الانحدار الخطي المتعدد Multiple Linear Regression

ان نموذج الانحدار المتعدد هو عبارة عن انحدار للمتغير التابع (Y) على العديد من المتغيرات المستقلة $X_1 , X_2 , ...X_K$ ويسمى هذا بنموذج الانحدار الخطي المتعدد ، Multiple Linear Regression .

ويهدف هذا المقال إلى توضيح كيفية تقدير نموذج الانحدار الخطي المتعدد ، ، ثم تحديد أهم افتراضات النموذج ، يضاف إلي ذلك بيان عدم وجود علاقة خطية تامة بين المتغيرات المستقلة وكيف أن المصفوفة ($X`X$) ، تكون مصفوفة غير شاذة (Singular–Non) إذا كان محددها لا يساوي صفرا . ثم يتم بعد ذلك تقدير معلومات النموذج ، تقدير التباين والتباين المشترك والانحراف المعياري لها للوصول إلى اختبار معاملات النموذج .

نموذج الانحدار الخطي المتعدد :

يستند النموذج الخطي المتعدد على افتراض وجود علاقة خطية بين متغير تابع Y_i وعدد من المتغيرات المستقلة $X_1,X_2,...X_K$ وحد عشوائي U_i ، ويعبر عن هذه العلاقة ، بالنسبة لـn من المشاهدات وk من المتغيرات المستقلة ، بالشكل آلاتي :

$$Y_i = B_0 + B_1X_{i1} + B_2X_{i1} + \dots + B_KX_{ik} + U_i \qquad \dots. (1)$$

وفي واقع الآمر فان هذه المعادلة هي واحدة من جملة معادلات يبلغ عددها (n) تكون نظام المعادلات آلاتي :

$$Y_1 = B_0 + B_0X_{11} + B_2X_{12} + \dots + B_KX_{1K} + U_1$$
$$Y_2 = B_0 + B_1X_{21} + B_2X_{22} + \dots B_KX_{2K} + U_2$$
$$. \quad . \quad .. .. \quad \dots \quad \dots \quad \dots \quad ..$$
$$.... .. \quad .. \quad .. \quad \dots \quad \dots \quad \dots \quad ..$$
$$Y_n = B_0 + B_1X_{n1} + B_2X_{n2} + \dots + B_KX_{nK} + U_n$$

هذه المعادلة تتضمن (k+1) من المعلومات المطلوب تقديرها علما بان الحد الأول منها (B_0) يمثل الحد الثابت ، الآمر الذي يتطلب اللجوء إلى المصفوفات والمتجهات لتقدير تلك المعلمات. عليه يمكن صياغة هذه المعادلات في صورة مصفوفات وكآلاتي :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1K} \\ 1 & X_{21} & X_{22} & \dots & X_{2K} \\ . & . & . & \dots & . \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ . \\ B_K \end{bmatrix} + \begin{bmatrix} U_0 \\ U_1 \\ . \\ U_n \end{bmatrix} \quad \dots. \ (2)$$

وباختصار

$$Y = XB + U$$

Y: متجه عمودي أبعاده (n+1) يحتوي مشاهدات المتغير التابع .

X : مصفوفة أبعادها ($n \times k+1$) تحتوي مشاهدات المتغيرات المستقلة يحتوي عمودها الأول على قيم الواحد الصحيح ليمثل الحد الثابت .
B: متجه عمودي أبعاده ($K+1 \times 1$) يحتوي على المعالم المطلوب تقديرها .
U: متجه عمودي أبعاده ($n \times 1$) يحتوي على الأخطاء العشوائية .
وبما أن المعادلة (1) هي العلاقة الحقيقية المجهولة والمراد تقديرها باستخدام الإحصاءات المتوفرة عن المتغير التابع ، Y ، والمتغيرات المستقلة ، $X_1, X_2, ..X_K$ ، فانه يستوجب تحقق الفروض الأساسية الخاصة بU_i التالية :

$U_i \sim N(0, \sigma^2 I n)$

والذي يعني أن U_i يتوزع توزيعا طبيعيا (N) متعدد المتغيرات لمتجه وسطه صفري (0) ومصفوفة تباين وتباين مشترك عددية هي (σ^2 In) .

فرضيات النموذج الخطي المتعدد :

عند استخدام طريقة OLS في تقدير نموذج الانحدار الخطي المتعدد ، فانه يجب توافر الافتراضات آلاتية :

1- القيمة المتوقعة لمتجه حد الخطا تساوي صفرا أي أن ، $E(U_i) = 0$:

$$E(Ui) = E\begin{bmatrix} U_1 \\ U_2 \\ . \\ U_n \end{bmatrix} = \begin{bmatrix} E(U_1) \\ E(U_2) \\ . \\ E(U_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ 0 \end{bmatrix}$$

2- تباين العناصر العشوائية ثابت ، والتباين المشترك بينها يساوي صفرا ، أي أن :

$$Cov(U) = E(UU') = \sigma^2 In$$

$$E(UU') = E\begin{bmatrix} U_1 \\ U_2 \\ . \\ U_n \end{bmatrix}\begin{bmatrix} U_1 & U2 & ... & U_n \end{bmatrix}$$

$$= E\begin{bmatrix} U_1^2 & U_1U_2 & ... & U_1U_n \\ U_2U_1 & U_2^2 & ... & U_2U_n \\ . & . & . & . \\ U_nU_1 & U_nU_2 & ... & U_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(U_1^2) & E(U_1U_2) & .... & E(U_1U_n) \\ E(U_2U_1) & E(U_2^2) & .... & E(U_2U_n) \\ . & . & ..... & . \\ E(U_nU_1) & E(U_nU_2) & .... & E(U_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{var}(U_1) & Cov(U_1U_2) & .... & Cov(U_1U_n0 \\ Cov(U_2U_1) & Var(U_2) & .... & Cov(U_2U_n) \\ . & . & .... & . \\ Cov(U_nU_1) & Cov(U_nU_2) & .... & Var(U_n) \end{bmatrix}$$

$\because$ var (Ui) = E(U_i^2) = σ^2

$\because$ Cov ($U_iU_j) = E(U_iU_j) = 0$, I # j

$$E(UU') = \begin{bmatrix} \sigma_1^2 & 0 & ...... & 0 \\ 0 & \sigma_2^2 & ...... & 0 \\ 0 & 0 & ...... & \sigma_n^2 \end{bmatrix}$$

حيث أن : $\sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & ..... & 0 \\ 0 & 1 & ...... & 0 \\ 0 & 0 & ...... & 1 \end{bmatrix}$$

$$= \sigma^2 In$$

وتسمى المصفوفة العددية أعلاه بمصفوفة التباين والتباين المشترك – Variance Covariance Matrix لحد الخطاU ، حيث تشكل العناصر القطرية في المصفوفة ، تباين قيم U بينما تبقى العناصر غير القطرية (أعلى واسفل القطر) مساوية للصفر لانعدام التباين المشترك والترابط بين قيم U_i .

3- ليس هناك علاقة خطية تامة بين المتغيرات المستقلة كما وان عدد المشاهدات يحجب أن يزيد على عدد المعلمات المطلوب تقديرها ، أي أن :

R (x) = k + 1 < n

حيث أن (r) رتبة مصفوفة البيانات ، (x) عدد المتغيرات المستقلة (k) زائدا (1) الحد الثابت ، وهي اصغر من عدد المشاهدات (n) . وهذه الفرضية ضرورية جدا لضمان أيجاد معكوس المصفوفة ($x'x$) ، إذ أن انتفاء هذا الفرض يجعل رتبة المصفوفة (X) اقل من (K+1) وبالتالي فان رتبة ($x'x$) التي تستخدم في الحصول على مقدرات OLS بدورها اقل من (K+1) ولايمكن أيجاد معكوسها بسبب ما يسمى بمشكل الارتباط الخطي المتعدد ، وبالتالي لايمكن الحصول على مقدرات المربعات الصغرى العادية ، OLS .

طرق تقدير معلمات النموذج :

في ضوء الفرضيات المذكورة أعلاه يمكن استخدام طريقة OLS في تقدير معلمات النموذج الخطي المتعدد ، ولهذا الغرض يمكن كتابة المعادلة (1) بصيغتها التقديرية كآلاتي :

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2}$$

ولما كان هدفنا هو الحصول على قيم كل من $\hat{B}_0, \hat{B}_1, \hat{B}_2$ التي تجعل مجموع مربعات الانحرافات اقل ما يمكن ، أي تصغير القيمة $\sum e_i^2$ (مبدا المربعات الصغرى) إلى اقل قيمة ممكنة ، أي :

$$\text{Min} \rightarrow \sum_{i=1}^{n} e_i^2$$

$$\because e_i = Y_I - \hat{Y}_i$$

$$\sum e_i^2 = \sum (Y_I - \hat{Y}_i)^2$$

ومن خلال التعويض عن $\hat{Y}_i$ بما يساويها واخذ المشتقات الجزئية بالنسبة إلى $\hat{B}_2, \hat{B}_1, \hat{B}_0$ ومساواتها بالصفر نحصل على :

$$\sum e_i^2 = \sum (Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2})^2$$

$$\frac{\delta e_i^2}{\delta \hat{B}_0} = 2\sum (Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2})(-1) = 0$$

$$-2 \sum (Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}) = 0$$

بالقسمة على (-2) وفك القوس ، نحصل :

$$\sum Y_i - n\hat{B}_0 - \hat{B}_1 \sum X_{i1} - \hat{B}_2 \sum X_{i2} = 0$$

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_{i1} + \hat{B}_2 \sum X_{i2} \qquad (3)$$

$$\frac{\delta \sum e_i^2}{\delta \hat{B}_1} = 2\sum (Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2})(-X_{i1}) = 0$$

$$-2\sum X_{i1}(Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}) = 0$$

بالقسمة (-2) وفك القوس ، نحصل :

$$\sum X_{i1}Y_i - \hat{B}_0 \sum X_{i1} - \hat{B}_1 X_{i1}^2 - \hat{B}_2 \sum X_{i1} x_{i2} = 0$$

$$\sum X_{i1}Y_i = \hat{B}_0 \sum X_{i1} + \hat{B}_1 \sum X_{i1}^2 + \hat{B}_2 \sum X_{i1} X_{i2} \qquad (4)$$

$$\frac{\delta \sum e_i^2}{\delta \hat{B}_1} = 2\sum (Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2})(-X_{i2}) = 0$$

$$-2\sum X_{i2}(Y_i - \hat{B}_0 - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}) = 0$$

بالقسمة على (-2) وفك القوس ، نحصل :

$$\sum X_{i2}Y_i - \hat{B}_0 \sum X_{i2} - \hat{B}_1 \sum X_{i1} X_{i2} - \hat{B}_2 \sum X_{i2}^2 = 0$$

$$\sum X_{i2}Y_i = \hat{B}_0 \sum X_{i2} + \hat{B}_1 \sum X_{i1} X_{i2} + \hat{B}_2 \sum X_{i2}^2 \qquad \mathbf{(5)}$$

وتمثل المعادلات (3) ، (4) و (5) المعادلات الطبيعية الثلاث التي تستخدم في تقدير المعالم الثلاثة المجهولة $\hat{B}_2, \hat{B}_1, \hat{B}_0$. أن هذه المعادلات ، يمكن حلها بإحدى الطرق آلاتية :

أولا : طريقة المحددات :

ويمكن أن تحل هذه المعادلات بواسطة قاعدة كرا يمر للحصول على قيم $\hat{B}_K$ من المعلمات وعلى النحو آلاتي :

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_{i1} + \hat{B}_2 \sum X_{i2}$$
$$\sum X_{i1} Y_i = \hat{B}_0 \sum X_{i1} + \hat{B}_1 \sum X_{i1}^2 + \hat{B}_2 \sum X_{i1} X_{i2}$$
$$\sum X_{i2} Y_i = \hat{B}_0 \sum X_{i2} + \hat{B}_1 \sum X_{i2} + \hat{B}_2 \sum X_{i2}^2$$
$$\begin{bmatrix} \sum Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \end{bmatrix} = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{bmatrix}$$

ومن النظام أعلاه ، يمكن أيجاد المحددات آلاتية :

$$|D| = \begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} Y_i & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} Y_i & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{vmatrix}$$

$$|N_1| = \begin{vmatrix} n & \sum Y_1 & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1} Y_i & \sum X_{i1} X_{i2} \\ \sum X_{i2} & \sum X_{i2} Y_i & \sum X_{i2}^2 \end{vmatrix}$$

$$|N2| = \begin{vmatrix} n & \sum X_{i1} & \sum Y_i \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1} Y_i \\ \sum X_{i2} & \sum X_{i1} X_{i2} & \sum X_{i2} Y_i \end{vmatrix}$$

$$\hat{B}_1 = \frac{|N_1|}{|D|} = \frac{\begin{vmatrix} n & \sum Y_i & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1} Y_i & \sum X_{i1} X_{i2} \\ \sum X_{i2} & \sum X_{i2} Y_i & \sum X_{i2}^2 \end{vmatrix}}{\begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} Y_i & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} Y_i & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{vmatrix}}$$

$$\hat{B}2 = \frac{|N_2|}{|D|} = \frac{\begin{vmatrix} n & \sum Y_i & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1} Y_i & \sum X_{i1} X_{i2} \\ \sum X_{i2} & \sum X_{i2} Y_i & \sum X_{i2}^2 \end{vmatrix}}{\begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} Y_i & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} Y_i & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{vmatrix}}$$

أما بالنسبة ل $\hat{B}_0$ فيتم الحصول عليه عن طريق :

$$\hat{B}_0 = \overline{Y} - \hat{B}_1 \overline{X}_1 - \hat{B}_2 \overline{X}_2$$

ثانيا : طريقة الانحرافات :

ويمكن تقدير معاملات الانحدار المتعدد باستخدام أسلوب الانحرافات أو ما يسمى بالمتوسطات ، أي انحرافات القيم الأصلية عن وسطها وكآلاتي :

ولهذا الغرض نأخذ نموذج يحتوي متغيرين مستقلين X1 و X2 :

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + e_i$$

وبأخذ المتوسط لهذه المعادلة :

$$\overline{Y} = \hat{B}_0 + \hat{B}_1 \overline{X}_1 + \hat{B}_2 \overline{X} 2 + \overline{e}_i \ , \ \overline{e}_i = 0$$

$$\hat{Y}_i - \overline{Y} = \hat{B}_1 (X_{i1} - \overline{X}_1) + \hat{B}_2 (X_{i2} - \overline{X}_2) + e_i$$

$$\hat{y}_i = \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + e_i$$

اثبات أن $\overline{Y} = \overline{Y}$:

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i$$

وبادخال $\sum$ على طرفي المعادلة اعلاه ، نحصل على :

$$\sum \hat{Y}_i = n\hat{B}_0 + \hat{B}_1 \sum X_i$$

وبالقسمة على n :

$$\frac{\sum \hat{Y}_i}{n} = \hat{B}_0 \frac{n}{n} + \hat{B}_1 \frac{\sum X_i}{n}$$

$$\overline{Y}_i = \hat{B}_0 + \hat{B}_1 \overline{X}_i \qquad .... (6)$$

$$\therefore \hat{B}_0 = \overline{Y} - \hat{B}_1 \overline{X}_i \qquad ...... (7)$$

$$\therefore \overline{Y} = \hat{B}_0 + \hat{B}_1 \overline{X}_1 \qquad ....(8)$$

وبطرح المعادلة (8) من المعادلة (6) نحصل :

$$\overline{Y}_i - \overline{Y} = \hat{B}_0 + \hat{B}_1 \overline{X}_i - \hat{B}_0 - \hat{B}_1 \overline{X}_i$$

وبعد الاختصار في الطرف الايمن ، نحصل :

$$\overline{\hat{Y}}_i - \overline{Y} = 0$$

ومنها يكون :

$$\therefore \overline{\hat{Y}}_i = \overline{Y}$$

or

$$y_i = \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + e_i \qquad (I=1,2,3,\ldots\ldots,n) \qquad \ldots(9)$$

وفي واقع الأمر فان المعادلة أعلاه هي واحدة من جملة معادلات يبلغ عددها n معادلة تكون نظام المعادلات التالي :

$$Y_1 = B_1 X_{11} + B_2 X_{12} + ... + B_K X_{1K} + e_1$$

$$Y_2 = B_1 X_{21} + B_2 X_{22} + ... + B_K X_{2K} + e2$$

$$\ldots. \qquad \ldots\ldots\ldots. \qquad \ldots\ldots\ldots$$

$$y_n = B_1 X_{n1} + B_2 X_{n2} + ... + B_K X_{nK} + en$$

ويمكن التعبير عن المعادلات أعلاه في هيئة مصفوفة وكما يلي :

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1K} \\ X_{21} & X_{22} & ... & X_{2K} \\ . & . & . & . \\ X_{n1} & X_{n2} & ... & X_{nk} \end{bmatrix} \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \\ . \\ \hat{B}_K \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ e_n \end{bmatrix}$$

حيث يمكن التعبير عن ذلك بصيغة المصفوفات :

$$Y = x\ \hat{B} + e$$

حيث أن :

Y: متجه عمودي أبعاده (n ×1) يحتوي على انحرافات قيم المتغير التابع .

X: مصفوفة أبعادها (n × k – 1) تحتوي على انحرافات قيم المتغيرات المستقلة حيث أنها لا تتضمن العمود الأول الذي يمثل الحد الثابت . حيث يمكن بذلك استخراج الحد الثابت $\hat{B}_0$ من خارج المصفوفة باستخدام القانون آلاتي :

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}_1 - \hat{B}_2 \bar{X}_2$$

or

$$\hat{B}_0 = \bar{Y} - (\hat{B}_1 \bar{X}_1 - \hat{B}_2 \bar{X}_2)$$

$\hat{B}$: متجه عمودي أبعاده (K – 1 ×1) تحتوي على المعالم المجهولة .

E : متجه عمودي أبعاده (n ×1) يحتوي على البواقي .

ويمكن التوصل الى مصفوفة الانحرافات باتباع الخطوات التالية :

باعادة كتابة المعادلة (4 . 7) على النحو الاتي :

$$e_i = y_i - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}$$

ولما كانت افضل طريقة للحصول على اصغر قيمة ممكنة للانحرافات تتم بواسطة تربيعها وبجعل مجموع مربعاتها اصغر ما يمكن . وبأخذ المشتقة الجزئية لها بالنسبة لكل من $\hat{B}_1, \hat{B}_1$ ومساواتها بالصفر نحصل على :

$$\sum e_i^2 = \sum (y_i - \hat{B}_1 x_{i1} - \hat{B}_2 x_{i2})^2$$

$$\frac{\delta \sum e_i^2}{\delta \hat{B}_1} = 2\sum (y_i - \hat{B}_1 x_{i1} - \hat{B}_2 X_{i2})(-x_{i1}) = 0$$

$$-2\sum X_{i1}(y_i - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}) = 0$$

وبالقسمة على (-2) وفك القوس ، نحصل على :

$$\sum X_{i1} y_i - \hat{B}_1 \sum X_{i1}^2 - \hat{B}_2 \sum X_{i1} X_{i2} = 0$$

$$\sum x_{i1} y_i = \hat{B}_1 \sum X_{i1}^2 + \hat{B}_2 \sum X_{i1} X_{i2}$$

... (10)

$$\frac{\delta \sum e_i^2}{\delta \hat{B}_2} = 2\sum (y_i - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2})(-X_{i2}) = 0$$

$$-2\sum x_{i2}(y_i - \hat{B}_1 X_{i1} - \hat{B}_2 X_{i2}) = 0$$

وبالقسمة على (-2) وفك القوس ، نحصل :

$$\sum X_{i2}y_i - \hat{B}_1\sum X_{i1}X_{i2} - \hat{B}_2\sum X_{i2}^2 = 0$$
$$\sum X_{i2}y_i = \hat{B}_1\sum X_{i1}X_{i2} + \hat{B}_2\sum X_{i2}^2 \quad \textbf{...(11)}$$

ويمكن صياغة المعادلتين أعلاه على شكل مصفوفة وكآلاتي :

$$\begin{bmatrix}\sum X_{i1}y_i \\ \sum X_{i2}y_i\end{bmatrix} = \begin{bmatrix}\sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i1}X_{i2} & \sum X_{i2}^2\end{bmatrix}\begin{bmatrix}\hat{B}_1 \\ \hat{B}_2\end{bmatrix}$$

ومن النظام أعلاه ، يمكن إعادة كتابته بالشكل التالي :

$$x'y = (x'x)^{-1}\hat{B}$$

وعليه فان تقدير المعالم باستخدام المصفوفة بأسلوب الانحرافات يأخذ الصيغة التالية :

$$\hat{B} = (X'X)^{-1}X'Y$$

وبعد احتساب المتجه $X'Y$ ومحدد المصفوفة $|X'X|$ الذي ينبغي أن لا يساوي صفرا نوجد مقلوب المصفوفة الذي هو عبارة عن $(X'X)^{-1} = \frac{adj(x'x)}{|x'x|}$ ومن ثم تطبيق القانون أعلاه . أما $\hat{B}_0$ فيمكن حسابه بموجب القانون آلاتي :

$$\hat{B}_0 = \bar{Y} - \hat{B}_1\bar{X}_1 - \hat{B}_2\bar{X}_2$$

هذا ويمكن استخراج القيم بالانحرافات دون الرجوع إلى البيانات الأصلية وكما مبين أدناه :

$$\sum y_i^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$\sum y_1^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$\sum y_2^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$\sum x_1 y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{n}$$

$$\sum x2y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{n}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n}$$

وبعد استخدام الحاسوب ، فقد اصبح من السهل على الباحث الاقتصادي أن يحصل على النتائج من خلال أجادته استخدام إحدى البرمجيات الإحصائية أمثال SPSS ، EXCEL ، ، ولايحتاج إلى استخدام الصيغ أعلاه في الجوانب التطبيقية ، ولكن تم عرضها هنا لمعرفة كيفية عمل الانحدار المتعدد.

اختبار الفرضيات لنموذج الخطي المتعدد :

يهدف هذا البحث إلى توسيع معارفنا الأساسية لنموذج الانحدار وذلك بأجراء اختبار معنوية الانحدار المتعدد والمقدر باستخدام توزيع اختبار إحصاءه F ومقارنته باختبار t ومن ثم تقييم كفاءة الأداء العام لنموذج الانحدار المتعدد R^2 ومقارنته بمعامل التحديد المقدر المعدل $\bar{R}^2$ ، وكذلك اختبار العلاقة بين F و R^2 من خلال جدول تحليل التباين ، ANOVA ، ثم علاقة R^2 بقيمة المتغير العشوائي ، $\sum e_i^2$.

اختبار معنوية المعالم (t) :

يستخدم اختبار t لتقييم معنوية تأثير المتغيرات المستقلة $x_1,x_2,...x_k$ في التغير التابع y في نموذج الانحدار المتعدد يعتمد على نوعين من الفروض :

فرضية العدم H_0 $B_1 = B_2 = B_3 ...= B_K = 0$

الفرضية البديلة H_1 $B_1 = B_2 \# B_3 \# ...B_K = 0$

وبعد احتساب قيمة (t) تقارن مع قيمتها الجدولية لتحديد قبول او رفض فرضية العدم ومن ثم تقييم معنوية معلمات النموذج المقدر ، والصيغة الرياضية لهذا الاختبار يمكن بيانها كما يلي :

ا – بالنسبة الى $\hat{B}_1$

$$t\hat{B}_1 = \frac{\hat{B}_1}{S_{\hat{B}_1}}$$

$$S_{\hat{B}_1} = \sqrt{S^2{}_{\hat{B}_1}}$$

$$S^2_{\hat{B}_1} = \operatorname{var}(\hat{B}_1) = S^2 e a_{11}$$

$$\operatorname{var}(\hat{B}) = S^2 e (x'x)^{-1}$$

$$S^2 e = \frac{e'e}{n-k-1} = \frac{Y'Y - \hat{B}'X'Y}{n-k-1} = \frac{\sum y^2 - (\hat{B}_1 \sum x_1 y + \hat{B}_2 \sum x_2 y)}{n-k-1}$$

ب – بالنسبة الى $\hat{B}_2$:

$$t_{\hat{B}_2} = \frac{\hat{B}_2}{S_{\hat{B}_2}}$$

$$S_{\hat{B}_2} = \sqrt{S^2{}_{\hat{B}_2}}$$

$$S^2{}_{\hat{B}_2} = \operatorname{var}(\hat{B}_2) = S^2 e a_{22}$$

$$S^2 e = \frac{e'e}{n-k-1}$$

معامل التحديد R^2 Multiple Coefficient of determination

ويعد مؤشر أساس في تقييم مدى معنوية العلاقة بين المتغير التابع (Y) والمتغيرات المستقلة (X_K) إذ (k ، ... ، k = 1) ، بعبارة أخرى هو مقياس يوضح نسبة مساهمة المتغيرات المستقلة في تفسير التغير الحاصل في المتغير التابع . ويمكن اشتقاقه باستخدام المصفوفات بالانحرافات كآلاتي :

$$\because y = x\hat{B} + e$$

$$e = y - \hat{B}$$

$$e'e = (y - x\hat{B})'(y - x\hat{B})$$

$$e'e = y'y - y'x\hat{B} - x'\hat{B}'y + \hat{B}'x'x\hat{B}$$

وبما أن التحديد الثاني الثالث قيمة واحدة كما وان كلا منها يمثل مبدلا للآخر فان :

$$\therefore e'e = y'y - 2\hat{B}x'y + \hat{B}'x'y + \hat{B}'x'x\hat{B}$$

$$\therefore \hat{B} = (x'x)^{-1}x'y$$

$$(x'x) = \hat{B} = x'y$$

$$e'e = y'y - 2\hat{B}'x'y + \hat{B}'x'y$$

$$e'e = y'y - \hat{B}'x'y$$

بذلك يمكن كتابة معادلة الانحرافات الكلية كآلاتي :

$$y'y = \hat{B}x'y - e'e$$

إذ أن :

$y'y$: تمثل الانحرافات الكلية .

$\hat{B}'x'y$: تمثل الانحرافات الموضحة من قيل خط الانحدار .

$e'e$: تمثلا الانحرافات غير الموضحة .

وبما أن معامل التحديد R^2 عبارة عن نسبة الانحرافات الموضحة من قيل خط الانحدار إلى الانحرافات الكلية ، Total variation ، فانه يمثل نسبة مجموع مربعات التغير في المتغيرات المستقلة إلى مجموع المربعات الكلية :

$$\therefore R^2 = \frac{\hat{B}x'y}{y'y} = \frac{\hat{B}'x'y}{\sum y^2}$$

$$R^2 = 1 - \frac{e'e}{y'y - n\bar{Y}^2}$$

$$R^2 = \frac{\hat{B}_1 \sum x_1 y + \hat{B}_2 \sum x_2 y}{\sum y^2}$$

أن إضافة متغيرات مستقلة جديدة إلى المعادلة يؤدي إلى رفع قيمة R^2 ، وذلك لثبات قيمة المقام وتغير قيمة البسط بمقدار $(\hat{B}xy)$ غير أن الاستمرار بإضافة المتغيرات المستقلة سيؤدي إلى انخفاض درجات الحرية (n-k-1) ، مما يتطلب استخراج معامل التحديد المعدل أو المصحح R^{-2} وعلى النحو الآتي :

$$\bar{R}^2 = \left[(1-R^2)\frac{n-1}{n-k-1}\right]$$

اختبار إحصائية F ، Statistics–F

يستهدف هذا الاختبار معرفة مدى معنوية العلاقة الخطية بين المتغيرات المستقلة $X_1, X_2, \ldots X_K$ على المتغير التابع Y ، وكما هو الحال في الانحدار البسيط فأنه يعتمد على نوعين من الفروض :

فرضية العدم H0 : وتنص على انعدام العلاقة بين كل متغير من المتغيرات المستقلة ,XK…X1,X2 وبين المتغير التابع Y ، أي :

$$H_0 : \hat{B}_1 = \hat{B}_2 = \hat{B}_K = 0$$

الفرضية البديلة H1 : وتنص على وجود علاقة معنوية بين المتغيرات المستقلة والمتغير التابع ، أي :

$$H_1 : \hat{B}_1 \# \hat{B}_2 \# \ldots \hat{B}_k \# 0$$

والصيغة الرياضية لهذا الاختبار هي :

$$F = \frac{\hat{B}'x'ylk}{e'e\ln - k - 1}$$

or

$$F = \frac{R^2 lk}{1 - R^2 \ln - k - 1}$$

وبعد احتساب قيمة F تقارن مع قيمتها الجدولية بدرجة حرية (k) و (n-k-1) للبسط والمقام ولمستوى معنوية معين . فإذا كانت القيمة المحتسبة اكبر من القيمة الجدولية ترفض H_0 وتقبل H_1 أي أن العلاقة المدروسة معنوية ، وهناك على الأقل متغير مستقل واحد من المتغيرات XK ذو تأثير في Y . أما إذا كانت القيمة المحتسبة اصغر من الجدولية فان ذلك يعني قبول H_0 أي أن العلاقة الخطية

المدروسة غير معنوية أي انه ليس ثمة تأثير من أي متغير من المتغيرات المستقلة على المتغير التابع .

جدول تحليل التباين ، ANOVA :

لغرض الوقوف على تأثير كل من X1 ، X_2 في المتغير التابع Y ، لابد من عمل جدول تحليل التباين لبيان اثر المتغيرين المستقلين X_1، X_2 في النموذج.

جدول تحليل التباين

مصدر التباين	مجموع مربعات الخطا	درجات الحرية	متوسط مربعات الخطا	اختبار F
الانحراف الموضح من قبل x_2, x_1	$\hat{B}'x'y$.	K	$\hat{B}'x'ylk$	$F = \frac{\hat{B}'x'ylk}{e'e\ln - k - 1}$
الانحراف غير الموضح	$e'e$	$n-k-1$	$e'e\ln - k - 1$	
الانحراف الكلي	$y'y$	$n-k$		

قياس حدود الثقة :

لاحتساب حدود الثقة لاية مشاهدة (نقطة) من مشاهدات خط الانحدار للمجتمع او بعبارة اخرى لحساب القيمة المتوسطة الحقيقية ال Y عند مستوى معنوية معين للمتغير المستقل في النموذج . نفترض ان النقطة المراد تقدير حدود ثقتها هي $E(Y_0)$. ولتقدير المجال الذي يمكن ان تقع فيه قيمة $E(Y_0)$ المقابلة لتشكيلة معينة من قيم المتغيرات المستقلة (K) يجب اشتقاق متباينة القيمة $E(Y_0)$.

$$X_0 = [1 X_{01} X_{02} \ldots X_{0K}]$$

$$\hat{Y}_0 = [1 X_{01} X_{02} \ldots X_{0K}] \begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \\ . \\ . \\ \hat{B}_K \end{bmatrix}$$

$$\hat{Y}_0 = \hat{B}_0 + \hat{B}_1 X_{01} + \ldots + \hat{B}_K X_{0K}$$

وباختصار :

$$\hat{Y}_0 = X_0 \hat{B}$$

ولغرض اشتقاق المتباينة الخاصة بتقدير فترات حدود الثقة للقيمة $E(Y_0)$ يجب اشتقاق وتباين القيمة $(\hat{Y}_0)$ وكالاتي :
لايجاد الوسط فاننا ناخذ القيمة المتوقعة ل $(\hat{Y}_0)$:

$$E(\hat{Y}_0) = E(X_0\hat{B})$$
$$E(\hat{Y}_0) = X_0E(\hat{B})$$
$$\because E(\hat{B}) = B$$
$$\therefore E(\hat{Y}_0) = X_0B$$

ولايجاد التباين :

$$\text{Var}\ (\hat{Y}_0) = E\{(\hat{Y}_0 - E(\hat{Y}_0)(\hat{Y}_0 - E(\hat{Y}_0)'\}$$
$$= E\{(\hat{Y}_0 - X_0B)(\hat{Y}_0 - X_0B)'\}$$
$$\because \hat{Y}_0 = X_0\hat{B}$$
$$\therefore \text{var}(\hat{Y}_0) = E\{(X_0\hat{B} - X_0B)'\}$$
$$= X_0\{E(\hat{B}-B)(\hat{B}-B)'\}X_0'$$
$$= \sigma^2 X_0(X'X)^{-1}X_0'$$

واذا رمزنا للقيمة التقديرية لتبلين قيمة حدود الثقة $\text{var}(\hat{Y}_0)$ بالرمز $S^2(\hat{Y}_0)$ ، فان :

$$S^2(\hat{Y}_0) = S^2X_0(X'X)^{-1}X_0'$$

وعليه فان حدود الثقة للقيمة $E(Y_0)$ تكون :

$$E(Y_0) = \hat{Y}_0 \mp t_{\alpha l2}.S(\hat{Y}_0)$$
$$E(Y_0) = X_0\hat{B} \mp T_{\alpha L2}.S(\hat{Y}_0)$$

اولا :التطبيق العملي للانحدار الخطي المتعدد

يستخدم الانحدار المتعدد لمعرفة الاثر او ا لعلاقة بين المتغيرات التفسيرية والمتغير المعتمد (متغير واحد) من خلال تقدير هذه العلاقة وبالشكل الاتي:

$$Y = f\,(X1, X2\ , X3) \qquad \ldots(1)$$

ففي المعادلة السابقة نريد ان نعرف تاثير المتغيرات X1 و X2 وX3 على المتغير Y ، ونستعين بعلم الاحصاء لمعرفة العلاقة السالفة الذكر .

والمعادلة التقديرية حسب نموذج الانحدار الخطي المتعدد (بالنسة للمتغيرات المذكورة في المعادلة 1) تكون وفق الاتي :

$$Y = a + bX1 + cX2 + dX3 + u \quad \ldots\ldots\ldots\ldots\ldots (2)$$

حيث تمثل :

a : معامل التقاطع او الحد الثابت

b ، c ، d : تمثل معاملات معادلة الانحدار الخطي المتعدد (ميول) .

U ، تمثل الخطا القياسي او الخطا العشوائي للنموذج المقدر

ولتقدير النموذج السابق نستخدم طريقة المربعات الاعتيادية (Ordinary Least Squares) ، والبرنامج الاحصائي SPSS يقوم بشكل تلقائي بحساب المعاملات (a و b و c و d)

ثانيا : معنوية معاملات معادلة الانحدار المتعدد

بعد الحصول على نتائج معادلة الانحدار (المعاملات a و b و c و d) يجب علينا ان نبين هل ان هذه المعاملات مقبولة من الناحية الاحصائية (معنوية احصائيا) مع التنويه ان المعنوية تكون لكل معامل على حدة ، لكي نحكم على معنوية معاملات الانحدار نستعين باختبار T او احصائية t ، ومستوى الاحتمالية المقابل لها ، وبرنامج SPSS يقوم تلقائيا باستخراج اختبار t ومستوى الاحتمالية المقابل لها.

ثالثا : المعنوية الاجمالية لنموذج الانحدار

هناك احصائيات تستخدم لمعرفة المعنوية الاجمالية للنموذج ومنها R R^2 R^{2-} ، الاول R هي معامل الارتباط البسيط (يقيس قوة ا لعلاقة بين متغيرين او اكثر) اما R^2 فهو يسمى معامل التحديد ، وهو يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الانحدار الخطي البسيط (متغير مستقل واحد مع متغير معتمد واحد) ، اما R^{2-} فهو يستخدم لتفسير القوة التفسيرية لنموذج الانحدار الخطي المتعدد (لانه ياخذ بنظر الاعتبار عدد المتغيرات المستقلة ولذلك يسمى بالمصحح ، لانه بالاصل مشتق من R^2) . وايضا تستخدم احصائية F للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين (مستوى احتمال).

رابعا : التطبيق العملي باستخدام البرنامج الاحصائي SPSS

اذا توفرت لدينا البيانات الاتية :

السنوات	الكمية Y	السعر X1	الدخل X2	سعر السلعة البديلة X3
1981	40	9	400	10
1982	45	8	500	14
1983	50	9	600	12

1984	55	8	700	13
1985	60	7	800	11
1986	70	6	900	15
1987	65	6	1000	16
1988	65	8	1100	17
1989	75	5	1200	22
1990	75	5	1300	19
1991	80	5	1400	20
1992	100	3	1500	23
1993	90	4	1600	18
1994	95	3	1700	24
1995	85	4	1800	21

تمثل هذه البيانات العلاقة بين الكمية من سلعة معينة (Y) والعوامل المؤثرة علها وهي السعر (X1) ، دخل المستهلك (X2) بالدولار ، سعر السلعة البديلة (X3) .
وحسب النظرية الاقتصادية هناك علاقة بين المتغير المعتمد وهو الكمية المطلوبة والمتغيرات التفسيرية (المستقلة)الاخرى وهي (السعر ، الدخل ، سعر السلعة البديلة)
ويمكن معرفة الاثر او ا لعلاقة بين المتغيرات التفسيرية والمتغير المعتمد من خلال تقدير هذه العلاقة وبالشكل (شكل العلاقة) الاتي:

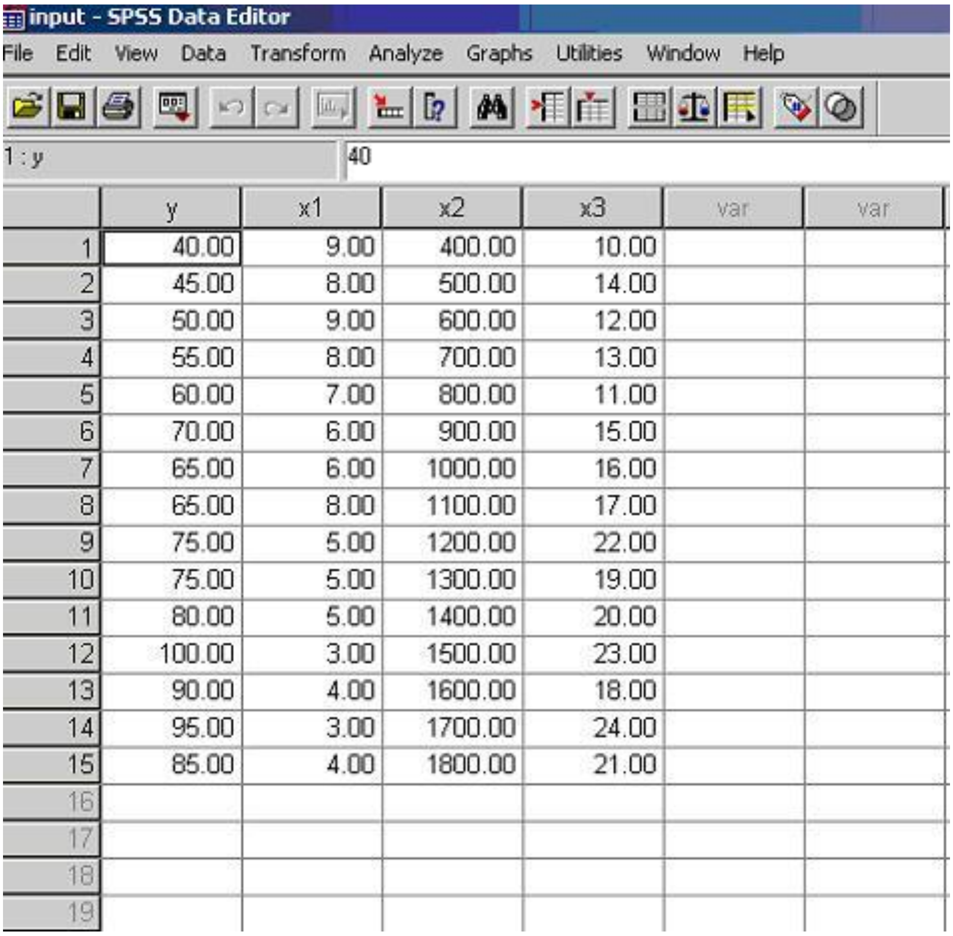
$$Y = f(X1, X2, X3) \qquad \ldots(1)$$

والمعادلة التقديرية حسب نموذج الانحدار الخطي المتعدد تكون وفق الاتي :

$$Y = a + bX1 + cX2 + dX3 + u \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (2)$$

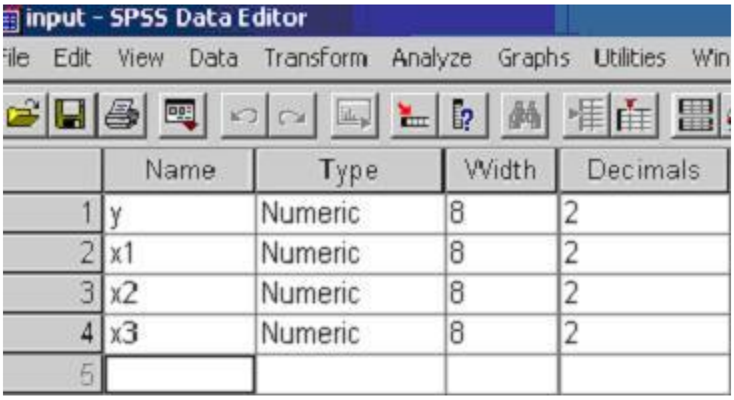
حيث تمثل :
a : معامل التقاطع او الحد الثابت
b ، c ، d : تمثل معاملات دالة معادلة الانحدار الخطي المتعدد
U ، تمثل الخطا القياسي او الخطا العشوائي للنموذج المقدر

ولتقدير النموذج السابق نستخدم طريقة المربعات الاعتيادية (Ordinary Least Squares)
للحصول على معاملات النموذج السالف الذكر في (2) .
ويمكن استخدام البرنامج الاحصائي الممتاز SPSS الاصدار 11.0 في الحصول على نتائج تقدير معادلة الانحدار الخطي المتعدد وكما يلي
اولا : نقوم بادخال البيانات في محرر بيانات SPSS



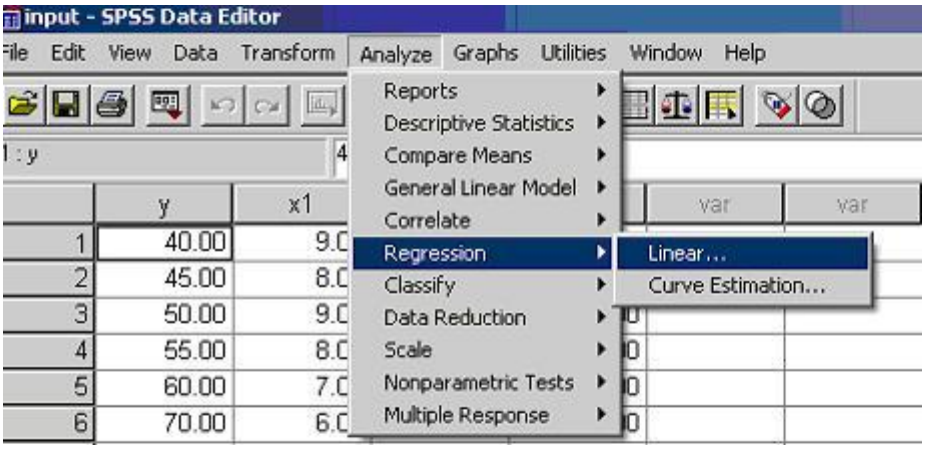
	y	x1	x2	x3	var	var
1	40.00	9.00	400.00	10.00		
2	45.00	8.00	500.00	14.00		
3	50.00	9.00	600.00	12.00		
4	55.00	8.00	700.00	13.00		
5	60.00	7.00	800.00	11.00		
6	70.00	6.00	900.00	15.00		
7	65.00	6.00	1000.00	16.00		
8	65.00	8.00	1100.00	17.00		
9	75.00	5.00	1200.00	22.00		
10	75.00	5.00	1300.00	19.00		
11	80.00	5.00	1400.00	20.00		
12	100.00	3.00	1500.00	23.00		
13	90.00	4.00	1600.00	18.00		
14	95.00	3.00	1700.00	24.00		
15	85.00	4.00	1800.00	21.00		
16						
17						
18						
19						

ثانيا : نقوم بتسمية المتغيرات كما في الشكل الاتي

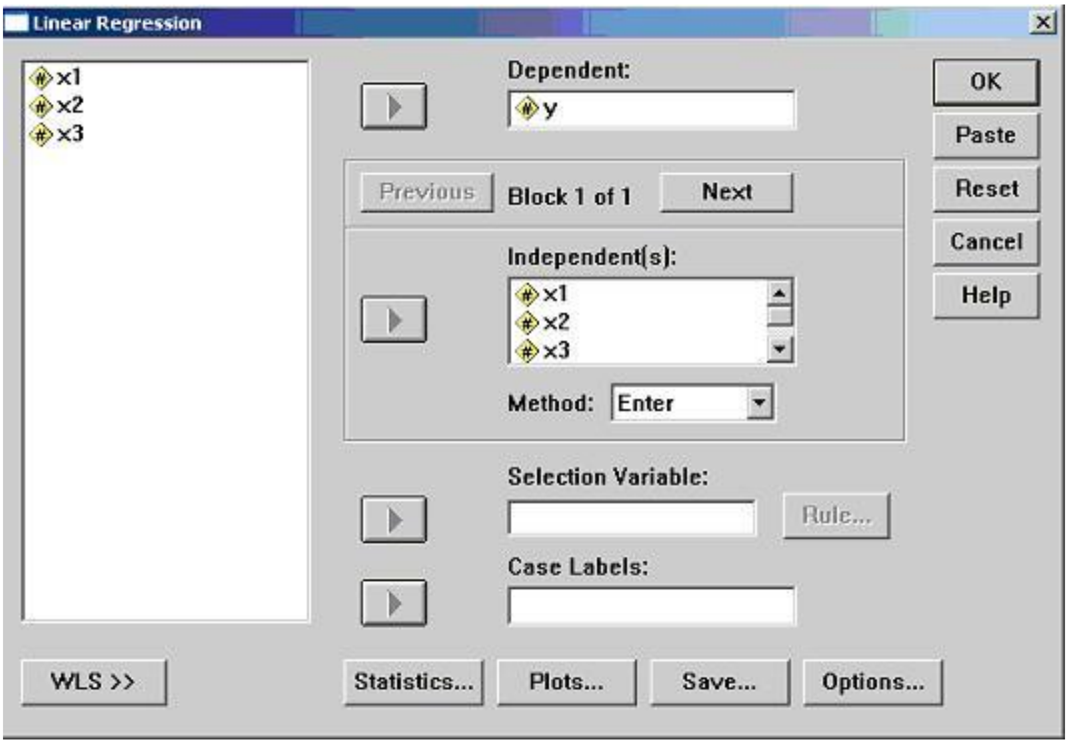


	Name	Type	Width	Decimals
1	y	Numeric	8	2
2	x1	Numeric	8	2
3	x2	Numeric	8	2
4	x3	Numeric	8	2
5				

ثالثا : نذهب الى قائمة analyze ونختار منها الامر Regression ومن القائمة الفرعية نختار Linear ، كما في الشكل الاتي :



رابعا : من نافذة تحليل الانحدار نقوم بتحديد المتغير المعتمد (Y) وننقله الى خانة المتغير المعتمد ، نحدد المتغيرات المستقلة وننقلها الى خانة المتغيرات المستقلة ، ثم ننقر OK كما في الشكل الاتي :



خامسا : سوف نحصل على شاشة المخرجات الآتية :

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X3, X2, X1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.938	4.52761

a. Predictors: (Constant), X3, X2, X1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4374.508	3	1458.169	71.133	.000[a]
	Residual	225.492	11	20.499		
	Total	4600.000	14			

a. Predictors: (Constant), X3, X2, X1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	79.106	19.782		3.999	.002
	X1	-4.928	1.611	-.563	-3.059	.011
	X2	1.590E-02	.007	.392	2.146	.055
	X3	.175	.637	.043	.275	.789

a. Dependent Variable: Y

تحليل النتائج التي حصلنا عليها من SPSS

اولا: التحليل الإحصائي

<u>الجدول الأول</u> يمثل طريقة الانحدار المستخدمة وهي طريقة Enter حيث يتبين ان البرنامج قام بادخال جميع المتغيرات المستقلة في معادلة الانحدار الخطي المتعدد.

<u>الجدول الثاني</u> : يوضح الجدول الثاني قيم معامل الارتباط الثلاثة وهي معامل الارتباط البسيط R حيث بلغ 0.97 ومعامل التحديد R^2 وهو يساوي 0.95 واخيرا معامل التحديد المصحح R^{2-} والذي بلغ 0.94 مما يعني ان المتغيرات المستقلة (التفسيرية) (السعر ، الدخل ، سعر السلعة الاخرى) استطاعت ان تفسر 0.94 من التغيرات الحاصلة في الكمية المطلوبة (Y) والباقي (0.06) يعزى الى عوامل اخرى.

الجدول الثالث: يمثل الجدول الثاني جدول تحليل التباين والذي يمكن المعرفة من خلاله على القوة التفسيرية للنموذج ككل عن طريق احصائية F وكما نشاهد من جدول تحليل التباين المعنوية العالية لاختبار F ‏($P < 0.0001$) . مما يؤكد القوة التفسيرية العالية لنموذج الانحدار الخطي المتعدد من الناحية الاحصائية .

الجدول الرابع : يبين الجدول الرابع والأخير قيم معاملات الانحدار للمقدرات والاختبارات المعنوية الاحصائية لهذه المعاملات ويمكن تلخيص هذا الجدول بالشكل الاتي

المتغير المعتمد	المتغيرات المستقلة			
Y	**الحد الثابت**	**X1**	**X2**	**X3**
قيمة المعامل	79.1	-4.93	1.6	0.17
قيم اختبار T	3.99	-3.059	2.146	0.275
المعنوية	0.002	0.01	0.055	0.789

من الجدول نستنتج ان المتغيرات المستقلة (سعر السلعة) كان معنوي من الناحية الاحصائية وحسب اختبار t (عند مستوى معنوية $P \le 0.05$) ، في حين كاد متغير الدخل ان يكون معنوي (عند مستوى معنوية $P \le 0.05$) . الا ان المتغير المستقل (سعر السلعة البديلة) لم يكن ذو تاثير معنوي في نموذج الانحدار المتعدد وحسب اختبار t .

التحليل الاقتصادي :

حسب منطق النظرية الاقتصادية ، الكمية المطلوبة من سلعة معينة ترتبط بعلاقة عكسية مع السعر ، وبعلاقة طردية مع الدخل ، وبعلاقة طردية مع سعر السلعة البديلة ، ومن النتائج التي حصلنا عليها نجد الاتي : (جميع الاشارات كانت مطابقة مع النظرية الاقتصادية)

ان معامل السعر كان (4.93-) وهذا مطابق لمنطق النظرية الاقتصادية ، مما يعني ان كل زيادة في السعر بمقدار دولار واحد سيؤدي الى انخفاض الكمية المطلوبة بمقدار 5 وحدات تقريبا (4.93) ، اما فيما يخص الدخل ، ايضا كان مطابق للنظرية الاقتصادية حيث كان (1. 6) مما يعني انه كل زيادة في الدخل بمقدار دولار واحد ستؤدي الى ارتفاع الكمية المطلوبة بمقدار (1.6) وحدة ، واخيرا بالنسبة لعامل السلعة البديلة ، نجد انه ايضا مطابق للنظرية الاقتصادية حيث بلغت

قيمته (0.17) ، أي انه اذا ازداد الكمية المطلوبة من السلعة بمقدار وحدة واحدة فان الطلب على السلعة البديلة سوف يزداد بمقدار 0.17 وحدة .

والله ولي التوفيق

والسلام عليكم ورحمة الله وبركاته

مع تحيات أخوكم: الشمري