

統計學 W8

簡單線性迴歸模型的應用

簡單線性迴歸模型的應用

案例一：大學生成績與起薪

調查 7 名新任職大學生之在校成績 (X) 與月起薪 (Y)

原始數據與統計加總

在校學業成績 (X)	90.8	76	79.7	62.8	74.1	83	64.8
起薪 (Y, 萬元)	3.4	2.8	2.8	2.2	2.5	3.1	2.3

$$\sum X_i = 531.2$$

$$\sum X_i^2 = 40895.42$$

$$\sum Y_i = 19.1$$

$$\sum Y_i^2 = 53.23$$

$$\sum X_i Y_i = 1474.43$$

題型 599. ☆☆☆

調查 7 名新任職的大學生之在校成績 (X) 與起薪 (Y) (月薪, 單位: 萬元) 如下:

在校學業成績 (X)	90.8	76	79.7	62.8	74.1	83	64.8
起薪 (Y)	3.4	2.8	2.8	2.2	2.5	3.1	2.3

可計算出:

$$\sum_i X_i = 531.2, \quad \sum_i Y_i = 19.1,$$

$$\sum_i X_i^2 = 40895.42, \quad \sum_i Y_i^2 = 53.23, \quad \sum_i X_i Y_i = 1474.43$$

- (1) 求 X 、 Y 的樣本相關係數。
- (2) 求 Y 對 X 的直線迴歸方程式 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ 。
- (3) 在顯著水準 $\alpha = 0.05$ 下, 檢定 $H_0: \beta_1 = 0$ 對 $H_1: \beta_1 \neq 0$ 。
- (4) 求判定係數 R^2 。
- (5) 若某生在校學業成績 85 分, 試求其起薪待遇的 95% 預測區間。

[台科大工管]

$$\frac{7}{\sum_{i=1}^7 Y_i^2} = \frac{7}{53.23} = 1.315 \quad \frac{1}{\left(\sum_{i=1}^7 Y_i\right)^2} = \frac{1}{1.1143}$$

相關係數與迴歸方程計算

(1) 樣本相關係數 r

$$S_{xx} = 584.997, S_{yy} = 1.1143, S_{xy} = 25.013$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \approx 0.9797$$

結論: 成績與起薪呈極強正相關。

(2) 直線迴歸方程式

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 0.0428$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \approx -0.52$$

$$\hat{Y} = -0.52 + 0.0428X$$

假設檢定與模型判定

(4) 判定係數

95.98%

模型解釋力極高

(3) 顯著性檢定 ($\alpha=0.05$)

$$H_0 : \beta_1 = 0$$

檢定

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = 10.927$$

$$t_{0.025, 5} = 2.571$$

結果：拒絕 H_0 。成績對起薪具顯著影響。

成績 85 分之起薪預測

加1時機!!!!

(5) 95% 預測區間

代入 $X = 85$:

$$\hat{y}_0 = -0.52 + 0.0428(85) = 3.118$$

預測區間 PI:

$$3.118 \pm 2.571 \cdot 0.0947 \cdot \sqrt{1 + \frac{1}{7} + \frac{(85 - 75.89)^2}{585}}$$

區間: [2.82, 3.42] (萬元)

案例二：房產稅與房屋價值

分析房產稅 (X) 與房屋價值 (Y) 之關係

題型 601. ☆☆☆

The value of a house can be estimated with a linear relationship involving yearly property tax. The value of six houses, recorded in thousands of dollars, and the property tax on each of them, recorded in dollars, is as follows:

Property tax, X	200	175	190	220	195	235
House value, Y	50	43	46	56	52	59

The summary statistics are:

$$\sum_{i=1}^6 X_i = 1,215, \quad \sum_{i=1}^6 Y_i = 306, \quad \sum_{i=1}^6 X_i^2 = 248,375, \quad \sum_{i=1}^6 Y_i^2 = 15,786$$

- (1) State the required assumptions for conducting linear regression analysis.
- (2) Test for a linear relationship between property tax and house value.
- (3) What is a 95% confidence interval for the change in house value for one unit change in property taxes?
- (4) Calculate coefficient of determination and interpret it.
- (5) Find a 95% confidence interval for the value of a new house that has a property tax of \$180.

P. 2.

[台大國企]

房屋數據加總統計

房產稅 X (\$)	200	175	190	220	195	235
房屋價值 Y (k\$)	50	43	46	56	52	59

$$\sum X_i = 1215$$

$$\sum X_i^2 = 248375$$

$$\sum Y_i = 306$$

$$\sum Y_i^2 = 15786$$

$$\sum X_i Y_i = 62590$$

前提假設與關係檢定



(1) 模型假設

1. 線性關係
2. 殘差常態性
3. 變異數同質性
4. 獨立性



(2) 線性關係檢定

斜率估計與判定係數

(3) 斜率信賴區間 (95%)

$$\hat{\beta}_1 = \frac{625}{2337.5} \approx 0.267$$

$$CI = 0.267 \pm 2.776 \cdot 0.037$$

區間: [0.164, 0.370]

意即: 稅收每增 \$1, 房價平均增 \$164-\$370。

(4) 判定係數 R^2

$$R^2 = 0.9284$$

房產稅變動可解釋房屋價值變動的 92.84%。模型擬合度極佳。

房產稅 \$180 之房屋價值預測



(5) 95% 預測區間 (X=180)

$$\hat{Y} = -3.144 + 0.267 (180) = 44.916$$

預測區間 PI:

$$44.9876 \pm 2.776 \cdot 1.795 \cdot \sqrt{1 + \frac{1}{6} + \frac{(180 - 202.5)^2}{2337.5}}$$

預測區間: [38.83, 51.00] (k\$)

案例三

信義區公寓價值分析

研究背景: 台北市政府員工針對信義區公寓之「坪數 (Size)」與「市場價值 (Value, NT Dollars)」進行迴歸分析。

迴歸方程:

$$\text{Value} = -1,859 + 65 \cdot \text{Size}$$



Predictor	Coef	SE Coef
Constant	-1859	231.4
Size	64.993	3.047

Analysis of Variance			
Source	DF	SS	MS
Regression	1	13548662082	13548662082
Residual Error	33	982687392	29778406
Total	34	14531349474	

- (1) How many apartments were in the sample?
- (2) Based on the regression equation, what is the estimated value of an apartment in XinYi District of size 20 (坪)? Is this a reasonable estimate?
- (3) Determine and interpret the coefficient of determination.
- (4) Determine the coefficient of correlation. How did you determine the sign of the correlation coefficient?
- (5) Conduct a test of hypothesis to determine if there is significant positive association between the market value of apartments and the size of the apartment in level ground. Use the 0.01 significance level. [政大國貿]

解 (1) $n - 1 = 34 \implies n = 35$

模型統計摘要與 ANOVA

Predictor	Coef	SE Coef
Constant	-1859	231.4
Size (坪)	64.993	3.047

Source	DF	SS	MS
Regression	1	13,548,662,082	13,548,662,082
Residual	33	982,687,392	29,778,406
Total	34	14,531,349,474	

資料來源: 政大國貿 考題

樣本數量與價值預測

(1) 樣本中共有多少公寓？

根據 Total DF = 34:

$$n - 1 = 34 \Rightarrow n = 35$$

結論：樣本數量為 **35 間** 公寓。

(2) 20 坪公寓之價值估計

$$V = -1,859 + 65(20)$$

$$V = -1,859 + 1,300 = -559$$

不合理。市場價值不可能為負數。這顯示 20 坪已超出模型有效預測範圍(外插法風險)。

判定係數、相關係數與檢定

(3) 判定係數 R^2 與 (4) 相關係數 r

$$R^2 = \frac{SSR}{SST} \approx 0.9324$$

解釋: 93.24% 的價值變動可由坪數解釋。

$$r = +\sqrt{0.9324} \approx 0.9656$$

因斜率 (64.993) 為正, 故 r 為正號。

(5) 顯著正向關聯檢定 ($\alpha=0.01$)

$$H_0: \beta_1 \leq 0 \text{ vs } H_1: \beta_1 > 0$$

$$t = \frac{64.993}{3.047} \approx 21.33$$

$$\text{查表: } t_{0.01, 33} \approx 2.445$$

t 值遠大於臨界值, 強烈拒絕 H_0 。坪數與房價具顯著正相關。