《迴歸分析》

試題評析	今年高考迴歸分析沒有出現計算繁瑣或觀念困難之題型，只要在考前有練習考古題，要得到滿分並非難事。不過，還是跟以往國考題目一樣，一些推論的部份所需之模型必要假設仍然沒有提供，故考生千萬別自顧計算而忘記最重要的必要假設。本卷基本分90以上。
高分命中	第一題：《高點迴歸分析熱門題庫書籍》，趙治勳老師編著，頁2-9~頁2-13。 第二題：《高點迴歸分析熱門題庫書籍》，趙治勳老師編著，頁2-61。 第三題：(一)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-9~頁2-10。 (二)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-21。 (三)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-12。 第四題：(一)至(三)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-42。 (四)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-40。 第五題：(一)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，A1例1(4)。 (二)《高點統計學總複習講義第一回》趙治勳老師編撰，例12。 第六題：(一)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-21。 (二)《高點迴歸分析熱門題庫書籍》趙治勳老師編撰，頁2-54<註>2.。

本試題可能使用之查表值如下：

$t_{9,0.05}=1.833$, $t_{9,0.025}=2.262$, $t_{10,0.05}=1.812$, $t_{10,0.025}=2.228$, $t_{18,0.01}=2.552$, $t_{18,0.02}=2.214$, $t_{19,0.01}=2.539$, $t_{19,0.02}=2.205$, $t_{20,0.01}=2.528$, $t_{20,0.02}=2.197$

$F_{1,5,0.05}=6.61$, $F_{1,6,0.05}=5.99$, $F_{2,5,0.05}=5.79$, $F_{2,6,0.05}=5.14$

一、若$n=10$, $\sum X_i=70$，$\sum Y_i=185$，$\sum X_i^2=652$，$\sum Y_i^2=3793$，$\sum X_iY_i=1537$，請計算表中所列(1)～(4)的簡單線性迴歸的參數估計值與標準誤。（16分）

參數	估計值	標準誤
β_0	$b_0=$ (1)	(2)
β_1	$b_1=$ (3)	(4)

答：

(一) $b_0=\bar{Y}-b_1\bar{X}=\dfrac{185}{10}-1.4938(\dfrac{70}{10})=8.0434$

(二) $\sqrt{V(b_0)}=\sqrt{(\dfrac{1}{n}+\dfrac{\bar{X}^2}{SS_X})MSE}=\sqrt{(\dfrac{1}{10}+\dfrac{(7)^2}{162})(1.1259)}=0.6732$

其中 $MSE=\dfrac{SSE}{n-k-1}=\dfrac{SST-SSR}{n-k-1}=\dfrac{SS_Y-b_1^2SS_X}{n-k-1}$

$=\dfrac{370.5-(1.4938)^2(162)}{10-1-1}=1.1259$

(三) $b_1=\dfrac{SS_{XY}}{SS_X}=\dfrac{242}{162}=1.4938$

(四) $\sqrt{V(b_1)}=\sqrt{\dfrac{MSE}{SS_X}}=\sqrt{\dfrac{1.1259}{162}}=0.0834$

二、下表所列為四個自變數X_1，X_2，X_3，X_4的複迴歸問題，分別以其中一個為因變數，而以其他三

個為自變數時，所得到的判定係數R^2。請計算(1)～(4)變異數膨脹因子（Variance Inflation Factor, VIF）。（8分）

因變數	判定係數R_2	變異數膨脹因子VIF
X_1	0.0842	(1)
X_2	0.9642	(2)
X_3	0.1141	(3)
X_4	0.5687	(4)

答：

(一) $\dfrac{1}{1-R_1^2}=\dfrac{1}{1-0.0842}=1.0919$

(二) $\dfrac{1}{1-R_2^2}=\dfrac{1}{1-0.9642}=27.933$

(三) $\dfrac{1}{1-R_3^2}=\dfrac{1}{1-0.1141}=1.1288$

(四) $\dfrac{1}{1-R_4^2}=\dfrac{1}{1-0.5687}=2.3186$

三、某公司受委託進行飲料公司飲料自動販賣機的定期維修。該公司發現自動販賣機的每半個月維修費支出（Y：以元為單位），與自動販賣機之銷售量（X：以千元為單位）密切相關，為進一步瞭解二者間之關係，隨機抽取若干台飲料自動販賣機，記錄其每半個月維修費支出及銷售量金額，使用簡單線性迴歸模式進行分析。假設取得之樣本為：

i	1	2	3	4	5	6	7	8	9	10	11
X_i	2.5	2.7	3.2	3.8	4.0	4.4	4.8	5.4	6.9	7.9	8.8
Y_i	63	60	64	69	70	76	93	102	111	128	133

(一)請利用最小平方法計算參數β_0與β_1之估計值。（10分）

(二)請計算判定係數R^2之值。（5分）

(三)請計算參數β_1估計值b_1之估計標準誤（standard error）。（5分）

答：

假設模型：$Y_i=\beta_0+\beta_1X_i+\varepsilon_i, i=1,2,\cdots,11$

$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

(一) $\hat{\beta}_1=\dfrac{SS_{XY}}{SS_X}=\dfrac{553.4545}{44.2073}=12.5195$

$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}=\dfrac{969}{11}-12.5195(\dfrac{54.4}{11})=26.1763$

(二) $R^2=\dfrac{SSR}{SST}=\dfrac{\hat{\beta}_1^2SS_X}{SS_Y}=\dfrac{(12.5195)^2(44.2073)}{7188.9091}=0.9638$

(三) $\sqrt{V(\hat{\beta}_1)}=\sqrt{\dfrac{MSE}{SS_X}}=\sqrt{\dfrac{28.8834}{44.2073}}=0.8083$

其中$MSE=\dfrac{SSE}{n-k-1}=\dfrac{SST-SSR}{11-1-1}=\dfrac{SS_Y-\hat{\beta}_1^2 SS_X}{9}=28.8834$

四、若因變數Y和三個自變數X_1，X_2，和X_3的複迴歸所得之ANOVA表如下：

Source	df	SS
Reg	3	1029.9
Res	18	575.8
Total	21	1605.7

且參數估計值等表列如下：

參數	估計值	標準誤	p-value	型一SS
β_0	48.63	18.56	0.0173	113654
β_1	-0.0258	0.172	0.8826	597.56
β_2	-0.507	0.235	0.0445	1.57
β_3	1.237	0.337	0.0018	430.76

(一)請計算檢定H_0：$E[Y]=\beta_0+\beta_1X_1+\beta_3X_3$ vs. H_a：$E[Y]=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3$之檢定統計值。（3分）

(二)請計算檢定H_0：$E[Y]=\beta_0+\beta_1X_1$ vs. H_a：$E[Y]=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3$之檢定統計值。（3分）

次若Y對X_3之迴歸的ANOVA表如下：

Source	df	SS
Reg	1	849.9
Res	20	(1)
Total	21	(2)

(三)請計算(1)、(2)及檢定H_0：$E[Y]=\beta_0+\beta_3X_3$ vs. H_a：$E[Y]=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3$之檢定統計值。（6分）

又若樣本數n=22，且因變數Y與自變數X_1、X_2、X_3之複迴歸所得的相關係數矩陣如下：

	Y	X_1	X_2	X_3
Y	1.0	-0.610	0.019	0.728
X_1		1.0	-0.082	-0.703
X_2			1.0	0.439
X_3				1.0

(四)請計算偏相關係數（partial correlation coefficients）$r_{Y1.2}$（3分）和$r_{Y2.13}$（6分）之值。

答：

假設模型：$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_1X_{3i}+\varepsilon_i, i=1,2,\cdots,22$

$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

(一)T.S.: $T^*=\dfrac{\hat{\beta}_2-0}{S(\hat{\beta}_2)}=0.0445$

(二)T.S.: $F^*=\dfrac{SSR(X_2,X_3|X_1)/2}{SSE(X_1,X_2,X_3)/22-3-1}=\dfrac{[SSR(X_2|X_1)+SSR(X_3|X_1,X_2)]/2}{SSE(X_1,X_2,X_3)/18}$

$=\dfrac{[1.57+430.76]/2}{575.8/18}=6.7575$

(三)(1)755.8 (2)1605.7

$$\text{T.S.:}\ F^* = \frac{SSR(X_1, X_2 \mid X_3)/2}{SSE(X_1, X_2, X_3)/22-3-1} = \frac{[SSR(X_1, X_2, X_3) + SSR(X_3)]/2}{SSE(X_1, X_2, X_3)/18}$$

$$= \frac{[1029.9 - 849.9]/2}{575.8/18} = 2.8135$$

(四) $r_{Y1.2} = \dfrac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1-r_{Y2}^2}\sqrt{1-r_{12}^2}} = \dfrac{(-0.61)-(0.019)(-0.082)}{\sqrt{1-(0.019)^2}\sqrt{1-(-0.082)^2}} = -0.6106$

$$r_{Y2.13} = \frac{r_{Y2.3} - r_{Y1.3}r_{12.3}}{\sqrt{1-r_{Y1.3}^2}\sqrt{1-r_{12.3}^2}} = \frac{(-0.488)-(-0.2014)(0.3546)}{\sqrt{1-(0.2014)^2}\sqrt{1-(0.3546)^2}} = -0.4549$$

其中 $r_{Y2.3} = \dfrac{r_{Y2} - r_{Y3}r_{23}}{\sqrt{1-r_{Y3}^2}\sqrt{1-r_{23}^2}} = \dfrac{(0.019)-(0.728)(0.439)}{\sqrt{1-(0.728)^2}\sqrt{1-(0.439)^2}} = -0.488$

$$r_{Y1.3} = \frac{r_{Y1} - r_{Y3}r_{13}}{\sqrt{1-r_{Y3}^2}\sqrt{1-r_{13}^2}} = \frac{(-0.61)-(0.728)(-0.703)}{\sqrt{1-(0.728)^2}\sqrt{1-(-0.703)^2}} = -0.2014$$

五、若已知n=23，且算出$(X'X)^{-1}=\begin{bmatrix} 18.93 & 1.28 & 1.70 & 2.78 \\ 1.28 & 0.33 & 0.47 & 0.34 \\ 1.70 & 0.47 & 1.02 & 0.76 \\ 2.78 & 0.34 & 0.76 & 0.74 \end{bmatrix}$，$b=\begin{bmatrix} 6.73 \\ -1.92 \\ 5.94 \\ 3.91 \end{bmatrix}$，與 MSE=3.64之線性複迴歸問題，則：

(一)請以bonferroni法計算β_1，β_2與β_3之94%聯合信賴區間（simultaneous confidence intervals）。（12分）

(二)請計算$2\beta_1+\beta_2+\beta_3$的96%信賴區間。（8分）

答：

假設模型：$Y_i = \beta_{0id} + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_1 X_{3i} + \varepsilon_i, i = 1,2,\cdots,23$

$$\varepsilon_i \sim N(0,\sigma^2)$$

$$S^2(b) = (X'X)^{-1} MSE$$

(一) β_1 之聯合信賴區間爲：

$$(\hat{\beta}_1 \pm t_{(19)\frac{0.06}{2\binom{3}{1}}} S(\hat{\beta}_1)) = (-1.92 \pm 2.539\sqrt{(0.33)(3.64)}) = (-4.7027, 0.8627)$$

β_2 之聯合信賴區間爲：

$$(\hat{\beta}_2 \pm t_{(19)\frac{0.06}{2\binom{3}{1}}} S(\hat{\beta}_2)) = (5.94 \pm 2.539\sqrt{(1.02)(3.64)}) = (1.0477, 10.8323)$$

β_3 之聯合信賴區間爲：

$$(\hat{\beta}_3 \pm t_{(19)\frac{0.06}{2\binom{3}{1}}} S(\hat{\beta}_3)) = (3.91 \pm 2.539\sqrt{(0.74)(3.64)}) = (-0.2571, 8.0771)$$

(二) $2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 2(-1.92) + 5.94 + 3.91 = 6.01$

$$S^2(2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) = 4S^2(\hat{\beta}_1) + S^2(\hat{\beta}_2) + S^2(\hat{\beta}_3)$$

$$+2(2)Cov(\hat{\beta}_1,\hat{\beta}_2)+2(2)Cov(\hat{\beta}_1,\hat{\beta}_3)+2Cov(\hat{\beta}_2,\hat{\beta}_3)$$

$$=[4(0.33)+1.02+0.74+4(0.47)+4(0.34)+2(0.76)](3.64)=28.5376$$

$2\beta_1+\beta_2+\beta_3$ 之96%信賴區間為：

$$((2\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3)\pm t_{(19)\frac{0.04}{2}}S(2\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3))=(6.01\pm 2.205\sqrt{28.5376})$$

$$=(-5.7692,17.7892)$$

六、去年一年中，某市區裝有冷氣機的四層樓公寓頂樓住戶，平均月付電費（Y：單位元）及每月平均溫度，但每月平均溫度分為低（10.0-15.0℃）、中（15.1-25.0℃）和高（25.1℃以上）三級，得到紀錄如下表：

月份	1	2	3	4	5	6
月平均溫度	低	低	中	中	中	高
月電費：Y	1088	1107	1178	1288	1455	1981

月份	7	8	9	10	11	12
月平均溫度	高	高	高	中	中	低
月電費：Y	2370	2659	2577	1937	1553	1170

研究人員設法探究月平均溫度與月電費Y之間的關係，考慮利用虛擬變數：

$$\mathrm{D}_{i1}=\begin{cases}1, \text{若月平均溫度=高，}\\ 0, \text{otherwise，}\end{cases}\qquad \mathrm{D}_{i2}=\begin{cases}1, \text{若月平均溫度=中，}\\ 0, \text{otherwise，}\end{cases}$$

建構並適配複迴歸模式：

$$Y_i=\beta_0+\beta_1 D_{i1}+\beta_2 D_{i2}+\varepsilon_i \qquad \forall_i=1, 2, ..., 12$$

但在編寫電腦程式時，誤使用以月平均溫度為因子之單因子變異數分析模式，得到下列之變異數分析表：

Source	df	SS
處理	2	3,182,320.7
誤差	9	621,360.2
總和	11	3,803,680.9

(一)請計算檢定H_0：$\beta_1=\beta_2=0$ vs. H_a：β_1與β_2不全為0的檢定統計值。（6分）

(二)請計算β_0，β_1與β_2之估計值。（9分）

答：

假設模型：$Y_i=\beta_0+\beta_1 D_{i1}+\beta_2 D_{i2}+\varepsilon_i, i=1,2,\cdots,12$

$$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

(一)T.S.: $F=\dfrac{MSR}{MSE}=\dfrac{3182320.7/2}{621360.2/9}=23.0469$

(二) $E(Y\mid D_1=0,D_2=0)=\beta_0$

$E(Y\mid D_1=1,D_2=0)=\beta_0+\beta_1$

$\Rightarrow \beta_1=E(Y\mid D_1=1,D_2=0)-E(Y\mid D_1=0,D_2=0)$

$E(Y\mid D_1=0,D_2=1)=\beta_0+\beta_2$

$$\Rightarrow \beta_2 = E(Y \mid D_1 = 0, D_2 = 1) - E(Y \mid D_1 = 0, D_2 = 0)$$

故 $\hat{\beta}_0 = \overline{Y}_{D1=0,D2=0} = 1121.6667$

$$\hat{\beta}_1 = \overline{Y}_{D1=1,D2=0} - \overline{Y}_{D1=0,D2=0} = 2396.75 - 1121.6667 = 1275.0833$$

$$\hat{\beta}_2 = \overline{Y}_{D1=0,D2=1} - \overline{Y}_{D1=0,D2=0} = 1482.2 - 1121.6667 = 360.5333$$