

※ 注意：全部題目均請作答於試卷內之「非選擇題作答區」，請標明題號依序作答。

1. 是非簡答題 (本大題共 25 分，每小題 5 分)：

下列敘述，對的請打○，錯的請打×，若敘述為錯者請說明理由 (理由佔該小題 4 分)：

- (1) 兩個現象或變數間若存在著很強的因果關係，則其相關係數 (correlation coefficient) 不可能為零。
- (2) 信賴區間與假說檢定對於拒絕虛無假說與否的結論必定相符。
- (3) 由一樣本所求得的 95% 信賴區間有 95% 的機率會包含母體的真實值。
- (4) 在對列聯表資料作卡方檢定時，必須遵循 Rule of Five 的原則，亦即列聯表中每格觀察次數不得小於 5。
- (5) P 值 (P -value) 為在對立假說下，得到比所觀察到的檢定統計量 (test statistic) 更極端的機率。

複選題：

2. (5分) 關於假說檢定的敘述何者正確?

- (A) 型一誤差和檢定力相加為一定數
- (B) 在固定樣本數下，吾人可以任意決定一檢定之型一誤差和檢定力的值
- (C) 在固定樣本數下，型一誤差越小，則檢定力傾向越小
- (D) 在固定型一誤差下，樣本數越大，則檢定力傾向越大
- (E) 檢定力為1的檢定並不存在

3. (5分) 關於 95% 信賴區間 (confidence interval) 的敘述何者正確?

- (A) 此信賴區間的長度會比 90% 信賴區間的長度來的短
- (B) 有 95% 的機會包含真實的參數
- (C) 隨機挑選一參數，則有 95% 的機會落在信賴區間裡
- (D) 亦可用來執行假設檢定，其型一誤差為 0.05
- (E) 樣本數越多，則信賴區間會愈短

4. (5分) 給定反應變數 Y 和解釋變數 $X = (X_1, \dots, X_p)$ ，關於迴歸分析的敘述何者正確?

- (A) 解釋變數的個數越多，則判定係數 (R^2) 越大
- (B) 解釋變數的個數越多，則校正的判定係數 (adjusted R^2) 越大
- (C) 解釋變數之間的相關性越強，則迴歸係數的估計越不穩定
- (D) 在 $p = 1$ 的情況下，若 X 和 Y 的相關係數絕對值越大，則判定係數 (R^2) 越大
- (E) 在 $p = 1$ 的情況下，判定係數 (R^2) 為 0.49 代表 X 和 Y 的相關係數為 0.7

計算題：

5. (10分) 欲檢定銅板的公正性，考慮假說檢定 $H_0: p = 0.5$ vs. $H_1: p \neq 0.5$ ，其中 p 為銅板出現正面的機率。令 X 為丟 9 次銅板出現正面的次數。若 $X \geq 8$ 或 $X = 0$ ，則拒絕虛無假說。
- (1) 請計算該檢定的型一誤差。
 - (2) 請計算該檢定在 $p = 0.8$ 下的檢定力。

問答題

6. 在流行病學因果關係的推論上，Hill's criteria 非常重要。如果想證明「抽菸與肺癌發生的關係」，請說明可能的研究設計、變項、測量方法，並請逐一說明因果判定有哪些必要基準與次要的基準。(15分)
7. 某人想針對「帶有 A 基因是否較容易發生孩童時期氣喘病」這個題目進行研究設計，請說明可能會出現的干擾因子 (confounder) 與修飾因子 (modifier) 各有哪些？該如何區分二者？並請說明該如何處理或分析干擾因子與修飾因子？(10分)

簡答題

8. 請列出流行病學中三種常見的偏差 (bias) (6分)
9. 混淆因子的定義為何 (3分)? 混淆因子 (confounder) 是系統誤差 (systematic error) 或隨機誤差 (random error) (2分)? 請列出三種控制混淆因子的方法 (6分)
10. 隨機誤差可以用何種估計量評估 (2分)? 請解釋此估計量所代表的意義 (3分)? 隨著研究樣本數的增加，系統誤差及隨機誤差分別會受到什麼樣的影響 (3分)?

試題隨卷繳回