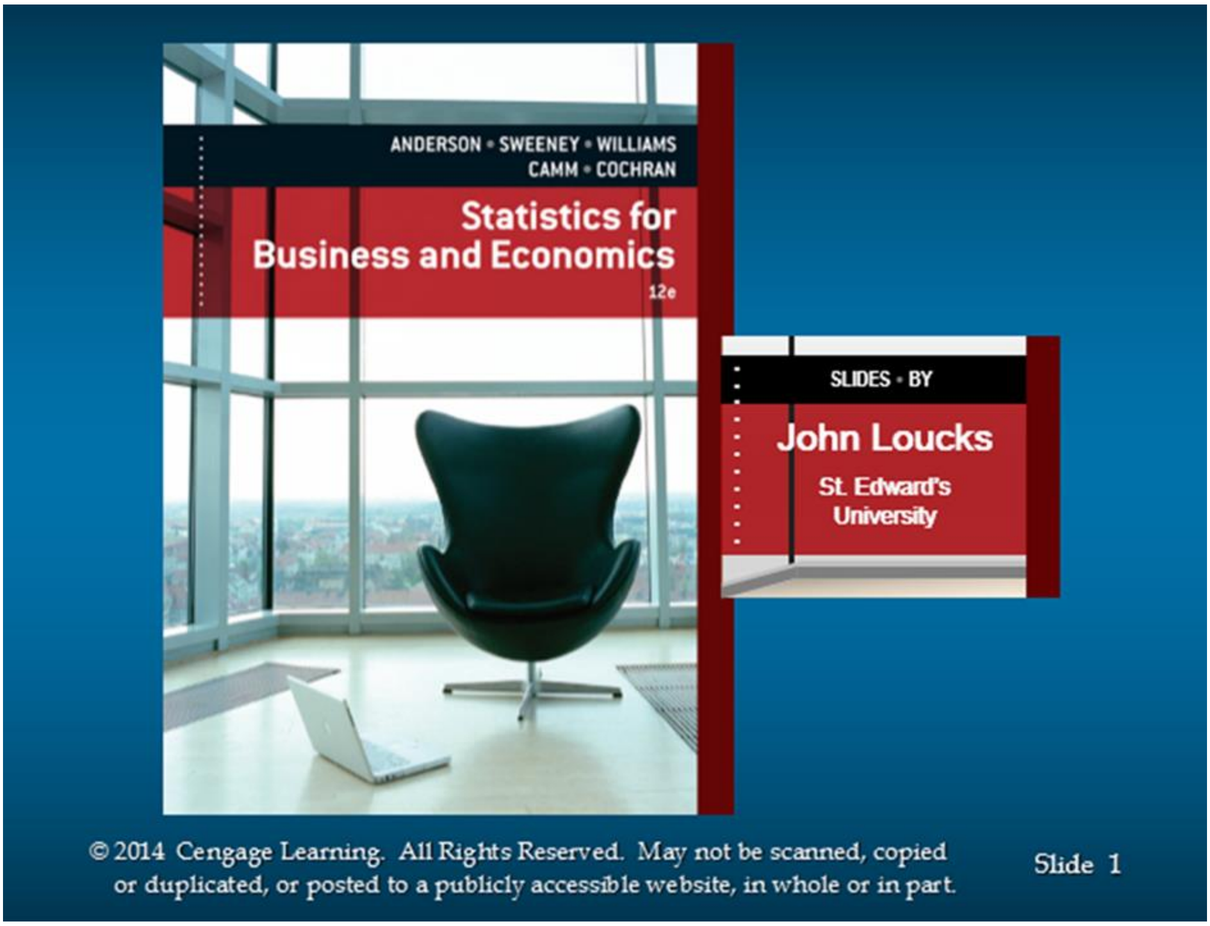
كلية إدارة الاعمال

الإحصاء 2 Statistics

الأستاذ الدكتور محمود محمد ديب طيوب

محاضرة رقم 3



الفصل الثاني للعام 2024–2025

الارتباط الخطي البسيط Simple linear correlation

1- مقدّمة:

نحتاج في كثير من الدراسات والأبحاث العلمية إلى تتبع وقياس العلاقة بين متغيّرين أو أكثر، ثم إلى تصوير هذه العلاقة في صيغة رياضية تتخذ شكل المعادلة وفقاً لنوع العلاقة التي نشاهدها ومن واقع البيانات الإحصائية الخاصة بهذه المتغيّرات مثال دراسة العلاقة بين عمر الأم ووزن المولود، أو دراسة وقياس العلاقة بين عدة جرعات من مادة سامة أعطيت في سياق التجربة ومدة حياتها، فالمتغيّران هنا يلعبان دوراً غير متناظر. فمثلاً المتغيّر x (الجرعات) يمثّل مسبباً والمتغيّر y الذي يمثل متغيّراً تابعاً.

يقصد بالارتباط بين متغيّرين وجود علاقة بينهما، بمعنى أنه إذا تغير أحد المتغيّرين في اتجاه معين زيادة أو نقصاناً، يميل المتغيّر الآخر إلى التغير في اتجاه معين كذلك، وبشكلٍ آخر

تعريف: إذا كان لكل قيمة من قيم المتغيّر x يوجد قيمة مقابلة لمتغيّر آخر y فإن الأزواج المرتّبة من هذه القيم تسمى مجتمعاً ذا بعدين ويسمى الزوج المرتّب (x,y) متغيّراً عشوائياً ذا بعدين. إن

.

الشروط اللازمة لدراسة العلاقة بين الظاهرتين المدروستين هي:

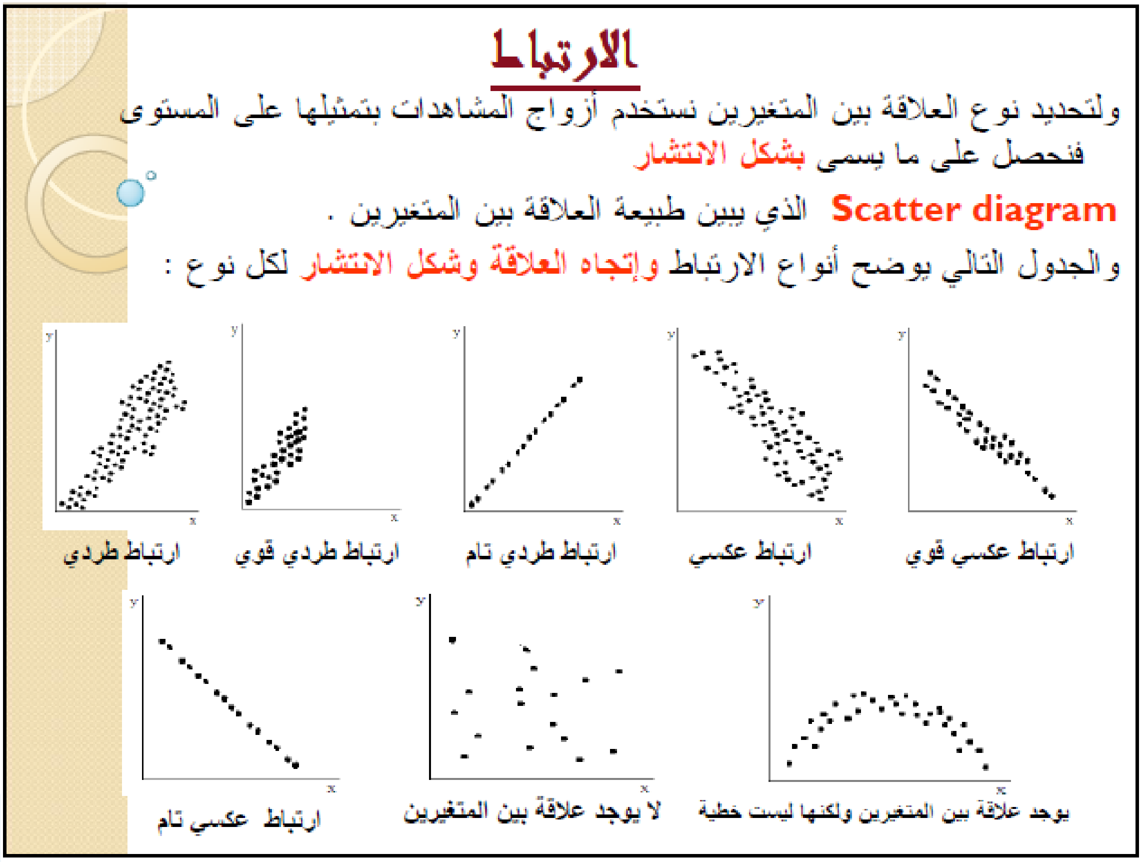
- أن تكون بين الظاهرتين المدروستين علاقة جدلية واضحة.
- أن تكون إحدى الظاهرتين ظاهرة مسببة والأخرى ناتجة.
- أن تكون كل من الظاهرتين قابلة للقياس بواسطة وحدة قياس معينة لكل منهما.
- أن تكون كل القياسات المأخوذة متقابلة من حيث الزمان أو المكان أو كلاهما معاً.

2-6- أنواع العلاقات بين الظواهر:

بعد التحقق من أن الشروط السابقة متوفرة بين الظاهرتين المدروستين ننتقل إلى مرحلة البحث عن نوعية ومتانة العلاقة بين تلك الظاهرتين لذلك نرمز بـ x للظاهرة المسببة ونرمز بـ y للظاهرة الناتجة وبالتالي نقول أن:

$$y = f(x)$$

إن هذه المعادلة يجب أن تعكس بدقة كافية العلاقة بين الظاهرتين x و y ولذلك فإن نوعها سيكون متعلقاً بنوع تلك العلاقة، ومن أهم أنواع العلاقات هي:



- - الارتباط البسيط (simple correlation):

لتحديد نوعية العلاقة بين المتغيّرين نعتمد على شكل الانتشار ونحاول رسم المنحني الذي يمر من أكثر النقاط أو بالقرب من معظمها مما يساعدنا في اختيار نوع التابع الذي يمثل العلاقة الارتباطية أفضل تمثيل،

قياس الارتباط

- تستخدم معاملات الارتباط لقياس درجة الارتباط بين متغيرين(ظاهرتين) .
- تعريف معامل الارتباط :

يعرف معامل الارتباط والذي يرمز له بالرمز r بأنه عبارة عن مقياس رقمي يقيس قوة ونوع الارتباط بين متغيرين , حيث تتراوح قيمته بين

(+1) و (-1) , أي أن

وتدل إشارة المعامل الموجبة على العلاقة الطردية ,

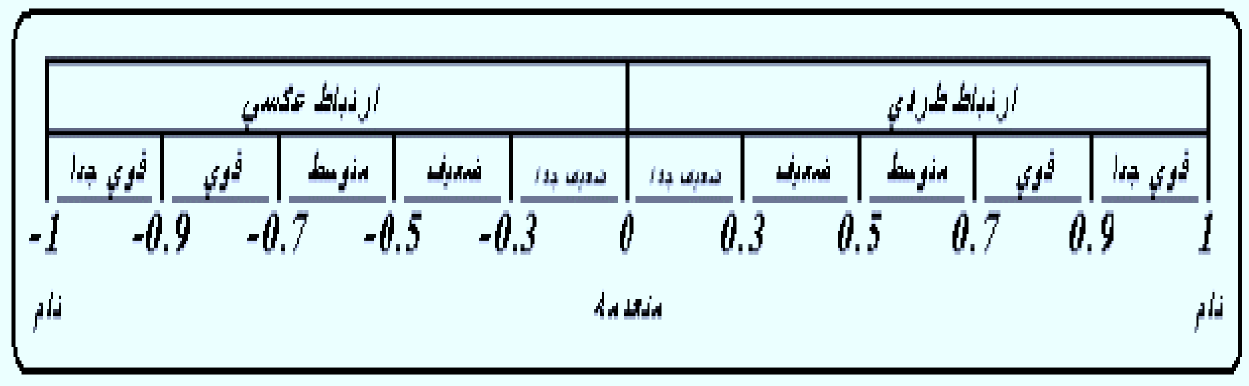
بينما تدل إشارة المعامل السالبة على العلا $-1 \leq r \leq +1$

قياس الارتباط

والجدول التالي يوضح أنواع الارتباط **واتجاه العلاقة** لكل نوع :

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
من 0.70 إلى 0.99	ارتباط طردي قوي
من 0.50 إلى 0.69	ارتباط طردي متوسط
من 0.01 إلى 0.49	ارتباط طردي ضعيف
0	لا يوجد ارتباط خطي

وما قيل عن الارتباط الطردي ينطبق على الارتباط العكسي (مع وضع إشارة سالبة)



1- معامل بيرسون للارتباط الخطي

- معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخداماً خاصة في العلوم الإنسانية والاجتماعية.

- عند تطبيق معامل بيرسون للارتباط يجب أن يكون كلا المتغيرين(y,x) بيانات كمية.

1 حساب معامل الارتباط الخطي لبيرسون:

يرمز لمعامل الارتباط بيرسون بـ r لأنه عبارة عن قيمة تقديرية لمعامل ارتباط للمجتمع R وقيمته تنحصر بين $-1 \le r \le +1$ وهي قيمة مطلقة لا يعبر عنها بواحدات قياس، وبصورةٍ عامة فإن معامل الارتباط للظواهر لا يمكن أن تصل إلى الواحد الصحيح وفيما يلي طرق حساب معامل الارتباط لبيرسون.

لتكن لدينا سلسلة القياسات التالية للمتغيّرين x و y :

$$x_i : x_1, x_2, x_3, x_4, , , , , x_i$$
$$y_i : y_1, y_2, y_3, y_4, , , , , y_i$$

طرق حساب معامل الارتباط:

Calculating a Correlation Coefficient

Calculating a Correlation Coefficient

	In Words	*In Symbols*
1.	Find the sum of the x-values.	$\sum x$
2.	Find the sum of the y-values.	$\sum y$
3.	Multiply each x-value by its corresponding y-value and find the sum.	$\sum xy$
4.	Square each x-value and find the sum.	$\sum x^2$
5.	Square each y-value and find the sum.	$\sum y^2$
6.	Use these five sums to calculate the correlation coefficient.	$r = \dfrac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$

Correlation Coefficient

Example continued:

Hours, x	0	1	2	3	3	5	5	5	6	7	7	10
Test score, y	96	85	82	74	95	68	76	84	58	65	75	50
xy	0	85	164	222	285	340	380	420	348	455	525	500
x^2	0	1	4	9	9	25	25	25	36	49	49	100
y^2	9216	7225	6724	5476	9025	4624	5776	7056	3364	4225	5625	2500

$\Sigma x = 54$ $\quad \Sigma y = 908$ $\quad \Sigma xy = 3724$ $\quad \Sigma x^2 = 332$ $\quad \Sigma y^2 = 70836$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{12(3724) - (54)(908)}{\sqrt{12(332) - 54^2}\sqrt{12(70836) - (908)^2}} \approx -0.831$$

There is a strong negative linear correlation.

As the number of hours spent watching TV increases, the test scores tend to decrease.

Activate Windows
Go to Settings

- حساب معامل الارتباط بالاعتماد على الانحراف المعياري:

يتم ذلك دون الحاجة إلى حساب الدرجات المعيارية بالعلاقة الآتية

هامة

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n.\sigma_x .\sigma_y}$$

مثال: من معطيات المثال نجد أن:

- نكوّن الجدول المساعد:

n	X	Y	$x - \bar{x}$	$y - \bar{y}$	$(x_i - \bar{x})(y - \bar{y})$
1	2	4	-3.8	-3.9	14.82
2	3	4	-2.8	-3.9	10.92
3	4	3	-1.8	-4.9	8.82
4	6	12	+0.2	+4.1	0.82
5	9	16	+3.2	+8.1	25.92
6	10	9	+4.2	+1.1	4.62
7	8	11	+2.2	+3.1	6.82
8	7	10	+1.2	+2.1	2.52
9	4	10	-1.8	+2.1	-3.78
10	5	9	-0.8	+1.1	-0.88
Σ	58	79			70.6

من معطيات المثال لدينا:

- المتوسط الحسابي لـ x $\bar{x} = 5.8$
- المتوسط الحسابي لـ y $\bar{y} = 7.9$

ومنه يكون معامل الارتباط:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n.\sigma_x .\sigma_y} = \frac{70.6}{10 \times 2.522 \times 4.482}$$

$$r = \frac{70.6}{113.036} = 0.625$$

- حساب معامل الارتباط عن طريق حساب الانحرافات عن المتوسط ومربعاتها:

يمكن تبسيط حساب معامل الارتباط والاكتفاء بحساب الانحرافات عن المتوسط الحسابي ومربع هذه الانحرافات وتفيد هذه الطريقة بحساب معامل الارتباط إذا كان عدد القياسات قليلاً.

مثال لتكن لدينا القياسات الآتية عن المتغيّرين x وy.

	1	2	3	4	5
x	5	7	8	6	9
y	7	7	9	8	9

المطلوب: حساب معامل الارتباط بين x وy.

الحل:

n	X	Y	$x-\bar{x}$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x_i-\bar{x})(y-\bar{y})$
1	5	7	-2	4	-1	1	2
2	7	7	0	0	-1	1	0
3	8	9	+1	1	+1	1	1
4	6	8	-1	1	0		0
5	9	9	+2	4	+1	1	2
Σ	35	40		10		4	5

1- حساب المتوسط الحسابي لـ x :

$$\bar{x}=\frac{\sum x}{n}=\frac{35}{5}=7$$

2- حساب المتوسط الحسابي لـ y:

$$\bar{y}=\frac{\sum y}{n}=\frac{40}{5}=8$$

3- حساب معامل الارتباط:

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

$$r=\frac{5}{\sqrt{10\times4}}=\frac{5}{\sqrt{40}}=0.79$$

- حساب معامل الارتباط من الدرجات الخام مباشرة:

أ - معامل بيرسون للارتباط الخطي

- حساب معامل بيرسون للارتباط الخطي :

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين

باستخدام الصيغة التالية: y, x

$$r_p=\frac{n\sum xy-(\sum x)(\sum y)}{\sqrt{(n\sum x^2-(\sum x)^2)(n\sum y^2-(\sum y)^2)}}$$

حيث :

$\sum_{i=1}^{n} x_i y_i$: مجموع حاصل ضرب x في y

$\sum x$: مجموع قيم المتغير x

$\sum y$: مجموع قيم المتغير y

$\sum x^2$: مجموع مربعات قيم المتغير x

$\sum y^2$: مجموع مربعات قيم المتغير y

$$r=\frac{n\sum x.y-(\sum x)(\sum y)}{\sqrt{\left[n\sum x^2-(\sum x)^2\right]\left[n\sum y^2-(\sum y)^2\right]}}$$

معامل بيرسون العزومي للارتباط الخطي

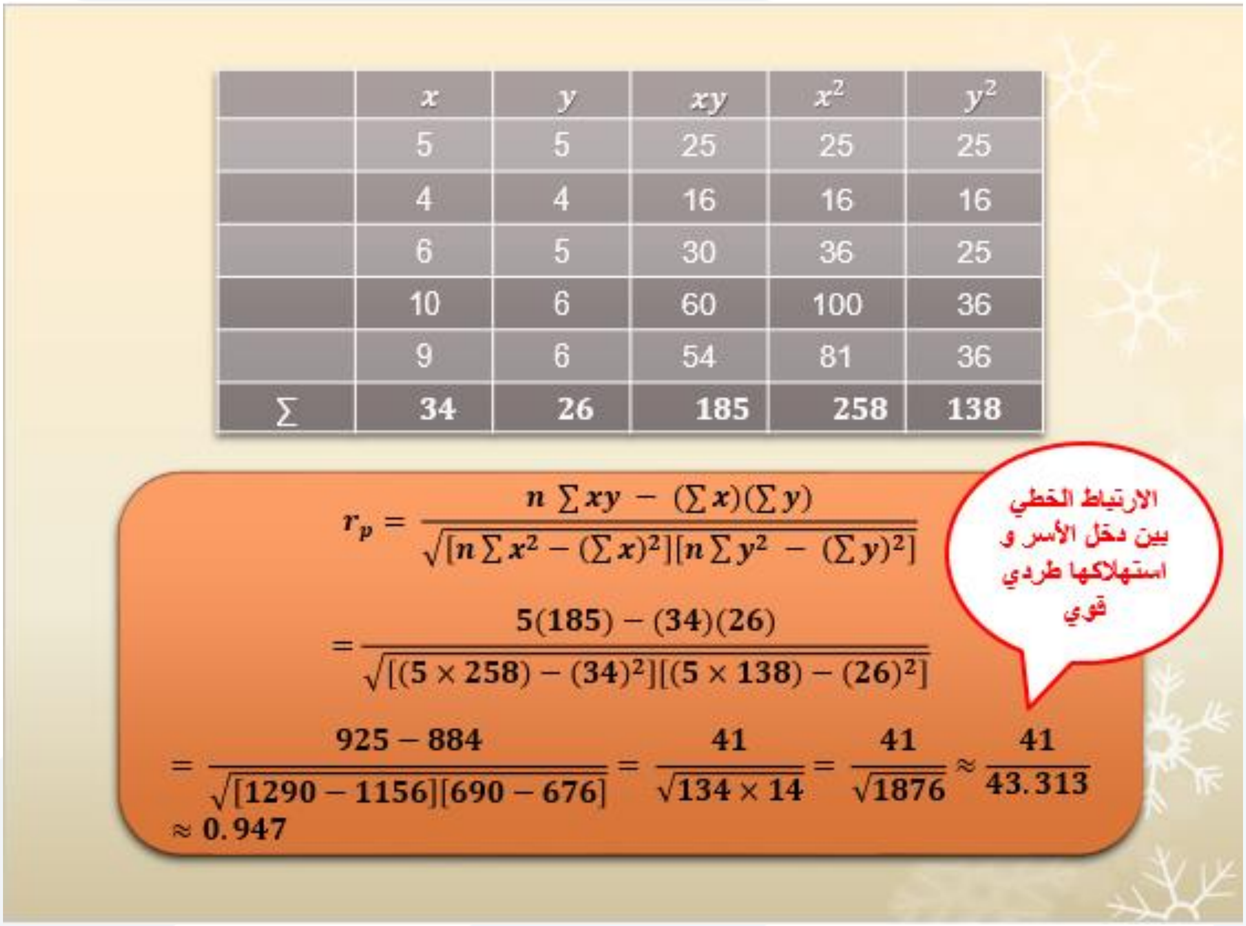
إذا كان لدينا n من البيانات في المتغيرين Y ,X
فإن معامل بيرسون للارتباط الخطي يحسب بالعلاقة :

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

مثال

لدراسة العلاقة بين الدخل x و الاستهلاك y بمئات الريالات في مدينة ما , أخذت عينة من الأسر فأعطت النتائج التالية:

x	5	4	6	10	9
y	5	4	5	6	6



	x	y	xy	x^2	y^2
	5	5	25	25	25
	4	4	16	16	16
	6	5	30	36	25
	10	6	60	100	36
	9	6	54	81	36
$\sum$	34	26	185	258	138

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$= \frac{5(185) - (34)(26)}{\sqrt{[(5\times 258) - (34)^2][(5\times 138) - (26)^2]}}$$

$$= \frac{925 - 884}{\sqrt{[1290 - 1156][690 - 676]}} = \frac{41}{\sqrt{134\times 14}} = \frac{41}{\sqrt{1876}} \approx \frac{41}{43.313}$$

$$\approx 0.947$$

مثال: يبين الجدول الآتي دخل وإنفاق عدد من الأسر

الأسر	1	2	3	4	5	6	7	8	9	10
الدخل x	5	7	6	8	10	12	9	6	14	10
الانفاق y	4	8	5	10	12	14	15	9	10	8

المطلوب: حساب معامل الارتباط من الدرجات الخام مباشرة.

جامعة المنارة
MANARA UNIVERSITY

الجدول المساعد لحساب معامل الارتباط الخطي لبيرسون

الحل:

n	X	Y	X^2	Y^2	X.Y
1	5	4	25	16	20
2	7	8	49	64	56
3	6	5	36	25	30
4	8	10	64	100	80
5	10	12	100	144	120
6	12	14	144	196	168
7	9	15	81	225	135
8	6	9	36	81	54
9	14	10	196	100	140
10	10	8	100	64	80
Σ	87	95	831	1015	883

ومنه نجد أن:

$$r = \frac{n\sum x.y - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

$$r = \frac{10\times 883 - 87\times 95}{\sqrt{\left[10\times 831 - 87^2\right]\left[10\times 1015 - 95^2\right]}}$$

$$r = \frac{8830 - 8265}{\sqrt{741\times 1125}} = \frac{565}{\sqrt{833625}}$$

$$r = \frac{565}{913.031} = 0.6188 \approx 0.62$$

مثال

يبين الجدول التالي درجات الرياضيات ويرمز لها بـ x ودرجات الإحصاء ويرمز بها بـ y:

درجات الرياضيات x	15	12	10	14	8	11	9	16	17	*18*
درجات الإحصاء y	13	11	8	12	6	8	7	10	12	*13*

والمطلوب حساب معامل الارتباط بين المتغيّرين x و y وذلك باعتماد الصيغ المختلفة:

Calculating a Correlation Coefficient

In Words	*In Symbols*
1. Find the sum of the x-values.	$\sum x$
2. Find the sum of the y-values.	$\sum y$
3. Multiply each x-value by its corresponding y-value and find the sum.	$\sum xy$

مجموع قيم المتغيرx

مجموع قيم المتغيرy

مجموع جداء قيم المتغيرين

Calculating a Correlation Coefficient

In Words	*In Symbols*
4. Square each x-value and find the sum.	$\sum x^2$
5. Square each y-value and find the sum.	$\sum y^2$
6. Use these five sums to calculate the correlation coefficient.	$r = \dfrac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$

مجموع مربعات قيم المتغير المستقل

مجموع مربعات قيم المتغير

علاقة حساب معامل الارتباط

الحل: لحساب معامل الارتباط يفضل إنشاء الجدول المساعد التالي:

الجدول المساعد لحساب معامل الارتباط بين متغيّرين لبيانات مبوّبة

i	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	xi^2	y_i^2	X . y
1	15	13	+2	+3	6	4	9	225	169	195
2	12	11	-1	+1	-1	1	1	144	121	132
3	10	8	-3	-2	6	9	4	100	64	80
4	14	12	+1	+2	2	1	4	196	144	168
5	8	6	-5	-4	20	25	16	64	36	48
6	11	8	-2	-2	4	4	4	121	64	88
7	9	7	-4	-3	12	16	9	81	49	63
8	16	10	+3	0	0	9	0	256	100	160
9	17	12	+4	+2	8	16	4	289	144	204
10	18	13	+5	+3	15	25	9	324	169	234
$\sum$	130	100	0	0	72	110	60	1800	1060	1372

الحل :

1- حساب المتوسط الحسابي لمتغيّر xi:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{130}{10} = 13$$

2- حساب الانحراف المعياري للمتغير xi:

$$\sigma_X = \sqrt{\frac{1}{n}\sum\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{1}{10}(110)} = 3,32$$

3- حساب المتوسط الحسابي للمتغير yi:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{110}{10} = 11$$

1- حساب الانحراف المعياري للمتغيّر yi:

$$\sigma_y=\sqrt{\frac{1}{n}\sum\left(y_i-\bar{y}\right)^2}=\sqrt{\frac{1}{10}(60)}=2.45$$

لحساب معامل الارتباط نعوض في المعادلات التالية:

- **المعادلة الأولى :**

$$R_{yx}=\frac{\Sigma\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{n.\sigma_x.\sigma_y}=\frac{72}{10\times 2.45\times 3.32}=\frac{72}{71.34}=0.88$$

المعادلة الثانية :

$$R_{yx}=\frac{\Sigma\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\Sigma\left(x_i-\bar{x}\right)^2}\sqrt{\Sigma\left(y_i-\bar{y}\right)^2}}=\frac{72}{\sqrt{110}\ \sqrt{60}}=\frac{72}{81.34}=0.88$$

- **المعادلة الثالثة:**

$$R_{yx}=\frac{n\Sigma x_i y_i-\Sigma x_i.\Sigma y_i}{\sqrt{n\Sigma x_i^2-(\Sigma x_i)^2}\sqrt{n\Sigma y_i^2-(\Sigma y_i)^2}}=$$

$$R_{yx}=\frac{10\times 1372-130\times 110}{\sqrt{10\times 1800-(130)^2}\sqrt{10\times 1060-(110)^2}}=\frac{720}{\sqrt{1100}\sqrt{600}}=\frac{720}{812.4}=0,88$$

- **مثال:**

سُجلت ست قراءات تقريبية لحجم الإنتاج وحجم صادرات النفط الخام بالمملكة العربية السعودية (بالمليار برميل) خلال عدة سنوات كما يلي:

حجم الصادرات (y)	2	2	2	1	1	1
حجم الإنتاج (x)	3	4	2	2	2	2

ادرس وجود علاقة ارتباط خطية بين حجم الإنتاج وحجم صادرات النفط الخام.

- **الحل :**

x	y	xy	x^2	y^2
3	2	6	9	4
4	2	8	16	4
2	2	4	4	4
2	1	2	4	1
2	1	2	4	1
2	1	2	4	1
Σ 15 $=\sum x$	9 $=\sum y$	24 $=\sum xy$	41 $=\sum x^2$	15 $=\sum y^2$

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$$r_p = \frac{6(24)-(15)(9)}{\sqrt{((6\times 41)-15^2)((6\times 15)-9^2)}} =$$

$$\frac{144-135}{\sqrt{(246-225)(90-81)}} = \frac{9}{\sqrt{189}} = \frac{9}{13.75} = 0.65$$

من الملاحظ أن علاقة الارتباط الخطي بين حجم الإنتاج وحجم صادرات النفط الخام علاقة طردية متوسطة.

3-4- خصائص معامل الارتباط:

1 . إن قيمة معامل الارتباط تتراوح ضمن المجال [+1 ، -1] أي $-1 \le r \le +1$.

2 . إذا أضفنا أو طرحنا عدداً ثابتاً مقداره C إلى كل من القياسات x_i و y_i فإن قيمة معامل الارتباط لا تتغير أي أن:

$$R'_{yx} = \frac{\sum_{i=1}\left[\left(x_i + C - \left(\bar{x} + C\right)\right)\right]\left[\left(y_i + C\right) - \left(\bar{y} + C\right)\right]}{n.\sigma_x.\sigma_y} = R_{xy} \qquad (11-6)$$

$$R'_{yx} = \frac{\sum_{i=1}^{n}\left(n - \bar{x}\right)\left(y - \bar{y}\right)}{n.\sigma_x.\sigma_y} = rx.y \qquad (12-6)$$

وذلك لأن العملية لا تؤثر على الانحرافات ولكن تجعل المتوسطين يزيدان بمقدار العدد الثابت.

3− إذا ضربنا أو قسمنا كلاً من القياسات x_i و y_i بعددين ثابتين C_1 و C_2 على التوالي فإن قيمة معامل الارتباط لا تتغيروذلك لأن هذه العملية تزيد من قيمتي $\bar{x}$ و σ_x و $\bar{y}$ و σ_y بمقدار C_1 و C_2 على التوالي وبذلك يكون:

$$R'_{yx} = \frac{C_1 C_2 \sum_{i=1}^{n} (X - \bar{X})(Y - \bar{Y})}{n C_1 \sigma_x C_2 \sigma_y}$$

$$R'_{yx} = \frac{C_1 C_2 \sum (X - \bar{X})(Y - \bar{Y})}{C_1 C_2 n \sigma_x \sigma_y}$$

مثال: بالعودة إلى معطيات المثال (1) نجد أن:

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = 72$$

$$\sigma_x = 3.32 ; \sigma_y = 2.45$$

$$n = 10$$

$$c_1 = 2, c_2 = 4$$

ومنه نجد أن:

$$r'_{yx} = \frac{2\times4\times72}{2\times4\times10\times3.32\times2.45} = 0.89$$

تفسير معامل ارتباط بيرسون:

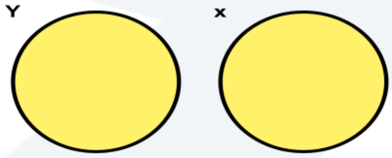
يتضح مما سبق أن معامل الارتباط بين متغيّرين هو قيمة مجردة تعبر عن العلاقة القائمة بين المتغيّرين حيث تنحصر قيمته بين −1، +1 ويعبر عادة عن قيمة معامل الارتباط بكسر عشري.

وهنا نلفت الانتباه من الوقوع في خطأ تفسير معامل الارتباط على أنه نسبة مئوية، فمثلاً معامل الارتباط (0.25) لا يعد نصف معامل ارتباط (0.50) ومعامل الارتباط (0.50) لا يعد نصف معامل ارتباط الذي قيمته واحد صحيح.

– كما أن الفرق بين معاملي الارتباط **(0.40)** **(0.60)** لا يساوي الفرق بين معاملي الارتباط **(0.70)** و **(0.90)** فمعامل الارتباط هو مقدار مجرد ولا يقاس على ميزان خطي ووحداته متساوية.

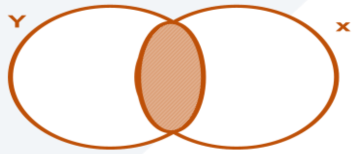
– كما **لا** يجب تفسير معامل الارتباط على أساس **وحدات الدرجات الأصلية حيث أن قيمة معامل الارتباط تكون مستقلة عن الوحدات التي تقاس بها المتغيّرات والقيم التي يأخذها كل منهما.** وأحياناً يعدّ الباحث أن معامل الارتباط الذي تنحصر قيمته بين 0.30 و 0.70 متوسط القيمة، أي يعبر عن علاقة ارتباطية متوسطة. ولكن هذه خاطئة من وجهة نظر الاستدلال الإحصائي فدلالة معامل الارتباط هي دالة لحجم العيّنة، حيث أن قيمة معامل الارتباط المرتفعة التي يحصل عليها الباحث باستخدام عينات صغيرة، ربما لا يكون لها أي معنى على الإطلاق من ناحية الاسالإحصائي على الارتباط في المجتمع الأصلي الذي سحبت منه العيّنات. **من الطرق المفيدة في تفسير القيم المختلفة لمعامل الارتباط (r) هو تربيع هذه القيم أي الحصول على قيمة (r^2) والمقدار r^2 هو النسبة بين التباين الكلي لأحد المتغيّرين والجزء من هذا التباين الذي يمكن التنبؤ به باستخدام المتغيّر الثاني.** فإذا كان معامل **الارتباط بين متغيّرين 0.707 مثلاً فإن $r^2 = 0.50 = 0.707$ وعندما $r=0.50$ فإن $r^2 = 0.25$ ولذلك فإنه يمكن عدّ أن معامل الارتباط 0.707 ضعف معامل الارتباط 0.50 حيث أن نسبة r^2 في الحالتين** 1 : 2 تقريباً.. **ويسمى أحياناً المقدار (r^2) معامل التحديد أو التباين المشترك بين المتغيّرين، لأن قيمته تعبر عن ذلك الجزء من التباين لأحد المتغيّرين الذي يمكن تحديده أو التنبؤ به باستخدام المتغيّر الآخر، فإذا كان معامل الارتباط $r = 0.80$ فإن $r^2 = 0.64$ وهذا يعني أن تباين مشترك بين المتغيّرين نسبته 64% وإذا كانت $r = -1$ أو $r = 1+$ يكون هناك تباين مشترك بين**

المتغيّرين نسبته 100% وإذا كانت $r = 0$ فإن $r^2 = 0$ ولا يكون هناك تباين مشترك بين المتغيّرين ويمكن توضيح ذلك بالأشكال التالية:

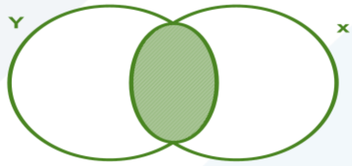


1- إذا كانت $r = 0$ ، $r^2 = 0$

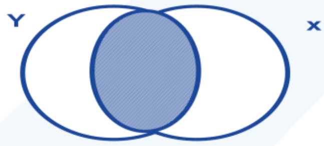
– إذا كانت $r = 0.30$ ، $r^2 = 0.90$



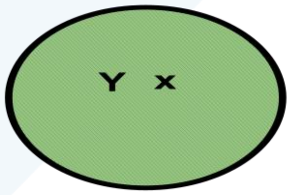
3- إذا كانت $r = 0.80$, $r^2 = 0.64$



– إذا كانت $r = 0.90$, $r^2 = 0.81$



5- إذا كانت $r = 1$, $r^2 = 1$



ملاحظة يمثل الجزء المظلل المحصور بين الدائرتين (جزء التقاطع) بالتباين المشترك بين المتغيّرين.

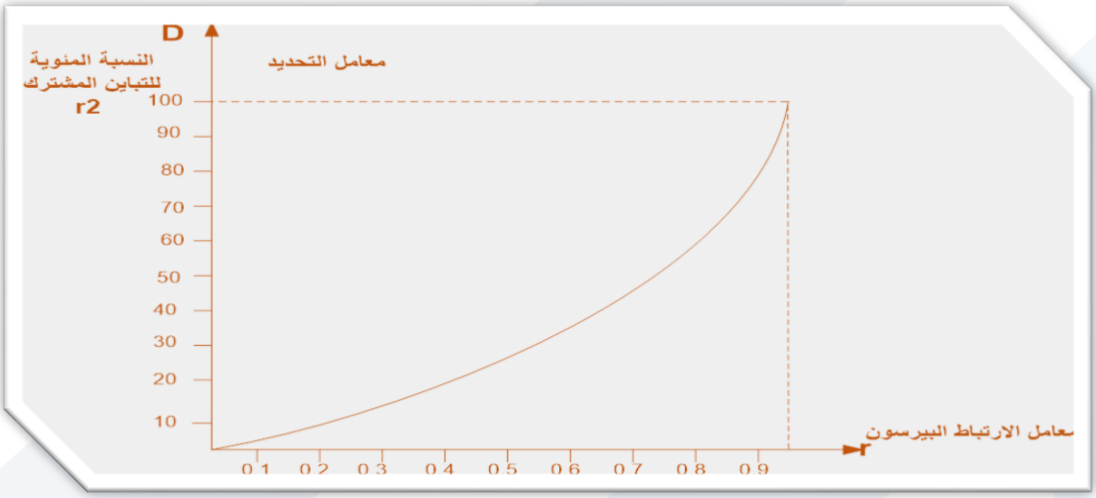
يسمى المقدار $Rs = 1- r^2$ بمعامل الاغتراب أو عدم التحديد *coefficient de Non determination* لأن قيمتها تعبر عن الجزء من التباين في أحد المتغيّرين الذي لا نستطيع التنبؤ به أو تحديده باستخدام المتغيّر الآخر. ونظراً لأن قيمة (r) تختلف عن قيمة (r^2) فإنه يجب على الباحث أن يحتاط عند تفسير قيمة معامل الارتباط بين متغيّرين ،الجدول التالي:

جامعة المنارة
MANARA UNIVERSITY

جدول 6−1 معامل الارتباط والنسبية المئوية للتباين المشترك

معامل الارتباط	النسبة المئوية للتباين المشترك r^2
0.1	1
0.2	4
0.3	9
0.4	16
0.50	25
0.60	36
0.70	49
0.80	64
0.90	81
1	1

نلاحظ أن معاملات الارتباط التي تتراوح قيمته بين 0.1، 0.3 نبين أن جزءاً صغيراً من التباين في (y) يقترن بتباين (x) حيث يتراوح بين 1% إلى 9% ويمكن توضيح ذلك بالشكل البياني:



شكل (6−1) يبين العلاقة بين معامل الارتباط r ومعامل التمديد D.

وتبين أن معامل الارتباط 0.50 الذي يعده كثير من العاملين في العلوم الاقتصادية أو الاجتماعية والتربوية والاجتماعية معاملاً مرتفعاً، يعني أن 25% من التباين في المتغيّر (y) يقترن بالتباين في المتغيّر (x)، وهذا يعني أيضاً أن 75% من التباين في (y) يقترن بعوامل أخرى تختلف عن المتغيّر (x) ومن هذا تبين أن الباحث يحتاج إلى معامل ارتباط مقداره (0.71)= $r = 0.5041 = 0.71$ لكي يعدّ أن نصف التباين في المتغيّر (y) يقترن بالتباين في المتغيّر (x).

5−4− الخطأ المعياري لمعامل الارتباط σ_r :

عبارة عن مدى بعد أو قرب قيمة معامل الارتباط المشتق من العيّنة التجريبية عن المعامل الحقيقي فيما لو أجريت الدراسة على أفراد المجتمع الأصلي،

فمثلاً إذا أجرى الباحث دراسة على عينة مؤلفة من 100 طالب وقام بحساب معامل الارتباط بين الدرجات التحصيلية في مادة الرياضيات وبين الاحصاء فكان معامل الارتباط مثلاً **r=0.62** فإن الخطأ المعياري لهذا الارتباط يساوي:

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

ومنه يكون:

$$\sigma_r = \frac{1-0.62^2}{\sqrt{100-1}} = \frac{1-0.384}{9.9498} = 0.06553$$

ونستنتج من ذلك أن معامل الارتباط الحقيقي يقع بين القيمتين التاليتين عند حدود ثقة 95% وشك 5%:

$$P = \left(0.62 \mp 1.96 \times 0.06553\right) = 0.95$$

أو عند 99% وشك 1%

$$P = \left(0.62 \mp 2.58 \times 0.06553\right) = 0.99$$

:

فيكون المجال عند 95% يساوي:

$$P = \left(0.62 \mp 0.1284\right) = 0.95$$

$$P\left(0.4915; 0.7484\right) = 0.95$$

وعند 99% يكون:

$$P = \left(0.62 \mp 0.169\right) = 0.99$$

$$P\left(0.451; 0.789\right) = 0.99$$

يلاحظ من ذلك بأن حدود معامل الارتباط الحقيقي واسعة نسبياً وغالباً ما تختلف عن المعامل التجريبي، وهذا ما يجعل معاملات الارتباط المستخرجة من عينات صغيرة ضعيفة الثبات، ويلاحظ أن معامل الارتباط يختلف عن معاملات النزعة المركزية أو التشتّت، في أن توزيعه ليس دائماً اعتدالياً، فالتوزيع لا يكون كذلك إلا حين تكون العيّنة كبيرة نسبياً أو في حالات صغر قيمة معامل الارتباط. أما معامل ارتباط كبير نسبياً (0.85 – 0.75) فإن توزيع معامل الارتباط يكون ملتوياً لذلك فإن حساب الانحراف المعياري (الخطأ المعياري) لمعامل الارتباط يكون قليل الفائدة.

جَـامعة المَـنارة
MANARA UNIVERSITY

معنوية معامل الارتباط

Hypothesis Testing for ρ

A hypothesis test can also be used to determine whether the sample correlation coefficient r provides enough evidence to conclude that the population correlation coefficient ρ is significant at a specified level of significance.

A hypothesis test can be one tailed or two tailed.

$$\begin{cases} H_0: \rho \geq 0 & \text{(no significant negative correlation)} \\ H_a: \rho < 0 & \text{(significant negative correlation)} \end{cases}$$ Left-tailed test

$$\begin{cases} H_0: \rho \leq 0 & \text{(no significant positive correlation)} \\ H_a: \rho > 0 & \text{(significant positive correlation)} \end{cases}$$ Right-tailed test

$$\begin{cases} H_0: \rho = 0 & \text{(no significant correlation)} \\ H_a: \rho \neq 0 & \text{(significant correlation)} \end{cases}$$ Two-tailed test

Activate W

The *t*-Test for the Correlation Coefficient

- Can be used to test whether the correlation between two variables is significant.
- The **test statistic** is ***r***.
- The **standardized test statistic**

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

follows a *t*-distribution with **d.f. = *n* – 2**.
- **In this text, only two-tailed hypothesis tests for *ρ* are considered.**

Using the *t*-Test for ρ

In Words	*In Symbols*
1. State the null and alternative hypothesis.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom.	d.f. = $n - 2$
4. Determine the critical value(s) and rejection region(s).	Use Table 5 in Appendix B.

Using the t-Test for ρ

In Words	*In Symbols*
5. Find the standardized test statistic.	$t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$
6. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, reject H_0. Otherwise fail to reject H_0.
7. Interpret the decision in the context of the original claim.	

Example: Finding the Correlation Coefficient

Calculate the correlation coefficient for the gross domestic products and carbon dioxide emissions data. What can you conclude?

GDP (trillions of \$), x	CO_2 emission (millions of metric tons), y
1.6	428.2
3.6	828.8
4.9	1214.2
1.1	444.6
0.9	264.0
2.9	415.3
2.7	571.8
2.3	454.9
1.6	358.7
1.5	573.5

Solution: Finding the Correlation Coefficient

x	y	xy	x^2	y^2
1.6	428.2	685.12	2.56	183,355.24
3.6	828.8	2983.68	12.96	686,909.44
4.9	1214.2	5949.58	24.01	1,474,281.64
1.1	444.6	489.06	1.21	197,669.16
0.9	264.0	237.6	0.81	69,696
2.9	415.3	1204.37	8.41	172,474.09
2.7	571.8	1543.86	7.29	326,955.24
2.3	454.9	1046.27	5.29	206,934.01
1.6	358.7	573.92	2.56	128,665.69
1.5	573.5	860.25	2.25	328,902.25
$\Sigma x = 23.1$	$\Sigma y = 5554$	$\Sigma xy = 15{,}573.71$	$\Sigma x^2 = 67.35$	$\Sigma y^2 = 3{,}775{,}842.76$

Solution: Finding the Correlation Coefficient

$\Sigma x = 23.1$ $\Sigma y = 5554$ $\Sigma xy = 15{,}573.71$ $\Sigma x^2 = 32.44$

$\Sigma y^2 = 3{,}775{,}842.76$

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

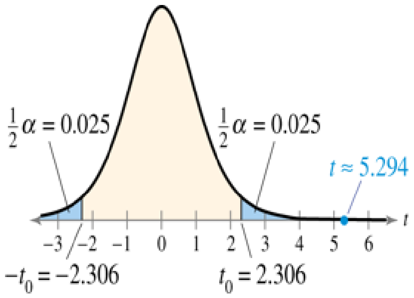
$$= \frac{10(15{,}573.71) - (23.1)(5554)}{\sqrt{10(67.35) - 23.1^2}\sqrt{10(3{,}775{,}842.76) - 5554^2}}$$

$$= \frac{27{,}439.7}{\sqrt{139.89}\sqrt{6{,}911{,}511.6}} \approx 0.882$$

r ≈ 0.882 suggests a strong positive linear correlation. As the gross domestic product increases, the carbon dioxide emissions also increase.

Solution: *t*-Test for a Correlation Coefficient

- H_0: $\rho = 0$
- H_a: $\rho \neq 0$
- $\alpha = 0.05$
- d.f. = $10 - 2 = 8$
- **Rejection Region:**



If you reject the hypothesis, then there is evidence to conclude a linear correlation

- **Test Statistic:**

$$t \approx \frac{0.882}{\sqrt{\frac{1-(0.882)^2}{10-2}}} \approx 5.294$$

- **Decision:** Reject H_0

At the 5% level of significance, there is enough evidence to conclude that there is a significant linear correlation between gross domestic products and carbon dioxide emissions.

© 2012 Pearson Education, Inc. All rights reserved.

أي باختصار يمكن اختبار دلالة معنوية معامل الارتباط بين x وy:

– الفرضية الابتدائية = $H_0 = r = 0$

– الفرضية البديلة $H_1 = r \neq 0$

– حساب قيمة دالة الاختبار (t) لستودينت:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.952}{\sqrt{\frac{1-0.952^2}{10-2}}} = 8.797$$

إذا علمت أن قيمة t_{x_2} الجدولية عند مستوى دلالة 5% تساوي 1.860 ومقارنة قيمة t

المحسوبة مع الجدولية نجد أن المحسوبة أكبر من الجدولية وبالتالي ترفض الفرضية الابتدائية والارتباط بين المتغيّرين دال إحصائياً وللتأكد أيضاً تحسب معامل التحديد $r^2 = 0.952^2 = 0.9063$ أي 90.63% من التباين المشترك.

– حساب معامل الاغتراب $r_s = 1- r^2$ أي $rs = 1-0.952^2 = 0.0936$

أي أن هناك $r_s\% = 0.0436 \times 10$ يساوي 9.369 %

دلالة معامل الارتباط:

إذا كان معامل الارتباط قريباً من ∓ 1 هناك علاقة خطيّة قوية بين المتغيّرين وإذا كان معامل الارتباط مساوياً للصفر لم يكن هناك علاقة خطيّة. أما قيم R_{yx} المتوسطة الأخرى فيجب اختبار دلالة معامل الارتباط لها. لإجراء هذا الاختبار نعدّ المجتمع ذا البعدين الذي أخذت منه العيّنة المكونة من n من الأزواج ونفترض أن معامل الارتباط فعندها يكون اختيار الدلالة R_{yx} يساوي:

$$t = \frac{Ryx}{\sqrt{\frac{1-R^2yx}{n-2}}}$$

القيمة الجدولية لـ t لأن $n=10$ وبالتالي تكون $t(x_2 = 0.05, 8) = 2.306$ ولاختبار دلالة معامل الارتباط نتبع الخطوات التالية:

. الفرضية الابتدائية: $H_0 : R_{yx} = 0$.

. الفرضية البديلة: $H_1 : R_{yx} \neq 0$

. مستوى المعنوية مثلاً: $\alpha = 0,05$

تحديد دلالة الاختبار:

$$t=\frac{r_{yx}}{\sqrt{\frac{1-r_{yx}^{2}}{n-2}}}=\frac{0.88}{\sqrt{\frac{(1-0.88)^{2}}{10-2}}}=\frac{0.88}{0.168}=5.24$$

من جدول ستودينت $t = [0.05\ ,\ 8] = 2.306$ بما أن t المحسوبة أكبر من $t_{\left(1-\frac{\alpha}{2},n-2\right)}$ نرفض الفرضية الابتدائية H_0 وبالتالي تغير وجود ارتباط معنوي بين المتغيّرين المدروسين.

اختبار معنوية معامل الارتباط:

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

$$t=0.88\sqrt{\frac{1-0.88^{2}}{10-2}}=5.24$$

كما يمكن وضع العلاقة بالصورة التالية: حساب معامل التحديد بدلالة قيمة اختبار معنوية معامل الارتباط

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

Testing the Significance of the Correlation Coefficient – Copier Sales Example

H_0: $\rho = 0$ (the correlation in the population is 0)

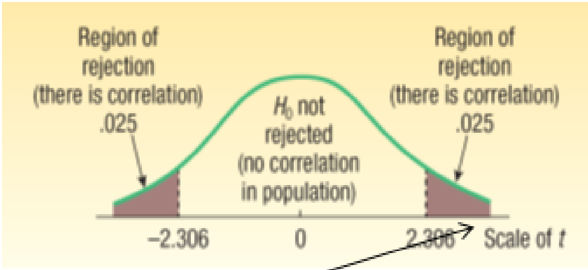
H_1: $\rho \neq 0$ (the correlation in the population is not 0)

Reject H_0 if:

$t > t_{\alpha/2,n-2}$ or $t < -t_{\alpha/2,n-2}$

$t > t_{0.025,8}$ or $t < -t_{0.025,8}$

$t > 2.306$ or $t < -2.306$



$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.759\sqrt{10-2}}{\sqrt{1-.759^2}} = 3.297$$