

一、是非題 (每題 2 分) 請於試卷內之「非選擇題作答區」標明題號依序作答。

1. 任一個常態分配 $N(\mu_1, a^2)$ 一定可以轉換成另一個常態分配 $N(\mu_2, b^2)$ 。
2. 敏感度(sensitivity)跟特異度(specificity)都是一種條件機率(conditional probability)。
3. 男女兩個不同性別的血壓不能以兩個個別的 boxplot 來比較分布狀況。
4. 如果要比較臺灣大學甲、乙兩學院教師血壓分布的異質性，可以使用兩個獨立樣本的 t 檢定 (independent two-sample t test)。
5. 變異數分析(Analysis of variance, ANOVA)可以檢定不同母體的母體期望值是否相等，也可以檢定不同母體的變異數是否相等。
6. 布阿松分配(Poisson distribution)一定是對稱的分布。
7. 皮爾森相關係數(Pearson's correlation coefficient)可以拿來評估兩個連續變數之間的任何關係。
8. 樣本數的大小不會影響信賴區間的寬度。
9. 一組樣本之樣本平均值的數值大小不會影響該樣本之母體期望值之信賴區間的寬度。
10. 無母數統計(nonparametric statistics)方法指的是不假設資料的母體分配來進行統計分析。

二、選擇題 80 分 (單選，1-10 每題 3 分，11-20 每題 5 分) 請於試卷內之「選擇題作答區」依序作答。

1. 某研究探討含高劑量塑化劑之食品對懷孕的影響，餵食了 45 隻懷孕母鼠含高劑量塑化劑之食品，其中 30 隻母鼠流產。另外 50 隻懷孕母鼠則食用不含塑化劑之食品，其中 25 隻母鼠流產。請問，若是高劑量塑化劑之食品對懷孕沒有影響，那麼一般母鼠的流產比例估計值以下何者最佳：(A) 30/45; (B) 25/50; (C) 30/95; (D) 55/95。
2. (延續上題)請問是否食用含高劑量塑化劑之食品會流產的勝算比(odds ratio, OR)應以以下何者來估計為最佳：(A) 0.5; (B) 1; (C) 1.5; (D) 2。
3. (延續上題)如果要計算 OR 的 95% 信賴區間，一般必須先計算 $\log(\text{OR})$ 的 95% 信賴區間，已知 $\log(\text{OR})$ 的估計值為 k ，標準誤(se)為 0.42，請問以下何者是 $\log(\text{OR})$ 的 95% 信賴區間：
 - (A) $(k - 1.96 \times \frac{0.42}{\sqrt{45}}, k + 1.96 \times \frac{0.42}{\sqrt{45}})$
 - (B) $(k - 1.96 \times \frac{0.42}{\sqrt{95}}, k + 1.96 \times \frac{0.42}{\sqrt{95}})$
 - (C) $(k - 1.96 \times 0.42, k + 1.96 \times 0.42)$
 - (D) $(k - 1.96 \times 0.42 \times \sqrt{95}, k + 1.96 \times 0.42 \times \sqrt{95})$
4. (延續上題)如果利用上一小題的答案來計算 OR 的 95% 信賴區間，若上一小題的 $\log(\text{OR})$ 的 95% 信賴區間以符號來表示為 (a, b) ，那麼以下何者是 OR 的 95% 信賴區間？
 - (A) (a, b) ; (B) $(\log(a), \log(b))$; (C) $(\exp(a), \exp(b))$; (D) $(2a, 2b)$ 。
5. 在一個探討吸菸得肺癌的研究中，如果得到的 OR 值的 95% 信賴區間為 $(1.22, 2.33)$ ，你認為以

見背面

下敘述何者正確？

- (A) 「吸菸得肺癌的機率」是「不吸菸得肺癌的機率」的 $(1.22+2.33)/2$ 倍。
(B) 「吸菸得肺癌的機率」比「不吸菸得肺癌的機率」高。
(C) 該信賴區間不包含 0，所以吸菸與得肺癌沒有統計上的顯著關係。
(D) 若檢定吸菸與得肺癌是否有關， p 值將會大於 0.05。
6. 某研究認為統一超商門市數目(以 X 表示，單位為間)，與營業收入(單位為億元)之間的迴歸關係式應為 $E(Y)=a+bX$ ；在 1997 至 2007 年之間門市數目由 1588 間成長為 4705 間。收集這 11 年之間的資料得迴歸模式為 $E(Y)=-14.566+0.228X$ ；請問若門市數目由 4000 間增加至 4100 間，則營業收入的成長最接近以下多少億元？
(A) $-14.566+0.228 \times 4000$ ；
(B) $-14.566+0.228 \times 4100$ ；
(C) $-14.566+0.228 \times 100$ ；
(D) $-14.566+0.228 \times 1$ 。
7. (延續上題)由上一題的迴歸模式你認為門市數目與營業收入之間
(A) 呈正相關；(B) 呈負相關；(C) 無相關；(D) 無法確定。
8. (延續上題)利用上題之迴歸模式，請問以下何者正確？
(A) 若門市數目能達到 10000 間，則營業收入可超過 2000 億元。
(B) 早期門市數目不到 1000 間時，營業收入不到 300 億元。
(C) 該模式中 -14.566 與 0.228 這兩個估計值可利用最小平方法求得。
(D) 門市數目與營業收入兩者之間的皮爾森相關係數為 0.228。
9. (延續上題)若要檢定門市數目與營業收入兩者之間是否有顯著相關，就等同於檢定以下何者？
(A) $H_0: a=0$ versus $H_1: a \neq 0$ ；
(B) $H_0: b=0$ versus $H_1: b \neq 0$ ；
(C) $H_0: a=0$ 且 $b=0$ versus $H_1: a \neq 0$ 且 $b \neq 0$ ；
(D) $H_0: a \neq 0$ 且 $b \neq 0$ versus $H_1: a=0$ 且 $b=0$ 。
10. (延續上題)上一題應使用以下自由度為多少的 t 檢定？
(A) 自由度為 11 的 t 檢定；
(B) 自由度為 10 的 t 檢定；
(C) 自由度為 9 的 t 檢定；
(D) 自由度為 8 的 t 檢定。
11. 有關標準差(standard deviation)及標準誤(standard error)，下列何者錯誤？
(A) 標準差越大，表示資料的散佈程度越大。
(B) 標準誤越大，表示資料的測量誤差(measurement error)越大。
(C) 樣本數增加，標準誤變小。
(D) 樣本數增加，標準差原則上不變。
(E) 標準差和標準誤有同樣的「單位」(比如公斤，公分等等)。

12. 有關檢力(power)，下列何者錯誤？
- (A) 樣本數增大，檢力亦增加。
 - (B) 樣本數趨近無限大，檢力亦趨近無限大。
 - (C) 一個不偏的(unbiased)統計檢定，其檢力不會小於其顯著水準(significance level)。
 - (D) 顯著水準越小，檢力亦小。
 - (E) 通常我們會希望一個研究的檢力至少有 80%以上。
13. 有關樣本平均值的信賴區間(confidence interval)，下列何者錯誤？
- (A) 樣本數增大，信賴區間變窄。
 - (B) 樣本數趨近無限大，信賴區間的寬度趨近於 0。
 - (C) 95%信賴區間比 99%信賴區間較寬。
 - (D) 信賴區間的上限(upper limit)大於其下限(lower limit)。
 - (E) 25%信賴區間的上限大於樣本平均值；25%信賴區間的下限小於樣本平均值。
14. 有關布阿松分佈(Poisson distribution)，下列何者錯誤？
- (A) 為離散分佈(discrete distribution)。
 - (B) 其可能出現數值為正整數(含零)，不可能出現負值。
 - (C) 只有一個參數(parameter)；此參數同時為該分佈之期望值及變異數。
 - (D) 將兩個獨立的布阿松分佈相加，可得到一個二項式分佈。
 - (E) 流行病學中死亡率及發生率的信賴區間計算，一般皆依賴布阿松分佈。
15. 有關無母數方法(non-parametric method)，下列何者錯誤？
- (A) 資料不符合常態分佈，仍可使用之。
 - (B) 資料不符合布阿松分佈(Poisson distribution)，仍可使用之。
 - (C) 無母數檢定方法的確切 p 值(exact p value)無法求算，只能計算近似值。
 - (D) 小樣本時，通常要考慮採用無母數方法。
 - (E) 常用的統計套裝軟體，通常亦有支援無母數方法的分析。
16. 欲檢定城市居民與鄉村居民(各抽樣 100 人)之膽固醇平均濃度有無不同，可採用下述何者？
- (A) Student t test
 - (B) McNemar test
 - (C) ANOVA (analysis of variance)
 - (D) ANOM (analysis of mean)
 - (E) Kruskal-Wallis test
17. 有關羅吉斯迴歸分析(logistic regression analysis)，下列何者錯誤？
- (A) 自變項(independent variable)可以為任何測度尺度的變項。
 - (B) 應變項(dependent variable)的測度尺度為二元性(binary)變項。
 - (C) 自變項與應變項皆不需符合常態分佈。
 - (D) 截距項(intercept)亦是其迴歸係數之一，其值可能為正，亦可能為負。
 - (E) 樣本數超過一萬個案時，受限於電腦計算能力，不建議使用羅吉斯迴歸分析。

18. 假設 X 服從 $\text{Binomial}(n, p)$ ，下列何者正確？
- (A) 只有兩種結果，成功或失敗
 - (B) $\text{Var}(X) = p(1 - p)$
 - (C) 當 $p = 0.5$ 時，實驗結果的變異是最小的
 - (D) 可以表示為 n 個任意的 $\text{Bernoulli}(p)$ 隨機變數和
 - (E) 根據 p 的數值，可能為右偏也可能為左偏的分布
19. 下列何者無法從 box plot 中得到？
- (A) 組距(range)
 - (B) 第一四分位數
 - (C) 第二四分位數
 - (D) 最小值(minimum)
 - (E) 平均數(mean)
20. 下列關於信賴區間(confidence interval)的敘述，何者正確？
- (A) 信心水準(confidence level)是根據資料所決定的
 - (B) 95%信心水準是指隨機抽取一參數值，則該數值有 0.95 的機率落入所建立的信賴區間
 - (C) 信心水準給定後，信賴區間就固定了，不會隨著資料而改變
 - (D) 信心水準固定下，樣本數越多，則信賴區間會較寬
 - (E) 可用來執行假設檢定

試題隨卷繳回