4

الباب الرابع: الارتباط والانحدار الخطي البسيط

Chapter 4: Correlation &Simple Linear Regression

سنتناول في هذا الفصل :

(١) مفهوم الإرتباط وأنواعه.

(٢) طرق حساب معاملات الارتباط المختلفة.

(٣) مفهوم الانحدار الخطي البسيط .

مقدمة عن الارتباط

تقابلنا كثيرا في الحياة العملية مواقف تتضمن متغيرين(ظاهرتين) وأكثر ويكون المطلوب معرفة ما إذا كان هناك علاقة بين هذه المتغيرات وما هو شكل هذه العلاقة ؟ وأيضا كيفية التنبؤ بأحد هذين المتغيرين في حالة معرفتنا بالمتغير الآخر .

فكثيرا ما تجدين في بعض المجلات معادلة الطول مع الوزن فإذا أردت أن تعرفي الوزن المثالي أدخلي طولك في المعادلة ليظهر وزنك المثالي ، وقد توصلوا إلى هذه المعادلة أو إلى هذه الصيغة بدراسة العلاقة ما بين المتغيرين الطول والوزن على مجموعة من الأفراد.

الارتباط

الارتباط: هو تعيين طبيعة وقوة العلاقة بين متغيرين أو عدمها

- معامل الارتباط هو مؤشر هذه العلاقه
- أول خطوه فى تحديد طبيعه العلاقه بين متغيرين هي رسم شكل الانتشار

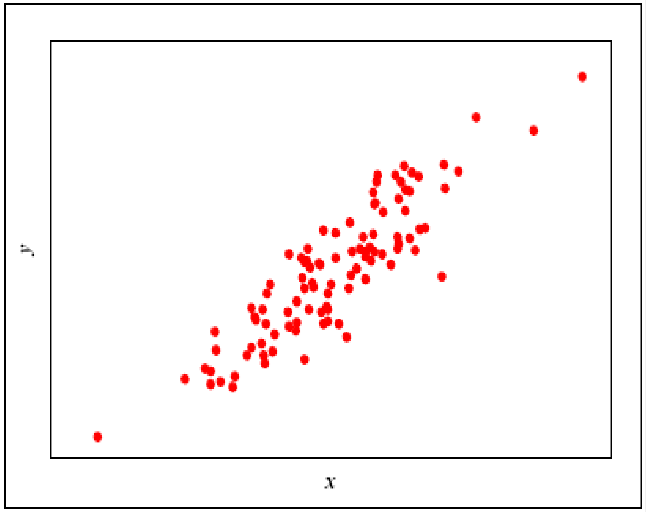
- إذا كان لدينا متغيران فقط . المتغير X والذي يسمى بالمتغير المستقل Independent variable
- يرافق المتغيرX متغير آخر Yويسمى بالمتغير التابع dependent variable

الارتباط

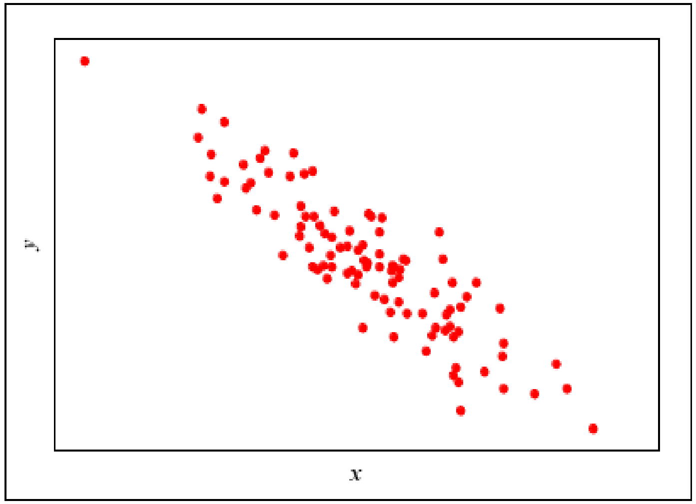
أنواع الارتباط

الارتباط الموجب (الطردي) (Positive Correlation) بأنه علاقة بين متغيرين (x, y) بحيث إذا تغير أحد المتغيرين فإن الآخر يتبعه في نفس الاتجاه..

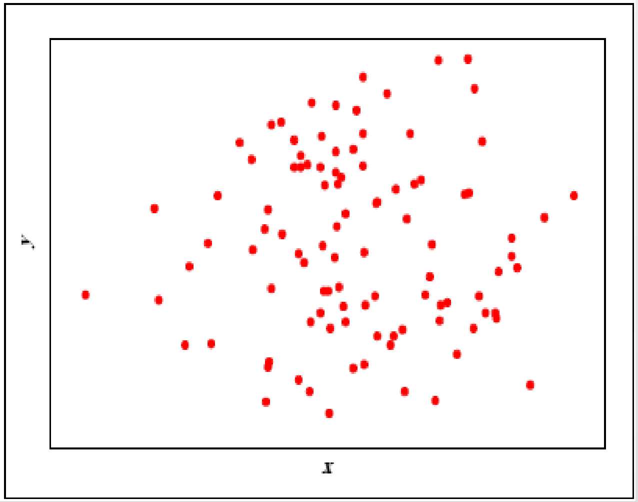
الارتباط السالب (العكسي) (Negative Correlation) بأنه علاقة بين متغيرين (x, y) بحيث إذا تغير أحد المتغيرين فإن الآخر يتبعه في الاتجاه المضاد.



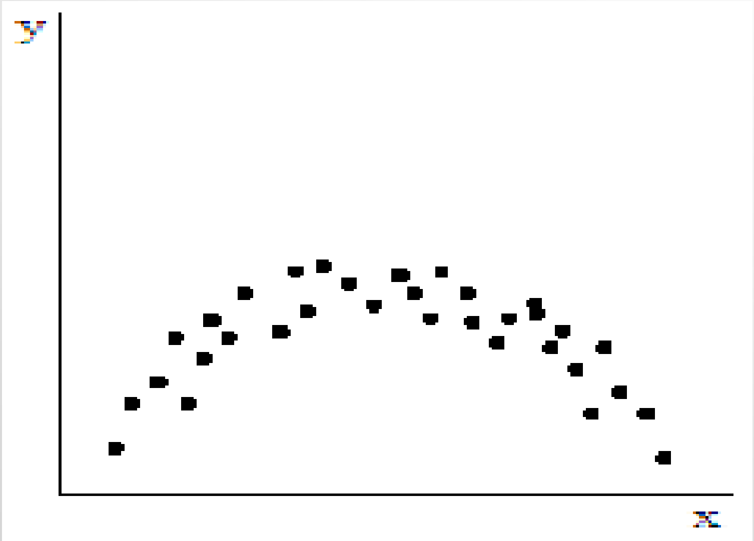
شكل الانتشار الخاص بالارتباط
الموجب (الطردي)



شكل الانتشار الخاص بالارتباط السالب
(العكسي)

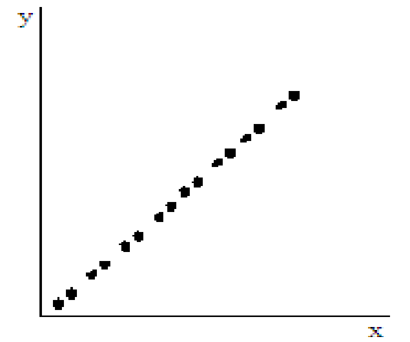


شكل الانتشار الخاص باستقلال
متغيرين (ظاهرتين)

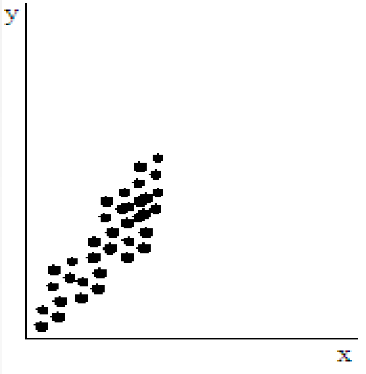


شكل الانتشار الخاص بالعلاقه الغير خطيه
بين متغيرين (ظاهرتين)

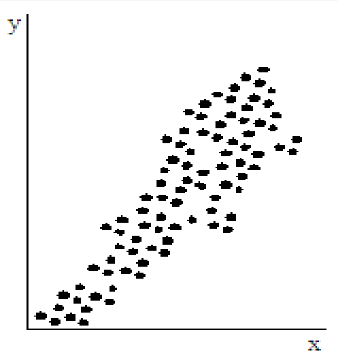
شكل الانتشار



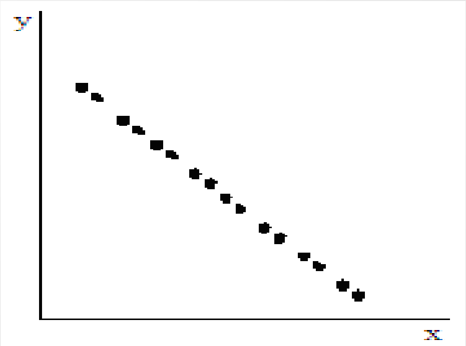
ارتباط طردي تام



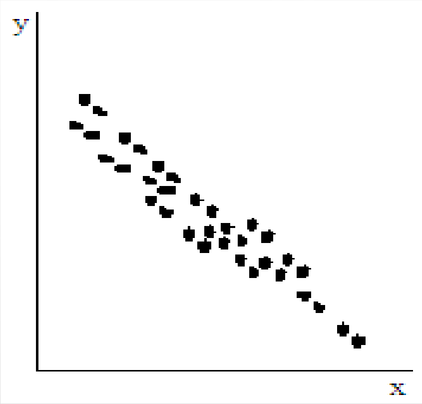
ارتباط طردي قوي



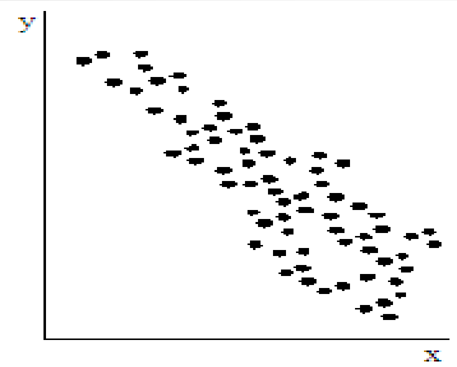
ارتباط طردي متوسط



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي متوسط

قياس الارتباط

- تستخدم معاملات الارتباط لقياس درجة الارتباط بين متغيرين(ظاهرتين) .
- <u>تعريف معامل الارتباط :</u>

يعرف معامل الارتباط والذي يرمز له بالرمز r بأنه عبارة عن مقياس رقمي يقيس قوة ونوع الارتباط بين متغيرين ، حيث أن

$$-1 \leq r \leq +1$$

وتدل إشارة المعامل الموجبة على العلاقة الطردية ،

بينما تدل إشارة المعامل السالبة على العلاقة العكسية .

والخط التالي يوضح أنواع الارتباط :

تام	قوي	متوسط	ضعيف	منعدم
١	٠,٧	٠,٥		٠

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

معامل بيرسون للارتباط الخطي

- عند تطبيق معامل بيرسون للارتباط يجب أن يكون كلا المتغيرين(y،x) بيانات كمية وذلك باستخدام الصيغة التالية:

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

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

- **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 :**

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

	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
$\sum$	$15 = \sum x$	$9 = \sum y$	$24 = \sum xy$	$41 = \sum x^2$	$15 = \sum y^2$

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

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

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

2 - معامل سبيرمان لارتباط الرتب

- نستخدم معامل سيبرمان لارتباط الرتب (Rank Correlation coefficient) اذا كانت البيانات وصفية ترتيبية او كمية.
- <u>طريقة حساب معامل سبيرمان لارتباط الرتب :</u>
- إذا فرضنا أن المتغير X له الرتب R_x وأن المتغير Y له الرتب R_y. وبفرض
- أن d ترمز لفرق الرتبتين، بمعنى $d = R_x - R_y$. فإن معامل سبيرمان لارتباط الرتب يعطى بالصيغة التالية:

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

حيث n هي عدد الأزواج المرتبة .

- **مثال :**
- **لدراسة علاقة ارتباط تقديرات الطلاب في مادة الإحصاء وتقديراتهم في مادة الرياضيات، اخترنا خمس طلاب وكانت تقديراتهم كما يلي :**

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

هل توجد علاقة ارتباط؟ ما نوعها ومدى قوتها؟

- **الحل:**

x	y	رتب x	رتب y	d	d^2
F	D	1	2	-1	1
A	C	5	3	2	4
C	B	3	4	-1	1
D	F	2	1	1	1
B	A	4	5	-1	1
			$\sum$	0	8
				$\sum d$	$\sum d^2$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{(6)(8)}{5(25-1)} = 1 - \frac{48}{120} = 1 - 0.4 = 0.6$$

نلاحظ وجود علاقة ارتباط طردية متوسطة بين تقديرات الطلاب في مادة الإحصاء وتقديراتهم في مادة الرياضيات.

- **مثال :**

عند تقييم مجموعة من الناقدين الرياضيين لعدد 10 من اللاعبين تبعاً للحمل التدريبي قبل المسابقة وترتيب هؤلاء اللاعبين بعد المسابقة كان الترتيب التالي :

اللاعب	A	B	C	D	E	F	G	H	I	J
رتبة الحمل التدريبي	5	9	10	2	8	7	4	1	6	3
رتبة اللاعب النهائية	4	8	10	2	9	6	3	1	7	5

فاحسب معامل الارتباط لدراسة العلاقة بين الحمل التدريبي والترتيب النهائي.

- **الحل:**

اللاعب	رتبة الحمل التدريبي (R_x)	رتبة الترتيب (R_y)	$d = R_x - R_y$	d^2
A	5	4	+1	1
B	9	8	+1	1
C	10	10	0	0
D	2	2	0	0
E	8	9	-1	1
F	7	6	+1	1
G	4	3	+1	1
H	1	1	0	0
I	6	7	-1	1
J	3	5	-2	4
				$\sum d^2 = 10$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{(6)(10)}{10(99)} = 1 - \frac{60}{990} = 1 - 0.06 = 0.94$$

هذا الارتباط طردي قوى، بمعنى أنه كلما زاد الحمل التدريبي كلما تم الحصول على ترتيب متقدم.

3 - معامل الاقتران (فاي)

- **معامل اقتران "فاى" يستخدم لقياس العلاقة بين متغيرين اسميين كل منهما ثنائي التقسيم، كالنوع (ذكر/أنثى) والإصابة بالمرض (مصاب/غير مصاب) والتدخين (مدخن/غير مدخن)...الخ.**

المجموع	X_2	X_1	X \ Y
a+b	b	a	Y_1
c+d	d	c	Y_2
	b+d	a+c	المجموع

معامل فاي للاقتران يعطى في الصورة التالية :

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

مثال :

أوجد قيمة معامل الاقتران بين النوع x (ذكر/ أنثى) والإصابة بمرض الاكتئاب Y

(مصاب/ غير مصاب) حسب البيانات التالية :

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

الحل :

نوجد أولاً **المجاميع الهامشية** كما في الجدول

الاكتئاب / النوع	مصاب	غير مصاب	المجموع
ذكر	12 a	8 b	20
أنثى	4 c	6 d	10
المجموع	16	14	30

وعليه فإن :

a= 12

b = 8

c = 4

d = 6

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

$$= \frac{72-32}{\sqrt{44800}} = \frac{40}{211.66} \approx 0.19$$

أي أنه توجد علاقة ضعيفة بين النوع والإصابة بمرض الاكتئاب .

الانحدار

- *والانحدار* هو أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار

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

- 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 الخطي البسيط

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

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

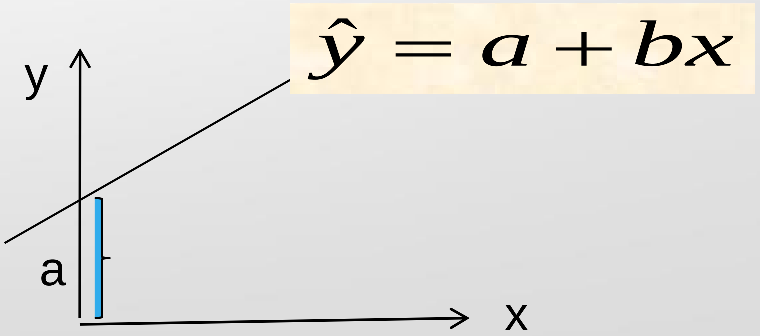
حيث a **: ثابت الانحدار أو الجزء المقطوع من محور y**

b **: ميل الخط المستقيم أو <u>معامل الانحدار</u>**

- **وتحسب القيمتان** a **و** b **من العلاقتين التاليتين:**

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

$$a = \frac{\sum y - b\sum x}{n}$$



الانحدار الخطي البسيط

ملاحظات مهمة:

- إشارة معامل الانحدارb تدل على نوع الارتباط(طردي أو عكسي)
- لإيجاد قيمة مقدرة جديدة $\hat{y}$ نعوض بقيمة معلومة للمتغير المستقل ولتكن x في معادلة تقدير خط الانحدار

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

- مثال :

لدراسة علاقة الاستهلاك المحلي (y) بالإنتاج (x) لمادة الإسفلت (بالمليون برميل) خلال عدة سنوات، أخذنا عشر قراءات تقريبية كما يلي :

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

أوجدي معادلة الانحدار الخطي البسيط، وتوقعي قيمة الاستهلاك عندما يصل إنتاج ١١ مليون برميل .

- الحل :

	x	y	x y	x^2
	10	6	60	100
	13	8	104	169
	15	9	135	225
	14	8	112	196
	9	7	63	81
	7	6	42	49
	6	5	30	36
	6	6	36	36
	5	5	25	25
	5	5	25	25
$\sum$	90	65	632	942
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$

$$b=\frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}=\frac{6320-(90)(65)}{9420-90^2}=\frac{6320-5850}{9420-8100}=\frac{470}{1320}=0.36$$

$$a=\frac{\sum y-b\sum x}{n}=\frac{65-(0.36\times 90)}{10}=3.26$$

:. معادلة خط الانحدار البسيط في هذه الحالة :

$$\hat{y}=3.26+0.36x$$

التقدير:

- ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج ١١ مليون برميل،

$$x = 11$$

- وبالتعويض في المعادلة السابقة نجد أن:

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

$$= 3.26 + (0.36 \times 11) = 7.22$$

أي أن الاستهلاك قد يصل إلى 7.22 مليون برميل خلال السنة.

تطبيق الانحدار في مجال السلاسل الزمنية

- أحد طرق تعيين الاتجاه العام للسلسلة الزمنية هو استخدام أسلوب الانحدار الخطي البسيط، باعتبار **أن الزمن (السنوات، الشهور،...الخ) متغير مستقل X، والمتغير التابع Y هو الظاهرة محل الدراسة.**
- ملاحظات:

- نعين للمتغير المستقل القيم ... ,x = 0, 1, 2 لتمثل وحدة الزمن.

- **مثال :**

البيانات التالية تمثل عدد الحقول المكتشفة (Y) خلال الأعوام 1991م إلى 2000م :

السنة	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
y	62	63	67	69	70	75	79	82	84	86

قدري معادلة الاتجاه العام الخطي، ثم توقعي عدد الحقول المكتشفة عام 2002م.

- **الحل:**

$$b=\frac{n\sum xy-(\sum x)(\sum y)}{\sqrt{(n\sum x^2-(\sum x)^2)}}=\frac{35530-(45*737)}{2850-45^2}=\frac{2365}{825}=2.87$$

$$a=\frac{\sum y-b\sum x}{n}=\frac{737-(2.87\times 45)}{10}=60.79$$

$$\hat{y}=60.79+2.87x$$

السنة	x	y	xy	x^2
1991	0	62	0	0
1992	1	63	63	1
1993	2	67	134	4
1994	3	69	207	9
1995	4	70	280	16
1996	5	75	375	25
1997	6	79	474	36
1998	7	82	574	49
1999	8	84	672	64
2000	9	86	774	81
$\sum$	45	737	3553	285
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$

- **ولتوقع عدد الحقول المتوقع اكتشافها عام 2002م**
- **حيث أن عام 2000 ← $x = 9$**

إذن عام 2002 ← $x_h = 11$

- **وبالتعويض في معادلة الاتجاه العام نجد أن:**

$$\hat{y}_h = 60.79 + 2.87 x_h$$

$$= 60.79 + 2.87(11) = 92.36 \approx 92 \text{ حقل}$$