

※ 注意：請於試卷內之「選擇題作答區」依序作答。

單選題（每題 5 分，共 40 分）

1. 將一組共 800 人的收縮壓資料 X （單位：mmHg），對每個人做轉換 $Y=(X-5)/2$ 。下述何者錯誤？
 - (A) Y 的平均值為（原平均值-5）/2
 - (B) Y 的標準差(standard deviation)為原標準差的一半
 - (C) Y 的變異數(variance)為原變異數的四分之一
 - (D) Y 的四分位差(interquartile range)為原四分位差的一半
 - (E) Y 的變異係數(coefficient of variation)比原先小
2. 在固定樣本數與效應量(effect size)不變的情況下，若將顯著水準 α 由 0.05 改為 0.01，且仍以「 $p < \alpha$ 即拒絕虛無假說」為規則，下述何者錯誤？
 - (A) 拒絕虛無假說會變得更困難（門檻更嚴格）
 - (B) 在虛無假說為真時，拒絕虛無假說的機率會下降
 - (C) 檢力(power)通常會下降
 - (D) 犯型二誤差(β)的機率通常會上升
 - (E) p 值會因為 α 改變而改變
3. 令 X, Y 為隨機變數(random variables)， a, b 為常數(constants)， $E(\cdot)$ 為期望值(expected value)， $\text{Var}(\cdot)$ 為變異數(variance)， $\text{Cov}(\cdot, \cdot)$ 為共變異數(covariance)。下列等式何者錯誤？
 - (A) $E(aX+bY)=aE(X)+bE(Y)$
 - (B) $\text{Var}(a+bX)=b^2\text{Var}(X)$
 - (C) $E(XY)=E(X)E(Y)$
 - (D) $\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X, Y)$
 - (E) $\text{Cov}(X, a)=0$
4. 某母體的平均數為 μ 、變異數為 σ^2 。研究者從此母體以相同方式、彼此獨立地重複抽取許多次大小皆為 n 的樣本，每次計算樣本平均數。下列何者最能正確描述樣本平均數的抽樣分佈(sampling distribution)之概念？
 - (A) 單一次抽樣得到的 n 個原始觀測值
 - (B) 單一次抽樣得到的樣本平均數
 - (C) 母體參數 μ 在不同樣本中的變動分佈
 - (D) 在固定 n 下，重複抽樣所得到的樣本平均數之機率分佈
 - (E) 針對 n 個原始觀測值取後放回抽樣(sampling with replacement)，重新抽取

見背面

出的新樣本

5. 急診室某類外傷到診人數可近似服從布阿松 (Poisson) 分佈，平均每小時到診數為 $\lambda=3$ 。在任一小時內，外傷到診人數「至少 1 人」的機率為何？
- (A) e^{-3}
(B) $1-e^{-3}$
(C) $1-e^{-1}$
(D) e^{-1}
(E) $1-e^{-3} - 3e^{-3}$
6. 下列何者最正確描述伯努利 (Bernoulli) 分佈、二項分佈(binomial)與布阿松分佈之間的關係？
- (A) 二項分佈是多次獨立伯努利試驗成功次數的分佈；且當 $n \rightarrow \infty$ 、 $p \rightarrow 0$ 、並使 $np \rightarrow \lambda$ 時，Binomial(n, p)可近似為 Poisson(λ)
(B) 伯努利分佈可視為布阿松分佈在 $\lambda=1$ 時的特例；二項分佈則是多個布阿松隨機變數的和
(C) 布阿松分佈可視為二項分佈在 $p=0.5$ 時的近似；伯努利分佈則是二項分佈在 $n=2$ 的特例
(D) 伯努利分佈描述每次試驗的失敗次數；二項分佈描述第一次成功前的失敗次數；布阿松分佈描述固定次數試驗中的成功次數
(E) 若 $X \sim \text{Binomial}(n, p)$ ，則 X 必可表示為 n 個相互獨立、參數皆為 $\lambda=p$ 的布阿松隨機變數之和
7. 在進行單樣本平均數的假說檢定時，若將顯著水準 (Significance level, α) 從 0.05 縮減為 0.01，在其他條件不變的情況下，下列敘述何者正確？
- (A) 型一錯誤 (Type I error) 的風險增加
(B) 檢定力 (Statistical power) 會增加
(C) 拒絕虛無假設 (H_0) 的難度增加
(D) 信賴區間的寬度會變窄
8. 某研究探討新型降血壓藥物的效果，得到 95% 信賴區間為 $[-2 \text{ mmHg}, 15 \text{ mmHg}]$ 。關於此結果，下列詮釋何者最為精確？
- (A) 有 95% 的把握該藥物具有顯著降壓效果
(B) 若虛無假設為「效果 = 0」，則在 $\alpha = 0.05$ 下無法拒絕虛無假設

接次頁

- (C) 該研究有 95% 的機率涵蓋了樣本平均數
- (D) 該藥物在 95% 的受試者身上都有療效

複選題 (每題 5 分, 共 20 分)

9. 下列有關迴歸分析(regression analysis)的描述, 何者有誤?
- (A) 若解釋變數(explanatory variable) X_1 的迴歸係數為正, 代表 X_1 與反應變數(response variable) Y 有正相關。
 - (B) 只要建立好迴歸模型, 就可以說明解釋變數(explanatory variable)與反應變數(response variable)之間有因果關係。
 - (C) 迴歸模型可以用來說明解釋變數(explanatory variable)與反應變數(response variable)之間變化的線性相關性。
 - (D) 判定係數(coefficient of determination) R^2 是用來測量模型的解釋能力, 越接近 1 代表模型的解釋能力越佳。
10. 請問, 在解釋迴歸分析的結果時, 下列敘述何者有誤?
- (A) t 檢定(t test)可以檢驗特定迴歸係數是否為 0, 若檢定結果推翻虛無假說(null hypothesis), 就可以說此係數所代表的解釋變數(explanatory variable)與反應變數(response variable)具有因果關係。
 - (B) F 檢定(F test)可以檢定迴歸模型中的有所迴歸係數是否同時為 0, 用以推論此迴歸模型是否有用, 若檢定結果推翻虛無假說(null hypothesis), 就可以說模型中的所有解釋變數(explanatory variable)與反應變數(response variable)之間皆存在相關性。
 - (C) 若迴歸模型中放入性別 (男性為 0, 女性為 1), 所估計的迴歸係數 β_{sex} , 解釋意義為男性相較與女性, 反應變數(response variable)的期望值會多 β_{sex} 。
 - (D) 若模型中放數入體重 (以每一公斤為單位), 此變數前的迴歸係數 β_{wt} 所代表的意義是, 解釋變數(explanatory variable)每增加一公斤, 反應變數(response variable) 會增加 β_{wt}
11. 下列有關線性迴歸(linear regression)的建構, 何者有誤?
- (A) 模型需假設解釋變數(explanatory variable)服從常態分配(normal distribution), 意即 $X \sim N(\mu_X, \sigma^2)$ 。
 - (B) 模型有變異數同質的假設, 也就是說每筆觀察值要有相同的變異數 σ^2 。

見背面

- (C) 線性迴歸分析(linear regression analysis)沒有樣本數與變數個數的限制，所以即便變數個數多於樣本數，也可以使用一般的線性迴歸模型。
- (D) 線性迴歸模型(linear regression model)中迴歸係數，可以利用最小平方法(least squared estimates, LSE)進行估計。

12. 下列不同的研究設計或研究問題中，哪些不應該使用線性迴歸分析(linear regression analysis)?

- (A) 某長照研究者，分析政府長照服務使用資料庫的資料，想探討政府提供的五種喘息服務的對減輕照顧者的照顧負荷(為一透過量表測量的負荷壓力分數)(Y)的影響程度為何(註：資料為歸人檔，一位被照顧者會化整為一筆紀錄)。
- (B) 某研究者針對懷孕婦女進行一前瞻性世代研究(prospective cohort study)，想探討懷孕婦女的飲食習慣、運動習慣、健康行為對懷孕期間血糖(Y)變化的影響，血糖會透過儀器進行連續性血糖監測(Continuous Glucose Monitoring, CGM)，為一連續數值，會透過特殊的統計模型將懷孕 10 個月的血糖數值轉為一個分數，解釋變數(explanatory variable)也都轉為一個代表數值，因此每一位孕婦只有一筆資料。
- (C) 某研究者想探討某藥物治療(X)對疾病復發與否(Y)是否有影響，從醫院的資料庫撈取病人的臨床資訊(如：年齡、性別、治療處置、生化檢查值等)進行研究分析。
- (D) 某研究者進行一配對病例對照研究(matched case-control studies)設計，想探討憂鬱症患者發病時與健康對照組的人，每週平均的生理活動量(此為一連續數值)(Y)是否有差異，此差異會受哪些因素影響。

問答題 (共 40 分)

※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

13. (15%) 某研究想了解睡眠時間長短(小時)與身高(公分)的關係，現搜集 40 筆資料進行簡單線性迴歸分析，下表為迴歸分析中 ANOVA 的表

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	4114.328	4114.328	28.270	<0.0001
殘差	38	5530.360	145.536		
總和	39	9644.688			

接次頁

迴歸分析的另一部分結果見下表

	係數	標準誤	95%信賴區間	t 統計	P-值
截距	106.190	11.026	(83.87, 128.51)	9.631	<0.0001
睡眠時間	8.208	1.544	(5.08, 11.33)	5.317	<0.0001

- (1) 請寫出迴歸模型並用迴歸分析的結果解釋睡眠時間長短與身高的關係。(5分)
 - (2) 請問睡眠時間長短與身高的相關係數是多少?(5分)
 - (3) 請解釋睡眠時間的迴歸係數的95%信賴區間的意義。(5分)
14. 在一項關於某種新興傳染病疫苗效果的世代研究 (Cohort Study) 中，研究人員追蹤了兩組參與者：
- 疫苗組 (Vaccinated): 200 人，追蹤期間有 10 人感染。
 - 對照組 (Unvaccinated): 200 人，追蹤期間有 40 人感染。
- (1) 請進行假說檢定，設定虛無假設 (H_0) 與對立假設 (H_1)，使用常態逼近法 (Z-test for two proportions) 計算檢定統計量，並判斷在 $\alpha = 0.05$ 下，兩組感染率是否有顯著差異?(5分)
 - (2) 請計算疫苗組對相比於對照組的相對風險(RR)與其95%信賴區間(已知 $Z_{0.975} \approx 1.96$)。(5分)
 - (3) 根據假說檢定與信賴區間的結果，請問兩者結果是否一致，並請判斷疫苗是否具有顯著的保護效果?(2分)
15. 某流行病學調查使用一種快速篩檢試劑檢測 100 名已知患有 A 疾病的患者，結果有 80 人呈現陽性。
- (1) 請計算該試劑的敏感度 (Sensitivity) 及其 95% 信賴區間 (提示：使用常態逼近法)。(4分)
 - (2) 若此疾病在社區中的盛行率 (Prevalence) 極低，即使該試劑的敏感度與特異度都很高，其陽性預測值 (PPV) 通常會偏高還是偏低?這對公共衛生篩檢政策有何啟示?(3分)
16. 在一項參與者多達 10 萬人的大型世代研究 (Cohort study) 中，發現某種飲食習慣會降低 0.5% 的糖尿病風險，統計檢定結果為 $p\text{-value} < 0.001$ 。請討論：
- (1) 為什麼在極大樣本下容易出現「統計上顯著 (Statistically significant)」但「臨床上不重要 (Clinically insignificant)」的現象?(4分)
 - (2) 作為流行病學家，你會如何整合信賴區間來解讀此結果?(2分)

試題隨卷繳回