

## Simple regression and correlation

- ١ - مقدمة
- ٢ - معادلة خط الانحدار
- ٣ - استخدام المصفوفات في الانحدار
- ٤ - مصادر الاختلاف في الانحدار الخطي
- ٥ - القيم المعدلة لأثر انحدار Y على X
- ٦ - الانحرافات المعيارية وحدود الثقة للتقديرات المختلفة
- ٧ - اختبار معنوية معامل الانحدار
- ٨ - المقارنة بين معاملي الانحدار
- ٩ - التوزيع ذو المتغيرين
- ١٠ - العلاقة بين معامل التحديد وخطأ التقدير
- ١١ - تقييم ملائمة نموذج التحليل
- ١٢ - الارتباط البسيط
- ١٣ - العلاقة بين معامل الانحدار ومعامل الارتباط
- ١٤ - اختبار معنوية معامل الارتباط وتقدير حدود الثقة له
- ١٥ - اختبار تساوى معاملي ارتباط
- ١٦ - اختبار تجانس عدد من معاملات الارتباط
- ١٧ - السبب والأثر في تحليل الارتباط والانحدار
- ١٨ - تبين الدالة الخطية
- ١٩ - ارتباط الرتب



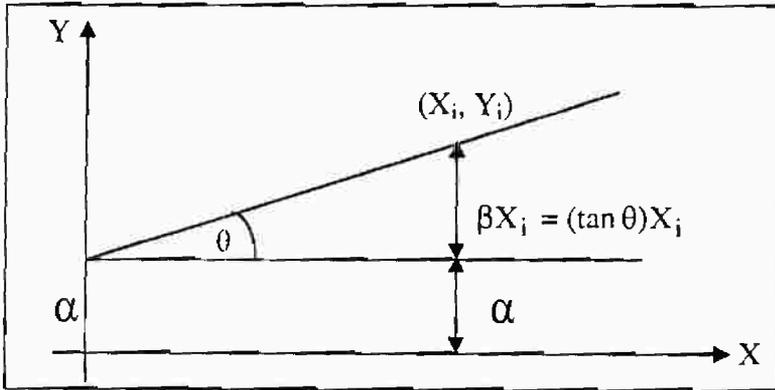
## ١-١٠ مقدمة

يطلق على الانحدار regression الارتداد أو الاعتماد ويستخدم لدراسة العلاقة بين متغيرين أحدهما  $Y$  والذي يعتمد في قيمته على متغير آخر  $X$ . ويطلق على  $Y$  المتغير التابع أو المعتمد dependent variable بينما يطلق على  $X$  المتغير المستقل explanatory variable وقد يسمى أيضاً بالمتغير التفسيري independent variable أى الذى يفسر التغيرات فى  $Y$ . ويوجد أمثلة كثيرة لذلك منها دراسة العلاقة بين الأسمدة  $X$  وكمية المحصول  $Y$ ، دراسة العلاقة بين العمر عند أول ولادة  $X$  ومحصول اللبن  $Y$ ، دراسة العلاقة بين الطول  $X$  والوزن  $Y$ ، دراسة العلاقة بين محيط الصدر  $X$  والوزن  $Y$ . وعادة ما يكون المتغير المستقل هو المتغير الأسهل فى القياس عن المتغير التابع. ويعبر عن العلاقة بين المتغيرين رياضياً  $Y = f(X)$ ، أى أن  $Y$  دالة  $X$  ويعبر عنها إحصائياً بالانحدار.

وقد تعبر المعادلة  $Y = 3 + 5X$  على سبيل المثال عن العلاقة بين المتغيرين، وهى تمثل معادلة خط مستقيم صورته العامة:

$$Y = \alpha + \beta X \quad (١-١٠)$$

حيث  $\alpha$  هى الجزء المقطوع من محور الصادات intercept، أى قيمة  $Y$  عندما تكون قيمة  $X$  مساوية للصفر،  $\beta$  تمثل ميل slope الخط المستقيم، أى ظل  $\tan$  الزاوية  $\theta$  التى يصنعها المستقيم فى الاتجاه الموجب لمحور السينات. كما يطلق على  $\beta$  معامل الانحدار أو الاعتماد regression coefficient. وشكل ١-١٠ يمثل تلك العلاقة بيانياً.

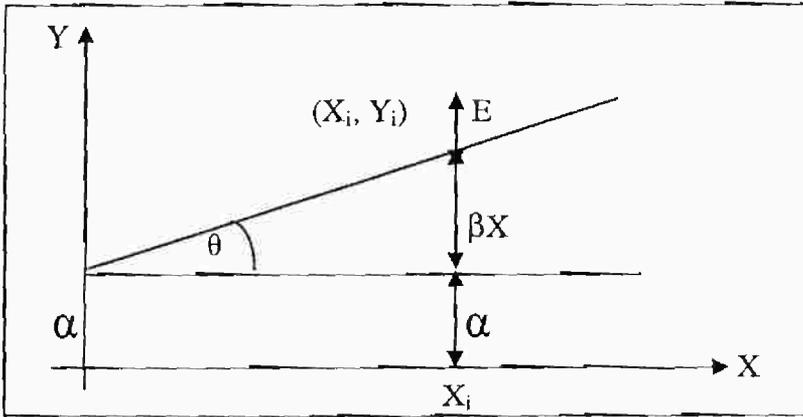


شكل ١-١٠ تمثيل العلاقة  $Y = \alpha + \beta X$  بيانياً

ومن الشكل يلاحظ أن لكل قيمة من المتغير المستقل  $X$  قيمة تقابلها للمتغير التابع  $Y$  وتقع كل قيم  $Y$  على خط مستقيم واحد.

ولدراسة العلاقة بين متغيرين وليكن أحدهما وزن الحيوان بالكيلوجرام والآخر العمر بالأسبوع، فإن أول ما يجب عمله هو تمثيل تلك العلاقة بيانياً على أن يمثل المحور السيني المتغير المستقل، وهو العمر في مثل هذه العلاقة، ويمثل المحور الصادي الوزن في شكل يعرف بشكل الانتشار scatter diagram ومنه يتضح إذا كانت العلاقة خطية أو غير ذلك. وإذا كانت العلاقة خطية، أى أن النقط في شكل الانتشار تتجمع حول خط مستقيم، حيث غالباً أن العلاقة لا تتحقق تماماً بمعنى أن بعض النقط تكون فوق الخط والبعض تحته، وقد تكون هناك عدة نقط على الخط. والمتغير التابع  $Y$  يمكن التعبير عنه بالعلاقة (٢-١٠) والشكل ٢-١٠ التاليين:

$$Y = \alpha + \beta X + \epsilon \quad (2-10)$$



شكل ٢-١٠ تمثيل العلاقة  $Y = \alpha + \beta X + \epsilon$  بيانياً

وتمثل  $\epsilon$  الخطأ الخاص المصاحب للمتغير  $Y$  ويطلق عليه الخطأ العشوائى أو التجريبى وقد يرجع إلى:

١ - خطأ في قياس المتغير  $Y$  errors of measurement

٢ - قد يكون هناك متغيرات أخرى غير  $X$  تؤثر على  $Y$  ولكن أهملت باعتبار أن  $X$  هو المتغير الأساسى محل الدراسة، ويعبر عنها بالمتغيرات المحذوفة omitted variables والتي تدخل ضمن مكونات الخطأ.

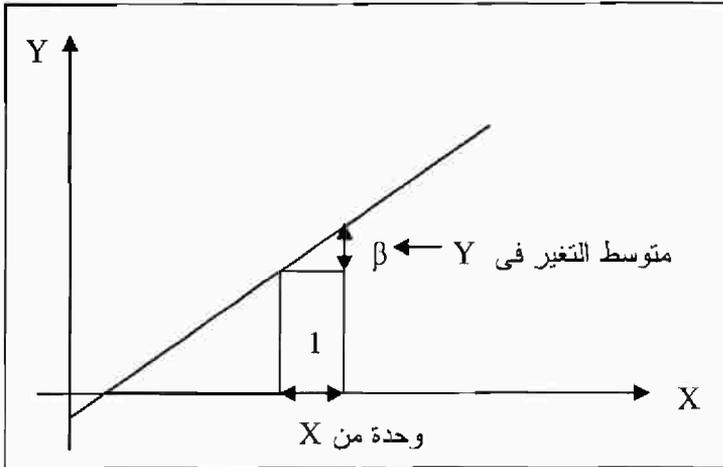
والعلاقة (٢-١٠) هنا تمثل علاقة حقيقية true relationship والهدف هو تقدير معالم هذه العلاقة وهما  $\alpha$  و  $\beta$ . وهذه العلاقة مثل للبيانات البيولوجية biological data لوجود عنصر الخطأ والذي لا يمكن التحكم فيه.

## ١٠-٢ معادلة خط الانحدار

يعرف معامل انحدار المتغير التابع  $Y$  على المتغير المستقل  $X$  بأنه متوسط التغير في المتغير التابع  $Y$  عندما يتغير المتغير المستقل  $X$  بمقدار الوحدة وذلك في مدى معين للمتغير  $X$ ، ويرمز له بالرمز  $\beta_{YX}$ .

فمثلاً عند دراسة العلاقة بين الوزن بالكيلوجرام والعمر باليوم في عجول التسمين كان معامل اعتماد (انحدار) الوزن على العمر هو 0.8 كج/يوم، أي  $\hat{\beta}_{YX} = 0.8$  كج/يوم. حيث  $\hat{\beta}_{YX}$  هو تقدير غير متحيز لمعامل الانحدار في العشييرة. وهذا يعني أنه بزيادة العمر بمقدار يوم فإن الوزن يزيد بمقدار 0.8 كج.

أما إذا كان معامل الانحدار  $\hat{\beta}_{YX} = -3$  وحدة من  $Y$  / وحدة من  $X$ ، فإن هذا يعني أنه بزيادة  $X$  بمقدار الوحدة فإن  $Y$  تنقص بمقدار 3 وحدات. ومعامل الانحدار لا بد وأن يكون مميزاً وقيمته تتراوح بين  $-\infty$  إلى  $+\infty$  ويمثل شكل ١٠-٣ تعريف معامل الانحدار هندسياً.



شكل ١٠-٣ تمثيل معامل الانحدار هندسياً

وترجع أهمية دراسة الانحدار إلى:

- ١ - معرفة ما إذا كانت  $Y$  تعتمد على  $X$  وللحصول على مقياس لتلك العلاقة.
- ٢ - التنبؤ بقيم  $Y$  عند معرفة قيم  $X$ ، وتفسير  $Y$  بواسطة  $X$  (وذلك في مدى معين).
- ٣ - تحديد شكل منحنى الانحدار regression curve

٤ - معرفة الأخطاء الحقيقية الموجودة في التجربة بعد التعديل (التصحيح) لأثر المتغير المستقل.

٥ - اختبار الفروض التي قد يضعها الباحثون حول العلاقة بين المسبب cause وتأثيره effect.

قد لا يفضل لسبب أو لآخر دراسة كل أفراد العشيرة وعلى ذلك يكتفى بدراسة العلاقة بين المتغيرين في عينة أو جزء (عينة) من هذه العشيرة وذلك بغرض الوصول إلى تقدير لكل من  $\alpha$ ،  $\beta$  والتي يعبر عنها بنموذج التقدير estimated model التالي:

$$Y = a + bX + e \quad (3-10)$$

حيث  $a$ ،  $b$  هما تقديران غير متحيزين لكل من  $\alpha$ ،  $\beta$  السابق الإشارة إليهما في (٢-١٠)،  $e$  تقدير لعنصر الخطأ غير المعروف. ولكي تتم عملية التقدير يلزم مجموعة من المشاهدات للمتغير  $X$  وما يقابلها من المتغير  $Y$ .

وفي حالة الانحدار الخطي البسيط simple linear regression يمكن رسم الخط بنقطتين وفي هذه الحالة لا يمكن تقدير خطأ، أما إذا كان هناك مجموعة من النقط فيلزم توفيق خط يسمى خط الانحدار regression line يمثل هذه العلاقة ومعادلته هي:

$$\hat{Y} = a + bX \quad (4-10)$$

وتقدير كل من  $a$ ،  $b$  يكون بإحدى الوسيلتين التاليتين أو غيرهما:

١- حساب الفروق (الأخطاء) بغض النظر عن الإشارة مع محاولة جعل مجموعها أقل ما يمكن. أو

٢- حساب مجموع مربع الأخطاء  $\sum e^2$  من النموذج (٤-١٠)، حيث  $e_i = Y_i - a - bX_i$  وجعله أقل ما يمكن. وهذه الطريقة هي الأكثر شيوعاً واستخداماً وتعرف بطريقة المربعات الصغرى least squares method.

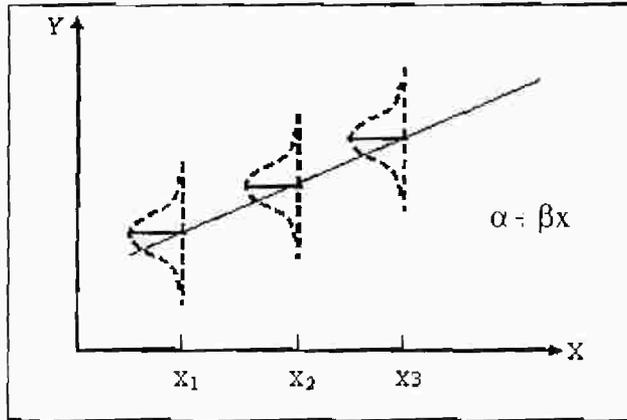
ولكى يتم ذلك توضع افتراضات لازمة لعملية التقدير وهي:

١- المتغير  $X$  ثابت fixed في المعاينات المتكررة repeated sampling أى ليس له توزيع احتمالي.

٢- لا يوجد ارتباط بين كل من المتغير المستقل  $X$  ومكون الخطأ  $e$ .

٣- لكل قيمة للمتغير  $X_i$  يوجد توزيع للمتغير  $Y$  متوسطه يقع على خط الانحدار  $\hat{\mu}_{y.x_i} = a + bX_i$ ، وتختلف المتوسطات للتوزيعات ولكن لها نفس التباين  $\sigma_{Y.X}^2$  وشكل ١٠-٤ يبين ذلك.

٤- الأخطاء  $e_i$  مستقلة وتتوزع طبيعياً بمتوسط يساوى الصفر وتباين  $\sigma_{Y.X}^2$  وهذا الفرض مهم عند اختبار معنوية الانحدار.



شكل ١٠-٤ التوزيع الطبيعي لقيم  $Y$  حول خط الانحدار  $\alpha + \beta x$  لبعض قيم  $X$

ولتطبيق طريقة المربعات الصغرى، أى لجعل مجموع مربعات الأخطاء  $\sum e^2$  أقل ما يمكن، يستخدم كل من التفاضل الجزئى وتفاضل دالة الدالة على النحو التالى:

$$Y_i = a + bX_i + e_i \quad \text{بما أن:}$$

$$e_i = Y_i - a - bX_i \quad \text{إذا}$$

$$e_i^2 = (Y - a - bX_i)^2 \quad \text{وبتربيع الطرفين:}$$

وبالجمع لكل قيم  $e_i$  من 1 إلى  $n$  حيث  $n$  عدد أزواج المشاهدات:

$$\sum e_i^2 = \sum (Y - a - bX_i)^2$$

بالتفاضل الجزئى لكل من  $a$ ،  $b$  ومساواة الناتج بالصفر:

$$\frac{\partial \sum e^2}{\partial a} = 2 \sum (Y_i - a - bX_i)(-1) = 0$$

$$\therefore \sum Y_i = na + b \sum X_i \quad (5-10)$$

$$\frac{\partial \sum e^2}{\partial b} = 2 \sum (Y_i - a - bX_i)(-X_i) = 0$$

$$\therefore \sum X_i Y_i = a \sum X_i + b \sum X_i^2 \quad (6-10)$$

وتسمى المعادلتان (5-10)، (6-10) بالمعادلتين الاعتياديتين normal equations وبحل هاتين المعادلتين أتياً يمكن الحصول على:

$$a = \bar{Y} - b\bar{X} \quad (7-10)$$

$$b_{yx} = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sum X^2 - (\sum X)^2/n} \quad (8-10)$$

ويمكن الوصول إلى صورة أخرى لمعامل الانحدار  $b$  إذا تم التعبير عن قيم المتغيرين كانحراف عن متوسطهما، أى استخدام  $y$  بدلاً من  $Y$  حيث  $x, y = Y - \bar{Y}$  بدلاً من  $X$  حيث  $x = X - \bar{X}$ . وبالتالي يمكن استنتاج أن:

$$b_{yx} = \frac{\sum xy}{\sum x^2} \quad (9-10)$$

وإذا قسم كل من بسط ومقام العلاقة (9-10) على درجات الحرية  $n-1$  فإن:

$$b_{yx} = \frac{\sum xy / (n-1)}{\sum x^2 / (n-1)} = \frac{\text{Cov}(X, Y)}{V(X)} \quad (10-10)$$

ويعرف  $\text{Cov}(X, Y)$  بالتغاير covariance وهو عبارة عن الجزء من التباين المشترك joint variance بين قيم المتغيرين  $X, Y$  وقيمتها قد تكون موجبة أو سالبة أو صفر.

ومعامل الانحدار  $b$  أو  $b_{yx}$  (معامل انحدار  $Y$  على  $X$ ) يأخذ أى قيمة ويستمد إشارته من إشارة التغيرات.

وتكون معادلة الخط المستقيم (معادلة الانحدار) عبارة عن:

$$\hat{Y} = a + bX \quad (11-10)$$

وبالتعويض عن  $a$  بقيمتها  $a = \bar{Y} - b\bar{X}$

$$\hat{Y} = \bar{Y} - b\bar{X} + bX = \bar{Y} + b(X - \bar{X}) = \bar{Y} + bx \quad (12-10)$$

أى أن:

$$\hat{y} = \hat{Y} - \bar{Y} = bx \quad (13-10)$$

ومن خصائص خط الانحدار:

- ١ - مجموع الانحرافات عن خط الانحدار يساوى صفر.
- ٢ - مجموع مربع الانحرافات عنه أقل ما يمكن.
- ٣ - نقطة تقاطع المتوسطين  $(\bar{X}, \bar{Y})$  تقع على هذا الخط.

مثال ١٠-١

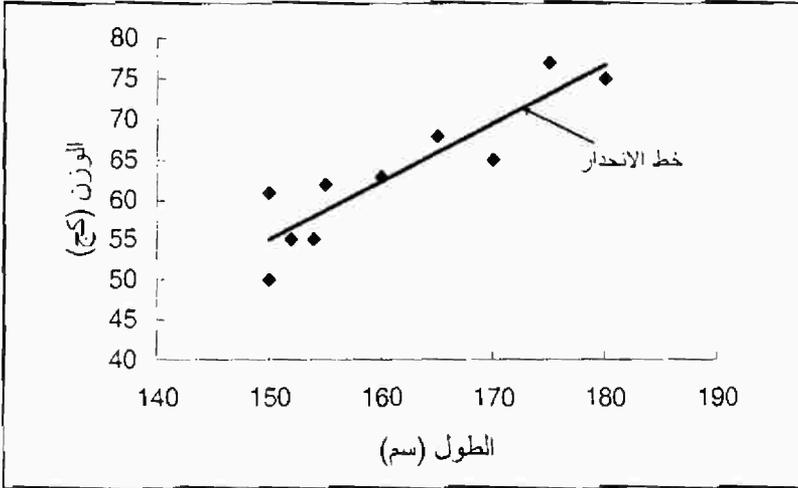
فى دراسة للعلاقة بين الطول  $X$  بالسنتيمتر والوزن  $Y$  بالكيلوجرام فى عشيرة ماء، كانت البيانات كالتالى:

الزوج	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
الطول	150	175	170	160	150	155	152	180	154	165
الوزن	50	77	65	63	61	62	55	75	55	68

المطلوب:

- ١ - تمثيل تلك العلاقة بيانياً.
- ٢ - حساب معامل انحدار الوزن على الطول
- ٣ - كتابة معادلة خط الانحدار ورسم خط الانحدار رسماً دقيقاً.

التمثيل البياني يمثله شكل ١٠-٥، ويلاحظ من دراسة شكل الانتشار أن العلاقة بين الطول والوزن علاقة أقرب ما تكون إلى الخطية (في المدى المدروس). وعلى ذلك يمكن تقدير معادلة الخط المستقيم (معادلة الانحدار)  $\hat{Y} = a + bX$  الذي يمثل هذه العلاقة بحيث يكون مجموع الانحرافات عن هذا الخط مساوية للصفر ومجموع مربع الانحرافات عن نفس الخط أقل ما يمكن، أي أن التقدير يتم بطريقة المربعات الصغرى.



شكل ١٠-٥ انحدار الوزن على الطول

وحيث إن الوزن قد يعتمد على الطول فإن الوزن  $Y$  هو المتغير التابع والطول  $X$  هو المتغير المستقل. ومن المعطيات:

$$n = 10$$

$$\bar{X} = 161.1 \quad \sum X = 1611 \quad \sum X^2 = 260595$$

$$\bar{Y} = 63.1 \quad \sum Y = 631 \quad \sum XY = 102415$$

وبتطبيق المعادلة (١٠-٨) فإن:

$$b_{yx} = \frac{102415 - \frac{(1611)(631)}{10}}{260595 - \frac{(1611)^2}{10}} = \frac{760.9}{1062.9} = 0.716 \text{ kg/cm}$$

وحيث إن:  $a = \bar{Y} - b\bar{X}$  فإن  $a = 63.1 - (0.716)(161.1) = -52.228 \text{ kg}$

ولرسم خط الانحدار (شكل ١٠-٥) تؤخذ نقطة تقاطع متوسطى المتغيرين والجزء المقطوع من المحور الصادى  $a$  ويتم التوصيل بين النقطتين للحصول على خط الانحدار الذى معادلته ستكون كما يلي:

$$\hat{Y} = -52.228 + 0.716X$$

وتسمى معادلة الخط المستقيم هذه معادلة الانحدار ويطلق عليها أيضاً معادلة التنبؤ prediction equation والتي يمكن حساب أى قيمة متوقعة لـ  $Y$  إذا علمت قيمة  $X$ ، فمثلا القيمة المتوقعة للوزن إذا كان الطول 185 سم هي:

$$\hat{Y} = -52.228 + (0.716)(185) = 80.232 \text{ kg}$$

والقيمة المتوقعة للوزن إذا كان الطول 172 سم هي:

$$\hat{Y} = -52.228 + (0.716)(172) = 70.924 \text{ kg}$$

ويمكن أن تقدر القيمة المتوقعة من (١٠-١٢) كما يلي:

$$\begin{aligned} \hat{Y} &= 63.1 + 0.716(x) = 63.1 + (0.716)(172 - 161.1) \\ &= 63.1 + (0.716)(10.9) = 70.904 \end{aligned}$$

والفرق الناتج بين 70.924 ، 70.904 راجع لخطأ التقريب rounding error.

مثال ١٠-٢

يمكن حل مثال ١٠-١ باستخدام PROC REG فى برنامج SAS، وتستخدم هذه الطريقة عند الرغبة فى الحصول على تقدير كل من  $a$  (الجزء المقطوع من المحور الصادى intercept) و  $b$  (معامل الانحدار slope) بالإضافة إلى تحليل التباين والذى سوف يتم تناوله فى الأجزاء التالية.

```
DATA WEIGHT;
INPUT HEIGHT WEIGHT @@;
CARDS;
150 50 175 77 170 65 160 63 150 61
155 62 152 55 180 75 154 55 165 68
PROC REG;
MODEL WEIGHT = HEIGHT;
RUN;
```

لاحظ أنه يمكن إضافة بعض الخيارات إلى الـ model منها على سبيل المثال:

$$\text{Model weight} = \text{height} / \text{XPX I};$$

وهذا يؤدي إلى الحصول على مصفوفة  $X'X$  ومعكوسها، أي  $(X'X)^{-1}$

نتيجة التحليل:

The REG Procedure  
Model: MODEL1  
Dependent Variable: weight

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	544.70676	544.70676	34.53	0.0004
Error	8	126.19324	15.77416		
Corrected Total	9	670.90000			

Root MSE	3.97167	R-Square	0.8119
Dependent Mean	63.10000	Adj R-Sq	0.7884
Coeff Var	6.29425		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-52.22693	19.66572	-2.66	0.0290
height	1	0.71587	0.12182	5.88	0.0004

الجزء المقطوع من

المحور الصادي (a)

معامل الانحدار b

لاحظ وجود بعض النتائج الأخرى والتي سوف يتم تناولها في الأجزاء التالية.

٣-١٠ استخدام المصفوفات في الانحدار

استخدام المصفوفات له كثير من الميزات من أهمها أن التعامل الرياضي فيها يكون أكثر اختصاراً وأوضح رؤية، أي أنه بمجرد كتابة المشكلة وحلها عن طريق المصفوفات فإنه يمكن تطبيق الحل على أي مشكلة بغض النظر عن عدد الحدود في معادلة الانحدار. والخطوات التالية تبين كيفية حل مثال ١٠-١ باستخدام المصفوفات.

سبق عرض المعادلة (٣-١٠) وهي  $Y = a + bX + e$ ، هذه المعادلة يمكن إعادة كتابتها باستخدام المصفوفات كالتالي:

$$Y = X\beta + \varepsilon \quad (14-10)$$

وباستخدام بيانات المثال 10-1 يمكن تكوين المصفوفات التالية:

$$Y = \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix}, X = \begin{bmatrix} 1 & 150 \\ 1 & 175 \\ \vdots & \vdots \\ 1 & 154 \\ 1 & 165 \end{bmatrix}, \beta = \begin{bmatrix} a \\ b \end{bmatrix}, \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

حيث:

Y : متجهة vector حجمه 10 x 1 يحتوى على قيم المتغير التابع

X : مصفوفة حجمها 10 x 2 تحتوى على العوامل المستقلة

$\beta$  : متجهة حجمه 2 x 1 تحتوى على المعالم المراد تقديرها وهى a ، b

$\varepsilon$  : متجهة حجمه 10 x 1 يمثل الخطأ

وبتطبيق المعادلة (14-10)  $Y = X\beta + \varepsilon$  يمكن الحصول على:

$$Y = \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} = \begin{bmatrix} a + 150b \\ a + 175b \\ \vdots \\ a + 154b \\ a + 165b \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

(10x1)      (10x1)      (10x1)

المعادلتان الاعتياديتان normal equations (10-1)، (10-2) يمكن التعبير عنهما بالمصفوفات كالتالى:

$$X'X\hat{\beta} = X'Y \quad (15-10)$$

وبالتعويض:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 150 & 175 & \dots & 154 & 165 \end{bmatrix} = \begin{bmatrix} 10 & 1611 \\ 1611 & 260595 \end{bmatrix}$$

(2x10)      (10x2)      (2x2)

والصورة العامة لهذا الجزء:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

(2 x n)                      (n x 2)                      (2 x 2)

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 150 & 175 & \dots & 154 & 165 \end{bmatrix} \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} = \begin{bmatrix} 631 \\ 102415 \end{bmatrix}$$

(2 x 10)                      (10 x 1)                      (2 x 1)

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

(2 x n)                      (n x 1)                      (2 x 1)

وبالتالى تكون الصورة العامة للمعادلة (١٠-١٥) كالتالى:

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad (١٦-١٠)$$

وبالتعويض

$$\begin{bmatrix} 10 & 1611 \\ 1611 & 2600595 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 631 \\ 102415 \end{bmatrix}$$

حل المعادلة (١٦-١٠) سوف يعطى تقديراً لكل من a، b بطريقة المربعات  
لصغرى. ولعمل ذلك لابد من الحصول على معكوس inverse المصفوفة التى فى  
الجانب الأيسر لهذه المعادلة أى إيجاد  $(X'X)^{-1}$  وبالتالى يكون:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (17-10)$$

بمعنى أن:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad (18-10)$$

ويمكن الحصول على هذا المعكوس كالتالى:

$$(X'X)^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix} \quad (19-10)$$

وبأخذ عامل مشترك للجانب الأيمن تصبح المعادلة (19-10) كالتالى:

$$(X'X)^{-1} = \frac{1}{n \sum (X_i - \bar{X})^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} \quad (20-10)$$

وبالتعويض

$$(X'X)^{-1} = \begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix}$$

وبتطبيق المعادلة (18-10) يكون تقدير كل من a, b كالتالى:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix} \begin{bmatrix} 631 \\ 102415 \end{bmatrix} = \begin{bmatrix} -52.22693 \\ 0.7158717 \end{bmatrix}$$

وهى نفس النتائج المتحصل عليها سابقاً مع بعض الاختلافات نتيجة للتقريب. ومن ذلك تتضح إمكانية وسهولة تطبيق طريق المصفوفات على أى أمثلة أخرى باستخدام نفس المعادلتين (18-10) و (20-10).

مثال ١٠-٣

حل مثال ١٠-١ باستخدام طريقة IML في برنامج SAS، وتستخدم هذه الطريقة عند الرغبة في استخدام المصفوفات في الحل

```
PROC IML;
X={1 150, 1 175, 1 170, 1 160, 1 150,
  1 155, 1 152, 1 180, 1 154, 1 165};
  XP = X`;
  XPX = XP*X;
  Y= {50, 77, 65, 63, 61, 62, 55, 75, 55, 68};
  YP = Y`;
  XPY = XP*Y;
  XPXINV=INV(XPX);
  SOL = XPXINV*XPY;
  PRINT XP YP XPX XPY XPXINV SOL;
  QUIT;
```

لاحظ:

طريقة كتابة المصفوفات حيث تكتب جميع عناصر المصفوفة داخل قوسين من النوع { }.  
تنتهي عناصر كل صف بفصله عادية.  
استخدام علامة ` للحصول على المقلوب transpose (X') واستخدام INV للحصول على المعكوس inverse.  
لا بد أن ينتهي البرنامج بكلمة quit.

نتائج التحليل:

XP	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
ROW1	1	1	1	1	1	1	1	1	1	1
ROW2	150	175	170	160	150	155	152	180	154	165
YP	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
ROW1	50	77	65	63	61	62	55	75	55	68
XPX	COL1	COL2	XPY	COL1						
ROW1	10	1611	ROW1	631						
ROW2	1611	260595	ROW2	102415						
XPXINV	COL1	COL2	SOL	COL1						
ROW1	24.517358	-0.151566	ROW1	-52.517358						
ROW2	-0.151566	0.0009408	ROW2	0.7158717						

## صندوق ١٠-١

- عندما يكون هناك متغيران أحدهما وليكن  $Y$  يعتمد على متغير آخر وليكن  $X$  فيعرف الأول بالتابع والثاني بالمستقل.
- ولكل من المتغيرين تباين ولكن بينهما أيضاً تباين مشترك يطلق عليه التباين  $covariance$ .
- يمكن تقنين العلاقة بين هذين المتغيرين باستخدام معادلة الخط المستقيم باعتبار  $X$  على المحور السيني و  $Y$  على المحور الصادي وحيث إن العلاقة بين المتغيرين عادة ما تكون غير تامة أى أن  $Y$  تحتوى على جزء غير راجع إلى  $X$  وهو ما يطلق عليه الخطأ وبالتالي فإن معادلة الخط المستقيم يضاف إليها مكون يعبر عن هذا الخطأ وتكون معادلة الانحدار هي:

$$Y = a + bX + e$$

حيث  $a$  ،  $b$  ، ثابتان،  $e$  الخطأ فى المشاهدة. كما هو الحال فى معادلة الخط المستقيم فإن  $a$  تمثل الجزء المقطوع من المحور الصادي بينما  $b$  هى ظل زاوية تقاطع الخط المستقيم مع المحور السيني.

- يقدر معامل انحدار  $Y$  على  $X$  من

$$b_{YX} = \frac{\sum XY - (\sum X \sum Y) / n}{\sum X^2 - (\sum X)^2 / n}$$

### ١٠-٤ مصادر الاختلاف في الانحدار الخطي

بالرجوع إلى البيانات التي في مثال ١٠-١ وجد أنه عندما كانت قيمة  $X$  للفرد الأول 150 سم فإن قيمة  $Y$  له 50 كج، بينما قيمة  $X$  للفرد الثاني 175 سم وقيمة  $Y$  له 77 كج. ومعنى ذلك أن الوزن أثقل للطول الأكبر، كما يلاحظ أنه بينما كانت قيمة  $X$  للفرد الخامس هي أيضاً 150 سم كان الوزن المقابل لها 61 كج. أى أن هناك أيضاً فرقا في الوزن حتى لو تساوت الأطوال. ويطلق على هذا الفرق غير المعروف مسبباته خطأ error أو من خط الانحدار from regression. وبالتالي فإن مصادر الاختلاف في الانحدار الخطي جزء منه راجع إلى اعتماد الوزن على الطول والجزء الآخر خاص بكل فرد. ويمكن بيان ذلك رياضياً كما يلي:

في الانحدار البسيط الذي يمثل العلاقة بين المتغيرين  $Y, X$ ، يمكن التعبير عن أى مشاهدة بالنموذج المبين في (١٠-٢) أى:

$$Y = \alpha + \beta X + \epsilon$$

حيث تمثل  $\alpha + \beta X$  متوسط المتغير  $Y$  والتي تقابل قيمة محددة للمتغير  $X$  ويرمز له  $\mu_{yx}$  وتمثل  $\epsilon$  مكون الخطأ error component، أى أن قيمة  $Y$  تمثل مجموع متغيرين أحدهما المتوسط والآخر الخطأ العشوائى. وفى النموذج التقديرى estimated model فإن:

$$Y = a + bX + e$$

$$Y = \bar{Y} + b(X - \bar{X}) + e$$

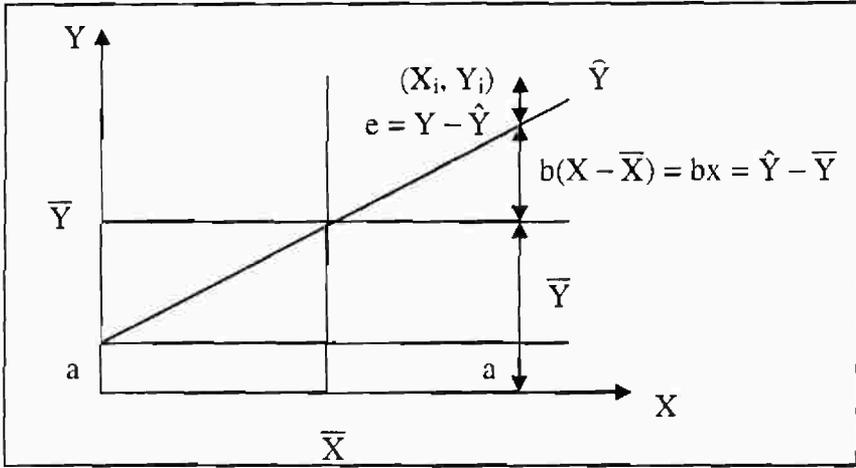
$$Y = \bar{Y} + (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$$

أى أن:

$$(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y}) \quad (١٠-٢١)$$

ويمكن توضيح ذلك بالشكل ١٠-٦.

ويتضح من الشكل ١٠-٦ إن انحراف القيمة عن متوسطها عبارة عن مجموع حدين. الأول يمثل انحراف القيمة المتوقعة عن المتوسط، والتي تمثل الجزء الراجع إلى اعتماد  $Y$  على  $X$  ويسمى due to regression أى  $(\hat{Y} - \bar{Y}) = b(X - \bar{X}) = bx$ . أما الحد الثانى فهو يمثل انحراف القيمة  $Y$  عن القيمة المتوقعة  $\hat{Y}$  والواقعة على خط الانحدار، وهذا الحد يمثل الخطأ العشوائى ويسمى from regression، أى  $e = d_{Y,X} = Y - \hat{Y} = Y - a - bX = y - bx$ .



شكل ١٠-٦ مصادر الاختلاف في Y

ويلاحظ أن مجموع انحرافات قيم  $\hat{Y}$  عن المتوسط يساوي صفر، أي  $\sum(\hat{Y} - \bar{Y}) = 0$ ، ومجموع انحرافات Y عن خط الاعتماد أيضاً يساوي صفر أي  $\sum(Y - \hat{Y}) = 0$ . وبالتالي من المعادلة (١٠-٢١) يمكن إثبات أن:

$$\sum(Y - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y - \hat{Y})^2 \quad (١٠-٢١)$$

أي أن:

$$TSS = RSS + ESS$$

حيث:

TSS عبارة عن مجموع المربعات الكلي (total sum of squares)

RSS عبارة عن مجموع المربعات الراجع لانحدار Y على X،  
sum of squares due to regression

ESS عبارة عن مجموع المربعات عن خط الانحدار،  
sum of squares from regression

هذه القيم يمكن الحصول عليها كالتالي:

$$RSS = \sum (\hat{Y} - \bar{Y})^2 = b^2 \sum x^2 = b \sum xy = \frac{(\sum xy)^2}{\sum x^2}$$

$$ESS = \sum y^2 - RSS = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2}$$

والجدول ١-١٠ يبين مصادر الاختلاف في Y لبيانات مثال ١-١٠

جدول ١-١٠ مصادر الاختلاف في قيم Y وتقسيم التباين إلى مصادره المختلفة

X	Y	$\hat{Y}$	$Y - \bar{Y}$	$\hat{Y} - \bar{Y}$	$e = Y - \hat{Y}$	$(\hat{Y} - \bar{Y})^2$	$(Y - \bar{Y})^2$
150	50	55.2	-13.1	-7.9	-5.2	62.41	27.04
175	77	73.1	13.9	10.0	3.9	100.00	15.21
170	65	69.4	1.9	6.3	-4.4	39.69	19.36
160	63	62.3	-0.1	-0.8	0.7	0.64	0.49
150	61	55.3	-2.1	-7.9	5.8	62.41	33.64
155	62	58.7	-1.1	-4.4	3.3	19.36	10.89
152	55	56.6	-8.1	-6.5	-1.6	42.45	2.56
180	75	76.6	11.9	13.5	-1.6	182.25	2.56
154	55	58.0	-8.1	-5.1	-3.0	26.01	9.00
165	68	65.9	4.9	2.8	2.1	7.84	4.41
<b>1611</b>	<b>631</b>	<b>631</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>	<b>542.86</b>	<b>125.16</b>

مثال ١-١٠

قسم مجموع المربعات الكلي في Y إلى مكوناته مستخدماً بيانات المثال ١-١٠.

مجموع المربعات الكلي:

$$TSS = \sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 40487 - \frac{(631)^2}{10} = 670.9$$

مجموع المربعات الزاجع إلى الانحدار:

$$RSS = \sum (\hat{Y} - \bar{Y})^2 = b^2 \sum x^2 = b \sum xy$$

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n} = 260595 - \frac{(1611)^2}{10} = 1062.9$$

$$RSS = (0.716)^2 (1062.9) = 544.9$$

مجموع مربعات الخطأ (من خط الانحدار):

$$ESS = \sum (Y - \hat{Y})^2 = TSS - RSS = 670.9 - 544.7 = 126.2$$

وقيم كل من RSS و ESS هي نفس القيم المتحصل عليها في جدول ١٠-١ الاختلافات بينها راجع إلى التقريب.

يمكن الحصول على نفس النتائج باستخدام المصفوفات كالتالي:

$$TSS = Y'Y - n\bar{Y}^2 \quad (٢٢-١٠)$$

$$RSS = b'X'Y - n\bar{Y}^2 \quad (٢٣-١٠)$$

$$ESS = TSS - RSS = Y'Y - b'X'Y \quad (٢٤-١٠)$$

وباستخدام بيانات مثال ١٠-١  
مجموع المربعات الكلي:

$$TSS = Y'Y - n\bar{Y}^2 = \begin{bmatrix} 50 & 77 & \dots & 55 & 68 \end{bmatrix} \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} - [(10)(63.1)^2] = 670.9$$

و هذا المجموع له 9 درجات حرية أي (n-1).

مجموع المربعات الزاجع إلى الانحدار:

$$RSS = [-52.22693 \quad 0.71587] \begin{bmatrix} 631 \\ 102415 \end{bmatrix} - [(10)(63.1)^2] = 544.534$$

وهذا المجموع له درجة حرية واحدة.

مجموع مربعات الخطأ (من خط الانحدار):

$$ESS = TSS - RSS = 670.9 - 544.534 = 126.357$$

و هذا المجموع له 8 درجات حرية أى (n-2).

ويعتبر متوسط مجموع مربعات الخطأ،  $(ESS)/(n-2)$ ، تقدير غير متحيز للتباين  $\sigma^2$  ويرمز له بالرمز  $S_{y.x}^2$ .

لاحظ أن نفس النتائج تم الحصول عليها عند استخدام اختيار PROC REG فى برنامج SAS فى فصل ١٠-٤ مع ملاحظة وجود خطأ التقريب.

### ١٠-٥ القيم المعدلة لأثر انحدار Y على X (Adjusted Y)

$$Y = \bar{Y} + bx + e \quad \text{كما ذكر سابقاً}$$

وعلى ذلك فقيمة Y المعدلة، ويرمز لها  $Y_A$ ، أى بعد إزالة الجزء الراجع للانحدار  $bx$  من قيمة Y هي:

$$\text{Adjusted } Y = Y_A = \bar{Y} + e = Y - bx \quad (10-25)$$

وتفيد القيم المعدلة فى المقارنة بين الأفراد أو بين المتوسطات المعدلة، أى بعد إزالة أثر المتغير المستقل. فمثلاً إذا كان معامل انحدار الوزن على العمر هو 0.8 كج لكل يوم وكان هناك حيوانان الأول عمره 140 يوم ووزنه 200 كج والثانى عمره 168 يوم ووزنه 220 كج، فأيهما أثقل وزناً؟ وللإجابة على ذلك يجب أولاً إزالة أثر الاختلاف فى العمر للحيوانين. فإذا كان متوسط العمر هو 105 يوماً ومتوسط الوزن 190 كج فإن:

$$Y_{A1} = Y_1 - bx_1 = 200 - 0.8(140 - 105) = 172 \text{ kg}$$

$$Y_{A2} = Y_2 - bx_2 = 220 - 0.8(168 - 105) = 169.6 \text{ kg}$$

ومعنى ذلك أنه لو أن الحيوانات كانا عند نفس العمر (105 يوم مثلاً) لكان التوقع أن يزن الحيوانان 172 و 169.6 كج، على التوالي. وعلى ذلك فالحيوان الأول أكثر وزناً من الحيوان الثانى بعد إزالة العمر وبالتالي يكون هو الأثقل وزناً.

وتجدر الإشارة هنا أنه بحساب معامل الانحدار فإنه يمكن استخراج كل العلاقة بين المتغيرين  $Y, X$  وما يترك فى  $Y$  بعد هذه العلاقة وهو  $e$  فإن العلاقة بينه وبين كل من المتغيرات الأخرى تساوى صفرأى أن:  $Cov(e, X) = Cov(e, \hat{Y}) = 0$ .

ويمكن التأكد من ذلك بحساب كل من  $\sum e\hat{Y}$  و  $\sum eX$  من جدول ١٠-١ والتى ستكون مساوية للصفر أيضاً.

### ١٠-٦ الانحرافات المعيارية وحدود الثقة للتقديرات المختلفة

#### Standard deviation and confidence limits of estimates

##### ١٠-٦-١ الانحرافات المعيارية

سبق اعتبار أن  $Y$  تقدير غير متحيز لـ  $\mu_{y.x}$  وكذلك  $\hat{Y}, b, a$  تقديرات غير متحيزة لكل من  $\mu_{Y.X}, \beta, \alpha$  على التوالي.

تباين  $\bar{Y}$  هو:  $\sigma_{\bar{Y}}^2 = \frac{\sigma^2}{n}$  وذلك من المعادلة (٣-١٣) بالبَاب الثالث.

تباين  $b$ :

$$\begin{aligned}\sigma_b^2 &= v\left(\frac{\sum xy}{\sum x^2}\right) \\ &= \frac{1}{(\sum x^2)^2} [V(x_1y_1 + x_2y_2 + \dots + x_ny_n)] \\ &= \frac{1}{(\sum x^2)^2} [x_1^2V(y_1) + x_2^2V(y_2) + \dots + x_n^2V(y_n)]\end{aligned}$$

وحيث إن:  $V(y_1) = V(y_2) = \dots = V(y_n) = \sigma^2$ ، فإن:

$$\begin{aligned}\sigma_b^2 &= \frac{1}{(\sum x^2)^2} [x_1^2 \sigma^2 + x_2^2 \sigma^2 + \dots + x_n^2 \sigma^2] \\ &= \frac{1}{(\sum x^2)^2} \sigma^2 \sum x^2 = \frac{\sigma^2}{\sum x^2} \quad (26-10)\end{aligned}$$

تباين الجزء المقطوع من محور الصادات  $\alpha$ :

يمكن بنفس المفهوم إثبات أن:

$$\sigma_a^2 = \sigma^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum x^2} \right) \quad (27-10)$$

وتباين متوسط العشرة الذي يقابل قيمة محددة لـ  $X$  أى  $\mu_{y.x}$  هو:

$$\sigma_{\mu_{y.x}}^2 = \sigma^2 \left( \frac{1}{n} + \frac{x_i^2}{\sum x_i^2} \right) \quad (28-10)$$

وتباين القيمة المتوقعة التى تقابل قيمة محددة لـ  $X$  هى:

$$\sigma_{\hat{Y}}^2 = \sigma^2 \left( 1 + \frac{1}{n} + \frac{x_i^2}{\sum x_i^2} \right) \quad (29-10)$$

حيث القيمة المتوقعة  $\hat{Y}$  هى تقدير لنقطة جديدة لـ  $Y$  التى تقابل قيمة محددة لـ  $X$ . وبأخذ الجذر التربيعى للمعادلات السابقة يمكن الحصول على الانحراف المعيارى للتقديرات المختلفة.

فمثلاً الانحراف المعيارى لمعامل الانحدار  $SE_b$  أو  $S_b$  هو

$$S_b = \frac{S_{y.x}}{\sqrt{\sum x^2}} \quad (30-10)$$

حيث  $S_{y.x}^2$  هى تقدير لـ  $\sigma^2$  علماً بأن  $S_{y.x}^2 = ESS/(n-2)$  كما ذكر من قبل. و  $ESS$  هى مجموع مربعات الخطأ و  $n$  عدد أزواج المتغيرات (المشاهدات).

وتعبر  $S_{y,x}$  عن الخطأ القياسي للتقدير أو الانحراف المعياري لـ  $Y$  باعتبار أن  $X$  ثابتة.

ومما سبق يمكن كتابة مصفوفة التباين والتغاير variance-covariance matrix للتقديرات المختلفة كالتالي:

$$V(\hat{\beta}) = \begin{bmatrix} V(a) & \text{Cov}(a,b) \\ \text{Cov}(a,b) & V(b) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2} & -\frac{\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} \\ -\frac{\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} & \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{bmatrix} \quad (31-10)$$

وبأخذ عامل مشترك  $\sigma^2$  فإن المعادلة (31-10) تصبح:

$$V(\hat{\beta}) = (X'X)^{-1} \sigma^2 \quad (32-10)$$

وللحصول على تقدير لمتوسط العشيرة (باستخدام جبر المصفوفات) عند قيمة معينة ولتكن  $X_0$  ضع  $X'_0 = [1 \ X_0]$ ، وبالتالي يمكن تقدير متوسط العشيرة عند قيمة  $X_0$  كما يلي:

$$\hat{\mu}_{y,x} = [1 \ X_0] \begin{bmatrix} a \\ b \end{bmatrix} = X'_0 b = b' X_0$$

وحيث إن تباين  $\hat{\mu}_{y,x}$  عبارة عن

$$V(\hat{\mu}_{y,x}) = V(a) + 2X_0 \text{cov}(a,b) + X_0^2 V(b)$$

فإن

$$V(\hat{\mu}_{y,x}) = [1 \ X_0] \begin{bmatrix} V(a) & \text{cov}(a,b) \\ \text{cov}(a,b) & V(b) \end{bmatrix} \begin{bmatrix} 1 \\ X_0 \end{bmatrix}$$

$$= X'_0 (X'X)^{-1} \sigma^2 X_0$$

$$= X'_0 (X'X)^{-1} X_0 \sigma^2$$

وهذه مطابقة تماماً للمعادلة (١٠-٢٩).

عندما تكون  $X_0 = 156$  (في مثال ١٠-١) فإن

$$\hat{\mu}_{y.x} = [1 \quad 156] \begin{bmatrix} -52.22693 \\ 0.71587 \end{bmatrix} = 59.45$$

$$V(\hat{\mu}_{y.x}) = X'_0 (X'X)^{-1} X_0 \sigma_{y.x}^2$$

وحيث إن  $\sigma_{y.x}^2 = 15.77416$  تعتبر تقديراً غير متحيز لـ  $\sigma_{y.x}^2$  فإن

$$V(\hat{\mu}_{y.x}) = [1 \quad 156] \begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix} \begin{bmatrix} 1 \\ 156 \end{bmatrix} (15.77416) \\ = 1.956841$$

وبالتالى

$$S_{\hat{\mu}_{y.x}} = \sqrt{1.956841} = 1.398871$$

تقدير تباين القيمة المتوقعة  $\hat{Y}$  عند قيمة معينة للمتغير  $X$ :

عندما يكون الغرض هو التنبؤ prediction بقيمة المتغير التابع  $Y$  عند قيمة معينة للمتغير  $X$  ولتكن  $X_0$  (وليس التقدير estimation كما هو الحال عند تقدير متوسط العشيرة عند نفس النقطة  $X$ ) فإن القيمة المتوقعة تكون هي نفسها مساوية لمتوسط العشيرة عند تلك النقطة ولكن بتباين أكبر هو:

$$V(\hat{Y}_{X=X_0}) = [1 + X'_0 (X'X)^{-1} X_0] \sigma_{y.x}^2$$

وتستخدم  $S_{\hat{Y}.X}^2$  فى حالة عدم معرفة  $\sigma_{y.x}^2$

١٠-٦-٢ حدود الثقة لكل من  $a$ ،  $b$ ،  $\hat{\mu}_{y.x}$ ،  $\hat{Y}$

بنفس المفهوم الذى استخدم لإيجاد حدود الثقة لكل من المتوسط والتباين فى الباب السادس فصل ٦-٢ يمكن إيجاد حدود الثقة كما يلى:

حدا الثقة لـ  $a$  هما:

$$(\bar{Y} - b\bar{X}) \pm t_{\alpha} S_{y.x} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}} \quad (٣٣-١٠)$$

حدا الثقة لـ  $b$  هما:

$$b \pm t_{\alpha} \frac{S_{y.x}}{\sqrt{\sum x^2}} \quad (٣٤-١٠)$$

حدا الثقة لـ  $\hat{\mu}_{Y.X}$ : أى حدا الثقة لمتوسط العشرة الذى يقابل قيمة معينة لـ  $X$  (حيث  $\mu_{Y.X} = \bar{Y} + bx$ ) هما:

$$\hat{Y} \pm t_{\alpha} S_{y.x} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}} \quad (٣٥-١٠)$$

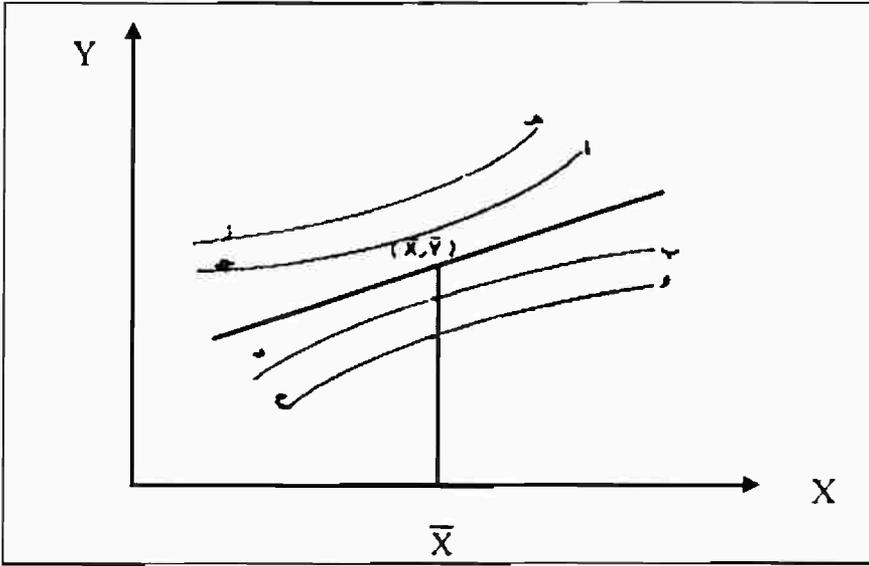
حيث  $\hat{Y} = \bar{Y} + bx$ .

وهذه تمثل حدى الثقة لخط الانحدار. وهذه القيم تشكل خطين وتسمى المساحة المحصورة بينهما بحزام الثقة confidence belt عند مستوى احتمالى  $\alpha$  (مستوى المعنوية).

أما حدا الثقة للقيمة المتوقعة  $\hat{Y}$  فهما:

$$(\bar{Y} - bx) \pm t_{\alpha} S_{y.x} \sqrt{1 + \frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}} \quad (٣٦-١٠)$$

وشكل ٧-١٠ يبين حزام الثقة حول  $\hat{\mu}_{Y.X}$  وحزام الثقة حول  $\hat{Y}$ .



شكل ١٠-٧ حزام الثقة حول  $\hat{\mu}_{Y.X}$  وهو أ ب ج د وحزام الثقة حول  $\hat{Y}$  وهو هـ و ز ح.

وفى المعادلات السابقة  $t_{\alpha}$  هي قيمة t الجدولية بدرجات حرية  $n-2$  ومستوى معنوية  $\alpha$ .

مثال ١٠-٥

احسب حدى الثقة لمعامل اعتماد  $Y$  على  $X$  باستخدام بيانات المثال ١٠-١ علماً بأن مستوى المعنوية 5%.

حدا الثقة هما:

$$b \pm t_{\alpha} \frac{S_{y..x}}{\sqrt{\sum x^2}}$$

حيث

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n}, \quad S_{Y.X} = \sqrt{\frac{\sum y^2 - [(\sum xy)^2 / (\sum x^2)]}{(n-2)}}$$

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

وبالتعويض فإن:

$$\sum x^2 = 260595 - \frac{(1611)^2}{10} = 1062.9$$

$$\sum y^2 = 40487 - \frac{(631)^2}{10} = 670.9$$

$$\sum xy = 102415 - \frac{(1611)(631)}{10} = 760.9$$

$$S_{Y.X} = \sqrt{\frac{670.9 - [(760.9)^2 / (1062.9)]}{(10 - 2)}} = \sqrt{\frac{670.9 - 544.7}{8}} = 3.97$$

$$S_b = \frac{S_{y.x}}{\sqrt{\sum x^2}} = \frac{3.97}{\sqrt{1062.9}} = 0.122$$

وبالتالى فإن حدى الثقة لمعامل الاعتماد (الانحدار) هما:

$$0.72 \pm (2.306)(0.122) = 0.72 \pm 0.28$$

أى أن الحد الأدنى للثقة هو 0.44 والحد الأعلى للثقة هو 1.00

#### مثال ١٠-٦

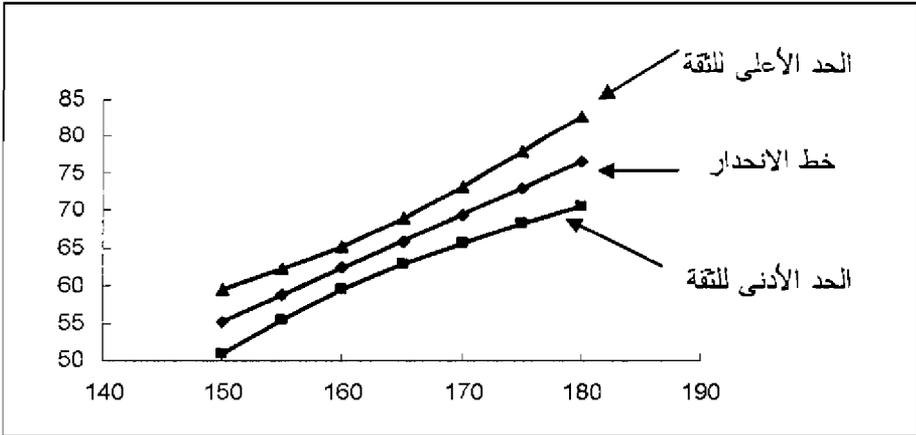
احسب حدود الثقة لخط الانحدار بمستوى معنوية 5% باستخدام بيانات المثال ١٠-١، وماذا تعنى ثم بين ذلك بيانياً.

لحساب حدود الثقة لخط الانحدار يلزم أولاً حساب حدود الثقة لعدد من  $\mu_{y.x}$  ثم تمثل هذه النقط بيانياً ويتم التوصيل فيما بينها فينتج خطين أحدهم يمثل الحد الأعلى لخط الانحدار والآخر يمثل الحد الأدنى لخط الانحدار.

حيث إن قيمة t الجدولية بمستوى معنوية 5% ودرجات حرية 8 هي 2.306 وقيمة  $S_{y.x} = 3.97$  وباستخدام المعادلة (١٠-٣٥) فإنه يمكن الحصول على النقاط التالية:

حدود الثقة		$\hat{Y} \pm t_{\alpha} S_{y.x} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}}$	$\hat{Y}$	X
الحد الأدنى	الحد الأعلى			
59.40	50.90	$55.15 \pm 4.25$	55.15	150
62.09	55.37	$58.73 \pm 3.36$	58.73	155
65.22	59.40	$62.31 \pm 2.91$	62.31	160
68.99	62.79	$65.89 \pm 3.10$	65.89	165
73.29	65.65	$69.47 \pm 3.82$	69.47	170
77.91	68.19	$73.05 \pm 4.86$	73.05	175
82.68	70.58	$76.63 \pm 6.05$	76.63	180

ويمكن تمثيل حدود الثقة لخط الانحدار لهذه البيانات بيانياً بالشكل التالي:



وحزام الثقة هذا يعني أن احتمال أن يحوى الحزام خط الاعتماد الحقيقى  $(\alpha + \beta X)$  هو 0.95.

### ٧-١٠ اختبار معنوية معامل الانحدار

يمكن اختبار الفرض القائل بأن معامل انحدار العشيرة  $\beta$  يساوى الصفر، أى لا توجد علاقة بين المتغيرين  $Y, X$  باستخدام اختبار  $t$  كما يلي:

فرض العدم:  $H_0: \beta = 0$ . وتحت صحة فرض العدم فإن  $\beta = 0$  وبالتالي:

$$t = \frac{b - \beta}{S_b} = \frac{b\sqrt{\sum x^2}}{S_{y.x}}$$

وتقارن هذه القيمة بقيمة  $t$  الجدولية بدرجات حرية  $n-2$  ومستوى معنوية وليكن

$\alpha$ .

أما إذا كان فرض العدم هو  $H_0: \beta = c$  حيث  $c$  هى قيمة معينة لمعامل انحدار العشيرة، فيكون الفرض البديل أن  $H_1: \beta \neq c$  وتكون

$$t = \frac{b - c}{S_b}$$

#### مثال ٧-١٠

اختبر معنوية معامل انحدار الوزن على الطول لبيانات مثال ١-١٠ باستخدام اختبار  $t$ .

فرض العدم:  $H_0: \beta = 0$  ومن مثال ١٠-٥  $S_b = S_{y.x} / \sqrt{\sum x^2} = 0.12177$  إذا

$$t = \frac{b - \beta}{S_b} = \frac{0.71587}{0.12177} = 5.88$$

ومن جدول  $t$  تظهر القيم  $t_{(8,0.01)} = 3.350$  ،  $t_{(8,0.05)} = 2.306$

وبالتالى يرفض فرض العدم وبالتالي فإن معامل الانحدار يختلف عن الصفر معنوياً، أى أن العلاقة بين  $Y, X$  معنوية بدرجة ثقة 99%.

يمكن اختبار معنوية معامل الانحدار باستخدام تحليل التباين عن طريق تكوين جدول تحليل التباين ANOVA table. فكما سبق وذكر أن مجموع المربعات الكلى فى  $Y$  يقسم إلى مكونين أحدهما راجع إلى الانحدار due to regression والآخر عن

خط الانحدار deviations from regression. وبالتالي فإن جدول ١٠-٢ لتحليل لتباين يمكن تكوينه كالتالي:

جدول ١٠-٢ تحليل التباين باستخدام تحليل الانحدار

SOV مصدر التباين	df	SS مجموع المربعات	MS متوسط المربعات	F
Due to regression راجع للانحدار	1	RSS = b'X'Y		RSS/S <sub>y.x</sub> <sup>2</sup>
From regression من الانحدار	n-2	ESS = ∑ y <sub>y.x</sub> <sup>2</sup> = Y'Y - b'X'Y	S <sub>y.x</sub> <sup>2</sup>	
<b>C. Total</b> الكل المصحح	<b>n-1</b>	<b>∑ y<sup>2</sup> = Y'Y - nȲ<sup>2</sup></b>		

وتقارن قيمة F المحسوبة بقيمة F الجدولية بدرجات حرية 1، n-2 ومستوى معنوية وليكن α.

وحيث إن t هي  $(b\sqrt{\sum x^2})/(S_{y.x})$  فإن قيمة t<sup>2</sup> هي  $(b^2 \sum x^2)/(S_{y.x}^2)$  وهي نفس قيمة F أي أن  $F = t^2$  في حالة الانحدار البسيط.

مثال ١٠-٨

كون جدول تحليل التباين واختبر معنوية معامل الانحدار باستخدام بيانات المثال ١٠-٤.

لاحظ أن جدول تحليل التباين المطلوب قد تم الحصول عليه باستخدام برنامج SAS عند حل مثال ١٠-٢، والجدول كان كالتالي:

SOV	df	SS	MS	F
Due to regression	1	544.7	5.447	34.52**
From regression	8	126.2	15.775	
<b>C. Total</b>	<b>9</b>	<b>670.9</b>		

قيم  $F$  الجدولية  $F_{(1,8,0.01)} = 11.26$  ،  $F_{(1,8,0.05)} = 5.32$

قيمة  $F$  المحسوبة أكبر من قيمة  $F$  الجدولية عند مستوى معنوية 5% ، 1% .

إذا العلاقة معنوية بدرجة ثقة 99% بين المتغيرين  $X$  و  $Y$ .

وحيث إن قيمة  $t$  المستخرجة في المثال ١٠-٧ هي 5.88 ومربعها هو 34.57 ، وهي نفس قيمة  $F$  (مع مراعاة الاختلاف البسيط بين مربع قيمة  $t$  وقيمة  $F$  نتيجة للتقريب) وكلتا الطريقتين متطابقتان تماماً.

### صندوق ١٠-٢

- معامل الانحدار هو متوسط التغير في المتغير التابع  $Y$  لكل وحدة تغير في المتغير المستقل  $X$ . وهو قيمة مميزة (أى كج، سم، أردب، وحدة سماء ... الخ) وتتراوح قيمته بين  $+\infty$  و  $-\infty$  .
- بينما تتيح دراسة الانحدار بحث العلاقة بين متغيرين يفترض أن أحدهما مستقل وثابت (أى بدون توزيع احتمالى) والآخر تابع وله توزيع احتمالى (طبيعى مثلاً) فإنه أيضاً يمكن من:
  - تقسيم التباين الكلى  $\sum y^2$  فى  $Y$  إلى جزء راجع إلى التغير فى  $X$  وهو  $b\sum xy$  وجزء آخر راجع إلى عوامل غير  $X$  وهو  $\sum y^2 - b\sum xy$  .
  - استنباط معادلة الانحدار  $\hat{Y} = \alpha + bX$  وبذلك يمكن التنبؤ بأى قيمة للمتغير التابع عند أى قيمة للمتغير المستقل فى حدود مدى الدراسة.
  - اختبار المعنوية للعلاقة  $by_{yx}$  وكذلك استنباط حدود الثقة عند مستوى احتمالى معين.

### ٨-١٠ المقارنة بين معاملي الانحدار

يستخدم اختبار  $t$  لمقارنة معاملي الانحدار في عينتين لمعرفة ما إذا كان هذان المعاملان هما تقدير لنفس معامل انحدار العشيرة  $\beta$  أم لا بنفس المفهوم المستخدم في مقارنة المجموع group comparison، كما هو موضح في الباب الثامن ويكون ذلك كما يلي:

$$H_0: \beta_1 - \beta_2 = 0 \text{ بمعنى } H_0: \beta_1 = \beta_2$$

وتحت صحة فرض العدم فإن الكمية  $\beta_1 - \beta_2 = 0$  تتبع توزيع  $t$  بدرجات حرية  $[(n_1 - 2) + (n_2 - 2)]$  وبالتالي فإن

$$t = \frac{(b_1 - b_2) - (\beta_1 - \beta_2)}{\sqrt{S_{b_1 - b_2}^2}} = \frac{b_1 - b_2}{S_{b_1 - b_2}} \quad (٣٧-١٠)$$

حيث

$$S_{b_1 - b_2}^2 = S_p^2 / \sum x_1^2 + S_p^2 / \sum x_2^2$$

$\sum x_1^2$  مجموع مربع الانحرافات عن المتوسط بالنسبة للمتغير  $X_1$  في العينة الأولى،

$\sum x_2^2$  مجموع مربع الانحرافات عن المتوسط بالنسبة للمتغير  $X_2$  في العينة الثانية،

$S_p^2$  تمثل التباين المشترك pooled variance والذي يعبر عنه كما يلي

$$S_p^2 = \frac{(\sum y_1^2 - b_1^2 \sum x_1^2) + (\sum y_2^2 - b_2^2 \sum x_2^2)}{(n_1 - 2) + (n_2 - 2)}$$

أي مجموع المربعات عن خط الاعتماد في العينة الأولى + مجموع المربعات عن خط الاعتماد في العينة الثانية مقسوماً على مجموع درجات الحرية للخطأ في العينتين.

كما يستخدم أيضاً تحليل التباين لإجراء هذا الاختبار ويمكن الرجوع في ذلك إلى .Steel and Torrie (1980)

قياسان  $Y, X$  أخذوا على أفراد عينتين وكانت البيانات التالية:  
العينة الأولى:

$$n_1 = 10, b_1 = 1.26 \text{ وحدة من } Y \text{ لكل وحدة من } X$$

$$\sum x_1^2 = 17.5, 2.3 = \text{مجموع المربعات عن خط الاعتماد}$$

العينة الثانية:

$$n_2 = 10, b_2 = 1.35 \text{ وحدة من } Y \text{ لكل وحدة من } X$$

$$\sum x_2^2 = 19, 4 = \text{مجموع المربعات عن خط الاعتماد}$$

اختبر الفرض القائل بأن معاملي الانحدار هما نفس التقدير لمعامل انحدار العشييرة.

الحل:

$$S_{y \cdot x}^2 = \frac{2.3 + 4}{10 + 10 - 4} = 0.39, H_0: \beta_1 - \beta_2 = 0$$

$$S_{b_1 - b_2} = \sqrt{0.04} = 0.2 \text{ وبالتالي } S_{b_1 - b_2}^2 = 0.39 \left( \frac{1}{17.75} + \frac{1}{19} \right) = 0.04$$

$$\therefore t = \frac{|1.35 - 1.26|}{0.2} = 0.045$$

قيمة  $t$  الجدولية  $t_{(16, 0.05)} = 2.12$  وهي أكبر من  $t$  المحسوبة (0.45) وبالتالي لا يرفض فرض العدم وعلى هذا فإن  $b_2, b_1$  هما تقديران لنفس معامل انحدار العشييرة  $\beta$  وتحسب  $\beta$  العامة للمثال لأنه ليس هناك سبب لوجود معاملي انحدار ولكن معامل انحدار واحد.

### ٩-١٠ التوزيع ذو المتغيرين

في كثير من الحالات يكون لكل من المتغيرين  $Y, X$  توزيع احتمالي بمعنى أن أزواج المتغيرين كل منهما مسحوب عشوائياً. فمثلاً كمية اللبن التي تعطىها البقرة في الموسم والعمر عند أول ولادة لكل بقرة إذا سحبت 20 بقرة عشوائياً وتم تسجيل العمر

وكمية اللبن لكل بقرة. أو عند دراسة العلاقة بين عدد لقطع دودة ورق القطن وكمية المحصول في عدد من القطع اختيرت عشوائياً. في مثل هذه الحالات تكون العينات مسحوبة من عشيرة ذات متغيرين، وفي هذه الحالة لا توضع قيود أو شروط على أى من المتغيرين، كأن يكون أحدهما ثابتاً *fixed* مثلاً. ويمكن حساب معاملين للانحدار أحدهما معامل انحدار المتغير الأول مثلاً على المتغير الثاني، والآخر معامل انحدار المتغير الثاني على المتغير الأول. ولو أنه في كثير من الأحيان يكون الاهتمام مقصوراً على حساب معامل انحدار واحد له معنى معين. وفي العينات المسحوبة من عشيرة ذات متغيرين فإنه يمكن تلخيص بيانات العينة المسحوبة وذلك بحساب متوسط وتباين كل من المتغيرين على حدة وكذلك حساب التغيرات بينهما ويمكن أن يعبر عن التغيرات بمعامل التحديد *coefficient of determination* أو الجذر التربيعي له والذي يعرف بمعامل الارتباط البسيط كما سيتضح فيما بعد.

مثال ١٠-١٠

في عينة من 8 أزواج كانت البيانات التالية:

رقم المشاهدة	١	٢	٣	٤	٥	٦	٧	٨
$Y_1$	6	4	0	10	14	2	12	8
$Y_2$	9	3	2	10	15	8	11	12

احسب معامل انحدار  $Y_1$  على  $Y_2$  وكذلك معامل انحدار  $Y_2$  على  $Y_1$  مع تقسيم التباين في كل من المتغيرين إلى مكوناته.

الحل:

$$\sum y_1^2 = 168 \quad \sum Y_1^2 = 560 \quad \bar{Y}_1 = 7 \quad \sum Y_1 = 56$$

$$\sum y_2^2 = 135.5 \quad \sum Y_2^2 = 748 \quad \bar{Y}_2 = 8.75 \quad \sum Y_2 = 70$$

$$\sum y_1 y_2 = 130 \quad \sum Y_1 Y_2 = 620$$

معامل انحدار  $Y_1$  على  $Y_2$  0.959 وحدة من  $Y_1$  لكل وحدة من  $Y_2$  ناتج من

$$b_{Y_1 Y_2} = \frac{\sum y_1 y_2}{\sum y_2^2} = \frac{130}{135.5} = 0.959$$

معامل انحدار  $Y_2$  على  $Y_1$  0.774 وحدة من  $Y_2$  لكل وحدة من  $Y_1$  ناتج من

$$b_{Y_2 Y_1} = \frac{\sum y_1 y_2}{\sum y_1^2} = \frac{130}{168} = 0.774$$

$$4.9 \quad \text{تباين المتغير الأول:} \quad S_{Y_1}^2 = \frac{168}{7} = 24 \quad \text{والانحراف المعياري 4.9}$$

$$4.4 \quad \text{تباين المتغير الثاني:} \quad S_{Y_2}^2 = \frac{135.5}{7} = 19.36 \quad \text{والانحراف المعياري 4.4}$$

$$\text{التغاير بين المتغيرين:} \quad \text{cov}(Y_1, Y_2) = \frac{130}{7} = 18.57$$

إذا اعتبر أن  $Y_2$  هو المتغير التابع فإن مجموع المربعات الراجع لانحدار  $Y_2$  على  $Y_1$  (RSS):

$$\text{RSS} = \frac{(\sum y_1 y_2)^2}{\sum y_1^2} = \frac{(130)^2}{168} = 100.6$$

$$\text{TSS} = 135.5 \quad \text{مجموع المربعات الكلي للمتغير } Y_2:$$

مجموع المربعات عن خط انحدار  $Y_2$  على  $Y_1$ :

$$\text{ESS} = 135.5 - 100.6 = 34.9$$

وعلى ذلك فإن:

$$\frac{\text{RSS}}{\text{TSS}} = \frac{(\sum y_1 y_2)^2 / \sum y_1^2}{\sum y_2^2} = \frac{(130)^2 / 168}{135.5} = 0.74$$

أما إذا اعتبر أن  $Y_1$  هو المتغير التابع فإن مجموع المربعات الراجع لانحدار  $Y_2$  على  $Y_1$ :

$$RSS = \frac{(\sum Y_1 Y_2)^2}{\sum Y_2^2} = \frac{(130)^2}{135.5} = 124.7$$

مجموع المربعات عن خط انحدار  $Y_1$  على  $Y_2$ :

$$ESS = 168 - 124.7 = 43.3$$

وعلى ذلك فإن:

$$\frac{RSS}{TSS} = \frac{(\sum Y_1 Y_2)^2 / \sum Y_1^2}{\sum Y_1^2} = \frac{(130)^2 / 135.5}{168} = 0.74$$

أما حاصل ضرب معاملي الانحدار

$$(b_{Y_1 Y_2})(b_{Y_2 Y_1}) = (0.959)(0.774) = 0.74$$

وبالتالي يوجد معادلتان للانحدار يمثلان خطي الانحدار يمر كل منهما بنقطة تقاطع متوسطي المتغيرين ولا يشترط أن ينطبق الخطين كل منهما على الآخر، إذ لا يتحقق ذلك إلا إذا كان مجموع مربع الانحرافات عن خط الانحدار مساوياً للصفر أي أن جميع النقط واقعة على خط الانحدار بدون أى أخطاء كما يلاحظ أن معامل انحدار  $Y_2$  على  $Y_1$  ليس هو مقلوب معامل انحدار  $Y_1$  على  $Y_2$ .

معادلة خط انحدار  $Y_2$  على  $Y_1$  هي:

$$\hat{Y}_2 = a_1 + b_{Y_2 Y_1} Y_1$$

حيث:

$$a_1 = 8.75 - (0.774)(7) = 3.33 \quad \text{وبالتعويض} \quad a_1 = \bar{Y}_2 - b_{Y_2 Y_1} \bar{Y}_1$$

$$\hat{Y}_2 = 3.33 + 0.774 Y_1 \quad \text{وعلى ذلك فإن معادلة خط الانحدار}$$

أما معادلة خط انحدار  $Y_1$  على  $Y_2$  فهي:

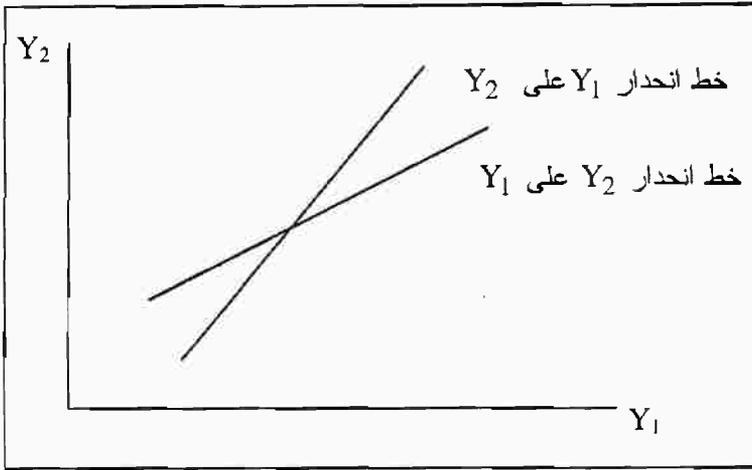
$$\hat{Y}_1 = a_2 + b_{Y_1 Y_2} Y_2$$

حيث:

$$a_2 = 7 - (0.959)(8.75) = -1.39 \quad \text{وبالتعويض} \quad a_2 = \bar{Y}_1 - b_{Y_1 Y_2} \bar{Y}_2$$

$$\hat{Y}_1 = -1.39 + 0.959 Y_2 \quad \text{وعلى ذلك فإن معادلة خط الانحدار}$$

وبين الشكل التالي خطى الانحدار .



ويلاحظ أيضاً أن النسبة بين مجموع المربعات الراجعة للانحدار إلى مجموع المربعات الكلي عندما كان  $Y_2$  هو المتغير التابع هي نفسها عندما كان  $Y_1$  هو المتغير التابع، أى أن هذه النسبة والتي تساوى 0.74 هي نفسها بغض النظر عن كون أى المتغيرين هو التابع وهذه النسبة تساوى حاصل ضرب معاملى الانحدار أى أن:

$$(b_{Y_1 Y_2})(b_{Y_2 Y_1}) = \frac{(\sum y_1 y_2)^2 / \sum y_1^2}{\sum y_2^2} = \frac{(\sum y_1 y_2)^2 / \sum y_2^2}{\sum y_1^2}$$

ويرمز لهذه الكمية بالرمز  $r^2$  ويطلق عليها معامل التحديد coefficient of determination أى أن:

$$r^2 = (b_{Y_1 Y_2})(b_{Y_2 Y_1}) = \frac{(\sum y_1 y_2)^2}{(\sum y_1^2)(\sum y_2^2)} \quad (38-10)$$

والجذر التربيعى لمعامل التحديد ( $r$ ) يطلق عليه معامل الارتباط (التلازم) البسيط. وكثيراً ما يعرف معامل الارتباط بأنه معامل اعتماد قياسي standard regression، أى أنه إذا قيست المشاهدات فى كل من المتغيرين بوحدة قياسية (أى مقسومة على انحرافها المعيارى) كان معامل اعتماد  $Y$  على  $X$  فى هذه الحالة يساوى معامل اعتماد  $X$  على  $Y$  يساوى الارتباط بينهما، فإذا رمز لمعامل الاعتماد القياسي هذا بالرمز  $b_{YX}$  فإن:

$$b_{YX}^{\circ} = \frac{\sum \frac{x}{\sigma_x} \sum \frac{y}{\sigma_y}}{\sum \frac{x^2}{\sigma_x^2}} = \frac{\sum xy}{\sum x^2} \cdot \frac{\sigma_x^2}{\sigma_x \sigma_y}$$

$$b_{YX}^{\circ} = b_{yx} \frac{\sigma_x}{\sigma_y} = b_{xy} \frac{\sigma_y}{\sigma_x} = r \quad (٣٩-١٠)$$

وسوف يتم فيما بعد شرح معامل الارتباط بالتفصيل.

١٠-١٠ العلاقة بين معامل التحديد  $r^2$  وخطأ التقدير  $S_{y.x}$

سبق بيان أن مجموع مربعات الانحرافات عن خط الانحدار وكان رمزه ESS أو  $d_{y.x}^2$  هو

$$ESS = \sum d_{y.x}^2 = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2} = \sum y^2 \left( 1 - \frac{(\sum xy)^2}{\sum x^2 \sum y^2} \right)$$

وحيث إن:

$$(b_{Y_1 Y_2})(b_{Y_2 Y_1}) = \frac{(\sum y_1 y_2)^2}{(\sum y_1^2)(\sum y_2^2)} = r^2$$

$$ESS = \sum y^2 (1 - r^2) \quad (٤٠-١٠)$$

وحيث إن ESS هي مجموع مربعات فإن قيمتها تساوى صفر على الأقل ولكي يتحقق ذلك يلزم أن تكون قيمة  $r^2$  ما بين الصفر والواحد الصحيح أى أن:

$$(0 \leq r^2 \leq 1) \quad (٤١-١٠)$$

وهذه إحدى خصائص معامل التحديد.

وعندما تكون  $n$  أى حجم العينة كبيراً فإن  $n-1$  تساوى تقريباً  $n-2$  وعلى ذلك فإن العلاقة (٤٠-١٠) تصبح:

$$\frac{ESS}{n-2} \cong \frac{\sum y^2}{n-1} (1-r^2)$$

أى أن:

$$S_{y.x}^2 \cong S_y^2 (1-r^2) \quad (٤٢-١٠)$$

وبأخذ الجذر التربيعى للطرفين:

$$S_{y.x} \cong S_y \sqrt{(1-r^2)} \quad (٤٣-١٠)$$

حيث  $S_{y.x}$  تمثل خطأ التقدير أما  $S_y$  فتمثل الانحراف المعياري للمتغير  $Y$ .  
وأيضاً:

$$\frac{S_{y.x}^2}{S_y^2} \cong 1-r^2 \quad (٤٤-١٠)$$

$$r^2 \cong 1 - \frac{S_{y.x}^2}{S_y^2} \cong \frac{S_y^2 - S_{y.x}^2}{S_y^2} \quad (٤٥-١٠)$$

وعندما تكون  $r^2 = 1$ ، وهو الحد الأقصى لقيمة معامل التحديد، فإن:  $S_{y.x}^2$  لا بد وأن تساوى الصفر. وعلى ذلك فإن قيمة  $r^2$  تساوى صفر فى حالة عدم وجود علاقة بين المتغيرين أى عندما يكون التباين هو نفسه التباين عن خط الانحدار وكما أن  $r^2$  تساوى الواحد الصحيح عندما تكون جميع النقط واقعة على خط الانحدار.

وإذا وصف معامل التحديد  $r^2$  كما سبق بأنه النسبة من التباين فى  $Y$  والتي ترجع إلى اعتماد  $Y$  على  $X$  فإن المقدار  $(1-r^2)$  يمثل النسبة من التباين فى أحد المتغيرين الخالية من تأثير المتغير الثانى. وعلى سبيل المثال إذا كانت  $r^2 = 0.74$  (كما فى مثال ١٠-١٠) فمعنى ذلك أن 74% من الاختلاف أو التباين فى أحد المتغيرين ترجع إلى الاختلافات فى قيم المتغير الآخر وأن  $1-0.74 = 0.26$  من التباين راجعة إلى الخطأ العشوائى أو التباين الغير منسوب إلى تباين قيم المتغير الثانى.

ويمكن اختبار مدى معنوية معامل التحديد باختبار فرض العدم  $H_0: \rho = 0$  ضد  
الفرض البديل  $H_0: \rho \neq 0$  حيث  $\rho$  عبارة عن معامل ارتباط العشيرة. ولاختبار ذلك  
الفرض تستخدم  $t$  بدرجات حرية  $n - 2$  كالتالى:

$$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \quad (10-16)$$

وعلى الرغم من أن معامل التحديد هو المقياس الأكثر استخداماً، إلا أن هذا لا  
يعطى دليلاً كافياً على أن  $r^2$  فعلاً تفسر قيمة الاختلافات فى المتغير التابع والراجعة  
إلى المتغير المستقل. ولذلك لابد من تقييم مدى ملائمة  $\text{evaluating the fit}$  النموذج  
المستخدم فى التحليل.

### 10-11 تقييم ملائمة نموذج التحليل $\text{Evaluating the fit}$

حتى الآن تم شرح النتائج الأساسية والتي تستخدم للوصول إلى استنتاجات فيما  
يتعلق بمفهوم النموذج الخطى البسيط. وهذه النتائج تكون صالحة وذات معنى حتى  
الآن فيما يتعلق فقط بالخطأ التجريبي (أو المتبقى residual) فى النموذج المستخدم.  
وكثيراً ما يجب تحليل مكونات هذا المتبقى من خلال عمل بعض الرسومات البيانية  
والتحليلية. وكما سبق فإن كبر قيمة معامل التحديد ومعنوية اختبار  $t$  له لا يؤكدان أن  
النموذج المستخدم يعتبر هو الأكثر ملائمة لتحليل البيانات محل الدراسة. والمثال  
التالى يوضح ذلك. ولابد من الأخذ فى الاعتبار أن التقييم الدقيق للمتبقى له أهمية فى  
التأكيد على المحافظة على الافتراضات الخاصة باستخدام نظرية المربعات الصغرى.

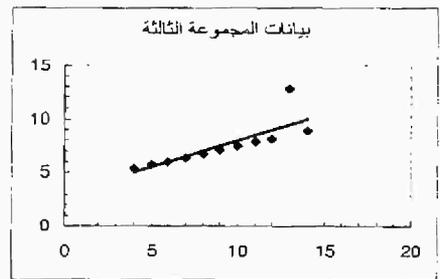
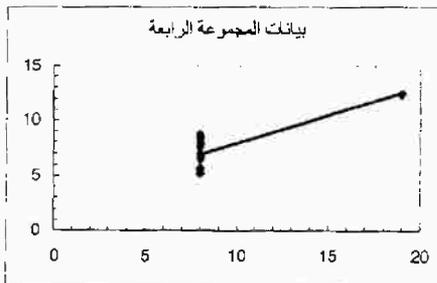
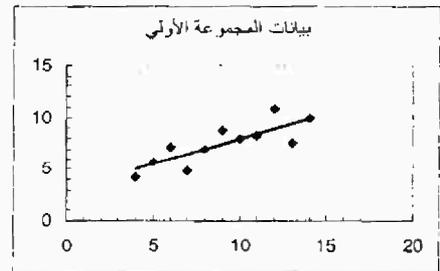
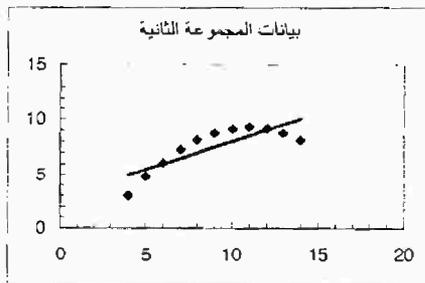
مثال 10-11

يوضح جدول 10-3 أربعة مجموعات من البيانات لمتغيرين  $X$ ،  $Y$  وهذه  
المجموعات لها نفس الإحصاءات. والشكل 10-8 يوضح الشكل الانتشارى لهذه  
البيانات مع خط الانحدار لكل مجموعة (المصدر: Chatterjee and Price, 1991).

يظهر تحليل بيانات هذا المثال باستخدام معادلة الانحدار الخطى البسيط أن معادلة  
الانحدار للمجموعات الأربع من البيانات متطابقة وهى  $\hat{Y} = 3 + 0.5X$  ومعامل  
التحديد قيمته  $r^2 = 0.67$  للأربعة مجموعات من البيانات. ولكن الشكل  
الانتشارى 10-8 وخط الانحدار لمجموعات البيانات كل على حده توضح أن  
النتائج التى اعتمدت على المتوسط ومعامل الانحدار ومعامل التحديد لم تتجح  
جميعها فى الكشف عن الاختلافات بين أنماط المجموعات الأربع من البيانات وبالتالي  
يكون التحليل غير صحيح.

جدول ١٠-٣ أربعة مجموعات من البيانات لمتغيرين  $X, Y$  لها نفس المتوسط

مجموعة ٤		مجموعة ٣		مجموعة ٢		مجموعة ١		
$Y_4$	$X_4$	$Y_3$	$X_3$	$Y_2$	$X_2$	$Y_1$	$X_1$	
6.58	8	7.46	10	9.14	10	8.04	10	
5.76	8	6.77	8	8.14	8	6.95	8	
7.71	8	12.74	13	8.74	13	7.58	13	
8.84	8	7.11	9	8.77	9	8.81	9	
8.47	8	7.81	11	9.26	11	8.33	11	
7.04	8	8.84	14	8.10	14	9.96	14	
5.25	8	6.08	6	6.13	6	7.24	6	
12.50	19	5.39	4	3.10	4	4.26	4	
5.56	8	8.15	12	9.13	12	10.84	12	
7.91	8	6.42	7	7.26	7	4.82	7	
6.89	8	5.73	5	4.74	5	5.68	5	
<b>7.5</b>	<b>9</b>	<b>7.5</b>	<b>9</b>	<b>7.5</b>	<b>9</b>	<b>7.5</b>	<b>9</b>	المتوسط



شكل ١٠-٨ شكل انتشاري مع خط الانحدار للأربعة مجموعات من بيانات مثال ١٠-١١.

لاحظ أن المخالفات الصغيرة small violations والخاصة بالتحليل بطريقة المربعات الصغرى لا تؤثر بدرجة كبيرة في استنتاجات التحليل بينما مخالفات نموذج التحليل المستخدم تؤدي إلى تغيير جذري في استنتاجات التحليل.

يعتبر تحليل المتبقي analysis of residual طريقة سهلة للكشف عن مدى ملائمة نموذج تحليل الانحدار. ويوجد طريقتان لتحليل المتبقي: الطريقة الأولى تعتمد على الرسم البياني حيث يمثل المحور الصادي القيم القياسية للمتبقي والمحور السيني عبارة عن القيم المتوقعة  $\hat{Y}$  أو قيم المتغير المستقل  $X_i$ ، أما الطريقة الثانية فهي تعتمد على تقسيم المتبقي إلى جزئين، الجزء الأول يسمى الخطأ النقي pure error والجزء الثاني يطلق عليه عدم الكفاية lack of fit.

#### ١٠-١١-١ تحليل المتبقي بطريقة الرسم البياني

من المعادلتين (١٠-٣) و (١٠-٤) وجدول تحليل التباين ١٠-٢ يمكن حساب قيمة الانحراف (المتبقي) عن خط الاعتماد لكل قيمة من المتغير المعتمد  $Y_i$  كالتالي:

$$e_i = Y_i - \hat{Y}_i$$

وهذه القيم يمكن تحويلها إلى قيم قياسية باستخدام  $e_{is} = e_i / S_{Y.X}$ .

بصفة عامة عندما يتم اختيار النموذج الصحيح فإن القيم القياسية للمتبقي تتراوح بين 2 إلى -2 وتوزع عشوائياً حول الصفر. والرسم البياني لابد أن لا يظهر نمط واضح للاختلافات في هذه القيم.

#### مثال ١٠-١٢

استخدم بيانات المثال ١٠-١١ لرسم القيم المتبقية القياسية  $e_{is}$  على المحور الصادي وكل من القيم المتوقعة  $\hat{Y}_i$  وقيم المتغير المستقل  $X_i$  للمجموعات الأربع من البيانات.  $S_{Y.X} = 1.24$  ،  $a = 3$  ،  $b_{Y.X} = 0.5$  للمجموعات الأربع من البيانات.

يوضح جدول ١٠-٤ قيم المتغير المستقل و القيم المتوقعة و القيم المتبقية القياسية لمجموعات البيانات الأربع المذكورة في مثال ١٠-١١، والشكلين ١٠-٩ و ١٠-١٠ يوضحان القيم المتبقية القياسية على المحور الصادي وقيم المتغير المستقل و القيم المتوقعة على المحور السيني للمجموعات الأربع، على الترتيب.

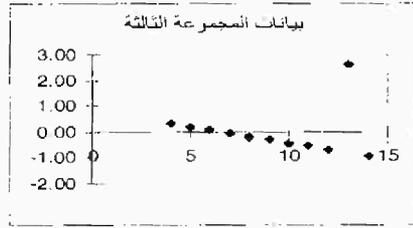
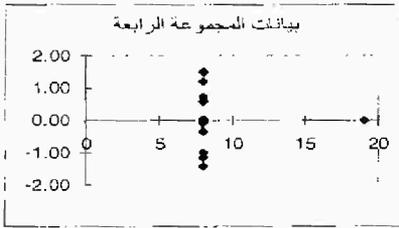
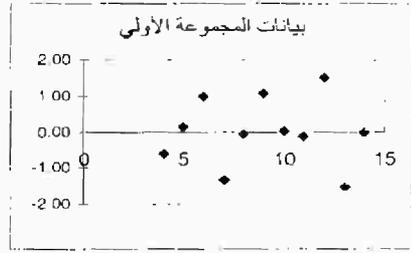
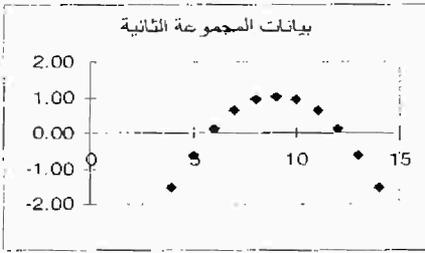
يتضح من الشكلين ١٠-٩ و ١٠-١٠ أن بيانات المجموعة الأولى فقط هي التي يمكن تحليلها بالنموذج  $Y = a + bX + e$  وكان توزيع القيم المتبقية القياسية لهذه المجموعة يبدو عشوائياً حول الصفر وتراوحت قيمه بين  $\pm 2$  ولم يظهر توزيع هذه القيم أي نمط معين بتغير قيم العامل المستقل أو القيم المتوقعة. بينما الأشكال البيانية الخاصة بمجموعات البيانات الثانية والثالثة والرابعة أظهرت جميعها نمطاً معيناً غير

عشوائى لتوزيع القيم المتبقية القياسية بتغير سواء العامل التابع أو القيم المتوقعة مما يدل على عدم صلاحية النموذج  $Y = a + bX + c$  فى تحليل هذه المجموعات من البيانات ولا بد من البحث عن نموذج آخر غير النموذج الخطى البسيط. ومعنى هذا أن هناك ربما عوامل أخرى تؤثر على المتغير التابع و/أو العلاقة بين  $X$  و  $Y$  ليست بالخطية ولكنها قد تكون أكثر تعقيداً.

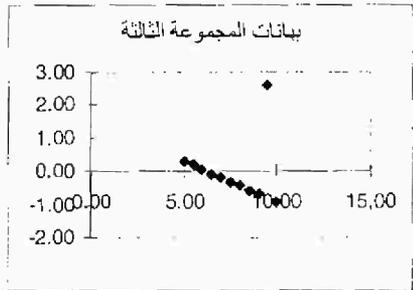
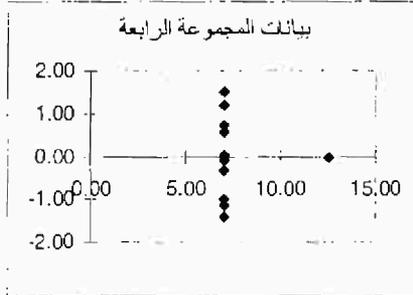
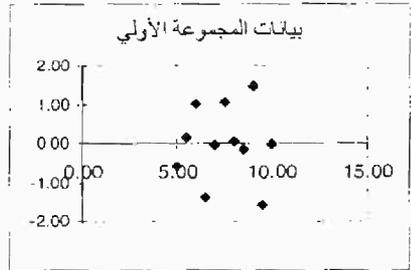
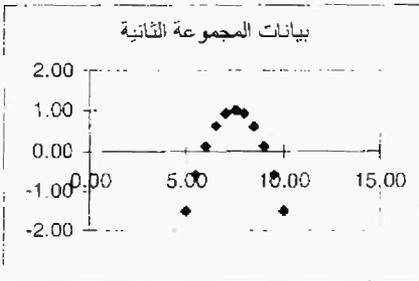
جدول ١٠-٤ قيم المتغير المستقل و القيم المتوقعة و القيم المتبقية القياسية للمجموعات الأربع من البيانات المذكورة فى مثال ١٠-١١

مجموعة ٤			مجموعة ٣			مجموعة ٢			مجموعة ١		
$e_{4s}$	$\hat{Y}_4$	$X_4$	$e_{3s}$	$\hat{Y}_3$	$X_3$	$e_{2s}$	$\hat{Y}_2$	$X_2$	$e_{1s}$	$\hat{Y}_1$	$X_1$
-0.34	7.0	8	-0.44	8.0	10	0.92	8.0	10	0.03	8.0	10
-1.00	7.0	8	-0.19	7.0	8	0.92	7.0	8	-0.04	7.0	8
0.57	7.0	8	2.62	9.5	13	-0.61	9.5	13	-1.55	9.5	13
1.49	7.0	8	-0.32	7.5	9	1.03	7.5	9	1.06	7.5	9
1.19	7.0	8	-0.56	8.5	11	0.61	8.5	11	-0.14	8.5	11
0.03	7.0	8	-0.94	10.0	14	-1.54	10.0	14	-0.03	10.0	14
-1.41	7.0	8	0.06	6.0	6	0.11	6.0	6	1.00	6.0	6
0.00	12.5	19	0.32	5.0	4	-1.54	5.0	4	-0.60	5.0	4
-1.16	7.0	8	-0.69	9.0	12	0.11	9.0	12	1.49	9.0	12
0.74	7.0	8	-0.06	6.5	7	0.61	6.5	7	-1.36	6.5	7
-0.09	7.0	8	0.19	5.5	5	-0.61	5.5	5	0.15	5.5	5

العلاقة بين متغيرين: الانحدار والارتباط البسيطان



شكل ٩-١٠ القيم المتبقية القياسية على المحور الصادي وقيم المتغير المستقل على المحور السيني للأربعة مجموعات من البيانات.



شكل ١٠-١٠ القيم المتبقية القياسية على المحور الصادي والقيم المتوقعة على المحور السيني للمجموعات الأربع من البيانات.

١٠-١١-٢ طريقة الخطأ النقي وعدم الكفاية Lack of fit and pure error

كما سبق فإن  $e_i = Y_i - \hat{Y}_i$  تمثل المتبقي عند  $X_i$  وهذه تمثل القيمة التي تختلف بها القيمة الحقيقية  $Y_i$  عن القيمة المتوقعة  $\hat{Y}_i$ . وأيضا سبق إيضاح أن  $\sum e_i = 0$ . تشمل قيمة المتبقي هذه على جميع المعلومات المتاحة والتي فشل النموذج المستخدم في التحليل في توظيفها لتفسير التباين في المتغير المعتمد  $Y$ . هذا المتبقي يمكن تقسيمه إلى جزئين الأول يعرف بالخطأ النقي pure error، كما سبق، ويرمز له بالرمز  $S_e^2$  والجزء الثاني يعرف بعدم الكفاية lack of fit ويرمز له بالرمز  $MS_L$  وهذا الجزء يمثل عدم ملائمة النموذج المستخدم في تحليل البيانات محل الدراسة. يجب ملاحظة أنه يلزم وجود مشاهدات متكررة لكل من  $X, Y$  حتى يمكن تحليل المتبقي. ولحساب ذلك افترض وجود بيانات مكررة يمكن تعريفها كالتالي:

$Y_{1n_1}, \dots, Y_{12}, Y_{11}$  عبارة عن عدد  $n_1$  من المشاهدات المكررة عند  $X_1$

$Y_{2n_2}, \dots, Y_{22}, Y_{21}$  عبارة عن عدد  $n_2$  من المشاهدات المكررة عند  $X_2$

$Y_{ju}$  عبارة عن مشاهدة  $u$  عند  $X_j$  حيث  $u = 1, 2, \dots, n_j$

$Y_{mn_m}, \dots, Y_{m2}, Y_{m1}$  عبارة عن عدد  $n_m$  من المشاهدات المكررة عند  $X_m$

$$\text{بمعنى أن } n = \sum_{j=1}^m \sum_{u=1}^{n_j} 1 = \sum_{j=1}^m n_j \text{ مشاهدة.}$$

مساهمة  $n_1$  من المشاهدات عند  $X_1$  في مجموع مربعات الخطأ النقي عبارة عن مجموع المربعات الداخلى لقيم  $Y_{1u}$  حول متوسطهم  $\bar{Y}_1$  بمعنى أن:

$$\sum_{u=1}^{n_1} (Y_{1u} - \bar{Y}_1)^2 = \sum_{u=1}^{n_1} Y_{1u}^2 - n_1 \bar{Y}_1^2 = \sum_{u=1}^{n_1} Y_{1u}^2 - \left( \sum_{u=1}^{n_1} Y_{1u} \right)^2 / n_1 \quad (٤٧-١٠)$$

تجميع pooling مجموع المربعات الداخلية الراجعة إلى كل المكررات يؤدي إلى الحصول على مجموع المربعات الراجع إلى الخطأ النقي

$$\sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \bar{Y}_j)^2 \quad (٤٨-١٠)$$

ودرجات الحرية عبارة عن:

$$n_e = \sum_{j=1}^m (n_j - 1) = \sum_{j=1}^m n_j - m \quad (٤٩-١٠)$$

وبما أن مجموع مربعات الخطأ النقي هو جزء من مجموع المربعات للمتبقى فإنه يمكن التعبير عن المتبقى للملاحظة  $u$  عند  $X_j$  كالتالي:

$$Y_{ju} - \hat{Y}_j = (Y_{ju} - \bar{Y}_j) - (\hat{Y}_j - \bar{Y}_j)$$

لاحظ أن جميع القيم المكررة عن أي  $X_j$  سوف يكون لهما نفس القيمة المتوقعة  $\hat{Y}_j$ . بتربيع جانبي المعادلة (٤٩-١٠) والجمع على كل من  $u$  و  $j$  يمكن الحصول على:

$$\sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \hat{Y}_j)^2 = \sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \bar{Y}_j)^2 + \sum_{j=1}^m \sum_{u=1}^{n_j} (\hat{Y}_j - \bar{Y}_j)^2 \quad (٥٠-١٠)$$

لاحظ أن الجانب الأيسر من هذه المعادلة عبارة عن مجموع مربعات المتبقى، أما الجانب الأيمن فيحتوي على مجموع مربعات الخطأ الحقيقي وهو

$$\sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \bar{Y}_j)^2$$

أما الجزء الثاني فهو مجموع مربعات عدم الكفاية .

### مثال ١٠-١٣

الجدول التالي يوضح عدد 24 مشاهدة بعضها مكرر. ومعادلة الانحدار الخطي البسيط لها  $\hat{Y} = 1.436 + 0.338X$ . يوضح جدول ١٠-٥ تحليل التباين لهذه البيانات، (المصدر: Draper and Smith, 1981).

مشاهدة	X	Y									
١	1.3	2.3	٧	3.3	1.8	١٣	4.7	5.4	١٩	5.3	2.1
٢	1.3	1.8	٨	3.7	3.7	١٤	4.7	3.2	٢٠	5.7	3.4
٣	2.0	2.8	٩	3.7	1.7	١٥	4.7	1.9	٢١	6.0	3.2
٤	2.0	1.5	١٠	4.0	2.8	١٦	5.0	1.8	٢٢	6.0	3.0
٥	2.7	2.2	١١	4.0	2.8	١٧	5.3	3.5	٢٣	6.3	3.0
٦	3.3	3.8	١٢	4.0	2.2	١٨	5.3	2.8	٢٤	6.7	5.9

## جدول ١٠-٥ تحليل بيانات مثال ١٠-١٣

SOV	df	SS	MS	F
Due to regression	1	6.326	6.926	6.569
From regression	22	21.192	$0.963 = S^2$	

الخطوة التالية هي حساب مجموع مربعات الخطأ النقي وبالطرح من مجموع مربعات الخطأ الراجع للانحدار يمكن الحصول على مجموع مربعات عدم الكفاية.

١ - مجموع مربعات الخطأ النقي لمكررات القيم Y عند  $X = 1.3$  عبارة عن  $0.125 = (0.5)(2.3 - 1.8)^2$  وهذه القيمة لها درجة حرية واحدة.

٢ - مجموع مربعات الخطأ النقي لمكررات القيم Y عند  $X = 4.7$  عبارة عن

$$(5.4)^2 + (3.2)^2 + (1.9)^2 - (3)[(5.4 + 3.2 + 1.9)/3]^2$$

$$= 43.01 - (10.5)^2/3 = 6.26$$

وهذه القيمة لها 2 درجة حرية.

ويمكن تلخيص حسابات مجموع مربعات الخطأ النقي في الجدول التالي:

df	$\sum_{u=1}^n (Y_{ju} - \bar{Y}_j)^2$	مستوى X
1	0.125	1.3
1	0.845	2.0
1	2.000	3.3
1	2.000	3.7
2	0.240	4.0
2	6.260	4.7
2	0.980	5.3
1	0.020	6.0
<b>11</b>	<b>12.470</b>	<b>المجموع</b>

وبالتالى يمكن إعادة كتابة جدول تحليل التباين ١٠-٥ فى جدول آخر ١٠-٦ والذى يظهر مجموع مربعات عدم الكفاية ومجموع مربعات الخطأ النقى بدرجات حرية = درجات الحرية من خط الاعتماد - درجات حرية عدم الكفاية، أى  $11 = 22 - 11$

جدول ١٠-٦ تحليل التباين لبيانات مثال ١٠-١٣ مع إظهار قيمة مجموع مربعات كل من عدم الكفاية والخطأ النقى

SOV	df	SS	MS	F
Due to regression	1	6.326	6.926	6.569
From regression	22	21.192	$0.963 = S^2$	
Lack of fit	11	8.722	$0.793 = MS_L$	0.699
Pure error	11	12.470	$1.134 = S_e^2$	
<b>C. Total</b>	<b>23</b>			

لاحظ أن  $F$  المحسوبة لعدم الكفاية حسبت من  $F = 0.793/1.134 = 0.699$  وهى غير معنوية بمستوى  $\alpha = 0.05$  وبالتالي فإن النموذج المستخدم يعتبر كافياً لتفسير التباين فى قيم  $Y$ .

مثال ١٠-١٤

يمكن استخدام برنامج SAS لإجراء اختبار عدم الكفاية lack of fit لبيانات مثال

١٠-١٣

```
DATA LACK;
INPUT X Y @@;
CARDS;
1.3 2.3 1.3 1.8 2 2.8 2 1.5 2.7 2.2 3.3 3.8
3.3 1.8 3.7 3.7 3.7 1.7 4 2.8 4 2.8 4 2.2
4.7 5.4 4.7 3.2 4.7 1.9 5 1.8 5.3 3.5 5.3 2.8
5.3 2.1 5.7 3.4 6 3.2 6 3 6.3 3 6.7 5.9
PROC SORT;
BY Y;
PROC RSREG;
MODEL Y = X / COVAR = 1 LACKFIT;
RUN;
```

لاحظ:

لا بد من أن يتم ترتيب البيانات على المتغير التابع ولذلك استخدم اختيار PROC SORT باستخدام Y كمتغير للترتيب.

استخدم اختيار PROC RSREG (quadratic response surface) بدلاً من PROC REG.

استخدم اختيار  $COVR = 1$  مع النموذج للإشارة إلى أن هناك متغيراً مستقلاً واحداً فقط.

استخدم اختيار lackfit مع النموذج للحصول على مجموع المربعات الراجع لكل من الخطأ النقي وعدم الكفاية.

النتائج:

## The RSREG Procedure

## Response Surface for Variable Y

Response Mean	2.858333
Root MSE	0.981503
R-Square	0.2298
Coefficient of Variation	34.3383

Regression	DF	Type I Sum		F Value	Pr > F
		of Squares	R-Square		
Covariates	1	6.324667	0.2298	6.57	0.0178
Linear	0	0	0.0000	.	.
Quadratic	0	0	0.0000	.	.
Crossproduct	0	0	0.0000	.	.
Total Model	1	6.324667	0.2298	6.57	0.0178
Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	11	8.723666	0.793061	0.70	0.7183
Pure Error	11	12.470000	1.133636		
Total Error	22	21.193666	0.963348		

صندوق ١٠-٣

- تسمى نسبة التباين الراجع للاعتماد إلى التباين الكلي  $(b\sum xy / \sum y^2)$  بمعامل التحديد  $r^2$ .
- لا يجب أخذ ارتفاع  $r^2$  ومعنويتها دليلاً على أن النموذج المفترض قد استخلص كل المعلومات في  $Y$ ، ففي كثير من هذه الحالات تكون مازالت معلومات في المتبقى لم يستخلصها النموذج.
- في الحالات التي يكون فيها أكثر من قيمة للمتغير المعتمد  $Y$  لنفس قيمة المتغير المستقل  $X$ ، فإنه يمكن تقسيم مجموع مربعات المتبقى إلى جزئين:

جزء راجع لكفاية النموذج الإحصائي المستخدم والآخر خطأ نقي، وبهذا يمكن أخذ فكرة عن مدى ملائمة النموذج الإحصائي وعن القدر الذي استخلص به المعلومات في  $Y$ .

١٠-١٢ الارتباط البسيط Simple correlation

(معامل الارتباط لبيرسون Pearson correlation coefficient)

يقيس معامل الارتباط شدة العلاقة بين متغيرين أو هو مقياس لدرجة تغير متغيرين معاً وهو تقدير غير متحيز لمعامل ارتباط المتغيرين في العشرة ويحسب من المعادلة التالية:

$$r_{Y_1 Y_2} = \frac{\sum (Y_1 - \bar{Y}_1)(Y_2 - \bar{Y}_2)}{\sqrt{\sum (Y_1 - \bar{Y}_1)^2} \sqrt{\sum (Y_2 - \bar{Y}_2)^2}}$$

$$= \frac{\sum y_1 y_2}{\sqrt{\sum y_1^2} \sqrt{\sum y_2^2}} \quad (١٠-٥١)$$

وبقسمة البسط والمقام على درجات الحرية  $n-1$  فإن:

$$r_{Y_1 Y_2} = \frac{(\sum y_1 y_2)/(n-1)}{\sqrt{(\sum y_1^2)/(n-1)}\sqrt{(\sum y_2^2)/(n-1)}} = \frac{\text{cov}(Y_1, Y_2)}{(S_{Y_1})(S_{Y_2})} \quad (٥٢-١٠)$$

حيث  $S_{Y_1}$  هي الانحراف المعياري للمتغير  $Y_1$ ،  $S_{Y_2}$  هي الانحراف المعياري للمتغير  $Y_2$ .

$$r_{Y_1 Y_2} = \frac{130}{\sqrt{168}\sqrt{135.5}} = 0.86 \text{ وفي مثال } ١٠-١٠ \text{ فإن معامل الارتباط}$$

ومعامل الارتباط هو الجذر التربيعي لحاصل ضرب معاملي الانحدار، أي أنه عبارة عن المتوسط الهندسي لهما أي:

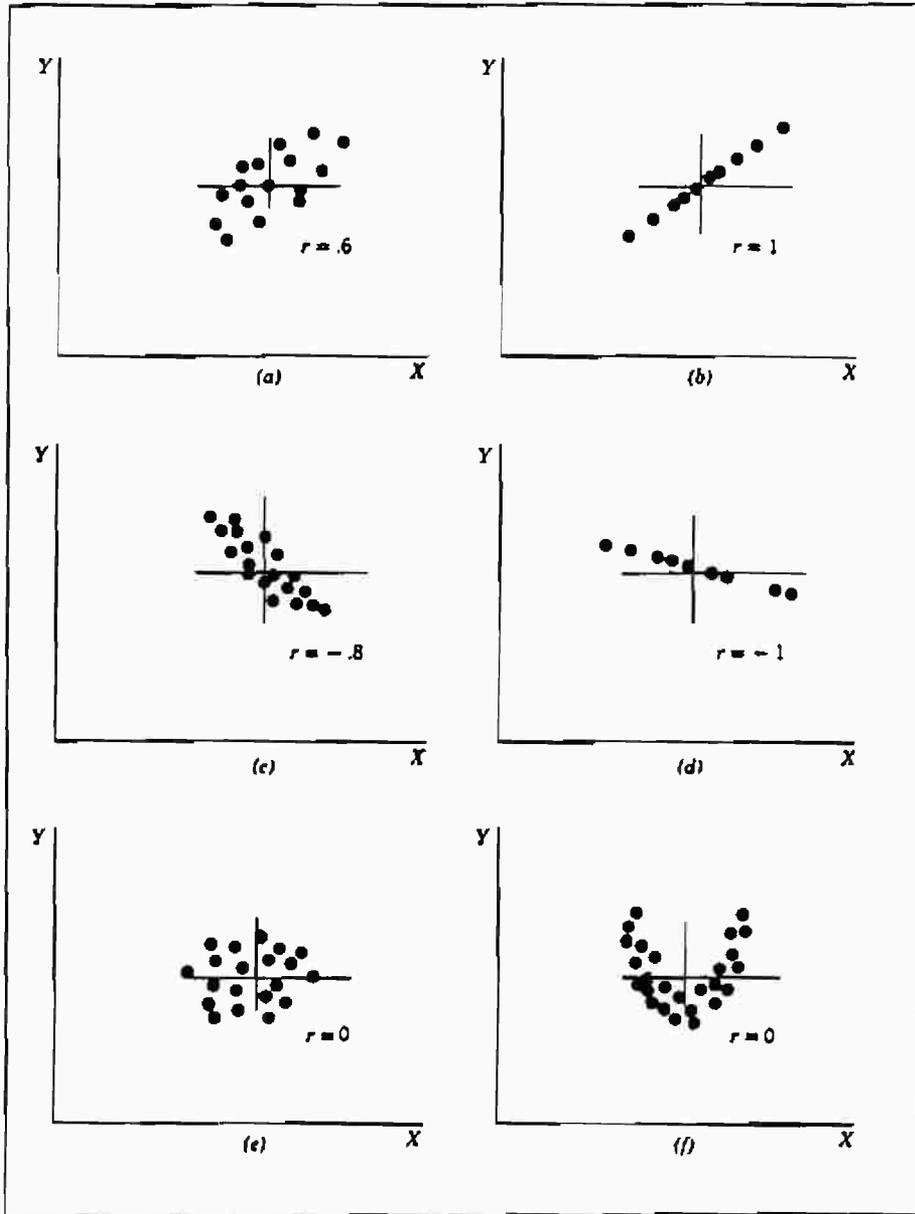
$$r = \sqrt{(b_{Y_1 Y_2})(b_{Y_2 Y_1})} \quad (٥٣-١٠)$$

$$r = \sqrt{(0.959)(0.774)} = 0.86 \text{ وفي مثال } ١٠-١٠ \text{ فإن معامل الارتباط}$$

وإذا كان معامل التحديد، وهو مربع معامل الارتباط، تتحصر قيمته بين الصفر والواحد الصحيح فإن معامل الارتباط تتحصر قيمته بين  $+1$ ،  $-1$  أي أن:

$$(-1 \leq r \leq +1) \quad (٥٤-١٠)$$

ويكون معامل الارتباط موجباً عندما تكون القيم الكبرى للمتغير الأول تقابلها القيم الكبرى للمتغير الثاني، أي أن تغير المتغيرين يكون في نفس الاتجاه. ويكون معامل الارتباط سالباً عندما تكون القيم الكبرى لأحد المتغيرين تقابلها القيم الصغرى للمتغير الآخر. ويمكن من الأشكال الانتشارية (١٠-١١ a, b, c, d, e, f) الاستدلال على الارتباط. الشكل (b) يمثل علاقة خطية كاملة موجبة  $r=1$ ، والشكل (d) يمثل علاقة خطية كاملة سالبة  $r=-1$ . الشكل (a) يمثل علاقة غير كاملة موجبة  $r=0.6$  والشكل (c) يمثل علاقة غير كاملة سالبة  $r=-0.8$ . عند مقارنة الشكلين (e) و (f) يلاحظ أن الارتباط قيمته صفر في الحالتين ولكن الشكل (f) يظهر بوضوح وجود علاقة قوية بين  $Y$  و  $X$  رغم أن  $r=0$  ومعنى ذلك أن معامل الارتباط في هذه الحالة لا يعنى عدم وجود علاقة بين المتغيرين ولكنه يعنى عدم وجود علاقة خطية.



شكل ١٠-١١ حالات مختلفة تمثل العلاقة بين متغيرين X و Y

وعلى عكس معامل الاعتماد الذي هو قيمة مميزة فإن معامل الارتباط كمية مجردة أو مطلقة، أي مستقلة عن وحدات القياس. كما يمكن حسابه بوحدات مختزلة

لكل من المتغيرين ولا يلزم لها التعديل بعد ذلك. والملاحظ من دراسات البيانات البيولوجية أن معامل الارتباط نادراً ما يكون أعلى من 0.9.

مثال ١٠-١٥

احسب معامل انحدار  $Y_2$  على  $Y_1$  وكذا معامل انحدار  $Y_1$  على  $Y_2$  ومعامل الارتباط بينهما للبيانات التالية، ثم مثل العلاقة بين المتغيرين بيانياً.

2	0	8	4	1	: $Y_1$
3	1	9	5	2	: $Y_2$

من البيانات:

$$n = 5 \quad \sum Y_1 = 15 \quad \sum Y_1^2 = 85$$

$$\sum Y_1 Y_2 = 100 \quad \sum Y_2 = 20 \quad \sum Y_2^2 = 120$$

$$b_{Y_1 Y_2} = \frac{\sum Y_1 Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sum Y_1^2 - \frac{(\sum Y_1)^2}{n}} = \frac{100 - \frac{(15)(20)}{5}}{85 - \frac{(15)^2}{5}} = 1$$

$$b_{Y_2 Y_1} = \frac{\sum Y_1 Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sum Y_2^2 - \frac{(\sum Y_2)^2}{n}} = \frac{100 - \frac{(15)(20)}{5}}{120 - \frac{(20)^2}{5}} = 1$$

معامل انحدار  $Y_1$  على  $Y_2$  عبارة عن واحد صحيح، أي وحدة من  $Y_1$  لكل وحدة من  $Y_2$ ، وكذلك معامل انحدار  $Y_2$  على  $Y_1$ .

معادلتى خطى الانحدار هما:

$$\hat{Y}_1 = -1 + Y_2$$

$$\hat{Y}_2 = 1 + Y_1$$

ومعامل الارتباط هو:

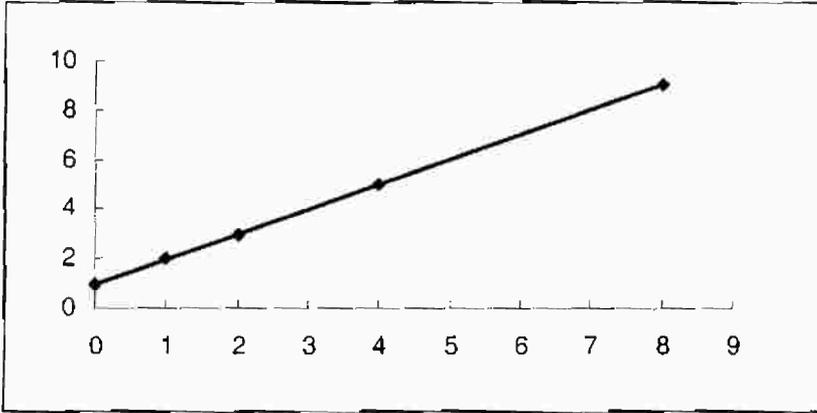
$$r_{Y_2 Y_1} = \frac{\sum Y_1 Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sqrt{\sum Y_1^2 - \frac{(\sum Y_1)^2}{n}} \sqrt{\sum Y_2^2 - \frac{(\sum Y_2)^2}{n}}}$$

$$= \frac{100 - \frac{(15)(20)}{5}}{\sqrt{85 - \frac{(15)^2}{5}} \sqrt{120 - \frac{(20)^2}{5}}} = 1$$

وكما سبق يمكن استخدام معاملي الانحدار في حساب معامل الارتباط كالتالي:

$$r = \sqrt{(b_{Y_1 Y_2})(b_{Y_2 Y_1})} = \sqrt{(1)(1)} = 1$$

والشكل ١٠-١٢ يمثل العلاقة بين المتغيرين بيانياً.



شكل ١٠-١٢ التمثيل البياني لخطي الانحدار في مثال ١٠-١٥

لاحظ أن جميع النقط تقع على خط واحد، خط انحدار  $Y_2$  على  $Y_1$  هو نفسه خط الانحدار  $Y_1$  على  $Y_2$  وهذا لا يحدث إلا عندما يكون الارتباط تاماً بين المتغيرين وهي الحالة التي تكون القيم المتوقعة predicted values هي نفسها القيم الفعلية actual values حيث  $S_{Y_1 Y_2}^2 = 0$ .

استخدام برنامج SAS فى حساب معامل انحدار  $Y_2$  على  $Y_1$  وكذا معامل انحدار  $Y_1$  على  $Y_2$  ومعامل الارتباط بينهما لبيانات مثال ١٠-١٥.

```
DATA RC;
INPUT Y1 Y2 @@;
CARDS;
1 2 4 5 8 9 0 1 2 3
PROC REG;
MODEL Y1 = Y2;
MODEL Y2 = Y1;
PROC CORR;
RUN;
```

لاحظ:

كل معامل انحدار يراد حسابه يوضع فى نموذج model منفصل .  
استخدام PROC CORR لحساب معامل الارتباط.

النتائج

The REG Procedure  
Model: MODEL1

Dependent Variable: Y1  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	40.00000	40.00000	Infy	<.0001
Error	3	0	0		
Corrected Total	4	40.00000			

Root MSE 0 R-Square 1.0000  
Dependent Mean 3.00000 Adj R-Sq 1.0000 Coeff Var 0

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-1.00000	0	-Infy	<.0001
Y2	1	1.00000	0	Infy	<.0001

Dependent Variable: Y2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	40.00000	40.00000	Infy	<.0001
Error	3	0	0		
Corrected Total	4	40.00000			

Root MSE	0	R-Square	1.0000
Dependent Mean	4.00000	Adj R-Sq	1.0000
Coeff Var	0		

Parameter Estimates

Variable	Parameter		Standard		Pr >  t
	DF	Estimate	Error	t Value	
Intercept	1	1.00000	0	Infy	<.0001
Y1	1	1.00000	0	Infy	<.0001

The CORR Procedure

2 Variables: Y1 Y2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Y1	5	3.00	3.16228	15.00	0	8.00
Y2	5	4.00	3.16228	20.00	1.00	9.00

Pearson Correlation Coefficients, N = 5  
 Prob > |r| under H0: Rho=0

	Y1	Y2
Y1	1.00000	1.00000
Y2	1.00000	1.00000

١٠-١٣ العلاقة بين معامل الانحدار ومعامل الارتباط

١- حيث إن معامل انحدار  $Y_2$  على  $Y_1$  هو

$$b_{Y_2 Y_1} = \frac{\text{cov}(Y_1, Y_2)}{S_{Y_1}^2}$$

فإنه بضرب كل من البسط والمقام في  $S_{Y_2}$

$$b_{Y_2 Y_1} = \left( \frac{\text{cov}(Y_1, Y_2)}{S_{Y_1}^2} \right) \left( \frac{S_{Y_2}}{S_{Y_2}} \right) = \left( \frac{\text{cov}(Y_1, Y_2)}{(S_{Y_1})(S_{Y_2})} \right) \left( \frac{S_{Y_2}}{S_{Y_1}} \right)$$

$$r b_{Y_2 Y_1} = \left( \frac{S_{Y_2}}{S_{Y_1}} \right) \quad (١٠-٥٥)$$

وحيث إن  $S_{y_2} = \sqrt{(\sum y_2^2)/(n-1)}$  ،  $S_{y_1} = \sqrt{(\sum y_1^2)/(n-1)}$

$$r b_{Y_2 Y_1} = r \sqrt{(\sum y_2^2)/(\sum y_1^2)} \quad (١٠-٥٦)$$

أى أن:

$$r = b_{Y_2 Y_1} \sqrt{(\sum y_1^2)/(\sum y_2^2)} \quad (١٠-٥٧)$$

وبالمثل فإن:

$$r = b_{Y_1 Y_2} \sqrt{(\sum y_2^2)/(\sum y_1^2)} \quad (١٠-٥٨)$$

٢- حيث إن معادلة خط انحدار  $Y_2$  على  $Y_1$  هي  $\hat{Y}_2 = \bar{Y}_2 + b(Y_1 - \bar{Y}_1)$  وبالتعبير عن  $Y_1$  ،  $Y_2$  بوحداتهم القياسية standard units أى:

$$Y_2^\circ = \frac{Y_2 - \bar{Y}_2}{S_{Y_2}} \quad , \quad Y_1^\circ = \frac{Y_1 - \bar{Y}_1}{S_{Y_1}}$$

وبالتعويض في معادلة خط الانحدار عن  $Y_1$  ،  $Y_2$  بوحداتهم القياسية فإن

$$\hat{Y}_2 S_{y_2} = b_{Y_2 Y_1} Y_1 S_{y_1}$$

$$\hat{Y}_2 = b_{Y_2 Y_1} \frac{S_{Y_1}}{S_{Y_2}} Y_1$$

$$\hat{Y}_2 = r Y_1 \quad (١٠-٥٩)$$

حيث  $\hat{Y}_2$  تمثل القيمة المتوقعة معبراً عنها بوحدات قياسية. وباستخدام الوحدات القياسية يصبح معامل الارتباط  $r$  هو نفسه معامل الانحدار والاختلافات بينهما تتلاشى.

٣- سبق بيان أن معامل الارتباط هو المتوسط الهندسى لمعاملى الانحدار، أى:

$$r = \pm \sqrt{(b_{Y_1 Y_2})(b_{Y_2 Y_1})}$$

وخطى الانحدار يتطابقان عندما تكون  $r = \pm 1$ ، ويكونان قريبين جداً من بعضهما عندما يكون معامل الارتباط قريب من  $\pm 1$ .

#### ١٠-١٤ اختبار معنوية معامل الارتباط وتقدير حدود الثقة له

إذا سحب عدد كبير من العينات العشوائية حجم كل منها  $n$  من عشيرة تتبع التوزيع الطبيعي لمتغيرين معامل الارتباط بينهما  $\rho = 0$  وحسب معامل الارتباط لكل عينة، فإن توزيع معاملات ارتباط العينات يقترب من التوزيع الطبيعي ويزداد اقتراباً منه بزيادة حجم العينة. وبالتالي فإن التوزيع العيني لمعاملات الارتباط يكون متماثلاً أيضاً حتى عندما يكون توزيع أحد المتغيرين غير طبيعي كأن يكون قيماً ثابتة  $fixed$  مثلاً يختارها الباحث على أن يكون توزيع المتغير الآخر طبيعياً. أما إذا كان معامل ارتباط العشيرة  $\rho \neq 0$  فإن التوزيع العيني لمعاملات الارتباط لا يتبع التوزيع الطبيعي وإنما يكون التوزيع ملتويًا  $skewed$  مهما كان حجم العينات. ويمثل شكل ١٠-١٣ التوزيع العيني لمعاملات الارتباط المسحوبة من عشيرة معامل الارتباط يساوى صفرًا وكذلك التوزيع العيني لمعاملات ارتباط مسحوبة من عشيرة معامل الارتباط بها لا يساوى الصفر.

صندوق ١٠-٤

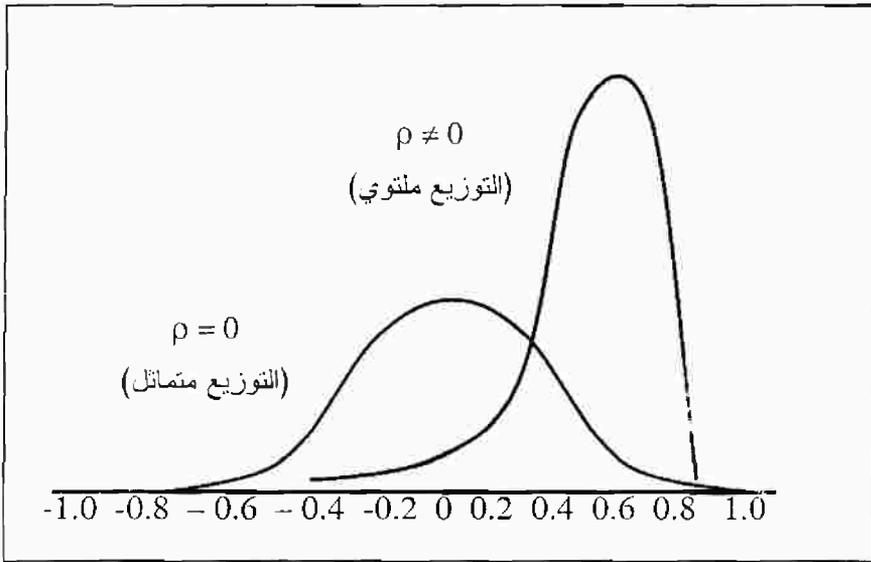
- معامل الارتباط ( $r$ ) بين متغيرين هو مقياس لشدة العلاقة بينهما. تتراوح قيمته بين  $+1$ ،  $-1$  وكلما بعدت قيمته عن الصفر ازدادت شدة العلاقة بين المتغيرين.
- معامل الارتباط كمية مطلقة لا تميز.
- العلاقة الرياضية بين معاملي الاعتماد والارتباط وثيقة جداً، ويمكن تعريف الأخير بأنه الجذر التربيعي لحاصل ضرب معامل اعتماد  $X$  على  $Y$  ومعامل اعتماد  $Y$  على  $X$ . ويعرف معامل الارتباط كذلك بأنه معامل الانحدار القياسي أو المعياري standard (أى بين المتغيرين بعد تعبيرهما  $(Y - \bar{Y}) / \sigma_Y$ ) لأحد المتغيرين على الآخر

$$\begin{aligned} r &= b_{Y_1 Y_2} / \sqrt{\sum y_2^2 / \sum y_1^2} \\ &= b_{Y_2 Y_1} / \sqrt{\sum y_1^2 / \sum y_2^2} \\ &= \frac{\sum y_1 y_2}{\sqrt{(\sum y_1^2)(\sum y_2^2)}} \end{aligned}$$

- يسمى مربع معامل الارتباط ( $r^2$ ) بمعامل التحديد أى النسبة من تباين المتغير المرتبط بالمتغير الآخر أو الممكن تفسيرها من قبل المتغير الآخر.
- تختبر معنوية  $r$  فى حالة فرض العدم ( $\rho = 0$ ) عن طريق اختبار  $t$  أو مباشرة عن طريق جداول خاصة.
- فى حالة ما إذا كان فرض العدم ( $\rho \neq 0$ ) فإن هذه الحالة تحتاج إلى تحويل خاص.

١٠-١٤-١ اختبار معنوية معامل الارتباط عندما يكون معامل الارتباط في العشيرة يساوى صفراً ( $\rho = 0$ )

يستخدم في مثل هذه الحالة، والتي يكون فيها توزيع  $r$  متماثلاً حول المنتصف (كما في شكل ١٠-١٣) طريقتان لاختبار معنوية العلاقة بين متغيرين. الأولى اختبار معنوية الانحدار باستخدام اختبار  $t$  ودرجات حرية  $n-2$  ومستوى معنوية وليكن  $\alpha$ . والثانية اختبار معنوية معامل الارتباط باستخدام الجداول الخاصة بذلك. وكل من الاختبارين يعطى نفس النتيجة ويمكن استخدام أى منهما.



شكل ١٠-١٣ التوزيع العيني لمعاملات الارتباط

أولاً: استخدام اختبار  $t$ :

$$H_0: \rho = 0$$

$$t = \frac{b}{S_b}$$

بدرجات حرية  $n-2$  ومن العلاقة (١٠-٥٦)

$$b = r \frac{S_{Y_2}}{S_{Y_1}}$$

$$\begin{aligned} \hat{r}_t &= r \frac{S_{Y_2}}{S_{Y_1}} / S_b = r \frac{S_{Y_2}}{S_{Y_1}} / \sqrt{\left[ \frac{\sum y_2^2 - r^2 \sum y_1^2}{(n-2)\sum y_1^2} \right]} \\ &= \frac{r S_{Y_2}}{S_{Y_1}} \sqrt{\frac{(n-2)\sum y_1^2}{\sum y_2^2(1-r^2)}} \end{aligned}$$

وحيث إن:

$$\frac{S_{y_2}}{S_{y_1}} = \sqrt{\frac{\sum y_2^2}{\sum y_1^2}}$$

$$\hat{r}_t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (٦٠-١٠)$$

حيث يعبر عن  $\frac{(1-r^2)}{(n-2)}$  بتباين معامل الارتباط.

مثال ١٠-١٧

في مثال ١٠-١٠ اختبر معنوية معامل الارتباط إذا كان معامل ارتباط العشيرة يساوى صفراً.

حيث  $n = 8$ ،  $r = 0.86$ ، درجات الحرية  $n - 2 = 8 - 2 = 6$  وحيث إن:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(0.86)(\sqrt{6})}{\sqrt{1-(0.86)^2}} = 8.09$$

وبما إن  $t_{(6,0.05)} = 2.447$ ،  $t_{(6,0.01)} = 3.707$  وهذه القيم أعلى من قيمة  $t$  المحسوبة (8.09) وبالتالي يرفض فرض العدم وهذا يعنى أن العلاقة بين المتغيرين معنوية بدرجة ثقة 99%.

ثانياً: وضعت جداول بمستوى معنوية 5% وأيضاً بمستوى معنوية 1% لاختبار معنوية معامل الارتباط مباشرة وذلك بدرجات حرية  $n - 2$  (جدول ٧ ملحق أ)، وتتوقف نتيجة الاختبار على حجم العينة وعلى قيمة  $r$  فمثلاً:

العينة الأولى:  $r_1 = 0.3$   $n_1 = 10$  درجات الحرية = 8

العينة الثانية:  $r_2 = 0.3$   $n_1 = 100$  درجات الحرية = 98

العينة الثالثة:  $r_3 = 0.7$   $n_3 = 10$  درجات الحرية = 8

على ذلك عند درجات حرية 8 كان معامل الارتباط والذي قيمته 0.3 غير معنوي بينما معامل الارتباط وقيمته 0.7 فهو معنوي بمستوى معنوية 5% عند نفس درجات الحرية. بينما معامل الارتباط (0.3) والذي كان غير معنوي عند درجات حرية 8 يكون معنوياً بمستوى معنوية 1% عند درجات حرية 98، ويجب ملاحظة أنه بزيادة درجات الحرية فإن الحد الأدنى لمعامل الارتباط المعنوي يقل.

والعلاقة المعنوية بين المتغيرين لا تعني بالضرورة أن أحد المتغيرين هو سبب التغير في المتغير الآخر (أي علاقة سببية) فقد يكون السبب راجع إلى مصادر أخرى. وإذا كانت  $r = 0.3^{**}$  وبالرغم من أن معامل الارتباط معنوي جداً - أي رفض فرض انعدم القائل بأنه ليس هناك علاقة بين المتغيرين وقبول الفرض البديل بوجود علاقة بينهما - إلا أن هذا يعني أن  $(0.3)^2 = 0.09$  من تباين  $Y_2$  يرجع للعلاقة بين المتغيرين  $Y_1$ ،  $Y_2$  وأن  $(1 - 0.09) = 0.91$  من التباين خالي من تأثير تلك العلاقة ويرجع إلى مصادر أخرى ذكرت فيما سبق عند الحديث عن عنصر الخطأ في فصل (10-1). وتبين الاختبارات الإحصائية أن هناك علاقة خطية بميل لا يساوي الصفر ومعنوي. ولكن هذا لا يعني أن معظم الاختلافات في  $Y_2$  هي نتيجة انحدارها على  $Y_1$  بل قد يكون هناك مصادر أخرى لتباين.

وجدير بالملاحظة أنه إذا كان معامل انحدار  $b_{yx}$  معنوياً عند مستوى معنوية معين كان أيضاً كل من  $r$ ،  $b_{xy}$  معنوياً عند نفس المستوى.

١٠-١٤-٢ اختبار معنوية معامل الارتباط عندما يكون معامل ارتباط العشيرة لا يساوي صفراً ( $p \neq 0$ )

إجراء مثل هذا الاختبار فإنه يفترض أن يكون توزيع كل من المتغيرين طبيعياً، أي يشترط أن تكون أزواج المتغيرات في العينة تمثل عينة عشوائية مأخوذة من عشيرة ذات متغيرين وهي السابق الإشارة إليها. وإذا كان  $\rho$  لا يساوي صفراً فإن التوزيع العيني لمعاملات الارتباط يكون ملتوياً ولا يجوز استخدام  $t$  والتي تستخدم فقط عندما يكون التوزيع متماثلاً أي عندما تكون  $\rho = 0$ . ولقد وجد Fisher أنه بتحويل  $r$  إلى قيمة أخرى هي  $Z$  تصبح الأخيرة ذات توزيع طبيعي. وتحسب  $Z$  كما يلي:

$$Z_r = \frac{1}{2} [\log_e(1+r) - \log_e(1-r)] \quad (10-11)$$

حيث  $\log_e$  هو اللوغاريتم الطبيعي أى  $\ln$  ، وتوزيع  $Z$  طبيعياً بمتوسط  $Z_p$  حيث  $Z_p = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$  وانحراف معياري  $\sigma_z = \frac{1}{\sqrt{n-3}}$ .

ووضعت جداول لتحويل  $r$  إلى ما يقابلها من قيم  $Z$  (جدول ٨ ملحق أ) وأخرى للتعبير عن قيم  $Z$  بما يقابلها من قيم  $Z$  (جدول ٩ ملحق أ) ويمكن تغير أى منهما للأخرى بدرجة كافية من الدقة.

وعلى ذلك فالكمية  $\frac{Z_r - Z_p}{\sqrt{n-3}}$  تتوزع طبيعياً أيضاً. وتستخدم جداول التوزيع الطبيعي أو جدول  $t$  عند درجات حرية  $\infty$  ويفضل ألا يقل حجم العينة عن 50 حتى يكون معامل الارتباط أكثر وثوقاً *reliable*.

مثال 10-18

أخذت عينة عشوائية من عشيرة ذات معبرين وكان معامل الارتباط 0.5، اختبر ما إذا كانت هذه العينة مسحوبة من عشيرة معامل الارتباط بها 0.6 علماً بأن حجم العينة كان 50 فرداً بدرجة ثقة 95%.

$$H_0: \rho = 0.6$$

إذا كانت  $r = 0.5$  ، فإن قيمة  $Z$  المقابلة لها هي  $Z = 0.549$

وعند قيمة  $\rho = 0.6$  قيمة  $Z$  المقابلة لها هي  $Z_p = 0.693$

$$\sigma_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{50-3}} = 0.146$$

وبالتالى

$$\frac{|Z_r - Z_p|}{\sigma_z} = \frac{|0.549 - 0.693|}{0.146} = 0.99$$

وهذه القيمة (0.99) تقارن بقيمة  $t_{(\infty, 0.05)} = 1.96$  وعلى ذلك لا يرفض فرض العدم بأن معامل ارتباط العينة 0.5 لا يختلف عن معامل ارتباط العشيرة 0.6 أى أن هذه العينة مسحوبة من عشيرة معامل الارتباط بها 0.6.

١٠-١٤-٣ تقدير حدود الثقة لمعامل ارتباط العشييرة باستخدام بيانات العينة

حيث إن التوزيع العيني لمعاملات الارتباط يصبح ملتويًا إذا كان معامل الارتباط العشييرة لا يساوى الصفر وحيث إن  $Z$  تتوزع طبيعياً بغض النظر عن حجم العينة بمتوسط  $Z_p$  وتباين  $\frac{1}{n-3}$  فإن حدود الثقة لـ  $Z_p$  هي:

$$\Pr(Z_r - Z_{\alpha/2}\sigma_Z \leq Z_p \leq Z_r + Z_{\alpha/2}\sigma_Z) = 1 - \alpha \quad (١٠-٦٢)$$

يمكن أن يعبر عن قيمة  $Z_{\alpha/2}$  بأنها قيمة  $t$  عند مستوى معنوية  $\alpha$  ودرجات حرية  $\infty$ .

وباستخدام (جدول ٩ ملحق أ) يمكن إيجاد قيم  $r$  المقابلة والتي تمثل حدود الثقة لمعامل ارتباط العشييرة.

مثال ١٠-١٩

ما هي حدود الثقة لمعامل ارتباط العشييرة إذا كان معامل ارتباط العينة 0.425 بحجم العينة 35 وذلك بدرجة ثقة 99% ؟

بما أن  $r = 0.425$  وباستخدام (جدول ٩ ملحق أ) فإن  $Z_r = 0.454$

$$t_{(\infty, 0.01)} = 2.576, \sigma_Z = \frac{1}{\sqrt{35-3}} = \frac{1}{32} = 0.177$$

∴ حدى الثقة لـ  $Z_p$  هما:

$$0.454 \pm (2.576)(0.177) = 0.454 \pm 0.455$$

أى هما -0.001 يمثل الحد الأدنى و +0.909 يمثل الحد الأعلى. وبتحويل  $Z$  إلى قيم  $r$  المقابلة يكون الحد الأدنى -0.001 والحد الأعلى 0.721 بدرجة ثقة 99% .

لاحظ أن حدى الثقة ليسا على نفس المسافة على كلا الجانبين لمعامل ارتباط العينة حيث التوزيع غير متماثل كما أنه إذا زاد حجم العينة فإن فترة الثقة تنكمش والعكس صحيح.

١٠-١٥ اختبار تساوى معاملى ارتباط لكل من عينتين مأخوذتين من نفس العشييرة معامل الارتباط بها لا يساوى صفراً

أى اختبار فرض العدم بأن الفرق بين معاملى الارتباط يساوى الصفر. ولكى يتم ذلك تحول كل من قيمتى معامل الارتباط إلى قيم  $Z$  المقابلة ثم تختبر معنوية الفرق بين قيمتى  $Z$  وما ينطبق على  $Z$  فهو ينطبق على  $r$ .

$$\sigma_{Z_1}^2 = \frac{1}{n_1 - 3} \text{ هو } Z_1 \text{ حيث تباين } Z_1$$

$$\text{و تحول } r_2 \text{ إلى } Z_2 \text{ وتباين } Z_2 \text{ هو } \sigma_{Z_2}^2 = \frac{1}{n_2 - 3}$$

وعلى ذلك فإن  $\sigma_{Z_1 - Z_2}^2 = \sigma_{Z_1}^2 + \sigma_{Z_2}^2$  والانحراف المعياري للفرق بين  $Z_1$  و  $Z_2$  هو

$$S_{Z_1 - Z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad (١٠-٦٣)$$

وحيث إن  $\frac{Z_1 - Z_2}{S_{Z_1 - Z_2}}$  تتبع التوزيع الطبيعي فيمكن مقارنتها بقيمة  $t$  عند مستوى معنوية  $\alpha$  ودرجات حرية  $\infty$ .

مثال ١٠-٢٠

فى تجربة كانت البيانات كالتالى:

العينة الأولى:  $n_1 = 15$  ،  $r_1 = 0.32$

العينة الثانية:  $n_2 = 35$  ،  $r_2 = 0.65$

اختبر ما إذا كان معاملا الارتباط للعينتين مسحوبين من نفس العشييرة التى معامل الارتباط بها لا يساوى صفراً.

بما أن  $r_1 = 0.32$  فإن  $Z_1 = 0.332$  ،

$r_2 = 0.65$  فإن  $Z_2 = 0.775$

والتباين لكل من  $Z_1$  و  $Z_2$  عبارة عن  $\sigma_{Z_1}^2 = \frac{1}{12}$  ،  $\sigma_{Z_2}^2 = \frac{1}{32}$

والانحراف المعياري للفرق بينهما من العلاقة (٦٣-١٠)

أى أن:

$$S_{Z_1-Z_2} = \sqrt{\frac{1}{12} + \frac{1}{32}} = 0.338$$

وبالتالى فإن

$$t = \frac{|0.332 - 0.775|}{0.338} = 1.31$$

وتقارن قيمة  $t$  المحسوبة بالقيمة  $t_{(\infty, 0.05)} = 1.96$  وبالتالى فإن  $Z_1$  و  $Z_2$  من نفس العشيرة وبالتالى فإن معاملى الارتباط هما لعينتين من نفس العشيرة.

١٦-١٠ اختبار تجانس عدد من معاملات الارتباط وتجميعهم فى قيمة واحدة كتقدير أثر صلاحية لمعامل ارتباط العشيرة

إذا كانت العينات مأخوذة من نفس العشيرة تحول قيم معاملات الارتباط للعينات لمختلفة إلى قيم  $Z$  المناظرة لها ثم يختبر الفرض بأن معاملات ارتباط العينات هي تقدير لنفس معامل ارتباط العشيرة باستخدام المعادلة التالية:

$$\sum W_i (Z_i - \bar{Z}_W)^2 \quad (٦٤-١٠)$$

التي تساوى

$$\sum W_i Z_i^2 - (\sum W_i Z_i)^2 / \sum W_i \quad (٦٥-١٠)$$

والتي تتوزع حسب توزيع مربع كاي  $\chi^2$  بدرجات حرية = (عدد معاملات ارتباط العينات - 1) حيث

$$W_i = 1/\sigma_{Z_i}^2 = (n_i - 3) \quad (٦٦-١٠)$$

أى أن  $W_i$  هي مقلوب تباين  $Z_i$  وتستخدم كنوع من عوامل الوزن weight حيث يعطى وزن أكبر للعينات كبيرة الحجم ووزن أقل للعينات الأقل حجماً. وإذا كانت قيمة مربع كاي غير معنوية فهذا يعنى أن قيم  $Z$  متجانسة وبالتالى فإن قيم  $r$  للعينات تكون مأخوذة من نفس العشيرة ويحسب المتوسط  $\bar{Z}_W$  حيث إنه يساوى مجموع قيم  $Z$  الموزونة بمقلوب تبايناتها على مجموع الأوزان أى أن:

$$\bar{Z}_W = \frac{\sum W_i Z_i}{\sum W_i} \quad (٦٧-١٠)$$

ثم تستخرج قيمة  $r$  المقابلة لقيمة  $\bar{Z}_W$  وتكون كتقدير لمعامل ارتباط العشيرة وتباين هذا المتوسط هو:

$$\sigma_{\bar{Z}_W}^2 = \frac{1}{\sum (n_i - 3)} \quad (٦٨-١٠)$$

وبذا يكون انحرافه المعياري

$$\sigma_{\bar{Z}_W} = \sqrt{\frac{1}{\sum (n_i - 3)}} \quad (٦٩-١٠)$$

ولقد وجد Fisher أن قيم  $Z$  تكون متحيزة biased فكل منها أكبر بهذه الكمية:

$$\frac{\rho}{2(n_i - 1)} \quad (٧٠-١٠)$$

ويظهر أثر هذا التحيز واضحاً عند الرغبة في الحصول على تقدير لمعامل الارتباط في العشيرة من عدد كبير من معاملات الارتباط للعينات المختلفة، حيث تتجمع هذه الكمية الصغيرة لكل ويظهر تأثيرها؛ ولذا يجب التصحيح لهذا التحيز بالعلاقة التالية:

$$\text{corrected } Z_i = \hat{Z}_i = Z_i - \left( \frac{\rho}{2(n_i - 1)} \right) \quad (٧١-١٠)$$

وإذا لم يكن معامل الارتباط في العشيرة معروفاً، وهذا ما يحدث غالباً، تستخدم  $r$  المقابلة  $\bar{Z}_W$  بدلاً من  $\rho$  ثم يحسب متوسط قيم  $Z$  المصححة والموزونة كما يلي:

$$\bar{Z}_W^\circ = \frac{\sum W_i Z_i^\circ}{\sum W_i} \quad (٧٢-١٠)$$

وتكون قيمة  $r$  المقابلة لقيمة  $\bar{Z}_W^\circ$  هي معامل الارتباط في العشيرة.

وإذا كان حجم العينات لا يختلف كثيراً فقد يهمل الوزن ويحسب المتوسط حيث يقسم مجموع قيم  $Z$  المقابلة لقيم  $r$  على عددها ثم تستخرج قيمة  $r$  المقابلة.

البيانات التالية مأخوذة من عشيرة ما:

العينة	معامل الارتباط	حجم العينة
١	0.62	30
٢	0.59	5
٣	0.65	12
٤	0.60	17

هل هذه العينات مأخوذة من عشيرة معامل الارتباط فيها وليكن  $\rho$ ، وما هو التقدير الأكثر صلاحية لهذا المعامل؟

العينة	معامل الارتباط $r_i$	حجم العينة	$Z_i$	الوزن $W_i = n_i - 3$	$W_i Z_i$	$W_i Z_i^2$
1	0.62	30	0.725	27	19.575	14.192
2	0.59	5	0.678	2	1.356	0.919
3	0.65	12	0.775	9	6.975	5.406
4	0.60	17	0.693	14	9.702	6.723
المجموع		64		52	37.608	27.240

ولاختبار فرض العدم بأن معاملات الارتباط في العينات المختلفة لها نفس معامل الارتباط في العشيرة فإن:

$$\chi^2 = \sum (n_i - 3) Z_i^2 - \frac{[\sum (n_i - 3) Z_i]^2}{\sum (n_i - 3)} \quad \text{حيث إن}$$

$$\text{وبالتعويض } \chi^2 = 27.240 - \frac{(37.608)^2}{52} = 0.04 \quad \text{بدرجات حرية } (4-1=3).$$

والقيمة الجدولية  $\chi^2_{(3,0.05)} = 7.81$ ، وعلى ذلك فإن  $\chi^2$  المحسوبة غير معنوية أي أن العينات التي قدر فيها معاملات الارتباط مأخوذة من نفس العشيرة. ويكون التقدير الأكثر صلاحية لمعامل الارتباط في العشيرة هو القيمة المقابلة  $\bar{Z}_W$ .

وحيث إن  $\bar{Z}_W = \frac{\sum W_i Z_i}{\sum W_i} = \frac{37.608}{52} = 0.72$  فإن قيمة  $r$  التى تقابلها هى 0.62.

ولتصحيح التحيز لتقدير معامل الارتباط فى العشيرة يستخدم التقدير 0.62 بدلاً من  $\rho$  فى الكمية  $\frac{\rho}{2(n_i - 1)}$  حيث إن معامل الارتباط فى العشيرة غير معروف كالتالى:

العينة	حجم العينة	الوزن $w_i = n_i - 3$	$r_i$	$Z_i$	$\frac{\rho}{2(n_i - 1)}$	$Z$ المصححة $Z^\circ$	$W_i Z_i^\circ$ المصححة والموزونة
١	30	27	0.62	0.725	0.011	0.714	19.278
٢	5	2	0.59	0.678	0.078	0.600	1.200
٣	12	9	0.65	0.775	0.028	0.747	6.723
٤	17	14	0.60	0.693	0.019	0.674	9.436
المجموع	64	52					36.637

وبالتالى فإن  $\bar{Z}_W = 36.637/52 = 0.704$  وقيمة  $r$  المقابلة هى 0.6 التى تعتبر تقديراً غير متحيزاً لمعامل الارتباط فى العشيرة.

### ١٠-١٧ السبب والأثر فى تحليل الارتباط والانحدار

#### Cause and effect in regression and correlation analysis

عند تفسير نتائج تحليل الانحدار والارتباط لابد أن يكون من الواضح تماماً أنه ليس بالضرورة أن نتائج التحليل تدل على السبب والأثر cause and effect، فعلى سبيل المثال فى إحدى الدراسات بلغ معامل الارتباط بين دخل الأسرة واستهلاك اللحوم فى فترة زمنية معينة 0.6، هذا ليس بالضرورة دليلاً على أن زيادة دخل الأسرة يؤدي إلى زيادة استهلاك اللحوم أو زيادة استهلاك اللحوم يؤدي إلى زيادة دخل الأسرة ولكن هذا المعامل يدل على أن المتغيرين يتغيران سوياً فى نفس الاتجاه وذلك قد يكون بسبب وجود متغير ثالث، وليكن ارتفاع مستوى المعيشة بصفة عامة، والذي يؤثر فى كل من الدخل واستهلاك اللحوم. وفى حالة ما إذا أمكن تثبيت هذا العامل

انماثل وتتعامل معه بطريقة إحصائية معينة فإنه في هذه الحالة فقط فإن معامل الارتباط قد يكون ذا مدلول معين. وهذا هو الهدف من دراسة الارتباط المتعدد أو الجزئي والذي سوف يتم تناوله فيما بعد. وبالمثل عند دراسة الانحدار البسيط فإن الحصول على معامل انحدار معنوي لا يدل أيضاً على السبب والأثر إلا إذا أخذت عوامل أخرى في الاعتبار.

وعلى الرغم من أن نتائج دراسة الانحدار والارتباط البسيط ليست دليلاً على السبب والأثر فإن تلك النتائج ربما تعطى بعض الاقتراحات التي قد تساعد في دراسة السبب والأثر فمثلاً وجد أن التدخين له علاقة ارتباط عالية بالإصابة بسرطان الرئة مما أوحى إلى ضرورة دراسة السبب والأثر بين المتغيرين لاستيضاح العلاقة انبيوكيميائية بينهما.

### ١٨-١٠ تباين الدالة الخطية Variance of a linear function

إذا كانت  $L_1 = Y_1 + Y_2$  فإن تباين  $L_1$  عبارة مجموع التباينين لكل من  $Y_1$  و  $Y_2$ .  
أي أن:

$$\sigma_{L_1}^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2 \quad (٧٣-١٠)$$

ذلك إذا كان  $Y_1$  و  $Y_2$  غير مرتبطين uncorrelated. وأيضاً فإن تباين الفرق بين هذين المتغيرين يساوي مجموع التباينين. أي أنه إذا كانت  $L_2 = Y_1 - Y_2$  فإن:

$$\sigma_{L_2}^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2 \quad (٧٤-١٠)$$

أي أن تباين الفرق يساوي تباين المجموع في حالة المتغيرات غير المرتبطة. ويمكن إثبات ذلك كما يلي:

حيث إن التباين لأي متغير وليكن  $L$  هو:

$$V(L) = \frac{\sum (L - \mu_L)^2}{N}$$

وحيث إن  $L = Y_1 + Y_2$  فإن  $\mu_L = \mu_{Y_1} + \mu_{Y_2}$  وبالتالي:

$$\begin{aligned}
 V(L) &= \frac{1}{N} \sum [Y_1 + Y_2 - (\mu_{Y_1} + \mu_{Y_2})]^2 \\
 &= \frac{1}{N} \sum [(Y_1 - \mu_{Y_1}) + (Y_2 - \mu_{Y_2})]^2 \\
 &= \frac{1}{N} [\sum (Y_1 - \mu_{Y_1})^2 + \sum (Y_2 - \mu_{Y_2})^2 + 2\sum (Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})] \\
 &= \frac{\sum (Y_1 - \mu_{Y_1})^2}{N} + \frac{\sum (Y_2 - \mu_{Y_2})^2}{N} + 2 \frac{\sum (Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})}{N}
 \end{aligned}$$

وبالتالى فإن

$$\sigma_L^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\text{cov}(y_1, y_2) \quad (٧٥-١٠)$$

وإذا كان المتغيران غير مرتبطين فإن الحد الأخير (التغاير covariance) يساوى صفراً.

وبالمثل يمكن إثبات أن تباين الفرق يساوى مجموع التباينين إذا كان المتغيران غير مرتبطين.

أما إذا كان المتغيران مرتبطين فإن تباين المجموع كما فى (٧٥-١٠) وحيث إن

$$\rho = \frac{\text{cov}(y_1, y_2)}{\sigma_{y_1} \sigma_{y_2}}$$

أى أن  $\text{cov}(y_1, y_2) = \rho \sigma_{y_1} \sigma_{y_2}$ ، وعلى ذلك تصبح المعادلة (٧٥-١٠) على الصورة التالية:

$$\sigma_{L_1}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\rho \sigma_{y_1} \sigma_{y_2} \quad (٧٦-١٠)$$

وعلى ذلك فإن الارتباط الموجب يزيد من تباين المجموع عن مجموع التباين بالمقدار  $2\rho \sigma_{y_1} \sigma_{y_2}$  والارتباط السالب يخفضه بنفس الكمية، أى

$$\sigma_{L_1}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 - 2\rho \sigma_{y_1} \sigma_{y_2} \quad (٧٧-١٠)$$

لاحظ هنا أن الارتباط الموجب ينقص تباين الفرق عن مجموع التباينين بينما الارتباط السالب يزيده عن مجموع التباينين. وفي تجارب الأزواج paired experiments، فإن تحليل t فإن الهدف الأساسي هو إيجاد الارتباط الموجب بين فردى الزوج وبالتالي ينقص تباين الفرق عن مجموع التباينين بالكمية  $2\rho\sigma_{Y_1}\sigma_{Y_2}$  وبنفس المفهوم يمكن إثبات أن:

$$\sigma_{(a_1Y_1+a_2Y_2)}^2 = a_1^2\sigma_{Y_1}^2 + a_2^2\sigma_{Y_2}^2 + 2\rho a_1 a_2 \sigma_{Y_1}\sigma_{Y_2} \quad (٧٨-١٠)$$

حيث إن  $a_1$  ،  $a_2$  ثابتان.

وللتعميم : إذا كانت  $L$  دالة خطية بالصورة  $L = a_1Y_1 + a_2Y_2 + \dots + a_nY_n$  فإن تباين  $L$  هو:

$$\begin{aligned} \sigma_L^2 &= a_1^2\sigma_{Y_1}^2 + a_2^2\sigma_{Y_2}^2 + \dots + 2a_1a_2\rho_{Y_1Y_2}\sigma_{Y_1}\sigma_{Y_2} \\ &+ \dots + 2a_{(n-1)}a_n\rho_{Y_{(n-1)}\cdot Y_n}\sigma_{Y_{(n-1)}}\sigma_{Y_n} \quad (٧٩-١٠) \\ &= \sum a_i^2\sigma_{Y_i}^2 + \sum \sum_{i \neq j} a_i a_j \rho_{Y_i Y_j} \sigma_{Y_i} \sigma_{Y_j} \end{aligned}$$

وإذا كانت المتغيرات غير مرتبطة فإن الحد الثاني من المعادلة في الطرف الأيمن يساوى صفراً وتصبح المعادلة:

$$\sigma_L^2 = \sum a_i^2\sigma_{Y_i}^2 \quad (٨٠-١٠)$$

وإذا كان هناك دالتان خطيتان في نفس المتغيرات، ولتكن الدالة الأولى  $L_3 = a_1Y_1 + a_2Y_2$  والدالة الثانية  $L_4 = b_1Y_1 + b_2Y_2$ ، فإن التباين بينهما:

$$\begin{aligned} \text{cov}(L_3, L_4) &= a_1b_1\sigma_{Y_1}^2 + a_2b_2\sigma_{Y_2}^2 \\ &+ (a_1b_2 + a_2b_1)\text{cov}(Y_1, Y_2) \quad (٨١-١٠) \end{aligned}$$

وللتعميم إذا كانت:

$$L_2 = b_1Y_1 + \dots + b_nY_n \quad , \quad L_1 = a_1Y_1 + \dots + a_nY_n$$

فإن:

$$\text{cov}(L_1, L_2) = \sum a_i b_i \sigma_i^2 + \sum_{i > j} (a_i b_j + a_j b_i) \text{cov}(Y_i, Y_j) \quad (٨٢-١٠)$$

### ١٠-١٩ ارتباط الرتب Rank correlation

في حالة المتغيرات التي لا تتبع توزيع معين أو التي تتبع توزيعاً غير معروف المعالم (الثوابت) فإن حساب  $r$  كتقدير لمعامل الارتباط في العشيرة يكون غير صالح وخاصة عندما يكون التوزيع ذو المتغيرين بعيداً عن التوزيع الطبيعي ولذلك فإن إجراء تحويل للمتغيرين أملاً في أن يكون توزيعهما المشترك وثيق الشبه للتوزيع الطبيعي ذي المتغيرين يجعل تقدير  $\rho$  ممكناً في بعض الأحيان وغير ممكن في البعض الآخر. ولكن هل المتغيران مرتبطان وهل يتغيران في نفس الاتجاه أم في اتجاهين متضادين؟ وقد تم فيما سبق إيضاح أنه في حالة اختبار فرض العدم بأن معامل الارتباط في العشيرة يساوي صفراً فإنه يمكن استخدام  $r$  على أن يتوزع أحد المتغيرين طبيعياً، أما عندما يكون توزيع المتغيرين غير طبيعي فإن أحسن طريقة للإجابة على السؤال السابق هو ترتيب كل من المتغيرين (تتازلياً وتصاعدياً) ثم يختبر التوافق بين الترتيبين. وفي حالة البيانات غير المرتبة فإن أول خطوة هي ترتيب كل من المتغيرين كل منهما على حده ويستخدم معامل ارتباط الرتب  $r_s$  والذي وضعه Spearman حيث:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (٨٣-١٠)$$

حيث  $d_i = Y_{1i} - Y_{2i}$  ،  $Y_{1i}$  ،  $Y_{2i}$  هما رتبة الفرد  $i$  في كل من المتغير الأول والثاني على التوالي. فإذا كانت  $\sum d_i^2$  تساوي صفراً فإن  $r_s = \pm 1$ .

ومعامل ارتباط الرتب تنحصر قيمته بين  $-1$  في حالة عدم التوافق المطلق و  $+1$  في حالة التوافق التام complete concordance. ولاختبار معنوية معامل ارتباط الرتب للعينات التي حجمها عشرة أزواج أو أقل يستخدم الحد الأدنى لقيمة  $r_s$  المعنوية على مستوى 5% ومستوى 1% كما وصفه Kendall (1970) في جدول ٧-١٠، حيث إن توزيع  $r$  وتوزيع  $r_s$  متساويان في حالة اختبار فرض العدم  $r = 0$ .

ويلاحظ أنه عندما يكون حجم العينة 4 مثلاً فإن الحد الأدنى لمعنوية  $r_s$  يكون أكبر من الواحد الصحيح عند مستوى 5% وأيضاً 1% وبالتالي فإن معنوية معامل

ارتباط الرتب في مثل هذه الحالة لن تتحقق حيث إن الحد الأعلى له يساوى الواحد الصحيح.

ويستخدم معامل ارتباط الرتب في الحالات التي يتعذر فيها قياس المتغيرات بطريقة كمية وإنما تعطى رتباً أو درجات، ولقد استنبط Kendall مقياساً آخر لدرجة التوافق بين المتغيرين لن يتم تناوله هنا في الوقت الحالي.

جدول ١٠-٧ معنوية معامل ارتباط الرتب للعينات التي حجمها عشرة أزواج أو أقل

حجم العينة	الحد الأدنى لمعنوية $r_s$	
	5%	1%
٤ أو أقل	-	-
٥	1.000	-
٦	0.886	1.000
٧	0.786	0.929
٨	0.738	0.857
٩	0.683	0.817
١٠	0.648	0.781
١١ أو أكثر	يستخدم جدول ٧ ملحق أ بدرجات حرية $n-2$	

مثال ١٠-٢٢

حساب معامل ارتباط الرتب واختبار معنويته.

رقم البقرة	الترتيب حسب كمية اللبن	الترتيب حسب المطابقة للنموذج	الفرق $d_i$	$d_i^2$
١	3	2	1	1
٢	4	3	1	1
٣	1	1	0	0
٤	2	5	-3	9
٥	5	4	1	1
المجموع			0	12

بتطبيق المعادلة (١٠-٨٣) فإن

$$r_s = 1 - \frac{(6)(12)}{(5)(25-1)} = 0.4$$

ومن القيمة الجدولية (حجم العينة = 5) يتضح أن معامل ارتباط الرتب غير معنوى أى ليس هناك علاقة بين الترتيبين.

### تمارين الباب العاشر

١-١٠ في عينة ما وجدت القيم التالية للعلاقة بين الوزن  $Y_2$  بالكيلوجرام والارتفاع  $Y_1$  بالسنتيمتر:

$$\begin{aligned} n &= 12 & \sum Y_1 &= 228 & \sum Y_2 &= 450 \\ \sum y_1 y_2 &= -542.5 & \sum y_1^2 &= 961 & \sum y_2^2 &= 1225 \end{aligned}$$

المطلوب حساب:

- ١ - معامل انحدار الوزن على الارتفاع ومعامل انحدار الارتفاع على الوزن.
- ٢ - معامل الارتباط.
- ٣ - تحليل التباين في الوزن إلى مكوناته.
- ٤ - إيجاد قيمة  $\hat{Y}_2$  عندما تكون  $Y_1$  تساوى 22, 19, 30 سنتيمتر.
- ٥ - تمثيل العلاقة بين المتغيرين بيانياً.

٢-١٠ في التمرين السابق اختبر الفرض بأن معامل انحدار الوزن على الارتفاع يساوى صفراً بمستوى معنوية 1% .

٣-١٠ في التمرين ١-١٠ اختبر الفروض الآتية:

$$H_0 : \rho = 0 \quad - 1$$

$$H_0 : \rho = -3 \quad - 2$$

٤-١٠ في تجربة نسمين كان يتم وزن 20 حمل أسبوعياً فإذا كان رمز الوزن  $Y$  والأسابيع  $X$  ونوافرت البيانات التالية:

$$\begin{aligned} \sum Y &= 400 & \sum Y^2 &= 9600 & \sum XY &= 5160 \\ \sum X &= 240 & \sum X^2 &= 3024 \end{aligned}$$

المطلوب حساب:

- ١ - معامل الانحدار  $b_{yx}$  ومعامل الارتباط بين المتغيرين.

٢- معامل التحديد.

٣- اختبار فرض العدم  $H_0: \beta = 0$  ،  $H_0: \rho = 0$  . وضع جدول تحليل التباين.

٤- اكتب معادلة التنبؤ وارسم خط الانحدار رسماً دقيقاً.

١٠-٥ ما هي حدود الثقة لكل من الانحدار والجزء المقطوع من محور الصادات في تمرين ١٠-٤ وما هي حدود الثقة لخط الانحدار؟ بين ذلك بيانياً.

١٠-٦ إذا توافرت البيانات التالية:

العينة الأولى : معامل الارتباط = 0.325	حجم العينة = 25
العينة الثانية : معامل الارتباط = 0.34	حجم العينة = 15
العينة الثالثة : معامل الارتباط = 0.31	حجم العينة = 22

هل هذه العينات مسحوبة من عشيرة واحدة وما هو تقديرك لمعامل الارتباط؟  
أحسب الاختبارات الإحصائية اللازمة.

١٠-٧ إذا كانت البيانات التالية مسحوبة من عشيرة طبيعية ذات متغيرين

2.3	2.1	2.1	1.8	1.7	1.1	1.0	0.1	$Y_1$
2.3	2.8	2.8	1.8	3.0	1.5	2.7	0.7	$Y_2$
6.1	4.9	4.4	4.4	3.4	2.9	2.7	2.6	$Y_1$
6.6	4.0	3.8	2.7	3.2	2.6	4.4	2.7	$Y_2$

وكان معامل ارتباط العشيرة يساوى 0.7 ومعامل اعتماد العشيرة  $Y_2$  على  $Y_1$  يساوى 1

المطلوب حساب:

١- معامل الارتباط بين المتغيرين ومعامل اعتماد  $Y_2$  على  $Y_1$  وارسم خط الانحدار رسماً دقيقاً.

٢- معامل التحديد.

٣- اختر أقل أربع قيم للمتغير  $Y_1$  وقيم  $Y_2$  المقابلة لها وأيضاً أكبر أربع قيم للمتغير  $Y_1$  وقيم  $Y_2$  المقابلة لتكوين عينة من 8 أزواج لهذه العينة. احسب المطلوب في ١- وقارن النتائج المتحصل عليها.

٤- قسم كل من قيم  $Y_2$  إلى مكوناتها ثم احسب  $\sum e_1 Y_1$  ،  $\sum e_1 Y_2$  وقارن بين  $\sum e_1^2$  ،  $\sum e_1 Y_1$  . اشرح معنى هذه النتائج.

$$١٠-٨ \text{ بين كيف أن تباين معامل الارتباط } = \frac{1-r^2}{n-2}$$

حيث  $r$  معامل الارتباط و  $n$  حجم العينة (عدد أزواج المشاهدات).

١٠-٩ قام اثنان من المحكمين بترتيب 10 بقرات حسب حالتها الجسمية كما يلي:

رقم البقرة	المحكم الأول	المحكم الثاني
١	8	10
٢	9	8
٣	7	9
٤	5	3
٥	1	4
٦	4	2
٧	2	1
٨	3	5
٩	6	6
١٠	10	7

احسب معامل ارتباط الرتب واختبر معنويته.