الفصـــــــــــل السادس

الارتباط والانحدار الخطي البسيط

1/6 مقـــــــــــدمة

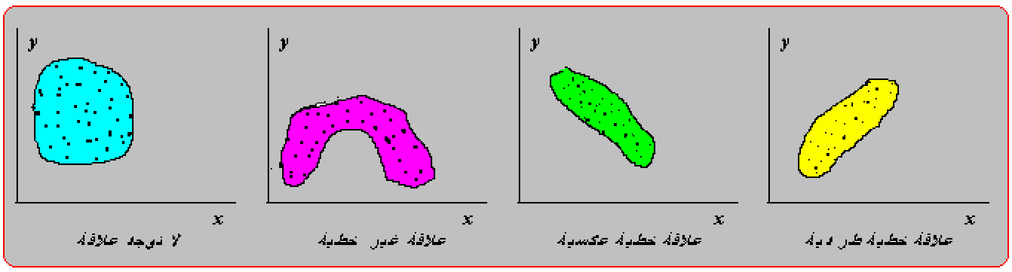
في الفصول الثلاث السابقة تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار، ومن الأمثلة على ذلك:

1- الإنفاق، والدخل العائلي.
2- سعر السلعة، والكمية المطلوبة منها.
3- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
4- تقديرات الطلاب في مقرر الإحصاء، وتقديراهم في مقرر الرياضيات.
5- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
6- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
7- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(x\ ,\ y)$ ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل(6-1)

شكل الانتشار لبيان نوع العلاقة بين x , y



2/6 الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي

هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

1/2/6 الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز r (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

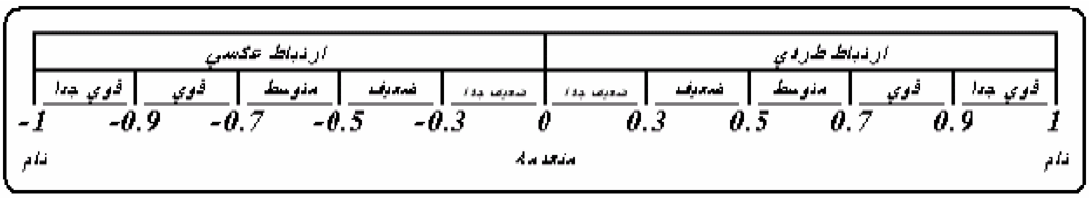
- نوع **العلاقة:ــ** وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.
2- إذا كانت إشارة معامل الارتباط موجبة $(r > 0)$توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .
3- إذا كان معامل الارتباط قيمته صفرا $(r = 0)$ دل ذلك على انعدام العلاقة بين المتغيرين.

- **قوة العلاقة:ــ** ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)، حيث أن قيمة معامل الارتباط تقع في المدى$(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

شكل (6-2)

درجات قوة معامل الارتباط



2/2/6 معامل الارتباط الخطى البسيط " لبيرسون" Pearson

في حالة جمع بيانات عن متغيرين كميين $(x \ , \ y)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" Pearson، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بن الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة ، تستخدم صيغة " بيرسون" التالية :

$$r = \frac{S_{xy}}{S_x\, S_y} = \frac{\dfrac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y-\bar{y})^2}{(n-1)}}} \qquad (٦-١)$$

حيث أن :

$S_{xy} = \sum (x-\bar{x})(y-\bar{y}) / (n-1)$: هو التغاير بين $(x \;,\; y)$،

$S_x = \sqrt{\sum (x-\bar{x})^2 / (n-1)}$: هو الانحراف المعياري لقيم (x)،

$S_y = \sqrt{\sum (y-\bar{y})^2 / (n-1)}$: هو الانحراف المعياري لقيم (y).

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} \qquad (٦-٢)$$

مثـــــال (6-1)

فيما يلي المساحة المتزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995حتى عام 2002.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

الحـــــل

بفرض أن (x) هي المساحة المتزرعة، (y) هي الكمية، ولحساب معامل الارتباط بين $(x \;,\; y)$ يتم تطبيق المعادلة **(6-2)**، وذلك على النحو التالي:

- حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{x} \;,\; \bar{y})$.

$$\bar{x} = \frac{\sum x}{n} = \frac{2108}{8} = 263.5 \;,\; \bar{y} = \frac{\sum y}{n} = \frac{5264}{8} = 658$$

- حساب المجاميع

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

$$\sum(x-\bar{x})^2 = 12040\ ,\ \sum(y-\bar{y})^2 = 23850\ ,$$
$$\sum(x-\bar{x})(y-\bar{y}) = -13528$$

إذا معامل الارتباط قيمته هي:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$
$$= \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

- يوجد ارتباط عكسي قوي بين المساحة المتزرعة، وكمية إنتاج اللحوم.

تبسيط العمليات الحسابية:

في بعض الأحيان، يكون استخدام صيغة المعادلة (**6-2**) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيما كسرية، من أجل ذلك يمكن تبسيط الصيغة (**6-2**) إلى صيغة أسهل تعتمد على مجموع القيم وليس على انحرافات القيم عن وسطها الحسابي، وهذه الصيغة هي:

$$\boxed{r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}} \qquad (٦-٣)$$

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

- حساب المجاميع:

x	y	xy	x^2	y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244

المجاميع المطلوبة
$\sum x = 2108\ ,\ \sum y = 5264$ $\sum xy = 1373536$

289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

$$\sum x^2 = 567498$$
$$\sum y^2 = 3487562$$

- حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة (**3-6**) أعلاه، نجد أن معامل الارتباط قيمته هي:

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

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

وهي نفس النتيجة السابقة:

3/2/6 معامل ارتباط الرتب (اسبيرمان) Spearman

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة "بيرسون" السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط اسبيرمان " Spearman ، ويعبر عنه بالمعادلة التالية :

$$\boxed{r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}} \qquad (٤-٦)$$

حيث أن d هي الفرق بين رتب مستويات المتغير الأول x ، ورتب مستويات المتغير الثاني y ، أي أن : $d = R_x - R_y$.

مثـــــــال (2-6)

فيما يلي تقديرات 10 طلاب في مادتي الإحصاء، والاقتصاد:

تقديرات إحصاء	أ	جـ+	د	د+	ب+	جـ+	أ+	ب	ب+	ب+
تقديرات اقتصاد	أ+	د	جـ	جـ	أ	ب	ب+	ب	جـ	ب

والمطلوب:

1- احسب معامل الارتباط بين تقديرات الطلبة في المقررين.

2- وما هو مدلوله ؟

الحـــــــــل

1- بفرض أن x هي تقديرات الإحصاء، y هي تقديرات الاقتصاد، يمكن حساب معامل الارتباط بينهما باستخدام المعادلة (**6-4**)، وذلك بإتباع الآتي:

الرتب	1	2	3	4	5	6	7	8	9	10
تقديرات إحصاء	أ+	أ	ب+	ب+	ب+	ب	جـ+	جـ+	د+	د
رتب x	1	2	(3+4+5)/3=4			6	(7+8)/2=7.5		9	10
تقديرات اقتصاد	أ+	أ	ب+	ب	ب	ب	جـ	جـ	جـ	د
رتب y	1	2	3	(4+5+6)/3=5			(7+8+9)/3=8			10

- إذا يمكن حساب المجموع: $\sum d^2$ كما يلي:

x	y	رتب x	رتب y	d	d^2
أ	أ+	2	1	1	1
جـ+	د	7.5	10	-2.5	6.25
د	جـ	10	8	2	4
د+	جـ	9	8	1	1
ب+	أ	4	2	2	1
جـ+	ب	7.5	5	2.5	6.25
أ+	ب+	1	3	-2	4
ب	ب	6	5	1	1
ب+	جـ	4	8	-4	16
ب+	ب	4	5	-1	1
					44.5

$$\sum d^2 = 44.5$$

- معامل الارتباط هو:

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

$$= 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$

$$= 1 - 0.2697 = 0.7303$$

2- مدلول معامل الارتباط :

بما أن $r = 0.703$ ، ويدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالب في مادة الإحصاء ، ومادة الاقتصاد .

ملحوظة:- يمكن استخدام صيغة معامل ارتباط "اسبيرمان" في حساب الارتباط بين متغيرين كميين، حيث يتم استخدام رتب القيم التي يأخذها المتغير، ونترك للطالب القيام بحساب معامل ارتباط الرتب بين المساحة والكمية في مثال (**1-5**) السابق، وعليه أن يقوم بتفسير النتيجة: (معاونة : $\sum d^2 = 148$)

)

3/6 الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على إنتاجية الدونم.
- دراسة أثر الإنتاج على التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- أثر الدخل على الإنفاق الاستهلاكي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

1/3/6 نموذج الانحدار الخطي

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$\boxed{y = \beta_0 + \beta_1 x + e} \qquad (٦-٥)$$

حيث أن:

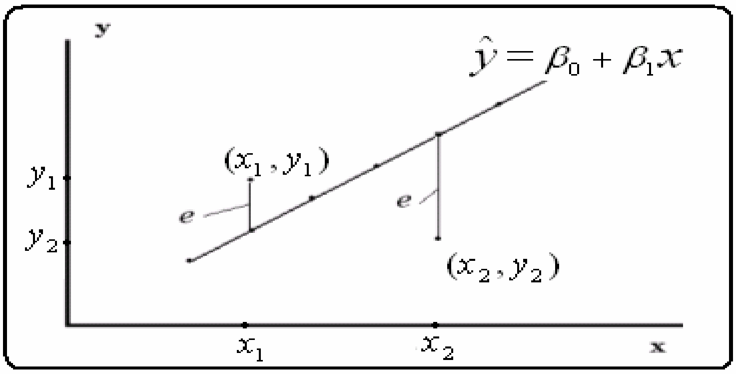
y : هو المتغير التابع (الذي يتأثر)

x : هو المتغير المستقل (الذي يؤثر)

b_0 : هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x = 0$

b_1 : ميل الخط المستقيم $(b_0 + b_1 x)$ ، ويعكس مقدار التغير في y إذا تغيرت x بوحدة واحدة.

e : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y} = b_0 + b_1 x$ ، أي أن : $e = y - (b_0 + b_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.



2/3/6 تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(b_1 \ , \ b_0)$ في النموذج (6-5) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum (y - (b_0 + b_1 x))^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\boxed{\begin{array}{l} \hat{\beta}_1 = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \ , \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{array}} \qquad (٦-٦)$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x، $\bar{y}$ هو الوسط الحسابي لقيم y، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{b}_0 + \hat{b}_1 x$، ويطلق على هذا التقدير " تقدير معادلة انحدار y على x.

مثال (6-3)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكجم، وذلك لعينة من العجول الرضيعة حجمها 10.

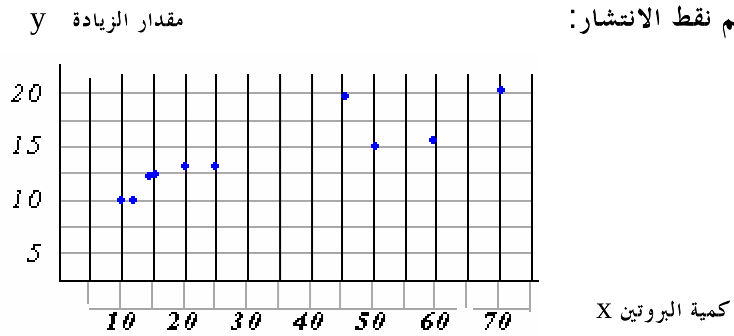
كمية البروتين	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن	10	10	12	12	13	13	19	15	16	20

والمطلوب :

1- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
2- قدر معادلة انحدار الوزن على كمية البروتين.
3- فسر معادلة الانحدار.
4- ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟
5- ارسم معادلة الانحدار على نقط الانتشار في المطلوب (1) .

الحـــل

1- رسم نقط الانتشار:



من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

2- تقدير معادلة الانحدار.

بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلتين في (**6**-**6**)، ومن ثم يتم حساب المجاميع التالية:

كمية البروتين x	الزيادة في الوزن y	$x\ y$	x^2
10	10	100	100
11	10	110	121
14	12	168	196
15	12	180	225
20	13	260	400
25	13	325	625
46	19	874	2116
50	15	750	2500
59	16	944	3481
70	20	1400	4900
320	140	5111	14664

المجاميع المطلوبة
$\sum x = 320$
$\sum y = 140$
$\sum xy = 5111$
$\sum x^2 = 14664$
إذا الوسط الحسابي:
$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$
$\bar{y} = \frac{\sum x}{n} = \frac{140}{10} = 14$

- بتطبيق المعادلة الأولى في (**6**-**6**) يمكن حساب $\hat{b}_1$ كما يلي:

$$\hat{b}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2}$$

$$= \frac{6310}{44240} = 0.1426$$

- بتطبيق المعادلة الثانية في (**6**-**6**) يمكن حساب $\hat{b}_0$ كما يلي:

$$\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} = 14 - (0.1426)(32) = 9.4368$$

- إذا معادلة الانحدار المقدرة، هي:

$$\hat{y} = 9.44 + 0.143x$$

3- تفسير المعادلة:

- الثابت $\hat{b_0} = 9.44$: يدل على أنه في حالة عدم استخدام البروتين قي التغذيـــة، فـــإن الوزن يزيد 9.44 كجم.
- معامل الانحدار $\hat{b_1} = 0.143$: يدل على أنه كلما زادت كمية البروتين جرام واحـــد، حدث زيادة في وزن العجل بمقدار 0.143 كجم، أى زيادة مقدارها 143 جرام.

4- مقدار الزيادة في الوزن عند $x = 50$ هو:

$$\hat{y} = 9.44 + 0.143(50) = 16.59$$

وأما ومقدار الخطأ العشوائي هو:

$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

5- رسم معادلة الانحدار على نقط الانتشار.

يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

x	50	10
$\hat{y}$	16.59	10.87

إذا معادلة الانحدار هي: