الفصل السابع

الإنحدار Regression

الانحدار أو يسمى التنبؤ Prediction وهو تقدير القيمة المستقبلية لمتغير واحد بناءً على معرفة قيم متغير أو أكثر، وهنالك عدة أنواع من معامل الإنحدار:

1) الإنحدار الخطي البسيط Simple Linear Regression تشير تسمية هذا المعامل "بسيط" الى أنه يتضمن متغير تابع y يعتمد على متغير واحد مستقل x وكلمة خطي تشير الى أن العلاقة ين المتغيرين y و x هي علاقة خطية.

2) الإنحدار المتعدد Multiple Linear Regression هذا النوع من الإنحدار يتضمن إعتماد المتغير y على أكثر من متغير مستقل مثل x1 و x2 ... الخ.

3) الإنحدار غير الخطي Non-Linear Regression إذا كانت العلاقة بين المتغير y والمتغيرات المستقلة غير خطية مثل علاقة أسية أو لوغاريتمية أو تربيعية ... الخ. وهنالك أنواع أخرى مثل الإنحدار الهرمي Hierarchical Regression والإنحدار التدريجي Stepwise Regression وغيرها.

الإنحدار الخطي البسيط Simple Linear Regression

الانحدار الخطي البسيط هو أداة إحصائية تستعمل لبيان العلاقة بين متغيرين كميين بحيث يمكن توقع قيمة المتغير التابع Dependent variable (y) من المتغير المستقل Independent variable (x) . على سبيل المثال، إذا كان الباحث يعرف العلاقة بين النسبة المئوية لتراكم المادة الجافة وإنتاجية الحنطة فانه يمكنه التنبؤ بالإنتاجية عن طريق الإنحدار الخطي البسيط بمجرد تحديد مستوى تراكم المادة الجافة، بصورة عامة يستعمل الإنحدار للأغراض الآتية:

1) تعد هذه الطريقة تقنية لنمذجة وتحليل البيانات العددية.

2) إستغلال العلاقة بين متغيرين للتنبوء بقيم أحد المتغيرات من خلال قيم المتغير الأخر.

3) التنبؤ وتقدير وإختبار فرضية ونمذجة العلاقات السببية.

هناك العديد من نماذج الانحدار ولكن النموذج الأكثر أهمية والأكثر شيوعاً في الاستعمال هو نموذج الانحدار الخطي بسيط، في هذا النموذج يوجد لدينا المتغير التابع y المعروف أيضاً باسم متغير الإستجابة والمتغير المستقل x المعروف أيضاً باسم المتغير المتنبئ، ويمكن ذكر النموذج على النحو التالي:

$$y = a + bx + e$$

إذ أن:

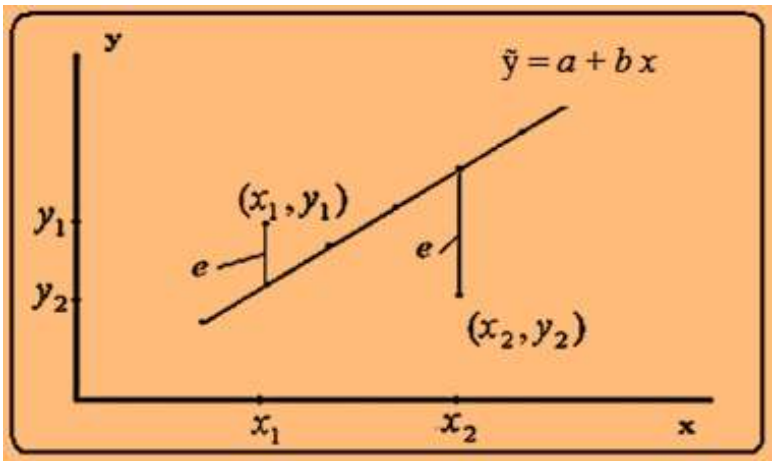
y : المتغير التابع.

x : المتغير المستقل.

a : ثابت الإنحدار وهو الجزء المقطوع من المحور العمودي y الذي يعكس قيمة المتغير التابع y في حالة عدم وجود قيمة للمتغير المستقل x ، بمعنى أخر $(x = 0)$.

b : معامل الإنحدار (الميل) وهو مقدار التغيير في y إذا تغيرت x وحدة واحدة، ويساوي منحدر الخط المستقيم $(a + b\,x)$.

e : الخطأ العشوائي الذي يشير إلى الفرق بين القيمة الفعلية للمتغير التابع y والقيمة المقدرة التي يرمز لها $\tilde{y} = a + b\,x$ ، وهذا يعني أن الخطأ العشوائي يساوي $e = y - (a + b\,x)$ ، ويمكن توضيح هذا الخطأ في الشكل البياني أدناه:



الفرق بين نموذجي الإنحدار الخطي البسيط النظري والفعلي

النموذج النظري (المفترض): نموذج يفترض أن العلاقة بين المتغيرين (x و y) تحكمها المعادلة الآتية:

$$y = a + bx + e$$

النموذج الفعلي (المقدر): نموذج عملي يقدره الباحث بجمع البيانات عن المتغيرين (x و y) ومن ثم حساب معاملات الانحدار (a و b) (الثابت والميل) ويعبر عن هذا النموذج بالمعادلة الآتية:

$$\hat{y} = \hat{a} + \hat{b}\, x + e$$

بعد تقدير النموذج يمكن حساب الخطأ العشوائي لكل عينة كمايلي:

$$\hat{e} = y - \hat{y} = y - (\hat{a} + \hat{b}x)$$

تقدير نموذج الانحدار الخطي البسيط

هنالك عدة طرق لتقدير أو حساب نموذج الانحدار الخطي البسيط وكل الطرق تعتمد على حساب قيم معاملات الانحدار (a و b)، وتعد طريقة المربعات الصغرى Least Squares Method (OLS) من أفضل الطرق لأنها تجعل مجموع مربعات الأخطاء العشوائية أقل ما يمكن، ولحساب القيمة المقدرة لمعامل الإنحدار البسيط للمتغير التابع y بدلالة المتغير المستقل x تطبق المعادلة الآتية:

$$\hat{y} = \hat{a} + \hat{b}\,x + e$$

ولحساب قيمة $\hat{b}$ كمايلي:

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

وتحسب قيمة $\hat{a}$ كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

مثال(1):

لدراسة علاقة الاستهلاك المحلي لمادة الإسفلت (بالمليون برميل) خلال عدة سنوات اخذنا عشر قراءات تقريبية كما يلي:

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

أوجد معادلة الانحدار الخطي البسيط وتوقع قيمة الاستهلاك عندما يصل إنتاج 16 مليون برميل؟

الحل :

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

$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$

x	y	xy	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**
90	65	632	**942**

$\hat{b} = \frac{10(632)-(90)(65)}{10(942)-(90)^2} = 0.36$

$\bar{y} = 6.5$

$\bar{x} = 9$

$\hat{a} = 6.5 - (0.36)(9) = 3.26$

$\hat{y} = 3.26 + 0.36\ x$

ولتوقع قيمة الاستهلاك نعوض في النموذج اعلاه x=16

$\hat{y} = 3.26 + 0.36\ (16)$

$= 9.02$

أي أن الاستهلاك قد يصل إلى 9.02 مليون برميل، أي ما يعادل 9020000 برميل خلال السنة

مثال(2):

فيما يلي بيانات إضافة السماد المركب لنباتات البطاطا (x)، ومقدار الزيادة في تركيز النشاء في درنات النبات (y)، عند قياس عينة من 10 نباتات لكل معاملة سماد.

Fertilizer (x)	10	11	14	15	20	25	46	50	59	70
Starch (y)	10	10	12	12	13	13	19	15	16	20

المطلوب :

1) ارسم نقاط الانتشار؟

2) قدر معادلة انحدار تركيز النشاء في الدرنات (y) على كمية السماد المضاف (x)؟

3) فسر معادلة الانحدار؟

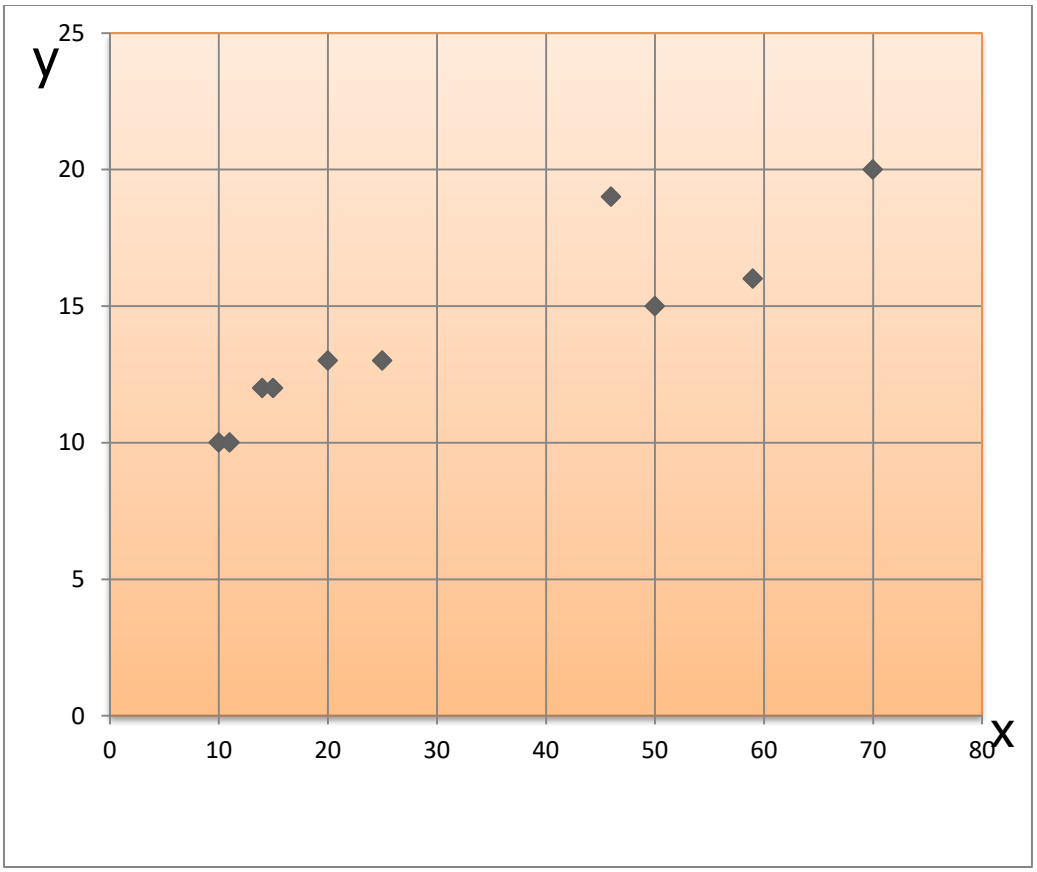
4) ما هو مقدار الزيادة في تركيز النشاء عند التسميد بمقدار 50 كغم.هـ$^{-1}$ (x = 50)؟

5) ما هي مقدار الخطأ العشوائي؟

6) ارسم معادلة الانحدار على نقط الانتشار، وما هي توقعاتك لشكل العلاقة؟

الحل:

1) رسم نقاط الانتشار



2) حساب نموذج الإنحدار المقدر بتطبيق المعادلة $\hat{y} = \hat{a} + \hat{b}\,x + e$ تتبع الخطوات الآتية:

– حساب قيم العلاقات بين x و y كما في الجدول الآتي:

x	y	xy	x^2	y^2
10	10	100	100	100
11	10	110	121	100
14	12	168	196	144
15	12	180	225	144
20	13	260	400	169

25	13	325	625	169
46	19	874	2116	361
50	15	750	2500	225
59	16	944	3481	256
70	20	1400	4900	400
$\sum \boldsymbol{x = 320}$	$\sum y = 140$	$\sum xy = 5111$	$\sum x^2 = 14664$	$\sum y^2 = 2068$

$$\sum x^2 = 14664 \qquad n = 10 \qquad \sum x = 320 \qquad \sum y = 140$$

$$\sum y^2 = 2068 \qquad \bar{y} = \frac{140}{10} = 14 \qquad \bar{x} = \frac{320}{10} = 32$$

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

$$= \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2} = \frac{6310}{44240} = 0.1426$$

الان يمكن حساب قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

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

اذن حساب قيمة النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 9.4368 + 0.1426\,x$$

3) تفسير معادلة الإنحدار:

1) معامل الإنحدار الثابت $\hat{a} = 9.4368$ يدل على أنه في حالة عدم إستعمال التسميد (x = 0) فان تركيز النشاء يزداد بمقدار 9.4368 ملغم..

2) معامل الإنحدار (الميل) $\hat{b} = 0.143$ يدل على أنه كلما زادت كمية السماد كغم. واحد فان هنالك زيادة تحدث في تركيز النشاء بمقدار 0.1426 ملغم..

4) ما هو مقدار الزيادة في تركيز النشاء عند التسميد بمقدار 50 كغم. (x = 50)؟

$$\hat{y} = 9.4368 + 0.1426\, x$$
$$\hat{y} = 9.4368 + 0.1426\ (50) = 16.59$$

5) لتقدير قيمة الخطأ العشوائي عند (x = 50) تستخرج قيمة تقاطع $y_{x=50}$ من المخطط البياني أعلاه والتي تساوي 15 ثم تطبق المعادلة كمايلي:

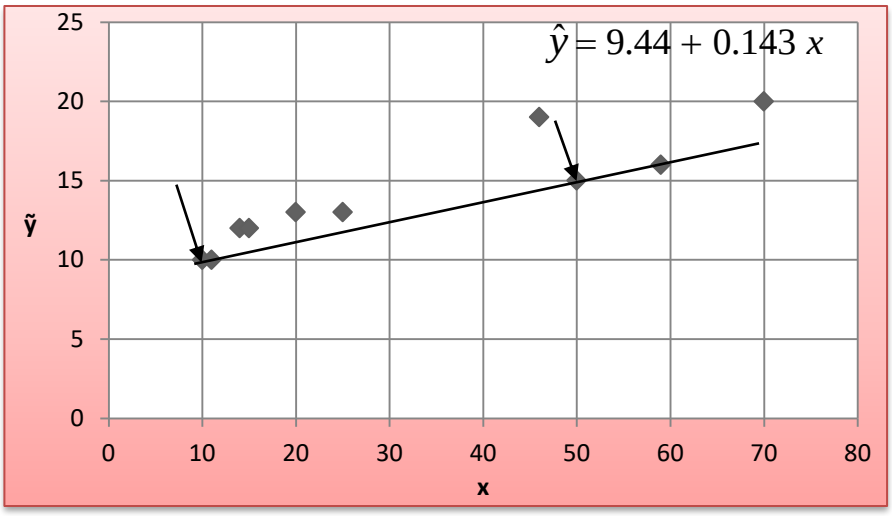
$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

6) رسم معادلة الانحدار على نقاط الانتشار:

ملاحظة: من البديهي يمكن رسم خط مستقيم إذا علم أي نقطتين على ذلك الخط، فاذا كانت:

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

لذا يمكن رسم الخط المستقيم كمايلي:



يستنتج من ذلك وجود علاقة الخط المستقيم بين المتغيرين

الإنحدار الخطي المتعدد *Multiple Linear Regression*

يعد الإنحدار الخطي المتعدد من الأساليب الإحصائية المتقدمة والتي تضمن دقة الإستدلال من أجل تحسين نتائج البحث عن طريق الإستخدام الأمثل للبيانات في إيجاد علاقات سببية بين الظواهر موضوع البحث .

والإنحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ قيم مستقبلية ، وهو عبارة أيضاً عن إنحدار للمتغير التابع (**Y**) على العديد من المتغيرات المستقلة $X_1 , X_2 , ...X_K$ لذا فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي يؤثر فيه عدة متغيرات مستقلة أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو االانتشار ، فبإمكاننا التنبؤ بالمستوى الرقمي في فعالية رمي المطرقة على سبيل المثال إعتماداً على دراسة حالات أخرى للرامي كالعمر الزمني والعمر التدريبي والمهارة والمواصفات الجسمية وغيرها .

إن الإنحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن أستخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادةً ما تكون مستمرة)

والمعادلة الخطية في الإنحدار الخطي المتعدد هي :

$$Y = a + b_1X_1 + b_2X_2 + \dots\dots\dots + e$$

حيث أن **Y** = المتغير التابع

a = قيمة ثابتة ***Constant*** أو ***Intercept***

b_1 = ميل ألإنحدار ***y*** على المتغير المستقل الأول

b_2 = ميل ألإنحدار ***y*** على المتغير المستقل الثاني

X_1 = المتغير المستقل الأول

X_2 = المتغير المستقل الثاني

ويمكن استخدام الإنحدار الخطي المتعدد في حالة توافر الشروط التالية :

1. أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .
2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات المستقلة والمتغير التابع .
3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل .

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية احصائياً مع التنويه بأن المعنوية تكون لكل معامل على حدة .

ولكي نحكم على معنوية معاملات الإنحدار نستعين باختبار ***T*** ومستوى الاحتمالية المقابل له وبالطبع فإن برنامج ***SPSS*** سيقوم تلقائياً باستخراج اختبار ***T*** ومستوى الاحتمالية المقابل له .

كما سيتم الحصول على إحصائيات تستخدم لمعرفة المعنوية الإجمالية للنموذج ومنها (R) ، (R^2) ، (R^{2-}) .

فالأول R هو معامل الأرتباط البسيط والذي يقيس قوة العلاقة بين متغيرين أو أكثر ، أما R^2 فهو يسمى بمعامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الإنحدار الخطي البسيط (متغير مستقل واحد مع متغير معتمد واحد) ، أما R^{2-} فهو يستخدم لتفسير القوة التفسيرية لنموذج الانحدار الخطي المتعدد (لأنه يأخذ بنظر الاعتبار عدد المتغيرات المستقلة ولذلك يسمى بالمصحح لأنه بالأصل مشتق من R^2) .

كما نستخدم أيضاً أحصائية F للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين .

معامل التحديد Determination Coefficient

هو مقياس لتقدير دقة معامل الانحدار ويرمز له R^2 وذلك لأنه يساوي مربع معامل الإرتباط البسيط ويأخذ هذا المعامل قيم بين 0 الى 1 أي أنه $0 \leq r^2 \leq 1$ وكلما إقتربت قيمة معامل التحديد من 1 فان ذلك يدل على قلة قيمة الخطأ العشوائي، مثال ذلك لو كانت قيمة معامل التحديد لمتغيرين تساوي 0.87 فهذا يفسر (يدل) على أن معادلة الإنحدار تفسر 87% من التغير الحاصل في المتغير التابع y حدثت

بسبب التغير الحاصل في المتغير المستقل x والباقي من التغير البالغ 13% حدث بسبب عوامل أخرى غير المتغير المستقل x، ويتم حساب R^2 على النحو التالي:

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$
$$= 2068 - 9.44(140) - 0.1426(5111) = 17.5714$$
$$SS_T = \sum y^2 - n(\bar{y})^2 = 2068 - 10(14)^2 = 108$$
$$SS_R = SS_T - SS_e = 108 - 17.5714 = 90.4286$$
$$R^2 = \frac{SS_R}{SS_T} = \frac{90.4286}{108} = 0.83$$

SS_e : مجموع مربعات الخطأ التجريبي.

SS_T : مجموع المربعات الكلية.

SS_R : مجموع مربعات الإنحدار.

مثال (2) :

أراد أحد الباحثين التنبؤ بالزيادة الحاصلة في تركيز المادة القلويدية الفعالة (y) في كالس نبات الداتورة عند إضافة تراكيز من Ascorbic acid (x) الى الوسط الغذائي للكالس، وعند تقدير 13 عينة من الكالس حصل الباحث على البيانات الآتية:

Ascorbic acid (x)	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.25	3.5
Alkaloids (y)	10	8	12	12	14	12	16	18	17	20	18	20	21

المطلوب :

1) ارسم نقاط الانتشار؟

2) قدر معادلة انحدار تركيز المادة القلويدية الفعالة (y) بدلالة تراكيز من Ascorbic acid (x)؟

3) فسر معادلة الانحدار؟

4) ما هو مقدار الزيادة في تركيز المادة القلويدية الفعالة عند إضافة تركيز 3.0 من Ascorbic acid (x = 3.0)؟

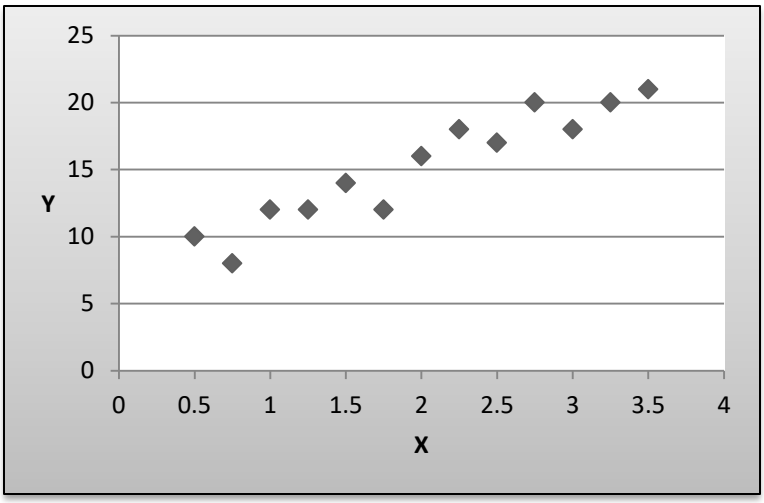
5) ما هي مقدار الخطأ العشوائي؟

6) ما هي قيمة معامل التحديد وكيف يفسر؟

7) ارسم معادلة الانحدار على نقط الانتشار، وما هي توقعاتك لشكل العلاقة؟

الحل:

1) رسم نقاط الانتشار؟



2) حساب نموذج الإنحدار المقدر بتطبيق المعادلة $\hat{y} = \hat{a} + \hat{b}\,x + e$ تتبع الخطوات الآتية:

- حساب قيم العلاقات بين x و y كما في الجدول الآتي:

n	x	y	x^2	y^2	xy
1	0.5	10	0.25	100	5
2	0.75	8	0.5625	64	6
3	1	12	1	144	12
4	1.25	12	1.5625	144	15
5	1.5	14	2.25	196	21
6	1.75	12	3.0625	144	21
7	2	16	4	256	32
8	2.25	18	5.0625	324	40.5
9	2.5	17	6.25	289	42.5
10	2.75	20	7.5625	400	55
11	3	18	9	324	54
12	3.25	20	10.5625	400	65
13	3.5	21	12.25	441	73.5

$\boldsymbol{n}$	$\sum x = 26$	$\sum y = 198$	$\sum x^2 = 63.375$	$\sum y^2 = 3226$	$\sum xy = 442.54$
$= \mathbf{13}$					

$n = 13$ $\sum xy = 442.54$ $\sum x = 26$ $\sum y = 198$

$\sum x^2 = 63.375$ $\sum y^2 = 3226$ $\bar{y} = \frac{198}{13} = 15.23$ $\bar{x} = \frac{26}{13} = 2$

- حساب قيمة $\hat{b}$ بتطبيق المعادلة كمايلي:

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(13)(442.54) - (26)(198)}{(13)(63.375) - (26)^2}$$

$$= \frac{605.02}{147.875} = 4.09$$

اذن قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 15.23 - (4.09)(2) = 7.05$$

النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 7.05 + 4.09\,x$$

3) تفسير معادلة الإنحدار:

1) معامل الإنحدار الثابت $\hat{a} = 7.05$ يدل على أنه في حالة عدم إستعمال Ascorbic acid (x = 0) فان تركيز المادة القلويدية الفعالة تزداد بمقدار 7.05 ملغم.100غم$^{-1}$.

2) معامل الإنحدار $\hat{b} = 4.09$ يدل على أنه كلما زاد تركيز Ascorbic acid وحدة واحد فان هنالك زيادة تحدث في تركيز المادة القلويدية الفعالة بمقدار 4.09 ملغم.100غم$^{-1}$.

4) ما هو مقدار الزيادة في تركيز المادة القلويدية الفعالة عند تركيز الحامض 3.0 ملغم.لتر$^{-1}$(x = 3.0)؟

$$\hat{y} = 7.05 + 4.09\,x$$
$$\hat{y} = 7.05 + 4.09\,(3.0) = 19.32$$

5) لتقدير قيمة الخطأ العشوائي عند (x=3.0) تستخرج قيمة تقاطع $y_{x=3.0}$ من المخطط البياني أعلاه والتي تساوي 18 ثم تطبق المعادلة كمايلي:

$$\hat{e}_{x=3.0} = y_{x=3.0} - \hat{y}_{x=3.0} = 18 - 19.32 = -1.32$$

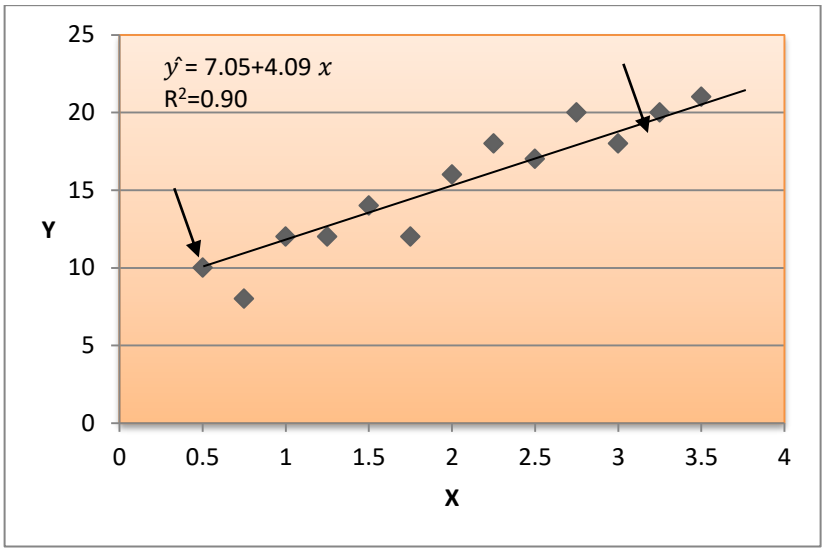
6) حساب قيمة معامل التحديد R^2 وكيف يفسر؟

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$
$$= 3226 - 7.05\,(198) - 4.09\,(442.54) = 20.1$$
$$SS_T = \sum y^2 - n(\bar{y})^2 = 3226 - 13(15.23)^2 = 210.6$$
$$SS_R = SS_T - SS_e = 210.6123 - 20.1114 = 190.5$$
$$R^2 = \frac{SS_R}{SS_T} = \frac{190.5}{210.6} = 0.90$$

7) رسم معادلة الانحدار على نقاط الانتشار:

ملاحظة: من البديهي يمكن رسم خط مستقيم إذا علم أي نقطتين على ذلك الخط، فاذا كانت:

x	3.0	0.5
$\hat{y}$	19.32	10.0



يستنتج من ذلك وجود علاقة الخط المستقيم بين المتغيرين

مثال (3):

البيانات الآتية تمثل المتغير y والمتغير x لتجربة ما كمايلي:

$$n = 7 \qquad \sum xy = 2093 \qquad \sum x = 103 \qquad \sum y = 144$$

$$\sum x^2 = 1531 \qquad \sum y^2 = 3012 \qquad \bar{y} = \frac{144}{7} = 20.5 \qquad \bar{x} = \frac{103}{7} = 14.7$$

المطلوب:

1- تقدير معادلة انحدار المتغير y بدلالة المتغير x .

2- تقدير المتغير y عندما يكون المتغير x يساوي 10.

3- حساب معامل التحديد وتفسير معناه.

الحل:

1) تقدير معادلة انحدار المتغير y بدلالة المتغير x .

- حساب قيمة $\hat{b}$ بتطبيق المعادلة كمايلي:

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(7)(2093) - (103)(144)}{(7)(1531) - (103)^2}$$

$$= \frac{-181}{108} = -1.68$$

- حساب قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 20.5 - (-1.68)(14.7) = 45.29$$

النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 45.29 - 1.68\,x$$

2) تقدير المتغير y عندما يكون المتغير x يساوي 10.

$$\hat{y} = 45.29 - 1.68\,x$$

$$\hat{y} = 45.29 - 1.68\,(10) = 28.49$$

3) حساب قيمة معامل التحديد R^2 وكيف يفسر؟

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$

$$= 3012 - 45.29\,(144) - (-1.68)(2093) = 6.48$$

$$SS_T = \sum y^2 - n(\bar{y})^2 = 3012 - 7(20.5)^2 = 70.25$$

$$SS_R = SS_T - SS_e = 70.25 - 6.48 = 63.77$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{63.77}{70.25} = 0.90$$

قيمة معامل التحديد لمتغيرين تساوي 0.90 وهذا يفسر (يدل) على أن معادلة الإنحدار تفسر 90% من التغير الحاصل في المتغير التابع y حدثت بسبب التغير الحاصل في المتغير المستقل x والباقي من التغير البالغ 10% حدث بسبب عوامل أخرى غير المتغير المستقل x.