

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)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、若以樣本 y 對 x 做線性迴歸，可得到迴歸估計式 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 。假設 x 、 y 之樣本平均及標準差分別為 \bar{x} 、 \bar{y} 、 s_x 、 s_y ，樣本相關係數為 r 。今先將 x 、 y 標準化，即：

$$x_i^* = \frac{x_i - \bar{x}}{s_x}, \quad y_i^* = \frac{y_i - \bar{y}}{s_y}$$

然後以 y^* 對 x^* 做線性迴歸，得到 $\hat{y}_i^* = \tilde{\beta}_0 + \tilde{\beta}_1 x_i^*$ 。試求：

(一) $\tilde{\beta}_0 = ?$ (10 分)

(二) $\tilde{\beta}_1$ 和 $\hat{\beta}_1$ 的關係。(10 分)

(三) r 和 $\tilde{\beta}_1$ 之關係。(5 分)

二、 X 為前測成績， Y 為後測成績。假設甲乙兩班的前、後測成績關係分別為

$$\text{甲：} Y = \beta_{01} + \beta_{11}X + \varepsilon$$

$$\text{乙：} Y = \beta_{02} + \beta_{12}X + \varepsilon$$

下表資料中 G 代表班別 ($G=1$ 為甲班， $G=0$ 為乙班)，令 XG 為 X 和 G 乘積。

Y	X	G	Y	X	G
5.3	4	1	3	3	0
10.4	9	1	15	8	0
9.2	8	1	9.4	5	0
10.1	9	1	13.1	6	0
7.3	6	1	9.1	3	0
4.3	3	1	17.7	11	0
9.7	9	1	7.3	7	0
6.3	6	1	10.2	10	0
6.6	5	1	19.4	12	0
9	9	1	13.6	9	0

我們以上表資料分別配適以下四組迴歸：M1：Y 對 X 迴歸；M2：Y 對 G 迴歸；M3：Y 對 X 和 G 複迴歸；M4：Y 對 X、G 和 XG 複迴歸。變異數分析結果如下：

M1: $\hat{Y} = 1.26 + 1.203 X$					M2: $\hat{Y} = 11.78 - 3.94 G$				
Source	DF	Adj SS	F-Value	P-Val	Source	DF	Adj SS	F-Value	P-Val
Regression	1	202.24	26.48	0.00	Regression	1	77.72	5.34	0.033
X	1	202.24	26.48	0.00	G	1	77.72	5.34	0.033
Error	18	137.49			Error	18	262.01	14.56	
Total	19	339.73							

M3: $\hat{Y} = 3.39 + 1.133 X - 3.26 G$					M4: $\hat{Y} = 2.52 + 1.251 X - 0.86 G - 0.343 XG$				
Source	DF	Adj SS	F-Value	P-Val	Source	DF	Adj SS	F-Value	P-Val
Regression	2	254.79	25.50	0.000	Regression	3	258.451	16.96	0.00
X	1	177.07	35.44	0.000	X	1	141.449	27.84	0.00
G	1	52.55	10.52	0.005	G	1	0.419	0.08	0.778
Error	17	84.94	4.997		XG	1	3.658	0.72	0.409
					Error	16	81.284	5.080	

在顯著水準 0.05 下，試求：

(一) 檢定「兩班的 Y 對 X 關係是否平行（斜率相同）」，即

$$H_0: \beta_{11} = \beta_{12} \text{ and } \beta_{01} \neq \beta_{02} \text{ vs. } H_1: \beta_{11} \neq \beta_{12} \text{ and } \beta_{01} \neq \beta_{02} \text{。 (10 分)}$$

(二) 檢定「兩班是否有相同之 Y 對 X 線性關係（相同的斜率及截距）」，即

$$H_0: \beta_{11} = \beta_{12} \text{ and } \beta_{01} = \beta_{02} \text{ vs. } H_1: H_0 \text{ 為非。 (15 分)}$$

F-分布臨界值表

$$P(F > F_{\alpha, v_1, v_2}) = \alpha = 0.05$$

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479

三、某資料有 40 個觀察值，因變數為 y_1, \dots, y_{40} ，自變數為 x_1, \dots, x_{40} ，迴歸模式 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 。

(一)其檢定之有效性是建立在對 ε_i 的那些假設下？(10 分)

(二)若 (x_1, \dots, x_{20}) 為男生體重， (x_{21}, \dots, x_{40}) 為女生體重， y 為其運動後心跳頻率。已知男生體重的變異量一般較女生大。今以 y 對 x 做簡單線性迴歸，可能會違反(一)中那些假設？(5 分)

(三)若 x_1, \dots, x_{10} 是 10 個人第 1 年之測量值， x_{11}, \dots, x_{20} 為其第 2 年測量值， x_{21}, \dots, x_{30} 為其第 3 年測量值， x_{31}, \dots, x_{40} 為其第 4 年測量值。以 y 對 x 做簡單線性迴歸的話，會違反(一)中那些假設？(5 分)

四、連續變數 Y 代表因變數藥效(越大代表成效越佳)，自變數 X 為類別變數，代表 A 、 B 、 C 三種藥物處方，三組人樣本數相同，各只接受其中一種處方。

(一)某軟體將 X 轉成以下虛擬變數(dummy variable) X_1 及 X_2 ：

$$X_1 = \begin{cases} 1 & \text{當 } X = A \\ 0 & \text{當 } X = B \\ 0 & \text{當 } X = C \end{cases}, \text{ 及 } X_2 = \begin{cases} 0 & \text{當 } X = A \\ 1 & \text{當 } X = B \\ 0 & \text{當 } X = C \end{cases}$$

然後以 Y 對 X_1 及 X_2 配適迴歸模式： $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ 。請就以下檢定結果比較三種藥物之藥效(如：何者顯著較佳，何者間無顯著差別)。(15 分)

Term	Coef	SE Coef	t-Value	P-Value
Constant	8.200	0.732	11.20	0.000
X_1	-2.10	2.10	-1	0.32
X_2	4.50	1.03	4.35	0.000

(二)另一種軟體轉虛擬變數的方式如下：

$$X_1^* = \begin{cases} 1 & \text{當 } X = A \\ 0 & \text{當 } X = B \\ -1 & \text{當 } X = C \end{cases}, \text{ 及 } X_2^* = \begin{cases} 0 & \text{當 } X = A \\ 1 & \text{當 } X = B \\ -1 & \text{當 } X = C \end{cases}$$

然後以 Y 對 X_1^* 及 X_2^* 配適迴歸模式： $Y = \beta_1 X_1^* + \beta_2 X_2^* + \varepsilon$ 。請就以下檢定結果比較三種藥物之藥效。(15分)

Term	Coef	SE Coef	t-Value	P-Value
Constant	8.200	0.732	11.20	0.000
X_1^*	-2.64	1.01	-2.61	0.009
X_2^*	3.50	1.01	3.46	0.000