

الارتباط والانحدار الخطي البسيط

Correlation and Simple Linear Regression

1- الارتباط الخطي

عرضنا في الفصول السابقة بعض المقاييس الإحصائية، مثل مقاييس النزعة المركزية والتشتت وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها من متغير واحد، ومنتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر وسنتناول في هذه المحاضرة دراسة وتحليل العلاقة بين متغيرين وذلك باستخدام بعض طرائق التحليل الإحصائي مثل تحليل الارتباط وذلك لدراسة العلاقة بين متغيرين (x, y) .

تعريف الارتباط:

الارتباط يعني وجود علاقة بين ظاهرتين أو متغيرين بمعنى أن التغير في أحد المتغيرين يؤدي إلى التغير في المتغير الآخر سواء بالزيادة أو بالنقصان. وإذا كان المتغيران يزيدان معاً أو ينقصان معاً فإن العلاقة بينهما طردية أما إذا كان أحدهما ينقص بزيادة المتغير الآخر فإن العلاقة بينهما عكسية.

مثال:

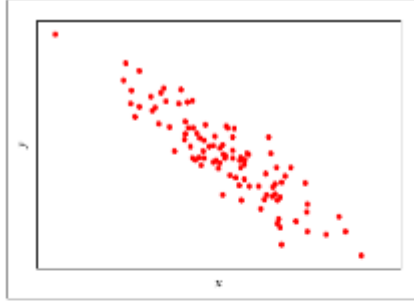
- دراسة العلاقة بين دخل الأسرة والمستوى التعليمي لرب الأسرة.
- دراسة العلاقة بين مستوى الإنتاجية وجودة المنتج في مصنع لإنتاج سلعة معينة
- دراسة العلاقة بين الكمية المطلوبة والسعر لسلعة معينة.
- دراسة العلاقة بين عمر الآلة وتكلفة صيانتها وهكذا ...

وأبسط الطرق لدراسة العلاقة بين المتغيرين أو الظاهرتين هو شكل الانتشار.

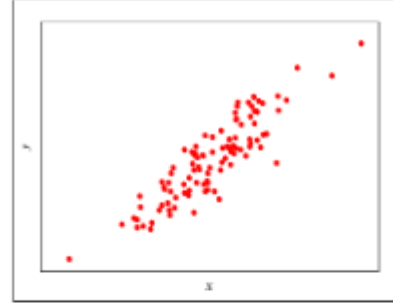
فإذا كان لدينا متغيرين وتم جمع بيانات عن أزواج قيم هذين المتغيرين فإنه يمكن تمثيلها بيانياً فيما يسمى بشكل الانتشار وتمثيل قيم الظاهرتين أو المتغيرين بشكل الانتشار بأخذ أشكال مختلفة على النحو الآتي:



جامعة
المنارة
MANARA UNIVERSITY



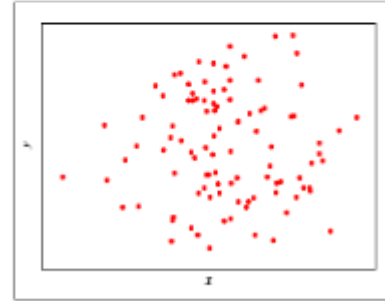
شكل الانتشار الخاص بالارتباط السالب
(العكسي)



شكل الانتشار الخاص بالارتباط
الموجب (الطردي)

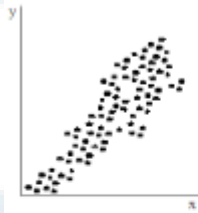


شكل الانتشار الخاص بالعلاقة الغير خطيه
بين متغيرين (ظاهرتين)

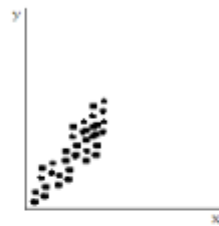


شكل الانتشار الخاص باستقلال
متغيرين (ظاهرتين)

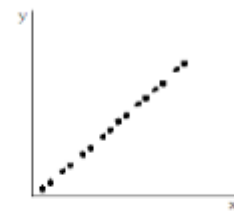
شكل الانتشار Scatter Plot



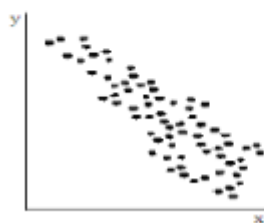
ارتباط طردي



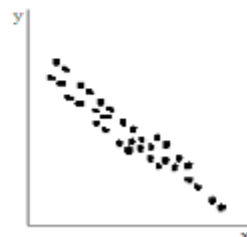
ارتباط طردي قوي



ارتباط طردي تام



ارتباط عكسي



ارتباط عكسي قوي



ارتباط عكسي تام

معامل الارتباط الخطي Correlation coefficient

معامل الارتباط هو مقياس كمي نسبي يستخدم لتحديد نوع وقوة العلاقة بين متغيرين ويرمز له

نوع العلاقة:

تأخذ ثلاثة صور مختلفة حسب إشارة معامل الارتباط كما يلي:

- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الآخر، والعكس.
- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الآخر، والعكس.

قوة العلاقة:

يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن الصفر، حيث إن قيمة معامل الارتباط تقع في المدى $(-1 \leq r \leq +1)$ ، وقد صنف بعض الاحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل الآتي:

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

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

معامل بيرسون للارتباط الخطي Pearson Linear Correlation Coefficient

- معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخداماً خاصة في العلوم الإنسانية والاجتماعية.
- ومستوى القياس المطلوب عند تطبيق معامل بيرسون للارتباط هو أن يكون كلا المتغيرين مقياس فترة أو نسبي أو بمعنى آخر أن تكون بيانات كلا المتغيرين (الظاهرتين) بيانات كمية.
- حساب معامل بيرسون للارتباط الخطي:

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين x و y باستخدام الصيغة الآتية:

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

مثال:

البيانات الآتية تمثل درجات 10 طلاب في مادتي الرياضيات والاحصاء:

درجات الرياضيات	80	80	70	60	90	90	90	70	80	90
درجات الاحصاء	60	70	70	50	90	80	80	50	90	60

والمطلوب: حساب معامل ارتباط بيرسون بين درجات المادتين:

الحل:

لإيجاد معامل ارتباط بيرسون نتبع الآتي:

- 1- نفرض أن x درجات الرياضيات، y درجات الإحصاء.
- 2- تكوين الجدول الآتي للحصول على المجاميع المطلوبة:

x	y	xy	x^2	y^2
80	60	4800	6400	3600
80	70	5600	6400	4900
70	70	4900	4900	4900
60	50	3000	3600	2500
90	90	8100	8100	8100
90	80	7200	8100	6400
90	80	7200	8100	6400
70	50	3500	4900	2500
80	90	7200	6400	8100
90	60	5400	8100	3600
$\sum = 800$	700	65900	65000	51000

3- حساب معامل ارتباط بيرسون بالمعادلة:

$$r_p = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} = \frac{10 \cdot (65900) - (800 \cdot 700)}{\sqrt{10 \cdot 65000 - (800)^2} \sqrt{10 \cdot 51000 - (700)^2}} = 0.636$$

2- معامل ارتباط الرتب (سبيرمان Spearman)

لحسب معامل ارتباط سبيرمان نتبع الآتي:

- 1- ترتيب القيم تصاعدياً أو تنازلياً
- 2- حساب رتب x ، رتب y . في حالة عدم وجود قيم مكررة يؤخذ الترتيب كرتبة أما في حالة وجود قيم مكررة نحسب الرتبة لكل قيمة مكررة كمتوسط لترتيب القيم المكررة، حيث رتبة القيم المكررة يساوي متوسط الترتيب لكل قيمة مكررة.
- 3- حساب: رتب y - رتب x $d = x$ مع ملاحظة أن $\sum d = 0$.
- 4- حساب معامل ارتباط الرتب (سبيرمان) بالمعادلة:

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

حيث n عدد أزواج القيم لـ (x, y) .

مثال:

البيانات الآتية توضح التقديرات التي حصل عليها 8 طلاب في مادتي الرياضيات والاحصاء:

تقديرات الإحصاء x	F	A	C	D	C	B	C	D
تقديرات الرياضيات y	D	C	B	F	D	A	D	F

والمطلوب: حساب الارتباط بين (تقديرات المادتين) ودرجة الارتباط.

الحل: البيانات وصفية وبالتالي نحسب معامل ارتباط الرتب سبيرمان بالخطوات الآتية:

1- تكوين الجدول الآتي للحصول على الرتب:

x	y	رتب x	رتب y	رتب y - رتب x d	
F	D	1	4	-3	9
A	C	8	6	2	4
C	B	5	7	-2	4
D	F	2.5	1.5	1	1
C	D	5	4	1	1
B	A	7	8	-1	1
C	D	5	4	1	1
D	F	2.5	1.5	1	1
				0	22

لعمل رتب x ورتب y يتم ترتيب التقديرات تنازلياً كما هو موضح بالجدول السابق ثم عمل رتب x ورتب y

رتبة التقديرات المكررة = متوسط الترتيب لكل تقدير مكرر.

2- حساب معامل ارتباط سبيرمان بين تقديرات المادتين بالمعادلة:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 22}{8(8^2 - 1)} = 0.74$$

أي أن الارتباط بين تقديرات المادتين طردي قوي.

ملاحظة:

- يستخدم معامل ارتباط بيرسون في حالة البيانات الكمية فقط
- يستخدم معامل ارتباط سبيرمان غالباً في حالة البيانات الوصفية كما يستخدم أيضاً في حالة البيانات الكمية
- معامل ارتباط بيرسون مقياس أكثر دقة معامل ارتباط سبيرمان يعتمد على ترتيب القيم وليس على لقيم الأصلية وبالتالي فهو مقياس تقريبي.

3 - معامل الاقتران (فاي)

معامل اقتران "فاي" يستخدم لقياس العلاقة بين متغيرين اسميين كليهما ثنائي التقسيم، كالنوع (ذكر\ أنثى) والإصابة بالأمراض (مصاب \ غير مصاب) إلخ

	X_1	X_2	المجموع
Y_1	a	b	a+b
Y_2	c	d	c+d
المجموع	a+c	b+d	

معامل فاي للاقتران يعطى في الصورة الآتية:

$$r_\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

مثال

أوجد قيمة معامل الارتباط بين النوع x (ذكر/ أنثى) والإصابة بمرض الاكتئاب Y (مصاب/ غير مصاب) حسب البيانات التالية :

النوع \ الاكتئاب	مصاب	غير مصاب
ذكر	12	8
أنثى	4	6

الحل :

نوجد أولاً المجاميع الهامشية كما في الجدول التالي :

النوع \ الاكتئاب	مصاب	غير مصاب	المجموع
ذكر	12	8	20
أنثى	4	6	10
المجموع	16	14	30

وعليه فإن :

$$\begin{aligned} a &= 12 \\ b &= 8 \\ c &= 4 \\ d &= 6 \end{aligned}$$

$$r_{\phi} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{12 \times 6 - 8 \times 4}{\sqrt{20 \times 10 \times 16 \times 14}} \approx 0.19$$

أي أنه توجد علاقة **ضعيفة** بين النوع والإصابة بمرض الاكتئاب.

3- الانحدار Regression

الانحدار هو أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار ، وله أنواع:

- **الانحدار الخطي البسيط:** فكلمة "بسيط" تعني أن المتغير التابع Y يعتمد على متغير مستقل واحد وهو X وكلمة "خطي" تعني أن العلاقة بين المتغيرين (X, Y) علاقة خطية.
- **الانحدار المتعدد:** إذا كان المتغير Y يعتمد على أكثر من متغير مستقل.
- **الانحدار غير الخطي:** إذا كانت العلاقة بين المتغير Y والمتغيرات المستقلة غير خطية كأن تكون من الدرجة الثانية أو أسية.

الانحدار الخطي البسيط

بعد تمثيل الأزواج المرتبة بالمستوي نحصل على شكل الانتشار فإذا أظهر الشكل الانتشاري للبيانات أن هناك علاقة خطية بين المتغيرين نقوم بتقدير خط الانحدار Y على X بواسطة العلاقة:

$$\hat{y} = a + bx$$

حيث a : ثابت الانحدار أو الجزء المقطوع من محور y .
 b : ميل الخط المستقيم أو معامل انحدار Y على X (أو Y/X).

وتحسب القيمتان a و b من العلاقتين التاليتين :

$$a = \frac{\sum y - b \sum x}{n} , b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

لإيجاد قيمة مقدرة جديدة \hat{y}_h نعوض قيمة معلومة للمتغير المستقل ولتكن x_h في معادلة تقدير خط الانحدار Y/X

$$\hat{y} = a + bx$$

ملاحظات هامة:

- ❖ ميل الخط يمثل كمية التغير في Y المناظرة للتغير في X بمقدار وحدة واحدة.
- ❖ إشارة معامل الانحدار تدل على نوع الارتباط (طردي أو عكسي).
- ❖ توجد علاقة بين معامل الانحدار ومعامل الارتباط الخطي.

مثال:

لدراسة العلاقة بين الكمية الكبرى للأكسجين المستنشق X و معدل ضربات القلب Y ومن أجل عشرة أشخاص وجدنا ما يأتي:

x	43	49	50	12	8	32	51	30	35	23
y	175	180	186	95	75	165	190	95	130	95

أوجد معادلة الانحدار الخطي البسيط وتوقع قيمة ضربات قلب شخص إذا علمت أن كمية الأكسجين المستنشقة هي $x = 30$.

الحل:

x	43	49	50	12	8	32	51	30	35	23	$\sum x$ = 333
y	175	180	186	95	75	165	190	95	130	95	$\sum y$ = 1386
$x.y$	7525	8820	9300	1140	600	5280	9690	2850	4550	2185	$\sum xy$ = 51940
x^2	1849	2401	2500	144	64	1024	2601	900	1225	529	$\sum x^2$ = 13237

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} = \frac{10 \times 51940 - 333 \times 1386}{10 \times 13237 - (333)^2} \approx 2.694$$

$$a = \frac{\sum y - b \sum x}{n} = \frac{1386 - (2.694) \times 333}{10} \approx 48.9$$

معادلة خط الانحدار البسيط في هذه الحالة:

$$\hat{y} = 48.9 + 2.694 x$$

ولتوقع قيمة ضربات قلب شخص استنشق كمية أكسجين قدرها $x_h = 30$ نعوض في المعادلة السابقة فنجد:

$$\hat{y}_h = 48.9 + 2.694 \times 30 = 129.7$$

أي عندما يستنشق أحد الأشخاص كمية من الأوكسجين وقدرها 30 نتوقع أن تكون ضربات القلب تساوي 130 ضربة بالدقيقة تقريباً.