معامل الارتباط

وعندما نقول **مقاييس العلاقة** نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر مثلا، كأن يكون الهدف معرفة هل هناك علاقة بين مستوى الإنتاجية وجودة المنتج في مصنع ما؟، أي هل كلما زادت الإنتاجية تقل جودة المنتج أو العكس .

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+١) وبين الارتباط السالب التام (–١) .

العلاقة الطردية بين المتغيرات: هو تعبير يشير إلى تزايد المتغيرين المستقل والتابع معا، فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجة تمثله هي (+١) .

العلاقة العكسية بين المتغيرات: هو تعبير يشير إلى تزايد في متغير يقابله تناقص في المتغير الآخر، فإذا كانت الإنتاجية منخفضة ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط سالب، وأعلى درجة تمثله هي (–١) .

إن معامل الارتباط التام الموجب (+١) يعنى التغير في اتجاه واحد في كلا الظاهرتين مع بقاء الأوضاع النسبية لوحدات الظاهرة ثابتة، سواء كان هذا التغير في اتجاه الزيادة (أي زيادة قيم الظاهرة الأولى تتبعها زيادة في قيم الظاهرة الأخرى)، أو في اتجاه النقص (أي نقص قيم الظاهرة الأولى يتبعها نقص في قيم الظاهرة الأخرى) .

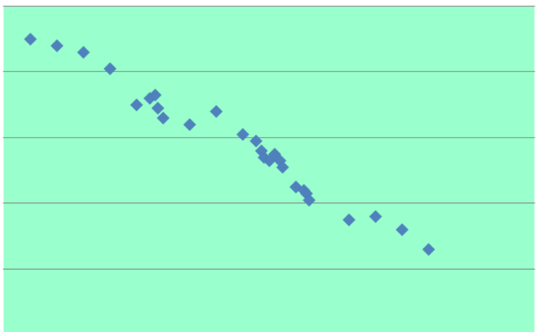
طرق التعرف على العلاقة بين متغيرين وحسابها

أولا: طريقة شكل الانتشار Scatter Diagram :

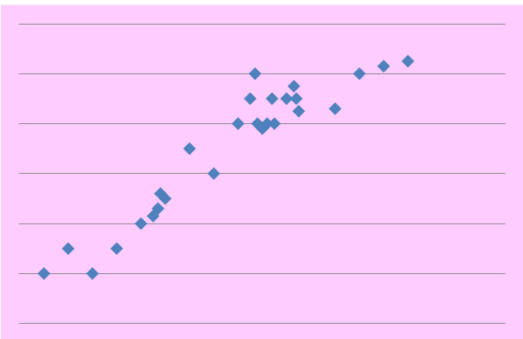
هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذه المحاضرة.

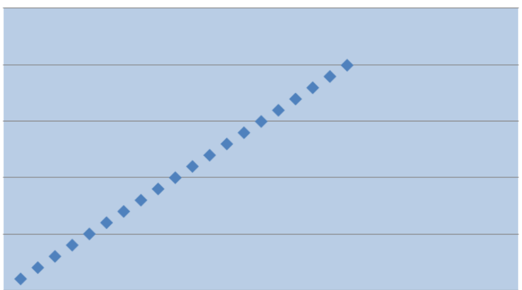
والمقصود **بشكل الانتشار** هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة. والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

ارتباط سالب

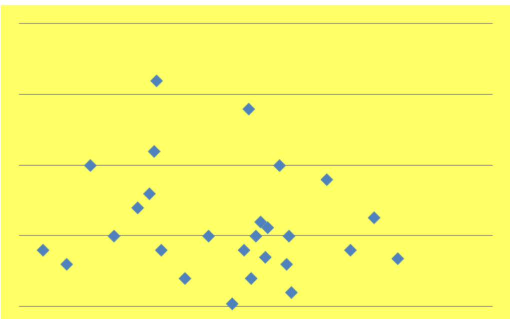


ارتباط موجب





ارتباط تام

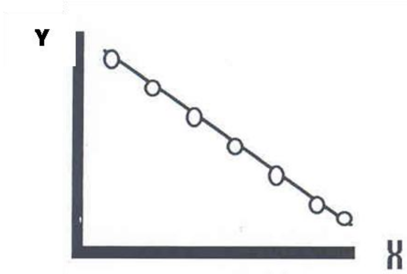


ارتباط صفري

الشكل الأول :

إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة. وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ). ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية. أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد فإن "الارتباط عكسي تام" كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.

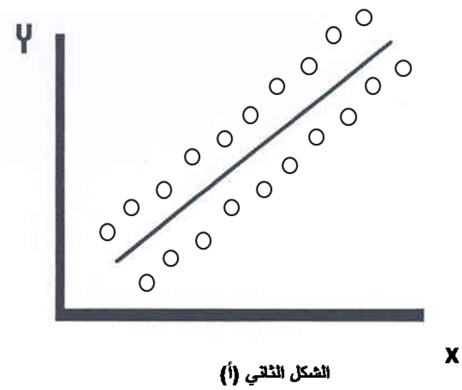
الشكل الأول (أ) ارتباط طردي تام (موجب)



الشكل الأول (ب) ارتباط عكسي تام (سالب)

الشكل الثاني :

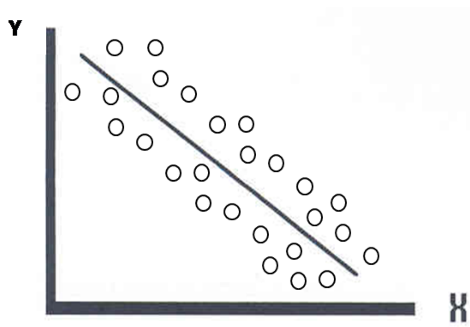
أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ، ب.



الشكل الثاني (أ)

ارتباط موجب قوي

(ارتباط خطي طردي)



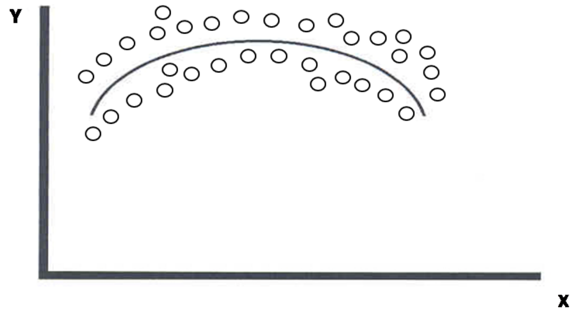
الشكل الثاني (ب)

ارتباط سالب قوي

(ارتباط خطي عكسي)

الشكل الثالث :

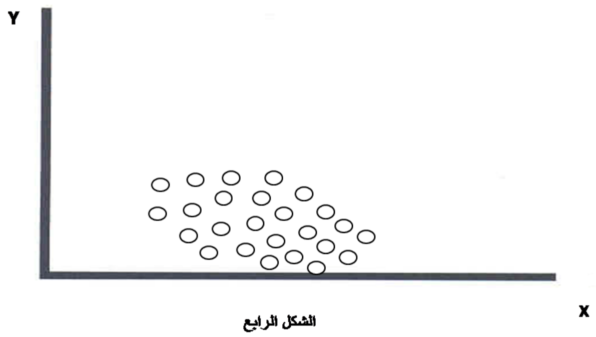
وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :



الشكل الثالث
(ارتباط غير خطي)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



الشكل الرابع
(لا توجد علاقة)

ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتبلط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + ١، – ١.

* فإذا كانت قيمة معامل الارتباط تساوي + ١ فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.

* وإذا كانت قيمة معامل الارتباط تساوي – ١ فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.

* وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.

* وكلما اقتربت قيمة معامل الارتباط من + ١ أو – ١ كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + ١ أو – ١ فإذا وصلت قيمة المعامل إلى + ١ أو – ١ كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + ١ ولا أصغر من – ١. ويمكن تمثيل قوة العلاقة بالشكل التالي:

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
من ٠.٧٠ إلى ٠.٩٩	ارتباط طردي قوي
من ٠.٥٠ إلى ٠.٦٩	ارتباط طردي متوسط
من ٠.٠١ إلى ٠.٤٩	ارتباط طردي ضعيف
0	لا يوجد ارتباط

يمكن تفسير الارتباط العكسي بنفس الطريقة مع المعاملات السالبة

معامل بيرسون للارتباط الخطى البسيط

Simple Correlation

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام (مثل الارتباط بين العمر والدخل) حيث يقاس العمر بالسنوات ويقاس الدخل بالعملة، بالريال أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي **<u>أربعين</u>** عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون **<u>خمسة آلاف</u>** ريال مثلاً.

وبالتالي فإن أفضل مقياس للارتباط بين مثل هذين المتغيرين – حسب رأي بيرسون – هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير. ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

ويتم حساب معامل ارتباط بيرسون من خلال العلاقــة التالية:

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

حيث :

تعني مجموع حاصل ضرب كل قيمة من X في Y.

تعني مجموع قيم المتغير X .

تعني مجموع قيم المتغير Y .

تعني مجموع مربع قيم المتغير X .

تعني مربع مجموع قيم المتغير X .

تعني مجموع مربع قيم المتغير Y .

تعني مربع مجموع قيم المتغير Y .

عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

مثال :

رغبة إحدى الشركات معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية لهم ، فقاموا بجمع معلومات عن هذا الموضوع وحصلوا على النتائج التالية :

الموظفين	أ	ب	ج	د	هـ	و	ز	ح	ط	ي
ساعات العمل X	٨	٢	٨	٥	١٥	١١	١٣	٦	٤	٦
مستوى الإنتاجية Y	٣	١	٦	٣	١٤	١٢	٩	٤	٤	٥

المطلوب: حساب معامل ارتباط بيرسون للبيانات السابقة .

الموظفين	ساعات العمل X	مستوى الإنتاجية Y	X^2	Y^2	XY
أ	٨	٣	٦٤	٩	٢٤
ب	٢	١	٤	١	٢
ج	٨	٦	٦٤	٣٦	٤٨
د	٥	٣	٢٥	٩	١٥
هـ	١٥	١٤	٢٢٥	١٩٦	٢١٠
و	١١	١٢	١٢١	١٤٤	١٣٢
ز	١٣	٩	١٦٩	٨١	١١٧
ح	٦	٤	٣٦	١٦	٢٤
ط	٤	٤	١٦	١٦	١٦
ي	٦	٥	٣٦	٢٥	٣٠
المجموع	٧٨	٦١	٧٦٠	٥٣٣	٦١٨

نطبق معادلة معامل ارتباط بيرسون كالتالي :

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

$$r = \frac{618 - \frac{(78)(61)}{10}}{\sqrt{\left(760 - \frac{(78)^2}{10}\right)\left((533) - \frac{(61)^2}{10}\right)}} = \frac{618 - 475.8}{\sqrt{(760 - 608.4)(533 - 372.1)}}$$

$$= \frac{1422}{\sqrt{(151.6)(160.9)}} = \frac{1422}{\sqrt{2439244}} = \frac{1422}{1562} = 0.91$$

وهذه النتيجة توضح أن درجة الارتباط = ٠.٩١ وهذه النتيجة تعتبر مؤشر على علاقة إيجابية قوية بين ساعات العمل ومستوى الإنتاجية.

ملاحظة مهمة :

من خواص معامل بيرسون للارتباط الخطي أنه لا يتأثر بالعمليات الحسابية التي تجري على المتغيرين x , y. بمعنى أنه لا يتأثر بالطرح (أو الجمع)، ولا بالقسمة (أو الضرب). أي إذا طرحنا (أو جمعنا) **قيمة معينة** من كل قيم x وقيمة أخرى من كل قيم y، أو قسمنا (أو ضربنا) قيم x على قيمة معينة وكل قيم y على قيمة أخرى فإن قيمة معامل الارتباط لا تتغير أي نحصل على القيمة نفسها.

معامل ارتباط الرتب

Rank Correlation

قد يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، فقد يكون المتغيران **وصفيين ترتيبين** Ordinal **أو أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً ترتيبياً**، أو أن يكون **المتغيران كميين**، ويكون اهتمام الباحث منصباً على الرتب أكثر من القيم. ففي انتخابات مجلس الشيوخ أو النواب الأمريكي مثلاً، يعتبر المرشح الأول هو من حصل على أعلى الأصوات بغض النظر عن عددها، والذي يحصل على عدد أصوات أقل منه مباشرة هو الثاني.. وهكذا.

فإذا كانت رتب المتغيرين تسير في الاتجاه نفسه؛ بمعنى أن الرتب الأعلى للمتغير الأول تناظرها رتب أعلى للمتغير الثاني كانت العلاقة طردية بينهما. وإذا كانت الرتب الأعلى للمتغير الأول تناظرها رتب أدنى للمتغير الثاني كانت العلاقة بينهما عكسية. ففي مثالنا السابق عن العلاقة بين دخل الناخب وعمره، كان الناخب الأكبر عمراً (بصفة عامة) هو الأعلى دخلاً، فمن الواضح أن العلاقة بينهما طردية، أما إذا كان الناخب الأكبر عمراً (بصفة عامة) هو الأقل مشاركة في العمل السياسي فإننا في هذه الحالة نكون أمام علاقة عكسية.

ولحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

حساب معامل سبيرمان لارتباط الرتب

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم ١، والقيمة الأعلى منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم ١، والقيمة الأقل منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز d) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

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

حيث : $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين

n هي عدد أزواج القيم

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي:

١ – مجموع الفروق بين الرتب يساوي صفر.

٢ – أن قيمة معامل ارتباط الرتب تنحصر بين – ١، + ١ فإذا كانت الرتبة رقم ١ للمتغير الأول تناظرها الرتبة ١ للمتغير الثاني، والرتبة ٢ للمتغير الأول تناظرها الرتبة رقم ٢ للمتغير الثاني، وهكذا.. فإن معامل ارتباط الرتب يساوي + ١ (ارتباط طردي تام بين الرتب).

وإذا كانت الرتبة رقم ١ (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – ١ (ارتباط عكسي تام بين الرتب).

٣ – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي

مثال:

البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي، ومدى ملاءمتها لحاجات الناس.

السؤال الأول	جيدة	مقبولة	ممتازة	جيدة	جيدة جداً	مقبولة	جيدة
السؤال الثاني	جيدة جداً	مقبولة	جيدة جداً	جيدة	جيدة	جيدة	ممتازة

والمطلوب: حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

السؤال الأول X	السؤال الثاني Y	رتب X	رتب Y	d الفرق بين الرتب	d^2 مربعات الفرق
جيدة	جيدة جداً	4	2.5	1.5	2.25
مقبولة	مقبولة	6.5	7	- 0.5	0.25
ممتازة	جيدة جداً	1	2.5	- 1.5	2.25
جيدة	جيدة	4	5	- 1.0	1.00
جيدة جداً	جيدة	2	5	- 3.0	9.00
مقبولة	جيدة	6.5	5	1.5	2.25
جيدة	ممتازة	4	1	3.0	9.00
المجموع				**Zero**	**26.0**

$$r_s = 1 - \frac{6\ (\sum d^2)}{n\ (n^2 - 1)} = 1 - \frac{6\ (26)}{7\ (49 - 1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

وهذا يعني أن الارتباط بين إجابات المبحوثين بالنسبة للسؤالين هو ارتباط طردي متوسط. وبالتالي فليس بالضرورة أن يكون رأي المجيبين في برامج الضمان الاجتماعي تعني ملاءمتها لحاجات الناس.

لدراسة العلاقة بين تقدير الطالبة في الإحصاء و تقديرها في الرياضيات ، اخترنا خمس طالبات و كانت تقديراتهم كالتالي:

تقدير الإحصاء X	A	C	D	F	A
تقدير الرياضيات y	B	C	B	D	A

رتب y	رتب x	y	x	d	d^2
A	B	5	4	1	1
C	C	3	3	0	0
D	B	2	4	-2	4
F	D	1	2	-1	1
A	A	5	5	0	0
∑					6

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

$$r_S = 1 - \frac{6 \times 6}{5(5^2-1)} = 1 - \frac{36}{5 \times 24} = 1 - \frac{36}{120}$$
$$= 1 - 0.3 = 0.7$$

معامل بوينت بايسيريال Point Biserial للارتباط

يستخدم لقياس الارتباط بين متغير كمي X و متغير اسمي Y مستويين (نعم –لا) أو (ذكر – أنثى) و غيرها.
اشارة معامل الارتباط ليس لها معنى في حالة المتغيرات الوصفية فتقاس قوة العلاقة و ليس اتجاهها.

أوجد قيمة معامل الارتباط بين مشاركة الطالبة في المحاضرة و درجتها في الاختبار للبيانات التالية :

المشاركة y	نعم	نعم	نعم	لا	لا
X درجة الاختبار	15	19	20	15	11

$$n_1 = 3$$

$$\bar{x}_1 = \frac{15+19+20}{3} = \frac{54}{3} = 18$$

$$n_2 = 2$$

$$\bar{x}_2 = \frac{15+11}{2} = \frac{26}{2} = 13$$

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{s_x}\sqrt{\frac{n_1 n_2}{n(n-1)}}$$

$$r_{pb} = \frac{18-13}{13}\sqrt{\frac{3\times 2}{5\times 4}} = \frac{5}{13}\sqrt{\frac{6}{20}} \approx 0.21$$

أوجد قيمة معامل الارتباط بين الإجابة على السؤال الإجباري y و الدرجة الإجمالية y لستة من الطلاب، حيث ١ تعني الإجابة على السؤال و ٠ تعني عدم الإجابة.

Y	1	1	1	0	0	0
X	14	16	19	11	7	8

3 $8 \over \quad = 8.67$	3 $\frac{19}{\quad} = 16.3$	$n_2 =$ $\bar{x}_2 = \frac{11+7+}{3}$	$S_x = 4.68$	$n_1 = 3$ $\bar{x}_1 = \frac{14+16+}{3}$

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{s_x}\sqrt{\frac{n_1 n_2}{n(n-1)}}$$

$$r_{pb} = \frac{16.3-8.67}{4.68}\sqrt{\frac{3\times3}{6\times5}} = 1.63 \times \sqrt{\frac{9}{30}} \approx 0.893$$

معامل الاقتران (معامل فاي) Phi

- يستخدم للعلاقة بين متغيرين اسميين كل منهما ثنائي التقسيم.
- اشارة معامل فاي ليس لها معنى فهو يقيس قوة العلاقة دون اتجاهها.

	X_1	X_2	Sum
Y_1	a	b	a+b
Y_2	c	d	C+d
Sum	a+c	b+d	a+b+c+d

$$r_{\emptyset} = \frac{a \times d \ - b \times C}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

أوجدي قيمة معامل الاقتران بين النوع (ذكر/ أنثى) و بين الاصابة بمرض الاكتئاب (مصاب/ غير مصاب) للبيانات التالية:

	lwhf	غير مصاب	المجموع
ذكر	12	7	19
أنثى	10	5	15
المجموع	22	12	36

$$r_{\emptyset} = \frac{12 \times 5 \ - 7 \times 10}{\sqrt{22 \times 12 \times 19 \times 15}} =$$

$$= \frac{60 - 70}{\sqrt{75240}} = \frac{-10}{274.299} = -0.037$$

معاملات الارتباط

كلا المتغيرين كمي	معامل بيرسون
متغير كمي+ متغير اسمي في مستويين	معامل بوينت بايسيريال
متغيرين كميين أو متغيرين وصفيين ترتيبين	معامل سبيرمان
متغيرين اسميين في مستوين	معامل فاي