، وقت بدون تعليمات المعلم Timeno

2. اضغط على Data view يوجد المتغيرين في ملف البيانات في عمودين ابدأ في إدخال البيانات.

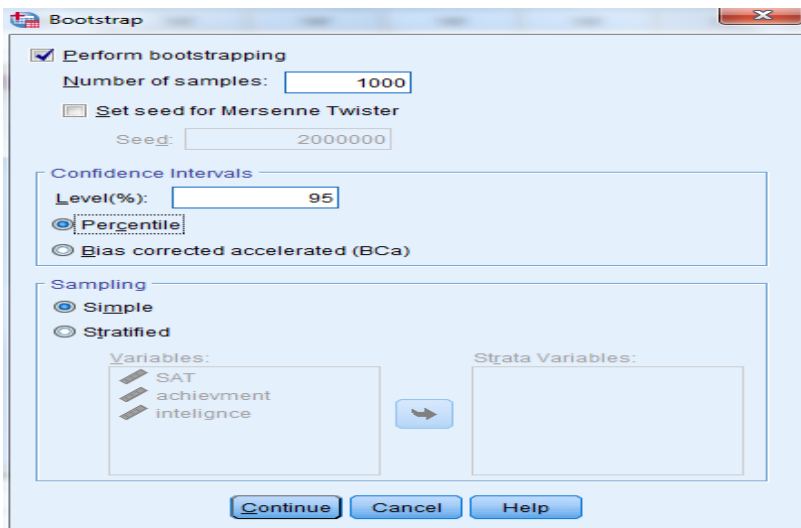
ثانياً : تنفيذ الامر: 1. اضغط Analyze → Compare Means → Paired samples T test تظهر الشاشة الآتية:



2. انقل Timewith إلى مربع 1 Variable ويمكن الضغط على المتغيرين معاً من خلال الضغط على مفتاح ctr ونقلهما إلى مربع Variable

3. انقل Timeno إلى مربع 2 Variable

4. اضغط على Bootstrap تظهر النافذة الآتية:



5. اضغط Perform bootstrapping :

- حدد عدد العينات التي ترغب في توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 1000 مثلاً وهي Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التي ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهي بسيطة.

6. اضغط Continue 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 إحصائي 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 المعياري: $SE = \frac{S}{N} = \frac{34.74}{\sqrt{8}} = 12.28$

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				
	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

t	df	Sig. (2-tailed)
2.806	7	.026

والفرق بين المتوسطين:

$$= 260.00 - 245.0 = 15$$

- الانحراف المعياري للفروق بين الدرجات: $SD = 15.11$
- والخطأ المعياري 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) $\text{Sig.} = 0.026$

وعليه يوجد % 2.6 فرصة في ان قيمة T تكون أكبر من القيمة المفترضة في الفرض الصفري، وعلى ذلك $0.05 < 0.026$ وبالتالي نرفض H_0 بمعنى توجد فروق ذات دلالة إحصائية عند 0.05 بين متوسط الزمن المستغرق في القراءة مع تعليمات المعلم و متوسط الزمن المستغرق في القراءة بدون تعليمات المعلم، والدلالة تكون في صالح المتوسط الأعلى حيث الوقت المستغرق بدون تعليمات المعلم.

- أخيراً 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 التي يمكن أن يقع فيها هذا المتوسط وهي: $CI_{95}(2.360, 27.6394)$ ، وعليه فإن قيمة متوسط الفرق تقع في هذا المدى وأن القيمة المفترضة في الفرض الصفري $= 0.0$ لا تقع في هذا المدى وعليه نرفض H_0 .

نتائج البوتستراب:

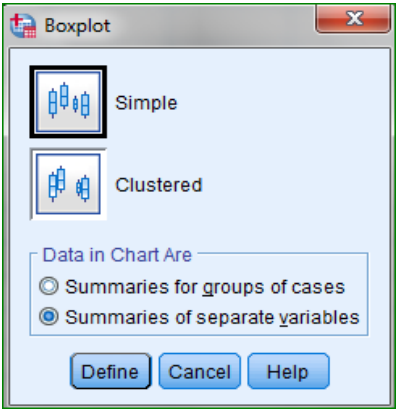
Bootstrap for Paired Samples Test

		Mean	Bootstrap ^a				
			Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval	
						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 هي المتولدة من 1000 عينة وهي لا تتضمن القيمة المتولدة في الفرض الصفري وعليه توجد فروق دالة إحصائية وفترات الثقة اضعف من فترات الثقة للعينة الوحيدة مما يدل على دقة التقديرات.



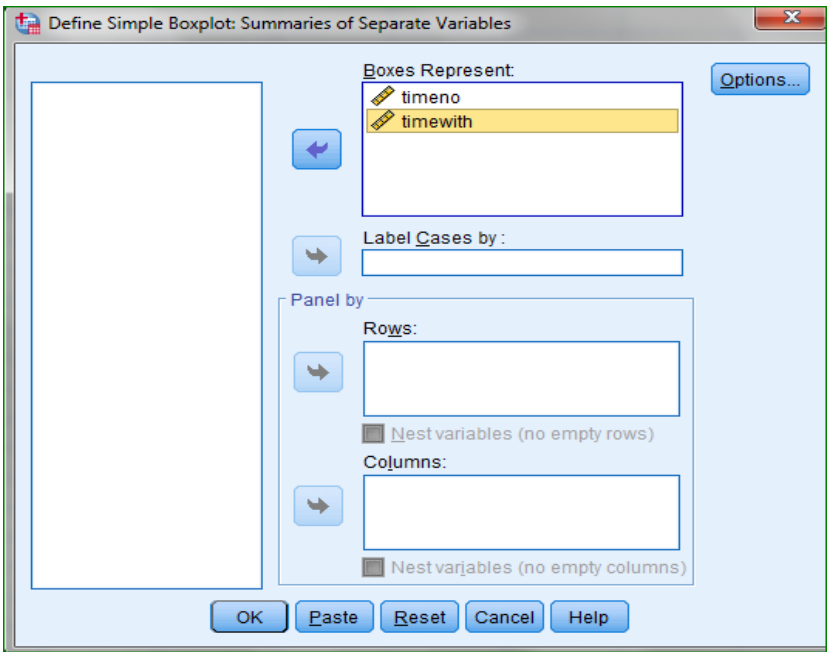
عرض النتائج بيانياً في ضوء **Boxplot** كالآتي:

Graphs→Legacy Dialogs →Boxplot.1

يعطى الشاشة الآتية:

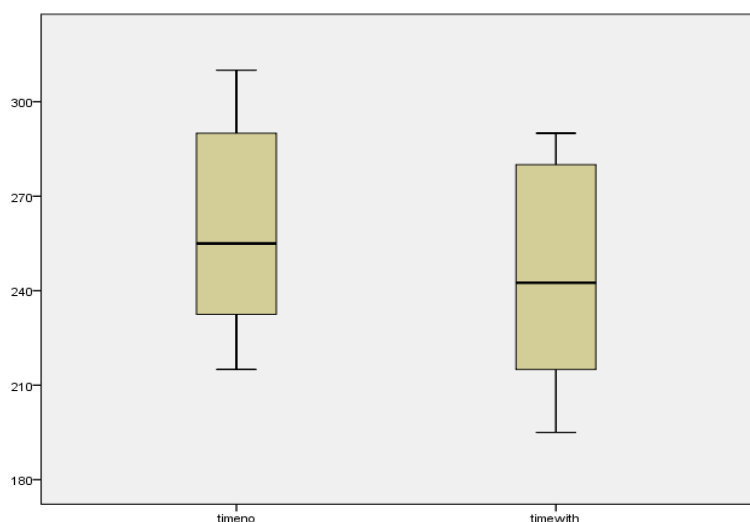
2. اضغط Simple، ثم اضغط Summaries of Separate Variables

3. اضغط Define:



5. . انقل المتغيرين إلى مربع Boxes Represent

6. . اضغط Ok يعطى الشكل الآتي:



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 الإرباعي 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 لتقييم ما إذا كانت توجد فروق بين متوسط مجتمع مجموعتين مستقلتين، وكل فرد في هذا الاختبار لا بد أن يكون له درجات على المتغيرين أحدهما متغير تصنيفي اسمي مستقل مثل الجنس أو المجموعة التجريبية والمجموعة الضابطة والآخر تابع متصل فتري 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 الثاني $\sigma_1^2 = \sigma_2^2$

اختبارات الفروض لقضية بحثية (Hinkle et al. 1994)

اهتم معلم بدراسة الفروق في تحصيل الرياضيات لطلاب الصف الثالث الابتدائي الذين يتعلموا تحت بيئتين مختلفتين، فتألفت العينة الأولى بيئة عالية التنظيم والتخطيط من حيث التدريس وفي حين تلقت العينة الثانية بيئة أقل تنظيمًا وتخطيطًا من حيث معينات التدريس، وبعد ستة أسابيع من التدريس الحصول على الدرجات التحصيلية للعينتين، واران المعلم اختبار ما إذا كانت توجد فروق في تحصيل الرياضيات بين البيئتين؟

الخطوات البحثية

1. سؤال البحث: هل يختلف تحصيل تلاميذ العينة الأكثر تنظيمًا في التدريس عن تحصيل

العينة الأخرى الأقل تنظيمًا التي تناولت تدريسيًا في بيئة أقل تنظيمًا؟ أو هل للبيئة التدريسية

أثر في تحصيل التلاميذ؟

2. متغيرات البحث: البيئة التدريسية: مستقل - اسمي - تصنيفي، تحصيل الرياضيات: تابع - فترى - متصل.

3. فرض البحث: توجد فروق في تحصيل الرياضيات بين تلاميذ البيئة الأكثر تنظيمًا وتلاميذ البيئة الأقل تنظيمًا.

4. تصميم البحث: باستخدام اختبار T في التصميمات البحثية التجريبية وشبه التجريبية والسببية المقارنة، وفي المثال السابق فإن الدراسة من النوع شبه التجريبي.

5. النموذج الإحصائي: بسيط حيث يوجد متغيرين وعليه إحصاء النموذج البسيط Bivariate statistics هو إحصاء بارامترى والاختبار المناسب: اختبار T المستقلة.

خطوات اختبارات 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 الأولى

• μ_2 متوسط مجتمع العينة الثانية

وهو اختبار ذو ذيلين.

2. الاختبار الإحصائي ومسلماته: الاختبار الإحصائي T المستقلة ومسلماته كالآتي:

- الأعتدالية Normality: يفترض أن توزيع المتغير التابع إعتدالي التوزيع وهذه المسلمة على جانب كبير من الأهمية خاصة للعينات الصغيرة لأن الخطأ المعياري يكون كبير بينما للعينات الكبيرة $N > 30$ يكون الخطأ المعياري 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 مستقلة

عن قياسات فرد آخر فى نفس المجموعة وإذا لم توجد استقلالية فإن قيمة P الاحتمالية لا يمكن الوثوق بها.

• **تساوى أو تجانس التباينات** Equal variances or homogeneity of variance :

تباين مجتمع المجموعة (العينة) الأولى مساوياً لتباين مجتمع العينة الثانية $\sigma_1^2 = \sigma_2^2$ ، ويرى (Privitera (2015) ان هذه المسلمة مرضية إذا كانت نسبة تباين المجتمع الأكبر إلى تباين المجتمع الأصغر أقل من 2.

والملاحظ أن الاستقلالية والعشوائية هى نواحى متعلقة بمنهجية الدراسة ويجب على الباحثين التحقق من مسلمات الاعتدالية وتجانس التباينات بعد جمع البيانات قبل إجراء اختبار T، وأحياناً يكتب فى تقارير البحوث تم التحقق من الاعتدالية للبيانات باستخدام اختبار كولموجروف-سيمنوف لدرجات المتغير التابع حيث إذا كانت قيمة الاختبار غير الدالة إحصائياً فإن التوزيع اعتدالى. وللتحقق من تجانس التباينات يتم إجراء اختبار Leven's F و إذا كانت $(P) < 0.05(\alpha)$ ، وعلى ذلك نقبل الفروض الصفرى وبالتالي يوجد تساوى التباينات.

• **المتغير التابع:** يكون من مستوى القياس الفترى أو النسبى.

تكون صيغتها كالتالى:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ ، $\bar{x}_2$ متوسط العينة الأولى والثانية على التوالى.
- s_1^2 ، s_2^2 تباين المجموعة الأولى والثانية على التوالى.
- n_1 ، n_2 حجم العينة الأولى والثانية على التوالى.

Pooled variance التباين المشترك

المعادلة السابقة تعرض الخطأ المعيارى $(S\bar{x}_1 - \bar{x}_2)$ لإحصاء T للعينات المستقلة وهذه الصيغة تكون صحيحة فقط فى المواقف التى تكون فيها أحجام العينتين متساوية $n_1 = n_2$ ولكن فى المواقف التى تكون فيها أحجام العينات مختلفة فإن الصيغة

السابقة تكون متحيزة ولذلك تكون غير مناسبة والتحيز يأتي من حقيقة أن المعادلة السابقة تعالج العينتين كما لو كان تباينهما متساوي $s_1^2 = s_2^2$ ، 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 σ^2 أفضل من العينات الصغيرة و أحد طرق تصحيح هذا التحيز في الخطأ المعياري هو دمج تباين العينتين في قيمة واحدة ويسمى التباين المدمج أو المشترك ويتم الحصول عليه بأخذ متوسطات أو دمج تباين العينتين باستخدام درجات الحرية كالتالي:

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

والبسط يعكس مجموع التباينات وكلاً منهما تم وزنه عن طريق درجات الحرية وعليه فإن اختبار T كالتالي:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

إذا:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

وعموماً عندما يكون أحجام العينات متساوية فإنه لا توجد فروق بين استخدام التباين كلاً على حدة واستخدام التباين المدمج في الصيغة السابقة حيث يكون القرار واحد ولكن إذا كان أحجام العينات غير متساوية فإنه التباين المشترك يحدث فروق قليلة. ولاختبار T درجات حرية:

$$df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

$$= 9 + 8 - 2 = 15$$

3. مستوى الدلالة الإحصائية وقاعدة القرار: افترض الباحث $\alpha = 0.05$ ولاختبار ذوذييلين و

$df = 15$ وبالبحت في جدول T فإن 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	
$n_1=9$	$n_2=8$
$\sum x_1=445$	$\sum x_2=551$
$\bar{x}_1=49.44$	$\bar{x}_2=68.88$
$\sum x_1^2=22865$	$\sum x_2^2=39009$
$S_1^2=10778$	$S_2^2=151.27$
$S_1=10.38$	$S_2=12.30$

إذا:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{49.44 - 68.88}{\sqrt{\frac{(9-1)107.78 + (8-1)151.27}{8+9-2} \left(\frac{1}{8} + \frac{1}{9}\right)}}$$

$$T = \frac{-19.44}{5.50} = -3.35$$

5. القرار والتفسير: حيث: $|-3.53| < T_{الدرجة 2.13}$ ، وعليه يُرفض H_0 وبالتالي

توجد فروق دالة إحصائية في تحصيل الرياضيات بين تلاميذ العينة الذين تلقوا تدريساً في بيئة أكثر تنظيماً وتلاميذ العينة ذو التدريس في البيئة الأقل تنظيماً.

القرار فى ضوء فترتى الثقة: يتم بناء فترات الثقة لـ CI_{95} :

$$\begin{aligned} CI_{95} &= (\bar{x}_1 - \bar{x}_2) \pm T_{c.v}(S_{\bar{x}_1 - \bar{x}_2}) \\ &= -19.44 \pm 2.13 \times 5.50 \\ &= (-31.81, 7.7) \end{aligned}$$

وعليه فإن قيمة صفر الموضوعية فى الفرض الإحصائى الصفري تقع خارج مدى فترة الثقة وبالتالي نرفض H_0

6. حجم التأثير: يتم حساب حجم التأثير لاختبار T المستقلة من المؤشرات الآتية:

أ- مؤشر كوهين d وتكون صيغته كالتالى:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_p^2}$$

حيث إن S_p^2 التباين المدمج:

$$\begin{aligned} S_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\ &= \frac{(9 - 1)107.78 + (8 - 1)151.27}{8 + 9 - 2} \\ &= \frac{862.24 + 1058.89}{15} \\ &= \frac{1921.13}{15} = 128.07 \\ S_p &= \sqrt{128.07} = 11.31 \\ d &= \frac{19.44}{11.31} = 1.75 \end{aligned}$$

وهو حجم تأثير كبير فى ضوء معايير Cohen.

ب- مؤشر مربع إيتا r^2 أو η^2 أو التباين المفسر:

$$\eta^2 = \frac{T^2}{T^2 + df} = \frac{T^2}{T^2 + n1 + n2 - 2}$$

$$= \frac{(3.53)^2}{(3.53)^2 + 8 + 9 - 2} = \frac{12.4609}{27.4609} = 0.4537$$

إذاً البيئة التدريسية فسرت 45.37 % من تباين التحصيل وهذا حجم تأثير كبير.

ج- مؤشر اوميغا² 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

الأتى:

1. افتح البرنامج تظهر الشاشة الافتتاحية.

2. تحت Type of power analysis اختار:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

لاحظ أننا اعتمدنا على التحليل البعدى Post hoc

3. تحت Test family اختار T- tests

4. تحت Statistical test اختار:

Test family	Statistical test
t tests	Means: Difference between two independent means (two groups)

5. ادخل المعالم Input parameters:

Input Parameters

Determine =>

Tail(s) Two

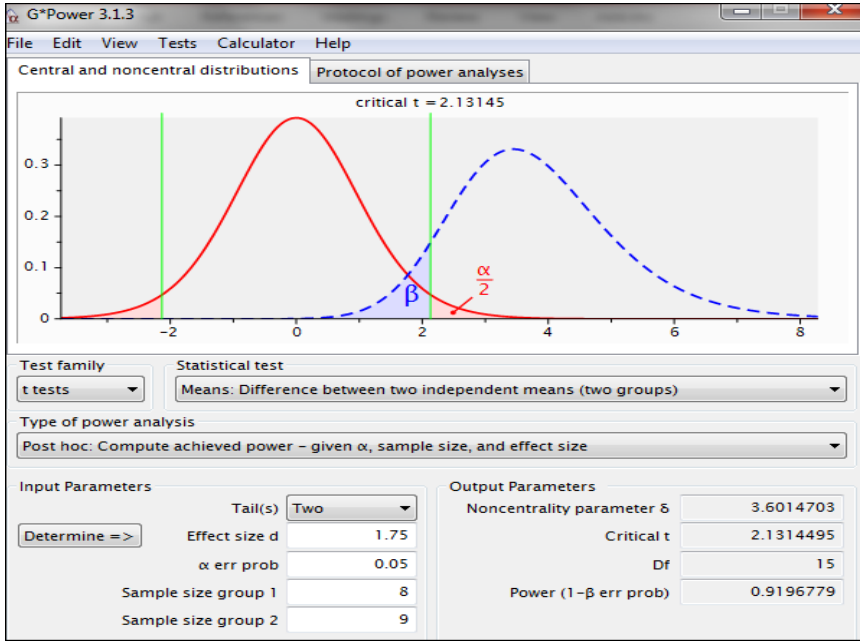
Effect size d 1.75

α err prob 0.05

Sample size group 1 8

Sample size group 2 9

6. اضغط Calculated تظهر شاشة الناتج كالاتى:



ويتضح أن القوة الإحصائية = 0.919 وهذا مستوى قوة مرغوبة وقد يرجع هذا إلى جودة التصميم البحثى وجودة البيانات.

عرض نتائج اختبار T المستقلة وفقاً لـ APA

لتقييم الفروق بين التلاميذ الذين يتعرضون لطريقة تدريس ما و أقرانهم الذين يتعرضون لطريقة تدريس أخرى 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 $(-31.18, 7.7)$ CI_{95} 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 العيى الاعتنالى 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 $(\sigma_{\bar{x}_1 - \bar{x}_2})$ 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 الصفري 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 مسلمة 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 تقريباً 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 $\alpha=0.02$ فإن القيمة الحرجة

$$F_{c.v} = 2.20 \text{ أو الجدولية:}$$

4. الحسابات:

$$F = \frac{105.96}{36.42} = 2.91 \frac{s_1^2}{s_2^2}$$

حتى لو اعتبرنا التباين الأصغر فى البسط و التباين الأكبر فى المقام فإن:

$$F = \frac{36.42}{105.96} = 0.34$$

وبالبحث فى جدول F حيث درجات حرية البسط (numerator) 40 و درجة حرية المقام

denominator 30 فإن F الحرجة = 2.30.

5. القرار : F المحسوبة (2.91) < F الحرجة (2.20)، وعليه يتم رفض H_0 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 $(\sigma_{\bar{x}_1 - \bar{x}_2})$ للفروق بين المتوسطات عند $\sigma_1^2 \neq \sigma_2^2$

عندما لا تتحقق مسلمة تجانس التباينات فإنه الإجراء البديل لاختبار الفرض

الصفري $H_0: \mu_1 = \mu_2$ 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 :

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{\frac{s_1^2}{n_1}}{n_1-1} + \frac{\frac{s_2^2}{n_2}}{n_2-1}}$$

مثال : أراد باحث التحقق من الفرض الإحصائي الآتي:

$$H_A: \mu_1 \neq \mu_2, H_0: \mu_1 = \mu_2$$

ولا يوجد تجانس في التباينات وكانت نتائج العينتين كالآتي:

الأولي	الثانية
$\bar{x}$	27.20
n	16
S^2	145.90
S	12.07
	7.09
	62.60
	31
	19.60

التحقق من مسلمة التجانس من خلال اختبار F كالآتي:

$$F = \frac{s_1^2}{s_2^2} = \frac{145.90}{62.60} = 2.33$$

وبالكشف في جدول F ملحق () بدرجتي حرية:

$$df_1 = n_1 - 1 = 16 - 1 = 15 \text{ (البسط)}$$

$$31 - 1 = 30 df_2 = n_2 - 1 = \text{(المقام)}$$

و عند $\alpha = 0.10$ يتضح أن: $F_{c.v} = 2.01$ وعليه نرفض H_0 وبالتالي لا يوجد تجانس بين σ_1^2

، σ_2^2 وعليه لحساب اختبار T يجب إجراء تعديل للخطأ المعياري و لدرجات الحرية كالآتي:

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{27.2 - 19.6}{\sqrt{\frac{145.90}{16} + \frac{62.60}{31}}}$$

$$= \frac{7.6}{3.34} = 2.28$$

وإجراء تعديل على درجات الحرية كالآتي :

$$\begin{aligned} df &= \frac{\left(\frac{145.90}{16} + \frac{62.60}{31}\right)^2}{\left(\frac{145.90}{16}\right)^2 / 15 + \left(\frac{62.60}{31}\right)^2 / 30} \\ &= \frac{(9.12 + 2.02)^2}{(9.12)^2 / 15 + (2.02)^2 / 30} \\ &= \frac{124.10}{5.54 + 0.14} = 21.85 \end{aligned}$$

ولا تقدر df كما فى الصيغة الآتية:

$$df = n_1 + n_2 - 2 = 6 + 31 - 2 = 45$$

وبالتالى لاختبار ذى ذيلين عند $\alpha = 0.10$ و درجات حرية تقريباً 22 فإن $T_{c.v} = 1.717$ وعليه

T المحسوبة (2.28) < الحرجة (1.717) وبالتالي نرفض H_0

المناعة أو الضلالة ضد عدم تحقق مسلمات T

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 تأثراً بعدم تحقق مسلماته بدرجة متوسطة مثل الاعتدالية أو تجانس التباينات، وكما نعلم أن عدم تجانس التباين يتم التغلب عليه بإجراء و تعديل فى صيغة $\sigma_{\bar{x}_1 - \bar{x}_2}$ و df، ويرى (2013) Howell أن الاستنتاج العام من الدراسات أنه إذا تساوت أحجام العينات ولم تتحقق مسلمة التجانس للتباينات فإنه يوجد تأثير صغير جداً على قيمة T و تكون فى حدود ± 0.02 عند $\alpha = 0.05$ 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 (2013) Howell عندما تكون أحجام العينات غير متساوية بدرجة كبيرة فيجب تطبيق مدخل Welch-Satterhwaite وفيها تستخدم صيغة درجات الحرية المعدلة، ويطرح تساؤل هو هل من الضروري التحقق من مسلمة التجانس للتباينات؟ وفى هذا الشأن يقترح (2007) Cai & Hayes 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 كلاً على حدة، و يروا من الأفضل استخدام التباينات كلاً على حدة حتى لو وجد تجانس لأن نسب الخطأ تبقى قريبة من $\alpha=0.05$ وعليه فإن استخدام T فى 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 فإذا كان n_1 وحجم كل عينة أكبر من 30 فإن اختبار T يستخدم 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 (2012) Huck أنه إذا لم تتحقق مسلمة الاعتدالية أو التجانس يوجد ثلاث بدائل متاحة هي:

- استخدام صيغة معينة لتعويض عدم تحقق الاعتدالية أو عدم التجانس كما سبق ذكره من إجراء تعديل للخطأ المعياري ودرجات الحرية.
- تعديل أو تحويل الدرجات الخام عن طريقة تحويل 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 T فى حالة ما إذا كشف اختبار F للتحقق من التجانس (عدم دلالة إحصائية) وإذا وجدت دلالة إحصائية فإن Wellch's T-test يستخدم بدلا من Student T-test.

أو تم تحويل البيانات باستخدام الجذر 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:

$$T = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

T لعينة مرتبطة:

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}}$$

T 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 لـ T 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 $\bar{x}_1, \bar{x}_2 - \bar{x}_1$ 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 T وعليه توجد احتمالية عالية لرفض

H_0 وعلى ذلك يكون الاختبار قوى إحصائياً و على ذلك فإن زيادة حجم العينة يزيد القوة الإحصائية لاختبار T .

والمقام يتغير كوظيفة لتباين العينة ففي T العينة واحدة فإن الانحراف المعياري S مقياس للتشتت و S_D في تجارب المجموعات المرتبطة ، S في تجارب المجموعات المستقلة، فكلما زاد التباين فإن المقام يزيد وبذوره تنقص قيمة T ولذلك فالعينات عالية التباين تقلل من القوة الإحصائية لـ T ولذلك فمن المرغوب تقليل تباين العينات كلما أمكن وجعلها متجانسة بقدر الامكان، وأحد الطرق لتقليل التباين هو إجراءات الضبط في التجربة وعليه ففي تجارب القياسات المرتبطة من المتوقع ان يكون التباين أقل من تباين التجارب المستقلة ولذلك فإن اختبار T المرتبطة أكثر قوة إحصائياً من T المستقلة.

على الرغم من أن الاختبار الإحصائي لمقارنة متوسطين T إلا إنه يمكن إجرائه باستخدام اختبار Z أو F ولكن اختبار Z يعطى نتائج متحيزة قليلاً حيث يمكن أن يؤدي إلى زيادة الخطأ من النوع الأول أكبر من مستوى الدلالة الإحصائية الذي يتبناه الباحث وهذا التحيز يزيد مع صغر حجم العينة، في المقابل فإن اختبار F غير متحيز ويكون الاستنتاج الخاص بـ H_0 في حالة استخدام F مماثل لاستنتاج اختبار T وبالتالي لا توجد إشكالية في استخدام T و F للمقارنة بين مجموعتين، وفي ضوء هذه الحقيقة فإن بعض الباحثون يفضلون استخدام اختبار F عند المقارنة بين متوسطى عينتين حتى أنه في 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 $p < 0.05$ ، $F_{(1,596)} = 8.51$ ، حيث إن $F_{(1,596)}$ هما درجتى الحرية ويكون: $df_1 = K-1$ حيث K عدد المجموعات أو العينات: $df_1 = 2 - 1 = 1$ ، درجة الحرية الثانية هي $df_c = N-K$ حيث N حجم العينة الكلى (ذكور و إناث): $df_2 = 598-2=596$

وفي بعض الأحيان يقوم الباحثون بعرض جدول تحليل التباين لـ F .

المقارنة بين متغيرات تابعة متعددة

أحياناً يتم جمع بيانات لعينة واحدة أو عینتین على أكثر من متغیر تابع ثم یقوم الباحث باختبار فرض صفری لكل متغیر تابع على حدة على الرغم أنه من الأفضل استخدام T^2 هو تلنج أو تحلیل التباين المتدرج MANOVA. ولكن الباحث یستخدم اختبار T أو F لكل متغیر تابع على حدة ویکتب فی تقرير البحث الآتی: وبإجراء اختبار T للفروق بین منخفضی ومرتفع التحصیل فی تقدير الذات و سمات الشخصية تبين بالنسبة لتقدير الذات $p < 0.001$, $T=5.28$ وبالنسبة للانسيابية $T_{(89)}=7.8$ $p < 0.001$. أو یكتب فی تقرير البحث: بإجراء اختبار F للفروق بین القیاس البعدي للمجموعة التجريبية والضابطة فی إبعاد سمات الشخصية هی للبعد الأول $F_{(1,642)}=32.474, p < 0.001$ و بالنسبة للبعد الثاني $F_{(1,642)}=(4.453), P < 0.0001$ ، وفي هذه الحالة یجب استخدام تصحیح Bonferroni Adjustment لأنه عندما یختبر الباحث فروض صفرية عديدة وكل فرض ینظر متغیر تابع فمن المتوقع تضخم الخطأ من النوع الأول حيث یحدث له زیادة عن القيمة التي افترضها الباحث فی بداية اختبار الفروض بطريقة غیر مباشرة، وأحد الطرق للتعامل مع هذه المشكلة والأكثر استخداماً هو تطبیق أسلوب تصحیح بونفیرونی وسوف یشار إليه أثناء تحلیل التباين.

تقدير حجم التأثير وتحليلات القوة لاختبار T

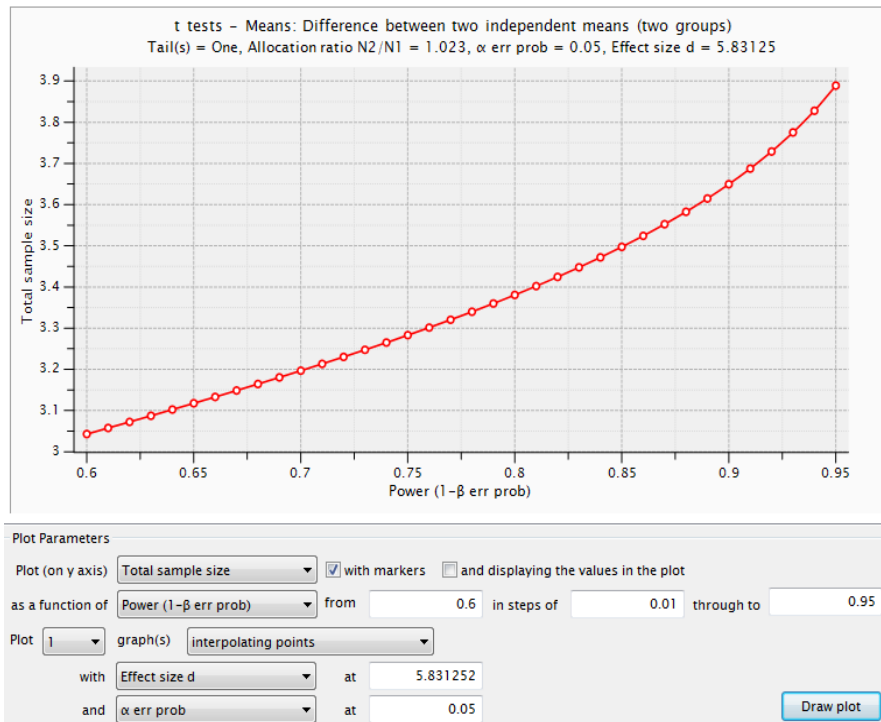
كما سبق وأوضحنا عند التعامل مع متوسط أو متوسطین باستخدام اختبارات الفروض فإن الباحث یجب أن یميز بین الدلالة الإحصائية والدلالة العملية ومن أهم مؤشرات حجم التأثير شیوعاً المصاحبة لاختبار T هی مؤشر كوهين (d) أو مؤشر μ^2 للمقارنة بین متوسطین أو مؤشر Partial eta squared (μ^2_p) المصاحب لاختبار F للمقارنة بین متوسطین ومؤشر (μ^2_p) هو مماثل لمؤشر r^2 لتقدير التباين المفسر. و فیما یلی أهم مؤشرات حجم التأثير المصاحب لاختبارات الفروض أو متوسطین باستخدام T ، F كما حددها Huck (2012) :

المؤشر	صغير	متوسط	كبير
كوهين d	0.20	0.50	0.80
إيتا μ^2	0.10	0.24	0.37
r_{11}^2 انتا	0.4	0.06	0.14
أوميغا 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 $\mu^2 p$ هي اختصارًا Squared Partial Eta

والمؤشرات الأكثر استخدامًا في تقارير البحوث هو مؤشر $d, \mu^2 p$.

والشكل التالي يوضح الرسم البياني لمنحنى العلاقة بين قوة الاختبار (محور السينات)، وحجم العينة الكلى:



تعقيب: في نهاية هذا الفصل لا بد من عرض الآتى:

- النتائج غير الدالة إحصائيًا لا تعنى أن H_0 حقيقى.
- الفرض الصفري 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

تكون للمجموعات الضابطة والتجريبية نفس الدرجات في بداية الدراسة ولكن يوجد العديد من الأسباب لأن نعتبر من الخطورة أن الاعتقاد بأن H_0 حقيقى إذا لم يرفض. وتوجد عدم منطقية لقبول H_0 لأمر يتعلق بجودة البيانات وذلك لحدوث نوع من التقلص Attenuation حيث لها ثبات منخفض، فالثبات المنخفض يقود إلى الفشل في رفض H_0 ولذلك فإن قبول H_0 نتيجة عدم الثبات للبيانات. والاعتبار الآخر لعدم رفض H_0 هو القوة الإحصائية وهذا متعلق بحجم العينة، فالفشل في الحصول على دلالة إحصائية قد يكون بسبب حجم العينة فربما تكون صغيرة جداً فربما يختلف متوسط مجتمع المجموعة الضابطة $\mu_{control}$ عن متوسط مجتمع المجموعة التجريبية μ_{exper} إختلافاً كبيراً ولكن الدراسة ينقصها الحساسية (القوة) للكشف عن هذه الفروق.

و أخيراً 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 في SPSS: مثال Hinkle et al.(1994)

أولاً إدخال البيانات:1. اضغط Variable View

2. اكتب مسمى المتغيرات عمود Name وهما:

المتغير المستقل: البيئة environment: بيئة أقل تنظيماً وتخطيطاً (1)، بيئة عالية التنظيم والتخطيط (2). والمتغير التابع: التحصيل achievement

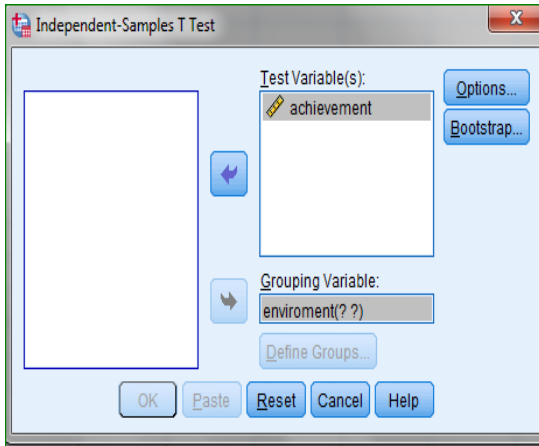
3. اضغط Data View تظهر شاشة البيانات

ثانياً: تنفيذ الأمر:

1. اضغط Analyze→Compare Means

→Independent-Samples

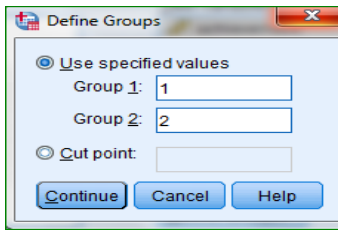
تظهر الشاشة الآتية:



2. اضغط achievement وانقله إلى

مربع Test Variables

3. اضغط environment وانقله إلى مربع Grouping Variable

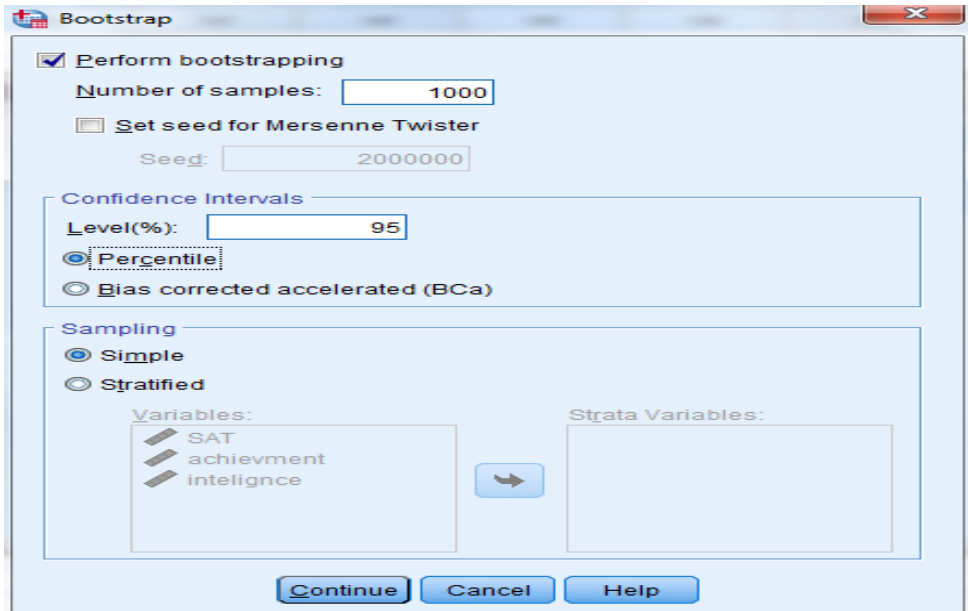


4. اضغط DefineGroups تظهر الشاشة الآتية:

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 التي ترغب في توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 1000 مثلاً وهي Default البرنامج، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التي ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهي بسيطة.

7. اضغط Continue ثم اضغط Ok

ثالثاً: تفسير المخرج: الجدول الأول: الإحصائيات الوصفية:

Group Statistics					
		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 المعياري Std. Deviation والخطأ المعياري للمتوسط $SE = \frac{10.38}{\sqrt{9}} = 3.46$: Std. Error Mean والواضح أن عدد أفراد المجموعة الأولى 9 والثانية 8 وأن متوسط تحصيل العينة الأولى أكبر من متوسط تحصيل العينة الثانية.

الجدول الثاني يتضمن إحصائيات اختبار T الجزء 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 الثاني:

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:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_A: \sigma_1^2 \neq \sigma_2^2$$

ويتضح أن قيمة $F=0.275$ و $\text{Sig } (p)=0.608 > 0.05$ وهى غير دالة إحصائياً وعليه نتحقق مسلمة استخدام اختبار T المستقلة وهى تجانس التباينات . وعليه تعتمد على نتائج اختبار T فى حالة Equal Variances Assumed وإذا رفض الباحث الفرض الصفري فعليه نعتد 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 $\text{Sig } (2tailed)=$ ، وبما أن: $0.003 < 0.05$ ، وعليه يُرفض H_0 وبالتالي يوجد فرق دال إحصائياً 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 ذو ذيل واحد أسهل من رفضه لاختبار ذو 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 بدلاً من هذا فإن التقديرات تكون عبارة عن مدى وهى أكثر واقعية من النقطة الوحيدة وهو للفترة التى لها حد أدنى وحد أعلى وتسمى حدود الثقة وهى تسمح بمعرفة كيف نشق فى ان متوسط المجتمع يكون داخل فترة معينة. وإذا لم تتضمن حدود الثقة القيمة المفترضة فى الفرض الصفري وهى فى هذه الحالة صفر فإن

الفرض الصفري يرفض، بما أن القيمة المتضمنة في الفرض الصفري تقع خارج منطقة رفض H_0 وعليه فإن حدود الثقة:

$$CI_{95} = (-31.151, -7.709)$$

وعليه لا تتضمن القيمة صفر بالتالي يرفض H_0 . ويمكن تغيير حدود الثقة إلى 99% من خلال اختيار Option في امر T.

نتائج البوتستراپ:

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

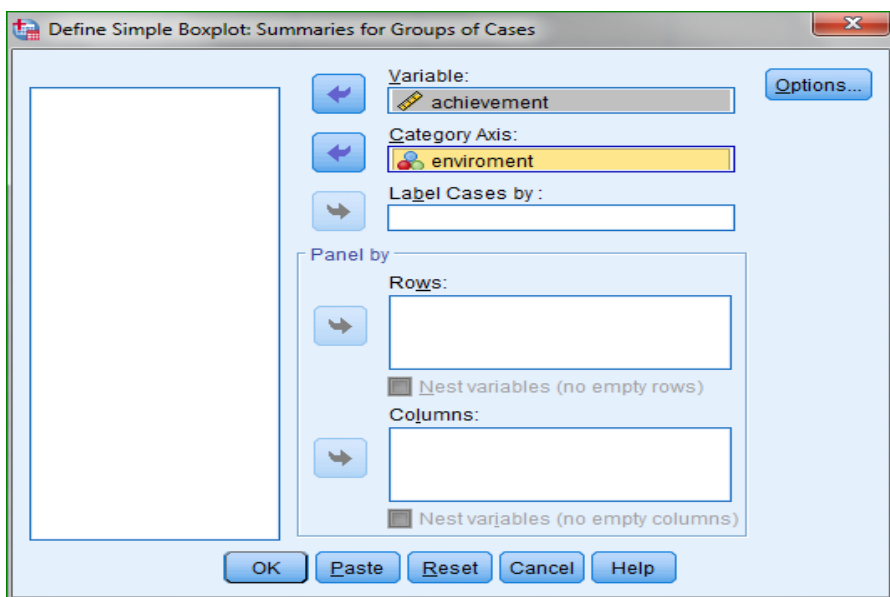
عليك المقارنة بين نتائج البوتستراپ ونتائج اختبارات الفروض الصفرية لبيانات العينة المجمعة من الأفراد، لاحظ لا اختلاف جوهري بين النتائج وذلك لصغر حجم العينة.
عرض النتائج بيانياً

يمكن عرض النتائج من خلال Histogram أو Boxplot كالاتي:

1. اضغط Graphs→ Legacy Dialogs→ Boxplot تظهر الشاشة الآتية:



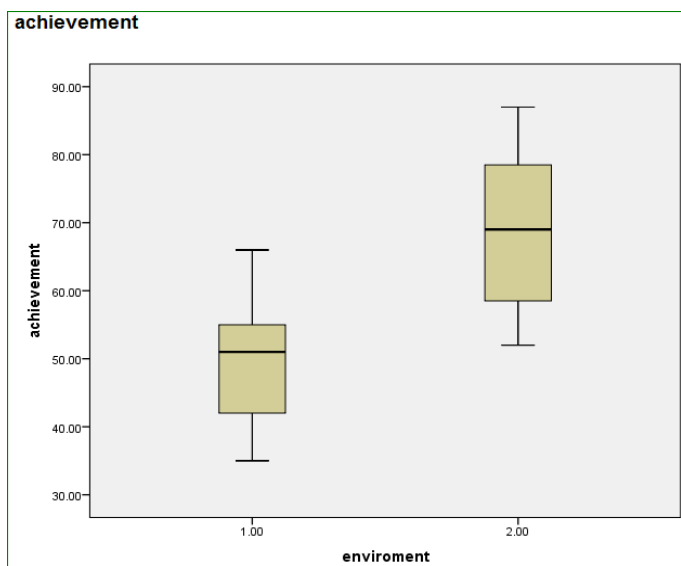
2. اضغط Simple ثم Define:



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 تحصيلاً.

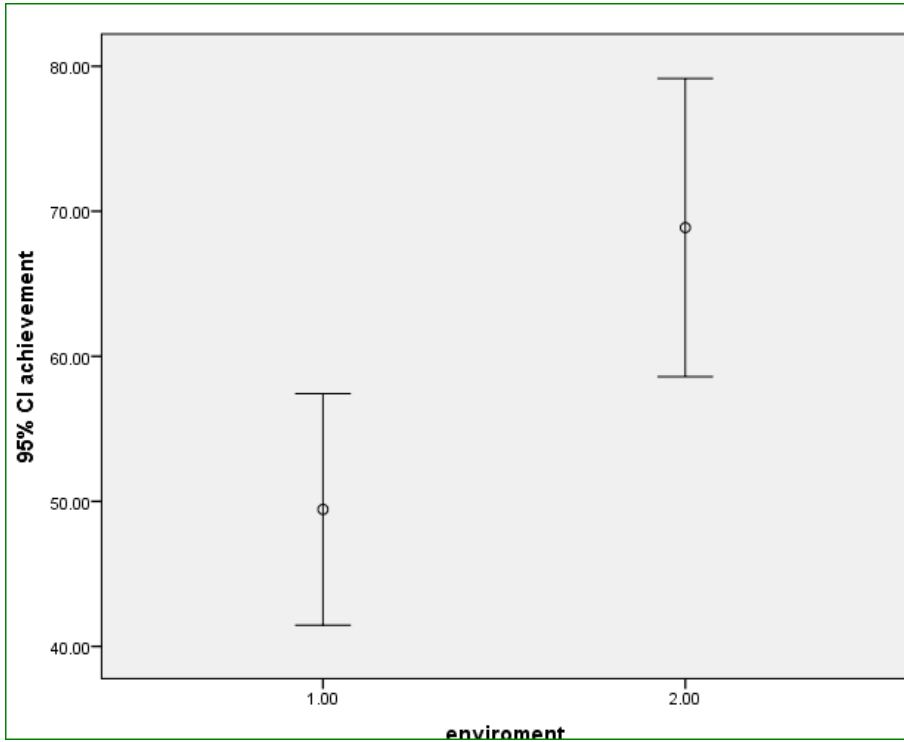
ويمكن عرض Error Bar كالآتي:

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

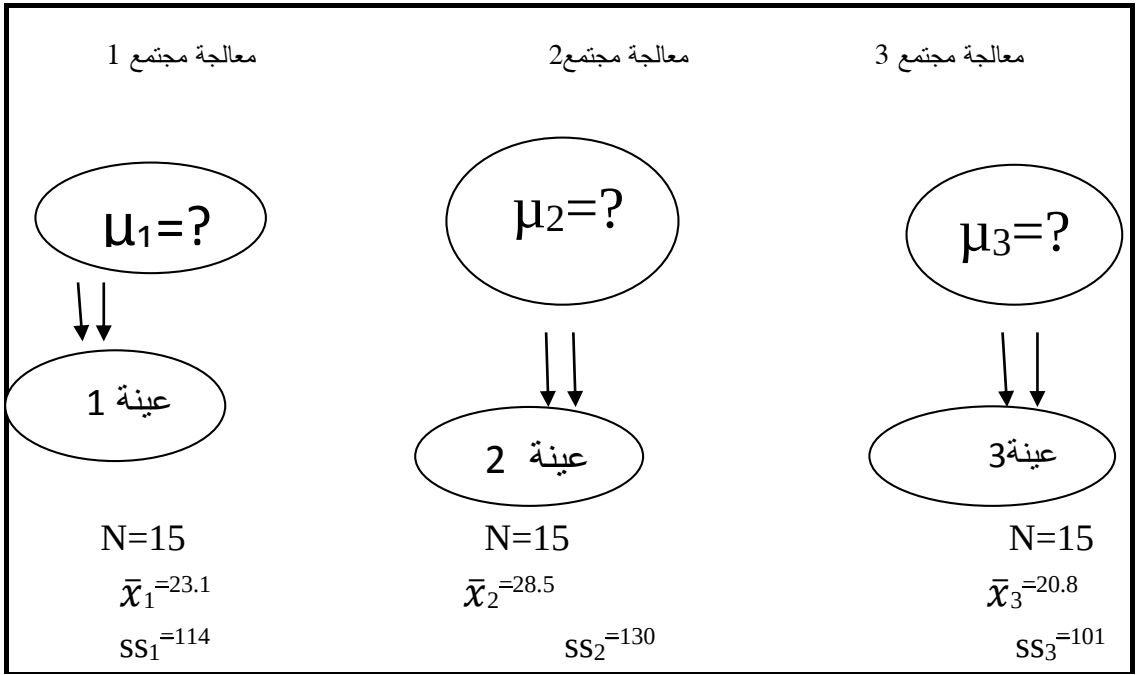
تحليل التباين احادى الاتجاه

One way analysis of variance (ANOVA)

فى الفصول السابقة تناولت اختبارات إحصائية استدلالية لاختبارات فروض حول متوسط مجتمع واحد أو متوسطين مثل اختبارات T و Z ، وفى هذا الفصل سوف نتسع المناقشة عندما يكون هدف الباحث إجراء اختبارات فروض تتعامل مع ثلاث متوسطات فأكثر فالأسلوب الإحصائى المستخدم فى هذه النوعية من الدراسات يطلق عليه تحليل التباين Analysis of variance أو اختصارًا ANOVA و أحيانًا يسمى 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 ولمعالجتي أو لمجتمعين فأكثر، ويتضح أن اختبار T يتشابه مع اختبار F فى ان كلاهما يؤدوا نفس الوظيفة وهى اختبار فروق بين متوسطات فى المجتمع ولكن اختبار T محدود فقط بمعالجتي 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

الأربع مجموعات. وفيما يلي شكل يوضح الموقف الامثل لاستخدام ANOVA احادى الاتجاه:



الشكل (1.20): الشكل الامثل لاستخدام ANOVA.

فى هذا الشكل نلاحظ وجود ثلاث عينات يمثلوا ثلاث مجتمعات والهدف من التحليل هو تحديد ما إذا كان يوجد فروق بين العينات الثلاثة لأن هذا يعطى دليل كافى 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.

لاحظ أن هذين التفسيرين يناظران الفرض الصفري (H_0) والفرض البديل (H_A) فى اختبارات الفروض.

ويستخدم هذا الاختبار مع التصميمات التجريبية بين المجموعات أو تصميم بين المجموعات المستقلة أحادية الاتجاه أو العامل One way between subject design حيث إن المتغير المستقل متعدد المستويات ويطلق عليه مستويات العامل و يشاع استخدام مفهوم عامل ليشير إلى المتغير المستقل أو متغير المعالجة فى

التصميمات التجريبية أو شبه التجريبية، بينما شروط المعالجات 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
$n_1 = 4$	$n_2 = 4$	$n_1 = 4$

على ذلك فإن الهدف من استخدام ANOVA هو عمل استنتاجات حول ما إذا كانت يوجد فروق بين متوسطات المجتمعات بغض النظر عن عدد العينات أو المجموعات المتضمنة ويصبح السؤال الرئيسى هو هل متوسطات المجتمعات المختلفة متساوية أو غير متساوية؟ وكما سبق الإشارة اثناء الحديث عن اختبار T يمكن استخدام تحليل التباين لاختبار دلالة الفروق بين متوسطين أيضاً كبديل لاختبار T ولذلك نعود فنقول أنه يمكن استخدام اختبار F للتحقق من الدلالة الإحصائية بين متوسطين فاكثر.

الفرق بين تحليل التباين البسيط وأنواع أخرى من ANOVA

فى هذا الفصل نعرض تحليل التباين البسيط ويشار إليه تحليل التباين احادى الاتجاه One way ANOVA. ولكن يوجد انواع أكثر تعقيداً لـ ANOVA كما سوف نتعرض موضع آخر مثل تحليل التباين ذو اتجاهين Two-way ANOVA وتحليل التباين للقياسات المتكررة Repeated measures ANOVA وتحليل التباين المتدرج Multivariate ANOVA على الرغم انهم جميعا يركزوا حول المتوسطات 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 كفى 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 Two-way ANOVA يعنى أن الاختبار يتعامل مع متغيرين مستقلين تصنيفين وتابع واحد متصل بينما Three-way ANOVA 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 وأحياناً يطلق عليه البعض تحليل التباين المتعدد Multiple ANOVA حيث 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، فالباحث يتضمن مجموعتين مختلفتين أحدهما تلقت أحد طرق العلاج بينما الأخرى تلقت الطريقة الثانية وقام بقياس كل مجموعات عبر فترات زمنية مختلفة والتصميم كالاتى:

- 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				
العلاج	بعد المعالجة مباشرة		بعد 6 شهور	
	الأولي	درجات	درجات	درجات
	الثانية	درجات	درجات	درجات

تصميم بحثي 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:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \dots = \sigma_n^2$$

واحياناً يطلق عليها Homoscedasticity وعدم توافر هذه المسلمة بمعنى عدم التجانس Heterogeneity of variance يؤدي 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، وإذا لم تتحقق هذه المسلمة فمن الأفضل استخدام إحصاء Welsh وإجراء اختبارات المقارنات البعدية Post-hoc tests يكون محل شك.

- **الاعتدالية:** يفترض ان تكون توزيع بيانات المتغير التابع اعتدالية فى مجتمعاتها، وهذه المسلمة على درجة كبيرة من الأهمية خاصة للعينات الصغيرة. والابتعاد بدرجة متوسطة عن الاعتدالية ليست مشكلة خطيرة، ودائمًا مع العينات المتوسطة والكبيرة تتوفر الاعتدالية. ويرى (Green & Salkind 2014) فى بعض المجتمعات غير الاعتدالية التوزيع فلحجم عينة 15 لكل مجموعة كافية لإعطاء تقديرات دقيقة لـ P . والقوة الإحصائية لاختبار F تنقص كلما كان توزيعات المجتمعات غير اعتدالية (ملتبسة - متفرطة)، ودائمًا يتم التحقق منها باستخدام اختبار كولموجوروف-سيميرنوف أو اختبار شابيرو-ويلك Shapiro-Wilk.

- **الاستقلالية:** يفترض ان كل فرد مستقل بذاته عن الآخرين واختيار أى فرد فى نفس المجموعة مستقل عن أى فرد آخر فى ذات المجموعة، كما أن اختيار أفراد أى مجموعة لا يؤثر على اختيار أفراد المجموعات الأخرى.

- **المعانة العشوائية Randomness:** 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 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 فأكثر للعامل وكل مستوى من العامل أو المتغير المستقل هو مجموعة وإذا كان متوسطات المجموعات متساوية فإنه من المحتمل ان يكون التباين بين المجموعات متساوي والفروق الكبيرة بين متوسطات المجموعات يؤدي إلى تباينات كبيرة بين متوسطات المجموعات، والفرض الصفري H_0 للمقارنة بين متوسطات عديدة:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

- 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 دائماً غير موجهه ويصاغ إذا حدث أن أحد متوسطات المجموعات أو أكثر تختلف تأثيرها على المتغير التابع عن متوسطات المجموعات الأخرى فيمكن أن يصاغ كالاتي:

$$H_A: \mu_1 \neq \mu_2 \neq \mu_3 \dots \neq \mu_k$$

$$H_A : \mu_1 \neq \mu_2 = \mu_3 = \dots = \mu_k$$

توجد فروق على الأقل لمتوسط واحد عن بقية المتوسطات للمجموعات، وعليه فإن للتجربة أو للمعالجة تأثير على المتغير التابع (تذكر أن الفرض البديل هو فرض البحث).

وكذلك يفترض اختبار F ان التباينات متساوية:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$$

- 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 الكلي $Total\ variability(SS_{total})$ للبيانات إلى مصدرين أساسين:

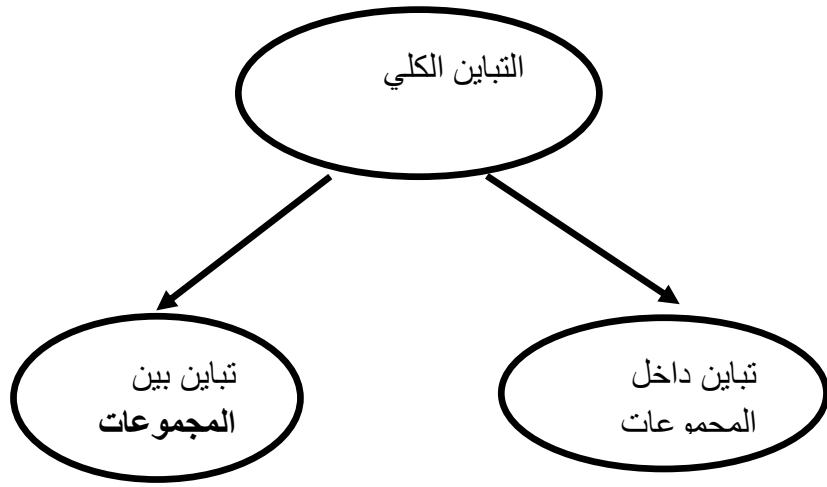
1. التباين أو الاختلاف داخل كل مجموعة ويسمى مجموع المربعات داخل المجموعات أو

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 الآتي:



الفروق نتيجة:

- تأثير المعالجة

الفروق نتيجة:

- عوامل غير منتظمة (الخطأ)

- عوامل عشوائية (فروق فردية)

تقدير التباين داخل المجموعات MS_{within}

أحيانًا يشار إليه بتباين الخطأ أو البواقي MS_e ، وفي تحليل التباين فإن التباين داخل المجموعات:

$$MS_W = \frac{SS_1 + SS_2 + SS_3 + \dots + SS_k}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) \dots + (n_k - 1)}$$

• SS تباين درجات المجموعة (مجموع المربعات).

• K عدد المجموعات.

ويمكن تبسيط المعادلة السابقة كالآتى:

$$MS_W = \frac{SS_1 + SS_2 + SS_3 + \dots + SS_k}{N - K}$$

حيث N حجم العينة الكلى فى كل المجموعات:

$$N = n_1 + n_2 + n_3 + n_4 + \dots + n_k$$

وبسط المعادلة السابقة يرمز له بالرمز SS_W كما أن $N-K$ هى درجات الحرية الأولى (داخل

المجموعات أو الخطأ أو البواقي) فى تحليل التباين وعليه:

$$MS_W = \frac{SS_W}{df_W}$$

حيث df_W درجات الحرية داخل المجموعات:

$$df_W = N - K = df_1 + df_2 + df_3 + \dots + df_k$$

$$df_1 = n_1 - 1, \quad df_2 = n_2 - 1$$

و df_W يطلق عليها df_{error} أو درجات الحرية للخطأ.

حيث مجموع المربعات داخل المجموعات:

$$SS_W = SS_1 + SS_2 + SS_3 + \dots + SS_k$$

وتقدر مجموع المربعات داخل المجموعات من الدرجات الخام كالآتى:

$$SS_W = \sum X^2 - \left(\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} + \dots + \frac{(\sum X_k)^2}{n_k} \right)$$

- $\sum x^2$ (T²) 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 T²

تقدير التباين بين المجموعات $MS_{Between}$

التباين الثانى فى تحليل التباين هو قائم على الاختلاف بين المجموعات وعليه فالمعادلة العامة

من الدرجات الخام كالآتى:

$$SS_B = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \dots + \frac{(\sum X_K)^2}{n_3} - \frac{(\sum X)^2}{N}$$

- $(\sum X_1)^2$ مربع مجموع درجات المجموعه الأول.
- $(\sum X_2)^2$ مربع مجموع درجات المجموعه الثانية.
- $(\sum X)^2$ مربع مجموع درجات كل المجموعات
- ويشار إلى $\sum X$ بـ T .

ويطلق على درجات الحرية بين المجموعات بـ درجات الحرية المعالجة:

$$df_{between} = df_B = K - 1$$

التباين الكلى SS_{total}

يقدر التباين الكلى من الصيغة الآتية:

$$SS_{total} = \sum X^2 - \frac{(\sum X)^2}{N}$$

- $\sum X^2$ مجموع مربعات كل درجات العينات المتضمنة فى التحليل.
- $(\sum X)^2$ 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:

$$F = \frac{MS_B}{MS_W}$$

$$MS_B = \frac{SS_B}{(K-1)} \quad \text{حيث:}$$

$$MS_w = \frac{SS_w}{(N - K)}$$

لذلك فإن القيمة الكبيرة لـ F 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 بدلاً 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 بدلاً من اختبار واحد لـ F فإن درجه تضخم $\alpha = 0.05$ تكون كالتى :

$$= 1 - (1 - 0.05)^{10} = 0.40$$

واحتمال ارتكاب الخطأ من النوع الأول للمقارنات بين المجموعات الخمسة ليس 0.05 كما هو مفترض انما 0.4 . وعليه فعلى الباحث التمييز بين Test wise Level و α 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 T لمقارنة كل الفروق بين المتوسطات:

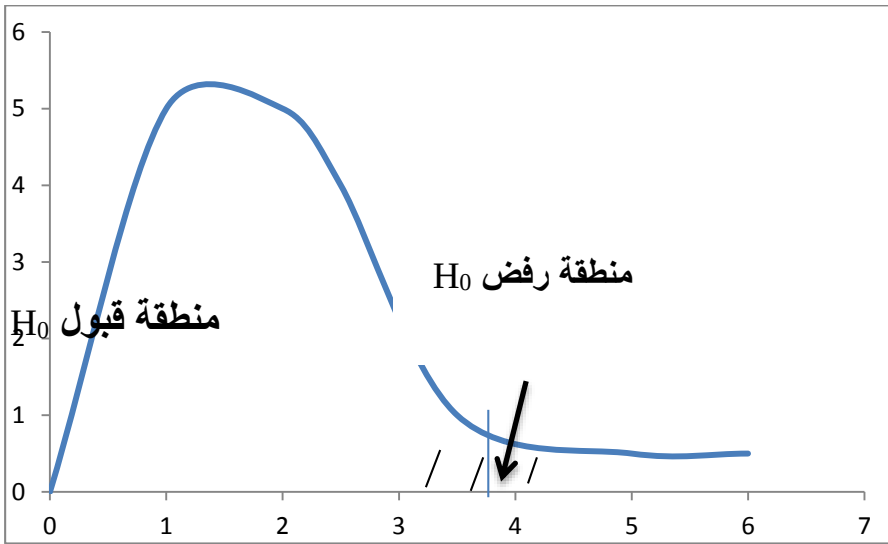
الاختبار الأول: مقارنة المعالجة الأولى مع الثانية، الاختبار الثانى: مقارنة المعالجة الأولى مع الثالثة، الاختبار الثالث: مقارنة المعالجة الثانية مع الثالثة. وعليه فإن ارتكاب الخطأ من النوع الأول يزداد:

$$= 1 - (1 - 0.05)^3$$

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 لـ F يقيسوا نفس التباين للفرض الصفري الحقيقى فإن قيمه F تكون قريبة من الواحد الصحيح وعلى ذلك فالتوزيع F فى المنطقة الموجبة ($0 \rightarrow 1,2,.....$) وناحية اليمين.

اختبارات الفروض لقضية بحثية (Pagano,2013)

قام باحث بقياس الضغوط النفسية لخمسة عشر طالب فى المرحلة الثانوية وقسمهم إلى ثلاث مجموعات، المجموعة الأولى تكونت عشوائيا 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 احدهما 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 والاختبار 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.

$H_A: \mu_1 \neq \mu_2 = \mu_3$

أو $H_A: \mu_1 \neq \mu_2 \neq \mu_3$

توجد فروق على الأقل بين متوسط أحد العينات الأولى والعينات الأخرى أو بين متوسطات العينات الثلاثة.

2. الاختبار الإحصائي ومسلماته: الاختبار هو F:

$$F = \frac{MS_B}{MS_{W(error)}}$$

في مجلدات يشار إلى داخل المجموعات بالخطأ وكأن تقول مجموع مربعات الخطأ أو درجة حرية الخطأ وهذا يكون في ضوء التعبير عن الدرجة المقاسة في ضوء النموذج البنائي Structural Model حيث يتعامل مع الدرجة ليست تامة الثبات بل تتضمن جزء من الخطأ.

ودرجات الحرية: $df_B = K - 1 = 3 - 1 = 2$

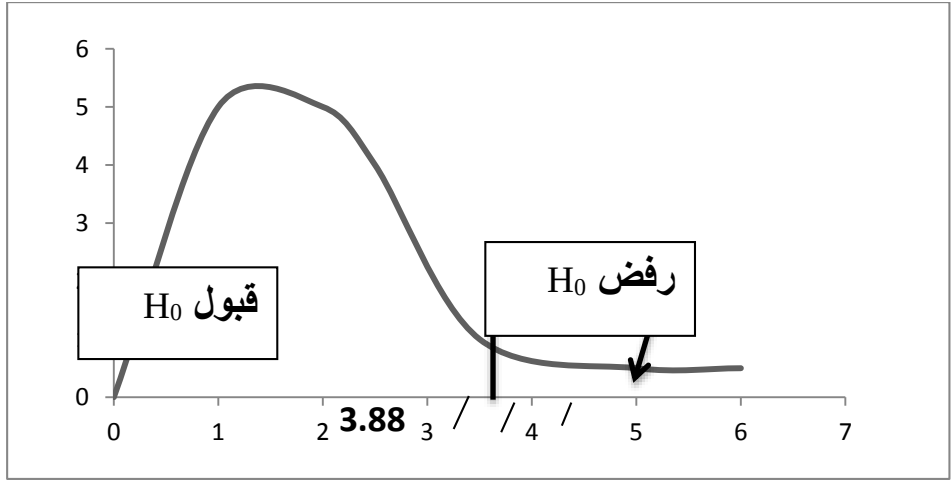
$$df_W = N - K = 15 - 3 = 12$$

$$df_{total} = N - 1 = 15 - 1 = 14$$

3. مستوى الدلالة الإحصائية وقاعده القرار: لمستوى دلالة إحصائية 0.01 ولدرجتي حريه (

البسط - الصف) $df_B = 2$ و(المقام - العمود) $df_W = 12$ وبالبحث في جداول F (ملحق)
(فان:

$F_{c.v}$ الحرجة = 3.88 وعليه فإن منطقتي الرفض لـ H_0



وعليه إذا كانت: F المحسوبة $< F$ الحرجة (3.88)، نرفض H_0

• 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
$\Sigma 20$	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 في المجموعات:

$$\Sigma x = T = 2+3+7+2+6+10+8+ +13+15= 125$$

حيث T مجموع درجات لأفراد في المجموعات الثلاثة .

• مجموع مربع كل الدرجات لكل المشاركين في المجموعات:

$$\Sigma X^2 = T^2 = 4+9+49+4+36+100+ +169+225 = 1299$$

حيث T^2 مجموع مربعات كل الدرجات للعينة الكلية .

$$\begin{aligned} \bar{X} = \bar{T}1 &= \frac{T1}{N} = \frac{\text{مجموع كل الدرجات}}{\text{حجم العينة الكلى}} \\ &= \frac{125}{15} = 8.333 \end{aligned}$$

- حساب مجموع المربعات بين المجموعات SS_B :

$$SS_B = \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} \right] - \frac{(T)^2}{N}$$

$$= \left[\frac{(20)^2}{5} + \frac{(40)^2}{5} + \frac{(65)^2}{5} \right] - \frac{(125)^2}{15}$$

$$= 203.333$$

- حساب مجموع مربعات داخل المجموعات (الخطأ) SS_W :

$$SS_W = (T)^2 - \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} \right]$$

$$= 1299 - \left[\frac{(20)^2}{5} + \frac{(40)^2}{5} + \frac{(65)^2}{5} \right] = 54$$

- حساب مجموع المربعات الكلى:

$$SS_{Total} = SS_W + SS_B$$

او يمكن تقديره من الصيغة الآتية:

$$SS_{Total} = T^2 - \frac{(T)^2}{N}$$

$$= 1299 - \frac{(125)^2}{15} = 257.333$$

$$SS_{Total} = 257.333 \quad \text{وعليه:}$$

$$257.333 = 54 + 203.333$$

$$SS_B = SS_T - SS_W \quad \text{وتُحسب } SS_B \text{ كالآتي:}$$

$$SS_W = SS_T - SS_B \quad \text{بينما}$$

- حساب التباين بين المجموعات:

$$MS_B = \frac{SS_B}{df_B} = \frac{203.333}{2} = 101.667$$

• حساب التباين داخل المجموعات:

$$MS_W = \frac{SS_W}{df_W} = \frac{54}{12} = 4.50$$

• 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 لـ F 3.88، وعليه

يرفض الباحث الفرض الصفري وبالتالي توجد فروق ذات دلالة إحصائية في متوسط درجات الضغوط النفسية بين المجموعات الثلاثة عند 0.05.

6. حجم التأثير: يتم حساب حجم التأثير في تحليل التباين من عدة مؤشرات كالآتي:

1. مؤشر إيتا² أو r² : ويقدر من الصيغة الآتية:

$$\eta^2 = \frac{SS_B}{SS_T}$$

ويشار إلى هذا المؤشر بنسبه الارتباط :

$$r^2 = \eta^2 = \frac{203.333}{257.333} = 0.790$$

وهذا يعنى ان المتغير المستقل فسر 79.0 % من تباين درجات الضغوط.

ويمكن تقدير η^2 من الصيغة الآتية:

$$\begin{aligned} \eta^2 &= \frac{1}{1 + \frac{df_W}{F \times df_B}} \\ &= \frac{1}{1 + \frac{12}{22.59 \times 2}} = 0.79 \end{aligned}$$

واشار (2013) Howell إلى أنه توجد صيغة لـ η^2 بحيث تكون مقياس لحجم التأثير وليس

كنسبة تباين مفسر وتسمى (RRE) Percent Reduced in Error

ب- مؤشر اوميغا تربيع ω^2 : البديل الأفضل لـ η^2 لتقدير حجم التأثير أو تقدير حجم التأثير التجريبي هو مؤشر ω^2 . ويوجد صيغتين لـ ω^2 اعتماداً 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 التجريبي كالاتى:

$$\omega^2 = \frac{SS_B - (K - 1)MS_W}{SS_{tot} + MS_W}$$

لاحظ أن مؤشر ω^2 يتميز عن مؤشر η^2 فى الآتى:

- يصحح مؤشر η^2 بتضمين 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 كالاتى:

$$\omega^2 = \frac{203.333 - (3 - 1)4.500}{257.333 + 4.500} = 0.742$$

وعلى ذلك فإن مواقف الضغوط فسرت 74.2 % من تباين درجات الضغوط وهذا حجم تأثير كبير .

وحدود قيم η^2 و ω^2 هى:

التفسير	النسبة أو القيمة
صغير	0.01 – 0.05
متوسط	0.06 – 0.13
كبير	≥ 0.14

لاحظ أن قيمة η^2 (0.792) أكبر قليلاً من قيمه ω^2 (0.742) لأن مؤشر ω^2 يعطى تقدير أكثر دقة ولذلك يوصى باستخدامه عن مؤشر η^2 على الرغم ان الفرق بينهما ضئيلة جداً.

ج - مؤشر d : هذا المؤشر طرحه (Steiger (2004 ويسمى Root – Mean square Standardized effect (RMSSE) ويقدر كالتى :

$$d = \text{RMSSE} = \frac{(\mu_j - \mu_G)}{\sigma}$$

- μ_j متوسط مجتمع كل مجموعة.
- μ_G متوسط متوسطات مجتمعات المجموعات Grand Mean.
- σ الانحراف المعيارى .

وهذه الصيغة فى المجتمع كالتى :

$$d = \sqrt{\left(\frac{1}{K-1}\right) \Sigma \left(\frac{\mu_j - \mu_G}{\sigma}\right)^2}$$

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 كالتى:

$$\Sigma(X_j - \bar{X}_G) = \Sigma (13 - 8.33)^2 + (8 - 8.33)^2 + (4 - 8.33)^2$$

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. أسفل Type of power analysis اختر:

Type of power analysis
Post hoc: Compute achieved power - given α , sample size, and effect size

لاحظ أننا اعتمدنا على التحليل البعدي Post hoc

3. أسفل Test family اختر F- test

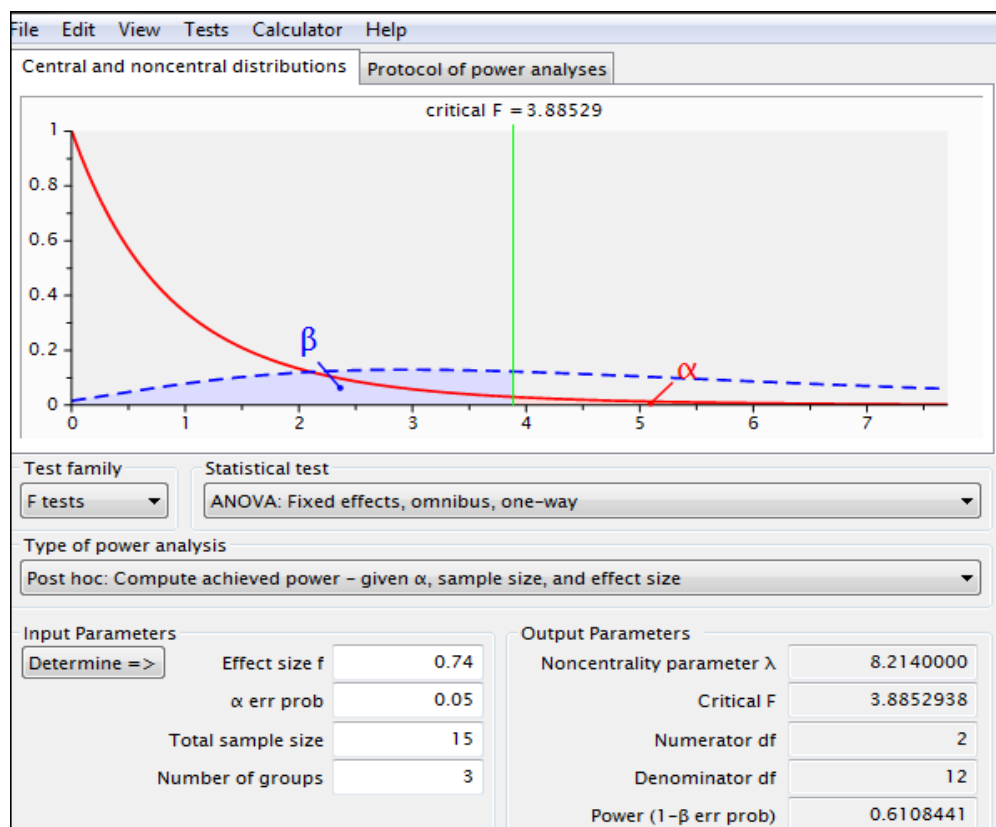
4. أسفل Statistical test اختر:

Test family	Statistical test
F tests	ANOVA: Fixed effects, omnibus, one-way

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 كالآتي:



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 كالاتى:

$$f = \sqrt{\frac{(k-1)(F-1)}{Nk}}$$

$$f = \sqrt{\frac{(3-1)(22.59-1)}{15 \times 3}} = \sqrt{\frac{2 \times 21.59}{45}} = 0.97$$

وبتعديل حجم التأثير فى المدخلات من 0.74 إلى 0.97 فإن المخرجات كالاتى:

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 طردياً مباشره مع حجم التأثير الحقيقى للمتغير المستقل كلما كان الحجم الحقيقى كبير تزيد القوة.
 - القوة ترتبط عكسياً مع تباين العينة، التباين المرتفع للعينة يؤدي إلى قوة المنخفضة للكشف عن الفروق الحقيقية.
- القوة وحجم العينة:** كلما زاد عدد الأفراد فى التجربة يؤدي إلى زياده قيمه F المحسوبة، حيث بزيادة N 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 يؤدي إلى زياده F وعليه تزيد احتماليه رفض H_0 وبالتالي زياده القوة الإحصائية للاختبار.
- القوة والتأثير الحقيقى للمتغير المستقل (حجم التأثير):** التأثير الحقيقى الكبير للمتغير المستقل يعطى قيم كبيره لـ $(\bar{X}_1 - \bar{X}_G)$ و $(\bar{X}_2 - \bar{X}_G)$ و $(\bar{X}_3 - \bar{X}_G)$ وعليه يؤدي إلى زيادة قيمه MS_B (البسط) وبذوره يزيد من قيمه F المحسوبة وهذا يؤدي إلى احتمال كبير لرفض H_0 بالتالى تزيد القوة للاختبار.
- القوة 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 زيادة SS_1 و SS_2 و SS_3 يؤدي إلى زيادة المقام لـ F وهو متوسط مجموع المربعات داخل المجموعات (الخطأ) وعليه فإن زيادة MS_W يؤدي إلى نقصان F المحسوبة وعلى ذلك فإن زيادة التباين داخل المجموعات يؤدي 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 دائماً أن تكون نسبة الخطأ من النوع الأول إلى الخطأ من النوع الثانى 4:1 وعليه فإن وضع $\alpha = 0.05$ فإن $\beta = 0.20$

2. تحديد مستوى القوة الإحصائية المرغوب: $P = 1 - 0.20 = 0.80$ وهذا المستوى يوصى به Cohen (1988) كأفضل مستوى مرغوب ويفضل ألا يقل عن 0.80.

3. تحديد أدنى حجم تأثير متوقع: يحدد في ضوء الدراسات السابقة، أو ماذا يتوقع الباحث من تأثير حقيقى للمتغير المستقل بفرض أن الدراسات السابقة لحجم تأثير ثلاثة طرق للمعالجة على متغير تابع هي $ES=0.74$. ومن ثم إذا اعتبرنا:

$$K=3, P = 0.80, ES = 0.74, \alpha = 0.05$$

فانه يتم تحديد حجم العينة في ضوء مدخلين:

1. **مدخل جداول (Cohen 1988):** فإن حجم العينة المطلوب لتحقيق هذا هو 40 لكل مجموعة وعليه فإن حجم العينة الاجمالى لتنفيذ التجربة: $40 \times 4 = 160$

2. المدخل الثانى هو استخدام التحليل القبلى للقوة الإحصائية Prior power باستخدام برنامج G.Power ولتحديد حجم العينة المطلوب لدراسة الفروق بين المجموعات الثلاثة فى الضغوط النفسية (المثال السابق):

1. افتح البرنامج تظهر الشاشة الافتتاحية:

2. أسفل Type of power analysis اختار

Type of power analysis
A priori: Compute required sample size - given α , power, and effect size

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:

Statistical test

ANOVA: Fixed effects, omnibus, one-way

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 كالاتى:

Test family: F tests

Statistical test: ANOVA: Fixed effects, omnibus, one-way

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters

Determine =>

Effect size f: 0.40

α err prob: 0.05

Power (1- β err prob): 0.80

Number of groups: 3

Output Parameters

Noncentrality parameter λ : ?

Critical F: ?

Numerator df: ?

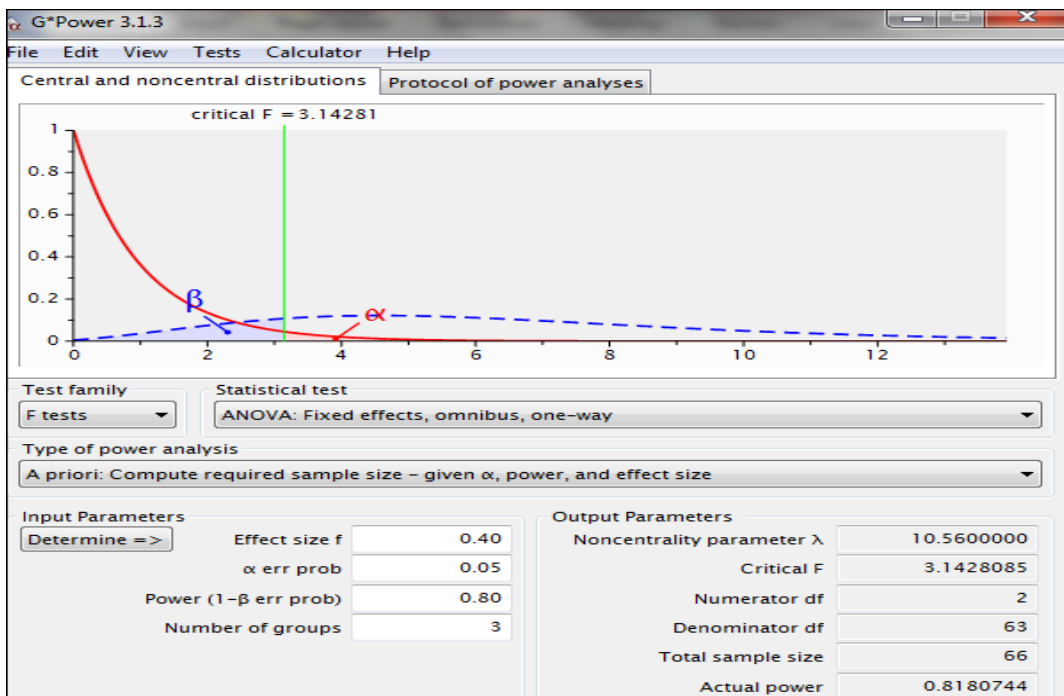
Denominator df: ?

Total sample size: ?

Actual power: ?

X-Y plot for a range of values

Calculate



لاحظ أن حجم العينة الاجمالي 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 بشأن قبول أو رفض H_0 . والعلاقة بين F و T:

$$F = T^2$$

وحيث إن اختبار T يقارن المسافة Distance بين متوسطى العينة (البسط) والمسافة المحسوبة للخطأ المعياري (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 هي :

1. تختبر نفس الفروض عند وجود معالجتين فقط وهي:

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

2. درجات الحرية لـ T و df لمقام اختبار F (df_w) واحدة فاذا كانت معالجتين لكل معالجة

سنة أفراد فإن $df \downarrow T$:

$$= n_1 + n_2 - 2 = 12 - 2 = 10$$

و df لمقام F (داخل المجموعات) هي:

$$df_w = N - K = 12 - 2 = 10$$

ودرجتى الحرية لـ F هي $df = (1, 10)$

3. يختلف توزيع T عن توزيع F فاذا كانت $df \downarrow T$ (18) و $df \downarrow F$

(1,18) فإن قيمه F الحرجة عند 0.05 وقيمة T الحرجة عند 0.05 كالآتي :

$$F = (2.10)^2 = (T \text{ الحرجة})^2 = 4.41$$

مثال لبيانات تقدر باستخدام T وباستخدام ANOVA

قضية (Hinkle et al. 1994):

اهتم باحث بإجراء مقارنة بين متوسط عدد المكالمات التي يتسلمها طلاب الجامعة المقيمين في

المدينة الجامعية ومقارنتها بعدد المكالمات لطلاب الجامعة المقيمين في منازلهم وتم اختيار

عينه عشوائية من المجتمعين العينة الأولى 10 طلاب والعينة الثانية 10 طلاب عند $\alpha =$

0.05

خطوات اختبارات الفروض باستخدام T, F: كما نعلم أنه يوجد متغيرين أحدهما مستقل وهو

مكان الإقامة بمستويين والمتغير التابع نسبي وهو عدد المكالمات والفروض الإحصائية لـ T

وايضاً 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

$$\Sigma T = 98 , T^2 = 912$$

$$T = \frac{3.5 - 6.30}{\sqrt{\frac{(10-1)(19.61) + (10-1)(24.01)}{10+10-2}} \left(\frac{1}{10} + \frac{1}{10}\right)} = \frac{-2.80}{\sqrt{4.36}} = -1.34$$

ولاختبار ذو ذيلين و $\alpha = 0.05$ فان:

$$T(الدرجة)(18) = \pm 2.101$$

وبما ان: $2.10 > 1.34$ وبالتالي يفشل الاختبار في رفض H_0 بحيث لا توجد فروق.

وبحساب F كالآتي:

$$SS_B = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} - \frac{(\sum T)^2}{N}$$

$$= \frac{(35)^2}{10} + \frac{(63)^2}{10} - \frac{(98)^2}{20}$$

$$= 519.40 - 480.20 = 39.20$$

$$SS_W = \sum T^2 - \left[\frac{(\sum X_1)^2}{10} + \frac{(\sum X_2)^2}{10} \right]$$

$$= 912 - \left[\frac{(35)^2}{10} + \frac{(63)^2}{10} \right]$$

$$= 912 - 51.40 = 392.40$$

$$SS_T = \sum T^2 - \frac{(\sum T)^2}{N}$$

$$= 912 - \frac{(98)^2}{10} = 912 - 480.20 = 431.80$$

$$MS_B = \frac{39.20}{1} = 39.20$$

وعليه فإن:

$$MS_W = \frac{392.60}{18} = 21.81$$

$$F = \frac{39.20}{21.81} = 1.80$$

وبالتالى:

وبالبحث فى جدول F بدرجة حرية المقام 8 ودرجة حرية البسط 1 ومستوى دلالة إحصائية

0.05 فان $F_{c.v} = 4.41$ الدرجة وعليه فإن:

$$F \text{ المحسوبة } (1.80) < F_{c.v} \text{ الدرجة } (4.41)$$

إذا نقبل الفرض H_0 بعدم وجود فروق في عدد المكالمات بين الطلاب ساكني المدن الجامعية والطلاب ساكني منازلهم.

إجراء تصحيح بونيفروني في حالة إجراء F مرات عديدة في نفس الدراسة

يمكن لباحث جمع بيانات لمتغيرين تابعين أو أكثر في نفس الدراسة ويوجد على الأقل ثلاث مجموعات فيقوم الباحث بحساب اختبار F لكل متغير تابع على حدة في الدراسة وهذا هو المؤلف في البحوث ويستخدم مستوى دلالة إحصائية واحدة لكل المقارنات للمتغيرات التابعة ولكن في 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 بونيفروني Bonferroni adjustment حيث يتم قسمة مستوى ألفا على عدد مرات إجراء ANOVA في الدراسة فإذا تم إجراء تحليل التباين ثلاث مرات فإن قيمة ألفا المعدلة

$$\text{هي: } = \frac{0.05}{3} = 0.016$$

وإذا تم إجراء تحليل التباين مرتين فإن ألفا المعدلة هي:

$$= \frac{0.05}{2} = 0.025$$

ويأخذ الباحث قراره في 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 الأبعاد.

كتابة تحليل التباين في تقرير البحث وفقاً لـ APA

في كتابة نتائج ANOVA في تقرير البحث فلا بد من عرض متوسطات المجموعات والانحرافات المعيارية كتابة في فقرة أو في جدول. على الرغم ان هذه الإحصائيات لا تدخل في حساب نسبه F ولكنها ضرورية. وفيما يلي تقرير الدراسة: " المتوسطات والانحرافات المعيارية تم عرضهما في جدول (). اشارت نتائج تحليل التباين إلى

وجود فروق دالة إحصائية بين التوقيعات الثلاثة فى الضغوط: $F(2,12) = 22.59$, $P < 0.05$, $\eta^2 = 0.792$

أو $F(2,12) = 22.59$, $P < 0.05$, $\omega^2 = 0.742$

والباحثون يفضلون كتابة ANOVA فى تقرير البحث كالاتى:

مصدر التباين	مجموع المربعات SS	درجات الحرية df	متوسطات المربعات MSS	نسبة F المحسوبة
بين المجموعات	203.33	2	101.667	22.59
داخل	54.00	12	4.500	-
الكلية	257.333	14	-	-

* داله إحصائياً عند $\alpha = 0.05$ ، $F_{c.v} = 3.88$ 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 إحصائياً: عندما يحصل الباحث على نتائج دالة إحصائياً من تحليل التباين الاحادى فهذا يعنى رفض H_0 والشئ 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 تفصيلاً أو خصوصية عن ما إذا كانت μ_1 أكبر من μ_2 أو أكبر من μ_3 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 رفض H_0 ونقص الصدق يجعل النتائج

غير مضمون تعميمها وهدف تحليل التباين هو التعميم على مستوى المجتمع من خلال متوسطات العينة.

- عدم رفض الفرض الصفري لا يعنى ان متوسطات كل متوسطات المجتمع متساوية.
- يجب إجراء تصحيح بونيفروني عند إجراء تحليلات عديدة لـ ANOVA 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 فى SPSS: مثال (Pagano(2013

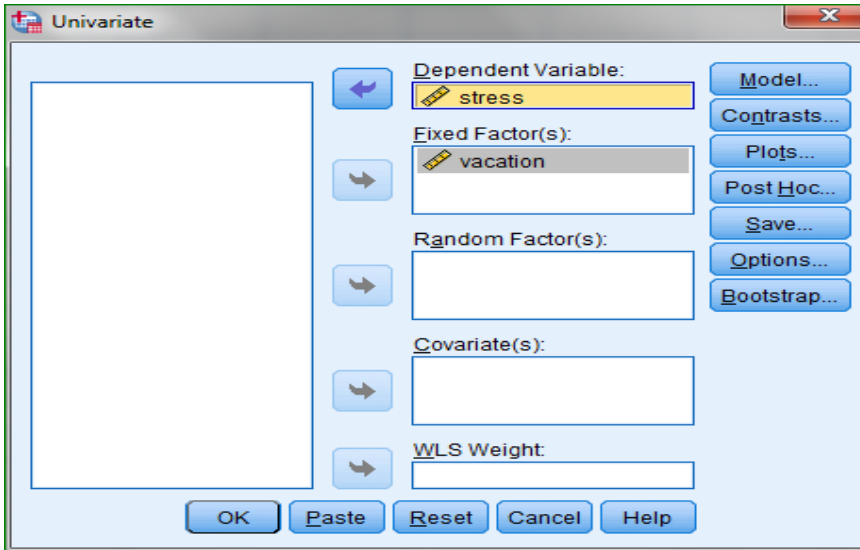
أولاً: إدخال البيانات: 1. أضغط Variable view، فى عمود Name حدد مسمى المتغيرات وهى كالاتى: **المتغير المستقل**: الأجازة Vacation ثم تكوده إلى ثلاث مستويات: (بعد العودة مباشرة من الأجازة (أول الترم)=1، منتصف الترم=2، نهاية الترم=3)، **المتغير التابع**: الضغوط النفسية Stress.

3. فى عمود Values حدد تعريف 1,2,3.

4. أضغط Data view تظهر شاشة البيانات بها عمودين. ابدأ فى إدخال المتغيرات.

ثانياً: تنفيذ الأمر: الطريقة الأولى: Analyze → General Linear Model

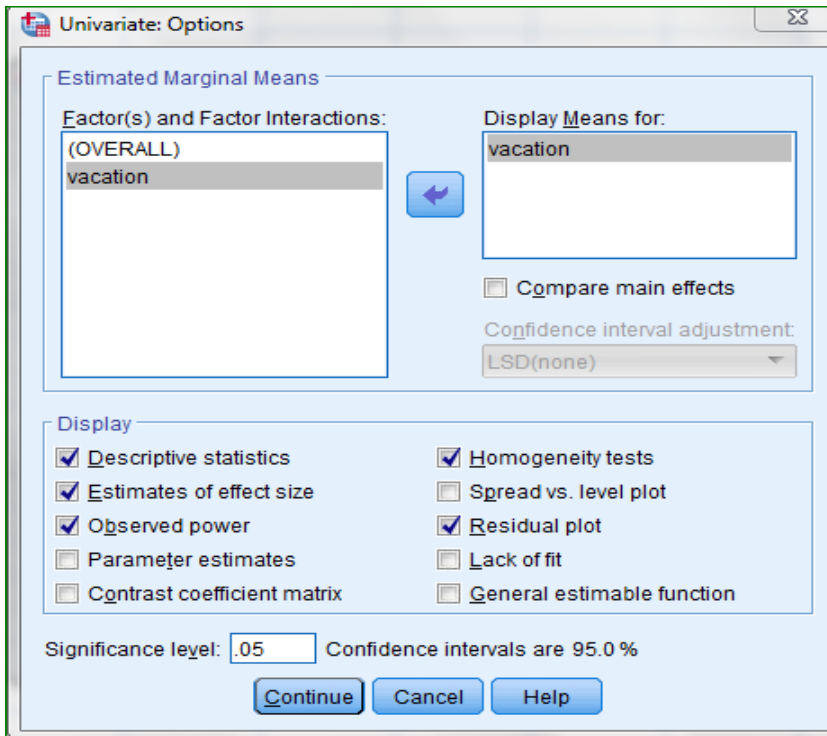
Univariate → تظهر الشاشة الآتية:



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:

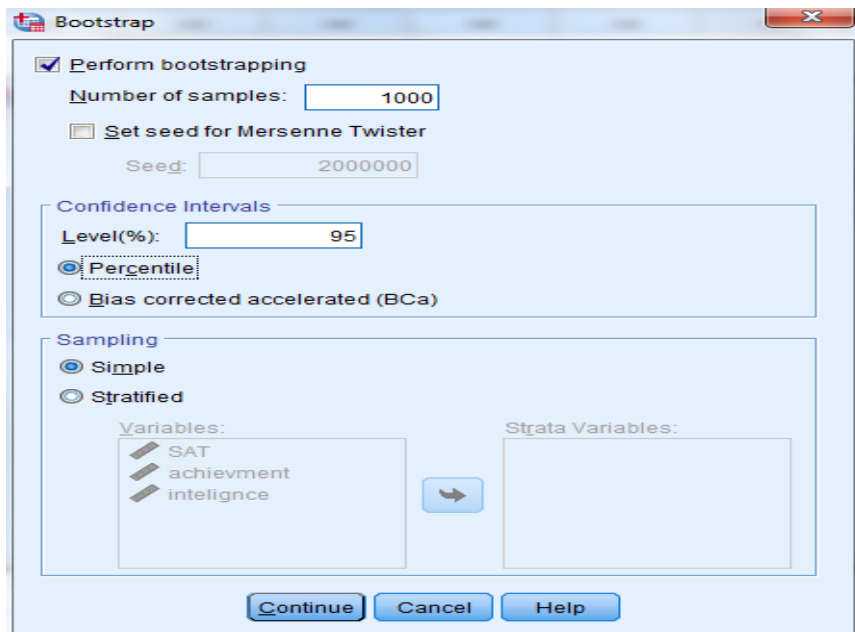


5. اضغط vacation في مربع Factor(s) and factor Inter وانقله إلى مربع Display Means

6. اضغط على Estimate of effect size, Descriptive statistics,

Residual plot, Observed power, Homogeneity tests,

7. اضغط على Bootstrap تظهر النافذة الآتية:



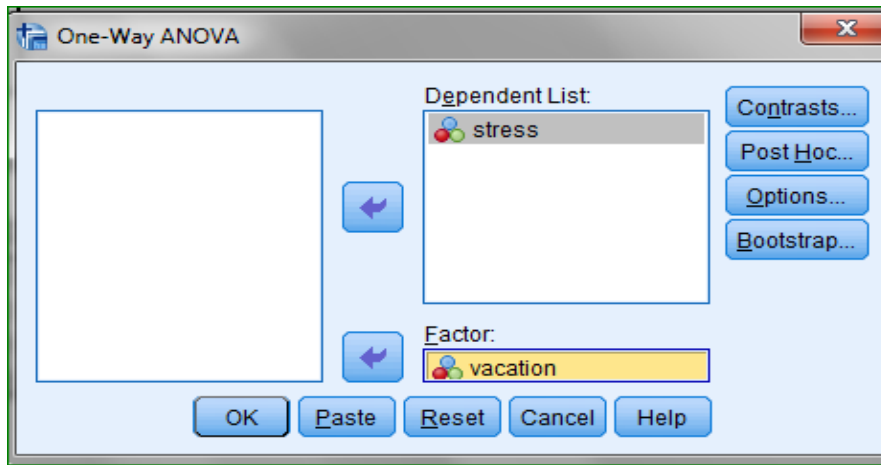
8. اضغط Perform bootstrapping :

- حدد عدد العينات التي ترغب في توليدها من قاعدة البيانات المتاحة وحدد عدد العينات المراد توليدها ويمكن تحديد عدد العينات بـ 200 مثلاً، حيث إن أحجام العينات الصغيرة إلى المتوسط من 10 إلى 100 عينة.
- حدد فترات الثقة التي ترغبها 99% لمستوى دلالة 0.05 أو 99% لمستوى دلالة 0.01
- حدد نوعية العينة بسيطة أو طبقية Stratified وهي بسيطة.

9. اضغط Continue ثم اضغط Ok

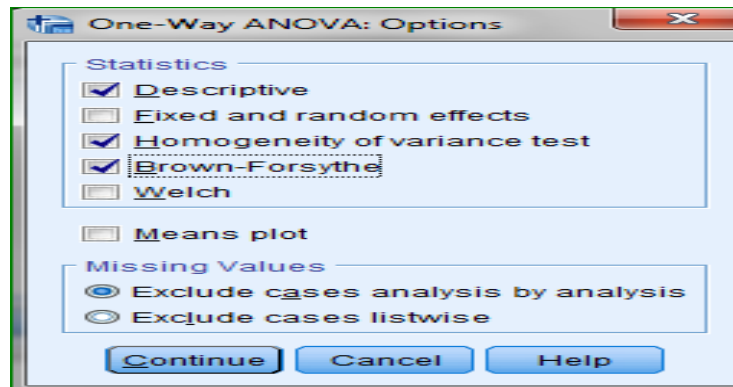
الطريقة الثانية: 1. اضغط Analyze → Compare means → One way ANOVA

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 variance test, واختبارى - Brown forsythe , welch

Between-Subjects Factors		
		N
vacation	1.00	5
	2.00	5
	3.00	5

4. اضغط Continue ثم اضغط OK

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 كالاتى:

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 المعياري Std. Deviation لدرجات كل عينة.

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.			

الجدول الثالث: اختبار Levin's اختبار

تجانس البيانات.

والوضح أن قيمة $F=0.686$ ،

$df_1=(c-1)=2$, $df_2=(15-3)=12$

والدلالة الإحصائية $sig=0.522$ وبما أن $p > 0.05(0.522)$ ، وعليه نقبل H_0

وبالتالي 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 ل Vacation هي $F=22.59$ ، $F=\left(\frac{101.667}{4.5}\right)$ ،

Error هي مجموع المربعات داخل المجموعات أو الخطأ أو البواقي بما أن $Sig=0.000$

وعليه $0.000 < 0.05$ ، وبالتالي نرفض H_0 وعليه توجد فروق ذات دلالة إحصائية فى

الضغوط بين المجموعات الثلاثة.

وأوضح أن مؤشر حجم التأثير (مربع إيتا) الجزئية هي: $\eta_p^2=0.790$ وهي حجم تأثير من

النوع القوي بمعنى أن الأجازة فسرت 79 % من تباين تقدير الذات.

ولاحظ أسفل هذا الجدول أعطى : $R^2_{adj} = 0.755$, $R^2_{square} = 0.790$ ، على اعتبار ان

تحليل التباين هو جزء من تحليل الانحدار وأن قيمة R^2 هي تقريباً نفسها قيمة η_p^2 وهذا هو

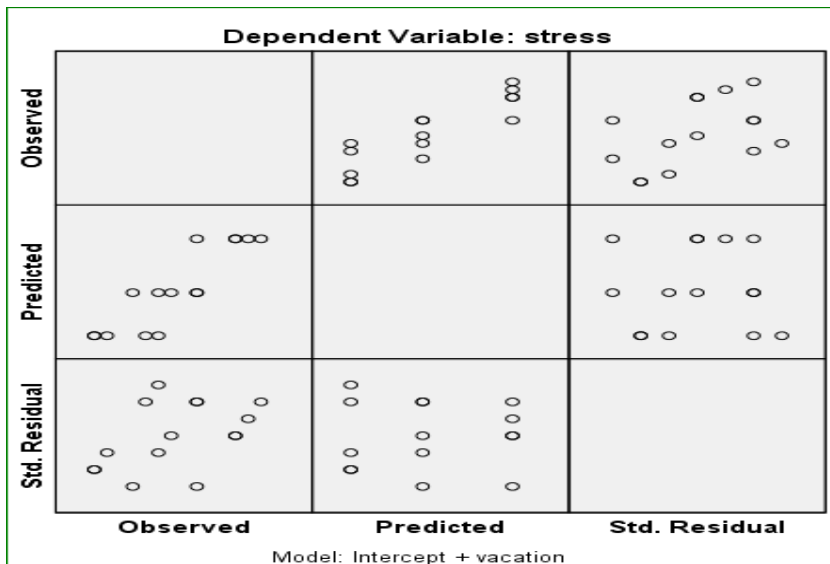
مؤشر لمطابقة النموذج Model fit.

وأعطى البرنامج الجدول الآتي:

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

وهي متوسطات كل عينة والخطأ المعياري للمتوسط.

ثم أعطى البرنامج scatter plot للمتغير التابع stress لكل من الدرجات المقاسة والدرجات المتنبأ بها وقيم البواقي المعيارية:



والمهم في هذا العرض البياني هو الصف الأوسط (الثاني) حيث تختبر مسلمة التوزيع العشوائي لأخطاء القياس Homoscedastiaty.

نتائج البوتستراب:

Bootstrap for Multiple Comparisons

Dependent Variable: stress

			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 البوتستراب 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

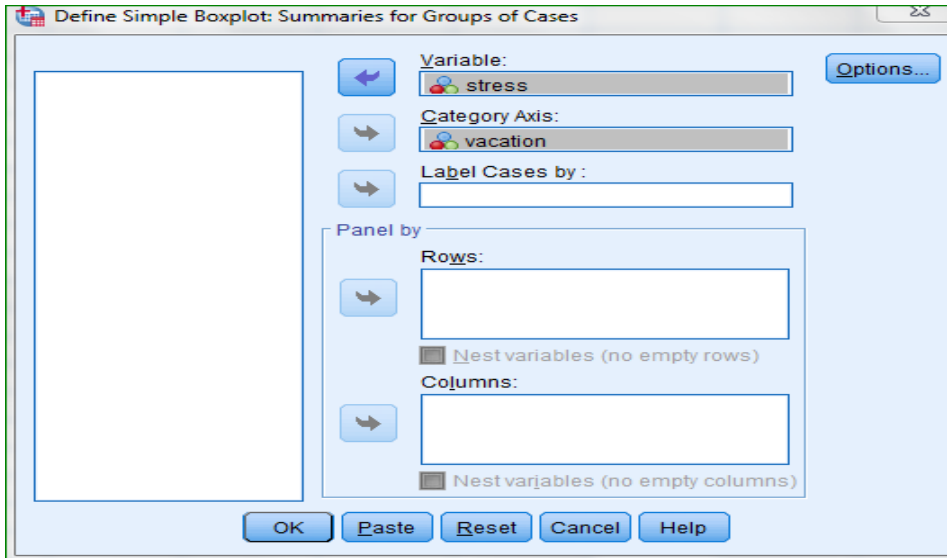
وكذلك جدول ANOVA كالتالي:

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 بيانياً

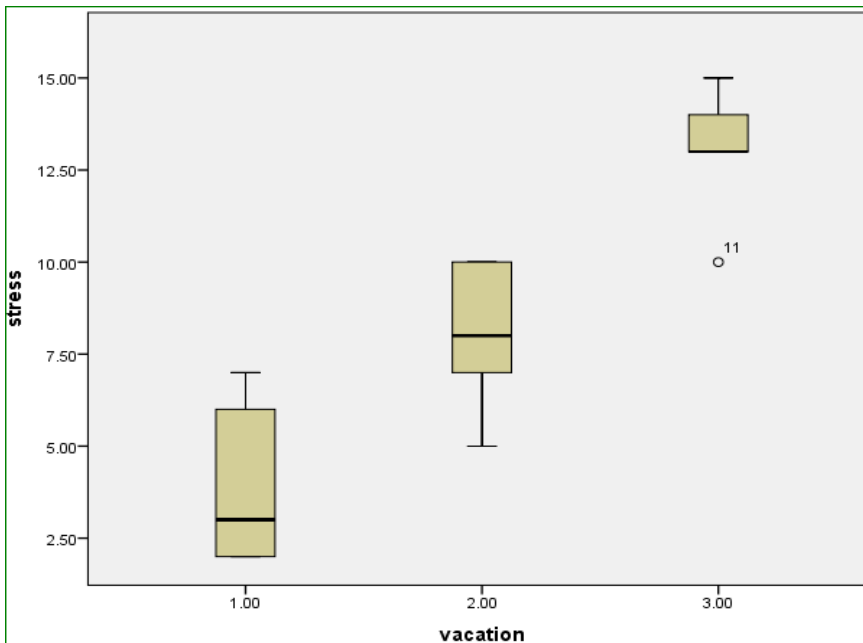
1. اضغط Graphics → Legacy Dialogs → Boxplot

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 في الفترات الثلاثة مختلف، حيث أعلاهم هو في نهاية الترم يليه منتصف الترم ثم في بداية الترم. وتعتبر قيمة الحالة 11 في توزيع نهاية الترم 3 قيمة متطرفة وهي القيمة 10 ولكنها غير ذلك في 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 إحصائياً تشير إلى ان كل المعالجات أو مستويات المتغير المستقل ليس لها نفس التأثير على المتغير التابع، وفى تحليل التباين يتم اختيار الفرض الصفرى $\mu_1=\mu_2=\mu_3$ بينما تكون قيمة الخطأ فى النوع الأول كما حددها الباحث مسبقاً ولتكن 0.05 أو 0.01 أو غيرها وهذا يسمى احياناً بـ Omnibus test، وفى 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 إحصائياً 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.

فافتراض أن باحثاً قارن بين ثلاثة عينات أو معالجات وافترض أن متوسطات العينات هى $\bar{X}_1=3$ ، $\bar{X}_2=5$ ، $\bar{X}_3=10$ وعلى ذلك يوجد ثلاثة فروق بين المتوسطات هى: فرق نقطتين بين $\bar{X}_1$ ، $\bar{X}_2$ ، فرق خمس نقاط $\bar{X}_1$ ، $\bar{X}_3$ ، وفرق سبعة نقاط بين $\bar{X}_2$ ، $\bar{X}_3$ ، فلو استخدم ANOVA لتقييم 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 ($\mu_1 \neq \mu_2$) ولكن ماذا عن فرق الخمس نقاط هل إنه كافى لإحداث دلالة إحصائية؟، وماذا عن الفرق بين $\bar{X}_1$, $\bar{X}_2$ هل هو دال إحصائياً أيضاً؟ 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 وإيهما غير دال إحصائياً. وعلى ذلك فإن الاختبارات البعدية في ANOVA يتم إجرائها فى حالة:

1. رفض الفرض الصفري 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 فمثلاً عند إجراء اختبار T للمقارنة بين مجموعتين ونرفض الفرض H_0 حيث T المحسوبة تزيد عن T الحرجة فإن الباحث يعمل فى نسبة PC وهى 0.05. وكثيراً 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 كالاتى:

$$\alpha_E = C\alpha \text{ (FW)}$$

• 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 قسمه α_E 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 ($\alpha E=0.05$, $C=10$) 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 إحصائياً.

وقد صنف (Hinkle et al. 1994) المقارنات المتعددة إلى:

أولاً: اختبار المقارنات المتعددة أو البعدية اللاحقة لـ F 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 A فى مقابل B و A فى مقابل C ومقارنة C فى مقابل B ولأربع مجموعات توجد ستة مقارنات، عدد أزواج المقارنات يتحدد بالآتى:

$$= \frac{4 \times 3 \times 2}{4} = \frac{24}{4} = 6$$

ومصطلح Non-pairwise or complex فيه يتم مقارنة متوسط ما فى مقابل باقى المتوسطات، وفى المجموعات A, B, C تتم مقارنة المجموعة A مع متوسط B, C معاً.

وهذه الفئة من المقارنات المزدوجة تتضمن:

أ- اختبارات فى حالة تساوى أحجام العينات فى المجموعات وتتضمن اختبارات توكى Tukey ونيومان-كيولز Newman-keuls.

ب- فى حالة عدم تساوى أحجام العينات فى المجموعات وتتضمن اختبار توكى - كرامير Tukey-kramer.

2. المقارنات المعقدة: تتضمن اختبار شيفيه Scheffe

ثانيًا: المقارنات المخططة **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 مسبقاً (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 مسبقاً 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 الصفري فى المقارنات المتعددة البعدية أصعب من إجراء المقارنات المخطط لها مسبقاً وعلى ذلك فإن المقارنات القبلية لها قوة إحصائية أعلى من المقارنات البعدية العديدة بين كل الأزواج الممكنة بين المجموعات.

وبتطبيق ذلك على تجربة قام بها باحث للمقارنة بين أربعة أنواع مختلفة من العلاج للاكتئاب A, B, C, D، فالباحث عنده توقع مسبقاً عن مقارنة بين طريقتين A, C فهذه مقارنة مخطط لها. ولكن إذا أراد الباحث إجراء مقارنات بين كل زوج من أنواع العلاج ستة مقارنات ففى هذه الحالة يعتمد الباحث على مقارنات بعدية ولكن الاجابة التقليدية أو الإجراء التقليدى هو إجراء مقارنات بين كل الأزواج الممكنة.

وحقيقه يوجد جدل حول ما إذا كان الباحث يقوم بإجراء المقارنات المتعددة فى حاله ما إذا كان قيمة F غير دالة إحصائياً فكانت الاجابة القديمة بدون وجود دلالة إحصائية فلا داعى أو من غير المناسب إجراء المقارنات المتعددة ولكن (2013) Howell 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 تماماً 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 ودلالاتها الإحصائية للمقارنات المتعددة صممت لدلالاتها الإحصائية بدون النظر أو الاعتبار لقيمة F، لاحظ أن التحليل المبدئي لـ ANOVA يطلق على قيمة F بـ Omnibus F-test، بل ان Wilcox (1987) أشار إلى أنه لا يوجد منطقية لتطبيق اختبار F على الإطلاق وبدلاً 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 كالاتى:

	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 كالاتى:

F	MS	df	SS	مصدر التباين
27.33	87440	4	349760	بين المجموعات
	32.00	35	1120.00	داخل المجموعات (الخطأ)

P<0.05

وعلى ذلك توجد فروق دالة إحصائية بين المجموعات الخمسة.

حجم التأثير: يقدر حجم التأثير من خلال مؤشر:

$$\eta^2 = \frac{SS_B}{SS_{total}} = \frac{3497.60}{4617.60} = 0.76$$

وهذا يعنى ان طريقة التنشئة فسرت 76% من تباين التحصيل.

والنقدير 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 احياناً Contrasts ويتم التخطيط لها قبل جمع البيانات

وتوجد العديد من هذه النوعية من المقارنات أهمها:

أولاً : 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 حساب T مع MSerror التى تدخل فى حساب تحليل التباين وتقييم اختبار T فى ضوء درجات الحرية الخطأ أو داخل المجموعات dfw (N-K) ولكن (2013) Howell 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 وتقييم T فى ضوء df = (N-2) وعندما يحدث عدم تجانس للتباينات وأحجام 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، فمثلاً مع إجراء مقارنتين أو اختبار T مرتين فإن قيمتها تتضخم إلى 0.02 عندما تكون $\alpha=0.01$.

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 $\alpha=0.05$ 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 دوراً كبيراً فى إحداث هذه الفروق.

وبعد ان تناولنا أحد طرق المقارنات القبلية نتطرق إلى طرق اخرى مثل المقارنات العمودية Orthogonal-contrasts واختبار Dunn-sidak و Bonferroni و Dunn's test من خلال مخرج برنامج SPSS نظر لتعقيدات حسابية.

تصحيح بونيفروني 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 تحفظاً، 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 $\alpha=0.5$ الخطأ من النوع الأول، فإن احتمال ان لا يزيد الخطأ من النوع الأول عن $0.15 = 0.05 \times 3$.

فاذا كانت C هي عدد المقارنات و α احتمال الخطأ من النوع الأول لكل مقارنة بين مجموعتين فان:

$$Fw \leq C\alpha$$

واذا تم وضع $\frac{\alpha}{C} = \alpha$ 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 هي:

$$\alpha = \frac{0.05}{3} = 0.0167$$

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 بونيفروني 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 و SAS اختبار بونيفروني مع اختبارات المقارنات البعدية ولذلك فإن استخدامه غير مناسب فى هذه الحالة لأنه يقع ضمن المقارنات القبلية المخطط لها

مسبقاً. ويتطبيق اختبار بونيفروني على المثال السابق للمقارنة بين المجموعة الخامسة والثالثة كالآتي:

$$= \frac{29 - 10}{\sqrt{\frac{2(32)}{8}}} = \frac{19}{\sqrt{8}} = 6.72$$

وبين المجموعة الثالثة والأولى:

$$= \frac{11 - 4}{\sqrt{\frac{2(32)}{8}}} = \frac{7}{\sqrt{8}} = 2.47$$

وبين المجموعة الثانية والثالثة:

$$= \frac{10 - 11}{\sqrt{\frac{2(32)}{8}}} = \frac{-1}{\sqrt{8}} = -0.35$$

وفي حالة إجراء أربع مقارنات فإن:

$$F_w = 3 \times 0.05 = 0.15$$

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 $\alpha = 0.0167$ و $df_w = 40$ $35 = 5 -$ وعليه فإن:

$$T \left(\frac{\alpha}{2}, 35 \right) = \pm 2.052$$

ومن ثم يرفض H_0 في حالة المقارنة بين المجموعة الخامسة والثالثة بينما يتم قبول الفرض H_0 في 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 بونيفرونى وهذا متاح فى برنامج SPSS وبمقارنة القوة بين اختبار بونيفرونى واختبار دان - سيداك فإن القوة الإحصائية تميل قليلاً لصالح دان - سيداك.

إجراءات بونيفرونى متعددة المراحل Multistage Bonferonni procedures

كما سبق ووضحنا ان اختبار بونيفرونى يستخدم عند اختبارات فروض متعدد لضبط نسبة الخطأ (Fw) وهذه الإجراءات لها ميزة فى وضع إطار محدد لخطأ Fw عند α عند إجراء الفروض الصفرية المحتملة بين أزواج المتوسطات وذلك مثل اختبار توكى ولكنه أقل تحفظاً 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 إحصائياً.

كتابة نتائج اختبارات المقارنات القبلية فى البحث وفقاً لـ APA

اتضح وجود تأثير دال إحصائياً بين المجموعات الخمسة، حيث اتضح وجود فروق دالة إحصائياً بين المجموعة الخامسة ($\bar{x} = 29$) مقارنة بالمجموعة الثالثة ($\bar{x} = 10$) حيث:

$$T(35) = 6.72, tc.v(\frac{\alpha}{3} = .0167) = \pm 2.52$$

ثانياً: المقارنات البعدية: عندما لا تكون المقارنات مخطط لها مسبقاً فإننا نستخدم اختبار المقارنات البعدية حيث يقوم الباحث بإجراء مقارنات بين كل الأزواج 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 بونيفروني فى حالة وجود عدد محدود من المقارنات المحدد مسبقاً. ويرى (2013) Howell 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 لفisher'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 اختبارات T المتعددة القبلية الذى تتاولناه 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 كالاتى:

$$Fisher (LSD) = T_{c.v} \sqrt{MS_w \left[\frac{1}{n_1} + \frac{1}{n_1} \right]}$$

وقيمة $T_{c.v}$ الحرجة من خلال درجات الحرية dfw مثلاً:

$$LSD_{(3,5)} = 2.052 \sqrt{32 \left[\frac{1}{8} + \frac{1}{8} \right]}$$

وبمقارنة T الحرجة يتم رفض H_0 ، ولكن البعض يوصى بعدم استخدام هذا الاختبار لأنه لا يقي نفسه من تضخم الخطأ من النوع الأول ولكن إذا كانت التجربة تتضمن ثلاثة مجموعات أو متوسطات فإن هذا الإجراء ممتاز لأن قيمة FW تبقى ثابتة عند α المفترضة، وهو أكثر اختبارات المقارنات البعدية ليبرالية وفي نهاية المتصل يعتبر شيفيه أكثر تحفظاً .

إحصاء Studentized Range Statistic (Q)

نعلم ان التوزيع الاحتمالي للمقارنات البعدية قائم على هذا الإحصاء ويتحدد هذا التوزيع بالاتي:

$$Q_r = \frac{\bar{X}_L - \bar{X}_u}{\sqrt{\frac{MS_w}{n}}}$$

$\bar{X}_L - \bar{X}_u$ فرق متوسطات المعالجات الأكبر والأصغر و k عدد المعالجات.

وإحصاء Q مشابهًا تمامًا لصيغة T في المقارنات القبلية.

$$T = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{2MS_W}{n}}}$$

الفرق الوحيد هو ان صيغة Q تتضمن $\sqrt{2}$ في المقام أى ان:

$$Q = \frac{T}{\sqrt{2}}$$

وعلى ذلك ففي توزيع Q يختلف عدد المجموعات بينما عدد الأفراد في المجموعات لا يتغير وإحصاء Q توزيع عيني وقيمته الحرجة تعتمد على عدد المجموعات (K) ودرجات الحرية المرتبطة بتقدير متوسطات مجموع المربعات داخل المجموعات (الخطأ) MS_w . وعليه فإن اختبارات المقارنات البعدية مثل توكي وتوكي-كرامير ونيومان-كيلوز تستخدم توزيع Q كتوزيعات عينية لضبط نسب تضخم الخطأ من النوع الأول.

ولحساب اختبار Q لمتوسطات المجموعات الخمسة السابقة : $\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4, \bar{X}_5$ على التوالى: 29, 24, 11, 10, 4

و كما نعلم ان: $MS_w = 32$, $df_w = 40 - 5 = 35$, $n = 8$, وان أعلى متوسط $\bar{X}_5 = 29$ وان أدنى متوسط $\bar{X}_1 = 4$, 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 جدول Q بـ $df = 35$, $k = 5$ يتضح ان:

$$Q_{(0.05)}(5, 35) = 4.07$$

وحيث إن : $Q_{\text{المحسوبة}}(4.07) < Q_{\text{الدرجة}}(12.5)$ ، وعليه نرفض H_0 أى يوجد فرق دالة إحصائية بين المتوسط الأعلى والمتوسط الأدنى. 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 (WSD) Wholly Significance Difference ويستخدم إحصاء Q للمقارنات ما عدا Q_{HSD} لتوكى هو القيمة العظمى لـ Q_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 لـ $Q_{c.v} = 4.07Q$ وعليه:

$$\begin{aligned}\bar{X}_i - \bar{X}_j &= Q \sqrt{\frac{MS_w}{n}} \\ &= 4.07 \sqrt{\frac{32}{8}} = 8.14\end{aligned}$$

فإن كان الفرق بين أى متوسطين أكبر من 8.14 فإن قيمة Q دالة 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 الفرق بين أزواج المتوسطات دالة إحصائيًا لأن:

$$| \bar{X}_i - \bar{X}_j | (8.14) > Q \text{ الحرجة}$$

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 أزواج المتوسطات وعليه فإن:

$$\bar{X}_1 = \bar{X}_2 = \bar{X}_3 \neq \bar{X}_4 = \bar{X}_5$$

والإشارة تساوى تشير إلى عدم رفض H_0 . ويمكن ترجمتها إلى:

$$(\bar{X}_1 = \bar{X}_2, \bar{X}_1 = \bar{X}_3, \bar{X}_2 = \bar{X}_3, \bar{X}_4 = \bar{X}_5)$$

$$\bar{X}_1 \neq \bar{X}_4, \bar{X}_1 \neq \bar{X}_5, \bar{X}_2 \neq \bar{X}_4, \bar{X}_2 \neq \bar{X}_5, \bar{X}_3 \neq \bar{X}_4, \bar{X}_3 \neq \bar{X}_5 \quad \text{بينما:}$$

وعليه نقبل الفرض البديل.

صيغة Tukey فى 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 الحرية 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 أزواج المتوسطات واتضح ان الفروق بين قيم المتوسطات دالة إحصائياً.

3. طريقة نيومان-كيولز Newman-keuls method: هذه الطريقة مشابهة لطريقة توكى ماعدا انها تعتمد على الفروق بين المتوسطات المرتبة. ودار حولها جدلاً كبيراً الا انها مازالت تستخدم فى برنامج SPSS حيث تتغير القيمة الحرجة لـ Q فى ضوء المدى لمجموعة المتوسطات فعندما يتم ترتيب المتوسطات للمجموعات من

الأدنى إلى الأعلى فإذا حدثت مقارنة بين المجموعة الأولى والثالثة فإن عدد المجموعات $K=3$ ، وليس $K=5$ كما هو الحال في اختبار توكي ويبدو ان ميزه اختبار نيومان-كيولز أنه اختبار قوى إحصائياً ولكن من عيوبه أنه يفتقد بعض الضبط α .

فعند المقارنة بين $\bar{X}_5$ ، $\bar{X}_1$ فإن $k = 5$ ، وعند المقارنة بين $\bar{X}_3$ ، $\bar{X}_1$ فإن $k = 3$ ، وعند المقارنة بين $\bar{X}_3$ ، $\bar{X}_2$ 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 توكي الذي يعتبر فيه قيمة Q الحرجة واحدة لكل المقارنات، وحسابات توكي هي نفسها حسابات نيومان-كيولز ماعدا الاختلاف في قيمة Q الحرجة وكلاً منهما يستخدم ومناسب للمقارنات المزدوجة Pairwise وإذا كانت نسبة F دالة إحصائياً في تحليل التباين وعندما تكون أحجام العينات متساوية.

وعليه فإنه توجد اختلافات في نتائج توكي عن نيومان-كيولز فيما يتعلق بالدلالة الإحصائية للفرق بين متوسطين حيث توجد فروق دالة إحصائياً باستخدام نيومان-كيولز مقارنة بطريقة توكي ولهذا السبب فإن طريقة نيومان-كيولز أكثر قوة إحصائياً من توكي ولكن هذه الميزة تقابل بمحدد لنيومان-كيولز وهو ان نسبة الخطأ من النوع الأول يحدث لها تضخم عن مستوى ألفا المفترض مسبقاً (α) ولكن هذا التضخم α يكون أقل من تضخمها في حالة استخدام المقارنات المتعددة باستخدام T ولكن باستخدام طريقة توكي تظل قيمة α ثابتة عند المستوى المفترض مسبقاً.

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 توكي ولكنه نادر الاستخدام.

5. اختبار بنجاميني - 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) بدلاً من FM وتساءلوا عن ماهى النسبة المئوية للنتائج الدالة إحصائياً ويمكن بعد ذلك ايجاد النتائج الدالة إحصائياً بالخطأ. فافترض توجد تسع مقارنات ووجدنا اربع تأثيرات دالة إحصائياً واحد هذه التأثيرات الدالة إحصائياً هى حقيقة خطأ من النوع الأول (رفض H_0 وهو فى المجتمع صحيح) ويعرف FDR بالآتى:

$$FDR = \frac{\text{عدد المقارنات الدالة وخطأ}}{\text{عدد المقارنات الدالة وهى حقيقة}} = \frac{1}{4} = 0.25$$

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 كالاتى:

$$Q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSw \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

مثال: افترض أن باحث درس تقدير الذات بين أربع مجموعات أو عينات وكانت نتائج ANOVA كالاتى:

المجموعة	الأولى	الثانية	الثالثة	الرابعة
$\bar{X}$	21.86	23	17.29	16.80
N	7	9	7	10

وملخص جدول ANOVA كالتالي:

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 الباحث $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$:

ولإجراء المقارنات المتعددة كالتالي:

بين المجموعة الأولى والثانية:

$$Q_{1-2} = \frac{21.86 - 23}{\sqrt{8.13 \left(\frac{\frac{1}{7} + \frac{1}{9}}{2} \right)}} = -1.12^*$$

بين المجموعة الأولى والثالثة:

$$Q_{1-3} = \frac{21.86 - 17.29}{\sqrt{8.13 \left(\frac{\frac{1}{7} + \frac{1}{7}}{2} \right)}} = 4.24^*$$

بين المجموعة الأولى والرابعة:

$$Q_{1-4} = \frac{21.86 - 16.80}{\sqrt{8.13 \left(\frac{\frac{1}{7} + \frac{1}{10}}{2} \right)}} =$$

بين المجموعة الثانية والثالثة:

$$Q_{2-3} = \frac{23.0 - 17.29}{\sqrt{8.13 \left(\frac{\frac{1}{9} + \frac{1}{7}}{2} \right)}} = 5.62^*$$

بين المجموعة الثانية والرابعة:

$$Q_{2-4} = \frac{23 - 16.18}{\sqrt{8.13 \left(\frac{\frac{1}{9} + \frac{1}{10}}{2} \right)}} = 6.69^*$$

المجموعة الثالثة والمجموعة الرابعة:

$$Q_{1-2} = \frac{17.29 - 16.80}{\sqrt{8.13 \left(\frac{1}{2} + \frac{1}{10} \right)}} = 0.49$$

ولحساب القيمة الحرجة لـ Q بدرجات حرية: $k=4, df = 33 - 4 = 29$

إذا $Q_{\text{الحرجة}} = 3.86$

وعليه فإن الفروق بين المجموعتين الأولى والثالثة، والمجموعة الأولى والرابعة، والمجموعة الثانية والثالثة، وبين الثانية والرابعة هي فروق دالة إحصائية، 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 إليه 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

فى كثير من المواقف يهتم الباحث باختبار فروض أكثر تعقيداً مقارنة بالمقارنات البسيطة بين كل زوج من المتوسطات ويعتبر اختبار شيفيه من أقدم الاختبارات وقائم على إجراء Tukey ويسمح بالمقارنة بين أى متوسط وباقى متوسطات كل الأزواج معا عند α . وليس مقارنات مزدوجة (كل متوسطين معاً). ولا ينصح شيفيه باستخدام لكل زوج على حدة وعلى الرغم عن ذلك فإن برنامج SPSS لا يأخذ بهذه النصيحة. وفى اختبار شيفيه فإن كل فرض يصاغ فى ضوء اتحاد خطى للمعاملات والمتوسطات وهذا الاتحاد يشار اليه بـ Contrast. وشكل الفرض الصفرى يكون كالاتى:

$$H_0: \sum_{k=1}^k C_K \mu_X = 0 \quad \text{حيث} \quad \sum_{k=1}^k C_K = 0$$

ويضاف المتوسط مضروباً 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

$$\text{المقارنات: } 3 \mu_2 = \mu_1 + \mu_3 + \mu_4$$

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:

$$\left(\frac{1}{3} - 1 + \frac{1}{3} + \frac{1}{3}\right) = 0$$

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 الضروري تصحيح هذه المعاملات وهي C_4, C_3, C_1 ويتم تقديرها كالاتى:

$$C_1 = \frac{7}{7 + 7 + 10} = 0.24$$

$$C_2 = -1 \left(\frac{9}{9} \right) = -1$$

$$C_3 = \frac{7}{7 + 7 + 10} = 0.29$$

$$C_4 = \frac{10}{7 + 7 + 10} = 0.42$$

وتكون صيغته طريقه شيفيه:

$$F = \frac{\sum (C_K \bar{X}_K)^2}{MS_w \left[\frac{C_K^2}{nK} \right]}$$

والتوزيع العيني هو اختبار F وتتحدد القيمة الحرجة :

$$\text{Critical value} = F_{c.v} \times (K - 1)$$

وعليه فإن القيمة الحرجة ل F عند $\alpha = 0.05$ ، $dF_1 = 4 - 1 = 3$ ، (البسط) $dF_2 =$

$$F_{c.v} = 2.93 \quad : \quad 33 - 4 = 24 \text{ (المقام) هي}$$

والقيمة الحرجة المستخدمة فى شيفيه:

$$F_{c.v} \times (K - 1) = 2.93 * 3 = 8.79$$

وعليه فإن قيمة شيفيه:

$$F = \frac{(C_1 \bar{X}_1 + C_2 \bar{X}_2 + C_3 \bar{X}_3 + C_4 \bar{X}_4)^2}{MS_w \left(\frac{(C_1)^2}{n_1} + \frac{(C_2)^2}{n_2} + \frac{(C_3)^2}{n_3} + \frac{(C_4)^2}{n_4} \right)}$$

$$F = \frac{(0.29 (2.86) + (-1)(23.0) + 0.29 (17.29) + 0.42 (16.80))^2}{8.13 \left(\frac{(0.29)^2}{7} + \frac{(-1)^2}{9} + \frac{(0.29)^2}{7} + \frac{(0.42)^2}{10} \right)}$$

$$= \frac{21.07}{1.24} = 16.96$$

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 (Hinkel et al. 1994).

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 وارتدت إجراء مقارنة واحدة 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) .

- وإذا كان لديك مقارنات قبلية عديد (قبل جمع البيانات) وليس بالضرورة بين كل زوج

فاستخدم إجراء Multistage Bonferroni

- إذا 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 خطأ 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	54	12	4.500		
المجموعات				22.59	3.88
الكلية	257.333	14			
المجموعة الأولى			المجموعة الثانية	المجموعة الثالثة	
		4	8	13	
		5	5	5	
					N

أولاً: المقارنات القبلية (المحددة قبل جمع البيانات):

1. إجراء اختبار T المتعددة.

أ - مقارنة المجموعة الأولى والثالثة:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2MS_w}{n}}}$$

$$T = \frac{4 - 13}{\sqrt{\frac{2 \times 1.5}{5}}} = -6.71$$

ب - المجموعة الثانية والثالثة:

$$T = \frac{8 - 13}{\sqrt{\frac{2 \times 4.5}{5}}} = -3.74$$

حيث قرر الباحث قبل إجراء التجربة إجراء المقارنات في حالة الحصول على دلالة

إحصائية لـ F وبالمبحث في جدول T بـ $\alpha = 0.05$ ، $df_w = N - K = 15 - 3 = 12$ ،

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 توكي (HSD): الخطوة الأولى: حساب قيمه Q وهى كالآتى:

أ. 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)،

$$df_w = 15 - 3 = 12, \alpha = 0.05, \text{ فان:}$$

$$Q_{c.v} = 3.77 \text{ (الحرجة)}$$

وعليه فإن قيمه Q المحسوبة للمقارنة الثلاثة $Q < Q_{c.v}$ (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 بقية 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		
P	T	P	Q	P	F	مقارنات المجموعات
						2*1
0.036		0.029	4.21		4.44	
0.00001	-	0.001006	9.49	0.00009	22.50	3*1
6.71						
0.001	-	0.008	5.27	0.010	6.49	3*2
3.73						

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 لاختبار شيفيه تتراوح أو أكبر 1.2 إلى 1.5 مرة من P لاختبار توكى لنفس المقارنات وهذا يدل على أن المقارنات القبلية (المخطط لها) أكثر قوة إحصائياً من المقارنات البعدية (شيفيه - توكى). وفى المقارنات البعدية فإن اختبار توكى أكثر قوة 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 مسبقاً فى فروضه أى المجموعات التى ربما تسهم فى حدوث الدلالة الإحصائية لتحليل التباين.

والاختبار بين شيفيه وتوكى يعتمد على فلسفه البحث، فإذا أراد الباحث الحصول على اختبار أكثر قوة إحصائياً لرفض (H_0) بمعنى الحصول على دلالة إحصائية بين المقارنات مع ضبط معقول للخطأ من النوع الأول فإن اختبار Tukey مفضل عن

Scheffe (2013), Pagano, ويرى (2014) Gravetter & Wallnau بأنه يمكن الحصول على دلالة إحصائية بين أزواج المتوسطات بواسطة استخدام توكي ولكن لا نحصل عليها باستخدام شيفيه، ولكن شيفيه أكثر أماناً لأن لديه وقاية كبيرة من تضخم الخطأ من النوع الأول. ولكن إذا أراد الباحث الذهاب أبعد 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 توكي _ كرامير في حالة عدم تساوى العينات ولكن هذا نادر الاستخدام فى التراث. كما يفضل استخدام طريقة نيومان - كيولز لأنها أكثر قوة إحصائياً من طريقة توكي.

وفيما يلى ملخص لطرق المقارنات المتعددة:

الإجراء	H_0	استخدامه	الإحصائية	التوزيع التحتي
توكي	$M_i = M_k$	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)
توكي - كرامير	$M_i = M_k$	ازواج المقارنات مع عدم تساوى n	$Q = \frac{\bar{X}_i - \bar{X}_k}{\sqrt{MS_w \frac{1/n_i + 1/n_k}{2}}}$	توزيع Q مع df(r, N-K)
شيفيه	$\sum C_k \mu_k = 0$	مقارنات زوجية مع عدم تساوى n ومقارنات معقدة	$F = \frac{(\sum C_k \bar{X}_k)^2}{MS_w \sum (\frac{C_k^2}{n_k})}$	توزيع F مع df (K - 1, N-K)
المقارنات المتعامدة المخطط لها	$\sum C_k \mu_k = 0$	مقارنات متعامدة	$F = \frac{(\sum C_k \bar{X}_k)^2}{MS_w \sum (\frac{C_k^2}{n_k})}$	توزيع F مع df (1, N-K)

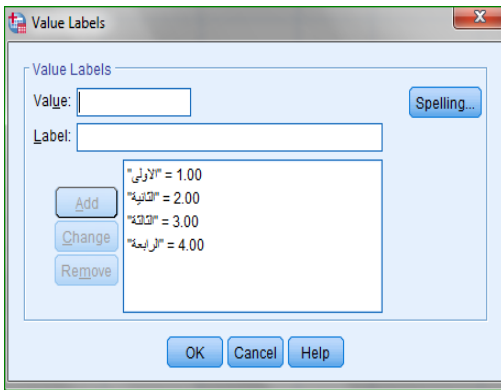
K عدد المجموعات، r عدد المقارنات، N حجم العينة الكلى، n حجم العينة فى لمجموعة الواحدة .

ويبدو ان إجراءات المقارنات البعدية مثل فيشر واختبار Duncan ونيومان - كيولز هي أقل استخدامًا في الدراسات، في حين أن اختبار توكي واختبار شيفيه هي أكثر استخدامًا وأحيانًا يستخدم الباحث مقارنات بعدية ويطلقوا عليها مقارنات مخططة أو العكس صحيح.

تنفيذ اختبار F مقرونًا بالمقارنات المتعددة البعدية في برنامج SPSS

أولاً: إدخال البيانات: 1. اضغط Variable view وحدد مسمى المتغيرات في عمود Name كالاتي: المتغير المستقل: طريقة التنشئة Method ويأخذ الكود الاتي (طريقة التنشئة الأولى =1، الثانية =2، الثالثة=3، الرابعة=4، الخامسة=5)، المتغير التابع التحصيل: achievement

2. في عمود Values أكتب الكود المقابل للمتغير Method كالاتي:

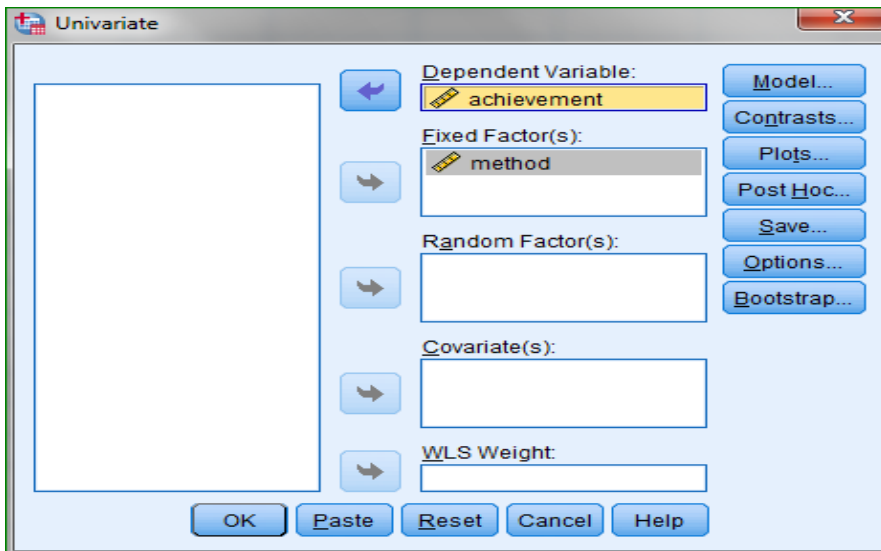


أ- اضغط على الخانة تظهر الشاشة الآتية:

ب - أكتب الكود 1 أمام Value ثم أكتب مسماة أمام Label ثم اضغط Add ثم أكتب 2 ومسماة وأضغط Add وهكذا.

ثانياً: تنفيذ F مقرونًا بالمقارنات البعدية:

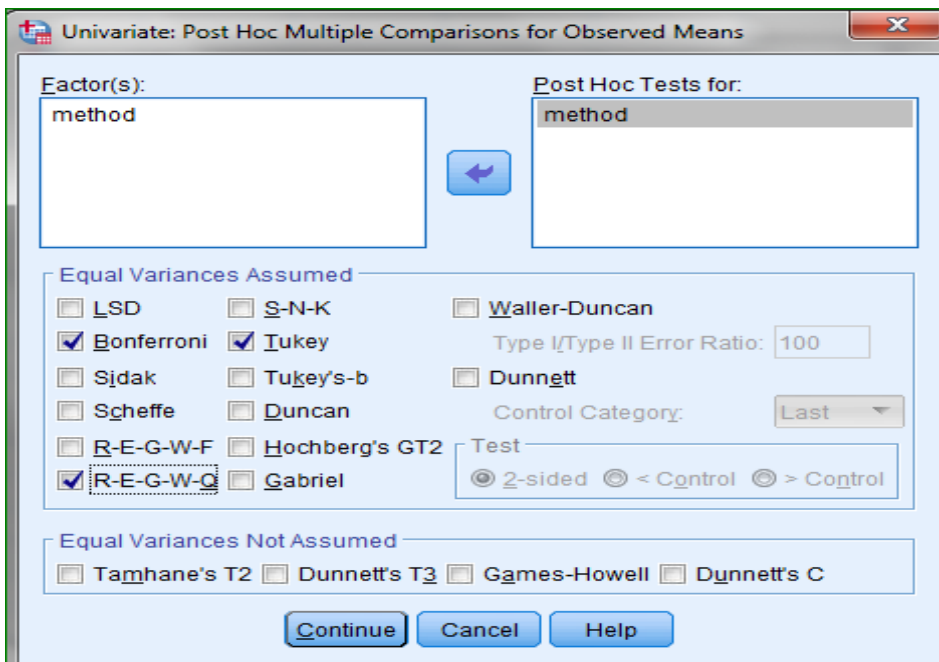
1. اتبع خطوات تنفيذ اختبار F من خلال Analyze→General Linear Model→Univariate



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 R.E.G.W.Q, Tukey, Bonferroni

6. اضغط Continue ثم اضغط Ok

Analyze → Compare Means → One –Way :أما فى حالة تنفيذ الأمر من خلال: ANOVA واتباع باقى الخطوات.

ويوجد العديد من طرق المقارنات المتعددة البعدية كما سبق عرضها ويفضل فى حالة تحقق تساوى البيانات استخدام R-E-G-W-Q, Tukey, وكذلك يفضل Bonferroni لأنها تتحكم فى الخطأ من النوع الأول ولا يفضل استخدام Newman - kauls لأنها لا تضبط تضخم الخطأ من النوع الأول وكذلك لا يفضل استخدام least- significance difference((LSD لأنها لا تتحكم فى الخطأ من النوع الأول أيضاً وهو مكافئ لإجراء اختبار T لكل زوج من المستويات .

ويفضل استخدام GT2,Gabriel و 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-Kramer وطريقة Newman kauls

ثالثاً: المخرج: اعطى جدول التحقق من مسلمة تجانس التباينات كالاتى:

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 كالاتى:

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 $P=0.000$ ، $F=27.418$ ، وعليه 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 كالتالى:

Multiple Comparisons							
Dependent Variable: achievement							
		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
(I) method	(J) method				Lower Bound	Upper Bound	
Tukey HSD	الاولى	القائبة	-6.0000-	2.81688	.231	-14.0987-	2.0987
		القائفة	-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-
	القائبة	الاولى	6.0000	2.81688	.231	-2.0987-	14.0987
		القائفة	-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-
	القائفة	الاولى	7.0000	2.81688	.117	-1.0987-	15.0987
		القائبة	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
		القائبة	13.8750 [*]	2.81688	.000	5.7763	21.9737
		القائفة	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	
	القائبة	19.0000 [*]	2.81688	.000	10.9013	27.0987	
	القائفة	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 كالاتى:

Bonferroni	الأولى	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	الأولى	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	الأولى	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	الأولى	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	الأولى		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	2	
Tukey HSD ^{a, b}	الأولى	8	4.0000	23.8750 29.0000
	الثانية	8	10.0000	
	الثالثة	8	11.0000	
	الرابعة	8		
	5.00	8		
	Sig.		.117	
Ryan-Einot-Gabriel-Welsch Range ^{b, c}	الأولى	8	4.0000	23.8750 29.0000
	الثانية	8	10.0000	
	الثالثة	8	11.0000	
	الرابعة	8		
	5.00	8		
	Sig.		.075	

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 إحصائية بين متوسطات تحصيل المجموعات الخمسة : $P=0.000$ ، $F=27.418$ ، $\eta_p^2=0.758$ ، $0.05 < \alpha$ ، 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 بإستخدام شيفيه حيث يسمح بمقارنة أى متوسط مع باقى الأزواج. ويمكن تنفيذه باتباع حل الخطوات السابقة وفى Post hoc

اضغط scheffe يظهر المخرج الآتي:

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 Sidak ,Dunncan

ويتم تنفيذها بإتباع الخطوات السابقة فى hoc Post اضغط على أى من هذه الاختبارات، وفيما يلي 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 مسبقاً 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 فمثلاً تخيل تصميم تجريبى بثلاث مجموعات تجريبية ومجموعة ضابطة فالمقارنة الأولى تتضمن مقارنة المجموعات التجريبية الثلاثة على حدة بالمجموعة الضابطة وعليه 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

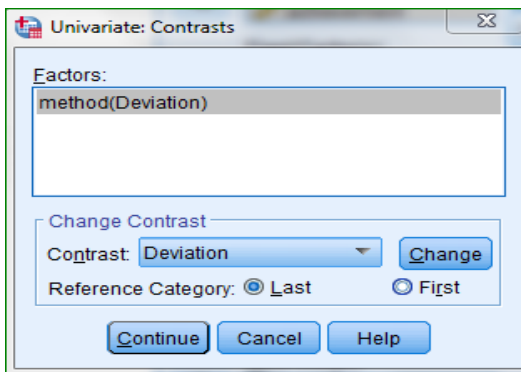
2. اضغط اختيار Contrast تظهر الشاشة

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 لباقي المستويات
Grand Mean

Simple : مقارنة متوسط كل مستوى أو مجموعة بمتوسط مستوى محدد مسبقاً 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 حدة.

Repeated : مقارنة متوسط كل مستوى بالذى يليه مباشرة فمثلاً (1, 2) و (2, 3) و (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: $\text{Contrast Estimate} = -11.575$ ، $\text{Sig} = 0.000$ ، وبما أن $0.05 < 0.000$ ، توجد فروق ذات دلالة إحصائية بين متوسط المجموعة الأولى ومتوسط باقى المجموعات معاً.

وهكذا لمتوسط المجموعة الثانية وباقى المتوسطات معاً حيث $\text{Sig} = 0.004 < 0.05$.

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 Upper Bound -15.192- -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 Upper Bound -9.192- -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 Upper Bound -8.192- -.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 مسبقاً مثل متوسط المجموعة الضابطة. وإذا اعتبرنا أن المجموعة الأولى الضابطة فيتم اختيار بديل Simple وتتبع الخطوات الآتية:

1. اضغط Contrast.
2. في مربع Change Contrast اضغط على السهم واختر بديل Simple.
3. أمام Reference Category اضغط First باعتبار أن المجموعة الضابطة لها كود 1
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 Upper Bound	.281 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 Upper Bound	1.281 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 Upper Bound	14.156 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 for Difference	
	Lower Bound	19.281

لاحظ تمت المقارنة بين المجموعة الأولى وكل متوسطات باقى المجموعات على حدة، حيث توجد فروق بين متوسطات تحصيل المجموعة الأولى والثانية $\text{Sig} = 0.04$ وكذلك بين الأولى والثالثة $\text{Sig} = 0.018$ وهكذا. وإذا كان الباحث يهدف إلى المقارنة بين كل متوسط المجموعة بمتوسط المجموعة التى تسبقها مباشرة فاتباع الخطوات الآتية:

1. فى المربع الذى امام Contrast اضغط على السهم واختار Diffirence ثم اضغط على Change

2. اضغط على Continue ثم اضغط على OK.

		Dependent Variable
method Difference Contrast		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 Upper Bound	.281 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	
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	
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 for Difference	
	Lower Bound	12.260

الفصل الثانى والعشرون

ما وراء التحليل

Meta-Analysis

ماهية ما وراء التحليل

فى عام 1952 اشار Hans Eysenck جدلاً واسعاً فى مجال علم النفس الاكلينكى 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. وللدرد على ادعاء Eysenck قام Smith & Glass (1977) 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 كمياً بما وراء التحليل Meta-analysis واستنتجا ان برامج العلاج النفسى فعالة، ولكن Eysenck (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 Eysenck والآخرين لهذا الأسلوب إلا أنها أصبحت أكثر انتشاراً وقبولاً كطريقة لتلخيص نتائج الدراسات الإمبريقية داخل العلوم الاجتماعية والسلوكية والطبية.

ومنذ ابتدع علماء النفس أسلوب ما وراء التحليل منذ 1970 اصبح أكثر استخداماً 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 منهجي وإحصائي شامل للوصول إلى الاستنتاجات من التراث البحثي وأحد الطرق لإحداث الدمج أو التوليف المنظم للتراث الإمبريقي (الدراسات الأولية) لظاهرة ما بهدف التكامل بين النواتج للوصول إلى تعميمات أو حل التناقضات بين الدراسات. ويُعرف ما وراء التحليل بأنه التحليل الإحصائي لمجموعة كبيرة من نتائج الدراسات المفردة بهدف احداث تكامل لنواتجها (Glass, 1976)، وما وراء التحليل طريقة لإجراء التحليل الإحصائي لنواتج البحوث لكثير من الدراسات المفردة المستقلة لموضوع محدد (Borenstein et al., 2009; Hedges & Olkin, 1985)، أو التوليف الكمي 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 البحثي لظاهرة ما (Card, 2012).

ويطلق على أسلوب ما وراء التحليل بالتوليف الكمي Quantitative synthesis ويستخدم أحياناً كمرادف للتوليف البحثي، ويتضمن أسلوب ما وراء التحليل مجموعة من الأساليب الإحصائية المستخدمة في المراجعة المنظمة للدراسات السابقة لتقليل التحيز كلما امكن. والإجراءات الإحصائية المستخدمة في أسلوب ما وراء التحليل يمكن تطبيقها لأي مجموعة بيانات. ويهدف ما وراء التحليل إلى الوصول إلى استنتاج وحيد وشامل لنتائج دراسات مختلفة لنفس الموضوع باستخدام نفس الطرق، ويساعد في فهم نتائج تجارب التدخل (المنهج التجريبي وشبه التجريبي)، ولا يساعد ما وراء التحليل في الحصول المعرفة التراكمية فقط بل يعطى مؤشر واتجاه واضح عن ماهي الفجوات أو الموضوعات في التراث البحثي التي هي بحاجة لمزيد من الدراسة والبحث في المستقبل.

الحاجة إلى ما وراء التحليل في العلوم السلوكية والاجتماعية

الملاحظ تراكم كم هائل من الدراسات في تراث علم النفس والعلوم الاجتماعية، فالحاجة اليوم ليس لمزيد من الدراسات الإمبريقية التي تتناول ظاهرة ما بل لإعطاء نظرة تقييمية

شاملة للكّم الهائل من البيانات المتراكمة على مدى أكثر من خمسين عامًا بمعنى إعادة 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 وإى عوامل أخرى 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 الإمبريقي من الدراسات.

ومن خلال المقارنة الكمية لنواتج الدراسات المتنوعة فإن ما وراء التحليل يكتشف معرفة جديدة غير واضحة في الدراسات المفردة وأحياناً يجيب على أسئلة لا يمكن حلها من الدراسات المفردة، فالدراسة المفردة غير قادرة على الحكم على مقارنة فعالية Effectiveness برنامج تدريبي سواء للأفراد مرتفعي ومنخفضي القدرة العقلية، ومن خلال مقارنة متوسط إحصاء حجم التأثير Cohen's d عبر المجموعات المختلفة للدراسات فإن ما وراء التحليل يجيب عن هذا السؤال، وكذلك يمكن دراسة المتغيرات المتفاعلة Moderator variables التي أحياناً يصعب دراستها من خلال الدراسات المفردة عن طريق ما وراء التحليل.

وفى الواقع لا توجد دراسة تامة بدون أخطاء، فكل الدراسات تتضمن أخطاء قياس فى كل قياساتها حيث من الصعب القول انها تامة الصديق البنائى خاصة فى العلوم السلوكية والنفسية حيث تتداخل وتتشابك فيها المتغيرات ببعضها البعض بالتالى فإن المفهوم فى حد ذاته ليس خالصاً نقياً (Hunter & Schmidt, 2004)، وبالإضافة إلى خطأ القياس فإن الدراسات تعاني من خطأ المعاينة حيث تتضمن الدراسة مجموعة صغيرة من المجتمع وهذا بدوره يؤدي إلى نتائج متعارضة ويبدو هذا واضحاً فى العلوم السلوكية والطبيعية، وتناول (Hedges 1987) باستخدام طرق ما وراء التحليل لدراسة الاختلاف لنتائج الدراسات فى 13 مجال فى الفيزياء الطبيعية و 13 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 مستقبلاً 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 المجالات النفسية الامريكية التى تتبع APA خاصة Psychological Bulletin فقد زاد استخدامه فى الدراسات المنشورة فى الفترة الاخيرة ويؤكد معظم

محررى المجالات على توظيف هذا الأسلوب فى دراساتهم كمدخل لمراجعة الدراسات السابقة لأن هذا بدوره يحسن جودة الاستنتاجات من التراث البحثى مما ينعكس إيجابياً على بناء نظرية متماسكة وأكثر قدرة على تفسير الظاهرة.

ويرى (1997) Hunt ان ما وراء التحليل أكثر تأثيراً فى ابحاث المجال الطبى من العلوم النفسية والاجتماعية، ويبدو أن فعالية ما وراء التحليل أكثر فى العلوم الطبية نظراً لطبيعة الابحاث التجريبية التى تعتمد على عدد محدود من الأفراد قد يصل إلى فرد واحد من ثم فإن ما وراء التحليل يسعى إلى احداث تكامل بين هذه الدراسات للوصول إلى استنتاج عام.

فاذا كان باحث مهتم بدراسة اثر المتغير المستقل X على المتغير التابع Y وبدأ فى جمع التراث المتاح فى هذا المجال، فالباحث يصنف كل نتائج الدراسات فى ثلاثة تصنيفات، الأول يتضمن الدراسات التى اثبتت وجود علاقة موجبة ودالة إحصائية، والثانى يتضمن تلك الدراسات التى توصلت إلى علاقة سالبة ودالة إحصائية، والثالث يتضمن تلك الدراسات التى توصلت إلى علاقة غير دالة إحصائية، وعليه فالباحث غير قادر للوصول إلى استنتاج من هذه الدراسات المنفصلة، ويرى أن هذا مدعاة لإجراء دراسته وبعد إجراء الدراسة ووصل إلى وجود اثر سالب ويفسر نتائج دراسته بان هذا يتفق مع دراسات ويتعارض مع دراسات اخرى ، ويصدر توصية بإجراء مزيد من الدراسات للوصول إلى تأكيد إلى هذه النتيجة ويظل البحث فى هذا المجال على هذه الوتيرة وهكذا دواليك وهذا لا يحدث تطور فى العلم، وعليه فإن الدراسات المفردة الأولية نادراً ما تحل التناقضات فى العلوم الاجتماعية والنفسية (Ellis, 2010).

ويرى (2004) Hunter & Schmidt ان الحقيقة المؤلمة ان معظم التراث البحثى اظهر نتائج متعارضة فبعضها وجد علاقات دالة إحصائية والبعض غير دالة إحصائية. ووجدت فى التراث النفسى والاجتماعى تقريباً 50-50% (Cohen, 1962; Sedlmeier & Gigerenzer, 1989) بمعنى أن نصف النتائج دالة إحصائياً ونصفها غير دالة إحصائياً، وهذا هو الحال الطبيعى فى معظم العلوم السلوكية والاجتماعية، وعليه فمن الصعب بناء نظريات ومعرفة متراكمة متماسكة، من ثم فإن

ما وراء التحليل يلعب دورًا كبيرًا في عملية البحث في العلوم النفسية والاجتماعية. المراجعة المنظمة الكيفية السردية وما وراء التحليل هما طريقتان متميزتان لتوليف الدراسات السابقة ويستخدمان لتحديد وتحليل وتوليف الأدلة الكمية من الدراسات السابقة.

ولا يمكن استخدامه لدراسات تتضمن موضوعات مختلفة في نفس ما وراء التحليل مثل دراسات تناول برامج لعلاج الاكتئاب ودراسات تناول الفروق بين الجنسين في الوحدة النفسية، فيشار إلى هذا بإشكالية "التفاح والبرتقال" في ما وراء التحليل بمعنى استخدامه لتوليف دراسات لا تتعامل مع نفس المفاهيم أو العلاقات ولكن هذا ما اتبعه Smith & Glass (1977) حيث دمج نواتج دراسات من معالجات مختلفة مثل العلاج النفسى المعرفى، العلاج السلوكى المعرفى، العلاج الجشتالتى وغيرها في دراسة واحدة للحصول على الفعالية العامة لهذه البرامج. وبالمثل ليس من الأفضل تضمين نواتج دراسات من تصميمات بحثية مختلفة التى تتضمن أساليب إحصائية مختلفة حتى لو تعاملت مع نفس الموضوع، فعلى سبيل المثال لا تتضمن دراسات لتأثير معالجات على الاكتئاب باستخدام تصميم المجموعة الضابطة والتجريبية مع دراسات من منهجية بحثية أخرى تتعامل مع الاكتئاب كدراسات الملاحظة أو العلاقة.

وعليه فإن نتائج من تصميمات بحثية مختلفة بمعالجات إحصائية مختلفة من الصعب دمجها في نفس ما وراء التحليل ولكن من الأفضل إجراء ما وراء تحليل منفصل للنواتج التجريبية وللنواتج دراسات العلاقة.

ما وراء التحليل في مقابل التحليل الأولي أو الثانوي

قبل تناول أسلوب ما وراء التحليل لا بد من التفريق بين مصطلحات ما وراء التحليل و التحليل الأولي Primary analysis والتحليل الثانوي Secondary analysis . ويرى Glass (1976) ان التحليل الأولي يحدث عندما يقوم باحث بجمع البيانات من العينة أو الشركات أو أى وحدة معاينة خلال تطبيق ادوات القياس ثم تحليلها للإجابة

على أسئلة بحثه أو دراسته وهذا ما يحدث فى معظم ابحاث الماجستير والدكتوراه وما بعد الدكتوراة.

اما **التحليل الثانوى** يشير إلى اعادة تحليل بيانات تم جمعها مسبقاً للإجابة عن أسئلة بحثية مختلفة أو الاجابة عن على نفس الأسئلة بطريقة منهجية اخرى مثل استخدام أساليب إحصائية اخرى لم تتاح عند إجراء الدراسة الأولية. ودراسات التحليل الثانوى يمكن أن يجريها نفس الباحث أو باحثين اخرين مختلفين إذا كان باستطاعتهم الحصول على البيانات الخام من صاحب الدراسة الأولية.

لاحظ أن التحليل الأولى والثانوى يتطلب الحصول على البيانات الخام الاصلية للدراسة. فى المقابل فإن **ما وراء التحليل** يتطلب التحليل الإحصائى لنتائج (ليست البيانات الخام) أكثر من دراسة (ليست لدراسة واحدة). فى هذا يتضح مظهرين لما وراء التحليل:

أولاً: يتضمن نتائج الدراسات كوحدة التحليل خاصة ان النتائج تكون فى صورة أحجام التأثير، والحصول على أحجام التأثير لا يتطلب الوصول للدرجات الخام حيث يمكن حسابها من البيانات المطبوعة فى الدراسات الأولية أو الثانوية المنشورة مثل المتوسطات والتباينات واحجام العينات ودرجات الحرية.

ثانياً: يعتبر تحليل لنتائج دراسات متعددة وعليه فإن الدراسة المفردة هى وحدة التحليل ويتراوح حجم الدراسات من دراستين كحد أدنى إلى مئات الدراسات فأكثر إذا كان متاح.

يهدف ما وراء التحليل إلى الوصول إلى استنتاج من عينة من الدراسات فى المقابل فإن التحليلات الأولية والثانوية تهدف للوصول إلى استنتاج من عينة من الأفراد، وبالتالي يعتبر ما وراء التحليل شكل من أشكال مراجعة التراث البحثى. وعليه فإن ما وراء التحليل ليس دراسة اولية أو ثانوية لحل قضية معينة أو للإجابة عن سؤال بحثى محدد انما فقط هو أسلوب إحصائى يسعى لتكامل النتائج الدراسات المختلفة لنفس الموضوع وباستطاعته ضبط خطأ المعاينة واى اخطأ اخرى غير مرغوبة لإعطاء أساس متكامل وشامل للاستنتاجات.

حقيقة من الصعب القول بوجود دراسة متقنة تمامًا Perfect study فكل الدراسات تعاني من بعض المحددات المنهجية وتتضمن خطأ 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 الضروري 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 الضروري تحديد مصطلحات وعينة الدراسة (طلاب مرحلة تعليمية أو عينة صعوبات تعلم وغيرها) جيداً 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) Searching literature أو Sampling:** وفى هذا الشأن يبدأ الباحث فى الحصول عن الدراسات الأكثر ارتباطاً ومن المناسب الحصول على عينة من الدراسات ممثلة تمثيلاً جيداً 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 إحصائية 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 وتقييم 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 مدركاً لمحددات ذلك على استنتاجات الدراسة ويجب أن يتميز بالشفافية **Transparency** أن يذكر الأساليب التى استخدمها فى البحث عن الادبيات البحثية. ويرى (2013) Kline 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**

Studies: وهذه المرحلة تتضمن قراءة التراث والوصول إلى استنتاجات عن طبيعة العلاقات أو التأثيرات بين المتغيرات. وتوضيح معايير الاستبعاد **Exclusion** أو التضمن 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 التضمن **Inclusion**.

• **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 وبيتعد 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:

الحقيقة	السائد
نشأ في العلوم الاجتماعية والسلوكية واصبح أكثر استخداماً في العلوم الطبية.	نشأ في ابحاث المجال الطبي
مناسب لمعظم طرائق البحث مثل البحث الارتباطي والسببي المقارن.	مناسب للدراسات التجريبية
يمكن أن يتضمن دراسات من تصميمات عينات غير عشوائية (منهج شبه تجريبي)	يجب أن يتضمن دراسات
مثل تصميم قياس قبلي - بعدى.	من تصميمات تجريبية تتضمن مجموعة ضابطة.
يمكن استخدامه لدراستين أو ثلاثة فقط.	
يمكن استخدامه لدراسات ذات أحجام عينات صغيرة وتستخدم اختبارات لتصحيح من تحيز العينة الصغيرة.	يتطلب دراسات كثيرة.
ما وراء التحليل لا يحسن من جودة الدراسات الاصلية.	يتطلب دراسات ذات عينات كبيرة.
	يتغلب على مشاكل الجودة (الصدق)
	للدراسات الاصلية

مميزات ما وراء التحليل

يملك ما وراء التحليل مميزات ومصادر قوة عديدة كالاتى (Borenstein et al., 2009; leandro, 2005; Littlel et al., 2008; Lisepy & Wilson, 2002)

- **الموضوعية:** ما وراء التحليل يتغلب على محددات التلخيص الكيفي التقليدي للدراسات السابقة ويمدنا بطريقة فعالة لتلخيص نتائج كم هائل من الدراسات والكشف عن علاقات لا يمكن ادراكها من الدراسات المفردة ويمدنا بشفافية غير متاحة في 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 تعاني 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 العلاقة أو التأثير ولذلك فإن طريقة العد ربما تخبرنا عن دلالة إحصائية موجبة أو سالبة ولكنها حقيقة ليس لها دلالة عملية. كما أن دراسات العينات الصغيرة غالباً تعاني من قوة إحصائية منخفضة للكشف عن التأثيرات الدالة إحصائياً بمعنى تعطى نتائج غير دالة إحصائياً بينما هي في طبيعتها دالة عملياً أو اكلنيكياً. وباستخدام ما وراء التحليل يمدنا بتقديرات عامة أكثر دقة نتيجة دمج الدراسات ذات أحجام العينات الصغيرة والكبيرة في دراسة واحدة وهذا من شأنه 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 بانه يتعامل مع مؤشر حجم التأثير من كل دراسة وليس الدلالة الإحصائية أو قيمة P لتحديد ما إذا كان حجم التأثير متنسق عبر الدراسات وكذلك تضمنين كل أحجام التأثير لكل الدراسات في توليف إحصائي واحد وهذا في غاية الأهمية لحساب حجم تأثير تلخيصي، على ذلك فإن ما وراء التحليل يذهب أبعد من لدلالة الإحصائية ويسمح ما وراء التحليل بتقدير مدى الاختلاف بين التأثيرات ويميز بين الاختلاف الحقيقي والاختلاف الزائف.

• **وراء التحليل إجراء مفيد في بناء معرفة علمية صادقة قائم على عملية منظمة**
لتلخيص النواتج البحثية تشمل معايير 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 إليها 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

معاً عبر الدراسات المختلفة يعطى تقديرات لحجم التأثير أكثر قوة إحصائياً من الدراسات المفردة.

○ **الحصول على تقديرات معالم أفضل Parameter Estimates**: نتيجة لخطأ المعاينة يمكن لأي دراسة ان تنتج تقدير غير دقيق للتأثير خاصة فى المجتمعات الكبيرة، فالمقاييس متعددة المفردات أكثر ثباتاً 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 المسؤولة عن الاختلاف بين النتائج أو أحجام التأثير مثل خصائص العينة (عمرها، جنسها)، نوعية المعالجة، مكان إجراء الدراسة، ونوعية القياسات (ملاحظة، تقرير ذاتي)، جودة تصميم البحث، مصدر تمويل الدراسة (خاص، حكومي)، 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 وقابلية لتعميمها إلى مواقف أخرى، لو ان الدراسات غير متسقة بين المؤسسات أو المواقف فلا بد من فحص وكشف مصادر الاختلاف. وتحليل التفاعل يمكن تحديد البرامج أو المعالجات أكثر نجاحاً. 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 (Little 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 الكمي 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 تقييم 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) وآخرون يلقى العديد من الانتقادات أو المحددات معظمها مرتبط بالتطبيقات غير المنطقية وسوء استخدام الأسلوب (Littell et al., 2008). ومن أهم الانتقادات الموجهة لهذا الأسلوب (Borenstien et al., 2009; Card, 2012; Littell et al., 2008; Rosenthal & DiMatteo, 2001; Lisepey & Wilson, 2002)، كالتالى:

- يحتاج إلى جهد وخبرة عالية لفهمه وتنفيذه من جمع الدراسات وتكويدها تقدير أحجام التأثير المناسبة وتحليلها إحصائياً وتفسيرها وذلك مقارنة بالمراجعة الكيفية السردية، وعليه يتطلب جهد إحصائى عالى لتنفيذه والامام بفنياته ولكن كأي أسلوب يبدو صعب عندما تقرأه أول مرة، ولكن معظم تطبيقات ما وراء التحليل لا تتطلب تحليلات معقدة، وأيضاً كتابة تقرير دراسة ما وراء التحليل تحتاج إلى مهارة عالية مقارنة بتقرير الدراسات الأولية

- بدون اتباع القواعد المتبعة لأسلوب ما وراء التحليل من جمع الدراسات المرتبطة وتكويدها وتحليلها تصبح عبارة عن تدريب إحصائي لا معنى له.
- **التحديد غير المناسب للمشكلة:** عند إجراء ما وراء التحليل لا بد تحديد نوعية الدراسات أو المعالجات المتضمنة فى التحليل وغياب المعنى المنطقى من تضمين الدراسات المختلفة التى تتضمن معالجات وعينات وتصميمات بحثية ونواتج مختلفة يعطى نتائج من الصعب تفسيرها ويعتبر هذا من الانتقادات الكلاسيكية لما وراء التحليل.
- يشبه البعض خلط دراسات مختلفة التصميمات والعينات كالخلط بين **التفاح والبرتقال** حيث تتباين الدراسات فى الجودة المنهجية والمقاييس والعينات وغيرها من المظاهر المنهجية، ولكن محلل أو مستخدم الأسلوب يعتبر ان هذا الخلطة مناسبة عندما يكون البعض مهتم بعمل سلطة فواكه، فمن غير المناسب ان يكون الباحث مهتم بالتفاح فقط أو يريد ان يميز التفاح عن البرتقال ولا بد من وجود منطقية نظرية قوية لإجراء ما وراء التحليل لإعطاء نتائج ذات معنى ويمكن التغلب على ذلك بدراسة العوامل المسؤولة عن تفسير التباين بين الدراسات المختلفة.
- **تقدير غير مناسب لجودة الدراسة:** ما وراء التحليل لدراسات منخفضة الجودة لها ثبات منخفض وتصميم بحثى ضعيف ومشاكل إجرائية ينتج عنها نتائج غير صادقة. ويطلق على هذه المشكلة " زبالة دخلت، زبالة خرجت". وللتغلب على هذه المشكلة لا بد من وضع معايير لانتقاء الدراسات فى ضوء تصميم البحث الجيد ونوعية المعالجات التى تتناول المتغيرات المستقلة لتضمين الدراسات فى التحليل ويسعى ما وراء التحليل ما إذا كان التنوع المنهجى يؤثر على النتائج من خلال تحليل المتفاعلات.
- **انخفاض الصدق الخارجى External validity :** كأى نوع من الدراسات فإن نتائج ما وراء التحليل يمكن تعميمها فقط إلى المجتمعات والمعالجات والمواقف التى تشبه المواقف التى تمت فى التجربة، فالمعالجات التجريبية التى تتم فى ضوابط منهجية صارمة وعينات متجانسة والتركيز على متغير مستقل وحيد فى الغالب وتدريب عالى الجودة للمعالج أو المتدرب، فالنتائج المتحصل عليها من هذه النوعية

من الدراسات من الصعب تعميمها في المواقف الطبيعية سواء في المدرسة أو الجامعة أو المستشفى أو غيرها. وعليه فنتائج ما وراء التحليل لهذه الدراسات تعاني من نفس المحدد وهو انخفاض القوة التعميمية في المواقف الطبيعية.

- **اعتماده على أحجام التأثير Effect sizes:** لدمج الدراسات كمياً فمن الضروري الحصول على قياسات تكون قابلة للمقارنة عبر الدراسات، ولذلك يحصل الباحث على مجموعة من أحجام التأثير لتحديد قوة أو حجم العلاقات بين المتغيرات. وفي ما وراء التحليل يحسب حجم التأثير لكل دراسة على حدة ثم تُوزن بحجم العينة ومتوسطها لتمدنا بالآثر العام Overall effect . وفي 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 لكي تكون مفهومة لدى الاكاديميين وصانعي 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 الذي يحاول ان يدمج بين فنيات المراجعة الكيفية والكمية. فالبحت الكيفي على درجة كبيرة من الأهمية حيث يمدنا برؤى وبمعلومات ثرية تسهم في فهم الظاهرة مقارنة بالبحث الكمي الإمبريقي. والتقدم العلمى يكون محدداً لو اعتمد فقط على البيانات الكمية المتحصل عليها من طرائق البحث الكمية.

- **المعاينة غير المناسبة وطرق جمع البيانات:** لا يعطى ما وراء التحليل انتباه كافي لمصادر الخطأ المحتمل والتحيز في معاينة (عينة الدراسات) ومعايير تقييم الدراسات

المدمجة، وعدم الانتباه لثبات وصدق البيانات المستخلصة للتحليل من الدراسات. ويعتمد تحليلات ما وراء التحليل في 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** : دائماً يتم نشر النتائج الدالة إحصائياً مقارنة بالدراسات ذات النتائج غير الدالة إحصائياً (الصفيرية) وهذا يعرف بتحيز تقرير النتائج عليه فإن الباحث يعتمد على حساب حجم التأثير للنتائج الدالة إحصائياً التي تكون متاحة للباحث وهذا بدوره يؤدي إلى استنتاجات مربكة ومتحيزة لأحجام التأثير، وهذا مرتبط بتحيز النشر حيث الدراسات الدالة نتائجها إحصائياً أكثر مقبولة للنشر في الدوريات العلمية مقارنة بنظيرتها غير الدالة إحصائياً بالتالي لا نستطيع الوصول إلى استنتاجات صادقة لأنه يعتمد على الدراسات المنشورة الدالة إحصائياً.

• **التحيز في انتقاء الدراسات التي تعتمد على لغة الباحث** بينما الدراسات في البلدان الأخرى بلغات أخرى تكون غير متاحة للباحث وكذلك مدى الإتاحة Availability حيث بعض الدراسات تحتاج إلى تكلفة عالية للحصول عليها وهذا غير متاح للباحث. فالباحث العربي لديه إتاحة ومرغوبة 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 إحصائياً ويتجاهل النتائج غير الدالة إحصائياً بالتالي يعطى نتائج متحيزة، ولكن المؤيدون يرون ان الأسلوب يتضمن العديد من الفنيات لتشخيص هذه المشكلة وهي تحيز الدراسات ومحاولة معالجتها. وهذه المشكلة ليست قاصرة فقط على ما وراء التحليل انما أيضاً قائمة في المراجعات الكيفية للدراسات السابقة.

• **مصادقية استخلاص البيانات Data extraction**: كثيراً 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 خطأ فى استخلاص بيانات 10 دراسات (37%) من 27 دراسة لما وراء التحليل، وهذه بدوره لا يعطى نتائج دقيقة بدون تجنب خطأ 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 وتكوين المعلومات من هذه الدراسات وتحليل نتائج البيانات. وتحديد المشكلة يتطلب تعريف وتحديد المتغيرات وطبيعة العلاقة بينهما وكيفية اختبارها إمبيريقياً، والمتغيرات فى العلوم الاجتماعية يجب تحديدها بطريقتين الأولى التعريف المفاهيمى أو النظرى وهى تحدد صفات المتغير وخصائصه، فمثلاً التعريف النظرى للذكاء هو القدرة على التعلم، الاستدلال، الفهم، وأى نشاط عقلى آخر،

فالتعريف النظرى شامل وعريض بينما التعريف الإجرائى أكثر تحديداً ويكون قابلاً للقياس ووصف لخصائص للأحداث الملاحظة لتحديد مدى وجود الظاهرة أو المفهوم.

ويرى (Lipsey & Wilson 2001) يجب أن تكون صياغة المشكلة واضحة ومباشرة وأكثر تفصيلاً مثل: هل البرامج القائمة على الوسائط المتعددة فعالة فى تحسين التحصيل؟، هل البرامج القائمة على استراتيجيات التعلم المنظم ذاتياً لها تأثيرات فعالة فى تنمية تقدير الذات أو أى نواتج أخرى مثل التحصيل أو الاتجاه أو التفكير الناقد وغيرها من النواتج الانفعالية والمعرفية والسلوكية؟

لاحظ أن صياغة المشكلة وتحديدها يساعد فى تحديد طبيعة المتغيرات المستقلة والتابعة، فالمتغير المستقل موضع الاهتمام والدراسة هو خصائص البرنامج بينما المتغيرات التابعة تقدير الذات، التحصيل، الاتجاه، التفكير الناقد. وهذا التحديد يعتبر بمثابة المرشد لتحديد الكلمات المفتاحية وبدورها المساعدة فى البحث عن التراث البحثى والدراسات السابقة.

ودائماً تصاغ عناوين ما وراء التحليل لدراسة تأثيرات المعالجات كالاتى:

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)

مثال: أثر العلاج السلوكى مقارنة بالعلاج المعرفى فى تخفيف أعراض الاكتئاب لدى نزلاء السجون.

تحديد أهداف وأسئلة أو موضوع البحث المراد الإجابة فى ما وراء التحليل وتتضمن عدة خطوات أهمها:

الموضوع والمجال Topic and scope

يرى (Littell et al. 2008) أن الخطوة الأولى لإجراء توليف البحوث هي تحديد الأسئلة الجوهرية التي سوف تكون بمثابة مرشد للباحث لجمع التراث البحثي ومن الضروري تحديد نطاق ومجال والمسلمات الأساسية لإجراء دراسات ما وراء التحليل.

والأسئلة والفروض الجيدة تأتي من مصادر عديدة منها الممارسة البحثية، الخبرة، والمراجعة المستمرة لتراث ظاهرة ما. والسؤال البحثي يجب أن يكون قابلاً للبحث، محدد بوضوح، ممتع، جديد، أخلاقي، ومناسب. فمثلاً هل البرنامج فعال؟ فهذا ليس سؤال محدد لأن التجربة أو البرنامج ليس فعالاً لكل الفئات بمعنى له شروط وظروف معينة، ولكن الصيغة الأفضل للأسئلة عن تأثير برنامج أو تدخل ما هو: هل المعالجة X_1 أكثر فعالية من المعالجة X_2 على المتغير التابع Y لمجتمع ما؟ وهذا من شأنه تحديد ما هي المعلومات التي يحاول الباحث الحصول عليها ومنها تحديد المجتمع والمعالجات والمواقف والنواتج، كذلك تحديد المكونات المستبعدة من الدراسات السابقة.

فمثلاً أحد المراجعات يهدف إلى تقدير أثر البرامج المدرسية على مدى واسع من النواتج مثل كفاءة الذات، العلاقات بين الأقران، التحصيل الأكاديمي، وهكذا. ومراجعة ما وراء تحليل أخرى تركز على أثر هذه البرامج المدرسية لدى التلاميذ منخفضي الدخل، ومراجعة ثالثة تركز على أثر البرامج المدرسية على مدى محدود من النواتج أو على ناتج واحد مثل التحصيل الأكاديمي، بالتالي هذه الأسئلة تعكس أطر مختلفة للبحث في التراث.

وكلما كان الباحث أكثر وضوحاً في تحديد الموضوع المراد دراسته في ما وراء التحليل كلما كانت الدراسات المجمعة من التراث البحثي أكثرًا تجانساً ولكن عدم التحديد الدقيق للموضوع المراد دراسته يؤدي إلى اتساع دائرة البحث في التراث ويحصل على دراسات غير متجانسة، وعليه تظهر مشكلة التفاح والبرتقال في ما وراء التحليل ولكن البعض يرى أن الهدف العام هو البحث في سلة الفواكه ومحاولة إيجاد العلاقات بينهم.

ومن الأفضل أن تكون القضية المراد بحثها محددة المفاهيم والأسباب والعواقب وأن تكون مستندة لنظرية أو نموذج لأن هذا يساعد الباحث في 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 الضروري 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 ويدورها تقلل التحيز الشخصى Subjectivity الذى يحدث لو لم تتوفر هذه المعايير، وكذلك الابتعاد عن تحيز الانتقاء للدراسات كلما أمكن. وإذا لم تحدد هذه المعايير مسبقاً فإن تضمين الدراسات يعتمد على التفضيل الشخصى والإتاحة للدراسات وعوامل أخرى وهذا بدوره يبتعد عن الموضوعية.

- تحدد طبيعة المجتمع المراد دراسته، فإذا 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 المعايير سوف يصل إلى نفس الـ 60 دراسة.

ويحدد (Lipsey & Wilson 2001) معايير صلاحية الدراسات أو معايير استبعاد أو تضمين الدراسات فى ما وراء التحليل بالآتى: الملامح المميزة لأهلية الدراسة، المستجيبين فى الدراسة، المتغيرات الأساسية، تصميمات البحث، المدى اللغوى والثقافى، زمن الدراسة، نوع النشر. ويحددها (Littell et al. 2008) بالآتى: التصميمات البحثية، المجتمعات، التدخلات أو البرامج، المقارنات، والنواتج المقاسة المراد تضمينها أو استبعادها. بينما يحددها (Card 2012) بالآتى: 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:

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 (Y, X) 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 والتسويق والتكؤ 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. فإذا 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. فمثلاً 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 أخرى.

• **المتغيرات المفتاحية أو النواتج** Key Variables or Outcomes : ماهى

المتغيرات الأساسية المتضمنة فى الموضوع أو المشكلة المراد دراستها؟ وفى 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 مصداقيتها وثباتها. وايضاً توقيت قياس النواتج فالاعتماد على القياسات البعيدة لا يخرنا عن مدى حدث استمرارية للتحسن ولذلك يهتم بالقياسات البعيدة 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. وفى دراسات التدخل لا بد من استخدام دراسات تجريبية يتم توزيع أفراد العينة عشوائياً 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

القبلي والبعدي ويمكن أن يستبعد الدراسات التي لم يتم التوزيع فيها عشوائى على المجموعة التجريبية والمجموعة الضابطة أو يمكن الاعتماد عليها لو استخدمت قياسات قبلية للمتغير التابع أو المتغيرات الأكثر ارتباطاً به. ويمكن الاعتماد على الدراسات ذات الجودة المنهجية العالية (توزيع عشوائى للمجموعات) أو الدراسات شبه التجريبية ذات القياسات القبليّة للمجموعة الضابطة والتجريبية للمتغيرات الأكثر ارتباطاً 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 التي 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 التضمن للدراسات التي 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 دوراً كبيراً فى اختلاف النتائج وتقود إلى استنتاجات خاطئة.

ودائماً يتم استبعاد دراسات المجموعة الواحدة قياس قبلي وبعدي 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 منهجياً 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 البحثي المستخدم فى الدراسات المتضمنة فى تحليل ما وراء فهل يعتمد على دراسات تجريبية فقط؟ أم دراسات شبه 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 نظراً لاختلاف المفاهيم من ثقافة إلى أخرى. فمثلاً معنى مفاهيم مثل السرقة أو الاحتيال أو السعادة فى البيئة العربية مختلف عن معناها فى البيئات الأخرى، وهذا بدوره يؤدي 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 مثلاً 1970 فهذا يمثل صعوبة كبيرة على الباحث من حيث الوقت

والتكلفة. يمكن للباحث على الدراسات فى المدة الزمنية من مثلاً 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 يختلف عبر الزمن وهذا يتطلب تضمين الدراسات التى أجريت خلال كل الفترات الزمنية وايضاً 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؟ فمثلاً 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 لمصادقية المراجعة للتراث البحثى.

وكلما كان الباحث أكثر مرونة فى تبني معايير منهجية فإن هذا يؤدى إلى نتائج غير دقيقة ولكن وضع معايير منهجية صارمة يؤدى أيضاً إلى الحصول على عينة صغيرة من الدراسات ويمكن استبعاد دراسات مفيدة من التحليل وهذا من شأنه يؤدى إلى الحصول على عينة من الدراسات غير ممثلة تمثيلاً جيداً لمجتمع الدراسات وبالتالي يحد من القدرة التعميمية للنتائج. وأجريت 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 تمدّه بالمعلومات الكافية عن المعايير

المنهجية التى يتبناها. والجودة المنهجية ليست شئ بسيط انما تتضمن عوامل عديدة العينة، العشوائية، الصدق، الثبات، التصميم، التكافؤ، وغيرها. فالمعايير المفضلة لدراسة لموضوع ما فقد تختلف عن دراسة موضوع آخر ولكن المعايير التى لا خلاف عليها هى قضايا ثبات البيانات والتصميم التجريبي المستخدم.

وبعض الباحثون وضعوا اطر وقوائم لتقدير الجودة المنهجية ولكن لا يوجد اتفاق بين الباحثين حول مكونات هذه الاطر وهى تتضمن الخصائص المنهجية الجيدة التى يجب أن تتسم بها الدراسات. وتوجد علاقة تبادلية بين قيود معايير التضمين للدراسات المتضمنة فى ما وراء التحليل والملاحح المنهجية للدراسات، فكلما زاد تنوع واتساع هذه المعايير كلما حصل الباحث على مدى واسع لدراسات ذات ملاحح منهجية وإجرائية متنوعة.

عمومًا معايير التضمين والاستبعاد المرنة والصارمة لها مميزات أهمها (Card, 2012):

- تضمين أكبر عدد من الدراسات الممكنة واستبعاد عدد محدود بينما المعايير الصارمة تستبعد عدد كبير من الدراسات وتضمن عدد قليل. والاختيارين يعكسوا نقطتى النهاية على متصل. والمعيار الاهم للتضمين هو مجتمع الدراسات التى بصدد الوصول إلى استنتاجات بشأنه، وكذلك معيار الجودة المنهجية فهل الباحث على استعداد لتضمين دراسات تعاني من خلل منهجى ما وهذا من شأنه يزيد من انتقاد أسلوب ما وراء التحليل بمعنى " زباله دخلت التحليل، زباله خرجت من التحليل" ولكن المعايير الصارمة تجعل استنتاجات النتائج لمجتمع محدود.
- المعايير المرنة تجعل الدراسات أكثر تنوعًا فى ملاححها المنهجية بينما المعايير الصارمة ينتج عنها عدد محدود من الدراسات ومتشابهة فى ملاححها وهذا يؤدى إلى قوة إحصائية غير مناسبة لتقويم متوسط حجم التأثير.
- **معلومات حجم التأثير** Effect Size Information: فى النهاية من المعايير الهامة توافر البيانات الكافية لحساب حجم التأثير، وفى معظم الدراسات تتوافر هذه

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 التي يعتمد عليها باحث ما وراء التحليل حتى يسمح لمراجع آخر يكود النواتج البحثية المتطلبة لإنجاز دراسة ما وراء التحليل.

ويتضمن البروتكول الاتي:

- غلاف يتضمن عنوان دراسة ما وراء التحليل، تفاصيل الاقتباس، والتواصل مع المؤلفين.
- مقدمة (الخلفية واهداف الدراسة).
- المنهجية وتتضمن معايير التضمن 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 أشكالاً عديدة مثل:

- الفروق بين متوسطات المجموعات (الفروق بين الذكور والإناث في الدافعية).
- العلاقات بين المتغيرات (العلاقة بين القلق والتحصيل).
- نسب لملاحظات أو ظواهر معينة في المجتمع (نسبة مدمنى الفيس بوك أو نسبة انتشار الجريمة في المجتمع).

تتعدد النواتج المراد تجميعها من الدراسات أهمها:

• **حجم التأثير:** تجمع نتائج الدراسات لاستخدامها في ما وراء التحليل في صورة حجم التأثير ولمقارنة النتائج بصورة لها معنى لا بد أن يستخدم نفس حجم التأثير لتأكيد كل النتائج حتى تكون قابلة لاستخدامها في ما وراء التحليل، وعليه فلا بد من تحديد شكل النواتج البحثية المناسبة لما وراء التحليل وتعرض في 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 أخرى مثل تحويل مؤشر d إلى مؤشر معامل الارتباط r .

على ذلك فالمكون العام لنوع التحليل سواء دمج أو مقارنة هو أحجام التأثيرات حيث تعكس ماهية أسلوب ما وراء التحليل. وتتنوع مؤشرات حجم التأثير الأكثر استخداماً في 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 إلا أنها في معظمها تعكس العلاقة بين متغيرين، والعلاقة المتحصل عليها من الدراسات الارتباطية تعكس درجة التلازم أو التغاير بين متغيرين ويمكن أن تكون مقياس تقريبي للعلاقات السببية Causal relations ولكن العلاقات من الدراسات التجريبية فهي مقياس حقيقي للسببية على الرغم من مهددات الصدق الداخلي في التصميمات التجريبية.

وأي من أنواع العلاقات هي مقياس لحجم التأثير المستخدم في ما وراء التحليل، ويوجد العديد من مؤشرات حجم التأثير متضمنة المتوسطات، النسب، ثبات المقاييس، درجات التغير الطولي (الدراسات الطولية) ولكنها أقل استخداماً في ما وراء التحليل. وأحجام التأثير المشتقة من دراسات ارتباطية، طولية، شبة تجريبية، تجريبية تعكس مدى من الأسئلة البحثية يمكن الإجابة 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 في ضوء مؤشرات النزعة المركزية مثل المتوسط، الوسيط، المنوال، والنسب.

وإذا توافرت هذه المؤشرات لنواتج الدراسات فيمكن إجراء ما وراء التحليل للدراسات. لاحظ أن هذه المؤشرات تكون لدراسات متنوعة تناولت نفس المتغير وغالبًا تستخدم نفس المقياس كمتوسط الذكاء كما قيس في اختبار ستانفورد- بينيه. ويمكن حساب مؤشرات النزعة المركزية لدراسات استخدمت مقاييس مختلفة لنفس المفهوم شرط تحديد التعريف الإجرائي للمتغير عبر كل الدراسات.

- **مقارنات المجموعات Group contrasts:** هذا النوع من نواتج البحوث يتضمن متغير واحد فأكثر تم قياسه لمجموعتين فأكثر من الأفراد ثم تتم المقارنة بين المجموعات ويتم عرض المتوسطات أو النسب لكل مجموعة. وفي هذا الإطار يوجد شكلين لمقارنة المجموعات وهي موضع اهتمام ما وراء التحليل:

- البحوث التجريبية أو المحاولات الاكلينيكية Experimental or clinical

trials: حيث تتم المقارنة بين مجموعة تعرضت لمعالجة ما تجريبية (مجموعة تجريبية) بمجموعة لم تتلقى معالجة أو معالجة تقليدية (مجموعة ضابطة) حيث تكون المقارنة على قيم المتغير الناتج (التابع) وتفسر على انها تأثير المعالجة.

- بحوث الفروق بين المجموعات Group differences research: تتم المقارنة

في ضوء خصائص المجموعة مثل المقارنة بين الذكور والإناث في التحصيل أو أي متغيرات أخرى تابعة أو المقارنة بين طلاب ذي صعوبات التعلم ومنخفضي التحصيل بأقرانهم العاديين في القدرة القرائية، وهذه النوعية من الدراسات تتبع من دراسات **المنهج السببي المقارن Causal comparative method**.

وهذه النوعية من الدراسات سواء التجريبية أو مقارنة المجموعات هي شائعة في التراث البحثي لما وراء التحليل وخاصة البحوث التجريبية التي تهدف إلى تقدير فعالية المعالجة أو التدخل (Lipsey & Wilson, 2001; Rosenthal, 1994). وعندما تتم المقارنة بين ثلاث مجموعات فأكثر فيتم دراسة كل مجموعتين معًا على حدة نظرًا لأن الفنيات المستخدمة للمجموعات الثلاثة معًا غير متاحة (Lipsey & Wilson, 2001).

- **العلاقة بين المتغيرات Association between variables**: هذا النوع من نواتج البحوث يعكس التغيرات 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 في المرحلة الثانوية، وتعتبر مؤشرات كأي تربيع أو معامل فأى أو لمدا λ أو نسبة Odds-ratio مؤشرات حجم التأثير، وفي 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 أدوات القياس مثل حساب معامل الارتباط بين درجات الاختبار واعادته Test-retest كمؤشر لمعامل الثبات، أو استخدام معامل الارتباط لتقدير الصدق التنبؤى Predictive validity بين درجات اختبار القدرات والتحصيل اللاحق في الجامعة ويستخدم للتحقق من صدق اختبار الاستعدادات أو القدرات.

- **بحوث الفروق الفردية Individual differences research**: وهذه النوعية من بحوث الارتباط هي الأكثر استخداماً وهي تهتم بدراسة التغيرات بين سمات مختارة للأفراد كالعلاقة بين ادمان الانترنت والوحدة النفسية، أو العلاقة بين التنشئة الاسرية والعنف المدرسى، أو العلاقة بين وقت المذاكرة درجات الطلاب في الرياضيات وغيرها.

وتعتبر دراسات ما وراء التحليل لأبحاث العلاقات أو الارتباطات شائعة الاستخدام في التراث وذلك لتوافر العديد من مؤشرات أحجام التأثير المناسبة (Hunter & Schmidt, 2004; Rosenthal, 1994)، وتعتبر قيمة معامل الارتباط مؤشراً 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 هي 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 غالباً 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 محل 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:** متضمناً 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 High school or Program or intervention أو 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.

فمثلاً عنوان دراسة ما وراء التحليل: أثر برنامج مستند على عادات العقل لتنمية التفكير المنظومى فإن الكلمات المفتاحية: عادات العقل والتفكير المنظومى وأحياناً كلمة برنامج.

وإذا كنت بصدد البحث فى قواعد البيانات الاجنبية فراعى الكلمات المترادفة مثل مفهوم المراهقة adolescence وقد تستخدم بمعنى Teen وعليه يتم البحث كالاتى (or Teen * adolescence) وكلمة البرنامج يمكن أن نبعث 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 كالاتى:

(Adolescence* or teen) And depression* and (Program* or service* or treatment* or intervention* or prevent*)

وهذا البحث أكثر تحديداً لأنه يمدنا بالدراسات التى تتضمن المراهقة والاكتئاب والبرنامج معاً.

أو "Adolescence, depression, and intervention"

بالتالى يسترجع الدراسات التى تضمنت هذه الكلمات معاً.

فكلمة * أو or ؟ تستخدم للبحث باستخدام كلمات مختلفة لها نفس المعنى فى التراث البحثى.

وكلمة "و" and تستخدم للبحث للعلاقات بين مفهومي مثل عادات العقل ودافعية الإنجاز، يمدنا بالدراسات التي تتضمن العلاقة بينهما.

معظم برامج البحث تتضمن كلمة لا "Not" أى استبعاد الدراسات التي تتضمن كلمة أو مصطلح معين. ولتحديد الدراسات المستبعدة يمكن استخدام كلمة Not. ولمحدودية مجال عملية البحث يمكن وضع العينة موضع اهتمام الدراسة ككلمة مفتاحية.

فيما يلي أمثلة لكلمات مفتاحية:

- **نوع العينة:** المراهقون، المراهقة، الذكور، الإناث، أطفال الأحداث، طلاب المرحلة الثانوية أو المدارس العليا، تلاميذ المرحلة الابتدائية، المعاقون سمعياً..... إلخ.
- **نوع التدخل Type of intervention أو البرنامج:** برنامج تدريبي، برنامج علاجي، برنامج إرشادي، برنامج تحدى، إلخ
- **نوع البحث Type of Research:** تقويم، تأثيرات، فعالية، تجربة، تجريبي مجموعة ضابطة، ارتباطي، إلخ

استراتيجية البحث الخلفي Backward Searches

بعد جمع وتراكم مجموعة كبيرة من الدراسات لتطبيق معايير 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 التي تتضمن معظم المعلومات المراد تكويدها وكذلك يمكن الاستشهاد ببعض الدراسات المذكورة فى الدراسات المراجعة والمناسبة للموضوع المراد درسته وايضاً من الضروري قراءة دراسات ما وراء التحليل السابقة التي اجريت فى نفس الموضوع ان وجدت أو موضوعات قريبة لموضوع بحثك. وعملية البحث فى الدراسات المناسبة التي وردت فى الدراسات المراجعة تسمى البحث الخلفي وهذا المدخل مفيد فى تحديد الدراسات القديمة ولكن يوجد عيب فى هذا المدخل هو احتمال التحيز حيث إن الدراسات التي تعطى نتائج مرغوبة (دالة إحصائية) أكثر احتمالية لتضمينها فى الدراسة

المجموعة 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 احياناً 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 قاعدة 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 غالباً 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).

مصادقية 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 تمثيلاً جيداً 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 لتقييم 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. أما 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 مصادقية عملية البحث:

- إجراء البحث فى التراث البحثى باستخدام عدة طرائق للبحث مثل قواعد البيانات المختلفة والمجلات 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 أقرروا بمدى

كفايتها ابدأ فى إجراء ما وراء التحليل ولو أشاروا إلى غياب بعض الدراسات فيجب إعادة استراتيجيات البحث مرة أخرى.

- 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 الإضافية، وإذا لم تحصل على المزيد من الدراسات فابدأ بإجراء استراتيجية البحث الأمامى فى الدراسات التى أقيمت بعد تحقيقها معايير التضمن والاستبعاد على ان تبدأ بالدراسات القديمة أولاً 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 الإضافية وإذا شعرت أنك لا تحصل على مزيد من الدراسات بنهاية عملية التأكيد فإنه يعطى درجة من الثقة ان عملية البحث كانت تقريباً شاملة وممثلة للتراث. وعلى الرغم من هذه الثقة فى شمولية عملية البحث فى الادبيات البحثية فإن Card (2012) ينصح بخطوتين اضافيتين:

- اجلس فى المكتبة أو على الانترنت وحاول ان تطلع على جدول محتويات المجالات الاكثر ارتباطاً بالموضوع فى السنوات الخمسة الاخيرة مثلاً وإذا لم تحصل على مزيداً من الدراسات فإن هذا يزيد الثقة فى شمولية ومصادقية 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 قاعدة بيانات للدراسات المتطلبية لما وراء التحليل

بعد جمع الدراسات السابقة فإن الخطوة التالية هي كيفية تنظيم عرض الدراسات، فعملية البحث والحصول على الدراسات السابقة تستغرق وقتًا وجهدًا إلا أن عملية التنظيم تيسر على الباحث إدراك العلاقات بين هذه الدراسات. والتنظيم الجيد يتطلب وضع إطار تنظيمي محدد يتضمن مكونات أساسية مثل أولاً: معايير التضمين والاستبعاد التي تستخدمها لتقييم الدراسات المجمعة وفيها تحدد الدراسات المستبعدة في ضوء كل معيار وعلى ذلك توضح للقارئ مبررات استبعاد دراسة ما مما يوفر الشفافية. ثانياً: يجب أن توجد قائمة محددة لطرائق البحث في التراث البحثي مثل PsycINFO و ERIC مثلاً حتى تتيح للآخرين إعادة عملية البحث. ويمكن الاستعانة بالدراسات النظرية لأنها مفيدة في تحديد دراسات إضافية وكذلك في تفسير نتائج ما وراء التحليل.

وعملية البحث تتكون من جزئين:

- إيجاد المراجع المنشورة التي تلبى معايير التضمين.
- الحصول على نسخ من هذه الدراسات لقراءتها وتكويدها في ضوء محكات التضمين أو الاستبعاد.

والمبادئ الأساسية للمعاينة تشير إلى أن العينات المتاحة Convenience samples من غير المحتمل أن تكون ممثلة تمثيلاً جيداً لمجتمع الدراسات الكبيرة، فالبحث في قاعدة البيانات PsycINFO ليس ضماناً للحصول على قدر كافٍ من الدراسات المتطلبية لما وراء التحليل، وتزداد هذه المشكلة في واقع البحث في البيئة العربية لعدم توافر قواعد البيانات الكافية المتضمنة للتراث البحثي خاصة في العلوم السلوكية والاجتماعية. وينصح (Littell et al. 2008) بالحصول على عينة ممثلة لمجتمع الدراسات وتجنب اشكالية تحيز النشر أو أي مصادر تحيز أخرى مرتبطة.

ولا مانع من الحصول على عينة طبقية من الدراسات للموضوع حيث تفيد في دراسات مجموعات فرعية، فمثلاً أراد باحث دراسة فعالية برنامج تدخل لتخفيف العسر القرائي فلا مانع من جمع دراسات لبرامج في المرحلة الابتدائية والمرحلة الإعدادية.

ولتنظيم جمع الدراسات السابقة المرتبطة يفضل بعض الباحثون وضع خرائط منظمة Systematic maps للتراث البحثي لإعطاء نظرة لعدد وانواع الدراسات التي تناولت دراسة الموضوع، وتساعد هذه الخرائط أو الاطر في تحديد وتصنيف الدراسات في ضوء مجتمعتها، 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 بروتوكول 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 البحثي 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 بروتوكول 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 بروتوكول التكويد والإجراءات المستخدمة لاستخدامه لتكويد الدراسات وتدريب المكودين.

بناء بروتوكول التكويد Developing A coding Protocol

فى ضوء دراسات ما وراء التحليل لا توجد صيغة موحدة لعملية تكويد الدراسات بل تختلف من باحث إلى آخر وحسب اهداف الدراسة ويتضمن بروتوكول التكويد خصائص الدراسات المراد إجراء ما وراء التحليل لها وهذا يتوقف عل طبيعة المعلومات فى الموضوع أو الظاهرة المراد دراستها.

ويحدد (2001) Lipsey & Wilson الهدف من التكويد هو بناء قاعدة بيانات للتحليل الإحصائى ومن الأفضل استخدام مفردات ذات النهاية المغلقة كلما امكن ويحاول الباحث ان يحدد مسبقاً الاختيارات المحتملة لكل مفردة فى استبيان التكويد حيث تسهل على المكود ملئ مفردات الاستبيان أو القائمة مثل مفردة حجم العينة، ولكن توجد مواقف لا يستطيع الباحث تحديد بدائل مفردات الاستبيان وعليه فهو مضطر إلى بناء مفردات ذات نهاية مفتوحة وفى هذه الحالة يلجأ الباحث إلى مجموعة من التصنيفات الفرعية أو الابعاد ولكن هذا يستغرق وقت وجهد من الباحث أو المكود. كلما كان المكود مدرباً كلما زادت قدرته على تجميع المفردات فى اطر أو تصنيفات متقاربة. ويفضل ان يكون بروتوكول التكويد بسيط وقصير كلما أمكن ومنظم فى اجزاء واضحة، ومن المهم أيضاً اعطاء تعريف كامل وواضح لكل مفردة وإرشادات لكيفية تكويدها لمساعدة المكود للتحكم فى الحالات الغامضة وغير المعتادة. وفى الممارسة العملية بناء بروتوكول تكويد يدوى Coding manual حيث يترجم ماهى المعلومات المراد الحصول عليها من كل مفردة وكذلك اعطاء توصيف للبدائل المختلف للإجابة واعطاء إرشادات للحالات التى بها غموض. وكتاب التكويد اليدوى يكون بمثابة الدليل المرشد للمكود عندما يواجهه صعوبات.

ويحدد (2012) Card اعتبارين يمكن تطبيقهما لمراجعات ما وراء التحليل هما:

- أسئلة البحث موضع الاهتمام: مثل إجراء الدراسات الأولوية من تحديد المتغيرات فإن التخطيط لما وراء التحليل يتطلب تحديد السؤال البحثى الذى على أساس يتم تكويد خصائص الدراسة أو توصيفات الدراسة Study

descriptors مثل (الطريقة أو المنهج، القياسات، العينات، المتغيرات المستقلة والتابعة، نوعية المعالجة، الإجراءات وغيرها)، فلو ان سؤال البحث منصباً 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 بلامح معينة تعطى أحجام تأثيرات أكبر من أحجام تأثيرات لدراسات بلامح مختلفة وبالتالي فالباحث بحاجة إلى معلومات كثيرة لخصائص الدراسات لتكويدها.

ويوجد نوعية من الأسئلة البحثية تتضمن متغيرات متفاعلة، فبعض محلى ما وراء التحليل يرغبون فى التنبؤ بعدم التجانس فى أحجام التأثير من خصائص الدراسات المكودة، وعليه فإن الباحث بحاجة إلى تأكيد 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 التغير 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، وعموماً يجب أن يوضع فى الاعتبار أربعة أنواع أساسية لإجراء ما وراء التحليل فى العلوم الاجتماعية: خصائص العينة، عملية القياس، التصميم البحثى، والمصدر. ويمكن عرضها كالاتى (Card, 2012; Valentine, 2009):

المظهر العام	المظاهر الخاصة	امثلة
العينة	إجراءات المعاينة	نوعية المجتمع، العينة الممثلة، العينة العشوائية، البلد.....الخ.
خصائص القياس	مصدر البيانات	الجنس، العمر، الحالة الاقتصادية الاجتماعية، الذكاء، الموقع....الخ
عملية القياس	ملاحظة ذاتي (الاستبيان - المقاييس)، ملاحظة، مقابلة وغيرها.	ملاحظة منظمة أم غير منظمة، تطبيق فردي أم جماعي، مقيد بوقت أو غير مقيد
قياسات معينة مستخدمة	الصورة كاملة أم مختصرة، مقياس مترجم أم من البيئة أو الثقافة.	قياسات معينة مستخدمة
خصائص التصميم	نوع التصميم البحثي	المجموعة الواحدة وقياس قبلي وبعدي، تصميم المجموعتين قياس قبلي وبعدي.
خصائص مميزة للتصميم	نوع المعالجة التي تلقتها المجموعة الضابطة، الفترة الزمنية بين القياسات الطولية.	نوع المعالجة التي تلقتها المجموعة الضابطة، الفترة الزمنية بين القياسات الطولية.
المصدر	النشر	دراسة منشورة أو غير منشورة، جودة النشر
عام النشر	عام النشر	السنة، تاريخ جمع البيانات.
التمويل	التمويل	ممولة أم غير ممولة، جهة التمويل.
خصائص الباحثين	الجامعة، الجنس، الجنسية.	الجامعة، الجنس، الجنسية.

جودة الدراسة	الصدق الداخلي	معاينة عشوائية، ضبط المتغيرات الدخيلة، التزييف، الانسحاب.
الصدق الخارجي	العينات من المجتمع الممثلة له، التجريب يناظر الواقع.	
الصدق البنائي	ثبات القياسات (معايير التضمنين)، خصائص القياسات.	

وفيما يلي عرض لاهم خصائص الدراسات (Card, 2012; Littell et al. 2008; Lipsey& Wilson, 2002; Wilson, 2009)

1. كود الدراسة: يتم وضع كود الدراسة رقمياً أو حرفياً واحيائاً يتم تحديد مؤلف الدراسة وسنة النشر.

2. المفاهيم Constructs: دائماً تركز الدراسات التجريبية على متغير تابع واحد فهي 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 محل ما وراء التحليل ماهي المتغيرات موضع اهتمام دراسته لتكويدها. فمثلاً ما وراء التحليل لدراسات برامج القراءة العلاجية فهي 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، وإى توصيف آخر للعينة مثل الذكاء. وحجم 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 وأحياناً يتم عرض النتائج فى ضوء مجموعات فرعية فمثلاً تعرض العلاقة بين التحصيل والتثنية الأسرية لعينة الريف ولعينة المدينة على حدة إذا كان هذا موضع اهتمام دراسة ما وراء التحليل حيث تقدر أحجام التأثير للذكور وللإناث كلاً على حدة مثلاً 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 Measurement: من الضرورى التميز بين المفاهيم الإجرائية أو القياسات Operational definitions فى الدراسة عن المفاهيم النظرية Theoretical contrusts لأنه توجد مقاييس مختلفة تقيس نفس المفهوم، فمثلاً 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 يؤدي إلى عدم تماثل عبر الدراسات، والأفضل الاعتماد على الدراسات التى اعتمدت على المقياس الأكثر استخدامًا، وأحيانًا يكود محلل ما وراء التحليل نتائج الدراسات التى اعتمدت على مقاييس متعددة لنفس المفهوم ثم يقدر متوسط حجم التأثير. ومن المهم تضمين الخصائص السيكمترية مثل معاملات الصدق والثبات كمعايير للتضمين فإذا كان معامل الثبات 0.8 فيكتب 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. وفى 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 توكيد 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 توكيد طبيعة المجموعة الضابطة كونها تلقت

معالجة تقليدية أو لم تتلقى معالجة على الإطلاق. 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 لتقييم تحيز النشر، و**سنة النشر** فى المجلة أو المؤتمر أو بحث دكتوراه وذلك لدراسة التحليل المتفاعل لمعرفة اتجاهات أحجام التأثير عبر فترات زمنية مختلفة، و**مصدر تمويل** الدراسة لأن هذا من أكبر العوامل التى تؤدى احياناً إلى تحيز النتائج، و**خصائص الباحث** مثل الجنس أو الثقافة أو المؤسسة التى ينتمى إليها لدراسة تحيز الباحثين بمعنى مدى اختلاف أحجام التأثير للدراسات التى أجراها الذكور عن الدراسات التى أجريت عن طريق الإناث.

تأكيد حجم التأثير Effect Size Coding

أى 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 والتأكيد 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 SPSS , STATA, SAS, G-Power وغيرها من البرامج. وطور Shadish, Robinson, & Lu (1999) برنامج حاسوبى Effect Size Calculator ومتاح فى الموقع ([http:// www.assess.com/](http://www.assess.com/)) . والمعلومات المراد توكيدها 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 إحصائياً ولكن يمكن تضمين أحجام الاثر للذكور وللإناث فى تحليل واحد لانهما مستقلين.
- عملية التأكيد تتطلب بيانات إحصائية عن حجم الاثر مثل المتوسطات، النسب، أحجام العينة، التباينات، الانحرافات المعيارية، الارتباطات القبلية والبعدية وقيمة P الاحتمالية.... الخ. وكذلك يتضمن بروتوكول التأكيد معلومات إحصائية مطلوبة 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).

تأكيد جودة الدراسة Coding Study quality

ينصح البعض ان يتضمن ما وراء التحليل تأكيد 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 الإحصائى Statistical Conclusion Validity وداخل كل نوع من

انواع الصدق الاربعة توجد مظاهر عديدة Shadish, Cook, & Campbell (2002):

- **الصدق الداخلى** Internal validity : يشير إلى أى درجة ان تصميم الدراسة أو المتغير المستقل يسمح باستنتاج السببية، بمعنى أن التغير الحادث فى المتغير التابع يرجع ويرجع فقط إلى المعالجة وليس إلى متغيرات اخرى دخيلة. يعتبر تصميم الدراسة من أكثر الامور تأثيرًا فى الصدق الداخلى فالدراسات التجريبية (المعاينة العشوائية) أكثر صدقًا داخليًا من الدراسات شبه التجريبية (لا عشوائية).

- **الصدق الخارجى** External validity : يشير إلى أى درجة نواتج الدراسة قابلة للتعميم لمجتمع الدراسة وكذلك لعينات مختلفة لنفس المجتمع وشروط اخرى غير الشروط التى اجريت فيها التجربة ولطرق قياس اخرى لنفس المفهوم موضع الدراسة. والدراسات الاكثر صدقًا خارجيًا هى التى اعتمدت على عينات عشوائية من مجتمع الدراسة وهذا الشرط يكاد يكون نادرًا فى الدراسات النفسية والاجتماعية. من الضرورى فى دراسات ما وراء التحليل تكويد ماذا كانت الدراسات المجمعَة حققت هذا المعيار من عدمه، ولكن إجراء ما وراء التحليل لدراسات مفردة محدودية الصدق الخارجى يؤدى إلى استنتاجات عالية الصدق الخارجى وهذا من أهم مميزات ما وراء التحليل.

- **صدق المفهوم** Construct validity : يشير إلى أى درجة المقاييس المستخدمة فى الدراسة تقيس البناء النظرى الذى ينوى الباحث قياسه. ولكن هذا المصطلح يشير فى هذا الصدد إلى خصائص القياس وهى الصدق والثبات للقياسات ولذلك يجب تكويد ثبات المقاييس للقياسات المتضمنة للمتغيرين فى التحليل لإجراء تصحيحات أحجام الأثر، وكذلك يمكن تضمين قياسات الصدق لأن من معايير التضمين ان يكون التعريف الإجرائى للمفهوم واضح ويمكن تصحيح الصدق غير الدقيق من خلال أساليب ما وراء التحليل، خاصة حينما يقدر من خلال معامل الارتباط (معامل الصدق) بين المقياس المستخدم فى الدراسة ومقياس معيارى متفق عليه لنفس المفهوم ويطلق عليه الصدق المحكى.

- **صدق الاستنتاج الإحصائي:** يشير إلى صدق الاستنتاجات الإحصائية فيما يخص العلاقة بين المسبب المفترض والتأثير أو الناتج من التجربة.

وجودة الدراسة بناءً تحتى ومتعدد المظاهر من تصميم الدراسة وانتقاء العينة وطبيعة المعالجة الإحصائية ويمكن أخذها في الاعتبار كمتغير وسيط في ما وراء التحليل.

ويقترح (2012) Card توصيف المظاهر المختلفة لجودة الدراسة المهمة في المجال أو التخصص واعتبارها متغيرات متفاعلة لأحجام الأثر بين الدراسات.

ماذا عن الدراسات ضعيفة المنهجية؟

كثير من الخبراء يستبعد الدراسات ضعيفة المنهجية من دراسة ما وراء التحليل ولكن (2004) Hunter & Schmidt يرى أن هذا غير مناسب ولا يفضل استبعادها لأن عدم المناسبة المنهجية Methodological inadequacies لا يمكن اعتبارها إمبريقياً ولا توجد دراسة بدون محددات وقصورات منهجية على الرغم أن هذه المحددات لا تسبب دائماً تحيز في النتائج.

الاختلافات المنهجية بين الدراسات تسبب اختلاف النتائج عبر الدراسات نتيجة على الأقل لخطأ المعاينة في الدراسات ويمكن اختبار عدم المناسبة المنهجية من خلال اعتبارها متغير متفاعل ويمكن دراسة أثرها في دراسة ما وراء التحليل وإذا لم توجد فروق بين الدراسات عالية المنهجية والدراسات ضعيفة المنهجية على ذلك لا توجد إشكالية عدم المناسبة المنهجية، وعليه يجب تضمين الدراسات الضعيفة المنهجية والجيدة المنهجية ثم دراسة الفروق بينهما ويكون القرار في النهاية في ضوء الدراسات عالية الجودة وإذا لم يوجد فروق بينهما (الضعيفة - الجيدة) يكون التحليل على كل الدراسات، وعليه فلا بد من تقييم المناسبة المنهجية إمبريقياً ولا يتم الحكم عليها في ضوء التقييمات الذاتية الشخصية حيث توصل Cooper (1998) أن معامل الاتفاق بين المقيدين للحكم على المناسبة المنهجية على أفضل الأحوال 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 منهجياً فى دراسة ما وراء التحليل.

توصيفات الدراسة Study Descriptors

كل الخصائص المكونة المتحصل عليها من الدراسة يطلق عليها مواصفات الدراسة وطرح (2001) Lipsey & Wilson 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 كالاتى:

- **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 مثلاً فى ما إذا يوجد اختلاف لأحجام الاثر عبر التخصصات المختلفة (علم النفس فى مقابل الصحة النفسية).

ويمكن تلخيص توصيفات الدراسة كالاتى:

القضايا الأساسية

مصدر العينة

- الخصائص الديموجرافية (المستوى الاقتصادي الاجتماعي، العمر، الجنس، المستوى التعليمي، مكان المعيشة وغيرها).
- الخصائص الشخصية (القدرات المعرفية، السمات الشخصية).
- الخصائص التشخيصية أو ملامح خاصة (نزول سجن، صعوبة تعلم، مصاب بمرض معين، معاق سمعياً وغيرها).
- المتغير المستقل (التدخل أو المعالجة أو البرنامج).
- النظرية القائمة عليها الدراسة أو البرنامج.
- خصائص المنظمة أو المؤسسة (عمرها، حجمها، الهيكل الإداري).
- 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 الداخلي (العوامل الدخيلة في التصميم).
- القوة الإحصائية.
- جودة القياسات (الصدق والثبات).
- كيفية تحليل البيانات.
- طبيعة المجموعة الضابطة (لم تتلقَ معالجة أو تلقت معالجة بديلة).
- طريقة توزيع أفراد العينة على مجموعات التصميم.
- دور المجرب أو الباحث في المعالجة أو البرنامج.

توصيفات المصدر

شكل النشر (مجلة، كتاب، مجلد، رسالة دكتوراه، تقرير حكومي أو مؤسسي)، تاريخ النشر، بلد النشر واللغة، تمويل الدراسة والجهة المانحة، خصائص الباحث (الجنس، التخصص، والدرجة العلمية).

ومن الضروري توكيد هذه المتغيرات وتختلف عملية التوكيد حسب الهدف من دراسة ما وراء التحليل وطبيعة الدراسات المتضمنة في التحليل.

تقويم عملية التوكيد Evaluatig Coding Process

بعد الانتهاء من تحديد خصائص الدراسة المراد توكيدها فإن الخطوة التالية اقرأ بعناية الدراسات لتوكيد 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 مرتبطتان ببروتكول التوكيد يحدداهم (Card, 2012; Lipsey & Wilson, 2002) بالآتى:

- الشفافية والاعادة للتوكيد Transparency & Replicability of Coding:

عند كتابة التقرير النهائى لدراسة ما وراء التحليل يجب اعطاء تفاصيل كافية عن عملية التوكيد حتى يعرف القارئ كيفية عمل قرارات التوكيد(الشفافية)، وكذلك إذا ما طبق باحث آخر نفس قواعد التضمين هل يصل إلى نفس الاستنتاج أو القرارات التوكيدية؟ ولتحقيق هذين المبدأين فمن الضروري اعطاء وصف كامل لكيفية توكيد كل دراسة كمياً، فبعض خصائص الدراسة تكود بطريقة مباشرة ولا تتطلب جهد من المكود ودائماً لا تختلف من مكود إلى آخر مثل متوسط عمر العينة، ولكن عندما تكون خصائص الدراسة أقل وضوحاً حيث تتطلب احكام توكيد عالية الاستدلال فمن الضروري كتابة تقرير كامل عن عملية التوكيد للتأكد من الشفافية والاعادة وهذا بدوره يسهم فى تقدير ثبات المكودين، فعلى سبيل المثال لتوكيد طبيعة القياسات المستخدمة وتصميمات الدراسات فيجب كتابة قيم مختلفة لهذا الكود التصنيفي.

- **ثبات التكويد Reliability of coding**: لتقويم عملية الاعدادة لنظام التكويد فمن الضروري 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 يوماً 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 دراماتيكي 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 كل من النوعين ضروريان لإجراء ما وراء التحليل.

- **الاتساق بين مكودين مختلفين Inter-coder** : حيث يكود باحثان أو مكودان نفس العينة من الدراسات المجموعة بصورة مستقلة ويفضل ان تكون عينة الدراسات من 20 إلى 50 دراسة للحصول على ثبات دقيق نسبياً 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 عالى Overestimate للثبات الحقيقى.

وباستخدام مدخلى تقدير ثبات المكودين فمن المفيد تقدير الثبات لكل خصائص دراسة تم تكويدها على حدة وايضاً 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 كالاتى:

$$AR = \frac{\text{عدد الاتفاقات}}{\text{عدد الدراسات}}$$

- عدد الاتفاقات: عدد الدراسات التى اخذت نفس الكود للخاصية.

- عدد الدراسات: عدد الدراسات التي 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 هي 70% . وهو بسيط فى حسابه ويعتبر مؤشراً لثبات التكويد ولكن توجد بعض المحددات هي قضية التحيز فيمكن أن يكون أحد المكودين متحيز لبدل 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 التي يعانى منها معامل الاتفاق

ويقدر المعامل Kappa من المعادلة الآتية:

$$K = \frac{\sum fo - \sum fe}{N - \sum fe}$$

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 مثلاً 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** Pearson Correlation(r): 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

المكود يمكن استخدام أحد الاختبارات لتحديد الفروق بين القياسات التكويدية باستخدام اختبار T المرتبطة.

بناء وتنظيم نظام أو بروتكول التكويد

بعد تحديد الخصائص المراد تكويدها، فالخطوة التالية التخطيط لتكويد الدراسات، والمرشد لعمل التخطيط هو بروتكول التكويد الذى يتضمن وسيط التكويد الشيت يدوى أو الكمبيوتر الذى يفرغ فيه المعلومات المكودة بالإضافة إلى كتاب التكويد الذى يتضمن تعليمات عملية التكويد. وفى النهاية نحصل على قاعدة بيانات قابلة للتحليل فى برامج ما وراء التحليل.

التكويد اليدوى Manual Coding

كتاب التكويد اليدوى هو مجموعة مفصلة من التعليمات تصف نوعية المعلومات المطلوب استخلاصها من البحوث والدراسات فى صورة كمية لإدخالها البرنامج الكمبيوترى لما وراء التحليل، وهذا الكتاب التكويدى يخدم ثلاثة أهداف الأول يمد المكود بمرشد لنقل المعلومات أو خصائص الدراسة إلى وسيط التكويد (الشكل الورقى) وهذه المعلومات فى كتاب التكويد تستخدم فى المواقف النموذجية كذلك فى المواقف التى تتضمن اشكاليات معينة. **الثانى** يهدف إلى توحيد المعلومات المتحصل عليها عبر المكودين بالتالى أى حدوث اتساق عن طريق تعليمات واضحة وشفافة لما يقوم به كل مكود. **الثالث** يعتبر مفتاح التكويد اليدوى بمثابة وثيقة لعملية التكويد التى تعتبر كالمرشد لقاعدة بيانات ما وراء التحليل.

ويعد مفتاح التكويد اليدوى الأولى قبل تكويد الدراسات وى معلومات يمكن أن تضاف 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 - ابتدائي - 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

توكيد حجم التأثير

لتوكيد حجم التأثير لكل دراسة يجب توكيد 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 المعياري للمجموعة التجريبية.....

الانحراف المعياري 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 الكتروني لإجراء ما وراء التحليل

كالآتي:

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

فالمغيرات 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

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 (2012) Borenstein et al. إلى أن مصطلح 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 استخداماً 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 الابداعي - 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 فمثلاً 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 كالاتى (Lipsey & Wilson, 2001; Borenstien et al., 2009):

$$ES_{um}(D) = \bar{x}_1 - \bar{x}_2 = M_1 - M_2$$

• $\bar{x}_1$ ، M_1 متوسط المجموعة الأولى

• $\bar{x}_2$ ، M_2 متوسط المجموعة الثانية

ولتقدير الخطأ المعيارى لفروق المتوسطات غير المعيارية لاستخدامها فى تصحيح أو موازنة حجم التأثير كالاتى (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 كالاتى:

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}$$

• S_1^2 تباين قيم المجموعة الأولى.

• S_2^2 تباين المجموعة الثانية.

مثال: افترض أن دراسة تجريبية هدفت إلى المقارنة بين المجموعة التجريبية والمجموعة الضابطة في العدوان وكان متوسط المجموعة الأولى $(\bar{x}_1) = 103$ ومتوسط المجموعة الثانية $(\bar{x}_2) = 100$ والانحراف المعياري للعينتين $S_1 = 5.5$ ، $S_2 = 4.5$ وحجم العينة للعينتين متساوى $n_1 = n_2 = 50$ وعليه فإن حجم التأثير:

$$SE_{um} = 103 - 100 = 3$$

والانحراف المعياري المدمج كالاتي:

$$S_{pooled} = \sqrt{\frac{(50 - 1) \times 5.5^2 + (50 - 1) \times 4.5^2}{50 + 50 - 2}} = 5.0249$$

والخطأ المعياري لفروق المتوسطات:

$$SE_{um} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 5.0249 \sqrt{\frac{1}{50} + \frac{1}{50}} = 1.0050$$

الإجراءات الحسابية: تتطلب حساب حجم التأثير من خلال حساب الفرق بين المتوسطين مباشرة وتوافر الانحراف المعياري للمجموعتين واحجام عينة. ولكن بفرض ان بعض الدراسات لم تذكر أحد هذه المؤشرات خاصة الانحراف المعياري في هذه الحالة يمكن حساب الانحراف المعياري 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)

هي المؤشرات البديلة لفروق المتوسطات المعيارية لتقدير حجم التأثير وتنقسم إلى المؤشرات الارتباطية ومؤشرات التباين المفسر.

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. وتقدر قيمته كالاتى:

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N(\sum X^2 - (\sum X)^2))(\sum Y^2 - (\sum Y)^2)}}$$

- $\sum XY$ مجموع حاصل ضرب X فى Y.
- $\sum X$ مجموع قيم X، $\sum Y$ مجموعة قيم Y.
- $\sum X^2$ مجموع مربعات X، $(\sum X)^2$ مربع مجموع X.
- $\sum Y^2$ مجموع مربعات Y، $(\sum Y)^2$ مربع مجموع Y ، N حجم العينة.

ويكون أدائه أفضل إذا كان توزيع البيانات اعتداليًا والعلاقة بين المتغيرين خطية. والارتباط الموجب يعنى أن الأفراد مرتفعى الدرجات على المتغير X لهم درجات مرتفعة على المتغير Y بينما الارتباط السالب الأفراد مرتفعى الدرجات على المتغير X لهم درجات منخفضة على المتغير Y .

وتعتبر قيمة معامل الارتباط مؤشراً لحجم التأثير، وهو مؤشر أكثر معيارية وأسهل فى تفسيره نتيجة اعتماده على معايير فى تحديد قوة العلاقة ويقع مداه من 1.0 (موجب) إلى -1.0 (سالب) عكس مؤشر كوهين " d " فليس له معايير واضحة لتفسيره (Rosenthal et al., 2000)، وتفسير قيمة r كما وضحا 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 هذه المعايير ليست ثابتة فى كل التخصصات أو المجالات. ويفسر r من خلال مربع قيمته (r^2) ويطلق عليه **معامل التحديد** Coefficient of determination، وعلى ذلك يكون مربع معامل الارتباط ضعيفاً إذا كانت $r^2 = 0.01$ ، متوسط $r^2 = 0.09$ ، وكبير $r^2 \geq 0.25$. فمثلاً إذا كان $r = 0.50$ فإن مربع معامل الارتباط $r^2 = 0.25$ بمعنى أن أحد المتغيرين فسر 25% من تباين المتغير الثانى.

وتباين الخطأ أو الخطأ المعياري لمعامل الارتباط تقريباً 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 تحويل أو مقياس فيشر transformation(Z_r) Fisher (ليس اختبار الدلالة الإحصائية Z) ويجرى التحليل على القيم المحورة، ويقدر Z_r كالتالي (Card, 2012; Lipsey & Wilson, 2001):

$$(ES_{Z_r})Z_r = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$$

حيث $\ln$ أو $\log_e$ اللوغاريتم الطبيعي.

وعليه فإن الدراسات التي تعتمد على معامل الارتباط كمؤشر لحجم التأثير في الدراسات الارتباطية في ما وراء التحليل يجب تحويله إلى الدرجة أو تحويل Fisher ثم تحويل القيم النهائية لفischer إلى معامل الارتباط لتفسير النتائج.

ويحول معامل ارتباط بيرسون إلى تحويل فيشر في ما وراء التحليل لأن التوزيع العيني لمعامل الارتباط في المجتمع يكون ملتوى ناحية اليمين للمعاملات السالبة وناحية اليسار للمعاملات الموجبة وعندما يكون قريب من الصفر فإن التوزيع العيني يكون تقريباً اعتدالي وتحسن الاعتدالية بزيادة حجم العينة ولأن التوزيع العيني يكون ملتوى كلما زادت قيمة معامل الارتباط فإنه من غير المناسب استخدام التوزيع الملتوى لاختبارات الفروض حول معامل الارتباط في المجتمع p إلا إذا كانت $=0$ p وهذا أدى بـ Fisher لإجراء تعديل فيشر Z لأنه توزيع منتظم Symmetric (Lipsey & Wilson, 2001, Hedges & Olkin, 1985).

وتباين تحويل Z_r :

$$V_z = \frac{1}{N - 3}$$

حيث N حجم العينة.

والخطأ المعياري لـ Z_r :

$$SE_{Z_r} = \sqrt{V_z}$$
$$= \sqrt{\frac{1}{N-3}}$$

وكلما زادت حجم العينة فإن مقام المعادلة السابقة يزيد وبالتالي تنقص قيمة الخطأ المعياري. ويجب إدخال الخطأ المعياري لتحويل فيشر إلى قاعدة البيانات أثناء عملية التكويد لاستخدامه في عملية موازنة أحجام التأثير، وعليه عند استخدام تحويل فيشر لا نستخدم تباين معامل الارتباط انما نعتد على تباين الدرجة Z_r . ولأن 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 التالي (Hedges & Olkin, 1985):

$$r = \frac{e^{2Z_r} - 1}{e^{2Z_r} + 1}$$

حيث r معامل ارتباط بيرسون، Z_r تحويل فيشر المقابل إلى r .

فعلى سبيل المثال في دراسة ما قدرت معامل الارتباط بـ 0.50 مع حجم عينة 100 بالتالي فإن تحويل فيشر Z_r كالتالي:

$$Z_r = 0.5 \times \log_e \left(\frac{1 + 0.5}{1 - 0.5} \right) = 0.5493$$

وتباين الدرجة Z_r :

$$V_z = \frac{1}{100 - 3} = 0.0103$$

والخطأ المعياري للدرجة Z_r :

$$SE_z = \sqrt{0.0103} = 0.1015$$

ولتحويل الدرجة Z_r إلى معامل الارتباط:

$$r = \frac{e^{(2 \times 0.5493)} - 1}{e^{(2 \times 0.5493)} + 1} = 0.50$$

مثال: أراد باحث دراسة العلاقة بين ادمان الانترنت والوحدة النفسية بالتالى جمع الدراسات التى تناولت هذه العلاقة من التراث فى البيئة العربية وهنا يعتبر معامل الارتباط مؤشر لحجم التأثير ثم يحول إلى تحويل Fisher وبعد حساب المتوسط العام لتحويل Z يحول مرة اخرى إلى معامل الارتباط.

حساب قيمة معامل الارتباط من قيم الاختبارات

بعض الدراسات الأولية المنشورة تتضمن نتائج الإحصائيات الاستدلالية دون حساب حجم التأثير ويفضل فى كثير من تحليلات ما وراء التحليل التعبير عن معظم الإحصائيات فى ضوء معامل الارتباط r وهذا الموقف يحدث لعدة اسباب منها:

1. تتضمن الدراسات الدلالة الإحصائية لمعامل الارتباط بدون قيمة معامل الارتباط وهذا الموقف نادر الحدوث.

2. بعض الباحثين يميلون إلى تقسيم المتغير المتصل إلى تصنيفين مثل مرتفع ومنخفض التحصيل أو مرتفعى ومنخفضى الدافعية ثم المقارنة بينهما على المتغير التابع باستخدام اختبار T أو F، ولكن هذا تقسيم اصطناعى وبدوره يؤدي إلى تقلص Attenuation قيمة حجم التأثير (Hunter & Schmidt, 2004)، ولذلك من المفيد حساب قيمة معامل الارتباط بل وتصحيحه من التقسيم الاصطناعى للمتغيرات المتصلة.

3. بعض الدراسات تُقسم أو تصنف كلاً من المتغيرين المتضمنين فى معامل الارتباط إلى تصنيفين ثم تحلل المتغيرين من خلال جداول الاقتران باستخدام إحصاء χ^2 وفى هذه 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

• $df = N - 2 = n_1 + n_2 - 2$ درجات الحرية:

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 $df = N - C$ ، حجم العينة، C عدد المجموعات ويساوى 2.

ويقدر قيمة r من اختبار T المرتبطة كالتالى:

$$r = \sqrt{\frac{T^2}{T^2 + df}}$$

• $df = N - 1$ درجات الحرية

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 تصنيفي (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):

$$r = \sqrt{\frac{g^2 n_1 n_2}{g^2 n_1 n_2 + (n_1 + n_2) df}}$$

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 (r ²)	.10(.01)	.30(.09)	.50(.25)
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 الداخلي

يهدف مقياس الاتساق الداخلي أو الثبات الداخلي للمقياس التي تحديد درجة التناسق بين مفردات المقياس، وتوجد العديد من مقاييس الاتساق الداخلي ولكن أكثرهم انتشاراً المعامل ألفا كرونباخ ويقدر في ضوء عدد المفردات في المقياس j ومتوسط الارتباطات بين هذه المفردات ($\bar{r}$) كالآتي:

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

متى يتم إجراء ما وراء التحليل لمقاييس الاتساق الداخلي؟ يرى (2012) Card موقفين لاستخدام الاتساق الداخلي في ما وراء التحليل هما:

الأول: عند تصحيح الثبات في الدراسات من نقص الثبات Unreliability .

الثاني: عندما يهدف الباحث إلى الوصول إلى متوسط المعامل ألفا كرونباخ لدراسات مختلفة استخدمت نفس المقياس وهذا في غاية الأهمية خاصة للمقاييس المعروفة في دراسة ظاهرة معينة مثل مقياس كاتل للشخصية أو اختبار ستانفورد بينية أو العوامل الخمسة الكبرى للشخصية وغيرها لتقييم استخدام هذه المقاييس في دراسة الظاهرة. وتوجد طرق مطروحة في التراث لإجراء ما وراء التحليل لمعامل ألفا كرونباخ منها ما قدمه (2006) Rodriguez & Maeda حيث تم تحويل المعامل ألفا لمؤشر لحجم التأثير كالتالي:

$$ES_{\alpha} = \sqrt[3]{1 - \alpha}$$

• α معامل الثبات في الدراسة.

والخطأ المعياري لهذا التحويل لمعامل ألفا كالآتي:

$$SE_{ES\alpha} = \sqrt{\frac{18j(N-1)(1-\alpha)^{\frac{2}{3}}}{(j-1)(9N-11)^2}}$$

• 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 أخرى حتى يسهل

تفسيره كالآتى:

$$\alpha = 1 - ES_{\alpha}^3$$

• ES_{α} معامل ألفا المحور فى ما وراء التحليل.

واخيراً توجد مجموعة من الاشكاليات أو التحديات عن إجراء ما وراء التحليل

لهذه النوعية من أحجام التأثير التى عرضناها فى هذا الفصل يحددها (Card,

2012; Lipsey & Wilson, 2001):

- عدم توافر البيانات المتوفرة لتقدير حجم التأثير، فمثلاً فى الدراسات الطولية

لا تتوفر للباحث معامل الارتباط بين مرتى القياس للمتغير عكس معامل الاتساق

الداخلى الذى تتوفر البيانات المطلوبة لحساب حجم التأثير. وعدم توفر البيانات

اللازمة لحساب حجم التأثير يتطلب الاتصال المباشر بمؤلفى هذه الدراسات.

- التحدى الثانى هو عدم الاتساق فى الطرق التحليلية وتقرير أحجام التأثير المناسب،

فمثلاً فى تحليل الانحدار عند حساب حجم التأثير حيث تتضمن الدراسات متغيرات

منبئات أو تغايرات متنوعة غير المتغيرين موضعى الاهتمام وتتضمن الدراسات

طرق تحليلية متنوعة فبعضها اعتمد على طريقة التحليل Enter وبعضها اعتمد

على Stepwise.

ايضاً عند إجراء ما وراء التحليل على التحليل العاىلى الاستكشافى حيث يعتمد على

تشبعات المفردات بالعوامل كمؤشر لحجم التأثير فيوجد اختلاف واضح فى الطرق

التحليلية فى ضوء طرق التقدير أو الاستخلاص (المكونات الأساسية- طرق التحليل

العاىلى) وطرق التدوير (مائل- متعامد) وكيفية تحديد عدد العوامل، وبكل تأكيد

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 وايضاً ضرورة 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

والمغيرات 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 الاستدلالي 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 وبالتالي يكون تقديره أكثر دقة وثبات من ذلك حجم التأثير القائم على حجم عينة صغير، وبالتالي 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 دوراً 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:

أولاً: إجراء تصحيحات لأحجام التأثير من المحددات المنهجية للدراسات (الفصل السابع).

ثانياً: 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 الضروري إعادة 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 مستقلة 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 (مثلاً 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 فى الدراسة التى تتناول مفاهيم مختلفة وهذا نادرًا ما يحدث لأنه لا يتناسب مع منهجية ما وراء التحليل التى 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 فى 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 فى ضوء مدى مناسبة التعريف الإجرائى فى الدراسة مع التعريفات الإجرائية فى 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 فى حساب متوسط حجم التأثير. فى 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 فى حساب حجم التأثير العام. من هنا تأتى أهمية عملية الموازنة Weighting للدراسات المتضمنة فى التحليل بما يتناسب مع دقتها وجودتها وحجم العينة.

مثال(2012 Card): افترض أن دراسة تتضمن عينة 10 أفراد وتوصلت إلى ان العلاقة بين X، Y تساوى 0.20 (تحويل فيشر = 0.203)، والدراسة الثانية اعتمدت على عينة 10000 فرداً و معامل الارتباط بين X، Y 0.30 (تحويل فيشر = 0.310)، ولكن قبل حساب متوسط العلاقة بين المتغيرين فى الدراستين فمن المهم التأمل فى دقة تقديرات أحجام التأثير للدراستين، فالدراسة الأولى اعتمدت على عينة 10 أفراد والخطأ المعيارى لتحويل فيشر Z_r لها كالاتى:

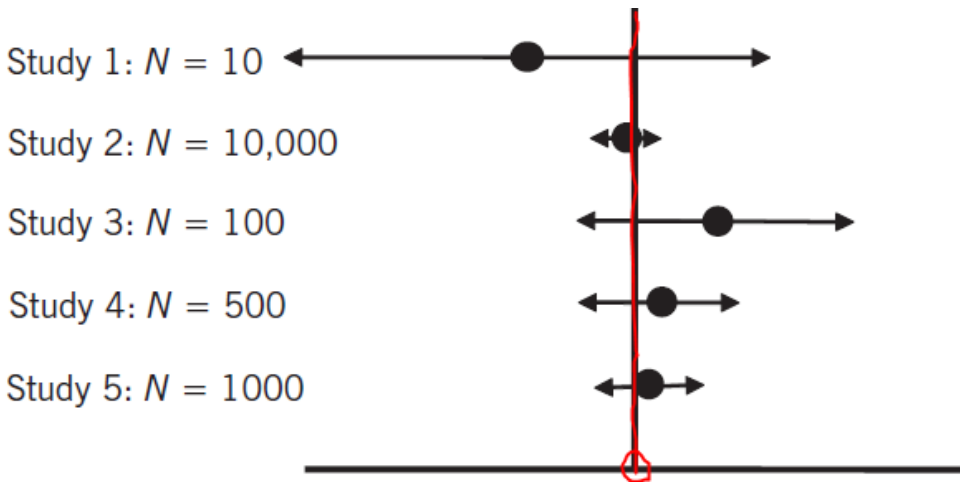
$$SE_{Z_r} = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{10-3}} = 0.378$$

فى حين ان الخطأ المعيارى لـ Z_r للدراسة الثانية:

$$SE_{Z_r} = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{10000-3}} = 0.010$$

ولاحظ أن الخطأ المعيارى فى حالة استخدام عينات كبيرة قليل مقارنة بنظيره فى حالة عينات صغيرة، وعليه فإن أحجام التأثير المتولدة من عينات كبيرة أكثر دقة وتقترب من حجم التأثير الحقيقى الفعلى فى المجتمع. والخطأ المعيارى يتناسب عكسياً مع حجم 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 نسبياً وهذا يؤدي 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 جداً فإن حجم التأثير فى هذه الدراسة (الدائرة السوداء) يكون قريب جداً من حجم التأثير الحقيقى فى المجتمع ومدى فترات الثقة يكون ضيقاً لأن هذه الدراسة لها خطأ معيارى صغير جداً. 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. فإذا كانت وزن الدراسة i هو w_i ، والخطأ المعيارى لحجم التأثير فى الدراسة SE_i ، فإن وزن الدراسة هو مقلوب أو معكوس تباين الخطأ المعيارى كالاتى:

$$w_i = \frac{1}{SE_i^2}$$

• SE_i الخطا المعياري 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.
- تحويل فيشر لكل معامل ارتباط Z_r .

- الخطأ المعياري لمعامل التحويل SE_{Zr} لفيشر فمثلاً الدراسة الأولى حجم العينة 228 وعليه فإن الخطأ المعياري لتحويل 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 كالتالى:

$$w_{study1} = \frac{1}{SE_i^2} = \frac{1}{(0.0693)^2} = \frac{1}{0.00480249} = 208.2253$$

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 محددین أساسین 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 الأكثر حدوثاً أو شيوعاً ولا يفضل على الإطلاق استخدامه فى تحليلات ما وراء التحليل لعرض القيمة النموذجية المنتصفية لأحجام التأثير. أما **الوسيط** Median وهو القيمة المنتصفية بعد ترتيب القيم تنازلياً أو تصاعدياً 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 يستبعد ما من طرفى التوزيع وهذا عكس المتوسط الأكثر تأثراً بالالتواء. ولكن نظراً لاعتماد الوسيط على قيمة أو قيمتين فلا يأخذ فى اعتباره كل

أحجام التأثير الموزونة للدراسات ولذلك يعتبر الوسيط مؤشر مكملاً للمتوسط. ومقياس النزعة المركزية الأكثر استخداماً في دراسات ما وراء التحليل المتوسط Mean وهو مجموع القيم مقسوماً على عددها وهو يتضمن كل أحجام التأثير الموزونة.

ويقدر المتوسط الموزون لحجم التأثير للدراسات المختلفة من الأوزان w_i وأحجام التأثير ES_i من كل دراسة i كالتالي (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 .
- $\overline{ES}$ أو M_{ES} متوسط حجم التأثير.

وعليه فإن متوسط حجم التأثير للدراسات هو مجموع حاصل ضرب حجم التأثير لكل دراسة في وزنها مقسوماً على مجموع أوزان الدراسات. لو كان وزن كل الدراسات مساوياً للواحد الصحيح أو كان وزن كل الدراسات متساوياً فإن المتوسط يكون مجموع أحجام التأثير مقسوماً على عددها.

والمعادلة السابقة تطبق على أي نوع من أحجام التأثير سواء معامل الارتباط وتحويره Z_r حيث يقدر المتوسط الموزون لتحوير فيشر ثم يحول مرة أخرى إلى معامل الارتباط، ومؤشرات فروق المتوسطات المعيارية مثل d ، ومؤشر Odds Ratio.

مثال: في الجدول السابق المتضمن 22 دراسة يتم حساب المتوسط الموزون كالتالي:

$$\begin{aligned} \overline{Z_r} &= \frac{\sum(w_i Z_{ri})}{\sum w_i} \\ &= \frac{\sum(.583 \times 208.12) + .201 \times 297.81 + \dots + .509 \times 482.45 + .651 \times 202.92}{\sum(208.12 + 297.81 + \dots + 482.45 + 202.92)} \\ &= \frac{2764.36}{7152.21} = 0.387 \end{aligned}$$

وعليه فإن معامل ارتباط بيرسون المقابل لقيمة متوسط معامل التحوير فيشر $\overline{Z_r}$ كالتالي:

$$r = \frac{e^{(2 \times 0.387)} - 1}{e^{(2 \times 0.387)} + 1} = 0.368$$

وعليه يوجد معامل ارتباط موجب من الدرجة المتوسطة بين العدوان ورفض الاقتران.

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 المثال 95% 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 إحصائياً 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 كالاتى (Hedges & Olkin, 1985):

$$SE_{\bar{ES}} = \sqrt{\frac{1}{\sum w_i}}$$

• w_i وزن كل دراسة

مثال: فى الجدول السابق يقدر الخطأ المعيارى كالاتى:

$$SE_{\bar{z}_r} = \sqrt{\frac{1}{\sum w_i}} = \sqrt{\frac{1}{7152.21}} = 0.118$$

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 تخط بين اختبار Z للدلالة

الإحصائية وتحويل فيشر (Z_r) من خلال قسمة قيمة المعلم (المتوسط الموزون) مقسومًا على الخطأ المعياري كالتالي:

$$Z = \frac{\overline{ES}}{SE_{\overline{ES}}}$$

• $\overline{ES}$ متوسط حجم التأثير لكل الدراسات.

• $SE_{\overline{ES}}$ الخطأ المعياري لمتوسط حجم التأثير.

وهذا الاختبار يتم تقييمه وفقًا للتوزيع الاعتدالي المعياري لاختبار Z وقيمتها الجدولية أو الحرجة عند مستوى دلالة إحصائية $\alpha = 0.05$ هي 1.96 ($p \leq 0.05$) ولمستوى دلالة إحصائية 0.01 هي 2.58 ($p \leq 0.01$)، وعليه لو كانت قيمة Z المحسوبة أكبر من القيمة الجدولية أو الحرجة بالتالي توجد دلالة إحصائية بمعنى أن العلاقة أو الفرق بين المتغيرين دالة إحصائيًا.

ويكون الفرض الصفري والبديل في هذه الحالة:

$$H_0: \overline{ES} = 0$$

$$H_A: \overline{ES} \neq 0$$

• $\overline{ES}$ متوسط حجم التأثير في المجتمع.

وتقدر فترات الثقة من الخطأ المعياري كالتالي (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 الأدنى.

• $\overline{ES}_{upper}$ 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 المعياري لمتوسط حجم التأثير.

مثال: بالرجوع إلى جدول البيانات فإن قيمة اختبار Z كالتالي:

$$Z = \frac{\overline{ES}}{SE_{\overline{ES}}} = \frac{0.387}{0.118} = 32.70, P < .001$$

بالتالى فإن متوسط حجم التأثير دال إحصائياً 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 إحصائياً بين العدوان ورفض الاقران.

ولحساب 95% فترات الثقة كالاتى:

$$\overline{ES}_{lower} = 0.387 - 1.96 \times 0.0118 = 0.363$$

$$\overline{ES}_{upper} = 0.387 + 1.96 \times 0.0118 = 0.410$$

ثم تحول قيمتى فترتى الثقة لتحويل فيشر إلى معامل الارتباط المقابل لتصبح الحد الأدنى $r = 0.348$ وهكذا للحد الأعلى $r = 0.388$ ، بالتالى فترتى الثقة 95% 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 كالاتى: اتضح من نتائج ما وراء التحليل لـ 22 دراسة تناولت العلاقة بين العدوان والرفض ان متوسط معامل الارتباط 0.368 وهى دالة إحصائياً حيث انها أكبر من الصفر ($P < .01$)، 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 الإحصائي الذي يرفض الفرض الصفري 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 في المجتمع، بكلمات أخرى وجود فروق أو انحرافات بين أحجام التأثير فإنها لا ترجع فقط إلى خطأ المعاينة بل ترجع إلى مصادر أخرى ربما تكون مرتبطة بخصائص الدراسات المختلفة في هذه الحالة لا يوجد تجانس ولكن تباين واختلافات.

انظر إلى الشكل (1.22) تلاحظ أن كل الدراسات لها حجم تأثير واحد متضمناً في الخط الراسي الذي يعرض حجم التأثير العام في المجتمع في هذا الموقف يوجد تجانس وان أي ابتعاد لحجم التأثير في الدراسة (الدائرة) يكون متوقع نتيجة خطأ المعاينة وليس لأي مصدر آخر وفي هذا الموقف يوجد تجانس. ولكن وجود اختلافات أكبر من خطأ المعاينة في هذه الحالة لا يصلح الخط المستقيم في الشكل (1.22) لعرض أحجام التأثير انما توجد عدة بدائل للتعامل مع هذا التوزيع غير المتجانس نتناولها لاحقاً.

وتقيم الدلالة الإحصائية للتجانس لأحجام التأثير من خلال إحصاء Q ويسمى اختبار Hedges Q test للتجانس وهذا الاختبار يعرض مدى الاختلاف في أحجام التأثير بين الدراسات المختلفة المتضمنة في ما وراء التحليل ويقدر من كالتالي (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 للدراسة i .
- $\overline{ES}$ متوسط حجم التأثير الموزون العام لكل الدراسات.
- درجات الحرية لاختبار Q:

$$df = k - 1$$

- K عدد الدراسات في 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 في المجتمع وهو الخط الراسي في الشكل (1.22).

ولكن كيف يمكن استنتاج وجود تجانس من عدمه؟ يعامل إحصاء Q في ضوء توزيع كاي تربيع χ^2 درجات حرية k-1 وعلى ذلك يجب البحث في جدول χ^2 لاستخراج قيمتها الجدولية وإذا كانت:

$$\chi^2_{\text{الجدولية}} \geq \text{قيمة Q المحسوبة}$$

نرفض الفرض الصفري بمعنى عدم وجود تجانس أو وجود اختلاف بين أحجام التأثير أكثر مما توقعناه نتيجة خطأ المعاينة.

مثال: في 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 في 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 في جدول χ^2 الاتي:

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 مثلاً 0.01 فإن χ^2 الجدولية = 38.932 اذًا:

χ^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 الأخذ في الاعتبار القوة الإحصائية لهذا الاختبار ونقص هذه القوة للاختبار فيجب أن يكون الباحث على حذر في تفسير النتائج غير الدالة إحصائياً كدليل للتجانس (الفرض الصفري).

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 يجرى تباین حجم التأثير المقاس من الدراسة إلى مكونين هما جزء يعزى إلى خطأ المعاينة لعينات الأفراد في الدراسة والثاني 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 في ما وراء التحليل الا أنه لا يشير إلى كيفية ان يكون هذا الاختلاف بين أحجام التأثير. ومن المؤشرات المفيدة لتحديد التجانس في ما وراء التحليل مؤشر I^2 ويفسر كنسبة مئوية للاختلاف بين أحجام التأثير الموجود بين الدراسات بالنسبة إلى الاختلاف الكلي بين أحجام التأثير (Higgins & Thompson, 2002; Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006، ويتم تقديره كالآتي:

$$I^2 = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{Q - (k - 1)}{Q} \times 100, \text{ when } Q > (k - 1)$$

- 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 تقريباً عند الصفر عندما تكون Q أقل من المتوقع تحت الفرض الصفري للتجانس. والمقام مكون من Q وهو عبارة عن التباين الكلى بين أحجام التأثير بينما البسط يتكون من الاختلاف الكلى بين أحجام التأثير مطروحاً منه الاختلاف المتوقع جراء خطأ المعاينة.

مؤشر I^2 هو النسبة بين الاختلاف بين الدراسات (الاختلاف الكلى مطروحاً 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 استخداماً 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 كالاتى:

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100 = \frac{291.17 - (22 - 1)}{291.17} = 92.8\%$$

وهذا يعبر عن اختلافات كبيرة جداً بين أحجام التأثير.

عدم الاستقلالية بين أحجام التأثير

فى هذا الفصل عرضنا لأهمية الاستقلالية لأحجام التأثير ومن المفضل الاعتماد على حجم تأثير واحد من الدراسة لأن حساب أكثر من حجم تأثير من نفس الدراسة يحطم مسلمة الاستقلالية. فى ما وراء التحليل يجب الاعتماد على العشوائية بقدر الامكان فى عينة الدراسات المتضمنة فى ما وراء التحليل نظراً لاستخدام اختبارات الفروض الإحصائية حول المتوسط الموزون لأحجام التأثير.

احجام التأثير المتعددة من القياسات المتعددة

أحد مصادر أحجام التأثير فى الدراسة ان الدراسة يمكن أن تتضمن مقاييس مختلفة لنفس المفهوم أو البناء فى الدراسة، فمثلاً فى الدراسات الـ 22 اعتمد أحد المؤلفين على العلاقة بين العدوان والرفض من الاقران من وجهة نظر زملائه ($r = .556$) وإيضاً من وجهة نظر المعلمين ($r = .338$)، واحد الدراسات اعتمدت على دراسة هذه العلاقة على فترات زمنية مختلفة مثلاً على مستوى تلاميذ الصف الثانى وعلى مستوى تلاميذ الصف الرابع.

فى هذه المواقف يرى Card (2012) يوجد اختياريين للحصول على حجم تأثير وحيد:

1. اختيار حجم التأثير الذى يقع فى اهتمام الدراسة التى يقوم بها الباحث وهذا فى ضوء معايير التضمين والاستبعاد التى اعتمد عليها الباحث فى الدراسة الأولى يضع الباحث معيار تقدير العلاقة فى ضوء تقديرات الاقران وليس تقدير المعلم. ومن الضرورى وضع المعيار لصناعة القرار قبل إجراء ما وراء التحليل وليس فى ضوء حجم التأثير.

2. وهو الأكثر استخداماً فى الدراسات وهو اخذ متوسط أحجام التأثير المتعددة فى نفس الدراسة لنفس المتغيرين موضع الاهتمام واستخدام متوسط أحجام التأثير كمؤشر وحيد 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 الاستدلالي الشرطي 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 خطأ المعاينة. وإذا وجدت اختلافات بين أحجام التأثير أكبر مما نتوقعه من خطأ المعاينة ونتيجة لخصائص الدراسات فإن نموذج تحليل ما وراء التحليل يسمى نموذج التأثيرات العشوائية Random effects mode ويسمى النموذج غير الشرطي 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 الصفري بمعنى اختلافات بين أحجام التأثير أكبر مما توقعناه جراء خطأ المعاينة فى الدراسات، وهذا الاختلاف يلقي 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 أخطاء 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 التأثير غير المتجانسة. كما ان تحليلات ما وراء التحليل تتعامل مع عينة صغيرة نسبياً من أحجام التأثيرات أو الدراسات تتراوح فى معظمها من 15 دراسة إلى 25 دراسة وهذا من شأنه ان تكون القوة الإحصائية لاختبار Q غير كافية لرفض الفرض الصفري 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 أخرى غير خطأ المعاينة لعينات الأفراد في الدراسات وهذه المصادر لا يمكن تحديدها. وفى هذه الحالة يتبنى الباحث نموذج التأثيرات العشوائية Random effects model حيث يفترض أن 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 الإضافى يتصرف مثل خطأ المعاينة حيث يؤدي إلى ابتعاد أو انحراف حجم التأثير المقاس عن حجم التأثير الحقيقى فى المجتمع وعليه فإن النموذج العشوائى يتعامل مع مصدرين للخطأ هما:

- خطأ معاينة على مستوى الدراسات Study Level حيث إن عينة الدراسات غير ممثلة تمثيلاً جيداً 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 تمثيلاً جيداً 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 أن 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 أن 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

البحثية- الإجراءات- حجم العينة- نوعية العينة، وغيرها) وكذلك خصائص العينة وهذه العوامل تجعل حجم التأثير منخفض أو كبير في الدراسات. وفي هذه الحالة إجراء موازنة Weighting لحجم التأثير في ضوء خطأ المعاينة للأفراد يكون مقبول ومناسب. والإجراءات الإحصائية لدراسة الاختلاف بين أحجام التأثير للدراسات المختلفة يسمى **نموذج التأثيرات المثبتة: تجزئة تباين أحجام التأثير Fixed effects model: Partitioning effect size variance**

الثالث: افترض أن التباين بين أحجام التأثير يرجع إلى مصادر أخرى غير خطأ المعاينة هي مكونة من مصدرين أحدهما عوامل منتظمة يمكن تحديدها (خصائص الدراسات) والآخر مصادر عشوائية، و لإجراءات الإحصائية لتحليل ما وراء التحليل تسمى **نموذج التأثيرات المختلطة Mixed effects model** حيث إن الاختلاف بين أحجام التأثير لا يرجع إلى خطأ المعاينة فقط بل إلى مكونين آخرين أحدهما منتظم (خصائص الدراسات) والآخر عشوائي (خصائص معاينة الدراسات من مجتمع الدراسات التي تناولت العلاقة بين المتغيرين). وهذا النموذج يتشابه مع نموذج التأثيرات المثبتة: تجزئة تباين أحجام التأثير في تحليل خصائص الدراسات كمتغيرات متفاعلة ولكنه يختلف عنه في أنه يسمح بتضمين المكون العشوائي في التحليل كتباين بواقى ليصبح في النموذج بعد تقدير المكون المنتظم، بكلمات أخرى يفترض ان تباين أحجام التأثير غير خطأ المعاينة يرجع إلى مكونين أحدهما منتظم حيث يعرض العلاقة بين خصائص الدراسات وحجم التأثير، والآخر يمثل الخطأ العشوائي نتيجة خطأ معاينة للدراسات الممثلة لمجتمع الدراسات التي تناولت المتغيرين ويجب تضمينه في الوظيفة الموازنية **Weighting function** عند تحليل حجم التأثير حيث يعاد حساب المتوسط الموزون وفترات الثقة بعد استخدام اوزان مختلفة وليس وزن واحد كما في النموذج المثبت.

نموذج التأثيرات المثبتة Fixed effects model

مصطلح التأثيرات المثبتة يشير إلى حقيقة ان مستويات المعالجة في التجربة تعتبر مثبتة ومصدر الاختلاف هي عينات أو مجموعات الأفراد والمتغير التابع. ويفترض

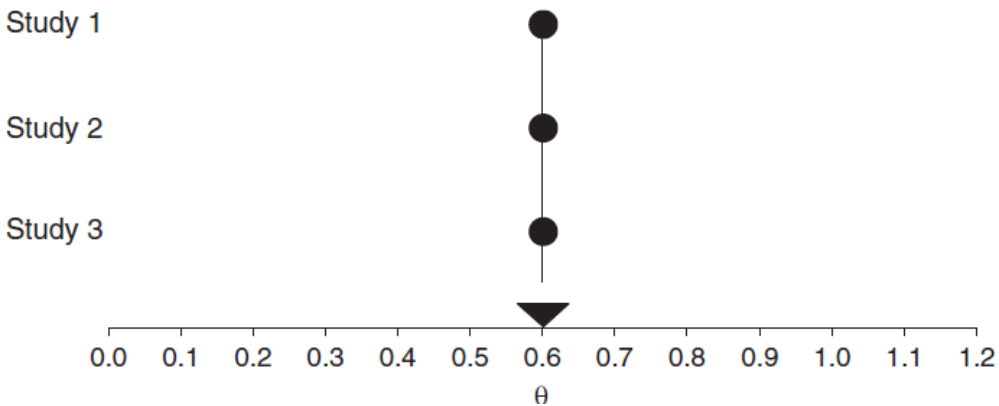
هذا النموذج ان حجم التأثير المقدر من الدراسة يناظر حجم التأثير الحقيقى فى المجتمع مع وجود خطأ عشوائى نتيجة عوامل الصدفة المرتبطة بخطأ المعاينة فى الدراسة الناتج عن الاختيار غير العشوائى للعينة من مجتمع الدراسة وايضاً نتيجة ان العينة غير ممثلة تمثيلاً جيداً لمجتمع العينة وكذلك لصغر حجم العينة. ولأسباب نظرية فربما لا يعتقد الباحث ان هذه الافتراضات مناسبة أو متوفرة لتوزيع أحجام التأثير. فعلى سبيل المثال يعتقد بوجود فروق عشوائية بين الدراسات ترجع إلى اختلافات فى الإجراءات والمواقع المختلفة التى اجريت فيها الدراسات وغيرها من العوامل التى تكون ابعد من خطأ المعاينة فى الدراسات، وربما يكون الاختلافات ترجع إلى عينة الدراسات فى ما وراء التحليل حيث لا تكون ممثلة لمجتمع الدراسات تمثيلاً جيداً وهذا يؤدى إلى خطأ معاينة خاص بالدراسات بالإضافة إلى خطأ المعاينة الخاص بعينات الأفراد فى الدراسة. ويرى (Borenstein et al. 2009) أن النموذج المثبت فى ما وراء التحليل يفترض ان أحجام التأثير للدراسات تقىس نفس حجم التأثير الحقيقى فى المجتمع بمعنى أن كل العوامل التى تؤثر فى حجم التأثير واحدة فى كل الدراسات ولذلك فإن حجم التأثير هو نفسه (مثبت) عبر كل الدراسات والاختلاف عن حجم التأثير الحقيقى يرجع إلى خطأ المعاينة.

وعلى ذلك فإن نموذج التأثيرات المثبتة يفترض ان كل الدراسات فى ما وراء التحليل تأتى من نفس المجتمع وهذا يعنى ان حجم التأثير الحقيقى يكون هو نفسه عبر كل الدراسات وفى هذه الحالة يفترض وجود تجانس بين أحجام التأثير المختلفة وان مصدر الاختلاف بين حجم التأثيرات يرجع إلى مكونات من داخل الدراسات مثل خطأ المعاينة أو خطأ القياس (Hegger, 2006). ويهتم هذا النموذج بالتباين داخل الدراسة نتيجة اختلاف المعاينات ولا يهتم بالتباين بين الدراسات

وفيفترض هذا النموذج (Kline, 2013):

- وجود مجتمع واحد متجانس من الدراسات مع حجم تأثير حقيقى وحيد.
- اختلاف حجم التأثير للدراسة عن حجم التأثير الحقيقى فى المجتمع يرجع إلى التباين داخل الدراسة بمعنى إلى خطأ المعاينة.

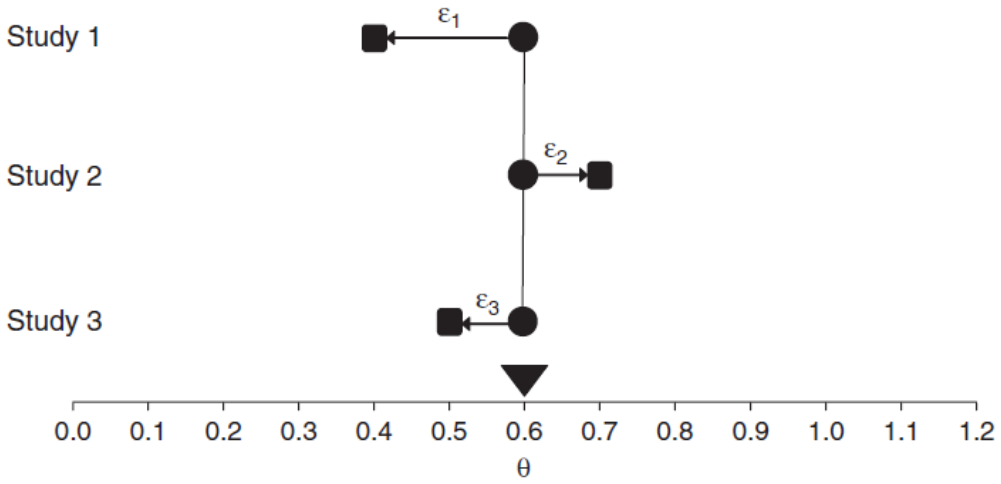
ويمكن توضيح ذلك بالشكل الاتي:



الشكل (2.22): نموذج التأثيرات المثبتة: التأثيرات الحقيقية.

فاذا افترضنا ان حجم التأثير العام أو الحقيقي في المجتمع هو 0.60 (كما معروض في المستطيل على المحور الافقى في الأسفل $\overline{0.6}$) ويرمز له بالرمز θ وحجم التأثير الحقيقي لكل دراسة معروض على الخط الراسي (دائرة). ويفترض نموذج التأثيرات المثبت ان حجم التأثير الحقيقي يجب أن يكون 0.60 وهذا معروض في الدوائر $\bullet$ التي تقع مباشرة على الخط المستقيم، ولكن هذا صعب حدوثه في الواقع التطبيقي الإمبريقي للدراسات لأن حجم التأثير يختلف من دراسة لأخرى نتيجة خطأ المعاينة في كل دراسة فلو كان هذا الخطاً مساوياً للصفر في الدراسة فإن حجم التأثير لها يكون مساوياً لحجم التأثير في الحقيقي في المجتمع.

ولكن في الممارسة العملية فإن حجم العينة في كل دراسة يكون محدد بحجم معين ولذلك فيوجد قدرًا من خطأ المعاينة يجعل حجم التأثير في الدراسة ليس هو نفسه حجم التأثير الحقيقي وقد تزداد قيمته أو تنقص. ويمكن توضيح ذلك بالشكل الاتي:



الشكل (3.22): نموذج التأثيرات المثبتة في ظل تأثير أخطاء المعاينة للدراسات.

لاحظ في هذا الشكل ان حجم التأثير الحقيقي لكل دراسة ما زال 0.60 (الدوائر) بينما حجم التأثير المقاس (المربعات) $\blacksquare$ اختلف عن حجم التأثير الحقيقي من دراسة إلى أخرى.

ففي الدراسة الأولى خطأ المعاينة $(\epsilon_1) = -0.20$ وعليه فإن حجم التأثير المقاس:

$$ES_1 = 0.60 - 0.20 = 0.40$$

وفي الدراسة الثانية خطأ المعاينة $(\epsilon_2) = 0.10$ وعليه فإن حجم التأثير المقاس:

$$ES_2 = 0.60 + 0.10 = 0.70$$

وفي الدراسة الثالثة خطأ المعاينة $(\epsilon_3) = -0.10$ وعليه فإن حجم التأثير المقاس:

$$ES_3 = 0.60 - 0.10 = 0.50$$

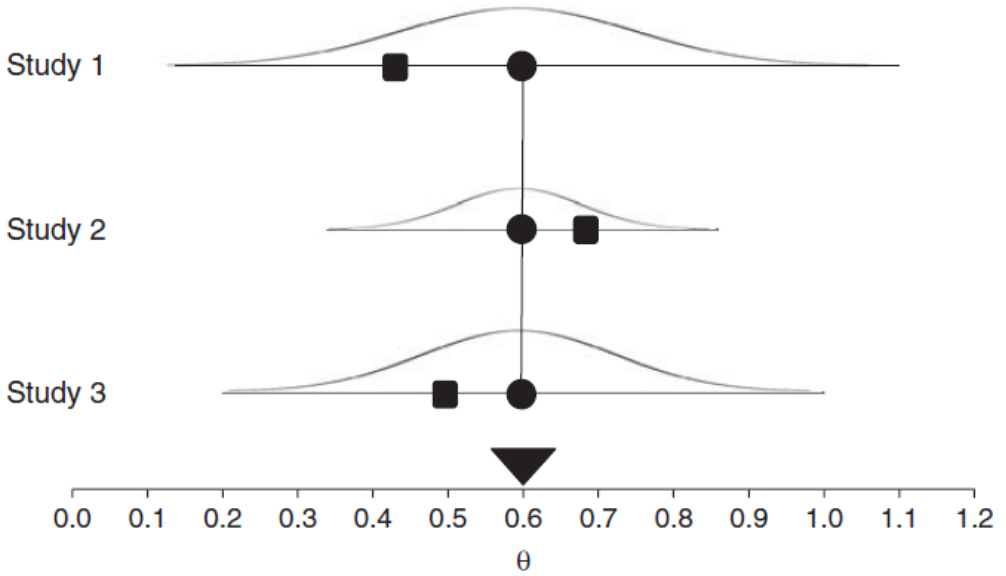
بصفة عامة فإن حجم التأثير المقاس في كل الدراسة ES_i كالاتي:

$$ES_i = \theta + \epsilon_i$$

• θ حجم التأثير العام الحقيقي في 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 كالاتى:



الشكل (4.22): التوزيع العينى لأخطاء المعاينة فى النموذج المثبت.

وفى هذا الشكل نلاحظ أن توزيع ابحام التأثير المقاسة حول حجم التأثير الحقيقى فى ضوء المنحنى الطبيعى، وفى الدراسة الأولى حيث حجم عينة صغير فمن المتوقع ان يقع حجم التأثير المقاس على جانبي حجم التأثير الحقيقى فى مدى من 0.20 إلى 1.00، فى المقابل فى الدراسة الثانية حيث حجم العينة كبير نسبياً فمن المتوقع ان يقع حجم التأثير المقاس فى مدى ضيق من 0.40 إلى 0.80.

نموذج التأثيرات العشوائية Random Effects Model

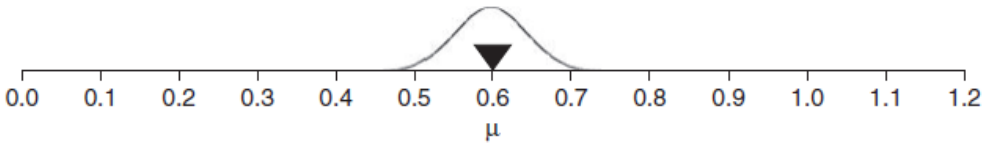
فى نموذج التأثيرات المثبتة افترضنا ان حجم التأثير الحقيقى نفسه فى كل الدراسات ولكن هذه المسلمة من الصعب تحقيقها. فلمجموعة من الدراسات المراد دمجها فى ما وراء التحليل فإنها ليست متشابهة أو متماثلة تماماً ان تعطى نفس حجم التأثير للعلاقة بين المتغيرين، فهذه الدراسات لها منهجيات وإجراءات وتصميمات واحجام عينات مختلفة بل مقاييس مختلفة، على ذلك فإن الدراسات مختلفة فى خصائصها وقيمة حجم التأثير مختلف من دراسة إلى اخرى.

فمثلاً 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 متوسطة 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 في 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.

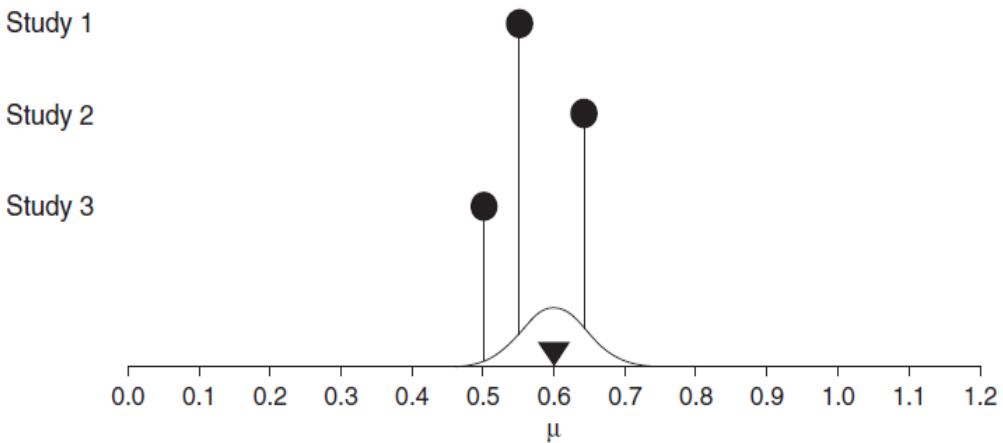
مثال: افترض أن دراسة ما وراء التحليل لثلاثة دراسات وان أحجام التأثير الحقيقية لهذه الدراسات هي:

$$\theta_1 = 0.50$$

$$\theta_2 = 0.55$$

$$\theta_3 = 0.65$$

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.

فى هذا الشكل لاحظ أن حجم التأثير الحقيقى العام فى المجتمع (μ) يساوى 0.60 (على الخط الافقى) بينما حجم التأثير الحقيقى للدراسة الأولى 0.50 بالتالى حدث ابتعاد لحجم التأثير للدراسة الأولى عن حجم التأثير الحقيقى العام فى المجتمع بحوالى 0.10 وهذا يمثل نوع من الخطأ لاحظ أن هذا الخطأ لا يرجع إلى خطأ المعاينة لأننا افترضنا ان خطأ المعاينة مساوى للصفر.

ولكن حجم العينة فى كل دراسة ليس ممثلاً للمجتمع تمثيلاً كافياً وعليه فإن خطأ المعاينة لا يساوى صفر. فاذا كان حجم التأثير الحقيقى للدراسة θ_i فإن حجم التأثير المقاس فى الدراسة ES_i يكون أكبر أو أصغر من حجم التأثير الحقيقى θ_i نتيجة خطأ المعاينة. ففى الشكل (6.22) الدراسة الثالثة لها حجم تأثير حقيقى 0.50 وخطأ المعاينة -0.10 بالتالى حجم التأثير المقدر للدراسة 0.40.

وفى الشكل (6.22) يتضح أن المسافة بين المتوسط العام الحقيقى فى المجتمع (μ) وحجم التأثير المقاس فى الدراسة يتكون من جزئين هما:

- المسافة بين المتوسط العام لحجم التأثير الحقيقى (μ) وحجم التأثير الحقيقى للدراسة θ_i ويطلق على هذا الخطأ أو الفرق زيتا ζ_i .

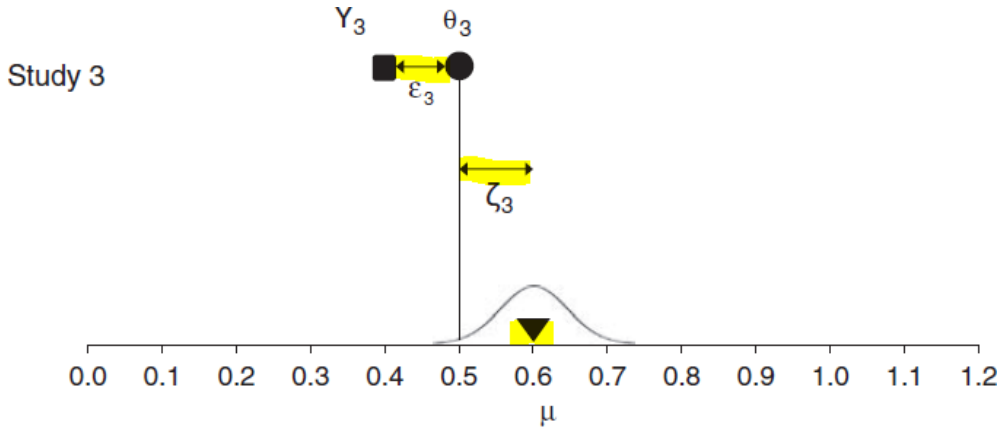
- المسافة بين حجم التأثير الحقيقى للدراسة θ_i وحجم التأثير المقاس للدراسة ES_i وهذا الفرق نتيجة خطأ المعاينة فى الدراسة ويطلق عليه الخطأ ايبسلون ε_i .

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 (زيتا) انحراف أو ابتعاد ثابت لحجم التأثير المقاس عن متوسط حجم التأثير الحقيقى.
- ε_i (ايبسلون) انحراف شرطى لحجم التأثير الحقيقى للدراسة نتيجة خطأ المعاينة.

ويمكن توضيح ذلك بالشكل الآتى (Borenstien et al., 2009):



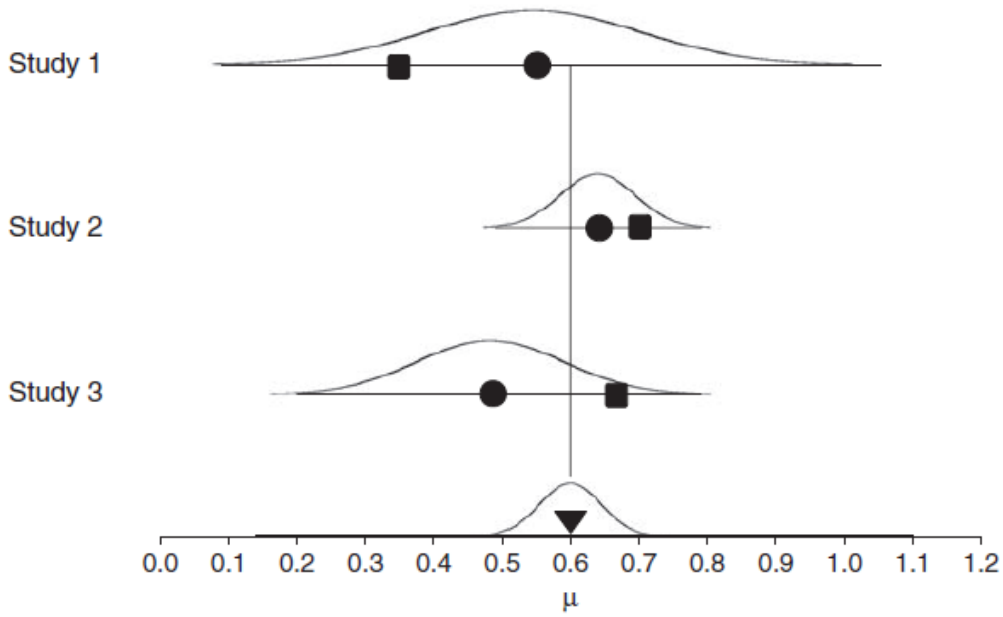
الشكل (7.22): حجم التأثير الحقيقي والمقاس فى نموذج التأثيرات العشوائية.

فالدراسة الثالثة متوسط حجم التأثير الحقيقى لمجتمع الدراسات $(\mu) = 0.60$ ، بينما حجم التأثير الحقيقى فى الدراسة $\theta_3 = 0.50$ ، 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 $(Y_3)ES_3$ 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:

$$(Y_3)ES_3 = 0.60 - 0.10(\zeta_3) - 0.10(\varepsilon_3) = 0.40$$

والمسافة من متوسط حجم التأثير الحقيقى للدراسات μ (المستطيل) إلى حجم التأثير الحقيقى للدراسة θ_i (الدوائر) يعتمد على الانحراف المعياري لتوزيع التأثيرات الحقيقية عبر الدراسات ويطلق عليه تاو τ أوتباينها τ^2 ويكون توزيعها اعتدالى كالتالى:



الشكل (8.22): توزيع أحجام التأثير الحقيقية للدراسات في نموذج التأثيرات العشوائية.

لاحظ أن أحجام التأثير الحقيقية للدراسات توزعت عشوائيًا حول متوسط التأثيرات الحقيقية في نموذج التأثيرات العشوائية بينما في نموذج التأثيرات المثبتة تم إجبار كل أحجام التأثير أن تقع على الخط الراسي.

الفروق بين نموذج التأثيرات المثبتة والعشوائية في ما وراء التحليل

توجد العديد من الفروق المفاهيمية، التحليلية عند حساب المتوسط الموزون ودلالاتها الإحصائية، والتفسيرية بين النموذجين كالاتي (Card, 2012; Kisamore & Brannick, 2008; Mengersen et al., 2013):

الفروق المفاهيمية النظرية Conceptual Differences

في النموذج المثبت الفرق بين حجم التأثير الوحيد الحقيقي في المجتمع θ وحجم التأثير المقاس في الدراسة نتيجة خطأ المعاينة ε_3 ، ويتلامس حجم التأثير الحقيقي للدراسات مع حجم التأثير الحقيقي في المجتمع، أي أن 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 τ وليس حجم تأثير وحيد فى المجتمع θ كما فى النموذج المثبت، ويفترض ان كل حجم تأثير مقياس يختلف عن متوسط مجتمع أحجام التأثير فى ضوء خطأ المعاينة على مستوى الأفراد بالإضافة إلى قيمة أخرى للتباين من مصادر أخرى 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 كالاتى:

$$S_T^2 = S_\theta^2 + S_i^2$$

- 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 نتيجة 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 (إبسيلون) انحراف شرطى لحجم التأثير الحقيقى للدراسة نتيجة خطأ المعاينة الأفراد فى الدراسات.

الفروق التفسيرية Interpretive Differences

ما وراء التحليل باستخدام النموذج المثبت يهدف إلى الوصول إلى استنتاجات شرطية Conditional Inferences حول مجموعة محددة من الدراسات المتضمنة في التحليل فقط وليس على مجتمع الدراسات التي تناولت المتغيرين بمعنى " أن متوسط حجم التأثير لـ 15 دراسة....." ولا يعمم على مجتمع الدراسات التي تناولت المتغيرين ويمكن القول إن نتائج نموذج التأثيرات المثبتة وصفية.

بينما ما وراء التحليل باستخدام نموذج التأثيرات العشوائية يهدف إلى الوصول لاستنتاجات غير مشروطة Unconditional Inferences ليست قاصرة على عينة الدراسات في التحليل بل إلى مجتمع الدراسات المحتملة التي تناولت دراسة العلاقة بين المتغيرين بمعنى أن " تراث الدراسات للعلاقة بين المتغيرين X , Y توصل إلى أن حجم التأثير أو العلاقة بين المتغيرين....."، ويمكن القول إن نتائج نموذج التأثيرات العشوائية استدلالية. ولذلك فإن نموذج التأثيرات العشوائية أكثر مناسبة في الدراسات النفسية التربوية الاجتماعية نتيجة تراكم كم هائل من الدراسات للوصول إلى استنتاجات أكثر تعميمية وليست قاصرة على عينة الدراسات في التحليل.

وعليه فإن نموذج التأثيرات العشوائية لا يفترض أن الدراسة مشتقة من نفس المجتمع كما هو الحال في نموذج التأثيرات المثبتة بل أنها تمثل مجتمع فريد في حد ذاتها، وفي هذه الحالة يوجد مصدرين للاختلافات أحدهما من الدراسة كما هو الحال في نموذج التأثيرات المثبتة والآخر من الاختلافات بين الدراسات ولذلك يسمى بنموذج عدم التجانس (Egger, 2006, Mengersen et al., 2013)

العمليات الحسابية لنموذج التأثيرات العشوائية

تناولنا في الفصل الثامن تحليل نموذج التأثيرات المثبتة من حساب المتوسط الموزون لأحجام التأثير في ضوء الخطأ المعياري، وحساب اختبار التجانس وفترات الثقة والدلالة الإحصائية باستخدام اختبار Z . أما عن العمليات الحسابية لنموذج التأثيرات العشوائية فإنها تختلف عن النموذج المثبت نظراً لوجود مكون آخر للتباين يتمثل في

المكون العشوائى ϵ_i ، وعليه فلا بد من إجراء موازنة لكل حجم تأثير فى ضوء تعديلين أحدهما خاص بالخطأ المعيارى المرتبط بخطأ المعاينة للأفراد والآخر مرتبط بالمكون العشوائى الذى ربما يرجع فى جزء منه إلى خطأ معاينة الدراسات المتضمنة فى ما وراء التحليل.

ويرى (2012) Card أن ما وراء التحليل للنموذج العشوائى يتضمن أربع خطوات عامة هى كالتالى:

أولاً: تقدير التجانس أو عدمه بين أحجام التأثير للدراسات من خلال اختبار Q.

ثانياً: تقدير التباين بين أحجام التأثير فى المجتمع.

ثالثاً: استخدام التباين لأحجام التأثير فى المجتمع للحصول على اوزان التأثيرات العشوائية Random-effects weights لأحجام التأثير المقاسة فى الدراسات.

رابعاً: استخدام اوزان التأثيرات العشوائية لتقدير متوسط التأثيرات العشوائية لأحجام التأثير والأخطاء المعيارية المرتبطة بها لاستخدامها فى حساب الدلالة الإحصائية وفترات الثقة.

1. تقدير التجانس من عدمه بين أحجام التأثيرات: يقدر التجانس من عدمه من خلال اختبار Q كالتالى:

$$Q = \sum w_i (ES_i - \overline{ES})^2 = \sum (w_i ES_i^2 - \frac{(\sum w_i ES_i)^2}{\sum w_i})$$

- 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 كاي تربيع الجدولية اتضح ان قيمتها أكبر من القيمة الجدولية لكاي تربيع، وعليه رفض الفرض الصفري 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 كالاتى (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

تأثير أكثر دقة وفى هذا التصحيح فإن حجم التأثير المصحح أو الدقيق هو دالة وظيفية نتيجة حجم التأثير المقاس وانحراف المعاينة فى كل دراسة.

وفى النموذج العشوائى فإن حجم التأثير يكون وظيفة لحجم التأثير المقاس بالإضافة إلى مكونين من عدم الدقة Imprecision وليس مكون واحد كما فى نموذج التأثيرات المثبتة هما:

- ζ_i (إساي) انحراف أو ابتعاد ثابت لحجم التأثير المقاس عن متوسط حجم التأثير الحقيقى نتيجة خطأ معاينة الدراسات فى ما وراء التحليل، ويقدر من خلال تباين مجتمع أحجام التأثيرات τ^2 .

- ε_i (زيتا) انحراف شرطى لحجم التأثير الحقيقى للدراسة نتيجة خطأ المعاينة الأفراد فى الدراسات، ويقدر من خلال الخطأ المعيارى SE. وعلى ذلك فإن اوزان التأثيرات العشوائية تتكون من تباين مجتمع أحجام التأثير بين الدراسات نتيجة عوامل مختلفة والخطأ المعيارى الخاص بكل دراسة.

وتقدر اوزان التأثيرات العشوائية كالاتى:

$$w_i^* = \frac{1}{\tau^2 + SE_i^2}$$

- w_i^* اوزان التأثيرات العشوائية.
- τ^2 تباين مجتمع أحجام التأثيرات.
- SE_i الخطأ المعيارى المرتبط بحجم التأثير لكل دراسة.

وعلى ذلك يتم تصحيح حجم التأثير لكل دراسة من خلال اوزان التأثيرات العشوائية. بالرجوع إلى جدول البيانات فى جدول (3.22) تم حساب w_i^* لكل دراسة ثم حساب حاصل ضرب وزن التأثير العشوائى w_i^* × حجم التأثير ES_i لكل دراسة.

وفى الجدول (3.22) بالنسبة للدراسة الأولى فإن التصحيح فى ضوء نموذج التأثيرات المثبتة كالاتى:

$$w_1 = \frac{1}{SE^2} = \frac{1}{(0.0693)^2} = 208.12$$

ولكن فى النموذج العشوائى تم حساب وزن جديد كدالة لكلاً من تباين مجتمع أحجام التأثير 0.0408 والخطأ المعيارى لكل دراسة $SE=0.0693$ وعليه يصبح تصحيح حجم التأثير للدراسة كالتالى:

$$w_i^* = \frac{1}{\tau^2 + SE_i^2} = \frac{1}{0.0408 + 0.0693^2} = 21.94$$

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 يميناً) 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 تقريباً 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):

$$\overline{ES} = \frac{\sum (w_i^* ES_i)}{\sum w_i^*}$$

ويتطبيق 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 تحويل حجم التأثير معامل الارتباط r إلى تحويل فيشر Z_i ثم قدر متوسط تحويلات فيشر لكل الدراسات $\overline{Z_r}$ بقيمة 0.338 ثم تعاد هذه القيمة مرة أخرى إلى متوسط معامل الارتباط وهى $\overline{r} = 0.326$. انظر نموذج التأثيرات المثبتة حيث إن قيمة $\overline{Z_r} = 0.387$ بينما $\overline{r} = 0.387$. 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 المعياري لمتوسط حجم التأثير كالاتى (Card, 2012; Lipsey & Wilson, 2001):

$$SE_{\overline{Z_r}} = \sqrt{\frac{1}{\sum(w^*)}} = \sqrt{\frac{1}{475.89}} = 0.0458$$

لاحظ أن الخطأ المعياري فى النموذج العشوائى أكبر من الخطأ المعياري فى النموذج المثبت 0.0118 (انظر الفصل الثامن).

وتقدر الدلالة الإحصائية من خلال اختبار Z (للدلالة وليس تحويل 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 إحصائياً بين العدوان ورفض الاقران.

لاحظ أن قيمة Z للنموذج العشوائى أقل بكثير من نظيرتها للنموذج المثبت 32.70 (انظر الفصل الثامن)، وهذا يرجع إلى أن الخطأ المعياري (المقام) فى النموذج العشوائى (0.0458) أكبر من الخطأ المعياري فى النموذج المثبت (0.0118) وهذا أدى إلى انخفاض قيمة Z فى النموذج العشوائى عن نظيراتها فى النموذج المثبت.

وتقدر فترات الثقة من الخطأ المعياري 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 الأدنى.
- $\overline{ES}$ متوسط حجم التأثير.
- $Z_{(1-\alpha)}$ القيمة الجدولية عند 0.05 أو 0.01 لاختبار ذو ذيلين.
- $SE_{\overline{ES}}$ الخطأ المعياري لمتوسط حجم التأثير.

ولكن قبل حساب فترات الثقة تحول قيمة متوسط تحويل فيشر إلى متوسط معامل الارتباط. ويقدر فترات الثقة CI 95% كالتالى:

$$\bar{r}_{lower} = \bar{r} - Z_{(1-\alpha)}SE_{\bar{r}} = 0.326 - 1.96 \times .0458 = .2363$$

$$\bar{r}_{upper} = \bar{r} + Z_{(1-\alpha)}SE_{\bar{r}} = 0.326 + 1.96 \times .0458 = .349$$

المقارنة بين نموذج التأثيرات المثبتة ونموذج التأثيرات العشوائية

قدم Borenstien et al.(2009) مقارنة جيدة ومستفيضة بين النموذجين فى النموذج المثبت نفترض أن حجم التأثير الحقيقى هو نفسة لكل الدراسات وأن متوسط حجم التأثير المقدر من الدراسات هو تقدير لحجم التأثير الحقيقى فى المجتمع وأن السبب الوحيد للاختلاف بين أحجام التأثير هو خطأ المعاينة (عينات الدراسات)، فى حين فى النموذج العشوائى نفترض أن حجم التأثير الحقيقى يختلف من دراسة إلى أخرى وأن الدراسات المتضمنة فى التحليل هى عينة عشوائية لأحجام التأثير للظاهرة موضع الدراسة وأن ملخص حجم التأثير هو تقدير لمتوسط أحجام التأثير. وفى النموذج المثبت مصدر الخطأ أو عدم التأكد هو نتيجة خطأ المعاينة بينما فى نموذج التأثيرات

العشوائية يوجد مصدرين للاختلاف هما خطأ المعاينة بالإضافة إلى خطأ ناتج عن الفروق بين الدراسات، وعليه فإن الأخطاء المعيارية وفترات الثقة وتباين أحجام التأثير في النموذج العشوائي أكبر أو أوسع من نظيراتها للنموذج المثبت.

أما الفرض الصفري في النموذج المثبت هو أن حجم التأثير في كل دراسة يساوي صفر بينما في النموذج العشوائي متوسط حجم التأثير لكل الدراسات يساوي صفر وبالتأكيد هذا ينعكس على عملية تعميم نتائج النموذجين.

ويفضل استخدام نموذج التأثيرات المثبتة إذا تحقق الشرطين:

- الدراسات المتضمنة في التحليل متماثلة أو متجانسة.
- الهدف من التحليل حساب حجم التأثير العام للدراسات المتماثلة دون تعميم النتائج لمجتمعات غير مجتمع عينة الدراسات.

فلو أرادت شركة أدوية التحقق من فاعلية عقار ما واستخدمت مئات المرضى من خلال المجموعة التجريبية والضابطة ولأن الشركة تجرى الدراسة على 100 مريض في نفس الوقت فالشركة تجرى عشرة دراسات في نفس الوقت على العينة وفي هذه الحالة فإن الدراسات العشرة متماثلة لأن العينة لها خصائص متشابهة ونفس المعالجة (العقار) ونفس القياسات ونفس الفريق البحثي، وعليه فإن هذه النوعية من الدراسات متماثلة أو متجانسة (الشرط الأول) وتقوم على نفس المجتمع فإن الهدف من دراسة ما وراء التحليل دراسة مدى فاعلية العقار على نفس العينة وتعمم النتائج على نفس المجتمع وليس مجتمعات مرضى أخرى (الشرط الثاني). ولكن يبدو هذا التجانس في دراسات العلوم الطبية ولكنه بعيد المنال في دراسات العلوم النفسية والاجتماعية.

وعلى النقيض من المثال السابق حيث يجمع الباحث بيانات أو دراسات من سلسلة من الدراسات في التراث البحثي لظاهرة ما لباحثين مستقلين بإجراءات وعينات ومنهجيات وبرامج مختلفة بمعنى أن الدراسات ليست متجانسة أو متكافئة ولذلك لا يمكن افتراض وجود حجم تأثير عام.

وإذا كان عددت الدراسات صغير جدًا في ما وراء التحليل فإن تقدير التباين بين الدراسات τ^2 يكون غير دقيق وعلى الرغم صلاحية استخدام نموذج التأثيرات العشوائية ولكن المعلومات المطلوبة لتنفيذ تكون ناقصة وفي هذه 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 Rosenthal & Rubin (1982) للتأثيرات المثبتة لسهولة فهمه وبساطة ولكن يؤدي إلى تقدير متحيز لحجم التأثير الحقيقي لأن البيانات الحقيقية لأحجام 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 لأحجام التأثير تولدت من مجتمعات مختلفة يؤدي 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 & Schmidt (1990, 2004) أو Field, Hedges & Vevea (1998) (2003).

راجع Hunter & Schmidt (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 نموذج 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 تطبيقي (فى: Ellis, 2010)

اهتم باحث بإجراء ما وراء التحليل للعلاقة بين ادمان الانترنت والتحصيل وجمع ثلاثة دراسات والبيانات كالاتى:

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 كالاتى:

$$M_{ES}(\bar{r}) = \frac{-0.48 \pm 0.58 + 0.05}{3} = -0.337$$

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 الدراساتين A, B.

ولكن المدخل الأكثر دقة ومنطقية وهو اعطاء وزن أو أهمية للدراسة بما يتناسب مع حجم العينة فمن غير المنطقي 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 في حجم عينها مقسوماً على حجم العينة الكلى كالآتى:

$$\bar{r} = \frac{\sum n_i r_i}{\sum n_i}$$

وعليه فإن متوسط حجم التأثير الموزون كالتالى:

$$= \frac{(80 \times -0.48) + (112 \times -0.58) + (32 \times 0.05)}{80 + 112 + 32}$$

$$\bar{r} = \frac{(-38.4) + (-65.0) + (1.6)}{224} = -0.454$$

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 يؤدي إلى انكماش حجم التأثير عن قيمته الحقيقية نتيجة 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:

$$ES_B(r)_{adjusted} = \frac{-0.58}{\sqrt{0.92}} = -0.605$$

اما الدراسة الثالثة لم يذكر الباحث معامل الثبات ألفا كرونباخ ولذلك يقدر على متوسط الثبات فى الدراستين A, B:

$$= \frac{0.70 + 0.92}{2} = 0.81$$

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 كالاتى:

$$ES(r)_{adjusted} = \frac{r}{\sqrt{\alpha_1 \alpha_2}}$$

وعليه فإن متوسط حجم التأثير الموزون المصحح من خطأ القياس أو نقص الثبات:

$$\begin{aligned} &= \frac{(80 \times -0.574) + (112 \times -0.605) + (32 \times 0.056)}{80 + 112 + 32} \\ &= \frac{(-45.92) + (-67.76) + (1.79)}{224} \\ &= -0.50 \end{aligned}$$

لاحظ أن متوسط حجم التأثير الموزون تغيرت قيمته من 0.337 إلى 0.454 نتيجة تصحيحة من خطأ المعاينة والى 0.50 نتيجة تصحيحة من خطأ المعاينة وخطأ القياس، وبالتالي 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: تحويل معامل الارتباط إلى تحويل فيشر Z_r وتحديد الدلالة الإحصائية ولكن (Hunter & Schmidt, 1990, 2004) أجروا ما وراء الحليل على قيم معامل الارتباط مباشرة دون تحويله إلى تحويل فيشر وفي هذه الحالة لا بد من حساب الخطأ المعياري المرتبطة بمتوسط حجم التأثير والخطأ المعياري لوصف انتشار التوزيع العيني لأحجام التأثير في المجتمع بمعنى أنه نوع خاص من الانحراف المعياري وفي هذا المثال فإن التوزيع العيني يتكون من ثلاثة أحجام تأثير، والبعض يفضل مصطلح التباين كتعبير عن الخطأ المعياري ولكن المصطلح الأكثر استخداماً هو الجذر التربيعي للتباين. وتباين عينة معاملات الارتباطات v_r تقدر من مجموع حاصل ضرب مربع الفرق بين كل حجم التأثير للدراسة ومتوسط حجم التأثير مضروباً في حجم العينة ثم يقسم الناتج على حجم العينة الكلي كالتالي (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 المعياري كالتالي:

$$SE_{\bar{r}} = \sqrt{\frac{v_r}{K}} = \sqrt{\frac{0.51}{3}} = 0.130$$

وتقدر الدلالة الإحصائية في ضوء اختبار تحويل معامل الارتباط إلى الدرجة المعيارية Z كالتالي:

$$Z = \frac{\bar{r}}{SE_{\bar{r}}}$$

$$= \frac{0.500}{0.130} = 3.85$$

وقيمة Z الجدولية لمستوى دلالة إحصائية 0.05 لاختبار ذو ذيلين 1.96 ولمستوى 0.01 هي 2.58 ، وبما أن قيمة Z المحسوبة أكبر من قيمتها الجدولية، وعليه يرفض الفرض الصفري بالتالى فإن النتائج دالة إحصائياً.

المدخل الثانى: تقدير فترات الثقة: تقدر 95% فترات الثقة من المعادلات الآتية:

$$CI_{95\% \text{ lower}} = \bar{r} - Z_{\alpha/2} SE_{\bar{r}} = -.500 - (1.96 \times .13) = -.755$$

$$CI_{95 \text{ upper}} = \bar{r} + Z_{\alpha/2} SE_{\bar{r}} = -.500 + (1.96 \times .13) = -.245$$

وبما أن قيمة المعلم فى الفرض الصفري هي صفر، وبما أن القيمة صفر لا تقع بين حدى الثقة الأدنى والأعلى، وعليه فإن النتائج دالة إحصائياً.

زيادة الخطأ المعيارى يجعل فترات الثقة أكثر اتساعاً 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 كالاتى:

$$\begin{aligned} Q &= \sum (n_i - 1) (r_i - \bar{r})^2 \\ &= ((80 - 1) \times (-.574 - -.50)^2) + ((112 - 1) \times (-.650 - -.500)^2) \\ &\quad + ((32 - 1) \times (.056 - -.500)^2) \\ &= (79 \times .005) + (111 \times .011) + (31 \times .309) \\ &= 0.395 + 1.221 + 9.579 \\ &= 11.195 \end{aligned}$$

وتختبر الدلالة الإحصائية من خلال اختبار كاي تربيع بدرجات حرية K-1 حيث K عدد أحجام التأثير، وبما أن قيمة اختبار كاي تربيع لدرجات حرية 2 ومستوى دلالة إحصائية 0.05 هي 5.991 ، بالتالي فإن قيمة Q المحسوبة أكبر من 5.991، وعليه نرفض الفرض الصفري 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 لبياننا هذا المثال لا يمكن إجراء تحليل التفاعل نظراً 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 النتيجة 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 كالاتى:

الأول: يتم تحويل أحجام التأثير الخام r إلى تحويل فيشر Z_r كالاتى:

$$Z_r = \frac{1}{2} \ln \frac{1+r}{1-r}$$

ويتم تقدير تباين الخطأ المعيارى كالاتى:

$$v_i = \frac{1}{n_i - 3}$$

ووزن الدراسة w_i مقلوب تباين الخطأ المعيارى كالاتى:

$$w_i = n_i - 3$$

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:

$$\begin{aligned}\bar{Z}_r &= \frac{\sum w_i z_i}{\sum w_i} = \frac{\sum (-0.48 \times 77) + (-0.58 \times 109) + (0.05 \times 29)}{77 + 109 + 29} \\ &= \frac{-111.03}{215} = -0.52\end{aligned}$$

وتباين المتوسط الموزون:

$$v = \frac{1}{215} = 0.005$$

والخطأ المعيارى للمتوسط كالآتى:

$$SE_{\bar{Z}} = \sqrt{0.005} = 0.071$$

وحساب اختبار Z (ليس تحويل فيشر) للدلالة الإحصائية كالآتى:

$$Z = \frac{\bar{Z}_r}{SE_{\bar{Z}}} = \frac{-0.52}{0.071} = -7.32$$

ووفترات الثقة كالآتى:

$$CI_{95\% \text{ lower}} = \bar{Z} - Z_{\alpha/2} SE_{\bar{Z}} = -0.52 - (1.96 \times 0.071) = -0.66$$

$$CI_{95\ upper} = \bar{Z} + Z_{\alpha/2} SE_{\bar{Z}} = -0.52 + (1.96 \times 0.071) = -0.38$$

ويتم رفض الفرض الصفري بمعنى وجود علاقة ارتباطية سالبة ودالة إحصائية بين التحصيل وادمان الانترنت.

واختبار الاختلاف بين أحجام التأثير كالاتى:

$$Q = \sum w_i (Z_i - \bar{Z})^2 = wZ^2 - \frac{(wZ)^2}{w}$$

$$= 68.97 - \frac{(-111.03)^2}{215} = 11.63$$

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 (نتيجة خطأ المعاينة) والتباين بين الدراسات τ^2 وتقدر قيمة τ^2 كالاتى:

$$\tau^2(v_\theta) = \frac{Q - (K - 1)}{(\sum w_i) - \frac{(\sum w_i^2)}{(\sum w_i)}}$$

$$= \frac{11.63 - (3 - 1)}{215 - \frac{18651}{215}} = \frac{9.63}{128.25} = 0.075$$

حساب الاوزان أو مقلوب تباين الخطأ المعيارى كالاتى:

$$w_i^* = \frac{1}{v_i^*}$$

حيث v_i^* كالاتى:

$$v_i^* = (v_i + 0.075)$$

$$= (0.005 + \text{خطأ المعاينة للعينات})$$

$$= 0.08$$

ويتم حساب المتوسط الموزون كالتالى:

$$\bar{z}_r^* = \frac{\sum w_i^* z_i}{\sum w_i^*} = \frac{-13.34}{32.34} = -0.41$$

وتباين المتوسط الموزون كالتالى:

$$v_i^* = \frac{1}{\sum w_i^*} = \frac{1}{32.34} = 0.31$$

والخطأ المعياري للمتوسط الموزون لمعامل الارتباط المحور فى نموذج التأثيرات العشوائية:

$$SE_{\bar{z}_r^*} = \sqrt{0.031} = 0.176$$

اختبار Z للدلالة الإحصائية كالتالى:

$$Z = \frac{\bar{z}_r^*}{SE_{\bar{z}_r^*}} = \frac{0.41}{0.176} = 2.33$$

لاحظ أن الاختبار دال إحصائياً عند 0.05 وليس دال إحصائياً عند مستوى دلالة إحصائية 0.01 وهذا يختلف عن نتائج النموذج المثبت.

وفترات الثقة:

$$CI_{95\% \text{ lower}} = \bar{Z}_r^* - Z_{\alpha/2} SE_{\bar{Z}_r^*} = -0.41 - (1.96 \times 0.176) = -0.66$$

$$CI_{95\% \text{ upper}} = \bar{Z}_r^* + Z_{\alpha/2} SE_{\bar{Z}_r^*} = -0.41 + (1.96 \times 0.176) = -0.38$$

ولتفسير النتائج يجب تحويل تحويل فيشر إلى معامل الارتباط كالتالى:

$$\bar{r} = \frac{e^{2\bar{Z}} - 1}{e^{2\bar{Z}} + 1}$$

بالتعويض بقيمة $\bar{Z}$ فى النموذج المثبت كالتالى:

$$\bar{r} = \frac{e^{2\bar{z}} - 1}{e^{2\bar{z}} + 1} = \frac{e^{2 \times -0.52} - 1}{e^{2 \times -0.52} + 1} = -0.48$$

بالتعويض بقيمة $\bar{z}_r^*$ في النموذج العشوائي:

$$\bar{r} = \frac{e^{2\bar{z}} - 1}{e^{2\bar{z}} + 1} = \frac{e^{2 \times -0.41} - 1}{e^{2 \times -0.41} + 1} = -0.39$$

وعليه فإن نتائج ما وراء التحليل في ضوء معامل الارتباط كالاتي:

النموذج المثبت: $-0.48(CI_{95} : -0.57, -0.36)$

النموذج العشوائي: $-0.39(CI_{95} : -0.64, -0.07)$

مقارنة بين المداخل

تم تحليل نتائج الدراسات الثلاثة باستخدام مدخل (Hunter & Schmidt (2004) ومدخل Hedges et al ويوجد فروق مختلفة بين المدخلين وفيما يلي معادلات المدخلين:

الجدول (6.22): المعادلات البديلة المستخدمة في المدخلين (Ellis, 2010).

Output	Generic	Hunter and Schmidt		Hedges et al.	
		d	r	d	r
Weighted mean ES	$\bar{ES} = \frac{\sum w_i ES_i}{\sum w_i}$	$\bar{d} = \frac{\sum w_i d_i}{\sum d_i}$	$\bar{r} = \frac{\sum n_i r_i}{\sum n_i}$	$\bar{d} = \frac{\sum w_i d_i}{\sum w_i}$	$\bar{z} = \frac{\sum w_i z_i}{\sum w_i}$
Variance of sample ESs	$v_i = \frac{1}{\sum w_i}$	$v_{d_i} = \frac{\sum w_i (d_i - \bar{d})^2}{\sum w_i}$	$v_{r_i} = \frac{\sum n_i (r_i - \bar{r})^2}{\sum n_i}$	$v_i = \frac{1}{\sum w_i}$	$v_i = \frac{1}{\sum w_i}$
z score	$z_i = \frac{ \bar{ES} }{SE_{\bar{ES}}}$	$z_i = \frac{ \bar{d} }{SE_{\bar{d}}}$	$z_i = \frac{ \bar{r} }{SE_{\bar{r}}}$	$z_i = \frac{ \bar{d} }{SE_{\bar{d}}}$	$z_i = \frac{ \bar{z} }{SE_{\bar{z}}}$
Homogeneity statistic	$Q = \sum w_i (ES_i - \bar{ES})^2$	$\chi^2_{k-1} = \sum \frac{(n_i - 1)(r_i - \bar{r})^2}{(1 - \bar{r}^2)^2}$		$Q = \sum w_i (d_i - \bar{d})^2$	$Q = \sum w_i (z_i - \bar{z})^2$

• ES حجم التأثير.

• $\overline{ES}$ متوسط حجم التأثير (r, d, zr).

• K عدد الدراسات أو أحجام التأثير.

• W_i وزن الدراسة.

• SE الخطأ المعياري.

مصدر المعادلات: (Hedges & Vevea, 1998; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001).

والجدول يعرض الفروق بين المدخلين في أربع معادلات أساسية هي متوسط حجم التأثير الموزون، التباين الخطأ المعياري، اختبار Z للدلالة، واختبار تقدير الاختلاف بين أحجام التأثير. فالصيغة العامة أو المعيارية Generic تم عرضها في العمود الأول أما الأعمدة الباقية توضح صيغ حسابات مداخل Hedges & Vevea (1998) و Hunter & Schmidt (2004). والفروق بين هذه المداخل كالتالي:

الأول: معامل الارتباط أو حجم التأثير بين المدخلين حيث حول Hedges معاملات الارتباط إلى تحويلات فيشر Z_r في حين اعتمد Hunter & Schmidt على معاملات الارتباطات مباشرة دون تحويل. فالارتباطات المحورة تصحح متوسط معامل الارتباط من التحيز السالب الصغير ولكن في نفس الوقت تسهم بقدر صغير من التحيز الموجب للنتائج. والاختيار بين تحليل معامل الارتباط 0.454- (خطأ المعاينة) أو 0.50- (خطأ المعاينة وخطأ القياس) وتحليل تحويلات معامل الارتباط (المثبت 0.48- والعشوائي 0.39-) وهذا تحيز سالب يعتمد على مدى تفضيل الباحث للتحيزات الموجبة والسالبة المحدودة للنتائج.

الثاني: التباين في البيانات الذي بدوره يؤثر على الأوزان التي تعطى للتقديرات وتستخدم في حساب تباين المعاينة والاختفاء المعيارية و فترات الثقة. فعند دمج معاملات الارتباط في ضوء مدخل H.S يكون قائم على حجم العينة لكل دراسة بينما الأوزان في ضوء مدخل Hedges في ضوء N-3 ، ولكن الشئ المفيد في مدخل H.S هو إجراء تصحيحات لأحجام التأثير في ضوء محددات الدراسة مثل نقص الثبات والصدق والتصنيفية الثنائية بينما في مدخل Hedges للنموذج العشوائي 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:

$$-0.454(CI_{95} : - 0.693 \text{ to } - 0.215)$$

مدخل Hunter & Schmidt المصحح فى ضوء تصحيح من نقص الثبات وخطأ المعاينة:

$$-0.500 (CI_{95} : - 0.755 \text{ to } - 0.245)$$

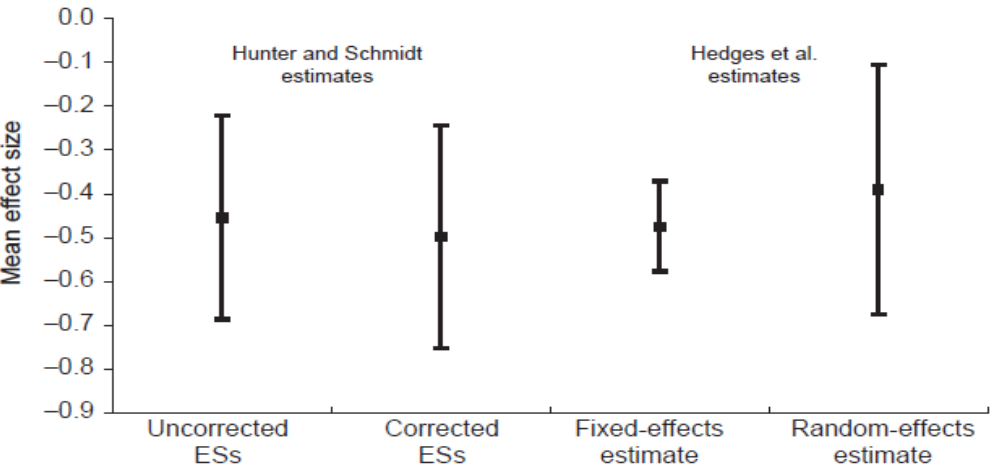
مدخل Hedges et al (التأثيرات المثبتة):

$$-0.478 (CI_{95} : - 0.572 \text{ to } - 0.365)$$

مدخل Hedges et al (التأثيرات العشوائية):

$$-0.391 (CI_{95} : - 0.639 \text{ to } - 0.068)$$

ويمكن عرض المتوسطات للمداخل الاربعة فى الشكل الاتى(Ellis(2010):



الشكل (9.22): عرض متوسط أحجام التأثير فى المداخل الاربعة.

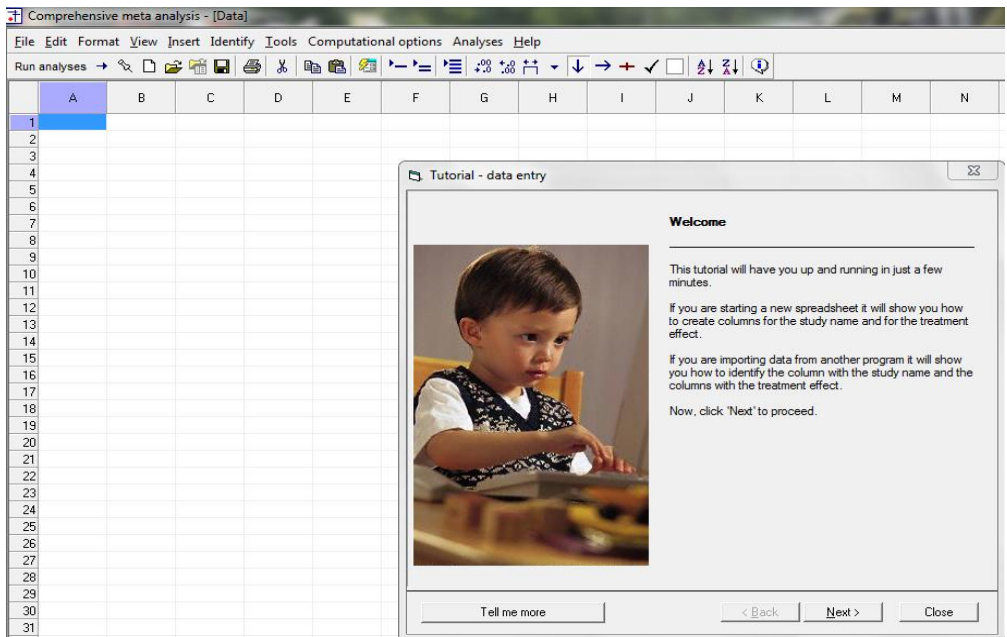
ويلاحظ الاختلاف الواضح بين المتوسطات فى المداخل الاربعة وتساءل Ellis (2010) ماهى النتائج الاكثر دقة؟ هذا السؤال نال اهتمام كثير من الباحثين امثال (Field, 2005; Hall & Brannick,2002; Schulze, 2004) ، فى ضوء دراسة المحاكاة لـ (Field 2005) توصل إلى ان طرق Hedges وزملائه تنتج

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 (2005) Field بان نضع معظم جهودنا فى تحليلات ما وراء التحليل فى ضوء مدخل Hunter & Schmidt وفى نهاية هذه المناقشة نؤكد على ما أشار إليه (2004) Schulze بان بعض المداخل أفضل من الأخرى لمهام مختلفة ولكن الطريقة الوحيدة الأفضل هى مجموعة من الإجراءات يجب توليفها. وقام (2002) Hall & Brannick بمقارنة نموذج التأثيرات العشوائية فى ضوء مدخل (1998) Hedges & Vevea و (1990) Hunter & Schmidt خصوصاً فيما يخص بمدى جدوى تحويل 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 Mark Lipsey و Matthias Egger و Cooper وغيرهم. ويتم تنزيل البرنامج على جهاز الكمبيوتر على سطح المكتب أو فى قائمة كل البرامج.

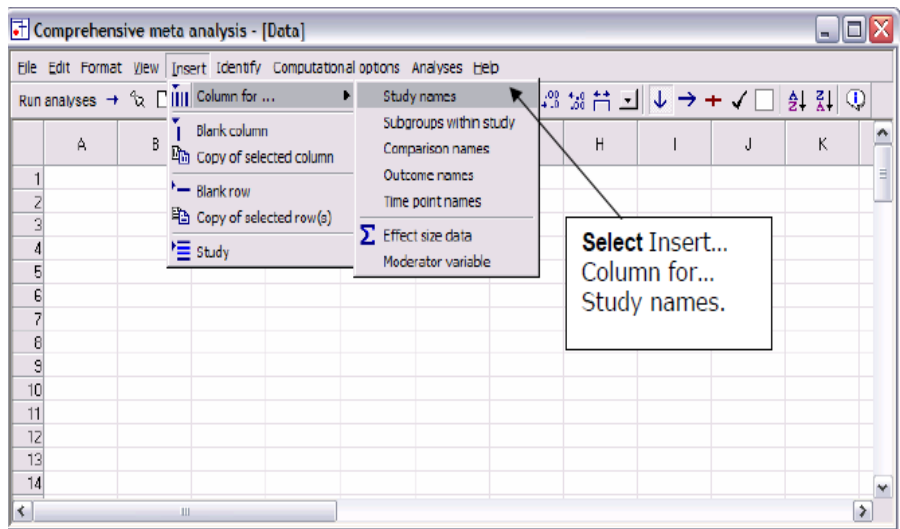
افتح البرنامج من خلال Start ثم All Programs ثم اضغط Comprehensive Meta-Analysis تظهر الشاشة الآتية:



• اضغط Close.

الجزء الأول: إدخال البيانات الأساسية والتحليل
 لإدخال البيانات إنشاء عمود لأسماء الدراسات كالتالي:

Insert.....Column for..... Study names

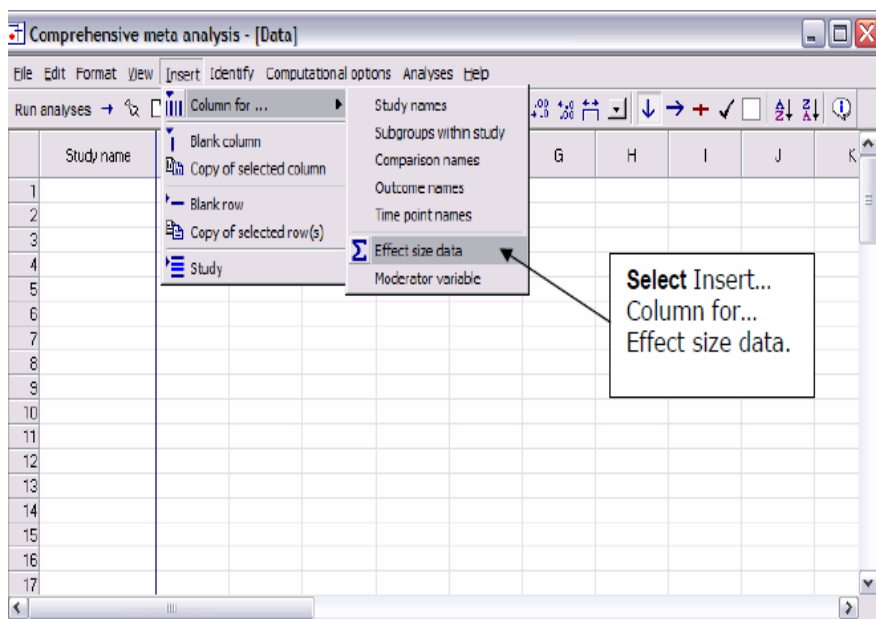


يقوم البرنامج بحشر أسماء الدراسات أو مؤلف الدراسات كالتالي:

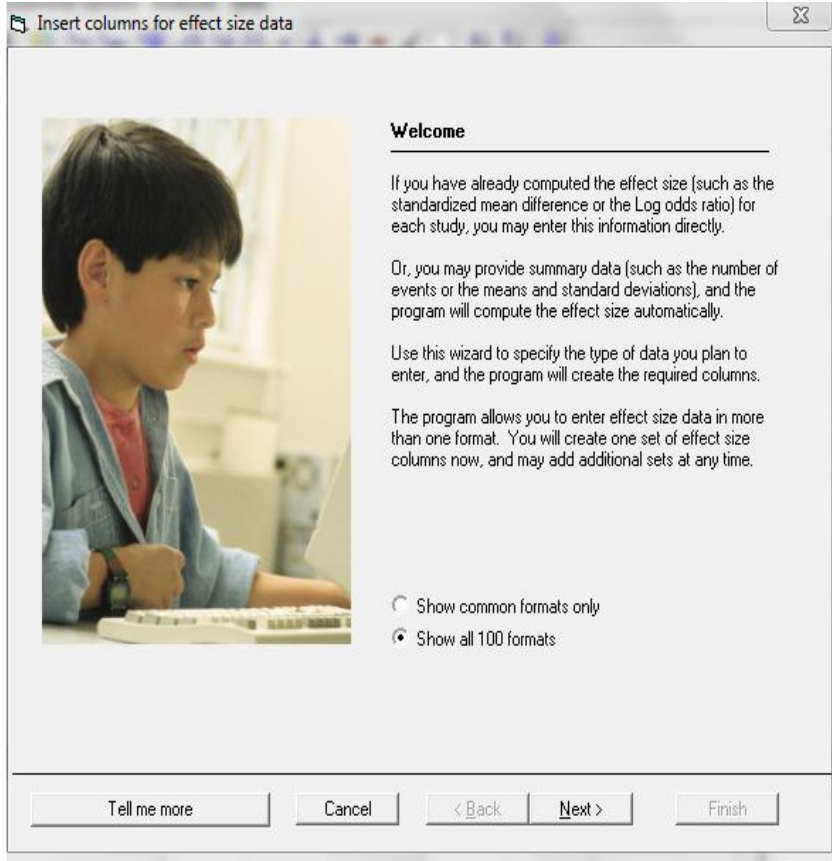
Comprehensive meta analysis - [Data]							
File Edit Format View Insert Identify Tools Computational options Analyses Help							
Run analyses →							
	Study name	B	C	D	E	F	G
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

إنشاء أعمدة لبيانات حجم التأثير كالتالى:

Insert.....Column for..... Effect Size data



تظهر الشاشة الآتية:



هذه الشاشة تمدنا بخيارات لإدخال بيانات حجم التأثير كالتالي:

الأول: لو قدر الباحث حجم التأثير مباشرة مثل مؤشرات فروق المتوسطات المعيارية أو لو غاريتم Odds ratio لكل دراسة فادخلها مباشرة في عمود حجم التأثير.

الثاني: 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 المطلوبة لذلك. ويسمح البرنامج بإدخال البيانات بأشكال عديدة.

• اضغط على اختيار Next > تظهر الشاشة الآتية:


Insert columns for effect size data





Types of studies included

On this panel, select the type of studies to be included in this meta analysis. This controls the types of data entry options to be displayed on the next panel.

If unsure, select the first option, which is appropriate for most analyses. You will be able to return to this panel and change the selection.

- ☒ Comparison of two groups, time-points, or exposures (includes correlations)
- ☐ Estimate of means, proportions or rates in one group at one time-point
- ☐ Generic point estimates
- ☐ Generic point estimates, log scale

Tell me more

Cancel

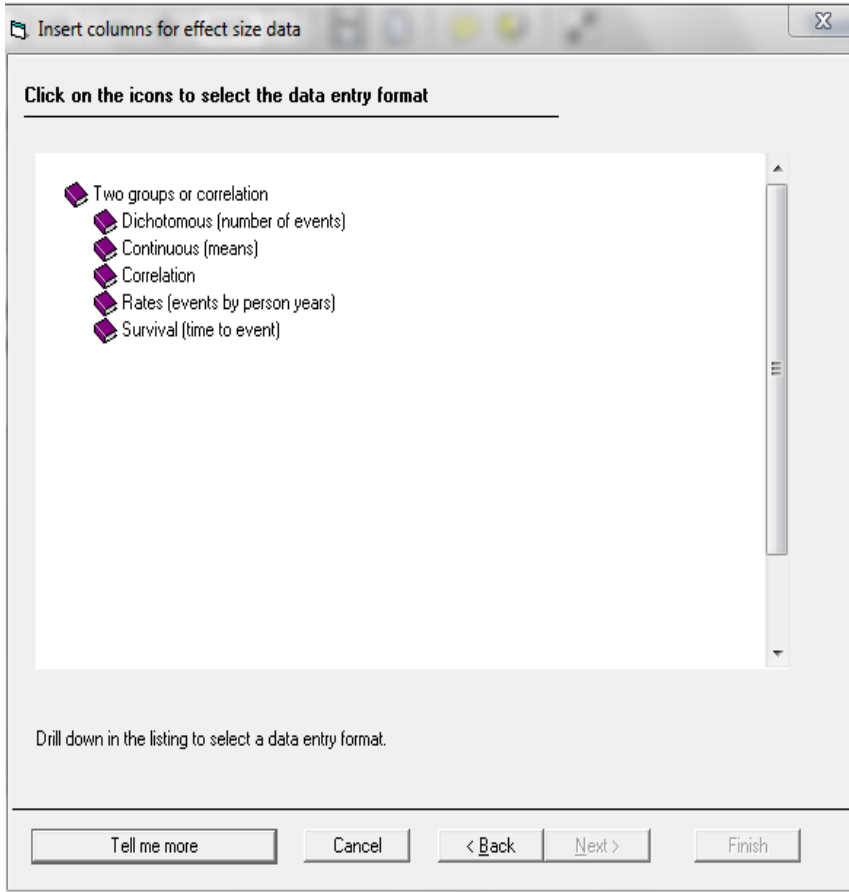
< Back

Next >

Finish

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:

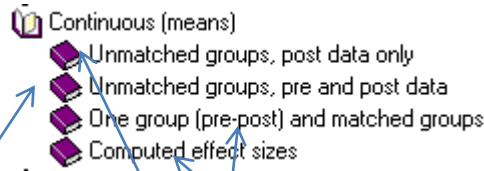
- Dichotomous (number of events)
 - Unmatched groups, prospective (e.g., controlled trials, cohort studies)
 - Matched groups, prospective (e.g., crossover trials or pre-post designs)
 - Unmatched groups, retrospective (e.g., case control studies)
 - Computed effect sizes

– المجموعات غير المرتبطة Unmatched groups وتحتها تتضمن الاحداث وحجم العينة في كل مجموعة، وعدم الأحداث والعينة في كل مجموعة، والاحداث أو عدم الاحداث في كل مجموعة، واختبار كاي تربيع وحجم العينة.

– المجموعات المترابطة Matched groups

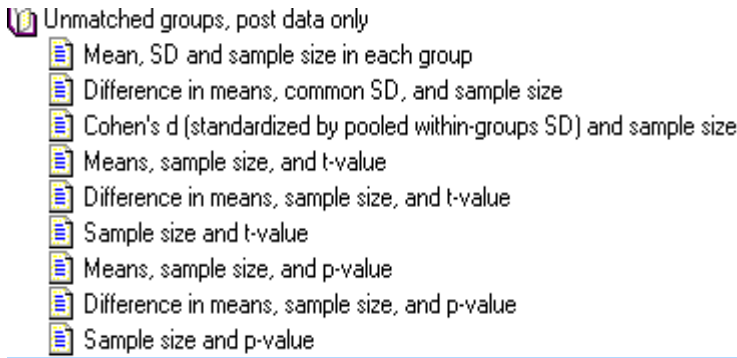
– أحجام التأثير المحسوبة.

- النواتج المتصلة Continuous: وتتضمن:

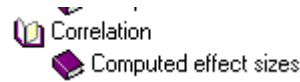


وتتضمن المجموعات المستقلة للبيانات البعدية فقط، بيانات المجموعة غير المتزاوجة وقياس قبلي وبعدي، وبيانات المجموعة الواحدة قبلي وبعدي والمجموعات المتزاوجة، وحجم التأثير.

وإذا ضغط الباحث على اختيار المجموعات المستقلة وقياس بعدي فقط تظهر الشاشة الخيارات الآتية:

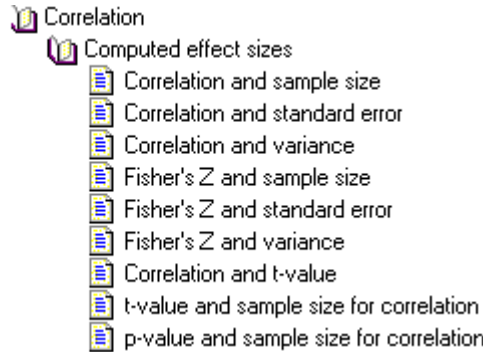


وعلى الباحث ان يختار شكل البيانات المراد ادخالها سواء المتوسط والانحراف المعياري SD وحجم العينة في كل مجموعة، أو مؤشر Cohen's d وحجم العينة أو حجم العينة وقيمة P الاحتمالية وغيرها.



• معامل الارتباط: قيم معامل الارتباط مباشرة. بالضغط على

Computed effect size تظهر اختيارات إدخال معامل الارتباط كالاتي:



• التقديرات:

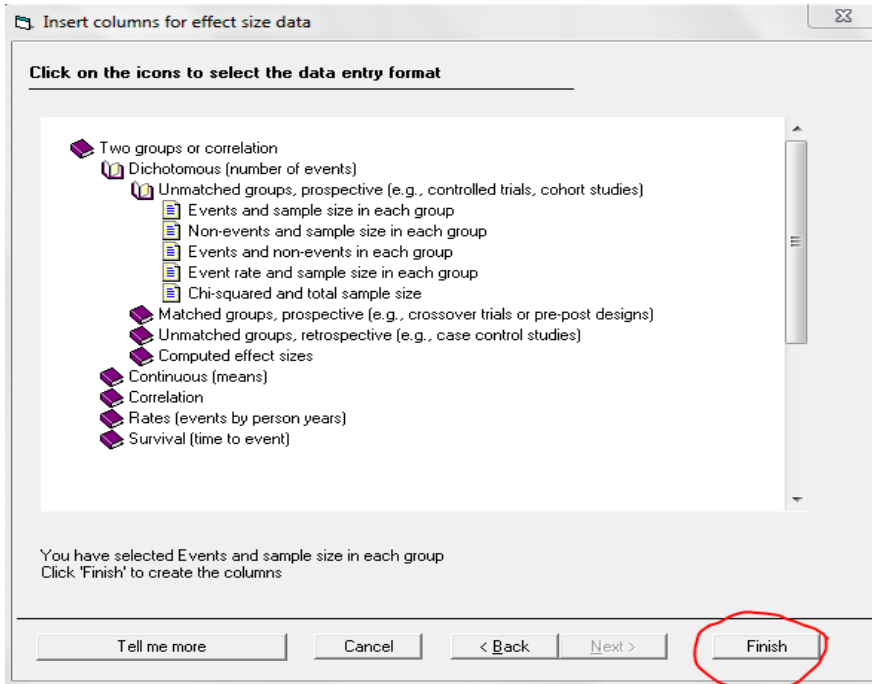
- Rates (events by person years)
 - Unmatched groups
 - Computed effect sizes

• البيانات الطولية:

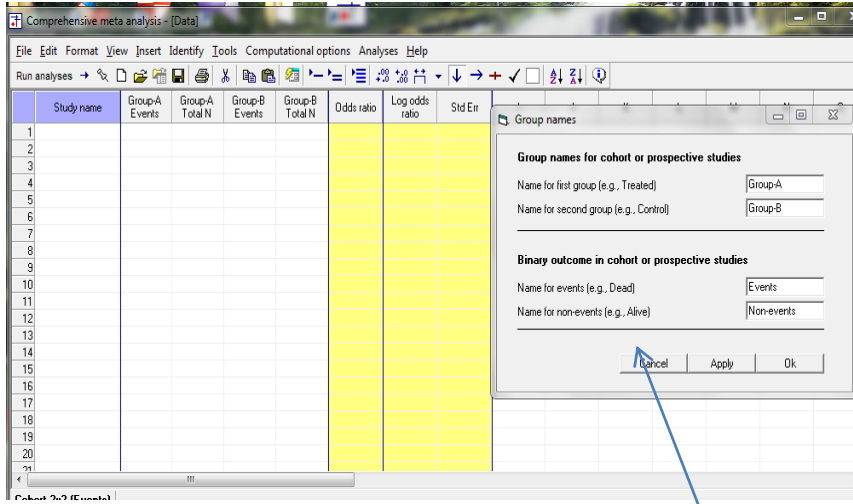
- Survival (time to event)
 - Computed effect sizes

فاذا اهتم الباحث بدراسة ما وراء التحليل لدراسات تجريبية حيث المتغير المستقل والتابع تصنيفي ثنائى كما فى الدراسات الطبية وفى هذه 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:



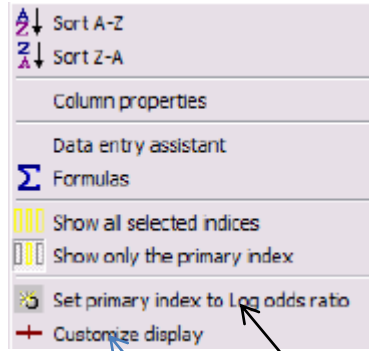
وفي المربع الفرعي يمكن تغيير مسمى المجموعات فبدلاً من Group A يمكن تسميتها بالمجموعة المعالجة Treated والمجموعة Group B يمكن تسميتها بالمجموعة الضابطة Control ثم اضغط Apply تظهر الشاشة الآتية:

	Study name	Treated Events	Treated Total N	controlled Events	controlled Total N	Odds ratio	Log odds ratio	Std Err
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								

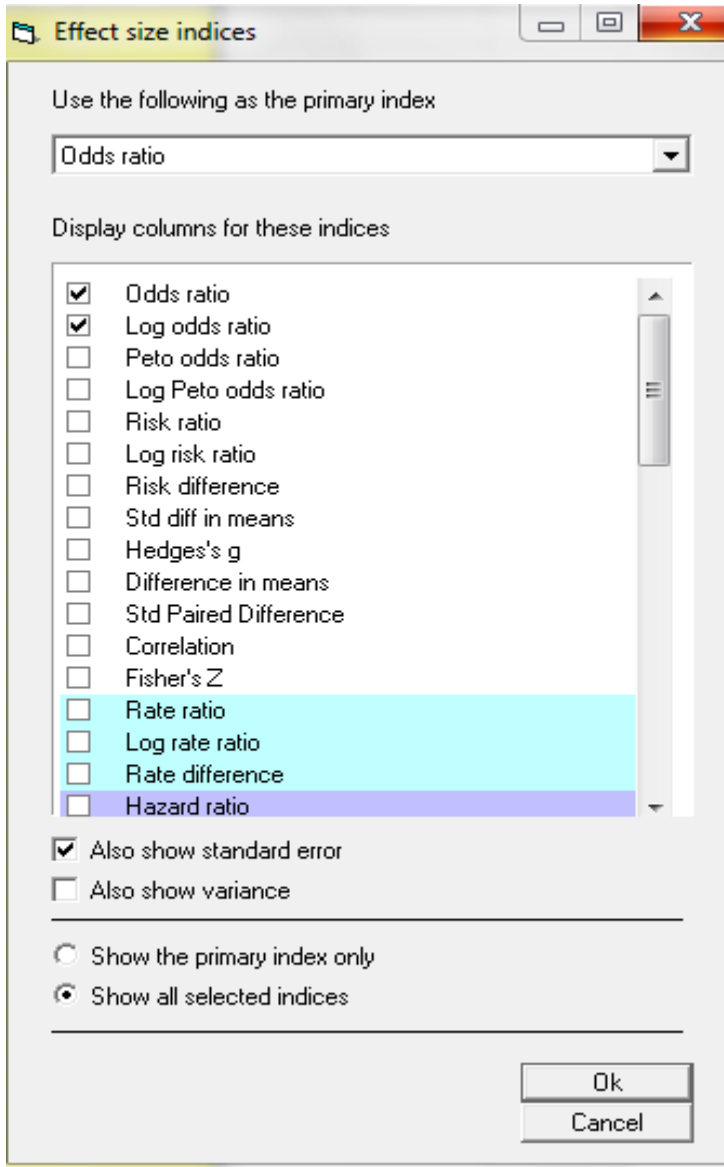
يحسب البرنامج حجم التأثير ويعرضها في الأعمدة الثلاثة.

ويمكن استعراض مؤشرات أخرى لحجم التأثير من خلال:

- اضغط كليك يمين على عمود Odds ratio تظهر قائمة كالآتي:



- اضغط set primary index أو Customized Display تظهر المؤشرات الآتية:



- ابدأ بادخل البيانات كما في الشكل الآتي واضغط على خانة Odds ratio يقوم بحساب مؤشرات حجم التأثير ويعرضها في الأعمدة الصفراء الثلاثة:

Comprehensive meta analysis - [Data]								
File Edit Format View Insert Identify Tools Computational options Analyses Help								
Run analyses →								
	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								

ولإظهار معادلات حساب مؤشرات حجم التأثير اضغط مرتين بالموس على خانة أو المربع 0.391 في عمود Odds ratio تظهر الشاشة الفرعية الآتية:

Data entry assistant

Data entry Odds ratio

Starting with
Cells in 2x2 table

Where cells are given as
A = Treated Events
B = Treated Total N - Treated Events
C = controled Events
D = controled Total N - controled Events

A = 4
B = 123 - 4 = 119
C = 11
D = 139 - 11 = 128

LogOddsRatio = Log((A * D) / (B * C))
LogOddsVariance = (1 / A + 1 / B + 1 / C + 1 / D)
LogOddsSE = Sqr(LogOddsVariance)
OddsRatio = Exp(LogOddsRatio)

LogOddsRatio = Log((4 * 128) / (119 * 11)) = -0.939
LogOddsVariance = (1/4 + 1/119 + 1/11 + 1/128) = 0.357

< Home >

ثم يقوم الباحث بإدخال بيانات دراسة جديدة ويقوم بالضغط على مربع إحصاء Odds ratio لحساب مؤشرات حجم التأثير كالآتي:

Comprehensive meta analysis - [Data]								
File Edit Format View Insert Identify Tools Computational options Analyses Help								
Run analyses →								
	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 كالاتى:

Comprehensive meta analysis - [Data]								
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
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 كالاتى:

Comprehensive meta analysis - [Analysis]

File Edit Format View Computational options Analyses Help

← Data entry ← Next table ← High resolution plot Select by ... + Effect measure: Odds ratio ←

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 الأخير 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 أى ان العلاقة بين المتغيرين دالة إحصائياً 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. وهذه الشاشة تتضمن عدة ايقونات منها:

عرض ملخص للنتائج

- اضغط Next table تظهر الجدول الاتى:

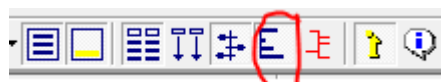
Comprehensive meta analysis - [Analysis]						
File Edit Format View Computational options Analyses Help						
← Data entry ↗ Next table High resolution plot Select by ... + Effect measure: Odds ratio						
Model	Effect size and 95% interval				Test of null (2-Tail)	
Model	Number Studies	Point estimate	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

فى هذا الجدول تم عرض اختبار عدم التجانس بين الدراسات Q وكانت قيمة 95.60 ويتم اختبار دلالتها فى ضوء اختبار كاي تربيع لدرجات حرية (9-1=8) ومستوى دلالة 0.50 وهى دالة إحصائياً حيث $P=0.000$ وعليه يوجد اختلاف بين أحجام التأثير بالتالى فإن استخدام النموذج المثبت غير مناسب والأفضل الاعتماد على نتائج نموذج التأثيرات العشوائية.

عرض اوزان الدراسات

يمكن عرض اوزان الدراسات من خلال الضغط على ايقونة:



أعرض إحصائيات الدراسات المفردة.

تعرض الإحصائيات الملخصة.

تعرض الإحصائيات الأساسية المخفية.

تعرض الاعداد غير المرئية.

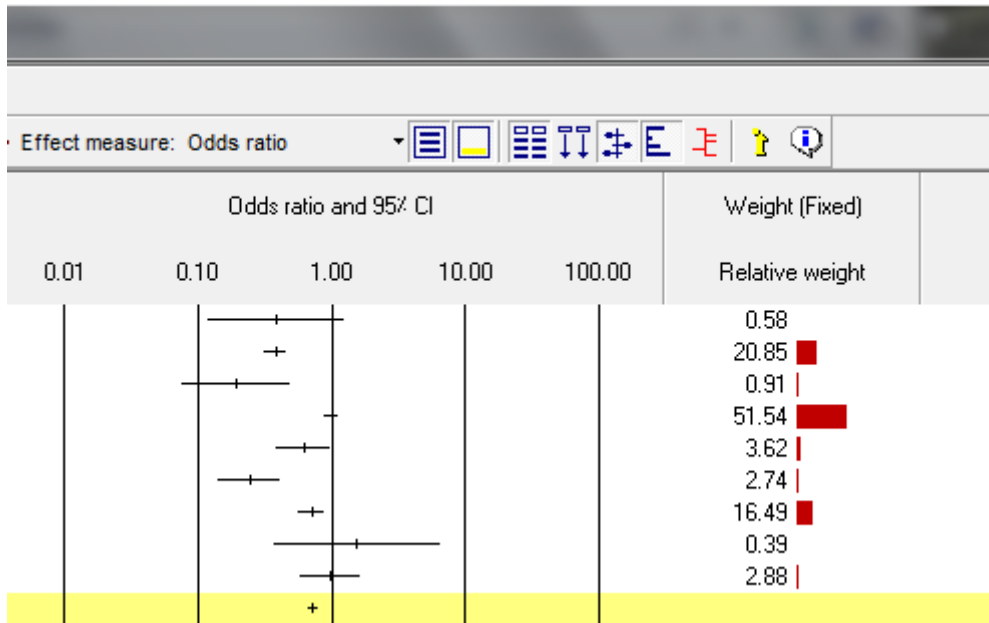
تعرض شكل Forest.

تعرض اوزان الدراسات.

تعرض البواقي.

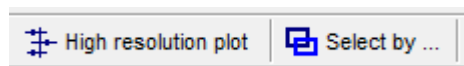
تعرض تفاصيل.

وبالضغط على  تعطى الاوزان النسبية للدراسات كالاتى:

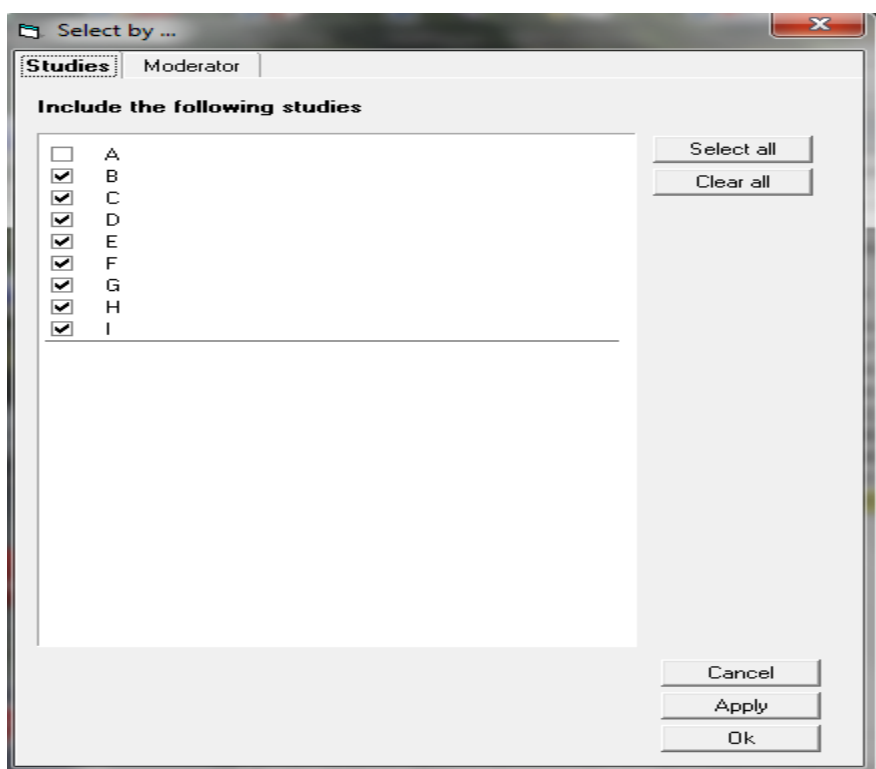


عرض نتائج ما وراء التحليل مع حذف إحدى الدراسات (تحليل الحساسية)

- اضغط على Select case:



يعطى الجدول الآتى:

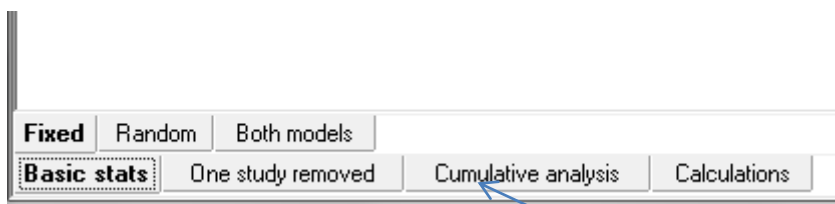


ثم اضغط على الدراسة A يستبعدا من التحليل.

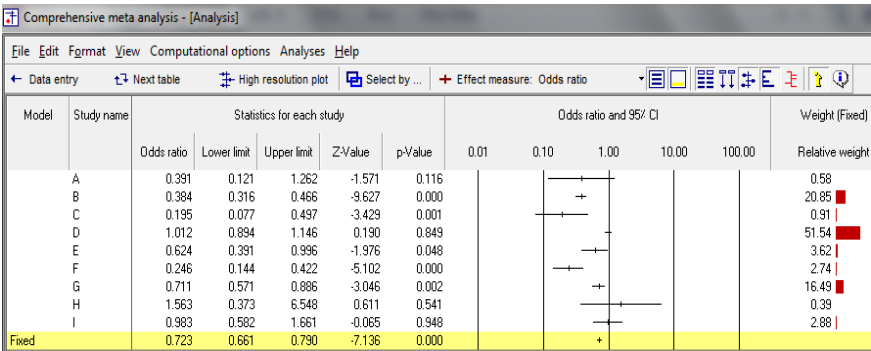
التحليل التراكمي Cumulative Meta-analysis

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:



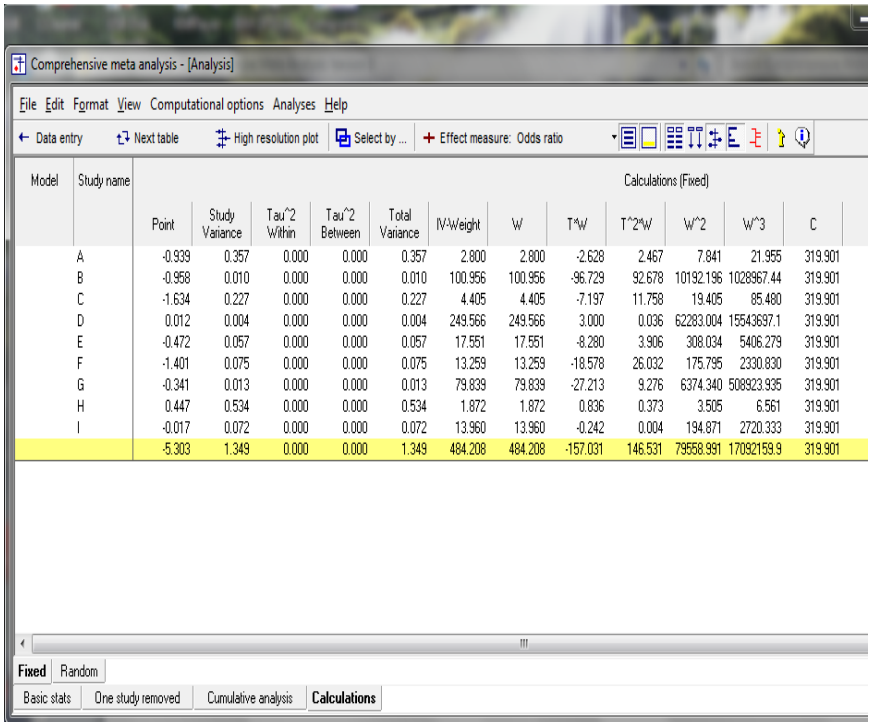
اضغط عليه يعطى النتائج الآتية:



يقوم البرنامج بعرض النتائج متراكمة حيث الصف الثاني يعرض نتائج الدراساتين الأولتين السابقتين والصف الثالث يعرض نتائج الدراسات الثلاثة السابقة وهكذا.

عرض الحسابات التفصيلية

- اضغط على Calculations أسفل الشاشة يعطى الحسابات التفصيلية الآتية:



is (Fixed)									
Q	Q df	I ²	B	K	Summary Point	Summary Variance	Group T ²	Group T ² Variance	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	
95.605	8.000	91.632	2879.790	9.000	-0.324	0.002	0.274	0.056	

إجراء التحليلات المتفاعلة

لإجراء التحليلات المتفاعلة باستخدام متغير تصنيفي (خاصية الدراسة) وليكن جنس العينة لا بد من إنشاء عمود المتغير المتفاعل كالتالي:

- اضغط مرتين متتاليتين بالموس على أى عمود لايضمن بيانات تظهر الشاشة التالية:

Column format

Name

Variable name

Column function

Data type

Decimals displayed

Alignment

Cancel

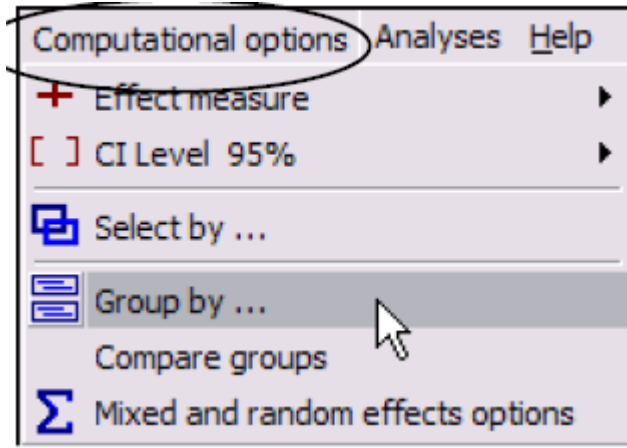
Ok

- امام Column function اضغط على السهم لأسفل واختار Moderator امام data type اختار Categorical.
- اكتب مسمى المتغير التصنيفي وليكن الجنس Sex.

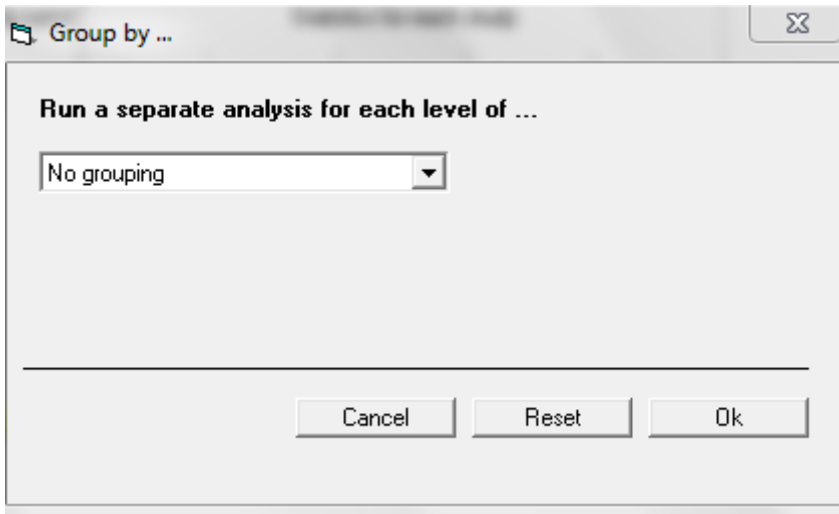
- اضغط OK يظهر مسمى المتغير فى شاشة البيانات.
- ابدأ فى إدخال بيانات المتغير المتفاعل كما يلى:

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									

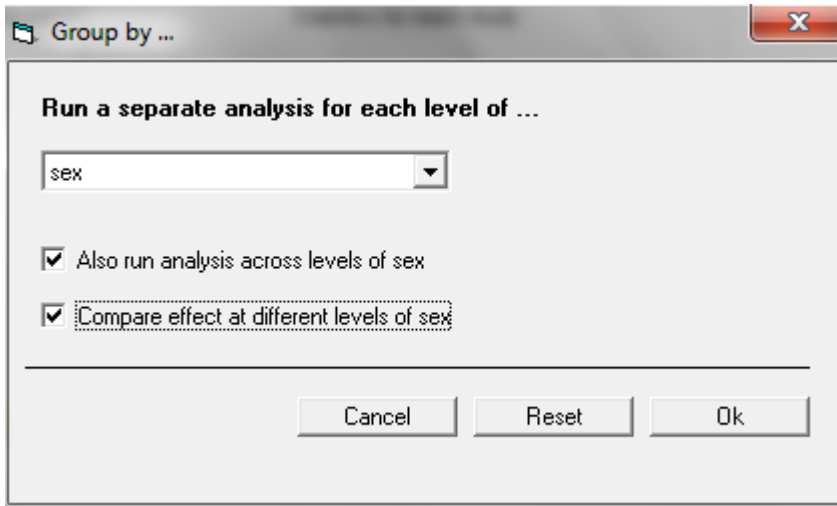
- اضغط Run analyses وبظهر المخرج اضغط على قائمة Computational options تظهر الشاشة الآتية:



- اضغط على Group by يعطى المربع التى:



- اضغط على السهم الأسفل فى امام No grouping 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 كالاتى:



- اضغط OK يظهر المخرج الآتي:

Comprehensive meta analysis - [Analysis]

File Edit Format View Computational options Analyses Help

← Data entry → Next table High resolution plot Select by ... + Effect measure: Odds ratio

Model	Group by sex	Study name	Statistics for each study					Odds ratio and 95% CI		
			Odds ratio	Lower limit	Upper limit	Z-Value	p-Value	0.50	1.00	2.00
	female	B	0.394	0.316	0.466	-9.627	0.000			
	female	D	1.012	0.894	1.146	0.190	0.849			
	female	F	0.246	0.144	0.422	-5.102	0.000			
	female	H	1.563	0.373	6.548	0.611	0.541			
Fixed	female		0.737	0.665	0.817	-5.829	0.000			
	male	A	0.391	0.121	1.262	-1.571	0.116			
	male	C	0.195	0.077	0.497	-3.429	0.001			
	male	E	0.624	0.391	0.996	-1.976	0.048			
	male	G	0.711	0.571	0.896	-3.046	0.002			
	male	I	0.983	0.582	1.661	-0.065	0.948			
Fixed	male		0.681	0.569	0.815	-4.184	0.000			
Fixed	Overall		0.723	0.661	0.790	-7.136	0.000			

Fixed Random Both models

Basic stats Calculations

لاحظ أن المخرج اعطى إحصائيات كل مجموعة على حدة المظللة بالدائرة بالإضافة إلى إحصائيات المجموعتين معاً (المشار إليه بالسهم).

اختيار نموذج الحسابات أو التحليل

لاختيار نموذج التحليل:

• اضغط Computational options واختار Mixed and random effects

تظهر الشاشة الآتية:

Mixed and random effects options

Combining studies within a subgroup

☐ Assume a common among-study variance component across subgroups (pool within-group estimates of tau-squared).

☒ Do not assume a common among-study variance component across subgroups (do not pool within-group estimates of tau-squared). This is the option used by RevMan.

Combining subgroups to yield an overall effect

☒ Combine subgroups using fixed effect model

☐ Combine subgroups using random effects model

Cancel Apply Ok

• اضغط Ok .

عرض الإحصائيات الإضافية لتحليل المجموعة

- اضغط على اختيار table Next يظهر الإحصائيات الإضافية لكل مجموعة بالإضافة إلى قيم التباين بين المجموعات وداخل المجموعات كالآتي:

Comprehensive meta analysis - [Analysis]

File Edit Format View Computational options Analyses Help

Data entry Next table High resolution plot Select by ... Effect measure: Odds ratio

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			

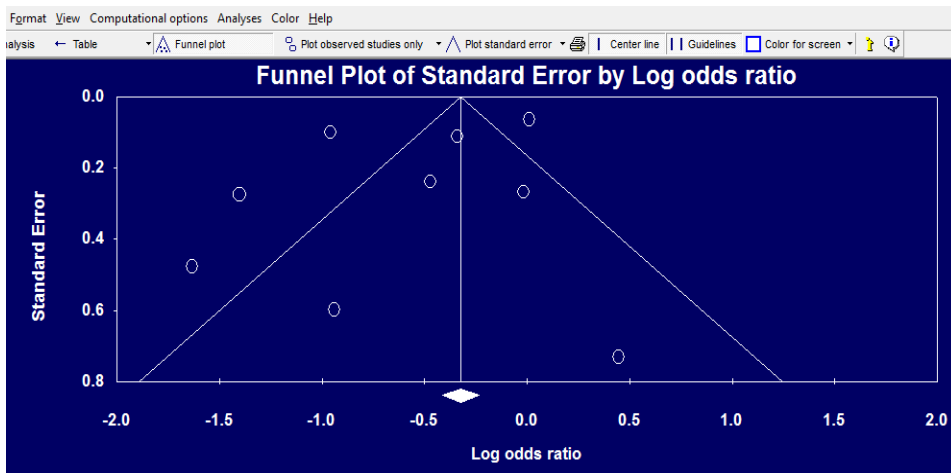
Fixed Random Both models

Basic stats Calculations

Tau-squared				
I-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

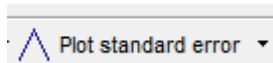
إجراء شكل القمع لتشخيص تحيز النشر

- اضغط على قائمة Analysis واختار Publication bias يعطى الشكل الآتى:

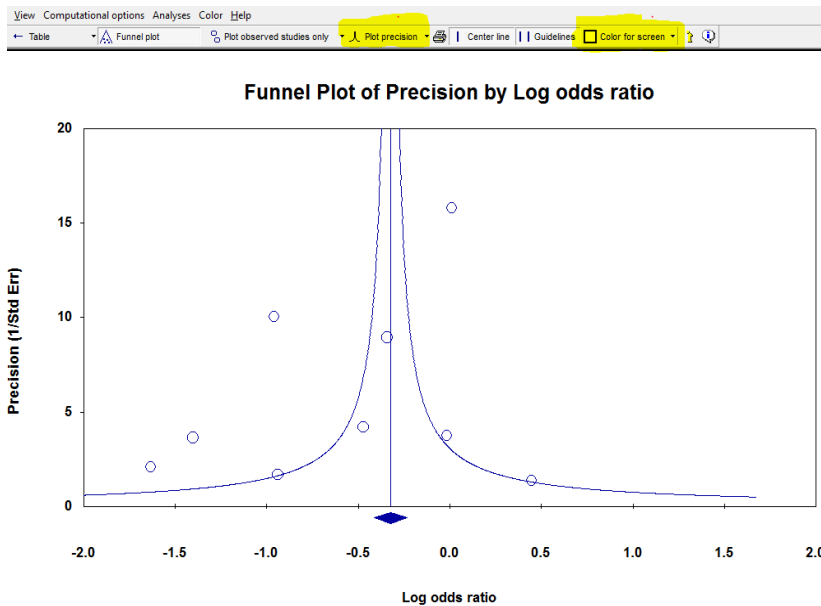


وتم عرض شكل القمع فى ضوء الخطأ المعياري و لوغاريتم Odd ratio ويتضح من الشكل وجود تحيز نشر نتيجة وقوع احداثيات عديدة خارج القمع وايضًا ان أحجام التأثير لا تتوزع اعتداليًا حول القمع. وإذا وقعت معظم الدراسات أسفل الشكل على جانب واحد من الشكل أكثر من جانب آخر فهذا يدل على عدم انتظامية شكل القمع بالتالى وجود تحيز للدراسات.

ويمكن عرض الشكل فى ضوء الدقة كالتالى:



-الضغط على Plots standard error أعلى الشكل واختيار Plot precision يظهر الشكل الاتي:



ويمكن تغيير لون الشكل من خلال الايقونة المظلمة بالأصفر Color of .screen

وتوجد طرق لتقدير التحيز النشر باستخدام إجراء احذف واملى Trim and Fill وفكرته قائمة على شكل القمع ولتنفيذه:

- اضغط على قائمة View فى الشكل السابق واختار Trim and Fill يظهر المخرج الاتي:

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
Adjusted values	0	0.72303	0.66141	0.79039	0.55612	0.37341	0.82823

Look for missing studies where?

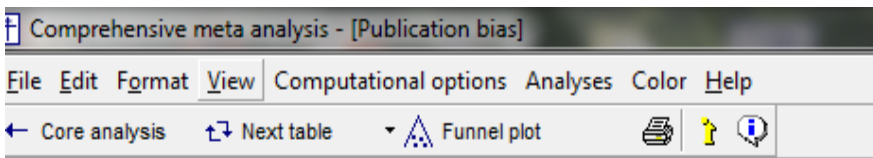
- ☐ Not specified
- ☒ To left of mean
- ☐ To right of mean

وبالضغط على الايقونة  أعلى الشاشة يقدم تقرير وتفسير للنتائج. ولا يوجد تأثير للدراسة الغائبة على تغير الشكل واعطى متوسط حجم التأثير وفترات الثقة لنموذج التأثيرات المثبتة والعشوائية.

معامل الارتباط الرتبي لتشخيص التحيز

وهو أحد طرق الاستدلال الإحصائي لتشخيص التحيز ويقدر من خلال معامل ارتباط كندال تاو b بين أحجام التأثير والاختفاء المعيارية المشتقة من أحجام العينات يقدر كالتالى:

- ضغط على قائمة View ثم اختار Rank correlation test 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

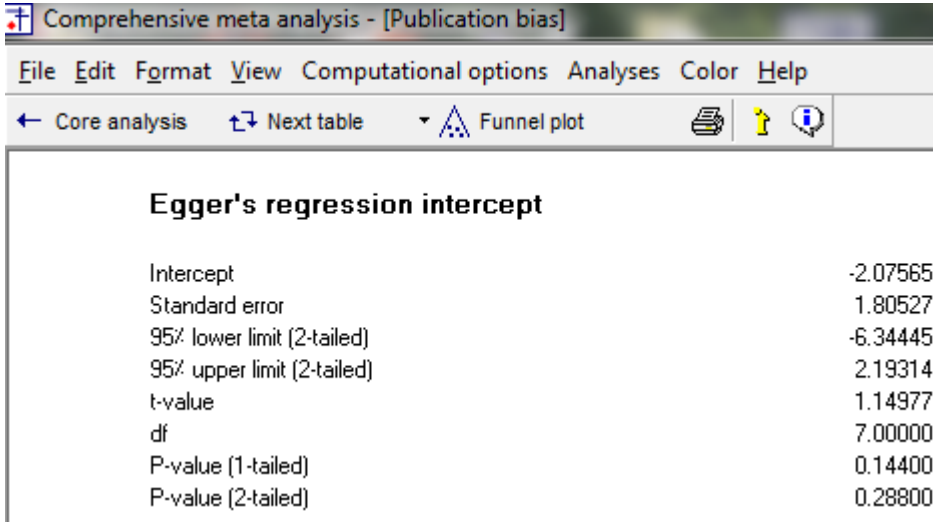
Tau	-0.02778
z-value for tau	0.10426
P-value (1-tailed)	0.45848
P-value (2-tailed)	0.91697

ويتضح عدم الدلالة الإحصائية حيث $P=0.458$ وعليه لا يوجد تحيز للنشر وهذا يتعارض مع تحليل شكل القمع وذلك لأن اختبار كندال ليس الاختبار الصادق فى هذه الحالة نظراً لصغر حجم عينة الدراسات ونقص القوة الإحصائية للاختبار ولا يمكن الاعتماد عليه فقط كدليل للحكم على تحيز النشر.

مدخل Egger Regression

هذا المدخل يعتمد على التنبؤ بأحجام التأثير من خلال مؤشر الدقة (مقلوب الخطأ المعياري) ويتم تنفيذ كالتالي:

- اضغط على قائمة View ثم اختار Regression test يظهر المخرج الآتي:



The screenshot shows the 'Egger's regression intercept' window in Comprehensive meta analysis. The window has a menu bar with File, Edit, Format, View, Computational options, Analyses, Color, and Help. Below the menu bar is a toolbar with buttons for Core analysis, Next table, and Funnel plot. The main area displays the following results:

Egger's regression intercept	
Intercept	-2.07565
Standard error	1.80527
95% lower limit (2-tailed)	-6.34445
95% upper limit (2-tailed)	2.19314
t-value	1.14977
df	7.00000
P-value (1-tailed)	0.14400
P-value (2-tailed)	0.28800

ويتم الحكم على وجود تحيز من خلال Intercept(B0) وقيمة -2.07 بفترات ثقة (-6.344, 2.1931) وقيمة P الاحتمالية 0.14400 وهى غير دالة إحصائياً 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 بالحدثة إلى حد 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 وأهم العوامل التي يمكن أن تسبب الاختلافات بين الدراسات لتضمينها في التحليل لإجراء التحليلات المتفاعلة.

• **الدقة في 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 اللوغاريتمي إلى حجم تأثير قابل للتفسير. ويجب أن يكون التفسير مرتبط بالنظرية سواء بتأييدها أو رفضها أو محاولة إدخال تحسينات أو تعديلات عليها كذلك ان وجدت اختلافات بين أحجام التأثير فيجب البحث عن العوامل التي احدثت هذه الاختلافات من خلال إجراء ANOVA أو ما وراء الانحدار.

تجنب اشكاليات عند كتابة عرض نتائج ما وراء التحليل

قدم (2012) Card عشرة اشكاليات شائعة عند عرض نتائج ما وراء التحليل وفيما يلي 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 النظرية للمراجعة أو للدراسة والتحليلات والنتائج المعروضة. أي تحليل يتم إجرائه يكون لسبب وهذا السبب يكون معروضًا في مقدمة الدراسة، وبعض الباحثين يعرضوا أحيانًا بعض التحليلات الاستكشافية ولكن يجب التركيز في المقام الأول على التحليلات التي تجيب على الأسئلة البحثية أو تخدم اهداف الدراسة.

2. **عدم اعطاء عرض وتفصيلات كافية عن المنهجية:** في دراسات ما وراء التحليل يجب عرض وصف كافي عن استراتيجيات البحث التي اعتمد عليها الباحث في الحصول على الدراسات، معايير التضمين أو الاستبعاد، إجراءات التأكيد لخصائص الدراسات، كيفية حساب حجم التأثير، واستراتيجية 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 وراء 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 للإجابة 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 للإجابة 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 وراء التحليل:** وهذا واضح عند كتابة جزء المناقشة فينسى الباحث الوصول إلى استنتاج واضح عن النتائج ما وراء التحليل للإجابة 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 تمثيلاً دقيقاً، فتوجد محددات فى الحصول على الدراسات وكذلك بعد استبعاد العديد من الدراسات فى ضوء معايير التضمين والاستبعاد فاصبحت عينة صغيرة نسبياً فى ضوء مجتمع الدراسات للمتغيرين المراد دراستهما ويجب عرض نتائج تقييم تحيز النشر وتحديد خصائص هذه الدراسات يحد من القدرة التعميمية لنتائج الدراسة.

7. **عدم عرض الملامح الوصفية للدراسات:** الفشل في اعطاء المعلومات الأساسية والوصفية عن الدراسات المتضمنة في ما وراء التحليل وهذه المعلومات أو خصائص الدراسات على درجة كبيرة من الأهمية سواء في استخدام بعض الخصائص في إجراء التحليلات المتفاعلة أو للوصول إلى استنتاجات من الدراسة ولذلك من الضروري عرض هذه الخصائص في بداية جزء النتائج.

8. **استخدم نموذج التأثيرات المثبتة في التحليل في 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 تقييم 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 في ما وراء التحليل وان نتائج هذه التحليلات مرتبطة بخصائص الدراسات ويجب عدم تفسير نتائجها بصورة سببية لأن طبيعة المتغيرات المتضمنة في هذه التحليلات غير تجريبية.

10. **لا تعتقد وجود طريق صحيح Right way لإجراء وكتابة ما وراء التحليل على الرغم من كل التوصيات الملموسة في هذا العرض** إلا انها توصيات مجردة ومطلقة ولكن من الضروري اتباع الصيغ المتبعة المحددة من قبل الجمعيات العلمية مثل الجمعية النفسية الأمريكية APA وكذلك المجالات العلمية في هذا الشأن.

وعلى الرغم من مميزات ما وراء التحليل كأسلوب لتحليل وتلخيص النواتج الكمية للتراث البحثي الخاص بموضوع ما ولكنه كأى أسلوب يشوبه العديد من المحددات والمشاكل ويمكن عرض البعض منها المرتبط بتفسير النتائج للممارسة ولصانعي السياسة والبحث. وبعض هذه المحددات كالتالى (Lipsey & Wilson 2001):

المناسبة المنهجية للأساس البحثي 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 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 تأثيرها على حجم التأثير يطلق عليها تحليلات الحساسية Sensitivity Analyses.

تداخل الملامح الجوهرية والمنهجية

يمكن تؤدي نتائج الدراسات الضعيفة منهجياً 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 هذه المتغيرات وأحجام 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 نتيجة 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 توكيدها للمساعدة فى تحديد العوامل التى تسبب هذا الاختلاف فرما 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 التضمنات هو ان التأثيرات فى ما وراء التحليل الخاصة بمدى كفاءة معالجة معينة لفئة معينة ليس من الضرورى أن تكون ممثلة للتأثيرات التى تحدث فى الممارسة الاكلينيكية فى المواقف غير البحثية.

واحد محددات تعميم النتائج ان الدراسات تتم فى مواقف معملية وليس من الضرورى ان تعميم نتائجها فى الممارسات الطبيعية فى المدرسة أو المستشفيات أو العيادات هذه 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 امبريقي كمي لتحديد النواقص المنهجية إذا تحقق التى:

الأول: ان يجمع الباحث مجموعة من الدراسات الشاملة التى تمثل مجتمع الدراسات تمثيلاً جيداً.

الثانى: تحديد ماهية العيوب المنهجية فى الدراسات.

الثالث: تطبيق ما وراء التحليل لكل الدراسات الجيدة والضعيفة منهجياً وليس كما هو سائد فى دراسات ما وراء التحليل استبعاد الدراسات الضعيفة منهجياً 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 تقييم 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 منهجياً والجيدة منهجياً. ومن الأفضل تحديد هذه 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 (2009) Borenstien et al. 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 فهذا 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 معاً؟ الاجابة تتوقف على طبيعة وهدف البحث.

ايضاً 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 هذه الأسئلة تتوقف على طبيعة السؤال البحثى.

هل يجب أن تكون الدراسات المتضمنة فى التحليل متشابهة؟ من وجهة النظر الإحصائية لا يوجد قيود على التماثل أو التشابه 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 فمثلاً فى الدراسات الدوائية والعلاج بادوية كلما كانت الدراسات المتضمن فى التحليل محكمة ومتجانسة ومعايير التضمين محددة تماماً لفئة معينة فإن هذا يعطى تقديرات دقيقة. ولكن فى معظم تحليلات ما وراء التحليل معايير التضمين تكون واسعة وليست مقيدة وهذا يعطى قدراً كبيراً 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، وايضاً بتضمين المتغيرات التصنيفية مثل الجنس او المرحلة التعليمية فهذا يفيد فى إجراء ما وراء تحليلات منفصلة فى الدراسة الواحدة وهذا يشار اليه ما وراء التحليل للمجموعات الفرعية - meta- Subgroup analysis ولو كان التأثير متماثل عبر المجموعات المختلفة عندئذ يقال ان للتأثير 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، فمثلاً ما وراء التحليل الذى يتضمن تقييم فعالية برنامج أو تدخل معين سيعتمد على الدراسات التجريبية وايضاً يعتمد على نوعية التجارب هل التصميمات التى تتضمن مجموعتين وإى انتقاء للمجموعتين للعينة عشوائية أو مقصودة، أم تصميم المجموعة الواحدة ويعتمد الباحث على الدراسات السببية الارتباطية كل هذا يتوقف على الهدف من الدراسة.

ويرى البعض تضمين كل الدراسات التجريبية سواء كانت حقيقية أو شبه 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 تأكيد 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 دمجها 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 الشأن 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 الاسباع 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 لدرستين أفضل وأكثر 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 التاكيد 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 إحصائياً ومن الأفضل مراجعتها كفيّاً ولكن هذا لا يعنى عدم إجراء ما وراء التحليل بل إجرائه وفى هذه 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 خطأ المعاينة (للمزيد يمكن الرجوع إلى Hunter & Schmidt, 2004).

معايير الجودة فى دراسة ما وراء التحليل

قدم (2009) Cooper قائمة تتضمن عدد من الخطوات الإرشادية أو المعايير التي يمكن أن تساعد على جودة الدراسة وتقييمها ومدى مقبوليتها للنشر وهي كالاتى:

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 للإجابة عن الأسئلة؟
- هل قواعد أو معايير التضمين أو الاستبعاد واضحة؟

جمع البيانات

- هل توجد استراتيجيات 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 الإحصائي المناسب؟
- هل حسب المتوسط الموزون في ضوء مقلوب تباين الخطأ المعياري؟
- هل حسب درجة الاختلاف بين أحجام التأثير؟
- هل أجرى تحليل التفاعلات إذا حدث اختلاف بين أحجام التأثير؟

اختبار ضلالة النتائج

- هل حسب تحليل الحساسية واستخدمه في تفسير النتائج؟
- هل تناول قضية البيانات الغائبة من خلال اختبارات تحيز النشر؟

عرض النتائج

- هل عرض النتائج في ضوء جداول وأشكال بشكل جيد؟
- هل الاستخلاصات من الدراسة واضحة وكاملة ومرضية؟

تفسير النتائج

- هل أصدر تعميمات ومحددات الدراسة؟
- هل تفسير النتائج مرتبطة بما أنتجته الدراسة من نتائج ومرتبطة بالقيمة العملية بمجال الدراسة؟

- هل حدد نواحي معينة بحاجة إلى دراسات أخرى؟

ومن خلال العرض السابق لفصول الكتاب يمكن لمعد هذا الكتاب من طرح قائمة
لكيفية قراءة وتقييم جودة دراسة ما وراء التحليل كالاتي:

- هل العنوان مصاغ بطريق توحى انها دراسة ما وراء التحليل؟
- هل الأسئلة مصاغة بطريقة مناسبة؟
- هل استراتيجية 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 البحثي؟
- 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 او عشوائى 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 والبيانات 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

- الصيد، 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.
- الصيد، عبد العاطي 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.
- الصيد، عبد العاطي احمد. (1994). الإحصاء التربوى. مذكرة غير منشورة كلية التربية، جامعة قناة السويس.
- الصيد، عبد العاطي 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.

عامر، عبد الناصر السيد.(2012). النماذج والإختبارات الإحصائية المستخدمة في البحث النفسي المصري والعربي. *المجلة المصرية للدراسات النفسية*، 22، 74، 371-355.

عامر، عبد الناصر السيد (2012). استخدام وتقدير حجم التأثير في الدراسات النفسية التربوية في المجلات العربية (1994 - 2011). *مجلة جامعة عين شمس للقياس والتقويم*، 2 ، 1-40 .

عامر، عبد الناصر السيد. (2014). تقدير القوة الإحصائية للدراسات النفسية في المجلات العربية (2000-2012). *مجلة دراسات تربوية ونفسية، جامعة الزقازيق*، 84، 213-244.

عامر، عبد الناصر السيد. (2018). *ما وراء التحليل للعلوم السلوكية والنفسية: الاسس النظرية والمنهجية 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 trials. *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 primer. *Psychological Bulletin*, 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). Hillsdale, NJ: Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. 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 central 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 anlysis*. 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: McGraw-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 Council 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](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., Keselson, 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 consequences 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 less 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., & Henzlen, 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 sciences(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). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton 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.campbellcollaboration.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: Allyn & 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: Understanding 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.