تحليل الانحدار الخطي Linear Regression Analysis

المحور الثالث:

نموذج الانحدار الخطي المتعدد (النموذج الخطي المتعدد)

(خطوات صياغة نموذج متعدد، تقدير معلمات النموذج، دراسة صلاحية النموذج)

3–تحليل الانحدار الخطي العام (النموذج الخطي المتعدد)

3|1 الكتابة الرياضية للنموذج:

يستند النموذج الخطي المتعدد في التحليل الساكن على افتراض وجود علاقة خطية بين متغير داخلي Y وعدد k $(k \geq 2)$من المتغيرات الخارجية المستقلة $X_1, X_2, ..., X_k$.

نعتبر النموذج التالي الذي يكتب بشكل مختصر كما يلي: $Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i : i = \overline{1,n}$

$\beta_0, \beta_1, \beta_2, ..., \beta_k$ هي معالم (وسائط) النموذج يطلب تقديرها.

ε_i هي الأخطاء العشوائية (الظاهرة غير المشاهدة).

3|2 الشكل المصفوفي لنموذج الانحدار الخطي المتعدد:

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_k X_{1k} + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + \varepsilon_2 \\ .... \\ .... \\ Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \varepsilon_n \end{cases} \Leftrightarrow (Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i : i = \overline{1,n}) =$$

ويمكن تحويل هذه الجملة إلى الشكل المصفوفي التالي:

$$\underset{[n\times1]}{Y} = \underset{[n\times(k+1)]}{X} \bullet \underset{[(k+1)\times1]}{\beta} + \underset{[n\times1]}{\varepsilon} \Leftrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} \cdots\cdots & X_{1k} \\ 1 & X_{21} & X_{22} \cdots\cdots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \cdots\cdots & X_{nk} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} .....[1]$$

3|3 الفرضيات الأساسية لطريقة المربعات الصغرى (OLS) **:**

لطريقة المربعات الصغرى العادية (Ordenary least squares method) عدة فرضيات أساسية، والإخلال بإحداها قد يؤدي إلى نتائج مظللة، خاصة عند دراسة صلاحية النموذج. وسنتطرق إلى هذه الفرضيات، كما سنرى كيفية معالجة بعض المشاكل الناتجة عن الإخلال بإحداهن. وفيما يلي عرض لهذه الفرضيات[1] :

الفرضية الأولى: خطية العلاقة بين المتغير التابع والمتغيرات المستقلة.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_k X_{ik} + \varepsilon_i : i = \overline{1,n} =$$

الفرضية الثانية: التوقع الرياضي للمتغيرات المفسرة المهملة معدوم، بمعنى: $E(\varepsilon^2) = 0$

---

[1]- جلاطو جيلا لي، **الإحصاء التطبيقي**، دار الخلدونية للنشر والتوزيع، القبة القديمة، الجزائر، الطبعة الأول ى2007، ص16-17.

نستنتج أن التوقع الرياضي لـ Y : هو: $\hat{Y} = E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k$

الفرضية الثالثة: تباين الأخطاء متجانس، بمعنى أن: $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 : i = \overline{1,n}$

الفرضية الرابعة: استقلالية الأخطاء، بمعنى أن: $Cov(\varepsilon_i, \varepsilon_j) = 0 : i \neq j$

وبناء على فرضيتي تجانس تباين الأخطاء واستقلالية البواقي، فان مصفوفة التباين – التباين المشترك تكون على النحو التالي:

$$\varepsilon\varepsilon' = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} (\varepsilon_1 \varepsilon_2 \dots \varepsilon_n) = \begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{pmatrix}$$

إذن: $$\begin{cases} Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 : i = \overline{1,n} \\ Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0 : i \neq j \end{cases} \Rightarrow \Omega_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0\dots & 0 \\ 0 & \sigma_\varepsilon^2 & 0\dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0\cdots & \sigma_\varepsilon^2 \end{bmatrix} = \sigma_\varepsilon^2 I_n$$

أي أن تباين البواقي يعطى بالعلاقة التالية: $$\boxed{\Omega_\varepsilon = Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I_n}$$

تسمى Ω_ε مصفوفة التباين – التباين المشترك للأخطاء

الفرضية الخامسة: عدم الارتباط بين المتغيرات المستقلة والبواقي: $Cov(X_i, \varepsilon_i) = E(X_i \varepsilon_i) = X_i E(\varepsilon_i) = 0$

الفرضية السادسة: **المتغير** العشوائي يتبع التوزيع الطبيعي $\varepsilon \to N(0, \sigma_\varepsilon^2 I_n)$

الفرضية السابعة: أن يكون حجم العينة أكبر بكثير من عدد المعالم المقدرة في النموذج: $n \rangle\rangle (k+1)$.

4|3 تقدير شعاع المعالم β وتباين الأخطاء σ_ε^2 بالطريقة (OLS) :

1|4|3 تقدير مركبات شعاع المعالم β :

ليكن النموذج $[1]$......$Y = X\beta + \varepsilon$ ، حيث ε كمية عشوائية غير معلومة.

وليكن $\hat{\beta}$ الشعاع المقدر للشعاع β عندئذ يكتب النموذج المقدر بالشكل: $\boxed{\hat{Y} = X\hat{\beta}}$

و نرمز لشعاع البواقي(Résidus) بالرمز $\hat{\varepsilon}$ ، وهو الفرق بين القيمة الحقيقية Y والقيمة المقدرة $\hat{Y}$.

بمعنى أن: $$\boxed{\hat{\varepsilon} = Y - \hat{Y} \Leftrightarrow \hat{\varepsilon} = Y - X\hat{\beta} .... [2]}$$

ولتقدير معالم النموذج $Y = X\beta + \varepsilon$.......[1] بمفهوم المربعات الصغرى العادية، نقوم بتدنئة مجموع مربعات البواقي الذي يكتب من الشكل: $\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$ حيث $\hat{\varepsilon}'$ هو منقول الشعاع $\hat{\varepsilon}$.

وبالتالي:
$$Min\sum_{i=1}^{n} \varepsilon_i^2 = Min\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

وبنفس الطريقة المطبقة في نموذج الانحدار الخطي البسيط، يمكن الوصول إلى النتيجة نفسها فيما يتعلق مقدرات النموذج،

$$\boxed{\hat{\beta} = (X'X)^{-1} X'Y..................[3]}$$

ويتميز الشعاع المقدر $\hat{\beta}$ بأنه عشوائي وذلك لارتباطه خطيا بالحد العشوائي ε.

3|4|1أ) الخصائص الإحصائية للمقدرات $\hat{\beta}$ و σ_ε^2 :

هل تعطي طريقة المربعات الصغرى العادية (OLS) أحسن تقديرات خطية، غير متحيزة، وذات أصغر تباين؟

نظرية Gauss–Markov[2]**:**

في عائلة التقديرات الخطية وغير المتحيزة فان مقدر طريقة المربعات الصغرى يتميز بأنه أحسن تقدير خطي غير متحيز وذو أصغر تباين، والتي تعني باللغة الانجليزية:" Best Linear Unbiased Estimators" "**BLUE**".

3|4|1ب) مصفوفة التباين – التباين المشترك للمقدر $\hat{\beta}$:

لدينا $(k+1)$معلم مقدر في النموذج. نرمز إلى مصفوفة التباين – التباين المشترك للشعاع المقدر $\hat{\beta}$ بالرمز $\Omega_{\hat{\beta}}$، وهي مصفوفة متناظرة وتعطى بالصيغة:

$$\boxed{\Omega_{\hat{\beta}} = \sigma_\varepsilon^2 (XX')^{-1}}$$

المصفوفة $\Omega_{\hat{\beta}}$ غير معروفة لأنها مرتبطة بتباين الأخطاء النظري σ_ε^2، مما يستوجب تقدير تباين الأخطاء للحصول على تباين البواقي.

3|4|2 تقدير تباين الأخطاء σ_ε^2 :

يكون من الضروري إيجاد مقدر تباين البواقي المجهول. فإذا كان $(k+1)$عدد المعلمات المطلوب تقديرها في النموذج الخطي [1] فإن:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-k-1)} = \frac{1}{(n-k-1)} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 ....[3']$$

وعليه :

$$\boxed{\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_\varepsilon^2 (X'X)^{-1}...................[4]}$$

[2] - Christian Labrousse**, Introduction à l'économie ،Dunod** ,Paris, 1985, p26

3|4|3 تطبيق على نموذج خطي متعدد (بمتغيرين مستقلين)

البيانات الاتية تمثل الناتج المحلي الإجمالي Y_i ويمثل X_1 قيمة الانتاج الزراعي، X_2 يمثل قيمة الانتاج الصناعي لإحدى الدول كما يأتي:

Y_i	X_1	X_2
66	38	47.5
43	41	21.3
36	34	36.5
23	35	18
22	31	29.5
14	34	14.2
12	29	21
7.6	32	10
8.9	34	15
11	38	18

الحل: من هذه البيانات نكون جدول جديد يحتوي على مربعات المتغيرات والضرب ما بينهما كما يأتي:

Y_i	X_1	X_2	Y_i^2	X_1^2	X_2^2	$Y_i X_1$	$Y_i X_2$	X_1X_2
66	38	47.5	4356.00	1444	2256.25	2508.0	3135.0	1805.0
43	41	21.3	1849.00	1681	453.69	1763.0	915.9	873.3
36	34	36.5	1296.00	1156	1332.25	1224.0	1314.0	1241.0
23	35	18	529.00	1225	324.00	805.0	414.0	630.0
22	31	29.5	484.00	961	870.25	682.0	649.0	914.5
14	34	14.2	196.00	1156	201.64	476.0	198.8	482.8
12	29	21	144.00	841	441.00	348.0	252.0	609.0
7.6	32	10	57.76	1024	100.00	243.2	76.0	320.0
8.9	34	15	79.21	1156	225.00	302.6	133.5	510.0
11	38	18	121.00	1444	324.00	418.0	198.0	684.0
$\sum Y_i = 243.5$	$\sum X_1 = 346$	$\sum X_2 = 231$	$\sum Y_i^2 = 9111.97$	$\sum X_1^2 = 12088$	$\sum X_2^2 = 6528.08$	$\sum Y_i X_1 = 8967.8$	$\sum Y_i X_2 = 7286.2$	$\sum X_1 X_2 = 8069.6$

3|4|3أ) طريقة المعادلات الاعتيادية او الطبيعية

من اشتقاق المعادلات الطبيعية او الاعتيادية في هذا الفصل والتي اخذت الشكل الاتي:

$$=\sum Y_i = n\beta_0 + \beta_1 \sum X_1 + \beta_2 \sum X_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$
$$=\sum X_1 Y_i = \beta_0 \sum X_1 + \beta_1 \sum X_1^2 + \beta_2 \sum X_1 X_2 \ldots\ldots\ldots\ldots (2)$$
$$=\sum X_2 Y_i = \beta_0 \sum X_2 + \beta_1 \sum X_2 X_1 + \beta_2 \sum X_2^2 \ldots\ldots\ldots\ldots (3)$$

نعوض بالقيم التي تم احتسابها في الجدول اعلاه كما يأتي:

$$243.5 = 10\beta_0 + 346\beta_1 + 231\beta_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$
$$8967.8 = 346\beta_0 + 12088\beta_1 + 8069.6\beta_2 \ldots\ldots\ldots\ldots (2)$$
$$7286.2 = 231\beta_0 + 8069.6\beta_1 + 6528.08\beta_2 \ldots\ldots\ldots\ldots (3)$$

حتى ننظم الحل يتم اول ايجاد قيمة معلمة من المعادلتين الاوليتين كما يأتي:

$$\boldsymbol{243.5 = 10\beta_0 + \beta_1 346 + \beta_2 231 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)}$$
$$\boldsymbol{8967.8 = 346\beta_0 + 12088\beta_1 + 8069.6\beta_2 \ldots\ldots\ldots\ldots. (2)}$$

لحذف معلمة الثابت β_0 من المعادلتين نقســم قيمة معلمة الثابت فب المعادلة رقم 2 على قيمة معلمة الثابت في المعادلة رقم 1 كما يأتي:346/10=34.6

نضرب القيمة المتحصل عليها 34.6 في المعادلة رقم 1 كما يأتي:

$$\boldsymbol{243.5(34.6) = 10(34.6)\beta_0 + 346(34.6)\beta_1 + 231(34.6)\beta_2}$$
$$\boldsymbol{8494.3 = 346\beta_0 + 11971.6\beta_1 + 7992.6\beta_2 \ldots\ldots\ldots (4)}$$

نطرح المعادلة رقم 4 من المعادلة رقم 2:

$$\boldsymbol{8967.8 = \cancel{346\beta_0} + 12088\beta_1 + 8069.6\beta_2 \ldots\ldots\ldots\ldots. (2)}$$
$$\boldsymbol{-8494.3 = \cancel{346\beta_0} + 11971.6\beta_1 + 7992.6\beta_2 \ldots\ldots\ldots (4)}$$
$$\boldsymbol{473.5 = 116.4\beta_1 + 77\beta_2 \ldots\ldots.. (5)}$$

الان لنحذف معلمة الثابت من المعادلة رقم 3 وكما أسلفنا في الطرقة اعلاه اذ نقسم معلمة الثابت في المعادلة رقم 3 على قيمة المعلمة في المعادلة 1 كما يأتي:231/10=23.1

نضرب هذه القيمة المتحصل عليها في المعادلة 1:

$$\boldsymbol{243.5(23.1) = 10(23.1)\beta_0 + 346(23.1)\beta_1 + 231(23.1)\beta_2}$$
$$\boldsymbol{5624.85 = 231\beta_0 + 7992.6\beta_1 + 5336.1\beta_2 \ldots\ldots. (6)}$$

نطرح المعادلة رقم 6 من المعادلة رقم 3 :

$$\boldsymbol{7286.2 = \cancel{231\beta_0} + 8069.6\beta_1 + 6528.08\beta_2 \ldots\ldots..\ldots.. (3)}$$
$$\boldsymbol{- \quad 5624.85 = \cancel{231\beta_0} + 7992.6\beta_1 + 5336.1\beta_2 \ldots\ldots\ldots\ldots. (6)}$$
$$\boldsymbol{1661.35 = 77\beta_1 + 1191.98\beta_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (7)}$$

لنحتسب قيمة المعلمة β_2 عن طريق حذف المعلمة β_1 وذلك باستخدام المعادلتين 5 و7 :

$$\boldsymbol{473.5 = 116.4\beta_1 + 77\beta_2 \ldots\ldots\ldots (5)}$$
$$\boldsymbol{1661.35 = 77\beta_1 + 1191.98\beta_2 .. (7)}$$

نقسم قيمة المعلمة β_1 في المعادلة رقم 7 على قيمتها في المعادلة رقم 5 :116.4/77=1.512

نضرب حاصل القسمة 1.512 في المعادلة رقم 7:

$$\boldsymbol{1661.35 = 77\beta_1 + 1191.98\beta_2 .. (7) * (1.512)}$$
$$\boldsymbol{1661.35(1.512) = 77\beta_1(1.512) + 1191.98\beta_2(1512) .. (7)}$$
$$\boldsymbol{2511.44 = 116.4\beta_1 + 1801.9\beta_2 \ldots\ldots\ldots. (8)}$$

نطرح المعادلة رقم 5 من المعادلة رقم 8:

$$\begin{aligned} 2511.44 &= 116.4\boldsymbol{\beta_1} + 1801.9\boldsymbol{\beta_2} \ldots\ldots\ldots.(8) \\ - \quad 473.5 &= 116.4\boldsymbol{\beta_1} + 77\boldsymbol{\beta_2} \ldots\ldots\ldots(5) \end{aligned}$$

$$\mathbf{2037.94 = 1724.9}\boldsymbol{\beta_2}$$

$$\therefore \boldsymbol{\hat{\beta}_2} = \frac{\mathbf{2037.94}}{\mathbf{1724.9}} = \mathbf{1.2}$$

نعوض قيمة المعلمة $\hat{\beta}_2$ في المعادلة رقم 8 كما يأتي :

$$\mathbf{2511.44 = 116.4}\boldsymbol{\beta_1} + \mathbf{1801.9(1.2)}$$

$$\mathbf{2511.44 = 116.4}\boldsymbol{\beta_1} + \mathbf{2162.28}$$

$$\mathbf{2511.44 - 2162.28 = 116.4}\boldsymbol{\beta_1}$$

$$\mathbf{349.16 = 116.4}\boldsymbol{\beta_1}$$

$$\therefore \boldsymbol{\hat{\beta}_1} = \frac{\mathbf{349.16}}{\mathbf{116.4}}\mathbf{2.98}$$

نعوض قيمة المعلمتين $\hat{\beta}_1$ و $\hat{\beta}_2$ في المعادلة رقم 1 كما يأتي :

$$\mathbf{243.5 = 10}\boldsymbol{\beta_0} + \mathbf{346(2.98) + 231(1.2)} \ldots\ldots\ldots \mathbf{(1)}$$

$$\mathbf{243.5 = 10}\boldsymbol{\beta_0} + \mathbf{1031.08 + 277.2 = 10}\boldsymbol{\beta_0} + \mathbf{1308.28}$$

$$\mathbf{243.5 - 1308.28 = 10}\boldsymbol{\beta_0}$$

$$\mathbf{-1064.78 = 10}\boldsymbol{\beta_0}$$

$$\therefore \boldsymbol{\hat{\beta}_0} = \frac{\mathbf{-1064.78}}{\mathbf{10}} = \mathbf{-106.478}$$

وبالتالي اخذ النموذج المقدر الصيغة الرياضية الاتية:

$$\hat{Y}_i = -106.478 + 2.98X_1 + 1.2X_2$$

3|4|3ب) طريقة المصفوفات: من صيغة المصفوفات الاتية: $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = [\acute{X}X]^{-1}\acute{X}Y$

$$\boxed{= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_2X_1 & \sum X_2^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum Y_i \\ \sum X_1Y_i \\ \sum X_2Y_i \end{bmatrix}}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 10 & 346 & 231 \\ 346 & 12088 & 8069.6 \\ 231 & 8069.6 & 6528.08 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 243.5 \\ 8967.8 \\ 7286.2 \end{bmatrix} = ؟$$

لتقدير معكوس المصفوفة $[\acute{X}X]^{-1}$ يجب ان نحتسب المحدد اولا ومن ثم مصفوفة المرافقات كما يأتي:

محدد المصفوفة D:

$$D = \left|\acute{X}X\right| \begin{vmatrix} 10 & 346 & 231 \\ 346 & 12088 & 8069.6 \\ 231 & 8069.6 & 6528.08 \end{vmatrix}$$

$$D = 10\begin{vmatrix} 12088 & 8069.6 \\ 8069.6 & 6528.08 \end{vmatrix} - 346\begin{vmatrix} 346 & 8069.6 \\ 231 & 6528.08 \end{vmatrix} + 231\begin{vmatrix} 346 & 12088 \\ 231 & 8069.6 \end{vmatrix}$$

$$D = 137929868.8 - 136544775.68 - 56918.4 = 1328174.72$$

$$\therefore D = 1328174.72$$

مصفوفة المرافقات $Adj\left[\acute{X}X\right]$ كما يأتي:

$$Adj\left[\acute{X}X\right] = \begin{bmatrix} +\begin{vmatrix} 12088 & 8069.6 \\ 8069.6 & 6528.08 \end{vmatrix} & -\begin{vmatrix} 346 & 8069.6 \\ 231 & 6528.08 \end{vmatrix} & +\begin{vmatrix} 346 & 12088 \\ 231 & 8069.6 \end{vmatrix} \\ -\begin{vmatrix} 346 & 231 \\ 8069.6 & 6528.08 \end{vmatrix} & +\begin{vmatrix} 10 & 231 \\ 231 & 6528.08 \end{vmatrix} & -\begin{vmatrix} 10 & 346 \\ 231 & 8069.6 \end{vmatrix} \\ +\begin{vmatrix} 346 & 231 \\ 12088 & 8069.6 \end{vmatrix} & -\begin{vmatrix} 10 & 231 \\ 346 & 8069.6 \end{vmatrix} & +\begin{vmatrix} 10 & 346 \\ 346 & 12088 \end{vmatrix} \end{bmatrix}$$

$$Adj\left[\acute{X}X\right] = \begin{bmatrix} 1379298688 & -349638.08 & -246.4 \\ -349638.08 & 11919.8 & -770 \\ -246.4 & -770 & 1164 \end{bmatrix}$$

$$\left[\acute{X}X\right]^{-1} = \frac{Adj\left[\acute{X}X\right]}{\left|\acute{X}X\right|} = \begin{bmatrix} \frac{1379298688}{1328174.72} & \frac{-349638.08}{1328174.72} & \frac{-246.4}{1328174.72} \\ \frac{-349638.08}{1328174.72} & \frac{11919.8}{1328174.72} & \frac{-770}{1328174.72} \\ \frac{-246.4}{1328174.72} & \frac{-770}{1328174.72} & \frac{1164}{1328174.72} \end{bmatrix}$$

$$\left[\acute{X}X\right]^{-1} = \frac{Adj\left[\acute{X}X\right]}{\left|\acute{X}X\right|} = \begin{bmatrix} 1038.49 & -0.263 & -0.00019 \\ -0.263 & 0.009 & -0.00058 \\ -0.00019 & -0.00058 & 0.00088 \end{bmatrix}$$

بما ان المصفوفة مربعة فان منقولها هو يعطي نفس الترتيب ولذا تبقى على نفس حالة عناصرها إذ أن من الواجب ان نقوم بتحويل الى منقول المصفوفة لشروطها في ايجاد المعكوس ومن الصيغ يمكن أن نحصل على تقديرات معالم المعادلة الخطية كما يأتي:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 1038.49 & -0.263 & -0.00019 \\ -0.263 & 0.009 & -0.00058 \\ -0.00019 & -0.00058 & 0.00088 \end{bmatrix} \begin{bmatrix} 243.5 \\ 8967.8 \\ 7286.2 \end{bmatrix} = \begin{bmatrix} -106.478 \\ 2.98 \\ 1.2 \end{bmatrix}$$

3|4|3 ت) طريقة المحددات (قاعدة كرامر)

كما أسلفنا يجب ان نستخرج محددات فرعية عن طريق ابدال عمود $\acute{\mathbf{X}}\mathbf{Y}$ في العمود الاول من المحدد الرئيسي للحصول على المحدد الفرعي الاول ونستبدله بدلا من العمود الثاني فيى المحدد الرئيسي للحصول على المحدد الفرعي الثاني ومن ثم نستبدله

بدلا من العمود الثالث لنحصل على المحدد الفرعي الثالث ومن ثم نحتسب او نقدر المعالم كما عملنا سابقا في التوجه النظري لهذه الطريقة كما يأتي:

$$\boldsymbol{D}=\left|\acute{X}X\right|=\begin{vmatrix}\mathbf{10} & \mathbf{346} & \mathbf{231}\\ \mathbf{346} & \mathbf{12088} & \mathbf{8069.6}\\ \mathbf{231} & \mathbf{8069.6} & \mathbf{6528.08}\end{vmatrix}$$

$$\mathbf{\acute{X}Y}=\begin{bmatrix}\mathbf{243.5}\\ \mathbf{8967.8}\\ \mathbf{7286.2}\end{bmatrix}$$

$$\boldsymbol{D_0}=\begin{vmatrix}\mathbf{243.5} & \mathbf{346} & \mathbf{231}\\ \mathbf{8967.8} & \mathbf{12088} & \mathbf{8069.6}\\ \mathbf{7286.2} & \mathbf{8069.6} & \mathbf{6528.08}\end{vmatrix}$$

$$D_1=\begin{vmatrix}10 & 243.5 & 231\\ 346 & 8967.8 & 8069.6\\ 231 & 7286.2 & 6528.08\end{vmatrix}$$

$$D_2=\begin{vmatrix}10 & 346 & 243.5\\ 346 & 12088 & 8967.8\\ 231 & 8069.6 & 7286.2\end{vmatrix}$$

$$\hat{\beta}_0=\frac{D_0}{D}=\frac{\begin{vmatrix}243.5 & 346 & 231\\ 8967.8 & 12088 & 8069.6\\ 7286.2 & 8069.6 & 6528.08\end{vmatrix}}{\begin{vmatrix}10 & 346 & 231\\ 346 & 12088 & 8069.6\\ 231 & 8069.6 & 6528.08\end{vmatrix}}=-106.478$$

وبالطريقة التي اتبعناها في الحصول على قيمة المحدد نحتسب ايضا قيمة المحددات

$$\hat{\beta}_1=\frac{D_1}{D}=\frac{\begin{vmatrix}10 & 243.5 & 231\\ 346 & 8967.8 & 8069.6\\ 231 & 7286.2 & 6528.08\end{vmatrix}}{\begin{vmatrix}10 & 346 & 231\\ 346 & 12088 & 8069.6\\ 231 & 8069.6 & 6528.08\end{vmatrix}}=2.98$$

$$\hat{\beta}_2=\frac{D_2}{D}=\frac{\begin{vmatrix}10 & 346 & 243.5\\ 346 & 12088 & 8967.8\\ 231 & 8069.6 & 7286.2\end{vmatrix}}{\begin{vmatrix}10 & 346 & 231\\ 346 & 12088 & 8069.6\\ 231 & 8069.6 & 6528.08\end{vmatrix}}=1.2$$

3|4|4 توزيعات المقدرات وبناء مجالات الثقة لها:

إذا افترضنا أن $\beta_j\ :\ j=\overline{0,k}$ معلم في النموذج فإن:

3|4|4أ) –**إذا كان** $n \leq 30$، **يكون التباين** σ_ε^2 **غير معروف:** $(\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\hat{\beta}_j} \rightarrow t(n-k-1,\alpha)$

أما مجال الثقة فيمكن استخلاصه بالطريقة التالية:

$$\Pr\left\{-t_{\alpha/2} \leq (\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\hat{\beta}_j} \leq t_{\alpha/2}\right\} = 1-\alpha : j = \overline{0,k}$$

ومنه :

$$\Pr\left\{\hat{\beta}_j - t_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j}\right\} = 1-\alpha : j = \overline{0,k}$$

إذن مجال الثقة للمقدر $\boldsymbol{\beta_j}$ هو: $\left[\hat{\beta}_j \pm t_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j}\right]$.

$t_{\alpha/2}$: هي القيمة الحرجة لتوزيع ستودنت بدرجة حرية $\nu = n-k-1$ ونسبة دلالة α.

3|4|4ب) – **إذا كان** $n \rangle 30$، **يكون التباين** σ_ε^2 **معروفا، أي أن:** $(\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\hat{\beta}_j} \rightarrow N(0,1)$

أما مجال الثقة فيمكن استخلاصه بالطريقة التالية:

$$\Pr\left\{-z_{\alpha/2} \leq (\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\hat{\beta}_j} \leq z_{\alpha/2}\right\} = 1-\alpha : j = \overline{0,k}$$

ومنه :

$$\Pr\left\{\hat{\beta}_j - z_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + z_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j}\right\} = 1-\alpha : j = \overline{0,k}$$

إذن مجال الثقة للمقدر $\boldsymbol{\beta_j}$ هو: $\left[\hat{\beta}_j \pm z_{\alpha/2}.\hat{\sigma}_{\hat{\beta}_j}\right]$.

ج–تتبع الكمية $(n-k-1)\hat{\sigma}_\varepsilon^2 / \sigma_\varepsilon^2$ توزيع χ^2 بـ درجة حرية $(n-k-1)$، أي أن:

$$(n-k-1)\hat{\sigma}_\varepsilon^2 / \sigma_\varepsilon^2 \rightarrow \chi_\alpha^2(n-k-1)$$

أما مجال الثقة فيمكن استخلاصه بالطريقة التالية:

$$\Pr\left\{ \chi_{\alpha/2}^2 \leq (n-k-1)\hat{\sigma}_\varepsilon^2 / \sigma_\varepsilon^2 \leq \chi_{1-\alpha/2}^2 \right\} = 1-\alpha$$

وبالتالي نحصل على ما يلي:

$$\boxed{\sigma_\varepsilon^2 \in \left[(n-k-1)\hat{\sigma}_\varepsilon^2 / \chi_{1-\alpha/2}^2 , (n-k-1)\hat{\sigma}_\varepsilon^2 / \chi_{\alpha/2}^2\right]}$$

$\chi_{\alpha/2}^2 , \chi_{1-\alpha/2}^2$: هما القيمتان الحرجتان لتوزيع χ^2 بدرجة حرية $\nu = n-k-1$ ونسبة دلالة α.

3|5 تحليل التباين والقدرة التفسيرية للنموذج:

3|5|1 معادلة تحليل التباين:

يمكننا النموذج $Y = X\beta + \varepsilon$ من قياس مدى تمثيل النموذج لمشاهدات العينة، وتستعمل البواقي كمقياس لجودة التوفيق لان قيمتها تعتمد على المتغير التابع Y. ويتغير Y حول وسطه كما يلي:

$$= Y_i = \hat{Y}_i + \hat{\varepsilon}_i \Leftrightarrow Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{\varepsilon}_i : i = \overline{1,n}$$

وبذلك نحصل على ما يلي: $(Y_i - \bar{Y})^2 = (\hat{Y}_i - \bar{Y})^2 + \hat{\varepsilon}^2 : i = \overline{1,n}$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2 \ldots\ldots\ldots\ldots\ldots\ldots[5]$$

ويمكن كتابة معادلة تحليل التباين على النحو التالي:

$$\boxed{Y^t Y = \widehat{Y^t}\hat{Y} + \widehat{\varepsilon^t}\hat{\varepsilon} \Leftrightarrow TSS = ESS + RSS \quad \cdots\cdots [5^!]}$$

جدول تحليل التباين $(ANOVA)$

مصدر التغير	مجموع المربعات	درجة الحرية	المتوسط
$X_1, X_2, \ldots, X_k$	$ESS = \sum(\hat{Y} - \bar{Y})^2$	k	SSR / k
البواقي	$RSS = \sum \hat{\varepsilon}_i^2$	$n-k-1$	$SSE / (n-k-1)$
المجموع	$TSS = ESS + RSS$	$n-1$	$SST / (n-1)$

3|5|2 معامل التحديد المتعدد(**Multiple Coefficient of Determination**):

هو نسبة تأثير كل المتغيرات المفسرة (المستقلة) على المتغير المفسر(التابع) .

ويعرف على انه[3] "نسبة التشتت المفسر إلى التشتت الكلي، وهو مؤشر على النسبة المئوية من التغير الكلي في المتغير التابع التي يمكن تفسيرها بدلالة المتغيرات المستقلة المدرجة في النموذج. ونرمز لهذا المعامل بالرمز R^2".

وهو يقيس جودة التوفيق والارتباط بين محددات الظاهرة المدروسة والظاهرة نفسها.

---

[3] شعوبي محمود فوزي، مرجع سابق، ص20.

▪ **الاشتقاق الرياضي لمعامل التحديد:**

بالتعريف :

$$\boxed{R^2=\frac{\hat{Y}'\hat{Y}}{Y'Y}=\frac{SSR}{SST}\Leftrightarrow R^2=1-\frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y}=1-\frac{SSE}{SST}\text{...........}[6]}$$

▪ **خواص معامل التحديد** R^2 :

1) من معادلة تحليل التباين يكون لدينا: $0\leq R^2\leq 1$.

2) يمكن التمييز بين حالتين كما يلي:

أ- إذا اقتربت قيمة R^2 من الصفر فهو دلالة على ضعف العلاقة بين المتغير التابع والمتغيرات المستقلة، وإذا كان $R^2=0$ فلا وجود للعلاقة بين المتغير التابع والمتغيرات المستقلة الداخلة في النموذج.

ب- وإذا كان $R^2=1$ فان ذلك يشير إلى أن (100%)من التغير في المتغير التابع يعود إلى التغير في المتغيرات المستقلة المدرجة في النموذج. لذلك كلما اقتربت قيمة R^2 من الواحد الصحيح، كلما زادت درجة الثقة في النموذج المقدر.

3) يعاب على المعامل R^2 انه لا يأخذ في الاعتبار عدد المشاهدات ولا عدد المتغيرات.

3|5|3 معامل التحديد المعدل($\overline{R}^2$):

إن إضافة متغير مستقل أو أكثر إلى النموذج المقدر يؤدي إلى زيادة في قيمة R^2 نتيجة لزيادة قيمة البسط في معادلة R^2 وبقاء المقام على حاله دون تغيير. ، وبالتالي قيمة R^2 لا تعبر عن القيمة الحقيقية له، ولتفادي هذا الإشكال أُقتُرح مؤشرُ آخر يأخذ بعين الاعتبار درجة الحرية في النموذج ، ويسمى معامل التحديد المعدل " $\overline{R}^2$" حيث :

$$\boxed{\overline{R}^2=1-(\frac{n-1}{n-k-1})\left(1-R^2\right)}$$

1. في العينات الصغيرة، $k\rangle 1$ فان: $\overline{R}^2\leq R^2$ أما في العينات الكبيرة فإن: $\lim\limits_{n\to+\infty}(\frac{n-1}{n-k-1})=1\Rightarrow \overline{R}^2\approx R^2$
2. يمكن أن تنخفض قيمة R^2 وذلك بإدخال متغيرة خارجية جديدة في النموذج.
3. يمكن أن يأخذ $\overline{R}^2$ قيما سالبة.

4|5|3 العلاقة بين R^2 واختبار F :

$$\boxed{F_C=(\frac{n-k-1}{k})(\frac{R^2}{1-R^2})}$$

يمكن تحديد طبيعة العلاقة بين R^2 و F من خلال بناء جدول تحليل التباين(ANOVA).

جدول(...): يوضح الشكل النهائي لجدول ($ANOVA$)، والعلاقة بين R^2 واختبار F .

Source	SS	$d.f$	MS	F_C
Regression	$ESS=\hat{Y}'\hat{Y}$	k	$R^2Y'Y/(k)$	$F_C=\left(\frac{R^2}{1-R^2}\right)\left(\frac{n-k-1}{k}\right)$
Residuel	$RSS=\hat{\varepsilon}'\hat{\varepsilon}$	$n-k-1$	$(1-R^2)Y'Y/(n-k-1)$	
Total	$TSS=Y'Y$	$n-1$	…………	

3|6 الاختبارات الإحصائية لمعنوية تقديرات المربعات الصغرى:

تناولنا فيما سبق استخدام أسلوب المربعات الصغرى لتقدير معلمات النموذج. إلا أن ذلك يتطلب إجراء بعض الاختبارات الإحصائية حول قيم المعلمات التي تم تقديرها للتأكد من معنويتها الإحصائية.

3|6|1 اختبار المعنوية الكلية للنموذج $(F-Fisher)$:

لاختبار المعنوية الكلية للنموذج الخطي المتعدد $Y=X\beta+\varepsilon$......[1]، أو بالأحرى اختبار ما إذا كانت المتغيرات المستقلة $X_1, X_2, X_3, ..., X_k$ لها تأثير جوهري على المتغير التابع Y، يمكن صياغة الفرضيتين الصفرية والبديلة لها على النحو التالي:

الفرضية الصفرية هي: $H_0 : \beta_1=\beta_2=....=\beta_k=0$

والفرضية البديلة هي: $H_1 : \exists\beta_i\neq 0; i=\overline{1,k}$

*__إحصاءة الاختبار__: $$F_C=\frac{R^2/k}{(1-R^2)/(n-k-1)}\rightarrow F_\alpha(k,n-k-1)$$

حيث يمكن حساب F_C وتحديد قيمة F_α المجدولة بالبحث في جدول القيم الحرجة لتوزيع F عند درجات حرية $(df=n-k-1)$ومستوى معنوية معين 0,05 أو 0,01 .

*__القرار__: يمكن إجراء هذا الاختبار باستخدام جدول تحليل التباين $(ANOVA)$.

- فإذا كان $F_C\rangle F_\alpha$ نرفض الفرضية H_O، بمعنى أن للنموذج ككل معنوية إحصائية بنسبة دلالة α .
- أما إذا كان $F_C\leq F_\alpha$ نقبل الفرضية H_O، بمعنى أن النموذج غير مقبول إحصائيا عند مستوى الدلالة α .

ويمكن اختبار المعنوية الكلية للنموذج اعتمادا على مفهوم P-value حسب القاعدة التالية[4]: " نرفض الفرضية الصفرية إذا كان $p\leq\alpha$ ".

◊ **P-value** : هو احتمال البيانات التي توضح عزوفا عن الفرضية الصفرية، عندما تكون هذه الفرضية صحيحة.

[4] David R. Anderson et autres, p692.

3|6|2 اختبار المعنوية الإحصائية للمعالم:

تركز هذه الفقرة على اختبار معنوية كل متغير مستقل منفردا في تفسير المتغير التابع. ومن الاختبارات الإحصائية التي يمكن تطبيقها في هذا الإطار هو التوزيع الطبيعي Z ، إذا كان حجم العينة كاف $(n\rangle 30)$. وقد يصعب في الواقع العملي على معنوية المتغيرات المستقلة منفردة. ويتم تحويل قيم أي متغير إلى وحدات (t) باستخدام الصيغة التالية:

$$t = (\hat{\beta}_j - \beta_j) / \hat{\sigma}_{\hat{\beta}_j} \rightarrow N(0,1)$$

وتصاغ الفرضية الصفرية والبديلة لها كما يلي: $\begin{cases} H_0 : \forall i = 1,n : \beta_i = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$:[5].

*<u>**إحصاءة الاختبار**</u>: $t_c = \hat{\beta}_i \big/ \hat{\sigma}_{\hat{\beta}_i} \rightarrow t_{\alpha/2}(n-k-1)$

* <u>**القرار**</u>: تتم مقارنة قيمة t_C المحسوبة بقيمة $t_{\alpha/2}$ المجدولة.

- فإذا كان: $|t_c| \geq t_{\alpha/2}$ ، نرفض الفرضية H_O بمعنى أن المعلم β_i يختلف معنويا عن الصفر بنسبة دلالة α .
- وإذا كان: $|t_c|\langle t_{\alpha/2}$ ، نقبل الفرضية H_O . أي أن المعلم β_i ليس له معنوية إحصائية عند مستوى الدلالة α .

كما يمكن اختبار المعنوية الإحصائية للمعلم β_i اعتمادا على مفهوم P-value حسب القاعدة التالية: " نرفض الفرضية H_O إذا كان $p \leq \alpha$ ".

[17]David R. Anderson et autres, p694.