

Chương 3: THỐNG KÊ CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG NGHIÊN CỨU

- ▷ **Số tuyệt đối thời kỳ** phản ánh quy mô, khối lượng của hiện tượng tại một *khoảng thời gian nhất định*.
Các số tuyệt đối thời kỳ của cùng một chỉ tiêu có thể cộng được với nhau để có trị số của thời kỳ dài hơn.

Ví dụ 38

Số lượng sản phẩm sản xuất ra của một doanh nghiệp trong:

- quý 1 năm 2018 là 10.000 SP,
- quý 2 năm 2018 là 20.000 SP.
- quý 3 năm 2018 là 30.000 SP.
- quý 4 năm 2018 là 40.000 SP.

⇒ Số lượng SP năm 2018 là: 100.000 SP.

- ▷ **Đơn vị tiền tệ** là đơn vị tính toán bằng các loại đồng tiền như đồng: Việt Nam, USD, Euro, ... để thể hiện giá trị của sản phẩm, dịch vụ.
Nó giúp cho việc tổng hợp, so sánh nhiều sản phẩm có giá trị sử dụng và đơn vị đo lường khác nhau, tuy nhiên, nó không có tính chất so sánh được qua thời gian!
- ▷ **Đơn vị thời gian lao động**: giờ công, ngày công,... được dùng để tính lượng lao động hao phí sản xuất ra những sản phẩm không thể tổng hợp hoặc so sánh với nhau bằng các đơn vị tính toán khác, hoặc những sản phẩm phức tạp do nhiều người cùng thực hiện qua nhiều giai đoạn.

3.1. Số tuyệt đối

Số tuyệt đối trong thống kê biểu hiện quy mô, khối lượng của hiện tượng kinh tế xã hội trong điều kiện *thời gian* và *địa điểm* cụ thể.

A. Các loại số tuyệt đối

- ▷ **Số tuyệt đối thời điểm** phản ánh quy mô, khối lượng của hiện tượng tại một *thời điểm nhất định*.
Ta không thể cộng các số tuyệt đối của cùng một chỉ tiêu tại hai thời điểm khác nhau, vì không có nghĩa.

Ví dụ 37

1. Số lao động của công ty A ngày 1-1-2019 là 80 người.
2. Giá trị tài sản cố định của công ty A vào ngày 1-2-2019 là 1 tỷ đồng, vào ngày 2-1-2019 là 1,1 tỷ đồng. Ta không thể nói giá trị TSCĐ của công ty trong 2 ngày là 2,1 tỷ?

B. Đơn vị tính:

- ▷ **Đơn vị hiện vật** là đơn vị tính toán phù hợp với đặc điểm của hiện tượng:
- mét, km, mét vuông, km², kg, tấn, lít, m³, số con, số người, ...
 - Đơn vị kép: Kw/h, tạ/ha, ...
 - Đơn vị quy chuẩn:
 - Máy lạnh có quy chuẩn 1,5HP, có nghĩa là các loại máy sử dụng đơn vị công suất khác sẽ được đưa về mã lực. (1 HP = 0,746 kW.)
 - Sản lượng lương thực (quy ra thóc) cả nước của tỉnh X năm 2000 là 3 triệu tấn.
- Ở đây thóc là sản phẩm quy đổi để quy đổi các loại cây lương thực khác ra thóc theo hệ số quy đổi của nhà nước (THÔNG TƯ SỐ: 815-TCTK/NN)
- 1kg thóc = 1kg ngô hạt khô = 3kg khoai tươi, sắn tươi
= 5 kg khoai riêng, khoai nước

3.2. Số tương đối

Số tương đối trong thống kê là chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng nghiên cứu.

Số tương đối so sánh hai mức độ cùng loại nhưng khác nhau về điều kiện không gian hoặc thời gian, hoặc so sánh hai hiện tượng khác loại nhưng có mối quan hệ với nhau

Số tương đối được biểu hiện bằng: số lần, %, người/km², ...

Ví dụ 39

1. $t = \frac{\text{doanh số bán hàng của công ty năm 2018}}{\text{doanh số bán hàng của công ty năm 2017}}$
2. Mật độ dân số (người/km²) = $\frac{\text{Dân số trung bình}}{\text{diện tích đất}}$
3. GDP trung bình đầu người (đ/người) = $\frac{GDP}{\text{dân số trung bình}}$

Các loại số tương đối

▷ **Số tương đối động thái** (tốc độ phát triển) là kết quả so sánh giữa hai mức độ của cùng hiện tượng nhưng khác nhau về thời gian.

Công thức:

$$t = \frac{y_1}{y_0}$$

Trong đó: t: Số tương đối động thái,

y_0 : mức độ của hiện tượng kỳ gốc,

y_1 : mức độ của hiện tượng kỳ nghiên cứu (kỳ báo cáo).

Ví dụ 40

Biết sản lượng sản xuất của công ty A qua hai năm như sau:

- Năm 2017: 100 tấn.

- Năm 2018: 145 tấn.

$$\Rightarrow \text{Tốc độ phát triển} = \frac{145}{100} = 1,45 = 145\%.$$

* Nếu ta tính số tương đối động thái với kỳ gốc y_0 thay đổi và kể ngay trước kỳ báo cáo, ta có các **số tương đối động thái liên hoàn**.
* Nếu ta tính số tương đối động thái với kỳ gốc y_0 cố định ta có các **số tương đối động thái định gốc**.

Ví dụ 42

Có tài liệu về doanh số bán hàng của công ty X qua các năm

Năm	2000	2001	2002	2003
Doanh số (tỷ đồng)	10,00	12,00	14,40	15,84

1. Tìm số tương đối động thái liên hoàn.

2. Tìm số tương đối động thái định gốc.

$$1. \frac{y_{2001}}{y_{2000}} = \frac{y_{2002}}{y_{2001}} = \frac{y_{2003}}{y_{2002}} =$$

$$2. \frac{y_{2001}}{y_{2000}} = \frac{y_{2002}}{y_{2000}} = \frac{y_{2003}}{y_{2000}} =$$

Ví dụ 43

Sản lượng cafe hạt A năm 2010 là 10 tấn, kế hoạch dự kiến năm 2011 là 15 tấn, thực tế năm 2011 đạt được là 12 tấn.

Số tương đối động thái: $t =$

Số tương đối nhiệm vụ kế hoạch: t_{NK}

Số tương đối hoàn thành kế hoạch: t_{HK}

$$\text{Mối liên hệ: } \frac{y_1}{y_0} = \frac{y_k}{y_0} \cdot \frac{y_1}{y_k}$$

Ví dụ 44

Nhà máy B chuyên sản xuất sản phẩm X. Năm 2003, nhà máy phần đầu hạ giá thành sản phẩm 2,5% và nâng cao sản lượng lên 10% so với năm 2002. Kết thúc năm 2003, nhà máy hoàn thành vượt mức kế hoạch hạ giá thành 2% và vượt mức K.H sản lượng 6%. Xác định biến động giá thành và biến động sản lượng năm 2003 so với năm 2002.

\Rightarrow sản phẩm sản xuất năm 2018 so với năm 2017 bằng 1,5 lần, hay tăng 50%.

Ví dụ 41

Vốn đầu tư xây dựng của 1 địa phương năm 2010 là 250 tỷ. Vốn đầu tư năm 2010 tăng 25% so với năm 2009. Tìm số vốn đầu tư năm 2009.

▷ Số tương đối kế hoạch

1. STĐ nhiệm vụ kế hoạch

Là tỷ lệ so sánh giữa mức độ kế hoạch đặt ra kỳ này với mức độ thực tế đạt được của chỉ tiêu ấy ở kỳ gốc

$$t_{NK} = \frac{y_k}{y_0}$$

trong đó

t_{NK} : STĐ nhiệm vụ kế hoạch,

y_k : mức kế hoạch.

y_0 : mức thực tế kỳ gốc

.

2. STĐ hoàn thành kế hoạch

Là tỷ lệ so sánh giữa mức độ thực tế đạt được trong kỳ nghiên cứu với mức độ kế hoạch đặt ra cùng kỳ của một chỉ tiêu nào đó.

$$t_{HK} = \frac{y_1}{y_k}$$

trong đó

t_{HK} : STĐ hoàn thành kế hoạch,

y_k : mức kế hoạch.

y_1 : mức thực tế kỳ nghiên cứu

$$\text{Mối liên hệ: } t = t_{NK} \cdot t_{HK} \Leftrightarrow \frac{y_1}{y_0} = \frac{y_k}{y_0} \cdot \frac{y_1}{y_k}$$

Giải

Xét chỉ tiêu giá thành, ta có

$$t_{NK} = \frac{y_k}{y_0} = 1 - 0,025 = 0,975 = 97,5\%.$$

$$t_{HK} = \frac{y_1}{y_k} = 1 - 0,02 = 0,98 = 98\%.$$

$$\Rightarrow t = t_{NK} \cdot t_{HK} = 0,975 \cdot 0,98 = 0,9555 = 95,55\%$$

Vậy, Giá thành sản phẩm X năm 2003 so với năm 2002 bằng 95,55%, nghĩa là giảm 4,45%.

Xét chỉ tiêu sản lượng, ta có

▷ **Số tương đối kết cấu** xác định tỉ trọng của mỗi bộ phận cấu thành tổng thể.

$$\text{Số tương đối kết cấu} = \frac{\text{trị số tuyệt đối bộ phận}}{\text{trị số tuyệt đối tổng thể}}$$

Ví dụ 45

Doanh nghiệp A có 2000 công nhân, trong đó có 1200 công nhân nam và 800 công nhân nữ.

Ta có số tương đối kết cấu như sau

$$\begin{aligned} \text{Tỉ lệ công nhân nam} &= \frac{1200}{2000} = 60\% \\ \text{Tỉ lệ công nhân nữ} &= \frac{800}{2000} = 40\% \end{aligned}$$

3.3. Các đặc trưng về tính hướng tâm

Trung bình

▷ **Giá trị trung bình** cung cấp một đại lượng đo lường vị trí trung tâm của dữ liệu.

Công thức

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

trong đó: - n số phần tử mẫu.

- x_i giá trị quan sát thứ i , $i \in \{1, \dots, n\}$.

Ví dụ 46

Thời gian giải một bài toán (tính theo phút) của 35 học sinh được ghi trong bảng sau

3	10	7	8	10	9	6	4	8	7	8	10
9	5	8	8	6	6	8	8	8	7	6	10
5	8	7	8	8	4	10	5	4	7	9	

▷ **Công thức trung bình có trọng số**

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}, \quad (2)$$

trong đó: - n số phần tử.

- x_i giá trị quan sát thứ i , $i \in \{1, \dots, n\}$.

- f_i trọng số thứ i , $i \in \{1, \dots, n\}$.

▷ **Trung bình cho trường hợp dữ liệu phân tổ**

$$\bar{x} = \frac{\sum_{i=1}^n m_i f_i}{\sum_{i=1}^n f_i}, \quad (3)$$

trong đó

- n số phần tử.

- m_i giá trị giữa của tổ i , $i \in \{1, \dots, n\}$. $m_i = \frac{x_{\max} - x_{\min}}{2}$.

- f_i tần số của tổ i , $i \in \{1, \dots, n\}$.

Ví dụ 47

Tài liệu về năng suất (tạ/ha) của 50 hộ nông dân trong bảng

Năng suất	Số hộ
30 - 35	5
35 - 40	10
40 - 45	20
45 - 50	12
50 - 55	3
Tổng	50

Cho bảng năng suất lao động trung bình của 800 lao động như sau

NSLD trung bình 1 lao động (tr đồng)	Số LD (f_i)	trị số giữa (x_i)	$x_i f_i$
100 - 200	74		
200 - 300	95		
300 - 400	142		
400 - 500	200		
500 - 600	165		
≥ 600	124		
Tổng	800		

Số trung bình điều hòa

Công thức

$$\bar{x} = \frac{\sum_{i=1}^n M_i}{\sum_{i=1}^n \frac{M_i}{x_i}}, \quad (4)$$

trong đó

- n : số phần tử mẫu.
- x_i : giá trị quan sát thứ i , $i \in \{1, \dots, n\}$.
- M_i : tổng giá trị quan sát ($M_i = x_i \cdot f_i$).

Ví dụ 48

Có tài liệu về năng suất lao động 3 tổ của 1 công ty như sau

Tổ	Năng suất lao động bình quân 1 công nhân (tần/người)	Sản lượng (tần)
1	11	220
2	12	264
3	13	312

Giải

$$\bar{x} = \frac{220 + 264 + 312}{\frac{220}{11} + \frac{264}{12} + \frac{312}{13}} \approx 12,0606$$
$$\left(\text{NS trung bình} = \frac{\text{tổng sản lượng}}{\text{tổng số công nhân}} \right)$$

Ví dụ 49

Bạn X đạp xe đạp từ A đến B với vận tốc $a(km/h)$, từ B đến C với vận tốc $b(km/h)$. Tính vận tốc trung bình từ A đến C.

Giải

Ví dụ 50

Công nhân X làm 1 loại sản phẩm. Lần thứ nhất làm mất 20 phút, lần thứ 2 làm mất 18 phút. Hỏi trung bình anh X làm một sản phẩm mất bao nhiêu thời gian.

Ví dụ 51

Khi quan sát 5 công nhân làm việc trong cùng một khoảng thời gian nhất định, ta thu được dữ liệu sau

- Công nhân 1 làm xong 1 sản phẩm hết 20 phút.
 - Công nhân 2 làm xong 1 sản phẩm hết 18 phút.
 - Công nhân 3 làm xong 1 sản phẩm hết 17 phút.
 - Công nhân 4 làm xong 1 sản phẩm hết 21 phút.
 - Công nhân 5 làm xong 1 sản phẩm hết 20 phút.
- Tính thời gian trung bình làm ra một sản phẩm của mỗi công nhân.

Giải

$$\bar{x} = \frac{5}{\frac{1}{20} + \frac{1}{18} + \frac{1}{17} + \frac{1}{21} + \frac{1}{20}} \approx 19,0841$$

Trung bình nhân

Số trung bình nhân thường được dùng trong trường hợp các lượng biến có quan hệ tích số với nhau. (thường dùng cho tốc độ phát triển, vd: tốc độ phát triển năng suất, dân số, ...).

▷ **Số trung bình nhân đơn giản**

$$\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (5)$$

trong đó

- n số phần tử mẫu.
- x_i giá trị quan sát thứ i , $i \in \{1, \dots, n\}$.

▷ **Số trung bình nhân có trọng số**

$$\bar{x} = \sqrt[k]{\prod_{i=1}^k x_i^{f_i}} = \sqrt[k]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}} \quad (6)$$

trong đó $\sum_{i=1}^k f_i = n$.

Ví dụ 52

Có tài liệu về giá trị sản xuất của một doanh nghiệp như sau:

- Tốc độ phát triển của năm 2016 so với năm 2015 là 115%.
 - Tốc độ phát triển của năm 2017 so với năm 2016 là 125%.
- Xác định tốc độ phát triển trung bình của doanh nghiệp từ năm 2015 đến 2017.

Giải $\bar{x} = \sqrt[3]{1,15 \cdot 1,25} = 1,118$

Ví dụ 53

Có tài liệu về giá trị sản xuất của một doanh nghiệp trong 10 năm như sau:

- 5 năm đầu mỗi năm có tốc độ phát triển là 110%.
- 3 năm kế tiếp mỗi năm có tốc độ phát triển là 115%.
- 2 năm cuối mỗi năm có tốc độ phát triển là 125%.

Tính tốc độ phát triển bình quân về giá trị sản xuất của doanh nghiệp trong 10 năm.

Dùng máy tính tính các đặc trưng mẫu (Fx 570VN Plus, Vinacal 570ES Plus II)

Bước chuẩn bị:

Xóa bộ nhớ: $\boxed{\text{Shift}} \Rightarrow \boxed{9} \Rightarrow \boxed{2} \Rightarrow \boxed{=}$

Khai báo cột tần số:

$\boxed{\text{Shift}} \Rightarrow \boxed{\text{Mode}} \Rightarrow \boxed{\downarrow} \Rightarrow \boxed{4}$ (chọn STAT) $\Rightarrow \boxed{1}$ (chọn ON)

Nhập dữ liệu

$\boxed{\text{Mode}} \Rightarrow \boxed{3}$ (chọn STAT) $\Rightarrow \boxed{1}$ (chọn 1-VAR)

nhập dữ liệu xong nhấn $\boxed{\text{AC}}$.

Xuất kết quả

$\boxed{\text{Shift}} \Rightarrow \boxed{1} \Rightarrow \boxed{4}$ (chọn var) \Rightarrow chọn kết quả cần xuất ($n, \bar{x}, \sigma_x, S_x$).

Mode

▷ **Mode** là giá trị xuất hiện thường xuyên nhất trong một nhóm dữ liệu.

Ví dụ 54

Không có Mode: 51 60 68 75 83 91
 Một trị Mode: 21 **30** **30** **30** 35 40 45 50
 Mode = **30**
 Hai trị Mode: 20 **21** **21** 30 33 **35** **35** 41
 Mode = **21** và Mode = **35**

Ví dụ 55

Điểm môn thi môn Lý thuyết thống kê của 1 lớp như sau

Điểm số	Số SV
4	10
5	15
6	30
7	52
8	15
9	2
10	1
Tổng	125

▷ Mode của bảng dữ liệu phân tổ đều

$$Mod = x_{Mod(min)} + h_{Mod} \cdot \frac{f_{Mod} - f_{Mod-1}}{(f_{Mod} - f_{Mod-1}) + (f_{Mod} - f_{Mod+1})}$$

Trong đó

- $x_{Mod(min)}$: giá trị dưới của tổ chứa Mod
(**tổ chứa Mod \equiv tổ có tần số lớn nhất**)
- h_{mod} : trị số khoảng cách tổ chứa Mode,
- f_{Mod} : tần số tổ chứa Mode,
- f_{Mod-1} : tần số tổ đứng trước tổ chứa Mode,
- f_{Mod+1} : tần số tổ đứng sau tổ chứa Mode.

Ví dụ 56

Cho tài liệu doanh số bán của 50 cửa hàng tiện lợi như sau

Doanh số (tr đồng)	Số cửa hàng
200 - 300	8
300 - 400	10
400 - 500	20
500 - 600	7
600 - 7 00	5
Tổng	50

$$Mod = \dots + \dots \cdot \frac{\dots - \dots}{(\dots - \dots) + (\dots - \dots)}$$

▷ Mode của bảng dữ liệu phân tổ không đều

$$Mod = x_{Mod(min)} + h_{Mod} \cdot \frac{f_{Mod} - f_{Mod-1}}{(f_{Mod} - f_{Mod-1}) + (f_{Mod} - f_{Mod+1})}$$

Trong đó

- $x_{Mod(min)}$: giá trị dưới của tổ chứa Mod
(**tổ chứa Mod \equiv tổ có mật độ phân phối $d_i = \frac{f_i}{h_i}$ lớn nhất**)
- h_{mod} : trị số khoảng cách tổ chứa Mode,
- f_{Mod} : tần số tổ chứa Mode,
- f_{Mod-1} : tần số tổ đứng trước tổ chứa Mode,
- f_{Mod+1} : tần số tổ đứng sau tổ chứa Mode.

Ví dụ 57

Cho tài liệu doanh số bán của 79 cửa hàng tiện lợi như sau

Doanh số (tr đồng)	Số cửa hàng (f_i)	Khoảng cách tổ (h_i)	Mật độ phân phối $d_i = \frac{f_i}{h_i}$
200 - 400	8		
400 - 500	12		
500 - 600	25		
600 - 800	25		
800 - 1000	9		
Tổng	79		

$$Mod = \dots + \dots \frac{\dots - \dots}{(\dots - \dots) + (\dots - \dots)}$$

Trung vị (Median, Med)

▷ **Trung vị** là giá trị đứng ở vị trí giữa khi các dữ liệu được sắp xếp theo thứ tự tăng dần.

Số trung vị chia dãy lượng biến làm hai phần, mỗi phần có số phần tử bằng nhau.

Với số quan sát lẻ, trung vị là giá trị chính giữa

$$Med = x_{\frac{n+1}{2}}$$

Với số quan sát chẵn, trung vị là giá trị trung bình của hai số ở giữa

$$Med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2}$$

Ví dụ 58

Số lượng sản phẩm lỗi bị trả lại trong vòng 21 ngày tại 1 cửa hàng điện máy là

3 4 4 9 8 8 6 4 7 9 1 3 5 3 5 9 8
6 3 7 1

- Tìm trung bình, trung vị, mode của tập dữ liệu trên.
- Ngày thứ 22 bị trả 9 sản phẩm. Tìm trung bình, trung vị, mode của tập dữ liệu mới.

Giải

$$1. \text{ Trung bình: } \bar{x} = \frac{3 + 4 + \dots + 7 + 1}{21} = \frac{113}{21} \approx 5,38.$$

$$- \text{ Trung vị: } n = 21 \Rightarrow \frac{21+1}{2} = 11. \text{ Med nằm tại vị trí 11.}$$

Xếp dữ liệu

1 1 3 3 3 3 4 4 4 5 5 6 6 7 7 8 8 8 9 9 9
Med

- Mode = 3.

▷ Trung vị của bảng dữ liệu phân tổ

$$Med = x_{Med(min)} + h_{Med} \cdot \frac{\frac{\sum f_i}{2} - S_{Med-1}}{f_{Med}}$$

Trong đó

- $x_{Med(min)}$: giá trị dưới của tổ chứa Med

(**tổ chứa Med** \equiv **tổ có tần số tích lũy đầu tiên** $\geq \frac{\sum f_i + 1}{2}$)

- h_{Med} : trị số khoảng cách tổ có Med,

- f_{Med} : tần số tổ có Med,

- S_{Med-1} : **tổng** tần số của các tổ đứng trước tổ có Med,

- $\sum f_i$: tổng tất cả các tần số.

Phân vị - Tứ phân vị

▷ **Phân vị thứ p** là một giá trị mà

- ít nhất có p phần trăm các quan sát có giá trị nhỏ hơn hoặc bằng giá trị này,

- và có ít nhất (1 - p) phần trăm các quan sát có giá trị lớn hơn hoặc bằng giá trị này.

Phương pháp tính phân vị thứ p

Bước 1: Sắp xếp dữ liệu theo thứ tự tăng dần.

Bước 2: Tính chỉ số i

$$i = \frac{p}{100} \cdot n$$

trong đó p là phân vị cần tính, n là số quan sát.

Bước 3:

- Nếu i không phải là một số nguyên, lấy phần nguyên trên của i là vị trí phân vị thứ p.

- Nếu i là số nguyên, phân vị thứ p là trung bình của các giá trị ở vị trí thứ i và i + 1

Ví dụ 60

Cho bảng khảo sát một mẫu gồm 12 sinh viên tốt nghiệp ngành quản trị kinh doanh như sau

SV	Lương khởi điểm (USD)	SV	Lương khởi điểm(USD)
1	3450	7	3490
2	3550	8	3730
3	3650	9	3540
4	3480	10	3925
5	3355	11	3520
6	3310	12	3480

Hãy tìm giá trị trung bình, trung vị của mức lương khởi điểm của các SV trên.

a. Hãy tính phân vị thứ 85.

b. Hãy tính phân vị thứ 50.

Giải. a.

Bước 1. Xếp xếp

3310 3355 3450 3480 3480 3490 3520 3540 3550 3650 3730 3925

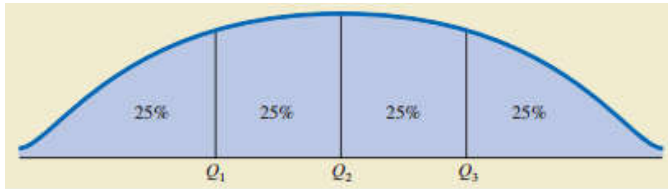
Bước 2. Tính i

$$i = \frac{85}{100} \cdot 12 = 10,2.$$

Bước 3. $i = 10,2$ không phải là số nguyên \Rightarrow vị trí của phân vị thứ 85 là vị trí 11. Ta có phân vị thứ 85 là 3730.

▷ **Tứ phân vị** chia dữ liệu thành 4 phần, mỗi phần chứa 25% số quan sát. Các điểm chia được gọi là tứ phân vị và được định nghĩa như sau

- Q_1 = tứ phân vị thứ nhất = phân vị thứ 25.
- Q_2 = tứ phân vị thứ hai = phân vị thứ 50.
- Q_3 = tứ phân vị thứ ba = phân vị thứ 75.



Ví dụ 61

Tìm tứ phân vị của dữ liệu trong ví dụ 60

3310 3355 3450 | 3480 3480 3490 | 3520 3540 3550 | 3650 3730 3925
 $Q_1=3465$ $Q_2 = 3505$ $Q_3=3600$
 Trung vị

▷ **Tứ phân vị của dữ liệu phân tổ có khoảng cách tổ**

Bước 1: xác định vị trí tứ phân vị 1 và 3.

Bước 2

- Tứ phân vị thứ nhất = $Q_1 = x_{Q_1(\min)} + h_{Q_1} \cdot \frac{\frac{1}{4} \sum f_i - S_{Q_1-1}}{f_{Q_1}}$
- Tứ phân vị thứ ba = $Q_3 = x_{Q_3(\min)} + h_{Q_3} \cdot \frac{\frac{3}{4} \sum f_i - S_{Q_3-1}}{f_{Q_3}}$

Trong đó

- $x_{Q_1(\min)}, x_{Q_3(\min)}$ giới hạn dưới của tổ chứa phân vị.
- h_{Q_1}, h_{Q_3} khoảng cách tổ chứa phân vị.
- S_{Q_1-1}, S_{Q_3-1} tần số tích lũy của các tổ đứng trước tổ chứa phân vị.
- f_{Q_1}, f_{Q_3} tần số tổ chứa phân vị.

Ví dụ 62

Tìm tứ phân vị 1, 3 trong ví dụ 59

3.4. Các đặc trưng về độ phân tán

▷ **Khoảng biến thiên**

Khoảng biến thiên (R) = Giá trị lớn nhất - Giá trị nhỏ nhất.

Ví dụ 63

Khoảng biến thiên trong ví dụ 60 là $3925 - 3310 = 615$

Khoảng biến thiên bị **ảnh hưởng** nhiều bởi các **giá trị đột biến**.

Ví dụ 64

Trong ví dụ 60, nếu có 1 SV có lương khởi điểm 10 000 USD, thì khoảng biến thiên sẽ là

$$10000 - 3310 = 6690,$$

con số này không mô tả tốt cho bảng dữ liệu đang xét.

Khoảng biến thiên càng nhỏ thì tổng thể càng đồng đều, số trung bình mang tính đại diện cao và ngược lại.

Ví dụ 65

Có hai tổ công nhân, mỗi tổ 5 người với mức năng suất lao động (Kg) như sau

Tổ 1: 200, 250, 300, 350, 400.

Tổ 2: 280, 290, 300, 310, 320.

Tìm giá trị trung bình, khoảng biến thiên của mỗi tổ

▷ Độ trải giữa (IQR)

$$IQR = Q_3 - Q_1, \quad (7)$$

trong đó Q_1, Q_3 là tứ phân vị thứ nhất và thứ 3.

Ví dụ 66

Cho tài liệu về tiền lương (triệu đồng) của hai tổ công nhân, mỗi tổ 11 người như sau

Tổ 1	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3,0	3,3	3,6	3,9
Tổ 2	1,9	2,0	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9

1. Tìm tứ phân vị của 2 tổ.

2. Tìm độ trải giữa của hai tổ.

$$R_1 = \dots > R_2 = \dots$$

⇒ mức lương trong tổ 1 biến thiên nhiều hơn mức lương trong tổ 2.

▷ **Độ lệch tuyệt đối trung bình** là số trung bình cộng của các độ lệch tuyệt đối giữa các biến và trị số trung bình cộng của các biến đó.

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

hoặc

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot f_i}{\sum_{i=1}^n f_i}$$

trong đó

- n số phần tử.
- x_i giá trị quan sát thứ i , $i \in \{1, \dots, n\}$.
- f_i trọng số thứ i , $i \in \{1, \dots, n\}$.
- \bar{x} : trung bình.

Ví dụ 67

Tính độ lệch tuyệt đối trung bình từng tổ trong ví dụ 65.

Chú ý:

- Độ lệch tuyệt đối trung bình càng nhỏ, tổng thể càng đồng đều, tính chất đại biểu của số trung bình \bar{x} càng cao.
- Độ lệch tuyệt đối trung bình có ưu điểm hơn khoảng biến thiên vì nó xét tất cả các biến.

Phương sai

▷ **Phương sai** là đại lượng đo lường sự phân tán của tất cả dữ liệu so với giá trị trung bình.

Công thức

Phương sai tổng thể σ^2 :

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \bar{x}^2 - \mu^2.$$

trong đó

- μ trung bình tổng thể,
- N số quan sát của tổng thể.

Phương sai mẫu s^2

	Phương sai mẫu	PS mẫu hiệu chỉnh
Không phân tổ	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
Phân tổ	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i - 1}$

trong đó

- \bar{x} trung bình mẫu,
- n số quan sát của mẫu.

Độ lệch chuẩn - hệ số biến thiên

▷ Độ lệch chuẩn

$$\text{Độ lệch chuẩn mẫu} = s = \sqrt{s^2}. \quad (8)$$

$$\text{Độ lệch chuẩn tổng thể} = \sigma = \sqrt{\sigma^2}. \quad (9)$$

Độ lệch chuẩn được đo bằng đơn vị như dữ liệu gốc, đó đó, độ lệch chuẩn dễ dàng sử dụng hơn.

▷ Hệ số biến thiên

Là đại lượng cho biết độ lệch chuẩn lớn hơn bao nhiêu lần so với trung bình, và được biểu diễn bằng đơn vị phần trăm.

$$\text{Hệ số biến thiên mẫu} = V = \frac{s}{\bar{X}}. \quad (10)$$

$$\text{Hệ số biến thiên tổng thể} = V = \frac{\sigma}{\mu}. \quad (11)$$

◀ ▶ 🔍 ↺ ↻ ⌂