

الانحدار الخطي البسيط

أ. النموذج الخطي للمتغيرين: أو ما يعرف بتحليل الانحدار البسيط، ويستخدم لاختبار الفروض حول العلاقة بين متغير تابع (Y) ومتغير مستقل أو مفسر (X). طبعاً بعد رسم شكل الانتشار والتأكد من وجود علاقة خطية تقريبية والتي تأخذ الشكل التالي:

$$Y_i = aX_i + b \dots\dots\dots 1$$

طبعاً أثناء رسم شكل الانتشار من غير الممكن أن تقع جميع النقاط على نفس الخط تماماً، وبالتالي هناك تشويش عشوائي أو حد الخطأ (white noise) وبالتالي نكتب المعادلة أعلاه رقم 1 كما يلي:

$$Y_i = aX_i + b + e_i \dots\dots\dots 2$$

ويفترض في حد الخطأ (e_i) أن يكون:

- ✓ موزع توزيع طبيعي.
- ✓ وسطه أو توقعه الرياضي يساوي صفر.
- ✓ تباينه ثابت.
- ✓ حدود الخطأ غير مترابطة فيما بينها.

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:

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

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 (a) و (b) كما يلي:

$$a = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \dots\dots\dots 4$$

$$b = \bar{Y} - a\bar{X} \dots\dots\dots 5$$

مثال: ليكن لديك المتغيرين X, Y

Y	40	44	46	48	52	58	60	68	74	80
X	6	10	12	14	16	18	22	24	26	32

المطلوب:

- ✓ - أرسم شكل الانتشار.
- ✓ أوجد المعادلة الخطية بين X و Y

✓ المعادلة: نقوم بالحسابات التالية:

n	Y_i	X_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
1	40	6	-12	-17	204	144
2	44	10	-8	-13	104	64
3	46	12	-6	-11	66	36
4	48	14	-4	-9	36	16
5	52	16	-2	-5	10	4
6	58	18	0	1	0	0
7	60	22	4	3	12	16
8	68	24	6	11	66	36
9	74	26	8	17	136	64
10	80	32	14	23	322	196
Σ	570	180	0	0	956	576

$$\bar{X} = \frac{180}{10} = 18$$

$$\bar{Y} = \frac{570}{10} = 57$$

من أجل تقدير المعالم (a ; b) للمعادلة التالية $Y_i = aX_i + b$ نقوم بالحسابات الآتية:

$$a = \frac{\sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{10} (X_i - \bar{X})^2} = \frac{956}{576} = 1.66$$

$$b = \bar{Y} - a\bar{X} = 57 - (1.66 \times 18) = 27.12$$

على هذا الأساس تصبح معادلة الانحدار الخطي البسيط المقدرة بين المتغيرين X و Y كما يلي:

$$Y_i = 1.66X_i + 27.12$$

ج. معنوية تقديرات المعالم a و b

من أجل اختبار المعنوية الإحصائية لتقديرات المعالم a و b يلزمنا حساب تباين كل منهما.

المعدلات التالية تعطي تقديرات غير متحيزة لتباين a و b.

حيث تباين المعلمة b تحسب كما يلي:

$$\sigma_b^2 = \frac{\sum_{i=1}^n e_i^2}{n - k} \times \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

في حين تباين المعلمة a تحسب كما يلي:

$$\sigma_a^2 = \frac{\sum_i^n e_i^2}{(n-k) \cdot \sum_{i=1}^n (X_i - \bar{X})^2}$$

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

مثال 2: أختبر معنوية كل من المقدرات a و b المحسوبة في المثال رقم 1.

من أجل عملية الاختبار نقوم بالحسابات الموضحة في الجدول التالي:

n	X _i	Y _i	$\hat{y}$	$e_i = (Y_i - \hat{y})$	e^2	X _i ²	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	6	40	37.08	2.92	8.5264	36	144	289
2	10	44	43.72	0.28	0.0784	100	64	169
3	12	46	47.04	-1.04	1.0816	144	36	121
4	14	48	50.36	-2.36	5.5696	196	16	81
5	16	52	53.68	-1.68	2.8224	256	4	25
6	18	58	57	1	1	324	0	1
7	22	60	63.64	-3.64	13.2496	484	16	9
8	24	68	66.96	1.04	1.0816	576	36	121
9	26	74	70.28	3.72	13.8384	676	64	289
10	32	80	80.24	-0.24	0.0576	1024	196	529
Σ	180	570	/	0.00	47.3056	3816	576	1634

أ. نقوم أولاً بحساب $\hat{y}$

$$\hat{y}_1 = 1.66(6) + 27.12$$

$$\hat{y}_1 = 37.08 \text{ وهكذا } \dots$$

يمثل $\hat{y}$ القيم المقدرة من

خلال معادلة الانحدار.

ب. نحسب الفرق أو حد

$$e_i = (Y_i - \hat{y}) \text{ الخطأ}$$

K يمثل عدد المقدرات أو

المعالم ؛ أي K=2

$\sigma_a^2 = \frac{\sum_i^n e_i^2}{(n-k) \cdot \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{47.3056}{(10-2) \cdot 576} = 0.01$		$\sigma_a = \sqrt{0.01} = 0.1$
$\sigma\sigma_b^2 = \frac{\sum_i^n e_i^2}{n-k} \times \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_i^n e_i^2}{n-k} \times \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{47.3056}{10-2} \times \frac{3816}{10 \cdot 576} = 3.92$		$\sigma_a = \sqrt{3.92} = 1.98$
		$t_b = \frac{a - a_0}{\sigma_b} = \frac{1.66}{0.1} = 16.6$
		$t_b = \frac{a - a_0}{\sigma_b} = \frac{1.66}{1.98} = 0.84$

نلاحظ أن كل من t_a و t_b أكبر من $t_{n-k=8}^{0.05} = 2.306$ المأخوذة من الجدول بدرجة حرية 8 ومستوى معنوية

5%، وبالتالي a و b معنوية إحصائية عند 5%.

د. اختبار جودة النموذج

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.

$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$		
مجموع مربع الانحرافات القيم المقدرة $\hat{Y}$	مجموع مربع الانحرافات القيم في Y عن وسطها	مجموع مربع الأخطاء
التغير المفسر في (Y)	التغير الإجمالي في (Y)	تغير البواقي في (Y)
SCE	SCT	SCR
$\frac{SCT}{SCT} = \frac{SCE}{SCT} + \frac{SCR}{SCT} \Rightarrow 1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$		
تسمى النسبة بين مجموع انحرافات القيم المقدرة ومجموع انحرافات القيم بمعامل التحديد؛ أي:		
$R^2 = \frac{SCE}{SCT} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$		
$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \Rightarrow$		
$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$		

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

مثال: أحسب معامل التحديد للمثال السابق.

n	$(Y_i - \bar{Y})^2$	$(\hat{Y}_i - \bar{Y})^2$	$(Y_i - \hat{Y}_i)^2$
1	289	396.8064	8.5264
2	169	176.3584	0.0784
3	121	99.2016	1.0816
4	81	44.0896	5.5696
5	25	11.0224	2.8224
6	1	0	1
7	9	44.0896	13.2496
8	121	99.2016	1.0816
9	289	176.3584	13.8384
10	529	540.0976	0.0576
Σ	1634	1587.2256	47.3056
	SCT	SCE	SCR

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{1587.22}{1634} = 0.9714$$

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{47.3056}{1634} = 1 - 0.029 \approx 0.9710$$

Dependent Variable: Y
 Method: Least Squares
 Date: 10/07/25 Time: 21:45
 Sample: 1 10
 Included observations: 10

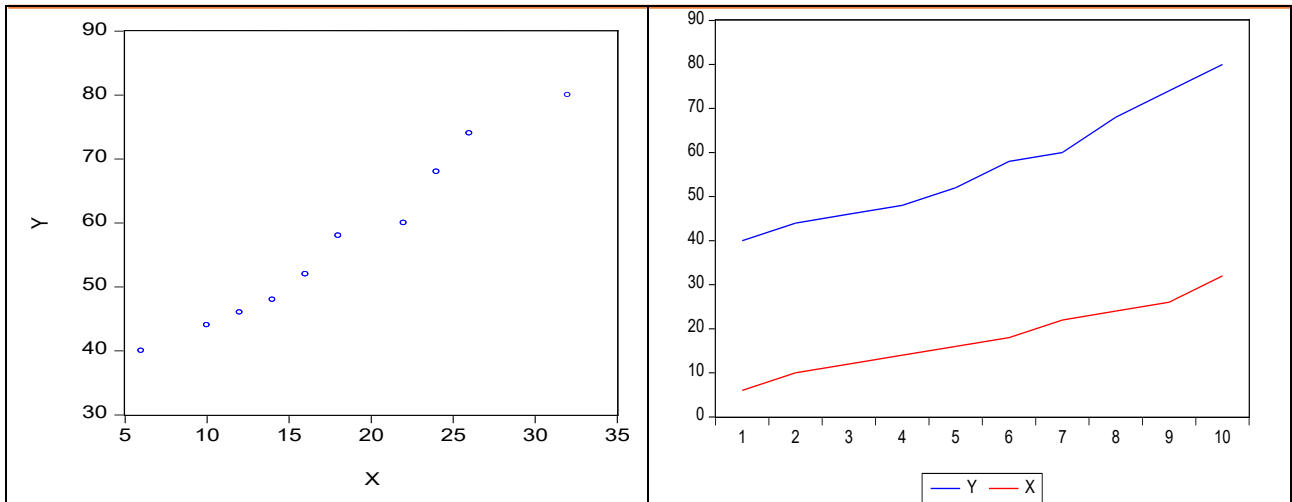
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	1.659722	0.101321	16.38082	0.0000
C	27.12500	1.979265	13.70458	0.0000
R-squared	0.971049	Mean dependent var	57.00000	
Adjusted R-squared	0.967430	S.D. dependent var	13.47426	
S.E. of regression	2.431706	Akaike info criterion	4.791920	
Sum squared resid	47.30556	Schwarz criterion	4.852437	
Log likelihood	-21.95960	Hannan-Quinn criter.	4.725533	
F-statistic	268.3312	Durbin-Watson stat	1.783613	
Prob(F-statistic)	0.000000			

Correlation		
	X	Y
X	1	0.9854
Y	0.9854	1

covariance		
	X	Y
X	57.6	95.6000
Y	95.6000	163.4

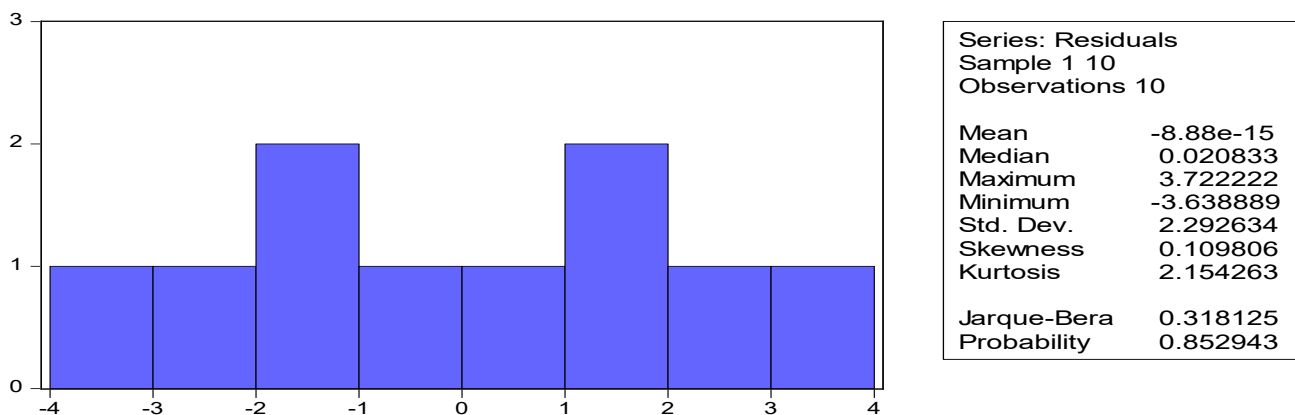
	X	Y
Mean	18.00000	57.00000
Median	17.00000	55.00000
Maximum	32.00000	80.00000
Minimum	6.000000	40.00000
Std. Dev.	8.000000	13.47426
Skewness	0.230583	0.437782
Kurtosis	2.123843	1.917612
Jarque-Bera	0.408469	0.807574
Probability	0.815271	0.667786
Sum	180.0000	570.0000
Sum Sq. Dev.	576.0000	1634.000
Observations	10	10

ملحق خاص لتحليل الانحدار البسيط: أ. لمزاودة



$$Y = 27.125 + 1.659722222222 * X$$

obs	Actual	Fitted	Residual	Residual Plot
1	40	37.0833333333333	2.9166666666667	.
2	44	43.7222222222222	0.2777777777778	.
3	46	47.0416666666667	-1.0416666666667	.
4	48	50.3611111111111	-2.3611111111111	.
5	52	53.6805555555556	-1.6805555555556	.
6	58	56.9999999999999	1.0000000000000	.
7	60	63.6388888888889	-3.6388888888889	.
8	68	66.9583333333333	1.0416666666667	.
9	74	70.2777777777778	3.7222222222223	.
10	80	80.2361111111111	-0.2361111111111	.



مثال 2:

	Y_i	X	$(Y_i - \bar{Y})$	$(X_i - \bar{X})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
--	-------	---	-------------------	-------------------	----------------------------------	---------------------

ملحق خاص لتحليل الانحدار البسيط: أ. لمزاودة

1	7389.99	8000	-2595.59	-3280	8513535	10758400
2	8169.65	9000	-1815.93	-2280	4140320	5198400
3	8831.71	9500	-1153.87	-1780	2053889	3168400
4	8652.84	9500	-1332.74	-1780	2372277	3168400
5	8788.08	9800	-1197.50	-1480	1772300	2190400
6	9616.21	11000	-369.37	-280	103423.6	78400
7	10593.45	12000	607.87	720	437666.4	518400
8	11186.11	13000	1200.53	1720	2064912	2958400
9	12758.09	15000	2772.51	3720	10313737	13838400
10	13869.62	16000	3884.04	4720	18332669	22278400
Σ	99855.75	11280	0	0	50104729	64156000

$$a = \frac{\sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{10} (X_i - \bar{X})^2} = \frac{50104729}{64156000} = 0.78 \quad \Rightarrow \quad Y_i = 0.78X_i + 1187.17$$

$$b = \bar{Y} - a\bar{X} = 9985.57 - (0.78 \times 11280) = 1187.17$$

e ²	(X-X) ²	y-y ²	$\hat{\text{var}}_e = \frac{\sum e^2}{n-2} = \frac{165231.3}{8} = 20653.91$ $\hat{\text{var}}_a = \frac{\hat{\text{var}}_e}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{20653.91}{64156000} = 0.000321$ $\sigma_a = \sqrt{\hat{\text{var}}_a} = \sqrt{0.000321} = 0.0179$ $t_a = \frac{\hat{a}}{\hat{\sigma}_a} = \frac{0.78}{0.0179} = 43.575 > t_8^{0.025} = 2.306$ نلاحظ أن t المحسوبة أكبر من الجدولية وبالتالي معلمة a معنوية.
1382.352	10758400	6737087.4481	
1407.75	5198400	3297601.7649	
55009.01	3168400	1331415.9769	
3099.149	3168400	1776195.9076	
1856.748	2190400	1434006.2500	
22788.92	78400	136434.1969	
2141.838	518400	369505.9369	
19897.92	2958400	1441272.2809	
16661.65	13838400	7686811.7001	
40986	22278400	15085766.7216	
165231.3	64156000	39296098.1839	