

等 別：三等考試

類 科：統計

科 目：迴歸分析

考試時間：2小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)作答時請參考表一、二、三、四、五及圖一。

一、考慮簡單線性迴歸模型如下：

$$E(Y | X = x) = \beta_0 + \beta_1 x \quad (1)$$

若解釋變數 X 的值替代為 $Z = aX + b$ ， $a \neq 0$ 且 b 為常數，則模型 (1) 改寫為：

$$E(Y | Z = z) = \gamma_0 + \gamma_1 z \quad (2)$$

(一)請比較 β_0 與 γ_0 、 β_1 與 γ_1 的關係。(10分)

(二)請問模型 (1) 與模型 (2) 的判定係數是否改變？(回答是或否即可) (2分)

二、若反應變數為 Y ，解釋變數為 X_j ， $j=1,2,\dots,p$ ，及 n 個觀測值。考慮線性迴歸模型如下：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i=1,2,\dots,n \quad (3)$$

其中 ε_i 為均數是 0，變異數是 σ^2 的隨機誤差項。若將模型 (3) 以向量及矩陣方式表達如下：

$$Y = X\beta + \varepsilon \quad (4)$$

(一)請分別定義 Y 、 X 、 β 及 ε 之向量及矩陣之表達式，並標示其行與列的大小。(8分)(二)試求模型 (4) 中， β 的最小平方估計式。(10分)

(三)證明題(二)所得的最小平方估計式為不偏的。(5分)

(四)若欲求得 β 的最大概似估計式，需對誤差 ε 有如何的假設？(2分)三、表一為民國 101 年縣市有關教育的資料（最後兩列分別為各變數值的加總與平方後之加總），圖一為其對應之兩兩變數散布圖矩陣（Scatter matrix），表二為這些變數之變異共變異矩陣（Variance-covariance matrix）。若考慮 X_3 及 X_4 放入模型中，表三為其估計結果。表四為僅考慮 X_4 放入模型中的估計結果。以題二中迴歸模型 (4) 的表達方式，表五為僅考慮 X_3 在模型中 $(X^T X)^{-1}$ 與 $(X^T Y)$ 的結果（上標 T 代表矩陣的轉置）。

請回答下列問題：

(一)在 Y 與 X_1 的散布圖中，可看到一個明顯的離群值（Outlier），請說明為那一個縣市？(2分)

(二)請計算所有變數之兩兩變數間的相關係數矩陣（Correlation matrix）。(10分)

(請接第二頁)

等 別：三等考試
類 科：統計
科 目：迴歸分析

- (三)若將題(一)中所發現的離群值排除後，再計算 Y 與 X_1 的相關係數。另外，若將該離群值排除，已知不會影響 X_3 及 X_4 的相關係數。請建議後續統計分析（包含迴歸分析）該如何處理此一離群值。（6分）
- (四)請說明表三中三個「 t statistic」的意義，及其值與所對應之 p value 所代表之結論。（6分）
- (五)請說明表三中「Residual standard error」的意義。（5分）
- (六)請說明表三中「F-statistic」的意義，及其值與所對應之 p value 所代表之結論。（5分）
- (七)請比較題(四)及題(六)的結論是否一致？無論一致與否，皆請說明為何能有這樣的結果。（5分）
- (八)表三與表四中所得 X_4 的迴歸係數估計皆為正的，是否可說明「國中生視力不良率愈高，大專以上學歷所占比例愈高；高視力不良率可提升國民的教育程度，因此視力不良率很高不是一件不好的事。」請評論引號中的話。（5分）
- (九)請說明為何表四的「Multiple R-squared」比表三的值小，但表四的「Adjusted R-squared」卻比表三的值大。（5分）
- (十)僅考慮 X_3 在模型中的簡單線性迴歸模型，請計算其截距與斜率的估計值。（6分）
- (十一)若考慮下列三個模型：
- $$Y = \beta_{01} + \beta_{31}X_3 + \beta_{41}X_4 + \varepsilon$$
- $$Y = \beta_{02} + \beta_{32}X_3 + \varepsilon$$
- $$Y = \beta_{03} + \beta_{43}X_4 + \varepsilon$$
- 那一個模型為最適模型？請寫出理由及所根據的準則。（8分）

等 別：三等考試
類 科：統計
科 目：迴歸分析

表一

	15歲以上民間人口之教育程度結構-大專及以上 ($Y, \%$)	平均每一教師教導學生數-國小 (X_1)	平均每一教師教導學生數-國中 (X_2)	視力不良率-國小 ($X_3, \%$)	視力不良率-國中 ($X_4, \%$)
新北市	37.9	15.0	14.4	54.6	77.6
臺北市	63.3	12.8	12.7	52.4	78.1
臺中市	39.0	16.3	14.0	53.4	78.1
臺南市	34.9	15.9	14.6	49.6	75.9
高雄市	37.6	16.0	14.0	51.2	74.1
宜蘭縣	29.2	13.6	13.4	43.3	68.2
桃園縣	35.1	17.1	14.3	49.5	74.1
新竹縣	34.8	15.2	12.8	47.0	70.1
苗栗縣	27.1	12.9	12.1	43.9	67.2
彰化縣	28.1	16.0	14.5	52.7	79.6
南投縣	27.7	11.8	13.1	41.4	66.8
雲林縣	24.7	13.5	13.8	44.9	65.8
嘉義縣	22.9	12.1	12.8	41.5	67.4
屏東縣	27.0	13.5	14.6	38.9	61.2
臺東縣	19.4	9.5	11.8	31.1	55.1
花蓮縣	29.4	10.9	12.3	36.6	59.5
基隆市	35.1	14.8	12.5	53.0	74.4
新竹市	46.1	16.7	13.2	49.7	73.3
嘉義市	49.8	17.5	14.7	54.0	78.7
總和	649.1	271.1	255.6	888.7	1345.2
平方和	24124.55	3957.55	3454.52	42366.01	96138.94

表二

	Y	X_1	X_2	X_3	X_4
Y	108.29	11.24	1.91	47.15	49.93
X_1	11.24	4.97	1.50	12.37	12.70
X_2	1.91	1.50	0.89	3.36	3.65
X_3	47.15	12.37	3.36	44.35	45.79
X_4	49.93	12.70	3.65	45.79	49.93

(請接第四頁)

等 別：三等考試
類 科：統計
科 目：迴歸分析

表三

	<i>Estimate</i>	<i>Std Err</i>	<i>t statistic</i>	<i>p value</i>
<i>Intercept</i>	-26.12	29.4966	-0.886	0.389
X_3	0.58	1.2380	0.468	0.646
X_4	0.47	1.1667	0.402	0.693

Residual standard error: 8.048 on 16 degrees of freedom
Multiple R-squared: 0.4683, Adjusted R-squared: 0.4018
F-statistic: 7.046 on 2 and 16 DF, p-value: 0.006388

表四

	<i>Estimate</i>	<i>Std Err</i>	<i>t statistic</i>	<i>p value</i>
<i>Intercept</i>	-36.63	18.6525	-1.964	0.06611
X_4	1.00	0.2622	3.813	0.00139

Residual standard error: 7.861 on 17 degrees of freedom
Multiple R-squared: 0.461, Adjusted R-squared: 0.4293
F-statistic: 14.54 on 1 and 17 DF, p-value: 0.00139

表五

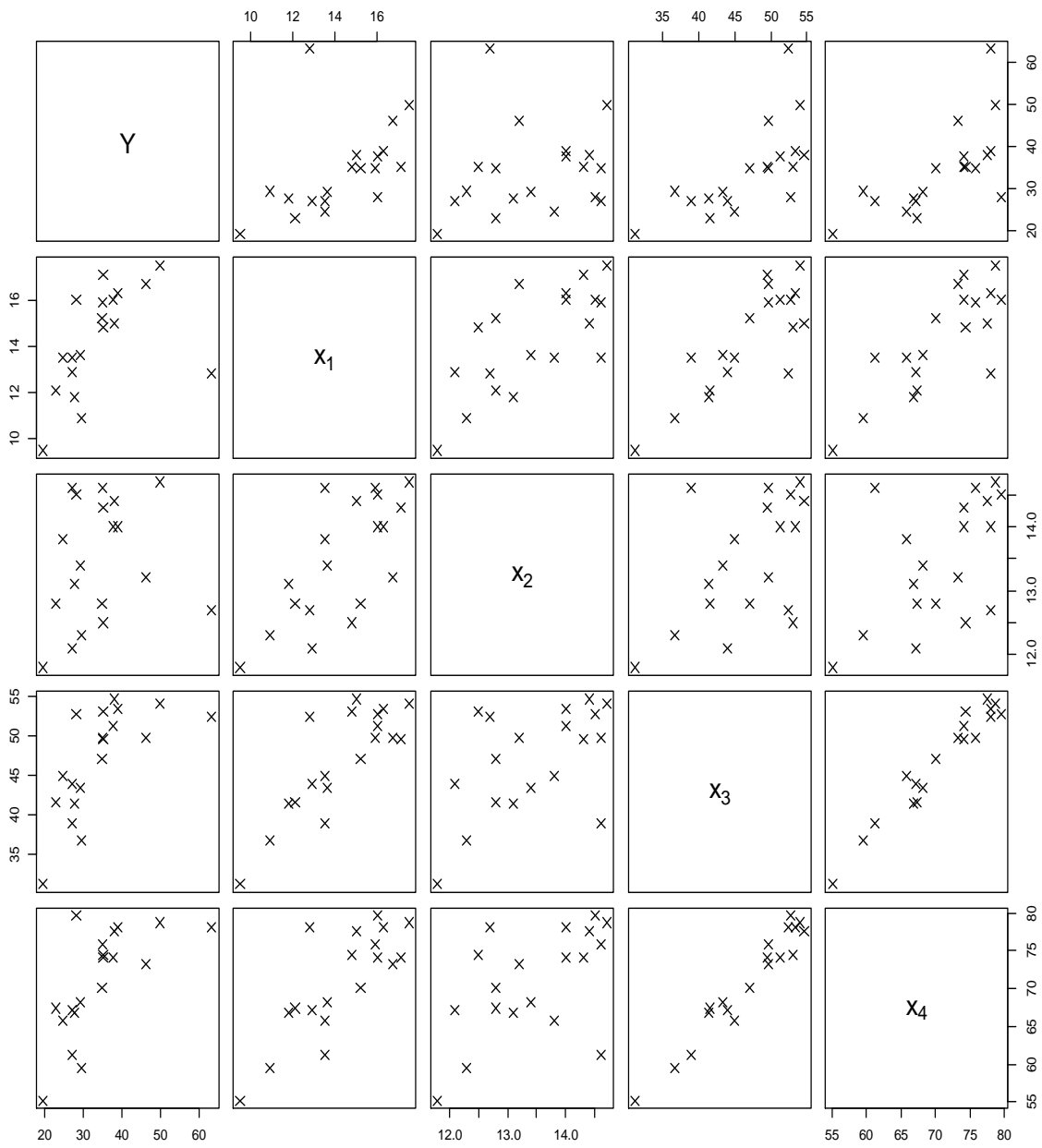
$$(X^T X)^{-1}$$

	<i>Intercept</i>	X_3
<i>Intercept</i>	2.7933940	-0.0585962
X_3	-0.0585962	0.0012528

$$(X^T Y)^T = (649.1 \quad 31209.5)$$

(請接第五頁)

等 別：三等考試
類 科：統計
科 目：迴歸分析



圖一