



10/16統計學實習課

114-1統計學實習課

1. A characteristic of interest for the elements is called a

- a. sample
- b. data set
- c. variable
- d. None of these alternatives is correct.

答案：c. variable (變數)

解釋：

在統計學中，**變數 (variable)** 是對感興趣的元素或個體的一個特徵或屬性。例如，如果我們研究一群人，年齡、性別或收入就是變數。

2. Income is an example of a variable that uses the

- a. ratio scale
- b. interval scale
- c. nominal scale
- d. ordinal scale

答案：a. ratio scale (比例尺度)

解釋：

- **比例尺度 (Ratio Scale)** 具有所有尺度的特性：分類、排序、等距，**且具有真正的零點 (True Zero)**。
- 收入 (Income) 屬於比例尺度，因為我們可以說 0 元代表「沒有收入」，而且我們可以進行有意義的比例計算 (例如 100 元是 50 元的兩倍)。

3. In a questionnaire, respondents are asked to mark their gender as male or female. The scale of measurement for gender is _____ scale.

- a. ordinal
- b. nominal
- c. ratio
- d. interval

答案：b. nominal (名目尺度)

解釋：

- **名目尺度 (Nominal Scale)** 僅用於分類和標記數據，類別之間沒有固有的順序或大小關係。
- 性別（男性、女性）是典型的名目數據，它們是不同的類別，但無法排序或比較大小。

4. Quantitative data

- a. are always nonnumeric
- b. may be either numeric or nonnumeric
- c. are always numeric
- d. None of these alternatives is correct.

答案：c. are always numeric (總是數值的)

解釋：

- **量化數據 (Quantitative Data)** 是以數字形式表示的數據，代表數量或數值（例如：年齡、身高、價格）。
- **質化數據 (Qualitative Data)** 是以文字或標籤形式表示的數據（例如：性別、職業、顏色）。因此，量化數據必須是數值的。

6. A statistics professor asked students in a class their ages. On the basis of this information, the professor states that the average age of all the students in the university is 24 years. This is an example of

- a. a census
- b. descriptive statistics
- c. an experiment
- d. statistical inference

答案：d. statistical inference (統計推論)

解釋：

- **統計推論 (Statistical Inference)** 是指利用從**樣本**（這班學生）獲得的資訊，對更廣泛的**母體**（全大學的學生）的特徵進行估計、假設檢定或預測。
- 教授根據「班級」的資訊推論「全大學」的情況，這就是一個推論過程。

7. The Department of Transportation of a city has noted that on the average there are 17 accidents per day. The average number of accidents is an example of

- a. descriptive statistics
- b. statistical inference
- c. a sample
- d. a population

答案：a. descriptive statistics (描述統計)

解釋：

- **描述統計 (Descriptive Statistics)** 是指收集、整理、呈現和總結數據，以描述其基本特徵。
- 「平均每天有 17 起事故」是對已收集數據（過去一段時間的事故數）的**數值性總結**，屬於描述統計的範疇。

8. The measure of location which is the most likely to be influenced by extreme values in the data set is the

- a. range
- b. median
- c. mode
- d. mean

答案：d. mean (平均數)

解釋：

- **平均數 (Mean)** 的計算涉及數據集中的每一個數值，因此它對極端值 (outliers) 非常敏感。一個極大或極小的數值會顯著地拉高或拉低平均數。
- **中位數 (Median)** 則只受中間數值的影響，對極端值具有魯棒性。

9. The variance can never be

- a. zero
- b. larger than the standard deviation
- c. negative
- d. smaller than the standard deviation

答案：c. negative (負數)

解釋：

- **變異數 (Variance)** 是數據與平均數差的平方的平均值。由於平方數必為非負數 (≥ 0)，因此變異數**不可能為負數**。
- 變異數可以為零 (如果所有數據都相同)。
- 當標準差 (σ) 介於 0 和 1 之間時，變異數 (σ^2) 會小於標準差 (例如 $\sigma = 0.5, \sigma^2 = 0.25$)。當 $\sigma > 1$ 時，變異數會大於標準差。

10. The p th percentile is a value such that at least p percent of the observations are

- a. less than or equal to this value
- b. less than this value
- c. more than or equal to this value
- d. more than this value

答案：a. less than or equal to this value (小於或等於此值)

解釋：

- **第 p 個百分位數 (p th percentile)** 的定義是：至少有 p 百分比的觀測值小於或等於這個值，同時，至少有 $(100 - p)$ 百分比的觀測值大於或等於這個值。選項 a 符合上半部分的標準定義。

11. Given the following information: Standard deviation = 8, Coefficient of variation = 64%.

The mean would then be

- a. 12.5
- b. 8
- c. 0.64
- d. 1.25

答案：a. 12.5

計算過程：

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}}$$

已知 $CV = 64\% = 0.64$ ，標準差 $(\sigma) = 8$ 。

$$0.64 = \frac{8}{\text{Mean}}$$

$$\text{Mean} = \frac{8}{0.64} = 12.5$$

12. In the class power-point file, we saw the random numbers generated from $U(0, 10)$, $N(5, 16)$, and $Exp(5)$. Judging from the histogram of data, which distribution has the largest variance?

- a. $U(0, 10)$
- b. $N(5, 16)$
- c. $Exp(5)$
- d. The variances are the same.

答案： c. Exp(5)

解釋：

題目中三個分配的變異數如下：

1. 均勻分配 $U(0,10)$

變異數公式為： $(b - a)^2/12 = (10 - 0)^2/12 = 100/12 \approx 8.33$

2. 常態分配 $N(5,16)$

變異數就是給定的數值 16

3. 指數分配 $Exp(5)$

若參數是「平均數 = 5」，則變異數為：

$$\text{Var} = 5^2 = 25$$

13. Refer to the above table, among the students who plan to go to graduate school, what percentage indicated "Other" majors?

- a. 15.75
- b. 45
- c. 54
- d. 35

答案：b. 45

根據表格：

- 計畫念研究所 (Yes) 的學生總數 = 280
- 其中主修 Other 的學生數 = 126

$$\text{百分比} = \frac{126}{280} \times 100\% = 45\%$$

14. Which course has the largest coefficient of variation?

- a. 數學甲
- b. 數學乙
- c. 化學
- d. 物理

答案：b. 數學乙

要找 變異係數最大 的科目

公式：

$$CV = \frac{\text{標準差 (s.d.)}}{\text{平均數 (mean)}} \times 100\%$$

科目	平均數	標準差	CV (%)
數學甲	36.68	18.72	51.0%
數學乙	34.36	25.97	75.5%
化學	38.88	27.00	69.4%
物理	28.75	21.50	74.8%

15. Which course's grades are closest to symmetry?

- a. 國文
- b. 物理
- c. 數學甲
- d. 英文

答案：a. 國文

解釋：

數據分佈對稱性的常見判斷方法是：

1. 比較 Mean (平均數) 和 Median (中位數)：越接近越對稱。
2. 觀察四分位數與中位數的距離： $Q_3 - \text{Median}$ 和 $\text{Median} - Q_1$ 越接近越對稱。
3. 觀察 Mean 和 Max/Min 的相對位置 (例如是否偏向某一極端)。

課程	Mean	Median	$Q_3 - \text{Median}$	$\text{Median} - Q_1$	差值
國文	43.56	44.00	$53 - 44 = 9$	$44 - 34 = 10$	1
物理	28.75	23.00	$41 - 23 = 18$	$23 - 12 = 11$	7
數學甲	36.36	34.00	$49 - 34 = 15$	$34 - 22 = 12$	3
英文	36.68	34.00	$56 - 34 = 22$	$34 - 16 = 18$	4

16. Fifteen percent of the students in a school of Business Administration are majoring in Economics, 20% in Finance, 35% in Management, and 30% in Accounting. The graphical device which can be used to present these data is

- a. a line graph
- b. only a bar graph
- c. only a pie chart
- d. both a bar graph and a pie chart

答案：d. both a bar graph and a pie chart (長條圖和圓餅圖皆可)

解釋：

這是一組質化數據 (主修專業) 及其百分比 (頻率分佈)。

- **圓餅圖 (Pie Chart)**：用於顯示一個整體 (全體學生) 中各類別 (各專業) 所佔的**比例**，非常適合百分比數據。
 - **長條圖 (Bar Graph)**：用於顯示各類別的**頻率或數量**，也適用於比較不同類別的百分比。
 - **線形圖 (Line Graph)**：主要用於顯示數據隨時間的變化，不適用於這類類別數據。
- 因此，兩者都適用。

17. The most common graphical presentation of quantitative data is a

- a. histogram
- b. bar graph
- c. relative frequency
- d. pie chart

答案：a. histogram (直方圖)

解釋：

- **直方圖 (Histogram)：**是將連續的量化數據分組後，以長方形的面積來表示各組數據的頻率，是呈現量化數據分佈最常用的圖形。
- **長條圖 (Bar Graph) 和圓餅圖 (Pie Chart)** 主要用於質化數據或類別數據。

18. In a cumulative relative frequency distribution, the last class will have a cumulative relative frequency equal to

- a. one
- b. zero
- c. the total number of elements in the data set
- d. 0.5

答案：a. one (一)

解釋：

累積相對頻率是將所有類別的相對頻率相加。由於相對頻率的總和（代表數據集所有部分）必須等於 1 (或 100%)，因此最後一個類別的累積相對頻率必定為 1。

19. A tabular method that can be used to summarize the data on two variables simultaneously is called

- a. simultaneous equations
- b. crosstabulation
- c. a histogram
- d. an ogive

答案：b. crosstabulation (交叉列表/列聯表)

解釋：

- **交叉列表 (Crosstabulation)** (或稱列聯表) 是一種表格方法，用於同時匯總和呈現兩個或多個變數的頻率分佈，例如第 13 題的「畢業主修」和「是否讀研」。

20. A graphical presentation of the relationship between two variables is

- a. an ogive
- b. a histogram
- c. either an ogive or a histogram, depending on the type of data
- d. a scatter diagram

答案：d. a scatter diagram (散佈圖)

解釋：

- **散佈圖 (Scatter Diagram)：**用於展示兩個**量化變數**之間關係的圖形，通常用於判斷變數間是否存在相關性。
- 直方圖 (Histogram) 和累積曲線 (Ogive) 都是單變數的分佈圖。

21. The most important statistical descriptive measure of the location of a data set is the

- a. mean
- b. median
- c. mode
- d. variance

答案：a. mean (平均數)

解釋：

- **平均數 (Mean)**：是最常用且最重要的集中趨勢量數。雖然它受極端值影響，但在大多數情況下，它是描述數據中心位置的最佳單一指標。

22. The scatter diagram can be used to check if two variables are correlated. Based on the following data, what can you say about the variables x and y ?

x = Number of Interceptions	y = Number of Points Scored
1	14
3	24
2	18
1	17
3	27

- a. Positive relationship
- b. No apparent relationship
- c. Negative relationship
- d. None of the above

答案：a. Positive relationship (正向關係)

- 數據點 (1, 14), (3, 24), (2, 18), (1, 17), (3, 27) 顯示，當 x 增加時， y 傾向於增加。
- **正向關係 (Positive Relationship)**：表示兩個變數同向變動，一個增加，另一個也傾向於增加。

23. The process of capturing, storing, and maintaining data is known as

- a. data manipulation.
- b. data mining.
- c. data warehousing.
- d. big data.

答案：c. data warehousing (資料倉儲)

解釋：

- **資料倉儲 (Data Warehousing)：**是指為支持決策制定過程而進行的數據的收集、存儲和管理。它專注於長期、歷史性的數據儲存和維護。
- **資料探勘 (Data Mining)：**是從大量數據中發現模式和知識的過程。

24. A cumulative relative frequency distribution shows the proportion of data items

- a. with values less than or equal to the upper limit of each class.
- b. with values less than or equal to the lower limit of each class.
- c. with values less than or equal to the upper limit of each class.
- d. with values less than or equal to the lower limit of each class.

答案：a. with values less than or equal to the upper limit of each class. (數值小於或等於每個組的上限)

解釋：

- **累積頻率**或**累積相對頻率**是將某一組及其**所有先前組**的頻率或相對頻率加總。
- 這個總和代表了數值**小於或等於該組上限**的所有數據項的比例或數量。選項 *a* 和 *c* 內容重複，都是正確的定義。

25. Your retirement account earned the following returns in the past five years: 9%, 3%, 3%, 2%, and 1%. Calculate the geometric mean return of your retirement account.

- a. 3.2%
- b. 3.6%
- c. 3.8%
- d. 4.0%

答案：b. 3.6%

幾何平均數 $G = [(1 + r_1)(1 + r_2) \cdots (1 + r_n)]^{1/n} - 1$

把題目數字代入：

$$G = [(1.09)(1.03)(1.03)(1.02)(1.01)]^{1/5} - 1 = 0.0355 = 3.55\%$$