المحور الثاني: نموذج الانحدار الخطي البسيط

المحور الثالث: نموذج الانحدار الخطي المتعدد

تمهيد:

تعتمد نماذج الانحدار على مقاييس العلاقة وتضفي عليها معنى أكثر عمقا في دراسة الظواهر المختلفة، وذلك بالوصول إلى معادلات (نماذج مقدرة) يمكن استخدامها في عمليات التقدير والتنبؤ.

تعريف الانحدار:

هو دراسة اعتماد المتغير التابع على متغير مستقل أو أكثر، وذلك لتقدير قيمة المتغير التابع من خلال الاستفادة من البيانات المتاحة التي تقدمها قيم المتغير المستقل أو المتغيرات المستقلة. ويؤكد هذا التعريف على أن نموذج الانحدار لا يكتفي بإيضاح العلاقة فقط، بل يسعى لكشف الاعتماد بين المتغيرين، بمعنى: أن أحدهما يؤثر في الآخر.

أهداف تحليل الانحدار:

يمكن القول بأن تحليل الانحدار يسعى إلى تحقيق هدف أو أكثر من الأهداف التالية:

- **الوصف**(Descriptive): من خلال تحديد شكل العلاقة بين المتغير التابع والمتغير أو المتغيرات المستقلة.
- **التقدير**(Estimation): من خلال تحديد أي من المتغيرات المستقلة كان ذا أثر أكبر في قيم المتغير التابع.
- **التنبؤ**(Prediction): تمهد المعطيات التي تقدمها قيم المتغير أو المتغيرات المستقلة للوصول إلى قيم المعلمات التي تستخدم في معادلة التنبؤ بقيم المتغير التابع للظاهرة محل الدراسة.
- **التحكم**(Controlling): وهو الإفادة من الأهداف السابقة في تعديل قيم المتغير المستقل بالزيادة أو النقص، بالشكل الذي يحقق الهدف المنشود في قيم المتغير التابع.

وفي هذا البرنامج سنتناول نماذج الانحدار التالية:

- ✓ الانحدار الخطي البسيط (Simple Linear Regression)،
- ✓ الانحدار الخطي المتعدد (Multiple Linear Regression).

المحور الثاني

نموذج الانحدار الخطي البسيط (Simple Linear Regression)

(تحديد قيمة كل معلمة للنموذج، اختبار الموثوقية، التنبؤ)

2–نموذج الانحدار الخطي البسيط (Simple Linear Regression):

الانحدار الخطي البسيط هو من الأساليب المعتمدة في قياس العلاقات الاقتصادية، يهتم بدراسة وتحليل أثر متغير مستقل واحد على متغير تابع. ويسمى بالخطي لأن الصيغة الممثلة للعلاقة خطية، ووصف بأنه بسيط لأن عدد المتغيرات المستقلة محل الدراسة متغير واحد فقط.

2|1 معادلة النموذج الخطي البسيط

يتناول هذا النوع من النماذج العلاقة بين متغير واحد تابع (Y) ومتغير واحد مستقل (X)، ويأخذ نموذج الانحدار الخطي البسيط الشكل التالي:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i :\ i = \overline{1,n} \ldots [1]}$$

Y_i: قيم المشاهدات في المتغير التابع،

X_i: قيم المشاهدات في المتغير المستقل

β_0: ثابت النموذج وهي نقطة التقاطع مع المحور العمودي،

β_1: الميل الحدي للانحدار،

ε_i: الخطأ العشوائي

<u>الشكل المصفوفي للنموذج</u>: لدينا

$$[1] \Leftrightarrow \begin{cases} Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n \end{cases}$$

ويمكن تحويل هذه الجملة إلى الشكل المصفوفي التالي:

$$\underset{[n\times1]}{Y} = \underset{[n\times2]}{X}\,\underset{[2\times1]}{\beta} + \underset{[n\times1]}{\varepsilon} \Leftrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \ldots [1]$$

<u>2|2 الفرضيات الأساسية للنموذج</u>

(1) التوقع الرياضي للأخطاء معدوم، بمعنى: $\forall i = \overline{1,n} :\ E(\varepsilon_i) = 0$

والتي تكافئ الفرضية التالية: $\hat{Y} = E(Y_i) = \widehat{\beta_0} + \widehat{\beta_1} X_i$

(2) تباين الخطأ العشوائي ثابت، بمعنى: $\forall i = \overline{1,n} :\ Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 = Var(Y_i)$

(3) استقلالية الأخطاء العشوائية، بمعنى: $\forall i \neq j: Cov(\varepsilon_i,\ \varepsilon_j) = Cov(Y_i,\ Y_j) = 0$

(4) المتغير X_i هو متغير غير عشوائي، بمعنى: $\forall i = \overline{1,n}: Cov(X_i,\ \varepsilon_i) = 0$

(5) قيم المتغير العشوائي (ε_i) تتوزع توزيعا طبيعيا حول وسطها الحسابي: $\varepsilon_i \backsim N(0,1)$

(6) أن يكون حجم العينة أكبر بكثير من عدد المعلمات المقدرة، بمعنى: $n \gg 2$

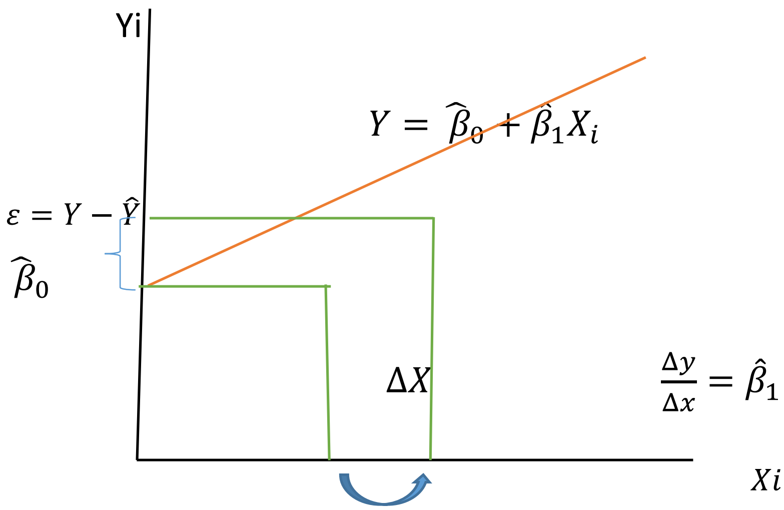
3|2 تقدير معلمات النموذج وتباين الأخطاء

بعد اختيار النموذج المناسب، نقوم بتقدير معلمتي هذا النموذج β_0 و β_1، واختبار معنويتها احصائيا. زمن أجل ذلك نختار الطريقة المناسبة والأكفأ لعملية التقدير وهي طريقة: "المربعات الصغرى العادية OLS " Ordenary Least Squered

1|3|2 تقدير المعلمات بطريقة المربعات الصغرى العادية

عند تمثيل المشاهدات (X.Y) في معلم متعامد تظهر لنا سحابة من النقط، ونحاول في الخطوة الموالية تقدير خط الانحدار الذي يشمل أكبر عدد من النقط ويمثل العلاقة بين X و Y أحسن تمثيل من خلال تهنئة مجموع مربعات الأخطاء العشوائية $\hat{\varepsilon}_i$. والشكل التالي سوف يوضح كيفية عمل طريقة المربعات الصغرى.

شكل: يوضح الهدف من طريقة المربعات الصغرى



المصدر: سلفادور دومينيك، الإحصاء والاقتصاد القياسي، الطبعة الثانية، الجزائر، ديوان المطبوعات الجامعية،1993، ص143 .

ويمكن الحصول على قيم $\boldsymbol{\beta_0}$ و $\boldsymbol{\beta_1}$، باستخدام طريقة المربعات الصغرى على النحو التالي:

نعلم أن: $\forall \boldsymbol{i} = \overline{\mathbf{1}, \boldsymbol{n}} : \ \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} = \boldsymbol{Y_i} - \hat{\mathbf{Y}}_{\mathbf{i}}$

ولتقدير معالم النموذج $\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i}$ نقوم بتصغير مجموع مربعات البواقي:

أي أن: $Min(SSE) = \ Min \sum \left(Y_i - \hat{Y}_i\right)^2$

واعتمادا على الاشتقاق الجزئي للدالة $L\left(\widehat{\beta}_0 , \widehat{\beta}_1\right) = \sum\left(Y_i - \hat{Y}_i\right)^2$ بالنسبة لكل من $\widehat{\beta}_0, \widehat{\beta}_1$ وإعدام هذه المشتقات الجزئية، وهو الشرط اللازم لبلوغ الدالة قيد الدراسة قيمتها الصغرى، يتم الحصول على الصيغ الرياضية للمقدرتين $\widehat{\beta}_0, \widehat{\beta}_1$:

على الشكل التالي:
$$\boxed{\begin{cases} \widehat{\boldsymbol{\beta}_1} = \dfrac{\sum(\boldsymbol{X_i}-\overline{\boldsymbol{X}})(\boldsymbol{Y_i}-\overline{\boldsymbol{Y}})}{\sum(\boldsymbol{X_i}-\overline{\boldsymbol{X}})^2} = \dfrac{\sum \boldsymbol{X_i Y_i} - \boldsymbol{n}\overline{\boldsymbol{X}}\overline{\boldsymbol{Y}}}{\sum \boldsymbol{X_i}^2 - \boldsymbol{n}\overline{\boldsymbol{X}}^2} \\ \widehat{\boldsymbol{\beta}_0} = \overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}_1}\overline{\boldsymbol{X}} \end{cases}}$$

$\overline{X}$: المتوسط الحسابي للمشاهدات (X) ، $\overline{Y}$: المتوسط الحسابي للمشاهدات (Y)

أما الشرط الكافي هو: أن يكون محدد مصفوفة المشتقات الجزئية من الرتبة الثانية بالنسبة للمعلمات المقدرة موجبا.

بمعنى:
$$H=\begin{vmatrix}\frac{\partial^2 L}{\partial\widehat{\beta_0}^2} & \frac{\partial^2 L}{\partial\widehat{\beta_1}\,\partial\widehat{\beta_0}}\\ \frac{\partial^2 L}{\partial\widehat{\beta_0}\,\partial\widehat{\beta_1}} & \frac{\partial^2 L}{\partial\widehat{\beta_1}^2}\end{vmatrix}=\begin{vmatrix}2n & 2\sum X_i\\ 2\sum X_i & 2\sum X_i^2\end{vmatrix}=4n\sum X_i^2-4(\sum X_i)^2$$

أي أن:
$$H=4n\sum(X_i-\overline{X})^2>0$$

2|3|1|أ) الصيغة المصفوفية للمعلمات المقدرة في النموذج المقدر

$$\varepsilon^t\varepsilon=\left(Y-X\hat{\beta}\right)^t\left(Y-X\hat{\beta}\right)$$

$$\varepsilon^t\varepsilon=Y^tY-Y^tX\hat{\beta}-X^t\hat{\beta}^t+X^t\hat{\beta}^tX\hat{\beta}$$

$$\varepsilon^t\varepsilon=Y^tY-2YX^t\beta^t+X^t\hat{\beta}^tX\hat{\beta}$$

وكما يأتي نأخذ المشتق الأول لمجموع مربعات البواقي $\boldsymbol{\varepsilon^t\varepsilon}$ بالنسبة للمعلمة $\widehat{\boldsymbol{\beta}}$ فيكون لدينا:

$$\frac{\partial\boldsymbol{\varepsilon^t\varepsilon}}{\partial\boldsymbol{\beta}}=-\mathbf{2}\boldsymbol{X^tY}+\mathbf{2}\boldsymbol{X^tX}\widehat{\boldsymbol{\beta}}=\mathbf{0}$$

نقسم طرفي المعادلة على 2 فتأخذ المعادلة الشكل الاتي: $\boldsymbol{X^tY}=\boldsymbol{X^tX}\widehat{\boldsymbol{\beta}}$

وبالتالي: $\boxed{\widehat{\boldsymbol{\beta}}=[\boldsymbol{X^tX}]^{-1}\boldsymbol{X^tY}}$

وعليه يمكن ان نكتب الصيغة الرياضية لتقدير المعلمتين بالشكل الاتي:

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}_0\\ \widehat{\boldsymbol{\beta}}_1\end{bmatrix}=\begin{bmatrix}\boldsymbol{n} & \sum\boldsymbol{X_i}\\ \sum\boldsymbol{X_i} & \sum\boldsymbol{X_i^2}\end{bmatrix}^{-1}*\begin{bmatrix}\sum\boldsymbol{Y_i}\\ \sum\boldsymbol{X_iY_i}\end{bmatrix}$$

2|3|1|ب) تطبيق: اذا كانت لدينا البيانات الاتية:

Yi	5	8	6	7	9	10
Xi	2	3	2	4	6	8

علما أن النموذج خطي بسيط بين المتغير المستقل X والمتغير التابع Y، قدر المعلمتين $\boldsymbol{\beta_0}$ و $\boldsymbol{\beta_1}$ للنموذج

$$\boldsymbol{Y_i}=\boldsymbol{\beta_0}+\boldsymbol{\beta_1X}+\boldsymbol{\varepsilon_i}$$

الحل: يمكن تكوين الجدول الآتي:

n	Yi	Xi	YiXi	X2
1	**5**	**2**	**10**	**4**
1	**8**	**3**	**24**	**9**
1	**6**	**2**	**12**	**4**
1	**7**	**4**	**28**	**16**
1	**9**	**6**	**54**	**36**
1	**10**	**8**	**80**	**64**
N=6	$\sum \boldsymbol{Y_i}$ =45	$\sum \boldsymbol{X_i}$=25	$\sum \boldsymbol{X_i}\,\mathbf{Y_i}$=208	$\sum \boldsymbol{X_i^2}$=133

ولتطبيق هذه الصيغة يمكن ان نستخدم البيانات في الطريقة الاولى لكي نتأكد من صحة الطريقة وان تقدير المعالم هو نفسه في هذه الطريقة كما يأتي: $\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \begin{bmatrix}6 & 25\\ 25 & 133\end{bmatrix}^{-1}\begin{bmatrix}45\\ 208\end{bmatrix}$

اول الخطوات في الحل يجب ان نحتسب معكوس المصفوفة وذلك باحتساب محددة المصفوفة ومن ثم مصفوفة المرافقة كما يأتي:

محددة المصفوفة هو: $|X^tX| = \begin{vmatrix}6 & 25\\ 25 & 133\end{vmatrix} = 6(133) - (25)^2 = 798 - 625 = 173$

مصفوفة المرافقة: $Adj[X^tX =] \begin{bmatrix}133 & -25\\ -25 & 6\end{bmatrix}$

نحتسب معكوس المصفوفة كما يأتي: $[X^tX]^{-1} = \frac{Adj[X^tX]}{|X^tX|} = \begin{bmatrix}\frac{133}{173} & \frac{-25}{173}\\ \frac{-25}{173} & \frac{6}{173}\end{bmatrix} = \begin{bmatrix}0.77 & -0.145\\ -0.145 & 0.035\end{bmatrix}$

بعد الحصول على معكوس المصفوفة يمكن تقدير المعلمتين بطريقة المصفوفات والتي تم اشتقاقها كما يأتي:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \begin{bmatrix}0.77 & -0.145\\ -0.145 & 0.035\end{bmatrix}^{-1} * \begin{bmatrix}45\\ 208\end{bmatrix} = \begin{bmatrix}4.543\\ 0.71\end{bmatrix}$$

وايضا توصلنا الى نفس نتائج طريقة المعادلات الاعتيادية

2|3|1|ت) طريقة المحددات قاعدة كرامر (Grammer's Role)

تعتمد هذه الطريقة على تكوين محددات فرعية اضافة للمحدد الرئيس وذلك باستخدام المتجه $\begin{bmatrix}\sum Y_i\\ \sum X_iY_i\end{bmatrix}$ واستبداله بأحد أعمدة المحدد الرئيس لحساب المحدد الفرعي أو الثانوي كما يأتي :

$$D = |X^tX| = \begin{vmatrix}n & \sum X_i\\ \sum X_i & \sum X_i^2\end{vmatrix} = n\left(\sum X_i^2\right) - \left(\sum X_i\right)^2$$

نستبدل العمود الاول من المحدد الرئيسي بالمتجه العمودي كما يأتي:

$$\begin{cases} D_0 = \begin{vmatrix}\sum Y_i & \sum X_i\\ \sum X_iY_i & \sum X_i^2\end{vmatrix} = \left(\sum Y_i\right)\left(\sum X_i^2\right) - \left(\sum X_i\right)\left(\sum X_iY_i\right) \\ D_1 = \begin{vmatrix}n & \sum Y_i\\ \sum X_i & \sum X_iY_i\end{vmatrix} = n\left(\sum X_iY_i\right) - \left(\sum Y_i\right)\left(\sum X_i\right) \end{cases}$$

وعليه يمكن ان نقدر المعلمتين بهذه الطريقة كما يأتي:

$$\begin{cases} \hat{\beta}_0 = \frac{D_0}{D} = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_iY_i)}{n(\sum X_i^2) - (\sum X_i)^2} \\ \hat{\beta}_1 = \frac{D_1}{D} = \frac{n(\sum X_iY_i) - (\sum Y_i)(\sum X_i)}{n(\sum X_i^2) - (\sum X_i)^2} \end{cases}$$

نطبق باستخدام بياناتنا السابقة وكما يأتي:

$$\begin{cases} \hat{\beta}_0 = \frac{D_0}{D} = \frac{(45)(133) - (25)(208)}{6(133) - (25)^2} = \frac{5985 - 5200}{173} = \frac{785}{173} = 4.543 \\ \hat{\beta}_1 = \frac{D_1}{D} = \frac{6(208) - (45)(25)}{6(133) - (25)^2} = \frac{1248 - 1125}{173} = \frac{123}{173} = 0.71 \end{cases}$$

2|3|2 تقدير تباين الخطأ العشوائي

المعلمتان β_0, β_1 غير عشوائيين، بخلاف المقدرتان $\widehat{\beta_0}, \widehat{\beta_1}$. وهو ما يتطلب دراسة خصائصهما باعتبارهما عشوائيتين.

سؤال: "**هل تعطي طريقة المربعات الصغرى العادية أفضل تقديرات خطية وغير متحيزة؟** "

الجواب: عندما يتضمن تقدير النموذج الخطي الفرضيات الستة سالفة الذكر، يكون تقدير المعلمات أفضل تقدير خطي غير متحيز $(BLUE)$ " Best Linear Uniased Estimators " ويعطينا:

$$\boxed{\boldsymbol{Var(\varepsilon_i) = \hat{\sigma}^2_{\hat{\varepsilon}} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2} = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-2}}}$$

وهو أصغر تباين للأخطاء، حيث: $n-2$ هو حجم العينة – عدد المعلمات المقدرة وهذا ما تحققه طريقة المربعات الصغرى العادية.

2|3|3 الخصائص الإحصائية للمقدرات

2|3|3أ) خاصية عدم التحيز: وتعني أن: $\begin{cases} E(\widehat{\beta_0}) = \beta_0 \\ E(\widehat{\beta_1}) = \beta_1 \end{cases}$ ومقدار التحيز هو: $\begin{cases} \widehat{\beta_0} - \beta_0 \\ \widehat{\beta_1} - \beta_1 \end{cases}$.

2|3|3ب) خاصية الكفاءة: وتعني ذات أصغر تباين.

$$\boldsymbol{Var(\widehat{\beta_1}) = \hat{\sigma}^2_{\widehat{\beta_1}} = \left(\frac{1}{\sum (X_i - \bar{X})^2}\right) \hat{\sigma}^2_{\hat{\varepsilon}}}$$

$$\boldsymbol{Var(\widehat{\beta_0}) = \hat{\sigma}^2_{\widehat{\beta_0}} = \left(\frac{1}{n} \times \frac{\sum X_i^2}{\sum (X_i - \bar{X})^2}\right) \hat{\sigma}^2_{\hat{\varepsilon}} = \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}\right) \hat{\sigma}^2_{\hat{\varepsilon}}}$$

تباينا المقدرتين غير معروفين لأنهما يرتبطان بتباين الخطأ النظري الذي هو الآخر مجهول

2|3|3ت) خاصية الاتساق: وتعني تحقيق الشرطين التاليين:

عندما يكون حجم العينة كبيرا $(+\infty)$ فإن: $\begin{cases} E(\widehat{\beta_0}) \backsim \beta_0 \quad ، \quad E(\widehat{\beta_1}) \backsim \beta_1 \\ V(\widehat{\beta_0}) \backsim 0 \quad ، \quad V(\widehat{\beta_1}) \backsim 0 \end{cases}$

2|4 خطوات توفيق نموذج الانحدار الخطي البسيط:

للحكم على صلاحية نموذج الانحدار الذي تم توفيقه للعلاقة بين المتغير التابع والمتغير المستقل، لابد وأن يتوفر في هذا النموذج مجموعة من الشروط، يمكن تقسيمها إلى:

2|4|1 الشروط النظرية (الاقتصادية)

2|4|1أ) اتفاق (أو منطقية) إشارات وقيم معاملات الانحدار:

يجب أن تكون إشارات وقيم المعالم منطقية مع الأساس النظري الذي يحكم الظاهرة محل الدراسة.

مثلا: نموذج انحدار العلاقة بين الدخل والاستهلاك: $C = a + bY$.

الشرطان المفروضان على معلمي النموذج، وفقا للنظرية الاقتصادية، هما:

- أن يكون $0 < b < 1$ حيث يمثل b الميل الحدي للاستهلاك.
- أن يكون $a > 0$ ويمثل a الجزء الثابت من الاستهلاك (حالة الدخل معدوم: $Y = 0$).

فعدم توافر هذين الشرطين يجعل نموذج الانحدار الذي تم توفيقه غير سليم من الناحية النظرية.

2|4|1|ب) القدرة التفسيرية للنموذج وقوة الارتباط وتحليل التباين

أولا–القدرة التفسيرية للنموذج: هي مدى قدرة المتغير المستقل في النموذج على تفسير التغيرات التي تحدث في المتغير التابع.

بمعنى آخر هي نسبة التغيرات التي تحدث في المتغير التابع، وتعزى إلى المتغير المستقل.

ويعرف أيضا أنه: " نسبة تأثير المتغير المستقل في المتغير التابع.

ويمكن حسابها من خلال معامل التحديد R^2 الذي يمكن الحصول عليه على النحو التالي:

لدينا: $Y_i = \hat{Y}_i + \hat{\varepsilon}_i \Leftrightarrow Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{\varepsilon}_i$

وبالتالي: $\sum(Y_i - \bar{Y})^2 = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum(\hat{\varepsilon}_i)^2$

بمعنى أن: $\boxed{\boldsymbol{TSS = ESS + RSS \dots (1)}}$

وهذه العلاقة مفيدة جدا لخدمة ما يتعلق بقياس القدرة التفسيرية، ولذا من المهم أن نفحص بعناية معنى كل حد من حدودها.

- $TSS = \sum(Y_i - \bar{Y})^2$: مجموع مربعات الانحرافات الكلية في المتغير Y_i؛ Total Sum of Squares
- $ESS = \sum\left(\hat{Y}_i - \bar{Y}\right)^2$: مجموع مربعات الانحدار المقدرة (المشروحة)؛ Explained Sum of Squares
- $RSS = \sum(\hat{\varepsilon}_i)^2 = \sum\left(Y_i - \hat{Y}_i\right)^2$: مجموع مربعات البواقي.

وبقسمة أطراف معادلة تحليل التباين (1) على الانحرافات الكلية TSS نجد: $(1) \Leftrightarrow \frac{ESS}{TSS} + \frac{RSS}{TSS} = 1$

وعليه يعرف معامل التحديد R^2 كما يلي: $R^2 = \frac{ESS}{\text{TSS}} = 1 - \frac{RSS}{TSS}$

تعريف: معامل التحديد R^2 هو نسبة مجموع المربعات المقدرة على مجموع المربعات الكلية.

ويقيس هذا المؤشر جودة التوفيق والارتباط بين المتغير المستقل X_i والمتغير التابع Y_i .

وهو أيضا، يعبر عن القدرة التفسيرية للنموذج، ومن الواضح أن: $0 \leq R^2 \leq 1$

ملاحظات:

- إذا كان $R^2 = 1$، فهناك جودة في التوفيق والارتباط بين المتغيرين X_i و Y_i، وأن القدرة التفسيرية للنموذج عالية؛
- إذا كان $R^2 = 0$، فإما أنه لا توجد قدرة تفسيرية في للنموذج، أو توجد قدرة تفسيرية، لكن النموذج غير خطي.

2|4|1|ت) قوة الارتباط بين متغيري النموذج:

يستخدم تحليل الارتباط في تقدير درجة الارتباط الخطي بين متغيرين، وتحديد اتجاه هذه العلاقة.

وتقاس قوة الارتباط بين المتغيرين اعتمادا على معامل الارتباط (r) المعرف كما يلي: $r = \mp\sqrt{R^2}$

- تنتمي قيمة معامل الارتباط (r)إلى المجال $[-1, +1]$؛
- فإذا كان: $r > 0$ فإن العلاقة بين المتغيرين طردية؛
- وإذا كان: $r < 0$ فإن العلاقة بين المتغيرين عكسية

ملاحظات:

1. إذا كان $r = 0$، لا توجد علاقة خطية بين المتغيرين.
2. إذا كان $r = \mp 1$، هناك ارتباط خطي تام بين المتغيرين.

3. هناك فرق جوهري بين معامل الارتباط (r) ومعامل التحديد (R^2) حيث:

يقيس (r)العلاقة بين المتغيرين X_i و Y_i، دون تحديد السببية.

في حين يقيس (R^2)العلاقة بين المتغيرين X_i و Y_i ، مع الاهتمام بالسببية (أحد المتغيرين يفسر الآخر).

2|4|2 الشروط الرياضية:

2|4|2|أ) اختبار جودة توفيق النموذج

الغرض من هذا الاختبار هو التوصل إلى قرار حول صلاحية النموذج في تمثيل العلاقة بين المتغير التابع والمتغير المستقل تمثيلا جيدا. ويحتاج الاقتصادي إلى اختبار الفرضيات لاتخاذ القرارات المناسبة التي تتعلق بالظاهرة المدروسة، ويعتمد اختبار الفرضيات على التوزيعات الاحتمالية: توزيعات فيشر، ستودنت، الطبيعي...الخ.

أولا–اختبار المعنوية الكلية للنموذج:

وهي اختبار النموذج بين المتغيرين التابع والمستقل في نموذج الانحدار باستخدام اختبار فيشر(**F–test**). ويقصد باختبار المعنوية الكلية، الإجابة عن السؤال التالي: " هل النموذج الخطي مقبول لتمثيل العلاقة بين المتغيرين أم لا؟ "

ولاختبار مدى قبول هذا النموذج إحصائيا نعتمد على إحصاءة فيشر المعرفة على النحو التالي:

- **الإحصاءة المحسوبة لفيشر :**

بمعنى أن:
$$F_C = \frac{MES}{MRS} = \frac{ESS/1}{RSS/(n-2)} = \frac{ESS(n-2)}{RSS} = \frac{(ESS/SST)(n-2)}{(RSS/SST)} = \left(\frac{R^2}{1-R^2}\right)(n-2)$$

أي أن:
$$\boxed{\boldsymbol{F_C = \frac{MES}{MRS} = \left(\frac{R^2}{1-R^2}\right)(n-2)}}$$

وهو ما يوضحه جدول تحليل التباين التالي:

Analysis of Variance table "ANOVA"

Model	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig*
Regression	*ESS*	*1*	*MES = ESS/1*	*MES/MRS*=$\left(\frac{R^2}{1-R^2}\right)(n-2)$	؟
Residual	*RSS*	*n–2*	*MRS = ESS/(n–2)*		
Total	*TSS*	*n–1*			

- **صياغة الفرضيات الإحصائية**

الفرضية الصفرية: $H_0 : \hat{\beta}_0 = \widehat{\beta_1} = 0$ (النموذج غير مناسب أو غير مقبول احصائيا)

الفرضية البديلة: $H_1 : \widehat{\beta_1} \neq 0$ (النموذج مناسب أو مقبول احصائيا)

ومن ثم: $F_c \backsim F(1, n-2, \alpha)$

- **<u>اتخاذ القرار</u>:**

بفرض p هي القيمة الاحتمالية لاختبار فيشر، و α هو مستوى الدلالة المطلوب،

- **إذا اعتمادنا الاحتمال p: نرفض الفرضية الصفرية H_0 إذا كان $p \leq \alpha$، ونقبل الفرضية البديلة H_1**

- **إذا اعتمادنا الإحصاءة F : نرفض الفرضية الصفرية H_0 إذا كان $F_c \leq F_\alpha$، ونقبل الفرضية البديلة H_1**

ثانيا–المعنوية الجزئية لمعالم النموذج:

وهي اختبار معنوية معامل الانحدار للمتغير المستقل وثابت الانحدار، من خلال اختبار ستودنت **(t–test)**.

الفرضية الصفرية: $H_0 : \widehat{\beta_0} = \widehat{\beta_1} = 0$ (المتغير المستقل X ليس له أثر معنوي على المتغير التابع Y)

الفرضية البديلة: $H_1 : \widehat{\beta_1} \neq 0$ (المتغير المستقل X له أثر معنوي على المتغير التابع Y)

إذا افترضنا أن $\boldsymbol{n} < 30$ ، $\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}$ غير معروف فإن: $\frac{\widehat{\beta_1}-\beta_1}{\hat{\sigma}_{\widehat{\beta_1}}} \backsim t(n-2, \alpha)$

- **مجال الثقة لمعلمات الانحدار:** عند اختبار الفرضيات المتعلقة بالمعلمات، فقد نرفض الفرضية H_0 أو نقبلها.

ففي حالة قبول H_0 معناه أن المعلمتين β_0 و β_1 معدومتان معنويا.

وفي ظل رفض الفرضية H_0 وقبول H_1: $pr\left\{-t_{\alpha/2} \leq \frac{\widehat{\beta_1}-\beta_1}{\hat{\sigma}_{\widehat{\beta_1}}} \leq t_{\alpha/2}\right\} = 1-\alpha$

عندئذ: $pr\{\widehat{\beta_1} - t_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}} \leq \beta_1 \leq \widehat{\beta_1} + t_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}}\} = 1-\alpha$

وبالتالي فإن مجال الثقة للمعلمة β_1 هو: $[\widehat{\beta_1} - t_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}} , \widehat{\beta_1} + t_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}}]$

لكن $\beta_1 = 0$، إذن: $\boxed{t_c = \frac{\widehat{\beta_1}}{\hat{\sigma}_{\widehat{\beta_1}}} \backsim t_{\alpha/2}}$

حيث: $t_{\alpha/2}$ هي القيمة الحرجة لتوزيع ستودنت بدرجة حرية $(n-1)$ ومستوى دلالة α.

$\hat{\sigma}_{\widehat{\beta_1}}$ هو الانحراف المعياري للمعلمة المقدرة $\widehat{\beta_1}$.

- **<u>اتخاذ القرار</u>:**

بفرض p هي القيمة الاحتمالية لاختبار ستودنت، و α هو مستوى الدلالة المطلوب،

- إذا اعتمادنا الاحتمال p: نرفض الفرضية الصفرية H_0 إذا كان $p \leq \alpha$، ونقبل الفرضية البديلة H_1
- إذا اعتمادنا الإحصاءة t : نرفض الفرضية الصفرية H_0 إذا كان $|t_c| \geq t_{\alpha/2}$، ونقبل الفرضية البديلة H_1

ملاحظة:

إذا كان $\left(\boldsymbol{n} \geq \boldsymbol{30} \text{ ، } \widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}\right.$ معروف$\left.\right)$ فإن: $\frac{\widehat{\beta_1}-\beta_1}{\hat{\sigma}_{\widehat{\beta_1}}} \backsim N(0,1)$

عندئذ نستعمل التوزيع الطبيعي: $pr\{\widehat{\beta_1} - z_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}} \leq \beta_1 \leq \widehat{\beta_1} + z_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}}\} = 1-\alpha$

وبالتالي فإن مجال الثقة للمعلمة β هو: $[\widehat{\beta_1} - z_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}} , \widehat{\beta_1} + z_{\alpha/2} \cdot \hat{\sigma}_{\widehat{\beta_1}}]$

لكن $\beta_1 = 0$، إذن: $z_c = \frac{\widehat{\beta_1}}{\hat{\sigma}_{\widehat{\beta_1}}} = \frac{\mathrm{B}}{\mathrm{Std}\text{–}Error} \backsim z_{\alpha/2}$

5|2 مرحلة التنبؤ واختبار الدقة التنبؤية

1|5|2 التنبؤ:

عندما يتم تقدير معلمتي النموذج فمن الممكن جدا حساب التنبؤ للقيمة المستقبلية للمتغير التابع في اللحظة $t+1$ واختبار القدرة التنبؤية للنموذج **على النحو التالي:**

لنفترض أن النموذج المقدر هو: $\widehat{Y_i} = \widehat{\beta_0} + \widehat{\beta_1} X_i$

وعندما تكون قيم X المستقبلية معروفة يعطى التنبؤ بالعلاقة التالية: $\widehat{Y_{i+1}} = \widehat{\beta_0} + \widehat{\beta_1} X_{i+1}$

2|5|2 اختبار الدقة التنبؤية للنموذج:

قد يكون النموذج مقبولا بالنظر إلى بعض الإحصاءات؛ معامل التحديد مرتفع، معلمات النموذج لها معنوية إحصائية كبيرة، لكن القدرة التنبؤية للنموذج محدودة، نتيجة احتمال وقوع تغيرات فجائية مثلا، لذا لابد من اختبار مدى قدرة النموذج على التنبؤ قبل استخدامه.

لهذا الغرض توجد عدة معايير لقياس قدرة النموذج على التنبؤ، نقدم أهمها:

2|5|2أ) معامل عدم التساوي لثايل ($\boldsymbol{Theile}$): ويحسب بالعلاقة: $UT = \sqrt{\frac{\sum_{i=1}^{n}(AT_i - FT_i)^2}{\sum_{i=1}^{n} {AT_i}^2}}$ حيث:

UT: هو معامل عدم التساوي، AT_i: نسبة التغير في القيم الحقيقية وتحسب ب: $AT_i = \left(\frac{Y_i - Y_{i-1}}{Y_i}\right) \times 100$

FT_i: نسبة التغير في القيم المقدرة وتحسب ب: $FT_i = \left(\frac{\hat{Y}_{i+1} - \hat{Y}_i}{\hat{Y}_{i+1}}\right)$، حيث: $0 < FT_i < 1$

وكل ما كانت قيمة معامل ثايل قريبة من الصفر، كانت القدرة التنبؤية للنموذج أفضل.

2|5|2ب) فترة الثقة للتنبؤ: المقصود هنا هو تقدير فترة الثقة للقيمة التنبؤية المتحصل عليها للمتغير التابع $\boldsymbol{Y_i}$.

من أجل ذلك نحسب أولا الخطأ المعياري للتنبؤ: $\boldsymbol{\hat{\sigma}_{Y_f} = \sigma_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})}{\sum \ (X_i - \bar{X})^2}^{2}}}$ **حيث:** $\boldsymbol{\sigma_\varepsilon = \frac{\sum \hat{\varepsilon}_i^2}{n-k}}$

وعليه يمكن حساب حدود مجال الثقة للتنبؤ كما يلي: $\boldsymbol{Y_{i+1} \in \left[\hat{Y}_{i+1} \pm \hat{\sigma}_{Y_f} \cdot t_{\frac{\alpha}{2};n-k}\right]}$