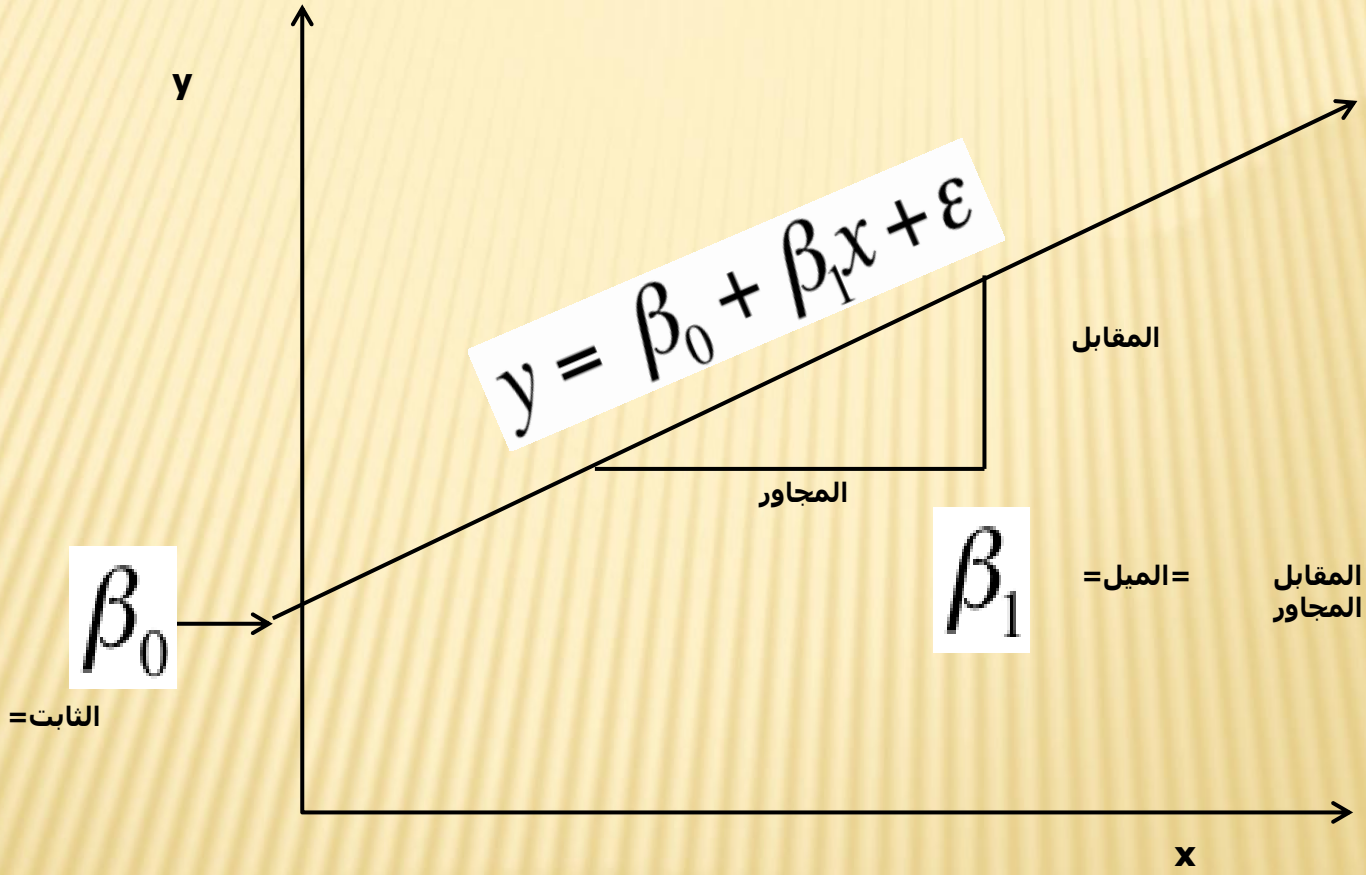
معامل الارتباط والانحدار الخطي

THE CORRELATION COEFFICIENT &THE LINEAR REGRESSION

الاستاذ المساعد
الدكتـــــور احمد الحسيني
ماجستير علوم مالية ومصرفية
2023-2022

معامل الارتباط وخط الانحدار



معامل الارتباط

مقدمة

يقصد بمعامل الارتباط Correlation Coefficient تعيين طبيعة وقوة العلاقة بين متغيرين أو أكثر، بمعني ايجاد مقدار العلاقة بين المتغيرات سواء أكانت سلبية أم إيجابية، وسوف تقتصر دراستنا على الارتباط بين متغيرين (ظاهرتين) فقط. ويمكن قياس العلاقة بين ظاهرتين بطرق مختلفة منها:

طريقة شكل الانتشار Scatter Diagram: وسيلة مبدئية تساعد في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية، تصلح إذا كان المتغيران كميين.

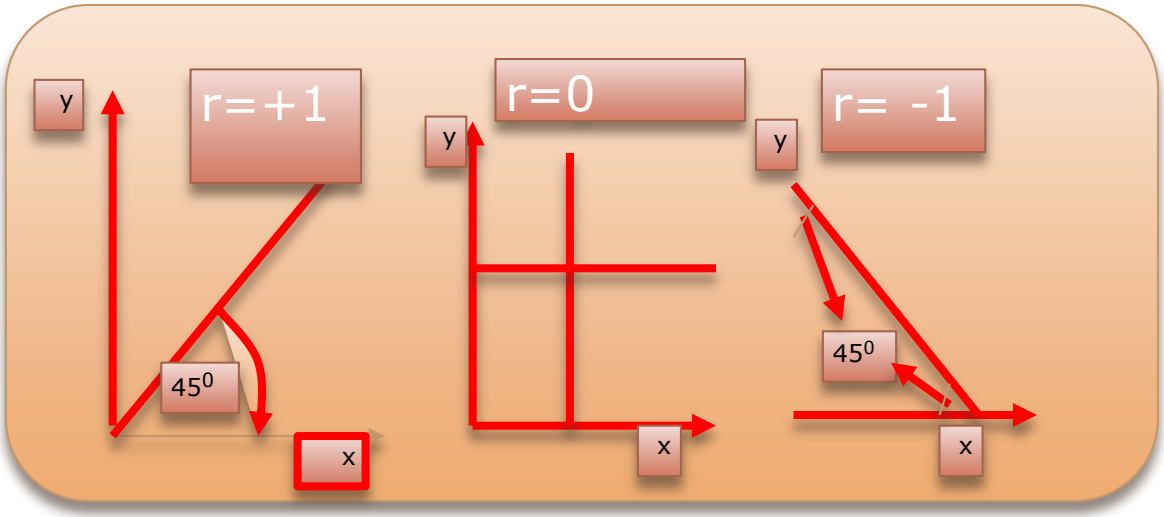
معامل بيرسون Pearson: ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين يختلفان في وحدات القياس أو في مستواهما العام، عن طريق حساب انحرافات كل من المتغيرين عن وسطهما الحسابي، وقسمة الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير.

معامل سبيرمان Spearman: أحد معاملات ارتباط الرتب Rank Correlation لإيجاد الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، سواء أكان المتغيران وصفيين أم كميين أم أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً.

معامل الارتباط

أنواع الارتباط

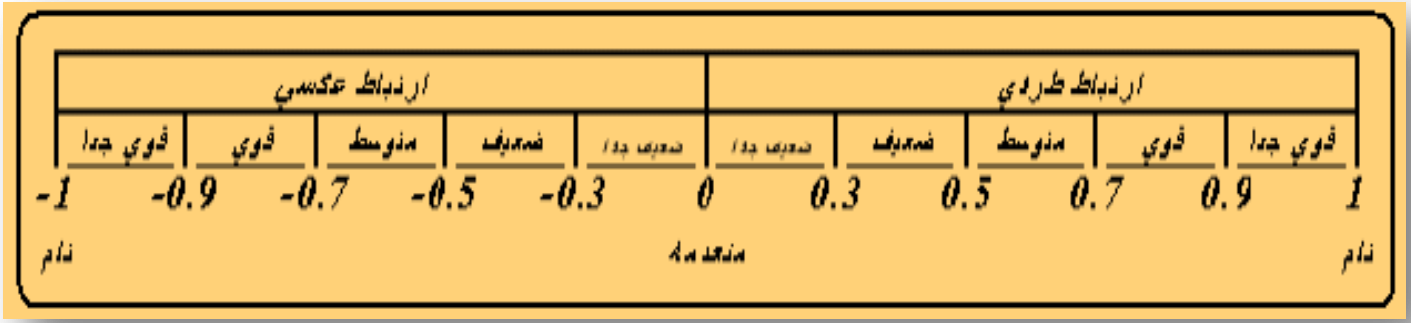
- ارتباط موجب (العلاقة الطردية)
- لا يوجد ارتباط (العلاقة منعدمة)
- ارتباط سالب (العلاقة العكسية)



معامل الارتباط

أنواع الارتباط

- دراسة العلاقة بين متغيرين اثنين فقط يعرف بالارتباط البسيط Simple Correlation، بينما دراسة العلاقة بين أكثر من متغيرين تعرف بالارتباط المتعدد Multiple Correlation
- الارتباط الموجب (الطردي) : Positive correlation اذا تغير أحد المتغيرين X أو Y فإن المتغير الآخر يتبعه بنفس الاتجاه، بمعنى أنه إذا زاد أو نقص أحدهما، زاد أو نقص الأخر.
- الارتباط السالب (العكسي) : Negative correlation إذا تغير أحد المتغيرين X أو Y فإن المتغير الآخر يتبعه بالاتجاه المضاد، بمعنى أنه إذا زاد أحدهما نقص الأخر، أو إذا نقص أحدهما زاد الأخر.



معامل الارتباط r

تعبير يشير إلى المقياس الإحصائي الذي يدل على:

مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية.

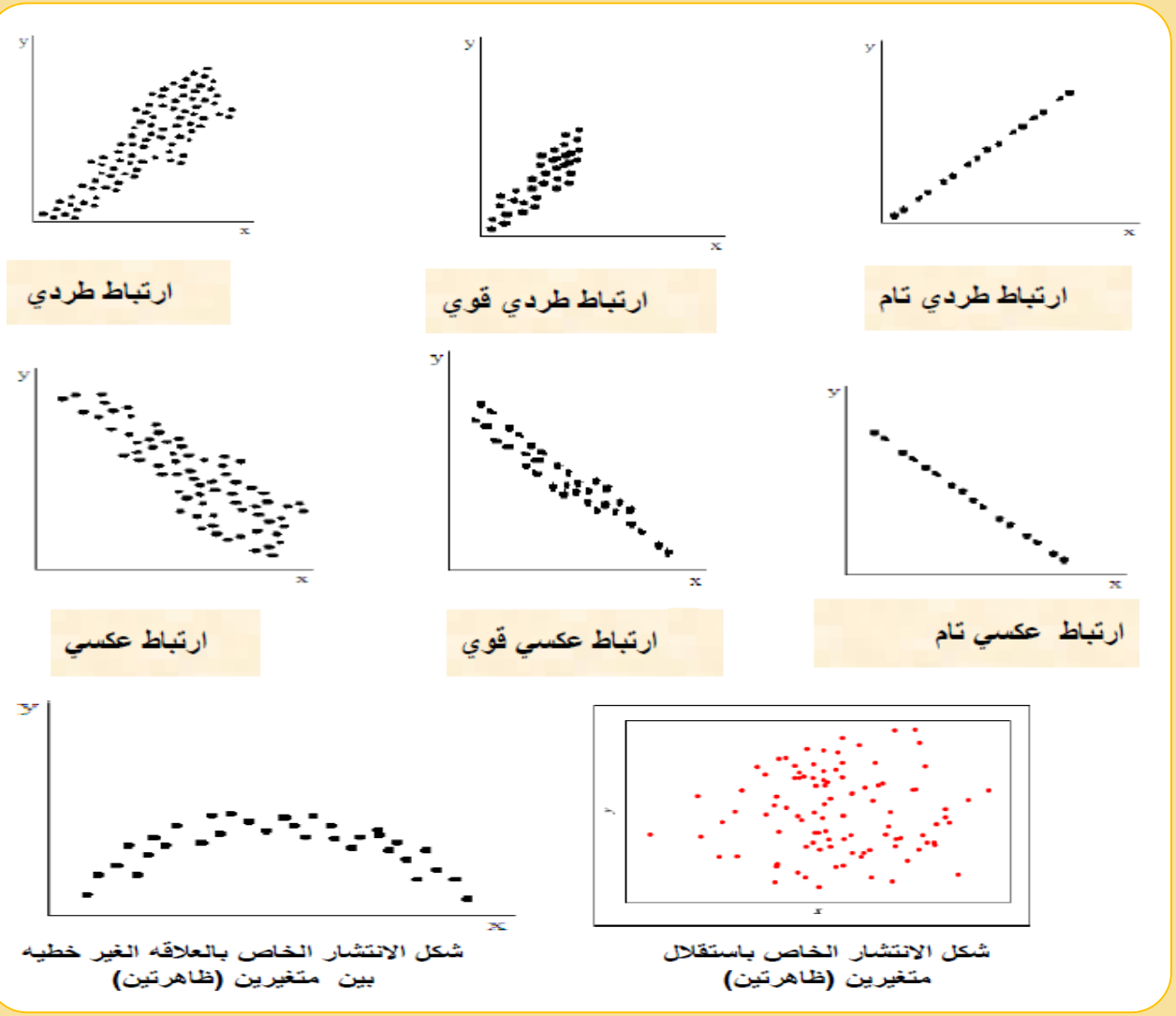
تتراوح قيمته بين الارتباط الموجب التام (+1) وبين الارتباط السالب التام (-1).

تدل إشارة المعامل الموجبة على العلاقة الطردية، بينما تدل إشارة المعامل السالبة على العلاقة العكسية .

معامل الارتباط

أولاً: طريقة شكل الانتشار Scatter Diagram

- شكل الانتشار هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي؛ حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى.
- طريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة



معامل الارتباط

ثانياً: معامل ارتباط بيرسون Pearson

- عند تطبيق معامل بيرسون للارتباط يجب أن تكون بيانات كلا المتغيرين (الظاهرتين) بيانات كمية.
- ويمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين باستخدام الصيغة المختصرة التالية:

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

- مثال: البيانات التالية تمثل المتغيرين X,Y ، الأول: حجم الإنتاج والثاني حجم الصادرات لأحد شركات البترول، المطلوب حساب معامل ارتباط بيرسون، ومعرفة مدى قوة العلاقة الخطية بينهما؟

Y	1	3	8	7	6	5	7	8	12	12
X	9	11	17	18	19	16	16	19	23	23

- كل ما نحتاجه لحساب معامل الارتباط الخطي لبيرسون بالصيغة المختصرة حساب مجموع مربعات قيم x ومجموع مربعات قيم y ومجموع حاصل ضربهما بعد معرفة $\sum X$ ، $\sum Y$ ، n .

حيث :

- $\sum XY$ تعني مجموع حاصل ضرب كل قيمة من X في Y
- $(\sum X)$ تعني مجموع قيم المتغير X
- $(\sum Y)$ تعني مجموع قيم المتغير Y
- $\sum X^2$ تعني مجموع مربع قيم المتغير X
- $(\sum X)^2$ تعني مربع مجموع قيم المتغير X
- $\sum Y^2$ تعني مجموع مربع قيم المتغير Y
- $(\sum Y)^2$ تعني مربع مجموع قيم المتغير Y
- n عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

معامل الإرتباط

معامل ارتباط بيرسون

x	y	xy	x^2	y^2
9	1	9	81	1
11	3	33	121	9
17	8	136	289	64
18	7	126	324	49
19	6	114	361	36
16	5	80	256	25
16	7	112	256	49
19	8	152	361	64
23	12	276	529	144
23	12	276	529	144
$\sum$ 171	69	1314	3107	585
$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$	$=\sum y^2$

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

$$r_p = \frac{10(1314) - (171)(69)}{\sqrt{((10\times 3107) - 171^2)((10\times 585) - 69^2)}} = \frac{13140 - 11799}{\sqrt{(31070 - 29241)(5850 - 4761)}} = \frac{1341}{\sqrt{18291089}}$$

$$= \frac{1341}{1411.30} = 0.95$$

العلاقة الخطية بينهما علاقة ارتباط طردية قوية

معامل الارتباط

ثالثاً: معامل ارتباط سبيرمان Spearman

- عند تطبيق معامل سبيرمان للارتباط يكون الاهتمام منصباً على الرتب أكثر من القيم سواء أكان كلا المتغيرين (الظاهرتين) كميين أم وصفيين أم أحدهما مخالف للآخر.
- ولحساب معامل سبيرمان لارتباط الرتب يتم ترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً لكلا المتغيرين.
- في حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة 1، والقيمة الأعلى منها مباشرة الرتبة 2 ..... وهكذا.
- في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة 1، والقيمة الأقل منها مباشرة الرتبة 2 ..... وهكذا.
- عند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة، ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط للقيم.
- بعد ترتيب المتغيرين يتم حسب الفروق بين رتب كل من المتغيرين، ونرمز للفروق بالرمز (d).
- بعدها يتم تربيع الفروق ، ونحصل على مجموعها أي نحصل على $\sum d^2$
- ثم حساب معامل سبيرمان لارتباط الرتب باستخدام الصيغة التالية:

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

حيث :

- $\sum d^2$ تعني مجموع مربعات الفروق بين رتب المتغيرين X و Y
- n عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

ويلاحظ:

- مجموع الفروق بين الرتب يساوي صفر.
- فإذا كانت الرتبة 1 للمتغير الأول تناظرها الرتبة 1 للمتغير الثاني، والرتبة 2 للمتغير الأول تناظرها الرتبة رقم 2 للمتغير الثاني، وهكذا. فإن معامل ارتباط الرتب يساوي + 1 (ارتباط طردي تام بين الرتب). والعكس (ارتباط عكسي تام بين الرتب).

معامل الارتباط

معامل ارتباط سبيرمان

لحساب معامل سبيرمان لارتباط الرتب يتم الآتي :

- ترتيب كل من المتغيرين ترتيباً تنازلياً لكلا السؤالين، ففي السؤال الأول يأخذ التقدير الأعلى الرتبة 1، والتقدير الأقل الرتبة 2 ... وهكذا بالنسبة للسؤال الثاني.
- عند تساوي التقديرات نعطي كل تقدير رتبة مختلفة، ثم نحسب متوسط هذه الرتب، وتعطى القيمة للتقدير.
- تحسب الفروق بين رتب كل من السؤالين d، ثم تربع وتجمع $\sum d^2$

❑ مثال: البيانات التالية تمثل إجابات عينة لسبعة طلاب حول برامج الأبتعاث، ومدى ملاءمتها للمتقدمين، والمطلوب حساب معامل سبيرمان لارتباط الرتب بين اجابات هذين السؤالين ؟

❑ تطبق المعادلة

السؤال الأول X	السؤال الثاني Y	رتب X	رتب Y	d الفرق بين الرتب	d^2 مربعات الفرق
جيد	جيد جداً	4	2.5	1.5	2.25
مقبول	مقبول	6.5	7	- 0.5	0.25
ممتاز	جيد جداً	1	2.5	- 1.5	2.25
جيد	جيد	4	5	- 1.0	1.00
جيد جداً	جيد	2	5	- 3.0	9.00
مقبول	جيد	6.5	5	1.5	2.25
جيد	ممتاز	4	1	3.0	9.00
المجموع				Zero	26.0

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

$$= 1 - \frac{6(26)}{7(49-1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

إذاً نوعية الارتباط طردي متوسط. وبالتالي فليس بالضرورة أن يكون رأي المجيبين تعني ملاءمة البرنامج للمتقدمين.

الانحدار الخطي

الانحدار الخطي Linear Regression

- الانحدار أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار.
- ويعرف المتغير الأول بالمتغير التابع dependent ويرمز له بالحرف Y، في حين يعرف المتغير الآخر بالمتغير المستقل Independent ويرمز له بالحرف X.
- تهدف دراسة الانحدار التنبؤ بقيمة متغير Y بمعرفة متغير X. إذاً الغاية من استخدام أسلوب تحليل الانحدار الخطي دراسة وتحليل أثر متغير كمي على متغير كمي آخر.
- ويعتمد نموذج الانحدار دائماً على علاقة السببية، بمعنى ان يكون التغير في المتغير المستقل مسبب رئيسي للتغير في المتغير التابع.

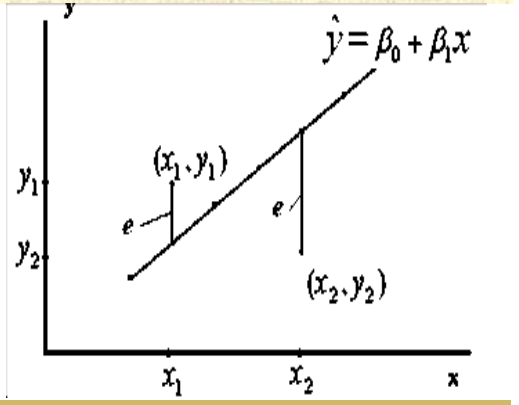
الانحدار الخطي البسيط

$$Y = a + b X$$

وتحسب a, b تبعاً للآتي :

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum (X)^2 - (\sum X)^2}$$

$$a = \frac{\sum Y - b\sum X}{n}$$



الانحدار الخطي

معادلة خط الانحدار

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

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 :

حيث :

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

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

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

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

نعوض

الانحدار الخطي

مثال تطبيقي للانحدار الخطي

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

أوجد معادلة الانحدار الخطي البسيط، وتوقع قيمة الاستهلاك عندما يصل إنتاج 16,000,000 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 :

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

الانحدار الخطي

مثال تطبيقي للانحدار الخطي

- ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج **16000000** برميل، نحول وحدة هذه القيمة من برميل إلى مليون برميل بالقسمة على مليون أي أن القيمة المستخدمة في توقع الاستهلاك هي $x_h = 16$
- وبالتعويض في المعادلة السابقة نجد أن:

$$\hat{y}_h = a + bx_h$$

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

أي أن الاستهلاك قد يصل إلى **9.02** مليون برميل، أي ما يعادل **9020000** برميل خلال السنة.

معامل الارتباط

معامل الاقتران (معامل فاي Phi)

- يستخدم معامل فاي لإيجاد العلاقة الارتباطية بين متغيرين اسميين في مستوين، كل منهما ثنائي التقسيم.
- ولحساب معامل الاقتران يتم ترتيب كل من المتغيرين الاسميين كالآتي:

	X_1	X_2	Sum
Y_1	a	b	a+b
Y_2	c	d	C+d
Sum	a+c	b+d	a+b+c+d

- ثم تطبق معادلة فاي

$$r_{\emptyset} = \frac{a \times d - b \times C}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

- مع ملاحظة أن اشارة معامل فاي ليس لها معنى، فهو يقيس قوة العلاقة الارتباطية دون اتجاهها.

معامل الارتباط

معامل الاقتران (معامل فاي Phi)

مثال: أوجد قيمة معامل الاقتران بين الجنس (ذكر/ أنثى)، وبين الاصابة بمرض الاكتئاب (مصاب/ غير مصاب) للبيانات التالية:

البيان	X1مصاب	X2 غير مصاب	Sum
Y1 ذكر	12 a	7 b	19
Y2 أنثى	10 c	5 d	15
Sum	22	12	36

ثم تطبق معادلة فاي

$$r_{\emptyset} = \frac{a \times d - b \times C}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$r_{\emptyset} = \frac{12 \times 5 - 7 \times 10}{\sqrt{22 \times 12 \times 19 \times 15}} =$$

$$= \frac{60 - 70}{\sqrt{75240}} = \frac{-10}{274.299} = -0.037$$

To Be Continued