4

الباب الرابع: الارتباط والانحدار الخطي البسيط

Chapter 4: Correlation & Simple Linear Regression

سنتناول في هذا الفصل :

(1) مفهوم الإرتباط وأنواعه.

(2) طرق حساب معاملات الارتباط المختلفة.

(3) مفهوم الانحدار الخطي البسيط وتطبيقاته .

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 متغيري الطول والوزن على مجموعة من الأفراد.

الارتباط

الارتباط : هو تعيين طبيعة وقوة العلاقة بين متغيرين أو عدمها.

- **معامل المشاركة** هو عبارة عن مقياس رقمي يقيس قوة العلاقة بين متغيرين أحدهما على الأقل له مستوى قياس اسمي حيث تتراوح قيمته بين (0) و (1) وهو يعكس مدى اقتران أو ارتباط المتغيرين ببعضهما.
- **معامل الارتباط** هو مؤشر هذه العلاقة
- أول خطوة فى تحديد طبيعة العلاقة بين متغيرين هي رسم شكل الانتشار
- إذا كان لدينا متغيران فقط . المتغير X وهو متغير يتم تحديده من قبل الباحث أو الشخص الذي يقوم بالدراسة وهو يسمى بالمتغير المستقل Independent variable
- يرافق المتغيرX متغير آخر Yويسمى بالمتغير التابع dependent variable وهو متغيرإحصائي لأن نتيجته غير محددة وتعتمد على قيم المتغير المستقل.

الارتباط

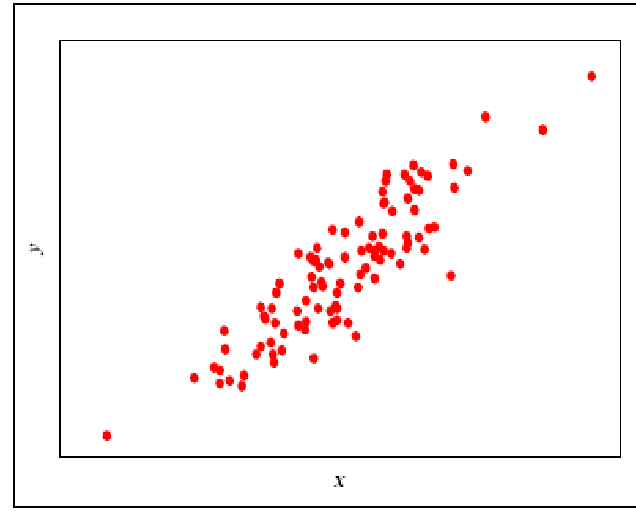
أنواع الارتباط

الارتباط الموجب (الطردي) (Positive Correlation) :

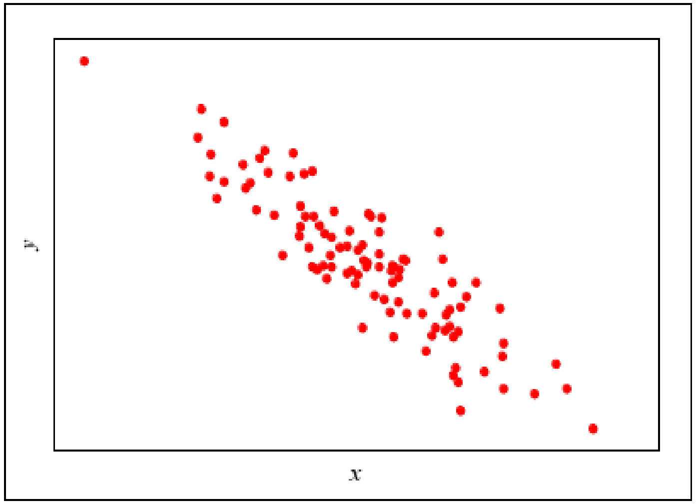
هنا العلاقة بين المتغيرين *(x, y)*تكون أنه إذا تغير أحد المتغيرين فإن الآخر يتبعه في *نفس الاتجاه.*

الارتباط السالب (العكسي) (Negative Correlation) :

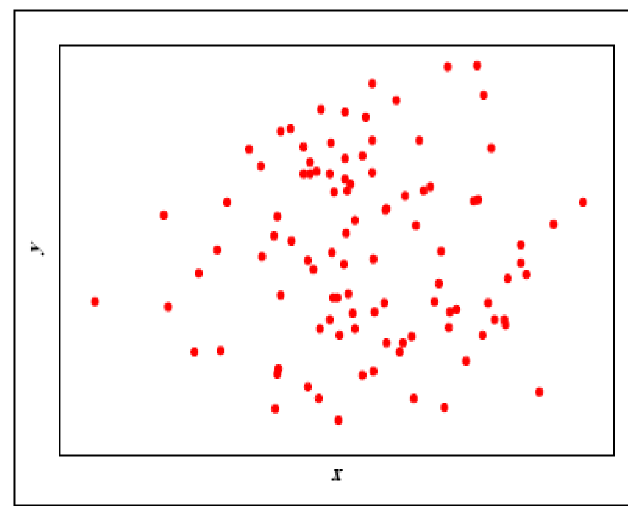
هنا العلاقة بين المتغيرين *(x, y)*تكون أنه إذا تغير أحد المتغيرين فإن الآخر يتغير في *الاتجاه المعاكس.*



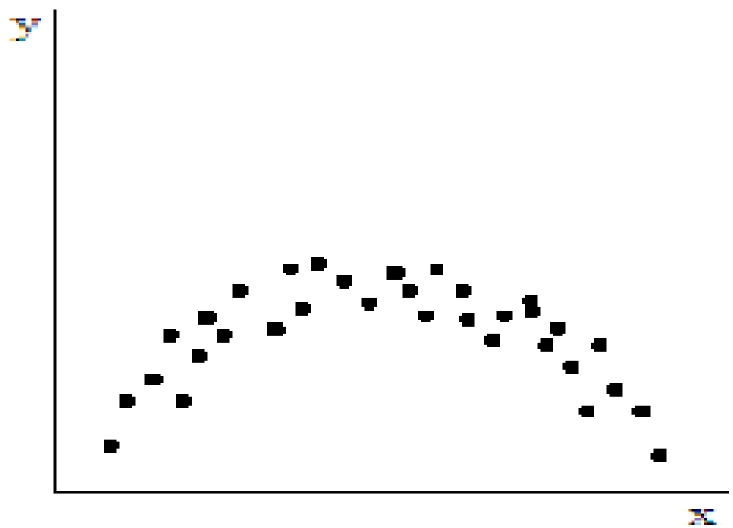
شكل الانتشار الخاص بالارتباط الموجب (الطردي)



شكل الانتشار الخاص بالارتباط السالب (العكسي)

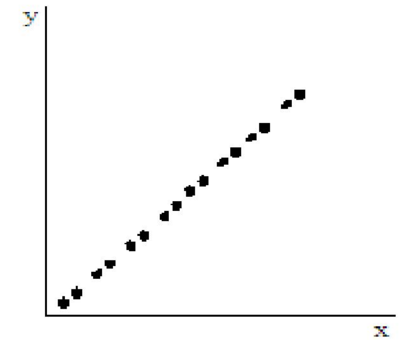


شكل الانتشار الخاص باستقلال متغيرين (ظاهرتين)

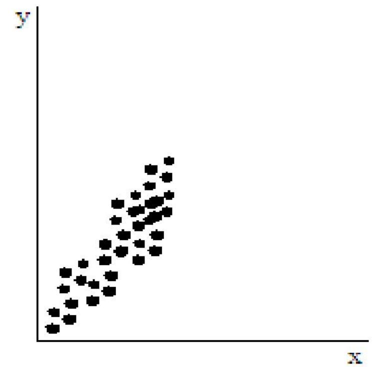


شكل الانتشار الخاص بالعلاقه الغير خطيه بين متغيرين (ظاهرتين)

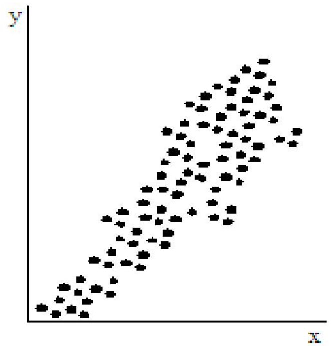
شكل الانتشار



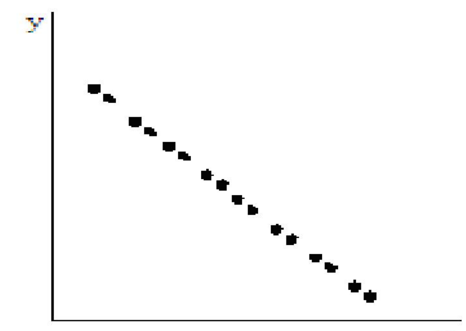
ارتباط طردي تام



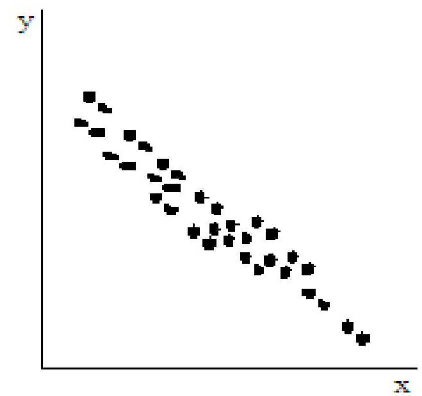
ارتباط طردي قوي



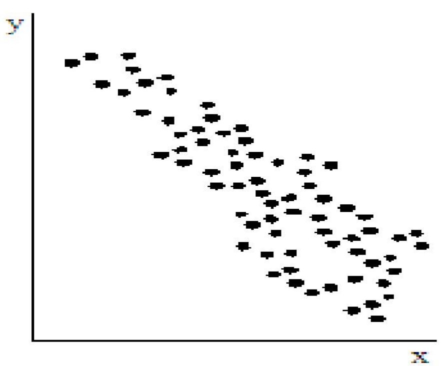
ارتباط طردي متوسط



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي متوسط

قياس الارتباط

- تستخدم معاملات الارتباط لقياس درجة الارتباط بين متغيرين(ظاهرتين) .
- 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 $-1 \leq r \leq +1$.

وتدل إشارة المعامل الموجبة على العلاقة الطردية ،

بينما تدل إشارة المعامل السالبة على العلاقة العكسية .

قياس الارتباط

الجدول 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.

١- معامل بيرسون للارتباط الخطي

- حساب معامل بيرسون للارتباط الخطي :

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين y, x 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.

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

	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 $=\sum y^2$

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

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

يتضح أن العلاقة بين حجم الإنتاج وحجم صادرات النفط الخام <u>علاقة طردية متوسطة</u>.

2 – معامل سبيرمان لارتباط الرتب

- نستخدم معامل سيبرمان لارتباط الرتب **(Rank Correlation coefficient)** إذا كان المتغيرين كليهما وصفي ترتيبي أو كليهما متغير كمي.

- **طريقة حساب معامل سبيرمان لارتباط الرتب :**

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 يُعطى 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
			Σ	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 قوي، بمعنى أنه كلما زاد الحمل التدريبي كلما تم الحصول على ترتيب متقدم.

ملاحظات هامة :

- ومما سبق يتضح أن معامل ارتباط الرتب يمكن حسابه سواءً أكانت البيانات كمية أو وصفية ترتيبية بينما معامل الارتباط (بيرسون) لا يمكن حسابه إلا على المتغيرات الكمية..
- يتميز معامل سيبرمان لارتباط الرتب بسهولة حسابه حتى لو كانت البيانات غير مرتبة.
- يعاب على معامل سيبرمان إهماله لفروق الأعداد عند حساب الرتب وبالتالي فهو أقل دقة.

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.

X / Y	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 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 التالي :

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

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

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

- الانحدار الخطي البسيط :

كلمة " 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:

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

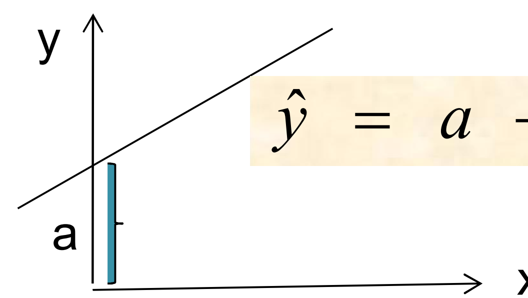
حيث a : ثابت الانحدار أو الجزء المقطوع من محور y

b : ميل الخط المستقيم أو معامل انحدار

- وتحسب القيمتان a و b من العلاقتين التاليتين:

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

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



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

الانحدار الخطي البسيط

ملاحظات مهمة:

- إشارة معامل الانحدار b 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

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

نعوض

- 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 الإنتاج 16 مليون برميل .

- الحل :

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

	x	y	xy	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
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$

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

∴ معادلة خط الانحدار البسيط في هذه الحالة :

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

تابع حل المثال:

- ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج **16 مليون برميل**، نعوض في المعادلة السابقة نجد أن:

$$x = 16$$

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

$$= 3.26 + (0.36 \times 16) = 9.02$$

أي أن الاستهلاك قد يصل إلى **9.02** 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 Yهو الظاهرة محل الدراسة.

- ملاحظات:

- نعين للمتغير المستقل x القيم ...,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: تقدير معادلة الاتجاه العام الخطي

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
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$

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

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

معادلة الاتجاه العام الخطي في هذا المثال

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

تابع حل المثال (حساب التوقع):

- ولتوقع عدد الحقول المتوقع اكتشافها عام 2002م

حيث أن 2000م ← x = 9

إذن 2002م ← $x_h = 11$

- وبالتعويض في معادلة الاتجاه العام نجد أن:

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

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

اختبار ذاتى

1-6-3 اختاري الإجابة المناسبة للفقرات التالية :

1-العلاقه بين متغيرين (x,y) بحيث إذا تغير أحد المتغير فإن الاخر يتبعه فى نفس الاتجاه هى علاقه

A. ارتباط طردى (إرتباط سالب)	B. ارتباط عكسى (إرتباط سالب)	C. إرتباط طردى (إرتباط موجب)	D. إرتباط عكسى (إرتباط موجب)

2- تتراوح قيمة معامل الارتباط (r) بين

A. $-1 \le r \le 1$	B. $-1 \le r \le 2$	C. $-2 \le r \le 1$	D. $-2 \le r \le 2$

3- لدراسه العلاقه الارتباط بين التدخين والتعليم عند الشباب أخذنا عينه من ألف شاب وقسمناهم حسب التدخين (مدخن أو غير مدخن) والتعليم (متعلم أو أمى) فإن أنسب معامل إرتباط هو معامل

A. الارتباط الخطى (بيرسون)	B. إرتباط الرتب (سبيرمان)	C. بوينت بايسيريال	D. الاقتران (فاى)

اختبار ذاتى

4-7-2 لدراسه الارتباط بين درجات الطلاب فى مادتى الاحصاء والرياضيات أخذت عينه من ست طلاب وكانت نتائجهم كالتالى:

درجات الاحصاء(x)	80	90	70	65	60	50
درجات الرياضيات(y)	95	70	85	65	60	45

4 ـ معامل الارتباط الخطى (بيرسون) بين درجات الطلاب فى ماده الاحصاء ودرجاتهم الرياضيات يساوى

إذا علمت أن $\sum y^2 = 31000$ $\sum x^2 = 29725$ $\sum y = 420$ $\sum x = 415$

A. 0.69	B.0.96	C. 0.75	D. 0.37

5- معامل إرتباط الرتب (سبيرمان) بين درجات الطلاب فى ماده الاحصاء ودرجاتهم فى الرياضيات يساوى

A. 0.73	B. 0.63	C. 0.93	D. 0.83

اختبار ذاتى

4-7-3 إذا أعطيت البيانات التاليه عن الدخل بمئات الريالات (x) وقيمه الاستهلاك بمئات الريالات- (y) بمئات الريالات لخمسه أشخاص :

$\sum xy = 512$ $\sum x = 68$ $\sum y = 37$ $\sum x^2 = 990$ $\sum y^2 = 277$ $n = 5$

6- معامل انحدار (b) الدخل على الاستهلاك يساوي

A. 1	B.0. 25	C. 0.13	D. 0.0

7- ثابت الانحدار a يساوي

A. 1.29	B. 5.56	C. 20.52	D. 10.95

8- تقدير الاستهلاك عندما يصل الدخل إلى 100 ريال

A. 16.33	B. 15.09	C. 21.96	D. 5.76

اختبار ذاتى

4-7-4 إذا كانت البيانات التالية تمثل كميه صادرات البترول (y) بالمليون برميل لأحدى الشركات فى أربعه أعوام من 2006م إلى 2009 م كالتالى:

$\sum xy = 512$ $\sum x = 6$ $\sum y = 320$ $\sum x^2 = 14$ $n = 4$

9- قيمه معامل الانحدار تساوي

A. 1.05	B. 12. 25	C. 10.85	D. 6.4

10- إذا علمت أن قيمه ثابت الانحدارتساوي a=70.4 فإن تقدير قيمه صادرات البترول عام 2011 م

A. 40.5	B. 92.5	C. 102. 4	D. 150.85