الانحدار الخطي المتعدد

أ. تعريف: يستخدم تحليل الانحدار المتعدد لاختبار الفروض حول العلاقة بين متغير تابع (Y) وإثنين أو أكثر من المتغيرات المستقلة $(X_1, X_2,\ldots)$، ويمكن كتابة نموذج الانحدار الثلاثي كما يلي:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \varepsilon_i \ldots\ldots\ldots 1$$

الفرض الإضافي (إلى فروض النموذج الخطي البسيط) أنه لا توجد علاقة خطية تامة بين المتغيرات المستقلة $(X_1, X_2,\ldots)$، ويمكن الحصول على تقديرات معالم المربعات الصغرى العادية بإيجاد النهاية الصغرى لمجموع مربعات البواقي.

$$\sum \varepsilon^2_i = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i})^2$$

ب. كتابة معادلة الانحدار: تعتبر طريقة المربعات الصغرى (OLS) هي أسلوب لتوفيق أفضل خط مستقيم لعينة مشاهدات أو قيم (XY) وهو يتضمن تصغير مجموع المربعات الانحرافات النقاط عن الخط إلى أدنى حد ممكن:

$$Min \sum_{i=1}^{n} (Y_i - \hat{y}_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

حيث تشير (Y_i) إلى المشاهدات الفعلية وتشير $(\hat{y}_i)$ إلى القيم المناظرة، حيث $e_i = Y_i - \hat{y}_i$

بالعودة للمعادلة رقم 1 نقوم بإيجاد كل من (b0) (b1) و(b2) كما يلي:

حيث: $\hat{b}_1$ و $\hat{b}_2$ **هي مقدرتان للمعلمتين الحقيقيتين** b_1 و b_2

ج. إختبار معنوية تقديرات المعالم b1 وb2

من أجل اختبار المعنوية الإحصائية لتقديرات المعالم **b1 وb2** يلزمنا حساب تباين كل منهما.

المعدلات التالية تعطي تقديرات غير متحيزة لتباين **b1 وb2**

حيث تباين المعلمة **b1** تحسب كما يلي:

$$\sigma^2_{b_1} = \frac{\sum_{i}^{n} e_i^2}{n-k} \times \frac{\sum_{i=1}^{n} (X_2 - \bar{X}_2)^2}{\sum_{i=1}^{n} (X_1 - \bar{X}_1)^2 \times \sum_{i=1}^{n} (X_2 - \bar{X}_2)^2 - [\sum_{i=1}^{n} (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]^2}$$

في حين تباين المعلمة **b2** تحسب كما يلي:

$$\sigma^2_{b_2} = \frac{\sum_{i}^{n} e_i^2}{n-k} \times \frac{\sum_{i=1}^{n} (X_1 - \bar{X}_1)^2}{\sum_{i=1}^{n} (X_1 - \bar{X}_1)^2 \times \sum_{i=1}^{n} (X_2 - \bar{X}_2)^2 - [\sum_{i=1}^{n} (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]^2}$$

على هذا الأساس تكون Var(b1)أو σ_b^2 و σ_a^2 أو Var(b2) هي الأخطاء المعيارية للتقدير، وكما هو معلوم فإن e_i موزع توزيع طبيعي وبالتالي b1 وb2 تكون هي الأخرى موزعة طبيعيا، وهذا ما يمكنا من استخدام توزيع t بدرجة حرية (n-k).

مثال: ليكن لديك المتغيرين X2 1X,Y

المطلوب:

- ✓ -أرسم شكل الانتشار.
- ✓ أوجد المعادلة الخطية بين X1,X2 وY

Y	40	44	46	48	52	58	60	68	74	80
X1	6	10	12	14	16	18	22	24	26	32
X2	4	4	5	7	9	12	14	20	21	24

i	Y	X1	X2	$Y-\bar{Y}$	$X_1-\bar{X}_1$	$X_2-\bar{X}_2$	$(Y-\bar{Y})(X_1-\bar{X}_1)$	$(Y-\bar{Y})(X_2-\bar{X}_2)$	$(X_1-\bar{X}_1)(X_2-\bar{X}_2)$	$(X_1-\bar{X}_1)^2$	$(X_2-\bar{X}_2)^2$
1	40	6	4	-17	-12	-8	204	136	96	144	64
2	44	10	4	-13	-8	-8	104	104	64	64	64
3	46	12	5	-11	-6	-7	66	77	42	36	49
4	48	14	7	-9	-4	-5	36	45	20	16	25
5	52	16	9	-5	-2	-3	10	15	6	4	9
6	58	18	12	1	0	0	0	0	0	0	0
7	60	22	14	3	4	2	12	6	8	16	4
8	68	24	20	11	6	8	66	88	48	36	64
9	74	26	21	17	8	9	136	153	72	64	81
10	80	32	24	23	14	12	322	276	168	196	144
$\sum$				0	0	0	**956**	**900**	**524**	**576**	**504**

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(Y-\bar{Y})(X_1-\bar{X}_1)\times\sum_{i=1}^{n}(X_2-\bar{X}_2)^2-\sum_{i=1}^{n}(Y-\bar{Y})(X_2-\bar{X}_2)\times\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{\sum_{i=1}^{n}(X_1-\bar{X}_1)^2\times\sum_{i=1}^{n}(X_2-\bar{X}_2)^2-[\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)]^2}$$

$$\hat{b}_1 = \frac{(956\times504)-(900\times524)}{(576\times504)-(524)^2} = 0.65$$

$$\hat{b}_2 = \frac{\sum_{i=1}^{n}(Y-\bar{Y})(X_2-\bar{X}_2)\times\sum_{i=1}^{n}(X_1-\bar{X}_1)^2-\sum_{i=1}^{n}(Y-\bar{Y})(X_1-\bar{X}_1)\times\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{\sum_{i=1}^{n}(X_1-\bar{X}_1)^2\times\sum_{i=1}^{n}(X_2-\bar{X}_2)^2-[\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)]^2}$$

$$\hat{b}_2 = \frac{(900\times576)-(956\times524)}{(576\times504)-(524)^2} = 1.109 \approx 1.11$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}_1 - \hat{b}_2\bar{X}_2 = 57-((0.65\times18)-(1.11\times12)) = 31.98$$

ومنه معادلة الإنحدار هي: $Y_i = 31.98 + 0.65X_{1i} + 1.11X_{2i}$

ملحق :

Dependent Variable: Y
Method: Least Squares
Date: 02/12/26 Time: 22:33
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.650051	0.250161	2.598527	0.0355
X2	1.109868	0.267434	4.150068	0.0043
C	31.98067	1.631796	19.59845	0.0000

R-squared	0.991634	Mean dependent var	57.00000
Adjusted R-squared	0.989243	S.D. dependent var	13.47426
S.E. of regression	1.397467	Akaike info criterion	3.750525
Sum squared resid	13.67040	Schwarz criterion	3.841300
Log likelihood	-15.75262	Hannan-Quinn criter.	3.650944
F-statistic	414.8492	Durbin-Watson stat	2.114085
Prob(F-statistic)	0.000000		

مثال 2: أختبر معنوية كل من المقدرات b1 و b2 المحسوبة في المثال رقم 1.

i	Y	X1	X2	$\hat{Y}$	ε_i	ε_i^2	$(Y-\bar{Y})^2$
1	40	6	4	40.32	-0.32	0.1024	289
2	44	10	4	42.92	1.08	1.1664	169
3	46	12	5	45.33	0.67	0.4489	121
4	48	14	7	48.85	-0.85	0.7225	81
5	52	16	9	52.37	-0.37	0.1369	25
6	58	18	12	57	1	1	1
7	60	22	14	61.82	-1.82	3.3124	9
8	68	24	20	69.78	-1.78	3.1684	121
9	74	26	21	72.19	1.81	3.2761	289
10	80	32	24	79.42	0.58	0.3364	529
						13.6704	**1634.00**

$$\sigma_{b_1}^2=\frac{\sum_{i}^{n}e_i^2}{n-k}\times\frac{\sum_{i=1}^{n}(X_2-\bar{X}_2)^2}{\sum_{i=1}^{n}(X_1-\bar{X}_1)^2\times\sum_{i=1}^{n}(X_2-\bar{X}_2)^2-[\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)]^2}$$

$$\sigma_{b_1}^2=\frac{13.67}{7}\times\frac{504}{(576\times504)-(524)^2}=0.06$$

$$\sigma_{b_2}^2=\frac{\sum_{i}^{n}e_i^2}{n-k}\times\frac{\sum_{i=1}^{n}(X_1-\bar{X}_1)^2}{\sum_{i=1}^{n}(X_1-\bar{X}_1)^2\times\sum_{i=1}^{n}(X_2-\bar{X}_2)^2-[\sum_{i=1}^{n}(X_1-\bar{X}_1)(X_2-\bar{X}_2)]^2}$$

$$\sigma_{b_2}^2=\frac{13.67}{7}\times\frac{576}{(576\times504)-(524)^2}=0.07$$

$\sigma_{b_2}=\sqrt{0.07}=0.26$ $t_{b_0}=\frac{b_2}{\sigma_{b_2}}=\frac{1.11}{0.26}=4.27$	$\sigma_{b_1}=\sqrt{0.06}=0.24$ $t_{b_1}=\frac{b_1}{\sigma_{b_1}}=\frac{0.65}{0.24}=2.70$

نلاحظ أن كل من t_{b1} و t_{b2} أكبر من $t_{n-k=7}^{0.05}=2.365$ المأخوذة من الجدول بدرجة حرية 7 ومستوى معنوية 5%، وبالتالي b1 وb2 ذو معنوية إحصائية عند 5%.

د. تحديد معامل التحديد المتعدد

كلما كانت المشاهدات أقرب إلى خط الانحدار كلما زاد التغير في (Y) الذي تفسره معادلة الانحدار المقدرة، فالتغير الإجمالي في (Y) يساوي التغير المفسر زائد تغير البواقي.

تسمى النسبة بين مجموع انحرافات القيم المقدرة ومجموع انحرافات القيم بمعامل التحديد؛ أي: $$R^2=\frac{SCE}{SCT}=\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2}$$
$$R^2=\frac{SCE}{SCT}=1-\frac{SCR}{SCT}\Rightarrow$$ $$R^2=1-\frac{\sum\varepsilon_1^2}{\sum(Y_i-\bar{Y})^2}$$

تتراوح قيمة معامل التحديد $0\le R^2\le 1$

مثال: أحسب معامل التحديد للمثال السابق.

$$R^2=1-\frac{\sum_{i=1}^{n}\varepsilon_i^{\,2}}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=1-\frac{13.67}{1634}=1-0.0083\simeq 0.9916=99.16\%$$

$$R^2_{Adj}=1-(1-R^2)\frac{n-1}{n-k}=1-(1-0.9916)\frac{10-1}{10-3}=1-(0.0084\times\frac{9}{7})=1-0.0108=0.9892-98.92\%$$

ملاحظة هامة:

ما الذي يحدث عند إضافة متغيرات جديدة ؟

معامل التحديد (R^2) لا ينخفض عادةً عند إضافة متغيرات مستقلة؛ غالبًا يزيد أو يبقى ثابتًا حتى لو كانت المتغيرات الجديدة ضعيفة الفائدة، وهذا قد يعطي انطباعًا مضللًا بأن النموذج تحسّن.

في هذه الحالة نقوم بحساب معامل التحديد المعدل (R^2_{Adj}) فقد يرتفع إذا أضاف المتغير الجديد تفسيرًا حقيقيًا، وقد ينخفض إذا كانت الإضافة ''حشوًا'' لا يحسن التفسير بشكل كافٍ مقابل زيادة التعقيد في الحسابات.

$$R^2_{Adj}=1-(1-R^2)\frac{n-1}{n-k}$$

اختبار المعنوية الكلية للانحدار: يمكن اختبار المعنوية الاجمالية للنموذج باستخدام نسبة التباين المفسر إلى التباين غير المفسر، يتبع هذا توزيع بدرجات حرية (k-1) و(n-k)، حيث n عدد المشاهدات وk عدد المعالم المقدرة.

$$F_{k-1,n-k}=\frac{\sum(\hat{Y}-\bar{Y})^2/(k-1)}{\sum\varepsilon_i^{\,2}/(n-k)}=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

إذا تجاوزت نسبة F المحسوبة قيمة F الجدولية عند مستوى معنوية 5% ودرجات الحرية المحددة نقبل الفرضية بأن معامل الانحدار ليست جميعها مساوية للصفر وأن معامل التحديد يختلف عن الصفر.

مثال: اختبر المعنوية الكلية للنموذج السابق.

$$F_{k-1,n-k}=\frac{\sum(\hat{Y}-\bar{Y})^2/(k-1)}{\sum\varepsilon_i^{\,2}/(n-k)}=\frac{1620.47/2}{13.67/7}=\frac{810.235}{1.9528}=414.9$$

$$F_{k-1,n-k}=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}=\frac{0.9916/2}{0.0084/7}=\frac{0.4958}{0.0012}=413.16$$

نلاحظ أن قيمة F المحسوبة أكبر من قيمة F الجدولية $F_{2,7}=4.74$ عند مستوى معنوية 5% ، وبالتالي نقبل الفرضية بأن معالم النموذج $b_1,b_2\neq 0$ وأن $R^2\neq 0$.

ملحق Eviews:

Dependent Variable: Y
Method: Least Squares
Date: 02/12/26 Time: 22:33
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.650051	0.250161	2.598527	0.0355
X2	1.109868	0.267434	4.150068	0.0043
C	31.98067	1.631796	19.59845	0.0000

R-squared	0.991634	Mean dependent var	57.00000
Adjusted R-squared	0.989243	S.D. dependent var	13.47426
S.E. of regression	1.397467	Akaike info criterion	3.750525
Sum squared resid	13.67040	Schwarz criterion	3.841300
Log likelihood	-15.75262	Hannan-Quinn criter.	3.650944
F-statistic	414.8492	Durbin-Watson stat	2.114085
Prob(F-statistic)	0.000000		