代號：31470
頁次：4－1

108年特種考試地方政府公務人員考試試題

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：㈠可以使用電子計算器。
㈡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。
㈢本科目除專門名詞或數理公式外，應使用本國文字作答。

注意：若無特別標示，本試卷採用顯著水準為 0.05 及 95%信心水準為原則。

附表 A：t 分布 $\alpha=0.025$ 與 $\alpha=0.05$ 右尾臨界值, df 為自由度

df	1	2	3	4	5	6	7	8	9	10
$t_{df,0.025}$	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228
df	11	12	13	14	15	16	17	18	19	20
$t_{df,0.025}$	2.201	2.179	2.160	2.145	2.131	2.120	2.110	2.101	2.093	2.086
df	21	22	23	24	25	26	27	28	29	30
$t_{df,0.025}$	2.080	2.074	2.069	2.064	2.060	2.056	2.052	2.048	2.045	2.042
df	1	2	3	4	5	6	7	8	9	10
$t_{df,0.05}$	6.314	2.920	2.353	2.132	2.015	1.943	1.895	1.860	1.833	1.812
df	11	12	13	14	15	16	17	18	19	20
$t_{df,0.05}$	1.796	1.782	1.771	1.761	1.753	1.746	1.740	1.734	1.729	1.725
df	21	22	23	24	25	26	27	28	29	30
$t_{df,0.05}$	1.721	1.717	1.714	1.711	1.708	1.706	1.703	1.701	1.699	1.697

附表 B：F 分布 $\alpha=0.05$ 右尾臨界值, df1 為分子自由度, df2 為分母自由度

$F_{df1,df2,0.05}$

df1\df2	1	2	3	4	5	6
1	161.45	18.51	10.13	7.71	6.61	5.99
2	199.50	19.00	9.55	6.94	5.79	5.14
3	215.71	19.16	9.28	6.59	5.41	4.76
4	224.58	19.25	9.12	6.39	5.19	4.53
5	230.16	19.30	9.01	6.26	5.05	4.39
6	233.99	19.33	8.94	6.16	4.95	4.28

一、考慮一簡單線性迴歸模型$Y_i = \alpha + \beta X_i + \varepsilon_i$, i=1,…,n, 其中Y_i為因變數，X_i為自變數，ε_i為誤差項且與X_i獨立。另外，也假設ε_i（i=1,…,n）具有獨立且相同的常態分布$N(0,\sigma^2)$，其中σ^2表變異數。（每小題 5 分，共 20 分）

㈠請導出參數α,β的最小平方估計式$\hat{\alpha},\hat{\beta}$，並證明其不偏性（unbiasedness）。

㈡如果其他假設不變，但$Var(\varepsilon_i) = \sigma^2 X_i^2$, i=1,…,n。說明由㈠導出之$\hat{\beta}$是否仍具有不偏性？在此情形下，是否可提供較佳的估計式（以式子說明概念或作法，無需列出詳細結果）？

㈢如果其他假設不變，但$Cov(\varepsilon_i,\varepsilon_{i+1}) = \rho\sigma^2$, i=1,…,n-1。說明由㈠導出之$\hat{\beta}$是否仍具有不偏性？試舉例說明何種類型的數據會較容易發現$\rho \neq 0$的情形。如何檢定$\rho = 0$（以式子說明概念或作法，無需列出詳細結果）？

㈣假設自變數X_i無法直接被觀察到，而是觀察到一個替代變數$W_i, i = 1,..., n$, $W_i = X_i + \delta_i$, δ_i為白噪音（white noise）與其他變數均獨立，且δ_i（i=1,…,n）具有獨立且相同的常態分布$N(0,1)$。此時若將W_i取代最小平方估計式$\hat{\beta}$中的X_i，並令所得之新估計式為$\hat{\beta}_w$。說明此$\hat{\beta}_w$是否仍具有不偏性？當 n 很大時，$\hat{\beta}_w$的漸近偏差為何？在此情形下是否可提供較佳的估計式（以式子說明概念或作法，無需列出詳細結果）？

二、在一調查薪資結構的研究中，吾人欲了解薪資（Y）與以下兩變數（X1, X2）的關係，其中 X1 表性別（女性為 F，男性為 M），X2 表區域別（分為 A, B, C 三個區域），收集資料如下表：

Y	6	4	3	4	4	2
X1	F	F	F	M	M	M
X2	A	A	B	B	C	C

一般來說，統計軟體的語法建立 Y 與兩變數的迴歸模型分析時，模式部分可寫為 Y~X1+X2（R 軟體）或 Y= X1 X2（SAS 軟體），或是直接點選 X1, X2 為自變數進行迴歸分析。請依據此精神與上述之資料，

㈠定義一個設計矩陣（design matirx），並說明此設計矩陣各個欄（column）的意義。寫下線性迴歸模型，以矩陣形式列出正規方程式（normal equation），解正規方程式求出參數估計值，列出三區域之兩兩比較薪資差異的估計值。（14 分）

㈡完成下面之 ANOVA 表。（8 分）

Analysis of Variance Table： Response： Y

變異來源	自由度（d.f.）	平方和（SS）	均方和（MS）	F 值 F value
迴歸				
殘差				
總和		8.833		

㈢計算性別薪資差異（男性對女性）的 95%信賴區間，估計一個男性在區域 A 的平均薪資及其 95%信賴區間。最後，根據 ANOVA 表格中 F 值說明其代表之意義。（10 分）

三、在一個關於放射線對腫瘤及壽命的影響研究中，研究人員利用老鼠設計了一項為期兩年的實驗。此實驗設計 30 隻老鼠每週照射不同劑量的放射線（劑量範圍為 1～10），並記錄其壽命（單位：週）。數據形式如下表：

X（劑量）	1	1	1	2	2	2	3	3	…	…	8	8	9	9	9	10	10	10
Y（壽命）	104	104	104	104	104	98	104	94	…	…	53	56	44	36	56	37	26	46

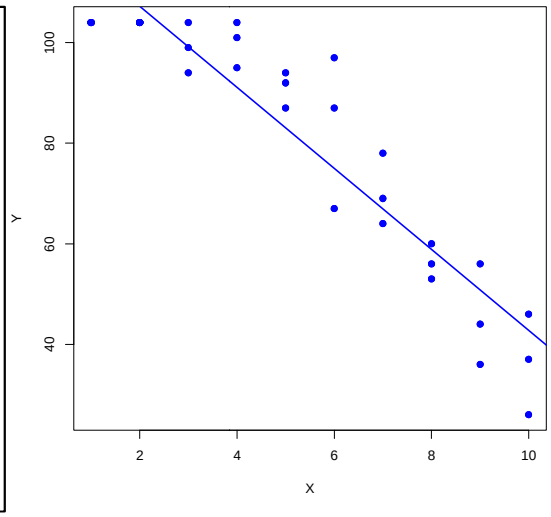
根據資料，研究人員完成一迴歸分析及配適圖如下：

```
Call:
lm(formula = Y ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-16.745  -5.830  -1.500   5.113  22.028

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 123.3111     3.6872   33.44  < 2e-16 ***
X            -8.0566     0.5942  -13.56 7.95e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.349 on 28 degrees of freedom
Multiple R-squared:  0.8678,
```



㈠根據分析結果，求 X 與 Y 之相關係數，完成下面之變異數分析表（ANOVA table）並說明此模型是否恰當？另，預測當 X=15 時之壽命，說明是否認同此預測值？（15 分）

變異來源	自由度（d.f.）	平方和（SS）	均方和（MS）	F 值 F value
迴歸				
殘差				---
總和			---	---

㈡由於實驗時間的限制，事實上有 8 隻老鼠壽命記錄在 104 週時還是活著的狀態。試問若預算足夠而得以完整觀察所有老鼠的壽命時（如實驗時間 3 年），則迴歸分析的參數估計會如何變動（可配合圖形說明），亦即實驗數據因經費限制而對於真實之「壽命與輻射劑量關係」的分析結果可能產生怎樣的影響？（5 分）

四、一組資料內含 Y 及 X_1～X_5 等變數，資料有 31 筆觀察值。為了進行變數選取，考慮 Y 對 X_1～X_5 之一階（first order）所有可能迴歸模式。經由分析整理得到下表：

no. of variables	X1	X2	X3	X4	X5	**adjr2**	**Cp**	no. of variables	X1	X2	X3	X4	X5	**adjr2**	**Cp**
1	0	0	0	0	1	**0.142**	**14.5**	3	1	0	1	0	1	**0.371**	**5.4**
1	0	0	1	0	0	**0.142**	**14.5**	3	1	0	0	1	1	**0.361**	**5.8**
1	0	0	0	1	0	**0.14**	**14.6**	3	0	1	1	0	1	**0.294**	**8.8**
1	1	0	0	0	0	**0.014**	**20.8**	3	0	0	1	1	1	**0.277**	**9.6**
1	0	1	0	0	0	**0.008**	**21**	3	0	1	0	1	1	**0.263**	**10.2**
2	0	0	1	0	1	**0.288**	**8.3**	3	1	1	0	0	1	**0.21**	**12.6**
2	0	0	0	1	1	**0.286**	**8.4**	3	1	1	1	0	0	**0.178**	**14**
2	1	0	0	0	1	**0.189**	**12.9**	3	1	0	1	1	0	**0.169**	**14.5**
2	1	0	1	0	0	**0.185**	**13.1**	3	1	1	0	1	0	**0.156**	**15.1**
2	1	0	0	1	0	**0.176**	**13.5**	3	0	1	1	1	0	**0.128**	**16.3**
2	0	1	0	0	1	**0.163**	**14.1**	4	1	1	1	0	1	**0.377**	**6.1**
2	0	1	1	0	0	**0.137**	**15.3**	4	1	0	1	1	1	**0.361**	**6.7**
2	0	0	1	1	0	**0.126**	**15.8**	4	1	1	0	1	1	**0.343**	**7.5**
2	0	1	0	1	0	**0.115**	**16.4**	4	0	1	1	1	1	**0.322**	**8.4**
2	1	1	0	0	0	**0.021**	**20.7**	4	1	1	1	1	0	**0.164**	**15.3**
3	1	0	1	0	1	**0.371**	**5.4**	5	1	1	1	1	1	**0.401**	**6**

㈠以 adjusted R^2 為準則，排序選取最佳三個模式。（6 分）

㈡以 Mallow's Cp 為準則，排序選取最佳三個模式。（6 分）

㈢採用 F 檢定法，說明向後消去法（Backward elimination, stay level=0.05）準則的選模過程，並列出所選取之模式。（10 分）

㈣除變數選擇外，針對模型 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\varepsilon$ 分析得到另一表。請以第一列的值解釋 dfb.X2(−0.154)及 dffit(−0.371)的用途及其大概的原理。（6 分）

Obs.	dfb.X1	dfb.X2	dfb.X3	dfb.X4	dfb.X5	dffit	cov.r	cook.d	hat
1	-0.101	-0.154	-0.23	0.201	-0.132	-0.371	2.008	0.024	0.396
2	0.1	0.083	0.072	-0.081	-0.044	0.177	1.608	0.005	0.226
3	-1.145	2.676	2.773	-2.481	1.735	4.332	0.001	0.902	0.23
…									
30	0.019	0.053	0.049	-0.046	-0.016	-0.079	1.636	0.001	0.223
31	0.063	0.07	0.037	-0.048	-0.048	-0.184	1.498	0.006	0.179