

單選題 (每題 5 分, 共 65 分) ※ 注意：請於試卷內之「選擇題作答區」依序作答。

1. 有關標準差(standard deviation)、標準誤(standard error)以及變異係數(coefficient of variation), 下列何者正確?
  - (A) 標準誤越大, 表示資料的散佈程度越大。
  - (B) 變異係數越大, 表示資料的散佈程度越大。
  - (C) 樣本數增加, 標準差變小。
  - (D) 樣本數增加, 變異係數變小。
  - (E) 樣本數趨近無限大, 變異係數亦趨近無限大。
2. 有關假說檢定, 下列何者正確?
  - (A) 樣本數增加, 顯著水準(significance level)亦增加。
  - (B) 樣本數趨近無限大, 拒絕虛無假說(null hypothesis)的機率趨近 1。
  - (C) 樣本數趨近無限大, 檢力(power)亦趨近無限大
  - (D) 一個不偏的(unbiased)統計檢定, 拒絕虛無假說的機率等於拒絕對立假說(alternative hypothesis)的機率。
  - (E) 將顯著水準設越大, 檢力亦越大。
3. 有關樣本平均值的信賴區間(confidence interval), 下列何者正確?
  - (A) 雙尾信賴區間的寬度與樣本數的平方根成反比。
  - (B) 樣本數大於 30, 雙尾信賴區間的寬度趨近於 0。
  - (C) 90%雙尾信賴區間比 95%雙尾信賴區間較寬。
  - (D) 單尾信賴區間的寬度是雙尾信賴區間寬度之半。
  - (E) 雙尾信賴區間的寬度是單尾信賴區間寬度之半。
4. 有關伯努立分佈(Bernoulli distribution)、二項式分佈(binomial distribution)以及布阿松分佈(Poisson distribution), 下列何者正確?
  - (A) 二項式分佈為對稱分佈(symmetric distribution)。
  - (B) 布阿松分佈為右偏分佈(positively skewed distribution)
  - (C) 伯努立分佈沒有參數(parameter)。
  - (D) 布阿松分佈有兩個參數, 分別為該分佈之期望值及變異數。
  - (E) 二項式分佈有兩個參數, 分別為該分佈之期望值及變異數。
5. 有關無母數方法(non-parametric method), 下列何者正確?
  - (A) 樣本數極小時(比如  $n < 5$ ), 不可使用。
  - (B) 樣本數大於 30, 不可使用。
  - (C) 資料為右偏分佈, 不可使用。
  - (D) 計算無母數檢定方法的確切 p 值(exact p value), 不需依賴中央極限定理(central limit theorem)。
  - (E) 常用的統計套裝軟體, 並不支援計算無母數檢定方法的確切 p 值, 但可計算近似值。
6. 欲檢定台灣地區北部、中部、南部、東部居民(各抽樣 30 人)之膽固醇平均濃度有無不同, 可採用下述何者(膽固醇濃度為右偏分佈)?
  - (A) Student t test
  - (B) ANOVA (analysis of variance)
  - (C) McNemar test
  - (D) Wilcoxon rank-sum test
  - (E) Friedman test

見背面

7. 有關布阿松迴歸分析(Poisson regression analysis)，下列何者正確？
- (A) 死亡率及發生率的迴歸分析，一般皆使用布阿松迴歸。
  - (B) 應變項(dependent variable)為二元性變項(binary variable)。
  - (C) 應變項為連續性變項(continuous variable)且其值大於 0。
  - (D) 自變項(independent variable)為離散性變項(discrete variable)。
  - (E) 補償項(offset)是迴歸係數之一，有待估計。
8. 有關變異數分析(ANOVA)，下列何者正確？
- (A) 用來檢定不同群體的變異數是否有差異。
  - (B) 只能用在組數大於等於 3 的情況。
  - (C) 背後的理論基礎為 multiple linear regression。
  - (D) 在特定條件下跟 2-sample t-test (unequal variance)一樣。
  - (E) 使用的前提假設是整組資料服從常態分配。
9. 有關 p-value，下列何者正確？
- (A) 是一個固定的數值，並沒有隨機性涉入其計算中。
  - (B) 其可能的數值介於 0 和 0.1 之間
  - (C) 一般來說，在虛無假設成立下會服從 Uniform(0,1)分配。
  - (D) p-value<0.00000001 代表對立假說為真。
  - (E) p-value=0.05 代表實驗 100 次，平均會犯錯 5 次。
10. 有關線性迴歸(linear regression)，下列何者正確？
- (A) 增加解釋變數，則判定係數(R-square)一定會增加。
  - (B) 校正的判定係數(adjusted R-square)的意義是變異解釋的比例。
  - (C) 解釋變數的變異數越大，則分析結果的變異數越大。
  - (D) 解釋變數之間的相關係數越小，則分析結果的變異數越大。
  - (E) 迴歸係數等於 5，代表該解釋變數增加一單位，則觀測到的反應變數會增加 5
11. 有關羅吉斯迴歸(logistic regression)，下列何者正確？
- (A) 專門用來分析二元解釋變數的資料。
  - (B) 使用最小平方法(least squares estimate)估計迴歸係數。
  - (C) 可以用來建立線性的分類模型。
  - (D) 迴歸係數的估計值一定界於 0 和 1 之間。
  - (E) 迴歸係數等於 5，代表該解釋變數增加一單位，則事件發生的機率變成 5 倍。
12. 有關存活曲線(survival curve)，下列何者錯誤？
- (A) 存活曲線越高，代表事件越不容易發生。
  - (B) 存活曲線的斜率一定小於 0。
  - (C) 曲線的面積代表事件發生時間的期望值。
  - (D) 曲線的高度代表事件發生時間超過某數值的機率。
  - (E) 我們可以在存活曲線上找出事件發生時間的中位數。
13. 有關假說檢定(hypothesis testing)，下列何者錯誤？
- (A) 可以用信賴區間執行
  - (B) 一般來說，型一誤差越大，則型二誤差會越小。
  - (C) 在其他條件不變下，型一誤差決定後，則型二誤差也決定了。

- (D) 樣本數越大，則型一誤差越小
- (E) 樣本數越大，則型二誤差越小

是非題 (每題 2 分，共 10 分) ※ 注意：請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

14. \_\_\_\_\_ 比較三種降血壓藥物的藥效是否相同時，一定需要考量多重比較的問題。
15. \_\_\_\_\_ 檢定兩種降血壓藥物的藥效是否相同時，可以使用變異數分析的 F 檢定，也可以使用兩群獨立樣本的 t 檢定，兩種檢定方法會得到一樣的 p 值、自由度、與結論。
16. \_\_\_\_\_ 進行變異數分析時，必須假設各組的母體變異數都相同。
17. \_\_\_\_\_ 要檢定 20 位高血壓患者經過治療之後是否有效控制了血壓，可使用兩組獨立樣本的 t 檢定，檢定前後平均血壓值是否相等。
18. \_\_\_\_\_ 要檢定不同性別是否有不同的吸菸率時，可以使用卡方檢定，也可以使用兩個獨立樣本的二項式比例檢定，兩種方法會有相同的結果。

計算問答題 (25 分) ※ 注意：請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

19. 某研究想要探討住院天數(以天計)與保險給付(以千元計)之間的關係，以下顯示的是十位病人的資料，住院天數為{1, 3, 6, 7, 2, 4, 12, 15, 5, 9}，保險給付為{5, 17.5, 18, 20, 6, 14, 42, 54, 17, 30}

得迴歸分析之電腦輸出如下：

摘要輸出				
迴歸統計				
R 平方	0.9611			
觀察值個數	10			
	係數	標準誤	t 統計	P-值
截距	0.63	1.85	(a)	0.741
住院天數	3.39	0.24	(b)	<0.001

- (1) (4 分)請依照號碼說明(a)及(b)如何計算：

(a) \_\_\_\_\_

(b) \_\_\_\_\_

- (2) (4 分)請說明表格中(b)的 t 統計量檢定的虛無假設及對立假設為何?
- (3) (3 分)請說明表格中依據 t 分配計算 P 值時，該 t 分配的自由度是多少?
- (4) (3 分)請根據以上的表格，說明如何求出住院天數與保險給付的樣本相關係數?
- (5) (3 分)請說明上述 R 平方( $R^2$ )代表的意義與用處。
- (6) (3 分)請根據以上的表格，寫出住院天數與保險給付的迴歸模式。
- (7) (5 分)請說明上述的迴歸分析中，必須假設保險給付服從什麼樣的常態分配?

試題隨卷繳回