

第 4 章

敘述統計之二

(統計數字呈現資料)

定義 4-1-1：母群體平均數 (population mean)

母群體全部資料的平均數稱為母群體平均數，並以符號 μ (唸成 mu) 表示

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

其中 N 為母群體資料個數。

定義 4-1-2：樣本平均數 (sample mean)

經抽樣取得的樣本資料的平均數稱為樣本平均數，並以符號 \bar{X} (唸成 X bar) 表示

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

其中 n 為樣本數目。

定義 4-1-3：中位數 (median)

$m = x_{(\frac{n+1}{2})}$ ，當 n 為奇數時。

016 · 統計學習題解答

$$m = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, \text{ 當 } n \text{ 為偶數時。}$$

定義 4-1-4：眾數 (mode)

資料中出現頻度最高的資料，稱為眾數。

定義 4-2-1：全距 (range)資料中的最大值與最小值差稱為這群資料的全距，以符號 R_n 表示。**定義 4-2-2：內距 (interquartile)**

丟棄資料中較小及較大的 25% 資料後，以剩餘的中間 50% 資料的最大值減最小值稱為內距。

定義 4-2-3：絕對平均偏差 (mean absolute deviation) n 筆資料 x_1, \dots, x_n 的絕對平均偏差以符號 MAD 表示，其計算公式如下

$$\text{MAD} = \frac{\sum |x_i - \bar{X}|}{n}$$

定義 4-2-4：變異數 (variance)

(1) 母群體變異數 (population variance)

對於母群體資料 x_1, \dots, x_n 而言，其變異數以符號 σ^2 (唸成 sigma squared) 表示，計算公式為

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

(2) 樣本變異數 (sample variance)

對於樣本資料 x_1, \dots, x_n 而言，其變異數以符號 S^2 來表示，計算公式為

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$$

定義 4-2-5：標準差 (standard deviation)

(1) 母群體標準差 (population standard deviation)：對於母群體資料 x_1, \dots, x_n 而言，其標準差以符號 σ 表示，計算公式則為母群體變異數的正平方根，也就是

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

(2) 樣本標準差 (sample standard deviation)：對於樣本資料 x_1, \dots, x_n 而言，標準差以符號 S 表示，計算公式則為樣本變異數的正平方根，也就是

$$S = \sqrt{S^2} = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}}$$

定義 4-3-1：百分位數 (percentiles)

對於資料 x_1, \dots, x_n ，我們稱 x_p 為這些資料的 $100p$ 百分位，如果全部資料有 $100p\%$ 比例的資料小於或等於 x_p ，同時也有 $100(1-p)\%$ 比例的資料大於或等於 x_p 。

定義 4-3-2：四分位數 (quartiles)

將由小至大排序的資料切割成四等分的切割點位置分別為

- (1) 25 百分位數 ($x_{0.25}$)，我們又稱它為第一四分位數 (the first quartile) 或下四分位數 (the lower quartile)。
- (2) 50 百分位數 ($x_{0.5}$)，我們又稱它為第二四分位數 (the second quartile)，事實上它就是中位數 (median)。
- (3) 75 百分位數 ($x_{0.75}$)，我們又稱它為第三四分位數 (the third quartile) 或上四分位數 (the upper quartile)。

近似鐘型分佈資料的經驗法則：

- (1) 大約有 68% 的資料在平均數左右一個標準差範圍內。
- (2) 大約有 95% 的資料在平均數左右二個標準差範圍內。
- (3) 大約有 99.7% 的資料在平均數左右三個標準差範圍內。

定義 4-4-1：標準化值 (standardized score)

以符號 Z 表示資料的標準化值，則母群體資料與樣本資料的標準化值分別定義如下：

- (1) 母群體資料的標準化值：

母群體資料 x 標準化後的標準化值為

$$Z = \frac{x - \mu}{\sigma}$$

- (2) 樣本資料的標準化值：

樣本資料 x 標準化後的標準化值為

$$Z = \frac{x - \bar{X}}{S}$$

4.1 連鎖便利商店某門市五位專職員工的月薪為 23,000、35,000、29,000、25,000、40,000。

- (1) 計算其平均數。
- (2) 計算其中位數。
- (3) 計算其全距。
- (4) 計算其變異數及標準差。

解：

以 X 表示月薪，且將這五筆資料看成為一個母群體，則

$$(1) \mu = \frac{\sum x_i}{n} = \frac{23,000 + \dots + 40,000}{5} = 30,400$$

$$(2) \frac{5}{2} = 2.5, \text{ 中位數為排序第 3 位數據，所以中位數為 } 29,000。$$

$$(3) \text{全距} = \text{最大值} - \text{最小值} = 40,000 - 23,000 = 17,000$$

(4) 若我們將這五筆資料視為母群體中的全部資料，則其變異數為

$$\begin{aligned} \sigma^2 &= \frac{\sum (x_i - \mu)^2}{N} \\ &= \frac{(23,000 - 30,400)^2 + \dots + (40,000 - 30,400)^2}{5} \\ &= 49,800,000 \end{aligned}$$

4.2 抽樣調查 20 個家庭並記錄其家中的人口數量如下：

3 2 4 2 5 1 4 6 3 3
1 3 4 3 4 2 5 6 3 3

- (1) 計算家庭人口數的中位數。
- (2) 計算家庭人口數的眾數。
- (3) 計算家庭人口數分佈的全距。

解：

(1) 首先將數據排序

1 1 2 2 2 3 3 3 3 3
3 3 4 4 4 4 5 5 6 6

020 · 統計學習題解答

$n=20$ ，所以中位數為第 10 筆與第 11 筆的平均數， $x_{(10)}=3$ ， $x_{(11)}=3$ ，所以中位數為 3 人。

(2) 首先將數據轉換成為頻度表

| x | 頻度 |
|-----|----|
| 1 | 2 |
| 2 | 3 |
| 3 | 7 |
| 4 | 4 |
| 5 | 2 |
| 6 | 2 |
| | 20 |

所以家庭人口數的眾數為 5

(3) 全距 = 最大值 - 最小值 = $6 - 1 = 5$ 人

4.3 抽樣調查網路咖啡廳中消費者的年齡數據如下：

15 19 23 18 17 14 15 16 18 25
21 30 14 16 17 15 14 13 16 29

- (1) 計算網路咖啡廳消費人口年齡的中位數。
- (2) 計算網路咖啡廳消費人口年齡的眾數。
- (3) 計算網路咖啡廳消費人口年齡的平均數。

解：

(1) 首先將數據排序

13 14 14 14 15 15 15 16 16 16
17 17 18 18 19 21 23 25 29 30

$n=20$ ，所以中位數為排序第 10 筆與第 11 筆的平均值， $x_{(10)}=16$ ，

$x_{(11)} = 17$ ，所以中位數為 $\frac{16+17}{2} = 16.5$ 。

(2) 本題數據的頻度表為

| x | 頻度 |
|-----|----|
| 13 | 1 |
| 14 | 3 |
| 15 | 3 |
| 16 | 3 |
| 17 | 2 |
| 18 | 2 |
| 19 | 1 |
| 21 | 1 |
| 23 | 1 |
| 25 | 1 |
| 29 | 1 |
| 30 | 1 |

本題中 14, 15, 16 頻度相同，所以眾數並不存在。

$$(3) \bar{X} = \frac{\sum x_i}{20} = 18.25$$

4.4 抽樣調查 20 位大專同學在校參加社團數目的數據如下：

0 2 3 2 0 1 1 1 3 4
3 2 2 1 1 1 1 0 1 2

(1) 計算大專同學在校參加社團數目的中位數。

(2) 計算大專同學在校參加社團數目的眾數。

解：

(1) 首先將資料排序

022 · 統計學習題解答

0 0 0 1 1 1 1 1 1 1
1 2 2 2 2 2 3 3 3 4

$n=20$ ，中位數為排序第 10 筆與第 11 筆數據的平均值， $x_{(10)}=1$ ， $x_{(11)}=1$ 。

所以中位數為 1。

(2) 本題數據的頻度表為

| x | 頻度 |
|-----|----|
| 0 | 3 |
| 1 | 8 |
| 2 | 5 |
| 3 | 3 |
| 4 | 1 |

所以，大專同學在校參加社團活動的數目之眾數為 1。這表示大多數大專同學選擇一個社團活動。

4.5 抽樣記錄某人清晨體溫數據如下：36.3, 35.6, 36.2, 36.5, 35.8, 36.4, 36, 35.8, 36.1, 35.8, 36.3, 35.9。

- (1) 計算其清晨的平均體溫。
- (2) 計算其清晨體溫的變異數及標準差。
- (3) 計算其清晨體溫的全距。

解：

$$(1) \bar{X} = \frac{\sum x_i}{n} = \frac{36.3 + \cdots + 35.9}{12} = 36.058$$

$$\begin{aligned} (2) S^2 &= \frac{\sum (x_i - \bar{X})^2}{n-1} \\ &= \frac{(36.3 - 36.058)^2 + \cdots + (35.9 - 36.058)^2}{11} \\ &= 0.081 \end{aligned}$$

$$S = \sqrt{0.081} = 0.2846$$

$$\begin{aligned} (3) \text{全距} &= \text{最高體溫} - \text{最低體溫} \\ &= 36.5 - 35.6 \\ &= 0.9 \end{aligned}$$

4.6 記錄某月份股市漲跌百分比數據如下：

| | | | | | | | | | |
|-------|-------|-------|--------|--------|-------|-------|-------|--------|-------|
| 0.05 | 0.03 | -0.02 | 0.002 | -0.04 | -0.02 | 0.005 | 0.001 | -0.005 | -0.01 |
| 0.007 | -0.04 | 0.01 | -0.009 | -0.01 | 0.03 | 0.001 | -0.06 | -0.006 | 0.001 |
| 0.003 | 0.01 | -0.03 | -0.006 | -0.003 | 0.02 | 0.005 | -0.02 | -0.006 | 0.01 |

- (1) 計算該月股市平均漲跌百分比。
- (2) 計算該月股市漲跌百分比的變異數及標準差。
- (3) 計算該月股市漲跌百分比的四分位數。
- (4) 如果今日股市上漲 5%，以標準化值來表示今日股市漲跌的相對位置。

解：

$$(1) \bar{X} = \frac{\sum x_i}{n} = \frac{0.05 + \cdots + 0.01}{30} = -0.003$$

$$\begin{aligned} (2) S^2 &= \frac{\sum (x_i - \bar{X})^2}{n-1} = \frac{[0.05 - (-0.003)]^2 + \cdots + [0.01 - (-0.003)]^2}{29} \\ &= 0.0005 \end{aligned}$$

$$S = 0.0221$$

(3) 先將漲跌百分比排序

| | | | | | | | | | |
|--------|--------|--------|--------|--------|-------|-------|-------|-------|--------|
| -0.06 | -0.04 | -0.04 | -0.03 | -0.02 | -0.02 | -0.02 | -0.01 | -0.01 | -0.009 |
| -0.006 | -0.006 | -0.006 | -0.005 | -0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.003 |
| 0.005 | 0.005 | 0.007 | 0.01 | 0.01 | 0.01 | 0.02 | 0.03 | 0.03 | 0.05 |

第一四分位（25 百分位）： $i = 0.25(30) = 7.5$ ，所以第一四分位為排序數據第 8 筆，也就是 $x_{0.25} = -0.01$ （下跌 1%）。

024 · 統計學習題解答

第二四分位（50 百分位或中位數）：為第 15 筆與 16 筆的平均值，所以

$$x_{0.5} = \frac{-0.003 + 0.001}{2} = -0.001。$$

第三四分位（75 百分位）： $i = 0.75 \times 30 = 22.5$ ，所以第三四分位為排序第 23 筆數據， $x_{0.75} = x_{(23)} = 0.007$ 。

(4) $5\% = 0.05$ 的標準化值為

$$\frac{0.05 - (-0.003)}{0.0221} = \frac{0.053}{0.0221} = 2.40$$

這表示今日股市的漲幅位於平均值以上 2.4 個標準差的相對位置，根據經驗法則，這是個很高的漲幅。

4.7 抽樣調查某次測驗 30 位考生的成績如下：

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 550 | 600 | 490 | 510 | 640 | 620 | 580 | 658 | 556 | 543 |
| 520 | 510 | 480 | 635 | 606 | 620 | 617 | 592 | 587 | 561 |
| 679 | 496 | 483 | 451 | 430 | 573 | 619 | 580 | 606 | 611 |

- (1) 計算樣本平均數。
- (2) 計算樣本變異數及標準差。
- (3) 計算考生成績的第 83 百分位數。
- (4)(3) 所計算的第 83 百分位的標準化值是多少？
- (5) 如果這次考試錄取 17%，你的成績為 606，請以標準化值判斷你的相對位置及是否會錄取。

解：

$$(1) \bar{X} = \frac{\sum x_i}{n} = \frac{550 + \dots + 611}{30} = 566.77$$

$$(2) S^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$$

$$= \frac{(550 - 566.77)^2 + \dots + (611 - 566.77)^2}{29}$$

$$= 4021.15$$

$$S = \sqrt{4021.15} = 63.41$$

(3)先將考生成績排序

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 430 | 451 | 480 | 483 | 490 | 496 | 510 | 510 | 520 | 543 |
| 550 | 556 | 561 | 573 | 580 | 580 | 587 | 592 | 600 | 606 |
| 606 | 611 | 617 | 619 | 620 | 620 | 635 | 640 | 658 | 679 |

$i=0.83(30)=24.9$ ，所以考生成績的第 83 百分位數為排序第 25 的數據。也就是 $x_{0.83}=620$ 。

(4)第 83 百分位的標準化值為

$$\frac{620 - 566.7}{63.41} = 0.841$$

(5)錄取 17%，所以錄取最低成績為所有考生成績的第 83 百分位數，其標準化值為 0.841（見(4)）。

你的分數為 606，其標準化值為

$$\frac{606 - 566.77}{63.41} = 0.619$$

從 $0.619 < 0.841$ 研判你可能無法錄取。

4.8 台北市某月份日平均溫度記錄如下（單位： $^{\circ}\text{C}$ ）：

| | | | | | | | | | |
|----|------|------|------|------|------|----|------|----|------|
| 21 | 20.5 | 26 | 28.7 | 31 | 32.5 | 35 | 34.3 | 27 | 26.8 |
| 25 | 24 | 22.9 | 29 | 29.7 | 30.9 | 32 | 33 | 34 | 31 |
| 29 | 27.8 | 26 | 25 | 26.6 | 28 | 29 | 30 | 31 | 29.9 |

- (1)計算該月的平均溫度。
- (2)計算該月溫度的變異數及標準差。
- (3)計算該月溫度的第三四分位。
- (4)某日溫度為 28°C ，則當天溫度在全月的相對位置為何。

解：

$$(1) \bar{X} = \frac{\sum x_i}{n} = \frac{21 + \cdots + 29.9}{30} = 28.55$$

026 · 統計學習題解答

$$\begin{aligned}
 (2) S^2 &= \frac{\sum (x_i - \bar{X})^2}{n-1} \\
 &= \frac{(21 - 28.55)^2 + \dots + (29.9 - 28.55)^2}{29} \\
 &= 13.926
 \end{aligned}$$

$$S = \sqrt{13.926} = 3.73$$

(3) 先將溫度排序

| | | | | | | | | | |
|------|------|------|------|----|------|----|------|------|------|
| 20.5 | 21 | 22.9 | 24 | 25 | 25 | 26 | 26 | 26.6 | 26.8 |
| 27 | 27.8 | 28 | 28.7 | 29 | 29 | 29 | 29.7 | 29.9 | 30 |
| 30.9 | 31 | 31 | 31 | 32 | 32.5 | 33 | 34 | 34.3 | 35 |

$i = 0.75(30) = 22.5$ ，所以第三四分位為排序第 23 的數據值。也就是 $x_{0.75} = 31^\circ\text{C}$ 。

(4) 28°C 的標準化值為 $\frac{28 - 28.55}{3.73} = -0.147$ ，這表示當天的溫度所在的位置為低於全月平均溫度 0.14 個標準差。