اختبار جودة النموذج

1.3 المقدمــــة:

تناولــنا فــي الفصــل السابق، دراسة العلاقة بين متغيرين (الانحدار الخطي البســيط)، وقد بينا صيغ اشتقاق معلمات النموذج بطريقة OLS، وطرق تقدير هذ المعلمــات. ولكن المنطق يحتم علينا عدم قبول هذا النموذج إلا بعد أجراء تقييم لهذ النتائج من الناحية الاقتصادية والإحصائية. ويتم هذا باستخدام نتائج التقدير في اختبار النظريات أو اتخاذ القرارات أو وضع السياسات.

ســنتناول فــي هذا الفصل، تقييم نموذج الانحدار المقدر، ولتحقيق ذلك سيتم اختبار المعنوية الاقتصادية والإحصائية لنتائج تقدير نموذج الانحدار.

اختبار الفرضيات، Test of hypothesis:

تعــرف الفرضية بأنه ادعاء قابل لأن يكون صحيحاً أو غير صحيح، وتثبت صــحتها فقــط من خلال الاختبار (Test). وقبل البدء بدراسة الكيفية التي يتم على أساســها اختبار الفرضية، لابد من دراسة العلاقة الاقتصادية التي تستند إلى مجموعه من الفروض الخاصة بالنظرية الاقتصادية، كالعلاقة بين الكميه المطلوبة من سلعه ما وسعر تلك السلعة. وحسب منطق النظرية الاقتصادية تعكس دالة الطلب علاقة عكسية (سالبه) بين المتغير المستقل (السعر) والكميه المطلوبة (المتغير المعتمد).

أما إذا أخذنا العلاقة بين الأنفاق الاستهلاكي والدخل، فأن منطق النظرية يشير إلى ان الميل الحدي للاستهلاك (MPC) يكون موجباً ولكنه اقل من الواحد الصحيح، $0<MPC<1$، وان هناك حد أدنى للاستهلاك في الأمد القصير، ولكن لا يمكن الجزم بصــحة أو عدم صحة ذلك إلا بعد أخذ البيانات وقياس العلاقة الاقتصادية واختبارها. وهذا ما سنتناوله في المباحث القادمة.

ويختبر نموذج الانحدار قبل كل شيء العلاقة بين المتغير المستقل (X) والتابع (Y) وذلك للتثبت من وجودها من خلال اختبار المعنوية الإحصائية للمعلمات المقدرة $\hat{B}_0$ و $\hat{B}_1$ كلاً على انفراد وفي هذا المجال توجد فرضيتان:

1- فرضية العدم: وتنص على عدم وجود علاقة بين المتغيرين X و Y أي ان:

$$H_0: B_0 = 0$$
$$B_1 = 0$$

2- الفرضية البديلة: وتنص على وجود علاقة بين X و Y ، أي ان:

$$H_1: B_0 \neq 0$$
$$B_1 \neq 0$$

2.3 اختبار قيمة t، t-Value Test:

ولأجل اختبار ما إذا كانت $B_0 = 0$ ، $B_1 = 0$ أم لا، يستخدم اختبار (t) عند مستوى معنوية معينة ودرجة حرية (n-k) والصيغة الرياضية لهذا الاختبار هي:

- بالنسبة لـ $\hat{B}_1$:

$$t_{\hat{B}_1} = \frac{\hat{B}_1}{S_{\hat{B}_1}} \quad \ldots(1.3)$$

حيث ان:

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

$$S^2_{\hat{B}_1} = \frac{S^2_{e_i}}{\sum x_i^2}$$

$$S^2_{e_i} = \frac{\sum e_i^2}{n-2}$$

حيث ان:

t: هــو اختــيار (t) عند مستوى معنوية معين ودرجة حرية (n-k) حيث (
عدد المشاهدات في العينة و(k) عدد المعالم.

$\hat{B}_1$: القيمة التقديرية لـــ B_1 الحقيقية.

$S\hat{B}_1$: الانحراف المعياري للمعلمة المقدرة $\hat{B}_1$.

$S^2_{\hat{B}_1}$: تباين $\hat{B}_1$.

$S^2_{e_i}$: تباين الخطأ.

ب- بالنسبة لـــ $\hat{B}_0$ فأن:

$$t_{\hat{B}_0} = \frac{\hat{B}_0}{S\hat{B}_0} \quad \ldots(2.3)$$

حيث أن:

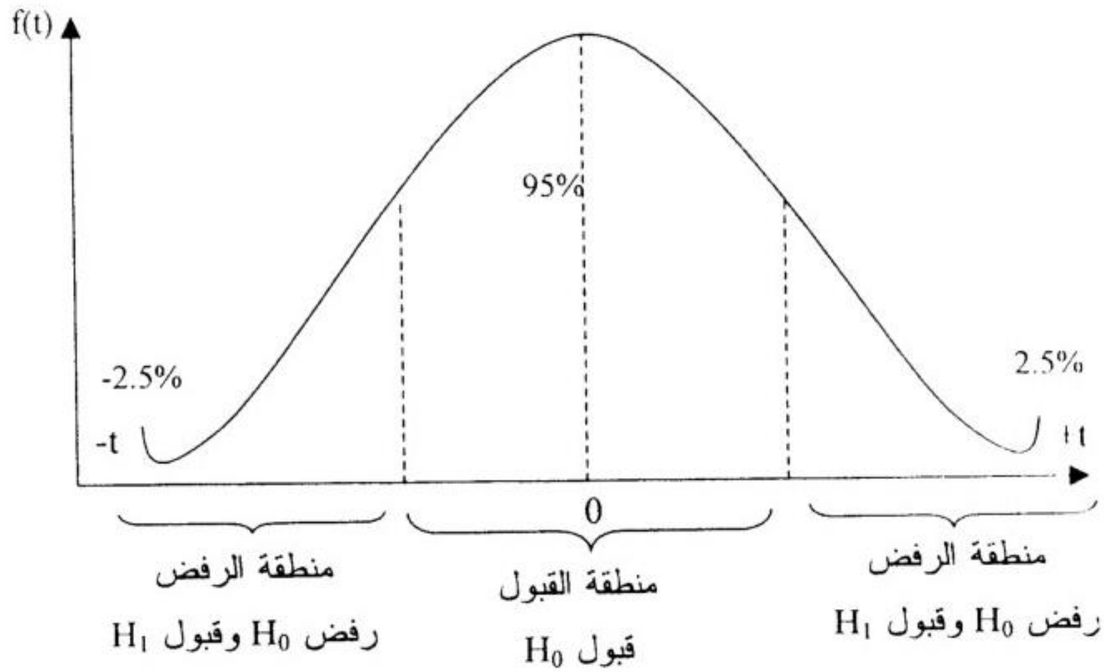
$$S\hat{B}_0 = \sqrt{S^2_{\hat{B}_0}}$$

$$S^2_{\hat{B}_0} = S^2_{e_i}\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right]$$

$$S^2_{e_i} = \frac{\sum e_i^2}{n-k}$$

وبعد احتساب قيمة (t) تقارن مع قيمتها الجدولية المعطاة في الجداول الخاصة بها عند درجات حرية (n-2) ومستوى المعنوية المطلوب (5%، 1%) لتحديد قبول أو رفض فرضـــية العدم. فإذا كانت قيمة t المحتسبة اكبر من قيمة t الجدولية ترفض فرضية العـــدم وتقبل الفرضية البديلة، بمعنى ان المعلمة ذات معنوية إحصائية. وبالعكس في حالـــة كـــون t المحتســـبة اقل من قيمتها الجدولية حيث تقبل فرضية العدم وترفض الفرضـــية الـــبديلة أي عـــدم معنوية المعلمة المقدرة. ويمكن توضيح ذلك بالشكل .(1.3)

شكل 1.3: قبول أو رفض الفرضية



ولاختـــبار معـــنوية المعلمات المقدرة $\hat{B}_0$ و $\hat{B}_1$ نعود الى بيانات المثال الوارد (1.2)، وباعتماد الصيغة الرياضية للاختبار، فأننا نحتاج الإحصاءات الآتية:

$\hat{Y}_i$	e_i	e_i^2
30.43333334	−4.83333334	23.36111118
36.38095239	−3.68095239	13.5494105
42.32857143	3.07142857	9.433673461
48.27619048	5.62380952	31.62723352
54.22380953	4.77619047	22.81199541
60.17142858	2.42857142	5.897959142
66.11904763	−1.11904763	1.252267598
72.06666667	−6.26666667	39.27111115
$\sum \hat{Y}_i = 410$	$\sum e_i = 0$	$\sum e_i^2 = 147.204762$

أ- بالنسبة الى $\hat{B}_1$:

$$S_{e_i}^2 = \frac{\sum e_i^2}{n-k} = \frac{147.204762}{8-2} = 24.534127$$

$$S_{\hat{B}_1}^2 = \frac{S_{e_i}^2}{\sum x_i^2} = \frac{24.534127}{672} = 0.036509117$$

$$S_{\hat{B}_1} = \sqrt{S_{\hat{B}_1}^2} = \sqrt{0.036509117} = 0.191073592$$

$$t_{\hat{B}_1} = \frac{\hat{B}_1}{S_{\hat{B}_1}} = \frac{1.486904762}{0.191073592} = 7.781843354$$

وبما ان قيمة (t) المحتسبة والبالغة (7.781) اكبر من قيمة (t) الجدولية عند مستوى معـــنوية (5%) ودرجة حرية (6) والبالغة (2.45)، عليه نرفض فرضية العدم ونقبل الفرضية البديلة أي معنوية المعلمة المقدرة $\hat{B}_1$.

ب- بالنسبة إلى $\hat{B}_0$:

$$S_{e_i}^2 = \frac{\sum e_i^2}{n-k} = \frac{147.204762}{8-2} = 24.534127$$

$$S_{\hat{B}_1}^2 = S_{e_i}^2 \left[\frac{1}{n} - \frac{\overline{X}^2}{\sum x_i^2} \right]$$

$$= 24.534127 \left[\frac{1}{8} - \frac{(18)^2}{672} \right]$$

$$= 24.534127 \left[0.125 - 0.482142857 \right]$$

$$= 14.89571996$$

$$S_{\hat{B}_0} = \sqrt{S_{\hat{B}_0}^2} = \sqrt{14.89571996} = 3.859497372$$

$$t\hat{B}_0 = \frac{\hat{B}_0}{S\hat{B}_0} = \frac{24.48571429}{3.859497372} = 6.344275415$$

وبمــا ان قــيمة (t) المحتســبة عــند مستوى معنوية (5%) ودرجة حرية (6)
(6.344) وهي اكبر من مثيلتها الجدولية البالغة (2.45)، عليه نرفض فرضية ال
وتقبل الفرضية البديلة، أي معنوية المعلمة المقدرة $\hat{B}_0$ عالية.

3.3 معامل التحديــد، (R^2) :Coefficient of Determination

هــو مقياس يوضح نسبة التغير في المتغير التابع (Y) الذي سببها التغيير ف
المتغــير المســتقل (X). أي نســبة الانحرافات الموضحة من قبل خط الانحدار ال
الانحرافات الكلية. ويمكن حساب هذا المعامل حسب الصيغ الآتية:

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \qquad \ldots(3.3)$$

أو

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2}$$

أو

$$R^2 = \frac{\hat{B}_1 \sum x_i y_i}{\sum y_i^2}$$

أو

$$R^2 = r^2$$

أو

$$R^2 = \frac{\hat{B}_1^2 \sum x_i^2}{\sum y_i^2}$$

أو

$$R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

ويمكن توضيح الصيغة الأخيرة من خلال الاشتقاق الآتي:

الانحرافات غير الموضحة + الانحرافات الموضحة من قبل خط الانحدار = الانحرافات الكلية

$$\sum(Y_i - \overline{Y})^2 = \sum(\hat{Y}_i - \overline{Y})^2 + \sum(Y_i - \hat{Y}_i)^2 \qquad \ldots(4.3)$$

وبقسمة طرفي المعادلة (4.3) على الانحرافات الكلية:

$$\frac{\sum(Y_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} + \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \overline{Y})^2}$$

$$1 = R^2 + \frac{\sum e_i^2}{\sum y_i^2}$$

$$\therefore R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

وتتراوح قيمة R^2 بين الصفر والواحد، أي ان: $0 \le R^2 \le 1$

حيـــث أن $R^2 = 1$ عـــندما تقـــع جمـــيع نقاط الانتشار على خط الانحدار المقدر أي $Y_i = \hat{Y}_i$ وهنا تكون العلاقة تامة.

وان $R^2 = 0$ (أو تقترب منه) عندما يكون خط انحدار العينة خطاً أفقياً أي $\hat{Y}_i = \overline{Y}$ ، ومعنى ذلك لا توجد علاقة بين المتغير التابع والمتغير المستقل.

وكـل مـن هاتيـن الحالتين نادرة الحدوث، ففي الأحوال العادية يفسر خط الانحـدار جـزءاً مـن التغيرات في Y وبالتالي يكون هناك جزءاً آخر غير مفسر بواسطة الخط ومن ثم نجد في أغلب الحالات $0 < R^2 < 1$.

ولمعـرفة مقدار ما يفسره المتغير المستقل، X، من التغير في المتغير التابع، Y نعـود إلى مثالنا الوارد في (1.2)، وبالاعتماد على الصيغ الرياضية الخاصة بـ R^2 فان:

$\hat{y}_i$	$\hat{y}_i^2$	y_i^2
−20.81666666	433.3336108	657.9225
−14.86904761	221.0885768	344.1025
−8.92142857	79.59188773	34.2225
−2.97380952	8.843543061	7.0225
2.97380953	8.843543121	60.0625
8.92142858	79.59188791	128.8225
14.86904763	221.0885774	189.0625
20.81666667	433.3336113	211.7025
$\sum \hat{y}_i = 0$	$\sum \hat{y}_i^2 = 1485.715238$	$\sum y_i^2 = 1632.92$

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{1485.715238}{1632.92} = 0.909851822 = \%90.98$$