

是非題(每題 2 分，共 40 分) 請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

1. 卡方檢定(Chi-square test)可以檢定兩個離散型隨機變數(discrete random variable)之間是否有相關性(association)。
2. 卡方檢定所得的 p 值表示兩個離散型隨機變數之間相關性的強度。
3. 卡方檢定所檢定的資料之母體服從卡方分配，是有母數統計。
4. 卡方檢定的檢定統計量(test statistic)在大樣本時，會近似連續型的隨機變數。
5. 卡方檢定可以使用在  $2 \times 2$  的列連表(contingency table)資料，也可以使用在  $3 \times 2$  的列連表。
6. 在假說檢定(hypothesis testing)中，p 值如果越小，表示統計的顯著性越大。
7. 在假說檢定中，如果 p 值比顯著水準(significance level)大，表示不能推翻虛無假說(null hypothesis)，虛無假說為真。
8. 在假說檢定中，p 值會受到樣本數的影響，樣本數極大時，p 值容易變小。
9. 同一筆資料進行雙尾檢定(two-sided test)時，其 p 值會比進行單尾檢定時的 p 值還大。
10. 假說檢定的結果是否顯著，與型一誤差率(type I error rate)及型二誤差率(type II error rate)有關，這兩種誤差越小，檢定結果越容易顯著。

請根據迴歸模型  $Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \varepsilon$ ,  $\varepsilon \sim N(0, \sigma^2)$  繼續回答第 11-20 題。

11. 樣本數越多，則  $\sigma^2$  越小。
12. 增加新的解釋變數一定可以降低  $\sigma^2$ 。
13. 若  $\sigma^2$  越小，則對  $\beta_1$  所建立的信賴區間越寬。
14. 在資料蒐集的過程中，可以觀察到  $\varepsilon$  的數值。
15. 若移除  $X_2$ ，則校正的判定係數(adjusted- $R^2$ )會變小。
16. 若移除  $X_2$ ，則對  $\beta_1$  所建立的信賴區間會變寬。
17. 若移除  $X_2$ ，則  $\beta_1$  的估計值會改變。
18. 若  $X_1$  和  $X_2$  的線性相關性越強，則參數估計式的變異數越小。
19. 若  $R^2 = 0$ ，則代表  $(X_1, X_2)$  無法解釋  $Y$ 。
20. 若  $X_1$  增加一單位，則新觀測到的  $Y$  會增加  $\beta_1$ 。

見背面

選擇題(每題3分，共15分)請於試卷內之「非選擇題作答區」標明題號依序作答。

21. 某研究公司想調查某一新型病毒在攝氏20度下的72小時存活率是否能超過50%。該公司先培養了100個內有該病毒的培養皿，其中9個在24小時內死亡，14個在24-48小時內死亡，28個在48-72小時內死亡。該公司打算利用假說檢定來探討存活率是否能超過50%，請問以下何者正確？
- (A)該公司的虛無假說為存活率能超過50%。
  - (B)該公司的虛無假說為存活率小於或等於50%。
  - (C)該公司的對立假說為存活率不到50%。
  - (D)該公司的對立假說為存活率等於50%。
22. (延續上題)請問72小時的存活率的估計值最接近以下何者？
- (A)28/100
  - (B)51/100
  - (C)49/100
  - (D)72/100
23. (延續上題)請問以下敘述何者不正確？
- (A)該存活率的95%信賴區間會包含50%。
  - (B)利用Z檢定所得之檢定統計量會介於±1.64之間。
  - (C)利用Z檢定所得之檢定統計量會介於±1.96之間。
  - (D)利用Z檢定所得之p值會小於0.05。
24. 王博士調查某間工廠中重金屬的暴露對裝配員體內血小板的影響，發現50位暴露在該重金屬下的裝配員中，有10位的體內血小板數目較一般人高。王博士另外隨機抽取沒有暴露在該重金屬下的45人，其中有5位的體內血小板數目較一般人高。如果要檢定該重金屬暴露是否與血小板數目增加有關，在顯著水準5%之下，請問以下何者正確？
- (A)可以檢定勝算比值(odds ratio, OR)是否為0。
  - (B)可以計算OR的95%信賴區間是否包含0。
  - (C)這筆資料所得之OR值為2，是常數，不是統計量。
  - (D)檢定統計量的p值如果比0.05小，表示兩者可能有相關。

25. (延續上題)請問上述算出的 OR 值的標準誤(standard error)最接近以下何者?

(A)  $\frac{1}{10} + \frac{1}{5} + \frac{1}{40} + \frac{1}{40}$

(B)  $10 + 5 + 40 + 40$

(C)  $\sqrt{\frac{1}{10} + \frac{1}{5} + \frac{1}{40} + \frac{1}{40}}$

(D)  $\sqrt{10 + 5 + 40 + 40}$

選擇題(每題5分,共35分)請於試卷內之「非選擇題作答區」標明題號依序作答。

26. 有關標準差(standard deviation)及標準誤(standard error),下列何者正確?

- (A) 樣本數增加,標準差亦變大。
- (B) 樣本數趨近無限大,標準差趨近零。
- (C) 標準差越大,表示資料的散佈程度越小。
- (D) 標準誤越大,表示資料的散佈程度越小。
- (E) 標準誤與原資料有同樣的單位(unit)。

27. 有關變異係數(coefficient of variation),下列何者正確?

- (A) 其值介於0和1之間。
- (B) 其值介於-1和1之間。
- (C) 其值與原資料有同樣的單位。
- (D) 其值的單位是原資料單位的倒數。
- (E) 其值越大,表示資料的散佈程度越大。

28. 有關機率分佈(probability distribution)與抽樣分佈(sampling distribution),下述何者正確?

- (A) 連續型隨機變數(continuous random variable)之機率密度函數(probability density function)的曲線下面積為機率(probability)。

見背面

- (B) 離散型(discrete)隨機變數之機率分佈函數的平方為機率。
- (C) 樣本統計量 (statistic)若為離散型，其抽樣分佈為連續型。
- (D) 樣本統計量若為連續型，其抽樣分佈為離散型。
- (E) 樣本統計量之抽樣分佈函數的曲線下面積等於樣本數。
29. 有關常態分佈(normal distribution)，下列何者錯誤？
- (A) 為對稱分佈。
- (B) 具有兩個參數(parameter)。
- (C) 大於中位數 (median)的機率等於 0.5。
- (D) 兩個獨立(independent)且期望值 (expected value)及變異數 (variance)皆相同的常態分佈之相除，亦為常態分佈。
- (E) 兩個獨立但期望值及變異數皆不相同的常態分佈之相加，亦為常態分佈。
30. 有關卡方分佈(chi-square distribution)，下列何者錯誤？
- (A) 為右偏(positively skewed)分佈。
- (B) 期望值等於自由度(degree of freedom)的平方。
- (C) 大於中位數的機率等於 0.5。
- (D) 兩個獨立且自由度相同的卡方分佈之相除，為 F 分佈。
- (E) 兩個獨立且自由度相同的卡方分佈之相加，亦為卡方分佈。
31. 有關 Student t 分佈，下列何者錯誤？
- (A) 中位數與眾數 (mode)皆為 0。
- (B) 自由度大於 1 時，期望值等於 0。
- (C) 自由度大於 2 時，變異數小於 1。
- (D) 自由度大於 4 時，其峰度(kurtosis)比標準常態分佈的峰度大。
- (E) Student t 分佈的平方，為 F 分佈。
32. X 為伯努立分佈 (Bernoulli distribution)其期望值為 a，Y 為布阿松分佈 (Poisson distribution)其期望值為 b，且 X 與 Y 不獨立。下列何者錯誤？
- (A) X 的變異數為  $a-a^2$ 。
- (B)  $X^2$  的期望值為 a。
- (C) Y 的變異數為 b。
- (D)  $Y^2$  的期望值為  $b-b^2$ 。
- (E)  $2X+3Y$  的期望值為  $2a+3b$ 。

問答題(共10分)

33. 某資料蒐集了新藥的劑量 $X$ (low, mediate, high)和血壓 $Y$ 。你想利用迴歸分析(regression analysis)了解[新藥的劑量是否對血壓有影響]，並且希望控制[在新藥劑量沒有影響的情況下誤判為有影響]的機率不超過0.05。請寫出適當的迴歸模型，並且在你的模型下寫出對應以上問題的虛無假說(null hypothesis)。

試題隨卷繳回