

1. 已知資料如下表所示：

y	10	20	30	40
$x_1$	3	5	9	11
$x_2$	6	3	10	5

試求迴歸方程  $\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

2. 已知資料如下表所示：

y	10	20	30	40
$x_1$	3	5	9	11
$x_2$	6	3	10	5

且二元迴歸方程為： $\hat{y} = 0.667 + 5x_1 + 4x_2$

- (1) 試檢定此迴歸方程是否可有效的解釋依變數  $y$ 。(  $\alpha = 0.05$  )
  - (2) 試求判定係數。
3. 承習題 2，請分別檢定自變數  $x_1$  與  $x_2$  對依變數是否具解釋力 (  $\alpha = 0.05$  ) ？並求  $\beta_1, \beta_2$  的 95% 信賴區間。
4. 試證明複判定係數  $R^2$  一定大於或等於調整判定係數  $R_a^2$ 。
5. 假設  $y$  對  $x_1, x_2, x_3, x_4, x_5$  作複迴歸，模式為：

$$\text{模式 I: } \hat{y} = \alpha + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5$$

經計算得到部分的變異數分析表：

變異來源	平方和	自由度	均方和	F
迴歸				
誤差	200			
總和	230	45		

- (1) 請完成變異數分析表。
  - (2) 試求判定係數，並且在  $\alpha = 0.05$  下，檢定  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  是否成立。(  $F_{0.05, 5, 40} = 2.4495$  )
  - (3) 若  $y$  對  $x_2, x_3, x_4, x_5$  作複迴歸，得到  $SSE = 210$ ，試問對模式 I 中檢定  $H_0: \beta_1 = 0$  是否顯著？(  $\alpha = 0.05, F_{0.05, 40} = 4.08$  )
6. 迴歸模型  $y = a + bz + cq + dw + ex + fr + \varepsilon$  中， $z, q, w, x$  與  $r$  為解釋變數，隨機抽取 26 個樣本建立迴歸方程式後，得出迴歸方程式為：

$$y = 1.2 + 3.0z - 1.5q - 0.9w + 3.5x + 2.6r$$

(0.3)
(1.3)
(0.8)
(1.2)
(2.3)
(0.3)

括號內為各迴歸係數之標準誤(Standard error)。已知來自迴歸的變異為 300，來自隨機誤差的變異為 200。試求下列各小題 (  $\alpha = 0.05$  )：

- (1) 檢定此迴歸模型是否合適。
  - (2) 求複相關、複判定係數與調整判定係數。
  - (3) 檢定解釋變數  $r$  之係數  $f$  是否為 2。
  - (4) 求解釋變數  $q$  之係數  $c$  的 95% 信賴區間，並檢定是否為 0。
7. 假設依變數  $y$  與兩個解釋變數  $x_1, x_2$  所開列之迴歸方程式如下所示：括號內為各迴歸係數之標準誤(Standard error)，樣本數  $n = 20$ 。

模式 I： $\hat{y} = 40.4 - 0.208x_1$   
(0.112)

模式 II： $\hat{y} = 12.2 + 3.22x_2$   
(0.57)

模式 III： $\hat{y} = 9.14 + 0.0364x_1 + 3.38x_2$   
(0.09) (0.7)

- (1) 請問三種模式以哪種模式較佳？
  - (2) 承上題，所選定的模式若將其中一個自變數維持不變，另一個自變數增加三個單位，請問依變數  $y$  可增加幾個單位？
8. 依據多年的房屋銷售經驗，博碩房屋仲介公司認為台北市信義區沒有電梯的公寓房價受到該公寓面積及屋齡的影響最大。為了對公寓房價的結構有更進一步的了解，博碩公司利用最近成交之 23 間屋齡 16 年以下房屋的資料，經統計軟體計算，得到下列部份的結果 ( $\alpha = 0.05$ )：

$y$ ：成交房價(單位：10 萬元)

$x_1$ ：房屋面積(單位：坪)     $x_2$ ：屋齡(單位：年)

Predictor	Coefficient	Standard Error	t statistic	P-Value
Intercept	57.35	10.01	(A)	0.000
$x_1$	3.54	0.62	(B)	0.000
$x_2$	-2.22	0.76	(C)	0.008

$s=11.96$      $R\text{-sq}=(D)$

$R\text{-sq(adj)}=71.5\%$

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Sum of Squares	$F$	$P$
Regression	2	8189.7	4094.9	(E)	0.0000
Errors	20	2861.0	143.1		
Total	22	11050.7			

- (1) 請問請  $x_1$  對依變數是否有解釋力？ $x_1$  的係數 3.54 的意義為何？
- (2) 請解釋  $t$ -statistic 的作用為何？
- (3) 請解釋  $F$  值的作用為何？
- (4) 請解釋  $P$ -Value 的意義為何？
- (5) 請解釋  $R\text{-sq}$  的意義為何？
- (6) 請問為何有了  $R\text{-sq}$  還要有  $R\text{-sq(adj)}$ ？

(7)請你計算上面表格(A)、(B)、(C)、(D)與(E)的值

(8)若針對一間 30 坪，屋齡 10 年的舊公寓售價和數十間都是 30 坪，屋齡 10 年的舊公寓之平均售價作預估，何者較為準確？為什麼？

9. 下表為統計軟體針對某迴歸問題所產生的報表：

Predictor	Coefficient	Stdev
Intercept	10.00	2.00
$x_1$	-2.00	1.50
$x_2$	6.00	2.00
$x_3$	-4.00	1.00

變異數分析表

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Regression		60		
Errors				
Total	19	140		

(1)根據上表請問一共抽取了幾個樣本？

(2)請問二元迴歸方程式為何？

(3)請計算複判定係數與複相關係數。

(4)求調整判定係數

(5)試檢定整體迴歸模型是否合適？( $\alpha = 0.05, F_{0.05,3,16} = 2.239$ )

(6)試檢定自變數  $x_3$  對依變數是否具影響力？( $\alpha = 0.05$ )

解

10. 以最小平方方法估測  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ ，統計軟體報表結果顯示：

Variable	Estimate	t statistic
Intercept	8.029	4.47
$x_1$	-0.996	-2.51
$x_2$	-0.718	-1.66
$x_3$	0.211	2.22

$$R^2 = 0.9775, F = 374.5, n = 30$$

請問：

(1)  $\beta_1$  之估計值的標準誤為何？

(2)請檢定  $\begin{cases} H_0: \beta_0 = 2 \\ H_1: \beta_0 \neq 2 \end{cases}$ 。 ( $\alpha = 0.05$ )

11. 針對 10 個觀測值，假設已知其二元迴歸方程如下所示：

$$\hat{y} = 29.3468 + 5.3128x_1 + 3.8344x_2$$

$$(0.2285) \quad (0.4332)$$

括號內為各迴歸係數之標準誤。

經計算得： $SST=7447.509$ ， $SSR=7373.952$

(1)請建立變異數分析表，並檢定整體迴歸模型是否合適？( $\alpha = 0.05$ )

(2)請檢定自變數  $x_1$  對依變數是否具解釋力？( $\alpha = 0.05$ )

(3)請檢定自變數  $x_2$  對依變數是否具解釋力？( $\alpha = 0.05$ )

12. 某項針對影響銷售量因素的研究進行迴歸分析，依變數與自變數分別代表如下： $y$  表銷售金額（百萬元）、 $x_1$  表人口密度（人/平方公里）、 $x_2$  表廣告費用（萬元）、 $x_3$  表銷售區域（0=北部，1=南部）、 $x_4$  表競爭者數量（家）、 $x_5$  表銷售員數量（人）。全部共有 100 個樣本，已知由電腦所跑出的部份資料如下所示：參數估計表：

	Coefficients	Standard Error	$t$ statistic	$P$ vaule
Intercept	7.99	2.57	3.10	0.0027
$x_1$	0.11	0.01	15.51	0.0000
$x_2$	-1.48	1.47	-1.00	0.3188
$x_3$	-2.43	1.03	-2.37	0.0205
$x_4$	3.02	0.25	12.22	0.000
$x_5$	1.32	1.47	0.90	0.3715

變異數分析表：

	$D.F$	$SS$	$MS$	$F$	$P$ vaule
Regression	5	11892.61			0.0000
Errors					
Total	72	12884.33			

(1)請完成變異數分析表。

(2)請問此複迴歸方程式為何？

(3)請計算複判定係數與複相關係數。

(4)求調整判定係數

(5)試檢定整體迴規模型是否合適？( $\alpha = 0.05$ )

(6)請問此迴歸模型引入哪幾個解釋變數即可？( $\alpha = 0.05$ )

13. 以下為統計軟體所跑出的結果。依變數與自變數分別代表如下： $y$  表軟體開發所需時間(月)、 $x_1$  表軟體程式碼行數(千行)、 $x_2$  表專案經理的工作年資(年)、 $x_3$  表舊軟體程式碼的可再用程度(百分比)。

參數估計表：

Variable	Parameter Estimate	Standard error	$t$ statistic	$P$ vaule
Intercept	0.533	1.495	0.357	0.726
$x_1$	0.112	0.038	2.919	0.010
$x_2$	0.758	0.286	2.652	0.017
$x_3$	-0.052	0.020	-2.550	0.021

變異數分析表：

Model	$D.F$	$SS$	$MS$	$F$	$P$ vaule
Regression	3	283.048	94.349	15.311	0.001
Errors	16	98.594	6.162		
Total	19	381.642			

試回答下列問題：

- (1)請問樣本數為何？
- (2)請寫出複迴歸方程式？
- (3)請計算複判定係數與調整判定係數？
- (4)請問軟體程式碼行數、專案經理的工作年資、舊軟體程式碼的可再用程度，在顯著水準 0.05 下，三個自變數是否可解釋軟體開發所需時間？
- (5)若自變數  $x_2$  為第一個被引入迴歸模式中之變數，已知：

$$SSR(x_1|x_2) = 32.895, SSE(x_1, x_2) = 138.675, SSE(x_2, x_3) = 151.098。$$

請計算軟體開發所需時間(y)與專案經理的工作年資( $x_2$ )迴歸模式的判定係數與調整判定係數

- (6)請利用偏判定係數來說明，軟體程式碼行數( $x_1$ )與舊軟體程式碼的可再用程度( $x_3$ )中，哪個自變數宜先引入？( $x_2$  已事先引入)
14. 假設  $n = 15$ ,  $SSR(x_1, x_2, x_3) = 3900$ ,  $SSR(x_1) = 2400$ ,
- $$SSR(x_2|x_1) = 1000, SSR(x_3|x_1, x_2) = 500, SSE(x_1, x_2, x_3) = 1100$$
- ，試求

(1)  $R_{123}^2$  與  $R_{y3.12}^2$ 。

(2)檢定  $\rho_{y123}$  與  $\rho_{y3.12}$  是否為 0。(  $\alpha = 0.05$  )

15. 已知有 19 筆觀察值，進行迴歸分析，下列為電腦所計算出的部分資料：

變異來源	平方和
迴歸	$SSR(x_1, x_2) = 844$
	$SSR(x_1) = 480$
	$SSR(x_2 x_1) = 364$
誤差	$SSE(x_1, x_2) = 156$

試求複相關係數  $r_{y12}$ ，偏相關係數  $r_{y2.1}$  與簡單線性相關係數  $r_{y1}$ 。

16. 若母體迴歸方程式為  $y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ 。請用最小平方法求出  $\beta_1$  及  $\beta_2$  的估計式為何？