

114年第二次專門職業及技術人員高等考試營養師、護理師、社會工作師考試、
114年專門職業及技術人員高等考試心理師、法醫師、語言治療師、
聽力師、牙體技術師、公共衛生師考試、高等暨普通考試驗光人員考試試題

等 別：高等考試
類 科：公共衛生師
科 目：生物統計學
考試時間：2 小時

座號：_____

※注意：可以使用電子計算器。

甲、申論題部分：(50分)

- (一)不必抄題，作答時請將試題題號及答案依照順序寫在申論試卷上，於本試題上作答者，不予計分。
(二)請以藍、黑色鋼筆或原子筆在申論試卷上作答。
(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、某醫學中心過去 1 年的住院病人，住院天數標準差為 3.1 天。從中隨機抽出 100 位住院病人，發現其住院平均天數為 4.3 天。

(一)試估計某醫學中心所有住院病人，過去 1 年住院平均天數之 99% 信賴區間為何？(10 分)

$$Z_{0.90} = 1.28, Z_{0.95} = 1.64, Z_{0.975} = 1.96, Z_{0.99} = 2.33, Z_{0.995} = 2.58$$

(二)假設另一所醫學中心住院平均天數為 4 天，試問本醫學中心所有住院病人之住院平均天數和另一醫學中心比較，是否有差異？並說明原因。
($\alpha = 0.01$) (5 分)

二、某學校使用三種不同教學方法教授生物統計學，欲評估三種教學方法的教學效果是否相同，三種教學方法之學期成績計算以後，得到如下變異數分析表：

變異來源	平方和	自由度	均方	F 值
組間	A	B	237.4	E
組內	C	D	71.72	
合計	1909.22	22		

(一)請寫出該檢定的虛無假設與對立假設。(5 分)

(二)請分別計算表格中 A、B、C、D 與 E 五個數值。(10 分)

三、某醫療行為的研究想要探討都市居民生病時，看醫生的比例是否高於鄉村居民。今隨機抽取 300 位都市居民，有 180 位表示生病時會去看醫生；另隨機抽取 200 位鄉村居民，有 60 位表示生病時會去看醫生，試問在顯著水準 0.05 下，都市居民生病時會去看醫生的比例，是否比鄉村居民比例高？(20 分)

乙、測驗題部分：(50分)

代號：2108

(一)本試題為單一選擇題，請選出一個正確或最適當答案。

(二)共40題，每題1.25分，須用2B鉛筆在試卡上依題號清楚劃記，於本試題或申論試卷上作答者，不予計分。

- 1 在線性迴歸中，欲檢定迴歸係數是否具有統計上的顯著意義時，需採取那一種檢定？
(A) Z 檢定 (B) t 檢定 (C) χ^2 檢定 (D) F 檢定
- 2 某次考試結束後，主辦單位想知道教育程度（學士/碩士/博士）是否為成績（0 分至 100 分）的預測因子。通常在將教育程度放入線性迴歸之前，若有需要做前置處理，會是下列何者？
(A)標準化 (B)取對數 (C)虛擬變數 (D)多項式轉換
- 3 當下面四個圖形的 y 軸單位都是百分比的時候，那一個圖形的百分比意義與其它三者不同？
(A)長條圖 (B)直方圖 (C)盒形圖 (D)次數多邊圖
- 4 下列何者是 two-sample t test 對應的無母數方法？
(A) Sign test (B) Wilcoxon signed rank test
(C) Wilcoxon rank sum test (D) Kruskal-Wallis test
- 5 今量測一群學生的身高（單位：cm）後，欲計算出這群學生的變異係數，試問其單位為何？
(A) cm^2 (B) cm (C) $\text{cm}^{0.5}$ (D)沒有單位
- 6 某研究者欲瞭解，當新興傳染病出現時，男性和女性願意在第一時間注射疫苗的比例是否有差異，在調查了 100 位民眾之後，結果發現，在受訪的 60 位男性中，有 30 位願意；在受訪的 40 位女性中，有 16 位願意，試問，在套用兩個母群體比例差的檢定公式時，標準誤該如何計算？
(A) $\sqrt{\frac{0.5 \times 0.5}{59} + \frac{0.4 \times 0.6}{39}}$ (B) $\sqrt{\frac{0.5 \times 0.5}{60} + \frac{0.4 \times 0.6}{40}}$
(C) $\sqrt{\frac{0.46 \times 0.54}{59} + \frac{0.46 \times 0.54}{39}}$ (D) $\sqrt{\frac{0.46 \times 0.54}{60} + \frac{0.46 \times 0.54}{40}}$
- 7 承上題，男性和女性兩組比例差的統計檢定量為何？
(A)-2.5 (B)-3.33 (C)-0.25 (D)-0.33
- 8 在寶可夢遊戲中，有一種設定是「異色寶可夢」，其外表的顏色與一般寶可夢不同，且出現機率低，會特別吸引玩家的捕捉意願。在手機遊戲中，官方舉辦特定活動時，會讓異色寶可夢出現的機率增加。假設官方表示，在某次活動中，每個玩家在遇見寶可夢時，有 20%的機率會碰到異色（每次的遇見都是獨立事件）。當活動結束後，某玩家說他總共遇見 100 隻，其中異色有 10 隻，試問，如果要協助他進行遇見異色的機率是否與官方數據相同，其計算所得的統計檢定量，其分布會趨近下列那一種分布？
(A) Z 分布 (B) t 分布 (C) χ^2 分布 (D) F 分布
- 9 Kaplan-Meier method 主要用於估計下列何種曲線？
(A)治療效果曲線 (B) ROC 曲線 (C)存活率曲線 (D)疾病盛行率曲線
- 10 在計算下列四種指標時，如果同時算出其信賴區間，試問，有幾個指標的區間可用來判定該變數是否具有統計上的顯著意義？①Odds ratio ②Relative risk ④Hazard ratio ④Regression coefficient
(A) 1 個 (B) 2 個 (C) 3 個 (D) 4 個
- 11 在 Kaplan-Meier 曲線中，出現下列那一種狀況時，曲線就會下降？
(A)有事件發生 (B)有設限資料產生
(C)有受試者退出研究，導致樣本數減少 (D)研究期間終了
- 12 在 Cox proportional hazards model 中所計算得到的 Hazard ratio，其意義為下列何者？
(A)事件發生的相對風險 (B)事件發生的實際機率
(C)存活機率的比值 (D)存活時間的比值
- 13 關於分層抽樣和集群抽樣之特性，下列敘述何者正確？
(A)兩者均為減少抽樣誤差
(B)兩者均為降低抽樣成本
(C)前者在減少抽樣誤差、後者在降低抽樣成本
(D)前者在降低抽樣成本、後者在減少抽樣誤差
- 14 下列那個方法所使用到的自由度，通常不會是整數？
(A) Two-sample t 檢定，假設母群體變異數不相等時
(B)計算三組平均值是否相等時，所採用的 ANOVA 的組內自由度
(C)計算線性迴歸時，檢定整個模型的 ANOVA 的誤差（error，或說殘差 residual）自由度
(D) Paired t 檢定，但有部分樣本只有前測資料、沒有後測資料時

- 15 在下列係數中，有幾個係數會具有特定的下界（最小值）或上界（最大值）？①迴歸係數 ②簡單線性迴歸的決定係數 ③變異係數 ④相關係數
(A) 0 個 (B) 1 個 (C) 2 個 (D) 3 個
- 16 將某數列的每一個數值都加上同一個常數之後，再重新計算下列指標，請問有幾個指標的數值不會改變？①平均數 ②四分位距 ③標準差 ④變異係數
(A) 0 個 (B) 1 個 (C) 2 個 (D) 3 個
- 17 下列這些類別資料的統計方法中，何者的計算過程不會算出卡方值？
(A) Test of independence (B) Test of goodness-of-fit
(C) McNemar's test (D) Fisher's exact test
- 18 羅吉斯迴歸 (Logistic regression) 主要用於預測何種類型的依變數？
(A) 連續型變數 (B) 二元類別型變數 (C) 時間序列資料 (D) 疾病發生率
- 19 一個研究若要偵測出較小的效應量，相較於偵測較大的效應量，需要增加樣本數還是減少樣本數？
(A) 增加樣本數 (B) 減少樣本數
(C) 維持不變，樣本數與效應量無關 (D) 隨機選擇
- 20 有兩人欲檢定男生和女生體重的變異數是否相同，在套用 F 檢定的公式時，其中一人將男生當成第一組、將女生當成第二組；另一人則恰好相反。假設男生和女生的樣本數與變異數均不相等，試問下列敘述何者正確？
(A) 兩人所計算出來的 F 值互為倒數
(B) 兩人檢定的結論可能會不相同
(C) 兩人所需要查詢的 F 檢定右尾臨界值是相同的
(D) F 右尾臨界值為 $F_{df1, df2, 1-\alpha}$
- 21 根據中央極限定理，當樣本數夠大時，樣本平均數的抽樣分布會趨近於何種分布？
(A) 卜瓦松分布 (B) 卡方分布 (C) 二項式分布 (D) 常態分布
- 22 若某疾病的盛行率為 1%，某檢測的敏感度為 90%，特異度為 95%，則檢測結果為陽性時，受檢者罹患該疾病的機率（陽性預測值）約為多少？
(A) 約 25.3% (B) 約 95% (C) 約 15.4% (D) 約 84.7%
- 23 關於二項式分布 (Binomial distribution) 與卜瓦松分布 (Poisson distribution) 的描述，下列敘述何者正確？
(A) 卜瓦松分布的期望值等於其變異數
(B) 當二項式分布的試驗次數 n 很小，且成功機率 p 很大時，二項式分布可以近似於卜瓦松分布
(C) 卜瓦松分布是常態分布的一個特例
(D) 二項式分布是由試驗次數 n 與平均次數 λ 來決定
- 24 關於抽樣分布 (Sampling Distribution) 的敘述，下列何者錯誤？
(A) 抽樣分布是從同一母群體中抽取所有可能樣本，並計算每個樣本的統計量後，這些統計量所形成的分布
(B) 抽樣分布的標準差稱為標準誤 (Standard Error)
(C) 當樣本數足夠大時，樣本平均數的抽樣分布會趨近於常態分布
(D) 樣本數越大，抽樣分布的變異性越大
- 25 在統計假設檢定中，P 值 (P-value) 的意義為何？
(A) 在虛無假設為真時，觀察到比目前統計量數值更極端的機率
(B) 拒絕虛無假設的機率
(C) 接受虛無假設的機率
(D) 犯下第二型錯誤的機率
- 26 在規劃一項研究時，影響樣本數估計的主要因素不包含下列何者？
(A) 顯著水準 (significance level) (B) 檢定力 (power)
(C) 效應量 (effect size) (D) 信賴區間 (confidence interval)
- 27 關於信賴區間 (Confidence Interval) 的敘述，下列何者錯誤？
(A) 在其他條件不變下，樣本數越大，信賴區間的寬度越窄
(B) 在其他條件不變下，信賴水準越高，信賴區間的寬度越窄
(C) 當我們說母群體平均數的 95% 信賴區間時，表示此區間有 95% 的機率可以涵蓋到母體平均數
(D) 研究者通常希望信賴區間越窄越好，以提供更精確的母體參數估計
- 28 若已知全國小三年級學童母群體平均身高為 135 公分，欲檢定某地區學童的平均身高是否異於全國，應使用何種統計方法？
(A) 獨立樣本 t 檢定 (B) 單樣本 t 檢定 (C) 配對樣本 t 檢定 (D) One-Way ANOVA 檢定

- 29 比較同一群病患在服用新藥前後血壓的差異，應使用何種檢定？
(A)配對樣本 t 檢定 (B)單樣本 t 檢定 (C)獨立樣本 t 檢定 (D) One-Way ANOVA 檢定
- 30 進行變異數分析時，若 F 檢定結果顯著，通常會進一步使用何種方法來分析各組間的差異？
(A)配對樣本 t 檢定 (B)相關分析 (C)卡方檢定 (D)事後多重比較
- 31 某藥廠宣稱其新開發的減肥藥能使體重至少減少 3 公斤。為驗證此說法，研究者隨機選取了 25 名志願者服用該藥，一個月後測量他們的體重變化。結果顯示平均體重減少了 2.5 公斤，樣本標準差為 1 公斤。請問在顯著水準 $\alpha=0.01$ 下，是否足以推翻藥廠的宣稱？（提示：此為單尾檢定，自由度為 24 時，t 值為 2.492 時 P 值=0.01）
(A)足以推翻，因為計算出的 t 值絕對值大於 2.492
(B)不足以推翻，因為計算出的 t 值絕對值小於 2.492
(C)足以推翻，因為樣本平均值低於 3 公斤
(D)無法判斷，因為缺乏母群體標準差
- 32 某研究中，欲比較三組病患（接受不同飲食療法）的焦慮程度分數。由於焦慮程度分數的資料分布嚴重偏斜，且樣本數相對較小，不符合常態分布假設。在這種情況下，下列那種統計方法最為適合用來檢定三組之間的差異？
(A)單因子變異數分析（One-way ANOVA）
(B)配對樣本 t 檢定（Paired-samples t test）
(C) Kruskal-Wallis test
(D) Wilcoxon rank sum test
- 33 某研究團隊想了解吸菸是否會影響肺功能。他們收集了 100 位吸菸者和 100 位非吸菸者的肺功能數據，並使用 t 檢定來比較兩組的平均肺功能。請問下列那一項是此研究中 t 檢定的虛無假設（null hypothesis）？
(A)吸菸者的平均肺功能低於非吸菸者 (B)吸菸者的平均肺功能高於非吸菸者
(C)吸菸者的平均肺功能與非吸菸者沒有差異 (D)吸菸會影響肺功能
- 34 關於配對資料（Paired Data）的分析方法與其配對的意義，下列敘述何者正確？
(A)配對資料分析通常用於比較兩個獨立樣本的差異
(B)配對資料的目的是為了增加樣本數，從而提高統計檢定力
(C)在配對資料分析中，主要關注的是同組配對資料的差異情形
(D)進行配對樣本 t 檢定時，如果樣本數足夠大，則不需要考慮同配對組的差異值是否符合常態分布
- 35 在迴歸分析中，判定係數 R 平方（R-squared）代表什麼？
(A)迴歸模型中的殘差平方和
(B)依變數的總變異中，可以由模型中自變數解釋的比例
(C)衡量自變數對依變數的影響強度
(D)迴歸模型的適合度，值越接近 0 越好
- 36 在簡單線性迴歸中，迴歸係數（slope）的意義為何？
(A)當自變數為零時，依變數的估計值
(B)依變數每增加一個單位，自變數的平均變化量
(C)自變數每增加一個單位，依變數的平均變化量
(D)依變數的總變異量中可由自變數解釋的比例
- 37 某研究團隊調查了體重（公斤）和跑步速度（公尺/秒）之間的關係。計算得出體重與跑步速度之間的皮爾森相關係數為-0.75。請問下列那一個敘述最能正確描述此相關關係？
(A)體重增加，跑步速度也會隨之增加 (B)體重增加，跑步速度傾向於減少
(C)體重與跑步速度之間沒有線性關係 (D)體重解釋了跑步速度變異的 75%
- 38 某研究建立了包含交互作用項的線性迴歸模型，用於推估兒童的發育指數（Developmental Index, DI），模型如下： $DI=70+0.5 \times \text{年齡（歲）}+2 \times \text{營養補充劑（0=無，1=有）}+0.1 \times \text{年齡} \times \text{營養補充劑}$ 。根據此模型，請問一位 5 歲的兒童，若有接受營養補充劑，其預期的發育指數（DI）為多少？
(A) 72.5 (B) 77.5 (C) 75.0 (D) 82.5
- 39 在存活分析中，關於「設限資料（censored data）」的描述，下列何者錯誤？
(A)設限資料會發生在研究結束時，個案尚未發生事件，但已無法繼續追蹤
(B)設限資料在分析時會被完全排除，不納入存活時間的計算
(C)失去追蹤的受試者，其事件發生時間可能在觀察期結束後，也可能在研究期間但未能被觀察到，皆可視為設限資料
(D)設限資料的存在是存活分析與其他統計分析方法的重要區別之一
- 40 判讀統計軟體輸出結果時，除了 P 值外，還應關注那些指標以評估統計顯著性及臨床意義？
(A)自由度與樣本數 (B)效應量與信賴區間 (C)標準差與變異數 (D)眾數與中位數