الجزء الثاني

الإحصاءات الوصفية الثنائية

- الفصل السادس: تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين: الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي
- الفصل السابع: الوصف العددي للبيانات الاسمية: مقاييس التطابق الثنائية
- الفصل الثامن: التطابق بين متغيرات تم قياسها على المستوى الترتيبي
- الفصل التاسع: التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار

الفصل السادس

تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين:

الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي

مقدمة

في الفصول المتعلقة بالجزء الأول من هذا الكتاب تناولنا طرق التحليل المتعلقة بمتغير واحد في وقت محدد، وتُعرف هذه الطريقة بطريقة الإحصاءات الوصفية الأحادية. ويمكن أن تكون لدينا بيانات متعلقة بمتغير واحد، مثل الدخل، نستطيع وصفه بطريقة ملائمة تسمح لنا بقياسه طبقاً لأهداف البحث ومستوى قياس المتغير، فعلى سبيل المثال؛ باستطاعتنا وصف الدخل الأسبوعي مقاساً بالدينار الليبي وذلك من خلال الطرق الثلاثة التالية:

1- الوصف الجدولي: وذلك من خلال توليد جدول تكراري نسبي.

2- الوصف البياني: وذلك من خلال توليد مضلع تكراري.

3- الوصف العددي أو الرقمي وذلك من خلال إجراء العملية الحسابية للحصول على المتوسط الحسابي والانحراف المعياري لضمان النزعة المركزية والتشتت لهذا التوزيع؟

إن كل واحد من هذه الإجراءات يركز تحديداً على متغير واحد نرغب في دراسته.

العلاقة بين المتغيرات:

بطبيعة الحال، في خضم عملية إجراء البحث في العادة نقوم بجمع البيانات حول عدد كبير من المتغيرات، كالدخل، والتعليم، والنوع، والوضع الصحي، وعدد ساعات مشاهدة الإذاعة المرئية، وعدد ضخم من المتغيرات الأخرى التي يرغب الباحث في دراستها. وكل واحد منا يمتلك الإحساس العام، للفكرة التي مفادها أن متغيرين قد يرتبطان بعضهما ببعض، أو أنهما يعتمدان على بعضهما البعض. ونحن نعرف أنه كلما كبر الأطفال زادوا طولاً، فالعمر والطول متغيران يرتبطان ببعضهما البعض. كما أننا أيضاً نعلم أنه كلما زاد دخلنا زاد نمط استهلاكنا، فالدخل والاستهلاك يرتبطان ارتباطاً وثيقاً. إن هذه الأمثلة توضح لنا المفهوم العام الذي نشعر به. إن هذه الفكرة تبين لنا أنه كلما تغيرَّت إحدى قيم المتغير من حالة إلى حالة أخرى تغيرِّت معها أيضاً قيم المتغير الآخر، عادة يختار الباحث متغيرين لاعتقاده أنهما يرتبطان بعضهما ببعض بطريقة متسقة. فعلى سبيل المثال، قد يُعْتَقَدُ أن دخل شخص يرتبط بشكل ما بالمكان الذي يقطن فيه. ولبحث هذه العلاقة يتطلب ذلك جمع بيانات من عينة من الناس، قد نجد من خلال دراسة هذه العينة أن الناس الذين يعيشون في أحدى المدن لديهم مستوى اقتصادي محدد، في حين أن الناس الذين يعيشون في مدينة مختلفة يميلون إلى مستوى اقتصادي مرتفع، بينما الناس الذين يعيشون في مدينة ثالثة يسجلون مستوى اقتصادي أكثر ارتفاعاً. إن مثل هذه النتائج تمكننا من استنتاج أن مكان الإقامة ومستوى الدخل يرتبطان حقاً ارتباطاً وثيقاً. فإذا كانت العلاقة بين الإقامة ومستوى الدخل علاقة ارتباطية، يعني هذا أننا نتوقع أن الأشخاص يختلفون في دخولهم وفقاً لمكان الإقامة. ومن هنا يمكننا مقارنة شخصين يعيشان في مناطق مختلفة فإننا نتوقع أنهما يحققان مستوياتٍ اقتصاديةً مختلفةً أيضاً. وعليه فإننا لا نتعامل مع الدخل كمتغير متميز كلياً، ولكننا بدلاً من ذلك، ننظر إليه بشكل ما على أنه مرتبط بمكان الإقامة للشخص. ولكي نبين مثل هذه العلاقة في البيانات التي تم جمعها، فإننا سنستخدم الإحصاءات الوصفية التي لا تلخص التوزيع لكل متغير بشكل منفصل فقط، ولكنها أيضاً تصف الطريقة التي تتغير بها قيم أحد المتغيرين والتي ترتبط بتغير القيم في المتغير الآخر.

دعنا نعطي موقفاً مخالفاً عندما يكون لدينا متغيران غير مرتبطين ببعضهما البعض.

بمعنى آخر، أنهما مستقلان. والمتغيران يكونان مستقلين إذا كان نمط التباين في الدرجات لأحدهما ليس مرتبطاً بنمط التباين في الدرجات للمتغير الآخر.

توجد لدينا عدة تقنيات، نطلق عليها إجمالاً الإحصاءات الوصفية الثنائية والتي يمكن أن تستخدم من قِبل الباحث لمعرفة ما إذا كانت هناك علاقة بين متغيرين، انظر الجدول التالي:

جدول (6ـ1) الإحصاءات الوصفية الثنائية

الطريقة	الأمثلة
- طرق الجدولة والتمثيل البياني، هذه الطرق تمدنا ببيانات بطريقة تبين احتمالية العلاقة بين متغيرين.	- الجدول الثنائي لبيانات مقولية "اسمية وترتيبية". رسم انتشاري لبيانات متصلة "المقياس ذو المسافات والنسبي".
- الطرق العددية، هذه الطرق هي عمليات رياضية تستخدم للتكميم لرقم أحادي، قوة العلاقة. وهذه الطرق يطلق عليها قياس العلاقات. وعندما يكون كلا المتغيرين قد تم قياسهما على الأقل على المستوى الترتيبي فهما يشيران أيضاً إلى اتجاه العلاقة.	(مستوى اسمي) Lambda, Cramerśv (قياس ترتيبي بقيم قليلة) Gamma, Kendalś tau-b Somerś d and tau-c (قياسات ترتيبية بقيم كثيرة) Spearmanś rank-orders Correlation coefficient (ذو مسافات ونسبي) Pearsonś correlation coefficient

المصدر: George Argyrous, op. cit. P. 136.

هذه الطرق متممة وتساند بعضها البعض، وهدف كل طريقة من هذه الطرق هو مدنا بالبينة أو الدليل لأي اعتماد قد يوجد بين متغيرين.

نمذجة العلاقة بين متغيرين [1]:

إن ما سوف نتعلمه في هذا الفصل والفصول اللاحقة له هي التقنيات المحددة لاستقصاء العلاقة بين متغيرين، إلا أنه في بادئ الأمر نحتاج إلى أن ننظر حول المستوى التصوري إلى ماذا يعني الارتباط بين متغيرين.

نفترض أننا قمنا بجمع بيانات حول متغيري الدخل ومكان الإقامة. إننا نقوم بإجراء هذا البحث معتقدين أن هناك علاقة بين هذين المتغيرين، آملين أن نكشف خصائص هذه العلاقة في البيانات التي قمنا بجمعها. حيث يمكننا تصوير أو نمذجة العلاقات الممكنة بين هذين المتغيرين - إن وجدت - بطرق متعددة. إن أبسط طريقة يرتبط بها متغيران يمكن أن تسبب العلاقة المباشرة والتي يمكن أن نبينها في الأشكال المحتملة التالية:

1- علاقة مباشرة أحادية الاتجاه "الدخل كمتغير تابع":

إن أنماط هذه العلاقات كطريق واحد يمتد من مكان الإقامة إلى الدخل (انظر شكل 6-1). ويمكن أن تكون لدينا نظرية من خلالها يمكننا القول إن المهنة وفرص الحياة تتباين عبر المدن، وبالتالي يؤثر ذلك على مستوى الدخل للأفراد الذين يعيشون في تلك المدن. في هذه الحالة بالذات يمكننا أن نجادل في أنه يوجد لدينا نمط الاعتماد مع الدخل كمتغير تابع ومكان الإقامة كمتغير مستقل.

مكان الإقامة (متغير مستقل) ← الدخل (متغير تابع)

شكل (6-1) علاقة أحادية الاتجاه "الدخل كمتغير تابع"

2- علاقة مباشرة أحادية الاتجاه "مكان الإقامة كمتغير تابع":

قد تكون هناك مجموعة أخرى من الباحثين الاجتماعيين الذين لا يوافقوننا على النموذج السابق، لأن أولئك الباحثين ينطلقون من بعد نظري آخر يسلم بوجود نمط الاعتماد بين المتغيرين، ولكن هذا النمط يتخذ اتجاهاً آخر. فالناس ذوي الدخول العالية يمكنهم أن يختاروا مكاناً مناسباً للإقامة، وبالتالي قد ينتقلون إلى المدينة ذات البيئة المرغوبة. وعليه فإن مكان الإقامة سيكون متغيراً تابعاً والدخل متغيراً مستقلاً. (انظر شكل 6 – 2).

الإقامة	←	الدخل
(متغير تابع)		(متغير مستقل)

شكل (6-2) علاقة أحادية الاتجاه "مكان الإقامة كمتغير تابع"

3- علاقة مباشرة ثنائية الاتجاه بين مكان الإقامة والدخل بشكل متعمد تبادليا:

أيضاً قد تكون هناك مجموعة ثالثة من علماء العلم الاجتماعي يمكنهم أن يتفقوا على أن المتغيرين مرتبطان بعضهما ببعض، ولكنهم يعتقدون أن كلا النمطين من السببية (العلّية) يعملان بكيفية تجعل المتغيرين يؤثر بعضهما في الآخر. أي علاقة تبادلية.

في هذا النموذج إنه من غير الملائم تصوير أن واحداً من هذين المتغيرين كمتغير مستقل والآخر كمتغير تابع. ولكن بدلاً من ذلك، يمكن النظر إليهما بأنهما معتمدان على بعضهما البعض تبادلياً.

مكان الإقامة	↔	الدخل
(متغير تابع)		(متغير تابع)

شكل (6-3) علاقة ثنائية الاتجاه بين مكان الإقامة والدخل بشكل تبادلي

وعليه فإنه من الأهمية بمكان أن لا يكون هناك تشويش في بيان العلاقة المعتمدة بالعلاقة المتقاربة Closely related ولكن ما يميز المشكلة هو تحديد ما المتغير التابع؟ وما المتغير المستقل. وتحدث هذه العملية فقط في حالة العلاقة أحادية الاتجاه وهي الطريقة المناسبة لتحديد أحد المتغيرين كمتغير مستقل والآخر كمتغير تابع. كذلك من الأهمية بمكان أن نتذكر عند اختيارنا لنماذج تستند على وجهات نظرية محددة حول طبيعة العالم الواقعي وسلوك البشر، فإن هذه النماذج يمكن أو لا يمكن أن تكون صحيحة. فالتحليل الإحصائي لا يبرهن البتة على أي من أنماط السببية التي أوضحناها سابقاً. وكل ما تبينه هذه التحليلات الإحصائية هو أن بعض العلاقات الإحصائية بين المتغيرات المشاهدة قد استندت على بيانات تم جمعها.

إن الطريقة التي تنظم وتفسر بها البيانات تبنى على النتائج التي تتوقف على هذه الافتراضات النظرية المسبقة. وإن نفس البيانات يمكن أن تبين لنا مجموعة من الفروق المعتمدة على افتراضات نظرية لما تمدنا به البيانات. فعلى سبيل المثال، لقد أوضحنا ثلاثة نماذج بسيطة لوصف العلاقة بين متغيرين وتجدر الإشارة إلى أن هناك نماذج أكثر تعقيداً والتي تتضمن العلاقة بين ثلاثة متغيرات أو أكثر. ولاكتشاف العلاقات الأكثر تعقيداً قد يقودنا الأمر للتعامل مع التحليل المتعدد الذي يبحث في العلاقات بين أكثر من متغيرين - سنؤجل الحديث في هذا الأمر إلى الجزء الرابع من هذا الكتاب - وعلى أي حال فإنه من الأهمية بمكان أن نضع في اعتبارنا عند تفسيرنا للنتائج الثنائية الحقيقة القائلة إن أي علاقة مشاهدة بين متغيرين يمكن أن تكون أكثر تعقيداً مقارنة بالنماذج البسيطة المتعلقة بالسبب والنتيجة كما وصفناها سابقاً.

والآن يمكننا أن نبين التقنيات المتعلقة بوصف العلاقات الثنائية وذلك من خلال الجدول الثنائي.

وصف البيانات المقولية باستخدام جداول التقاطع:

دعنا نفترض أننا نقوم بدراسة لاستقصاء العلاقة بين مكان الإقامة ومستوى الدخل، إننا قد نشك في وجود علاقة بين هذين المتغيرين وقد يقودنا هذا الشك إلى أبعد من ذلك، للاعتقاد بأن العلاقة السببية الأحادية تبدأ من مستوى الدخل إلى مكان الإقامة. بمعنى آخر يمكننا أن نجادل في أن مستوى الدخل هو المتغير المستقل، وأن مكان الإقامة هو المتغير التابع. ويتوجب علينا أن نتذكر أنها الطريقة الافتراضية الوحيدة لما نتوقع أن نجده. إن العالم الحقيقي، أو على الأقل البيانات التي تم جمعها من هذا العالم، قد لا تتفق مع هذا التوقع. إن هذين المتغيرين قد لا يرتبطان بقدر ما يكونان مستقلين عن بعضهما البعض. والسؤال هنا هو: كيف يمكن لنا أن ننظم البيانات التي تم جمعها لنبين ما إذا كان النموذج الذي نستخدمه نموذجاً صحيحاً أو ما إذا كان المتغيران في حقيقة الأمر مستقلين عن بعضهما.

إن متغير مكان الإقامة قد تم تعريفه إجرائياً من خلال طرح سؤال على (720) شخصاً بخصوص ما إذا كانوا يعيشون في مناطق ريفية أو حضرية (مقياس اسمي) في حين قد تم تعريف الدخل تعريفاً إجرائياً على مستوى المقياس الترتيبي وذلك من خلال تحديد ما إذا كان الشخص يكسب أكثر أو أقل من متوسط الدخل العام؛ إن هؤلاء الأشخاص الذين يكسبون أقل من المتوسط العام قد تم تصنيفهم إلى ذوي الدخول المنخفضة أما الذين يكسبون أكثر من المتوسط العام، فقد تم تصنيفهم إلى فئة ذوي الدخول العالية. إن هذا البحث الذي نجريه يقدم لنا (720) زوج من الأعداد، وإن كل شخص سوف تخصص له قيمته لتشير إلى المكان الذي يعيشون فيه. وقيمة أخرى تشير إلى المجموعة الاقتصادية التي ينتمي إليها. والسؤال هنا: كيف يمكن لنا أن ننظم هذه الأعداد المتعلقة بِـ (1440) شخصاً بطريقة تظهر لنا أي علاقة موجودة بين الدخل ومكان الإقامة؟. يمكننا في هذه الحالة أن نستخدم الطرق الأحادية التي تعلمناها في الفصول السابقة من هذا الكتاب ولبناء توزيعات تكرارية منفصلة لكل متغير. (جدول 6 - 2 وجدول 6 - 3).

جدول (6-2) التوزيعات التكرارية لمكان الإقامة

مكان الإقامة	التوزيع التكراري
حضري	397
ريفي	323
المجموع	$\sum 720$

جدول (6-3) التوزيع التكراري لمستوى الدخل

مستوى الدخل	التوزيع التكراري
دخل منخفض	372
دخل عالي	348
المجموع	$\sum 720$

إنه من الواضح أن هذه التوزيعات الأحادية المنفصلة في الجدولين السابقين لا تساعدنا البتة كثيراً وذلك لأنه من الصعب أن نلاحظ ما إذا كانت العلاقة موجودة بين متغيرين، والتي هي هدف الدراسة. ولكي نتحصل على أي علاقة محتملة يمكن أن توجد بين متغيرين تم قياسهما ببيانات مقولية فإنه ينبغي علينا استخدام الجدول الثنائي والذي يُعرف أيضاً بجدول التوافق أو جدول التقاطع (Crosstab). فالجدول الثنائي يعرض لنا التوزيعات التكرارية المشتركة لمتغيرين، إن جدول التقاطع لهذه البيانات قد تم جمعهما (نظرياً) يوضحها الجدول (6 – 4).

جدول رقم (6 ـ 4) العلاقة بين مكان الإقامة ومستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	230	397
ريفي	205	118	323
المجموع	**372**	**348**	$\sum 720$

المصدر: George Argyrous, Statistics for Social & Health Research, op. cit, P. 140.

من خلال جدول التوافق الذي يبين التوزيع المشترك لمتغيرين يمكننا أن نقرأ الدرجات التي أعطيت لكل حالة من حالات المتغير بشكل تلقائي. وبالنظر إلى هذا الجدول على سبيل المثال، يمكننا أن نرى أن هناك 205 حالة تقطن في المناطق الريفية ولديها دخول منخفضة. وبما أن الجداول الثنائية تصف بيانات بطريقة تُظهر هذا التوزيع المشترك، كما أنها تسمح لنا باستقصاء فيما إذا كان هذان المتغيران مرتبطين.

وتجدر الإشارة إلى أن هناك مجموعة من القواعد التي يتوجب على الباحث اتباعها لبناء الجداول الثنائية:

- لابد أن يكون للجدول عنوان ملائم، فجدول التقاطع يجب دائماً أن يكون له عنوان مع بيان قيم كِلاَ المتغيرين.
- الإشارة إلى مصدر البيانات، ويمكن أن يكون المصدر قبل أو بعد الجدول أو هامش مرتبط بالجدول كما هو وارد في الجدول السابق.
- وضع المتغيرات في مكانها المناسب في الصفوف والأعمدة - عادة ما يوضع المتغير المستقل في العمود والمتغير التابع في الصف - إذا كان هناك أي سبب للاعتقاد أن المتغيرين ليسا مرتبطين ببعضهما البعض ولكن أحد المتغيرين تابع على الجانب الآخر

(علاقة اتجاهية واحدة)، فالمتغير المستقل يجب أن يرتب عبر الأعمدة في حين أن المتغير التابع ينظم طبقاً للصفوف. في هذا المثال الذي أمامنا قد خصصنا أن يكون الدخل متغير مستقل (العمود) في حين أن متغير مكان الإقامة متغير تابع (الصف). إننا نقوم بذلك لاعتقادنا، وعلى ضوء النموذج المستخدم الذي مفاده أن دخل الفرد يحدده مدى العوامل التي في النهاية تؤثر على اختيار الناس أين يعيشون. وعلى أي حال، يمكننا فقط أن نـجادل بشكل شرعي أن هناك علاقة عكسية: فالدخل متغير (تابع) ومكان الإقامة متغير (مستقل). إذا ما اعتقدنا في هذا النموذج، فإننا بالتالي يمكننا بناء جدول ثنائي حيث يكون الدخل في الصفوف ومكان الإقامة عبر الأعمدة. وبطريقة أخرى يمكننا أن نـجادل أن الاعتماد يكون متبادلاً وعليه فالقضية لا تعني شيئاً ما إذا كان المتغير في الصفوف أو عبر الأعمدة. ولمناقشة استخدام جداول التقاطع كوسيلة لوصف البيانات، يتوجب علينا أن نكون على دراية ببعض المصطلحات:

- **حجم وأبعاد الجدول**: يعرف حجم الجدول بعدد الفئات لمتغير الصفوف مضروب في عدد الفئات لمتغير العمود. في هذا المثال، لدينا مجموعتان من الفئات هما: مكان الإقامة (حضري-ريفي)، وفئتا الدخل (منخفض- عالٍ) يولدان لنا جدول 2×2. وإذا تم قياس الدخل على مقياس يحتوي على أربع نقاط مثل: (دخل منخفض جداً، منخفض، عالٍ، عالٍ جداً) فإننا بالتالي سنتحصل على جدول تكون أبعاده 2×4.
- **خلايا الجدول**: كل مربع في الجدول سيحتوي على عدد الحالات التي لديها مجموعة متآلفة من القيم لمتغيرين، وفي هذه الحالة نطلق عليها جدول الخلية.
- **هوامش الجدول:** إن مدونات مجموع الأعمدة يطلق عليها هوامش العمود وبنفس الكيفية فإن مدونات مجموع الصفوف يطلق عليها هوامش الصف.

جداول التقاطع ذات التوزيعات النسبية [2] :

بإمكان الباحث أن يطوِّر قدراته ليرسم أي علاقة محتملة موجودة في البيانات وذلك بحساب التوزيعات النسبية، بدلاً من الأعداد المطلقة من الحالات في كل فئة مشتركة. وفي كل خلية من خلايا الجدول نبين المعلومات في إطار النسب إما لمجاميع الأعمدة أو لمجاميع الصفوف. إن التوزيعات النسبية تقوم أساساً على مجموع الأعمدة كما هو موضح في الجدول التالي:

جدول (6ـ5) العلاقة بين مكان الإقامة والدخل (N = 720)

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	$\frac{167}{372} \times 100 = 45\%$	$\frac{230}{348} \times 100 = 66\%$	$\frac{397}{720} \times 100 = 55\%$
ريفي	$\frac{205}{372} \times 100 = 55\%$	$\frac{118}{348} \times 100 = 34\%$	$\frac{323}{720} \times 100 = 45\%$
المجموع	$\frac{372}{372} \times 100 = 100\%$	$\frac{348}{348} \times 100 = 100\%$	$\frac{720}{720} \times 100 = 100\%$

إن قيمة 55% تمثل عدد أولئك الذين يعيشون في المناطق الريفية ولديهم مستويات اقتصادية منخفضة وهي نسبة تمثل العدد الكلي لأولئك الذين يكسبون الدخول المنخفضة:

$$\frac{205}{372} \times 100 = 55\%$$

وكما يمكننا أيضاً، وبشكل آخر أن نبين التوزيعات النسبية باستخدامنا لمجموع الصفوف كما هو مبين في الجدول رقم (6) الذي يحتوي على الحسابات المتعلقة بالقاطنين في المناطق الحضرية. ويمكننا بوجه السرعة أن نرى من هذه العملية أن 42 % من الناس كنسبة فقط من العدد الكلي لأولئك الناس الذين يعيشون في المناطق الحضرية ولديهم مستوى اقتصادي منخفض. وفي بعض الأحيان فإن الفرق في التوزيعات التكرارية للجداول يمكننا أن نجمعها لتعطينا البيانات الخام والنسب المئوية المناسبة وذلك بإضافة أعمدة أو صفوف أخرى.

جدول (6-6) العلاقة بين مكان الإقامة والدخل (N = 720)

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	$\frac{167}{397} \times 100 = 42\%$	$\frac{230}{397} \times 100 = 58\%$	% 100
ريفي	% 63	% 37	% 100
المجموع	**% 52**	**% 48**	**% 100**

إن أنسب بناء يعتمد على محتوى البيانات التي تهم المشاهدين وعلى أية حال، وبالرغم من المجاميع التي من خلالها يتم حساب التوزيعات النسبية، فإنه من الأهمية بمكان، أن نضيف حجم العينة الحقيقي حتى يمكننا إعادة تحويل النسب إلى الأرقام المطلقة إذا ما طُلب منا ذلك.

إجراءات توليد جداول التقاطع Cross tables باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Descriptive Statatics ⟸ Crosstabs

2- انقر على المتغير في القائمة التي تشكل صفوف الجدول والتي تكون في هذه الحالة Place of residence.

3- انقر على ◄ التي تشير إلى القائمة المحددة المُعْنَوَنَة Row(s) نقوم بلصق Place of residence في القائمة المحددة للصفوف Row(s).

4- انقر على المتغير في القائمة التي تشكل أعمدة الجدول والتي تكون في هذه الحالة Income Level.

5- انقر على ◄ التي تشير إلى القائمة المحددة المُعْنَوَنَة Column(s) نقوم بلصق Income Level في القائمة المحددة للأعمدة Column(s).

6- انقر فوق OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Place of residence Income Level Crosstabulation Count

		Income Level		TOTAL
		Low	**High**	
Place of residence	urban	167	230	397
	Rural	205	118	323
TOTAL		**372**	**348**	**720**

شكل (6-4) مخرجات Crosstabs باستخدام Spss

يمكن للباحث من خلال استخدام برنامج Spss أن يتوسع في توليد جدولٍ يحتوي على التوزيعات النسبية relative والمطلقة absolute.

كما يظهر في الشكل التالي:

Place of residence Income Level Crosstabulation

		Income Level		TOTAL
		Low	High	
Place of residence	urban Count	167	230	397
	% Within Place of residence	42.1 %	57.9 %	100.0 %
	% Within Income Level	44.9 %	86.1 %	55.1 %
	Rural Count	205	118	323
	% Within Place of residence	63.5 %	36.5	100.0 %
	% Within Income Level	55.1 %	33.9 %	44.9
TOTAL	Count	372	348	720
	% Within Place of residence	51.7 %	48.3 %	100.0 %
	% Within Income Level	100.0 %	100.0 %	100.0 %

المصدر: George Argyrous, op.cit., PP. 143 - 144

شكل (6-5) مخرجات SPSS لجداول التقاطع بالتوزيعات النسبية

تجدر الإشارة إلى أنه من خلال جدول واحد يمكن تقديم كل المعلومات التي تم توليدها بشكل منفصل في الجداول 4، 5، 6.

تفسير جداول التوافق : نمط وقوة العلاقة :

يمكننا في هذه الجزئية أن نقدم أداة مهمة لوصف جداول التوافق في البحوث الاجتماعية وأهمية هذه الأداة تكمن في أننا في مجال البحث الاجتماعي نقوم بجمع بيانات كثيرة وهي بيانات منفصلة وبيانات مقولية مقاسة على مستوى المقياس الاسمي أو الترتيبي. وعندما تُحول مجموعة من البيانات الخام إلى جدول التقاطع فإن المهمة تكمن في كيفية تفسيرها. ولتقيِّيم ما إذا كانت هذه البيانات تُظهر لنا أن هناك علاقة موجودة بين المتغيرين. وعند تفسير أي علاقة واضحة في جدول التوافق فإننا في واقع الأمر ننظر إلى خاصيتين أساسيتين:

1- نمط العلاقة.
2- قوة العلاقة.

ولكي يتبين شكل العلاقة يبقى من الأهمية بمكان أن نسلط الضوء على خلية النموذج Model Cell لكل عمود، انظر جدول رقم (6 – 7).

جدول رقم (6ـ 7) العلاقة بين مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167 (%45)	230 (%66)	397
ريفي	205 (%55)	118 (%34)	323
المجموع	**372**	**348**	**720**

وعند إلقاء الضوء على خلية النموذج لكل خلية في العمود يمكننا القول إن هناك علاقة. وبالنظر إلى التوزيعات النسبية فإنه من الواضح أن 55 % من الذين يكسبون

دخولاً منخفضة يعيشون في مناطق ريفية، في حين نجد أن نسبة 66 % من الذين يكسبون دخولاً عالية يعيشون في مناطق حضرية. ومن هنا يمكننا أن نقترح تفسيراً مفاده أن هناك علاقة بين الدخل ومكان الإقامة، ويشير نمط هذه العلاقة إلى أن أولئك الذين يكسبون دخلاً منخفضاً يتركزون في الريف، في حين إن أولئك الذين يكسبون دخولاً عالية يتركزون في المدن.

كذلك يمكننا أن نقيّم قوة هذه العلاقة وذلك بالنظر إلى نسبة الحالات في كل عمود تحصلنا عليها من خلال خلية النموذج في كل عمود. كذلك يمكننا أن نرى أنه بينما فئة النموذج للذين يكسبون دخولاً منخفضة يعيشون في الريف، كذلك توجد نسبة عالية من ذوي الدخول المنخفضة يعيشون في مناطق أخرى.

كذلك الحال، بينما نجد أن الغالبية من الذين يكسبون دخولاً عالية يعيشون في مناطق حضرية، فإن هناك نسبة كبيرة منهم لا يعيشون في هذه المناطق الحضرية، وتُبَيِّنُ لنا هذه النتيجة أن العلاقة بين الدخل ومكان الإقامة علاقة ليست قوية.

تفسير الجداول المتقاطعة عندما يكون لدينا متغيران مقاسان على المستوى الترتيبي:

في الجزء الأول من هذا الفصل تم التركيز على مناقشة بناء جداول التوافق والكيفية التي يمكن بها تفسير العلاقة في مثل هذه الجداول. ولقد تناولنا مثالاً يكون فيه أحد المتغيرين مقاساً على المستوى الاسمي (مكان الإقامة) ومتغير آخر تم قياسه على المستوى الترتيبي (الدخل). إن القواعد والإجراءات التي تعلمناها من هذا المثال السابق يمكن تطبيقها في بناء جدول التوافق لمتغيرات يمكن قياسها على أي مستويات مؤتلفة. عندما يقاس كلا المتغيرين على المستوى الاسمي، ومع ذلك، فإن التفسيرات في نمط العلاقة الموجود في جدول التوافق يمكن أن يتخذ خطوة أبعد ليدمج في مناقشة علاقة الاتجاه والاتساق.

1- اتجاه العلاقة Direction of Relationship:

تحدثنا في السابق عن العلاقة بين الدخل ومكان الإقامة وتوصلنا إلى نتيجة مفادها أن الدخل يرتبط بمكان الإقامة كأن يميل ذوو الدخول المحدودة أكثر إلى السكن في الأماكن

الريفية، بينما يميل ذوو الدخول العالية أكثر إلى العيش في المناطق الحضرية وجاءت هذه النتيجة لأننا تعاملنا مع متغير (مكان الإقامة) الذي تم قياسه على المستوى الاسمي وبالتالي لا نستطيع أن نتحدث عن الزيادة أو النقصان في الدخل من حيث كونه مرتبطاً بالزيادة أو النقصان في مكان الإقامة. وحسبنا أن نشير إلى أنه من غير المنطقي الحديث حول مكان الإقامة من حيث كونه أكبر أو أصغر فمكان الإقامة يتباين عبر الحالات ولكنه لا يزيد أو ينقص كمياً.

وعندما يقاس كلا المتغيرين على الأقل على المستوى الترتيبي فإنه بإمكاننا أن نتحدث عن اتجاه العلاقة من حيث كونها سالبة أو موجبة. فالعلاقة الموجبة تعني أن الزيادة في درجات وحدة المقياس (x) يقابلها زيادة في وحدة المقياس بالنسبة لمتغير (y) في حين إن العلاقة السالبة تعني أن الزيادة في درجات وحدات المقياس بالنسبة لمتغير (x) يقابلها نقص في وحدة المقياس بالنسبة لمتغير (y). فعلى سبيل المثال، قد نرغب في معرفة العلاقة بين الدخل ومشاهدة الإذاعة المرئية لشخص ما، بدلاً من العلاقة بين الدخل ومكان الإقامة، إن كمية ما يشاهده هذا الفرد يمكن قياسها من خلال طرح السؤال التالي: ما إذا كان الشخص يشاهد الإذاعة المرئية، أو لا يشاهدها البتة، يشاهدها بعض الليالي، أو معظم الليالي.

إن مثل هذا السؤال تم قياسه على المستوى الترتيبي أي أن كلا المتغيرين الآن تم قياسهما على المستوى الترتيبي، فإذا ما استطعنا أن نجد علاقة موجودة فعلاً في هذه الحالة يمكننا الحديث عن اتجاه العلاقة. نفترض أننا قمنا بجمع معلومات حول 300 شخص وتم قياس دخل هؤلاء الناس، (هذه المرة تم قياس الدخل على مقياس يتكون من ثلاث نقاط: دخل منخفض - دخل متوسط - دخل عال). وكمية مشاهدة الإذاعة المرئية، بإمكاننا بشكل أولي أن نقلص 600 حالة وهي تمثل البيانات الخام للبحث إلى جدول ثنائي. إننا نشك إذا ما وجد نمط الاعتماد بين هذين المتغيرين الذي سوف يبدأ من الدخل إلى مشاهدة الإذاعة المرئية. وعليه يمكننا أن نرتب الجدول على أن يكون الدخل (متغير مستقل) عبر الأعمدة ومشاهدة الإذاعة المرئية (كمتغير تابع) على الصفوف.

إن كل القواعد التي تعلمناها في بداية هذا الفصل والمتعلقة ببناء جداول التوافق لازالت قابلة للتطبيق. إلا أن هناك قاعدة إضافية مهمة عندما يقاس المتغيران على المستوى الترتيبي. وعندما يحتوي جدول التوافق على متغيرين مقاسين على المستوى الترتيبي، يستوجب ذلك علينا ترتيب الجداول بطريقة تزيد فيها قيم المتغير المستقل عبر الصفحة من اليمين إلى اليسار في حين أن قيم المتغير التابع تنقص أسفل الصفحة.

جدول رقم (6ـ8) توزيع مشاهدة الإذاعة المرئية حسب الدخل

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 71.5٪	15 15٪	10 10.5٪	100
بعض الوقت	20 19.0٪	70 70٪	10 10.5٪	100
معظم الوقت	10 9.5٪	15 15٪	75 79.0٪	100
المجموع	**105** **100.0٪**	**100** **100.0٪**	**95** **100.0٪**	**300**

ولكي نستطيع تفسير الجدول كما تم في المثال السابق يستوجب ذلك علينا أن نلقي الضوء على خلية النموذج لكل عمود وبالتالي يمكننا أن نرى بشكل سريع أن هناك علاقة بين هذين المتغيرين، أي أن الزيادة في الدخل يقابلها زيادة في كمية مشاهدة الإذاعة المرئية. ومن هنا يمكننا القول بأن هناك علاقة موجبة. وإذا كانت خلايا النموذج كلها تتراصف على طول القطر الآخر من الأعلى إلى أبداً إلى منخفض إلى معظم الليالي فإن هذا الجدول سيصف علاقة سالبة.

2- علاقة الاتساق Consistency of the Relation:

بالإضافة إلى اتجاه العلاقة عندما تعاملنا مع المتغيرين اللذين تم قياسهما على المستوى الترتيبي، فإنه بالتالي يمكننا أن نبحث فيما إذا كانت العلاقة علاقة اتساق. لاحظ أن كل خلايا النموذج في جدول (8) قد رتبت على طول القطر الموجب، وبالتالي يوجد تعاقب في العلاقة عبر كل مدى القيم، إن مثل نمط الاعتماد هذا يطلق عليه "علاقة الاتساق". إلا أنه على الطرف الآخر يمكننا مشاهدة النتيجة التي يحتويها الجدول رقم (6 – 9).

ومن خلال بيانات هذا الجدول يمكننا إعادة القول بأن هناك علاقة بين المتغيرين ويمكن وصفها بأنها علاقة ليست متسقة.

جدول رقم (6-9)

العلاقة بين مشاهدة الإذاعة المرئية حسب الدخل (علاقة غير متسقة)

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 %71.5	10 %10.0	15 %15	100
بعض الوقت	20 %19.0	10 %10.0	70 %70	100
معظم الوقت	10 %9.5	75 %71.5	15 %15	100
المجموع	**105** **%100.0**	**95** **%100.0**	**100** **%100.0**	**300**

من خلال هذا الجدول يمكننا أن نلاحظ في نهاية الفقرة المتعلقة بقياس الدخل، أنه كلما زاد الدخل زادت معه كمية مشاهدة الإذاعة المرئية. لكننا نلاحظ أن العلاقة بين المتغيرين هي علاقة عكسية، كلما تحركنا أبعد على فقرات مقياس الدخل.

الخلاصة:

لقد حاولنا في هذا الفصل أن نبين بشكل واسع بناء وتفسير الجداول الثنائية. ومن خلال هذا الفصل تبين لنا أهمية هذه الجداول في وصف البيانات المقولية بطريقة تكشف لنا ما إذا كانت هناك علاقة بين متغيرين يخضعان للاستقصاء؛ كما ناقشنا في هذا الفصل القواعد المحددة والإجراءات لتحويل البيانات الخام المجمعة إلى جداول تقاطعية مندمجة. كما بيّنّا الأسس التي على ضوئها يمكن تفسير أي علاقة يبينها جدول التوافق، وذلك كما اتضح لنا في كل الجداول التي حاولنا من خلالها تقييّم نمط وقوة العلاقة؛ أيضاً حاولنا من خلال هذا الفصل قياس متغيرين مقاسين على المستوى الترتيبي لإعطائنا أيضاً أبعاداً إضافية للعلاقة التي يمكن أن نكشفها باستخدام جداول التوافق، أعني اتجاه العلاقة وما إذا كانت العلاقة متسقة أم لا.

أسئلة للمراجعة:

1- من الجداول التالية: احسب النسب المتعلقة بالعمود:

أ:

التابع	المستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
المجموع	**75**	**110**	**185**

ب:

التابع	المستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
المجموع	**71**	**70**	**60**	**201**

2- من كل جدول من الجداول التالية: احسب النسبة المئوية للصف:

أ:

التابع	المستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
المجموع	75	110	185

ب:

التابع	المستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
المجموع	**71**	**70**	**60**	**201**

3- عينة افتراضية لأطفال من أستراليا، كندا، سنغافورة وبريطانيا تم مقارنتهم فيما يتعلق بكمية مشاهدة الإذاعة المرئية:

كمية مشاهدة الإذاعة المرئية	البلد				
	كندا	**أستراليا**	**بريطانيا**	**سنغافورة**	**المجموع**
منخفضة	23	25	28	28	104
متوسطة	32	34	39	33	138
عالية	28	30	40	35	133
المجموع	83	89	107	96	375

المطلوب: هل يمكنك القول إن كمية مشاهدة الإذاعة المرئية (التلفزيون) مرتبط ببلد الإقامة؟

4- عينة مؤلفة من 162 رجلاً بين 40 و 65 سنة تم اختيارهم لمعرفة الوضع الصحي لكل رجل. وتم طرح السؤال التالي على كل منهم لمعرفة ما إذا كان يدخن بشكل روتيني. وجاءت نتيجة هذه العينة بالشكل التالي:

الوضع الصحي	عادة التدخين		المجموع
	يدخن	لا يدخن	
سيء	34	13	47
متوسط	19	22	41
جيد	9	35	44
جيد جداً	3	27	30
المجموع	**65**	**97**	**162**

المطلوب:

- بيّن المتغير المستقل والمتغير التابع ومستوى قياس كل منهما؟
- هل يمكنك تصور أي علاقة محتملة بالنظر إلى أحد المتغيرين كمتغير تابع والآخر كمتغير مستقل، مع تبرير إجابتك؟
- من هذا الجدول احسب نسب العمود.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications, London ,2001, pp. 137 - 141.

2- 2. Ibid, PP. 141 - 149.

ثانيا: المصدران الرئيسيان لهذا الفصل :

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications, London, 2001 .

2- Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الفصل السابع

الوصف العددي للبيانات الاسمية :
مقاييس التطابق الثنائية

مقدمة

في الفصول السابقة أوضحنا الكيفية التي من خلالها تُبنى جداول التقاطع التي تُعْرَف بأنها: الطريقة التي تُنَظَّمُ بها البيانات المقولية بطريقة تكشف لنا ما إذا كانت هناك علاقة بين المتغيرات. ولتوضيح ذلك نسوق المثال التالي:

جدول (7ـ 1) مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167 (%45)	230 (%66)	397 (%55)
ريفي	205 (%55)	118 (%34)	323 (%45)
المجموع	**372 (%100)**	**348 (%100)**	**720 (%100)**

المصدر: George Argyrous, op.cit, p. 152

عند تحليل جداول التقاطع ننظر ما إذا كانت هناك علاقة موجودة ونهدف من وراء جداول التقاطع إلى طرح سؤالين مرتبطين:

1- ما هو نمط هذه العلاقة؟

2- وكم قوة هذه العلاقة؟

يمكننا أن نرى في هذا الجدول نمط العلاقة التي تشير إلى أولئك الذين يكسبون دخلاً منخفضاً نسبياً يعيشون في مناطق ريفية في حين يعيش أولئك الذين يكسبون دخلاً عالياً نسبياً في مناطق حضرية.

بمعنى آخر، الزيادة في الدخل تتجه إلى أن ترتبط بتغير مكان الإقامة من الريف إلى المناطق الحضرية. كذلك يمكننا أيضاً وصف هذه العلاقة بمصطلحات لفظية: كالعلاقة القوية. فالتباين في الدخل الأدنى إلى الأعلى يرتبط بكمية محددة من التباين في مكان الإقامة، من الريف إلى الحضر، ولكن هذا التباين ليس بكمية كبيرة من التغير.

وتجدر الإشارة إلى أنه قد نكون أكثر موضوعية إذا ما كانت لدينا طريقة لقياس قوة العلاقة في جداول التقاطع بدلاً من أن نعتمد على العين المجردة للحكم فيما إذا كان هناك تباين من شخص لآخر، لذلك فإنه من الأفضل أن تكون لدينا طريقة لقياس قوة العلاقة، لكي تمدنا بنفس الإجابة، بغض النظر عما يدليه الشخص من حكم. وهذا بالتحديد الوظيفة الأساسية لمقاييس التطابق. مثال، قد يختلف اثنان في حكمهما اليوم فيما إذا كان الجو حاراً نسبياً. فشخص يرى أن الجو حار نسبياً، في حين شخص آخر قد يحكم بأن الجو حار جداً، بينما شخص ثالث لديه شعور بأن درجة الحرارة تميل إلى البرودة. كلنا أشخاصاً نمر بنفس الخبرة المتعلقة بدرجة حرارة هذا اليوم، إلا أن انطباعاتنا الذاتية لهذه الخبرة تكون مختلفة. إذا ما نظرنا على أية حال، فكلنا نشير إلى مقياس الحرارة وننظر إلى درجة الحرارة التي وصلت إلى 20 درجة. هذا الأمر قد نوافق عليه جميعاً وإن مقياس درجة الحرارة يبين لنا نفس الدرجة بغضّ النظر عمن ينظر إليه. فمقياس درجة الحرارة مقياس موضوعي باعتباره مقياس مكمم يستند على معيار عام.

وبنفس الطريقة قد نجد مجموعة مختلفة من الناس يمكن أن تنظر إلى جدول التقاطع

ولفظياً تُقيِّم قوة العلاقة بطرق مختلفة، فمقاييس التطابق يمكن أن تمدنا بمؤشر حاسم لقوة العلاقة التي سوف تعطي نفس الإجابة لكل شخص(1).

1- مقاييس التطابق كإحصاءات وصفية:

تُعرف مقاييس التطابق: بأنها إحصائيات وصفية تقيس العلاقة بين المتغيرات. كما تشير إلى مفاهيم كمية وإلى أي مدى أن التغير في قيمة أحد المتغيرات يرتبط بالتغير قي قيمة المتغير الآخر. بمعنى آخر يشير التطابق إلى "الاعتماد" عندما تزيد سنوات العمر هل يزيد أيضاً الطول أو يقل؟ هل التغير في المعتقدات الدينية يتطابق مع التغير في الاتجاه والمواقف حول عقوبة الإعدام؟ وكما نعرف أن العلاقات قد نحصل عليها من خلال الرسوم البيانية والجداول التي تبين لنا بطريقة ما العلاقة التي يمكن أن توجد بين متغيرين. كذلك باستطاعتنا، بالإضافة إلى هذه الطرق الوصفية البسيطة أن نحسب مقاييس التطابق للتكميم الفعلي للانطباعات التي تحصلنا عليها من خلال هذه الأدوات.

إن الأمر المهم الذي ينبغي أن نتذكره حول مقاييس التطابق أن هذه العلاقة تساعدنا في وصف البيانات، بدلاً من الركون إلى الانطباعات المرئية لجداول التقاطع أو الرسوم البيانية، هذه الأدوات يمكنها في أحسن الظروف أن تمدنا برقم مفرد لقوة ونمط التطابق.

إن المشكلة مع هذه المقاييس هي تحديد الظروف الملائمة التي يمكن أن تقدم بها هذه المعلومات. فإذا كانت الظروف الملائمة لا تنطبق فحينئذٍ يمكن أن تكون هذه المقاييس العددية مقاييس مضللة، وهنا ينبغي أن تعتمد على الرسوم البيانية والجداول المرتبطة بالوصف اللفظي الملائم لوصف علاقة.

ولوضع مفهوم التطابق في الواقع العملي تكون مشكلة غامضة فالعمل مع مقاييس التطابق يمكن أن تكون تجربة غير مجدية وذلك لوجود عدد كبير من المقاييس للاختيار منها، وكلٌ لها فرادتها وقصورها، وفي الغالب فإن هذه المقاييس قد لا تقود إلى نفس النتيجة. فعلى سبيل المثال، أن الكثير من مقاييس العلاقة تكون حساسة فيما يتعلق بالقرار الذي يتم به تصميم المتغير كمتغير مستقل وما هو المتغير التابع. مثل هذه المقاييس

هي مقاييس غير متماثلة Asymmetric. فالمقاييس غير المتماثلة تكون مفيدة عندما نعتقد أن العلاقة لأحد المتغيرات الذي يعتمد على متغير آخر. ولكن على الجانب الآخر، فإنه إذا ما كان لدينا شك حول العلاقة على أنها علاقة متبادلة، فإننا في هذه الحالة نستخدم مقاييس التماثل Symmetric التي تأخذ نفس القيمة بغض النظر عن المتغير الذي تم تحديده بأنه متغير مستقل أو الذي تم تحديده بأنه متغير تابع. الجدول التالي يعطينا بعض التوجهات للاختيار بين المقاييس الأكثر شيوعاً، التي سنتناولها بالتفصيل في الفصول اللاحقة. إن نقطة البداية للقياس هي تحديد المستوى الذي على ضوئه يتم قياس المتغير.

جدول (7ـ2): مقاييس التطابق Meaoures of association

مستوى القياس	مقاييس التطابق	
الاسمي	لامبيدا	Lambda
	جودمان – كروشكل تاو	Goodman and Krushal tau
	كراميرزV	Cramerś V
الترتيبي	سومرز d	Somerś d
	جاما	Gamma
	كندلز تاو b	Kendallś tau-b
	كندلز تاو c	Kendallś tau-c
	ارتباط سبيرمان	Spermanś rank order Correlation Coefficient
ذو المسافات/ النسبي	ارتباط بيرسون (r)	Pearsonś(r) Product-moment Correlation Coefficient

وتجدر الإشارة إلى أنه عندما يقاس متغيران على مستويات مختلفة، فإنه كقاعدة عامة، فإن اختيار القياس يعتمد على أدنى اثنين من مستويات القياس، على سبيل المثال، إذا ما أردنا أن نستقصي فيما إذا كانت هناك علاقة بين النوع "اسمي" والرضا الوظيفي "ترتيبي" فإن أدنى مستوى القياس لهذين المتغيرين هو المقياس الاسمي، وبالتالي يجب

علينا أن نتقيد بمدى المقاييس التي يمكننا حسابها من تلك المقاييس التي أوردناها في الجدول السابق والتي تلي المقياس الاسمي. ولبناء مقياس التطابق فإنه من المفضل مراعاة الخصائص التالية [2]:

1- من الضرورة بمكان أن تأخذ مقاييس التطابق قيمة 1 (أو –1 حيثما كان ضرورياً) في مواقف يكون فيه التطابق كاملاً. ولسوء الحظ ليس الأمر كذلك، وأن ذلك يسبب الكثير من الإحباط المرتبط باستخدام مقاييس التطابق، فبعض مقاييس التطابق يمكن أن تأخذ قيماً أكبر من 1، بينما مقاييس أخرى مثل (جاما) يمكن أن تأخذ قيمة 1 عندما لا يوجد تطابق كامل.

2- إنه من الضرورة أيضاً أن تأخذ مقاييس التطابق قيمة 0 في مواقف لا يوجد فيها تطابق. ولسوء الحظ ليست كل المقاييس تنطبق عليها هذه الخاصية المثالية. إن بعض المقاييس يمكن أن تأخذ قيمة (0) حتى عندما يوجد تطابقُ بيّن وجلّيُّ للعين المجردة.

3- عندما يقاس كلا المتغيرين على الأقل على المستوى الترتيبي فإن علامة + أو – يجب أن تشير إلى اتجاه التطابق: فيما إذا كانت الزيادة في قيمة أحد المتغيرَيْن متطابقة مع الزيادة "تطابق موجب" أو النقصان "تطابق سالب" في قيمة المتغير الآخر.

فيما يتبقى من هذا الفصل سوف نتناول بالنقاش مقاييس التطابق لمتغيرين عندما يكون أحدهما أو كلاهما قد تم قياسه على المستوى الاسمي. وقبل فعل ذلك، إنه من الأهمية بمكان أن نضع في اعتبارنا أن كل هذه المقاييس تشير إلى التطابق، ولكنها ليست بالضرورة أن تبين ما إذا كان أحد هذه المتغيرات يسبب التغير في متغير آخر. فقد يشك الباحث نظرياً أن أحد المتغيرين يسبب التغير في متغير آخر، إلا أن الإحصاء الذي سوف نتعلمه لا يبرهن على مبدأ السببية، ولكن الأمر يتعلق فقط بتقديم شواهد تعزز الافتراضات النظرية. على سبيل المثال يمكننا أن نلاحظ العلاقة بين عدد طيور اللاقلاق في منطقة معينة ومعدل التكاثر في تلك المنطقة، ويمكننا أن نحسب القياس الذي يكمم

هذه العلاقة الإحصائية. إلا أنه ليس باستطاعتنا أن ننتقل من العملية الإحصائية إلى النتيجة التي مفادها أن طيور اللاقلاق قد تسبب معدل التكاثر *.

2- مقاييس التطابق للمتغيرات الاسمية:

وكما بينا على التو مقياس التطابق فإنه باستطاعتنا أن نفكر في مؤشر عددي يشير إلى قوة العلاقة. وهذه المقاييس تمتد ما بين طرفَيْن أحدهما يشير إلى حالة التطابق التام. والآخر إلى عدم وجود تطابق. وباستخدامنا للمثال المتعلق بالدخل ومكان الإقامة. يمكننا بيان ما إذا كان لدينا متغيران متطابقان بشكل تام.

جدول (7-3): مكان الإقامة وفقاً لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	0	397	397
ريفي	323	0	323
المجموع	**323**	**397**	**720**

يتبين لنا من هذا الجدول أنه إذا كان هناك أشخاص لديهم دخل منخفض فإن ذلك يتيح لنا أن نتوقع بشكل قاطع أنهم يعيشون في مناطق ريفية. فالدخل لهذه المجموعة من الحالات هو متنبئ تام لمكان الإقامة: فمعرفة مستوى الدخل لشخص ما يسمح لنا بالتكهن بشكل تام أين يعيش هذا الشخص. ولكن بطريقة أخرى، فإن التغير من دخل منخفض إلى دخل عال دائماً سوف يتطابق مع تغير في مكان الإقامة من الريف إلى الحضر. ومن خلال التطابق التام يمكننا القول إن كل تباينٍ في المتغير التابع (مكان

---

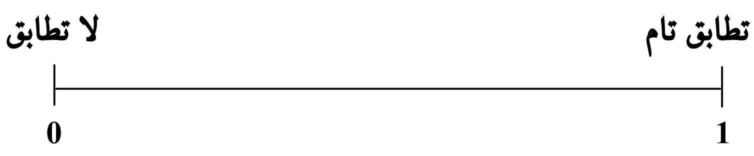
* طيور اللاقلاق: هي مجموعة متنوعة من الطيور التي تغوص في الماء بحثاً عن الطعام، من عائلة Ciconiidae وتكثر خصوصاً في نصف الكرة الشرقي وتمتاز هذه الطيور بأرجل طويلة ومناقير مستقيمة.

الإقامة) تم تفسيره بواسطة التباين في المتغير المستقل (الدخل): فالفرق بين هاتين الحالتين فيما يتعلق بمكان الإقامة يمكن تفسيره بمجرد الإشارة إلى الفرق في دخول هؤلاء الأشخاص. وعلى النقيض من ذلك فإن الجدول التالي يتعلق بحالة لا يوجد بها تطابق.

جدول (7-4): لا يوجد تطابق

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	178 (%55)	219 (%55)	397
ريفي	145 (%45)	178 (%45)	323
المجموع	**323**	**397**	**720**

إن التباين في الدخل في هذا المثال، يشير إلى أنه لا توجد علاقة للتباين في مكان الإقامة. فقد جاءت النسبة متساوية لدى الدخول المنخفضة للذين يعيشون في المدينة كما هو الحال لأولئك الذين لديهم دخول عالية. وكل واحد من هذه الفئات للمتغير المستقل هي بالضبط مساوية لنفس نمط الإجابات الموجودة في المتغير التابع. إن هاتين الحالتين اللتين تشيران إلى عدم التطابق والتطابق التام من المتغيرين اللذَيْن يقعان بشكل معاكس في نهاية المقياس. فإن حالة لا تطابق أعطيت قيمة 0 وإن علاقة التطابق التام قد أعطيت قيمة 1. انظر الشكل التالي:



شكل (7-1) ميزان مقاييس التطابق الاسمية

إننا في واقع الأمر لا يمكننا جمع بيانات تكون مطابقة لأحد هذين الطرفين المتطرفين. إلا أنهما ببساطة يقومان بمهمة نقاط مرجعية. فالبيانات عادة تقع في مكان ما بين المثال الذي سنقوم بتوضيحه.

جدول (7-5) مكان الإقامة وفقا لمستوى الدخل: تكرارات مشاهدة

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	230	397
ريفي	205	118	323
المجموع	**372**	**348**	**720**

وبالنظر إلى جدول التطابق يتبين لنا وجود بعض العلاقة بين هذين المتغيرين ولكنه من الوضوح أيضاً أن هذه الحالة ليست حالة تطابق تام. فإذا أردنا أن نعطي العلاقة القوية في هذا الجدول عدد يكون بين 0 و 1، فالصفر (0) يشير إلى عدم وجود تطابق في حين يشير رقم (1) إلى تطابق تام.

إن حساب قيمة لامبيدا يعطينا الموقع للعدد الدقيق على طول المتصل في الشكل (1) بحيث نعرف أن النتيجة الفعلية تقع على طول المتصل، ونقوم بذلك من خلال قياس المسافة الإحصائية بين الجدول الذي يحتوي على البيانات الفعلية التي نشاهدها، وأن كل واحد من الاحتمالين للمواقف المتطرفة لعدم وجود تطابق ووجود تطابق تام.

لامبيدا هي واحد من مجموعة المقاييس التي يطلق عليها نسبة التخفيض في الخطأ Proportional reduction in Error (PRE).

منطقة نسبة التخفيض في الخطأ (PRE):

إن منطق هذه المقاييس يتطلب منا أن نقوم بنوعين مختلفين من التنبؤ حول درجات الحالات المدروسة. إن أول هذه التنبؤات ينبغي علينا أن نتجاهل المعلومات المتعلقة بالمتغير المستقل، في هذه الحالة سنقوم بعدة أخطاء بالتنبؤ بالدرجة على المتغير التابع. أما التنبؤ الثاني، فإننا نأخذ في الاعتبار درجة الحالة على المتغير المستقل لمساعدتنا في التنبؤ بالدرجة على المتغير التابع. فإذا كان هناك تطابق بين المتغيرين، فإننا بالتالي نولّد أخطاء قليلة عندما نأخذ في الاعتبار المعلومات المتعلقة بالمتغير المستقل.

إن نسبة التخفيض في الخطأ (PRE) لمقاييس التطابق تعبر عن نسبة التخفيض في الخطأ بين اثنين من التنبؤات. وبتطبيق هذه الأفكار العامة لحالة المتغيرات المقاسة على المستوى الاسمي سيجعل هذا المنطق أكثر وضوحاً، فالمتغيرات الاسمية، تقوم أولاً بالتنبؤ بالفئة التي سوف تقع على المتغير التابع (y) بينما تتجاهل المتغير المستقل (X). ولما كنا نتنبأ على نحو أعمى في هذه الحالة، فإننا سنرتكب أخطاء كثيرة (هذا يعني أنه في الغالب سنتنبأ بقيمة حالة على المتغير التابع بشكل غير صحيح).

إن التنبؤ الثاني يسمح لنا أن نأخذ المتغير المستقل في الاعتبار. فإذا كان هناك تطابق بين المتغيرين، فإن المعلومات التي يوفرها المتغير المستقل سوف تخفض أخطاء التنبؤ. وكلما كان التطابق بين المتغيرين قوياً، كان هناك تخفيض كبير في الأخطاء. وفي حالة التطابق التام لن تكون هناك أخطاء على الإطلاق عندما نقوم بعملية التنبؤ بدرجة y من درجة على X. وعندما لا يوجد تطابق بين المتغيرين، على الجانب الآخر، فإن المعرفة بالمتغير المستقل لن تحسن من دقة تنبؤاتنا. فإننا سوف نولد تماماً أخطاء كثيرة في التنبؤ بمعرفتنا بالمتغير المستقل مثلما فعلنا بدون معرفة المتغير المستقل.

والمثال التالي سوف يجعل هذه المبادئ أكثر وضوحاً. افترض أنك وجدت في موقف مألوف للتنبؤ بما إذا كان كل واحد من 100 شخص الذينَ قد قابلتهم سيكونون طوالاً أو قصاراً أكثر من 5 قدم و9 أنش (5.9) في القامة تحت ظرف أنك لا تمتلك أية معرفة حول هؤلاء الناس على الإطلاق. ولما كانت لا تتوفر أية معلومات حول هؤلاء الناس، فإن التنبؤات ستكون في الغالب خاطئة.

افترض الآن أنك مررت بهذه المحنة مرتين: ولكن في الجولة الثانية كان لديك معلومات حول جنس الشخص الذي تتنبأ بقامته. ولما كانت القامة مرتبطة بالنوع، وأن الإناث في المتوسط أقصر من الذكور، فإن الإستراتيجية الأفضل ستكون بأن التنبؤ سيكون أن كل الإناث قصيرات، وأن كل الذكور طوال. إلا أنه يمكننا القول، بأننا لازلنا نولد أخطاء في الجولة الثانية. ولكن إ ذا كان المتغيران مرتبطين فإن عدد الأخطاء سيكون أقل. بمعنى أنه إذا تم استخدام المعلومات حول المتغير المستقل سوف يخفض عدد الأخطاء.

فإذا رجعنا إلى متغير الدخل ومكان الإقامة ووجدنا أنهما قادران على التكهن حول أين يعيش شخص ما من خلال معرفتنا بمستوى دخله، وكلما كانت العلاقة قوية كان توقعنا صحيحاً.

إن كل مقاييس نسبة التخفيض في الخطأ تتبع نفس الإجراء. إننا نحاول أن نتنبأ كيف أن الحالات ستتوزع في الجدول الثنائي تحت شرطين:

1- يمكننا التكهن بتوزيع الحالات على طول المتغير التابع بدون أية معرفة في درجات هذه الحالات للمتغير المستقل.

2- يمكننا التكهن بتوزيع الحالات على طول المتغير التابع بمعرفتنا لدرجات هذه الحالات للمتغير المستقل، ولكي نبين كيف يمكننا أن نقوم بهذه التكهنات دعنا نفترض أن 720 شخصاً في المسح الذي نقوم به قد اصْطَفُّوا جميعاً خارج الحجرة ثم بدأ كل واحد بالتحرك تلو الآخر وقبل دخول أي شخص إلى الحجرة ينبغي علينا التخمين (التنبؤ) بما إذا كان هؤلاء يعيشون في مناطق ريفية أو حضرية (التنبؤ بدرجاتهم على المتغير التابع). ولفعل مثل هذه التوقعات فإننا تحصلنا فقط على جزء يسير من المعلومات التي تشير إلى أن الغالبية من كل الـ 720 شخصاً يعيشون في مناطق حضرية.

إن السؤال الذي يمكن طرحه هنا: ما هو التخمين الذي يمكننا فعله قبل دخول أي شخص إلى الحجرة؟ إن معرفتنا فقط بأن الغالبية من الناس يعيشون في مناطق حضرية

هي أفضل تخمين للتكهن بأنّ كل 720 شخصاً يعيشون في مناطق حضرية. بمعنى آخر، عندما لا تكون لدينا معلومات أخرى، فإنه يمكن تخمين المتوسط! إن هذا في حقيقة الأمر هو استخدام نموذج لا تطابق الذي بيناه في جدول (4). كقاعدة للتنبؤ. فإذا لم تكن هناك علاقات كبيرة بين هذين المتغيرين فإن قاعدة التنبؤ سوف تولد لنا أخطاء قليلة جداً.

وعليه فإنه كلما كنا قريبين من أنماط الحالات الفعلية التي تشبه نموذج لا تطابق فقليل جداً من الأخطاء يمكن عملها عند استخدام هذه القاعدة للتنبؤ أين يعيش شخص ما؟ ففي المثال الذي بين أيدينا إذا ما توقعنا أن كل 720 شخصاً يعيشون في المناطق الحضرية، فإننا قد نعمل خطأ في التنبؤ يصل إلى 323 هذا هو الرقم لعدد الناس من المناطق الريفية التي لم نفلح في التكهن بأنهم يعيشون في مناطق حضرية، ونطلق على هذا الخطأ خطأ رقم (1) (E_1). $E_1 = 323$.

دعنا الآن نفترض أن نفس 720 شخصاً تم إخراجهم من الحجرة وطُلب منهم أن يدخلوا مرة ثانية إلى الحجرة بشكل عشوائي واحد تلو الآخر. وسنجد في هذه المرة أنه قبل دخول أي شخص قد تم إخبارنا فيما إذا كان هؤلاء الناس لديهم دخل منخفض أو عالٍ وفي شكنا أنه لا يوجد تطابق بين الدخل ومكان الإقامة فإن مثل أولئك الناس ذوي الدخول المنخفضة يعيشون في مناطق ريفية وأن الناس ذوي الدخول العالية يعيشون في مناطق حضرية فإنه بإمكاننا التكهن بأن الشخص ذا الدخل المنخفض يعيش في الريف وأن الشخص ذا الدخل العالي يعيش في المدينة. إننا في هذه الحالة وبشكل فعال مستخدمون نموذج التطابق التام كما أشرنا إليه في الجدول (3).

وباتباعنا لقاعدة التكهن فإننا سنولد 285 خطأً. هذه الأخطاء جاءت من 167 لأولئك الذين يكسبون دخولاً منخفضة من المناطق الحضرية الذين تم تصنيفهم خطأً على أنهم يعيشون في الريف، هذا بالإضافة إلى 118 منخفضاً يكسبون دخلاً عالياً من المناطق الريفية الذين تم التكهن خطأً بأنهم يعيشون في المدن. ونطلق على مثل هذا الخطأ خطأ (2) (E_2). $E_2 = 167 + 118 = 285$.

إن السؤال الذي يمكن طرحه هو ما إذا كنا قد ولّدنا أخطاء قليلة عندما أعطيت لنا

معلومات إضافية حول الدرجة لكل حالة على المتغير المستقل. هل شكنا حول احتمالية التطابق بين المتغيرين تخفض معدل الخطأ عند توليدنا لهذه التنبؤات أم لا؟ إن التقليل في الأخطاء يكون 38 = 285 - 323. إذاً نحن ولدنا 38 من الأخطاء القليلة باستخدامنا لقاعدة التنبؤ التطابق التام أكثر منه عند استخدامنا لقاعدة التنبؤ لعدم التطابق مشيرين إلى أنه توجد بعض العلاقة في البيانات.

من خلال لامبيدا يمكننا حساب هذا التخفيض في الأخطاء كنسبة لـ (E_1).

$$\lambda = \frac{E_1 - E_2}{E_1}$$

حيث إن: E_1 هي عدد الأخطاء بدون معلومات حول المتغير المستقل.

E_2 عدد الأخطاء بمعلومات حول المتغير المستقل وكعملية تناسب فإن معدل الخطأ قد تم خفضه:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{323 - 285}{323}$$
$$= 0.12$$

وعليه، بعد حصولنا على المعلومات حول مستوى دخل شخص ما (المتغير المستقل) فإنه بإمكاننا أن نقلل من الأخطاء إلى الحد الأدنى عندما نتكهن بـِ(أين يعيش شخص ما بـِ 12 %)؟ إن هذه النتيجة تبين لنا الميزة الكبيرة لمقاييس التخفيض في الخطأ (PRE): إن هذه المقاييس تقيس شيئاً ذا معنى يغير في التخفيض من معدلات الخطأ، وعليه يمكننا أن نتحصل على تفسيرات محددة. وكما نرى في شكل رقم 2 أن النتيجة وضعت التوزيعات المشاهدة للبيانات أكثر قرباً من أقصى لا تطابق من وجود تطابق تام.

شكل (7-2)

ومن هنا نجد أن لامبيدا تعطينا طريقة قاطعة، إنه بالرغم من وجود بعض العلاقة بين هذين المتغيرين، إلا أنها ليست علاقة قوية جداً. فنحن عادةً ما نتحدث بأن التطابق بين المتغيرات إما أن يكون ضعيفاً أو متوسطاً أو قوياً. وعليه لا يوجد خط فاصل دقيق يحدد متى تكون قيم مقاييس نسبة تخفيض الخطأ (PRE) ضعيفة؟ ومتى تكون قوية؟ إلا أنه يمكننا الاسترشاد بالتالي:

جدول (7ـ6) تفسير قيم لامبيدا

المدى (±)	القوة النسبية
0.0 – 0.2	ضعيفة جداً، جديرة بالإهمال
0.2 – 0.4	ضعيفة، تطابق منخفض
0.4 – 0.7	تطابق متوسط
0.7 – 0.9	قوية، عالية، تشير إلى تطابق
0.9 – 1.0	عالية جداً، علاقة قوية جداً

3- خصائص لامبيدا [3]:

تمتلك مقاييس التطابق للامبيدا خصائص معينة، بعضها يكون مرغوباً، في حين أن بعضها الآخر لا يكون كذلك من حيث التطبيق. وسوف نصف باختصار الخصائص الرئيسية في التالي:

أ- قيمة لامبيدا تصل دائماً إلى 1 عندما تشير البيانات إلى وجود تطابق تام. وإذا ما نظرنا إلى الطريقة التي تبنى بها لامبيدا فإن لامبيدا سيكون لها خصائص مقبولة في حالة التطابق التام الذي يصل إلى 1. فإذا كان لدينا تطابق تام بين المتغيرين، فإن البيانات سوف تتطابق تماماً مع قاعدة التنبؤ الثانية التي ينتج عنها عدم وجود أخطاء. وعندما لا تكون هناك أخطاء من خلال المعلومات المتعلقة بالمتغير المستقل

($E_2 = 0$) حينئذٍ ستكون معادلة لامبيدا ببساطة:

$$\lambda = \frac{E_1}{E_1} = 1$$

ب- إن قيمة لامبيدا غالباً تساوي 0 عندما تظهر البيانات أنه لا يوجد تطابق. فإذا لم يكن هناك تطابق في البيانات، فإن النتيجة المشاهدة سوف تطابق تماماً "نموذج لا تطابق"، وباستطاعتنا التكهن باستخدام "نموذج لا تطابق" الذي يولد لنا أخطاء ($E_2 = 0$). وهذه الأخطاء سوف تولد لنا قيمة 0 للامبيدا.

ج- في بعض الأحيان ستكون لامبيدا مساوية لِـ 0 عندما تُظهر البيانات بعض التطابق. مع أن لامبيدا دائماً سوف تكون مساوية لِـ 0 عندما لا يوجد تطابق. والعكس ليس دائماً بالضرورة صحيحاً، ففي بعض الأحيان عندما تكون لامبيدا مساوية لِـ 0 فإنه قد يكون حقاً هناك تطابق. وهذا يمثل القصور الأساسي في استخدام لامبيدا والذي سنشير إليه في موضعه في نهاية هذا الفصل.

د- تعتبر لامبيدا مقياساً للتطابق اللاامتماثل Asymmetric. وهذا يعني أن قيمة لامبيدا سوف تكون مختلفة اعتماداً على الطريقة التي تم بها التفكير حول المتغيرين أيهما متغير مستقل وأيهما متغير تابع. بمعنى آخر، إذا ما حاولنا في المثال السابق أن نتكهن بمستويات الدخول استناداً على أين يعيش الأفراد مفضلين ذلك على الطريقة الأخرى، فإن قيمة لامبيدا ستتغير. وعليه عند استخدامنا للامبيدا ينبغي علينا أن نكون واضحين حول طبيعة العلاقة التي نعتقد أنها تربط هذين المتغيرين معاً. إن هذه الطريقة تجعل من لامبيدا طريقة مفيدة عندما يكون لدينا أسباب قوية للاعتقاد أنه توجد طريقة واحدة للعلاقة بين المتغيرين تسير في اتجاه محدد.

ولتوضيح كل ذلك دعنا نقلب المتغيرات في المثال السابق وأن نتعامل مع مكان الإقامة كمتغير مستقل، ومستوى الدخل كمتغير تابع. إن هذا الأمر يتطلب منا بناء جدول توافق حيث نضع متغير مكان الإقامة عبر الأعمدة والدخل وفقاً للصفوف (جدول 7 – 7).

جدول (7-7) مكان الإقامة وفقاً لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	205	372
ريفي	230	118	348
المجموع	**397**	**323**	**720**

وبدون سابق أية معرفة لتوزيع المتغير المستقل (مكان الإقامة) فإنه باستطاعتنا التكهن بأن كل الحالات التي تقع في المجموعة ذات المستوى الاقتصادي المنخفض، لذلك فإن سوء التصنيف لـِ 348 شخصاً الذين يكسبون في الحقيقة دخلاً اقتصادياً عالياً:

$$E_1 = 348$$

وإذا ما أردنا أن نتكهن بمستويات الدخول لدى الناس، إلا أننا في هذه المرة لدينا معلومات فيما إذا كان هؤلاء يعيشون في مناطق ريفية أو حضرية، عندئذ يكون التكهن بأن كل الناس من المناطق الحضرية لديهم دخول عالية، ومن هنا نكون قد وقعنا في 118 خطأ. عليه فإن المعدل الكلي للخطأ عندما استخدمنا نموذج التطابق التام هو:

$$E_2 = 167 + 118 = 285$$

وبالتالي تصبح قيمة لامبيدا باستخدام نموذج العلاقة:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{348 - 285}{348}$$

$$= 0.18$$

بمعنى آخر فإن التطابق يعتبر - إلى حد ما - ضعيفاً، إلا أن هذا التطابق قد يمدنا بالإحساس بأننا قادرون على أن نتوقع مستوى دخل شخص ما استناداً إلى المكان الذي يعيش فيه. فالتطابق بين هذين المتغيرين يعتبر إلى حد ما أقوى عندما نظرنا إلى مستوى الدخل كمتغير تابع بدلاً من مكان الإقامة.

إجراءات توليد لامبيدا Lambda باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ descriptive statistics ⟸ Crosstabs

2- انقر على المتغير في القائمة التي سوف تشكل صفوف الجدول، والتي تكون في هذه الحالة Place of Residence..

3- انقر على ◀ التي تشير إلى المساحة المعنونة Row(S)، تقوم بلصق Place of Residence في القائمة المحددة للصفوف Row(S) .

4- انقر على Income Level.

5- انقر على ◀ التي تشير إلى المساحة المعنونة Column(s). تقوم بلصق Income Level في القائمة المحددة للأعمدة Column(s).

6- انقر على زر Statistics، يقوم بإعطائنا صندوق Statistics: Crosstabs وفي أعلى زاوية من الشمال يمكننا رؤية المساحة المعنونة Nominal Data هذه مقاييس التطابق المتوفرة عندما يكون على الأقل أحد المتغيرات مقاساً على المستوى الاسمي. في هذا المثال Place of residence قد تم قياسه على المستوى الاسمي.

7- اختر Lambda بالنقر على الصندوق الموجود أمامها وضع علامة T في الصندوق لتبين أنه قد تم اختيار لامبيدا.

8- انقر على Continue.

9- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Directional Measures

	Value	Asymp std. Error[a]	Aprrox. T[b]	Aprrox. Sig
Nominal by Lambda Symmetric	.151	.047	3.061	.002
Place of residence Dependent	.118	.056	1.976	.048
Income Level Dependent	.181	.052	3.184	.001
Goodman Place of residence Dependent	.045	.015		.000 c
And Kraskal-tau Income Level Dependent	.045	.015		.000 c

a. Not assuming the null hypothesis.
b. using the asymptotic standard error assuming the null hypothesis.
c. Based on chi-sguare approximation

في هذه المخرجات تم استخدام عملية التماثل عندما لا يكون لدينا أي سبب لأن نشك في أن واحداً من هذين المتغيرين يكون معتمداً على الآخر، بقدر ما يكون هذان المتغيران - تبادلياً - معتمدَيْنْ على بعضها البعض، ففي هذا المثال قيمة التماثل تساوي 0.151 تقع في مكان ما بين القيم اللاامتماثلة 0.118 و 0.181. إن عملية استخدام اللاتماثل تكون لديها قيمتان احتماليتان استناداً على أي من المتغيرين يعتقد أنه متغير تابع.

كذلك تحتوي هذه المخرجات على قيمة متعلقة بمقياس اسمي. ولقياس التطابق ألا وهو جودمان وكرشكال تاو Goodman and Kruskal tau والذي يحتوي على قيمة صغيرة للتطابق إذا ما قورن بـِ لامبيدا Lambda في الأعمدة الأخرى من هذا المربع تحتوي على معلومات ليست ذات علاقة بهذا الفصل، وإنما ستكون هذه المعلومات ذات أهمية في مواضع أخرى من هذا الكتاب خاصة عند التعامل مع الاستدلال من العينة على المجتمع [4].

حدود استخدام لامبيدا [5] :

على الرغم من الطريقة السهلة لحساب لامبيدا إلا أن هناك مشكلة عادة ما تقابلنا عند استخدام لامبيدا. إن هذه المشاكل التي تناولناها عند الحديث عن خصائص لامبيدا يمكن أن يكون لديها قيمة 0 حتى ولو كانت العلاقة موجودة بين المتغيرين. (قد تكون واضحة بمجرد النظر إلى جدول التقاطع) إن سبب هذه المشكلة يكمن في أن البيانات تنحرف بشكل كبير على طول المتغير التابع (إن قيمة لامبيدا ستكون 0 عندما تكون فئة النموذج للمتغير التابع مساوية لكل الفئات المتعلقة بالمتغير المستقل).

ولكي نرى ذلك عملياً، سوف نقوم بتحليل البيانات الواردة في الجدول التالي (7 - 8):

في هذا المثال النظري قد تم طرح السؤال المتعلق بما إذا كانت دائرة التشغيل تقوم بجهد كبير للتخفيف من مشكلة البطالة.

جدول (7-8):

هل يجب على دائرة التشغيل أن تبذل جهداً أكبر للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	110 % 18	168 % 42	278
نعم	490 % 82	232 % 58	722
المجموع	600	400	1000

وبالنظر إلى جدول التقاطع فإنه يمكننا القول إنه توجد علاقة محدودة. فكلما زادت نسبة الأفراد الذين تكون أعمارهم تحت 45 سنة يوافقون على السؤال المطروح حول سياسة التشغيل أكثر من أولئك الذين تكون أعمارهم 45 سنة فأكثر. وبوضوح، يمكننا أن نرى أن هناك بعض الاعتماد بين المتغيرين، والذي يمكننا أن نصفه بتعبيرات لفظية بالقول إن هناك علاقة جيدة بعض الشيء إلى متوسطة في القوة.

وعليه إذا أردنا أن نحاول تكميم هذه العلاقة باستخدام لامبيدا فإننا نحصل على تطابق متناسب مساوياً لـِ 0. (لاحظ الإجابة للمتغير التابع لكل الحالات 1000 هي "نعم" والتي هي أيضاً كذلك الإجابة لكل من الفئتين للمتغير المستقل: فغالبية الناس تحت الـِ 45 سنة أجابوا بنعم، وأن الغالبية من الناس ذوي الأعمار 45 سنة فما فوق قد أجابوا هم أيضاً بنعم).

إن هذا التوزيع المنحرف في إطار المتغير التابع سوف يولد لنا لامبيدا تساوي 0، حتى كان واضحاً لدينا من خلال العين المجردة أن هناك وجوداً لبعض التطابق بين المتغيرين. ولكي نبين ذلك فإننا نحتاج أولاً: إلى حساب عدد الأخطاء عند التكهن بدون معلومات حول المتغير المستقل (العمر)، إننا نتوقع أن كل حالات الـِ 1000 سوف تقع

في فئة "نعم"، حيث إن هذا الأمر سوف يقلل معدل الخطأ حيث يصل معدل الخطأ في هذه الحالة إلى 278 خطأ: $E_1 = 278$.

أما إذا كانت لدينا معلومات حول المتغير التابع فإننا سوف نبقى على نفس الأخطاء. فإذا ما نظرنا إلى إجابة المبحوثين الأولى أقل من 45 سنة فإننا نتوقع أن 600 حالة قد أجابت "بنعم" (110خطأ).

ثانياً: يمكننا التوقع أن 400 شخص الذين تبلغ أعمارهم 45 سنة فما فوق قد أجابوا "بنعم" (168 خطأ).

وهذا المجموع الذي يصل إلى 278 خطأ هو نفس التوقعات بدون معرفة عمر المبحوث: $E_2 = 278$. وأن قيمة لامبيدا ستكون:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{278 - 278}{278}$$

$$= 0$$

ومن هنا نجد أن لامبيدا قد فشلت في الحصول على علاقة مشاهدة والتي كانت واضحة لنا بالعين المجردة. إن هذا الأمر قد يلقي بعض الضوء على قاعدة مهمة: متى كانت لامبيدا مساوية لـِ 0 ينبغي معاينة التوزيع التكراري النسبي لنقرر فيما إذا كانت هذه التكرارات تعكس في الواقع لا علاقة أو فيما إذا كانت هذه العلاقة جاءت نتيجة لانحراف التوزيع المتعلق بالمتغير التابع. إذا فحصنا نسب العمود تقودنا إلى الاستنتاج أن قيمة 0 للامبيدا هي ناتجة من توزيع منحرف (كما في هذه الحالة) توجد لدينا ثلاثة خيارات:

1- ينبغي على الباحث ألا يشغل باله بمقاييس التطابق وعليه أن يتقيد بجدول التطابق وما يحتويه من توزيعات نسبية، وأن يكون الأساس في نتيجته هو النظر إلى العلاقة في ذاتها فقط. ويتطلب هذا الأمر من الباحث أن يقوم ببعض الأحكام الذاتية، ولكن طالما أن جدول التقاطع يسمح للقراء أن يقيموا بأنفسهم، فإن ذلك لا يشكل أي مشكلة في بناء هذه المحاولة باستخدام التوزيعات النسبية كشواهد. إن هذه

التوزيعات في بعض الأحيان "تتحدث بنفسها". فحساب إحصاءات متقدمة أكثر قد يؤدي إلى اختفاء معلومات مهمة في كتلة من الأعداد المشكوك فيها.

2- ينبغي على الباحث حساب مقاييس أخرى للتطابق حيث توجد مجموعة أخرى من مقاييس التطابق للبيانات المقاسة على المستوى الاسمي والتي يمكن استخدامها إذا ما واجهت الباحث بعض المشكلات المتعلقة بلامبيدا مثل مقاييس جودمان - كروشكل تاو Goodman-Kruskal tau. إن هذا المقياس يشبه مقياس لامبيدا في كونه مقياس لا تماثل للتطابق حيث تتراوح درجاته بين 0 و 1 ولكنه في الجانب الآخر لا يشبه لامبيدا لكونه لا يستخدم إجابة النموذج للمتغير المستقل عند إجراء عملية التنبؤات، ولكنه على العكس يستخدم بدلاً من ذلك التوزيع التكراري للحالات عبر كل الفئات للمتغير المستقل. ولما كان هذا الأمر أقل حساسية لانحراف التوزيعات الهامشية من لامبيدا بالتالي فهو خيار ملائم عندما تسبب عملية الانحراف قيمة لامبيدا المساوية لصفر.

ومن المقاييس الأخرى للتطابق هو Cramer`s V. وهذا المقياس دائماً يولد قيمة أكبر من 0 عندما يكون هناك تطابق بين المتغيرين. إلا أنه في الواقع لا يمتلك تفسيرات بسيطة فيما يتعلق بنسبة التخفيض في الخطأ (PRE).

وعليه فإنه لا يمكن استخدامه لتقييم قوة العلاقة في أي جدول تقاطع. إلا أنه يمكن أن يكون مفيداً عند مقارنة قوة العلاقات الثنائية عبر مجموعة من الجداول. والمعادلة المستخدمة لِـ Cramer`s V هي:

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

حيث إنّ:

X^2 هي إحصاءات مربع كاي لجدول التقاطع.

K تشير إلى عدد الصفوف أو إلى عدد الأعمدة أيهما أصغر. إنّ Crame`s V هي إحدى الخيارات الموازية مع لامبيدا، للبيانات الاسمية.

3- إن الخيار الثالث يكمن في معايرة الجدول حتى يكون مجموع الصفوف كلها متساوية. إن مثل هذا الإجراء إجراء معقد إلى حد ما، فعملية المعايرة في الجدول تتضمن محاولة التقليل من التباين في البيانات التي جاءت من خلال التوزيع المنحرف في المتغير التابع، بينما لا يزال الاحتفاظ بالتباين عبر فئات المتغير المستقل. وفي تقرير يستخدم هذا الإجراء ينبغي أن يكون واضحاً أن لامبيدا لا يمكن حسابها ببيانات الصف، وبحساب لامبيدا فإن هوامش الصفوف قد عُيّرت ليصل مجموعها إلى 100. وعند التعامل مع لامبيدا فإننا نقنن هوامش الصف بحيث يصل مجموع كل صف إلى 100. وهذا الأمر يتطلب حساب نسب الصف، التي يمكن التعامل معها كما لو كانت هي الأعداد الحقيقية للحالات. هذا يعني، أننا قمنا بحساب نسبة مجموع "نعم" للمبحوثين الذين هم تحت الـِ 45 سنة، وحساب نسبة أولئك الذين تصل أعمارهم إلى 45 سنة أو تزيد. ونقوم بعمل الشيء نفسه لإجابة أولئك الذين أجابوا "بلا". انظر جدول (7 – 9).

جدول (7ـ9):

هل يجب على دائرة التشغيل أن تبذل جهدا كبيرا للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	$\frac{110}{278} \times 100 = \%40$	$\frac{168}{278} \times 100 = \%60$	% 100
نعم	$\frac{490}{722} \times 100 = \%68$	$\frac{232}{722} \times 100 = \%32$	% 100

وبعد ذلك يمكننا استخدام أعداد النسب كما لو حسبت كحالات فعلية كما في الجدول التالي:

جدول (7ـ10):

هل يجب على دائرة التشغيل أن تبذل جهدا كبيرا في التقليل من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فأكثر	
لا	40	60	100
نعم	68	32	100

تذكر أن هذه النسب: 40 تمثل 40 % من 278 لمجموع إجابات "لا". وهكذا. إلا أننا قد تعاملنا مع هذه النسب كما لو كانت حالات فردية. وهذا يعني أن الحجم الكلي للعينة هو 200 بدلاً من 1000: الذين أجابوا "بنعم" يصل عددهم إلى 100 والذين أجابوا "بلا" يصل عددهم إلى 100. وباستخدامنا لهذه البيانات من جدول المعايرة (10)، يمكننا إعادة حساب لامبيدا بدون معرفة حول المتغير المستقل، فيمكننا تصنيف كل الإجابات الـِ 200 إما "بنعم" أو "بلا"، وعليه يمكننا أن نتسبب في أخطاء تصل إلى 100.

$$E_1 = 100$$

أما إذا كانت لدينا معرفة حول المتغير المستقل فإنه يمكن أن نصل إلى التوقعات التالية، فإذا بدأنا بأولئك الأشخاص الذين تقل أعمارهم عن 45 سنة، يمكننا أن نتنبأ بأولئك الذين أجابوا "بنعم"، ولما كان هذا التنبؤ يعطينا أقل معدل في الخطأ (40 خطأ) ولأولئك الذين تصل أعمارهم 45 سنة فما فوق، يمكننا التكهن بأنّ كل هؤلاء قد أجابوا "بلا"، وعليه فإننا بذلك قد ولدنا 32 خطأ:

$$E_2 = 40 + 32 + 72$$

ومن هنا تكون قيمة لامبيدا مساوية لـِ:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{100 - 72}{100}$$

$$= 0.28$$

وعندما نقرب هذه القيمة (0.28) في المائة (100)، فإن قيمة لامبيدا تشير إلى تطابق ضعيف إلى متوسط، فيما يتعلق بنسبة التخفيض في الخطأ (PRE). وعليه، فإنه يمكن تفسير لامبيدا بالإقرار بأن معرفتنا بالعمر تساعد قدرتنا على أن نتنبأ بالاتجاه حول الجهد المبذول من دائرة التشغيل للتخفيف من البطالة بنسبة 28 %.

أسئلة للمراجعة:

1- بين الفرق بين لا تماثل Asymmetric و متماثل Symmetric، وما هي أنسب هذه المقاييس للاستخدام في موقف يكون فيه متغيران تبادلياً معتمدان على بعضهما البعض Mutually dependent؟

2- ما هو الشيء المهم عندما تقوم بحساب لامبيدا لنقرر ما إذا كان أحد المتغيرين يرجح أن يكون معتمداً على المتغير الآخر، وإذا كان الأمر كذلك، حدد ما هو المتغير التابع وما هو المتغير المستقل؟

3- احسب لامبيدا من البيانات التالية، وبين قوة أي علاقة؟

أ:

تابع	مستقل 1	مستقل 2	المجموع
1	30	60	90
2	45	50	95
مجـ	75	110	185

ب:

تابع	مستقل 1	مستقل 2	مستقل 3	المجموع
1	56	40	10	106
2	15	30	50	95
مجـ	71	70	60	201

ج:

تابع	مستقل			المجموع
	1	2	3	
1	70	40	10	106
2	50	45	38	133
3	43	30	14	87
مجـ	163	115	62	326

4- يرغب باحث في معرفة الاتجاه نحو التعليم المختلط. وقد تم بحث 3000 طالبٍ وطالبةٍ، وجاءت النتيجة كالتالي:

الاتجاه	النوع		المجموع
	ذكور	إناث	
أوافق	849	367	
لا أوافق	191	1593	

احسب قيمة لامبيدا وفسرها.

5- احسب قيمة Cramerś V للبيانات التالية:

أ- $X^2 = 3.5, N = 20$

2 = Culmns (الأعمدة)، 4 = rows (الصفوف)

Rows = 2, Culmns, $X^2 = 9.8, N = 90$ = 4

Rows = 3, Culmns, $X^2 = 12, N = 800$ = 3

6- مسح لعدد 50 طالبةً من الإناث و50 طالباً من الذكور، طرح عليهم سؤال هل تعرف تاريخ وضع حجر الأساس لمشروع النهر الصناعي العظيم؟

الذكور أجابوا "بنعم" 29

الإناث أجابوا "بلا" 21

الإناث أجابوا "بنعم" 22

الذكور أجابوا "بلا" 28

- رتب هذه البيانات في جدول تقاطع Crosstabs. وبين المتغير المستقل والمتغير التابع.
- احسب قيمة لامبيدا لهذه البيانات؟

7- دراسة توصلت إلى نتيجة مفادها أن التطابق بين متغيرين باستخدام Cramerś V، 0.34. وفي دراسة سابقة قامت بقياس التطابق بين نفس هذين المتغيرين باستخدام V كانت النتيجة تتراوح بين 0.15 إلى 0.21. كيف لهؤلاء الباحثين تفسير هذه النتائج؟

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, P.153.

2- Ibid, PP. 154 - 155.

3- Ibid, PP. 159 - 161.

4- Ibid, P. 163.

5- George Argyrous, opt. Cit., PP. 165 - 196.

6- And, Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010. pp. 295 .

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

2- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

3- William E. Wagner, III, using SPSS for Social Statistics and Research Methods 2nd ed, Pine forge Press An Inprint of Sage, USA, 2010.

الفصل الثامن

التطابق بين متغيرات تم قياسها على المستوى الترتيبي

الوصف العددي للبيانات الترتيبية

مقدمة

نحاول في هذا الفصل التركيز على مقاييس التطابق لمتغيرات تم قياسها على المستوى الترتيبي لنسبة التخفيض من الخطأ (PRE) التي هي مشابهة لمقياس لامبيدا من حيث المنطق الأساسي لهذه المقاييس والكيفية التي يتم بها تفسير البيانات. وعند حديثنا عن لامبيدا فإننا نحاول من خلالها التنبؤ بقيمة حالة فردية أخذت للمتغير المستقل. والهدف من وراء ذلك هو أولاً: افتراض أنه لا يوجد تطابق بين المتغيرين. وثانياً: بافتراض أنه يوجد تطابق تام بين المتغيرين. وبمقارنة معدلات الخطأ المتعلق بكل قاعدة من قواعد التنبؤ فإننا نستطيع تقييم العلاقة الفعلية التي تحتويها مجموعة البيانات التي تم جمعها.

وتجدر الإشارة إلى أننا سنتبع نفس الإجراء مع البيانات الترتيبية، مع الأخذ في الاعتبار أننا سوف نستخدم بيانات إضافية حول المتغيرات المعطاة للتعامل معها وفقاً لمستوى البيانات الترتيبية، تختلف البيانات الترتيبية عن البيانات الاسمية، غير أننا في البيانات الترتيبية على دراية بالكيفية التي رتبت بها الحالات. وعليه فإننا نحاول أن نتنبأ

بالوضع الترتيبي لحالات الأزواج Pairs of Cases، وقاعدة مقاييس التطابق على أساس النجاح في تنبئنا لهذه الرتب. فإذا ما تعاملنا مع المثال الموضح في جدول (1) فإنه بإمكاننا أن نستمر في تكميم العلاقة القوية الموجبة ونلاحظ من خلال حساب مقاييس التطابق للبيانات الترتيبية ذات الصلة أنه توجد لدينا مجموعة من مقاييس التطابق المتعلقة بنسبة التخفيض في الخطأ التي يمكن حسابها من هذا الجدول مع تباين بسيط في طرقها الخاصة؛ وبذلك فإن كل المقاييس المتعلقة بالبيانات الترتيبية والتي سوف نناقشها لديها خصائص مشتركة تستند على التميِّز بين الأزواج المتوافقة Concordant Pairs والأزواج غير المتوافقة Disconcordant Pairs.

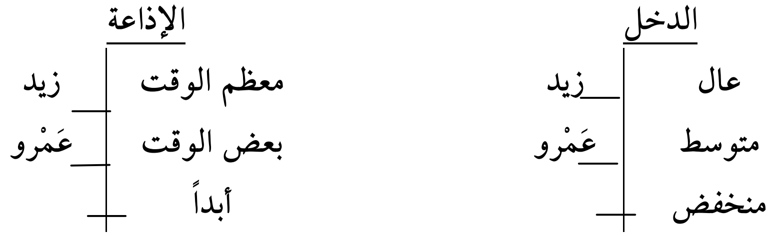
جدول (8-1): توزيع مشاهدة الإذاعة المرئية وفقاً للدخل

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 %71.5	15 %15.0	10 %10.5	100
بعض الوقت	20 %19.0	70 %70.0	10 %10.5	100
معظم الوقت	10 %9.5	15 %15.0	75 %79.0	100
المجموع	**105**	**100**	**95**	**300**

المصدر: George Argyrous, op.cit, p. 173

1- الأزواج المتوافقة:

نفترض أن واحداً من 75 في فئة الدخل العالي في جدول (1) يشاهد الإذاعة المرئية معظم الوقت ويدعى زيداً، وأن واحد من 70 في فئة الدخل المتوسط يشاهد الإذاعة المرئية بعض الوقت ويدعى عَمْراً. إن هذين الشخصين يمكن ترتيبهما بحيث يخالف أحدهما الآخر على كل واحد من هذين المتغيرين. انظر شكل (8 – 1).



شكل (8-1) ترتيب الزوج المتوافق

إن ترتيب زوج هاتين الحالتين يمكن تلخيصه في الجدول التالي:

جدول (8-2): حالات الزوجين المتوافقين

المتغير المستقل: الدخل	المتغير التابع: الإذاعة المرئية
زيد رتب فوق عَمْرو (لديه أعلى دخل)	زيد رتب أعلى من عَمْرو (مشاهدة أكثر للإذاعة المرئية)

وعليه فإن هاتين الحالتين قد تم ترتيبهما بشكل متساوٍ على كل واحد من هذين المتغيرين. وهذا قد يبدو غريباً في طرحه: حيث يمكن طرح السؤال التالي وهو: كيف يمكن أن يكون ترتيبهما متساوٍ إذا كان لكل منهما قيم مختلفة؟ إن القضية هي أن ترتيبهما متساوياً: إن زيداً قد رتب فوق عَمْرو على كلا المتغيرين. ونصف أزواج مثل هذه الحالات كزوج متوافق (Nc). ونعني بالزوج المتوافق: هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي رتب بشكل متساوٍ على كلا المتغيرين.

لقد أخذنا حالتين من كل الحالات البالغ عددها 300 والتي تشكل الزوج المتوافق. إن السؤال المطروح هو: هل نقوم بحساب العدد الكلي للأزواج المتوافقة التي يحتويها الجدول؟ ولفعل ذلك انظر إلى الخلايا المظَلَّلة في جدول التوافق التي من خلالها تم سحب كل من زيد وعَمْرو.

جدول (8ـ3) توزيع مشاهدة الإذاعة المرئية وفقا للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

في مناقشتنا أعلاه قد تم صياغة الزوج المتوافق بمماثلة زيد الذي هو واحد من الحالات التي تبلغ 75 حالة ولديه دخل عالٍ ويشاهد الإذاعة المرئية معظم الوقت. بعَمْرو الذي هو واحد من 70 حالة ذات الدخول المتوسطة ويشاهد الإذاعة المرئية بعض الوقت.

في واقع الأمر يمكننا أن نُقْرِنَ عَمْراً مع كل واحد وأي واحد من 70 حالة في الخلية التي تحتوي على دخل متوسط / ومشاهدة الإذاعة بعض الوقت يولد لنا 70 حالة من الأزواج المتوافقة: عَمْرو مضافاً إلى كل حالة من السبعين حالة الموجودة في الخلية الوسطى من الجدول (شاملة عَمْراً). وعندئذٍ يمكننا عمل الشيء نفسه لكل واحد من 74 حالة التي لديها دخل عالٍ وتشاهد الإذاعة المرئية معظم الوقت. إن هذا سوف يعطينا في المجموع النهائي 75 حصة في 70 أزواج متوافقة: $5250 = 70 \times 75$.

وبالنظر إلى جدول (8 – 4) يمكننا أن نرى أن هناك مجموعات متحدة التي بدورها تشكل الأزواج المتوافقة.

جدول (8ـ4): توزيع مشاهدة الإذاعة المرئية وفقاً للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

إن كل واحد من 75 حالة في أدنى الخلية اليسرى أيضاً تم ترتيبه فوق كل حالة من الحالات 15 في خلية أبداً / الدخل المتوسط كلاهما لديه دخل عالٍ ويشاهد الإذاعة المرئية أكثر. وبهذا سوف نضيف الأرقام التالية للأزواج المتوافقة:

$$75 \times 15 = 1125$$

في الحقيقة أن أي حالة سوف تشكل زوجاً متوافقاً مع أي حالة أخرى في الخلية التي هي فوق وإلى اليسار منها في الجدول وبالتالي فإن العدد الإجمالي للأزواج المتوافقة سيكون مبيناً في الجدول (8 – 5).

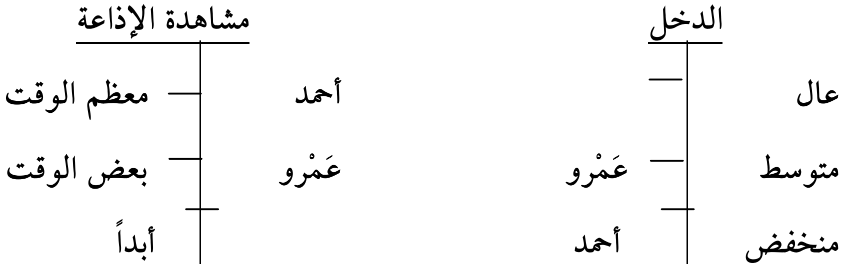
جدول (8-5): حساب الأزواج المتوافقة

75	15	10	
20	70	10	(75×70) +75×15) +75×20) + (75×75 = 13.500
10	15	75	
75	15	10	
20	70	10	$(10 \times 15) + (10 \times 75) = 900$
10	15	75	
75	15	10	
20	70	10	$(15 \times 20) + (15 \times 75) = 1425$
10	15	75	
75	15	10	
20	70	10	$(70 \times 75) = 5250$
10	15	75	

$$N_C = 13.500 + 900 + 1425 + 5250 = 21.075$$

2- الأزواج غير المتوافقة :

دعنا الآن نأخذ واحداً من عشرة أشخاص الذين لديهم دخل منخفض ويشاهدون الإذاعة المرئية معظم الوقت، ويدعى أحمد، ومقارنته بعَمْرٍو (هو واحد من 70 حالة ذوي الدخول المتوسطة ويشاهدون الإذاعة المرئية بعض الوقت)، فإن الترتيب لن يكون متساوياً لكلا المتغيرين فأحمد تم ترتيبه تحت عَمْرٍو فيما يتعلق بالدخل، ولكنه على الطرف الآخر رتب فوق عَمْرٍو فيما يتعلق بمشاهدة الإذاعة المرئية، انظر شكل (8 – 2). إن مثل هذه الحالات نطلق عليها الأزواج غير المتماثلة (Nd). ونعني بالزوج غير المتماثل هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي يرتب فيه متغير واحد بشكل مختلف في ترتيبه للمتغير الآخر.



شكل (8 ـ 2) ترتيب الزوج غير المتوافق

هذه الحالة سوف تشكل زوجاً غير متوافق مع الحالة الأخرى في الجدول أي في أي خلية أعلى وإلى اليمين. ولحساب العدد الإجمالي للأزواج غير المتوافقة ينبغي علينا أن نبدأ بالخلية اليسرى أسفل الجدول (8 – 6) وتماثلها مع كل الخلايا فوق وإلى اليمين منها.

جدول (8-6): حساب الأزواج غير المتوافقة

75	15	10	
20	70	10	$(10\times70) + (10\times15) + (10\times10) + (10\times10) = 1050$
10	15	75	
75	15	10	
20	70	10	$(20 \times 15) + (20 \times 10) = 500$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) + (15 \times 10) = 300$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	

$$N_d = 1050 + 500 + 300 + 700 = 5220$$

مقاييس التطابق للمتغيرات الترتيبية:

كل نسب التخفيض في الخطأ في مقاييس التطابق (PRE) تستخدم الفرق بين الأزواج المتوافقة والأزواج غير المتوافقة كأساس لتقييم ما إذا كان هناك تطابق وتحديد اتجاه هذا التطابق. إن السبب وراء النظر حول هذه الأزواج المتوافقة وغير المتوافقة أنها تعطينا معلومات يمكننا استخدامها في التنبؤ بما إذا كان لدينا متغيران متطابقان بشكل

موجب بعدئذٍ فإن جدول التقاطع سوف يحتوي أكثر الأزواج المتوافقة منه من الأزواج غير المتوافقة، والعكس بالعكس عندما يكون لدينا تطابق سالب.

وتجدر الإشارة إلى أنه عندما يكون هناك تطابق موجب بين المتغيرين فإن البيانات سوف تحتوي كثيراً من الأزواج المتوافقة وقليلاً من الأزواج غير المتوافقة. وإذا كان الأمر كذلك، فإن علمنا بأن شخصاً ما ترتيبه فوق شخص آخر فيما يتعلق بالدخل عندئذٍ يمكننا أن نتوقع أن ذاك الشخص يحتل كذلك مرتبة أعلى فوق الشخص الآخر فيما يتعلق بتوزيع مشاهدة الإذاعة المرئية.

التطابق الموجب: $\mathrm{N_c} - N_d > 0$

أما في حالة العلاقة السالبة بين المتغيرين فإن البيانات سوف تحتوي على كثير من الأزواج غير المتوافقة ومن هنا يمكننا الوصول إلى تكهن عكسي: فمعرفتنا بشخص رتب فوق شخص آخر فيما يتعلق بالدخل سيقودنا إلى أن نتكهن بأن هذا الشخص يحتل مرتبة أدنى من الشخص الآخر فيما يتعلق بتوزيعات مشاهدة الإذاعة المرئية:

التطابق السالب: $\mathrm{N_c} - N_d < 0$

أما فيما يتعلق بعدم وجود تطابق بين المتغيرين فإن البيانات ستحتوي تماماً على كثير من الأزواج المتوافقة والأزواج غير المتوافقة ومن هنا نكون غير قادرين على زيادة قدرتنا في التكهن بفئة المتغير التابع من خلال معرفتنا بفئة المتغير المستقل التي تقع فيها هذه الفئة:

لا تطابق: $\mathrm{N_c} - N_d = 0$

وتجدر الإشارة هنا إلى أن هناك أربعة أسس لنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية: سومرز Somerś d d وجاما Gamma وكنلدتاو Kendalś tau-b b وكنلدتاو Kendalś tau-c C. كل هذه المقاييس متشابهة في كونها تعطينا تفسيرات لنسبة التخفيض في الخطأ، كما أن كل هذه المقاييس تستخدم الفرق بين Nc و Nd كأساس لتقييم قوة العلاقة. إلا أن الفرق بين هذه المقاييس يكمن في كيف يمكننا معايرة هذا الفرق. ولمناقشة هذه المقاييس دعنا نبدأ بأبسطها المتمثل في جاما.

جاما Gamma :

تعتبر جاما من المقاييس الشائعة لـِ (PRE) لقياس التطابق بين متغيرين تم قياسهما على الأقل على المستوى الترتيبي وتم ترتيب هذين المتغيرين في جدول ثنائي. كما تعتبر جاما مقياساً متماثلاً للتطابق. وعليه فإن القيمة المحسوبة ستكون واحدة بغض النظر عن الكيفية التي يتحدد فيها المتغير المستقل والطريقة التي يحدد بها المتغير التابع. بمعنى آخر، إذا قلبنا الصفوف والأعمدة في الجدول بمعنى أن يكون الدخل تحت عند الصفوف ومشاهدة الإذاعة المرئية يكون عبر الأعمدة. فإن حساب جاما لن يتأثر. وعليه فإن جاما لن تكون حساسة لنموذج معين نعتقد أنه يصف العلاقة بين المتغيرين.

إن معادلة جاما تبين لنا الفرق بين عدد الأزواج المتوافقة والأزواج غير المتوافقة كنسبة من العدد الكلي للأزواج المتوافقة والأزواج غير المتوافقة وباستخدام البيانات من المثال السابق فإن حساب جاما يكون كالتالي:

$$G = \frac{N_C - N_d}{N_C + N_d} = \frac{21.075 - 2550}{21.075 + 2550}$$

$$= 0.78$$

وتشير هذه النتيجة إلى تطابق قوي موجب بين هذين المتغيرين، تلك النتيجة التي تعزز النتيجة التي وصلنا إليها استناداً إلى التحليل النظري لجدول التطابق. إنه من الوضوح أن المدى الممكن لقيم جاما تتراوح بين 1– و +1. فعندما تكون جاما 1– فإنها تشير إلى تطابق تام سالب: فمعرفة أن الحالة مرتبة فوق الحالة الأخرى على واحد من المتغيرين تشير إلى أنه يجب أن يرتب تحت المتغير الآخر. إن مثل هذه النتيجة نتحصل عليها إذا كان لدينا فقط الأزواج غير المتوافقة كما يوضحه الجدول التالي:

جدول (8ـ7)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام سالب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	0	0	%100
بعض الأوقات	0	%100	0
كل الأوقات	%100	0	0

أما على الجانب الآخر، إذا كان لدينا فقط أزواج متماثلة فإن قيمة جاما ستكون +1 مشيرة إلى تطابق موجب تام يعني: معرفتنا بترتيب حالة فوق الحالة الأخرى على المتغير المستقل تشير إلى أن هذه الحالة يجب أيضاً أن تكون مرتبة فوق المتغير التابع. إن مثل هذا الموقف ينعكس في الجدول التالي:

جدول (8ـ8)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام موجب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%100	0	0
بعض الأوقات	0	%100	0
معظم الأوقات	0	0	%100

عندما تكون قيمة جاما صفر فهي تشير إلى عدم وجود تطابق. فإذا كان هناك فقط مجموعة كبيرة من الأزواج المتماثلة مثل وجود مجموعة كبيرة من الأزواج غير المتماثلة، عندئذٍ فإن معرفة الترتيب على طول متغير واحد لا يعطينا أي دليل في كيفية ترتيبه على المتغير الآخر. ويتضح هذا الموقف جلياً في الجدول التالي:

جدول (8-9)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: لا تطابق

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%50	0	%50
بعض الأوقات	0	%100	0
معظم الأوقات	%50	0	%50

تشير الجداول السابقة إلى توضيح النقاط الثلاثة المتطرفة على المقياس المعياري لقياس قوة التطابق بين متغيرين ترتيبيين. كما هو موضح في الشكل التالي:

تطابق تام موجب　　　　لا تطابق　　　　تطابق تام سالب

1+　　　　0　　　　1−

شكل (8-3) مدى قيمة جاما

وبشكل واضح فإن البيانات الواردة في هذا المثال لا تتطابق مع أي واحد من هذه المواقف الثلاثة المتطرفة. إذاً السؤال الذي يطرح الآن هو أي من قواعد التنبؤ ستكون قريبة من النتائج التي تحصلنا عليها فعلياً؟ إنه من الوضوح أن جدول التطابق التام الموجب يمثل واحدة من تلك البيانات الفعلية الأقرب تشابهاً. فقيم جاما التي تحصلنا عليها (0.78) لا نجدها تماماً 1+ ولكنها قريبة منه أكثر منه إلى 0 أو −1.

وتمتاز جاما بشيوعية استخدامها نتيجة لسهولة حسابها نسبياً، إلا أن هذه الميزة قد أبطلت باستخدام برنامج العقل الآلي مثل (SPSS) الذي من خلاله يمكن حساب كل المقاييس بكل سهولة ويسر. إلا أنه قد يساورنا الشك أن عنصراً آخر قد يكون مرتبطاً بشيوعية استخدام جاما ولديه القدرة على توليد قيمة عالية لقوة التطابق مقارنة بمقاييس

تطابق أخرى ترتيبية. وتجدر الإشارة إلى القول بأن جاما تعتريها بعض القيود ينبغي علينا أن نكون على وعي بها:

أولهما: أن مقياس جاما هو فقط مقياس متماثل، وعليه لا يأخذ ميزة المعلومات المقدمة من خلال البيانات التي نعتقد أنها نموذج أكثر ملاءمةً لوصف العلاقة التي تكون معتمدة على طريقة واحدة ويتمثل هذا القيد الرئيسي الآخر عندما يكون التطابق تاماً سوف تولد لنا جاما قيمة +1 أو –1 والعكس ليس دائماً صحيحاً: فقيمة جاما + أو –1 لا تشير دائماً إلى تطابق تام. فقد يكون بإمكاننا أن نولد قيمة لجاما –1 أو +1 لجدول تقاطع حتى عندما يكون هناك بشكل واضح أقل من التطابق التام في البيانات. ويحدث هذا الأمر عندما تكون العلاقة غير مُتسقة. هنا ينبغي علينا اتباع القاعدة، وعليه فإنه قبل استخدام جاما لبيانات جدول ثنائي ينبغي علينا فحصها لتقييم ما إذا كانت العلاقة علاقة مُتسقة. كلا هذين القيدين في حقيقة الأمر ناشئُ من نفس خاصية حساب جاما. وهذا يتضمن عجزاً مرتبطاً بجاما يتضمن الحالات المتعادلة في معادلتهما. وعليه توجد ثلاثة أنماط من الحالات المتعادلة:

1- الحالات المتعادلة على المتغير المستقل (Tx) هذه الأزواج للحالات التي لديها نفس القيمة للمتغير المستقل ولكنها في ذات الوقت لديها قيم مختلفة للمتغير التابع. وهذا في الواقع أن أي حالتين في نفس عمود جدول التقاطع ولكنهما في صفوف مختلفة. في المثال السابق هذه الأزواج من الحالات التي لديها نفس الدخل ولكنها تشاهد كمية مختلفة من الإذاعة المرئية.

2- الحالات المتعادلة على المتغير التابع (Ty) فإن هذه الأزواج للحالات لديها نفس القيمة للمتغير التابع ولكنها في ذات الوقت لديها قيم مختلفة للمتغير المستقل. ومن الناحية العملية، فإن أي حالتين في نفس الصف بجدول التقاطع ولكنها في أعمدة مختلفة. في المثال الذي بين أيدينا هذه الأزواج من الحالات تشاهد نفس كمية الوقت ولكن لديها دخل منخفض.

3- حالات متعادلة على كلا المتغيرين (Tyx) هذه الحالات لديها نفس القيمة في كلا المتغيرين. فهذه أزواج من الحالات سحبت من نفس الخلية في الجدول. ففي المثال

الذي أوردناه فإن هذه الأزواج من الحالات لديها نفس الدخل وتشاهد نفس الكمية من الوقت "الإذاعة المرئية".

وأخيراً توجد مقاييس أخرى متعلقة بنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية التي نسعى إلى تعويض هذا القصور المرتبط بجاما وذلك بإضافة بعض أو كل هذه الحالات المتعادلة في عملية حساباتها [2].

سومرز Somerś d d:

يعتبر مقياس سومرز d مقياساً غير متماثل للتطابق وبذلك يكون هذا المقياس ذا حساسية لأي متغير يوصف بأنه متغير مستقل أو يوصف آخر بأنه متغير تابع. وعليه فإنه من المفيد عندما نشعر بالعلاقة بين متغيرين أن أفضل وصف لهذه العلاقة يتوقف على نموذج اعتماد الاتجاه الواحد. إن المنطق وراء استخدام سومرز d يعتمد في الأساس على فكرة أن الحالتين اللتين تتباينان فيما يتعلق بالمتغير المستقل ولكنهما لا تتباينان فيما يتعلق بالمتغير التابع (هذه الحالات متعادلة على المتغير التابع) تعكس لا تطابق. في المثال الذي تعاملنا معه، فالأزواج المتعادلة على المتغير التابع ولكنها ليست كذلك على المتغير المستقل فإن تلك الأزواج من الحالات تختلف فيما يتعلق بالدخل ولكن مشاهدة التطابق تماماً نفس كمية الإذاعة المرئية. وتحسب قيمة سومرز d كنسبة كل الحالات المتوافقة وغير المتوافقة مضافاً إليها الأزواج المتعادلة على المتغير التابع:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y}$$

ولحساب عدد الحالات المتعادلة فإننا نأخذ كل خلية، بَدْءًا من أعلى اليسار ونضربها في عدد الحالات التي تحتويها في عدد الحالات في الخلايا إلى اليمين منها. انظر جدول (8 – 10) وبتعويض هذه الحسابات في معادلة سومرز d نحصل على القيمة التالية:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y} = \frac{21.075 - 2550}{21.075 + 2550 + 6350}$$

$$= 0.62$$

إن هذه القيمة 0.62 تشير إلى تطابق موجب، متوسط بين هذين المتغيرين فالزيادة في الدخل مرتبطة بالزيادة في مشاهدة الإذاعة المرئية.

جدول (8-10) حساب الحالات المتعادلة على المتغير التابع

75	15	10	
20	70	10	$(75\times15) + (75\times10) = 1875$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) = 150$
10	15	75	
75	15	10	
20	70	10	$(20 \times 70) + (20 \times 10) = 1600$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	
75	15	10	
20	70	10	$(10 \times 15) + (10 \times 75) = 900$
10	15	75	
75	15	10	
20	70	10	$(15 \times 75) = 1125$
10	15	75	

$$T_y = 1875 + 150 + 1600 + 700 + 900 + 1125 = 3506$$

لاحظ أن معادلة سومرز d تقريباً تكون مساوية لمعادلة جاما ما عدا أن المقام في معادلة سومرز d لإعداد المتغير التابع تكون متعادلة. وكنتيجة لذلك عندما يكون لدينا مثل هذه الحالات المتعادلة، فإن قيمة سومرز d ستكون دائماً قيمتها أقل من قيمة جاما.

بمعنى آخر، بتجاهل الحالات المتعادلة فإن جاما يمكن أن يبالغ في قوة التطابق بين المتغيرين في علاقة غير متماثلة، لاسيما عندما تكون هناك مجموعة كبيرة من الحالات المتعادلة وحيث إنّ سومرز d هو مقياس غير متماثل للتطابق فإنه بإمكاننا في الواقع حساب خيارين لهذا المقياس، لأي جدول تقاطع. بمعنى آخر، يمكننا حساب سومرز d من خلال متغير واحد كمتغير مستقل والمتغير الآخر كمتغير تابع، كما أنه بإمكاننا قلب هذين المتغيرين بكيفية أخرى وحساب قيمة سومرز d مرة ثانية. ففي المثال السابق قمنا بحساب سومرز d متعامِلِينَ مع الدخل كمتغير مستقل ومشاهدة الإذاعة المرئية كمتغير تابع، وبما أن النموذج النظري لهذه العلاقة يصور لنا أن السببية تسير في ذلك الاتجاه فقد يكون شخص ما لديه نظرية مختلفة التي تفترض بطريقة ما أن مشاهدة الإذاعة المرئية تحدد مستوى الدخل استناداً إلى ذلك يصبح في مقدوره حساب قيمة سومرز d كخيار آخر مستخدماً الدخل كمتغير تابع متحصلاً على قيمة مختلفة لـِ سومرز d.

كندل - تاو Kendallś tau-b: b:

يعتبر كندل تاو b مقياساً للتطابق غير متماثل لنسبة التخفيض في الخطأ (PRE) لبيانات ترتيبية تم ترتيبها في جدول ثنائي. والميزة الرئيسية المرتبطة بمقياس كندل تاو b تكمن في استخدامه للمعلومات التي توفرها الحالات المتعادلة على المتغير التابع والمتغير المستقل:

ومن خلال المبادئ الرياضية يمكننا ملاحظة أن تاو b تمثل الوسط الهندسي للقيم البديلة لسومرز d. وبالرغم من أن هذا المفهوم قد يكون مشوشاً إلى حد ما نظراً لأن سومرز d من خلال التعريف هو مقياس غير متماثل. ولما كانت قيمة تاو b هي الوسط الهندسي لمقياس سومرز d فإن هذه القيمة ستكون في مكان ما بين القيمتين المتعلقتين بسومرز d التي يمكن حسابها من أي جدول تقاطع أمكن استخدامه. وتجدر الإشارة، إلى أن قيمة تاو b يتراوح مداها بين -1 و +1 حيث الأرقام المتعلقة بالصفوف تكون مساوية لعدد الأعمدة. وعليه وبشكل عام فإن تاو b تستخدم في حالات خاصة.

كندل تاو - C Kendallś tau-c:

يعتبر كنلدتاو C مقياساً متماثلاً لنسبة التخفيض في الخطأ للبيانات يشبه كثيراً مقياس تاو b. ويستخدم تاو C في مواقف تكون فيها رغبة الباحث في قياس التماثل في جدول يحتوي على أعداد غير متساوية من الصفوف والأعمدة التي تحتوي على عدد كبير من الحالات المتعادلة. إن المعادلة الدقيقة لـِ tau-c:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)}$$

حيث إنّ:

K عدد الصفوف أو عدد الأعمدة أيهما أصغر،

N مجموع عدد الحالات.

في المثال السابق لدينا نفس عدد الصفوف مساوياً لعدد الأعمدة (ثلاثة) لذلك فإن K = 3

وأن قيمة tau-c ستكون:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)} = \frac{2(3)(21.075 - 2550)}{300^2(3-1)}$$

$$= 0.62$$

ويتضح مما سبق أن هذه النتيجة مساوية لقيمة سومرز d التي تم حسابها فيما سبق ولذلك فإن هذين البديلين من المقاييس تتفق مع بعضها البعض في النتائج [3].

إجراء مقاييس التطابق باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Analyze ⟸ Descriptive ⟸ Statistics ⟸ Crosstabs

2- انقر فوق Frequency of TV. Watching.

3- انقر فوق ◀ الذي يشير إلى Row(S) في القائمة المحددة للمتغيرات Variables List.

4- انقر على المتغير الذي يشكل الصفوف في الجدول، في هذه الحالة Income Level.

5- انقر على ◀الذي يشير إلى Column(s) في القائمة المحددة للمتغيرات Variable List ستقود هذه العملية إلى لصق Income Level في Column(s) في القائمة المحددة للمتغيرات Income Level.

6- انقر على المربع الذي يكتب عليه Statistics، فيظهر مربع Statistics box: Crosstabs، وهنا تجد مساحة معنونة بـِ: Ordinal التي توفر قائمة بمقاييس التطابق المتوفرة لهذا المستوى من القياس.

7- يختار الباحث معامل Gamma و Somerś dبالنقر على الصندوق التالي لهما أو القريب إليهما. ثم وضع علامة T على صندوق المعامل المطلوبة للإشارة إلى الإحصاءات المختارة.

8- انقر فوق Continue.

9- أنقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

تفسير مقاييس التطابق من مخرجات SPSS:

يمكن للباحث أن يتحصل على نتائج مقاييس التطابق من مخرجات SPSS. وهذه المقاييس هي جزء من أوامر جدول التقاطع Crosstabs Command.

والشكل التالي يبين مخرجات SPSS لإحصاءات جداول التوافق Crosstabs Statistics.

Directional Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Sommerś d Symmetric Frequency of TV. Watching	.618	0.043	14.192	.000
Dependent	.618	0.043	14.192	.000
Income Level dependent	.618	0.043	14.192	.000

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis.

Symmetric Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Gamma	.784	0.043	14.192	.000
N. of Valid Cases	300			

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis

المصدر: George Argyrous, op.cit, P. 184

شكل (8-4): مخرجات SPSS لجداول التقاطع: مخرجات إحصائية

إن حزمة SPSS تنتج لنا القيم المتعلقة بـِ Gamma، وسومرز d: Sommerś d، في مربعين منفصلين يمثل أحدهما مقياساً متماثلاً والآخر مقياساً غير متماثل (يطلق عليه مقياس مباشر Directional Measures في حزمة SPSS. وفي كلتا الحالتين نجد أن القيمتين يولدهما برنامج SPSS، في العمود المعنون "بالقيمة" Value في كل مربع، فهي نفس القيم التي تم حسابها يدوياً. وتعكس هذه القيم تطابقاً متوسطاً، إلى تطابق قوي بين هذين المتغيرين لهذه الحالات المدروسة. وإذا وجد تطابق سالب، فإن الإشارة ناقص (−) سوف تطبع أمام القيمة.

معامل سبيرمان للرتب [4]:

كما نعرف أن ارتباط بيرسون (r) يقيس درجة العلاقة الخطية Linear relationship بين متغيرين عندما تكون البيانات (قيم X و y) تحتوي على درجات عددية مقاسة على المستوى ذي المسافات أو النسبي، وعلى أية حال، فقد تطورت مُعاملات أخرى للعلاقات غير الخطية NonLinear relationship لأنماط أخرى من البيانات، وأن أحد أهم هذه المقاييس قد أطلق عليه ارتباط سبيرمان. ويستخدم ارتباط سبيرمان في موقفين أولهما تستخدم معامل سبيرمان لقياس العلاقة بين X و y، عندما يكون هذان المتغيران مقاسين على مستوى المقياس الترتيبي، وثانيهما تعتبر معامل سبيرمان للرتب بديلاً ذا قيمة لمعامل بيرسون (r) حتى عندما تكون الدرجات الخام الأصلية مقاسة على المستوى ذي المسافات / والنسبي.

إذاً يستخدم معامل سبيرمان في موقفين هامين هما:

1- عندما تكون البيانات الأصلية بيانات ترتيبية أي عندما ترتب قيم X و y. في هذه الحالة، يمكننا ببساطة تطبيق معادلة ارتباط بيرسون (r) لمجموعة الرتب.

2- يستخدم معامل ارتباط سبيرمان عندما يرغب الباحث في مقياس اتساق العلاقة بين X و y. في هذه الحالة، تحول الدرجات الأصلية أولاً إلى رتب؛ وبعد ذلك تستخدم معادلة معامل بيرسون (r) للرتب، باعتبار أن معادلة بيرسون تقيس الدرجة على

النحو الذي تنطبق فيه الرتب على الخط المستقيم a straight Line، كما أن معادلة بيرسون (r) تقيس درجة الاتساق في العلاقة للدرجات الأصلية، وبالمصادفة، عندما يوجد اتساق في علاقة ذات اتجاه واحد بين المتغيرين فإننا نطلق على هذه العلاقة علاقة التماثل Monotonic. عليه، فإن ارتباط سبيرمان يمكن استخدامه لقياس درجة العلاقة التماثلية بين متغيرين. في أي من الحالتين، فإن ارتباط سبيرمان يمكن بيانه بالرمز r_s لنميزه عن ارتباط بيرسون. إن العملية الكاملة لحساب سبيرمان تحتوي على درجات مرتبة. والمثال التالي يوضح هذه العملية:

إن البيانات التالية تبين العلاقة التماثلية التامة بين X و y. عندما تزيد قيمة X. فإن قيمة y بالتالي تتجه نحو النقصان. ولحساب ارتباط سبيرمان، ينبغي على الباحث بادئ ذي بدء ترتيب القيم المتعلقة بـِ X و y، وبعد ذلك يقوم بحساب ارتباط بيرسون لهذه الرتب.

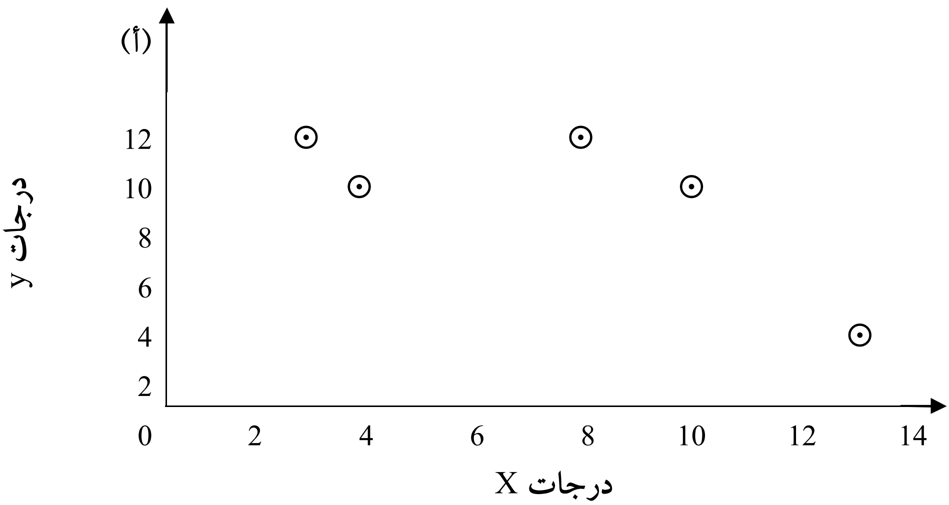
جدول (8-11)

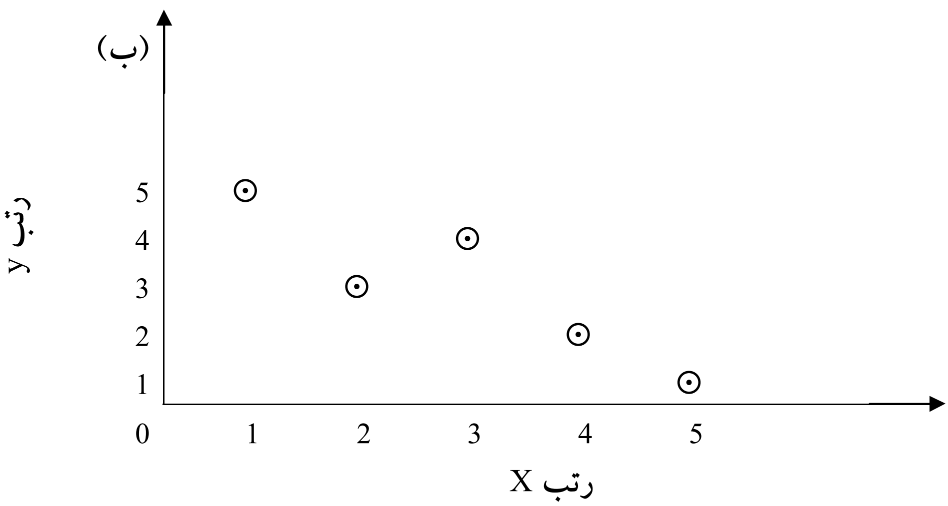
الرتب			البيانات الأصلية	
xy	Y	x	Y	X
5	5	1	12	3
6	3	2	10	4
12	4	3	11	8
8	2	4	9	10
5	1	5	3	13

$\sum Xy = 36$

إن شكل الانتشار للبيانات الأصلية والرتب يوضحها الشكل (8 – 5):

ولحساب الارتباط، فإننا نحتاج إلى حساب SS لِـ X و y و Sp. تجدر الإشارة إلى أن كل القيم يتم حسابها مع الرتب، وليس مع البيانات الأصلية. إن رتب X هي ببساطة الأعداد الصحيحة Integers، 1، 2، 3، 4، 5 ويصل مجموع هذه القيم إلى $\sum X = 15$ و $\sum X^2 = 55$.





شكل رقم (8-5) شكل الانتشار والرتب

وأن SS لرتب X هي،

$$SSx = \sum X^2 - \frac{(\Sigma x)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

لاحظ أن رتب y مطابقة لرتب X. هذا يعني وجود الأعداد الصحيحة، 1، 2، 3، 4، 5 عليه، فإن SSy ستكون مطابقة لـِ SSx:

$$SSy = \sum y^2 - \frac{(\Sigma y)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

ولحساب قيمة SP، فإننا نحتاج إلى: $\sum X$، $\sum y$ و $\sum Xy$ للرتب. إن قيم Xy يحتويها الجدول مع الرتب، كما هو واضح في الجدول (11)، إضافة إلى مجموع كل من X_s و y_s الذي يصل إلى 15. وباستخدامنا لهذه القيم نتـحصل على:

$$Sp = \sum xy - \frac{(\Sigma x)(\Sigma y)}{N}$$

$$= 36 - \frac{(15)(15)}{5}$$

$$= 36 - 45$$

$$= -9$$

وأن الارتباط النهائي لسبيرمان:

$$r_s = \frac{sp}{\sqrt{(ss_x)(ss_y)}}$$

$$= \frac{09}{\sqrt{(10)(10)}}$$

$$= -0.9$$

ويشير ارتباط سبيرمان إلى أن هذه البيانات تبين اتجاه علاقة قوية سالبة (قريبة من العلاقة التامة).

رتب الدرجات المتعادلة Ranking Tied Scores:

عند تحويل الدرجات الأصلية إلى رتب لارتباط سبيرمان، قد يصادف الباحث درجتان أو أكثر متطابقة. وعندما يكون هناك درجتان أو أكثر متطابقة في القيم، فإن الرتب لهذه القيم ينبغي أن تكون متساوية أيضاً. ولإنجاز ذلك ينبغي إتباع الإجراء التالي:

1- تسجل الدرجات في شكل منظم من أصغر الدرجات إلى أكبرها متضمنة القيم المتعادلة في القائمة.

2- تخصيص الترتيب أول، ثانٍ... الخ، لكل وضع في القائمة المنظمة Ordered List.

3- عندما يصادف أن هناك درجتَيْن أو أكثر تكون متعادلة، هنا ينبغي حساب المتوسط لهذه الأوضاع المرتبة. ويخصص قيمة هذا المتوسط كرتبة نهائية لكل درجة.

إن الإجراء المتبع لإيجاد رتب الدرجات المتعادلة يمكن توضيحه من خلال التالي:

الدرجات	الوضع الترتيبي	الرتبة النهائية	
3	1	1.5	المتوسط لـِ 1، 2
3	2	1.5	
5	3	3	
6	4	5	
6	5	5	المتوسط لـِ 4، 5، 6
6	6	5	
12	7	7	

المعادلة الخاصة بارتباط سبيرمان:

بعد أن يقوم الباحث بترتيب القيم المتعلقة بـِ X والقيم المتعلقة بـِ y، فإن عملية حساب SS، و SP عملية ضرورية جداً، أولاً ينبغي على الباحث أن يلاحظ أن رتب X ورتب y هي حقاً مجرد مجموعة من الأعداد الصحيحة: 1، 2، 3، 4،....... n. ولحساب المتوسط لهذه المجموعة من الأعداد الصحيحة يمكننا الحصول على النقطة الوسطى Midpoint لهذه المجموعة من خلال: $M = (n+1) \ /2$. وبشكل مشابه فإنه يمكننا حساب SS لهذه المجموعة من الأرقام الصحيحة من خلال:

$$SS = \frac{n(n^2 - 1)}{12}$$

أيضاً، وبما أن رتب X ورتب y هي متساوية القيم، فإن SSX سوف تكون متطابقة مع Ssy.

ولما كانت حسابات الرتب بالإمكان تبسيطها، ولما كان ارتباط سبيرمان يستخدم بيانات مرتبة، فإن هذه التبسيطات يمكن دمجها في العملية الحسابية النهائية لارتباط سبيرمان. وبدلاً من استخدام معادلة بيرسون بعد ترتيب البيانات، يمكن للباحث أن يضع الرتب مباشرة في المعادلة البسيطة التالية والتي تعطي نفس النتيجة التي تحصلنا عليها باستخدامنا لمعادلة بيرسون:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

حيث إنّ:

D الفرق بين ترتيب X وترتيب y لكل فرد. وينبغي الإشارة هنا إلى أن هذه المعادلة الخاصة يمكن استخدامها فقط بعد أن تحول الدرجات الأصلية إلى رتب. وتستخدم هذه المعادلة فقط عندما لا يكون هناك أي رتب متعادلة. أما إذا كان هناك مجموعة قليلة نسبياً من القيم المتعادلة، يمكن في هذه الحالة استخدام هذه المعادلة، الخاصة، لكن هذه المعادلة ستفقد دقتها كلما زادت الأرقام المتعادلة. ولتوضيح تطبيق هذه المعادلة الخاصة نسوق المثال التالي:

الرتب		الفرق	
X	Y	D	D^2
1	5	4	16
2	3	1	1
3	4	1	1
4	2	−2	4
5	1	−4	16

$$\sum D^2 = 38$$

وباستخدام المعادلة الخاصة لارتباط سبيرمان نتحصل على النتيجة التالية:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(38)}{5(25 - 1)}$$

$$= 1 - \frac{228}{120}$$

$$= 1 - 1.90$$

$$= -0.90$$

لاحظ أن هذه النتيجة هي بالضبط نفس النتيجة التي توصلنا إليها من خلال استخدامنا لمعادلة بيرسون بالرتب.

ولتفسير سبيرمان راهو، باعتباره مؤشر لقوة التطابق بين المتغيرات، حيث يقع مداه من 0 (لا يوجد تطابق)، إلى $\pm$ (تطابق تام). وأن قيمة التطابق التام تعني عدم وجود أي اختلافات في الرتب بين المتغيرين (إذا كانت كل الحالات قد رتبت تماماً بشكل منظم ومتساوٍ على كلا المتغيرين).

أما قيمة العلاقة السالبة ($r_S = -1.0$) فهي تعني وجود اختلاف تام. (إذا كانت الحالة رتبت بشكل عالٍ على واحد من المتغيرين كانت أقل على المتغير الآخر... الخ).

إنّ قيمة سبيرمان راهو في هذا المثال وصلت إلى $r_S = -0.90$ تشير إلى علاقة قوية بين هذين، وأن إشارة السالب تشير إلى أن هذه العلاقة علاقة سالبة. وإذا ما تم تربيع هذه القيمة $r_S = -0.90^2$ تساوي $r_S = 0.81$ ، وعليه فإن أخطاء التنبؤ التي تم توليدها، تم تخفيض نسبة 81 %، أي أنه عندما نتنبأ برتبة موضوع ما على المتغير التابع (y)، فإن رتب الموضوع الآخر X تم أخذها في الاعتبار.

الخلاصة:

حاولنا في هذا الفصل أن نبين العمليات التي يتم فيها حساب مقاييس التطابق لنسبة التخفيض من الخطأ (PRE) عندما يكون كلا المتغيرين قد تم قياسهما على الأقل على المستوى الترتيبي. وتجدر الإشارة في هذا السياق إلى أنه لا توجد قاعدة سهلة لتحديد أفضلية أي من هذه المقاييس على الآخر.

إن جزءاً من هذه المشكلة يتعلق بمفهوم التطابق نفسه وإن حقيقة مفهوم التطابق يتم تعريفه إجرائياً بطرق مختلفة فعلى سبيل المثال أن مقياس جاما وتاو هي مقاييس متماثلة في حين أن مقياس سومرز d هو مقياس غير متماثل ولذلك فإن الخيار بالدرجة الأولى يتوقف على النموذج الموجه للعلاقة التي يعتقد الباحث فيها. ولذا فإنه من الناحية العملية فإن هذه المقاييس عادة ما تشير إلى نفس الاتجاه وتقود الباحث للحصول على نتائج متشابهة.

أسئلة للمراجعة:

1- إذا كان هناك نقصان في قيمة متغير يرتبط بزيادة في قيمة المتغير الآخر، فما هو اتجاه هذه العلاقة؟

2- لماذا لا يحدث التطابق بين متغيرين بين كونهما إما تطابق موجب أو تطابق سالب عندما يكون على الأقل واحد من المتغيرات تم قياسه على المستوى الاسمي؟

3- من الجداول التالية: احسب عدد الأزواج المتوافقة. مفترضاً أن الأعداد على حاشية كل جدول تشير للقيم على مقياس ترتيبي:

أ-

3	2	1	
12	24	60	1
8	14	32	2

ب-

3	2	1	
12	24	60	1
8	14	32	2

ج-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3

د-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3
22	14	9	3	4

4- من الجدول التالي احسب عدد الأزواج غير المتوافقة مفترضاً أن الأعداد في الحاشية تشير إلى قيم مقياس ترتيبي:

أ-

3	2	1	
12	24	60	1
8	14	32	2

ب-

3	2	1	
12	24	60	1
8	14	32	2

ج-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3

5- من البيانات التالية التي تبين العلاقة بين التدخين ومستوى الصحة لدى مجموعة من الناس:

مستوى الصحة العامة	عادة التدخين		
	لا يدخن	**يدخن**	**المجموع**
سيئة	13	34	47
جيدة بعض الشيء	22	19	41
جيدة	35	9	44
جيدة جداً	27	3	30
المجموع	**97**	**65**	**162**

المطلوب:

أ- بالنظر إلى توزيع الصفوف هل بالإمكان استبيان التطابق بين هذين المتغيرين؟ وما هو اتجاه هذا التطابق؟ هل هذا الاتجاه يظهر عند حساب مقياس التطابق؟

ب-احسب قيمة جاما Gamma، وسومرز Somerś d d وبين ما النتيجة التي توصلت إليها حول العلاقة بين التدخين ومستوى الصحة العامة لدى هؤلاء الناس؟

6- من البيانات التالية احسب قيمة جاما وسومرز d. وفسر هذه النتائج التي تتوصل إليها:

غياب الأم من البيت	مستوى الإنجاز لدى الطفل			المجموع
	ضعيف	جيد	ممتاز	
باقية في البيت (لا تعمل)	20	58	22	100
تعمل جزئياً	15	62	23	100
تعمل طوال النهار	12	62	26	100
المجموع	**47**	**182**	**71**	**300**

7- من البيانات التالية استخدم سبيرمان الخاصة لحساب الارتباط. وأن هذه الدرجات D^2 يساوي 0، 4، 0، 1، 1. وأن مجموع $\sum D^2 = 6$.

الدرجة X	الدرجة y	الرتبة X	الرتبة y	D
5	12	2	2	0
7	18	3	5	2
2	9	1	1	0
15	14	5	4	−1
10	13	4	3	−1

8- من البيانات التالية احسب ارتباط بيرسون مبيناً الخطوات المتبعة؟

الأشخاص	x	Y	Yx	
1	0	0	0	SSX=40
2	2	1	2	SSy=54
3	8	10	80	
4	6	9	54	
5	4	6	24	
	20	30	160	

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, PP. 173 - 176.

2- Ibid, PP. 179 - 178.

3- Ibid, PP. 180 - 182.

4- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010, PP. 540 - 547.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

الفصل التاسع

التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار

مقدمة

تناولنا في الفصلين السابقين طرق وصف البيانات عندما يكون أحد المتغيرات على الأقل قد تم قياسه على المستوى الاسمي أو الترتيبي، وقد تم وصف هذه البيانات في شكل جداول تقاطع. ومن خلال جداول التقاطع يتبين لنا التوزيع المشترك لمتغيرين نستطيع من خلالهما ملاحظة ما إذا كان هناك تطابق بين المتغيرين. وعند فحص التوزيعات التكرارية النسبية في الجدول، يقودنا إلى أن نشك في أنَّ هذين المتغيرين يرتبط بعضهما بالبعض الآخر، مما يحدو بنا كخطوة لاحقة بحساب مقاييس التطابق التي تمكننا من الحصول على قيمة رقمية محددة لمثل هذا الشك. وتجدر الإشارة إلى أنه إذا كانت البيانات لهذين المتغيرين تحت الدراسة قد تم جمعها على المستوى ذي المسافات والنسبي، الذي يعني عادة وجود عدد كبير من القيم، الأمر الذي يصبح فيه التعامل مع جداول التوافق غير مناسب لوصف التوزيع. ومن هنا فإن التقنية المرادفة لجداول التوافق للبيانات المناسبة المقاسة على المستوى ذي المسافات / النسبي هي الرسم الانتشاري.

الرسم الانتشاري Scatter Plots:

إنه من الصعوبة بمكان أن نرتب البيانات ذات المسافات المتساوية والنسبية في جداول تقاطع؛ فالبيانات ذات المسافات المتساوية والنسبية عادة لا تقع في عدد صغير من الفئات المنفصلة صغير السن، كبير السن... الخ كالبيانات ذات حجم صغير أو كبير.

إن مثل هذه البيانات بطبيعة الحال يمكن اختزالها إلى قيم قليلة، ولكن هذا الاختزال سيكون على حساب المعلومات. وحيث إن هناك عادةً مجموعة من القيم للمتغيرات التي تم قياسها على المستوى ذي المسافات والنسبي، فإن جدول التوافق سوف يحتوي على عدد كبير من الصفوف والأعمدة، كما توجد عدة قيم في البيانات. وإذا نظرنا إلى التوزيع العمري بالسنوات لسكان بلد ما فإننا نحتاج إلى أكثر من 100 صف من البيانات لكي نأخذها في الاعتبار، فالحقيقة القائلة إن العمر سوف يتوزع على مدى واسع. إن رسم البيانات في شكل انتشاري يمكننا من الحصول على مدى واسع من القيم التي توفر لنا عادة أفضل طريقة لتنظيم مثل هذه البيانات للحصول على انطباع أولي فيما إذا كان هناك وجود لأي علاقة.

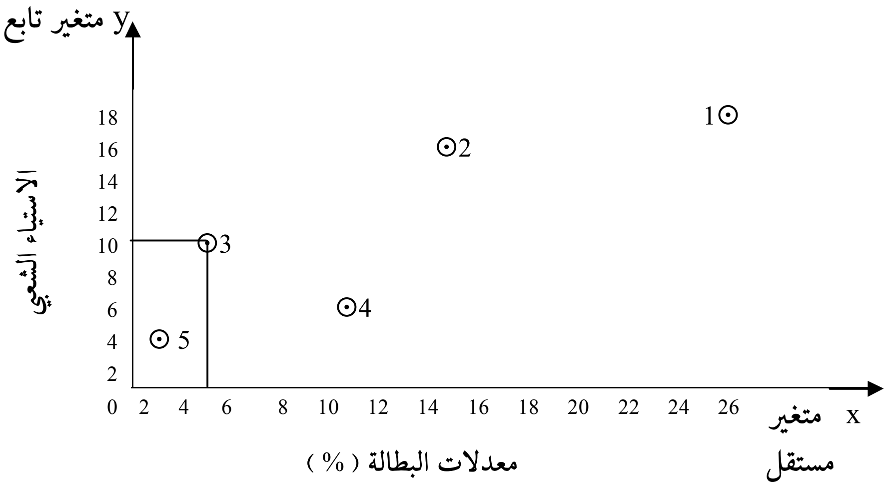
إن رسم الانتشار يبين لنا مجموعة من القيم المتآلفة بحيث إنّ كل حالة تسجل على كلا المتغيرين بشكل تلقائي. إن رسم الانتشار يبين لنا التوزيع المشترك لمتغيرين متصلين ومتناسقين على رسم الانتشار الذي يشير إلى القيم لكل حالة تدون لكل واحد من المتغيرين. فعلى سبيل المثال، قد نرغب في معرفة العلاقة بين معدلات البطالة ومستوى الاستياء الشعبي عن المدن والتي يمكن الحصول عليها من الإحصاءات الرسمية المتعلقة بمعدلات البطالة. فالبطالة لمتغير مستقل X وعدد حالات الاستياء الشعبي لمتغير تابع y لخمس مدن.

جدول (9 ـ 1) معدلات البطالة والاستياء الشعبي في خمس مدن

المدن	معدلات البطالة	معدلات الاستياء
1	25	17
2	13	15
3	5	10
4	10	5
5	2	4

المصدر : George Argyrous,, Statistics for Social and Health Research, With a Guide to Spss, op.cit , p. 233

وبترتيب هذه المعلومات في رسم انتشاري شكل (9 – 1) تجعل من السهولة بمكان قراءتها لتحديد فيما إذا كانت هناك علاقة.



شكل رقم (9 ـ 1) الرسم الانتشاري لبيانات البطالة والاستياء الشعبي

لقد جرت العادة أن يوضع المتغير التابع (y) على المحور الرأسي والمتغير المستقل (X) على المحور الأفقي عند بناء شكل الانتشار. وإذا ما نظرنا إلى أي واحدة من هذه النقاط 1-5 ورسمنا خط مستقيم من الأسفل إلى المحور الأفقي فإنه بإمكاننا أن نجد معدل البطالة في تلك المدينة. وبشكل مشابه برسمنا لخط مستقيم عبر المحور الرأسي يمكننا قراءة عدد حالات الاستياء الشعبي، فالخط المتقاطع لمدينة 3 كما تم توضيحه في هذا الإجراء يشير إلى أن معدلات البطالة لهذه المدينة هو 5 % وأن عدد حالات الاستياء الشعبي هي أيضاً 10 أحداث من الاستياء الشعبي. وبالنظر إلى هذا الشكل فإننا بداهة نرى أن هناك تطابق، لأننا استطعنا أن نتصور ميل الخط ماراً خلال هذه النقاط الخمس. وأن اتجاه التطابق يشار إليه من خلال ما إذا كانت ميول الخط التخيلي أعلى (موجب) أو أسفل (سالب). في هذه الحالة التي أمامنا فإن ميل الخط يكون موجباً مشيراً إلى أن الزيادة في معدلات البطالة متطابقة مع الزيادة في عدد حالات الاستياء الشعبي[1].

خط الانحدار Regression Line:

يعني الانحدار ببساطة مطابقة الخط خلال الرسم الانتشاري للحالات التي تكون مطابقة بشكل جيد للبيانات. وأن أي خط يمكن توضيحه من خلال المعادلة الرياضية التالية المتعلقة بالخط المستقيم. حيث إن: $y = a \pm bX$.

حيث أن: y المتغير التابع.

و X المتغير المستقل.

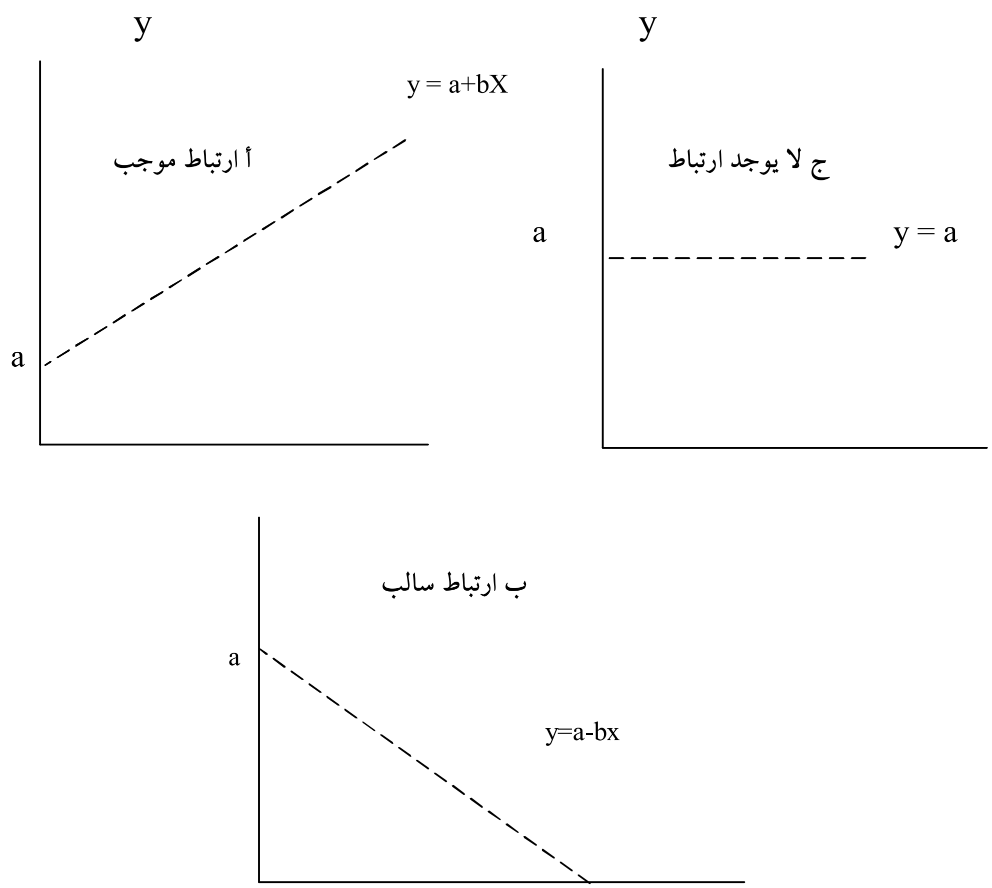
و a قيمة ثابتة وهي المساقة على المحور y من البداية إلى النقطة التي تقطع محور y فهي قيمة y عندما تكون X تساوي 0.

B تشير إلى ميل الخط.

+ تشير إلى التطابق الموجب.

- تشير إلى التطابق السالب.

من خلال هذه المعادلة يمكننا القول بأنَّ الخط المستقيم يمكن تعريفه من خلال عاملين أساسيين أولهما: بداية بنقطة على طول المحور الرأسي، a، وثانيهما هو ميل الخط من نفس هذه النقطة b ± وأن قيمة b هي القيمة التي نرغب فيما إذا كان أي ميل إما أن يكون موجباً أو يكون سالباً، كما تشير لبعض الارتباط بين المتغيرين. ففي الشكل (9 – 2) فإننا نلاحظ ثلاثة خطوط مختلفة تعكس قيمة b في ثلاثة مواقف بديلة لارتباط موجب، وارتباط سالب، ولا يوجد ارتباط.

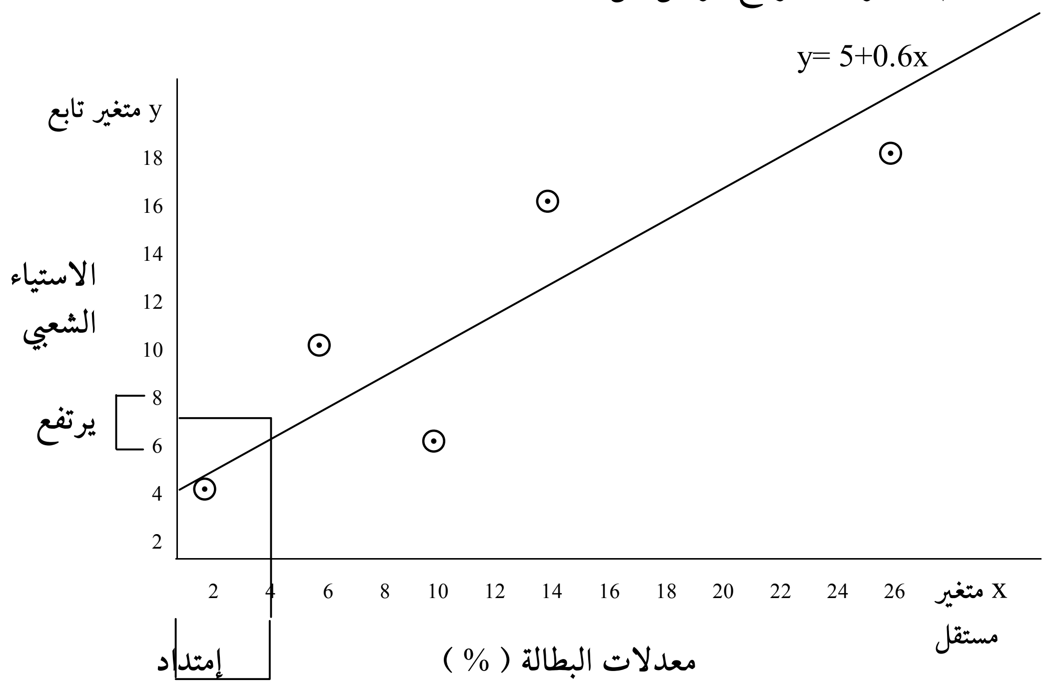


شكل رقم (9 ـ 2): ثلاث خطوات تبين ارتباط موجب أ وارتباط سالب ب ، (ج) لا يوجد ارتباط 0

وبالنظر إلى البيانات المتعلقة بالمدن الخمس يمكننا رسم مجموعة من الخطوط المستقيمة من خلال الشكل الانتشاري، وأن كل واحد من هذه الخطوط سوف يحتوي على معادلته الفريدة والخاصة به. فعلى سبيل المثال، في شكل (9 – 2) قد تم رسم خط يبدو لنا أنه يتوافق مع البيانات بشكل جيد. ويمكن أن نطلق على هذا الخط خط 1 وبدلاً من ذلك يمكننا التعبير عنه بمعادلة رياضية:

$$y = 5 + 0.6X$$

إن السؤال المطروح هو من أين جاءت هذه المعادلة؟



شكل رقم (9-3)

القيمة لـِ: a (5) وهي النقطة على محور y عندما يبدأ الخط. وهي عدد حالات الاستياء الشعبي الذي نتوقع أن نجده في مدينة لديها معدل بطالة صفر (0).

- علامة + تعني أن الخط لديه ميل موجب الذي يشير إلى ارتباط موجب بين هذين المتغيرين.
- أما قيمة 0.6 لِـ b فهي تشير إلى الميل، أو معامل خط الانحدار.

إن معامل خط الانحدار تشير إلى كم من الاستياء الشعبي سيزيد إذا ما زادت معدلات البطالة بـِ 1 % ولما كان الميل لأي خط مستقيم هو الارتفاع على الخط الممتد، فإنه بإمكاننا حساب قيمة (1) لأي ارتفاع في الاستياء الشعبي مثل زيادة 3 بين 5-8. حينئذٍ يمكننا قراءة الزيادة المقابلة في معدلات البطالة، التي تعطينا امتداد خمس (5) وبقسمة الارتفاع على الامتداد (run) فإن الميل سيكون كالتالي:

$$b = \frac{\text{الارتفاع}\,rise}{\text{الامتداد}\,run} = \frac{3}{5}$$

$$= 6.$$

ن الخط الذي أوضحناه الآن يعطينا مدى القيم المتوقعة للاستياء الشعبي اعتماداً على قيمة معدلات البطالة وإن الفرق بين القيمة المتوقعة والقيمة الفعلية للاستياء الشعبي حول معدل بطالة محدد، يطلق عليه المتبقي residual أو مصطلح الخطأ Error Term. إن المتبقي أو مصطلح الخطأ هو الفرق بين القيمة التي نتحصل عليها من خلال المشاهدة للمتغير التابع وبين القيمة المتوقعة للمتغير نفسه بواسطة أو عبر خط الانحدار. لاحظ أن الخط المستقيم لن يمر خلال كل النقاط في الشكل الانتشاري. في حقيقة الأمر أن الخط الجيد من الممكن أنه لا يمس أي من هذه النقاط: عادةً ما تُوجد فجوة بين كل رسم بياني وخط الانحدار باستثناء النقطة التي تقع على الخط ومن هنا تكون لدينا قيمة متبقية. على سبيل المثال، أن الخط الذي لدينا يتكهن بأن المدينة رقم 4 بمعدل بطالة 10 %، وعدد الاستياء الشعبي في هذه المدينة سيكون كالتالي:

$$y = 5 + 0.6X = 5 + 0.6(10)$$

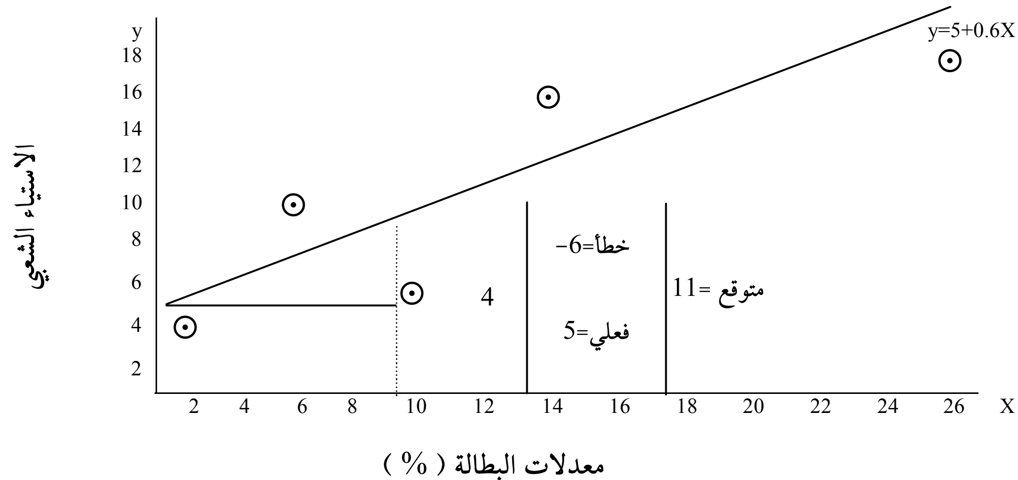
$$= 11$$

وبدلاً من ذلك، يوجد لدينا خمس حالات من الاستياء الشعبي لمدينة 4 بمعدل بطالة يصل إلى 10 %. وبالتالي فإن الخطأ (e) عند هذه النقطة يصل إلى 6-:

$$e = y_{\text{الفعلية}} - y_{\text{المتوقعة}} = 5 - 11$$

$$= -6$$

والشكل التالي يوضح هذه المسألة:



شكل (9-4) الدرجات الفعلية والدرجات المتوقعة

واستناداً على الخط الذي تم رسمه فإننا نلاحظ من خلال العين المجردة أن الخط إلى حد بعيد يطابق هذه البيانات. إن أفضل خط هو ذلك الخط الذي يقلل من البواقي ويجعلها قليلة قدر الإمكان. أي الخط الذي يقلل إلى حد أدنى البواقي.

إن تحليل الانحدار يستخدم الفكرة الرئيسية وهي: أقل مربعات الانحدار Least Square Regression (LSR) لتقليص الفارق بين القيمة المتوقعة لِـ y والقيم الفعلية لِـ y (مربعة) على أن يكون هذا الفرق صغيراً قدر الإمكان (عادة ما نربع البواقي أكثر من أن نقوم بجمعها حيث إن مجموع البواقي لأي خط تمر خلاله النقطة التي يكون فيها المتوسط

لكلا المتغيرين التابع والمستقل مساوياً لصفر. ولتقليل أثر العلامة الموجبة والسالبة فإنه يتم تربيع البواقي بحيث يصبح تعاملنا مع الأعداد الموجبة).

إن أقل مربعات الانحدار هي تلك القاعدة التي تبين لنا رسم الخط خلال الرسم الانتشاري والذي يقلل مجموع البقايا المربعة. إنه بإمكاننا إيجاد أقل مربعات خط الانحدار خلال عملية التجربة والخطأ فإنه بإمكاننا أن نستمر في رسم الخطوط خلال شكل الانتشار مع إيجاد المعادلات الخاصة والبواقي حتى نصل في نهاية الأمر إلى الخط الذي يقلل من هذه البواقي. ولحسن الحظ يوجد لدينا بديل عندما نستخدم القاعدتين الاثنتين حتى نستطيع اشتقاق أقل مربعات خط الانحدار بشكل مباشر. إن خط المربعات الصغرى للانحدار ينبغي أن تمر خلال النقطة التي تتساوي فيها المتوسطات للمتغير التابع والمتغير المستقل ($\bar{x}$ و $\bar{y}$). والمتوسط لعدد حالات البطالة $\bar{x}$ تكون:

$$\bar{x} = \frac{\sum x}{N} = \frac{55}{5}$$

$$= 11$$

وأن متوسط معدل الاستياء الشعبي $\bar{y}$

$$\bar{y} = \frac{\sum y}{N} = \frac{55}{5}$$

$$= 10.2$$

وعليه فإن أقل مربعات الانحدار خط الانحدار سوف تمر خلال النقاط المتناسقة (11 و 10.2) ويمكننا تعريف ميل خط الانحدار b من خلال المعادلة التالية:

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

وحيث إنَّ هذه المعادلة تبين الفكرة الأساسية التي يحتاجها الخط للتقليل من فروق التربيع بين القيم الفعلية والقيم المتوقعة، من هنا فإنه من السهولة بمكان حساب قيمة b وفقاً للمعادلة التالية:

$$b = \frac{N\,\Sigma(xy) - (\Sigma x)(\Sigma y)}{N\,\Sigma x^2 - (\Sigma x)^2}$$

جدول (9-2) حساب ميل خط الانحدار لأقل المربعات

المدن	معدلات البطالة X	حالات الاستياء y	X^2	y^2	Xy
1	25	17	625	289	425
2	13	15	69	225	195
3	5	10	25	100	50
4	10	5	100	25	50
5	2	4	4	16	8
	$\sum X55$	$\sum y = 51$	$\sum x^2 = 923$	$\sum y^2 = 655$	$\sum Xy = 728$

$$b = \frac{N\sum(xy) - \sum(x)\sum(y)}{N\sum x^2 - (\sum x)^2} = \frac{5(728)-(55)(51)}{5(923)(55)^2}$$

$$= \frac{3640-2805}{4615-3025}$$

$$= +\,0.53$$

إن قيمة b يطلق عليها معامل الانحدار، وهي قيمة ذات أهمية لأنها تحدد مقدار أي ارتباط بين متغيرين. فمعامل الانحدار يشير إلى كم من الوحدات في المتغير التابع ستتغير إذا ما تغيرت أي وحدة في المتغير المستقل.

وتجدر الإشارة هنا إلى أنه بعدما حددنا خط الانحدار خلال نقطة محددة (متوسط X و y) وكذلك من خلال إعطاء الاسم الأخير من خلال حساب ميل الخط خلال هذه النقطة، فإنه باستطاعتنا إعطاؤها وصفاً كاملاً من خلال اشتقاق القيمة المتعلقة بـِ a. ومن هنا يمكننا استخدام المعادلة التالية التي تستخدم كلا المَعْلَمَيْن المتعلقين بخط الانحدار الذي

تم بيانه (تلك التي تمر خلال المتوسط المتعلق بـِ X والمتوسط المتعلق بـِ y وبالتالي يكون لها ميل مساوياً لـِ b:

$$a = \bar{y} - b\bar{x}$$

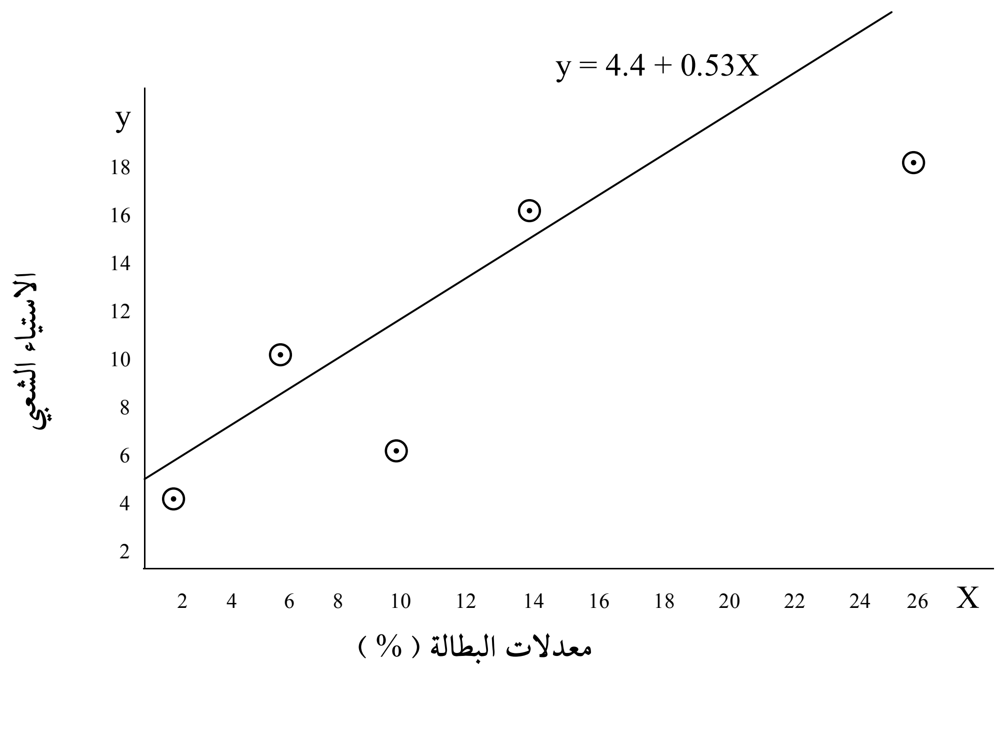
وعليه فقيمة a ستكون:

$$a = \bar{y} - b\bar{x} = 10.2 - 0.53(11)$$

$$= 4.4$$

وعليه يمكننا تعريف أفضل خط مقابل لمثل هذه الحالات من خلال المعادلة التالية:

$$y = 4.4 + 0.53X$$



شكل (9ـ5) خط الانحدار لأقل المربعات (OLS)

السؤال الذي يمكن طرحه هنا هو: ماذا يخبرنا خط الانحدار لأقل المربعات حول العلاقة بين معدلات البطالة وحالات الاستياء الشعبي لمثل هذه المجموعة من الحالات؟ للإجابة على هذا السؤال يمكننا القول:

- توجد علاقة موجبة بين المتغيرين: فالزيادة أو النقصان في معدل البطالة يرتبط مع الزيادة أو النقصان في عدد حالات الاستياء الشعبي.
- يمكننا قياس العلاقة الموجبة: فالزيادة في معدل البطالة لـِ: 1% يرتبط بالزيادة بـِ 0.53 لحالات الاستياء الشعبي.

إنه بإمكاننا الآن استخدام هذه المعادلة لغرض التنبؤ بعدد حالات الاستياء الشعبي في مدينة ما، فمن المرجح أن يكون لديها معدل محدد من البطالة، على سبيل المثال إذا ما علمنا أن مدينة أخرى لديها معدل بطالة يصل إلى 18 % فإن أفضل تكهن يمكن الوصول إليه هو أنها تمر بتجربة 13.9 حالة استياء شعبي [2].

معامل ارتباط بيرسون (r) :

قد تبين لنا من خلال المثال السابق أن قيمة b هي مؤشر ما إذا كانت هناك علاقة بين متغيرين تم قياسهما على مستوى المقياس ذي المسافات والنسبي، وكذلك تبين اتجاه مثل هذا الارتباط. ولكن السؤال هل تشير أيضاً هذه القيمة إلى قوة العلاقة؟ هل قيمة b التي تساوي 0.53 تشير إلى تطابق، قوي أو متوسط أو ضعيف؟ إنَّ هذه القيمة لسوء الحظ لا تشير إلى ذلك.

إن القضية الأساسية هي أن وحدات القياس تتباين من موقف إلى موقف آخر. فعلى سبيل المثال، إذا استخدمنا التناسب بدلاً من نقاط النسب لقياس معدلات البطالة، فعوضاً عن استخدام الحسابات 22، 20، 15، 10، 9 يمكننا استخدام 0.22، 0.20، 0.15، 0.1، 0.09 فإن القيمة المتوقعة لـِ: b ستكون 53 بدلاً من 0.53.

إن العلاقة الفعلية التي نود معرفتها لم تتغير، وإنما تغيرت فقط وحدات القياس. بمعنى آخر، إن قيمة b قد تأثرت ليس فقط بقوة الارتباط، وإنما أيضاً بوحدات القياس. وعليه لا توجد طريقة لمعرفة ما إذا كانت أي قيمة محددة لـِ: b تشير إلى ارتباط ضعيف أو متوسط أو قوي. وللتغلب على هذه المشكلة يمكننا تحويل قيمة b إلى مقياس معياري للارتباط يطلق عليه معامل ارتباط بيرسون (r). إن معامل بيرسون (r) يتراوح مداه دائماً بين −1 و +1 بغض النظر عن الوحدات الفعلية التي تم قياس المتغيرات عليها. والمعادلة المستخدمة لارتباط بيرسون (r) هي:

المعادلة (1)

$$r = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{[(x-\overline{x})^2][(y-\overline{y})^2]}}$$

أو

المعادلة (2)

$$r = \frac{N\,\Sigma(xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N\,\Sigma x^2 - (\Sigma x)^2][N\,\Sigma y^2 - (\Sigma y)^2]}}$$

وإذا نظرنا إلى المثال الذي تم التعامل مع مكوناته من خلال حساب قيمة b جدول (2) فإذا ما استبدلنا الإحصاءات من هذا الجدول وتعويضها بالمعادلة التالية سنتحصل على:

$$r = \frac{N\,\Sigma(yx) - (\Sigma x)(\Sigma y)}{\sqrt{[N\,\Sigma x^2 - (\Sigma x)^2][N\,\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{5(728-55(51)}{\sqrt{[5(923)-(55)^2][5(655)-(51)^2]}}$$

$$= \frac{3640-2805}{[4615-3025][3275-2601]}$$

$$= 0.81$$

إما إذا استخدمنا المعادلة الأولى لإيجاد ارتباط بيرسون (r) فإننا نتحصل على نفس القيمة كما هو موضح في التالي:

1	2	3	4	5	6	7
X المدن	y البطالة	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
25	17	+14	+6.8	95.2	196	46.24
13	15	+2	+4.8	9.6	4	23.04
5	10	-6	-0.2	1.2	36	0.04
10	5	-1	-5.2	5.2	1	27.04
2	4	-9	-6.2	55.8	81	38.44
$\Sigma x = 55$	$\Sigma y = 51$	$\sum(x-\bar{x}) = 00$	$\sum(y-\bar{y}) = 00$	167	$\sum(x-\bar{x})^2 = 318$	$\sum(y-\bar{y})^2 = 138.8$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[(x-\bar{x})^2][(y-\bar{y})^2]}}$$

$$= \frac{167}{\frac{\sqrt{(318)(134.8)}}{\sqrt{42866.4}}}$$

$$= \frac{167}{207.42}$$

$$= 0.81$$

إن قيمة r تبين لنا فقط قوة التطابق بل أيضاً اتجاه هذا التطابق. إن قيمة 0.81 تشير إلى الارتباط بين هذين المتغيرين لمثل هذه المجموعة من الحالات فهي قيمة موجبة وقوية.

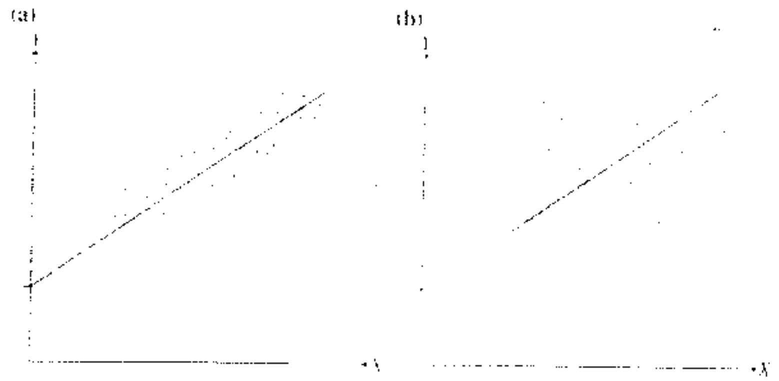
تفسير التباين: معامل التحديد أو التقدير:

قد بينا في هذا الفصل في حينه استخدام خط الانحدار للتكهن بعدد حالات الاستياء الشعبي في مدينة ما من خلال معطى معين لمعدل البطالة. إلا أننا قد رأينا أن هناك دائماً

"هامش للخطأ" في هذا التنبؤ، اعتماداً على مدى قرب النقاط وتجمعها حول خط الانحدار. من هنا يمكننا استخدام خط الانحدار لنبين أن زيادة معينة في X ستقود إلى زيادة كبيرة في y، ولكن إذا كان هناك خطأ كبير بين خط الانحدار والنقاط الفعلية للبيانات فالاحتمال القوي أن التنبؤات سوف تكون خاطئة بقدر كبير أكثر إذا قورنت بالموقف عندما تكون الدرجات متجمعة بشكل واضح حول خط الانحدار.

وكما يظهر في الشكل التالي أنه بالرغم من أن نفس خط الانحدار يتطابق بشكل واضح لكلتا المجموعتين من النقاط فإن ثقتنا ستكون كبيرة في قدرتنا للتنبؤ بقيمة (a) أكثر من قدرتنا على التنبؤ بقيمة (b) ويرجع السبب في ذلك إلى أن خط الانحدار في (a) قادر على تفسير قدر كبير في نسبة التباين في (y) أكثر منه في خط الانحدار (b).

وعليه فإننا بحاجة إلى بعض القياس لكي يبين لنا كم من التباين في المتغير التابع يتم تفسيره بواسطة خط الانحدار.



شكل (9-6)

إنه من حسن الحظ أنه يمكننا أن نقوم بذلك ببساطة من خلال تربيع قيمة (r) للحصول على قيمة معامل التحديد r^2، فالتباين المفسر من خلال خط الانحدار ذا صلة بالتباين المفسر في الحالة التي يوجد بها تطابق.

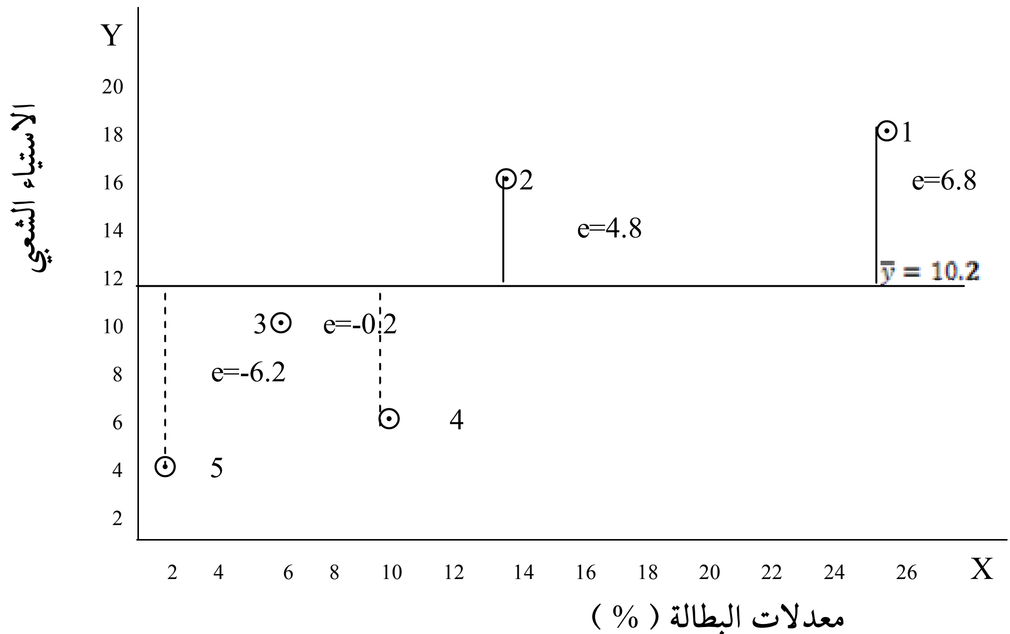
$$r^2 = (0.8!)^2$$
$$= 0.65$$

ويمكننا تفسير معامل التحديد كمقياس غير متماثل للتقليل من نسبة الخطأ في مقياس التطابق أي أن معامل التحديد تشبه إلى حد كبير مقاييس نسبة التقليل من الخطأ Proportional reduction in Error (PRE) التي تم التعرض لها في موضعه في الفصلين السابقين وبالتالي فإن معامل التحديد لديه منطق مشابه لمنطق لامبيدا، إلا أن تطبيقات معامل التحديد تتعلق بالمتغيرات المقاسة على المستويين ذي المسافات والنسبي.

ومن هنا يمكننا القول بأن معامل التحديد يُستخدم لتحديد الدلالة على مقدار مؤاءمة خط أقل التربيعات مع بيانات العينة الذي يسعى الباحث لاختبارها. وتُعرف معامل التحديد بأنه "النسبة بين مجموع مربعات الانحراف الراجح إلى الانحدار إلى مجموع مربعات الانحراف الكلية"[3]، ومن هنا نجد أن معامل التحديد أو التقدير يبين لنا نسبة الاختلاف الكلية في المتغير التابع (y) والتي تم تفسيرها من انحدار y على X.

وتجدر الإشارة إلى أنه أيضاً باستطاعتنا القيام بالتنبؤ حول القيم المتوقعة للمتغير التابع بدون أية معلومات حول المتغير المستقل (X). كما أننا إذاً يمكننا القيام بعملية التنبؤ عندما تكون لدينا معرفة حول المتغير المستقل (X)، ومقارنة معادلات الخطأ فعلى سبيل المثال، إذا أردنا أن نخمن في عدد حالات الاستياء الشعبي في كل مدينة وأن كل ما نعرفه أن متوسط عدد حالات الاستياء الشعبي لكل المدن الخمس هو 10.2، فإن أفضل توقع يمكن الوصول إليه هو أن عدد حالات الاستياء الشعبي لكل مدينة هو 10.2 بغض النظر عن المعدل الفعلي للبطالة [4].

نقوم برسم خط أفقي مستقيم لهذه القيمة كخط انحدار خلال شكل الانتشار. انظر الشكل التالي:



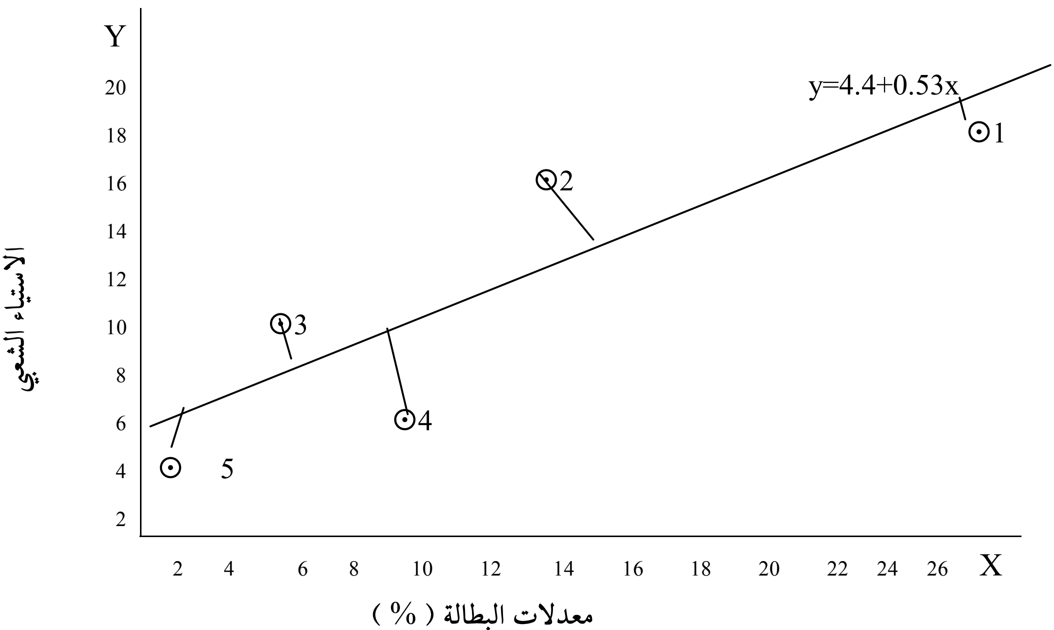
شكل (9-7) خط الانحدار بدون معلومات حول المتغير المستقل

إن هذا الخط الأفقي هو الخط الذي تم رسمه عندما لم يكن هناك ارتباط بين هذين المتغيرين (معدل البطالة وعدد حالات الاستياء الشعبي). إن معرفة ما إذا كان معدل البطالة عالياً أو منخفضاً لن يقودنا إلى تغيير توقعاتنا حول حالات الاستياء الشعبي عن المتوسط.

ففي بعض الأحيان قد يأتي هذا الخط بشكل قريب من العلاقة. ففي مدينة 3 يمكننا أن نرى أن هذا الخط يتنبأ بمعدل بطالة يصل إلى 5 %، والتي ستكون 10.2 من حالات الاستياء الشعبي.

فهنا نجد أنه في الحقيقة أن 10 حالات من الاستياء الشعبي. تولد لنا خطأ (e) في هذه المدينة يصل إلى 0.2 فقط. وعلى أية حال، فإنه في شواهد أخرى يمكننا أن نسبب خطأ أكبر باستخدام هذا الخط. ففي مدينة 1 على سبيل المثال، يصل معدل البطالة فيها

إلى 25 % مرة ثانية فإننا نتنبأ بأنه في هذه المدينة تصل حالة الاستياء الشعبي فيها إلى 10.2 لكنه في الواقع نجد أن حالات الاستياء الشعبي في هذه المدينة يصل إلى 17 حالة استياء شعبي. ومن هنا نجد أننا قد ارتكبنا خطأ في التكهن بمعدلات الاستياء الشعبي يصل إلى 6.8. والآن إذا قارنا هذه الأخطاء مع تلك الأخطاء التي فعلناها عندما قمنا بعملية التنبؤ وفقاً لخط انحدار أقل المربعات LSR. فهل هذا الخطأ فعلياً يحسن من توقعاتنا؟ انظر شكل (8).



شكل (9ـ8) خط الانحدار عندما يكون لدينا معرفة بالمتغير المستقل (X)

من هذا الشكل يمكننا أن نلاحظ أنه إذا كان هناك ارتباط وثيق بين هذين المتغيرين، وأن خط انحدار أقل المربعات سيقلل من معدل الخطأ. وبالتالي فإن الهوة بين نقاط هذه البيانات والخط ستكون قليلة جداً باستخدامنا لخط انحدار أقل المربعات أكثر منه باستخدام الخط الأفقي استناداً على فرضية لا يوجد ارتباط.

إن استخدام معامل التحديد كأحد مظاهر خط الانحدار هو الأساس في توضيح ذلك. فقيمة (r^2) تساوي 0.65 تشير إلى أن خط الانحدار لأقل المربعات Least Squares regressim Line تفسر 65% من التباين في المتغير التابع نسبة إلى التباين الذي تم تفسيره من خلال الخط الأفقي وهذه العملية تبين لنا التقليل الجوهري في معدل الخطأ(e).

يتعين علينا هنا أن نميز بين قيمة r و r^2 باعتبارهما قيمتان مرتبطتان بعضهما ببعض. ويعتبر معامل الارتباط مقياس معياري للعلاقة بين متغيرين؛ فمعامل الارتباط يُشير إلى المدى الذي يحدث فيه التغير في أحد المتغيرات يرتبط بالتغير في متغير آخر. عليه فإن قيمة r (تشبه b) التي هي في الأساس أداة للتنبؤ. أما على الجانب الآخر، فإن معامل التحديد يُعتبر مقياساً لتقليل النسبة في الخطأ (PRE)، في كمية التباين المفسر من خلال خط الانحدار، عليه فإن معامل التحديد يمد الباحث بإحساس كم لديه من الثقة يمكن الاعتماد عليها في صحة التنبؤ الذي يقوم به [5].

التباين الكلي: المفسر وغير المفسر:

يمكننا بشكل محدد تعريف التقليل في الخطأ الذي ينتج عندما تأخذ المعلومات حول المتغير المستقل في الاعتبار. هناك مجموعان مختلفان يمكن إيجادهما وبعد ذلك مقارنتهما بالتباين الكلي لِـ y لبناء إحصاء سيشير إلى التحسن في التنبؤ.

1- مجموع التباين المفسر Explained Variance ويوضح لنا مجموع التباين المفسر، يحسن قدرتنا على التنبؤ بـِ y عندما تؤخذ المعلومات المتعلقة بالمتغير المستقل (X) في الاعتبار. ويمكننا الحصول على هذا المجموع من خلال طرح $\bar{y}$ (تكهن درجة y بدون معلومات حول X) من الدرجات المتنبأ بها بواسطة معادلة الانحدار (درجة $\hat{y}$ أو y تم التنبؤ بها عندما كانت المعلومات متوفرة حول X). لكل حالة تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية كالتالي: $\sum(\hat{y}-\bar{y})$. وأن النتيجة المتحصل عليها يمكن بعد ذلك مقارنتها بالتباين الكلي في y لنتحقق إلى أي مدى تكون معرفتنا بـِ X قد حسنت من قدرتنا على التنبؤ بـِ y.

والمعادلة المتعلقة بذلك يمكن بيانها رياضياً كالتالي:

$$r^2 = \frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2}$$ = التباين المفسر / التباين الكلي

إن معامل التحديد أو التقدير أو r^2 هي نسبة التباين الكلي في y الذي تم تفسيره من خلال X. وتشير r^2 بالتحديد إلى المدى الذي يساعد متغير X على التنبؤ لفهم أو تفسير متغير y.

2- مجموع التباين غير المفسر: لمعرفة التباين غير المفسر فإننا نقوم بطرح درجات $\hat{y}$ المتنبأ بها من درجات y الفعلية لكل حالة من الحالات؛ وبعد ذلك تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية في الآتي:

$$\sum(y - \hat{y})^2$$

إن المجموع المتحصل عليه سيقيس كمية الخطأ في التنبؤ في y التي تبقى حتى بعد أن أخذنا متغير X في الاعتبار. إن نسبة التباين الكلي في y غير المفسرة لـِ X يمكن أن نتحصل عليها من خلال طرح قيمة r^2 من 1.00.

وتجدر الملاحظة هنا أنه كلما كانت العلاقة قوية بين X و y كانت قيمة التباين المفسر أكبر والعكس بالعكس. أنه في حالة العلاقة التامة $r = \pm 1.00$ فإن التباين غير المفسر سيكون صفر وأن r^2 ستكون 1.00 وهذا يشير إلى أن X تفسر كل التباين في y، وأننا بهذا بإمكاننا التنبؤ بقيمة y من X بدون خطأ (e). وعلى الجانب الآخر، فإذا كانت X و y لا يرتبطان بعضهما ببعض ($r = 00.0$)، فإن التباين المفسر سيكون (0) وأن r^2 ستكون كذلك 0.00 في مثل هذه الحالة، فإننا سنتوصل إلى نتيجة مفادها أن X لا يفسر أي شيء من التباين في y وأننا سنكون غير قادرين على التنبؤ بـِ y. ففي المثال الذي تعاملنا معه في هذا الفصل تحصلنا على $r = 81$ وبعد تربيع هذه الفئة تحصلنا على معامل تحديد تصل إلى 0.65. ($r^2 = 0.65$) والتي تشير إلى أن 65 % من التباين الكلي في حالات الاستياء الشعبي في هذه المدن تم تفسيره. ومن هنا نجد أن 35 % من التباين

الكلي في y لم يتم تفسيره من قِبل X، وقد يرجع السبب في ذلك إلى تأثير بعض المتغيرات المؤتلفة، أو إلى خطأ القياس، أو الصدفة العشوائية (6).

القياس والمتغيرات الديميوية Dummy Variobles:

يُعتبر معامل الارتباط والانحدار من أقوى التقنيات الإحصائية، لذلك فإنه في أغلب الأحيان تستخدم هذه التقنيات لتحليل العلاقة بين المتغيرات التي لم يتم قياسها على المستوى ذي المسافات / النسبي. إن هذه الممارسة بشكل عام، لا تمثل أية مشكلة إذا تعلق الأمر بالمتغيرات الترتيبية المتصلة التي لديها مدى واسع من الدرجات المحتملة، بالرغم من أن مثل هذه المتغيرات لا تمتلك نقاط الصفر الحقيقية، ومسافات متساوية من درجة إلى درجة. وقد بينا ذلك عند دراسة ارتباط سبيرمان "راهو" Spearmans` rho. كذلك يستخدم الباحثون في العلم الاجتماعي الارتباط والانحدار عند التعامل مع البيانات الترتيبية المضغوطة Collapsed. إن هذه المتغيرات تمتلك عدداً من الدرجات المحدودة (عادة بين 2 إلى 5)، مثل المسح الذي يبحث في الاتجاه نحو عقوبة الإعدام. إن مثل هذا الانتهاك لمستوى القياس لا يشكل بوضوح مشكلة طالما أن النتائج تعالج بدرجة مناسبة من الحذر.

إن هذه المرونة لدى الباحثين في العلم الاجتماعي ينبغي ألا تتسع لتشمل المتغيرات المقاسة على المستوى الاسمي. إن حساب معامل الارتباط أو الانحدار لمتغيرات مثل الحالة الاجتماعية، أو الطائفة الدينية تعتبر غير ذات جدوى؛ وذلك لأن درجات المتغيرات الاسمية ليست أرقام ولا تمتلك أي خاصية رياضية. إنه بإمكاننا أن نخصص درجة للديانة، الإسلام بدرجة (2) وغير الإسلام بدرجة (1). ولكن الدرجة المتحصلة للديانة الإسلام ليس ضعف الدرجة التي تم تحصيلها للديانة غير الإسلام. وبالتالي فإن الدرجات المتعلقة بالمتغيرات المقاسة على المستوى الاسمي هي مجرد عبارات وليست أرقام.

ولما كانت هذه المتغيرات متغيرات غير رياضية، فإنه من غير المنطق أن نقوم بحساب ميل الخط (b) أو نناقش ما إذا كانت هناك علاقات موجبة أو سالبة.

ومن هنا نجد أن الكثير من المتغيرات المهمة في حياتنا اليومية مثل: النوع، الحالة الاجتماعية، العرق، هي متغيرات اسمية لا يمكن تضمينها في معادلة الانحدار أو التحليل العلائقي، تلك التقنيات الأرقى والأقوى المؤثرة في مجال بحوث العلم الاجتماعي.

تجدر الإشارة إلى أنه من حسن الطالع أن الباحثين قد طوّروا طريقةً لمعالجة هذه المعضلة لكي نضيف المتغيرات المقاسة على المستوى الاسمي وذلك من خلال خلق ما يسمى بالمتغيرات الديميوية Dummy Variables. إن المتغيرات الديميوية يمكن أن تكون على أي مستوى من مستويات القياس بما فيها المتغيرات الاسمية، وإن لديها بالضبط فئتين، الأولى ترمز كـ: (0) والأخرى كـ: (1) وبالتالي ينبغي التعامل بهذه الكيفية مع المتغير المقاس على المستوى الاسمي لكي يكون ذا معني عندما يقدم كمتغير ثنائي يمكن تضمينه لمعادلة الانحدار. على سبيل المثال، يمكننا عند التعامل مع متغير النوع أن يخصص الرمز (0) للذكور، ويخصص الرمز (1) للإناث، العرق: أوروبي (0)، أسيوي (1) (ربما يحتاج الباحث إلى متغيرات ديميوية لإضافة أعراق أخرى أو انتماءات أخرى)؛ الانتماء الديني الإسلام قد يعطى له رمز (2) والمسيحية رمز (0) (مرة ثانية قد يحتاج الباحث إلى متغيرات ديميوية لأخذها في الحسبان لدى ديانات أخرى).

ولتوضيح ذلك، نفترض أننا نقوم بدراسة حول العلاقة بين العرق والتعليم (يقاس التعليم بعدد السنوات التي تم إنجازها). فإذا تم تخصيص رمز (0) للشخص ذو العرق الأبيض، والرمز (1) للشخص ذو العرق الأسود، فإنه بإمكاننا حساب معادلة ميل الخط y . وتكتب هذه المعادلة مستخدمين العرق كمتغير مستقل وحينها نبحث في العلاقة بين هذين المتغيرين. مرة ثانية، نفترض أننا قمنا بقياس عينة مستوى التعليم لدى العرق الأسود وخصصنا رمز (1). للعرق الأسود، والرمز (0) للعرق الأبيض، فإننا سنتحصل على معادلة الانحدار التالية:

$$y = a + bX$$

$$y = (12.0) + (-0.5)X$$

إذا كان التعليم متغير تابع y، والعرق متغير مستقل (X)، فإن خط الانحدار يعبر

العمودي Vertical لشكل الانتشار للنقطة، حيث y=12.0 والقيمة $b= -0.5$ للميل وهي تشير إلى علاقة سالبة بمعنى أن المبحوث ذو الخلفية السوداء في هذه العينة يسجل متوسط أقل من السنوات الدراسية مقارنة بنظيره الأبيض. وأن النتيجة ستكون واحدة إذا ما تغيرت طريقة الترميز.

فإذا كان الرمز (1) قد أعطي للرجل ذو الأصول العرقية البيضاء، وصفر (0) للرجل ذو الأصول السوداء، فإن قيمة (b) ستبقى بالضبط كما هي، وذلك لأن عملية الترميز المتعلقة بالمتغيرات الديمومية هي عملية اعتباطية، كما هو الحال في المتغيرات الترتيبية. فالباحث يتطلب منه أن يكون واضحاً حول ما تشير إليه القيم المتعلقة بالمتغيرات الديمومية.

كذلك يمكننا أيضاً استخدام معامل بيرسون (r) لتقييم قوة واتجاه العلاقة بالمتغيرات الديمومية. فإذا تحصلنا على سبيل المثال، على قيمة $r = -0.23$ بين العرق والتعليم، فإننا بالتالي نستطيع القول بأنه توجد علاقة سالبة ضعيفة إلى متوسطة بين هذين المتغيرين لهذه العينة. وباتساق الإشارة المتعلقة بميل الخط (b)، يمكننا أيضاً القول بأن التعليم يقل كلما زادت العرقية أي التحرك من العرق الأبيض إلى العرق السود. كذلك في هذا المقام، يمكننا استخدام معامل التحديد، وذلك بالقول، بأن العرق يفسر 5 % ($r^2 = 0.23^2 = 0.05$) من التباين في التعليم [7].

إجراءات الرسم البياني، والارتباط والانحدار باستخدام SPSS:

أولاً: توليد الرسم الانتشاري Scatter Plot مع خط الانحدار regression Line:

1- من القائمة الموجودة في الجزء العلوي من الشاشة أختر:

Graphs ⇦ Scatter (يعطيك صندوق الحوار Scatter plot).

2- انقر على Define (يعطيك مربع Simple scatter plot).

3- انقر على Number of civil disturbances.

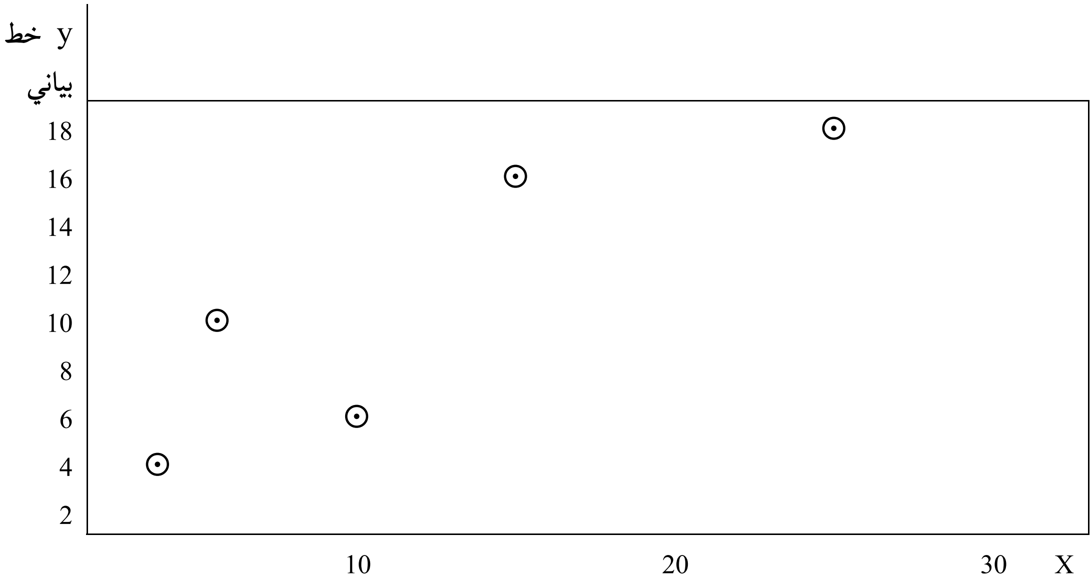
4- انقر على ◄ الذي يشير إلى y Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق Number of civil disturbances كمتغير تابع يمكن وضعه على y Axis).

5- انقر على unemployment rate (معدل البطالة).

6- انقر على ◄ الذي يشير إلى X Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق un employment rate كمتغير مستقل يمكن وضعه على X Axis في القائمة المحددة للمتغير المستقل).

7- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:



شكل (9 ـ 9) مخرجات SPSS للرسم الانتشاري

إجراءات إضافة خط الانحدار للرسم البياني باستخدام SPSS:

1- انقر نقرتان على الرسم البياني Scatter plot في نافذة Viewer. هذه العملية تقود إلى الرسم البياني في نافذة Chart Editor Window.

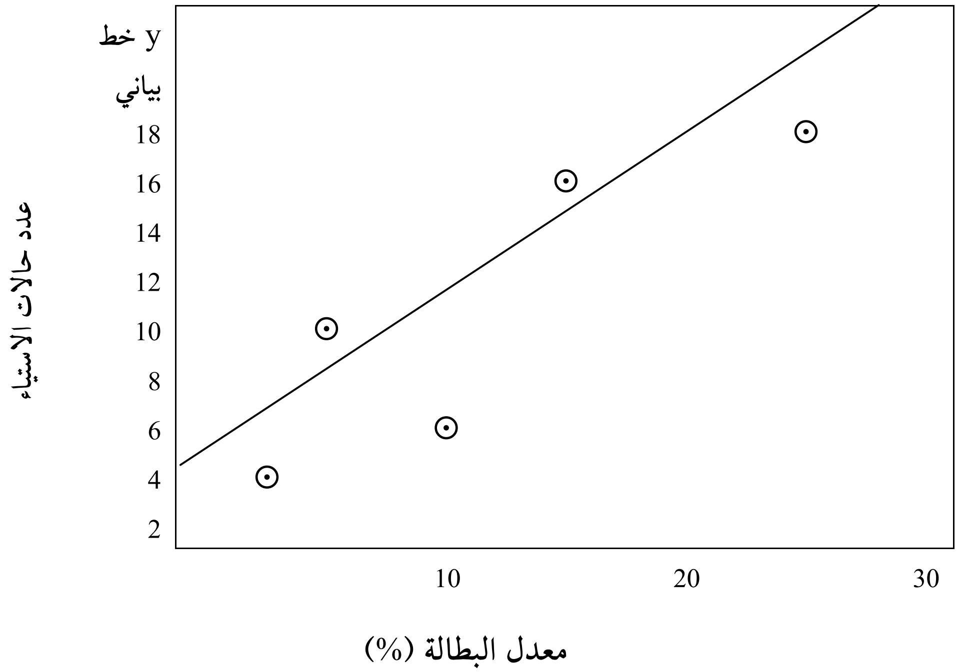
2- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Chart ⟸ Options (هذه العملية تقود إلى Scatter plot options ومن خلال ذلك يمكنك رؤية المساحة المعنونة بـِ: Fit Line التي لديها خيار معنون بـِ Total بعلامة ✓ بجانبه).

3- انقر على علامة ✓ بجانب الـِ Total. هذه العملية تضع علامة ✓ في المربع لبيان أنه تم الاختيار.

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:



شكل (9 ـ 10) مخرجات SPSS للرسم الانتشار مع خط الا نحدار

إجراءات الانحدار Regression مع تقدير المنحنى Curve Estimation باستخدام برنامج SPSS :

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(تقود هذه العملية إلى مربع الحوار Linear Regression dialog box).

2- انقر على Number of civil disturbances.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Number of Civil disturbances كمتغير تابع).

4- انقر على Unemployment rate (متغير مستقل).

5- انقر على ◄التي تشير إلى Indepedent(s) في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Unemployment rate كمتغير مستقل).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Regression

Variables Entered / Removed[b]

Model	Variable Entered	Variable Removed	Method
1	Unemployment Rate[a]		Enter

a. All requested Variables Entered

b. Dependent Variable: Number of Civil disturbances

Model Summery

Model	R	R Square	Adjusted R Squared	Std. Error of Estimate
1	.807[a]	.651	5.34	3.96

a. Predictors (Constant), unemployment Rate

ANOVA[b]

Model	Sum of squares	Df	MEAN square	F	Sig
Regression	87.701	1	37.701	5.586	.099[a]
Residual	47.099	3	15.700		
Total	134.800	4			

a. Predictors (Constant), unemployment Rate

b. Dependent Variable: Number of Civil disturbances

Correlation[a]

Model	Unstandardized coefficents		standardized coefficents	t	Sig
	B	Std Error	Beta		
(constant)1	4.423	3.013	.807	1.465	.239
Unemployment rate	.525	.222		2.364	.099

a. Dependent Variable: Number of Civil disturbances

المصدر: George Argyrous, op.cit, p. 218

شكل رقم (9ـ11) مخرجات SPSS للانحدار

تفسير مخرجات Spss للانحدار:

في مربع ملخص The Model Summery، نجد أن قيمة معامل ارتباط بيرسون (r) تساوي 807. ومعامل التحديد R square تساوي 651. وهاتان القيمتان هما نفس القيمتين اللتين تم حسابهما يدوياً. إن الجزء المهم من مربع معامل الارتباط Coefticients هو العمود المعنون بِـ B تحت معامل اللامعيارية Unstandardized coefficents ويعطينا هذا العمود المعلومات التالية:

- القيمة المتعلقة بِـ y-Intercept (والتي يطلق عليها a في التحليل أعلاه). إلا أن مخرجات SPSS يُطلق عليها كمية ثابتة Constant. وهي 4.423.

- ميل خط الانحدار The Slope of the regression Line الذي يشير إلى معامل الارتباط بمعدل البطالة 525.. مرة ثانية إن هذه النتائج هي نفسها التي تحصلنا عليها عند قيامنا بالعمليات الحسابية اليدوية.

اختبار t لمعامل الارتباط:

في الجزء الأول من هذا الفصل قد تم حساب خط الانحدار، ومعامل الارتباط الإحصائية المرتبطة بمجموعة من الحالات التي تم قياسها على المستوى ذي المسافات والنسبي. وقد تم التعامل مع هذه الإحصاءات الوصفية في إطار استقصاء العلاقة بين معدلات البطالة وحالات الاستياء الشعبي عبر مجموعة من المدن. والنتيجة التي توصلنا إليها كانت كالتالي:

(معدل البطالة) 0.53 + 4.4 = الاستياء الشعبي.

ومن خلال هذه الإحصاءات يمكننا الوصول إلى نتيجة مفادها إلى أن هناك تطابقاً موجباً وقويًّاً بين حالات الاستياء ومعدلات البطالة غير أنه يمكن القول، بأن هذه النتيجة المتحصل عليها من خلال عينة الدراسة قد لا تعكس ما يحدث في كل هذه المدن الخمس.

وكما هو الحال مع أي من الإحصاءات الوصفية الأخرى يمكن حسابها من عينة ما، فإننا بحاجة إلى أن نقرر ما إذا كان معامل الارتباط الذي يصف بيانات هذه العينة يعكس المجتمع الذي سحبت منه هذه العينة. فقد لا يكون هناك ارتباط بين هذين المتغيرين في المجتمع المكون لهذه المدن الخمس (ru)، وإنما يوجد فقط خطأ المعاينة الذي أدى بنا إلى اختيار هذه المدن الخمس التي قد لا تشبه باقي المدن الأخرى. وعليه، فإننا نحتاج إلى إجراء اختبار استدلالي لقيمة معامل الارتباط الذي تحصلنا عليه (r=0.81)[8].

في اختبار t لمعامل ارتباط بيرسون (r) يصاغ الفرض الصفري لهذا الاختبار: إنه لا يوجد ارتباط في المجتمع. في حين يصاغ الفرض البديل، إنه يوجد بعض الارتباط في المجتمع:

$$Ho : ru = 0$$

$$Hi : ru \neq 0$$

وبشكل واضح، فإن معامل ارتباط العينة قد وصل إلى: $r = 0.81$ وهي قيمة غير متطابقة مع الفرض الصفري. لكن السؤال المطروح هنا هو: هل باستطاعتنا رفض الفرض الصفري الذي مفاده: لا ارتباط في المجتمع على أساس نتيجة هذه العينة؟ ما هي الاحتمالية للحصول على عينة للمدن الخمس بارتباط يصل إلى $r = 0.81$ بين الاستياء الشعبي ومعدل البطالة من مجتمع يكون الارتباط فيه يساوي صفراً (0). وللحصول على هذه الاحتمالية، ينبغي علينا إجراء اختبار t مستخدمين المعادلة التالية:

$$t = \frac{r - ru}{sr}$$ العينة

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

وبالتعويض للقيم المرتبطة بـِ: r و r^2 لهذه المعادلة، فإننا نتحصل على:

$$S_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.81)^2}{5-2}}$$

$$= 0.34$$

$$t = \frac{0.81-0}{0.34}$$

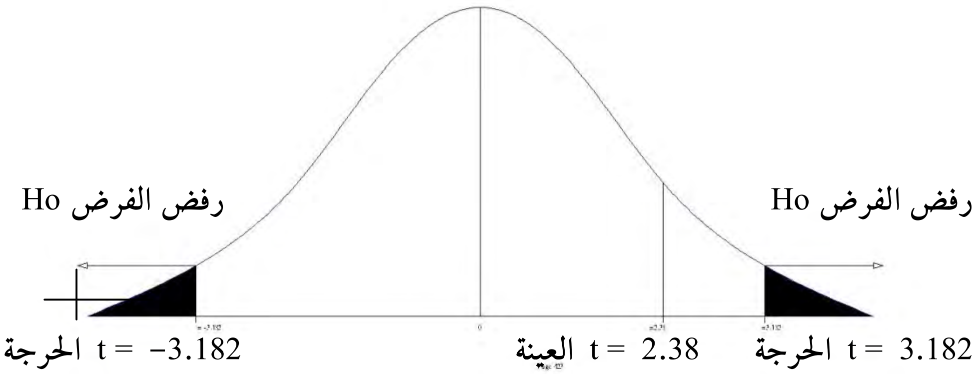
$$= 2.38$$

ولتحديد القيمة الحرجة لـِ t يمكننا الرجوع أو الاسترشاد بالجدول الذي يحتوي على القيم الحرجة لـِ t حيث إِنَّ درجة الحرية في هذا المثال تساوي N-2.

	مستوى الدلالة لاختبار أحادي الجانب				
	0.10	0.05	0.02	0.01	0.005
	مستوى الدلالة لاختبار ثنائي الجانب				
	0.20	0.10	0.05	0.02	0.01
df					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
.	…..	…..	…..	…..	…..
.	…..	…..	…..	…..	…..
.	…..	…..	…..	…..	…..
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.011
14	1.345	1.761	2.145	2.624	2.917
15	1.341	1.753	2.131	2.602	2.947
16	…..	…..	…..	…..	…..
∞	1.282	1.645	1.960	2.326	2.576

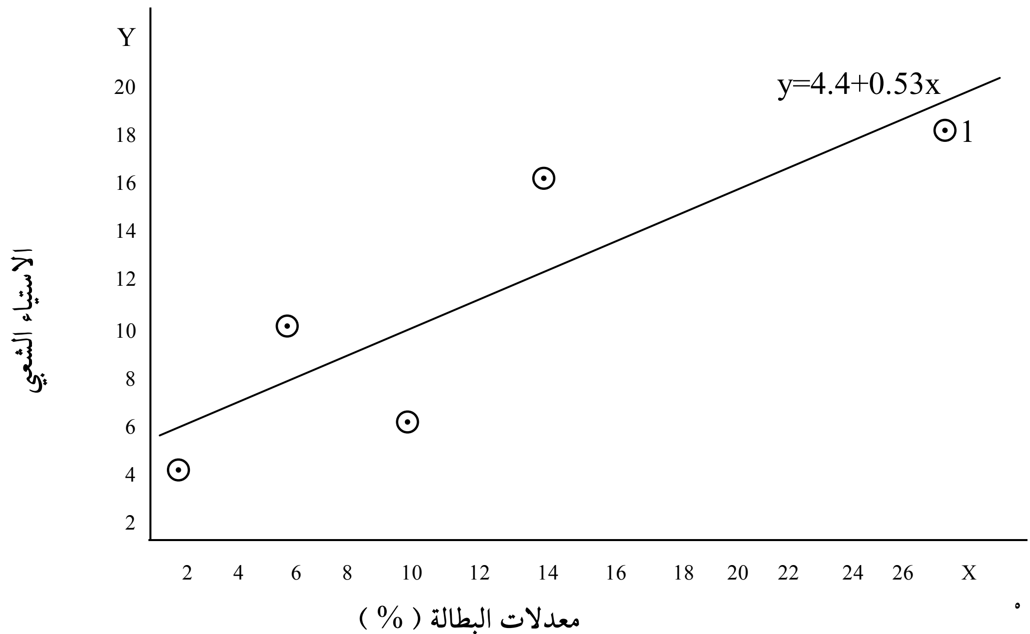
بمستوى ألفا 0.05 واختبار ثنائي الجانب، فإن درجة t الحرجة تكون كالتالي:
$\alpha = 0.05, df = 3$) t=3.182 الحرجة

وإذا وضعنا هذه المعلومات في شكل بياني (انظر الشكل رقم 11). فإننا بذلك نكون غير قادرين على رفض الفرض الصفري الذي مفاده لا يوجد ارتباط.



شكل (9-12) الدرجات الحرجة ودرجات العينة

إنه من الأهمية بمكان أن نتوقف لنفكر فيما حدث. ففي العينة التي بين أيدينا تحصلنا على ارتباط قوي بين المتغيرين X و y أي بين معدلات البطالة وحالات الاستياء الشعبي، ولقد أوضح لنا الاختبار الاستدلالي، أنه بالرغم من نتيجة هذه العينة التي ربما قد جاءت نتيجة لعامل الصدفة عند عملية المعاينة من المجتمع حيث إنّ هذين المتغيرين غير مرتبطين. ولنرى لِمَ لَمْ نستطع الإقرار بأن نتيجة العينة تعكس لنا هذه العلاقة في المجتمع؟ فإنه يصبح من الضرورة بمكان أن ننظر إلى شكل الانتشار لهذه البيانات مرة أخرى.



شكل (9 ـ 13) خط الانحدار لأقل المربعات (LSR)

يمكننا ملاحظة أن خط الانحدار قد تأثر بشكل كبير بالدرجة المتعلقة بمدينة رقم (1) التي يصل معدل البطالة فيها إلى 25 % يقابلها 17 حالة استياء شعبي. ولما كنا نتعامل مع عينة صغيرة (N=5)، فإن الحالة المتطرفة لاشك أنها تؤثر في النتيجة الكلية للعينة. فإذا كانت هذه الدرجة الواحدة مختلفة، كذلك خط الانحدار سيكون هو الآخر مختلفاً. وبما أن هذا الأمر محتمل، فمن غير الممكن أن يكون الارتباط القوي ذا دلالة عند التعامل مع العينات الصغيرة جداً [9].

اختبار الدلالة لارتباط بيرسون (r) باستخدام SPSS:

إجراء اختبار t لمعامل الارتباط الثنائي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Correlate ⟸ Bivariate

2- انقر على Number of civil disturbances.

3- انقر على ◄ مشيراً إلى المتغيرات في القائمة المحددة variable(s) هذه العملية تقوم بلصق Number of Civil disturbances في القائمة المحددة للمتغيرات (variable(s).

4- انقر على Unemployment rate.

5- انقر على ◄ مشيراً إلى القائمة المحددة للمتغيرات variable(s) (هذه العملية تقود إلى لصق Unemployment rate في القائمة المحددة للمتغيرات(variable(s)).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Correlations

Correlations

		Unemploment rate	Number of civil disturbances
Unemploment rate	Pearson Correlation sig.(2-tailed) N	1.000 5	.807 5
Number of civil disturbances	Pearson Correlation sig.(2-tailed) N	.807 5	1.000 5

المصدر: George Argyrous, op.cit, p. 425

شكل (9ـ14) مخرجات الارتباط الثنائي

تفسير اختبار الدلالة لمعامل ارتباط بيرسون (r) من خلال مخرجات SPSS:

إن اختبار الدلالة لارتباط بيرسون (r) يتولد كجزء من المخرجات عندما يجري تحليل الانحدار. إن الإجراء الذي تم إتباعه في الجزء الأول من هذا الفصل لتوليد إحصاءات الانحدار يولد لنا المعلومات الضرورية لإجراء اختبار الدلالة لهذه الإحصاءات.

تجدر الإشارة إلى أن هناك طرق أخرى بديلة من خلالها يمكننا توليد معامل الارتباط بين متغيرين وارتباط درجة t، ومستوى الدلالة.

ومن خلال أوامر الارتباطات بحزمة SPSS الثنائية Bivariate Correlation، يمكن للباحث توليد معامل الارتباط دون المعلومات الإضافية التي تأتي مع تحليل الانحدار الكامل[10].

من خلال النظر إلى مربع مخرجات الارتباط الثنائي يتبين لنا الارتباط بين المتغيرين. أولهما معدل البطالة مع نفسه. وثانيهما عدد حالات الاستياء الشعبي مع نفسه أيضاً. وكل واحد من هذين المتغيرين قد ولدًا معامل ارتباط 1.000. وهذه المعلومات تعتبر ضرورية طالما أن أي متغير بالضرورة يرتبط مع نفسه.

ففي الصف الأول من المربع يوضح لنا الارتباط بين معدل البطالة وحالات الاستياء الشعبي (.807)، وأن دلالة هذا الارتباط تصل إلى 099. وهذا يشير إلى أنه بالرغم من أن معامل الارتباط عالٍ، فهو ليس دالاً على مستوى (.05). وعليه يمكننا القول بأن هذه النتيجة جاءت كنتيجة لتباين المعاينة. أما الصف الثاني من المربع فيعطينا معامل الارتباط بين عدد حالات الاستياء الشعبي ومعدل البطالة وهو نفس المعامل المبين في الصف الأول من المربع.

اختبار t لمعامل ارتباط سبيرمان rho:

من خلال المثال السابق تعرفنا على الإجراءات المتعلقة بحساب معامل ارتباط بيرسون (r) ببحث ارتباط بيرسون (r) في استقصاء التطابق بين متغيرين تم قياسهما على المستوى ذي المسافات والنسبي. أما فيما يتعلق بارتباط سبيرمان، فإن هذا الإجراء يكون

ملائماً عندما يكون أحد المتغيرات على الأقل تم قياسه على المستوى الترتيبي. وإذا ما نظرنا بشكل دقيق إلى الإجراءات المتبعة في حساب هاذين النمطين من معامل الارتباط، فإننا نجد أنهما تقريباً متماثلان في الإجراءات، إلا أن الفرق بينهما هو أن معامل ارتباط (r) يمكن حسابه من البيانات الخام، في حين يعتمد ارتباط الرتب في حسابه على البيانات المرتبة.

تجدر الإشارة إلى أن اختبار الدلالة لكلا النمطين هو اختبار واحد، بمعنى، أن المعادلة لحساب درجة t للعينة هي واحدة بغض النظر، عما إذا كنا نختبر دلالة معامل ارتباط بيرسون (r) أو معامل الرتب $r_S(rhO)$.

$$t = \frac{r_S - P}{S_r} \quad \text{العينة}$$

$$S_r = \sqrt{\frac{1 - r_S^2}{n-2}}$$

إذا أراد الباحث حساب قيمة $r_S(rhO)$ فعليه أولاً حساب الفرق في الرتب لكل حالة من الحالات المدروسة، (D)، وبعد ذلك تربيع هذه الفروق. وبعد إنهاء هذه العمليات الحسابية ينبغي عليه تعويض تلك العمليات الحسابية في معادلة سبيرمان راهو (rhO).

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

فإذا توصلنا إلى ارتباط بين متغيرين يصل إلى $r_s = -0.8$ من بيانات تحتوي على $\sum D^2 = 1225.5$ لعدد 16 حالة.

فإننا بذلك يمكننا اتباع الإجراءات الخمسة المتعلقة باختبار الفروض:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: لا يوجد ارتباط بين X و y

Ho: P = 0

Hi: يوجد ارتباط بين X و y

Ho: P ≠ 0

الخطوة الثانية: اختيار اختبار الدلالة:

حيث إننا نبحث في الارتباط بين متغيرين تم قياسهما على المستوى الترتيبي. وإن البيانات تم وصفها من خلال حساب ارتباط سبيرمان rho للرتب. فإن الاستدلال المناسب هو اختبار t لمعامل الارتباط.

الخطوة الثالثة: حساب درجة العينة:

لقد تم حساب معامل ارتباط سبيرمان راهو بين المتغيرين X و y لعدد 16 حالة، وجاءت قيمة الارتباط 0.80− = r_S

$$r_S = \frac{6\sum D^2}{n(n_2 - 2)} = \frac{1225.5}{16(16^2 - 2)}$$

$$= -0.80$$

ولكي نرى ما إذا كانت هذه النتيجة قد جاءت للتباين العشوائي random variation عند إجراء عملية المعاينة من مجتمع حيث إنّ هذين المتغيرين لا يرتبطان بعضهما ببعض، فإننا بالتالي نحتاج أولاً إلى حساب الخطأ المعياري لتوزيع المعاينة لـِ rho.

$$S_r = \sqrt{\frac{1 - r_S^2}{n - 2} = 1 - \frac{(-0.80)^2}{16 - 2}} \ .$$

$$= 0.16$$

وعليه، فإن درجة t للعينة ستكون كالتالي:

$$t_{\text{العينة}} = \frac{r_S - P}{S_r} = \frac{-0.80}{0.16}$$

$$= -5$$

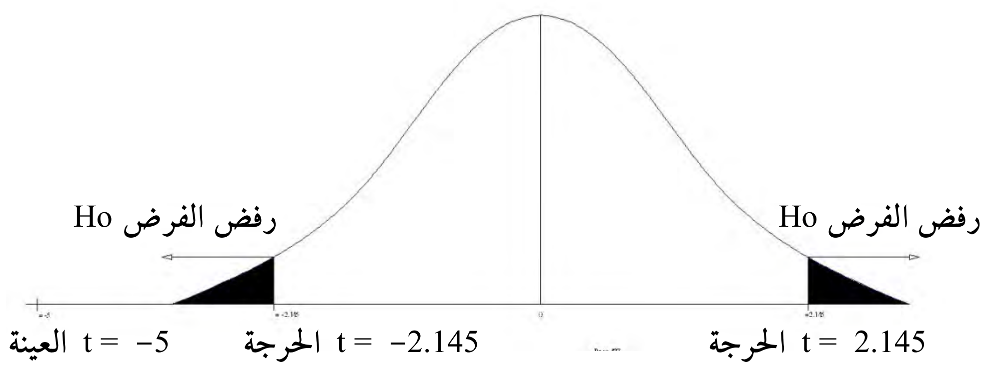
الخطوة الرابعة: حساب الدرجة (الدرجات) والمنطقة الحرجة:

بدلالة $\propto\ = 0.05$ لاختبار ثنائي الجانب، يمكننا الإشارة إلى جدول توزيع t للقيم الحرجة لكي تقرر القيمة الحرجة لـِ t بدرجة حرجة $df = 16 - 2 = 14$.

$$t = \pm 2.145\ (\propto = 0.05\ , df = 14)$$

الخطوة الخامسة: اتخاذ القرار:

إذا ما قمنا برسم درجات العينة والدرجات الحرجة في الشكل البياني التالي، فإنه يمكننا رفض الفرض الصفري. وعلى أساس هذه العلاقة القوية التي تـحصلنا عليها من خلال العينة المدروسة، فإنه ليس بإمكاننا القول بأنه لا يوجد ارتباط في المجتمع (11).



شكل (9-15) درجة العينة والدرجة الحرجة

أسئلة للمراجعة:

1- لماذا ينبغي على الباحث رسم شكل الانتشار للبيانات قبل إجراء تحليل الانحدار؟

2- ماذا تشير a و b في تحليل الانحدار؟

3- من خلال معادلات الانحدار بين اتجاه العلاقة؟

$y = 30 + 42 X$

$y = 30 - 0.38 X$

$y = -0.5 + 0.38 X$

$y = -0.5$

$y = -0.5 X$

4- اشرح الفرق بين ارتباط بيرسون (r) ومعامل خط الانحدار (b).

5- من البيانات التالية ارسم شكل الانتشار؟

a)

X	5	6	9	10	10	13	15	18	22	27
y	35	26	30	22	28	28	20	21	15	18

b) من خلال هذا الرسم الانتشاري، بين الإشارة التي تسبق المعامل (على سبيل المثال، علاقة موجبة أو علاقة سالبة).

6- ما هي توقعات أقل المربعات Least Sqnare لِـ y عندما تكون X = 12 = y = 27.165 - 0.15X.

7- إذا كانت معادلة خط الانحدار $y = 40 + 0.7X$ لخط انحدار تم رسمه لبيانات متعلقة بعدد سنوات الحياة المتوقعة Life expectancy ومصروفات الدولة على الرعاية الصحية لكل فرد من السكان لمجموعة من الدول النامية. وقد اعتمد عدد سنوات الحياة المتوقعة كمتغير تابع. ومصروفات الدولة على الرعاية الصحية كمتغير مستقل. وأن معادلة خط الانحدار كالتالي: $y = 40 + 0.7X$.

- ما هو عدد سنوات الحياة المتوقعة إذا قامت الدولة بصرف مبلغ قدره 30.000 دينار على الرعاية الصحية لكل فرد؟
- ما هو عدد سنوات الحياة المتوقع إذا لم تقم الدولة بصرف أية مبالغ مالية على الرعاية الصحية عندما تكون $y = 40$، $X = 0$.
- هل يمكنك القول بأنه توجد علاقة قوية بين المتغيرين؟

8- البيانات التالية تبين عدد الأطفال ومساهمة الزوج في أعمال المنزل (بيانات تصورية):

العائلة:	1	2	3	4	5	6	7	8	9	10	11	12
عدد الأطفال:	1	1	1	1	2	2	3	3	4	4	5	5
عدد الساعات الأسبوعية التي يقضيها الزوج في العمل المنزلي	1	2	3	5	3	1	5	0	6	3	7	4

- أوجد شكل الانتشار وماذا يعني ذلك؟
- أوجد قيمة a، b.
- ارسم معادلة خط الانحدار y على X.

9- مسح اجتماعي توصل إلى معامل ارتباط بين عدد السنوات الدراسية التي يقضيها المبحوث، واتجاهاته نحو تولي المرأة مهام إدارية عليا.

قيمة الارتباط (r) تساوي $r = 0.54$، وحجم العينة المدروسة في هذا المسح كانت 140 حالة، ودلالة هذا الارتباط كان اختبار t الذي وصلت قيمته إلى 7.54.

ما هي النتيجة التي يمكنك الوصول إليها حول طبيعة هذه العلاقة بين هذين المتغيرين؟

10- من البيانات التالية: ارسم شكل الانتشار ؟ أوجد معادلة خط الانحدار y على X. إذا كانت X = 16. ما هي القيمة التقديرية لـِ y.

$$b = 0.87, a = 3.775$$

$$\hat{y} = --- + --- X$$

X	18	14	10	15	7	12	13	8	9	17	15	12
y	20	11	14	16	10	10	17	11	12	20	18	12

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss Sage Publications ,London, 2001, P. 201.
2- Ibid, PP. 201 - 209.
3- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003، ص 379.
4- George Argyrous, op.cit., PP. 209 - 212.
5- Ibid, P. 213.
6- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, 2ed, Wadsworth Cengage Learning, UAS, 2010, PP. 343 - 345.
7- Ibid, PP. 349 - 350 and George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the Megraw Hill, Companies, Inc. USA, 1992, P.279.
8- George Argyrous, op.cit., P. 421.
9- Ibid, P. 424.
10- Ibid, P. 424.
11- Ibid, PP. 425 - 428.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciences, 8th ed, USA, 2010.
2- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss, Sage Publications ,London, 2001.
3- George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the MCGraw- Hill, Companies, Inc. USA, 1992.
4- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, Wadsworth Cengage Learning, UAS, 2010.
5- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003.