الارتباط الخطي البسيط

1- مفهوم العلاقة والارتباط:

لا أحد يستطيع أن ينكر أنه كلما تقدم الطفل في السن كلما تبع ذلك زيادة في بنيته الجسمية (الطول، الوزن...). فنستنتج هنا أن هناك علاقة بين متغيري "السن" و"البنية الجسدية"، يمكن القول بناء على ذلك أيضا أن متغير "البنية" هو نتاج أو محصلة متغير "السن".

مثال آخر: نلاحظ أن درجة الضغط ترتفع تبعا لارتفاع مستوى الهبوط تحت الماء، فعند عمق محدد تتناسب درجة ضغط محددة أيضا. فهناك بين هاتين الدرجتين أو القيمتين (العمق والضغط) علاقة واضحة تسمى تامة، بل وتسمى قانونا أيضا.

أما بالنسبة للارتباط، فلا بأس أيضا أن نوضحه من خلال المثال التالي: إذا كنا بصفة عامة نعتقد بأن وزن الفرد وطوله مرتبطان؛ على اعتبار أنه كلما كان الشخص أطول كان وزنه أكبر، فإنه ينبغي تسجيل أن هذه العلاقة ليست تامة، فهناك أشخاص قصيرون لكنهم بدينون، أو العكس طويلون لكنهم نحيفون. ففي هذه الحالة نحن نتحدث عن ارتباط وليس عن علاقة، لأنه وعلى العكس من المثال الأول، فإن هذه الحالة ليست مثبتة في جميع الأحوال.

وهو ما أدى بـ D'Hainaut إلى التصريح بأن معظم العلاقات بين القيم في العلوم الفيزيائية هي عبارة عن قوانين، بينما معظم العلاقات بين المتغيرات في العلوم الإنسانية هي عبارة عن ارتباطات، وهو أحد مصادر – كما يخلص الباحث – عدم الدقة وصعوبة التنبؤ في تلك العلوم.

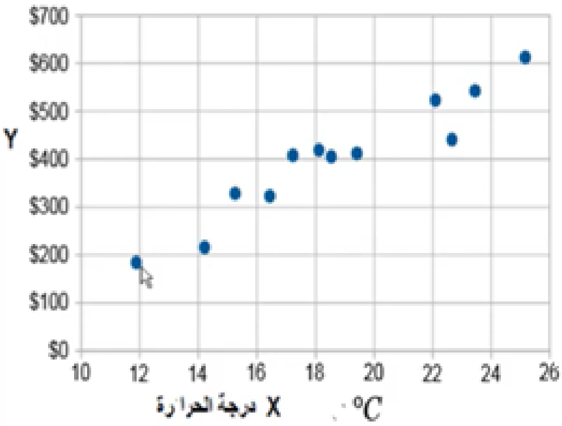
2- أنواع الارتباط بين المتغيرات:

- **اتجاه الارتباط:** يوجد هناك ارتباط موجب وارتباط سالب، عند الحصول على قيمة موجبة لمعامل الارتباط فإن هناك علاقة **طردية** بين المتغيرات المدروسة، أي أن الزيادة في المتغير الأول تتبعها زيادة أيضا في المتغير الثاني (كلما ارتفع التركيز تحسن التحصيل الدراسي). بينما يدل الحصول على قيمة سالبة لمعامل الارتباط على وجود علاقة **عكسية** بين المتغيرين إذ أن الزيادة في المتغير الأول نجم عنها نقصان في المتغير الثاني (كلما زادت الغيابات انخفض التحصيل).

- **قوة الارتباط:** أغلب معاملات الارتباط تنحصر قيمها بين (+1 و -1)، فإذا بلغت قيمة معامل الارتباط +1 فإن الارتباط بين المتغيرين طردي تام وهو أقوى أنواع الارتباط بين المتغيرات، وعلى العكس من ذلك، إذا بلغت قيمته -1 فإن الارتباط عكسي تام، ولا وجود لأي ارتباط بين المتغيرين في حالة بلوغ قيمة المعامل 0.

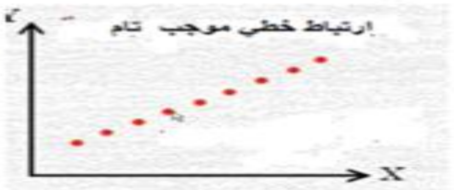
مثال على تمثيل بيانات على لوحة الانتشار

Y مبيعات الايس كريم	X درجة الحرارة
$215	14.2°
$325	16.4°
$185	11.9°
$332	15.2°
$406	18.5°
$522	22.1°
$412	19.4°
$614	25.1°
$544	23.4°
$421	18.1°
$445	22.6°
$408	17.2°



لوحة النتشار- الارتباط الخطي التام

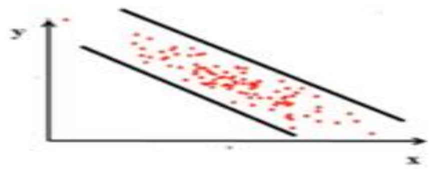
طردي



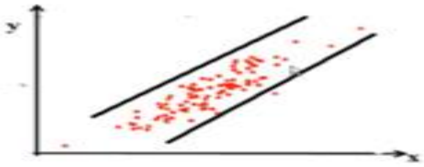
عكسي

لوحة الانتشار-الارتباط الخطي

ارتباط خطي جزئي سالب

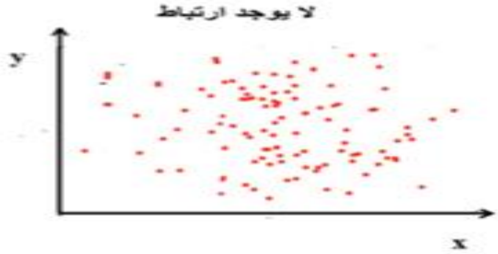


ارتباط خطي جزئي موجب

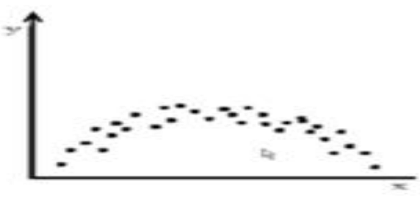


لوحة الانتشار لايوجد ارتباط خطي

لا يوجد ارتباط



لايوجد ارتباط خطي



3- العلاقة الخطية ومعامل الارتباط:

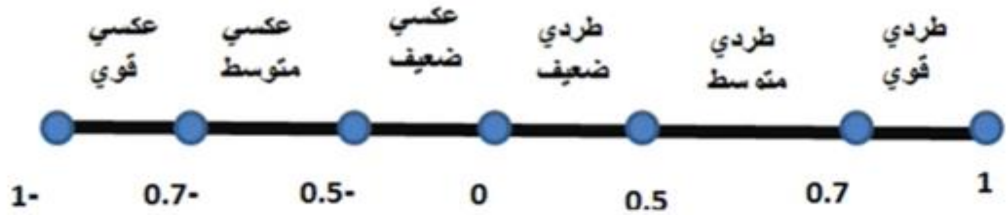
على الرغم من أنها ليست بديلا عن تطبيق الاختبارات الإحصائية التي تفحص الارتباط بين المتغيرات، إلا أنه يمكن معرفة قوة الارتباط (قوي – ضعيف – معدوم) واتجاهه (طردي – عكسي) عن طريق رسم ما يسمى لوحة الانتشار.

يتم من خلالها تمثيل المتغيرين بيانيا على محورين أفقي وعمودي وتوزيع قيم كلا المتغيرين على اللوحة لنحصل على طريقة لانتشار تلك القيم يمكننا أن نستنتج على ضوئها وجود أو عدم وجود ارتباط بين المتغيرين المدروسين.

دلالات القيم المختلفة لمعامل الارتباط

قيمة معامل الارتباط	الدلالة
1	طردي تام
من 0.7 الى اقل من1	طردي قوي
من 0.5 الى اقل من 0.7	طردي متوسط
اقل من 0.5	طردي ضعيف
0	لايوجد ارتبط خطي

وما قيل عن الرتباط الطردي يقال عن العكسي اذا كانت الاشارة سالبة والخط التالي يوضح ذلك



معامل ارتباط بيرسون Bravais-Pearson

يعتبر معامل الارتباط بيرسون أحد أهم الاختبارات الإحصائية البارامترية التي تفحص قوة واتجاه العلاقة بين متغيرين بياناتهما كمية، وتتراوح قيمته بين (-1 و+1). ويستعمل هذا المعامل عند افتراض أن أي تغير في المتغير X يتبعه تغير في المتغير Y، سواء بالزيادة أو بالنقصان.

ويتطلب استعمال معامل بيرسون توفر الشروط التالية:

- أن تكون بيانات المتغيرين كمية.
- بما أنه من الاختبارات البارامترية، فينبغي أن تتوزع قيم المتغيرين توزيعا اعتداليا.
- أن تكون العلاقة بين المتغيرين خطية (ينصح تمثيل العلاقة بيانيا قبل حسابه).
- كما يشترط الكثير من الباحثين أن لا يقل حجم العينة عن 50 فردا (لضمان اقتراب توزيع البيانات من الاعتدالية).

هناك ثلاثة طرق لحساب معامل الارتباط بيرسون:

- من خلال الدرجات المعيارية.
- من خلال الانحرافات المعيارية.
- ومن خلال الدرجات الخام.

1 3 . حساب معامل بيرسون باستخدام الدرجات الخام:

- معادلة معامل بيرسون:

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

حيث: rp = رمز معامل الارتباط

n = حجم العينة

x et y = المتغيران

معامل ارتباط سبيرمان.

أن معامل الارتباط بيرسون يستعمل لمعرفة قوة العلاقة الخطية بين درجات متغيرين كميين، إلا أن تلك العلاقة بين المتغيرات في البحوث النفسية والتربوية لا تكون دائما خطية؛ فإذا أراد باحث مثلا التمييز بين ثلاثة مستويات للطموح لدى تلاميذ التعليم الثانوي، فإنه سيجد صعوبة في إعطاء بيانات رقمية لهذه المستويات، الأمر الذي سيحتم عليه ترتيبها حسب شدتها مثلا (مستوى طموح مرتفع – متوسط – منخفض). وهذه العملية ستقوده إلى اعتماد اختبارات إحصائية تعتمد على رتب المتغيرات وليس على قيمها الكمية.

يبرز من أهم وأفضل هذه الاختبارات الإحصائية معامل ارتباط شارل إدوارد سبيرمان (*Spearman Correlation Coefficient*) والذي يستعمل عندما لا تكون العلاقة بين المتغيرين المدروسين خطية.

يعتبر معامل ارتباط سبيرمان من الاختبارات الإحصائية اللابارامترية، ويطبق في الحالتين التاليتين:

- على العكس من معامل بيرسون، يستعمل معامل سبيرمان عندما لا يكون حجم العينة كبيرا (من 10 إلى 30 فردا).
- عندما تكون عملية تحويل البيانات الكمية إلى رتب ممكنة (ترتيب تصاعدي للبيانات)، أو عندما تكون البيانات المحصل عليها تتبع مستوى قياس ترتيبي.
- ويضيف باحثون أنه لا ينبغي أن يكون عدد الرتب المتكررة كبيرا.

23 - حساب معامل الارتباط سبيرمان:

- معادلة معامل سبيرمان:

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

حيث: rs: رمز معامل الارتباط

d: الفرق بين الرتب في كلا المتغيرين

n: حجم العينة.

لتوضيح حالات وطريقة حساب معامل سبيرمان، نورد الأمثلة الثلاثة التالية:

مثال: البيانات التالية تمثل عدد مرات التغيب (x) عن محاضرات مقياس الإحصاء والتحصيل (y) في هذا المقياس لدى عينة من 10 طلبة.

xy	y^2	x^2	y	x	n
30	9	100	3	10	1
12	144	1	12	1	2
15	1	225	1	15	3
32	64	16	8	4	4
21	49	9	7	3	5
20	100	4	10	2	6
15	225	1	15	1	7
36	36	36	6	6	8
30	4	225	2	15	9
28	361	4	19	2	10
249	**993**	**621**	**83**	**59**	**Σ**

وبالتعويض في المعادلة السابقة نحصل على:

$$r = \frac{10\,(249) - (59)(83)}{\sqrt{[10\,(621) - (59)^2][10\,(993) - (83)^2}}$$

$$r = -0.83$$

فالعلاقة بين المتغيرين قوية وعكسية سالبة، أي أنه كلما زاد التغيب عن الدروس انخفض مستوى التحصيل الدراسي والعكس صحيح.

- **مثال** : نريد معرفة قوة العلاقة بين العلامات التي تحصل عليها هؤلاء التلاميذ في كل من مادتي الرياضة (x) والموسيقى (y):

n	x	y	رتب x	رتبy	d	d^2
1	32	20	8	1	7	49
2	35	18	7	2	5	25
3	47	17	6	3	3	9
4	48	14	5	4	1	1
5	50	13	4	5	-1	1
6	53	10	3	6	-3	9
7	56	9	2	7	-5	25
8	60	5	1	8	-7	49
Σ						168

وبالتعويض في المعادلة السابقة، نحصل على قيمة معامل الارتباط التالية:

$$rs = -1$$

والتي تدل على وجود ارتباط عكسي تام بين المتغيرين.

- **مثال** : قد تكون بعض بيانات المتغيرين متكررة، ففي هذه الحالة نعطي الرتبة المتوسطة لكل القيم التي لها نفس القيمة، كما في المثال التالي الذي يتعلق بفحص الارتباط بين نتائج عينة من الطلبة في مقياسي الإحصاء والمنهجية:

n	x	y	رتب x	رتبy	d	d^2
1	75	72	7.5	1	6.5	42.25
2	75	55	7.5	4.5	3	9
3	76	60	6	2	4	16
4	77	58	4	3	1	1
5	78	54	2	6.5	-4.5	20.25
6	77	55	4	4.5	-0.5	0.25
7	80	54	1	6.5	-5.5	30.25
8	52	52	4	8	-4	16
Σ						135

الملاحظ أن هناك عدة قيم متكررة وأعطيت لها الرتبة المتوسطة من خلال قسمة رتبها المتتالية على عدد مرات تكرارها، وبتطبيق نفس المعادلة السابقة نحصل على قيمة الارتباط التالية بين هذين المتغيرين:

$rs = 0.61$ ارتباط موجب متوسط

- **مثال** : لا تكون البيانات في جميع الأحوال عبارة عن معطيات عددية، قد تكون أيضا بيانات وصفية، كما في المثال التالي الذي يمثل تقديرات تلاميذ في مادتي الاحصاء و الفيزياء

الجدول التالي يبين تقديرات عينة من الطلبة لمادتي الاحصاء(x) وتقديراتهم في مادة الفيزياء (y) . احسب معامل ارتباط سبيرمان. واذكر الدلالة

x	y	رتب x	رتب y	الفرق بين الرتب D	D^2
جيد	جيد	2	2	0	0
ممتاز	جيدجدا	5	3.5	1.5	2.25
جيدجدا	جيد جدا	3.5	3.5	0	0
جيدجدا	ممتاز	3.5	5	-1.5	2.25
ضعيف	ضعيف	1	1	0	0
				$\sum D^2$	4.5

$$r = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$r = 1 - \frac{6(4.5)}{5(25-1)} = .0.775$$

وهو ارتباط طردي قوى بين المتغيرات

مثال ليكن جدول المتغيرات التالي

X	12	14	16	12	13	17	12	15	19
Y	63	75	57	81	54	60	34	64	58

المطلوب هل هناك علاقة ارتباطية ذات دلالة إحصائية بين المتغيرين X و Y

1-تحديد الفرضيات:

الفرضية الصفرية : $H_0: r=0$ لا توجد علاقة ارتباطية بين المتغيرين

الفرضية البديلة : $H_1: r \neq 0$ توجد علاقة ارتباطية بين المتغيرين

2تحديد اتجاه الفرضية البديلة غير موجهة (أي بطرفين)

2 = 3/ (3+2+1)

X	12	12	12	13	14	15	16	17	19
	1	2	3	4	5	6	7	8	9
رتبةX	2	2	2	4	5	6	7	8	9

Y	63	75	57	81	54	60	34	64	58
	34	54	57	58	60	63	64	75	81
رتبةY	1	2	3	4	5	6	7	8	9

Y

الافراد	X	Y	رتبةX	رتبةY	الفرق d	d^2
1	12	63	2	6	-4	16
2	14	75	5	8	-3	9
3	**16**	57	7	3	4	16
4	12	81	2	9	-7	49
5	13	54	4	2	2	4
6	17	60	8	5	3	9
7	12	34	2	1	1	1
8	15	64	6	7	-1	1
9	19	58	9	4	5	25
n= 9						**$\sum d^2 = 130$**

3 حساب المحسوبة rs

rs=1-6(130)/9(81-1)=1-1.083 =**-0.083**

تحديد مستوي الدلالة α=0.05

درجة الحرية df=n=9 تحويل البينات الكمية الي بيانات رتيبة

حساب rs الجدولية

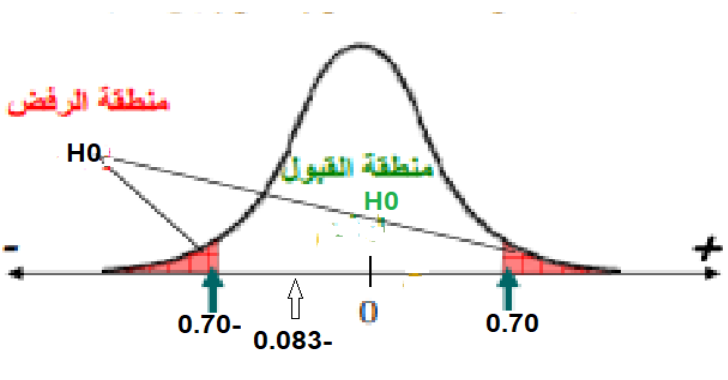
0.700 = rs (0.05,9)

4 اتخاذ القرار

بما ان rs = 0.083- المحسوبة اقل من rs = 0.700 الجدولية فإننا نقبل الفرضية الصفرية ونرفض الفرضية البديلة

5 النتيجة لا توجد علاقة ارتباطية ذات دلالة إحصائية بين المتغيرين

6 **التفسير** لا يوجد تفسير للعلاقة الارتباطية لين المتغيرين طالما ان الارتباط كان غير دال احصائيا



الانحدار الخطي

- خط انحدار التربيعات الصغرى

يوجد مؤشرات b و a حيث $y = ax + b + e$

مع ان e هو متغير عشوائي له متوسط معدوم ومتباينة ثابتة σ^2

عندما يكون لدينا سلسلة من الازواج

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n)$$

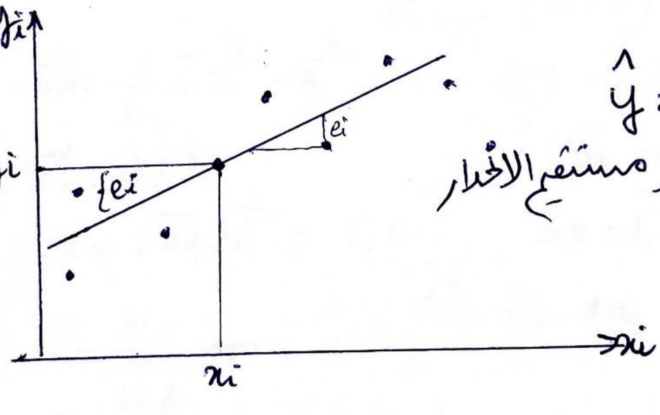
نستطيع أن نكتب كل ملاحظة $y_i = ax_i + b + e_i$

e_i هي عينة e وتسمى البواقي أو (residu)

تقدير المؤشران b و a باستعمال طريقة التربيعات الصغرى والتي نهدف بواسطتها إيجاد a و b التي تصغر مجموع التربيعات للبواقي e_i

$$e_i = y_i - ax_i - b$$

$$\sum e_i^2 = \sum (y_i - ax_i - b)^2$$



$$\hat{y} = ax + b$$

المستقيم المقدر أو مستقيم الانحدار

$$a = \frac{Cov(x, y)}{S_x^2}$$

وهذا المستقيم يمر دائماً وأبداً بالنقطة المتوسطة $(\bar{x}, \bar{y})$

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

$$b = \bar{y} - \frac{Cov(x,y)}{S_x^2} \cdot \bar{x}$$

مثال حسبنا بالنسبة لخمس أشجار مثمرة حجم المحصول من الثمار بالهيئات و بالنسبة للثمار الفاسدة بال%

- احسب المحصول المتوسط
- الانحراف المعياري للمحصول، ومعامل التباين
- البحث عن معادلة الانحدار الخطي $\hat{y} = ax + b$

x	y
8	59
6	58
11	56
22	51
14	50

$n = 5$

	x	y	x^2	y^2	$x \cdot y$
	8	59	64	3481	472
	6	58	36	3364	348
	11	56	121	3136	616
	22	51	484	2601	1122
	14	50	196	2500	700
Σ	61	274	901	15082	3258

الوسط الحسابي $\bar{x}$ و $\bar{y}$

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{5} \cdot 61$$

$$\bar{x} = 12{,}2$$

$$\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{5} \cdot 274$$

$$\bar{y} = 54{,}8$$

التباين S_x^2

$$S_x^2 = \frac{1}{n}(x_i - \bar{x})^2$$

$$S_x^2 = \frac{1}{n}\sum x^2 - \bar{x}^2 = \frac{901}{5} - (12{,}2)^2 = 31{,}36$$

$$S_x^2 = 31{,}36$$

الانحراف المعياري S_x

$$S_x = \sqrt{31{,}36} = 5{,}6$$

$$S_x = 5{,}6$$

معامل التباين

$$V = \frac{S_x}{\bar{x}} \cdot 100$$

$$V = \frac{5{,}6}{12{,}2} \cdot 100$$

$$V = 46\%$$

مستقيم الانحدار a

$$a = \frac{Cov(x,y)}{S_x^2} =$$

$$Cov(x,y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}$$

$$Cov(x,y) = \frac{3258}{5} - (12,2 \cdot 54,8)$$

$$Cov(x,y) = -16,96 < 0$$

المستقيم متناقص

$$a = -\frac{16,96}{31,36} = -0,541$$

$(12,2 , 54,8) = (\bar{x}, \bar{y})$ المستقيم يمر بالنقطة الوسطية

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

$$b = 54,8 - (-0,541) \cdot 12,2 = 54,8 + 6,588 = 61,4$$

$$\boxed{y = -0,541\, x + 61,4}$$

معادلة المستقيم المقدر
أو مستقيم الانحدار

الميل $a = -0,541$ اي سالب اذن المستقيم متناقص هذا
ما يفسر النسبة المئوية للثمار الفاسدة كبير جدا على الاشجار