استكمالا للمحاضرات السابقة

العلاقة بين الانحراف المعياري والانحراف المتوسط

اذا كان التوزيع غير متماثل عندها

$$\text{mean deviation} = \frac{4}{5}(\text{standard deviation})$$

$$MD = \frac{4}{5}S$$

ملاحظه : جميع مقاييس التشتت التي اخذناها في الكورس الاول تسمى مقاييس التشتت المطلقة .

النوع الثاني من مقاييس التشتت هو :

مقاييس التشتت النسبية Relative dispersion measures

تعرف مقاييس التشتت السابقة بانها مقاييس تشتت مطلقة , اما مقاييس التشتت النسبية التي سنتطرق لها في هذه المحاضرة هي مقاييس التباين لمجموعه من القيم بغض النظر عن وحدة القياس الخاصة بها وهي تفيد للمقارنات بين مجموعة بيانات او تجارب مختلفه بالقياس

1- معامل الاختلاف Coefficient of Variance (C.V)

افرض ان { $\bar{x}$ } يمثل الوسط الحسابي لمجموعة قيم وان { S } يمثل الانحراف المعياري لها، عندئذ يعرف معامل الاختلاف على النحو التالي:

$$C.V. = \frac{S}{\bar{X}} * 100$$

ان معامل الاختلاف يعتبر أفضل انواع معاملات التشتت كونه يعتمد على أفضل مقياس نزعة مركزية وأفضل مقياس تشتت وعند اجراء مقارنة بين قيم مجموعتين تتم مقارنة معامل اختلاف الأولى مع معامل اختلاف الثانية وعندئذ يقال عن المجموعة بانها أكثر تجانسا إذا كان معامل اختلافها اقل من الأخرى.

مثال: إذا كان متوسط درجات طلبة الصف الأول في امتحان مادة الرياضيات 69 درجة وبانحراف معياري قدره 19.3 في حين كان متوسط درجاتهم في امتحان الإحصاء 75 درجة بانحراف معياري قدره 25.5 في أي من الامتحانين كان مستوى أداء الطلبة أكثر تقاربا.

الحل: نجد معامل الاختلاف لكل امتحان

$$C.V.(stat) = \frac{S}{\bar{X}} * 100 = \frac{25.5}{75} * 100 = 34\%$$

$$C.V.(math) = \frac{S}{\bar{y}} * 100 = \frac{19.3}{69} * 100 = 28\%$$

وحيث ان معامل الاختلاف في امتحان الرياضيات اقل من معامل الاختلاف في امتحان الإحصاء وعليه فان مستوى أداء الطلبة في امتحان الرياضيات كان اكثر تقاربا .

مثال 2:
الجدول التالي يمثل اوزان واطوال خمسة اشخاص تحت الدراسة المطلوب بيان اي من مجموعتي البيانات اكثر تشتت (اقل تجانس) مجموعة الوزن او الطول ؟

عدد الاشخاص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:
نستخدم معامل الاختلاف للمقارنة
يتم ايجاد الوسط الحسابي بالصيغة االتاليه

$$\bar{X} = \frac{\sum X_i}{n}$$

والانحراف المعياري

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$$

لكلا المجموعتين وبعدها يتم حساب معامل الاختلاف وكما موضح في الجدول التالي

البيانات	الوسط الحسابي $\bar{X}$	الانحراف المعياري s	$C.V. = \frac{S}{\bar{X}} * 100$
الوزن	65	3.7417	5.75%
الطول	160	4.2071	2.62%

بما انه معامل الاختلاف لبيانات الاوزان اكبر من معامل اختلاف بيانات الطول فان التشتت النسبي لبيانات الاوزان اكبر من التشتت النسبي لبيانات الاطوال . اي ان بيانات الاوزان اقل تجانس من بيانات الاوزان .

2- الدرجة المعيارية او القياسية Standard Score :

في كثير من الأحيان نحتاج الى تحويل قيم المتغير العشوائي X الى شكل اخر يدعى الشكل المعياري او الشكل القياسي لهذه القيم، وتستخدم الدرجة المعيارية للمقارنة بين مفردتين من مجموعتين مختلفتين , حيث كلما كانت قيمة الدرجة المعيارية اكبر كانت النتيجة افضل والعكس صحيح .
وتعرف الدرجة المعيارية Z لأي قيمة من قيم X على النحو التالي:

$$Z_i = \frac{x_i - \bar{x}}{s} \qquad ، \quad i = 1، 2، 3، \ldots ، n$$

مثال: حصل طالب على درجة 84 في الامتحان النهائي بالرياضيات علما بان الوسط الحسابي في امتحان الرياضيات لجميع الطلبة هو 76 وبانحراف قياسي 10 اما في امتحان الفيزياء فكانت درجة نفس الطالب هي 90 وبوسط حسابي 82 وانحراف قياسي 16، في أي الموضوعين كانت قابلية هذا الطالب اعلى؟

الحل: عند مقارنة درجة الامتحانين مباشرة نجد ان درجته في الفيزياء (90) اعلى من درجته في الرياضيات (84) . ولكن عند تحويل هاتين الدرجتين الى درجات قياسية نجد ان:

$$Z_m = \frac{84-76}{10} = 0.8$$ للرياضيات

$$Z_f = \frac{90-82}{16} = 0.5$$ للفيزياء

ومن هذا يتضح ان قابليته في الرياضيات اعلى من الفيزياء وهو عكس ما توصلت اليه المقارنة السابقة.

Correlation الارتباط

الارتباط هو طريقة احصائية تستخدم لتحديد ما اذا كانت هنالك علاقة بين المتغيرات مثلا هل هناك علاقة بين التدخين ومرض معين في الرئة .

Scatter Plot رسم الانتشار

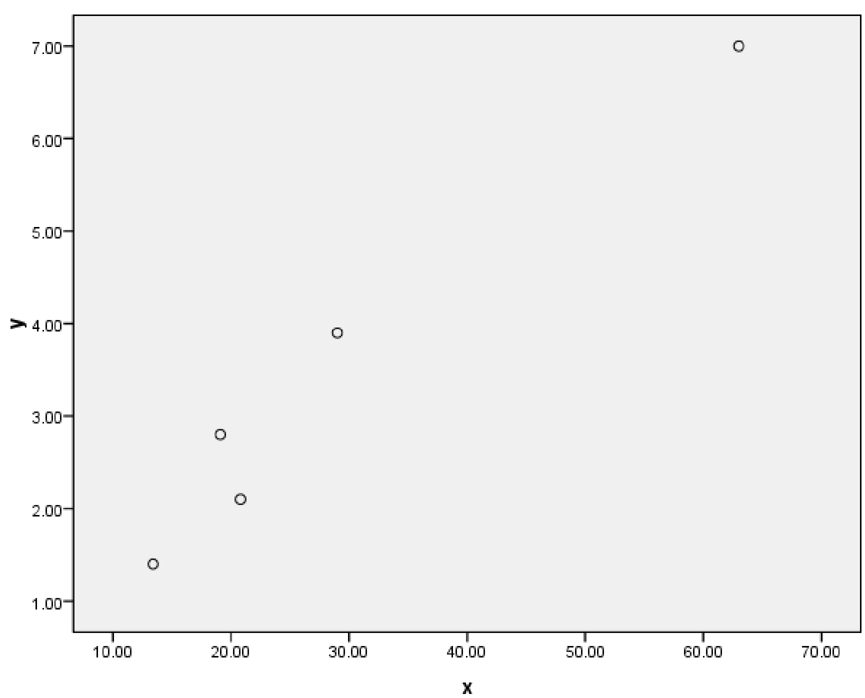
رسم الانتشار

وسيلة مبدئية يعرف الباحث من خلالها وجود ارتباط او عدمه ونوع الارتباط بين المتغيرين ان وجد **,** وهو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى.

مثال : انشئ مخطط الانتشار لبيانات المتغيرين الموضحة بالجدول ادناه

8.5	13.4	19.1	20.8	29	63	X
1.5	1.4	2.8	2.1	3.9	7	Y

الحل



1-معامل الارتباط البسيط Simple correlation coefficient :

هو مقياس لدرجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة . ويرمز له بالرمز r_{xy} ، r_{yx} او r

مميزات معامل الارتباط

1- تنحصر قيمة معامل الارتباط بين $+1$ و -1

2- تشير قيمة معامل الارتباط r الى قوة العلاقة الخطية بين المتغيرين , بينما تشير (اشارة) r الى اتجاه العلاقة بين المتغيرين وكما موضح ادناه :

- اذا كانت قيمة (r=0) هذا يدل على عدم وجود ارتباط خطي بين المتغيرين x وy .
- اذا كانت قيمة r=1 فهذا يدل على وجود ارتباط موجب تام (ارتباط طردي تام)
- اذا كانت قيمة r تقع (من 0.1 الى اقل من 0.4) فهذا يدل على وجود ارتباط طردي ضعيف.
- اذا كانت قيمة r تقع بين (0.4 الى اقل من 0.7) فهذا يدل على وجود ارتباط طردي متوسط.
- اذا كانت قيمة r تقع بين (0.7 الى اقل من 1) فهذا يدل على وجود ارتباط طردي قوي
- اذا كانت r=-1 فهذا يدل على وجود ارتباط سالب تام (ارتباط عكسي تام)
- اذا كانت قيمة r تقع (من- 0.1 الى اقل من- 0.4) فهذا يدل على وجود ارتباط عكسي ضعيف
- اذا كانت قيمة r تقع بين (-0.4 الى اقل من -0.7) فهذا يدل على وجود ارتباط عكسي متوسط
- اذا كانت قيمة r تقع بين (0.7 - الى اقل من 1-) فهذا يدل على وجود ارتباط عكسي قوي.

ملاحظة

الارتباط الموجب (الطردي) positive correlation يعني زيادة في احد المتغيرين تؤدي الى زيادة في المتغير الثاني

الارتباط السالب (العكسي) Negative correlation يعني زيادة في احد المتغيرين تؤدي الى نقصان في المتغير الثاني

الارتباط ورسم الانتشار

من رسم الانتشار للبيانات وطريقة انتشار القيم نستدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. كما موضح في بعض الاشكال ادناه

الشكل الأول:

إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة هذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية.
أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد) فإن "الارتباط عكسي تام " ومثال على ذلك العلاقة بين السرعة والزمن.

الشكل الثاني:

أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط فتكون العلاقة خطية (موجبة أو سالبة)

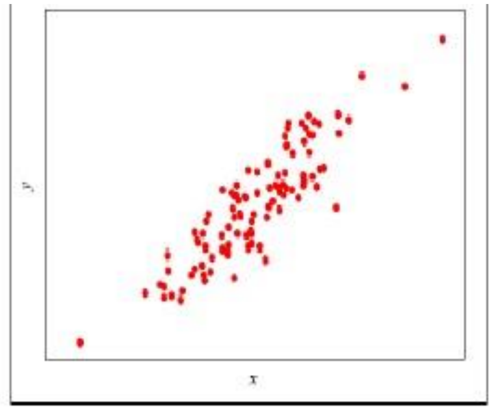
الشكل الثالث:

وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي Non Linear Correlation ":

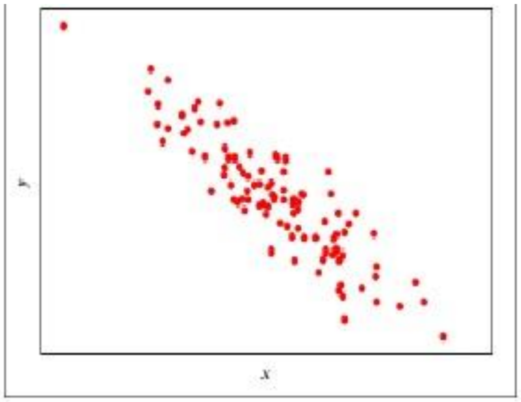
الشكل الرابع:

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله .

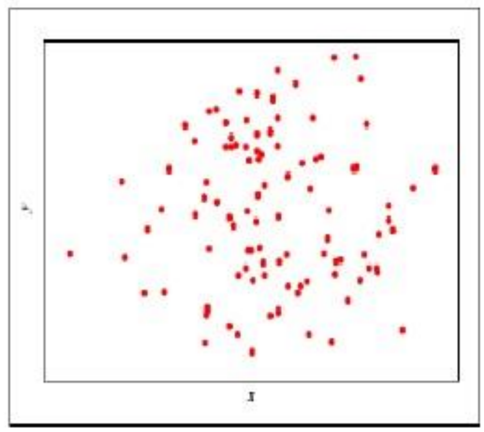
كما موضح بالرسوم ادناه



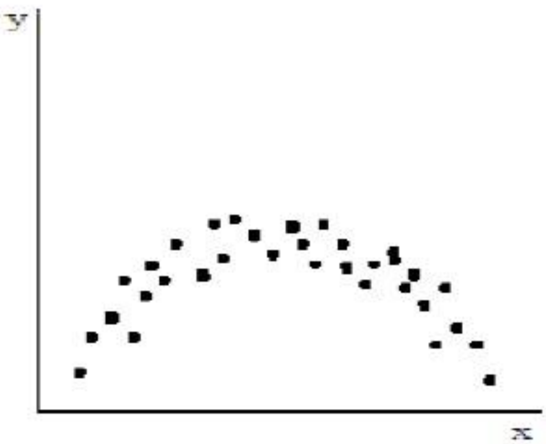
شكل الانتشار الخاص بالارتباط
الموجب (الطردي)



شكل الانتشار الخاص بالارتباط السالب
(العكسي)

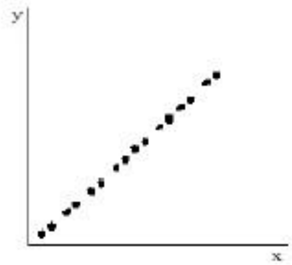


شكل الانتشار الخاص باستقلال
متغيرين (ظاهرتين)

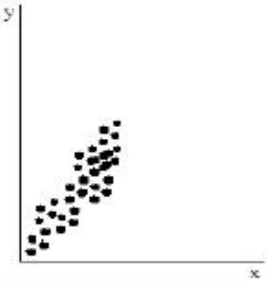


شكل الانتشار الخاص بالعلاقه الغير خطيه
بين متغيرين (ظاهرتين)

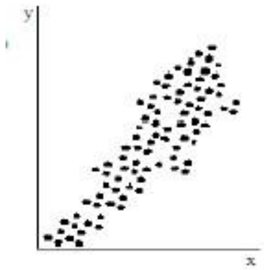
شكل الانتشار Scatter Plot



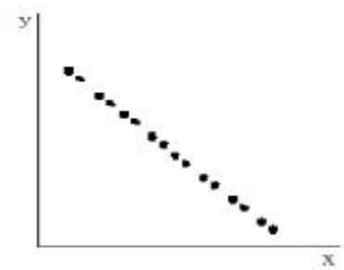
ارتباط طردي تام



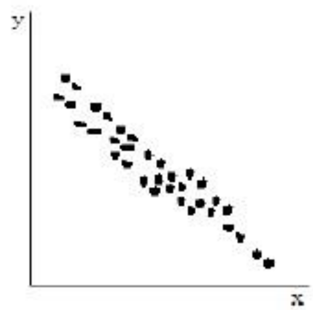
ارتباط طردي قوي



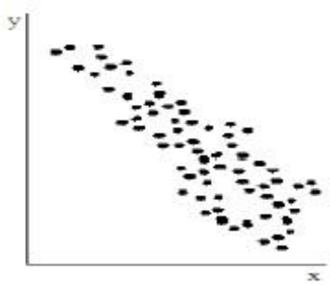
ارتباط طردي



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي

حساب معامل الارتباط البسيط

اذا كانت العلاقة بيم المتغيرين x , y علاقة خطية فإننا نستطيع تحديد درجة الارتباط بينهما باستخدام الصيغة التالية

$$r_{xy}=\frac{\operatorname{cov}(x,y)}{S_xS_y}=\frac{S_{xy}}{S_xS_y}\qquad ......(*)$$

حيث ان s_{xy} تمثل التباين المشترك بين المتغيرين x,y ويحسب بالصيغة التالية

$$\operatorname{cov}(x,y)=S_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}$$

وان s_x,s_y تمثل الانحراف المعياري للمتغيرين x,y

حيث ان

$$s_x=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$$

$$s_y=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}}$$

ملاحظة: كل من s_x,s_y موجبان دائما , فأن اشارة معامل الارتباط تعتمد على اشارة s_{xy} فاذا كانت موجبة فأن اشارة معامل الارتباط موجبة واذا كانت s_{xy} سالبة فان معامل الارتباط سالب

الان بالتعويض في معادلة (*) نحصل على صيغة معامل الارتباط كالاتي

$$r_{xy}=\frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}}}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

او الصيغه الثانية

$$r_{xy}=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sqrt{\sum x_i^2-n\bar{x}^2}\sqrt{y_i^2-n\bar{y}^2}}$$

او الصيغة الثالثة

$$r_{xy}=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-(\sum x_i)^2}\sqrt{n\sum y_i^2-(\sum y_i)^2}}$$

لحساب معامل الارتباط r يمكن استخدام احدى هذه الصيغ .

مثال: البيانات التالية تمثل الكمية المعروضة من سلعة معينة وسعر الوحدة الواحدة منها، يطلب حساب معامل الارتباط الخطي البسيط بين الكمية المعروضة والسعر ثم تفسير النتيجة بعد ذلك.

السعر (x)	2	3	5	4	5	6	3	5	4
الكمية المعروضة (y)	3	5	7	8	9	11	6	8	6

الحل/ سيتم استخدام الصيغة الاولى والصيغة الثالثة لحل المثال

الصيغة الاولى

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$\bar{x} = 4 \qquad \bar{y} = 7$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$		x_i^2	y_i^2	$x_i y_i$
2	***3***	-2	-4	4	16	8		4	9	6
2	***5***	-2	-2	4	4	4		4	25	10
5	***7***	1	0	1	0	0		25	49	35
4	***8***	0	1	0	1	0		16	64	32
5	***9***	1	2	1	4	2	or	25	81	45
6	***11***	2	4	4	16	8		36	121	66
3	***6***	-1	-1	1	1	1		9	36	18
5	***8***	1	1	1	1	1		25	64	40
4	***6***	0	-1	0	1	0		16	36	24
36	**63**			16	44	24		**160**	**485**	**276**

$$r_{xy} = \frac{24}{\sqrt{16}\sqrt{44}} = \frac{24}{26.533} = 0.905$$

باستخدام الطريقة الثالثة

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

$$r_{xy}=\frac{9(276)-(36)(63)}{\sqrt{9(160)-(36)^2}\sqrt{9(485)-(63)^2}}$$

$$\frac{2484-2268}{\sqrt{1440-1296}\sqrt{4365-3963}}=\frac{216}{240}=0.90$$

معامل الارتباط يساوي 0.9 وهذا يشير الى وجود علاقة ايجابية قوية وهذا يعني ان زيادة الكمية المعروضة يرافقها ارتفاع في سعر هذه الكمية.

مثال : الجدول التالي يمثل بيانات تم الحصول عليها من دراسة حول عدد حالات الغياب والدرجات النهائية لسبعة طلاب تم اختيارهم عشوائيا من صف احصائي .

الطلاب	عدد الغيابات x	الدرجة النهائية y
1	6	82
2	2	86
3	15	43
4	9	74
5	12	58
6	5	90
7	8	78

الحل

$$r_{xy}=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^{2}-(\sum x_i)^2}\sqrt{n\sum y_i^{2}-(\sum y_i)^2}}$$

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
6	82	492	36	6724
2	86	172	4	7396
15	43	645	225	1849
9	74	666	81	5476
12	58	696	144	3364
5	90	450	25	8100
8	78	624	64	6084
57	**511**	**3745**	**579**	**38993**

$$r_{xy} = \frac{7(3745)-(57)(511)}{\sqrt{7(579)-(57)^2}\,\sqrt{7(38993)-(511)^2}} = -0.94$$

تشير قيمة (r= -0.94) الى وجود علاقة سلبية قوية جدا بين الدرجه النهائية للطالب وعدد حالات الغياب للطالب.

2- ارتباط الرتب **Rank correlation**:

يعرف معامل ارتباط الرتب بأنه المعامل الذي يقيس درجة الارتباط ما بين صفتين، وقد اقترح هذا المعامل من قبل Sperman.

ليكن لدينا المتغيرين X_i و Y_i من النوع الوصفي (أي لا يمكن قياسهما بوحدات كمية) ولكن يمكن ترتيبهما تنازلياً او تصاعدياً وفق معيار معين. ان صيغة معامل ارتباط الرتب لـ سبيرمان (في حال عدم وجود تكرار للقيم) هي:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

بحيث ان $d_i = x_i - y_i$

<u>مثال:</u>

كانت تقديرات ستة طلاب في مادتي الإحصاء والرياضيات كما يلي:

<u>الإحصاء:</u>	متوسط	جيد	ضعيف	مقبول	جيد جداً	ممتاز
<u>الرياضيات:</u>	مقبول	متوسط	جيد	ضعيف	ممتاز	جيد جداً

جد معامل الارتباط البسيط بين تقدير الطالب في امتحان الإحصاء وتقديره في امتحان الرياضيات.

<u>الحل:</u>

نرتب التقديرات ترتيب تصاعدي ومن ثم نخصص رتباً تمثل قيم سلسلة اعداد طبيعية:

التقديرات: ضعيف، مقبول، متوسط، جيد، جيد جداً، ممتاز

الرتب: 1 2 3 4 5 6

ثم نعود لتخصيص هذه الرتب للتقديرات الاصلية كل حسب موقعها وكما موضح في الجدول التالي:

تقديرات X_i	تقديرات Y_i	رتب X_i	رتب Y_i	$d_i = X_i - Y_i$	d_i^2
متوسط	مقبول	3	2	1	1
جيد	متوسط	4	3	1	1
ضعيف	جيد	1	4	3-	9
مقبول	ضعيف	2	1	1	1
جيد جداً	ممتاز	5	6	1-	1
ممتاز	جيد جداً	6	5	1	1
المجموع					**14**

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6*14}{6(6^2-1)} = 1 - \frac{14}{35} = 1 - 0.4 = 0.6$$

أما في حالة تكرار قيم أحد المتغيرين او كلاهما فان معامل ارتباط الرتب لسبيرمان سوف يحسب بالشكل التالي:

مثال: الاتي تقديرات لكفاءة عشرة من العاملين في أحد المصانع من حيث آدائهم في تشغيل نوعين من المكائن الحديثة، يطلب حساب معامل الارتباط لسبيرمان.

النوع الاول	جيد	متوسط	جيد جداً	متوسط	ممتاز	ضعيف	مقبول	متوسط	جيد	جيد
النوع الثاني	متوسط	جيد	جيد	مقبول	جيد جداً	مقبول	ضعيف	متوسط	متوسط	ممتاز

الحل: نبدأ بترتيب التقديرات تصاعدياً ثم نخصص للترتيب رتب تمثل قيم سلسلة اعداد طبيعية ولكل متغير.

التقديرات X	ضعيف	مقبول	متوسط	متوسط	متوسط	جيد	جيد	جيد	جيد جداً	ممتاز
الرتب X	1	2	3	4	5	6	7	8	9	10

التقديرات Y	ضعيف	مقبول	مقبول	متوسط	متوسط	متوسط	جيد	جيد	جيد جداً	ممتاز
الرتب Y	1	2	3	4	5	6	7	8	9	10

بعد ذلك سيتم حساب معدل الرتب المقابلة للتقديرات المكررة وإعادة تخصيص هذا المعدل لهذه التقديرات ولكل متغير وكما يلي:

تقدير المتوسط لـ $\bar{X} = \frac{3+4+5}{3} = 4$

تقدير الجيد لـ $\bar{X} = \frac{6+7+8}{3} = 7$

تقدير المقبول لـ $\bar{Y} = \frac{3+2}{2} = 2.5$

تقدير المتوسط لـ $\bar{Y} = \frac{4+5+6}{3} = 5$

تقدير الجيد لـ $\bar{Y} = \frac{7+8}{2} = 7.5$

X_i	Y_i	رتب X_i	رتب Y_i	$d_i = X_i - Y_i$	d_i^2
جيد	متوسط	7	5	2	4
متوسط	جيد	4	7.5	3.5-	12.25
جيد جداً	جيد	9	7.5	1.5	2.25
متوسط	مقبول	4	2.5	1.5	2.25
ممتاز	جيد جداً	10	9	1	1
ضعيف	مقبول	1	2.5	1.5-	2.25
مقبول	ضعيف	2	1	1	1
متوسط	متوسط	4	5	1-	1
جيد	متوسط	7	5	2	4
جيد	ممتاز	7	10	3-	9
المجموع					**39**

بعد ذلك سيتم التعديل على قانون سبيرمان بأضافة الكمية $\frac{m(m^2-1)}{12}$ الى المقدار $\sum d_i^2$ حيث ان m تمثل عدد مرات تكرار الصفة.

ان عدد مرات تكرار التقدير متوسط بالنسبة الى X هو 3=m، فان:

$m(m^2-1)/12=3(9-1)/12=2$

عدد مرات تكرار التقدير جيد بالنسبة الى X هو 3=m، فان:

$m(m^2-1)/12=3(9-1)/12=2$

عدد مرات تكرار التقدير مقبول بالنسبة الى Y هو 2=m، فان:

$m(m^2-1)/12=2(4-1)/12=0.5$

عدد مرات تكرار التقدير متوسط بالنسبة الى Y هو m=3، فان:

$m(m^2-1)/12=3(9-1)/12=2$

عدد مرات تكرار التقدير جيد بالنسبة الى Y هو m=2، فان:

$m(m^2-1)/12=2(4-1)/12=0.5$

وعليه فان التعديل الكلي هو:

$2+2+0.5+2+0.5=7=k$

وبذلك فان:

$$r_s = 1 - \frac{6\left(\sum_{i=1}^{n} d_i^2 + k\right)}{n(n^2-1)} = 1 - \frac{6(39+7)}{10(100-1)}$$
$$= 1 - \frac{276}{990} = 0.721$$

ملاحظة :

ان قيمة معامل الارتباط للرتب تتراوح قيمته ما بين (+1,-1)، أي ان:

$$-1 \le r_s \le +1$$

المحاضرة (5)

معامل الارتباط الجزئي Partial correlation ~~cof~~ coefficient

في بعض الدراسات يوجد هناك عدد من المتغيرات (ثلاثة او اكثر) مرتبطة فيما بينها بعلاقة رياضية مثل انفاق الاسرة يكون مرتبط بدخلها الشهري وعدد افرادها.

ففي هذه الحالة ولغرض حساب معامل الارتباط بين المتغيرين الاثنين باستبعاد أثر المتغير الثالث نلجأ الى حساب معامل الارتباط الجزئي ويمكن حساب معامل الارتباط الجزئي باستخدام الصيغة التالية:

الصيغة الاولى:

$$\cancel{r_{12\cdot3}}\ r_{12\cdot3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

متغير اول

متغير ثاني

متغير ثالث

حيث ان:

$r_{12\cdot3}$ يمثل معامل الارتباط الجزئي للمتغيرين الاول والثاني باستبعاد اثر المتغير الثالث

r_{12}: يمثل معامل الارتباط البسيط بين المتغيرين الاول والثاني

r_{23}: يمثل معامل الارتباط البسيط بين المتغيرين الثاني والثالث

هذه الصيغة تستخدم في حالة وجود ثلاثة متغيرات ونرغب في حساب معامل الارتباط الجزئي بين المتغيرين الاول والثاني باستبعاد اثر المتغير الثالث.

الصيغة الثانية :

هذه الصيغة تستخدم في حالة وجود اربعة متغيرات ونرغب في حساب معامل الارتباط الجزئي بين المتغيرين الاول والثاني باستبعاد اثر المتغيرين الثالث والرابع .

$$r_{12.34} = \frac{r_{12.3} - (r_{14.3})(r_{24.3})}{\sqrt{(1-r^2_{14.3})(1-r^2_{24.3})}}$$

حيث ان

$r_{12.34}$: معامل الارتباط الجزئي بين المتغيرين الاول والثاني باستبعاد اثر المتغيرين الثالث والرابع .

$r_{12.3}$ معامل الارتباط الجزئي بين المتغيرين الاول والثاني باستبعاد المتغير الثالث

$r_{14.3}$ معامل الارتباط الجزئي بين المتغيرين الاول والرابع باستبعاد اثر المتغير الثالث

$r_{24.3}$ معامل الارتباط الجزئي بين المتغيرين الثاني والرابع باستبعاد اثر المتغير الثالث

مثال :

اذا كان لدينا ثلاث متغيرات X_1 تمثل الدخل الشهري و X_2 تمثل انفاق الاسرة الشهري و X_3 تمثل عدد افراد الاسرة وحصلنا على البيانات التالية

$r_{12} = 0.91$, $r_{13} = 0.39$, $r_{23} = 0.62$

جد معامل الارتباط الجزئي للمتغيرين الاول والثاني باستبعاد اثر المتغير الثالث [جد $r_{12.3}$]

الحل /

$$r_{12\cdot 3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

$$= \frac{0.91 - (0.39)(0.62)}{\sqrt{(1-(0.39)^2)(1-(0.62)^2)}} = 0.925$$

مثال ٤

اذا كان لدينا البيانات التالية

$r_{23} = 0.2$, $r_{14} = 0.5$, $r_{13} = 0.6$, $r_{12} = 0.7$

$r_{34} = 0.4$, $r_{24} = 0.3$

جد معامل الارتباط الجزئي للمتغيرين الاول والثاني باستبعاد اثر المتغيرين الثالث والرابع. [$r_{12\cdot 34}$]

الحل /

$$r_{12\cdot 34} = \frac{r_{12\cdot 3} - (r_{14\cdot 3})(r_{24\cdot 3})}{\sqrt{(1-r_{14\cdot 3}^2)(1-r_{24\cdot 3}^2)}}$$

يجب ان نجد اولاً $r_{12\cdot 3}$ ثم نجد $r_{14\cdot 3}$, $r_{24\cdot 3}$

$$r_{12\cdot 3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.7 - (0.6)(0.2)}{\sqrt{(1-(0.6)^2)(1-(0.2)^2)}}$$

$$= 0.74$$

$$r_{14\cdot3} = \frac{r_{14} - (r_{13})(r_{43})}{\sqrt{(1-(r_{13})^2)(1-r_{43}^2)}}$$

$$= \frac{0.5 - (0.6)(0.4)}{\sqrt{(1-(0.6)^2)(1-(0.4)^2)}} = 0.35$$

$$r_{24\cdot3} = \frac{r_{24} - (r_{23})(r_{43})}{\sqrt{(1-r_{23}^2)(1-r_{43}^2)}}$$

$$= \frac{0.3 - (0.2)(0.4)}{\sqrt{(1-(0.2)^2)(1-(0.4)^2)}} = 0.27$$

الان نجد $r_{12\cdot34}$

$$r_{12\cdot34} = \frac{0.74 - (0.35)(0.27)}{\sqrt{(1-(0.35)^2)(1-(0.27)^2)}}$$

$$= 0.72$$

ارتباط الصفات **Correlation between attributes**

ان موضوع ارتباط الصفات يقترن بوجود توزيع تكراري مزدوج لمتغيرين من النوع الوصفي ويسمى هذا الجدول بجدول التوافق contingency table حيث ان صفوف التوزيع والبالغ عددها k تمثل مستويات المتغير x واعمدته البالغ عددها m تمثل مستويات المتغير y.

X \ y	y_1	y_2	....	y_m	
x_1	f_{11}	f_{12}	...	f_{1m}	$T_{1.}$
x_2	f_{21}	f_{22}	...	f_{2m}	$T_{2.}$
.	.	.	...	....	....
.	.	.			
x_k	f_{k1}	f_{k2}	...	f_{km}	$T_{k.}$
total	$T_{.1}$	$T_{.2}$	...	T.m	N

اذ يتعذر على الباحث حساب الارتباط بينهما بصيغة معامل ارتباط الرتب لسبيرمان انما يستوجب ذلك استخدام مقياس اخر يسمى معامل التوافق **Coefficient of contingency**

معامل التوافق **Coefficient of contingency**

يعرف معامل التوافق بانه مقياس لقيمة العلاقة بين متغيرين في جدول توافق ذو مرتبة k*m، فاذا فرضنا ان C تمثل معامل التوافق وان r يمثل المجموع الكلي لنواتج قسمة مربع التكرار في كل خلية على حاصل ضرب مجموعي التكرارين العمودي والافقي المقابلين لتلك الخلية التي يقع فيها التكرار، عندئذ:

$$C = \sqrt{\frac{r-1}{r}}$$

وفيما يلي خطوات إيجاد قيمة هذا المعامل، ويمكن تحقيق هذه الخطوات باستخدام الصفوف او الاعمدة فان ذلك لا يؤثر على قيمة C. ولنأخذ الصفوف مثلاً:

بالنسبة الى الصف الأول فان:

$$r_1 = \frac{f_{11}^2}{T_{1.} * T_{.1}} + \frac{f_{12}^2}{T_{1.} * T_{.2}} + \frac{f_{13}^2}{T_{1.} * T_{.3}} + \dots \frac{f_{1m}^2}{T_{1.} * T_{.m}} = \frac{1}{T_{1.}} \sum_{i=1}^{m} \frac{f_{1i}^2}{T_{.i}}$$

$$r_2 = \frac{f_{21}^2}{T_{2.} * T_{.1}} + \frac{f_{22}^2}{T_{2.} * T_{.2}} + \frac{f_{23}^2}{T_{2.} * T_{.3}} + \ldots \frac{f_{2m}^2}{T_{2.} * T_{.m}} = \frac{1}{T_{2.}} \sum_{i=1}^{m} \frac{f_{2i}^2}{T_{.i}}$$

.

.

.

$$r_k = \frac{f_{k1}^2}{T_{k.} * T_{.1}} + \frac{f_{k2}^2}{T_{k.} * T_{.2}} + \frac{f_{k3}^2}{T_{k.} * T_{.3}} + \ldots \frac{f_{km}^2}{T_{k.} * T_{.m}} = \frac{1}{T_{k.}} \sum_{i=1}^{m} \frac{f_{ki}^2}{T_{.i}}$$

وعندئذٍ فان:

$$r = r_1 + r_2 + r_3 + \ldots + r_k = \sum_{j=1}^{k} r_j$$

مثال: الجدول التالي يبين عدد حوادث الطرق التي تعرض لها شاحنات منشأة لنقل البضائع خلال فترة زمنية معينة موزعة حسب نوع الحادث وحالة الجو، يطلب حساب معامل التوافق لهذا التوزيع:

نوع الحادث / حالة الجو	دهس	اصطدام	انقلاب	المجموع
صحو	25	12	9	**46**
ممطر	10	50	35	**95**
ضباب	20	45	40	**105**
المجموع	**55**	**107**	**84**	**246**

الحل/

$$r_1 = \frac{1}{46}\left(\frac{(25)^2}{55} + \frac{(12)^2}{107} + \frac{(9)^2}{84}\right) = 0.297$$

$$r_2 = \frac{1}{95}\left(\frac{(10)^2}{55} + \frac{(50)^2}{107} + \frac{(35)^2}{84}\right) = 0.419$$

$$r_3 = \frac{1}{105}\left(\frac{(20)^2}{55} + \frac{(45)^2}{107} + \frac{(40)^2}{84}\right) = 0.29$$

$$\therefore r = r_1 + r_2 + r_3 = 297 + 0.419 + 0..431 = 1.147$$

$$\therefore C = \sqrt{\frac{r-1}{r}} = \sqrt{\frac{1.147-1}{1.147}} = 0.358$$

<u>ملاحظة:</u> ان قيمة معامل التوافق تتراوح بين الصفر والواحد الصحيح $(0 \le C \le 1)$ وتقترب قيمة C من الواحد عندما تكون قيمة r كبيرة جداً بحيث ان $r-1 \approx r$، وتكون قيمة C مساوية للصفر عندما $r = 1$، وإذا كانت $0 < r < 1$ عندئذٍ لا يمكن إيجاد قيمة C.

الفصل الثاني

مفهوم الانحدار

THE CONCEPT OF REGRESSION

(1:2) تعريف تحليل الانحدار : The Concept of Regression

ان تحليل الانحدار عبارة عن وسيلة احصائية (Statisticl tool) يستخدم لتحليل العلاقة (relationship) بين متغير مستقل واحد اواكثر (Independent variables) ومتغير تابع (Dependent) .

ويعتبر تحليل الانحدار من اكثر الطرق الاحصائية استعمالا في مختلف العلوم لانه يصف العلاقة بين المتغيرات على هيئة معادلة .

فالمعادلة التي تضم متغيرا مستقلا واحداً تسمى معادلة الانحدار الخطي البسيط Linear regression equation.

وتكتب كالاتي :

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_1$$

بينما المعادلة التي تضم عدة متغيرات مستقلة فتسمي معادلة الانحدار الخطـي المتعدد Multiple linear regression equation

وتكتب كالاتي

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_1 + \hat{B}_2 X_2 + \ldots + \hat{B}_m X_m$$

حيث ان

Y_i = المتغير التابع او المعتمد

X_i = المتغير المستقل

B_0 = نقطة تقاطع خط او مستوى الانحدار بالمحور Y

B_i = معامل الانحدار الجزئي

e_i = قيمة الخطأ العشوائي

(2:2) استعمالات تحليل الانحدار : Uses of regression

يستعمل تحليل الانحدار لعدة اغراض منها : —

(1) وصف البيـانـات Data description

يمكن تلخيص ووصف مجموعة البيانات لدى الباحث بايجاد معادلة الانحـدار التي تصف تلك البيانات

(٢) تقـديـر المعلمات :(Parameter estimation)

يمكن ايجاد تقدير للمعلمة او المعلمات المجهولة التي تصف البيانات ومنها يمكـن الاستدلال على اهمية وقوة واتجاه العلاقة بين المتغيرات

(٣) التنبـــؤ (Prediction)

كما يمكن تقدير الاستجابة والتنبؤ بها بما يفيد كثيرا في التخطيط واتخاذ القرارات

(٤) السيطرة (Control)

ان النماذج الانحدارية قد تستخدم لغرض السيطرة . فعند ايجاد المعادلة التـي تصف البيانات فانه يمكن السيطرة على قيم المتغير التابع بتغيير قيم المتغيرات المستقلة

(4:2) انواع الانحدار Types of Regression

يمكن تصنيف الانحدار لغرض تبسيط دراسته في هذا الكتاب الى الاتي :-

(١) انحـــدار خطـــي Linear regression

ويقسم الى : -

(أ) انحدار خطي بسيط Simple linear regression ويضم متغيـــرا مستقلا واحداً . ومن امثلة ذلك : -

1. دراسة العلاقة بين دخل العائلة (X) والانفاق على الطعام (Y) لثلاثين عائلة من مدينة بغداد
2. دراسة العلاقة بين العمر (X) وضغط الدم (Y) لخمسين شخصاً في مدينـــة القاهرة.

معادلة الانحدار البسيط هي :-

$$Y = B_0 + B_1 X_{i1} + e_i$$

(ب) انحـــدار خطي متعـــدد Multiple linear regression

ويضم عدة متغيرات مستقلة . ومن امثلة ذلك : -

1. دراسة تأثيرات كل من وزن الجسم (X_1) والعمر (X_2) والطول (X_3) والموقف من التدخين (X_4) على المجموع الكلي للكوليسترول (Y) .
2. دراسة العلاقة بين : ثمن الكيلو الواحد من اللحم البقري (X_1) والدخل الممكن التصرف به (X_2) وعدد السكان (X_3) والكميات المستهلكة من اللحـــم البقري (بالكيلو) (Y_i)

ومعادلة الانحدار الخطي المتعدد بصورة عامة هي :

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_m X_m + e_i$$

(2) انحدار غيـــر خطـــي Non linear regression or Curvilinear regression.

ويقسم ايضا الى قسمين

(أ) انحـــدار غيـــر خطي بسيط Simple Curvilinear

ويضم متغيرا مستقلا واحدا ايضا ومن امثلة ذلك : -

دراسة العلاقة بين المسافة اللازمة لايقاف السيارة (Y) ايقافاً تاماً وسرعة السيارة اثناء سيرهـا (X_i)

ومعادلته بصورة عامة هي :-

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_i^2 + \ldots + e_i$$

ولكن هناك معادلات اخرى غير خطية قد تأخذ الشكل التالي : -

$$Y_i = B_0 \, e^{B_1 X} \mu$$

وهذه يمكن تحويلها الى معادلة خطية باستخدام التحويل اللوغاريتمي مثلا كما سيأتي ذكر ذلك في الفصول القادمة .

(ب) انحـدار غيـر خطـي متعـدد Multiple Curvilinear

ويضم عدة متغيرات مستقلة ومن أمثلة ذلك :

1 - دراسة العلاقة بين مقدار التأمين على الحياة Y وكل من درجة الخطورة (X_1) والعمر X_2 لعشرين طبيباً مثلا .

2 - دراسة العلاقة بين درجة تصلب الشرايين Y ونسبة الكولسترول X_1 والوزن (X_2) في الارانب ومعادلته لمتغيرين مستقلين مثلا) هي

$$Y = B_{00} + B_{10}X_1 + B_{01}X_2 + B_{11}X_1X_2 + B_{20}X_1^2 + B_{02}X_2^2 + \ldots$$

هذا وهناك معادلات اخرى غير خطية مثل

$$Y = B_0 X_1^{B_1} X_2^{B_2} X_3^{B_3} e$$

ولكن مثل هذه المعادلات يمكن تحويلها الى معادلات خطية باستخدام التحويل اللوغاريتمي مثلا .

أنواع الانحدار

- انحدار خطي
  - انحدار خطي بسيط
    ويضم متغير مستقل واحد
    ومعادلته هي
    $Y_i = B_0 + B_1 X_{i1} + e_i$
  - انحدار خطي متعدد
    ويضم عدة متغيرات مستقلة
    ومعادلته هي
    $Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} -- + e_i$
- انحدار غير خطي
  - غير خطي بسيط
    متغير مستقل واحد
  - متعدد
    عدة متغيرات مستقلة

بشكل عام للمعادلة

$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2}^2 --- + e_i$

فيمكن ان تكون المعادلة بشكل اخر

$Y_i = B_0 e^{B_1 X_i}$ مثلا

وغيرها من الاشكال غير الخطية

ملاحظة: الشكل الخطي يعني ان الاس للمعادلة يكون من الدرجة الاولى (اي المعادلة لا تحتوي على اس اكثر من واحد)

الانحدار الخطي البسيط

SIMPLE LINEAR REGRESSION

(1:3) في حالة عدم وجود تكرار لقيم المتغير المستقل (X_i)

(No repeated measurement)

(1) وصـف البيانـات

تتكون البيانات عادة منBمن ازواج المشاهدات(X_{11}، Y_1)(X_{21}، Y_2) ... (X_{n1}، Y_n)
و ترتب كما في الجدول التالي :

المشاهدات	X_{i1}	Y_i
1	X_{11}	Y_1
2	X_{21}	Y_2
3	X_{31}	Y_3
⋮	⋮	⋮
n	X_{n1}	Y_n

حيث ان

X_{i1} = القيمة i للمتغير المستقل X_1

Y_i = القيمة i للمتغير المعتمد او التابع Y

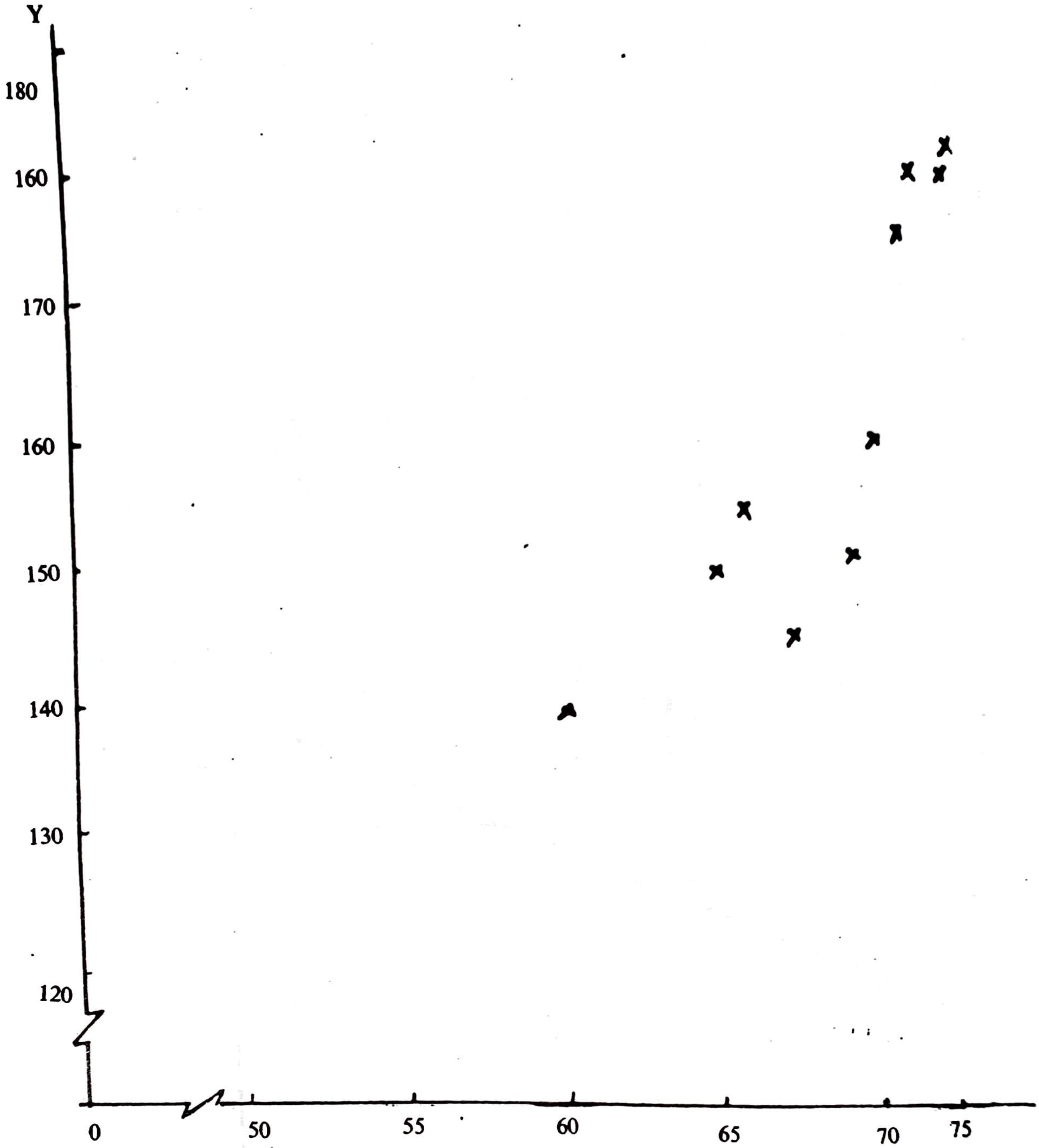
مثال (1:3) : البيانات التالية تمثل القوة الحصانية (X_{i1}) الى اقرب كيلوواط والسرعة القصوى (Y_i) الى اقرب كيلومتر / ساعة لعينة مؤلفة من عشرة سيارات سباق

جدول (1:3) بينات المثال (1:3)

السرعة Y_i	القـوة X_{i1}	ازواج المشاهدات
160	70	1
150	65	2
180	72	3
140	60	4
182	74	5
155	66	6
152	68	7
180	73	8
145	67	9
175	71	10

(2) التمثيـل البيانـي :

ان ازواج المشاهدات في المثال (1:3) يمكن تمثيلها نقاطاً على ورقة بيانية ذات محورين X و Y كما في شكل (1:3)



شكل (1:3) التمثيل البياني للمثال (1:3)

ان الرسم اعلاه يدعى : رسم الأنتشار Scatter diagram

(3) النموذج الخطي : The linear model

ان العلاقة بين X و Y يمكن التعبير عنها كدالة خطية تدعى معادلة خط الانحدار البسيط Simple linear regression equation كالاتي :-

$$Y_i = B_0 + B_1X_{i1} + e_i \qquad i = 1, 2, \dots n$$

حيث ان :

Y_i = المتغير المعتمد او التابع او مقدار الاستجابة

X_{i1} = المتغير المستقل

B_0, B_1 = ثوابت تدعى معلمات الانحدار . Regression Parameters

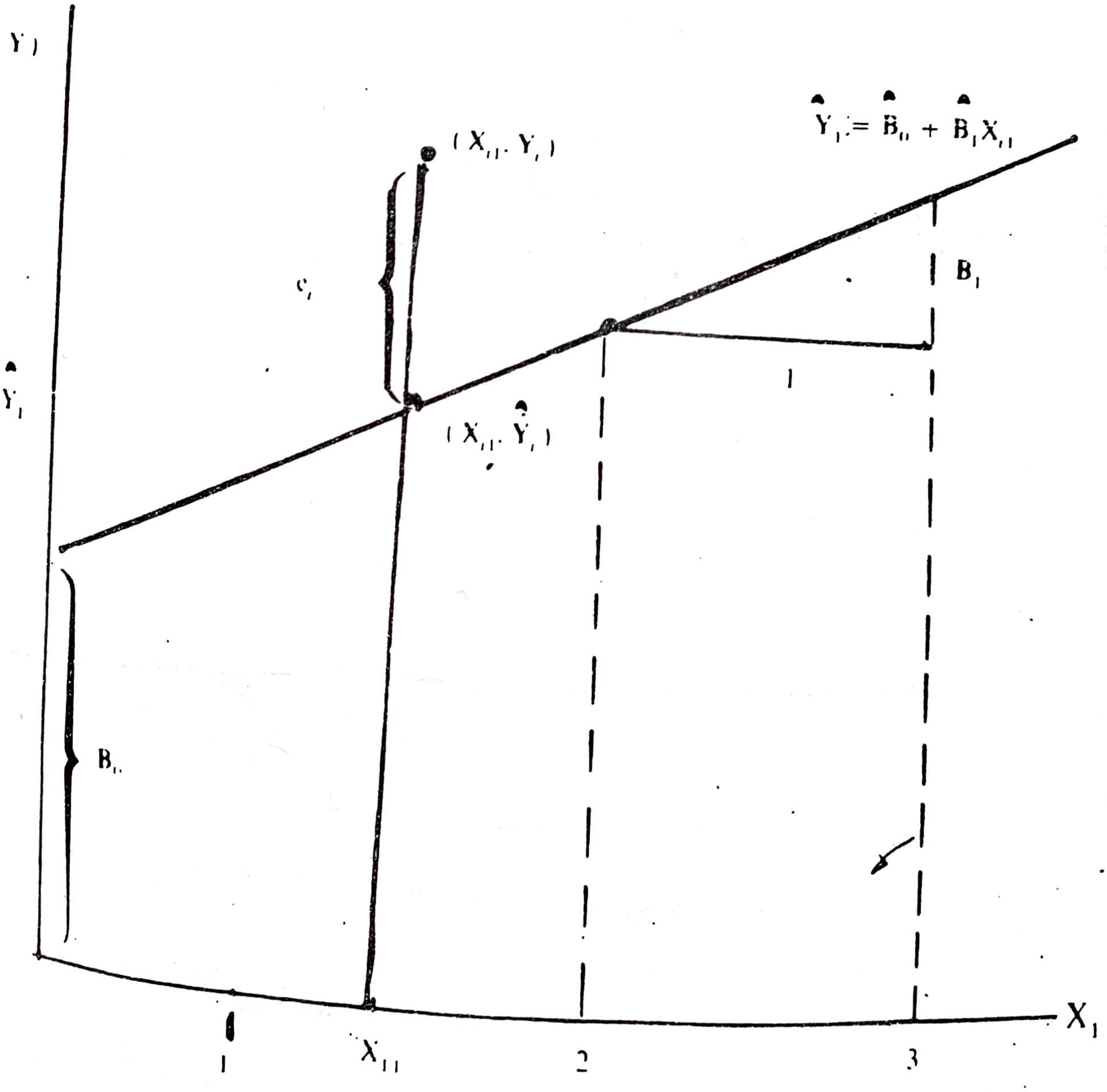
B_0 = نقطة تقاطع خط الانحدار بالمحور Y

ان B_0 تعطي متوسط الاستجابة عندما تكون قيمة X_{i1} تساوي صفراً

B_1 = معامل الانحدار لـ Y على X_{i1} وهي تمثل مقدار التغير في Y عند زيادة وحدة واحدة من X_{i1}

e_i = قيمة الخطأ العشوائي

كما موضح ذلك بالرسم التالي : -



لاحظ بان اشارة معامل الانحدار B_1 تدل على اتجاه العلاقة بين المتغيريـــن X وY فلوكانت اشارتها موجبة دل ذلك على ان العلاقة بين X و Y موجبة اي عند زيادة وحدة واحدة من X فان Y تزداد بمقدار B_1 . اما اذا كانت اشارة B_1 سالبة بمعنى ان العلاقة بين X و Y سالبة اي ان عند زيادة وحدة واحدة من فان Y تنقص بمقدار B_1 .

ان النموذج الخطي او الرياضي لازواج المشاهدات يكون كالاتي :

$$Y_1 = B_0 + B_1 X_{11} + e_1$$
$$Y_2 = B_0 + B_1 X_{21} + e_2$$
$$\vdots$$
$$Y_n = B_0 + B_1 X_{n1} + e_n$$