

是非題（每題 2 分）

1. 標準差(standard deviation)可以用來推估母體的分散程度，標準誤(standard error)則可用來推估樣本平均值的估計準確程度。
2. 變異數可以說明母體分散程度，但不能說明母體是否為對稱或是右偏分布。
3. 信賴區間越寬，信心水準越低。
4. 信賴區間可以描述樣本平均值在該區間的信心程度，不能描述樣本平均值落在該區間的機率。
5. 勝算比值(odds ratio, OR)跟卡方檢定一樣，都可以描述兩個二元變數之間的關係強度。
6. 兩組獨立樣本的 t 檢定，如果變異數不相同，則不能使用變異數分析(analysis of variance)進行檢定。
7. 如果已知母體為一對稱分布，則樣本中位數也可以估計母體期望值。
8. 如果已知台灣家戶所得為一右偏分布，則樣本中位數比樣本平均值更適合描述母體的中心。
9. 中央極限定理(central limit theorem)可以用來求得母體期望值的抽樣分配。
10. 中央極限定理(central limit theorem)僅適用於連續型隨機變數分配中所抽出之大樣本數據，不適用於離散型隨機變數分配中所抽出之大樣本數據。

※ 注意：選擇題請於試卷內之「選擇題作答區」依序作答。

單選題（每題 3 分）

11. 以下有關布阿松分配(Poisson distribution)的敘述何者最不恰當?
 - (1)布阿松分配可用於描述罕見疾病之發生人數的分布。
 - (2)在某些情況下，布阿松分配可用來逼近二項式分配(Binomial distribution)。
 - (3)服從布阿松分配的隨機變數必為正整數，其可能的值為 1, 2, 3, ... 到無限大。
 - (4)布阿松分配的期望值跟變異數會相等。
12. 某公司接受委託調查某國小學童今年一月是否罹患流感的情況，已知低年級學童與高年級學童的發生機率不同，分別為 p_1 與 p_2 。請問以下何者最不恰當?
 - (1)有一對兄弟分別就讀低年級與高年級，則這對兄弟中恰有 1 人得病的機率是 $p_1(1-p_1)+p_2(1-p_2)$ 。
 - (2)在 100 位低年級學童當中，罹患流感的人數會是二項式分配(Binomial distribution)。
 - (3)該公司調查 200 位高年級學童當中，罹患流感的比例，這個比例值是一種樣本平均值(sample mean)。
 - (4)如果該公司調查 100 位低年級學童當中，罹患流感的比例，這個比例值是一個統計量(statistic)。
13. (延續上題)該公司調查 100 位低年級學童，發現其中有 80 人罹患流感，請問以下何者是 p_1 的 95% 信賴區間?
 - (1) $(0.8-1.96 \times 0.8 \times 0.2/100, 0.8+1.96 \times 0.8 \times 0.2/100)$
 - (2) $(0.8-1.96 \times 0.8 \times 0.2/10, 0.8+1.96 \times 0.8 \times 0.2/10)$
 - (3) $(0.8-1.96 \times 0.8 \times 0.2, 0.8+1.96 \times 0.8 \times 0.2)$
 - (4) 以上皆非。

見背面

14. 某研究擬探討快速跑步一公里之後，20-30歲的男性與女性的每分鐘心跳增加次數是否有所不同。隨機抽取20位男性與20位女性之後，分別測量每個人跑步前與跑步後的心跳次數差異，男性的平均增加次數為19次，變異數為25；女性的平均增加次數為24次，變異數為36。如果利用迴歸模式 $Y = a + bX + \varepsilon$ 進行分析，請問以下哪一個敘述最恰當？
- (1) 解釋變數 X 為心跳增加次數，應變數 Y 為性別。
 - (2) 迴歸模式中的 Y 服從常態分配，期望值為0，變異數為 σ^2 。
 - (3) 若檢定迴歸係數 b 是否為0，則檢定統計量服從自由度為38的t分布。
 - (4) 如果檢定迴歸係數 b 是否為0，所得到的結果是不顯著，那麼利用成對的t檢定(paired t test)來進行分析也會得到不顯著的結果。
15. 王同學收集100位隨機抽選之健康女性的體重(W ，單位為公斤)，與膽固醇($Chol$ ，單位為mg/100 cc)的資料，這些女性的體重數值分布於51至62之間，膽固醇數值分布於121.47至171.19之間。建立了體重與血壓的迴歸模式 $E(Chol) = 316.25 - 3.06 \times W$ ，請問以下何者最不恰當？
- (1) 上述迴歸模式為簡單線性迴歸模式。
 - (2) 由上述迴歸模式可知，每增加一公斤，膽固醇數值平均改變量為3.06 mg/100 cc。
 - (3) 體重與膽固醇之間的線性相關係數應為負值。
 - (4) 王小姐體重46公斤，可利用上述迴歸模式估計王小姐的膽固醇為 $316.25 - 3.06 \times 46$ 。

單選題 (每題5分)

16. 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95%信賴區間，其寬度趨近0。
 - (C) 皮爾森相關係數 (Pearson correlation coefficient) 估計值趨近一個-1與1間之定值。
 - (D) 假說檢定拒絕虛無假說的機率趨近100%。
 - (E) 樣本標準差趨近一個大於0之定值。
17. 有關迴歸分析 (regression analysis)，下述何者錯誤？
- (A) 應變項 (dependent variable) 為連續變項且其值大於0時，可考慮將之取對數後，進行線性迴歸分析 (linear regression)。
 - (B) 應變項為計數資料 (count data) 時，可考慮進行布阿松迴歸分析 (Poisson regression)。
 - (C) 自變項 (independent variable) 可以為任何測度尺度的變項。
 - (D) 複線性迴歸分析 (multiple linear regression) 的判別係數 (coefficient of multiple determination)，不論自變項個數多寡，其值介於0與1之間。
 - (E) 複布阿松迴歸分析 (multiple Poisson regression) 之自變項個數超過50時，受限於電腦計算能力，通常即無法執行。

接次頁

18. 有關機率分佈 (probability distribution) , 下述何者錯誤?
- (A) 標準常態分佈 (standard normal distribution) 的平方, 即為自由度為 1 的卡方分佈 (chi-square distribution)。
 - (B) 不論自由度為何, 卡方分佈曲線下之總面積皆為 1。
 - (C) 布阿松分佈 (Poisson distribution) 的期望值 (expected value) 等於其變異數 (variance)。
 - (D) 伯努立分佈 (Bernoulli distribution) 是二元性 (binary) 且右偏 (positively skewed)。
 - (E) 抽樣分佈 (sampling distribution) 是樣本統計量 (statistic) 之機率分佈。
19. 有關樣本數計算, 下述何者錯誤?
- (A) 樣本數計算所需參數 (parameter), 要用客觀方法取得, 不可仰賴先驗知識 (*a priori* knowledge)。
 - (B) 一般統計方法在資料搜集完整後才進行, 樣本數計算則是在研究設計階段時即進行。
 - (C) 若希望信賴區間估計越窄, 所需樣本數即越多。
 - (D) 若將顯著水準 (level of significance) 訂的越大, 所需樣本數即越少。
 - (E) 可以使用統計套裝軟體進行樣本數計算。
20. 有關無母數方法 (non-parametric method), 下列何者錯誤?
- (A) 樣本數極小時 (比如 $n < 5$), 仍可使用。
 - (B) 資料分佈有三個峰 (peak) 以上時, 仍可使用。
 - (C) Kruskal-Wallis test 是無母數方法。
 - (D) Wilcoxon signed rank test 是無母數方法。
 - (E) 無母數方法計算複雜, 常用的統計套裝軟體並不支援。
21. 有關隨機抽樣 (random sampling) 與隨機指派 (randomization), 下列何者錯誤?
- (A) 可利用電腦產生亂數, 達到隨機的目的。
 - (B) 病例對照研究法 (case-control study) 中, 對照組的選取可以不採用隨機抽樣的方式。
 - (C) 臨床試驗 (clinical trial) 中, 病患若已有編號, 則可以採用系統性隨機指派 (systematic randomization) 的方式, 比如每十位病患取一人進入研究。
 - (D) 隨機抽樣目的在樣本具有代表性 (representativity), 隨機指派目的在不同治療組別間之可比性 (comparability)。
 - (E) 不論隨機抽樣與隨機指派, 皆無法避免隨機誤差 (random error)。
22. 下述何者是配對病例對照研究 (matched case-control study) 之迴歸分析方法?
- (A) conditional logistic regression。
 - (B) Cox regression。
 - (C) Poisson regression。
 - (D) regression to the mean。
 - (E) meta-regression。

23. 以下有關變異數分析(ANOVA)之敘述，何者錯誤？
- (A) 假設不同組資料的變異數一致
 - (B) 是在檢定不同組資料的平均數是否一致
 - (C) 也可以用來檢定特定兩組的平均數是否一致
 - (D) 檢定統計量一定大於 0
 - (E) 檢定統計量的虛無分配為 F 分配
24. 以下有關邏輯斯迴歸(logistic regression)之敘述，何者錯誤？
- (A) 針對二元(binary)的反應變數所使用的模型
 - (B) 解釋變數可以是任意的資料型態
 - (C) 迴歸係數大於零代表該解釋變數會增加事件發生的機會
 - (D) 迴歸係數的意義是 odds ratio
 - (E) 使用最大概似法(maximum likelihood estimator)估計迴歸係數
25. 以下有關線性迴歸(linear regression)之敘述，何者錯誤？
- (A) 解釋變數可以是任意的資料型態
 - (B) 解釋變數彼此間的相關性越強，估計越不穩定
 - (C) 解釋變數的變異越大，估計越穩定
 - (D) 解釋變數越多，模型所得到的判定係數(R-square)越大
 - (E) 解釋變數越多，模型所得到的校正判定係數(adjusted R-square)越大
26. 以下有關布阿松迴歸(Poisson regression)之敘述，何者錯誤？
- (A) 反應變數必須為大於等於零的正整數
 - (B) 不同反應變數的量測單位必須一致
 - (C) 解釋變數可以是任意的資料型態
 - (D) 迴歸係數代表 $\log(\text{mean ratio})$
 - (E) 使用最大概似法(maximum likelihood estimator)估計迴歸係數
27. 以下有關存活分析(Survival Analysis)之敘述，何者錯誤？
- (A) 存活曲線越高代表存活機率較大
 - (B) 存活曲線越高代表存活時間的中位數較大
 - (C) 存活曲線的斜率小於等於零
 - (D) 所有的存活函數起始點都一樣，但是觸碰到零的時間點不見得相同
 - (E) 存活曲線越陡的地方代表越不容易發生事件

問答題 (每題 5 分)

28. 假設有某種新藥，其療效 μ 必須大於某一標準值 μ_0 衛生署才會認定為有療效。若從藥廠的觀點要讓利益最大化，希望能控制[有療效的藥被認定為無療效]的機率不超過 0.27，請寫出相對應的虛無假說和對立假說。