

## 16.7. 探究不同情況對數據的離差之影響 (Exploring the Effect of the Different Operations on the Dispersion of the Data)

- 課程入面提到嘅“不同情況”主要包括：
  - 在數據中加入或剔除一項數據
  - 對所有數據一齊“加”或“減”同一個數值（其實“減”即係“加一個負數”）
  - 對所有數據一齊“乘”或“除”同一個數（其實“除”即係“乘一個細過 1 嘅數”）
  - 對所有數據又乘又加。

- 而要探究唔同情況對數據離差嘅影響，我哋首先當有一組由細排到大嘅數據喺度。

$$x_1, x_2, x_3, x_4, \dots, Q_1, \dots, Q_3, \dots, x_n$$

- 數據嘅分佈域 =  $x_n - x_1$
- 四分位數間距 =  $Q_3 - Q_1$
- 另設：平均數 =  $\bar{x}$       標準差 =  $\sigma$

### 16.7.1. 在數據中加入一項數據

- 當喺數據中加入一項數據嘅時候，我哋要留意嘅主要係數據排喺咩位度。
  - 除非新數據嘅值係大過原有“最大”或細過同“最細”嘅數據，否則分佈域唔會受到影響。
  - 而加入新數據後，四分位數嘅位置同值都可能有變。但一般嚟講，如果原有嘅數據量已經好多，咁多一個數據對四分位數唔會有明顯嘅影響。
  - 至於對標準差嘅影響，我哋要先睇返條公式：

$$\text{標準差} = \sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

$$\text{方差} = s = \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

- ◆ 如果“新數據同平均數嘅距離二次方”（即 $(x - \bar{x})^2$ ）大過方差，咁新嘅方差同新嘅標準差就會大咗。
  - 唔明？仲記唔記得講過方差係“數據同平均數嘅距離二次方”嘅平均數？咁如果新嘅一個數係大過舊方差，咁新方差咪會大咗囉！

### 16.7.2. 在數據中剔除一項數據

- 當數據中剔除一項數據的時候，我哋要留意嘅主要係數據排喺咩位置。
  - 除非剔除數據嘅值係唯一“最大”或“最細”嘅數據，否則分佈域唔會受到影響。
  - 而剔除一項數據後，四分位數嘅位置同值都可能有變。但一般嚟講，如果原有嘅數據量已經好多，咁少一個數據對四分位數唔會有明顯嘅影響。
  - 至於對標準差嘅影響，用返前面嘅推論方法。
    - ◆ 如果“剔除嘅數據同平均數嘅距離二次方”（即 $(x - \bar{x})^2$ ）大過方差，咁新嘅方差同新嘅標準差就會細咗。（留意係細咗！）
      - 唔明？你諗吓如果喺咁多個 $(x - \bar{x})^2$ 當中剔除一個較大嘅值，咁最後計返咁多個 $(x - \bar{x})^2$ 嘅平均數（即方差）係咪會細咗！
      - 仲唔明？咁你可以諗吓“有集中嘅數據同分散嘅數據喺一齊，咁如果我將較集中嘅數據剔除，咁 D 數據係咪會變得更分散”。

### 16.7.3. 對數據的每一項加上一個共同常數

- 當所有數據都加同一個數值（例如 a）嘅時候，數據會變成：

$$x_1+a, x_2+a, x_3+a, x_4+a, \dots, Q_1+a, \dots, Q_3+a, \dots, x_n+a$$

- 新數據嘅分佈域 =  $(x_n+a) - (x_1+a) = x_n - x_1$
- 新數據嘅四分位數間距 =  $(Q_3+a) - (Q_1+a) = Q_3 - Q_1$
- 新數據嘅平均收 =  $\bar{x} + a$
- 至於標準差，我哋要先睇返條公式：

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

- ◆ 留意“每個數都 +a”，而“平均數亦 +a”，所以每個  $(x_1 - \bar{x})$  嘅冇變到。
- ◆ 因此標準差嘅值都有變。
- ☆ 總結：將所有數據加同一個數係唔會改變數據組嘅分佈域、四分位數間距、標準差。
  - ◆ 其實咁都好合理。因為所有數據一齊郁，佢哋之間嘅“距離”都有變，所以離差都唔會有變。

### 16.7.4. 對數據的每一項乘以一個共同常數

- 當所有數據都乘同一個數值（例如 b）嘅時候，數據會變成：

$$bx_1, bx_2, bx_3, bx_4, \dots, bQ_1, \dots, bQ_3, \dots, bx_n$$

- 新數據嘅分佈域 =  $bx_n - bx_1 = b(x_n - x_1) = b \times$  舊嘅分佈域
- 新數據嘅四分位數間距 =  $bQ_3 - bQ_1 = b(Q_3 - Q_1) = b \times$  舊嘅四分位數間距
- 新數據嘅平均收 =  $b\bar{x}$
- 至於標準差，由條公式我哋亦可推論出：新數據嘅標準差 =  $b \times$  舊嘅標準差
- ☆ 總結：將所有數據都乘同一個數值，新數據組嘅分佈域、四分位數間距、標準差亦要乘返同一個數值。
  - ◆ 當數據被倍大或者縮小嘅時候，數據間嘅距離（即離差）亦會被倍大或縮小。

### 16.7.5. 對數據的每一項“又乘又加”

- 而有關將數據又乘又加嘅影響，我哋可以咁睇。
  - 唔理你係“先乘  $b$  後加  $a$ ”定“先加  $a$  後乘  $b$ ”，其實效果都係“乘完  $b$  再加一個數”。
  - ◆ 咁係因為  $b(x + a) = bx + ab = bx + c$
  - 而“乘完  $b$  再加一個數”當中只有“乘  $b$ ”會對數據組離差有影響。
  - 根據前面所講嘅嘢，我哋可以推論出：
    - ◆ 新數據嘅分佈域 =  $b \times$  舊嘅分佈域
    - ◆ 新數據嘅四分位數間距 =  $b \times$  舊嘅四分位數間距
    - ◆ 新數據嘅平均收 =  $b\bar{x} +$  “一個數”
      - 上面嘅“一個數”係咩就要睇係“先乘  $b$  後加  $a$ ”定“先加  $a$  後乘  $b$ ”。
    - ◆ 新數據嘅標準差 =  $b \times$  舊嘅標準差