

是非題（每題 2 分）請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

1. 若要比較臺灣大學甲、乙兩學院教師的性別分布，可以使用兩個盒鬚圖(boxplots)來比較。
2. 在盒鬚圖(boxplots)中，中位數(median)等於第一個四分位(first quartile)及第三個四分位(third quartile)的平均值。
3. 已知臺灣大學甲、乙兩學院教師血壓分布同屬常態分配，期望值相同，如果想要檢定兩個變異數是否相等，可以使用兩個獨立樣本的 t 檢定(independent two-sample t test)。
4. 卡方檢定與 Fisher' s exact 檢定都是無母數統計檢定。
5. 假設 Y 代表吳先生選學生會長的得票數，Y 服從二項式分配(Binomial distribution)，其中總投票人數為 N，成功機率為 p。如果總投票人數為 15000 人，可以使用常態分布來估計吳先生得票數超過一萬人的機率。
6. 如果鉛暴露與否跟罹患貧血的勝算比值(odds ratio, OR)比 1 大，可以推測若進行卡方檢定會達統計顯著。
7. 為探討男女性別的 B 型肝炎帶原率是否相同，可以隨機抽取各 100 位男女，進行 B 型肝炎檢驗，再利用勝算比值(odds ratio, OR)來檢定。
8. 使用兩個獨立樣本的 t 檢定(independent two-sample t test)所得的 p 值，跟使用變異數分析(Analysis of variance, ANOVA)檢定所得的 p 值在理論上會完全一樣。
9. 如果陳先生以同樣的六組號碼下注愛心彩卷連續七期都沒中獎，則第八期再以同樣的六組號碼下注，得獎機率會比較高。
10. 二項式分配(Binomial distribution)是伯努力分布(Bernoulli distribution)的特例。

※ 注意：請於試卷內之「選擇題作答區」依序作答。

單選題（每題 3 分）

1. 某研究探討工作人員在超市中將貨品上架所需的時間，收集了 15 筆資料，每筆資料包含了完成上架的架子數(X)，以及花費的時間(以分鐘計，Y)。這研究完成了兩個迴歸模式分析如下表：

模式一：

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -0.0938 + 0.4071x$$

$$se(\hat{\beta}_0) = 0.144 \quad se(\hat{\beta}_1) = 0.008 \quad R^2 = 0.9947$$

模式二：

$$\hat{y} = \hat{\beta}_1 x = 0.4026x$$

$$se(\hat{\beta}_1) = 0.004 \quad R^2 = 0.9983$$

請問在兩個迴歸模式當中，若分別檢定係數  $\beta_1$  是否為 0，結果為？

- (1) 在模式一中顯著，在模式二中顯著。
- (2) 在模式一中顯著，在模式二中不顯著。
- (3) 在模式一中不顯著，在模式二中顯著。
- (4) 在模式一中不顯著，在模式二中不顯著。

見背面

2. (延續上題)在上一題的檢定中你認為以下敘述何者最恰當?
- (1)在模式一中的 t 檢定，自由度為 14。
  - (2)如果要進行模式選擇(model selection)，第二個模式比較好，因為其  $R^2$  的值比較大。
  - (3)兩個模式下的係數估計值可以利用最小平方方法(least square estimate, LSE)或最大似似估計值(maximum likelihood estimate, MLE)求得。兩者答案會完全相同。
  - (4)由以上表格無法判斷架子數  $X$  跟時間  $Y$  成正相關或是負相關。
3. (延續上題)利用以上表格比較兩個模式，你認為以下何者最不恰當?
- (1)如果估計完成 100 個架子的上貨所需的時間，則模式一所得的估計值會比模式二所得的估計值略低。
  - (2)有截距項的模式一才是簡單線性迴歸模式；模式二是迴歸模式，但不是簡單線性迴歸模式。
  - (3)表格中模式一跟模式二對時間  $Y$  的期望值假設不同。
  - (4)模式一跟模式二都必須假設時間  $Y$  為常態分布。
4. 某研究探討長期食用地溝油對健康的影響，利用了 25 隻老鼠每日餵食混合了 4 cc 地溝油的食物，一年之後有 15 隻老鼠出現動作協調能力退步的現象。另外 25 隻老鼠每日餵食一般食品，一年之後有 5 隻老鼠出現動作協調能力退步的現象。請問以下何者最不恰當?
- (1)如果檢定食用地溝油一年是否影響動作協調能力，可使用卡方檢定。
  - (2)如果檢定食用地溝油一年是否影響動作協調能力，可檢定勝算比值(odds ratio, OR) 是否為 0。
  - (3)如果檢定食用地溝油一年是否影響動作協調能力，可使用兩個比例的  $Z$  檢定 (two-proportion  $Z$  test)。
  - (4)以上三種檢定都必須假設這 50 隻老鼠的動作協調能力表現互不影響，彼此獨立。
5. (延續上題)針對上一小題檢定的結果，在顯著水準 0.05 的條件下，請問以下何者最不恰當?
- (1)餵食地溝油一年的老鼠比上沒有餵食地溝油的老鼠，其動作協調能力達統計上顯著不同。
  - (2)餵食地溝油一年的老鼠比上沒有餵食地溝油的老鼠，其動作協調能力顯著變差。
  - (3)不同餵食習慣的老鼠中，動作協調能力變差的比率的 95%信賴區間互不重疊。
  - (4)如果每日餵食只有混合了 2 cc 地溝油的食物，則老鼠動作協調能力變差的比率會比較小。

單選題 (每題 5 分)

6. 樣本數增加趨近無限時，下述何者錯誤?
- (A)樣本平均值的分佈，趨近常態分佈。
  - (B)樣本平均值的分佈，其變異數趨近零。
  - (C)樣本平均值趨近族群平均值。
  - (D)樣本標準差的 95%信賴區間，其寬度趨近零。
  - (E)樣本標準差趨近零，其分佈的變異數亦趨近零。
7. 樣本數很小時 (比如  $n < 5$ )，下述何者錯誤?
- (A)卡方檢定不適用。
  - (B)無母數 (non-parametric) 統計方法，仍可適用。
  - (C)以樣本中位數 (median) 估計族群中位數，具有穩健性 (robustness)。

- (D)以樣本平均值估計族群平均值，可能會有偏差 (bias)。
- (E)樣本平均值的 95%信賴區間，可能很寬。
8. 有關線性迴歸分析 (linear regression)，下述何者錯誤？
- (A)截距項 (intercept) 亦是其迴歸係數之一，不可強制令其為零。
- (B)自變項 (independent variable) 的迴歸係數，可闡釋為斜率 (slope)，然必要時可強制令其為任意數，比如 1。
- (C)將自變項大幅平移 (比如加 10000)，再進入分析，斜率不會改變。
- (D)將自變項取餘弦 (cosine)，再進入分析，仍為線性迴歸。
- (E)將依變項取餘弦，再進入分析，仍為線性迴歸。
9. 有關假說檢定，下述何者錯誤？
- (A) 對立假設為真時，假說檢定能夠拒絕虛無假設的機率，稱之為檢力 (power)。
- (B) 虛無假設為真時，假說檢定卻拒絕虛無假設的機率，稱之為型一誤差率 (type I error rate)。
- (C) 一個不偏的假說檢定，其檢力必不小於型一誤差率。
- (D) 在給定的顯著水準 (significance level) 下，樣本數增加趨近無限時，型一誤差率趨近零。
- (E) 在給定的顯著水準下，樣本數增加趨近無限時，檢力趨近 100%。
10. 有關布阿松分佈 (Poisson distribution)，下述何者錯誤？
- (A) 其期望值 (expected value) 及變異數 (variance) 相等。
- (B) 由於是右偏分佈，以樣本平均值估計其期望值，會有偏差。
- (C) 為離散型 (discrete) 分佈。
- (D) 兩個獨立的布阿松分佈相加，仍為布阿松分佈。
- (E) 兩個獨立的布阿松分佈相乘，其期望值為原兩分佈期望值之相乘。
11. 有關常態分佈 (normal distribution) 及卡方分佈 (chi-square distribution)，下述何者錯誤？
- (A) 常態分佈之平均值 (期望值)，中位數，眾數 (mode)，三者皆相等。
- (B) 兩個以上常態分佈的線性組合 (linear combination)，仍為常態分佈。
- (C) 常態分佈的平方，即為自由度為 1 的卡方分佈。
- (D) 兩個獨立的自由度為 1 的卡方分佈相加，即為自由度為 2 的卡方分佈。
- (E) 兩個獨立的自由度為 1 的卡方分佈相除，為 F 分佈。
12. 有關伯努立分佈 (Bernoulli distribution) 及二項式分佈 (binomial distribution)，下述何者錯誤？
- (A) 伯努立分佈是二元性 (binary) 分佈。
- (B) 伯努立分佈之期望值，為二元事件之一的機率。
- (C) 伯努立分佈之變異數，為二元事件各自機率的相乘。
- (D) 期望值為 0.5 的伯努立分佈，變異數最大。
- (E) 兩個獨立的伯努立分佈相加，為二項式分佈。

## 複選題（每題5分）

13. 考慮解釋變數 $X = (X_1, \dots, X_p)$ 及二元(binary)反應變數 $Y$ 。以下針對邏輯斯迴歸(Logistic Regression)的敘述，何者錯誤？
- (A) 對 $Y$ 的條件期望值建立線性模型
  - (B)  $X_j$ 必須為連續變項才可使用
  - (C) 使用最大概似估計(maximum likelihood estimate)以估計參數
  - (D) 雖然 $Y$ 不服從常態分配，但還是可以用常態分配計算檢定迴歸係數時的 p-value
  - (E)  $X$ 的迴歸係數可解釋為贏輸比(odds ratio)
14. 考慮解釋變數 $X = (X_1, \dots, X_p)$ 及連續反應變數 $Y$ 。以下針對線性迴歸(Linear Regression)的敘述，何者錯誤？
- (A) 對 $Y$ 的條件期望值建立線性模型
  - (B) 假設誤差項不隨 $X$ 的變化而變化
  - (C) 使用最大概似估計(maximum likelihood estimate, MLE)以估計參數
  - (D) 亦可使用最小平方估計(least square estimate, LSE)，會得到和 MLE 不一樣的估計值
  - (E) 用 $X_1$ 和 $(X_1 - 5)$ 會得到不一樣的迴歸係數估計值
15. 下列關於 ANOVA 的敘述，何者錯誤？
- (A) 用來檢定多組平均數是否全相同的統計方法
  - (B) 假設不同組內的變異數相同
  - (C) 推翻虛無假說後，吾人即可認定每一組的平均數都不一樣
  - (D) 亦可以用線性迴歸模型完成 ANOVA 檢定
  - (E) 檢定統計量可以表示為[組間變異和組內變異的比值]
16. 下列關於分配(distribution)的敘述，何者錯誤？
- (A) 獨立的常態(Normal)分配相加一定是常態分配
  - (B) 獨立的 Bernoulli 分配相加一定是 Binomial 分配
  - (C) t 分配的中位數(median)和眾數(mode)一樣
  - (D) F 分配的數值可正可負
  - (E) 布阿松(Poisson)分配的平均數越大則變異係數(coefficient of variation)越大

## 問答題(10分)

白醫生宣稱血液中的白血球個數( $X$ )可用來判斷個人是否患有某疾病 $D$ ( $D = 1$ 代表有疾病， $D = 0$ 代表沒有疾病)。他更進一步建議利用準則[若 $X < c$ ，則預測 $D = 1$ ]以預測疾病的狀態。你打算利用 Receiver Operating Characteristic (ROC)曲線分析來驗證白醫師的建議。

- (1) 根據以上的描述，定義並描述如何繪製 ROC 曲線。
- (2) 根據以上的描述，請寫出 ROC 曲線下面積的機率表達式。

試題隨卷繳回