

1.(15%)(配分：3%/小題)

(a) 統計學(statistics)

Statistics is the discipline (art and science) that concerns the collection , organization , analysis , interpretation , and presentation of data.

統計學是一門針對資料做收集、整理(處理)、分析、解釋和呈現的科學。

(b) 統計推論(statistical inference)

Statistical inference is the process of using data obtained from a sample to make estimates and test hypotheses about the characteristics of a population.

統計推論是從樣本獲得的數據對母體參數進行估計和假設的檢定、過程
(底線只寫推論/預測/推估/推測扣 1 分)

(c) 樣本空間(sample space)

The sample space for a random experiment is the set of all experimental outcomes.

樣本空間是一個隨機實驗 所有可能結果的集合

(底線一個沒寫到扣 1 分)

(d) 聯合機率(joint probability)

The probability of the intersection of two events is called joint probability .

條件機率是兩個事件同時發生的機率

(e) 獨立事件(independent event)

Two events A and B are independent if $P(A|B) = P(A)$ or $P(B|A) = P(B)$

當 A,B 兩事件滿足：給定事件 B 之下，A 事件的條件機率等於 A 事件的機率。

或 給定事件 A 之下，B 事件的條件機率等於 B 事件的機率 ($P(A|B)=P(A)$ 或 $P(B|A)=P(B)$)，則稱 A,B 兩事件獨立

(寫兩個事件發生的機率互不影響給 1 分)

2.(45%)

(a)(10%,寫出一個類型 1.5%,一個例子 0.5%)

nominal,例:姓名,學號

ordinal,例:名次,滿意度

interval,例:溫度,西元年份

ratio,例:長度,金錢

(b)(5%)想了解母體的特性，藉由樣本收集進行分析或統計推論。

(c)(5%)Relative Frequency Method: assigning probabilities based on experimental or historical data. That is when data are available to estimate the proportion of the time the experimental outcome will occur if the experiment is repeated a large number of times.

在重複多次的實驗中,實驗結果發生的機率為出現該結果之次數與實驗的總次數之比

(底線一個沒寫到扣 1 分)

(d)(5%)

.When the data are believed to approximate a bell-shaped distribution (e.g., normal distribution), the empirical rule can be used to determine the percentage of data values that must be within a specified number of standard deviations of the mean. Approximately (68%, 95%, 99%) of the data values will be within (one, two, three) standard deviation of the mean.

當資料近似鐘型分配，則約有(68%,95%,99%)的資料點落在距離平均(1,2,3)個標準差的範圍內

(寫常態分配/底線少寫一個扣 1 分)

(e)(20%)先說明變數的測量尺度，進行敘述統計時，分圖表呈現(第二章)及統計量的計算(第三章)兩部份。

(A) (寫出類別/連續變數 3 分,沒有特別寫出分類但在(B)(C)有分類 2 分)

此資料中，類別型變數有 2 個:性別，具有三個類別「男、女、未告知」、入學管道，具有三個類別「繁星推薦、個人申請、考試入學)。

連續型變數有 2 個: 身高及體重。身高及體重可依分析需求進行分組(離散化)，轉成類別型變數。

(B) 圖表呈現:(寫出方法 1 分,目的 1 分,同一項算一種方法)

(1)計算次數分佈表(相對次數分佈表、百分次數分佈表)，以了解類別變數中，每個類別的觀察個數。連續變數也可藉由資料分組 (即，數值由小到大排序後，分組) 計算其次數分佈表。

(2)繪製長條圖及餅圖，呈現類別型變數的次數分佈表(相對次數分佈表、百分次數分佈表)。

(3) 繪製點圖、盒型圖、直方圖、莖葉圖，呈現連續型變數的分佈，以了解變數是否為對稱、右偏、左偏或其它分佈。

(4) 針對兩類別變數(或已分組後之連續型資料)計算交叉表(Crosstabulation)，以了解兩變數之聯合次數，及是否有 Simpson's Paradox 的現象。

(5) 繪製兩連續變數之散佈圖(於上加入趨勢線 Trendline)，以了解兩變數之關係(正向、負向、無關係，或其它非線關係)及偵測離群值。

(6) 繪製並排長條圖、堆疊長條圖來呈現兩類別變數之交叉表資訊，以若察各組之次數或比例。

(C) 統計量的計算:(寫出方法 1 分,目的 1 分,同一項算一種方法)

(1) 針對每一連續型變數，可計算平均數、加權平均數、中位數、幾何平均數、眾數、百分位數及四分位等等，來了解資料之中心趨勢及位置。

(2) 針對每一連續型變數，可計算範圍、四分位距、變異數、標準差、變異係

數等等，來了解資料之變異程度。

(3) 針對每一連續型變數，可計算偏態係數，來了解資料之分佈形狀。

(4) 針對每一連續型變數，將之轉成 z 分數，可依一般法則($Q1-1.15*IQR$, $Q3+1.5*IQR$)，檢測是否有離群值。

(5) 針對每一連續型變數，計算五數綜合，並以盒形圖呈現，以了解資料之分佈及是否有離群值。

(6) 針對兩個連續型變數，計算共變異數、相關係數，以了解此兩變數是否有線性關係。

3.

(a)

Green condition

Gender	Too fast	Fine	Total
Male	35	65	100
Female	40	60	100
Total	75	125	200

Male: $\frac{35}{100} = 35\%$, Female: $\frac{40}{100} = 40\%$

∴ Female 比例較高 (1)

(b) Male: $\frac{10}{50} = 20\%$, Female: $\frac{1}{10} = 10\%$

∴ Male 比例較高 (1)

(c) Male: $\frac{25}{50} = 50\%$, Female: $\frac{39}{90} \approx 43.33\%$

∴ Male 比例較高 (1)

(d) (a) 和 (b), (c) 得到 相反的結論 (3)

這是 辛普森悖論 (Simpson's Paradox) 的例子

(3)

4.

(a)

排序: 12, 15, 15, 18, 18, 19, 20, 21, 22, 22,
24, 24, 26, 26, 27, 29, 42, 64

$$L_{25} = \frac{25}{100} (18+1) = 4.75 \quad (1)$$

$$Q_1 = 18 + 0.75 \times (18 - 18) = 18 \quad (1)$$

$$L_{50} = \frac{50}{100} (18+1) = 9.5 \quad (1)$$

$$Q_2 = 22 + 0.5 (22 - 22) = 22 \quad (1)$$

$$18 < 22 \quad (1)$$

(b)

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

$$= \frac{1}{17} \left[13346 - \frac{1}{18} \times 444^2 \right] \approx 140.8235 \quad (3)$$

$$\Rightarrow S \approx 11.867 \quad (2)$$

(c)

$$L_{75} = \frac{75}{100} (18+1) = 14.25$$

$$Q_3 = 26 + 0.25(27-26) = 26.25 \quad (1)$$

$$IQR = 26.25 - 18 = 8.25 \quad (1)$$

$$\text{Lower limit} = 18 - 1.5 \times 8.25 = 5.625 \quad (1)$$

$$\text{Upper limit} = 26.25 + 1.5 \times 8.25 = 38.625 \quad (1)$$

Yes. 2 outliers: 42, 64 (1)

(d)
$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{18} \times 444 \approx 24.67$$

Public school districts 平均有 24.67% 的學生使用 iPads, 但有些 Public school districts 高達 64% 的學生使用 iPads

(2)

5.

(a)

① Let I = treatment-caused injury

① D = death from injury

① N = injury caused by negligence

① M = malpractice claim filed

① A = payment made in claim

$$P(I) = 0.04, \quad P(D|I) = \frac{1}{7}, \quad P(N|I) = 0.25$$

$$P(M|N) = \frac{1}{7.5} \approx 0.1333$$

$$P(A|M) = 0.5$$

(b)

$$P(N) = P(I) P(N|I)$$

$$= 0.04 \times 0.25 \quad \text{④}$$

$$= 0.01 \quad \text{①}$$

(c)

$$P(D) = P(I) P(D|I)$$

$$= 0.04 \times \frac{1}{7} \quad (4)$$

$$\approx 0.0057 \quad (1)$$

(d)

$$P(A, M|N)$$

$$= P(A|M) P(M|N)$$

$$= \frac{1}{2} \times \frac{1}{7.5} = \frac{1}{15} \approx 0.0667$$

(4) (1)