المحور الرابع: نموذج الانحدار الخطي المتعدد[1](Multiple Regression)

يفترض نموذج الانحدار الخطي البسيط وجود متغير مستقل واحد فقط لتفسير المتغير التابع. لكن في الواقع الاقتصادي، غالبًا ما يتأثر المتغير التابع بعدة متغيرات مستقلة وليس بمتغير واحد فقط.

لذلك، سنتناول في هذا الفصل دراسة نموذج الانحدار الخطي المتعدد، وهو النموذج الذي يعتمد على أكثر من متغير مستقل لتفسير المتغير التابع. كما سنتطرق إلى كيفية تقدير معالم هذا النموذج، وتحليل خصائصه، إضافة إلى اختبار دلالته الإحصائية واستخدامه في التنبؤ.

أولا: الكتابة العامة للنموذج

يمكن كتابة نموذج انحدار خطي عام أو متعدد على الشكل التالي:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + ... + \beta_k X_{ki} + u_i, \quad i = 1,2,...,n$$

- X_1 ، X_2 ،... X_k : متغيرات مستقلة (أو شارحة) عددها k .
- β_0 ، β_1 ، β_2 ،... β_k تسمى معالم النموذج وعددها: $(k+1)$.
- n طول العينة.
- u_i متغير عشوائي يعبر عن الأخطاء الموجودة في النموذج (الأخطاء في تفسير المتغير التابع Y) ولذلك تسمى الأخطاء، أو أيضا البواقي لأنه يمكن كتابتها عند كل مشاهدة i على الشكل التالي:

$$u_i = Y_i - \beta_0 - \beta_1 X_{1i} - \beta_1 X_{2i} - ... - \beta_k X_{ki}, \quad i = 1,2,...,n$$

الكتابة المصفوفية (Matrix form) للنموذج:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_1 X_{21} + ... + \beta_k X_{k1} + u_1 \\ Y_2 &= \beta_0 + \beta_1 X_{12} + \beta_1 X_{22} + ... + \beta_k X_{k2} + u_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{1n} + \beta_1 X_{2n} + ... + \beta_k X_{kn} + u_n \end{aligned}$$

[1]. تم الاعتماد في هذا المحور على محاضرات الأستاذ قوري يحي عبد الله، مطبوعة في الاقتصاد القياسي موجهة لطلبة ليسانس تخصص اقتصد كمي، جامعة بومرداس، 2017 / 2018

$$\underset{(n,1)}{Y} = \underset{(n,k+1)}{X}\ \underset{(k\times 1)}{\beta} + \underset{(n\times 1)}{u}$$

مع:

$$Y=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X=\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \quad \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad u=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

ثانيا: فرضيات النموذج

- الفرضية الأولى: النموذج خطي (Linear Model) بالنسبة لمعالم النموذج، وشعاع الأخطاء ويعني هذا أنه يمكن كتابته على الشكل:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + ... + \beta_k X_{ki} + u_i, \quad i = 1,2,...,n$$

- الفرضية الثانية: قيم المتغير X_{ji} $\left(j = 1,2,..,k, i = 1,2,...,n\right)$ حقيقية مشاهدة بدون أخطاء (غير عشوائية).

- الفرضية الثالثة: الأمل الرياضي للأخطاء معدوم، أو بعبارة أخرى: متوسط الأخطاء معدوم:

$$E(u_i) = 0, \quad \forall i = 1,.....n$$

- الفرضية الرابعة: تجانس أو ثبات تباين الأخطاء Homoscedasticity: وتعني أن تشتت الأخطاء حول المتوسط ثابت، ويعبر عنها رياضيا بالكتابة:

$$V(u_i) = E(u_i^2) = \sigma_u^2 \quad , \forall i = 1.....n$$

- الفرضية الخامسة: عدم وجود ارتباط ذاتي بين الأخطاء: بمعنى أن التباينات المشتركة لأخطاء الملاحظات المختلفة تكون معدومة، وهذا على طول العينة، ونعبر عنها رياضيا كما يلي:

$$Cov(u_s, u_t) = E(u_s u_t) = 0 \quad , \forall s \neq t \qquad s,t = 1.....n$$

الفرضية الرابعة والخامسة يمكن التعبير عنهما معا باستعمال مصفوفة تباين–تباين مشترك للأخطاء والتي يرمز لها بالرمز Ω_u كما يلي:

$$\Omega_u = E(uu') = \begin{pmatrix} E(u_1u_1) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2u_2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(u_nu_1) & E(u_nu_2) & 0 & E(u_nu_n) \end{pmatrix}$$

$$\Omega_u = \begin{pmatrix} V(u_1) & Cov(u_2,u_1) & \cdots & Cov(u_n,u_1) \\ Cov(u_1,u_2) & V(u_2) & \cdots & Cov(u_n,u_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_1,u_n) & Cov(u_2,u_n) & \cdots & V(u_n) \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{pmatrix} = \sigma_u^2 I_n$$

- الفرضية السادسة: الأخطاء مستقلة عن المتغيرات الشارحة: X_j ، أي:

$$Cov(X_{ji}, u_i) = 0, \forall j = 1,2,...,k, \quad i = 1, \ ..., \ n,$$

- الفرضية السابعة: فرضية التوزيع الطبيعي للأخطاء: وهي تعني أن الأخطاء تتوزع وفق القانون الإحتمالي الطبيعي. هذه الفرضية مع الفرضية الأولى والثانية والثالثة يمكن اختصارها رياضيا كما يلي:

$$u_i \rightarrow N(0, \sigma_u^2)$$

- الفرضية الثامنة: عدم وجود ارتباط خطي بين المتغيرات المستقلة وهو ما يستلزم أن محدد المصفوفة $X'X$ غير معدوم وبالتالي يمكن ايجاد المصفوفة $(X'X)^{-1}$.

- الفرضية التاسعة: $\frac{1}{n}X'X$ تؤول إلى مصفوفة محدودة غير فردية.

- الفرضية العاشرة: عدد المشاهدات أكبر بكثير من عدد المتغيرات المستقلة (بما في ذلك الحد الثابت).

$$(n > k+1)$$

المحور الرابع: نموذج الانحدار الخطي المتعدد(Multiple Regression)

ثالثا: تقدير معالم النموذج

طريقة المربعات الصغرى العادية

تقوم طريقة المربعات الصغرى العادية على تصغير مجموع مربعات البواقي أو الاخطاء وهي تلك الفروق الموجودة بين القيم الحقيقية Y والقيم المقدرة $\hat{Y}$.

ليكن نموذج الانحدار الخطي العام في كتابته المصفوفية والمكون من k متغير مستقل و n مشاهدة.

$$Y = X\beta + u$$

مقدرات المربعات الصغرى العادية $\hat{\beta}$ تعطى بالعبارة:

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

مع: X' يشير إلى منقول المصفوفة X.

ملاحظات:

- في الكتابة $\hat{\beta}=(X'X)^{-1}X'Y$ يمكن ملاحظة أن:

$$X'X=\begin{bmatrix} n & \sum_{i=1}^{n}X_{1i} & \sum_{i=1}^{n}X_{2i} & \cdots & \sum_{i=1}^{n}X_{ki} \\ \sum_{i=1}^{n}X_{1i} & \sum_{i=1}^{n}X_{1i}^2 & \sum_{i=1}^{n}X_{1i}X_{2i} & \cdots & \sum_{i=1}^{n}X_{1i}X_{ki} \\ \sum_{i=1}^{n}X_{2i} & \sum_{i=1}^{n}X_{2i}X_{1i} & \sum_{i=1}^{n}X_{2i}^2 & \cdots & \sum_{i=1}^{n}X_{2i}X_{ki} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \sum_{i=1}^{n}X_{ki} & \sum_{i=1}^{n}X_{ki}X_{1i} & \sum_{i=1}^{n}X_{ki}X_{2i} & \cdots & \sum_{i=1}^{n}X_{ki}^2 \end{bmatrix},\ X'Y=\begin{bmatrix} \sum_{i=1}^{n}Y_i \\ \sum_{i=1}^{n}X_{1i}Y_i \\ \sum_{i=1}^{n}X_{2i}Y_i \\ \vdots \\ \sum_{i=1}^{n}X_{ki}Y_i \end{bmatrix},\ \hat{\beta}=\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

مثال:

ليكن النموذج الموالي ذو المتغير التابع Y_i والمتغيرات المستقلة X_{1i} ، X_{2i} ، خلال خمس سنوات فقط من أجل تبسيط الحسابات (عمليا ينبغي تكبير حجم العينة).

i	Y_i	X_{1i}	X_{2i}
1	8	3	5
2	1	1	4
3	8	5	6
4	3	2	4
5	5	4	6

- اكتب النموذج بشكل مصفوفي.

-قدر معالم النموذج.

-اكتب معادلة الانحدار الخطي المتعدد للنموذج المقدر.

الحل:

عدد المتغيرات المستقلة في هذا المثال هو: $k=2$: بإضافة الحد الثابت للنموذج نتحصل على: $k+1=3$ ، عدد المشاهدات: $n=5$.

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_1X_{2i}+u_i,\quad i=1,2,...,5$$

المحور الرابع: نموذج الانحدار الخطي المتعدد(Multiple Regression)

i	y_i	x_{1i}	x_{2i}	x_{1i}^2	x_{2i}^2	$x_{1i}x_{2i}$	$x_{1i}y_i$	$x_{2i}y_i$
1	8	3	5	9	25	15	24	40
2	1	1	4	1	16	4	1	4
3	8	5	6	25	36	30	40	48
4	3	2	4	4	16	8	6	12
5	5	4	6	16	36	24	20	30
$\sum$	25	15	25	55	129	81	91	134

i	y_i	$\hat{y}_i$	e_i	e_i^2
1	8	5	3	9
2	1	1.5	-0.5	0.25
3	8	8.5	-0.5	0.25
4	3	4	-1	1
5	5	6	-1	1
$\sum$				11.5

$$\underbrace{\begin{pmatrix}8\\1\\8\\3\\5\end{pmatrix}}_{Y}=\underbrace{\begin{pmatrix}1&3&5\\1&1&4\\1&5&6\\1&2&4\\1&4&5\end{pmatrix}}_{X}\underbrace{\begin{pmatrix}\beta_0\\\beta_1\\\beta_2\end{pmatrix}}_{\beta}+\underbrace{\begin{pmatrix}u_1\\u_2\\u_3\\u_4\\u_5\end{pmatrix}}_{u}$$

- الكتابة المصفوفية:

$$Y=X\beta+u$$

- تقدير معالم النموذج:

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

$$X'X=\begin{pmatrix}n&\sum_{i=1}^{n}X_{1i}&\sum_{i=1}^{n}X_{2i}\\\sum_{i=1}^{n}X_{1i}&\sum_{i=1}^{n}X_{1i}^2&\sum_{i=1}^{n}X_{1i}X_{2i}\\\sum_{i=1}^{n}X_{2i}&\sum_{i=1}^{n}X_{2i}X_{1i}&\sum_{i=1}^{n}X_{2i}^2\end{pmatrix}=\begin{pmatrix}5&15&25\\15&55&81\\25&81&129\end{pmatrix}$$

$$\Rightarrow(X'X)^{-1}=\begin{pmatrix}5&15&25\\15&55&81\\25&81&129\end{pmatrix}^{-1}=\frac{1}{20}\begin{pmatrix}534&90&-160\\90&20&-30\\-160&-30&50\end{pmatrix}$$

$$X'Y=\begin{pmatrix}\sum_{i=1}^{n}Y_i\\\sum_{i=1}^{n}X_{1i}Y_i\\\sum_{i=1}^{n}X_{2i}Y\end{pmatrix}=\begin{pmatrix}25\\91\\134\end{pmatrix}$$

$$\hat{\beta}=(X'X)^{-1}(X'Y)=\frac{1}{20}\begin{pmatrix}534&90&-160\\90&20&-30\\-160&-30&50\end{pmatrix}\begin{pmatrix}25\\91\\134\end{pmatrix}=\begin{pmatrix}5\\2.5\\-1.5\end{pmatrix}$$

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_0\\\hat{\beta}_1\\\hat{\beta}_2\end{pmatrix}=\begin{pmatrix}5\\2.5\\-1.5\end{pmatrix}$$

المحور الرابع: نموذج الانحدار الخطي المتعدد(Multiple Regression)

- **معادلة الإنحدار الخطي المتعدد للنموذج المقدر:**

$$\hat{Y}_i = 5 + 2.5X_{1i} - 1.5X_{2i}, \quad i = 1,2,3,4,5$$

رابعا: تقدير تباين الأخطاء

تقدير تباين الأخطاء $\hat{\sigma}_u^2$ بطريقة المربعات الصغرى العادية يفضي إلى:

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}$$

$$\hat{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{n-k-1} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}$$

خامسا: مصفوفة التباين (التباين المشترك للمقدرات)

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_u^2 (X'X)^{-1} = \begin{pmatrix} \hat{\sigma}^2_{\hat{\beta}_0} & Cov(\hat{\beta}_0,\hat{\beta}_1) & Cov(\hat{\beta}_0,\hat{\beta}_2) & \cdots & Cov(\hat{\beta}_0,\hat{\beta}_k) \\ Cov(\hat{\beta}_1,\hat{\beta}_0) & \hat{\sigma}^2_{\hat{\beta}_1} & Cov(\hat{\beta}_1,\hat{\beta}_2) & \cdots & Cov(\hat{\beta}_1,\hat{\beta}_k) \\ Cov(\hat{\beta}_2,\hat{\beta}_0) & Cov(\hat{\beta}_2,\hat{\beta}_1) & \hat{\sigma}^2_{\hat{\beta}_2} & \cdots & Cov(\hat{\beta}_2,\hat{\beta}_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k,\hat{\beta}_0) & Cov(\hat{\beta}_k,\hat{\beta}_1) & Cov(\hat{\beta}_k,\hat{\beta}_2) & \cdots & \hat{\sigma}^2_{\hat{\beta}_k} \end{pmatrix}$$

نلاحظ من خلال هذه الكتابة أن تباينات المقدرات $\left(\hat{\sigma}^2_{\hat{\beta}_j}\right)$ تتموضع في محور المصفوفة بينما تمثل باقي العناصر التباينات المشتركة بين المقدرات، يمكن أن نكتب تباين المقدر $\hat{\beta}_j$ من الشعاع $\hat{\beta}$ كما يلي:

$$\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2_u a_{jj}$$

مع: a_{jj} هو العنصر الواقع في السطر رقم j والعمود رقم j من المصفوفة: $(XX)^{-1}$.

الانحراف المعياري لـ $\hat{\beta}_j$ يكتب على الشكل:

$$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{\hat{\sigma}^2_u a_{jj}} = \hat{\sigma}_u \sqrt{a_{jj}}$$

مثال (المثال السابق):

- احسب مقدر تباين الأخطاء $\hat{\sigma}^2_u$.
- أحسب مقدر مصفوفة تباين تباين–مشترك للمقدرات $\hat{\Omega}_{\hat{\beta}}$.

الحل:

- حساب مقدر تباين الأخطاء:

$$\hat{\sigma}^2_u = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}$$

$$\hat{\sigma}_u^2 = \frac{11.5}{5-2-1} = \frac{11.5}{2} = 5.75$$

- حساب مقدر مصفوفة تباين تباين–مشترك للمقدرات $\hat{\Omega}_{\hat{\beta}}$.

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_u^2 \left(X'X\right)^{-1} = \begin{pmatrix} \hat{\sigma}^2_{\hat{\beta}_0} & Cov(\hat{\beta}_0,\hat{\beta}_1) & Cov(\hat{\beta}_0,\hat{\beta}_2) \\ Cov(\hat{\beta}_1,\hat{\beta}_0) & \hat{\sigma}^2_{\hat{\beta}_1} & Cov(\hat{\beta}_1,\hat{\beta}_2) \\ Cov(\hat{\beta}_2,\hat{\beta}_0) & Cov(\hat{\beta}_2,\hat{\beta}_1) & \hat{\sigma}^2_{\hat{\beta}_2} \end{pmatrix}$$

$$\hat{\Omega}_{\hat{\beta}} = 5.75\begin{pmatrix} 26.7 & 4.5 & -8 \\ 4.5 & 1 & -1.5 \\ -8 & -1.5 & 2.5 \end{pmatrix} = \begin{pmatrix} 153.5 & 25.8 & -46 \\ 25.87 & 5.75 & -8.6 \\ -46 & -8.6 & 14.3 \end{pmatrix}$$

بالمطابقة نستنتج أن:

$$\hat{\sigma}^2_{\hat{\beta}_0} = 153.5,\ \hat{\sigma}^2_{\hat{\beta}_1} = 5.75,\ \hat{\sigma}^2_{\hat{\beta}_2} = 14.3,\ Cov(\hat{\beta}_0,\hat{\beta}_1) = Cov(\hat{\beta}_1,\hat{\beta}_0) = 25.8,$$

$$Cov(\hat{\beta}_0,\hat{\beta}_2) = Cov(\hat{\beta}_2,\hat{\beta}_0) = -46,\ Cov(\hat{\beta}_1,\hat{\beta}_2) = Cov(\hat{\beta}_2,\hat{\beta}_1) = -8.6$$

سادسا:جدول تحليل التباين

كما في نموذج الانحدار الخطي البسيط يمكن كتابة معادلة تحليل التباين في نموذج الانحدار الخطي المتعدد كما يلي:

$$\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n}\hat{u}_i^2$$

$$\text{TSS} \quad = \quad \text{ESS} \ + \ \text{RSS}$$

TSSترمز إلى مجموع المربعات الكلية، ESSمجموع المربعات الشارحة، RSS مجموع مربعات البواقي.

المحور الرابع: نموذج الانحدار الخطي المتعدد(Multiple Regression)

جدول تحليل التباين في حالة الانحدار الخطي المتعدد يكون على الشكل التالي:

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات
المتغيرات المستقلة	$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$	k	ESS / k
البواقي	$RSS = \sum_{i=1}^{n} \left(\hat{u}_i\right)^2$	$n-k-1$	$RSS / (n-k-1)$
المجموع	$TSS = ESS + RSS$ $= \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$	$n-1$	$\neq$

➢ معامل التحديد

في حالة الإنحدار الخطي المتعدد، يمكن أيضا اختبار جودة التوفيق بحساب معامل التحديد المتعدد R^2، فهو يشير إلى النسبة التي يمكن تفسيرها من التغير الكلي في المتغير التابع Y بدلالة المتغيرات المستقلة المدرجة في المعادلة، ويستعمل كمقياس لجودة التوفيق في نموذج الانحدار المحتوي على k متغير مستقل، ولحسابه يمكن إتباع نفس الطريقة المستعملة في النموذج الخطي البسيط:

لدينا:

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

يمكن حساب R^2 على الشكل:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

يمكن أيضا كتابة معامل التحديد R^2 بدلالة شعاع المقدرات $\hat{\beta}$ كما يلي:

$$R^2 = \frac{\boldsymbol{\beta}' (\mathbf{X}'\mathbf{Y}) - n(\bar{Y})^2}{\mathbf{Y}'\mathbf{Y} - n(\bar{Y})^2}$$

حيث أن:

$$\mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{n} Y_i^2$$

وبذلك يصيح القانون كالتالي:

$$R^2 = \frac{\beta'(\mathbf{X}'\mathbf{Y}) - n(\bar{Y})^2}{\sum Y_i^2 - n(\bar{Y})^2}$$

- معامل التحديد المصحح (المعدل)

إضافة متغيرات مستقلة أخرى للنموذج لا يمكن أن تقلل من قيمة R^2 ، بل يمكن أن تزيد من قيمته نظرا لثبات قيمة TSS، وارتفاع قيمة ESS. ولذلك يستحسن تعديل R^2 إلى معامل التحديد المعدل $\left(\bar{R}^2\right)$ وذلك بعد الأخذ بعين الاعتبار درجات الحرية كما يلي:

$$\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)} = 1 - \frac{\hat{u}'\hat{u}/(n-k-1)}{Y'Y/(n-1)} = 1 - \frac{n-1}{n-k-1}\left(1-R^2\right)$$

إذن عندما تكون درجة الحرية (Degrees of Freedom) ضعيفة وهي الفرق بين عدد المشاهدات وعدد المتغيرات المستقلة (العينات الصغيرة نسبيا)، يتم تعديل معامل التحديد R^2 باستعمال $\bar{R}^2$.
يمكن أن نلاحظ أنه عندما يكون عدد المشاهدات n كبير نسبيا فإن $\bar{R}^2$ يؤول إلى R^2 .

$$\lim_{n\to\infty} \bar{R}^2 = 1 - \lim_{n\to\infty} \frac{n-1}{n-k-1}\left(1-R^2\right) = R^2$$

ملاحظات:

- R يسمى معامل الارتباط المضاعف.
- $0 \le R^2 \le 1$، عندما يكون R^2 قريب من الواحد فهذا يدل على جودة التوفيق والعكس صحيح.
- إذا كان: $RSS = 0$ فهذا يعني أن النموذج عبارة عن خط مستقيم ولا توجد أخطاء في النموذج، أما إذا كانت: $ESS = 0$ فهذا يعني أن المتغيرات المستقلة ليس لها أي تأثير على المتغير التابع.

المحور الرابع: نموذج الانحدار الخطي المتعدد(Multiple Regression)

مثال: (المثال السابق):

- ارسم جدول تحليل التباين.
- احسب معامل التحديد ومعامل التحديد المعدل.

الحل:

- رسم جدول تحليل التباين:

لدينا:

$$TSS = ESS + RSS$$
$$RSS = \sum \hat{u}_i^2 = 11.5$$

مجموع المربعات الكلية (TSS) يمكن حسابها كما في الجدول الموالي:

i	Y_i	$(Y_i - \bar{Y})^2$
1	8	9
2	1	16
3	8	9
4	3	4
5	5	0
		$\sum (Y_i - \bar{Y})^2 = 38$

$$TSS = \sum (Y_i - \bar{Y})^2 = 38$$
$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = TSS - RSS = 38 - 11.5 = 26.5$$

جدول تحليل التباين يكون على الشكل:

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات
المتغيرات المستقلة	$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 = 26.5$	$k = 2$	ESS / k $= 26.5 / 2 = 13.25$
البواقي	$RSS = \sum_{i=1}^{n} \left(\hat{u}_i\right)^2 = 11.5$	$n - k - 1$ $= 2$	$RSS / (n - k - 1)$ $= 11.5 / 2 = 5.75$
المجموع	$\text{TSS} = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = 38$	$n - 1$ $= 4$	$\neq$

- حساب معامل التحديد:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} = \frac{26.5}{38} = 0.697$$

- حساب معامل التحديد المعدل:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right) = 1 - \frac{4}{2}(1 - 0.697) = 0.394$$

سابعا: بناء مجال الثقة للمعالم

يمكن عند مستوى معنوية (α) إيجاد مجال ثقة لكل معلم β_j:

$$\beta_j \in \left[\hat{\beta}_j \pm \hat{\sigma}_{\hat{\beta}_j} \times t_{n-k-1,1-\frac{\alpha}{2}} \right] \quad \forall j = 0,1,...,k$$

$t_{n-k-1,1-\frac{\alpha}{2}}$: القيمة (المقروءة) الحرجة لتوزيع استيودنت عند درجة حرية $n-k-1$ و مستوى معنوية (α).

ثامنا: اختبار الفرضيات

- **اختبار معنوية المعلمات**

اختبار معنوية معلمة من معالم النموذج يقصد بها إختبار هل هذه المعلمة معدومة أو غير معدومة في النموذج، يمكن كتابة فرضية العدم H_0 على الشكل:

$$H_0 : \beta_j = 0$$

ويمكن كتابة الفرضية البديلة H_1 على الشكل:

$$H_1 : \beta_j \neq 0$$

بما أن العلاقة بين Y و X_j قائمة على أساس النموذج الخطي، فإن قبول الفرضية H_0 يعني بأن خط الإنحدار هو عبارة عن خط أفقي، وبالتالي لا توجد علاقة معنوية بين المتغيرين Y و X_j.

الإحصائية المحسوبة:

$$t_c = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \rightarrow t_{n-k-1} \quad \forall j = 0,1,...,k$$

قاعدة القرار:

إذا كانت: $|t_c| = \frac{|\hat{\beta}_j|}{\hat{\sigma}_{\hat{\beta}_j}} > t_{n-k-1,1-\frac{\alpha}{2}}$ عند مستوى معنوية $(\%\alpha)$ ودرجة حرية $(n-k-1)$ ففي هذه الحالة نرفض الفرضية H_0 ونقبل الفرضية H_1، وبالتالي المعلم β_j معنوي إحصائيا.

بالمقابل: إذا كانت: $|t_c| = \frac{|\hat{\beta}_j|}{\hat{\sigma}_{\hat{\beta}_j}} \leq t_{n-k-1,\frac{\alpha}{2}}$ عند مستوى معنوية $(\%\alpha)$ ودرجة حرية $(n-k-1)$ ففي هذه الحالة نقبل الفرضية H_0 ونرفض الفرضية H_1، وبالتالي المعلم β_j ليس له معنوية إحصائية، وبالتالي لا توجد علاقة معنوية بين المتغيرين Y و X_j.

➢ اختبار المعنوية الكلية للنموذج

يهدف هذا الاختبار إلى معرفة مدى المعنوية الكلية للنموذج أي معنوية جميع المتغيرات المستقلة في شرح المتغير التابع في نفس الوقت.

$$\begin{cases} H_0 : \beta_1 = \beta_2 = ... = \beta_k \\ H_1 : \exists \beta_j \ / \quad \beta_j \neq 0 \end{cases}$$

تكتب الإحصائية المحسوبة F_c بالاعتماد على جدول تحليل التباين (نسبة التباين المفسر إلى التباين الغير المفسر) كما يلي:

$$F_c = \frac{ESS/q}{RSS/(n-k-1)} = \frac{R^2/q}{\left(1-R^2\right)/(n-k-1)} \rightarrow F_{(q,n-k-1),\alpha}$$

مثال: (المثال السابق):

- اختبر معنوية المعالم: β_0، β_1، β_2.
- اختبر الفرضية:

$$\begin{cases} H_0 : \beta_1 + \beta_2 = 2 \\ H_1 : \beta_1 + \beta_2 \neq 2 \end{cases}$$

الحل:

- اختبار معنوية: β_0.

$$\begin{cases} \beta_0 = 0 \\ \beta_0 \neq 0 \end{cases}$$

$$t_c = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \rightarrow t_2$$

تحت الفرضية H_0:

$$t_c = \frac{|\hat{\beta}_0|}{\hat{\sigma}_{\hat{\beta}_0}} = \frac{5}{\sqrt{153.5}} = 0.40$$

نلاحظ أنه عند مستوى معنوية (5%) ودرجة حرية (2) $t_c = 0.40 < t_{2,0.975} = 4.30$ وبالتالي نقبل الفرضية H_0 ونرفض الفرضية H_1 وهو ما يعني أن المعلم β_0 غير معنوي في النموذج.

- اختبار معنوية: β_1.

$$\begin{cases} \beta_1 = 0 \\ \beta_1 \neq 0 \end{cases}$$

$$t_c = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \rightarrow t_2$$

تحت الفرضية H_0:

$$t_c = \frac{|\hat{\beta}_1|}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{2.5}{\sqrt{5.75}} = 1.04$$

نلاحظ أنه عند مستوى معنوية (5%) ودرجة حرية (2) ، $t_c = 1.04 < t_{2,0.975} = 4.30$ وبالتالي نقبل الفرضية H_0 ونرفض الفرضية H_1 وهو ما يعني أن المعلم β_1 غير معنوي في النموذج.

- اختبار معنوية: β_2:

$$\begin{cases} \beta_2 = 0 \\ \beta_2 \neq 0 \end{cases}$$

$$t_c = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}} \rightarrow t_2$$

تحت الفرضية H_0:

$$t_c = \frac{|\hat{\beta}_2|}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{|-1.5|}{\sqrt{14.3}} = 0.39$$

نلاحظ أنه عند مستوى معنوية (5%) ودرجة حرية (2) $t_c = 0.39 < t_{2,0.975} = 4.30$ وبالتالي نقبل الفرضية H_0 ونرفض الفرضية H_1 ، وهو ما يعني أن المعلم β_1 غير معنوي في النموذج.

- اختبار الفرضية:

$$\begin{cases} H_0 : \beta_1 + \beta_2 = 2 \\ H_1 : \beta_1 + \beta_2 \neq 2 \end{cases}$$

لدينا:

$$t_c = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 2}{\hat{\sigma}_{\hat{\beta}_1 + \hat{\beta}_2}} \to t_2$$

مع:

$$\hat{\sigma}_{\hat{\beta}_1 + \hat{\beta}_2} = \sqrt{\sigma^2_{\hat{\beta}_1} + \sigma^2_{\hat{\beta}_2} + 2Cov(\hat{\beta}_1, \hat{\beta}_2)} = \sqrt{5.75 + 14.3 + 2(-8.6)} = 1.68$$

$$t_c = \frac{2.5 - 1.5 - 2}{1.68} = -0.59$$

نلاحظ أنه عند مستوى معنوية $(\%5)$ ودرجة حرية (2) $|t_c| = 0.59 < t_{2,0.975} = 4.30$ وبالتالي نقبل الفرضية H_0 ونرفض الفرضية H_1، وبالتالي مجموع المعلمتين β_1، β_2 لا يختلف معنويا عن العدد الحقيقي 2.

- اختبار المعنوية الكلية:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \\ H_1 : \exists \beta_j \ / \ \beta_j \neq 0 \end{cases}$$

$$F_c = \frac{ESS/q}{RSS/(n-k-1)} = \frac{R^2/q}{\left(1-R^2\right)/(n-k-1)} \to F_{(q,n-k-1),\alpha}$$

$$F_c = \frac{ESS/q}{RSS/(n-k-1)} = \frac{26.5/2}{11.5/2} = 2.3$$

نلاحظ أنه عند مستوى معنوية $(\%5)$ ودرجات حرية $(q=2)$ و $(n-k-1=2)$ على الترتيب أن: $F_c = 2.3 < F_{(2,2),0.95} = 19$ ، وبالتالي نقبل الفرضية H_0 ونرفض الفرضية H_1، أي النموذج غير معنوي.

تاسعا: التنبؤ

بما أن المتغيرات المستقلة X_j معرفة خارج النموذج الذي تم تقديره، وبالتالي إذا افترضنا أن X_j معرفة من أجل المشاهدة $n+h$ $(h=1,2,....)$ في هذه الحالة يمكن التنبؤ بقيمة Y_{n+h} وذلك كما يلي:

$$Y_{n+h}=\beta_0+\beta_1X_{1n+h}+\beta_1X_{2n+h}+...+\beta_kX_{kn+j}+u_{n+h},\quad h=1,2,..$$
$$\hat{Y}_{n+h}=\hat{\beta}_0+\hat{\beta}_1X_{1n+h}+\hat{\beta}_1X_{2n+h}+...+\hat{\beta}_kX_{kn+j}\quad h=1,2,..$$

الكتابة المصفوفية تكون على الشكل:

$$Y_{n+h}=X_{n+h}\hat{\beta}+u_{n+h}\quad h=1,2,..$$
$$\hat{Y}_{n+h}=X_{n+h}\hat{\beta}\quad h=1,2,..$$

يمكن أيضا إيجاد مجال ثقة لـ Y_{n+h}:

$$Y_{n+h}\in\hat{Y}_{n+h}\pm t_{n-k-1,\frac{\alpha}{2}}\times\hat{\sigma}_u\times\sqrt{1+X_{n+h}\left(X'X\right)^{-1}X'_{n+h}}$$

مثال: المثال السابق:

- اوجد القيمة التنبئية لـ Y من أجل المشاهدة $i=6$ علما أن: $X_{16}=7,\ X_{26}=9$.
- أوجد مجال ثقة لهذه القيمة التنبئية.

الحل:

- **إيجاد القيمة التنبئية:**

$$\hat{Y}_{5+1}=\hat{Y}_6=\beta_0+\beta_1X_{16}+\beta_1X_{26}=5+2.5\times7-1.5\times9=9$$
$$Y_{n+h}\in\hat{Y}_{n+h}\pm t_{n-k-1,\frac{\alpha}{2}}\times\hat{\sigma}_u\times\sqrt{1+X_{n+h}\left(X'X\right)^{-1}X'_{n+h}}$$
$$X_{n+h}\left(X'X\right)^{-1}X'_{n+h}=\begin{pmatrix}1 & 7 & 9\end{pmatrix}\begin{pmatrix}26.7 & 4.5 & -8\\ 4.5 & 1 & -1.5\\ -8 & -1.5 & 2.5\end{pmatrix}\begin{pmatrix}1\\7\\9\end{pmatrix}=8.2$$

- **مجال ثقة:**

$$Y_6\in9\pm4.30\times\sqrt{5.75}\times\sqrt{1+8.10}\Leftrightarrow Y_6\in\left[-22.10,\ 40.10\right]$$