١٤ الانحدار الخطى المتعدد

Multiple linear regression

١ – مقدمة

٢ – الانحدار المتعدد فى حالة متغيرين مستقلين

٣ – حدود الثقة لمعاملى الانحدار β_1 ، β_2

٤ – تقسيم الاختلافات فى Y إلى مكوناتها

٥ – تحليل التباين واختبار F

٦ – اختيار أفضل معادلة انحدار

٧ – الارتباط المتعدد والجزئى

١٤-١ مقدمة

عند دراسة الانحدار الخطى البسيط كان هناك متغير مستقل واحد (X) يؤثر على متغير تابع (Y)، ولكن كثيراً ما تكون هناك عدة متغيرات مستقلة تؤثر على المتغير التابع. فعند دراسة العوامل التى تؤثر على كمية اللبن مثلاً أو وزن الحملان عند الفطام أو غيرهما يكون هناك عوامل كثيرة وليس عاملاً واحداً. فإذا ما أخذ فى الاعتبار متغير مستقل واحد فقط وأهملت باقى المتغيرات المؤثرة المستقلة فهذا من شأنه أن يؤدى إلى زيادة مكون الخطأ مما قد يؤدى إلى نتائج بعيدة عن الدقة حيث تفقد اختبارات المعنوية حساسيتها وذلك راجع إلى زيادة خطأ التقدير $S_{y.x}$ السابق الإشارة إليه فى الباب العاشر. وحيث أن اختبار t يعبر عن النسبة بين معامل الانحدار (b) إلى الانحراف المعيارى له $[(S_{y.x})/(\sqrt{\sum x^2})]$، والتى تقل بزيادة الانحراف المعيارى، مما قد يؤدى إلى نتائج غير معنوية. فى حين أنها قد تكون معنوية إذا لم يتم إغفال المتغيرات الأخرى. إلا أن هناك نقطة أخرى يجب وضعها فى الاعتبار وهى درجات الحرية، فزيادة عدد المتغيرات المستقلة يؤدى إلى نقص فى درجات حرية الخطأ وهذا أيضاً يؤدى إلى نقص فى دقة الاختبار إذا ما كانت بعض هذه المعاملات غير معنوية. والتركيز فى هذا الباب سوف يكون على دراسة الانحدار والارتباط المتعددين اللذين يختصان بوصف العلاقات الخطية بين أكثر من متغيرين لتقدير التأثير المتحد combined effect لعدة متغيرات مستقلة على متغير واحد، ودراسة إمكانية التنبؤ بواسطة معادلة الانحدار التى تعطى أدق قيمة للمتغير التابع (Y) بدلالة المتغيرات المستقلة المعروفة، مع إهمال بعض العوامل المستقلة ذات التأثير الضئيل وغير المعنوى والتى بإهمالها لن تتأثر نتيجة التنبؤ. وكذلك دراسة إمكانية ترتيب العوامل المستقلة حسب أهميتها بناءاً على اختبارات المعنوية.

وسوف تتناول الدراسة فى هذا الباب العلاقة بين ثلاث متغيرات فقط (اثنين مستقلين وآخر تابع) وذلك بهدف إعطاء نبذة مختصرة عن طريقة التقدير ومفهوم كل من معاملات الانحدار المتعددة والجزئية وكذلك كل من معاملات الارتباط الكلية والجزئية واختبارات المعنوية، علما بأنه فى حالة أكثر من ثلاث متغيرات فليس هناك من مبادئ جديدة ولكنها امتداد لها.

١٤-٢ الانحدار المتعدد فى حالة متغيرين مستقلين

يمكن التعبير عن المتغير التابع Y الذى يؤثر فيه متغيرين مستقلين X_1، X_2 بالنموذج التالى:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (١٤-١)$$

حيث β_1 تمثل معامل الانحدار (الاعتماد) الجزئى partial regression coefficient للمتغير Y على المتغير X_1، أى متوسط مقدار التغير فى Y عندما تتغير X_1 بمقدار الوحدة مع ثبات X_2. وتمثل β_2 معامل الانحدار الجزئى للمتغير Y على المتغير X_2، أى متوسط مقدار التغير فى Y عندما تتغير X_2 بمقدار الوحدة مع ثبات X_1. أما بقية مكونات النموذج فهى كما سبق تعريفها فى الباب العاشر.

ويعرف الانحدار المتعدد multiple regression بأنه متوسط التغير فى Y بمقدار β_1، β_2 نتيجة تغير كل من X_1، X_2 بمقدار الوحدة، على الترتيب.

ولتقدير معالم (ثوابت) parameters هذا النموذج (١٤-١) يستخدم النموذج التقديرى التالى:

$$\hat{Y} = a + b_1X_1 + b_2X_2 \qquad (١٤\text{-}٢)$$

حيث a، b_1، b_2 تقديرات غير متحيزة لكل من α، β_1، β_2 على الترتيب. ويستخدم فى ذلك طريقة المربعات الصغرى التى سبق الإشارة إليها وبالتالى:

$$Y = a + b_1X_1 + b_2X_2 + e \qquad (١٤\text{-}٣)$$

١٤-٢-١ طريقة المربعات الصغرى Method of least squares

من النموذج (١٤-١):

$$\varepsilon = (Y - \alpha - \beta_1X_1 - \beta_2X_2)$$

$$\varepsilon^2 = \sum(Y - \alpha - \beta_1X_1 - \beta_2X_2)^2$$

$$Q = \sum\varepsilon^2 = \sum(Y - \alpha - \beta_1X_1 - \beta_2X_2)^2 \qquad (١٤\text{-}٤)$$

حيث $\sum\in^2$ تعبر عن مجموع مربعات الخطأ، أى مجموع مربع الانحرافات عن خط الانحدار. وتقدر α، β_1، β_2 بحيث يتحقق الشرطان التاليان:

١ - مجموع الأخطاء يساوى صفر، أى $\sum\varepsilon = 0$

٢- مجموع مربعات الخطأ $\sum\in^2$ أقل ما يمكن، ويكون ذلك باستخدام التفاضل الجزئى للمقدار $\sum\in^2$ بالنسبة لكل من α، β_1، β_2 كل على حدة، وفى كل مرة توضع نتيجة التفاضل مساوية للصفر كما يلى:

$$\frac{\partial Q}{\partial \alpha}\Big|_{a, b_1, b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-1) = 0$$

$$\frac{\partial Q}{\partial \beta_1}\Big|_{a,b_1,b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-X_1) = 0$$

$$\frac{\partial Q}{\partial \beta_2}\Big|_{a,b_1,b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-X_2) = 0$$

وتتكون مجموعة المعادلات التالية:

$$\begin{aligned}\sum Y &= na + b_1\sum X_1 + b_2\sum X_2\\ \sum X_1Y &= a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2\\ \sum X_2Y &= a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2\end{aligned} \qquad (١٤-٥)$$

وتسمى هذه المجموعة من المعادلات بالمعادلات الاعتيادية أو الطبيعية normal equations. وبحل هذه المعادلات آنياً يمكن الحصول على تقدير لكل من α، β_1، β_2 كالتالى:

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 \qquad (١٤-٦)$$

وبالتعبير عن X، Y كانحرافات عن متوسطاتهما فإن:

$$b_1 = \frac{\sum x_1y\sum x_2^2 - \sum x_2y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} \qquad (١٤-٧)$$

$$b_2 = \frac{\sum x_2y\sum x_1^2 - \sum x_1y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} \qquad (١٤-٨)$$

ويمكن إعادة كتابة المعادلة (١٤-٢) باستخدام جبر المصفوفات كالتالى:

$$Y = X\beta + \varepsilon \qquad (١٤-٩)$$

حيث:

$\varepsilon' = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]$ ، $Y' = [Y_1 \quad Y_2 \quad \cdots \quad Y_n]$ ،

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} \end{bmatrix} ،\beta' = [\alpha \quad \beta_1 \quad \beta_2]$$

ومن (١٤-٩):

$$Q = \sum \varepsilon^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$
$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

حيث إن $Y'X\beta = \beta'X'Y$ لأن كل منهما عبارة عن مصفوفة $1x1$.

وباستخدام طريقة المربعات الصغرى فإن:

$$\frac{\partial Q}{\partial \beta_2}\Big|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

ومنها:

$$X'X\hat{\beta} = X'Y$$

وبالتالى:

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (١٤-١٠)$$

حيث إن $(X'X)^{-1}$ عبارة عن مقلوب $(X'X)$ كما سبق شرحه فى الباب العاشر.

١-٢-٢ الافتراضات الخاصة بدراسة العلاقة بين أكثر من متغيرين:

١ - المتغيرات المستقلة مقاسة بدون خطأ وليس لها توزيعات احتمالية وعددها يكون أقل من حجم العينة (عدد المشاهدات).

٢ - الأخطاء فى قيم المتغير التابع مستقلة وتتوزع طبيعياً بمتوسط يساوى الصفر وتباين يساوى σ_e^2. وهذا الافتراض مهم وضرورى فى حالات اختبارات المعنوية.

٣ - لا يوجد أى ارتباط بين أى من الأخطاء وكل من المتغيرات المستقلة. أى أن $\operatorname{cov}(X_i, e_j) = 0$ لكل من i ، j.

مثال ١٤-١

من البيانات التالية أحسب معادلة الانحدار المتعدد.

4	7	8	6	2	:Y
5	4	3	2	0	$:X_1$
4	3	0	1	2	$:X_2$

من البيانات السابقة يمكن حساب ما يلى:

$\sum X_1 = 14$ $\sum X_2 = 10$ $\sum Y = 27$

$\sum X_1^2 = 54$ $\sum X_2^2 = 30$ $\sum Y^2 = 169$

$\sum x_1^2 = 14.8$ $\sum x_2^2 = 10$ $\sum y^2 = 23.2$

$\sum X_1 X_2 = 34$ $\sum X_1 Y = 84$ $\sum X_2 Y = 47$

وبالتعويض فى المعادلات الاعتيادية فى (١٤-٥):

$$5a + 14b_1 + 10b_2 = 27$$

$$14a + 54b_1 + 34b_2 = 84$$

$$10a + 34b_1 + 30b_2 = 47$$

ومن هذه المعادلات:

$a = 5$

$b_1 = 1.125$ وحدة من Y لكل وحدة من X_1 باعتبار X_2 ثابتة.

$b_2 = -1.375$ وحدة من Y لكل وحدة من X_2 باعتبار X_1 ثابتة.

وبالتالى فإن معادلة التنبؤ:

$$\hat{Y} = 5 + 1.125X_1 - 1.375X_2$$

وإذا اعتبر أن b_1، b_2 معاملا انحدار بسيط فإن معامل انحدار Y على X_1

$$b_{y.x_1}=\frac{\Sigma X_1 Y-\frac{\Sigma X_1 \Sigma Y}{n}}{\Sigma X_1^2-\frac{(\Sigma X_1)^2}{n}}=\frac{84-\frac{(14)(27)}{5}}{54-\frac{(14)^2}{5}}=0.57$$

أى 0.57 وحدة من Y لكل وحدة من X_1. بينما معامل الانحدار الجزئى بين Y و X_1 هو 1.125 مع تثبيت X_2.

وبنفس الطريقة معامل انحدار Y على X_2:

$$b_{y.x_2}=\frac{\Sigma X_2 Y-\frac{\Sigma X_2 \Sigma Y}{n}}{\Sigma X_2^2-\frac{(\Sigma X_2)^2}{n}}=\frac{47-\frac{(10)(27)}{5}}{30-\frac{(10)^2}{5}}=-0.7$$

أى 0.7- وحدة من Y لكل وحدة من X_2. بينما معامل الانحدار الجزئى بين Y، X_2 باعتبار X_1 ثابتة هو 1.375-. ومن ذلك يتضح أن معاملات الانحدار البسيطة مختلفة القيمة عن معاملات الانحدار الجزئية.

١٤-٣ حدود الثقة لمعاملى الانحدار β_1 ، β_2

حدى الثقة لـــ β_1 هما:

$$b_1 \pm tS_{b_1} \qquad (١٤-١١)$$

حدى الثقة لـــ β_2 هما:

$$b_2 \pm tS_{b2} \qquad (١٤-١٢)$$

حيث:

$$S_{b_1}=\sqrt{\frac{S^2\sum x_2^2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2}} \text{ ، } S_{b_2}=\sqrt{\frac{S^2\sum x_1^2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2}} \text{ ،}$$

S_b، S_{b_2} هما الانحراف القياسى standard error لكل من معاملى الانحدار b_1، b_2 على الترتيب، t هى قيمة t الجدولية بدرجات حرية $(n-k-1)$ حيث n هى

عدد المشاهدات، k عدد معاملات الانحدار المقدرة ومستوى معنوية وليكن α، S^2 هى متوسط مربعات الخطأ.

١٤-٤ تقسيم الاختلافات فى Y إلى مكوناتها

حيث إن:

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

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

فإنه يمكن أيضاً التعبير عن Y بعد التعويض عن a بقيمتها حيث:

$$Y = \overline{Y} + b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + e$$

وأيضا فإن:

$$\hat{Y} = \overline{Y} + b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2)$$

وعلى ذلك فإن:

$$Y - \overline{Y} = b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + e$$

$$\hat{Y} - \overline{Y} = b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2)$$

حيث إن:

$(Y - \overline{Y})$ عبارة عن انحراف قيمة Y عن متوسطها الحسابى.

$(\hat{Y} - \overline{Y})$ هو انحراف القيمة المتوقعة المقابلة لقيمة معينة لـــ X عن المتوسط الحسابى للمتغير Y.

$(Y - \hat{Y})$ هو انحراف القيمة Y عن القيمة المتوقعة.

أى أنه يمكن تقسيم انحراف القيمة عن متوسطها إلى مكونين أحدهما راجع إلى انحدار (اعتماد) Y على كل من X_1، X_2 والذى يطلق عليه الجزء الراجع للانحدار due to regression، والمكون الأخر راجع إلى الانحراف عن خط الانحدار والذى يطلق عليه عن خط الانحدار from regression، ويعبر عن $(Y - \hat{Y})$ أو (e) أو $(d_{y \cdot x_1 x_2})$.

مثال ١٤-٢

قسم الاختلافات فى Y إلى مكوناتها فى المثال ١٤-١

	Y	X_1	X_2	$\hat{Y}$	$(Y-\overline{Y})$	$(Y-\hat{Y})$	$(\hat{Y}-\overline{Y})$
	2	0	2	2.250	-3.4	-0.250	-3.150
	6	2	1	5.875	0.6	0.125	0.475
	8	3	0	8.375	2.6	-0.375	2.975
	7	4	3	5.375	1.6	1.625	-0.025
	4	5	4	5.125	-1.4	-1.125	-0.275
المجموع	**27**	**14**	**10**	**27.000**	**0**	**0**	**0**

١٤-٥ تحليل التباين واختبار F

تعتبر الكمية $\sum(Y-\hat{Y})^2/(n-k-1)$ (والتى يرمز لها بالرمز S^2 ويعبر عنها بمتوسط مربعات الخطأ أو بمتوسط المربعات الراجع إلى الانحراف عن خط الانحدار) تقديرا غير متحيز لـ σ_e^2 وذلك فى الانحدار المتعدد حيث يقسم مجموع مربعات احرافات قيم المتغير التابع Y عن متوسطها إلى مكونين:

١ - مجموع المربعات الراجع إلى خط الانحدار SS due to regression

٢ - مجموع المربعات عن خط الانحدار SS from regression

وتبين المعادلة التالية ذلك حيث إن:

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

ومنها:

$$\begin{aligned}\sum(Y-\overline{Y})^2 &= \sum(\hat{Y}-\overline{Y})^2 + \sum(Y-\hat{Y})^2 + 2\sum(\hat{Y}-\overline{Y})(Y-\hat{Y}) \\ &= \sum(\hat{Y}-\overline{Y})^2 + \sum(Y-\hat{Y})^2\end{aligned}$$

حيث إن الحد الأخير $[2\sum(\hat{Y}-\overline{Y})(Y-\hat{Y})]$ يساوى الصفر.

أى أن:

$$TSS = RSS + ESS$$

حيث:

$$TSS = \sum y^2 = \sum(Y-\overline{Y})^2$$

$$RSS = \sum(\hat{Y}-\overline{Y})^2 = \sum\hat{y}^2 = b_1\sum x_1y + b_2\sum x_2y$$

$$ESS = \sum(Y-\hat{Y})^2 = \sum y^2 - RSS$$

مثال ١٤-٣

قسم مجموع المربعات فى Y إلى مكوناته فى مثال ١٤-٢.

مجموع المربعات الكلى TSS:

$$\sum y^2 = 169 - \frac{(27)^2}{5} = 23.2$$

ومجموع المربعات الراجع إلى خط الانحدار RSS:

$$\begin{aligned} RSS &= \sum(\hat{Y} - \overline{Y})^2 \\ &= b_1 \sum x_1 y + b_2 \sum x_2 y \\ &= b_1 \left[\sum X_1 Y - \frac{\sum X_1 \sum Y}{n}\right] + b_2 \left[\sum X_2 Y - \frac{\sum X_2 \sum Y}{n}\right] \\ &= 0.57[84 - (14)(27)/5] - 0.7[47 - (10)(27)/5] = 19.075 \end{aligned}$$

مجموع المربعات الراجع إلى الانحراف عن خط الانحدار ESS:

$$ESS = 23.2 - 19.075 = 4.125$$

وكما كان الحال فى الاعتماد البسيط فإنه يمكن الحصول على ESS أيضاً بالحساب حيث إن: $ESS = \sum(Y - \hat{Y})^2$، ومن مثال ١٤-٢ فإن:

$$ESS = (-0.25)^2 + (0.125)^2 + \cdots + (-1.125)^2 = 4.125$$

وهى نفس النتيجة التى تم الحصول عليها بالطرح دون الرجوع إلى تقدير القيم المتوقعة ثم حساب انحرافات القيم الفعلية عن قيمها المتوقعة والتربيع والجمع على كل مفردات العينة.

ولاختبار معنوية الانحدار المتعدد باستخدام اختبار F يستخدم الجدول التالى:

SOV	df	SS	MS	F
راجع إلى الانحدار Due to reg.	k	RSS	RSS/k	$\frac{RSS/k}{S^2}$
عن خط الانحدار From reg.	$n-k-1$	ESS	$S^2 = ESS/(n-k-1)$	
Total الكلى	$n-1$	$\sum y^2$		

وقيمة F المحسوبة تقارن بقيمة F الجدولية بدرجات حرية (k)، $(n-k-1)$ ومستوى معنوية α، وفى هذه الحالة يعتبر اختبار مشترك join test يختبر معنوية معاملات الانحدار معاً وبعد ذلك يمكن تحديد أى معاملات الانحدار هذه تختلف معنوياً عن الصفر.

مثال ١٤-٤

فى مثال ١٤-١ كون جدول تحليل التباين واختبر الفرض $H_\circ : \beta_1 = \beta_2 = 0$.

SOV	df	SS	MS	F
راجع إلى الانحدار	2	19.075	9.5375	4.62^{ns}
عن خط الانحدار	2	4.125	2.0625	
الكلى	**4**	**23.2**		

^{ns} (P> 0.05)

وقيمة F المحسوبة (4.62) أقل من قيمة F الجدولية بدرجات حرية 2، 2 ومستوى معنوية 5% وبالتالى لا يكون هناك مبرراً لرفض فرض العدم أى أن العلاقة بين Y وكل من X_1، X_2 معاً غير معنوية.

ولاختبار معنوية أثر إضافة متغير ثالث عند دراسة العلاقة بين متغيرين حتى يمكن إهماله أو إضافته إلى النموذج فإنه يمكن إجراء الاختبار التالى:

حيث أن النموذج فى حالة العلاقة بين متغيرين اثنين فقط:

$$\hat{Y} = a' + b_1' X_1$$

والنموذج فى حالة العلاقة بين ٣ متغيرات:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

ويحسب مجموع المربعات الراجع للانحدار RSS فى كل من النموذجين والفرق بنهما يكون راجعاً إلى إدخال المتغير الثالث إلى النموذج وله درجة حرية واحدة. وعلى ذلك فمتوسط مربعاته هو نفسه مجموع مربعاته والذى يقسم على متوسط مربعات الخطأ وذلك لاختبار معنوية معامل الانحدار الجزئى Y على X_2 ويمكن توضيح ذلك بالمثال التالى.

مثال ١٤-٥

إذا كان مجموع المربعات الراجعة للخطأ = 3، $n = 50$، $\sum x_1 y = 2$، $\sum x_1^2 = 16$ ، متوسط المربعات الراجعة للانحدار = 0.5. كون جدول تحليل التباين واختبر فرض العدم $H_o : \beta_1 = \beta_2 = 0$، ثم اختبر إذا كانت β_2 معنوية أم لا.

SOV	df	SS	MS	F
راجع إلى الانحدار	2	1	0.5	8.33^{**}
عن خط الانحدار	47	3	0.06	

وحيث إن قيمة F المحسوبة معنوية بدرجة ثقة 99% فهذا يعنى أن العلاقة بين المتغيرات، العلاقة بين Y من ناحية و X_1، X_2 معاً من ناحية أخرى، معنوية إحصائياً وليست راجعة إلى الصدفة أو الأخطاء العشوائية.

ولاختبار فرض العدم $H_o : \beta_2 = 0$ يحسب مجموع المربعات الراجعة للانحدار Y على X_1، X_2 معاً ثم يحسب مجموع المربعات الراجعة للانحدار Y على X_1 وبالطرح يمكن الحصول على مجموع المربعات الراجعة نتيجة إضافة المتغير X_2 حيث $\sum(x_1 y)^2 / \sum x_1^2 = (2)^2 / 16 = 0.25$، وبالتالى فإن مجموع المربعات نتيجة إضافة X_2 إلى النموذج $1 - 0.25 = 0.75$، ويكون جدول تحليل التباين كالتالى:

SOV	df	SS	MS	F
الراجع للانحدار على X_1، X_2	2	1		
الراجع للانحدار على X_1 بغض النظر عن X_2	1	0.25		
الراجع للانحدار نتيجة لإضافة X_2	1	0.75	0.75	$\frac{0.75}{0.06} = 12.25$
عن خط الانحدار	47	3	0.06	

وعلى ذلك يرفض فرض العدم الخاص بعدم وجود علاقة بين المتغيرين Y، X_2 فى وجود X_1 ، أى أنه توجد علاقة معنوية بينهما بدرجة ثقة 99% وبالتالى فإنه لا يمكن إغفال إضافة المتغير X_2 إلى النموذج الإحصائى.

مثال ١٤-٦

حل مثال١٤-١ باستخدام برنامج SAS.

```
DATA MULTREG;
INPUT X1 X2 Y @@;
CARDS;
0 2 2 2 1 6 3 0 8
4 3 7 5 4 4
PROC REG;
MODEL Y = X1 X2;
RUN;
```

نتائج التحليل:

Model: MODEL1

Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	19.07500	9.53750	4.624	0.1778
Error	2	4.12500	2.06250		
C Total	4	23.20000			

Root MSE	1.43614	R-square	0.8222
Dep Mean	5.40000	Adj R-sq	0.6444
C.V.	26.59520		

Parameter Estimates

Parameter Variable	DF	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	5.000000	1.30725995	3.825	0.0621
X1	1	1.125000	0.42912910	2.622	0.1199
X2	1	-1.375000	0.52205808	-2.634	0.1190

وتوجد عدة طرق تمكن من تقرير إبقاء أو استبعاد متغير مستقل معين من معادلة الانحدار المتعدد. ومن هذه الطرق طريقة الاستبعاد الخلفى backward elimination، طريقة الاختيار الأمامى forward selection وطريقة الخطوة خطوة stepwise procedure. وفى جميع هذه الطرق يلزم استخدام أى من برامج التحليل الإحصائى مثل SAS أو غيره حيث إنه من الصعب تنفيذ هذه الطرق باستخدام الآلات الحاسبة العادية.

١٤-٦ اختيار أفضل معادلة انحدار Selecting the best regression equation

لجعل معادلة الانحدار المتعدد أكثر فائدة فى أغراض التنبؤ بسلوك المتغير التابع عند تغير أى من المتغيرات المستقلة فإنه لابد من أن تتضمن معادلة الانحدار كل المتغيرات المستقلة الممكنة والتى يمكن أن تؤثر على المتغير التابع. فلو فرض وجود متغير تابع وعدد 10 متغيرات مستقلة معنى ذلك أنه يمكن تكوين عدد $[{}_{10}C_1+{}_{10}C_2+{}_{10}C_3+\cdots+{}_{10}C_{10}]$ أى 1023 معادلة انحدار مختلفة، وهذا يجعل من الصعب مقارنة هذه المعادلات ببعضها البعض، هذا بالإضافة إلى تكلفة الحصول على المعلومات الخاصة بكل المتغيرات المستقلة. وبالتالى فإن المتغيرات ذات التأثير المعنوى فقط هى التى يمكن أن تتضمنها معادلة الانحدار وبالتالى اختيار أفضل معادلة انحدار.

١٤-٦-١ طريقة الاستبعاد الخلفى Backward elimination method

يتم فى هذه الطريقة وضع جميع المتغيرات المستقلة فى معادلة الانحدار ثم محاولة اختيار أفضل معادلة انحدار تحتوى على أقل عدد ممكن من المتغيرات المستقلة. وتجرى هذه الطريقة كالتالى:

١- تقدير معادلة انحدار تتضمن كل المتغيرات المستقلة.

٢- حساب قيمة F الجزئية (partial F) لكل متغير اشتملت عليه معادلة الانحدار والتى تعتمد فى حسابها على متوسط مجموع المربعات الجزئى partial mean square.

٣- تقارن أقل قيمة من قيم F الجزئية (بفرض أنها F_L) بقيمة F الجدولية (بفرض أنها F_0) بمستوى معنوية سبق اختياره. وبالتالى:

أ - إذا كانت قيمة F_L أصغر من قيمة F_0 يتم إلغاء المتغير المستقل قرينها. ثم يعاد حساب معادلة الانحدار مرة أخرى بدون هذا المتغير وتكرر نفس الخطوات مرة أخرى.

ب- إذا كانت قيمة F_L أكبر من أو تساوى قيمة F_0 فإن معادلة الانحدار فى هذه الحالة تمثل المعادلة النهائية.

١٤-٦-٢ طريقة الاختيار الأمامي forward selection method

فى هذه الطريقة يتم إضافة المتغيرات المستقلة الواحد تلو الآخر كالتالى:

١- تحسب قيم معاملات الارتباط البسيط بين المتغير التابع وكل من المتغيرات المستقلة. يتم اختيار أول متغير مستقل وهو المتغير الذى له أعلى معامل ارتباط بسيط مع المتغير التابع (افترض أنه X_H) بمستوى معنوية p_H، يقارن مستوى المعنوية هذا مع مستوى معنوية سبق تحديده وليكن SLE (significant level of entry) وبالتالى:

أ - إذا كانت قيمة p_H أكبر من قيمة SLE فإن المتغير X_H قرين هذه القيمة لا يمكن وضعة فى معادلة الانحدار وبالتالى لا يمكن حساب معادلة الانحدار فى هذه الحالة.

ب- إذا كانت قيمة p_H أقل من أو تساوى قيمة SLE فإن المتغير X_H يمكن أن تشتمل علية معادلة الانحدار.

٢ - يتم حساب معاملات الانحدار الجزئية بين المتغير التابع وكل من المتغيرات المستقلة المتبقية والمصححة للمتغير(أو المتغيرات) التى تضمنتها معادلة الانحدار مع حساب احتمالات معنوياتها. يتم اختيار المتغير المستقل الذى له أعلى معامل ارتباط جزئى مع المتغير المعتمد (وليكن X_H) بمستوى معنوية p_H ، وبالتالى:

أ - إذا كانت قيمة p_H أكبر من قيمة SLE فإن المتغير X_H قرين هذه القيمة لا يمكن وضعة في معادلة الانحدار وبالتالي لا يوضع هذا المتغير في معادلة الانحدار وتكون معادلة الانحدار متضمنة المتغير السابق فقط.

ب- إذا كانت قيمة p_H أقل من أو تساوى قيمة SLE فإن المتغير X_H يمكن أن تشتمل علية معادلة الانحدار. ويتم تقييم المتغير التالى ... وهكذا.

١٤-٦-٣ طريقة الخطوة خطوة stepwise method

هذه الطريقة هى نفس طريقة الاختيار الأمامى مع الاختلاف فإن المتغير المستقل لذى يتم اختياره ليس بالضرورة أن يظل باقيا فى معادلة الانحدار حيث أنه يتم إعادة تقييم المتغيرات التى سبق أن تضمنتها معادلة الانحدار عند كل مرة يضاف إليها متغير جديد أى باستخدام طريقة الاستبعاد الخلفى.

مثال ١٤-٧

أورد Draper and Smith (1981) بيانات عن استخدام أربعة مواد كيماوية مختلفة (X_1، X_2، X_3، X_4) معبرا عنها كنسبة من الخامات التى تستخدم فى صناعة الأسمنت وإثرها على كمية الطاقة الناتجة (Y) بالكالورى لكل جرام أسمنت

مصنع. والمطلوب تقدير أفضل معادلة انحدار متعدد باستخدام طريقة الاستبعاد الخلفى backward elimination وطريقة الاختيار الأمامى forward selection وطريقة الخطوة خطوة stepwise procedure وكانت البيانات كالتالى:

X_1	X_2	X_3	X_4	Y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

الحل باستخدام البرنامج الإحصائى SAS

```
DATA CEMENT;
INPUT X1-X4 Y @@;
CARDS;
7  26 6 60 78.5 1 29 15 52 74.3 11 56 8 20 104.3
11 31 8 47 87.6 7 52 6  33 95.9 11 55 9 22 109.2
3 71 17 6 102.7 1 31 22 44 72.5 2 54 18 22 93.1
21 47  4 26 115.9 1  40 23 34 83.8 11 66 9 12 113.3
10 68 8 12 109.4
PROC CORR NOPROB NOSIMPLE; VAR Y; WITH X1 X2 X3 X4;
PROC STEPWISE;
MODEL Y = X1 X2 X3 X4/F B STEPWISE;
RUN;
```

لاحظ:

١ – استخدام أمر prco corr لحساب معامل الارتباط البسيط بين المتغير المعتمد وكل من المتغيرات المستقلة رغم عدم الحاجة لوضع أمر لحساب ذلك ولكن وضع الأمر بغرض إظهار قيم معاملات الارتباط. لاحظ استخدام noprop عند عدم الرغبة فى مستويات المعنوية المصاحبة لكل معامل ارتباط. استخدام nosimple لعدم طبع معلومات توصيف البيانات مثل المتوسط، المجموع ... الخ لعدم الحاجة لهذه المعلومات.

٢ – استخدام proc stepwise عند الرغبة في اختيار أفضل معادلة انحدار.

٣ – يكتب أمر model مع وضع جميع المتغيرات المستقلة فى النموذج وإضافة الاختيارات المختلفة للثلاث طرق السابق شرحها حيث b تمثل backward و f تمثل forward مع العلم أنه يمكن كتابة الكلمة كاملة بدلا من الاختصار.

٤ – يمكن إضافة اختيارات مختلفة لدرجة المعنوية إلى أمر model وذلك عند الرغبة فى تغيير مستوى المعنوية فى الحالات المختلفة.

نتائج التحليل:

```
                     Correlation Analysis

    4 'WITH' Variables:  X1        X2        X3        X4
    1 'VAR'  Variables:  Y
        Pearson Correlation Coefficients  / N = 13

                                Y
                    X1       0.73072

                    X2       0.81625

                    X3      -0.53467

                    X4      -0.82131

          Forward Selection Procedure for Dependent Variable Y

Step 1 Variable X4 Entered  R-square = 0.67454196   C(p)= 138.73083349
             DF    Sum of Squares   Mean Square      F      Prob>F
Regression    1    1831.89616002  1831.89616002   22.80    0.0006
Error        11     883.86691690    80.35153790
Total        12    2715.76307692
```

```
            Parameter    Standard        Type II
Variable    Estimate     Error          Sum of Squares      F      Prob>F
INTERCEP   117.56793118  5.26220651    40108.47690796    499.16   0.0001
X4          -0.73816181  0.15459600     1831.89616002     22.80   0.0006

Bounds on condition number:          1,            1
---------------------------------------------------------------------------
Step 2   Variable X1 Entered  R-square = 0.97247105 C(p) = 5.49585082

             DF   Sum of Squares    Mean Square       F      Prob>F
Regression    2    2641.00096477   1320.50048238   176.63   0.0001
Error        10      74.76211216      7.47621122
Total        12    2715.76307692

            Parameter      Standard        Type II
Variable    Estimate       Error         Sum of Squares     F        Prob>F
INTERCEP   103.09738164    2.12398361 17614.67006622   2356.10   0.0001
X1           1.43995828    0.13841664   809.10480474    108.22   0.0001
X4          -0.61395363    0.04864455    1190.92463664  159.30   0.0001

Bounds on condition number:      1.064105,      4.256421
---------------------------------------------------------------------------
Step 3  Variable X2 Entered R-square = 0.98233545   C(p) = 3.01823347

             DF   Sum of Squares    Mean Square      F     Prob>F
Regression   3     2667.79034752    889.26344917  166.83 0.0001
Error        9       47.97272940      5.33030327
Total       12     2715.76307692

           Parameter     Standard        Type II
Variable   Estimate      Error         Sum of Squares    F         Prob>F
INTERCEP   71.64830697  14.14239348    136.81003409    25.67      0.0007
X1          1.45193796   0.11699759    820.90740153   154.01      0.0001
X2          0.41610976   0.18561049     26.78938276     5.03      0.0517
X4         -0.23654022   0.17328779      9.93175378     1.86      0.2054

Bounds on condition number:      18.94008,      116.3601
---------------------------------------------------------------------------
No other variable met the 0.5000 significance level for entry into the
model.
---------------------------------------------------------------------------
     Summary of Forward Selection Procedure for Dependent Variable Y
       Variable     Number   Partial    Model
Step   Entered        In       R**2      R**2         C(p)       F         Prob>F
 1       X4           1      0.6745    0.6745    138.7308   22.7985    0.0006
 2       X1           2      0.2979    0.9725      5.4959 108.2239    0.0001
 3       X2           3      0.0099    0.9823     3.0182     5.0259    0.0517
---------------------------------------------------------------------------
```

Backward Elimination Procedure for Dependent Variable Y

Step 0 All Variables Entered R-square = 0.98237562 C(p) = 5.00000000

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	4	2667.89943757	666.97485939	111.48	0.0001
Error	8	47.86363935	5.98295492		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	62.40536930	70.07095921	4.74551686	0.79	0.3991
X1	1.55110265	0.74476987	25.95091138	4.34	0.0708
X2	0.51016758	0.72378800	2.97247824	0.50	0.5009
X3	0.10190940	0.75470905	0.10909005	0.02	0.8959
X4	-0.14406103	0.70905206	0.24697472	0.04	0.8441

Bounds on condition number: 282.5129, 2489.203

Step 1 Variable X3 Removed R-square = 0.98233545 C(p) = 3.01823347

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	2667.79034752	889.26344917	166.83	0.0001
Error	9	47.97272940	5.33030327		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	71.64830697	14.14239348	136.81003409	25.67	0.0007
X1	1.45193796	0.11699759	820.90740153	154.01	0.0001
X2	0.41610976	0.18561049	26.78938276	5.03	0.0517
X4	-0.23654022	0.17328779	9.93175378	1.86	0.2054

Bounds on condition number: 18.94008, 116.3601

Step 2 Variable X4 Removed R-square = 0.97867837 C(p) = 2.67824160

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	2657.85859375	1328.92929687	229.50	0.0001
Error	10	57.90448318	5.79044832		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	52.57734888	2.28617433	3062.60415610	528.91	0.0001
X1	1.46830574	0.12130092	848.43186034	146.52	0.0001
X2	0.66225049	0.04585472	1207.78226562	208.58	0.0001

```
Bounds on condition number:      1.055129,      4.220516
------------------------------------------------------------------------
All variables left in the model are significant at the 0.1000 level.
------------------------------------------------------------------------
  Summary of Backward Elimination Procedure for Dependent Variable Y

       Variable    Number  Partial   Model
Step   Removed       In     R**2      R**2      C(p)       F          Prob>F
  1      X3          3    0.0000    0.9823    3.0182    0.0182    0.8959
  2      X4          2    0.0037    0.9787    2.6782    1.8633    0.2054
------------------------------------------------------------------------

                   Stepwise Procedure for Dependent Variable Y

Step 1  Variable X4 Entered  R-square = 0.67454196 C(p) =138.73083349

             DF  Sum of Squares    Mean Square      F       Prob>F
Regression   1    1831.89616002   1831.89616002   22.80    0.0006
Error        11    883.86691690     80.35153790
Total        12   2715.76307692

            Parameter         Standard      Type II
Variable    Estimate          Error        Sum of Squares     F        Prob>F

INTERCEP   117.56793118    5.26220651   40108.47690796  499.16     0.0001
X4          -0.73816181    0.15459600     1831.89616002  22.80     0.0006

Bounds on condition number:              1,            1
------------------------------------------------------------------------

Step 2  Variable X1 Entered  R-square = 0.97247105  C(p) = 5.49585082

              DF  Sum of Squares   Mean Square      F       Prob>F
Regression    2    2641.00096477  1320.50048238   176.63    0.0001
Error         10      74.76211216     7.47621122
Total         12    2715.76307692

            Parameter     Standard        Type II
Variable    Estimate      Error         Sum of Squares      F         Prob>F
INTERCEP   103.09738164   2.12398361   17614.67006622   2356.10    0.0001
X1           1.43995828   0.13841664     809.10480474    108.22    0.0001
X4          -0.61395363   0.04864455    1190.92463664    159.30    0.0001

Bounds on condition number:      1.064105,      4.256421
------------------------------------------------------------------------

Step 3  Variable X2 Entered  R-square = 0.98233545 C(p) = 3.01823347
```

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	2667.79034752	889.26344917	166.83	0.0001
Error	9	47.97272940	5.33030327		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	71.64830697	14.14239348	136.81003409	25.67	0.0007
X1	1.45193796	0.11699759	820.90740153	154.01	0.0001
X2	0.41610976	0.18561049	26.78938276	5.03	0.0517
X4	-0.23654022	0.17328779	9.93175378	1.86	0.2054

Bounds on condition number: 18.94008, 116.3601

Step 4 Variable X4 Removed R-square = 0.97867837 C(p) = 2.67824160

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	2657.85859375	1328.92929687	229.50	0.0001
Error	10	57.90448318	5.79044832		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	52.57734888	2.28617433	3062.60415609	528.91	0.0001
X1	1.46830574	0.12130092	848.43186034	146.52	0.0001
X2	0.66225049	0.04585472	1207.78226562	208.58	0.0001

Bounds on condition number: 1.055129, 4.220516

All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	X4		1	0.6745	0.6745	138.7308	22.7985	0.0006
2	X1		2	0.2979	0.9725	5.4959	108.2239	0.0001
3	X2		3	0.0099	0.9823	3.0182	5.0259	0.0517
4		X4	2	0.0037	0.9787	2.6782	1.8633	0.2054

١٤-٧ الارتباط المتعدد والجزئى Multiple and partial correlation

يفترض، كما هو الحال فى حالة الارتباط البسيط، أن المتغيرات عشوائية أى فى عينة عشوائية مسحوبة من عشيرة متعددة المتغيرات تتوزع طبيعيا multivariate normal distribution، وعلى ذلك فلن يكون الاهتمام بأى من المتغيرات مستقل وأى منها تابع.

ويعرف معامل الارتباط الجزئى بين متغيرين بأنه هو الارتباط بين هذين المتغيرين فى مجموعة من الأفراد أو المشاهدات كل منها له نفس المتغير الثالث (أو المتغيرات الأخرى فى حالة أكثر من ثلاث متغيرات)، أى معامل الارتباط بين الأول والثانى على اعتبار أن المتغير الثالث ثابت أو بعد التصحيح لتأثيره. وقيمة هذا المعامل تنحصر أيضاً بين 1- ، 1+.

ومعاملات الارتباط الجزئية فى حالة ثلاث متغيرات هى:

$$r_{1,2.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)}\sqrt{(1-r_{23}^2)}}$$

$$r_{1,3.2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{32}^2}}$$

$$r_{2,3.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{13}^2)}} \qquad (١٣-١٤)$$

حيث $r_{12} = r_{21}$ ... وهكذا. ومعنى $r_{12.3}$ أنه معامل الارتباط الجزئى بين المتغير الأول والمتغير الثانى مع بقاء المتغير الثالث ثابتاً.

ومن ذلك فإنه لحساب معاملات الارتباط الجزئية بين ثلاث متغيرات عشوائية يلزم حساب معاملات الارتباط البسيطة بين كل زوج منهم أى r_{12}، r_{13}، r_{23}. ويستخدم جدول ٧ ملحق أ لاختبار معنوية معاملات الارتباط البسيط. أما معامل الارتباط المتعدد R فيقيس شدة العلاقة بين أحد المتغيرات وجميع المتغيرات الأخرى معاً وقيمته تنحصر بين الصفر والواحد الصحيح. ويمكن حساب معامل الارتباط المتعدد بين المتغير 1 من جهة والمتغيرين 2، 3 من جهة أخرى كما يلى:

$$R_{1(2,3)} = \sqrt{1-(1-r_{12}^2)(1-r_{13.2}^2)} \qquad (١٤-١٤)$$

كما يحسب من مجموع المربعات فى تحليل الانحدار حيث:

$$R = \sqrt{\frac{RSS}{TSS}} \qquad (١٤-١٥)$$

وحيث إن:

$$R^2 = RSS/TSS$$
$$= \sum \hat{y}^2 / \sum y^2$$

فإن:

$$\sum \hat{y}^2 = R^2 \sum y^2 = RSS$$

وعليه فإن مجموع المربعات عن خط الانحدار هو:

$$ESS = TSS - RSS = \sum y^2 - R^2 \sum y^2 = (1-R)^2 \sum y^2$$

وتكون قيمة F لاختبار فرض العدم أن المتغيرين 2، 3 معاً لا يؤثران على المتغير 1 هى:

$$F = \frac{R^2 \sum y^2 / k}{(1-R^2)\sum y^2 /(n-k-1)} = \left(\frac{n-k-1}{k}\right)\left(\frac{R^2}{1-R^2}\right) \qquad (١٤-١٦)$$

وبالتالى فإنه يستخدم أيضاً اختبار F لاختبار معنوية معامل الارتباط المتعدد.

مثال ١٤-٨

استخدم بيانات المثال ١٤-١ لحساب معاملات الارتباط الجزئية بين المتغيرين الأول والثالث ومعامل الارتباط المتعدد بين المتغير 1 من جهة مع 2، 3 من جهة أخرى واختبر معنوية كل منهم ولسهولة التعبير أعد تسمية y بـــ X_3.

حساب كل من معاملات الارتباط البسيط كما يلى:

$$r_{13} = \frac{\sum X_1X_3 - \sum X_1 \sum X_3 / n}{\sqrt{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}\sqrt{\sum X_3^2 - \frac{(\sum X_3)^2}{n}}}$$

$$= \frac{84 - \frac{(27)(14)}{5}}{\sqrt{54 - \frac{(14)^2}{5}}\sqrt{169 - \frac{(27)^2}{5}}} = 0.45$$

وبالمثل

$$r_{23} = -0.46 \text{ ، } r_{12} = 0.49$$

معامل الارتباط الجزئى:

$$r_{13.2} = \frac{0.45 - (0.49)(-0.46)}{\sqrt{1-(0.49)^2}\sqrt{1-(-0.46)^2}} = 0.87$$

ودرجات الحرية 2

أما معامل الارتباط المتعدد (معادلة ١٤-١٤):

$$R_{1(23)} = \sqrt{1-[1-(0.49)^2][(1-(0.87)^2]} = 0.91$$

ومنها $R^2 = 0.82$ بمعنى أن حوالى 82% من التباين فى المتغير Y يمكن تفسيره بالنموذج الخطى فى المتغيرين المستقلين X_1، X_2.

ويمكن الحصول على معامل الارتباط المتعدد أيضا باستخدام الجذر التربيعى لناتج قسمة مجموع المربعات الراجعة للانحدار على مجموع المربعات الكلى كالتالى:

$$R_{1(23)} = \sqrt{\frac{19.075}{23.2}} = 0.91$$

ولاختبار معنوية معامل الارتباط المتعدد تحسب قيمة F حيث:

$$F = \frac{5-2-1}{2}\left[\frac{(0.91)^2}{1-(0.91)^2}\right] = 4.8$$

والفرق بين قيمة F والتى تساوى 4.8 فى حالة الارتباط المتعدد، 4.6 فى حالة الانحدار المتعدد إنما يرجع إلى خطأ التقريب.

وقيمة F المحسوبة أقل من قيمة F الجدولية وعلى ذلك فالعلاقة بين المتغيرات الثلاثة غير معنوية.

ويمكن حساب R^2_{adj}، أى R^2 المعدلة لعدد المتغيرات التفسيرية (المستقلة)، من المعادلة التالية:

$$R^2_{adj} = 1 - \left(\frac{n-1}{n-k-1}\right)\left(1 - R^2\right) = 1 - \left(\frac{5-1}{5-2-1}\right)(1 - 0.82) = 0.64$$

صندوق ١٤-١

- الانحدار المتعدد هو امتداد طبيعى للانحدار البسيط حيث يكون هناك متغير تابع واحد (Y) وعدة متغيرات مستقلة $(X_1, X_2, \cdots, X_p)$ بدلا من واحد فقط. وتصبح معادلة التنبؤ $\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_pX_p$ حيث b_1 هى معامل الانحدار الجزئى للمتغير Y على المتغير المستقل X_1 ... وهكذا لبقية الـــ b's.
- وبالتوازى مع الانحدار البسيط فإن التباين الكلى فى Y يقسم إلى أجزاء كل منها يرجع إلى متغير مستقل بعينه. ويحسب مجموع المربعات الراجع لمتغير مستقل معين بحساب المتبقى فى النموذج بدون هذا التغير منقوصا منه المتبقى فى النموذج بعد إضافة هذا المتغير.
- يلاحظ أن حاصل جمع مجموع انحرافات هذه الأجزاء الراجعة للانحدار لا تساوى مجموع المربعات الراجع إلى الاعتماد على كل المتغيرات المستقلة مجتمعة إلا إذا كانت المتغيرات المستقلة مستقلة عن بعضها تماما (أى معامل الارتباط بين كل اثنين منها = صفر).
- وحيث أنه فى الحالة الواقعية يمكن أن يزداد عدد المتغيرات المستقلة كثير، فهناك طرق للمساعدة فى تقرير أى من المتغيرات المستقلة يمكن حذفه من النموذج الإحصائى دون فقد معلومات فقدا معنويا. من هذه الطرق: الاستبعاد الخلفى، الاختيار الأمامى، الخطوة خطوة.

تمارين الباب الرابع عشر

١٤-١ أورد (1982) Hill التمرين التالى حيث أخذت عينة من 24 دجاجة وكانت X_1 تمثل الوزن الحى بالمائة جرام، X_2 تمثل عمق الصدر بالسنتيمتر وتمثل Y الوزن بعد الطبخ بالمائة جرام وكانت البيانات كالتالى:

رقم الدجاجة	Y	X_1	X_2	رقم الدجاجة	Y	X_1	X_2
١	6.03	16.44	8.3	١٣	7.55	18.40	9.0
٢	6.09	16.76	8.5	١٤	5.83	15.46	7.5
٣	6.67	17.22	8.8	١٥	5.93	15.20	8.2
٤	6.25	16.04	8.5	١٦	6.65	17.15	8.7
٥	5.65	15.74	8.5	١٧	6.24	15.60	8.1
٦	6.91	18.06	9.0	١٨	5.50	16.95	8.0
٧	7.81	20.60	8.7	١٩	7.22	17.22	8.1
٨	5.68	16.20	9.0	٢٠	7.17	18.36	8.7
٩	7.14	17.30	8.6	٢١	5.81	17.50	8.0
١٠	7.10	18.85	8.1	٢٢	7.07	16.66	8.0
١١	6.29	17.63	8..0	٢٣	5.51	14.88	8.6
١٢	6.70	17.78	8.6	٢٤	5.77	16.25	8.2

المطلوب:

١- تقدير معادلة الانحدار واختبار معنوية معاملات الانحدار المختلفة.

٢- تقدير معاملات الارتباط الجزئية والارتباط المتعدد واختبار معنوية كل منها.

١٤-٢ أجرى Ostle, 1963 تجربة على الأرانب لدراسة المتغيرات التى قد يكون لها علاقة بدرجة تصلب الشرايين (Y) degree of atherosclerosis والتى قدرت بقيم صحيحة تتراوح بين 0 إلى 4 وأخذت قياسات للمتغيرات التالية:

رقم الأرنب	Y	X_1	X_2	X_3	X_4	X_5
١	2	30	424	2.46	0.90	18
٢	0	30	313	2.39	0.91	10
٣	2	35	243	2.75	0.95	30
٤	2	35	365	2.19	0.95	21
٥	3	43	396	2.67	1.00	39
٦	2	43	356	2.74	0.79	19
٧	3	44	346	2.55	1.26	56
٨	0	44	156	2.58	0.95	28
٩	4	44	278	2.49	1.10	42
١٠	1	44	349	2.52	0.88	21
١١	1	44	141	2.36	1.29	56
١٢	1	44	245	2.36	0.97	24
١٣	1	44	395	2.15	1.01	27
١٤	3	45	297	2.56	1.11	45
١٥	2	45	310	2.62	0.94	20
١٦	3	45	151	3.39	0.96	35
١٧	4	45	370	3.57	0.88	15
١٨	4	45	379	1.98	1.47	64
١٩	3	45	463	2.06	1.05	31
٢٠	4	45	316	2.45	1.32	60
٢١	4	45	280	2.25	1.08	36
٢٢	0	49	139	2.20	1.36	59
٢٣	4	49	245	2.05	1.13	37
٢٤	1	49	373	2.15	0.88	25
٢٥	3	51	224	2.15	1.18	54
٢٦	4	51	677	2.10	1.16	33
٢٧	4	51	424	2.10	1.40	59
٢٨	0	51	150	2.10	1.05	30

حيث

X_1 : متوسط الجرعة اليومية للكولسترول بالجرام

X_2 : متوسط كولسترول سيرم الدم بالمليجرام لكل 100 مل.

X_3 : وزن الجسم الابتدائى initial body weight بالكيلوجرام.

X_4 : النسبة بين الوزن النهائى والوزن الابتدائى للجسم.

X_5 : متوسط كمية الغذاء المأكول بالجرام لكل كيلوجرام وزن ابتدائى.

المطلوب:

١ – إجراء تحليل الانحدار المناسب.

٢ – ما هى أفضل معادلة انحدار عند مستويات معنوية مختلفة ولتكن 5%، 15% .