واجب مادة الإحصاء

ارتباط بيرسون وسبيرمان

اعداد الطالب / محمد متولي

ارتباط بيرسون وسبيرمان

ما هو الارتباط؟ (ارتباط بيرسون)

وهناك عدد الربط بين -1 و +1 أن يقيس درجة العلاقة بين المتغيرين (تسميها سين وصاد). وثمة قيمة إيجابية للارتباط يعني ارتباط إيجابي (العاشر من القيم الكبيرة عادة ما تكون مرتبطة بالقيم الكبيرة والصغيرة نعم قيم العاشر عادة ما تكون مرتبطة الصغيرة قيم صاد). القيمة السلبية للارتباط ينطوي على آثار سلبية أو عكسية الارتباط (العاشر من القيم الكبيرة عادة ما تكون مرتبطة الصغيرة قيم نعم ، والعكس بالعكس).

صيغة للارتباط بيرسون

لنفترض أن لدينا متغيرين سين وصاد ، مع وسائل XBAR وYBAR انحرافات قياسية على التوالي والعاشر ود ي د على التوالي. الارتباط هو بوصفها

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{(n-1)S_X S_Y}$$

هناك بعض طرق مختصرة ، ولكن بشكل عام فإن الصيغة مملة ونبلغ الكمبيوتر القيام بكل هذا العمل.

متى يكون وجود ارتباط إيجابي؟

لنفترض أن قيمة س فوق المتوسط ، والتي يرتبط بها من نعم القيمة أيضا فوق المتوسط. ثم المنتج

$$\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

وسيكون المنتج اثنين من الأرقام الإيجابية التي من شأنها أن تكون إيجابية. إذا كانت القيمة العاشر وصاد ، هي قيمة أقل من المتوسط ، فإن المنتج أعلاه ستكون سلبية عدد اثنين ، الأمر الذي من شأنه أيضا أن يكون إيجابيا.

وبالتالي ، وجود ارتباط إيجابي هو دليل على وجود اتجاه عام بأن القيم الكبيرة العاشر ترتبط بالقيم الكبيرة والصغيرة نعم قيم العاشر ترتبط الصغيرة قيم ي.

متى يكون وجود ارتباط سلبي؟

لنفترض أن قيمة س فوق المتوسط ، والتي يرتبط بها من نعم وكانت قيمة بدل دون المتوسط. ثم المنتج

$$\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

وسيكون نتيجة إيجابية وسلبية في عدد من شأنها أن تجعل المنتج سلبية. س إذا كانت قيمة أقل من المتوسط ، ونعم وكانت قيمة أعلى من المتوسط ، فإن المنتج أعلاه سوف يكون سلبيا.

ولذلك ، ترابط سلبي هو دليل على وجود اتجاه عام بأن القيم الكبيرة العاشر ترتبط الصغيرة قيم نعم الصغيرة وقيم العاشر ترتبط بقيم كبيرة ي.

مثال

لنكن وأحصى وجود ارتباط بين معامل 1 دقيقة عشرات APGAR (العاشر) ، و 5 في الدقيقة APGAR عشرات (نعم). فيما يلي جدول يبين بعض من وسيطة calcuations.

X_i	Y_i	$\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)$
3	3	(3-7.9)*(3-8.8)=-4.9*-5.8=28.42
3	6	(3-7.9)*(6-8.8)=-4.9*-2.8=13.72
6	9	(6-7.9)*(9-8.8)=-1.9*0.2=-0.38
7	9	(7-7.9)*(9-8.8)=-0.9*0.2=-0.18
7	10	(7-7.9)*(10-8.8)=-0.9*1.2=-1.08
8	3	(8-7.9)*(3-8.8)=0.1*-5.8=-0.58
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	9	(8-7.9)*(9-8.8)=0.1*0.2=0.02
8	10	(8-7.9)*(10-8.8)=0.1*1.2=0.12
9	9	(9-7.9)*(9-8.8)=1.1*0.2=0.22
9	9	(9-7.9)*(9-8.8)=1.1*0.2=0.22
9	9	(9-7.9)*(9-8.8)=1.1*0.2=0.22
9	10	(9-7.9)*(10-8.8)=1.1*1.2=1.32
9	10	(9-7.9)*(10-8.8)=1.1*1.2=1.32
9	10	(9-7.9)*(10-8.8)=1.1*1.2=1.32
9	10	(9-7.9)*(10-8.8)=1.1*1.2=1.32
9	10	(9-7.9)*(10-8.8)=1.1*1.2=1.32
10	10	(10-7.9)*(10-8.8)=2.1*1.2=2.52
10	10	(10-7.9)*(10-8.8)=2.1*1.2=2.52

$$\bar{X}=7.9, S_X=1.75$$

$$\bar{Y}=8.7, S_Y=1.99$$

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{(n-1)S_X S_Y}$$

$$r=\frac{52.62}{(24-1)\times1.75\times1.99}$$

$$r=0.66$$

تفسير معامل الارتباط.

معامل الارتباط التدابير قوة علاقة خطية بين المتغيرين.

معامل الارتباط دائما بين -1 و +1. توثيق الارتباط هو + / -1 ، واقرب الى الكمال علاقة خطية. هنا هو كيف يمكن لي أن تميل إلى تفسير الارتباطات.

- -1.0 الى -0.7 قوية الارتباط السلبي.
- -0.7 الى -0.3 ضعف الارتباط السلبي.
- -0.3 ل+0.3 رابطة ضئيلة أو معدومة.
- +0.3 لضعف ارتباط إيجابي +0.7.
- +0.7 ل+1.0 ارتباط إيجابي قوي.

هذه القاعدة ، بطبيعة الحال ، إلى حد ما تعسفيا. لبعض الحالات ، أنا mught الانتقال الوقف أقرب إلى القيم 0 (على سبيل المثال ، 0 ،.2 و 0.6) ، وحالات أخرى ، وربما نقل قطع القيم اقرب الى 1 (على سبيل المثال ، 0.4 و 0.8).

ومن الأمثلة على ارتباط إيجابي قوي.

العلاقة بين لزوجة الدم والخلايا معبأة حجم 0.88.

نلاحظ أن كميات صغيرة تميل إلى أن تكون منخفضة اللزوجة والأحجام الكبيرة تميل إلى أن تكون لزوجة عالية.

ومن الأمثلة على ضعف ارتباط إيجابي.

العلاقة بين لزوجة الدم وfibrogen هو 0.46.

نلاحظ أن هناك أيضا اتجاه إلى قيم صغيرة fibrogen منخفضة اللزوجة والقيم الكبيرة fibrogen عالية اللزوجة. هذا الاتجاه ، مع ذلك ، هو أقل وضوحا مما كان عليه الحال في المثال السابق.

مثال رابطة ضئيلة أو معدومة.

العلاقة بين لزوجة الدم والبلازما -0.10 البروتين.

انخفاض مستويات بروتين يرتبط كل من اللزوجة العالية والمنخفضة. مستويات عالية من البروتين ، ترتبط هي أيضا مع كل من اللزوجة العالية والمنخفضة.

مصفوفة الارتباط.

عندما يكون لديك أكثر من المتغيرات ، يمكنك ترتيب الارتباطات بين كل زوج في المصفوفة.

في أسفل هذه الصفحة مثال على استخدام البيانات لزوجة الدم.

لإنشاء هذا الجدول ، حدد تحليل | ربط | يحتوي على متغيرين من الإحصائي للعلوم الاجتماعية القائمة.

التقريب ويساعد وجود ارتباط النموذج.

في أسفل الصفحة نفسها مصفوفة الارتباط ، مضروبة في 100 واعتقلت اثنين من أرقام كبيرة.

لدينا أيضا إزالة بعض من المعلومات دخيلة.

-- -- الارتباط المعاملات -- --

	فيس	PCV	فرع التحقيقات المالية	PROT
VISCOS	100	88	46	10-
PCV	88	100	42	16-
FIBROGEN	46	42	100	5-
البروتين	10-	16-	5-	100

Scatterplot المصفوفة.

كما يمكنك ترتيب scatterplots في نمط مماثل.

لإنشاء هذا الرسم البياني ، حدد الرسوم البيانية من التبعثر الإحصائي للعلوم الاجتماعية القائمة ثم حدد مصفوفة من مربع الحوار.

<u>تفسير الارتباطات</u>.

ينبغي عليك توخي الحذر في عدم الربط overinterpret المعاملات. لا تفترض أن علاقة السببية على قدم المساواة. كما نتوخى الحذر بشأن كيفية جمع البيانات. وهناك ضيقا يقتصر العينة يمكن أن يؤدي إلى انكماش في الارتباط.

ولا تعبر عن علاقة السبب والنتيجة.

مبيعات الروم وعدد من وزراء الميثوديات يرتبط ارتباطا إيجابيا ، ولكن عدد كبير من الوزراء لا يشجع الروم الشرب.

فهل من ثالث المتغيرات التي تؤثر في كل من المبيعات والروم الميثوديات وزراء؟

إن المثال السابق ، فإن كلا من مبيعات الروم وعدد من وزراء الميثوديون ارتباط مع عدد من الأشخاص في الولايات المتحدة ، كما يزيد من عدد من الناس ، وأنها تؤدي إلى زيادة في الطلب على كل من وزراء الميثوديات والروم.

إذا كنت للتعديل وفقا لعدد من الناس ، مثلا عن طريق الكمبيوتر مبيعات الروم وعدد من الوزراء للفرد الواحد ، فإن الرابطة سوف تختفي.

وهناك أمثلة كثيرة فيها ارتباط وثيق بين متغيرين ويمكن تفسير ثالث عامل. يبحث دائما عن تفسير بديل للعلاقة.

على سبيل المثال ، القش الغلة ارتباطا سلبيا على متوسط درجة الحرارة في فصل الربيع. ويبدو أن هذا counterintuitive. ولكن من السهل فهم بمجرد أن ندرك أن غلة العلف تعتمد اعتمادا كبيرا على سقوط الأمطار في فصل الربيع. الأمطار والربيع عادة أكثر برودة جاف الربيع.

تقييد المدى.

إذا كانت واحدة من المتغيرات مصطنعة يقتصر المدى ، فإن الارتباط سيكون على شفا الصفر.

العلاقة المتبادلة بين m inute1 و 5 دقائق APGAR عشرات هو 0.66.

إذا كان لنا أن مجموعة البيانات التي تحد من الرضع لمدة دقيقة واحدة مع APGAR< (5) ، ثم ينخفض الى 0.25 ارتباط.

هناك الكثير من النقاش حول مدى أهمية الشركة العشرات في التنبؤ الفرد النجاح في الكلية. معظم الكليات معلومات عن الشركة العشرات من الطلاب ، واتخاذ تدابير لنجاحها ، مثل الصف نقطة طالبة في المتوسط خلال العام.

هذه البيانات ، ولكن غير مؤكد وتوفر أدلة على العلاقة بين الشركة وعشرات الدرجات. معظم الكليات تقيد المسجلين لديها أعلى من طائفة معينة للالشركة. بالنسبة لبعض الكليات ، وهذا يمكن أن يؤدي إلى مجموعة محدودة للغاية من الشركة العشرات. إذا كانت هذه البيانات التي تظهر ضعف الارتباط ، ليس من الواضح ما إذا كان هذا يحدث بسبب القيود المصطنعة في الشركة مجموعة من النقاط.

أفضل ، ولكن ربما غير عملي ، وسيلة لتقييم هذه الحالة الكلية لجميع الوافدين على الاعتراف بغض النظر عن الشركة وبعد ذلك نرى ما إذا كان هناك ارتباط بين الشركة وعشرات برنامج العمل العالمي.

معامل الارتباط بيرسون

التعريف : تدابير قوة من علاقة خطية بين المتغيرين.

الافتراضات : كل من المتغيرات (غالبا ما يطلق على سين وصاد) الفاصلة / النسبة تقريبا ، وعادة ما توزع ، ويحتوي عاى متغيرين مشتركة التوزيع العادي.

الخصائص : معامل الارتباط بيرسون وعادة ما تدل عليه *ص* (رو) ، ويمكن أن يتخذ على القيم من 1.0 إلى -1.0. -1.0 فيه خير السلبية (العكسية) ارتباط ، ولا علاقة 0.0 ، و 1.0 هو ارتباط إيجابي الكمال.

الإحصاءات ذات الصلة : س ^{2} (تسمى معامل ص التصميم أو مربع) ويمكن أن تفسر على أنها نسبة الفرق في نعم هذا هو الوارد في العاشر

الاختبارات : إحصائية أهمية *ص* تختبر باستخدام اختبار (ت). الفرضيات لهذا الاختبار هي :

ح $_0$: رو = 0

ح $_{أ}$: رو <> 0

ع المنخفض القيمة لهذا الاختبار (أقل من 0.05 على سبيل المثال) يعني أن هناك ما يدل على رفض فرضية باطلة لصالح الفرضية البديلة ، أو أن هناك علاقة ذات دلالة احصائية بين المتغيرين.

ملاحظة : هذا الاختبار وتعادل لاختبار أي منحدر في الانحدار الخطي البسيط الإجراء.

موقع الغمزات : معامل ارتباط بيرسون وجدت في المواقع التالية :

1. الانحدار والارتباط -- وتنتج كل من إجراءات الارتباط بيرسون ورماح معاملات الارتباط. تي الإحصائية لاختبار أهمية يحسب ص. آر ^{2} أيضا.

2. الانحدار والارتباط -- الانحدار الخطي البسيط تقارير معامل ارتباط بيرسون ، واختبار (ت). آر ^{2} أيضا.

3. الانحدار والارتباط -- الارتباط مصفوفة الإجراءات تنتج مصفوفة الارتباطات لعدد من أزواج من المتغيرات في وقت واحد ، ويشمل ع القيمة للاختبار أو أهمية *ص*.

الرسوم البيانية : ويتمثل جزء هام من تفسير (ص) هو الالتزام scatterplot للبيانات. Scatterplots متاحة من رسوم بيانية الخيار ، وذلك كجزء من الانحدار الخطي البسيط والارتباط الرسومية في مصفوفة الخيار في الانحدار والارتباط.

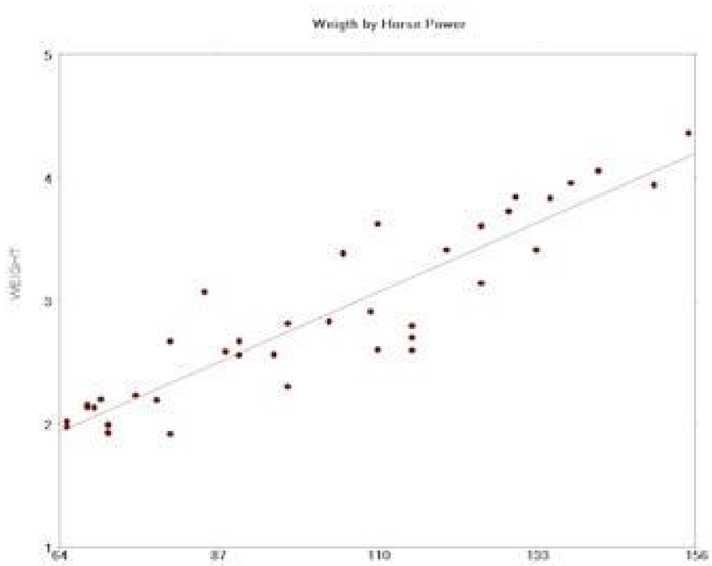
على سبيل المثال : استخدم الارتباط إجراء حساب *لص* المتغيرين حصان (حصان) ووزنه في [الغمزات](#) "السيارات" في قاعدة البيانات. نتائج من الغمزات (جزئيا) :

المتغيرات المستخدمة : إتش بي ، وعدد حالات الوزن المستخدمة : 38 ص بيرسون (الإرتباطات المعامل) س = 0.9172 - 0.8413 = ساحة اختبار الفرضية لتحديد أهمية العلاقة :

حاء (باطل) : = 0 المنحدرات أو حاء (باطل) : ص = 0 (بيرسون)

ر 36 = 13.81425 مع دال إلى واو ع <0.001

(ع) انخفاض القيمة يعني أن انحدار لا = 0). الرامح تصنيف معامل الارتباط = 0.9071 (رماح) = 12.93361 ر 36 دال إلى واو ع <0.001 ألف scatterplot هذه البيانات ويبين ارتباط إيجابي -- سيارات أعلى حصان تميل يزن أكثر من ذلك :



مثال على كتابة هذه النتائج :

السرد : "تم إجراء تقييم للعلاقة خطية بين حصان ووزن السيارة باستخدام ارتباط بيرسون".

النتائج : "تحليل باستخدام معامل الارتباط بيرسون يدل دلالة إحصائية علاقة خطية بين حصان ووزن السيارة ص (36) = 0.92 ، ع <0.001 لهذه البيانات ، وتعني (التنمية المستدامة) للحصان هو 101.7 (26.4) والوزن 2.86 (0.71). "

تحذير : هناك محاولات لاستنتاج السبب والنتيجة وجود ارتباط عند مراقبة. ومع ذلك ، فإن القدرة على احالة السببية تعتمد على خلق تجربة المصممة خصيصا لتوفير هذا النوع من الاستدلال.

مواضيع ذات صلة : رماح 's معامل الارتباط هو غير بارامترية النظير لr. انظر أيضا الانحدار الخطي البسيط ، الانحدار المتعدد ، الانحدار متعدد الحدود.

مسألة عن الارتباط

في بداية دورة تمهيدية الهندسة ، و 10 طالبا حصلوا على التجارب السابقة في تحديد قدرة الأولية الرياضية. يتضمن الجدول التالي قائمة الطلاب قبل الاختبار الدرجة والدرجة النهائية في الصف :

عدد الطلاب	التجارب السابقة	طبعا الصف
1	45	92
2	23	86
3	50	97
4	46	95
5	33	87
6	21	76
7	13	72
8	30	84
9	34	85
10	50	98

1. حساب معامل الارتباط بيرسون (ص) على هذه البيانات.

ص =

2. ما هو اختبار الاحصائية المستخدمة لتحديد ما إذا كانت هذه القيمة (ص) هو دلالة إحصائية؟

3. هو الارتباط في هذه البيانات الإحصائية. لماذا؟

4. عرض scatterplot للبيانات. هل تظهر البيانات ارتباطا خطيا. تفعل يبدو أن هناك أي الخارجة القيم؟

5. لنفترض أن أحد الطلاب th11 أضيفت إلى البيانات ، مع التجربة السابقة برصيد 40 ، ودورة من 70 فصلا. كيف هذا الغرض ص؟

معامل الارتباط بيرسون ، ص

الربط تقنية للتحقيق في العلاقة بين اثنين من المؤشرات الكمية والمتغيرات المتواصلة، على سبيل المثال، والعمر وضغط الدم. معامل ارتباط بيرسون قال (ص) هي مقياس لمدى قوة الارتباط بين المتغيرين.

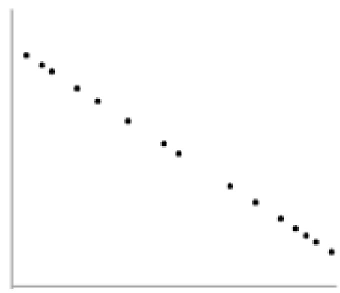
الخطوة الأولى في دراسة العلاقة بين المتغيرات المستمرة هو رسم التشرذم مؤامرة للمتغيرات للتأكد الخطي. معامل الترابط لا ينبغي أن يحسب إذا العلاقة ليست خطية لأغراض الربط فقط ، لا على سبيل الحقيقة محور المتغيرات التي يتم التخطيط لها. ومع ذلك ، تقليديا ، المستقل (أو التفسيرية) متغير هو تآمر على المحور السيني (أفقيا) وتعتمد (أو الاستجابة) المتغير هو تآمر على العمودي (الرأسي).

فإن اقتراب التشرذم من النقاط هو خط مستقيم ، وكلما زادت قوة العلاقة بين المتغيرات. كذلك ، لا يهم ما هي وحدات القياس المستخدمة.

قيم معامل الارتباط بيرسون

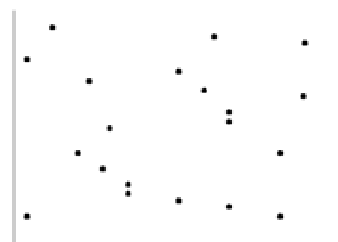
معامل ارتباط بيرسون قال (ص) لمواصلة (الفاصلة) بيانات تتراوح -1 إلى +1 :

ص = **-1**



بيانات مثالية تقع على خط مستقيم مع سلبي المنحدر

ص = **0**



لا علاقة خطية بين المتغيرات

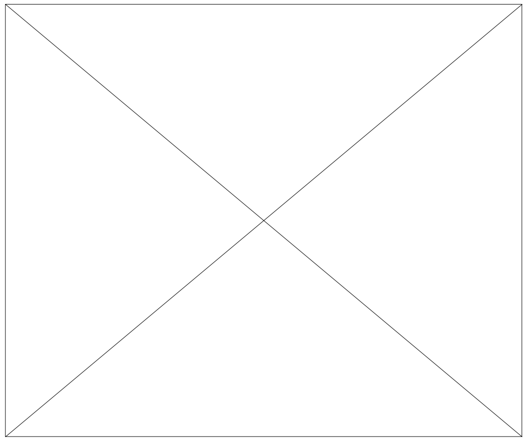
ص = **+1**

بيانات مثالية تقع على خط مستقيم مع إيجابية المنحدر

ارتباط إيجابي يدل على أن كل من المتغيرات زيادة أو نقصان معا ، في حين ترابط سلبي يشير إلى أنه يزيد من متغير واحد ، وبالتالي فإن انخفاض أخرى ، والعكس بالعكس.

مثال Scatterplots

تحديد القيمة التقريبية للمعامل ارتباط بيرسون. هناك 8 الخرائط ، واختيار الجواب الصحيح ، سوف تنتقل تلقائيا في الرسم البياني التالي.



نصيحة : إن مربع من معامل ارتباط يشير إلى نسبة تباين متغير واحد 'وأوضح' من قبل الآخرين (انظر كامبل & ماشين ، 1999 لمزيد من التفاصيل).

ملاحظة

الاختبار تي يستخدم لتحديد ما اذا كان معامل الارتباط تختلف اختلافا كبيرا عن الصفر ، وبالتالي أن هناك دليلا على وجود علاقة بين المتغيرين. ثم هناك الافتراض أن البيانات عن التوزيع العادي عينات عشوائية. وإذا لم يكن هذا صحيحا ، فإن النتائج قد تكون باطلة. إذا كان الأمر كذلك ، فمن الأفضل استخدام معامل رماح من رتبة ارتباط (غير بارامترية المتغيرات). انظر & ماشين كامبل (1999) التذييل A12 للحسابات وأكثر من هذا النقاش.

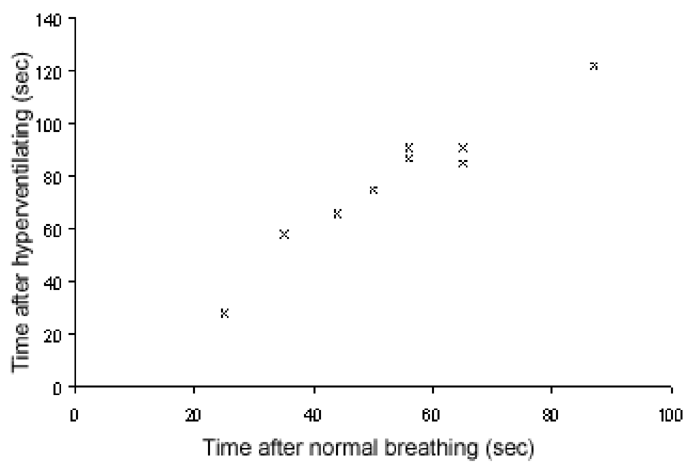
ومن المثير للاهتمام أن نلاحظ أنه مع عينات أكبر ، وانخفاض قوة الارتباط ، على سبيل المثال (ص) = 0.3 ، ويمكن أن يكون دلالة إحصائية (أي ع <0.01). ومع ذلك ، فإن هذا يدل على قوة ذات معنى الارتباط؟

ملاحظة : فقط لان اثنين من المتغيرات ذات الصلة ، لا يعني بالضرورة أن واحدة مباشرة لأسباب أخرى!

وعملت على سبيل المثال:

تسعة طلاب عقدت في التنفس ، ومرة واحدة بعد أن يتنفس بشكل طبيعي والاسترخاء لمدة دقيقة واحدة ، وhyperventilating لمرة واحدة بعد دقيقة واحدة. ويبين الجدول متى (ثانية) أنهم كانوا قادرين على التنفس عقد هل هناك وجود علاقة بين المتغيرين؟

موضوع	أ	باء	جيم	مد	هاء	واو	ع	حاء	أنا
طبيعي	**56**	**56**	**65**	**65**	**50**	**25**	**87**	**44**	**35**
Hypervent	**87**	**91**	**85**	**91**	**75**	**28**	**122**	**66**	**58**



ويبين الرسم البياني التشرذم مؤامرة (تعادل ماجستير في إكسل) من البيانات ، مما يدل على مدى معقولية افتراض خطية العلاقة بين المتغيرات.

Hyperventilating مرات تعتبر أن المتغير ، لذلك هي تآمر على المحور الرأسي.

من الناتج الإحصائي للعلوم الاجتماعية و Minitab ترد أدناه :

الإحصائي للعلوم الاجتماعية حدد تحليل> الارتباط> الثنائية Minitab variate :

Correlations

		NORMAL	HYPER
NORMAL	Pearson Correlation	1	.966**
	Sig. (2 tailed)	.	.000
	N	9	9
HYPER	Pearson Correlation	.966**	1
	Sig. (2 tailed)	.000	.
	N	9	9

**. Correlation is significant at the 0.01 level (2-tailed).

الارتباطات : عادي ، Hypervent

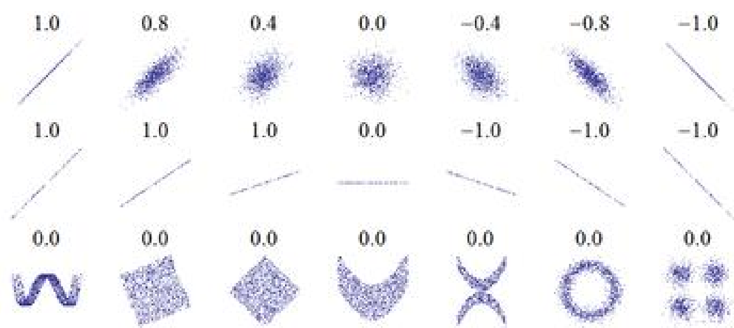
ارتباط بيرسون للمعلمين و Hypervent = 0.966

ف القيمة = 0.000

وفي الختام ، مطبوعات تبين أن قوة العلاقة بين المتغيرات التي هي عالية جدا (ص = 0.966) ، وإلى أن معامل ارتباط عالية جدا يختلف كثيرا عن الصفر (ف <0.001). كما يمكننا ان نقول ان 93 ٪ (0.966 ^{2}) من التباين في hyperventilating مرات يفسر التنفس الطبيعي.

ارتباط

هذه المادة عن معامل الارتباط بين المتغيرين.



عدة مجموعات من *(خ ، ذ)* نقطة ، مع ارتباط معامل *س* و *ص* في كل مجموعة. علما بأن علاقة ضجيج ويعكس اتجاه علاقة خطية (السطر العلوي) ، ولكن ليس انحدار العلاقة (وسط) ، كما أن جوانب كثيرة من العلاقات غير الخطية (القاع). ملاحظة : هذا الرقم في المركز منحدر من 0 ولكن في هذه الحالة يكون معامل الارتباط غير معروف ، لأن الفرق هو *نعم* الصفر.

في الإحصاءات ، وارتباط (وكثيرا ما يقاس معامل ارتباط ، ρ) تدل على قوة واتجاه علاقة *خطية* بين المتغيرات العشوائية. وهذا هو على النقيض من استخدام تعبير بالعامية ، والتي تدل على أي علاقة ، وليس بالضرورة الخطية. إحصائية عامة في الاستخدام ، أو المشاركة في *ترابط* العلاقة تشير إلى رحيل اثنين من المتغيرات العشوائية الاستقلال. بهذا المعنى الواسع هناك العديد من المعاملات ، وقياس درجة الارتباط ، وتكييفها لطبيعة البيانات.

بيرسون منتجات معامل لحظة

المادة الرئيسية: بيرسون المنتجات معامل الارتباط لحظة

وهناك عدد من معاملات مختلفة تستخدم لمختلف الحالات. أشهرها هي بيرسون المنتجات معامل الارتباط لحظة ، التي تم الحصول عليها عن طريق قسمة التغير المتكافئ من جانب اثنين من المتغيرات المنتج من الانحرافات المعيارية. على الرغم من اسمها ، لأول مرة من قبل فرنسيس غالتون. [1]

الخصائص الرياضية

معامل الارتباط $\rho_{\text{سين وصاد}}$ بين المتغيرات العشوائية *سين وصاد* مع القيم المتوقع μ μ $_{\text{العاشر وصاد}}$ والانحرافات المعيارية و σ σ $_{\text{س ص}}$ وتعرف على النحو التالي :

$$\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y},$$

هاء حيث هي القيمة المتوقعة لتشغيل وسائل cov التغاير. وتستخدم على نطاق واسع هو البديل التأشير

$$\mathrm{corr}(X,Y)=\rho_{X,Y}\ .$$

منذ هاء $\mu_{\text{س}}$ = *(س)* ، $_{\text{س}}$ ^{2} = σ هاء *[(س* -- هاء *(العاشر)* $^{(2)}$ = هاء *(س* $^{2)}$ -- هاء ^{2} *(س)* ، وكذلك *لصاد* ، ومنذ *هاء [(س* -- *هاء (س)) (ص* -- *هاء (نعم))]* = *هاء (سين صاد)* -- *هاء (العاشر) هاء (ص)* ، ويمكن أيضا أن يكتب لنا

$$\rho_{X,Y}=\frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\ \sqrt{E(Y^2)-E^2(Y)}}.$$

تعريف الارتباط إلا إذا كان كل من انحرافات قياسية محدودة وكلاهما صفر غير. بل هو نتيجة طبيعية للCauchy التفاوت بين شفارتز أن ارتباط لا يمكن أن تتجاوز 1 في القيمة المطلقة.

1 الارتباط هو في حالة وجود علاقة خطية زيادة -1 ، في حالة وجود علاقة خطية يتناقص ، وبعض في القيمة بين غيرها في جميع الحالات ، مما يشير إلى درجة الاعتماد الخطية بين المتغيرات. أقرب المعامل إما 1 أو -1 ، وقوة الارتباط بين المتغيرات.

إذا كانت المتغيرات المستقلة ثم الربط 0 ، ولكن العكس ليس صحيحا لأن تكتشف إلا معامل الارتباط الخطي بين المتغيرات التابعة. هنا مثالا : لنفترض أن المتغير العشوائي *العاشر* موحدة وزعت على الفاصلة من -1 إلى 1 ، *ص* = *س* $^{2.}$ ثم *نعم* تماما تحددها *العاشر* ، بحيث *سين وصاد* تتوقف ، ولكن الربط الصفر ، فهي غير مربوط. ومع ذلك ، في حالة خاصة عندما *سين وصاد* هي طبيعية مشتركة ، أي ما يعادل uncorrelatedness الاستقلال.

وهناك ارتباط بين متغيرين هو تمييع في وجود خطأ في القياس التقديرات حول واحد أو كل من المتغيرات ، وفي هذه الحالة disattenuation يوفر أكثر دقة المعامل.

نماذج الربط

إذا كان لدينا سلسلة من قياسات *ن سين وصاد* مكتوبة كما *وذ خ* *ط ط* حيث *ط* = 1 ، 2 ،... ، *ن* ، ثم بيرسون المنتجات معامل الارتباط لحظة يمكن أن تستخدم لتقدير العلاقة بين *العاشر* و *نعم*. بيرسون المعامل المعروفة أيضا باسم "العينة معامل الارتباط". معامل ارتباط بيرسون بعد ذلك أفضل تقدير للارتباط من *سين وصاد* . معامل ارتباط بيرسون هو مكتوب :

$$r_{xy}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y},$$

أين $\bar{x}$ و $\bar{y}$ هي عينة من وسائل *للسين وصاد* ، *وس ق ق ذ* هي عينة الانحرافات المعيارية *للسين وصاد* ، وهذا المبلغ هو من 1 إلى *ط* = *ن*. ومع ارتباط السكان، فإننا يمكن أن تعيد كتابة هذا النحو

خطأ!

$$r_{xy}=\frac{\sum x_iy_i-n\bar{x}\bar{y}}{(n-1)s_xs_y}=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-(\sum x_i)^2}~\sqrt{n\sum y_i^2-(\sum y_i)^2}}.$$

مرة أخرى ، تماما كما هي الحال في علاقة متبادلة مع السكان ، والقيمة المطلقة للالعينة الارتباط يجب أن تكون أقل من أو تساوي 1. رغم ما تقدم مربح تقترح صيغة واحدة لحساب تمرير خوارزمية عينة المتبادلة ، وتشتهر لعدم الاستقرار العددية (انظر أدناه للحصول على شيء أكثر دقة).

مربع معامل ارتباط العينة ، وهي المعروفة أيضا معامل التصميم ، هو جزء بسيط من الفرق *في ذ* ط التي استأثرت بها خطية من يصلح *لذ خ ط ط*. هذا هو المكتوب

$$r_{xy}^2=1-\frac{s_{y|x}^2}{s_y^2},$$

حيث *ق ذ | س* ^{2} هي مساحة للخطأ من الانحدار الخطي من على ذ *خ ط ط* به المعادلة *ص* = *أ* + ب *س* :

$$s_{y|x}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - a - bx_i)^2,$$

ق ذ و [2] هو الفرق ذ :

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

علما أنه منذ العينة معامل الارتباط هو متماثل في *وذ خ ط ط* ، سنحصل على نفس القيمة لصالح من *لذ خ ط ط* :

$$r_{xy}^2 = 1 - \frac{s_{x|y}^2}{s_x^2}.$$

هذه المعادلة كما يعطي فكرة بديهية معامل الارتباط العالي الأبعاد. تماما كما سبق ووصف العينة معامل الارتباط هو جزء بسيط من الفرق استأثرت به صالح من 1 الأبعاد الخطية submanifold على مجموعة من الأبعاد ناقلات 2 *(خ ط ، ط ذ)* ، حتى نتمكن من وضع لمعامل الارتباط صالح *للم* الأبعاد الخطية submanifold لمجموعة من الأبعاد ناقلات *ن* . على سبيل المثال ، إذا كان لنا أن يصلح طائرة = *ض* + *أ* + ب × *قبرصي* لمجموعة من البيانات *(خ ط* ، *ط ذ* ، *ض ط)* ثم معامل الارتباط *لض س* و *ص* هو

$$r^2 = 1 - \frac{s_{z|xy}^2}{s_z^2}.$$

توزيع معامل الارتباط لقد درست جمهورية أرمينيا فيشر [2] [3] ، وحزب العدالة والتنمية Gayen. [4]

تفسير الهندسية

لتركز على البيانات (أي البيانات التي تم نقل عينات من قبل وذلك يعني أن يكون متوسط صفر) ، ومعامل الارتباط ويمكن أيضا أن ينظر إليها على أنها جيب التمام للزاوية بين البلدين ناقلات العينات المسحوبة من اثنين من المتغيرات العشوائية .

بعض الأطباء يفضلون uncentered (غير متوافقة مع بيرسون) ومعامل الارتباط. انظر المثال التالي للمقارنة.

وكمثال على ذلك ، لنفترض وتوجد خمس دول لديها من ناتجها القومي الإجمالي 1 و 2 و 3 و 5 و 8 مليارات دولار ، على التوالي. لنفترض أن نفس هذه الدول الخمس (في نفس النظام) وتوجد لديها بنسبة 11 ٪ ، 12 ٪ ، 13 ٪ ، 15 ٪ و 18 ٪ من الفقر. ثم ترك سين وصاد الأمر (5) التي تحتوي على العناصر الناقلة للبيانات المذكورة أعلاه : س = (1 ، 2 ، 3 ، 5 ، 8) ، ص = (0.11 ، 0.12 ، 0.13 ، 0.15 ، 0.18).

من خلال الإجراء المعتاد لإيجاد زاوية بين ناقلات (انظر النقطة المنتج) ، ومعامل الارتباط *uncentered* هو :

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{2.93}{\sqrt{103}\sqrt{0.0983}} = 0.920814711.$$

علما بأن البيانات الواردة أعلاه كانت تعمدت أن يرتبط تماما : *ص* = 0.10 + 0.01 *س*. معامل ارتباط بيرسون لذلك يجب أن يكون واحد بالضبط. تركز البيانات) عن طريق تحويل هاء س (خ) = 3.8 و ذ به هاء (ذ) = 0.138) غلة س = (-2.8 ، -1.8 ، -0.8 ، 1.2 ، 4.2) وذ = (-0.028 ، -0.018 ، -0.008 ، 0.012 ، 0.042) ، والتي كما كان متوقعا.

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{0.308}{\sqrt{30.8}\sqrt{0.00308}} = 1 = \rho_{xy},$$

الدافع للشكل معامل الارتباط

دافع آخر للارتباط من تفتيش طريقة بسيطة الانحدار الخطي. كما هو وارد أعلاه ، هو ناقل العاشر من المتغيرات المستقلة ، خ ط ، ص ، والتي تعتمد من المتغيرات ، ذ ط ، ومجرد علاقة خطية بين سين وصاد المطلوب ، وذلك من خلال طريقة المربعات الأقل على تقدير صاد :

$$Y = X\beta + \varepsilon.$$

ومن ثم ، فإن معادلة أقل الساحات خط يمكن استخلاص أن تكون على شكل :

$$(Y-\bar{Y})=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2}(X-\bar{X})$$

والتي يمكن إعادة ترتيبها في شكل :

$$(Y-\bar{Y})=\frac{r\,s_y}{s_x}(X-\bar{X})$$

(ص) حيث لديه دراية شكل المذكورة أعلاه :

$$\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-(\sum x_i)^2}\ \sqrt{n\sum y_i^2-(\sum y_i)^2}}.$$

تفسير حجم وجود ارتباط

ارتباط	السلبية	الإيجابية
صغير	-0.3 الى -0.1	0.1 الى 0.3
المتوسط	-0.5 الى -0.3	0.3 إلى 0.5
كبير	-1.0 الى -0.5	0.5 إلى 1.0

عدة مؤلفين وعرض المبادئ التوجيهية لتفسير وجود ترابط المعامل. كوهين (1988) ، [5] وقد لاحظ أن جميع هذه المعايير في بعض الطرق التعسفية ، ولا ينبغي أيضا لاحظ بدقة. وذلك لأن تفسير وجود ارتباط معامل يعتمد على السياق والمقاصد. وهناك ارتباط من 0.9 قد يكون منخفضا جدا إذا ما تحقق من القانون البدنية باستخدام أدوات عالية الجودة ، ولكن يمكن أن تعتبر عالية جدا في مجال العلوم الاجتماعية حيث يمكن أن تكون هناك مساهمة أكبر من عوامل التعقيد.

على هذا المنوال ، من المهم أن نتذكر أن "الكبير" و "صغيرة" وينبغي ألا يؤخذ مرادفات "جيد" و "السيئة" من حيث أن تحديد وجود ارتباط له حجم معين. على سبيل المثال ، وجود ارتباط أو -1.0 من 1.0 تشير إلى أن اثنين من تحليل المتغيرات يعادل حجم مودولو. علميا ، وهذا يدل على أكثر مما كان نتيجة

تافهة عميق. على سبيل المثال ، النظر في اكتشاف العلاقة بين 1.0 كم طوله قدم مجموعة من الناس وعدد من بوصة من أسفل القدمين إلى قمة الرأس.

الحساسية لتوزيع البيانات

السكان نسخة معامل ارتباط بيرسون المحدد للحظات ، ويحتوي على متغيرين وجود أي احتمال التوزيع التي السكان التغاير يعرف والهامشية السكان الفروق محددة وغير الصفر. في حالة ويحتوي على متغيرين من التوزيع الطبيعي ، ومعامل الارتباط المشتركة التي تميز توزيع ما دامت الفروق الهامشية وسيلة معروفة. يحتوي على متغيرين الأخرى بالنسبة لمعظم التوزيعات هذا ليس صحيحا. ومع ذلك ، فإن معامل ارتباط كبير من المعلومات عن درجة التبعية بين خطي عشوائي كميات بغض النظر عما إذا كان المشترك التوزيع الطبيعي.

فإن العينة معامل الارتباط هو أقصى تقدير احتمال من السكان معامل الارتباط للبيانات يحتوي على متغيرين طبيعية ، وبشكل مقارب محايدة وفعالة ، مما يعني أن ما يقرب من المستحيل بناء أكثر دقة بكثير من تقديرات العينة معامل الارتباط في حال طبيعية ، وبيانات حجم العينة كبيرة أو معتدلة. غير عادية للسكان ، ومعامل ارتباط العينة تقريبا لا تزال غير منحازة ، ولكن قد لا تكون فعالة. العينة معامل الارتباط هو تقديرات متسقة للسكان معامل الارتباط ما دام عينة الوسائل الفروق ، والتغاير تتسق (الذي يضمن فيه القانون أعداد كبيرة يمكن تطبيقها).

الاستدلال الإحصائي لمعامل الارتباط بيرسون هو حساسية لتوزيع البيانات. الفحوص الدقيقة ، والتجارب مقارب على أساس فيشر التحول يمكن تطبيقها إذا كانت البيانات التي عادة ما يقرب من توزيعها ، ولكنها قد تكون مضللة على خلاف ذلك. في بعض الحالات ، وألبس الحذاء يمكن تطبيقها لبناء الثقة ، وتبديل الاختبارات يمكن تطبيقها على إجراء التجارب الفرضية. هذه المنظمات غير بارامترية النهج قد يعطي نتائج أكثر في بعض الحالات التي تكون فيها الحياة الطبيعية يحتوي على متغيرين لا تعقد. بيد أن مستوى هذه النصوص تعتمد على نهج والتبادل. التحليل الطبقي هو أحد السبل لعدم توفر الظروف الطبيعية يحتوي على متغيرين بسبب التكتيل ، وتقييم أثر عوامل الخطر على النتيجة في حين عقد آخر متغير باستمرار. [6]

ارتباط التدابير غير بيرسون علاقة لها حساسية لتوزيع البيانات. السكان نسخ ارتباط التدابير القائمة على quantiles أو صفوف دائما محددة. على عينة من التقديرات على أساس أن تكون متسقة طالما الكامنة عينة quantiles تتسق.

ارتباط معظم التدابير عام هي ثابتة على استخدام الموقع وحجم التحولات الحدية التوزيعات. وهذا هو ، إذا ما أردنا تحليل العلاقة بين *سين وصاد* ، والربط تتأثر تحويل *العاشر لوب* × + ج + *ص ل* dY ، حيث *أ* ، *ب* ، *ج* ، *د* ، والثوابت هي. وهذا ينطبق على معظم ارتباط الإحصاءات ، فضلا عن السكان متناظرة.

غير بارامترية معاملات الارتباط

العينة معامل الارتباط ليست قوية ، بمعنى أنه يراعي المتطرفة في مجموعة البيانات. بارامترية عدم ارتباط المعاملات ، مثل شي مربع ، نقاط ارتباط مزدوج التسلسل ، رماح 'ρs ، قال كندال τ ، وغودمان وKruskal's امبدا قد تؤدي على نحو أفضل من العينة معامل الارتباط عندما المتطرفة موجودة. هذه الأساليب كثيرا ما تكون أقل دقة من العينة ارتباط في حال عدم وجود المتطرفة. بشكل عام نلاحظ أن هذه الإحصاءات غير بارامترية المتوقع قيم مختلفة عن بعضها البعض ، وكذلك من معامل ارتباط بيرسون ، وحتى لعينات كبيرة. لأنها تختلف التقديرات السكانية البرامترات ، بصفة عامة لا يمكن مقارنة مباشرة. وهي بصفة عامة ينبغي أن ينظر في اتخاذ تدابير بديلة للرابطة ، وليس كبديل تقدير من السكان معامل الارتباط.

تدابير أخرى من التبعية بين المتغيرات العشوائية

المعلومات التي قدمها أحد معامل الارتباط ليست كافية لتحديد هيكل التبعية بين المتغيرات العشوائية. معامل الارتباط تماما يحدد هيكل التبعية إلا في حالات خاصة جدا ، على سبيل المثال عندما التراكمي توزيع المهام هي توزيعات طبيعية متعددة. (انظر الرسم البياني أعلاه.) في حالة توزيع الاهليلجية انها سمة (المفرطة) الحذف من كثافة متساوية ، مع ذلك ، لا تميز تماما اعتماد الهيكل (على سبيل المثال ، أ ر المتعدد توزيع درجات الحرية في تحديد مستوى من الذيل الاعتماد).

للحصول على المزيد من تدابير لالعامة التابعة في البيانات (غير الخطية) ومن الأفضل استخدام ارتباط نسبة تكون قادرة على كشف أي ما يقرب من التبعية الوظيفية ، أو الكون القائم على تبادل المعلومات / مجموع الارتباط وهو قادر على كشف المزيد العامة التابعة لها. هذه الأخيرة هي التي يشار إليها أحيانا عدة تدابير متبادلة لحظة ، وبالمقارنة مع تلك التي تنظر في هذه اللحظة فقط pairwise2 nd) أو التربيعية) التبعية.

فإن polychoric علاقة ارتباط آخر يطبق على ترتيبي البيانات التي تهدف إلى تقدير الترابط بين المتغيرات theorised الكامنة.

إحدى الطرق للقبض على أكمل نظرا لاعتمادها على هيكل النظر رابط بينهما.

الارتباط مصفوفات

الارتباط مصفوفة المتغيرات العشوائية *ن* × 1 ،... ، *س ن ن* × هو النموذج الذي *ط ن* ، *ي* الدخول كور (*س ط* ، *ي س*). إذا كانت التدابير المستخدمة هي نتاج علاقة بين لحظة معاملات ، مصفوفة الارتباط هو نفسه فإن التغاير مصفوفة من المتغيرات العشوائية الموحد *العاشر* (*ط*) التنمية المستدامة (*س لط* = 1 ،... ، *ن*. وبالتالي فإنها ليست بالضرورة ايجابية بين نصف مؤكد المصفوفة.

الارتباط المصفوفة المتماثلة بسبب الترابط بين *العاشرة والعاشرة ط ي* هي نفس العلاقة بين *العاشرة والعاشرة ي ط*.

عام المفاهيم الخاطئة

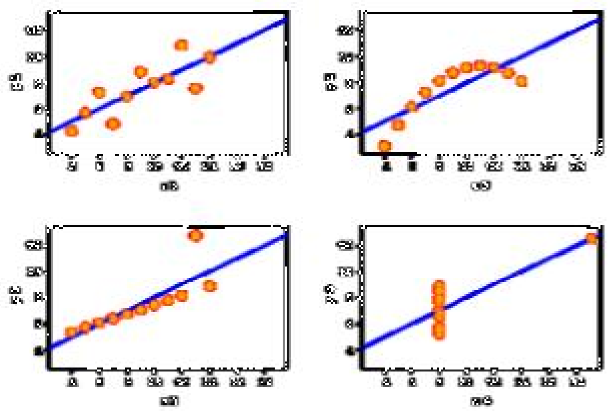
الارتباط والعلاقة السببية

المادة الرئيسية : الارتباط لا يعني السببية

التقليدي القائل بأن "لا ينطوي على علاقة السببية" يعني ان الارتباط لا يمكن استخدامها لاستنتاج علاقة سببية بين المتغيرات [7]. وهذا الرأي الذي لا ينبغي أن يؤخذ على أنه يعني أن العلاقات المتبادلة لا يمكن أن تشير إلى العلاقات السببية. ومع ذلك ، فإن الأسباب الكامنة والترابط ، إن وجدت ، قد تكون غير مباشرة وغير معروفة. ونتيجة لذلك ، إقامة علاقة بين اثنين من المتغيرات ليست شرطا كافيا لإقامة علاقة سببية (في أي من الاتجاهين.

وهناك علاقة بين طول العمر وطفل في حد سببية شفافة ، ولكن وجود ارتباط بين الحالة المزاجية والصحية للشعب في أقل من ذلك. هل تؤدي إلى تحسن المزاج وتحسين الصحة ، أو لا يؤدي إلى صحة جيدة مزاج جيد ، أو على حد سواء؟ أم أن بعض العوامل الأخرى وراء كل من؟ وبعبارة أخرى ، وجود ارتباط يمكن اتخاذها كدليل لاحتمال علاقة سببية ، ولكن لا يمكن الإشارة إلى العلاقة السببية ، إن وجدت ، قد يكون.

والارتباط الخطي



أربع مجموعات من البيانات ذات العلاقة 0.816

في حين أن ارتباط بيرسون تدل على قوة علاقة خطية بين المتغيرين ، وقيمته وحده قد لا يكون كافيا لتقييم هذه العلاقة : قد تكون هناك تفاعلات غير *الخطية* بين المتغيرات.

الصورة الموجودة على اليمين scatterplots من Anscombe 'الرباعية ، ومجموعة من أربعة أزواج من مختلف المتغيرات الناجمة عن فرانسيس Anscombe. [8] ذ المتغيرات الأربعة لها نفس متوسط (7.5) ، والانحراف المعياري (4.12) ، والارتباط (0.816) خط الانحدار (*ذ* = 3 + 0.5 *خ)*. ولكن ، وكما يمكن ملاحظته على هذه الاراضي ، وتوزيع للمتغيرات مختلفة للغاية. الأولى (أعلى اليسار) ، ويبدو أن توزع في العادة ، ويتطابق مع ما يتوقع احد عند النظر في اثنين من المتغيرات ارتباطا وبعد تولي طبيعتها. والثاني (أعلى يمين) لا توزع عادة ، بينما واضح في العلاقة بين متغيرين ويمكن ملاحظة ذلك ، ليست خطية ، ومعامل ارتباط بيرسون ليست ذات صلة. وفي الحالة الثالثة (أسفل اليسار) ، وعلاقة خطية مثالية ، باستثناء واحد الخارجة التي تمارس نفوذا كافيا لخفض معامل الارتباط في الفترة من 1 إلى 0.81. وأخيرا ، فإن المثال الرابع (أسفل اليمين) ويظهر مثال آخر عندما الخارجة تكفي لانتاج معامل الارتباط ، رغم أن العلاقة بينهما ليست خطية المتغيرات.

هذه الأمثلة تشير إلى أن معامل الارتباط ، وكأنها تختصر إحصائية ، لا يمكن أن تحل محل الفرد دراسة البيانات. نلاحظ أن الأمثلة التي تكون في بعض الأحيان ان لإثبات أن ارتباط بيرسون يفترض أن البيانات تتبع التوزيع العادي ، ولكن هذا ليس صحيحا. [9] ألف التوزيع العادي لا تحتوي إلا على خطي

العلاقات بين المتغيرات ، الأمر الذي يعني أن معامل الارتباط تماما يلخص قوة التفاعل في التوزيع العادي.

الحوسبة معامل الارتباط بيرسون

صيغة للارتباط بيرسون يتخذ أشكالا عديدة. وهناك صيغة شائعة الاستعمال هو مبين أدناه. الصيغة تبدو معقدة بعض الشيء ، ولكنها اتخذت خطوة خطوة كما هو مبين في الأمثلة العددية ، هي حقا بسيطة للغاية.

$$r=\frac{\sum XY-\frac{\sum X\sum Y}{N}}{\sqrt{\left(\sum X^2-\frac{(\sum X)^2}{N}\right)\left(\sum Y^2-\frac{(\sum Y)^2}{N}\right)}}$$

وهناك صيغة أبسط تبحث يمكن استخدامها إذا كانت الأرقام حولت عشرات ض :

$$r=\frac{\sum z_x z_y}{N}$$

حيث ض س س هو المتغير حولت عشرات ض و ض ذ هو متغير ص ض حولت العشرات.
الحوسبة بيرسون (ص) في البرنامج الإحصائي للعلوم الاجتماعية
ص الحوسبة بيرسون في اكسل

معامل ارتباط الرتب: (Conefficient Rank Correlation)

هذا المعامل يعرف بمعامل ارتباط سبيرمان (Spearman) أو معامل ارتباط الرتب (رتب القيم الأصلية وليس القيم) ولذا تختلف قيمته عن قيمة معامل بيرسون (للقيم الأصلية وليس لرتبها) وهو أقل دقة من معامل ارتباط بيرسون ويتعامل مع البيانات الرقمية وغير الرقمية للترتيب مثل جيد، جيد جدا،

... ويرمز له بالرمز r_s وهو ضمن الإحصاءات غير المعلمية ذات التوزيع الحر وقيمته موجبة أقل أو تساوي الواحد الصحيح وتحسب قيمته من الصيغة الرياضية علماً بأن:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

حيث d الفرق بين رتبه حسب المتغير الأول x ورتبه حسب المتغير الثاني y (الفرق بين رتب القيم لكل زوج من البيانات) وفي حالة التساوي يأخذ المتوسط الحسابي (فإذا كانت لقيمتين الرتبتين 7 ، 8 فيأخذ متوسط 7 ، 8 وتصبح الرتب لكل منها 7.5 بدل عن 7 ، 8) ، n عدد الأزواج للقيم فإذا كان لدينا مجموعة من الأفراد وجرى ترتيبهم حسب صفتين لكل فرد من المجموعة x ، y فإن $d_i = y_i - x_i$.

مثال:

تقدم عشرة طلاب لامتحان المرحلة الثانوية وكانت معدلات نتائجهم حسب الصف والمدرسة كالتالي والمطلوب حساب معامل سبيرمان للارتباط.

معدل الطالب في الصف (X)	73	80	70	66	89	71	65	88	92	74
مدل الطالب في المدرسة (Y)	69	78	66	70	92	64	55	90	88	72

الحل:

نكون جدول نبين فيه رتب كل من X (المعدل في الصف) و X (المعدل في المدرسة) والفرق d ومربع الفرق d^2 كالتالي:

X	Y	Rank X	Rank Y	d	d^2
73	69	6	7	– 1	1
80	78	4	4	0	0
70	66	8	8	0	0
66	70	9	6	3	9
89	92	2	1	1	1
71	64	7	9	– 2	4
65	55	10	10	0	0
88	90	3	2	1	1
92	88	1	3	– 2	4
74	72	5	5	0	0

بتطبيق القانون أعلاه:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$r_s = 1 - \frac{6 \times 20}{10(100-1)}$$

$$r_s = 1 - \frac{120}{990}$$

$$r_s = 1 - 0.121$$

$$r_s = 0.879$$

دلالة معامل الارتباط:

اختبار مدى المعنوية r_s (القيمة متوسطة وليست صفر أو ±1) وعندما تكون حجم العينة أكبر من وأقل من 30 (صغيرة) نقارنها مع المحسوبة من الجدول عند $\alpha/2$ وعندما تكون حجم العينة أكبر أو يساوي 30 فنوجد قيمة Z ونقارنها مع الجدولية حيث قيمة Z = قيمة معامل ارتباط الرتب مضروباً في الجذر التربيعي للعدد $n - 1$.

باعتبار أن المجتمع ذا البعدين X, Y والمأخوذ منه العينة من الأزواج المرتبة وبفرض أن ρ معامل ارتباط المجتمع فيكون r تقديراً للمعامل ρ. ولا بد من افتراض أن 0 = ρ لنحصل على اقتران احتمال(r) حسب النظرية:

إن جميع العينات ذات حجم n والممكنة مأخوذة من مجتمع ذي بعدين ويخضع للتوزيع المعتدل ومعامل ارتباطه $\rho = 0$ ، وأن r يعبر عن معاملات ارتباطات تلك العينات فإن:

$$t = \frac{r}{\sqrt{(1-r^2)(n-2)}}$$

يخضع لتوزيع t بدرجات حرية $n - 2$.

وفي حال ρ مجهولة فنأخذ بالنظرية التالية:

إذا أخذت عينات حجم كل منها n من مجتمع ذي بعدين وذي معامل ارتباط ρ وعرفنا الإحصاء Z كالتالي:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

فان Z يقرب من التوزيع الطبيعي ذي المعدل

$$\mu_z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}$$

والانحراف المعياري

$$\sigma_z = \frac{1}{\sqrt{n-3}}$$

تستخدم النظرية لإيجاد فترة ثقة ١٠٠% $(1-\alpha)$ لمعامل الارتباط ρ وذلك بإيجاد فترة ثقة لـ μ_z ومن ثم تحول لفترة ثقة لـ ρ من:

$$p(-z_{1-\alpha/2} \le \frac{Z-\mu_z}{\sigma_z} \le z_{1-\alpha/2}$$

ومنه نستنتج أن

$$Z - z_{1-\alpha/2}.\sigma_z \, , Z + z_{1-\alpha/2}.\sigma_z$$

وهي فترة الثقة $100\%(1-\alpha)$ لـــ μ_z ومن جدول تحويل r إلى Z نجد فترة الثقة المطلوبة ل (ρ)

ولنبين ذلك على مثالنا هنا:

لنختبر الفرضية $\rho = 0.8$ على مستوى معنوية 0.05 ومن ثم نحسب فترة ثقة 95% لمعامل الارتباط ρ.

الفرض $H_o : \rho \neq 0.8$ ، $H_o : \rho = 0.8$ حيث $\alpha = 0.05$

بالرجوع للجدول عند $\alpha = 0.05/2$ ، $n = 10$ نجد أن r_s الجدولية (r^*_s)

الحل باستخدام SPSS

مثال آخر: نفس المثال السابق مع البيانات التالية:

الحـــــل

معدل الطالب في الصف (X)	73	80	70	66	88	71	65	88	92	74
مدل الطالب في المدرسة (Y)	64	78	64	70	92	64	55	90	88	72

مثال آخر لحساب الدلالة وفترة الثقة

مثال ثالث (تقدير الأداء)

لقد درست في البند السابق معامل ارتباط بيرسون كمقياس لقوة الارتباط الخطي بين متغيرين، ورأيت أنه لحساب معامل ارتباط بيرسون لا بد من وجود بيانات رقمية للمتغيرين قيد الدراسة،ولكن هناك بعض الحالات التي يصعب فيها الحصول على قياسات رقمية؛ فإذا أردت التمييز بين ثلاثة أنواع من الجبن من حيث المذاق فإنه يصعب إعطاء بيانات رقمية، ولكن من السهل ترتيب الأنواع الثلاثة حسب مذاقها ولذلك يُراد مقياساً للارتباط يعتمد على رتب المتغيرات، وليس على القيم الرقمية لها. من أهم هذه المقاييس هو "معامل ارتباط سبيرمان."

قبل البدء بتعرف معامل ارتباط سبيرمان، يجب أن يذكر هنا أنه من الممكن استخدام معامل ارتباط سبيرمان حتى مع وجود بيانات رقمية. قبل إعطاء تعريف لمعامل ارتباط سبيرمان، تحتاج إلى تعريف بعض المصطلحات.

إذا كانت) س1، ص1)، (س2، ص2)،.......، (س ن، ص ن) مجموعة من الأزواج المرتبة، وإذا رتبت قيم المتغير س بحيث تعطي أصغر رقم فيها الرتبة (1)، والرقم الذي يليه الرتبة(2) ،وهكذا. وقمت بالعملية نفسها بالنسبة لقيم المتغير ص، واستخدمت الرمز فر ليدل على الفرق بين رتبة سر ورتبة صر.

أي أنف = 1رتبة س – 1رتبة ص1

ف = 2رتبة س – 2رتبة ص........2وهكذا.

فإنه يعرف معامل ارتباطسبيرمان كما يلي-:

رس – 1 =

6 فر 2

ن) ن (1 - 2

حيث أن: ف : رتب س - رتب ص (المتناظرة(

ف : 2مجموع مربعات فروق الرتب.

ن : عدد المشاهدات.

خواص معامل ارتباط بيرسون -:

1- ر[واذا كانت [ر] = 1 تكون العلاقة تامة اي النقاط على استقامةواحدة .
2- اذا كانت 0.35] > ر 1 > [فان العلاقة قوية.
3- اذا كانت 0] > ر 1 > [تكون العلاقة طردية غيرتامة.
4- اذا كانت] > 1- ر 0 > [فان العلاقة عكسية غير تامة.

كيفية ايجاد الرتب -:

1- نرتب الإعداد المعطاة اما تصاعديا او تنازليا] مثال : 3 ، 8 ، 12 ، 5 ، 12 ، 5 ، [12 نرتب هذه الاعداد تصاعديا مثلا فتصبح [(3 ، 5 ، 5 ، 8 ، 12 ن 12 ،[(12
2- يأخذ الرقم الذي لا يوجد له نظير الرتبة حسب موقعه في الترتيب.
3- اما الاعداد التي لها نظير مثلا 5 فيكون له رتبتان 2 و 3 معدل الرتبهو 2+3 / 2 = 2.5 ويعطى رقم 5 هذه الرتبة . ورقم 12 ياخذ الرتب 5 و 6 و 7 ويكونمعدل الرتب 5+6+7 / 63 =

تمرينات ومسائل

1- جد معامل ارتباط سبيرمان للبيانات التالية:

س	ص

10	2
19	18
8	4
32	18
4	5

2- جد معامل ارتباط سبيرمان للبيانات التالية:

س	**ص**
5	8
3	8
9	18
10	17
2	12
2	17
9	2
9	17

3- إذا كان معامل ارتباط سبيرمان بين س و ص يساوي 0.82 وكان س4 = *س-10 جد معامل ارتباط سبيرمان بين س* و ص,

حلول الأسئلة / السؤال الأول-:

س	**ص**
10	2
19	18
8	4
32	18
4	5

الترتيب تنازلي

رس (1 - 25) × 5 / 8.5 × 6 - 1 =

1- 17 / 40

2- 23 / 40 = 0.575

رتب س

س	**ص**	ف	ف2

3	5	-2	4
2	1.5	0.5	0.25
4	4	0	0
1	1.5	-0.5	0.25
5	3	2	4
			8.5

السؤال الثاني-:

س	ص
5	8
3	8
9	18
10	17
2	12
2	17
9	2
9	17

ترتيب تصاعدي

رس $= 1 - 6 \times 62 / 8 \times (64 - 1)$

$= 1 - 31/42$

$= 11 / 42 = 0.262$

س	ص	ف	ف2
4	2.5	1.5	2.25
3	2.5	0.5	0.25
6	8	-2	4
8	6	2	4
1.5	4	-2.5	6.25
1.5	6	-4.5	20.25
6	1	5	25
6	6	0	0
			62.0

معامل ارتباط سبيرمان

بمعنى لو طلب منك اوجد الدلاله الاحصائيه على طول تتجه الى قيمه **sig**.

وهي تفترض فرضين :

1- الفرض الصفري = وهذا الفرض الله يطولي بعماركم يقول (لايوجد دلالة احصائيه).

2- الفرض البديل = وحبيبنا الفرض البديل يقول أنه لا مالك لوا (يوجد دلالة احصائيه).

نعرف ان الفرض الصفري يعني عدم وجود علاقة والفرض البديل يعني وجود علاقة.

المرحله الثانيه وهو تطبيق الاختبار كالاتي :

ننظر الى قيمه **sig** وعندنا احتمالين أما أن يكون أكبر أو اصغر من **0.05**

الفرض الاول :

في حال كانت قيمة **sig** أصغر من **0.05** فإننا نرفض الفرض الصفري ونقبل البديل والفرض البديل قلنا سابقاً انه يعني انه العلاقة دالة احصائياً.

الفرض الثاني :

في حال كانت قيمة **sig** أكبر من **0.05** فإننا نقبل الفرض الصفري ونرفض البديل وهذا يعني ان العلاقة ليست دالة احصائياً.

ناخذ مثال :

في المثال التالي اجب ، هل العلاقة دالة احصائياً ولماذا ؟

بيرسون = **0.89** والدلالة = **0.006** (والدلاله هنا هي المقصود بها **sig** في جدول الـــ **spss**).

الحل :

الفرض الصفري = لايوجد دلالة

الفرض البديل = داله احصائياً

قيمة **sig**

0.006 وهي أصغر من **0.05**

إذا سوف نرفض الفرض الصفري ونقبل بالفرض البديل، وهذا يعني أن العلاقة داله احصائياً.

الجزء الثاني وهو فيه في جانبين فقط

أين تجد **sig** وكيف تحدد معامل الارتباط المناسب.

نلاحظ التالي :

نختار معامل الارتباط المناسب ننظر لراس الجدول البيانات نجد فئات المستوى التعليمي ويقابلها في الجهه الاخرى فئات الدخل.

مثال بيانات :

اسمي واسمي = (معامل كرايمر ، معامل فاي، معامل التوافق، معامل لامدا)

اسمي وترتيبي = (معامل الارتباط، معامل فيتا، معامل الثنائي للرتب).
ترتيبي وترتيبي = (معامل سبيرمان، معامل جاما، معامل كندال).
اسمي وكمي = (معامل الارتباط التسليلي، معامل ايتا، معامل ايبسلون).
ترتيبي وكمي = (معامل الارتباط المتسليل المتعدد).
كمي وكمي = (معامل بيرسون)

ولابد ان نعرف ان الكمي يحتوي نسبي وفئوي يعني رقم مثل (العمر) وفئات مثل **10- 20**.
ولازم ايضا يكون واضح الفرق بين الكمي والتريبي وقضيه المفاضله بينهم ... الخ.
وايضا لازم نفرق بين قضيه وجود علاقة سواء ايجابيه او سلبيه قوية .. الخ، وبين قضيه الدلاله الاحصائيه
اذا على هذا الاساس بتحدد المعامل الخاص بالارتباط لحل المساله مثلاً.
او لاخذ نتيجته مثل ماتشوفون في الجدول ثلاث معاملات على اي نتيجه سوف تعتمد بناء على ماذكرناه سابقاً.
وهنالك نقطه مهمة بما يتعلق بمعامل سبيرمان وجاما فهم خصائص كل معامل يجعلك اكثر قدره على تحديد المعامل المناسب مثل ما الذي ينفع في حاله العينه كبيره وفي حاله كانت اقل من .... لابد هنا من فهم الخصائص.
اما الدلالة الاحصائيه لو طلب توضيح هل العلاقة داله احصائياً نذهب فوراً للدائرة وهي **sig** ونشوف القيمة المقابله لمعامل بيرسون
الدلالة هنا كم ؟
0.000
وهي طبعا اصغر من **0.05**

الحواشي والمراجع

1. ^ رودجرز ، و Nicewander جي ، ومؤسسة التعاون (1988). "ثلاثة عشر طرق للنظر في معامل الارتباط". *الاميركي الإحصائيين* 42 : 59-66. دوى : 2685263/10.2307.
2. ^ فيشر لجمهورية أرمينيا (1915). "التردد توزيع قيم معامل الارتباط في عينات من عدد كبير من السكان غير مسمى". *Biometrika* 10 (4) : 507-521. دوى : biomet/10.4.507/10.1093.
3. ^ فيشر لجمهورية أرمينيا (1921). "وحول وجود خطأ محتمل معامل الارتباط استنتاجها من عينة صغيرة" (الشعبى). *Metron* 1 (4) : 3-32. http://hdl.handle.net/2440/15169. استرجاعها على 25-03-2009.
4. ^ Gayen ، وحزب العدالة والتنمية (1951). "تردد توزيع المنتج في معامل الارتباط لحظة عينات عشوائية من أي حجم الاستفادة من عوالم غير عادية". *Biometrika* 38 : 219-247. دوى : biomet/38.1-2.219/10.1093.
5. ^ كوهين ، J. (1988). *التحليل الإحصائي للسلطة العلوم السلوكية* (e2 éd.)
6. ^ Multivariable التحليلات دليل عملي للأطباء السريريون. الطبعة nd2. ميتشل كاتز حاء. جامعة كاليفورنيا في سان فرانسيسكو. ردمك - 13 : 9780521549851. ردمك - 10 : X052154985 وستقوم شعبة الإعلام : X052154985/10.2277
7. ^ الدريتش ، جون (1995). "واستطرد قائلا الإرتباطات الحقيقي في بيرسون وعيد ميلاد المسيح". *العلوم الإحصائية* 10 (4) : 364-376. http://www.jstor.org/stable/2246135.
8. ^ Anscombe ، فرانسيس جيه (1973) الرسوم البيانية في التحليل الإحصائي. *الإحصائيين الأمريكية* ، 27 ، 17-21.
9. ^ جى رودجرز ، ومؤسسة التعاون Nicewander. عشر طرق للنظر في معامل الارتباط. الأمريكية إحصائي ، 42 (1) :59 - 66 ، فبراير 1988.

وعلاوة على القراءة

- كوهين ، .J، P. كوهين ، والغرب ، والمنشورة في الجريدة الرسمية & ايكن ، ليرة سورية (2003). *التطبيقية المتعددة الانحدار / تحليل لارتباط العلوم السلوكية.* (rd3 الطبعة) هيلزدال نيوجيرسي : Erlbaum ورانس أسوشيتس.

Dissertation and Thesis Methods and Statistical Consulting (877-437-8622)

College Textbooks

Math Textbooks

Statistics Textbooks