معامل الارتباط:

:Pearson Linear Correlation coefficient

معامل الارتباط الخطي لبيرسون و يرمز له بـ R هو عبارة عن مقياس لقوة العلاقة الخطية بين متغيرين ، و هو يعكس مدى تماسك التأثير الناتج عن التغير في قيم المتغير x على التغير في قيم المتغير y و قيمة معامل الارتباط الخطي تكون دائماً بين 1- و 1+ فقيمته الموجبة دلالة على أن العلاقة الخطية بين x و y طردية أي تزايد قيم الأول يؤدي لتزايد قيم الثاني (إذا كانت R قريبة من الواحد : فالعلاقة قوية ، وإذا كانت قريبة من $\frac{1}{2}$ فتكون متوسطة ، وإذا كانت قريبة من الصفر فهي ضعيفة). الشكل(6−3) .

أما إذا كانت قيمته سالبة R<0 فإن العلاقة بين المتغيرين تكون سالبة أو عكسية. الشكل (6−4). فمثلاً نتوقع أن تكون R لمشاهدات x (الطول) مع مشاهدات y (الوزن) لمجموعة من الأشخاص موجبة أو طردية كما أننا نتوقع أن

تكون R لمشاهدات y (سعر السيارة) مع x (عمر السيارة) قوية سالبة أو عكسية أي ينقص سعر السيارة مع زيادة عمرها .

و لمعامل ارتباط بيرسون عدة صيغ متكافئة فإذا فرضنا أزواج المشاهدات لمتغيرين x,y هي :

$$(x_1 , y_1),\ (x_2 , y_2)\ , ... ,\ (x_n , y_n)$$

فنعرف معامل الارتباط لبيرسون R بأنه متوسط جداءات القيم المعيارية للمتغيرين x,y أي:

$$R = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{S_x}\right).\left(\frac{y_i-\bar{y}}{S_y}\right) \qquad (1-6)$$

حيث S_y, S_x القيم المعيارية لمشاهدات y,x على الترتيب

$$S_x{}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i{}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)$$

$$S_y{}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i{}^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)$$

وللعلاقة السابقة عدة صيغ متكافئة يمكن استخدامها في حساب معامل الارتباط منها :

$$R = \frac{\sum_{i=1}^{n} x_i . y_i - n.\bar{x}.\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i{}^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} y_i{}^2 - n.\bar{y}^2}} \qquad 2-6$$

وتكتب أيضا بالصيغة الآتية:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad 3-6$$

مثال (6−2)

لدراسة العلاقة بين الكمية الكبرى للأكسجين المستنشق X ومعدل ضربات القلب Y ومن أجل عشرة أشخاص أخذنا أول عشر أزواج من القيم في الجدول (6−1) وجدنا من الشكل (6−1) أنه هناك علاقة إيجابية قوية بالاعتماد على الصيغة (6−3) لعلاقة الارتباط نلخص عملية إيجاد معامل الارتباط في الجدول الآتي :

الجدول (6−2)

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
43	175	9.7	36.4	353.08	94.09	1324.96
49	180	15.7	41.4	649.98	246.49	1713.96
50	186	16.7	47.4	791.58	278.89	2246.76
12	95	-12.3	-43.6	928.68	453.69	1900.96
8	75	-25.3	-63.6	1609.08	640.09	4044.96
32	165	-1.3	26.4	-34.32	1.69	696.96
51	190	17.7	51.4	904.47	313.29	2641.96
30	95	-3.3	-43.6	143.88	10.89	1900.96
35	130	1.7	-8.6	-14.62	2.89	73.96
23	95	-10.3	-43.6	449.08	106.09	1900.96

بجمع قيم الاعمدة نجد :

$$\sum_{i=1}^{n} x_i = 333 \quad , \quad \sum_{i=1}^{n} y_i = 1386$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 \quad , \quad \sum_{i=1}^{n} (y_i - \bar{y}) = 0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 5786.2$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 2148.1 \quad , \quad \sum_{i=1}^{n} (y_i - \bar{y})^2 = 18446.4$$

و منه :

$$\bar{x} = \frac{333}{10} = 33.3 \quad , \bar{y} = \frac{1386}{10} = 138.6$$

بالتبديل في الصيغة (6-3)

$$R = \frac{5786.2}{\sqrt{2148.1}\sqrt{18446.4}} = \frac{5786.2}{6301.1} = 0.92$$

ملاحظة :

إن استخدام الصيغة (6-2) أسهل في التطبيقات العملية . و لنقوم بحساب R مرة ثانية و باستخدام الجدول المساعد :

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
43	175	7525	1849	30625
49	180	8820	2401	32400
50	186	9300	2500	34596
12	95	1140	144	9025
8	75	600	46	5625
32	165	5280	1024	27225
51	190	9690	2601	36100
30	95	2850	900	9025
35	130	4550	1225	16900
23	95	2185	520	9025
$\sum_{i=1}^{n} x_i = 333$	$\sum_{i=1}^{n} y_i = 1386$	$\sum_{i=1}^{n} x_i \cdot y_i = 51940$	$\sum_{i=1}^{n} x_i^2 = 13237$	$\sum_{i=1}^{n} y_i^2 = 201546$

الجدول (6-3)

$$R = \frac{51940-10\,(33.3)(138.6)}{\sqrt{13237-10\,(33.3)^2}\sqrt{210546-10\,(138.6)^2}}$$

$$= \frac{5786.2}{\sqrt{2148.2}\,\sqrt{18446.4}} = 0.92$$

تدل الإشارة الموجبة والقيمة القريبة من الواحد على أن علاقة الارتباط بين كمية الأكسجين ومعدل ضربات القلب طردية وقوية أي زيادة كمية الأكسجين المستنشق يؤدي لزيادة ضربات القلب في الواقع العلاقة هنا ليست سببية، ويجب التركيز أن الارتباط لا يعني السببية.

اختبار الفروض حول معامل الارتباط الخطي :

:Testing of Hypotheses for Linear Correlation Coefficient

بعد حساب معامل الارتباط الخطي للعينة المعطاة نطرح السؤال الآتي هل قيمة R المحسوبة من الصيغ السابقة تدل على أن هناك علاقة بين المتغيرين في المجتمع الذي سحبت منه العينة ، بمعنى آخر ما هي القيمة التي إذا كانت قيمة R أكبر منها يكون هناك علاقة ارتباط وإذا كانت قيمة R أصغر منها تكون العلاقة ضعيفة ، ومن ثمّ لا يوجد ارتباط خطي بين قيم X وقيم Y للإجابة عن هذا السؤال نقوم باختبار فرض العدم H_0 كما يأتي:

المتغيران غير مرتبطين خطياً H_0 ضد الفرض البديل H_1 و هو ذو طرفين كمايلي (المتغيران مرتبطان خطياً) : H_1.

في الحقيقة الفرض البديل H_1 يمكن أن يكون على الصيغة

المتغيران مرتبطان خطياً بشكل موجب : H_1

أوالمتغيران مرتبطان خطياً بشكل سالب : H_1

فإذا استخدمنا الرمز ρ للدلالة على معامل الارتباط الخطي للمجتمع ، فنكتب

فرضية العدم $H_0 : \rho = 0$

مقابل الفرض البديل ذي طرفين $H_1 : \rho \neq 0$

أو مقابل الفرض البديل بطرف واحد إمّا $H_1 : \rho > 0$

وإمّا $H_1 : \rho < 0$

لاختبار الفروض السابقة H_0 , H_1 لا بد من الأمور الآتية:

1- حساب الاحصاء t_0 لمعامل الارتباط R للعينة حيث t_0 تحسب تحت صحة فرضية العدم H_0 كما يأتي :

$$t_0 = \frac{R-\rho}{S_R} = \frac{R-0}{\sqrt{\frac{1-R^2}{n-2}}} = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \qquad (4-6)$$

و لهذا الإحصاء توزيع ستودنت بدرجات حرية تساوي $\gamma = n-2$ حيث $S_R = \sqrt{\frac{1-R^2}{n-2}}$ هو الخطأ المعياري عند اعتبار R مقدر لـ ρ .

2- تحديد مستوى المعنوية α . قد تكون 0.01 , 0.02 , 0.05 ,....حسب أهمية البحث.

3- نستخدم جدول توزيع ستودنت بدرجة حرية $\gamma = n-2$ لتعيين القيمة الحرجة $t_{1-\alpha/2}(\gamma)$ التي تحصر على يسارها و تحت منحني الكثافة لستودنت بدرجة γ مساحة مقدارها $1-\alpha/2$.

4- ثم نتخذ القرار المناسب وفق الآتي:

a) إذا كانت القيمة المطلقة للقيمة المحسوبة t_0 أكبر من القيمة الحرجة $(|t_0| > t_{1-\alpha/2}(\gamma))$ نرفض الفرضية H_0 و نقبل H_1 ذات الطرفين ويكون الارتباط معنوياً.

b) إذا كانت القيمة المطلقة للقيمة المحسوبة t_0 أصغر من القيمة الحرجة $(|t_0| < t_{1-\alpha/2}(\gamma))$ نقبل الفرضية H_0 ، و نعتبر أن قيمة R ليست معنوية ، أي ليست هناك علاقة ارتباط بين المتغيرين.

مثال (6-3)

بالعودة للمثال السابق (6-2) حيث حسبنا معامل ارتباط بيرسون للعينة ، و كان $R = 0.92$ وهو قيمة تقديرية لـ ρ معامل ارتباط x و y نقوم باختبار معنوية معامل الارتباط R عند مستوى معنوية $\alpha = 0.05$ مثلا.

الحل:

1- نصيغ الفرض الإحصائي

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

2- نحسب قيمة t_0 من العلاقة (6-4) و تحت فرضية H_0

$$t_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{0.92\sqrt{10-2}}{\sqrt{1-(0.92)^2}} \cong \frac{2.6}{\sqrt{0.15}} \cong 6.7$$

3- نعين منطقة قبول H_0 و هي المجال المعطى بالشكل:

$$\left[-t_{1-\alpha/2}(n-2), t_{1-\alpha/2}(n-2)\right] = [-t_{0.025}(8), t_{0.975}(8)]$$

و القيمة $t_{0.975}(8) = 0.23$ نجدها في سطر 8 و عمود 0.025 في جدول توزيع ستودنت.

إذن $t_0 = 6.7$ لا تقع في المجال $[-2.306\,, 2.306\,]$ أو $|t_0| > 2.306$ أي t_0 المحسوبة تقع في منطقة رفض H_0 ، ومن ثَمَّ الفرضية $\rho = 0$ غير صحيحة، و ذلك بدرجة ثقة أكبر من 95% ، و هذا يعني أننا على ثقة مقدارها 95% بأن هناك علاقة خطية بين الكمية الكبرى للأكسجين المستنشق و معدل ضربات القلب .

6−3 معامل سبيرمان لارتباط الرتب

Spearman's Rank Correlation Coefficient :

إن معامل الارتباط الخطي لبيرسون الذي سبق ذكره يمكن أن يستخدم لقياس الارتباط الخطي بين متغيرين كميين ، عندما تكون ملاحظات كُلٍّ من x و هي (x_i) و y وهي (y_i) مقادير كمية. فلا يمكن استخدامه إذا كانت المشاهدات ليست كمية (اسمية أو رتبية) ، لذلك دعت الحاجة لإيجاد مقياس للارتباط في حال كانت قيم أحد المتغيرين أو كليهما غير كمية. و من هذه المقاييس معامل سبيرمان لارتباط الرتب الذي يمكن استخدامه لقياس الارتباط بين المتغيرات التي يمكن ترتيب قيمها أي إذا كانت المتغيرات رتبية أو فئوية .

ليكن لدينا مجموعة مكونة من n من الأزواج المرتبة لمشاهدات العينة

$$(x_1\,, y_1)\,, (x_2\,, y_2)\ \ , \dots \dots \dots\,, (x_n\,, y_n\,)$$

و لنرمز بـ $r\,(\,x_i\,)$ لرتبة المشاهدة x_i في العينة $x_1\,, x_2\,, \dots, x_n$ (أي ترتيبها في العينة بعد ترتيب العينة تصاعديا) وبـ $r\,(\,y_i\,)$ لرتبة المشاهدة y_i في العينة $y_1\,, y_2\,, \dots, y_n$.

يتم استخدام رتب قيم المتغير x و رتب قيم المتغير y بشكل تصاعدي أو تنازلي معاً بشرط تطبيق نفس الطريقة لكلا المتغيرين. و في حال تساوي مشاهدتين $x_i = x_{i+1}$ نعطي لكل منها نفس الرتبة و تساوي متوسط رتبتي القيمتين .

وبعد استخراج رتب قيم المتغيرات نحسب معامل سبيرمان لارتباط الرتب كما يأتي:

a) معامل سبيرمان لارتباط الرتب هو معامل الارتباط الخطي لبيرسون لرتب المتغيرين أي معامل سبيرمان لارتباط الرتب يعطى من العلاقات المتكافئة الآتية:

$$r_s = \frac{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(w_i - \bar{w})}{\sqrt{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}\sqrt{\sum_{i=1}^{n}(w_i - \bar{w})^2}} \qquad (5-6)$$

$$r_s = \frac{\sum_{i=1}^{n}\gamma_i w_i - n\bar{\gamma}\bar{w}}{\sqrt{\sum_{i=1}^{n}\gamma_i^2 - n\,(\bar{\gamma})^2}\sqrt{\sum_{i=1}^{n}w_i^2 - n\,(\bar{w})^2}} \qquad (6-6)$$

$$r_s = \frac{n\sum_{i=1}^{n}\gamma_i w_i - \sum_{i=1}^{n}\gamma_i \sum_{i=1}^{n}w_i}{\sqrt{n\sum_{i=1}^{n}\gamma_i^2 - (\sum_{i=1}^{n}\gamma_i)^2}\sqrt{n\sum_{i=1}^{n}w_i^2 - (\sum_{i=1}^{n}w_i)^2}} \qquad (7-6)$$

حيث

$$\bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i \text{ ، } \bar{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i \text{ ، } w_i = r\,(y_i) \text{ ،} \gamma_i = r\,(x_i)$$

b) اذا لم يكن هناك تشابه بين رتب قيم كُلٍ من المتغيرين x وy فإننا نحسب معامل سبيرمان لارتباط الرتب من الصيغة البسيطة المكافئة الآتية :

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (8-6)$$

حيث: $d_i = r\,(x_i) - r\,(y_i)$ هو الفرق بين رتبتي x_i و y_i. و يجوز استخدام العلاقة البسيطة هذه و الحصول على قيمة تقريبية في حال وجود عدد

قليل من الرتب المتساوية . أما إذا كانت عدد الرتب المتساوية في أحد المتغيرين أو كليهما كبيراً فيجب استخدام العلاقة الأولى فقط .

مثال (6-4) :

لدراسة العلاقة بين كمية التدخين x والمقيسة بمتوسط عدد السجائر اليومية و شدة الإصابة بسرطان الرئة y أخذنا عينة عشوائية من عشرة أشخاص من المدخنين الذين أصيبوا بمرض سرطان الرئة ، و سجلنا مشاهدات كل من المتغيرين x وy في الجدول الآتي :

x	5	10	15	15	20	25	30	30	30	35
y	A	B	A	B	C	E	D	C	E	E

حيث مشاهدات المتغير Y هي :

A = خفيفة جداً ، B= خفيفة ، C = إصابة متوسطة ، D = إصابة شديدة ،

E = شديدة جداً .

أوجد معامل ارتباط الرتب لسبيرمان .

الحل : أولاً نوجد رتب قيم كل المتغيرات

x_i	5	10	15	15	20	25	30	30	30	35
$r(x_i)$	1	2	3.5	3.5	5	6	8	8	8	10

y_i	A	A	B	B	C	C	D	E	E	E
$r(y_i)$	1.5	1.5	3.5	3.5	5.5	5.5	7	9	9	9

نلخص العمليات الحسابية في الجدول (6-3) لحساب r_s من الصيغة البسيطة (6-8):

x	Y	رتبة x $r(x) = \gamma$	رتبة y $r(y) = w$	$d = r(x) - r(y)$	d^2
5	A	1	1.5	- 0.5	0.25
10	B	2	3.5	- 1.5	2.25
15	A	3.5	1.5	2	4
15	B	3.5	3.5	0	0
20	C	5	5.5	- 0.5	0.25
25	E	6	9	- 3	9
30	D	8	7	1	1
30	C	8	5.5	2.5	6.25
30	E	8	9	-1	1
35	E	10	9	1	1
					$\sum_{i=1}^{10} d_i^{\ 2} = 25$

الجدول (6-3)

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^{\ 2}}{n(n^2-1)} = 1 - \frac{6\,(25)}{10\,(100-1)} = 1 - \frac{150}{990} = +0.848$$

نحسب معامل ارتباط الرتب لسبيرمان باستخدام الصيغة (6-5) أي نطبق معامل الارتباط الخطي لبيرسون على بيانات الرتب في العمودين الثالث و الرابع من الجدول(6-2) السابق :

$$r_s = \frac{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(w_i - \bar{w})}{\sqrt{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}\sqrt{\sum_{i=1}^{n}(w_i - \bar{w})^2}} = \frac{67}{\sqrt{80}\,\sqrt{79}} = +0.843$$

بمقارنة النتيجتين نلاحظ هناك فرقاً (بسيطاً) و الاختلاف بين النتيجتين يُعزى لوجود عدد كبير من الرتب المتشابهة ولاسيّما رتب قيم المتغير y و من قيمة معامل الارتباط هذه $r_s = 0.843$ يتضح أن هناك علاقة طردية قوية نوعاً ما بين المتغيرين x و y ، و تلك العلاقة تعني أن شدة الإصابة بسرطان الرئة تعود لزيادة كمية التدخين.