

※ 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)

- 卡方檢定是無母數統計方法，這是因為沒有對資料母體進行分布的假設。
- 變異數分析(ANOVA)跟 T 檢定一樣，可以在假設兩個母體變異數相等的情況下，檢定兩個獨立的母體期望值是否相等。
- 勝算比(odds ratio)若大於 0，表示風險高，若小於 0，表示風險低。
- 勝算比(odds ratio)可以從資料的  $2 \times 2$  的列聯表求得，也可以利用羅吉斯迴歸(logistic regression)求得。
- 皮爾斯相關係數(Pearson's correlation coefficient)跟勝算比(odds ratio)都是用來查看兩個變數相關的統計量，前者用在連續變數，後者用在二元變數。
- 皮爾斯相關係數(Pearson's correlation coefficient)的範圍可以從 -1 到 1，值越大表示相關性越高。
- XY 散佈圖(scatter plot)可以用來檢視兩個連續變數之間的相關，包含線性及非線性關係。
- 當樣本數很大時，Z 檢定的使用就可以利用 T 檢定來取代。
- 要檢定兩個比率  $p_1$ ,  $p_2$  是否相等時，可以直接計算  $p_1$ ,  $p_2$  各自的信賴區間，再看兩個信賴區間是否重疊，就可以知道檢定結果是否顯著。
- 卡方檢定是單尾檢定(one-sided test)，這是因為卡方檢定的拒絕區(rejection region)在卡方檢定統計量之抽樣分布的右邊。

選擇題(每題 3 分)

- 某研究調查民眾對健保的滿意度，隨機抽樣 100 人，其中 64 人表示滿意。若在顯著水準  $\alpha=0.05$  下，若欲檢定民眾對健保的滿意度是否達八成，請問以下敘述何者最恰當？
  - 檢定結果達統計顯著，因此民眾對健保的滿意度未達八成。
  - 虛無假設為民眾對健保的滿意度未達八成。
  - 這個檢定可使用 T 檢定。
  - 以上檢定為雙尾檢定(two-sided test)。
- 某個病例對照研究(case-control study)隨機抽樣 55 位高血壓患者及 60 位健康民眾，詢問每個人是否有吸菸習慣，得數據如下表。

	沒有吸菸	有吸菸	合計
高血壓患者	25	30	55
健康民眾	40	20	60

請問，若以 OR 來探討有吸菸習慣是否增加高血壓的風險，則 OR 為

- $\frac{25 \times 20}{30 \times 40}$

見背面

(B)  $\frac{30 \times 40}{25 \times 20}$

(C)  $\frac{55 \times 65}{115}$

(D)  $\frac{25/30}{40/20}$

13. (延續上題)有關該研究，請問以下敘述何者最恰當？  
(A)高血壓患者吸菸比率之估計值為 30/50。  
(B)由 OR 值大小可以得知，若檢定吸菸與高血壓是否有相關時的檢定結果是否顯著。  
(C)因為  $OR > 1$ ，所以吸菸者有高血壓的機率比沒有吸菸者有高血壓的機率高。  
(D)因為  $OR < 1$ ，所以吸菸是高血壓的危險因子。
14. 某研究探討台大學生的平均血壓，隨機抽樣 100 人，分別測量其血壓之後，得平均值 108 毫米汞柱，得變異數 144。請問以下何者是台大學生平均血壓的 95%信賴區間？  
(A)  $108 \pm 1.96 \times 144$   
(B)  $108 \pm 1.96 \times \sqrt{144}$   
(C)  $108 \pm 1.96 \times 144 / 100$   
(D)  $108 \pm 1.96 \times \sqrt{144 / 100}$
15. (延續上題)若在顯著水準  $\alpha = 0.05$  下，檢定台大學生的平均血壓是否為 100 毫米汞柱，請問以下哪一個結論最恰當？  
(A)檢定結果得到統計顯著，所以台大學生的平均血壓為 100 毫米汞柱。  
(B)使用 Z 檢定，因 Z 值很大，所以拒絕虛無假設。  
(C)這個檢定跟利用信賴區間進行推論會有一樣的結論。  
(D)因為  $108 > 100$ ，所以這是單尾檢定。

計算問答題

16. 請以適當之統計詞彙，如機率分佈(probability distribution)，樣本分佈(sampling distribution)，標準差(standard deviation)，或標準誤(standard error)等，描述下列現象  
(A) 智商有很大的個體差異(individual variability)。(5分)  
(B) 某研究之迴歸係數有很大的不穩定性(instability)。(5分)  
(C) 某變數的散佈程度(dispersion)很高。(5分)  
(D) 某民意調查之隨機誤差(random error)在可接受範圍內。(5分)  
(E) 某預測模式的表現(performance)很好。(5分)

接次頁

17. 某研究進行病例對照研究(case-control study)，收集 100 例肺癌病例，依其居住地，每一病例配對一位同鄰里之個案為其對照。該研究以標準化問卷，調查病例及對照之吸菸，喝酒，嚼食檳榔等習慣。試說明你將如何分析此筆資料，包括你打算以何方式將資料輸入電腦，打算採用哪種迴歸分析，自變項(independent variable)為何？應變項(response variable)為何？如何闡釋你得到的迴歸係數(regression coefficient)？以及如何檢定吸菸，喝酒，嚼食檳榔間，對於肺癌的發生，是否有交互作用(interaction)？(10 分)
18. 給定一隨機樣本  $X_1, \dots, X_n \sim N(\mu, 1)$ ，並考慮檢定  $H_0: \mu = \mu_0$  vs.  $H_A: \mu > \mu_0$ 。令  $\bar{X}$  為樣本平均數，並採用以下的檢定 T 來執行假設檢定：當  $\bar{X} - \mu_0 > c$  時拒絕  $H_0$ 。定義  $\Phi$  為標準常態分配的累積機率函數，且定義  $z_\alpha$  為滿足  $\Phi(z_\alpha) = 1 - \alpha$  之數， $\alpha \in (0, 1)$ 。請用以上符號依序回答以下有關檢定 T 的問題。
- (A) 請計算 T 的型一誤差(Type-I error) (5 分)
  - (B) 請決定  $c$  使得 T 的型一誤差為 0.1 (5 分)
  - (C) 假設  $\mu = \mu_0 + \delta$ ,  $\delta > 0$ 。請計算 T 的檢定力(power)，並列出兩種可以增加 T 的檢定力的方法。(10 分)
  - (D) 在  $\delta$  固定下，至少需要多少樣本才能使 T 的檢定力超過 0.9？(5 分)
  - (E) 在  $n$  固定下， $\mu$  和  $\mu_0$  的差距( $\delta$ )至少要多少，T 的檢定力才能超過 0.8？(5 分)

試題隨卷繳回