

- 一、單選題 45 分(1-15 題每小題 3 分，請於試卷內之「選擇題作答區」依序作答。)
- 大勝鄉今年的鄉長選舉共有三位候選人，各代表 A、B、C 三個政黨。某民調公司想要預估各黨候選人今年的得票率。已知去年 A、B、C 三個政黨的得票率分別為 45%、35%、20%。現隨機抽選 120 位鄉民，其中有 60、45、15 位鄉民表示將投給 A、B、C 三個政黨。根據這個調查，請問以下哪個敘述最恰當？
 - 未來 A 政黨的得票率將達到 $60/120=50\%$ 。
 - 未來 C 政黨的得票率將超過 20%。
 - 未來任一位候選人的得票率不會超過 50%。
 - 未來投票結果可能跟這份民調不一樣。
 - (延續上題)以下何者是未來 B 政黨的得票率的 95%信賴區間？
 - $37.5\% \pm 1.96 \times \sqrt{\frac{37.5\% \times 62.5\%}{120}}$
 - $37.5\% \pm 1.96 \times \sqrt{\frac{37.5\% \times 62.5\%}{45}}$
 - $37.5\% \pm 1.96 \times \sqrt{37.5\% \times 62.5\%}$
 - $35\% \pm 1.96 \times \sqrt{\frac{35\% \times 65\%}{45}}$
 - (延續上題)如果要根據這個調查檢定三個政黨目前的支持率是否各為 $1/3$ ，請問應該進行以下何種檢定？
 - 卡方相關性檢定。
 - 卡方適合度檢定。
 - 卡方變異數分析。
 - 兩個獨立母體的比例檢定。
 - (延續上題)如果要進行上述小題的檢定，必須計算每個觀察值所對應的期望值，請問以下何者正確？
 - 期望值分別為 60、45、15。
 - 期望值分別為 $120 \times 45\%$ 、 $120 \times 35\%$ 、 $120 \times 20\%$ 。
 - 期望值分別為 $100 \times 120/240$ 、 $85 \times 120/240$ 、 $55 \times 120/240$ 。
 - 期望值分別為 40、40、40。
 - (延續上題)有關上述檢定統計量的自由度，請問以下何者正確？
 - 自由度為 1。
 - 自由度為 2。
 - 自由度為 3。
 - 自由度為 4。
 - 某研究想要探討一個人的血壓如何受到體重的影響，請問以下哪一個簡單線性迴歸模式 (simple linear regression model) 最恰當？
 - 血壓 = $a + b \times$ 體重， a 、 b 皆為迴歸係數。模式
 - 體重 = $a + b \times$ 血壓， a 、 b 皆為迴歸係數。
 - $E(\text{血壓}) = a + b \times$ 體重， $E(\text{血壓})$ 為血壓的期望值。
 - $E(\text{體重}) = a + b \times$ 血壓 + ε 。其中 ε 為誤差項。

7. (延續上題)如果要利用上述簡單線性迴歸模式檢定血壓是否受到體重的影響,請問以下何者正確?
- (A) 檢定虛無假設 $a=0$, 對立假設 $a \neq 0$ 。
 - (B) 檢定虛無假設 $b=0$, 對立假設 $b \neq 0$ 。
 - (C) 檢定虛無假設 $(a=0, b=0)$, 對立假設 $(a \neq 0, b \neq 0)$ 。
 - (D) 檢定虛無假設 $(a=0, b=0)$, 對立假設 $(a \neq 0, b > 0)$ 。
8. (延續上題)有關上述簡單線性迴歸模式之統計分析,請問以下何者錯誤?
- (A) 常態分配的假設是因為誤差項通常被假設為常態分配。
 - (B) 因為誤差項通常被假設為常態分配,所以體重也是常態分配。
 - (C) 上述簡單線性迴歸模式中誤差項的變異數是個常數。
 - (D) 上述簡單線性迴歸模式中的變異數不會隨著體重改變。
9. 為了比較化療藥物 A 和 B 的副作用,針對 50 位服用化療藥物 A 的病人,其中 27 人之藥物副作用被評估為「大」;另外根據 50 位服用化療藥物 B 的病人,其中 30 位之藥物副作用被評估為「大」。請估計服用化療藥物 B 而產生副作用相較於化療藥物 A 產生副作用的勝算比值(odds ratio)為
- (A) $(27 \times 20) / (23 \times 30)$
 - (B) $(23 \times 30) / (27 \times 20)$
 - (C) $(27 \times 30) / (23 \times 20)$
 - (D) $(23 \times 20) / (27 \times 30)$
10. (延續上題)如果想要檢定化療藥物 A 和 B 產生副作用的比例是否相同,以下敘述何者不正確?
- (A) 可以計算兩種藥物產生副作用的比例差異的信賴區間,看看是否包含 0 來完成檢定。
 - (B) 可以利用卡方檢定來檢定兩種藥物是否同樣容易產生副作用。
 - (C) 可以利用兩個獨立母體比例的檢定,來檢定兩種藥物產生副作用的機率是否一樣。
 - (D) 可以檢定上一小題的勝算比值是否為 0。
11. 有關簡單線性迴歸模式之統計分析,請問以下何者錯誤?
- (A) 簡單線性迴歸模式研究的是一個反應變數(response variable)和一個解釋變數(covariate)之間的所有關係。
 - (B) 反應變數彼此之間必須是獨立關係。
 - (C) 解釋變數的期望值與反應變數有關。
 - (D) 反應變數彼此之間不一定是相同的常態分配。
12. 以下有關中央極限定理(central limit theorem)的敘述何者最為恰當?
- (A) 中央極限定理討論的是樣本平均值在大樣本情況下的抽樣分配。
 - (B) 如果原樣本的母體是常態分配,那麼只要樣本數夠大,極限分配就會是原來母體的常態分配。
 - (C) 如果原樣本的母體不是常態分配,只要樣本數夠大,極限分配就會是常態分配,期望值與變異數都跟原來母體的期望值與變異數相同。
 - (D) 中央極限定理逼近到常態分配的速度與原來母體的期望值有關。

13. 有關 Receiver Operating Characteristic (ROC) 曲線的敘述，下列何者錯誤。
- (A) 通常用來評估一個分類準則的分類能力
 - (B) ROC 曲線不受測量單位的影響，畫出來都一樣
 - (C) ROC 曲線和通過(0, 0)的 45° 線一般不會有交點
 - (D) ROC 曲線越靠近(0, 1)代表鑑別力越強
14. (延續上題) 下列何者錯誤？
- (A) ROC 曲線可用來比較不同分類準則的鑑別能力
 - (B) 曲線下面積為 1 時表示有完美的鑑別力
 - (C) 曲線下面積為 0.5 時表示完全沒有鑑別力
 - (D) 不同的 ROC 曲線可能有相同的曲線下面積
15. 考慮解釋變數 X 及記數(counting data)反應變數 Y 。以下針對布阿松迴歸(Poisson regression) regression) 的敘述，何者錯誤？
- (A) 使用最大概似法(maximum likelihood)做估計
 - (B) 是針對 Y 的條件期望值建立線性模型
 - (C) 可以利用常態分配逼近的方式建立迴歸係數的信賴區間
 - (D) X 的迴歸係數為正數時，越大的 X 傾向伴隨著越大的 Y

二、單選題 35 分(16-22 題每小題 5 分，請於試卷內之「選擇題作答區」依序作答。)

16. 樣本數增加趨近無限，下述何者錯誤？
- (A) 平均值的估計會趨近一個定值。
 - (B) 標準差的估計會趨近一個定值。
 - (C) 標準誤趨近零。
 - (D) 在給定的一個顯著水準下，型一誤差率趨近零。
 - (E) 研究所需經費會隨樣本數增加而增加，因此實際研究很難有無限大的樣本。
17. 有關假說檢定，下述何者錯誤？
- (A) 顯著水準的數值訂的越小，表示假說檢定越嚴格，因此檢力也會越小。
 - (B) 某甲以擲骰子(共六面)的方式進行假說檢定，當 6 點之面朝上時，即拒絕虛無假說，則此假說檢定的型一誤差率為 1/6。
 - (C) 某乙以擲銅板(共兩面)的方式進行假說檢定，當正面朝上時即宣稱達到統計顯著性，則此假說檢定的檢力為 50%。
 - (D) 適當的組合多個正五面體骰子，以及多個銅板，可以達到顯著水準為 0.05 以及檢力為 50% 的假說檢定。
 - (E) 貝氏學派研究者(Bayesian)可以將先驗知識納入假說檢定之中。

見背面

18. 有關信賴區間，下述何者錯誤？
- (A) 95%信賴區間比 50%信賴區間還寬。
 - (B) 50%確切 (exact) 信賴區間的公式，能夠保證按其公式所計算出的信賴區間上下限，其間涵蓋真值的機率不小於 0.5。
 - (C) 有多種方法可以計算大樣本 (asymptotic) 信賴區間 (比如 Wald、score 及 likelihood ratio 等方法)。小樣本時他們的結果通常都差不多，樣本數大時，才可看出不同的結果。
 - (D) 以信賴區間來呈現研究結果，比僅呈現假說檢定的 P 值還好。因為信賴區間既有定性亦有定量，而 P 值僅有定性。
 - (E) 貝氏學派研究者可以將先驗知識納入信賴區間估計中。
19. 有關樣本數計算，下述何者錯誤？
- (A) 通常是在研究結案時 (資料完整搜集後，after the fact) 為之，如此才能得到確實的樣本數數值。
 - (B) 通常可以分成兩種情境：依據假說檢定所希望的檢力大小，以及依據信賴區間估計所希望的區間寬窄。
 - (C) 小樣本研究所需樣本數，可能要依賴確切樣本數計算公式。雖然計算較繁雜，但可避免大樣本方法的謬誤。
 - (D) 樣本數計算最好能夠配合敏感度分析 (sensitivity analysis)，以免參數的選定，不符實境。
 - (E) 樣本數計算通常有現成的套裝軟體可供選用，可盡量使用之。
20. 有關常態分佈 (normal distribution)，下述何者錯誤？
- (A) 為單峯、對稱、鐘型的連續分佈。
 - (B) 常態分佈之機率密度函數，其曲線下面積為機率。
 - (C) 兩個獨立的 (independent) 不同平均值的常態分佈之混合 (mixture)，亦為常態分佈。
 - (D) 兩個相依的 (dependent) 相同平均值的常態分佈之相加，亦為常態分佈。
 - (E) 標準常態分佈的平方，即為自由度為 1 的卡方分佈。
21. 有關無母數 (non-parametric) 統計檢定，下述何者錯誤？
- (A) 資料不符合常態分佈時仍可使用之。
 - (B) 小樣本研究可使用之。
 - (C) 大樣本研究若使用無母數統計檢定，會有型一誤差膨脹 (inflation) 的問題。
 - (D) Wilcoxon signed rank test 是無母數統計。
 - (E) Friedman test 是無母數統計檢定。
22. 有關線性迴歸分析 (linear regression)，下述何者錯誤？
- (A) 自變項 (independent variable) 可為二元性 (binary) 變項。
 - (B) 將自變項放大 (比如乘以 10)，整個模式的適合度 (goodness of fit) 變差。
 - (C) 將依變項 (dependent variable) 平移 (比如加 10)，截距項 (intercept) 會因此而有所改變。
 - (D) 有多個自變項的複線性迴歸分析，其參數估計法仍可採用最小平方法。
 - (E) 複線性迴歸分析的判別係數 (coefficient of multiple determination) 是一個介於 0 與 1 之間的沒有單位的數字。

問答題(20分)

1. 你被委託研究有無接受治療(X_1)以及年齡(X_2)是否會影響存活時間(T)。你考慮使用 Cox Proportional Hazard (PH) Model 來回答問題。
 - (1) 根據以上的描述，假設治療的成效與年齡之間沒有交互作用，請詳細描述該 Cox PH Model 的數學表達式。
 - (2) 請詳細解釋在上題的模型中， X_1 的迴歸係數的意義為何。
 - (3) 假設有一個新的治療方法(X_3)。請描述該如何用 Cox PH Model 以檢定[控制年齡下，新舊治療沒有差異]。你必須寫出使用的模型以及相對應的假說。
 - (4) 請描述該如何用 Cox PH Model 以檢定[新的治療方法與年齡之間沒有交互作用]。你必須寫出使用的模型以及相對應的假說。

