

الباب الرابع: الارتباط والانحدار الخطي البسيط

Chapter 4: Correlation & Simple Linear Regression

سنناول في هذا الفصل :

- (١) مفهوم الارتباط وأنواعه.
- (٢) طرق حساب معاملات الارتباط المختلفة.
- (٣) مفهوم الانحدار الخطي البسيط وتطبيقاته.



+2,000
+5,000
+1,500
+1,125
+1,062

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 تجددين في بعض المجالات 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 *Correlation*

- الارتباط: هو تعيين طبيعة وقوة العلاقة بين متغيرين أو عدمها
- معامل الارتباط Correlation Coefficient هو مؤشر هذه العلاقة
- أول خطواته في تحديد طبيعته العلاقة 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

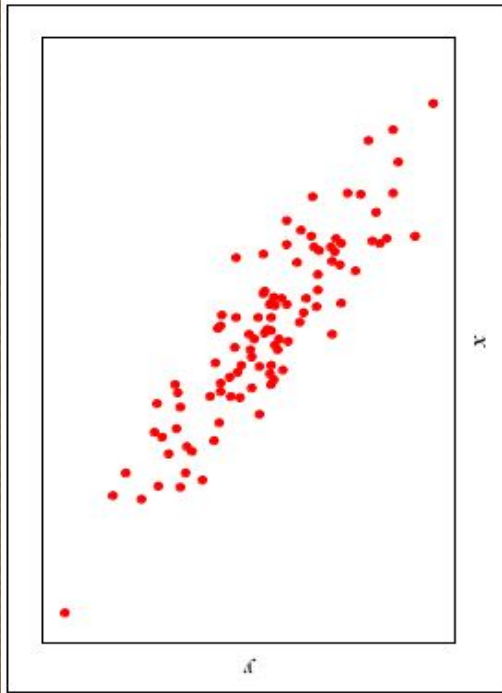
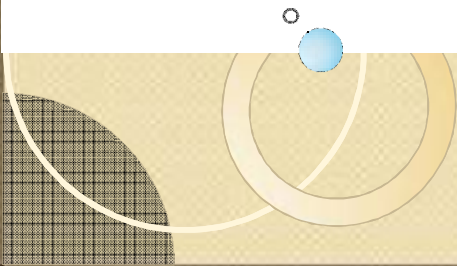
- يرافق المتغير X متغير آخر Y ويسمى بالمتغير التابع dependent variable وهو متغير تابع لأن نتيجته غير محددة وتعتمد على قيم المتغير المستقل

الارتباط

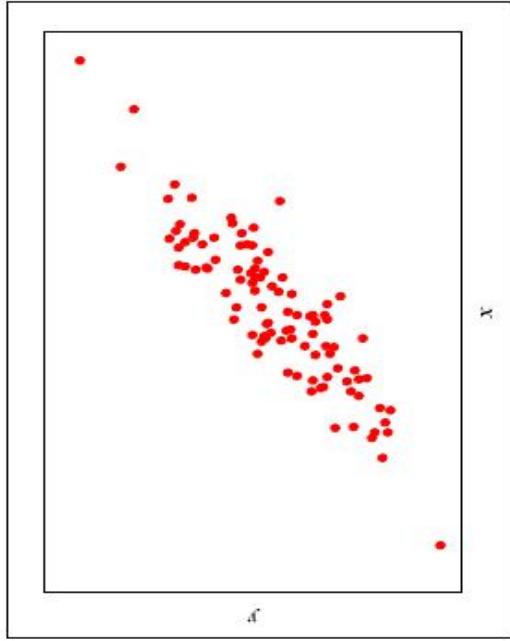
أنواع الارتباط

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

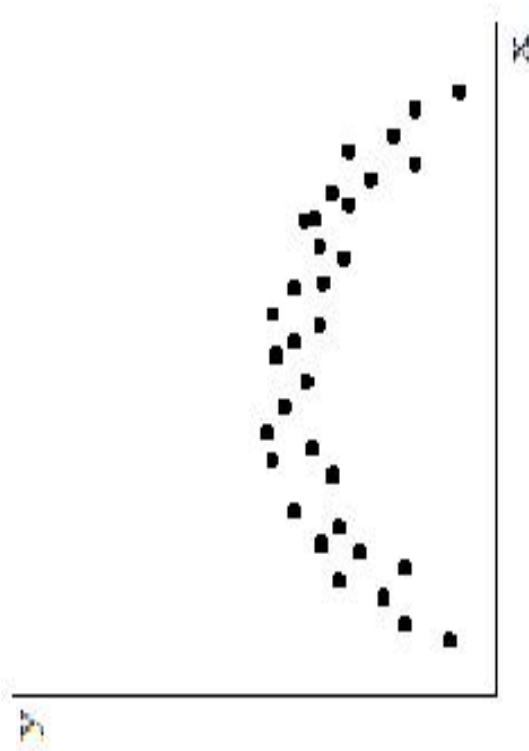
الارتباط الموجب (الطردى) (Positive Correlation) بأنه علاقة بين متغيرين (x, y) بحيث إذا تغير أحد المتغيرين فإن الآخر يتبعه في نفس الاتجاه.



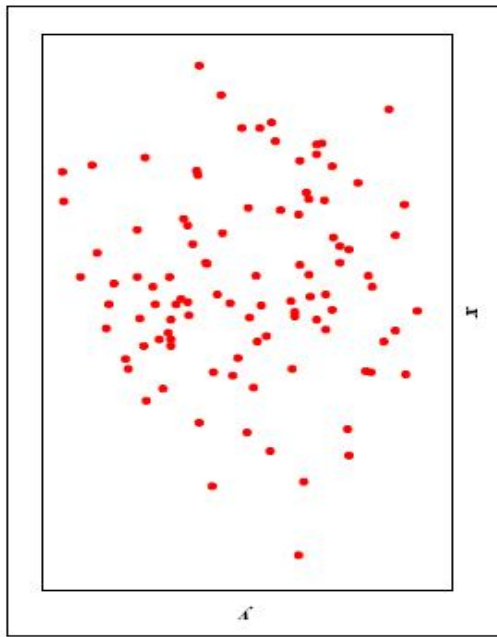
شكل الانتشار الخاص بالارتباط السالب (العكسي)



شكل الانتشار الخاص بالارتباط الموجب (الطردى)



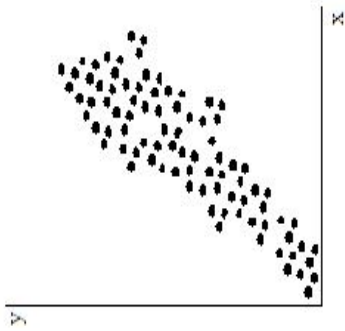
شكل الانتشار الخاص بالعلاقة الغير خطيه بين متغيرين (ظاهرتين)



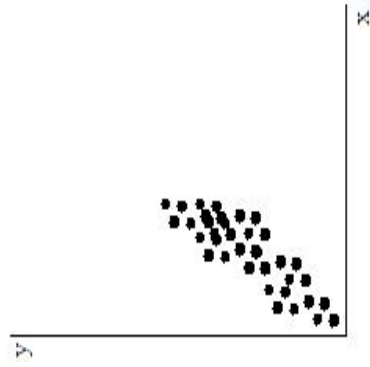
شكل الانتشار الخاص باستقلال متغيرين (ظاهرتين)

Scatter Plot

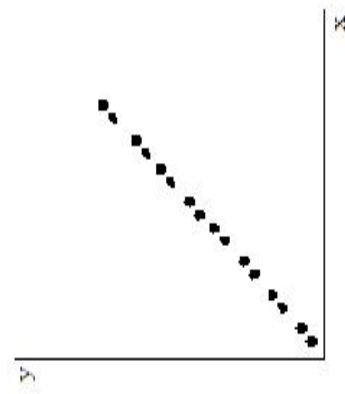
شكل الانتشار



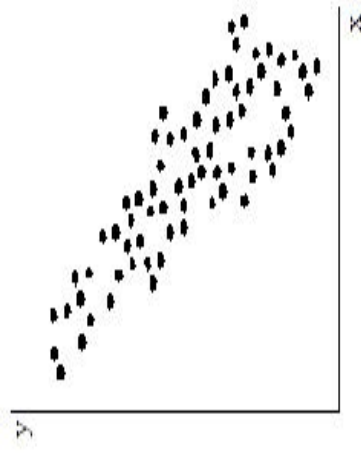
ارتباط طردي



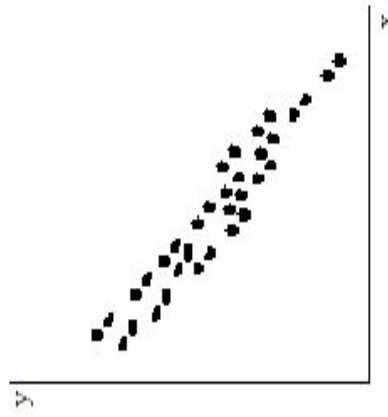
ارتباط طردي قوي



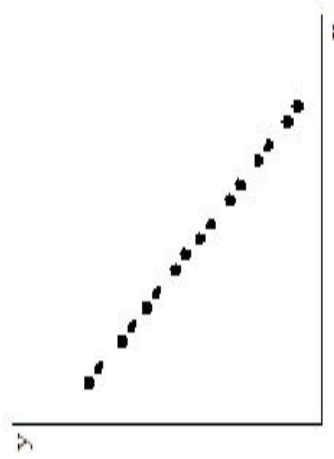
ارتباط طردي تام



ارتباط عكسي



ارتباط عكسي قوي



ارتباط عكسي تام

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 الارتباط بين متغيرين ، حيث تتراوح قيمته بين

$$(-1) \text{ و } (+1) ، \text{ أي أن } -1 \leq r \leq +1$$

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

- يمكن حساب العديد من معاملات الارتباط ويعتمد ذلك على مستوى القياس (السمي - ترتيبي - فترة - نسبي) للمتغيرات التي تبدو مرتبطة .

قياس الارتباط

وبالجدول التالي يوضح أنواع الارتباط واتجاه العلاقة وشكل الانتشار لكل نوع :

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)

١- معامل بيرسون للارتباط الخطي

Pearson Linear Correlation Coefficient

- معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخداماً خاصة في العلوم الإنسانية و الاجتماعية .
- و مستوى 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 .

أ - معامل بيرسون للارتباط الخطي

- حساب معامل بيرسون للارتباط الخطي :

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين X و Y ، 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 :

$$\begin{aligned} & \sum_{i=1}^n x_i y_i : \text{مجموع حاصل ضرب } X \text{ في } Y \\ & \sum x : \text{مجموع قيم المتغير } X \\ & \sum y : \text{مجموع قيم المتغير } Y \\ & \sum x^2 : \text{مجموع مربعات قيم المتغير } X \\ & \sum y^2 : \text{مجموع مربعات قيم المتغير } Y \end{aligned}$$

• مثال (4 - 2):

سُجِّلَتْ ست قراءات تقريبية لحجم الإنتاج وحجم صادرات النفط الخام بالمملكة العربية السعودية (بالمليار برميل) خلال عدة سنوات كما يلي:

حجم المبيعات (y)	2	2	2	1	1	1
حجم الإنتاج (x)	3	4	2	2	2	2

الدرس وجود علاقة ارتباط خطية بين حجم الإنتاج وحجم صادرات النفط الخام.

• الحل :

$$r_p = \frac{6(24) - (15)(9)}{\sqrt{((6 \times 41) - 15^2)((6 \times 15) - 9^2)}} = \frac{144 - 135}{\sqrt{(246 - 225)(90 - 81)}} = \frac{9}{\sqrt{189} \cdot 13.75} = 0.65$$

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

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

• مثال (4 - 1) :

لدراسة علاقة الصادرات بالميزان التجاري خلال عدة سنوات، أخذنا عشر قراءات تقريبية لقيمة صادرات المملكة العربية السعودية (x) وقيمة الميزان التجاري (y) بعشرات المليارات ريال كما يلي:

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

احسب معامل الارتباط الخطي، ما مدى قوة العلاقة الخطية؟

• الحل :

$$r_p = \frac{10(1314) - (171)(69)}{\sqrt{((10 \times 3107) - 171^2)((10 \times 585) - 69^2)}} = \frac{13140 - 11799}{\sqrt{(31070 - 29241)(5850 - 4761)}} = \frac{1341}{\sqrt{18291089}} = \frac{1341}{1411.30} = 0.95$$

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
Σ	171	1314	3107	585
= Σ x		= Σ y	= Σ xy	= Σ x ² = Σ y ²

من الملاحظ أن علاقة الارتباط الخطي بين قيمة صادرات المملكة العربية السعودية وقيمة الميزان التجاري موجودة وهي علاقة ارتباط طردية قوية.

الانحدار

١. التنبؤ Prediction

هو تقدير القيمة المستقبلية لمتغير واحد بناءً على معرفة قيم متغير آخر

و يفيدنا في:

- تحديد شكل العلاقة بين المتغيرين رياضياً وبيانياً) خط الانحدار (.
- توضيح اتجاه العلاقة بين المتغيرين .
- التنبؤ بقيمة أحد المتغيرين بدلالة المتغير الآخر .

الانحدار Regression

- والانحدار 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 بالمستوى نحصل على شكل الانتشار فإذا أظهر الشكل الانتشاري للبيانات أن هناك علاقة خطية بين المتغيرين نقوم بتقدير خط الانحدار \hat{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 \hat{Y} على X (أو Y/X)
 - وتحسب القيمتان a و b من العلاقتين التاليتين :
- حيث:

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

الانحدار الخطي البسيط

- 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 البسيط

• ملاحظات هامة :

❖ ميل الخط يمثل كمية التغير في Y المناظرة للتغير في X بمقدار وحدة واحدة .

❖ إشارة معامل الانحدار تدل على نوع الارتباط (طردى أو عكسي)

❖ توجد علاقة بين معامل الانحدار ومعامل الارتباط الخطي

- مثال (4 - 8) :
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 :

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

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

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

∴ معادلة خط الانحدار البسيط في هذه الحالة :

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
Σ	90	632	942
= Σ x		= Σ y	= Σ xy
		= Σ x ²	

تابع حل مثال (4 - 8)

- 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:

$$\begin{aligned}\hat{y}_h &= a + bx_h \\ &= 3.26 + 0.36(16) = 9.02\end{aligned}$$

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

4-4-3 تطبيق الانحدار في مجال السلاسل الزمنية

أسلوب السلاسل الزمنية

يتم رصد البيانات التي تعبر عن ظاهرة ما عند نقاط زمنية متتالية.

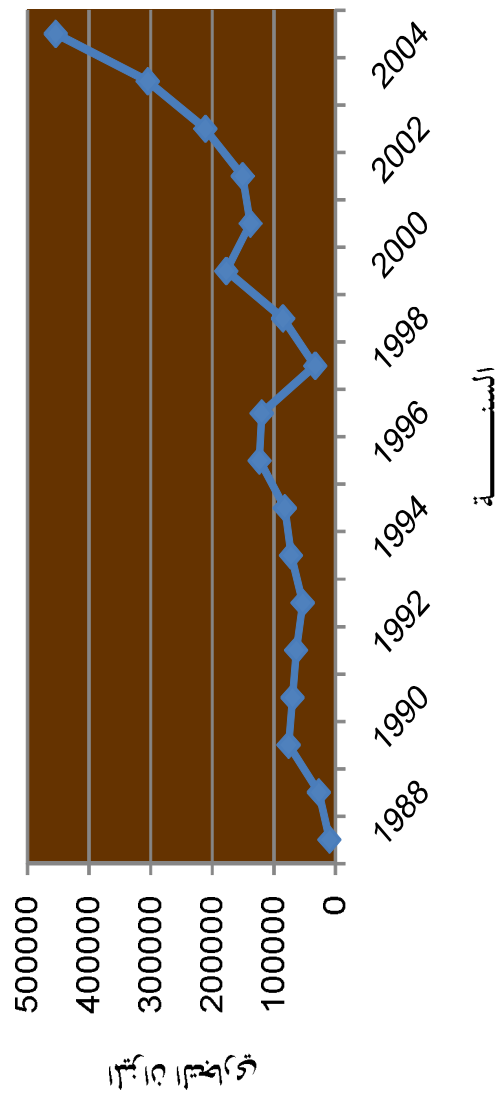
أمثلة:

- كمية الصادرات السنوية
- حجم التعاملات الربع السنوي في البورصة
- عدد الجرائم اليومية في إحدى البلاد

4-3 تطبيق الانحدار في مجال السلاسل الزمنية

شكل السلسلة الزمنية

يعتبر شكل السلسلة الزمنية من أبرز الأشكال المستخدمة في كثير من المجالات العلمية والاقتصادية والاجتماعية، حيث تتميز بعض الظواهر بالتطور خلال الزمن، مثل أسواق البورصة والأسهم، وسعر النفط. تطور أعداد المعتمرين والحجاج في موسمي رمضان والحج والنمو السكاني السنوي. فيما يلي مجموعة من أشكال السلاسل الزمنية لبعض الظواهر من واقع الحياة:

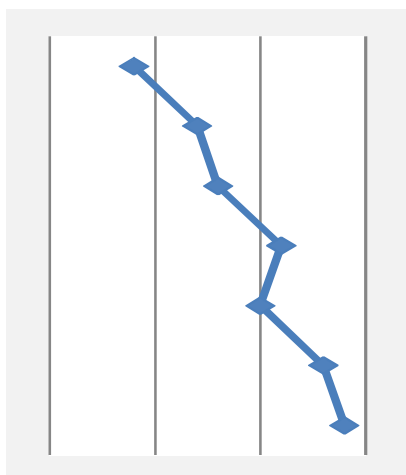


4-4-3 تطبيق الانحدار في مجال السلاسل الزمنية

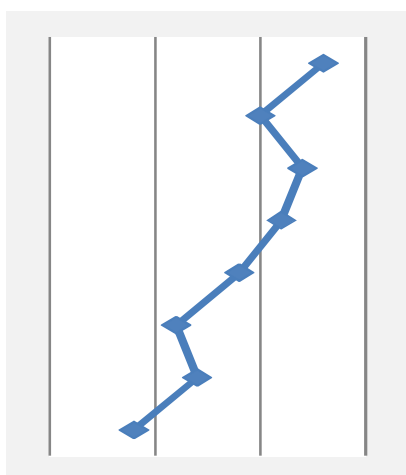
يستخدم شكل السلسلة الزمنية للتعرف على مكونات السلاسل الزمنية والتي تتمحور في أربع مكونات:

- الاتجاه العام: وهو اتجاه التطور الذي تأخذه السلسلة الزمنية خلال فترة طويلة من الزمن بالرغم من التذبذبات الموجودة بها، ويكون التطور إما بالزيادة أو بالنقصان، وبعض السلاسل لا يوجد لها اتجاه.
- التغيرات الموسمية: وهي التغيرات التي تتكرر بانتظام خلال فترة زمنية أقل من السنة، عادة في المواسم.
- التغيرات الدورية: وهي التغيرات التي تحدث في فترات زمنية أكثر من سنة وعادة كل خمس أو عشر سنوات.
- التغيرات العرضية: وهي التغيرات التي تحدث نتيجة حوادث فجائية غير متوقعة مثل الفيضانات والأعاصير والحروب... الخ.

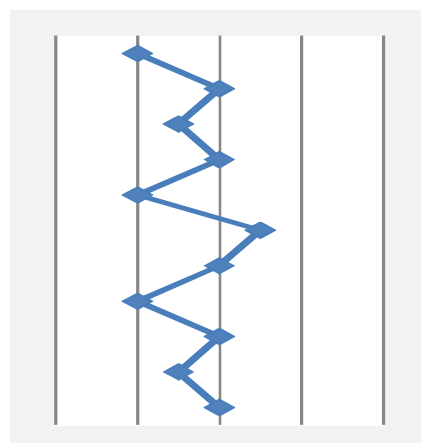
4-4-3 تطبيق الانحدار في مجال السلاسل الزمنية



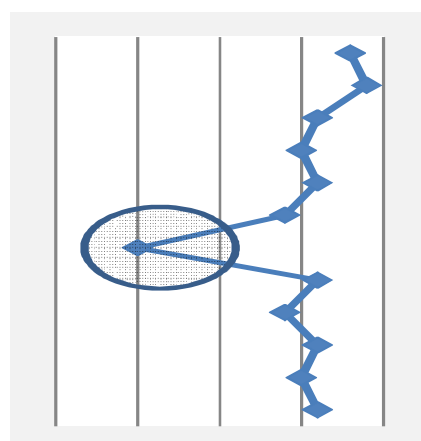
سلسلة ذات اتجاه عام بالزيادة



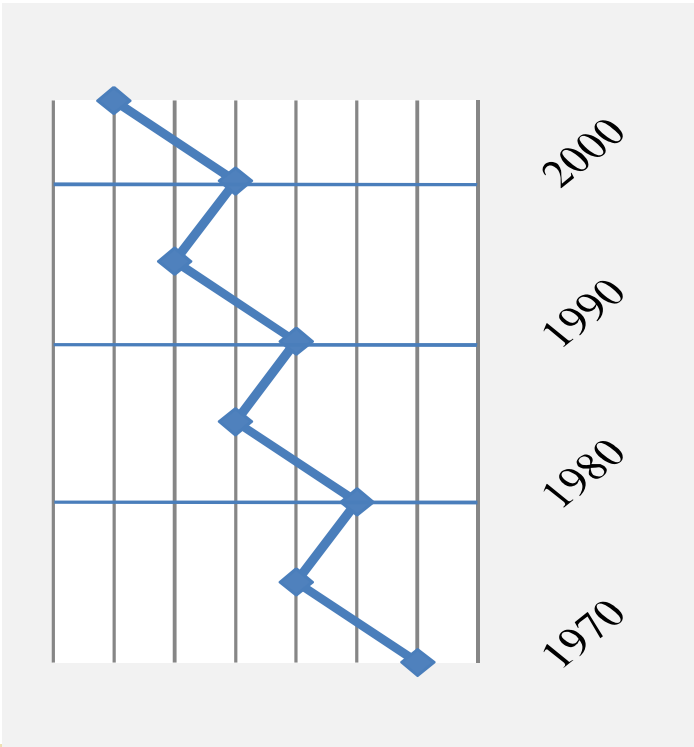
سلسلة ذات اتجاه عام بالنقصان



سلسلة ليس لها اتجاه عام



سلسلة ذات عامل عرضي



سلسلة ذات اتجاه زيادة وتغيرات موسمية

سلسلة ذات اتجاه زيادة تغيرات دورية

4-3 تطبيق الانحدار في مجال السلاسل الزمنية

- يتطلب تحليل السلاسل الزمنية عادة تحليل المكونات الأربعة ولكن قد تختلف السلاسل الزمنية من التغيرات الموسمية والدورية والعرضية، ويبقى تعيين الاتجاه العام .

- يختلف شكل الاتجاه العام في السلاسل الزمنية حسب طبيعة البيانات، وأحد أنواع الاتجاه العام هو الاتجاه الخطي .

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

- ملاحظات:

- تعيين للمتغير المستقل القيم $x = 0, 1, 2, \dots$ لتمثل وحدة الزمن.

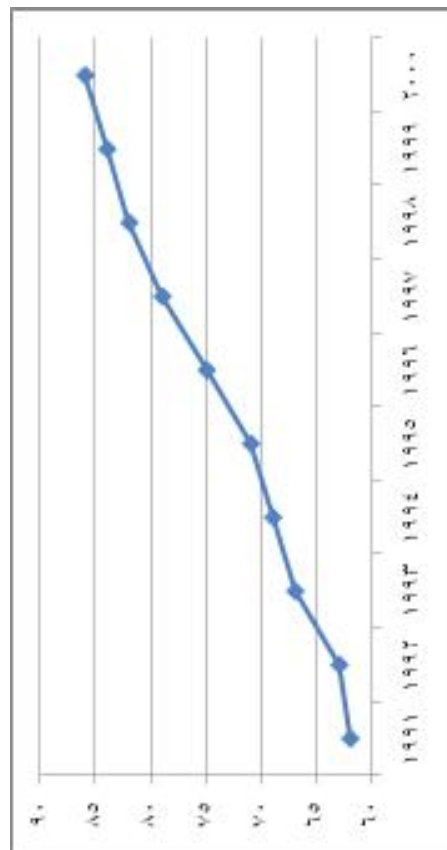
- تدل إشارة معامل الانحدار b على نوع الاتجاه العام .

- مثال (4 - 9) : 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

حدد نوع الاتجاه العام، ثم قدر معادلة الاتجاه العام الخطي، ثم توقع عدد الحقول المكتشفة عام 2002م.

- الحل:



يدل الاتجاه العام على الزيادة في قيمة عدد الحقول المكتشفة.

• تابع حل المثال (٩-٤) : تقدير معادلة الاتجاه العام الخطي ، وحساب التوقع

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

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

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

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
Σ	45	737	3553	285
	Σ x	Σ y	Σ xy	Σ x ²

∴ معادلة الاتجاه العام الخطي في هذه المثال

تابع حل المثال (٩-٤) : تقدير معادلة الاتجاه العام الخطي ، وحسابه المتوقع

• ولتوقع عدد الحقول المتوقع اكتشافها عام 2002م نعوض بقيمة

تدل على هذا الزمن؛ حيث أن 2000م $\leftarrow x = 9$

إذن 2002م $\leftarrow x_h = 11$

• وبالتعويض في معادلة الاتجاه العام نجد أن:

$$\begin{aligned}\hat{y}_h &= 60.79 + 2.87x_h \\ &= 60.79 + 2.87(11) = 92.36 \approx 92 \text{ حقل}\end{aligned}$$

الحمد لله
الذي هدانا لهذا
ما كنا لنهتدي لولا
هدى الله لنا

