

※ 本大題請於試卷內之「選擇題作答區」依序作答。

單選題(每題5分，共45分)

- 有關機率(probability)及機率密度(probability density)，下列何者錯誤？
 - 機率之數值介於0與1之間。
 - 機率密度之數值非負(nonnegative)。
 - 機率除以樣本數(sample size)即為機率密度。
 - 機率無單位(unit)。
 - 機率密度之單位為原隨機變數(random variable)之單位的倒數。
- 有關四分位差(interquartile range)，下列何者正確？
 - 其數值無單位。
 - 其數值非負。
 - 樣本數增加，其數值亦增加。
 - 樣本數增加，其數值變小。
 - 其數值越大代表鑑別力(resolution power)越大。
- 有關兩個變數間的相關係數(correlation coefficient)，下列何者錯誤？
 - 其值無單位。
 - 其值介於-1和1之間。
 - 兩變數同乘10，兩者間相關係數不變。
 - 一個變數維持不變，另一變數乘4，兩者間相關係數變為2倍。
 - 兩變數同乘-10，兩者間相關係數不變。
- 有關布阿松分佈(Poisson distribution)，下述何者錯誤？
 - 兩個獨立的(independent)布阿松分佈相加，其期望值(expected value)為原來兩個布阿松分佈期望值之相加。
 - 兩個獨立的布阿松分佈相加，其變異數(variance)為原來兩個布阿松分佈變異數之相加。
 - 兩個獨立的布阿松分佈相加，其期望值為原來兩個布阿松分佈變異數之相加。
 - 兩個獨立的布阿松分佈相加，其變異數為原來兩個布阿松分佈期望值之相加。
 - 兩個相依的(dependent)布阿松分佈相加，其變異數為原來兩個布阿松分佈間之共變異數(covariance)。
- 有關伯努力分佈(Bernoulli distribution)和二項式分佈(binomial distribution)，下述何者錯誤？
 - 兩個獨立但不相同的(non-identical)二項式分佈相加，其期望值為原來兩個二項式分佈期望值之相加。
 - 兩個獨立但不相同的二項式分佈相加，其變異數為原來兩個二項式分佈變異數之相加。
 - 三個獨立且相同的(identical)伯努力分佈相加，為二項式分佈。
 - 三個獨立但不相同的伯努力分佈相加，為三項式分佈(trinomial distribution)。

- (E) 三個獨立且相同的二項式分佈相加，為二項式分佈。
6. 有關常態分佈(normal distribution)，下述何者錯誤？
- (A) 兩個獨立的、期望值不同、變異數相同的常態分佈相加，為雙峯分佈(bimodal distribution)。
 - (B) 兩個獨立的、期望值相同、變異數不同的常態分佈相加，為對稱分佈(symmetric distribution)。
 - (C) 兩個獨立的、期望值及變異數皆相同的常態分佈相減，為常態分佈。
 - (D) 一個常態分佈乘上 10 再減 10000，為對稱分佈。
 - (E) 一個常態分佈平方後再乘上-1，為左偏分佈(negatively skewed distribution)。
7. 有關隨機樣本(random sample)的樣本平均值(sample mean)，下列敘述何者錯誤？
- (A) 樣本平均值為族群平均值(population mean)的不偏估計(unbiased estimate)。
 - (B) 樣本平均值為族群平均值的確切估計(exact estimate)。
 - (C) 樣本平均值為族群平均值的一致性估計(consistent estimate)。
 - (D) 樣本數增大 100 倍，樣本平均值的變異數減小 100 倍。
 - (E) 樣本數趨近無限大，樣本平均值的抽樣分佈(sampling distribution)趨近常態分佈。
8. 有關線性迴歸(linear regression)的敘述，下列何者錯誤？
- (A) 可以用來執行 ANOVA。
 - (B) 主要是針對連續型的反應變數使用。
 - (C) 在常態分配假設下，maximum likelihood estimate (MLE)等於 least squares estimate (LSE)。
 - (D) 解釋變數的個數越多，所得到的判定係數 R^2 越大。
 - (E) 解釋變數的個數越多，所得到的校正判定係數(adjusted- R^2)越大。
9. 有關線性迴歸模型 $Y = \beta_0 + X_1\beta_1 + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ ，下列何者正確？
- (A) 可以透過對 X_1 的線性轉換以改變 σ^2 的大小。
 - (B) 可以透過樣本數的增加以改變 σ^2 的大小。
 - (C) 若 X_1 的變異數越大，則對 β_0 所建立的信賴區間越窄。
 - (D) 若 σ^2 越大，則對 β_1 所建立的信賴區間越窄。
 - (E) 判定係數 R^2 (coefficient of determination)越大，代表 Y 和 X_1 的相關係數(Pearson correlation coefficient)越大。

※ 本大題請於試卷內之「非選擇題作答區」標明題號依序作答。

計算問答題(共 55 分)

10. 根據國民健康署某年度針對臺灣地區新生兒母乳哺育情況的調查，透過電話訪問 1600 位六個月以下新生兒的母親，其中有 720 位母親表示是純母乳哺育，480 位母親表示是混合母乳哺育(混合母乳定義為純母乳混合奶粉、果汁、葡萄糖水等食品)。根據上述調查，請問

接次頁

- (1)「六個月以下新生兒純母乳哺育」的比例為何？(4分)
- (2)請提出「六個月以下新生兒純母乳哺育比例」的95%信賴區間(confidence interval)，並解釋此區間的意義。(請詳列計算過程，並計算至小數點下第四位)(10分)
- (3)請問上一題在計算「六個月以下新生兒純母乳哺育比例的95%信賴區間」時，「六個月以下新生兒純母乳哺育比例」是服從哪一個分配(distribution)？為什麼？(4分)
- (4)請利用假設檢定的方式回答，臺灣地區「六個月以下新生兒純母乳哺育比例」是否高於全球平均值0.428？(請詳列計算過程並計算至小數點下第四位，同時寫清楚虛無假說、對立假說、顯著水準、檢定統計量、判讀方式與結論)(10分)
- (5)請指出在上一題的假設檢定中會犯的錯誤有哪些，並說明這些錯誤的定義。(4分)
- (6)什麼是假設檢定中的p值(p-value)？(3分)

11. 你拿到一組資料，蒐集了某新藥的劑量 X_1 (Placebo, Low, Mediate, High)和受試者服用過後疼痛的減緩程度 Y 。此研究另外蒐集了受試者的年齡 X_2 (30歲-79歲)和性別 X_3 ($X_3 = 0$ 為男性， $X_3 = 1$ 為女性)為可能的影響因子。你希望利用迴歸分析了解服用新藥是否對疼痛的減緩程度有幫助。過去的經驗認為年齡並不是一個線性的影響因子，所以建議用類別變數(categorical variable)來處理年齡的影響(分成5組：30-39、40-49、50-59、60-69、70-79)。

- (1)請根據以上敘述寫出適當的迴歸模型。(7分)
- (2)若樣本數為 n ，則此模型的SSR(regression sum of square)的自由度是多少？(3分)
- (3)在你的模型之下，一個45歲的女性，服用Mediate劑量後預期的疼痛減緩程度為何？(5分)
- (4)在你的模型之下，對應[新藥的劑量對疼痛的減緩程度沒有影響]的虛無假說為何？(5分)