

調整的判定係數(adjusted coefficient of determination):

使用判定係數 $R^2 = SSR/SST = 1 - SSE/SST$ 會隨著解釋變數個數的變多而變大(因 SSE 變小)，雖然有些解釋變數並無顯著的解釋能力。因此另定義調整的判定係數

$$\bar{R}^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}。$$

調整的判定係數提供一個較佳的比較準則來選取解釋變數的個數。當解釋變數的個數不多時，判定係數 R^2 與調整的判定係數 \bar{R}^2 差別不大，但當解釋變數的個數變多時，兩者就會有顯著的不同。

STAT 2622 統計資料分析 作業十 (不用交)

1. 假設有 20 筆(X,Y)的資料，其最小平方迴歸線為 $Y = 10.28 + 4.92X$ 且 $S_{xx} = 34.784, S_{yy} = 911.52$ ，解出下列問題：

- (a) X 和 Y 的相關係數。
- (b) 建立 ANOVA 表。
- (c) 求出判定係數。
- (d) 使用 F 檢定來檢定斜率 $H_0: \beta=0$ 對 $H_1: \beta \neq 0$ (檢定顯著水準 5%)。
- (e) 使用 t 檢定來檢定斜率 $H_0: \beta=0$ 對 $H_1: \beta \neq 0$ (檢定顯著水準 5%)。

2. 台灣 22 個行政區之空氣污染指數(PSI)平均值、 $PM_{10}(ug/m^3)$ 和 $O_3(ppb)$ 臭氧)之資料如下：

行政區	1	2	3	4	5	6	7	8	9	10	11	12
污染指數	44	52	52	54	51	53	53	58	58	58	63	61
PM_{10}	35	52	45	54	48	46	47	58	57	63	61	64
O_3	39	23	28	28	26	30	30	28	27	26	26	29
行政區	13	14	15	16	17	18	19	20	21	22		
污染指數	64	64	64	63	65	67	60	44	38	38		
PM_{10}	71	73	72	74	73	72	58	40	33	34		
O_3	26	30	30	26	28	30	34	27	25	29		

- (a) 以污染指數為應變數 PM_{10} 和 O_3 為解釋變數，求最小平方迴歸超

平面。

(b) SSE 、 SSR 和 SSE 各為多少?

(c) 給定檢定顯著水準 5%，使用 F 檢定來檢定多元迴歸模式是否有解釋能力。

(d) 給定檢定顯著水準 5%，檢定 PM_{10} 對污染指數是否有解釋能力。

(e) 當 $PM_{10} = 60, O_3 = 60$ 時，求空氣污染指數(PSI)平均值的估計值。

3. 假設採用 5 個解釋變數來建構一個多元線性迴歸模式，得到

ANOVA 表如下:

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

(a) 完成上面的 ANOVA 表

(b) $R^2 = ?$

(c) 給定檢定顯著水準 5%，檢定 $H_0: \beta_1 = \dots = \beta_5 = 0$ 是否顯著?

4. 酒能誤人肇事，衛生署公佈血液酒精濃度與肇事危險關係密切，

提出下列數據:

血液中酒精	肇事危險性	體重(kg)			
濃度(%)		59	68	77	86

飲用的啤酒罐數(每罐 355cc)

0	1	0	0	0	0
0.05	2	2.9	3.3	3.7	4.4
0.08	6	4.5	5.3	6.7	6.8
0.11	10	6.4	7.3	8.4	9.6
0.15	25	8.6	10	11.3	12.6
0.17	50	9.8	11.3	12.8	14.4

- (a) 求出血液中酒精濃度(反應變數)對體重和飲用的啤酒罐數(二個解釋變數)的最佳迴歸超平面。
- (b) 某人體重 65 公斤，飲啤酒 5 罐，其血液中酒精濃度大約是多少?
- (c) 求出肇事危險性(反應變數)對體重和飲用的啤酒罐數(二個解釋變數)的最佳迴歸超平面。
- (d) 某人體重 65 公斤，飲啤酒 5 罐，其肇事危險性大約是多少?
- (e) 求出肇事危險性(反應變數)對血液中酒精濃度、體重和飲用的啤酒罐數(三個解釋變數)的最佳迴歸超平面。