

2-1 一維數據分析

重點整理

只有一個變量的數據，稱為一維數據。

例：甲、乙、丙三人身高分別為 168 公分、173 公分、164 公分，可以用一維數據 168, 173, 164 表示。

一、平均數

用來描述一群資料集中趨勢的指標，稱為平均數。

1. 算術平均數：

將 n 個數據相加後，再除以 n ，即可得到這 n 個數據的算術平均數（有時簡稱為平均數）。也就是說，數據 $x_1, x_2, x_3, \dots, x_n$ 的算術平均數為

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}。$$

例：3, 2, 8, 7 這 4 個數的算術平均數為 $\frac{3+2+8+7}{4} = 5$ 。

2. 加權平均數：

將 n 個數據分別乘以權數，相加以後，再除以總權數，即可得到這 n 個數據的加權平均數。假設有 n 個數據 $x_1, x_2, x_3, \dots, x_n$ 的權數分別是 $w_1, w_2, w_3, \dots, w_n$ ，則這 n 個數據的加權平均數為

$$W = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \dots + x_n w_n}{w_1 + w_2 + w_3 + \dots + w_n}。$$

例：小明 5 個學科的每週上課時數與段考分數如下表，試以每週上課時數為權數求加權平均。

科目	國文	英文	數學	物理	歷史
每週上課時數(小時)	4	4	4	2	2
段考成績(分)	85	83	80	78	82

將這 5 個學科的分數分別乘以時數，相加以後再除以時數總和，得

$$W = \frac{85 \times 4 + 83 \times 4 + 80 \times 4 + 78 \times 2 + 82 \times 2}{4 + 4 + 4 + 2 + 2} = \frac{1312}{16} = 82(\text{分})。$$

3. 幾何平均數：

(1) 將 n 個數據相乘後，再開 n 次方，即可得到這 n 個數據的幾何平均數。

也就是說，數據 $x_1, x_2, x_3, \dots, x_n$ (皆為正數) 的幾何平均數為

$$G = \sqrt[n]{x_1 x_2 \dots x_n}。$$

例：試求 24, 18, 27, 32 的幾何平均數。(四捨五入取到小數點後第一位)

按計算機可得 $G = \sqrt[4]{24 \times 18 \times 27 \times 32} \approx 24.71720573$ ，約為 24.7。

(2) 幾何平均數的概念可以用來計算平均成長率。

例：由於供需失調，某葉菜的價格連續三天的漲幅分別為 40 %、32.3 %、8 %（與前一天比較），試求此葉菜這三天價格的平均成長率（平均漲幅）。

這 3 天的價格分別是前一天的 140 %、132.3 %、108 %，

求它們的幾何平均再減 1，

$$\sqrt[3]{1.4 \times 1.323 \times 1.08} - 1 = 1.26 - 1 = 0.26 = 26 \%。$$

二、百分位數

1. 百分位數的意義：

一群數據的第 m 百分位數為 P_m (m 是正整數， $1 \leq m \leq 99$)，其意義是至少有 $m\%$ 的數據小於或等於 P_m ，且至少有 $(100 - m)\%$ 的數據大於或等於 P_m 。

例：一群數據的第 80 百分位數 $P_{80} = 60$ ，表示至少有 80 % 的數據小於或等於 60，至少有 20 % 的數據大於或等於 60。

例：中位數就是 P_{50} 。

2. 百分位數的算法：

假設一組資料有 n 筆數據，計算 P_m (第 m 百分位數) 時，先將這 n 筆數據由小而

大排序，並計算 $I = n \times \frac{m}{100}$ 。

(1) 當 I 不為整數時，取大於 I 的最小整數 M ，則 P_m 就是第 M 筆數據的值。

(2) 當 I 為整數時， P_m 為第 I 筆與第 $I + 1$ 筆數據的平均值。

例：有一組 50 筆數據的資料，由小到大排序依次為 4，8，12，16，……，
4*n*，……，196，200，恰為等差數列。試求：

(1) P_{85} 。 (2) P_{40} 。 (3) 中位數。

$$(1) I_1 = 50 \times \frac{85}{100} = 42.5, \text{ 取 } M = 43, \text{ 第 } 43 \text{ 筆數據為 } 172, \text{ 故 } P_{85} = 172。$$

$$(2) I_2 = 50 \times \frac{40}{100} = 20, \text{ 故 } P_{40} \text{ 是第 } 20 \text{ 筆與第 } 21 \text{ 筆數據的平均，}$$

$$\text{即 } 80、84 \text{ 的平均，得 } P_{40} = \frac{80 + 84}{2} = 82。$$

$$(3) \text{ 中位數即 } P_{50}, \text{ 取 } I_3 = 50 \times \frac{50}{100} = 25,$$

故 P_{50} 是第 25 項 100、第 26 項 104 的平均 102。

三、標準差

變異數與標準差可以衡量一群數據的差異性（數據分散的程度）。

假設一組數據 $x_1, x_2, x_3, \dots, x_n$ 的算術平均數為 μ ，定義離均差為 $x_i - \mu$ ，也就是每個數據與平均數的差值，它可能大於 0、等於 0 或小於 0。

1. 變異數 (σ^2)：

離均差平方的平均值，即 $\sigma^2 = \frac{1}{n}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2]$ 。

2. 標準差(σ):

變異數的正平方根, 即 $\sigma = \sqrt{\frac{1}{n}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2]}$ 。

例: 一組數據 1, 2, 3, 4, 5, 試求變異數與標準差。

這 5 個數的平均數 $\mu = \frac{1+2+3+4+5}{5} = 3$, 故得

變異數 $\sigma^2 = \frac{1}{5}[(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2] = 2$,

標準差 $\sigma = \sqrt{2}$ 。

四、數據的伸縮與平移

1. 伸縮:

將每個數據 x_i 同時乘以一個非零常數 a , 得到新的數據 y_i , 也就是 $y_i = ax_i$ 。

例: 數據 2, 7, 5, 3 同時乘以 2, 得 4, 14, 10, 6。

2. 平移:

將每個數據 x_i 同時加(或減)一個定數 b , 得到新的數據 y_i , 也就是 $y_i = x_i + b$ 。

例: 數據 3, 12, 9, 1 同時加 3, 得 6, 15, 12, 4。

3. 伸縮且平移:

將每個數據 x_i 同時乘以一個非零常數 a 後, 再同時加(或減)一個定數 b , 得到新的數據 y_i , 也就是 $y_i = ax_i + b$ 。

例: 數據 2, 7, 5, 3 同時乘以 2 再加 3 ($y_i = 2x_i + 3$) 後, 得 7, 17, 13, 9。

4. 伸縮平移對平均數(μ)與標準差(σ)的影響:

設原始資料 $x_1, x_2, \dots, x_n$, 令 $y_i = ax_i + b$, 其中 $i = 1, 2, \dots, n$ 。則:

(1) 數據 y_i 的平均數 $\mu_y = a\mu_x + b$ 。(平均數也要伸縮、平移)

(2) 數據 y_i 的標準差 $\sigma_y = |a|\sigma_x$ 。(標準差不受平移影響)

例: 已知數據 x_i 分別為 1, 2, 3, 4, 5, 平均數 $\mu = 3$, 標準差 $\sigma = \sqrt{2}$ 。

假設 $y_i = 2x_i + 3$, 試求數據 y_i 的平均數與標準差。

y_i 的平均數 $\mu_y = 2 \times 3 + 3 = 9$,

y_i 的標準差 $\sigma_y = |2|\sqrt{2} = 2\sqrt{2}$ 。

五、標準化數據

不同性質的數據經過標準化以後, 才能分析比較。

1. 標準化數據: $X_i = \frac{x_i - \mu}{\sigma}$ 。

例: 已知數據 1, 2, 3, 4, 5 的平均數 $\mu = 3$, 標準差 $\sigma = \sqrt{2}$ 。試求 1 與 4 的標準化數據。

1 的標準化數據為 $\frac{1-3}{\sqrt{2}} = -\sqrt{2}$ (約為 -1.4)。

4 的標準化數據為 $\frac{4-3}{\sqrt{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$ (約為 0.7)。

2. 標準化數據的性質:

標準化數據的算術平均數為 0, 標準差為 1。

例題 1 算術平均數

- (1) 試求 3, 5, -3, 10, -8, 11 的算術平均數。(5 分)
- (2) 小華連續五週的數學週考成績為 80, 90, 75, 100, 90, 試求這些成績的算術平均數。(5 分)

解

$$(1) \frac{1}{6} \times (3 + 5 - 3 + 10 - 8 + 11)$$

$$= \frac{1}{6} \times 18 = 3$$

$$(2) \frac{1}{5} \times (80 + 90 + 75 + 100 + 90)$$

$$= \frac{1}{5} \times 435 = 87$$

例題 2 加權平均數

小雯手中持有 A、B、C、D、E 五家公司的股票若干，已知上週五與本週五的股價及小雯持股數如下表，試求：

	A	B	C	D	E
上週五股價(單位：元/股)	38	36	72	18	19
本週五股價(單位：元/股)	36	40	76	20	20
小雯持股數(單位：千股)	3	3	2	9	8

- (1) 上週五小雯持股的平均股價。(5 分)
- (2) 本週五小雯持股的平均股價。(5 分)

解 小雯持股總數為 $3 + 3 + 2 + 9 + 8 = 25$ (千股)

(1) 持股均價須用加權平均計算

$$\frac{1}{25} \times (38 \times 3 + 36 \times 3 + 72 \times 2 + 18 \times 9 + 19 \times 8)$$

$$= \frac{1}{25} \times 680 = 27.2 \text{ (元)}$$

(2) 同理，本週五持股均價

$$\frac{1}{25} \times (36 \times 3 + 40 \times 3 + 76 \times 2 + 20 \times 9 + 20 \times 8)$$

$$= \frac{1}{25} \times 720 = 28.8 \text{ (元)}$$

例題 3 幾何平均數(一)

- (1) 試求 18, 28, 147 的幾何平均數。(5 分)
(2) 試求 36, 40, 150 的幾何平均數。(5 分)

解 (1) $G = \sqrt[3]{18 \times 28 \times 147}$
 $= \sqrt[3]{2 \times 3^2 \times 2^2 \times 7 \times 3 \times 7^2}$
 $= 2 \times 3 \times 7$
 $= 42$

(2) $G = \sqrt[3]{36 \times 40 \times 150}$
 $= \sqrt[3]{2^2 \times 3^2 \times 2^3 \times 5 \times 2 \times 3 \times 5^2}$
 $= \sqrt[3]{2^6 \times 3^3 \times 5^3}$
 $= 2^2 \times 3 \times 5$
 $= 60$

例題 4 幾何平均數(二)

某公司營業額連續三年的成長率依序為 10%、33.1%、21%，試求平均成長率。(10 分)

解 這三年的營業額分別是前一年的 110%，133.1%，121%

平均成長率為

$$\begin{aligned} & \sqrt[3]{1.1 \times 1.331 \times 1.21} - 1 \\ &= \sqrt[3]{1.1 \times 1.1^3 \times 1.1^2} - 1 \\ &= 1.21 - 1 \\ &= 0.21 \end{aligned}$$

故平均成長率為 21%

例題 5 百分位數(一)(累積相對次數折線圖)

右圖為一組一維數據的累積相對次數折線圖，試求：

(1) 第 75 百分位數(P_{75})。(5 分)

(2) 第 40 百分位數(P_{40})。(5 分)

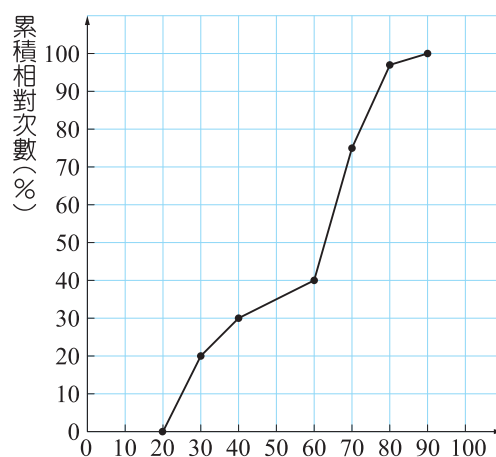
解 觀察累積相對次數折線圖，可知

橫軸 70 所對應的相對次數是 75 %

橫軸 60 所對應的相對次數是 40 %

(1) 第 75 百分位數 $P_{75} = 70$

(2) 第 40 百分位數 $P_{40} = 60$



例題 6 百分位數(二)

為配合畢業生申請國外研究所需要，我國許多大學已開始採用「等第績分平均(GPA)」來評估學生成績。所謂 GPA (Grade Point Average)，是透過轉換機制，將原始成績轉換為 0 至 4 分，0 代表不及格，其餘分數愈高，成績愈佳。已知某校一年級 873 名學生成績如下表，試求：

GPA	0	1	2	3	4
人數	63	104	269	312	125
累計人數	63	167	436	748	873

(1) 第 80 百分位數(P_{80})。(5 分)

(2) 第 25 百分位數(P_{25})。(5 分)

解 (1) $873 \times 80 \% = 698.4$

故 P_{80} 為累計第 699 人的成績，觀察已知條件，得 $P_{80} = 3$

(2) 同理， $873 \times 25 \% = 218.25$

取 P_{25} 為累計第 219 人的成績，得 $P_{25} = 2$

例題 7 變異數、標準差、數據的平移伸縮

- (1) 已知一組數據 2, 14, 11, 5, 20, 14, 試求：
- ① 變異數。(3 分)
 - ② 標準差。(3 分)
- (2) 已知一組數據 $x_i: 1, 3, 5, 7, 9$, 試求依下列規則變換後的新數據：
- ① $y = -2x + 30$ 。(2 分)
 - ② $y = 10(x - 2)$ 。(2 分)

解

(1) 算術平均數 $\mu = \frac{1}{6} \times (2 + 14 + 11 + 5 + 20 + 14) = 11$

① 變異數 $\sigma^2 = \frac{1}{6} \times [(-9)^2 + 3^2 + 0^2 + (-6)^2 + 9^2 + 3^2] = \frac{1}{6} \times 216 = 36$

② 標準差 $\sigma = \sqrt{36} = 6$

(2) ① $1, 3, 5, 7, 9$ 經 $y = -2x + 30$

轉換為 28, 24, 20, 16, 12

② $1, 3, 5, 7, 9$ 經 $y = 10(x - 2)$

轉換為 -10, 10, 30, 50, 70

例題 8 修正原始資料後平均數的改變

已知高一甲班有學生 32 人，某次段考數學科平均成績為 75 分。後來發現兩個錯誤，一個是輸入成績時，誤將 96 分登錄為 63 分，另一個是算錯總分，原來 45 分的考卷重新計算總分後，應得 60 分。試求修正這兩個成績以後，該班的數學平均成績。(10 分)

解 首先計算原始總分為 $75 \times 32 = 2400$

其次修正後總分為 $2400 - 63 + 96 - 45 + 60 = 2448$

故得該班數學平均成績為 $2448 \div 32 = 76.5$ (分)

例題 9 數據平移與伸縮對平均數與標準差的影響

這次段考因為數學成績不理想，經各班任課老師討論後，決定以原始分數除以 2 再加 50 分的方式調整。已知高一某班段考平均 38 分，標準差 5.6 分，試求該班調整後的：

(1) 平均分數。(5 分)

(2) 標準差。(5 分)

解 設原始分數為 x ，調整後分數為 y

$$\text{調整公式為 } y = \frac{x}{2} + 50$$

$$(1) \mu_y = \frac{38}{2} + 50 = 69 \quad \therefore \text{調整後的平均分數為 69 分}$$

$$(2) \sigma_y = \left| \frac{1}{2} \right| \sigma_x = \frac{1}{2} \times 5.6 = 2.8 \quad \therefore \text{調整後的標準差為 2.8 分}$$

例題 10 標準化數據

某公司在 A ， B 兩地分別招考員工，成績如右表。已知小明與小華都參加考試，請根據右表求解下列問題：

(1) 小明在 A 地參加考試得 74 分，試求小明的標準化數據。(5 分)

(2) 小華在 B 地參加考試，將得分標準化以後，得標準化數據為 0.5，試求小華的考試分數。(5 分)

	算術平均數	標準差
A 地	72	16
B 地	78	12

解 (1) 小明得分標準化數據為 $\frac{74 - 72}{16} = \frac{2}{16} = 0.125$

(2) 設小華的考試分數為 x

$$\frac{x - 78}{12} = 0.5 \Rightarrow x - 78 = 6 \Rightarrow x = 84$$

$\therefore$ 小華的考試分數為 84 分