

※ 考生請注意：本試題不可使用計算機。請於答案卷作答，於本試題紙上作答者，不予計分。

一、選擇題(每題 4 分，共 10 題，共 40 分) 請選出最佳答案

1. 下列何者主要不是用來衡量資料的離散(dispersion)程度？
  - (A) 標準差
  - (B) 全距
  - (C) 四分位數距
  - (D) 中位數
  
2. 抽樣 36 個人並記錄其體重(公斤)，以下為其計算出來的統計量：平均數 = 60；全距 = 30；眾數 = 54；變異數 = 36。則該資料的變異係數 (coefficient of variation) 為
  - (A) 0.1
  - (B) 0.5
  - (C) 5
  - (D) 9
  
3. 從一個龐大的母體中隨機抽取 400 個觀測值，其樣本平均數及樣本標準差分別為 320 和 120，則平均數的標準誤為
  - (A) 120
  - (B) 16
  - (C) 6
  - (D) 0.333
  
4. 某實驗事件 G 出現的機率是 1/4，若該實驗進行了 3 次，事件 G 並未發生。則在第 4 次實驗時，事件 G
  - (A) 一定發生
  - (B) 有 3/4 的機率會發生
  - (C) 可能發生
  - (D) 不可能發生
  
5. 氣象播報員宣稱台南市七月份的平均溫度等於或低於 32 度。抽取 32 個七月份溫度的樣本，則正確的假設檢定應表示為
  - (A)  $H_0: \mu \geq 32$      $H_a: \mu < 32$
  - (B)  $H_0: \mu \leq 32$      $H_a: \mu > 32$
  - (C)  $H_0: \mu \neq 32$      $H_a: \mu = 32$
  - (D)  $H_0: \mu < 32$      $H_a: \mu > 32$

6.  $X$  為一隨機變數，其機率函數為  $f(X) = X/9$ ， $X=2, 3$  或  $4$ ，則  $X$  的期望值為
- (A) 3.111  
(B) 3.222  
(C) 3.333  
(D) 3.666
7. 某大學生由她家到學校的時間是 40 到 60 分鐘的均勻分配，則她會剛好在 50 分鐘到校的機率是
- (A) 0.50  
(B) 0.20  
(C) 0.05  
(D) 0
8. 若  $A$  及  $B$  為獨立事件，其中  $P(A) = 0.1$  及  $P(B) = 0.6$ ，則  $P(A|B) =$
- (A) 0.06  
(B) 0.1  
(C) 0.1667  
(D) 0.5
9. 在 400 個選舉人樣本中，320 個表示她們支持現任市長。則不支持現任市長比例的 95% 信賴區間是
- (A) 0.161 到 0.239  
(B) 0.122 到 0.278  
(C) 0.761 到 0.839  
(D) 0.822 到 0.878
10. 某超商經理想要估計顧客使用該超商聯名信用卡與使用某著名信用卡的平均信用消費金額之差異。有關的資訊如下：

	超商聯名信用卡	著名信用卡
樣本數	60	70
樣本平均數	\$110	\$95
母體變異數	\$120	\$140

則兩信用卡顧客平均消費金額差異的 95% 信賴區間估計是

- (A) 5.20 至 24.80  
(B) 7.16 至 22.84  
(C) 11.08 至 18.92  
(D) 13.04 至 16.96

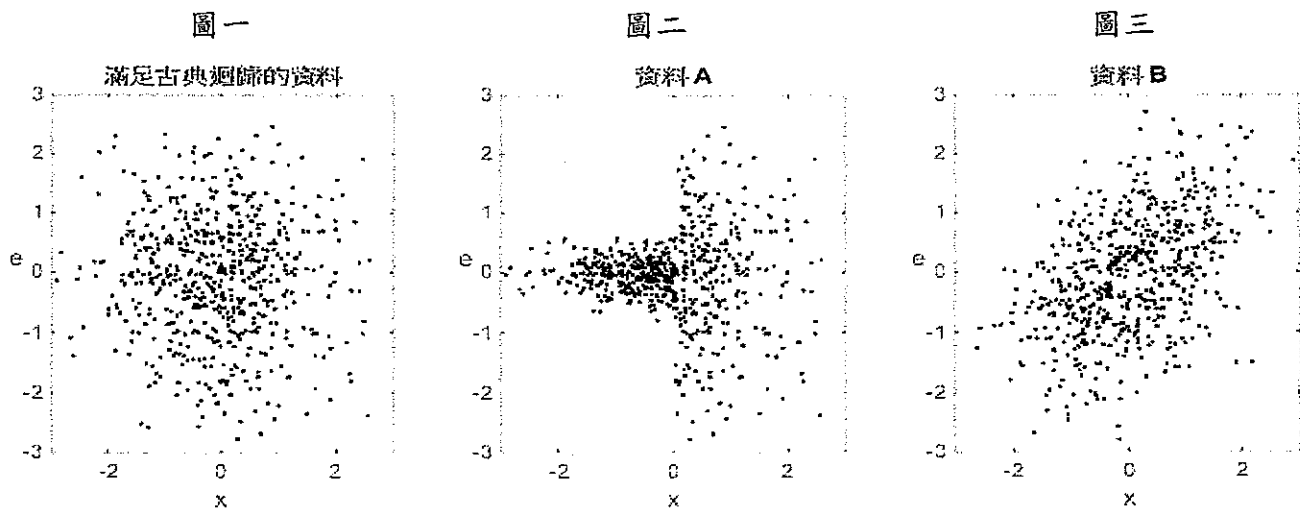
## 二、問答題(兩大題，共 60 分) 請詳加作答

1. 考慮以下簡單線性迴歸模型，

$$y = \alpha + \beta x + e,$$

其中  $y$  為被解釋變數， $x$  為解釋變數， $e$  為誤差項，而  $\alpha$  與  $\beta$  分別為截距項與斜率。目前有兩組資料 A 與 B 等待分析。

(1)[10 分] 若誤差項  $e$  能被觀察到，且和解釋變數  $x$  呈現以下散佈的形式。請問，資料 A (圖二) 主要違反古典線性迴歸模型中的那一(些)假設？資料 B (圖三) 又違反那一(些)假設？(提示：為讓考生更容易回答，我們把滿足古典線性迴歸模型的散佈形式舉例於圖一)



(2) 在真實世界中，誤差項無法觀測到，而只能觀察到  $y$  與  $x$ 。請問

(a)[8 分] 應如何分別檢定資料 A 與資料 B 是否違反這一(些)假設？

(b)[6 分] 若針對資料 A 以最小平方法來估計迴歸模型中的參數，會得 BLUE 的特性嗎？若不能，主要是哪個特性不符，可如何修正？

(c)[6 分] 若針對資料 B 以最小平方法來估計迴歸模型中的參數，會有什麼問題？可如何解決這些問題？

2. 小成與小宮一同申請了台南市的房價資料，共有 500 個觀察值。兩人的資料完全一樣，並想分析高鐵以東與其他地區的房價差異。

(1)[10 分] 小成考慮以下迴歸模型：

$$y = \alpha + \beta E + e,$$

其中  $y$  為單位價格(十萬元每平方公尺)，解釋變數  $E=1$  表示該筆房屋座落在高鐵以東，而  $E=0$  表其它； $\alpha$  與  $\beta$  為要估計的兩係數。利用最小平方法得到的估計結果： $\hat{\alpha} = 7.2$  標準誤為 0.24,

$\hat{\beta} = -2.2$  標準誤為 0.11，判定係數  $R^2 = 0.32$ 。請問高鐵以東的平均房價為何？其它地區的平均房價為何？高鐵以東與其它地區的平均房價差價為何？差距顯著嗎？

(2) [5 分] 承前一題，若小宮考慮以下迴歸模型：

$$y = \gamma + \delta W + e,$$

其中  $W=0$  表示該筆房屋座落在高鐵以東，而  $W=1$  表其它； $\gamma$  與  $\delta$  為要估計的兩係數。請問最小平方方法得到的係數估計值為多少，所得到的判定係數又是多少？(請說明如何得到這些結果，否則不予計分)

(3) [5 分] 承前兩題，他們的朋友鸞鸞也拿到相同資料，並考慮以下迴歸模型：

$$y = \eta W + \theta E + e,$$

其中  $W$  與  $E$  的定義如前兩題； $\eta$  與  $\theta$  則為要估計的兩係數。請說明上述迴歸有沒有線性重合的問題嗎？若有，應如何解決？若沒有，鸞鸞迴歸中的係數估計結果為何？(不說明不予計分)

(4) [10 分] 鸞鸞仔細地由資料中發現房價在高鐵以東與其它地區的變異數明顯不同。請問此時用(1)小題中小成的估計結果來檢定「兩區域均數相同」是否不妥？為什麼？鸞鸞應如何充分使用這資訊於「兩區域均數相同」的檢定中？(請詳列所提出方法的操作步驟)