

單選題 (每題 5 分, 共 35 分)

- 下列有關假設檢定 (hypothesis testing) 的敘述, 何者正確?
 - 型一誤差是當 H_1 (對立假說) 為真時錯誤接受 H_0 (虛無假說) 的機率
 - 當型一誤差增加時, 型二誤差也隨之增加
 - 在顯著水準為 0.1 之下的假設檢定結果和用 95% 信賴區間判斷的結果, 兩者一定會一致
 - 當顯著水準為 0.01 時, 表示我們所能允許錯誤地拒絕 H_0 的機率為 1%
 - P 值為 0.051 代表對立假說是錯誤的
- 下列何者不是中央趨勢 (central tendency) 的指標?
 - 幾何平均 (geometric mean)。
 - 調和平均 (harmonic mean)。
 - 裁剪平均 (truncated mean)。
 - 均方根 (root mean square)。
 - 平均差 (mean absolute deviation)。
- 下列有關變異係數 (coefficient of variation), 下述何者錯誤?
 - 其值無單位。
 - 其值介於 0 與 1 間。
 - 將原始數據乘以 5, 變異係數不改變。
 - 樣本數增加趨近無限時, 樣本變異係數 (sample coefficient of variation) 趨近族群變異係數 (population coefficient of variation)。
 - 將原始數據除以樣本平均值後再計算樣本標準差, 即為樣本變異係數。
- X 及 Y 表示隨機變數 (random variable), a 及 b 表示常數 (constant), E 表示期望值 (expected value), Var 表示變異數 (variance), Cov 表示共變異數 (covariance)。下列等式何者錯誤?
 - $E(a+X^2)=a+E^2(X)+Var(X)$
 - $E(X+XY)=E(X)+E(X)E(Y)+Cov(X, Y)$
 - $Var(a-X)=Var(X)$
 - $Var(X-Y)=Var(X+Y)=Var(-X+Y)=Var(-X-Y)$
 - $Cov(X+aY, bX-a)=bVar(X)+abCov(X, Y)$
- 樣本數增加趨近無限時, 下述何者錯誤?
 - 樣本平均值 (sample mean) 的分佈, 趨近常態分佈 (normal distribution)。
 - 樣本平均值趨近族群平均值 (population mean)。
 - 樣本變異數 (sample variance) 的分佈, 趨近自由度為 1 的卡方分佈 (chi-square distribution)。
 - 樣本變異數趨近族群變異數 (population variance)。
 - 樣本平均值的標準誤 (standard error of the mean) 趨近 0。
- 有關機率分佈, 下述何者錯誤?
 - 將自由度為 n 的卡方分佈開平方, 可得變異數為 n 的常態分佈。
 - 不論分子自由度及分母自由度為何, F 分佈 (F distribution) 曲線下之總面積皆為 1。
 - 不論期望值為何, 布阿松分佈 (Poisson distribution) 皆為右偏 (positively skewed) 分佈。
 - 相等期望值下, 布阿松分佈的變異數比二項式分佈 (binomial distribution) 的變異數較大。
 - 相等期望值下, 二項式分佈的變異數比伯努立分佈 (Bernoulli distribution) 的變異數較大。

7. 相同樣本數下，下列哪個抽樣調查(survey sampling)方法，有最小的變異數？
- (A) 簡單隨機抽樣(simple random sampling)。
 - (B) 分層抽樣(stratified sampling)。
 - (C) 集束抽樣(cluster sampling)。
 - (D) 系統性抽樣(systematic sampling)。
 - (E) 立意抽樣(purposive sampling)。

計算問答題 (共 65 分)

8. 某一病例對照研究欲調查口服避孕藥是否影響日後罹患子宮內膜癌的風險。研究數據顯示，在 120 位子宮內膜癌患者中，有 8 位曾經使用過口服避孕藥，而對照組的 300 位參與者中，有 10 位曾經使用過同款的口服避孕藥。
- (1) 請問有使用與沒使用口服避孕藥相比，其罹患子宮內膜癌的勝算比 (odds ratio) 為何？(2 分)
 - (2) 承(1)，此勝算比的 95%信賴區間為何？(5 分)
 - (3) 請進行適當的假設檢定，在顯著水準為 0.05 的情況下，檢視「曾經使用口服避孕藥是否會增加子宮內膜癌的風險」，並提出你的結論。(8 分)
9. 某醫院欲比較一新開發的電腦自動化診斷方法與實際的醫師診斷對於判別人體是否受到病毒感染有無存在顯著差別。蒐集的資料顯示，在 5,000 位來醫院就醫的民眾裡，自動化診斷方法找出 45 位受到特定病毒感染，而醫師判讀有 32 位受到感染，其中 28 位民眾被兩種方法皆判定為感染者。
- (1) 請問兩種診斷方法的一致性(%)為何？(3 分)
 - (2) 此類資料型態適用哪一種檢定進行假設檢定？(4 分)
 - (3) 以(2)中作答的檢定，在顯著水準為 0.05 的情況下，針對「病毒感染的診斷是否會受診斷方法所影響」進行假設檢定，並提出你的結論。(8 分)
10. 你拿到一組資料，蒐集了某降血壓新藥不同劑量 X_1 (1=等級 1, 2=等級 2, 3=等級 3, 4=等級 4, 5=等級 5)和高血壓 Y 的數據(樣本平均數 120, 樣本標準差 10)，共有 108 筆數據。此研究另外蒐集了可能的影響因子年齡 X_2 (30-79)和性別 X_3 (0 為男性, 1 為女性)。你希望利用迴歸分析了解新藥對血壓是否有影響。又過去的經驗認為不同的性別之下，年齡對血壓的影響並不相同，因此建模時必須考慮性別和年齡的交互作用(interaction)。
- (1) 請根據以上敘述寫出適當的迴歸模型。在你的模型之下，50 歲女性接受等級 3 的藥物治療後的期望高血壓為何？(7 分)
 - (2) 在你的模型之下，針對[性別和年齡無交互作用]的虛無假說為何？此假說的檢定統計量的虛無分配為何？(7 分)
 - (3) 在你的模型之下，針對[等級 1 和等級 2 的劑量對高血壓的影響無顯著差異]的虛無假說為何？此假說的檢定統計量的虛無分配為何？(7 分)
 - (4) 在你的模型之下，針對[劑量對高血壓無影響]的虛無假說為何？此假說的檢定統計量的虛無分配為何？(7 分)
 - (5) 若此模型的判定係數(coefficient of determination)為 $R^2 = 0.6$ ，則校正判定係數(adjusted- R^2)為多少？(7 分)

試題隨卷繳回