- الفصل الخامس -

الارتباط الخطي
Linear correlation

الارتباط الخطي يقترن بوجود متغيرين أو أكثر، ترتبط مع بعضها بعلاقات خطية معينة. على سبيل المثال « العلاقة بين طول الشخص (سم) ووزنه (كغم) » و « العلاقة بين تحصيل الطالب المتخرج ومعدل درجاته في الثانوية » و « العلاقة بين نسبة الشفاء من مرض معين وكمية الجرعة من الدواء للمريض وعمره » وغيرها من الأمثلة الأخرى.

فإذا كان التغير في احد المتغيرات يؤثر في تغير متغير آخر او مجموعة متغيرات أخرى يقال ان هذه المتغيرات مرتبطة فيما بينها، واذا كان المتغيرين مرتبطين (او مجموعة المتغيرات المرتبطة) يتغيران بنفس الاتجاه (زيادة او نقصان) في أحدهما تؤدي الى (زيادة او نقصان) في الأخر عندئذٍ يقال إن الارتباط مابينهما هو ارتباط موجب، مثال ذلك « زيادة طول الشخص يتوقع أن تقابلها زيادة في وزنه ».

اما اذا كان المتغيرين المرتبطين (او مجموعة من المتغيرات المرتبطة) يتغيران باتجاه معاكس (زيادة اونقصان) في احدهما تؤدي الى (نقصان او زيادة) في الأخر عندئذٍ يقال ان الارتباط ما بينهما هو ارتباط سالب، مثال ذلك « زيادة في سعر الوحدة من سلعة معينة يتوقع ان تؤدي الى انخفاض في الطلب على تلك السلعة »

ويمكن ان يكون الارتباط ارتباطاً تاماً « perfect » اذا كان التغير في أحد المتغيرين يتناسب مع التغير في المتغير الأخر مثال ذلك « أن الارتباط بين درجة الحرارة المئوية ودرجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار ان التغير في الاول يتناسب مع التغير في الثاني.

ان الاساس في دراسة الارتباط وحساب قيمته يستند الى العلاقة السببية التي تربط المتغيرات مع بعضها، بمعنى أخر وجود علاقة منطقية تفسر سبب الارتباط بينهما (المتغيرات). فمثلاً « تغير سلوك المستهلك تجاه زيادة أستهلاكه من سلعة معينة ناجم عن عامل او جملة عوامل جعلته يغير سلوكه منها

(زيادة دخله الشهري، زيادة عدد افراد عائلته، ارتفاع اسعار سلع بديلة لهذه السلعة، تقليص انفاقه على سلع أخرى، ... الخ).

في حين لا يوجد هناك اي منطق او سبب للعلاقة ما بين زيادة عدد الحوادث في الطرق من جهة وارتفاع استهلاك الفرد كمعدل من الطاقة الكهربائية بالرغم من امكانية حساب قيمة الارتباط بينهما الا انها قيمة ليس لها معنى او تفسير.

يتم حساب الارتباط من خلال معامل يدعى معامل الارتباط « correlation coefficient »

معامل الارتباط: هو درجة او قيمة العلاقة التي تربط متغيرين او أكثر مع بعضها، وهي قيمة حقيقية خالية من وحدات قياس المتغيرات المرتبطة بعلاقة.

انواع الارتباط:

هناك انواع عديدة من الارتباط كل منها يختص بنوع معين من العلاقات.

١- الارتباط الخطي البسيط: Simple Linear coefficient

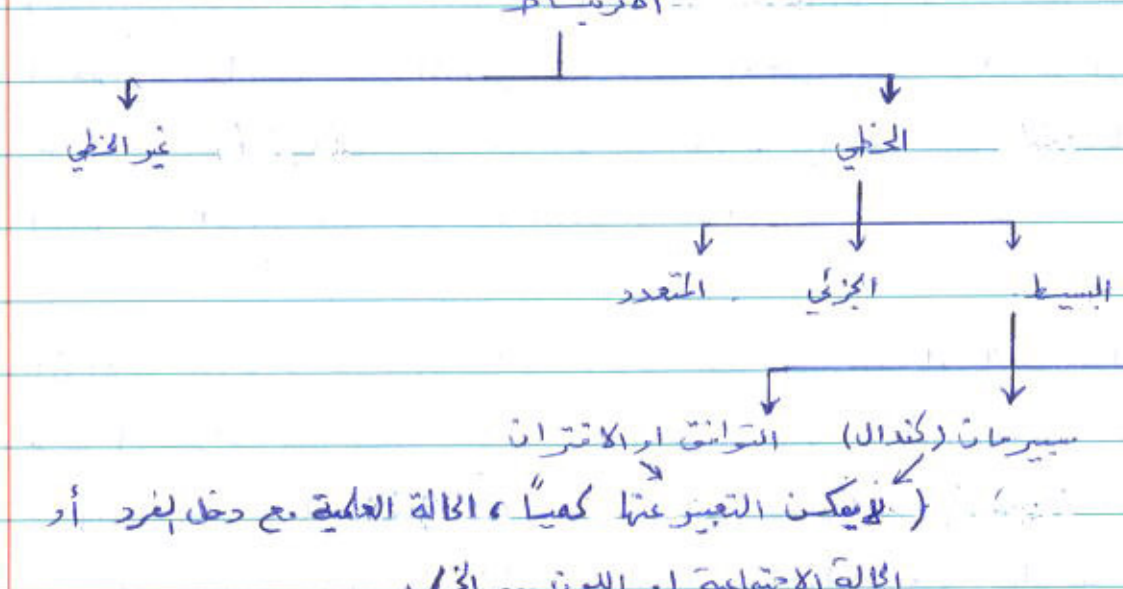
٢- الارتباط الجزئي: Partial correlation Coefficient

٣- الارتباط المتعدد: Multiple Correlation Coefficient

٤- ارتباط الرتب: Rank correlation

٥- ارتباط الصفات: correlation between attributes

* مدخلة: مقياس الارتباط يختلف بإختلاف الظواهر المناسبة المطلوب ايجاد العلاقة بينها.



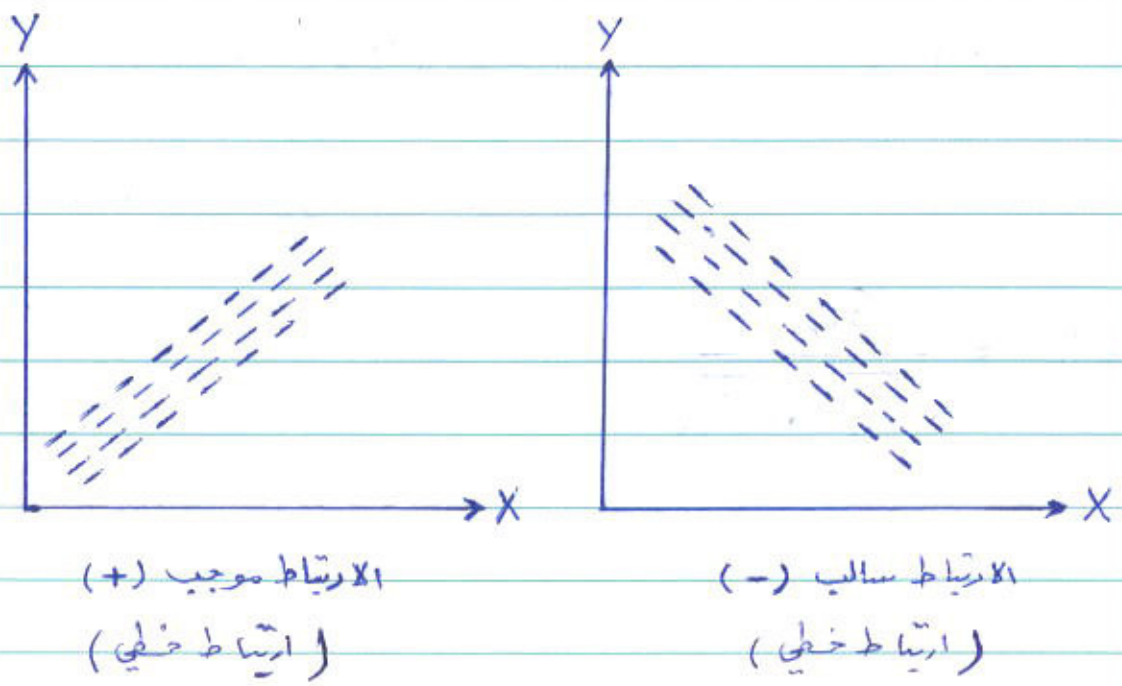
الارتباط الخطي البسيط
Simple Linear correlation

يعرف الارتباط الخطي البسيط بأنه " الدرجة أو القيمة العددية للعلاقة بين متغيرين فقط هما (X) و (Y).
ويمكن قياس العلاقة الارتباطية بين المتغيرين على مرحلتين، وكالآتي:

أ- الشكل الانتشاري: Scatter Diagram

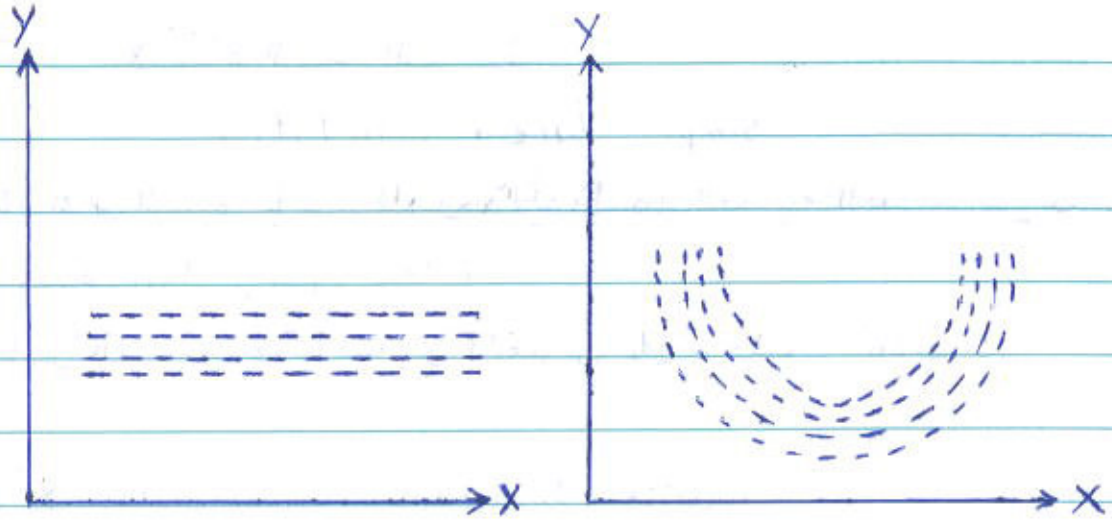
يُعد الشكل الانتشاري ابسط طريقة لعرض بيانات متغيرين يفترض بينهما علاقة ارتباطية وهو عبارة عن انتشار النقاط في المستويين (X) (Y). إذ يمثل الأحداثي السيني قيمة (X) و الأحداثي الصادي قيمة (Y). ويتم من خلاله تكوين فكرة اولية حول اتجاه وقوة العلاقة بين المتغيرين.

- على افتراض ان لدينا متغيرين هما (X) و (Y) على اساس عينة عشوائية من المشاهدات قوامها (n)، فان ازواج القيم لهذين المتغيرين تكتب على الشكل الآتي:
$[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ وعند تمثيل ازواج قيم المتغيرين بواسطة الشكل الانتشاري، فإننا نحصل على احد الاشكال الآتية التي من خلالها يتم التعرف على طبيعة العلاقة وقوتها بين المتغيرين.



الارتباط موجب (+)
(ارتباط خطي)

الارتباط سالب (-)
(ارتباط خطي)



عدم وجود ارتباط بين المتغيرين　　ارتباط غير خطي (علاقة منحنية)

ب- معامل الارتباط البسيط Simple Correlation Coefficient :

يعرف معامل الارتباط البسيط بأنه « القيمة العددية للعلاقة الارتباطية الخطية بين متغيرين فقط » ويأخذ معامل الارتباط البسيط عدد من الأشكال والصيغ الرياضية، ويرمز له بالرمز (r) أو (r_{xy}) او (r_{12}) أو (r_p).

طرق حساب معامل الارتباط البسيط لبيرسون :

وهو مؤشر احصائي يستخدم لقياس القوة الارتباطية الخطية بين متغيرين كميين، اي (يمكن قياسهما كمياً) مثال ذلك « قياس العلاقة بين الدخل الشهري للاسرة (X) وحجم انفاقها الشهري (Y)، ويعود الفضل الاول للعالم الانكليزي « كارل بيرسون » ((1867 - 1936)) في وضع الصيغة العامة لهذا المقياس.

معامل الارتباط البسيط وفق صيغة بيرسون كالآتي :

$$1 * r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\ \sqrt{S_{yy}}}$$

حيث إن :

$$S_{xy} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$S_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad , \quad S_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

إن الصيغة اعلاه تم ايجادها وفقاً لطريقة « الانحرافات » Deviations
وهناك صيغ بديلة لكل من [S_{xy} ، S_{xx} ، S_{yy}] وهي كالآتي :

$$1 - S_{xy} = \sum_{i=1}^{n} X_i Y_i - n \bar{X} \bar{Y} \quad \longleftrightarrow \quad \text{(البديل الاول)}$$

$$S_{xx} = \sum_{i=1}^{n} X_i^2 - n \bar{X}^2 \quad , \quad S_{yy} = \sum_{i=1}^{n} Y_i^2 - n \bar{Y}^2$$

$$2 - S_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n} \quad \longleftrightarrow \quad \text{(البديل الثاني)}$$

$$S_{xx} = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \quad , \quad S_{yy} = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

حيث إن :

n : تمثل عدد ازواج القيم (X_i و Y_i) لكلا المتغيرين .

$\bar{X}$ ، $\bar{Y}$: تمثل المتوسطات الحسابية للمتغيرين (X) و (Y) .

$\sum_{i=1}^{n} X_i Y_i$: تمثل حاصل ضرب قيم المتغيرين (X) و (Y) .

$\sum_{i=1}^{n} X_i^2$: تمثل مجموع مربعات قيم المتغير (X) .

$\sum_{i=1}^{n} Y_i^2$: تمثل مجموع مربعات قيم المتغير (Y) .

ملاحظات :-

$$1- \sum_{i=1}^{n} X_i Y_i \neq \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)$$

$$2- \sum_{i=1}^{n} X_i^2 \neq \left(\sum_{i=1}^{n} X_i\right)^2$$

$$3- \sum_{i=1}^{n} Y_i^2 \neq \left(\sum_{i=1}^{n}\right)^2$$

2* الصيغة الثانية لايجاد معامل الارتباط وتسمى الطريقة المباشرة

$$r_p = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - (\sum X)^2\right]\left[n\sum Y^2 - (\sum Y)^2\right]}}$$

خصائص معامل الارتباط البسيط لبيرسون

1- إن قيمة معامل الارتباط البسيط (r_p) تقع ضمن المجال $(-1 \leq r_p \leq +1)$

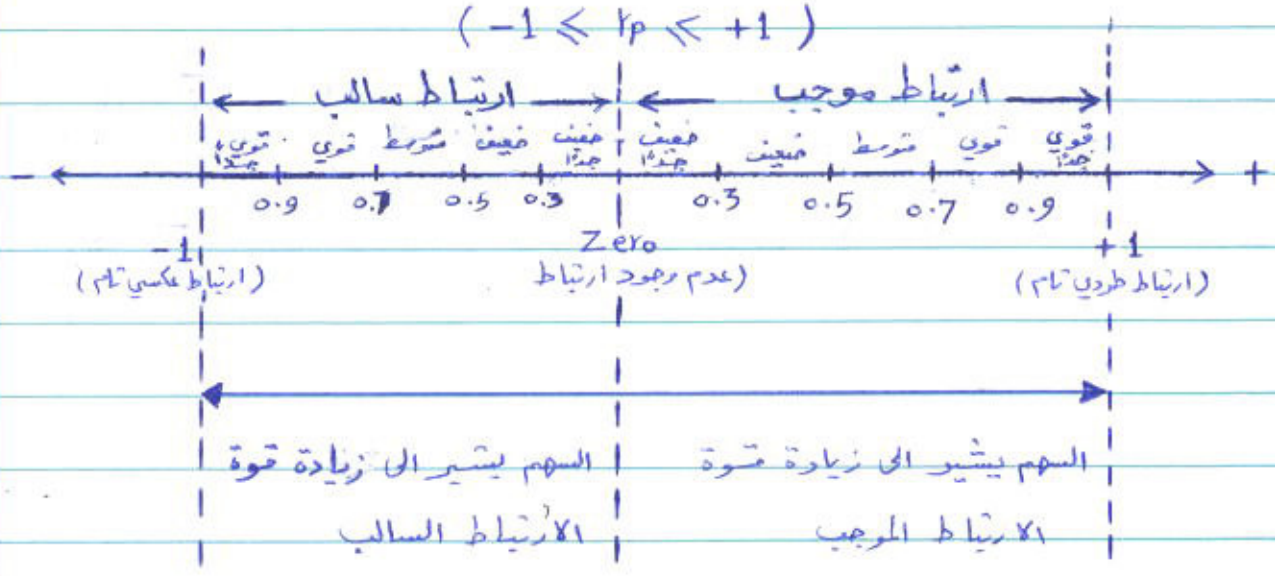
إذ ان :

$r_p = +1$ ← تعني ان الارتباط بين المتغيرين (X ، Y) ارتباط طردي تام.

$r_p = -1$ ← تعني ان الارتباط بين المتغيرين (Y ، X) ارتباط عكسي تام.

$r_p = 0$ ← تعني ان الارتباط بين المتغيرين (X ، Y) ارتباط معدوم.

$(-1 \leq r_p \leq +1)$



2- إن اية تحويلات خطية تجري على قيم أحد المتغيرين (X) أو (Y) أو كليهما لا تؤثر على قيمة معامل الارتباط البسيط (r_p) المستخرج للقيم المحددة.

3- تتأثر قيمة معامل الارتباط لبيرسون بالقيم الشاذة (Outlier Values)، مما يتوجب على متخذ القرار، توخي الدقة عند تفسير قيمة هذا المعامل.

مثال (1) :- البيانات التالية تمثل ازواج القيم لاوزان (كغم) خمسة طلاب وأطوالهم (سم).

وزن الطالب (X)	75	72	69	67	63
طول الطالب (Y)	175	170	172	165	162

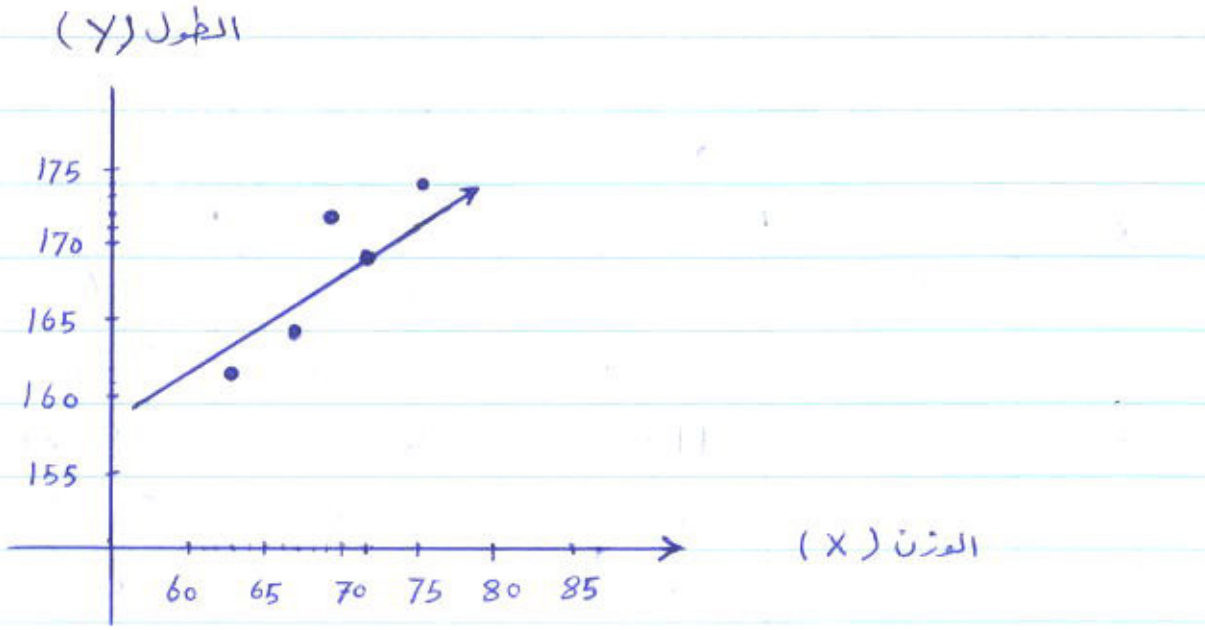
المطلوب :

1- ارسم الشكل الانتشاري (Scatter diagram) لازواج القيم.

2- احسب معامل الارتباط البسيط لبيرسون بين اوزان الطلبة واطوالهم باستخدام صيغة الانحرافات (Deviations) والصيغة البديل الأول.

الحل :

1- رسم الشكل الانتشاري :



2 - ايجاد معامل الارتباط البسيط لبيرسون :

أ- استخدام صيغة الانحرافات .

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-Y)^2$
75	175	5.8	6.2	35.96	33.64	38.44
72	170	2.8	1.2	3.36	7.84	1.44
69	172	-0.2	3.2	-0.64	0.04	10.24
67	165	-2.2	-3.8	8.36	4.84	14.44
63	162	-6.2	-6.8	42.16	38.44	46.24
346	844	0	0	89.2	84.8	110.80
	$\bar{X}=69.2$,		$\bar{Y}=168.8$	S_{xy}	S_{xx}	S_{yy}

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\ \sqrt{S_{yy}}} = \frac{89.2}{\sqrt{84.8}\ \sqrt{110.8}}$$

$$r_p = \frac{89.2}{(9.21)(10.526)} = \frac{89.2}{96.944} = +0.92$$

نلاحظ من النتيجة إن العلاقة بين وزن الطالب (X) وطوله (Y) علاقة طردية وقوية جداً وهذا يعني إن الزيادة في طول الطالب (Y) ، ستؤدي وبنفس الاتجاه الى زيادة وزنه (X).

ب- استخدام الصيغة البديل الاول:

X	Y	X.Y	X^2	Y^2
75	175	13125	5625	30625
72	170	12240	5184	28900
69	172	11868	4761	29584
67	165	11055	4489	27225
63	162	10206	3969	26244
346	844	58494	24028	142578
$\bar{X}=69.2$	$\bar{Y}=168.8$	ΣXY	ΣX^2	ΣY^2

يتم ايجاد قيم كل من [S_{xy} ، S_{xx} ، S_{xy}] ، على النحو الآتي :

$$S_{xy} = \Sigma XY - n\bar{X}\bar{Y}$$
$$= 58494 - 5(69.2)(168.8) = 89.2$$

$$S_{xx} = \Sigma X^2 - n\bar{X}^2$$
$$= 24028 - 5(69.2)^2 = 84.8$$

$$S_{yy} = \Sigma Y^2 - n\bar{Y}^2$$
$$= 142578 - 5(168.8)^2 = 110.8$$

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{89.2}{\sqrt{84.8}\sqrt{110.8}}$$

$$= \frac{89.2}{(9.21)(10.526)} = \frac{89.2}{96.944} = +0.92$$

مثال (2) البيانات الآتية تمثل ازواج القيم لمصاريف الجيب الاسبوعية (X) لثمانية طلاب في سن المراهقة، وعلاماتهم الفصلية (Y) في مادة الرياضيات .

مصروف الطالب (X)	2	4	10	8	2	13	5	12
علامته الفصلية (Y)	97	90	60	55	92	44	82	40

المطلوب : احسب معامل الارتباط البسيط لبيرسون، باستخدام الصيغة البديل الثاني :

١- نقوم بحساب المعلومات التالية ، وحسب الجدول الآتي :

X	Y	X.Y	X^2	Y^2
2	97	194	4	9409
4	90	360	16	8100
10	60	600	100	3600
8	55	440	64	3025
2	92	184	4	8464
13	44	572	169	1936
5	82	410	25	6724
12	40	480	144	1600
56	560	3240	526	42858
ΣX	ΣY	ΣXY	ΣX^2	ΣY^2

نقوم بايجاد قيم كل من [S_{xy} ، S_{xx} ، S_{yy}] ، وعلى النحو الآتي :

$$\therefore S_{xy} = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}$$

$$= 3240 - \frac{(56)(560)}{8} = 3240 - 3920$$

$$= -680$$

$$S_{xx} = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$= 526 - \frac{(56)^2}{8} = 526 - 392$$

$$= 134$$

$$S_{yy} = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$$

$$= 42858 - \frac{(560)^2}{8} = 42858 - 39200$$

$$= 3658$$

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{-680}{\sqrt{134}\sqrt{3658}}$$

$$= \frac{-680}{(11.576)(60.481)} = \frac{-680}{700.128} = -0.97$$

نلاحظ ان العلاقة بين مصروف جيب الطالب (X) وعلامته الفصلية (Y) علاقة عكسية وقوية جداً وهذا يعني ان زيادة مصروف جيب الطالب في سن المراهقة سيشغله عن الدراسة، مما يؤدي إلى حصوله على علامات واطئة في مادة الرياضيات.

مثال (3) :

أ- احسب معامل الارتباط البسيط لبيرسون (r_p) باستخدام الصيغة المناسبة للبيانات التي تمثل اربعة ازواج من قيم المتغيرين (X) و (Y).

المتغير (X)	5	1	0	2
المتغير (Y)	6	3	**2**	5

ب- احسب معامل الارتباط البسيط لبيرسون (r_p) لازواج قيم المتغيرين (Y^*) و (X^*)، بعد اجراء التحويل التالي، معلقاً على النتيجة.

$$X^* = 3X \quad , \quad Y^* = Y + 2$$

الحل :-

أ- نقوم بحساب معامل الارتباط البسيط لبيرسون (r_p) للقيم الاصلية للمتغيرين (X) و(Y) على النحو الآتي :

X	Y	XY	X^2	Y^2
5	6	30	25	36
1	3	3	1	9
0	2	0	0	4
2	5	10	4	25
8	16	43	30	74
$\bar{X} = 2$	$\bar{Y} = 4$	$\sum XY$	$\sum X^2$	$\sum Y^2$

نقوم بايجاد قيم كل من [S_{xy} ، S_{xx} ، S_{yy}] على النحو الآتي :

$$S_{xy} = \Sigma XY - n\bar{X}\bar{Y}$$
$$= 43 - 4(2)(4) = 43 - 32 = 11$$
$$S_{xx} = \Sigma X^2 - n\bar{X}^2$$
$$= 30 - 4(2)^2 = 30 - 16 = 14$$
$$S_{yy} = \Sigma Y^2 - n\bar{Y}^2$$
$$= 74 - 4(4)^2 = 74 - 64 = 10$$

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$
$$= \frac{11}{\sqrt{14}\sqrt{10}} = \frac{11}{(3.742)(3.162} = \frac{11}{11.832}$$
$$= +0.93$$

يتضح ان العلاقة طردية وقوية جداً

ب- نقوم بحساب معامل الارتباط البسيط لبرسون (r_p) للقيم الجديدة للمتغيرين (X^*) و (Y^*) والتي يمكن الحصول عليها من خلال اجراء التحويل التالى وفقاً للعلاقتين الآتيتين:

$X^* = 3X$ ، $Y^* = Y + 2$

X^*	Y^*	X^*Y^*	X^{*2}	Y^{*2}
15	8	120	225	64
3	5	15	9	25
0	4	0	0	16
6	7	42	36	49
24	24	177	270	154
$\bar{X}^* = 6$	$\bar{Y}^* = 6$	ΣX^*Y^*	ΣX^{*2}	ΣY^{*2}

$$\therefore S_{xy} = \Sigma X^*Y^* - n\bar{X}^*\bar{Y}^*$$
$$= 177 - 4(6)(6)$$
$$= 33$$

$$S_{xx} = \sum X^{*2} - n\bar{X}^{*2}$$
$$= 270 - 4(6)^2 = 126$$
$$S_{yy} = \sum Y^{*2} - n\bar{Y}^{*2}$$
$$= 154 - 4(6)^2 = 10$$

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{33}{\sqrt{126}\sqrt{10}}$$

$$= \frac{33}{(11.225)(3.162)} = +0.93$$

نلاحظ من النتيجة اعلاه انها مساوية تماماً إلى نتيجة معامل الارتباط البسيط (r_p) للقيم الاصلية، مما يؤكد بأن قيمة معامل الارتباط البسيط لاتتأثر بالتحويل المستخدم.

مثال: اوجد معامل الارتباط بين المتغيرين (X) و (Y) حسب البيانات الآتية: (4)

X :	480	490	500	510	520
Y :	280	290	330	290	310

بأستخدام الصيغة المباشرة:

الحل:- نقوم بتكوين الجدول الآتي:

X	Y	XY	X^2	Y^2
480	280	134400	230400	78400
490	290	142100	240100	84100
500	330	165000	250000	108900
510	290	147900	260100	84100
520	310	161200	270400	46100
2500	1500	750600	1251000	451600

$\bar{X} = 500$ ، $\bar{Y} = 300$

$$r_p = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{5(750600) - 2500(1500)}{\sqrt{[5(1251000) - (2500)^2][5(451600) - (1500)^2]}}$$

$$= \frac{3753000 - 3750000}{[6255000 - 6250000][2258000 - 2250000]}$$

$$= \frac{3000}{\sqrt{40000000}} = \frac{3000}{6324.555}$$

$r_p = +0.47$ العلاقة طردية قوية

الانحدار الخطي
Linear Regression

يعرف الانحدار او تحليل الانحدار (regression analysis) بأنه "مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات المعتمدة في العلاقة ، وغالباً ما تسمى العلاقات من هذا النوع بنماذج الانحدار .

يهدف تحليل الانحدار الى تقدير العلاقة الرياضية الخطية التي تربط متغيرين أو أكثر من بعضها ، وفي تحليل الانحدار يوجد نوعين من المتغيرات هما :

1- المتغير التابع (dependent Variable) : هو المتغير الذي تتأثر قيمه في حالة تغير قيم المتغير المستقل ولا يتأثر أو تتأثر به ، مثال ذلك : ان انفاق الاسرة الشهري يتأثر بدخلها الشهري في حين ان العكس غير منطقي ، مثال آخر : فأن سرعة الانجاز لسباح تتأثر بمساحة كف يده وطولها في حين ان العكس غير منطقي .

2 المتغير المستقل (independent Variable) ويسمى بالمتغير المتنبأ به predictor أو المتغير التوضيحي explained Variable .

أن الانحدار الخطي له صلة وثيقة بمفهوم الارتباط الخطي :

الانحدار الخطي : يهتم بتقدير العلاقة الخطية بين المتغيرات .

الارتباط الخطي : يهتم بتقدير القيمة العددية للعلاقة نفسها .

تنقسم العلاقات الخطية بين المتغيرات الى قسمين رئيسين هما :

1- العلاقات المحددة deteministic relations : وتعني ان التغير الحاصل في متغير تابع هو نتيجة لتغير متغير مستقل او (مجموعة متغيرات مستقلة) محدد أو (محددة) ولا يوجد غيره او (غيرها) .

مثال ذلك : العلاقة بين (المسافة - السرعة - الزمن) هي علاقة محددة بحيث يمكن ايجاد اي منها بدلالة الآخرين ، فالسرعة المطلوبة لتقطع مسافة محددة ما هي الا حاصل قسمة المسافة على الزمن المستغرق ، كذلك فأن الزمن المطلوب لتقطع تلك المسافة ما هو الا حاصل قسمة المسافة على السرعة . وهذا يعني ان لكل قيمة من قيم المتغير المستقل يقابلها قيمة واحدة فقط للمتغير التابع .

2 - العلاقات العشوائية Random relations : ان العلاقة العشوائية تعني ان التغير الحاصل في متغير تابع ليس نتيجة لتغير متغير مستقل او (مجموعة متغيرات مستقلة) فحسب بل الى جملة متغيرات أخرى مجهولة غير مسيطر عليها من قبل الباحث.

مثال ذلك: الانفاق الشهري للاسرة لا يتأثر فقط بدخلها الشهري وعدد افرادها فحسب بل هناك جملة من المتغيرات الأخرى المؤثرة على الانفاق منها (الذوق - العادات والتقاليد - العلاقات الاجتماعية مع أسرة أخرى، وغيرها)، وهذا يعني ان لكل قيمة من قيم المتغير المستقل او (المتغيرات المستقلة) سوف لن يقابلها قيمة واحدة فقط للمتغير التابع وانما مجموعة من القيم التي كل منها يتأثر بواحد أو أكثر من المتغيرات المجهولة. أن موضوع الانحدار يعالج العلاقات العشوائية وليس العلاقات المحددة، فأن العلاقة الخطية بين المتغيرات تعني عملية ربط هذه المتغيرات بنموذج أنحدار يتضح فيه المتغير التابع (المتأثر) والمتغير المستقل (المؤثر) او مجموعة متغيرات مستقلة (المؤثرة).

الانحدار الخطي البسيط
Simple Linear Regression

يعرف الانحدار الخطي البسيط ؛ بانه عملية تقدير العلاقة الخطية بين متغيرين ، أحدهما مستقل (مؤثر) والآخر تابع (المؤثر به).

الخطي : يعني ان العلاقة بين المتغيرين (Y و X) علاقة خطية معبر عنها بخط مستقيم يقطع المحور الافقي والعمودي.
البسيط ؛ يعني ان المتغير التابع (Y) يعتمد على المتغير المستقل (متغير واحد) وهو (X).
تحليل الانحدار الخطي : هو دالة رياضية يمكن منها حساب أحد المتغيرين بدلالة المتغير الآخر ، وأبسط حالة لهذه الدالة عندما يكون لدينا متغير (مستقل) واحد فقط يرتبط مع متغير (التابع) بعلاقة خط مستقيم . وهذا يعني ان هنالك دالة رياضية تجمع هذه المتغيرات مع بعضها ، وهذه الدوال (المعادلات) كمايلي:
معادلات الانحدار الخطي البسيط :

1- معادلة انحدار Y على X (Y/X) تعني ⇐ $Y = f(X)$

2- معادلة انحدار X على Y (X/Y) تعني ⇐ $X = f(Y)$

فاذا كانت العلاقة على شكل (Y / X) فهذا يعني دالة $Y = f(X)$ ، حيث ان:

X : تمثل المتغير العشوائي المستقل .

Y : تمثل المتغير العشوائي التابع .

الهدف من الانحدار الخطي البسيط :

1- تقدير قيم عددية لمعالم نموذج الانحدار الذي يتكون من (a , b).

2- الحصول على صيغة رياضية توضح العلاقة بين متغيرين إن وجدت مما يساعد في التقدير (التنبؤ) وتأخذ العلاقة شكل (خطية — غير خطية).

ان ايجاد معالم النموذج (تقدير قيم (a , b)) نحن نتكلم عن علاقة عشوائية تربط متغيرين ، مما يعني بهدف ايجاد خط مستقيم يوضح هذه العلاقة التي تحدد شكل واتجاه هذا المستقيم .

ان أهم الطرق لايجاد ذلك المستقيم هي :

طريقة المربعات الصغرى Least Squares : هي طريقة لتطبيق خط مستقيم على مجموعة من النقاط $(X_1, Y_1), (X_2, Y_2) \cdots (X_n, Y_n)$ ، بحيث يكون مجموع مربعات الاخطاء (($\sum (Y_i - \hat{Y})^2$)) أقل ما يمكن ، وذلك من خلال نموذج الانحدار الخطي في شكل المعادلة الخطية الآتية :

نموذج الانحدار الخطي :- هو دالة رياضية يمكن من خلالها حساب أحد المتغيرين بدلالة الأخر ، وابسط حالة لهذه الدالة عندما يكون لدينا متغير مستقل واحد فقط يرتبط مع متغير تابع بعلاقة خط مستقيم وكمايلي :

$$Y = a + bX + e$$

Y :- وتمثل المتغير التابع (الذي يتأثر)

a :- قيمة الثابت (ثابت الانحدار) وتمثل الجزء المقطوع من المحور الرأسي (الصادي) (Y) ، الذي يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل X ، في حالة $X = 0$.

b :- ميل الخط المستقيم (معامل الانحدار) يعكس مقدار التغير في Y اذا تغيرت X وحدة واحدة.

X :- المتغير المستقل (المؤثر) .

e :- (Error) الذي يمثل البواقي (الخطأ العشوائي) العوامل الأخرى التي تؤثر في المتغير (Y) غير المتغير المستقل (X) ، الفرق بين القيمة الفعلية لـ (Y) والقيمة المقدرة ($\hat{Y}$) .

حيث يمكن ايجاد قيم كل من (e, b, a) ، كمايلي :

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

$$a = \bar{Y} - b\bar{X}$$

$$e = Y - \hat{Y}$$

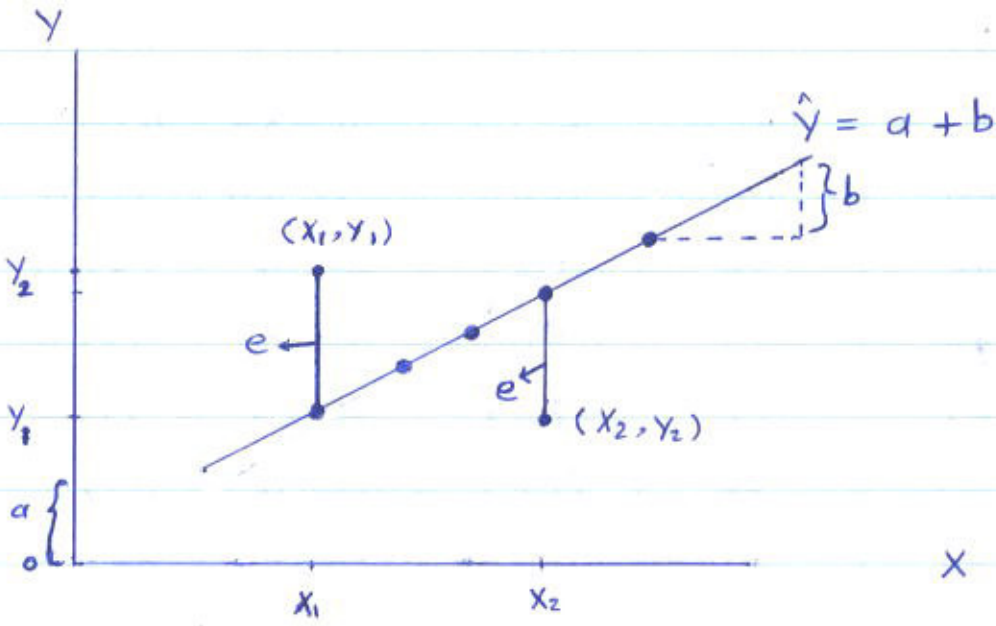
حيث إن :

$\bar{X}$: تمثل الوسط الحسابي لمشاهدات المتغير المستقل (X) $\quad X_1 , X_2 , \dots , X_n$

$\bar{Y}$: تمثل الوسط الحسابي لمشاهدات المتغير التابع (Y) $\quad Y_1 , Y_2 , \dots , Y_n$

وبتعويض قيمتي (a ، b) في نموذج او معادلة خط الانحدار $\hat{Y} = a + bX$

نحصل على معادلة أنحدار (Y / X) ، اي ايجاد معادلة انحدار المتغير المستقل X والمتغير التابع Y أو بالعكس (X / Y).



! – نموذج الانحدار الخطي بيانياً –

النموذج المفترض :

$$Y = a + bX + e$$

هو نموذج نظري نفترض انه يحكم العلاقة بين متغيرين (X , Y) كما في المعادلة أعلاه.

النموذج المقدر :

$$\hat{Y} = a + bX$$

هو نموذج فعلي يقدره الباحث من خلال جمع (العينة) بيانات عن المتغيرين (Y , X) ومن ثم تقدير (حساب) معاملات الانحدار (a ، b) الثابت والميل كما في المعادلة أعلاه.

وبعد تقدير النموذج يمكن حساب الخطأ العشوائي من خلال :

$$e = Y - \hat{Y}$$

حيث إن :

Y : تمثل القيمة الحقيقية من قيم المتغير التابع (القيمة المقابلة او المناظرة).

$\hat{Y}$: تمثل القيمة المقدرة (المتنبأ بها).

مثال (١) : أحسب معادلة الانحدار (Y/X) بأستخدام طريقة المربعات الصغرى لدرجات ثمانية طلبة في مقرري مادة الاحصاء و الرياضيات .

X درجات الاحصاء	85	65	60	45	80	65	85	95
Y درجات الرياضيات	77	67	62	52	72	57	82	87

الحل :- نعمل الجدول الآتي :

X	Y	X^2	Y^2	XY
85	77	7225	5929	6545
65	67	4225	4489	4355
60	62	3600	3844	3720
45	52	2025	2704	2340
80	72	6400	5148	5760
65	57	4225	3249	3705
85	82	7225	6724	6970
95	87	9025	7569	8265
580	556	43950	39656	41660

$$\bar{X} = \frac{\Sigma X}{n} = \frac{580}{8} = 72.5$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{556}{8} = 69.5$$

$$b = \frac{\Sigma XY - n\,\bar{X}\,\bar{Y}}{\Sigma X^2 - n\,\bar{X}^2} = \frac{41660 - 8(72.5)(69.5)}{43950 - 8(72.5)^2}$$

$$b = \frac{1350}{1900} = 0.71$$

$$a = \bar{Y} - b\,\bar{X} = 69.5 - 0.71(72.5)$$

$$= 18.025$$

ت معادلة انحدار (Y / X) هي كمايلي:

$$\hat{Y} = a + bX$$
$$= 18.025 + 0.71X$$

وفي حالة طلب السؤال تقدير درجة الرياضيات اذا كانت درجة الاحصاء (80) فيمكن تعويض $X = 80$ ، في المعادلة وايجاد قيمة $\hat{Y}$ التقديرية وكمايلي:

$$\hat{Y} = 18.025 + 0.71(80) = 74.8 \simeq 75$$

وهذا يعني تنبوء او تقدير درجة الرياضيات عندما $X = 80$.

أما الخطأ العشوائي (المعياري) في التقدير فهو كمايلي:

$$e = Y - \hat{Y} = 72 - 75 = -3$$

القيمة المقدرة المتنأبها ← ← القيمة المشاهدة المقابلة للدرجة 80 في قيم Y

مثال (٢): اوجد معادلة انحدار (Y / X) للبيانات الموضحة في الجدول الآتي:

X	4	10	9	12	8	5
Y	2	6	8	11	5	4

المطلوب:

2- ايجاد تقدير قيمة Y المقابلة عندما $X = 9$.

3- ايجاد الخطأ العشوائي في تقدير Y عندما تكون قيمة $X = 9$.

الحل:- نعمل الجدول الآتي:

X	Y	XY	X^2
4	2	8	16
10	6	60	100
9	8	72	81
12	11	132	144
8	5	40	64
5	4	20	25
48	36	232	430

1- ايجاد معادلة انحدار (Y/X) حسب المعادلة الآتية:

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

$$b = \frac{\Sigma xy - n\bar{x}\bar{y}}{\Sigma x^2 - n\bar{x}^2} \quad ، \quad a = \bar{y} - b\bar{x}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{48}{6} = 8 \quad ، \quad \bar{y} = \frac{\Sigma y}{n} = \frac{36}{6} = 6$$

$$b = \frac{332 - 6(8)(6)}{430 - 6(8)^2} = 0.96$$

$$a = 6 - 0.96(8) = -1.68$$

$$\therefore \hat{y} = -1.68 + 0.96x$$

2- القيمة التقديرية (المتنبأ بها) للمتغير y عندما $x = 9$ ، هي تعويضها في المعادلة:

$$\hat{y} = -1.68 + 0.96(9) = 6.96$$

3- من خلال ملاحظة الجدول السابق يتبين لنا ان مشاهدات المتغير (x) ومشاهدات المتغير (y) المقابلة لها، نجد ان المشاهدة التي تقابل $x = 9$ هي $y = 8$ وقد وجدنا في (2) تقدير قيمة $\hat{y}$ عندما $x = 9$ وهي $\hat{y} = 6.96$ نقوم بتعويضها بالمعادلة الآتية:

$$e = y - \hat{y}$$
$$= 8 - 6.96 = 1.04$$

∴ الخطأ العشوائي في التقدير هو: 1.04

مثال (3) يعطي الجدول الآتي معطيات 12 مشاهدة من (X ، Y) وكالآتي :

X	18	14	10	15	7	12	13	8	9	17	15	12
Y	20	11	14	16	10	10	17	11	12	20	18	12

المطلوب :

1- رسم لوحة شكل الانتشار .

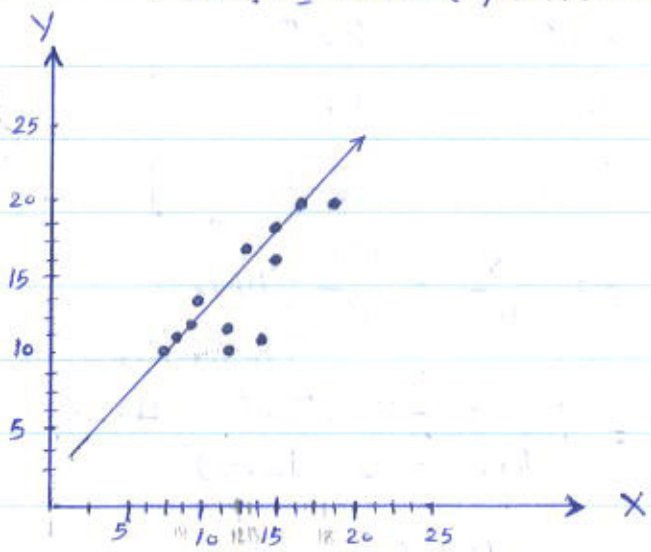
2- ايجد معادلة خط الانحدار (Y / X).

3- اذا كانت قيمة X=16 ، ولم توجد قيمة مناظرة لها من قيم Y ، فما هي القيمة التقديرية لـ $\hat{y}$.

4- ما هو الخطأ العشوائي في تقدير قيمة Y اذا كانت قيمة X=10 .

الحل :-

1- نحصل على لوحة الانتشار برصد النقاط لـ (X ، Y) المعطاة في الجدول .



2- معادلة خط انحدار (Y / X) نعمل الجدول الآتي :

X	Y	XY	X^2
18	20	360	324
14	11	154	196
10	14	140	100
15	16	240	225
7	10	70	49
12	10	120	144
13	17	221	169
8	11	88	64
9	12	108	81
17	20	340	289
15	18	270	225
12	12	144	144
150	171	2255	2010

$$\bar{X} = \frac{\Sigma X}{n} = \frac{150}{12} = 12.5 \quad , \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{171}{12} = 14.25$$

وبالتعويض بالمعادلة :

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{2255 - 12(12.5)(14.25)}{2010 - 12(12.5)^2}$$

$$b = \frac{225 - 2137.5}{2010 - 1875} = \frac{117.5}{135} = 0.87$$

$$a = \bar{Y} - b\bar{X} = 14.25 - 0.87(12.5)$$

$$= 3.37$$

وبالتعويض في معادلة خط الانحدار :

$$\hat{Y} = a + bX$$

$$= 3.37 + 0.87X$$

3- نعوض قيمة $X=16$ في معادلة خط الانحدار فنجد القيمة التقديرية لقيمة $\hat{Y}$.

$$\hat{Y} = 3.37 + 0.87(16)$$

$$= 17.29 \simeq 17$$

4 - نعوض قيمة $X=10$ في معادلة خط الانحدار فنجد قيمة $\hat{Y}$ التقديرية ومن ثم نجد قيمة الخطأ العشوائي في التقدير عندما $X=10$.

$$\hat{Y} = 3.37 + 0.87\,(10) = 12.07 \simeq 12$$

$$\therefore e = Y - \hat{Y} = 14 - 12 = 2 \quad \Leftarrow \text{الخطأ العشوائي هو}$$

نلاحظ إن قيمة $X=16$ هي قيمة تقديرية افتراضية لا وجود لها ضمن قيم المتغير X ولا تقابلها قيمة مناظرة (مقابلة) في قيم المتغير Y .

مثال (4) : البيانات التالية تمثل الكمية المطلوبة Y من سلعة معينة وسعر الوحدة الواحدة منها X ، والمطلوب تقدير معادلة خط انحدار (Y/X).

Y	3	5	6	4	6	4	9	8	9	6
X	11	8	7	8	6	9	5	5	4	7

الحل :- نقوم بعمل الجدول الآتي :

X	Y	XY	X^2
11	3	33	121
8	5	40	64
7	6	42	49
8	4	32	64
6	6	36	36
9	4	36	81
5	9	45	25
5	8	40	25
4	9	36	16
7	6	42	49
70	60	382	530

$$\bar{X} = \frac{\Sigma X}{n} = \frac{70}{10} = 7 \quad ، \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{60}{10} = 6$$

نحدد قيمة كل من (a ، b) وكما يلي:

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{382 - 10(7)(6)}{530 - 10(7)^2}$$

$$= \frac{-38}{40} = -0.95$$

$$a = \bar{Y} - b\bar{X} = 6 - (-0.95)(7)$$
$$= 12.65$$

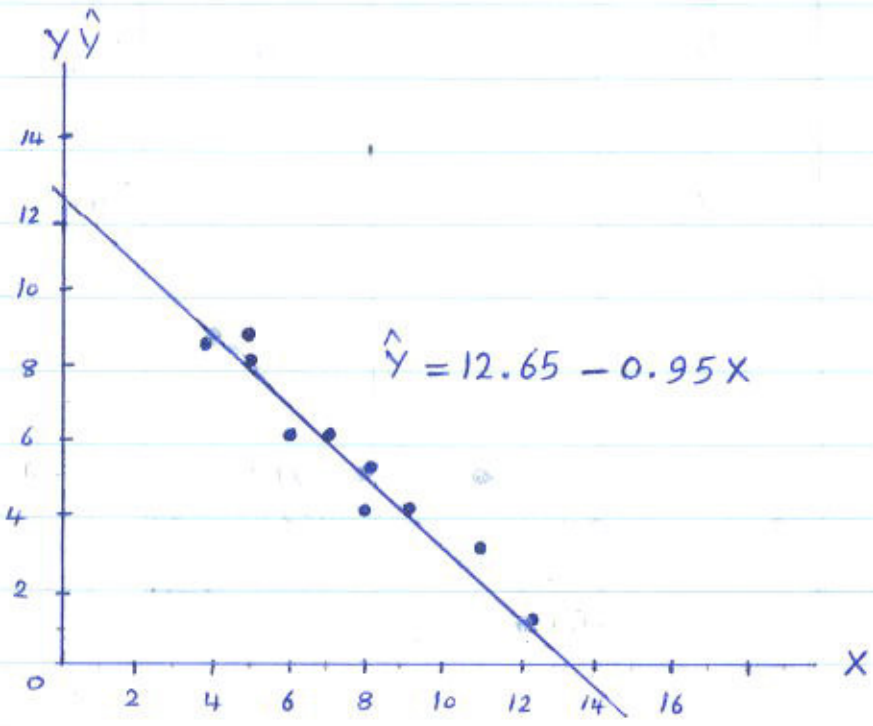
$$\therefore \quad \hat{Y} = a + bX$$
$$= 12.65 + (-0.95)X$$
$$= 12.65 - 0.95X$$

نلاحظ ان العلاقة بين المتغير X و Y علاقة عكسية (سالبة)، وهذا يفسر إنه كلما زاد السعر بمقدار (0.95) دينار قلت الكمية المطلوبة بمقدار (12.65).

ولغرض رسم المعادلة اعلاه نختار ثلاثة قيم الى (X) ونعوضها في المعادلة $\hat{Y}$ ويتم تعيينها في المستوي (X Y) وكما يلي،

$X = 4$

$\hat{Y} = 12.65 - 0.95(4) = 8.85$ ، $\Rightarrow$ ($X = 8 \Rightarrow 5.05$ ، $X = 12 \Rightarrow 1.25$) افتراضية



X Y
(4 ، 8.85)
(8 ، 5.05)
(12 ، 1.25)

مثال (5) : البيانات التالية تمثل درجات 12 طالب في امتحان مادة الرياضيات (X) وامتحان مادة الاحصاء (Y) وان درجة الامتحان القصوى هي 10 درجات .

المطلوب : ايجاد معادلة خط انحدار (Y/X) ، (X/Y) .

X مادة الرياضيات	2	3	9	8	7	10	5	6	3	6	0	1
Y مادة الاحصاء	0	2	7	7	5	9	3	6	4	4	1	0

الحل : نقوم بعمل الجدول الآتي :

X	Y	XY	X^2	Y^2
2	0	0	4	0
3	2	6	9	4
9	7	63	81	49
8	7	56	64	49
7	5	35	49	25
10	9	90	100	81
5	3	15	25	9
6	6	36	36	36
3	4	12	9	16
6	4	24	36	16
0	1	0	0	1
1	0	0	1	0
60	48	337	414	286

$$\bar{X} = \frac{\Sigma X}{n} = \frac{60}{12} = 5 \quad , \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{48}{12} = 4$$

معادلة خط انحدار (Y/X) :

$$b = \frac{337 - 12(5)(4)}{414 - 12(5)^2} = \frac{97}{114} = 0.85$$

$$a = 4 - 0.85(5) = -0.25$$

$$\therefore \hat{y} = a + bx$$
$$= -0.25 + 0.85x$$

2- معادلة خط انحدار (X/Y):

$$b = \frac{\sum xy - n\bar{y}\bar{x}}{\sum y^2 - n\bar{y}^2} = \frac{337 - 12(4)(5)}{286 - 12(4)^2}$$

$$= \frac{97}{94} = 1.03$$

$$a = \bar{x} - b\bar{y} = 5 - 1.03(4) = 0.88$$

$$\therefore \hat{x} = a + by$$
$$= 0.88 + 1.03y$$

لرسم معادلتي خط الانحدار اعلاه في نفس شكل الانتشار، لمعادلة Y/X نختار ثلاثة قيم الى X ونعوضها في معادلة $\hat{y}$ لغرض ايجاد قيم y المقابلة.

X : 1 , 5 , 10

$\hat{Y}$: 0.6 , 4 , 8.25

وكذلك بالنسبة لمعادلة خط انحدار X/Y.

Y : 1 , 5 , 8 ← فرضية

$\hat{X}$: 1.91 , 6.03 , 9.12