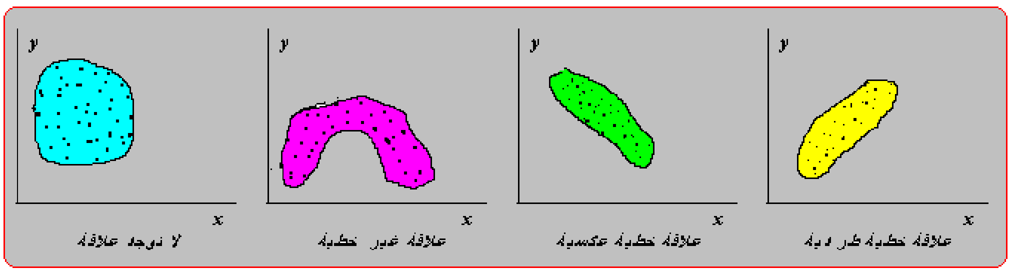
الارتباط والانحدار الخطي البسيط

نعلم أن إذا 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- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.
4- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
5- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(y\ ,\ x)$ ، 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 الانتشار لبيان نوع العلاقة بين **y , x**



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.

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 ρ (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:- 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:- 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:

معامل الارتباط الخطى البسيط " لبيرسون" Pearson

في حالة جمع بيانات عن متغيرين كميين $(y \ , \ x)$، **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 :

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

تبسيط العمليات الحسابية:

في بعض الأحيان، يكون استخدام صيغة المعادلة (5-1) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيما كسرية، من أجل ذلك يمكن تبسيط الصيغة (5-1) إلى صيغة أسهل تعتمد على مجموع القيم وليس على انحرافات القيم عن وسطها الحسابي، وهذه الصيغة **هي:**

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

مثال: رغب احد البنوك معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية لهم ،فقاموا بجمع معلومات عن هذا الموضوع وحصلوا على النتائج التالية.

الموظفين	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
ساعات العمل **X**	6	4	6	13	11	15	5	8	2	8
مستوى الإنتاجية **y**	5	4	4	9	12	14	3	6	1	3

والمطلوب: حساب معامل الارتباط بين ساعات العمل ومستوى الإنتاجية، وما هو مدلوله ؟

الحــل

بفرض أن (x) هي ساعات العمل ، (y) هي مستوى الإنتاجية ، ولحساب معامل الارتباط بين (x , y) يتم تطبيق المعادلة السابقة، وذلك على النحو التالي:

- حساب المجاميع:

x	y	xy	x^2	y^2
8	3	24	64	9
2	1	2	4	1
8	6	48	64	36
5	3	15	25	9
15	14	210	225	196
11	12	132	121	144
13	9	117	169	81
6	4	24	36	16
4	4	16	16	16
6	5	30	36	25
78	61	618	760	533

المجاميع المطلوبة
$\sum x = 78 \quad , \quad \sum y = 61$ $\sum xy = 618$ $\sum x^2 = 760$ $\sum y^2 = 533$

- حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة أعلاه، نجد أن معامل الارتباط قيمته هي:

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

$$= \frac{618 - \frac{(78)(61)}{10}}{\sqrt{\left(760 - \frac{(78)^2}{10}\right)\left(533 - \frac{(61)^2}{10}\right)}}$$

$$= \frac{142.2}{\sqrt{(151.6)(160.9)}} = \frac{142.2}{156.2} = 0.91$$

وهذه النتيجة توضح أن درجة الارتباط=0.91 وهي تعتبر مؤشر على علاقة ايجابية قوية بين ساعات العمل ومستوى الانتاجية.

استخدام برنامج SPSS لحساب معامل الارتباط

لحساب معامل الارتباط باستخدام **SPSS** نتبع الخطوات الآتية:

الانحدار الخطى 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 الإنتاج على التكلفة.

- دراسة أثر كمية البروتين التي يتناولها الإنسان على الزيادة في الوزن.
- دراسة أثر الدخل على الإنفاق الاستهلاكي.
- دراسة أثر حجم القروض على عدد الأنشطة.
- وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

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:

$$\boxed{y = \beta_0 + \beta_1 x + e}$$

حيث أن:

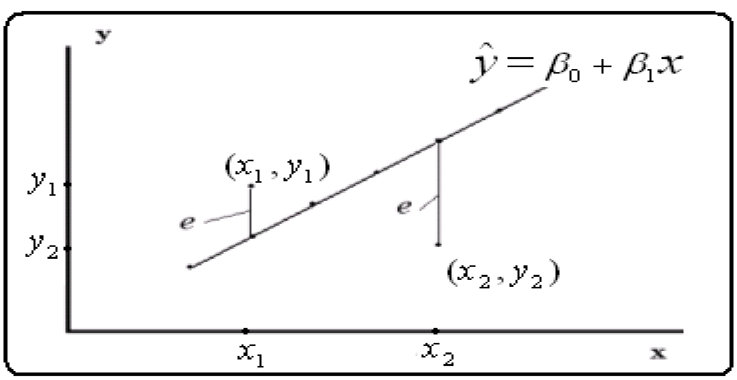
y : هو المتغير التابع (الذي يتأثر)

x : هو المتغير المستقل (الذي يؤثر)

β_0 : هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x = 0$

β_1 : ميل الخط المستقيم $(\beta_0 + \beta_1 x)$ ، ويعكس مقدار التغير في y إذا تغيرت x بوحدة واحدة.

e : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y} = \beta_0 + \beta_1 x$ ، أي أن : $e = y - (\beta_0 + \beta_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.



تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(\beta_1 \ , \ \beta_0)$ في النموذج (4-5) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية

$\sum e^2 = \sum (y - (\beta_0 + \beta_1 x))^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} ,$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

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

يمكن تطبيق طريقة المربعات الصغرى إذا تحققت الشروط التالية:

- القيمة المتوقعة للحد الخطأ يجب أن تكون مساوية للصفر(**E(ei)=0**). حيث أن **E** تعني القيمة المتوقعة.
- يجب أن لاتكون هناك علاقة بين الحد الخطأ والمتغير المستقل (**cov(ei,xi)=0**). حيث أن **cov** تعني الارتباط.
- لايوجد ارتباط ذاتي بين حدود الخطأ **corr (ei,ej)=0** حيث أن **corr** تعني ارتباط.
- أن يكون تبيان حد الخطأ ثابتاً للمشاهدات أي عدم وجود تباين مختلف للمشاهدات (**heteroskedasticity**)، **var(ei) = 0**.

مثـال:

فيما يلي بيانات عن قيم القروض الممنوحة بالمليون ريال، ومقدار الزيادة في عدد الأنشطة، وذلك خلال 10 سنوات.

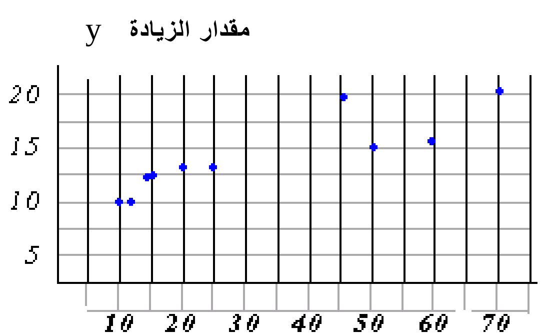
قيمة القروض	10	11	14	15	20	25	46	50	59	70
الزيادة في عدد الأنشطة	10	10	12	12	13	13	19	15	16	20

والمطلوب :

1- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
2- قدر معادلة انحدار عدد الأنشطة على قيمة القروض.
3- فسر معادلة الانحدار.
4- ما هو مقدار الزيادة في عدد الأنشطة اذا كان قيمة القروض 50 مليون؟ وما هو مقدار الخطأ العشوائي؟
5- ارسم معادلة الانحدار على نقط الانتشار في المطلوب (1) .

الحـل

1- رسم نقط الانتشار:



قيمة القروض x

من المتوقع أن يكون لقيمة القروض أثر طردي (إيجابي) على مقدار الزيادة في عدد الأنشطة.

2- تقدير معادلة الانحدار.
بفرض أن x هي قيمة القروض، y هي مقدار الزيادة في عدد الأنشطة، يمكن تطبيق المعادلتين في السابقتين، ومن ثم يتم حساب المجاميع التالية:

قيمة القروض x	الزيادة في عدد الأنشطة 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 $\hat{\beta}_1$ كما يلي:

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

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

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

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

- إذا معادلة الانحدار المقدرة، هي:

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

3- تفسير المعادلة:

- الثابت $\hat{\beta}_0 = 9.44$: يدل على أنه في حالة عدم صرف قروض، فإن عدد الأنشطة تزيد **9.44**.
- معامل الانحدار $\hat{\beta}_1 = 0.143$: يدل على أنه كلما زادت قيمة القروض مليون ريال، حدث زيادة فيعدد الأنشطة بمقدار **0.143** نشاط،
- مقدار الزيادة في عدد الأنشطة عند $x = 50$ هو:

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

وأما ومقدار الخطأ العشوائي هو:

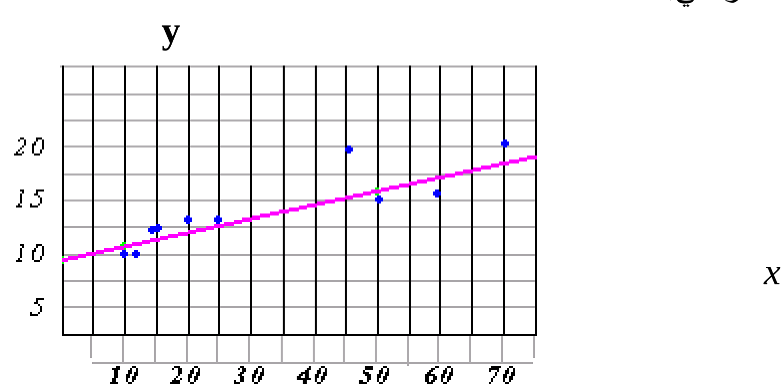
$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

4- 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:



3/3/5 استخدام برنامج SPSS لتقدير نموذج الانحدار الخطي البسيط

تمرين

رغب احد البنوك معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية لهم ،فقاموا بجمع معلومات عن هذا الموضوع وذلك بسحب عينة من 10 موظفين وحصلوا على النتائج التالية.

الموظفين	a	b	c	d	e	f	g	h	i	j
ساعات العمل X	**6**	**4**	**6**	**13**	**11**	**15**	**5**	**8**	**2**	**8**
مستوى الإنتاجية y	**5**	**4**	**4**	**9**	**12**	**14**	**3**	**6**	**1**	**3**

والمطلوب:باستخدام برنامج **SPSS** :

- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
- قدر معادلة انحدار مستوى الإنتاجية على ساعات العمل.

الحل :

2- مشاكل الانحدار

تتمثل هذه المشاكل في وجود الارتباط الذاتي لحدود الخطأ والتباين المختلف لحدود الخطأ والارتباط القوي للمتغيرات المستقلة. ويجب تصحيح هذه الحالات قبل تطبيق طريقة المربعات الصغرى.

6-2-1 الارتباط الذاتي لحدود الخطأ

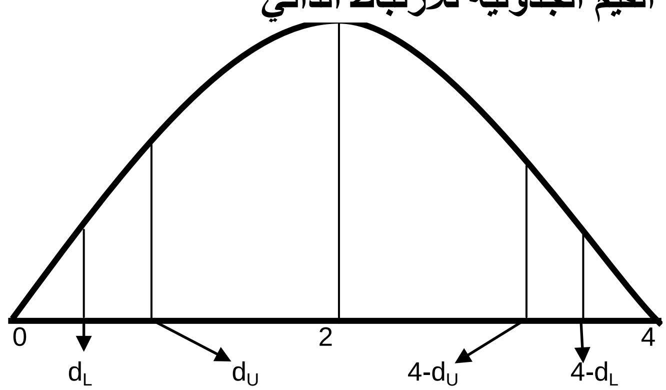
يمكن اكتشاف وجود الارتباط الذاتي لحدود الخطأ من خلال حساب معامل دوربن واتسون (DW) الذي يحسب من العلاقة:

$$DW = \sum(e_i-e_{i-1})^2/\sum e_i^2$$

يقارن DW بالقيم الجدولية. فإذا كان أصغر من d_L يستدل على وجود الارتباط الذاتي، وإذا وقع بين d_L و d_U لايمكن رفض أو قبول فرضية وجود الارتباط الذاتي، وإذا كان أكبر من d_U و أصغر من $4-d_U$ يستدل على عدم وجود الارتباط الذاتي، وإذا وقع بين $4-d_U$ و $4-d_L$ لا يمكن رفض أو قبول الفرضية، وإذا كان أكبر من $4-d_L$ يستدل على وجود الارتباط الذاتي (الشكل 11). في حال وجود الارتباط الذاتي يحسب معامل التصحيح ρ من العلاقة التالية وتصحح البيانات ثم يعاد حساب معاملات الانحدار.

$$Y_i = ß_1 + \rho Y_{i-1} + ß_2\,(x_i-x_{i-1})$$

الشكل 11 توضيح القيم الجدولية للارتباط الذاتي



وعليه يمكن اتباع الخطوات التالية لاكتشاف الارتباط الذاتي وتصحيحه:

- يجرى انحدار بين المتغير التابع والمتغير المستقل أو المتغيرات المستقلة.
- تحسب قيم الانحدار للمتغير التابع من خلال معاملات الانحدار.
- يحسب حد الخطأ للمشاهدات (ei) من خلال طرح القيم المحسوبة للمتغير التابع من قيم المشاهدات الأصلية، والفرق في حدود الخطأ (ei-ei-1)، ثم يتم تربيع الحدود (ei) و (e_i-e_{i-1}) وجمعها لكامل السلسلة.
- يحسب معامل دوربن واتسون ويقارن مع القيم الجدولية.
- اذا كان هناك ارتباط ذاتي يحسب معامل التصحيح وتصحح المشاهدات.
- يجرى الانحدار من جديد.

2-2-6 التباين المختلف لحدود الخطأ

يمكن اكتشاف وجود مثل هذه المشكلة من خلال تطبيق فحص ف لغولد فيلد كوانت على الشكل التالي:

- رتب المشاهدات تصاعدياً حسب المتغير الذي تعتقد أنه يشكل هذه المشكلة (يمكن إجراء ارتباط بين حد الخطأ والمتغيراتالمستقلة لتحديد ذلك ويتم اختيار المتغيرذو معامل الارتباط الاعلى).
- اسقط خمس المشاهدات من متوسط العينة لزيادة قوة الفحص.
- نفذ الانحدار على المجموعتين الناتجتين واحسب قيمة الفحص ف (F) على الشكل التالي:

$$F(n_1-k_1, n_2-k_2) = (1/n_1-k_1*\sum e_{i1}^2)/(1/n_2-k_2*\sum e_{i2}^2)$$

حيث أن n و k ترمز إلى عدد المشاهدات والمتغيراتالمستقلة على التوالي. إذا كانت قيمة ف المحسوبة أصغر من قيمة ف الجدولية نقبل فرضية عدم وجود تباين مختلف

وإذا كانت أكبر من قيمة ف الجدولية يكون عكس ذلك. وفي حال وجود تباين مختلف تقسم المشاهدات على الجذر التربيعي للمتغيرالذي يسبب هذا التباين.

3-2-6 الارتباط القوي للمتحولات المستقلة

في حال وجود هذه المشكلة يفضل إعادة جمع بيانات العينة. وفي حال وجود الارتباط التام بين المتغيراتلايمكن إجراء تقديرات الانحدار. أما في حال وجود الارتباط القوي الجزئي يمكن تقدير الانحدار. ولتلافي جزء من المشكلة يمكن تطبيق طريقة الفروقات المطبقة في تصحيح الارتباط الذاتي.

تمرين:

إذا توفرت لدينا البيانات الآتية لسلعة ما

الكمية المطلوبة y	السعر(ريال) x1
40	9
45	8
50	9
55	8
60	7
70	6
65	6
65	8
75	5
75	5
80	5
100	3
90	4
95	3
85	4

المطلوب:

1 – أوجد معامل الارتباط بين السعر والكمية المطلوبة.

2 – اختبر معنوية معامل الارتباط.

3 – قدر معادلة انحدار y على x واختبر معنوياتها