إحصاء ١١١

STAT 111

قسم الإحصاء – كلية العلوم

جامعة الملك عبد العزيز

الباب الرابع

الارتباط والانحدار الخطي البسيط

الارتباط:

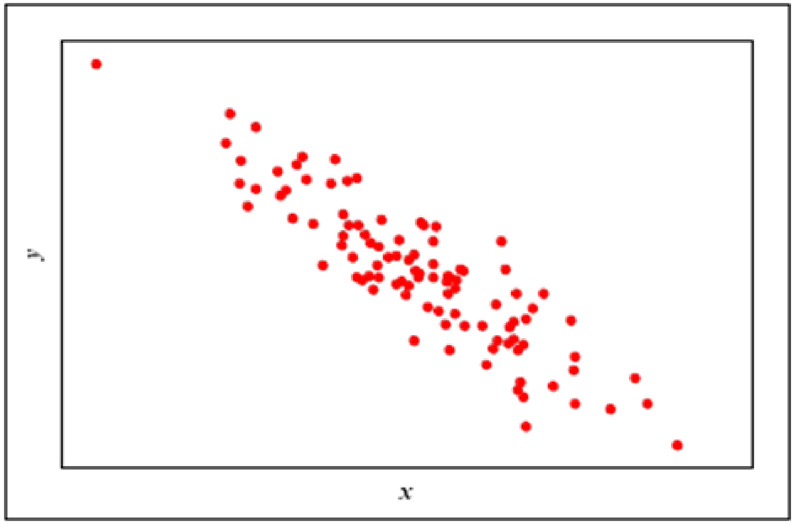
الارتباط هو تعيين طبيعة وقوة العلاقة بين متغيرين .أما معامل الارتباط فهو مؤشر هذه العلاقة. وسنتعرض هنا لموضوع الارتباط البسيط وهو الارتباط بين ظاهرتين (متغيرين) فقط الاول يسمى المتغير المستقل وهو متغير يتم *تحديده من قبل الباحث أو الشخص* الذي يقوم بالدراسة، والثاني المتغير التابع وهومتغيرعشوائي، لأن *نتيجته غير محددة وتعتمد على قيم المتغير المستقل.*

أشكال الارتباط

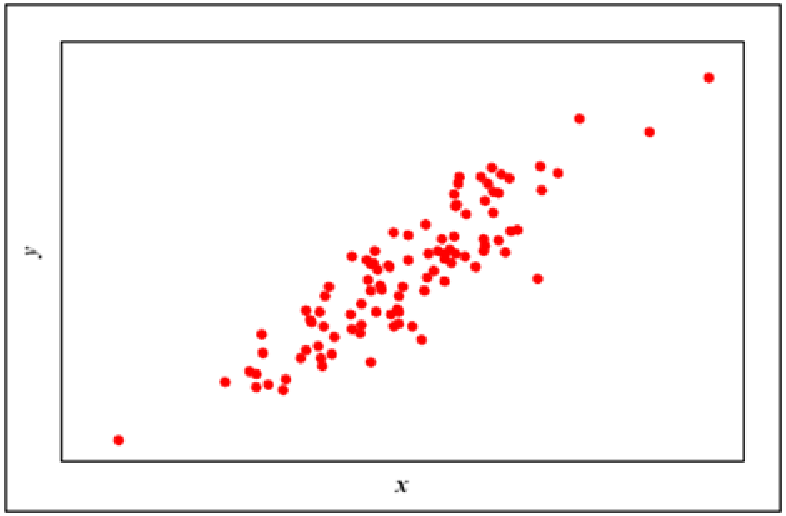
الارتباط الموجب (الطردي) بأنه علاقة بين متغيرين (x, y) بحيث إذا تغير أحد المتغيرين فإن الآخر يتبعه في ***نفس الاتجاه***.

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

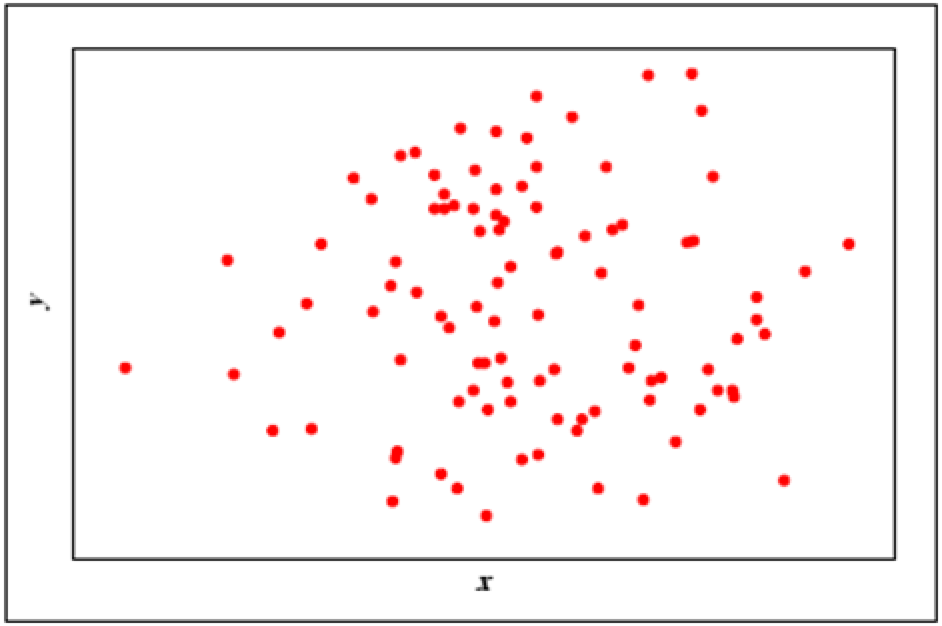
**ويمكن استخدام الرسم من خلال استخدام شكل الانتشار في تحديد طبيعة العلاقة بين المتغيرين، كما هو موضح في الاشكال التاليه:



شكل الانتشار الخاص بالارتباط السالب (العكسي).



شكل الانتشار الخاص بالارتباط الموجب (الطردي).



شكل الانتشار الخاص باستقلال متغيرين (ظاهرتين).

قياس الارتباط

تستخدم معاملات خاصة تسمى معاملات الارتباط لقياس *درجة الارتباط بين متغيرين (ظاهرتين) لهما مستوى قياس كمي أو ترتيبي*.

معامل الارتباط:

يعرف معامل الارتباط والذي يرمز له بالرمز r بأنه عبارة عن مقياس رقمي يقيس نوع وقوة الارتباط بين متغيرين, حيث تتراوح قيمته بين (+1) و (-1) , أي أن $-1 \le r \le +1$

وتدل إشارة المعامل *الموجبة* على *العلاقة الطردية*, بينما تدل إشارة المعامل *السالبة* على *العلاقة العكسية*. ويوضح الجدول التالي تفصيل دلالة قيم r:

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

***** ومايقال عن الارتباط الطردي ينطبق على الارتباط العكسي (مع وضع إشارة سالبة).***

1. معامل بيرسون للارتباط الخطي

يعتبر معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخداماً وخاصةً في العلوم الإنسانية والاجتماعية. ومستوى القياس المطلوب عند تطبيق معامل بيرسون للارتباط هو أن يكون كلا المتغيرين (الظاهرتين) بيانات كمية.

حساب معامل بيرسون للارتباط الخطي

يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين $Y,\ X$ باستخدام الصيغة التالية:

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

مثال:

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

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

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

$$r_p = \frac{n\left(\sum xy\right)-\left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\left(\sum x^2\right)-\left(\sum x\right)^2\right]\left[n\left(\sum y^2\right)-\left(\sum y\right)^2\right]}} = \frac{6(24)-(15)(9)}{\sqrt{\left[(6\times 41)-15^2\right]\left[(6\times 15)-9^2\right]}}$$

$$= \frac{144-135}{\sqrt{[246-225][90-81]}} = \frac{9}{\sqrt{21\times 9}} = \frac{9}{\sqrt{189}}$$

$$\approx \frac{9}{13.75} \approx \mathbf{0.65}$$

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

٢. معامل بوينت بايسيريال للارتباط

يستخدم لقياس علاقة الارتباط بين متغير كمي (X) ومتغير اسمي (Y) (ذي مستويين) كالإجابة (نعم – لا)، أو الجنس (ذكر – أنثى)...الخ، وهذا المعامل مستنتج من معامل ارتباط بيرسون.

٣. معامل سبيرمان لارتباط الرتب

يستخدم هذا المعامل أساساً إذا كان قياس المتغيرين كليهما مقياس ترتيبي أو كليهما متغيرين كميين ذو قيم كبيرة لتسهيل الحسابات. ويتم حسابه من خلال أعطاء كل قيمة من قيم كل متغير رتبة (R) حيث أقل قيمة تأخذ الرتبة ١ والتي تليها الرتبة ٢ وهكذا، ومن ثم نطبق الصيغة التاليه:

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

حيث n هي عدد الأزواج **المرتبة** و $d = R_x - R_y$.

مثال:

لدراسة علاقة ارتباط تقديرات الطلاب في مادة الإحصاء وتقديراتهم في مادة ألرياضيات، اخترنا خمسة 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 طرديه متوسطة بين تقديرات الطلاب في مادة الإحصاء وتقديراتهم في مادة الرياضيات.

مثال:

في دراسة لمعرفة العلاقة بين عدد الحقول المكتشفة وطول الأنابيب (بالكيلومتر) الناقلة للنفط الخام بالمملكة العربية السعودية خلال عدة سنوات، سجلت سبع قراءات على النحو التالي:

عدد الحقول (x)	55	54	56	61	62	63	67
طول الأنابيب (y)	21960	22027	23006	23008	23020	23125	23120

هل توجد علاقة ارتباط بين عدد الحقول وطول الأنابيب؟

الحل:

x	y	رتب x	رتب y	d	d^2
55	21960	2	1	1	1
54	22027	1	2	-1	1
56	23006	3	3	0	0
61	23008	4	4	0	0
62	23020	5	5	0	0
63	23125	6	7	-1	1
67	23120	7	6	1	1
			$\sum$	0.0	4

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{(6)(4)}{7(49-1)} = 1 - \frac{24}{336} = 1 - 0.07 = 0.93$$

نلاحظ وجود علاقة ارتباط طرديه قوية بين عدد الحقول المكتشفة وطول الأنابيب الناقلة للنفط الخام.

٣. معامل الاقتران (فاي)

معامل اقتران "فاى" يستخدم أساساً لقياس العلاقة بين متغيرين اسميين كل منهما ثنائي التقسيم، كالنوع (ذكر/أنثى) والإصابة بالمرض (مصاب/غير مصاب) والتدخين (مدخن/غير مدخن)...الخ. بفرض أن المتغيرين معرفين على صورة جدول ثنائي مزدوج كما يلي:

	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) للبيانات التالية:

النوع \ مرض الاكتئاب	مصاب	غير مصاب
ذكر	12	8
أنثى	4	6

الحل:

نوجد أولا المجاميع الهامشية كما في الجدول التالي:

النوع \ مرض الاكتئاب	مصاب	غير مصاب	المجموع
ذكر	12	8	**20**
أنثى	4	6	**10**
المجموع	**16**	**14**	**30**

وعليه فإن:

$$a=12,\quad b=8,\quad c=4,\quad d=6$$

$$r_{\varnothing}=\frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}=\frac{12\times 6-8\times 4}{\sqrt{20\times 10\times 16\times 14}}=\frac{72-32}{\sqrt{44800}}=\frac{40}{211.66}\approx 0.19$$

أي أنه توجد علاقة ضعيفة بين النوع والإصابة بمرض الاكتئاب.

الانحدار

الانحدار هو أسلوب يمكن بواسطته *تقدير قيمة المتغير التابع (Y) إذا علمت قيمة المتغير المستقل (X) عن طريق معادلة الانحدار*, وسوف نهتم فقط بالانحدار الخطي البسيط وهو الذي يكون بين متغيرين فقط احدهما مستقل X والاخر تابع Y.

** معادلة الانحدار تعرف على الصورة التالية:

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

حيث:

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

b: معامل الانحدارأو ميل الخط المستقيم.

<u>وتحسب القيمتان a و *b* من العلاقتين التاليتين:</u>

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

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

ولإيجاد أي قيمة مقدرة جديدة $\hat{y}_h$ ، نعوض بقيمة معلومة للمتغير المستقل ولتكن X_h في معادلة خط الانحدار.

ملاحظات هامة

- إشارة معامل الانحدار تدل على نوع الارتباط (طردي أو عكسي).
- توجد علاقة بين معامل الانحدار ومعامل الارتباط الخطي.

مثال:

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

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

أوجد معادلة الانحدار الخطي البسيط، وتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج 11,000,000 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
	$=\sum x$	$=\sum y$	$=\sum x\,y$	$=\sum x^2$

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

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

إذاً معادلة خط الانحدار البسيط في هذه الحالة: $\hat{y}=3.26+0.36x$ ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج 11000000 برميل، نحول وحدة هذه القيمة من برميل إلى مليون برميل بالقسمة على مليون أي أن القيمة المستخدمة في توقع الاستهلاك هي $x_h = 11$ وبالتعويض في المعادلة السابقة نجد أن:

$$\hat{y}_h = a + bx_h = 3.26 + 0.36(11) = 7.22$$

أي أن الاستهلاك المحلي قد يصل إلى 7.22 مليون برميل عندما يصل الإنتاج إلى 11 مليون برميل.

تطبيق الانحدار في مجال السلاسل الزمنية

يتطلب تحليل السلاسل الزمنية عادة تحليل المكونات الأربعة السالف ذكرها في الفصل الثاني، ولكن قد تخلو السلاسل الزمنية من التغيرات الموسمية والدورية والعرضية، ويبقى تعيين الاتجاه العام . ويختلف شكل الاتجاه العام في السلاسل الزمنية حسب طبيعة البيانات، وأحد أنواع الاتجاه العام هو الاتجاه الخطي.

سلسلة ذات اتجاه عام خطي متزايد

سلسلة ذات اتجاه عام خطي متناقص

أحد طرق تعيين الاتجاه العام الخطي هو استخدام أسلوب الانحدار الخطي البسيط، باعتبار أن (الزمن) السنوات، الفصول، ...الخ (متغير مستقل (x)، والمتغير التابع (y)هو الظاهرة) أو المتغير (محل الدراسة).

ملاحظات:

- تُعين للمتغير المستقل القيم $x = 0, 1, 2, \ldots$ لتمثل وحدة الزمن.
- تدل إشارة معامل الانحدار b على نوع الاتجاه العام (زيادة أو نقصان).

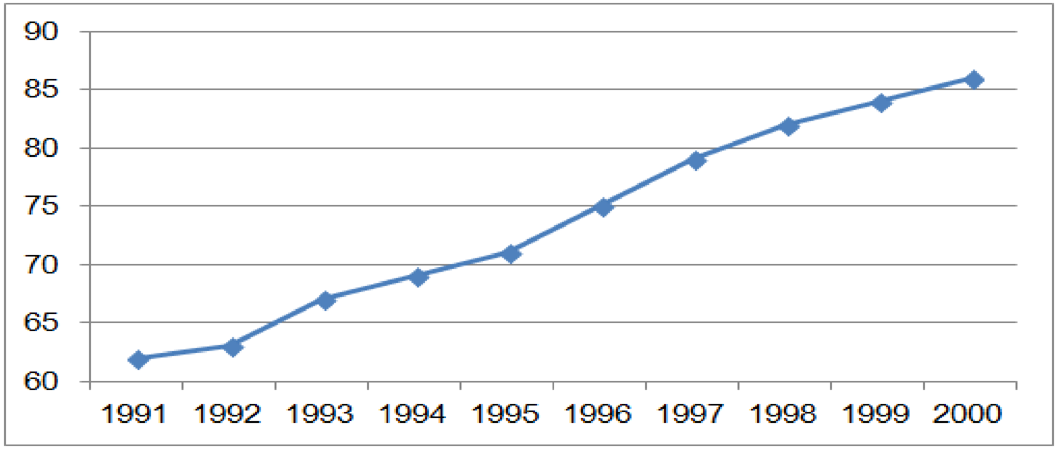
مثال:

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

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

الحل:



يدل الاتجاه العام على الزيادة فى قيمة عدد الحقول المكتشفة.

السنة	x	y	$x\,y$	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
	$=\sum x$	$=\sum y$	$=\sum x\,y$	$=\sum x^2$

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

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

$\therefore$ معادلة الاتجاه العام الخطية في هذا المثال $\hat{y}=60.79+2.87x$ ،

ولتوقع عدد الحقول المتوقع اكتشافها عام 2002م نعوض بقيمة تدل على هذا ألزمن حيث أن:

$$2000\text{م} \leftarrow x=9 \text{ إذن } 2002\text{م} \leftarrow x_h=11$$

وبالتعويض في معادلة الاتجاه العام نجد أن:

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