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 .

Notes

❑..

❑..

❑..

❑..

❑..

❑..

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.

Notes

❑..

❑..

❑..

❑..

❑..

❑..

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 هى 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 وهومتغير احصائى لأن نتيجته غير محددة وتعتمد على قيم المتغير المستقل.

Notes

❑..

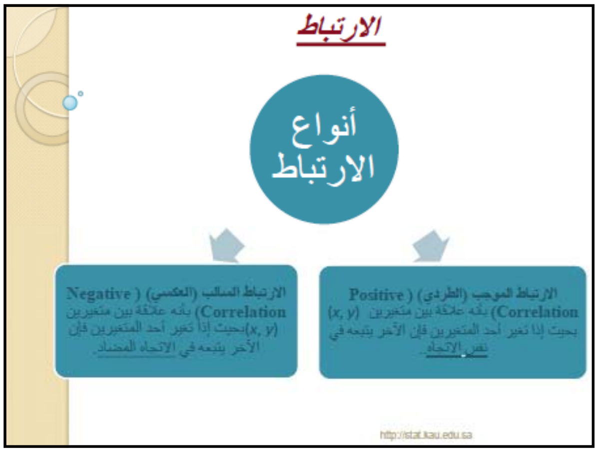
❑..

❑..

❑..

❑..

❑..



Notes

❑ ……………………………………………………………

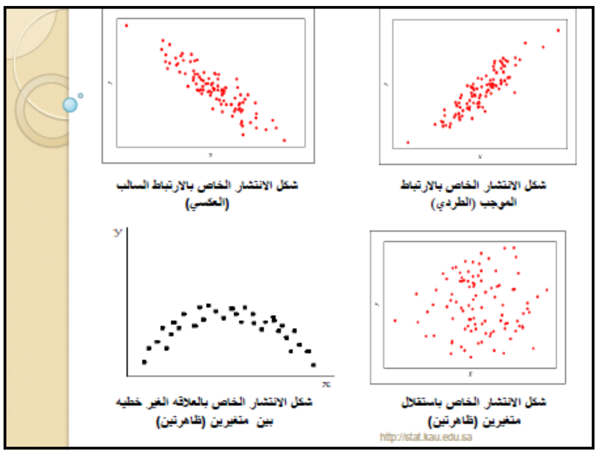
❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………



Notes

❑ ……………………………………………………………

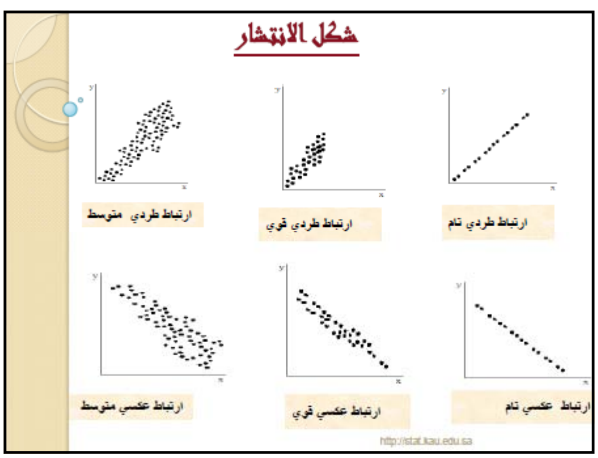
❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………



Notes

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

❑ ……………………………………………………………

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

وتدل إشارة المعامل الموجبة على العلاقة الطردية ،

بينما تدل إشارة المعامل السالبة على العلا $-1 \le r \le +1$

http://stat.kau.edu.sa

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

قياس الارتباط

والجدول 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)

http://stat.kau.edu.sa

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

1- معامل بيرسون للارتباط الخطي

- معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخداماً خاصةً في العلوم الإنسانية والاجتماعية.
- عند تطبيق معامل بيرسون للارتباط يجب أن يكون كلا المتغيرين(x،y) بيانات كمية.

http://stat.kau.edu.sa

Notes

- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………
- ❑…………………………………………………………

حساب معامل بيرسون للارتباط الخطي

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين x، y باستخدام الصيغة التالية:

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

حيث:

$\sum xy$: مجموع حاصل ضرب x في y

$\sum x$: مجموع قيم المتغير (أو الظاهرة) x

$\sum y$: مجموع قيم المتغير (أو الظاهرة) y

$\sum x^2$: مجموع مربعات قيم المتغير (أو الظاهرة) x

$\sum y^2$: مجموع مربعات المتغير (أو الظاهرة) y

n : عدد المفردات

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

- 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 :

	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 $=\sum y^2$

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

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

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

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

http://stat.kau.edu.sa

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

2 – معامل سبيرمان لارتباط الرتب

- نستخدم معامل سبيرمان لارتباط الرتب (Rank Correlation coefficient) إذا كان المتغيرين كليهماوصفي ترتيبي أو كليهما متغيركمي.
- 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 .

http://stat.kau.edu.sa

Notes

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

❑ ....

- 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 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 سبيرمان لارتباط الرتب بسهولة حسابه حتى لو كانت البيانات غير مرتبة.
- يعاب على معامل سبيرمان 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 :

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

http://stat.kau.edu.sa

Notes

❑...

❑...

❑...

❑...

❑...

❑...

مثال :

أوجد قيمة معامل الاقتران 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 a	8 b	20
أنثى	4 c	6 d	10
المجموع	16	14	30

وعليه فإن :

a= 12

b = 8

c = 4

d = 6

$$r_\phi = \frac{(ad)-(bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{(12\times6)-(8\times4)}{\sqrt{20\times10\times16\times14}}$$

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

أي أنه توجد علاقة ضعيفة بين النوع والإصابة بمرض الاكتئاب .

http://stat.kau.edu.sa

Notes

❑...

❑...

❑...

❑...

❑...

❑...

4 - معامل بوينت بايسيريال للارتباط

معامل بوينت بايسيريال

(Point Biserial correlation coefficient)

يستخدم لقياس علاقة الارتباط بين متغير كمي X ومتغير اسمي Y(ذي مستويين) كالإجابة (نعم – لا)، أو الجنس (ذكر أنثى)...الخ.

مثال:

نستخدم معامل بوينت بايسيريال لدراسة الارتباط بين إجابة الطالبة على السؤال الإجباري y(أجابت/لم تجب) ،وبين الدرجة الإجمالية x ، لمجموعة من الطالبات في اختبار الإحصاء.

http://stat.kau.edu.sa

Notes

❑...

❑...

❑...

❑...

❑...

❑...

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.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

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:

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

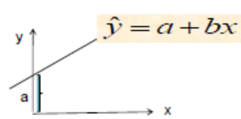
حيث a : ثابت الانحدار أو الجزء المقطوع من محور y

b : ميل الخط المستقيم أو معامل انحدار

- وتحسب القيمتان a و b من العلاقتين التاليتين:

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

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



Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

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

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

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

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

.: معادلة خط الانحدار البسيط في هذه الحالة :

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

http://stat.kau.edu.sa

Notes

❑..

❑..

❑..

❑..

❑..

❑..

تابع حل المثال

- ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج 16 مليون برميل
- وبالتعويض 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.

http://stat.kau.edu.sa

Notes

❑..

❑..

❑..

❑..

❑..

❑..

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.

Notes

❑..

❑..

❑..

❑..

❑..

❑..

- 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.

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

الحل: تقدير معادلة الاتجاه العام الخطي

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

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

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

معادلة الاتجاه العام الخطي في هذه المثال

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

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

تابع حل المثال : حساب التوقع

- ولتوقع عدد الحقول المتوقع اكتشافها عام 2002م نعوض بقيمة تدل على هذا الزمن؛ حيث أن 2000م ← x = 9

إذن 2002م ← $x_h=11$

- وبالتعويض في معادلة الاتجاه العام نجد أن:

$$\hat{y}_h=60.79+2.87x_h$$
$$=60.79+2.87(11)=92.36\approx 92$$ حقل

Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………