

是非題 (每題 2 分) ※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

1. 由盒子圖 (boxplot) 可以看出一組資料是否有偏(skewed)，與是否為常態分佈。
2. 母體期望值的 95% 信賴區間代表的是該期望值落在這個區間的機率為 0.95。
3. 簡單線性迴歸分析之常態分配的假設是因為解釋變數具有常態分佈。
4. 就算沒有常態分配的假設，也可以求得簡單線性迴歸模式中迴歸係數的最小平方法估計值。
5. 中央極限定理指的是在樣本數很大的時候，樣本平均值的抽樣分配會逼近標準常態分配。
6. 疾病盛行率(prevalence)、敏感度(sensitivity)、特異度(specificity)都是條件機率。
7. 某研究收集了男女幼童各 100 人從出生開始到 12 歲每年的身高。該研究可以利用全部資料，以身高為反應變數(response variable)，年齡為自變數(independent variable)的簡單線性迴歸來探討幼童身高的生長曲線。
8. (延續上題)該研究可以針對每個幼童 12 歲時的身高與出生身高的差異，再利用兩個獨立樣本的 t 檢定來檢定男女幼童在 12 年間增加身高的平均值是否相等。
9. (延續上題)該研究可以針對男童從出生開始到 12 歲的每年身高，以變異數分析(ANOVA)檢定每年平均身高是否相等。
10. (延續上題)該研究可以只針對出生的身高，利用兩個獨立樣本的 t 檢定來檢定男女在出生時身高的平均值是否相等。

選擇題 20 分 (每題 4 分) ※ 注意：選擇題請於試卷之「選擇題作答區」依序作答。

1. 有一研究調查 A, B 兩個鄉鎮的居民之飲食習慣的種類，得下表：

	飲食習慣甲	飲食習慣乙	飲食習慣丙	合計
A 鄉鎮	50	30	20	100
B 鄉鎮	50	80	70	200
合計	100	110	90	300

若使用卡方檢定會得到統計量 19.39，p 值小於 0.0001，請問該卡方檢定的自由度為

- (A) 1
  - (B) 2
  - (C) 3
  - (D) 6
2. (延續上題)在計算卡方檢定統計量時，必須針對每一個觀察值計算其期望值，請問表格中「A 鄉鎮、飲食習慣乙」的觀察值 30 的期望值最接近以下何者？
- (A) 33.3
  - (B) 36.7
  - (C) 66.7
  - (D) 73.3

見背面

3. (延續上題)上述檢定的結論為以下何者？
- (A) 不管是哪一個鄉鎮的人，三種飲食習慣分布的比例都不同，這個不同達統計顯著。
  - (B) 不管是哪一個鄉鎮的人，三種飲食習慣分布的比例都沒有不同。
  - (C) 不同鄉鎮的人，有不同的飲食習慣分布，這個不同達統計顯著。
  - (D) 不同鄉鎮的人，並沒有不同的飲食習慣分布。
4. (延續上題)在該卡方檢定的虛無假設下，B 鄉鎮飲食習慣為第一種的比例最接近何者？
- (A) 0.25
  - (B) 0.35
  - (C) 0.45
  - (D) 0.55
5. (延續上題)以下敘述何者正確？
- (A) 因為  $p$  值不只小於 0.05，還小於 0.001，所以這個檢定的結果比  $0.01 < p \text{ 值} < 0.05$  的情形還要更顯著。
  - (B) 因為上述檢定的  $p$  值小於 0.001，因此顯著水準必小於 0.05。
  - (C) 此時的卡方檢定所參考的卡方分布其拒絕區為該卡方分布的兩側。
  - (D) 此時的卡方檢定是一種無母數檢定。

選擇題 60 分 (每題 5 分) ※ 注意：選擇題請於試卷之「選擇題作答區」依序作答。

6. 某班學生共 100 人，其中 90 人為近視。您遇到該班兩名學生，其中一名近視，另一名沒近視。這種情況的機率最接近下列何值？
- (A) 0.81
  - (B) 0.18
  - (C) 0.10
  - (D) 0.09
  - (E) 0.01
7. 抽樣 100 人，欲檢定飯前與飯後，膽固醇濃度有無不同，可採用下述何者？
- (A) paired t test
  - (B) McNemar test
  - (C) ANOVA (analysis of variance)
  - (D) ANOM (analysis of mean)
  - (E) Kruskal-Wallis test

接次頁

8. 下列何者錯誤？
- (A) 樣本數增加，檢力增加。
  - (B) 樣本數增加，標準誤變小。
  - (C) 樣本數增加，95%信賴區間(confidence interval)變窄。
  - (D) 樣本數趨近無限大，95%信賴區間寬度趨近零。
  - (E) 樣本數趨近無限大，不論虛無假設或對立假設都會達到統計顯著。
9. 病例對照研究(case-control study)，可用下述哪種分析方法？
- (A) logistic regression。
  - (B) Fourier analysis。
  - (C) meta-analysis。
  - (D) Cox regression。
  - (E) Poisson regression。
10. 有關隨機抽樣，下列何者錯誤？
- (A) 可利用電腦產生亂數，達到隨機的目的。
  - (B) 可利用亂數表，達到隨機的目的。
  - (C) 族群中的每一個案，可有不一樣的抽樣機率。
  - (D) 隨機抽樣下，仍然有系統誤差(systematic error)。
  - (E) 隨機抽樣下，仍然有隨機誤差(random error)。
11. 若  $X$  為常態分佈，下列何者錯誤？
- (A)  $X$  為連續性分佈。
  - (B)  $X$  為對稱分佈。
  - (C)  $2X$  為常態分佈。
  - (D)  $X+2$  為常態分佈。
  - (E)  $X^2$  為常態分佈。
12. 下列何者是無母數方法(non-parametric method)？
- (A) Wilcoxon signed-rank test
  - (B) ANOVA (analysis of variance)
  - (C) ANCOVA (analysis of covariance)
  - (D) Student t test
  - (E) paired t test

見背面

13. 吾人想研究新藥的劑量(P: placebo, R: 標準劑量, H: 高劑量)對存活時間(Y)的影響。若以P為對照組,考慮 dummy variable  $X_1 \in \{0,1\}$ 和 $X_2 \in \{0,1\}$ 滿足 $X_1 + X_2 = 1$ ,  $X_1 = 1$ 代表R,  $X_2 = 1$ 代表H。假設抽樣20人,並考慮線性迴歸模型 $E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ 。請問下列何者為服用高劑量的估計存活時間?
- (A)  $\beta_1$
  - (B)  $\beta_2$
  - (C)  $\beta_0 + \beta_2$
  - (D)  $\beta_1 + \beta_2$
  - (E)  $\beta_0 + \beta_1 + \beta_2$
14. (延續上題) 下列何者為對 $\beta_1$ 適當的解釋?
- (A) 服用標準劑量的平均存活時間
  - (B) 服用高劑量的平均存活時間
  - (C) 相較於 placebo, 服用標準劑量平均可以多存活的存活時間
  - (D) 相較於高劑量, 服用標準劑量平均可以多存活的存活時間
  - (E) 以上皆非
15. (延續上題) 欲檢定新藥是否有效, 應考慮下列何種假設檢定?
- (A)  $H_0: \beta_1 = 0$
  - (B)  $H_0: \beta_2 = 0$
  - (C)  $H_0: \beta_1 = 0$  以及  $H_0: \beta_2 = 0$
  - (D)  $H_0: \beta_1 = \beta_2 = 0$
  - (E)  $H_0: \beta_1 = \beta_2$
16. (延續上題) 欲檢定高劑量與標準劑量的影響是否有差異, 應考慮下列何者假設檢定?
- (A)  $H_0: \beta_1 = 0$
  - (B)  $H_0: \beta_2 = 0$
  - (C)  $H_0: \beta_1 = 0$  以及  $H_0: \beta_2 = 0$
  - (D)  $H_0: \beta_1 = \beta_2 = 0$
  - (E)  $H_0: \beta_1 = \beta_2$
17. (延續上題) 研究者認為服藥的年齡對存活時間也是一個重要的影響因素。若吾人多考慮一個年齡變數並重新建立迴歸模型, 下列敘述何者正確?
- (A) 一般來說,  $X_1$  以及  $X_2$  的迴歸係數估計值並不會發生變化
  - (B)  $X_1$  以及  $X_2$  的迴歸係數估計值會改變, 但是相對應的假設檢定其結論並不會改變
  - (C) 判定係數( $R^2$ )一定會增加
  - (D) 校正的判定係數(adjusted  $R^2$ )一定會增加
  - (E) 以上皆非