

※ 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。

是非題（每題 2 分）

考慮兩個迴歸模型 $M1: Y = \beta_0 + X_1\beta_1 + \varepsilon_1$, $\varepsilon \sim N(0, \sigma^2)$ 以及 $M2: Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ ，並使用 LSE (least square estimate) 進行統計推論。

1. 在 M1 下，若 σ^2 越小，則 LSE 對 β_1 的估計越精準。
2. 在 M1 下，若 X_1 的變異數越小，則 LSE 對 β_1 的估計越精準。
3. 在 M2 下，若 X_1 和 X_2 的相關係數(Pearson's correlation coefficient)的絕對值越大，則 LSE 對 β_1 的估計越精準。
4. M1 的 R^2 (coefficient of determination)大於 M2 的 R^2 。
5. M1 的 SST(total sum of squares)大於 M2 的 SST。
6. M1 殘差(residual)的平均數大於 M2 殘差的平均數。

單選題（每題 5 分）

7. 有關標準差(standard deviation)及標準誤(standard error)，下列何者錯誤？
 - (A) 樣本標準差(sample standard deviation)是族群標準差(population standard deviation)的一致性估計量(consistent estimator)。
 - (B) 樣本標準差越大，表示族群的散佈程度(dispersion)越大。
 - (C) 標準誤越大，表示抽樣分佈(sampling distribution)的散佈程度(dispersion)越大。
 - (D) 樣本數(sample size)增加為 4 倍，標準誤減半。
 - (E) 將每個樣本的數據皆乘以 4，樣本標準差數值不變。
8. 有關樣本平均值(sample mean)的信賴區間(confidence interval)，下列何者錯誤？
 - (A) 族群標準差越大，信賴區間越寬。
 - (B) 標準誤越大，信賴區間越寬。
 - (C) 樣本數相同時，樣本標準差越大，信賴區間越寬。
 - (D) 根據族群標準差所求得之信賴區間，其平均寬度不受樣本數影響。
 - (E) 99%信賴區間比 95%信賴區間較寬。
9. 有關布阿松分佈(Poisson distribution)，下列何者錯誤？
 - (A) 適用於稀有疾病(rare disease)。因此某族群某病發生人數的期望值若大於 30，布阿松分佈即不適用。
 - (B) 為離散分佈(discrete distribution)。
 - (C) 為右偏分佈(positively skewed distribution)。

見背面

- (D) 給定期望值(expected value)，其機率分佈(probability distribution)即可完全決定。
- (E) 將兩個獨立(independent)，但期望值不同的布阿松分佈相加，仍為布阿松分佈。
10. 有關伯努力分佈(Bernoulli distribution)和二項式分佈(binomial distribution)，下述何者錯誤？
- (A) 皆為離散分佈。
- (B) 可能出現數值為正整數(含零)，不可能出現負值。上限分別為1(伯努力分佈)及n(二項式分佈)。
- (C) 伯努力分佈的期望值即為出現1的機率。
- (D) n個獨立且期望值相同的伯努力分佈相加，為二項式分佈。
- (E) 二項式分佈除以n，為伯努力分佈。
11. 有關常態分佈(normal distribution)，下述何者錯誤？
- (A) 期望值=中位數(median)=眾數(mode)。
- (B) 將兩個獨立，但期望值不同、變異數(variance)也不同的常態分佈相加，仍為常態分佈。
- (C) 將兩個期望值相同，但變異數不同的常態分佈以等機率混合(mixture)，仍為常態分佈。
- (D) 將期望值為0、變異數為1的常態分佈取平方，為卡方分佈(chi-square distribution)。
- (E) 給定期望值及變異數，常態分佈的機率密度函數(probability density function)即可完全決定。
12. 有關盒鬚圖(box-and-whisker plot)，下述何者錯誤？
- (A) 可呈現中位數。
- (B) 可呈現四分位差(interquartile range)。
- (C) 可呈現強影響點(influential point;或稱槓桿點，leverage point)。
- (D) 可呈現離群點(outlying point)。
- (E) 統計套裝軟體可自動繪製，不需人工計算。
13. 下列有關假設檢定(hypothesis testing)的敘述，何者錯誤？
- (A) 若想對同一群人檢測用藥前後的血壓是否有差異，可以針對血壓值進行配對樣本檢定(paired t-test)。
- (B) 兩獨立樣本t檢定(two-sample t-test)可以用於檢定兩個比例(proportion)的差異。
- (C) 變異數分析(ANOVA)可以用於檢定多組相互獨立樣本的母體平均數是否具有顯著差異。
- (D) 95%信賴區間(95% confidence interval)也是檢定方法的一種。
- (E) 假設檢定中，若顯著水準(significance level)設定為0.01，代表我們所能允許錯誤地拒絕虛無假說的機率為1%。

計算問答題

14. 某醫院針對 100 位重度肥胖者 ($BMI > 35$) 正實行一減重計畫。六個月後，30 位參與者成功減下原先體重的 20%。
- (1) 請問在這個樣本中，「六個月內成功減去原體重 20%」的比例為何？以下將這個比例簡稱為此減重計畫的成功率。(2 分)
 - (2) 請問此成功率之抽樣分佈 (sampling distribution) 會符合何種分配 (distribution)？依據為何？(3 分)
 - (3) 請計算此成功率之 95% 信賴區間 (confidence interval)，並解釋區間所代表的意義。(請詳列計算過程，並計算至小數點下第四位)(7 分)
 - (4) 假設我們已知這個減重計畫在全台灣的平均成功率為 20%，請問在這個樣本中所觀察到的成功率超過 30% 的機率為何？請計算至小數點下第四位。(5 分)
 - (5) 承上，這個樣本中所觀察到的成功率是否顯著高於全台灣平均值？請提出虛無假說、對立假說、顯著水準、檢定統計量、判讀方式與結論。(7 分)
 - (6) 在此題的假設檢定中，請定義型一錯誤 (type I error)。(2 分)
 - (7) 在此題的假設檢定中，請定義型二錯誤 (type II error)。(2 分)
 - (8) 請解釋假設檢定中 p 值 (p-value) 的意義為何。(2 分)
15. 你拿到一組資料，蒐集了不同人的出生地區 X_1 (1=北部，2=中部，3=東部，4=西部，5=南部) 和身高 Y 。此研究另外蒐集了可能的影響因子年齡 X_2 (30 歲-79 歲) 和性別 X_3 (0 為男性，1 為女性)。你希望利用迴歸分析了解出生地區是否對身高有影響。又過去的經驗認為不同的性別之下，年齡對身高的影響並不相同，因此建模時應當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，北部和南部對身高影響的差異是多少？(8 分)
 - (2) 在你的模型之下，一個 50 歲南部出生的男性，期望的身高是多少？(5 分)
 - (3) 在你的模型之下，針對[地區對身高無影響]的虛無假說為何？(5 分)
 - (4) 承上題，此假說的檢定統計量的虛無分配為何？(5 分)

試題隨卷繳回