

## Bài 13. Các số đặc trưng đo xu thế trung tâm

### A. Lý thuyết

#### 1. Số trung bình và trung vị

##### a) Số trung bình

Số trung bình (số trung bình cộng) của mẫu số liệu  $x_1, x_2, \dots, x_n$ , kí hiệu là  $\bar{x}$  được tính bằng công thức:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

**Ví dụ:** Kết quả thống kê số điểm đạt được sau mỗi lần bắn của một xạ thủ được ghi lại trong bảng sau:

7	9	8	9	7	10	9	9	7	8
---	---	---	---	---	----	---	---	---	---

Tính điểm số trung bình qua các lần bắn của xạ thủ.

##### Hướng dẫn giải

Đếm số phát súng xạ thủ đã bắn trong bảng trên, ta thấy xạ thủ đã bắn tổng cộng là 10 phát. Ta suy ra  $n = 10$ .

Lần thứ nhất xạ thủ bắn được 7 điểm. Do đó ta có  $x_1 = 7$ .

Lần thứ hai xạ thủ bắn được 9 điểm. Do đó ta có  $x_2 = 9$ .

Tương tự, ta được  $x_3 = 8, x_4 = 9, x_5 = 7, x_6 = 10, x_7 = 9, x_8 = 9, x_9 = 7, x_{10} = 8$ .

Suy ra, điểm số trung bình qua các lần bắn của xạ thủ là:

$$\frac{7 + 9 + 8 + 9 + 7 + 10 + 9 + 9 + 7 + 8}{10} = 8,3 \text{ (điểm)}$$

### Chú ý:

Trong trường hợp mẫu số liệu cho dưới dạng bảng tần số thì số trung bình được tính theo công thức:

$$\bar{X} = \frac{m_1x_1 + m_2x_2 + \dots + m_kx_k}{n}$$

trong đó  $m_k$  là tần số của giá trị  $x_k$  và  $n = m_1 + m_2 + \dots + m_k$ .

**Ví dụ:** Kết quả điều tra về số con của một số hộ gia đình trong một tổ dân phố được ghi lại trong bảng sau:

Số con	0	1	2	3	4
Số hộ gia đình	4	4	8	3	1

Hỏi trung bình mỗi hộ gia đình trong tổ dân phố có bao nhiêu con?

### Hướng dẫn giải

Tổng số hộ gia đình là:  $n = 4 + 4 + 8 + 3 + 1 = 20$  (hộ gia đình).

Trung bình mỗi hộ gia đình trong tổ dân phố có số con là:

$$\frac{0.4 + 1.4 + 2.8 + 3.3 + 4.1}{20} = 1,65 \text{ (con)}$$

**Ý nghĩa:** Số trung bình là giá trị trung bình cộng của các số trong mẫu số liệu, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

### b) Trung vị

Trong trường hợp mẫu số liệu có giá trị bất thường (rất lớn hoặc rất bé so với đa số các giá trị khác), người ta không dùng số trung bình để đo xu thế trung tâm mà dùng trung vị.

Để tìm trung vị (kí hiệu là Me) của một mẫu số liệu, ta thực hiện như sau:

+ Sắp xếp các giá trị trong mẫu số liệu theo thứ tự không giảm.

+ Nếu số giá trị của mẫu số liệu là số lẻ thì giá trị chính giữa của mẫu là trung vị. Nếu là số chẵn thì trung vị là trung bình cộng của hai giá trị chính giữa của mẫu.

**Ví dụ:** Theo dõi thời gian giải một bài toán của 4 học sinh, giáo viên nhận thấy có 2 em giải bài toán trong 2 phút; 1 em giải bài toán trong 3 phút và 1 em giải bài toán trong 7 phút. Hãy tìm số trung bình và trung vị của mẫu số liệu trên.

### Hướng dẫn giải

+ Số trung bình là:  $\frac{2.2 + 3.1 + 7.1}{4} = 3,5$  (phút)

+ Trung vị:

Sắp xếp số liệu theo thứ tự không giảm:

2      2      3      7

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa lần lượt là 2 và 3.

Vậy trung vị của mẫu số liệu là:  $\frac{2+3}{2} = 2,5$ .

**Nhận xét:** Trong mẫu số liệu được sắp xếp trên, số phần tử ở bên trái trung vị và số phần tử ở bên phải trung vị bằng nhau và bằng 2. Một học sinh giải bài toán

mất 7 phút cao hơn hẳn số trung bình, đây chính là giá trị bất thường. Nếu ta thay thời gian giải bài toán của học sinh giải mất 7 phút thành 8; 9; 10;... (phút) thì trung vị vẫn không thay đổi trong khi số trung bình sẽ thay đổi.

**Ý nghĩa:** Trung vị không bị ảnh hưởng bởi giá trị bất thường trong khi số trung bình bị ảnh hưởng bởi giá trị bất thường. Vì vậy, khi mẫu số liệu có giá trị bất thường, người ta thường dùng trung vị đại diện cho các số liệu thống kê.

**Ví dụ:** Đo chiều cao (đơn vị cm) của 9 học sinh lớp 10A và được kết quả như bảng sau:

149	153	155	153	150	188	148	151	150
-----	-----	-----	-----	-----	-----	-----	-----	-----

Tìm số trung bình và trung vị của mẫu số liệu trên. Trong hai số đó, số nào phù hợp hơn để đại diện cho chiều cao của 9 học sinh lớp 10A?

### Hướng dẫn giải

+ Số trung bình là:  $\frac{149 + 153 + 155 + 153 + 150 + 188 + 148 + 151 + 150}{9} = 155,22$

(cm)

+ Trung vị:

Sắp xếp số liệu theo thứ tự không giảm:

148    149    150    150    151    153    153    155    188

↓

Giá trị chính giữa

Vậy trung vị của mẫu số liệu là: 151 (cm).

+ Giữa hai số trung bình và số trung vị như trên, số trung vị bằng 151 (cm) phù hợp hơn để đại diện cho chiều cao của 9 học sinh lớp 10A vì trong mẫu số liệu có một em học sinh có chiều cao 188cm, đây là giá trị bất thường.

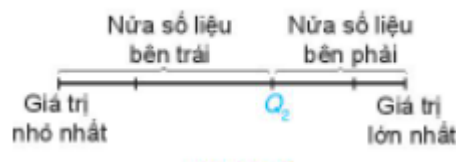
## 2. Tứ phân vị

Tứ phân vị dùng để xác định ngưỡng để phân loại các số liệu có trong mẫu số liệu.

Để tìm các tứ phân vị của mẫu số liệu có  $n$  giá trị, ta làm như sau:

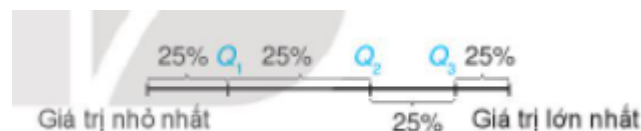
- + Sắp xếp mẫu số liệu theo thứ tự không giảm.
- + Tìm trung vị. Giá trị này là  $Q_2$ .
- + Tìm trung vị của nửa số liệu bên trái  $Q_2$  (không bao gồm  $Q_2$  nếu  $n$  lẻ). Giá trị này là  $Q_1$ .
- + Tìm trung vị của nửa số liệu bên phải  $Q_2$  (không bao gồm  $Q_2$  nếu  $n$  lẻ). Giá trị này là  $Q_3$ .

$Q_1$ ,  $Q_2$ ,  $Q_3$  được gọi là các **tứ phân vị** của mẫu số liệu.



**Chú ý:**  $Q_1$  được gọi là tứ phân vị thứ nhất hay tứ phân vị dưới,  $Q_3$  được gọi là tứ phân vị thứ ba hay tứ phân vị trên.

**Ý nghĩa:** Các điểm  $Q_1$ ,  $Q_2$ ,  $Q_3$  chia mẫu số liệu đã sắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần đều chứa 25% giá trị.



**Ví dụ:** Hàm lượng Protein (đơn vị gam) trong 100g của một số loại thực phẩm được cho trong bảng sau:

8,6	7,9	4,1	6,1	27,5	4	0,9	7,9	3,2	1,7
1,1	1,5	0,8	1,2	1,5	2	0	0,7	2,2	24,2

Hãy tìm các tứ phân vị. Các tứ phân vị này cho ta thông tin gì?

### Hướng dẫn giải

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

0 0,7 0,8 0,9 1,1 1,2 1,5 1,5 1,7 2 2,2 3,2 4 4,1 6,1 7,9 7,9 8,6 24,2  
27,5

Hai giá trị chính giữa

+ Vì  $n = 20$  là số chẵn nên  $Q_2$  là trung bình cộng của hai giá trị chính giữa. Do đó:

$$Q_2 = (2 + 2,2) : 2 = 2,1.$$

+ Ta tìm  $Q_1$  là trung vị của nửa số liệu bên trái  $Q_2$ :

0 0,7 0,8 0,9 1,1 1,2 1,5 1,5 1,7 2

Hai giá trị chính giữa

Vì lúc này  $n = 10$  là số chẵn nên  $Q_1$  là trung bình cộng của hai giá trị chính giữa. Do đó:

$$Q_1 = (1,1 + 1,2) : 2 = 1,15.$$

+ Ta tìm  $Q_3$  là trung vị của nửa số liệu bên phải  $Q_2$ :

2,2 3,2 4 4,1 6,1 7,9 7,9 8,6 24,2 27,5

Hai giá trị chính giữa

Vì lúc này  $n = 10$  là số chẵn nên  $Q_3$  là trung bình cộng của hai giá trị chính giữa.  
Do đó:

$$Q_3 = (6,1 + 7,9) : 2 = 7.$$

Vậy các tứ phân vị của mẫu số liệu là  $Q_1 = 1,15$ ;  $Q_2 = 2,1$ ;  $Q_3 = 7$ .



Các tứ phân vị cho ta hình ảnh phân bố của mẫu số liệu. Khoảng cách từ  $Q_1$  đến  $Q_2$  là  $2,1 - 1,15 = 0,95$  trong khi khoảng cách từ  $Q_2$  đến  $Q_3$  là  $7 - 2,1 = 4,9$ . Điều này cho thấy mẫu số liệu tập trung với mật độ cao ở bên trái của  $Q_2$  và mật độ thấp ở bên phải của  $Q_2$ .

### 3. Một

Mốt của mẫu số liệu là giá trị xuất hiện với tần số lớn nhất.

**Ý nghĩa:** Có thể dùng một để đo xu thế trung tâm của mẫu số liệu khi mẫu số liệu có nhiều giá trị trùng nhau.

**Ví dụ:** Kết quả thống kê điểm số bài kiểm tra giữa kỳ của một số học sinh lớp 10B được cho trong bảng sau:

3	5	6	7	7	7	8	8	9	10
---	---	---	---	---	---	---	---	---	----

Tìm một cho mẫu số liệu này.

#### Hướng dẫn giải

Số học sinh đạt điểm 3: 1 học sinh.

Số học sinh đạt điểm 5: 1 học sinh.

Số học sinh đạt điểm 6: 1 học sinh.

Số học sinh đạt điểm 7: 3 học sinh.

Số học sinh đạt điểm 8: 2 học sinh.

Số học sinh đạt điểm 9: 1 học sinh.

Số học sinh đạt điểm 10: 1 học sinh.

Vì số học sinh đạt điểm 7 là lớn nhất (có 3 học sinh) nên một của mẫu số liệu này là 7.

### **Nhận xét:**

+ Một có thể không là duy nhất. Chẳng hạn, với mẫu số liệu sau:

6 9 5 9 9 7 5 5 9 5

Ta thấy các số 5; 9 đều xuất hiện với số lần lớn nhất (4 lần) nên mẫu số liệu này có hai một là 5 và 9.

+ Khi các giá trị trong mẫu số liệu xuất hiện với tần số như nhau thì mẫu số liệu không có một. Chẳng hạn, với mẫu số liệu sau:

6 8 6 7 8 6 7 7 8

Ta thấy các giá trị 6; 7; 8 trong mẫu số liệu đều xuất hiện với tần số như nhau (3 lần) nên mẫu số liệu này không có một.

+ Một còn được định nghĩa cho mẫu dữ liệu định tính (dữ liệu không phải là số). Ví dụ trong buổi biểu quyết chọn một trong ba bạn Hoa, Bình, Tú làm bí thư của lớp 10C, bạn thư ký của lớp đã tổng kết được kết quả biểu quyết như sau:



Tên bạn được chọn	Hoa	Bình	Tú
Số phiếu	18	12	10

Trong mẫu dữ liệu này, số phiếu chọn “bạn Hoa” nhiều nhất, được gọi là **mốt**.

## B. Bài tập tự luyện

### B1. Bài tập trắc nghiệm

**Câu 1.** Một xạ thủ bắn 30 viên đạn vào bia kết quả được ghi lại trong bảng phân bố như sau

Điểm	6	7	8	9	10
Số lần	4	3	8	9	6

Khi đó điểm trung bình cộng là (kết quả làm tròn đến hàng phần trăm)

A. 8,33;

B. 8,34;

C. 8,31;

D. 8,32.

**Hướng dẫn giải**

**Đáp án đúng là: A**

$$\text{Ta có } \bar{x} = \frac{6.4 + 7.3 + 8.8 + 9.9 + 10.6}{30} = 8,33333333...$$

Điểm trung bình cộng làm tròn đến hàng phần trăm là 8,33.

**Câu 2.** Điểm kiểm tra môn Toán của 10 học sinh được cho như sau

6; 7; 7; 6; 7; 8; 8; 7; 9; 9. Số trung vị của mẫu số liệu trên là

A. 6;

B. 7;

C. 8;

D. 9.

### Hướng dẫn giải

**Đáp án đúng là: B**

Ta sắp xếp số liệu theo thứ tự không giảm như sau: 6; 6; 7; 7; 7; 7; 8; 8; 9; 9.

Dãy số trên có tất cả 10 giá trị, 2 giá trị chính giữa bằng 7. Vậy số trung vị của mẫu số liệu trên là  $(7 + 7):2 = 7$ .

**Câu 3.** Cho mẫu số liệu thống kê: 5; 2; 1; 6; 7; 5; 4; 5; 9. Một của mẫu số liệu trên bằng

A. 6;

B. 7;

C. 5;

D. 9.

### Hướng dẫn giải

**Đáp án đúng là: C**

Ta có giá trị 5 xuất hiện nhiều nhất trong mẫu số liệu trên. Vậy một bằng 5

## B2. Bài tập tự luận

**Câu 1.** Tìm số trung bình, trung vị, một và tứ phân vị của mỗi mẫu số liệu sau đây:

a) Tuổi thọ (tính theo năm) của 6 loại bóng đèn được ghi lại như sau:

19    20    31    18    26    24

b) Số cân nặng (tính theo kg) của 9 học sinh lớp 10D được ghi lại như sau:

37    37    40    42    45    40    42    42    55

### Hướng dẫn giải

a) Tuổi thọ trung bình của 6 loại bóng đèn là:  $\frac{19+20+31+18+26+24}{6} = 23$

(năm)

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

18    19    20    24    26    31

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa lần lượt là 20 và 24.

Vậy trung vị của mẫu số liệu là:  $\frac{20+24}{2} = 22$ .

+ Vì các giá trị 18; 19; 20; 24; 26; 31 trong mẫu số liệu đều xuất hiện với tần số như nhau (1 lần) nên mẫu số liệu này không có mốt.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị  $Q_2$  chính là trung vị của mẫu số liệu. Do đó ta có  $Q_2 = 22$ .

Ta tìm  $Q_1$  là trung vị của nửa số liệu bên trái  $Q_2$ :

18    19    20  
          ↓  
Giá trị chính giữa

Giá trị chính giữa lúc này là 19. Ta suy ra  $Q_1 = 19$ .

Ta tìm  $Q_3$  là trung vị của nửa số liệu bên phải  $Q_2$ :

24    26    31  
          ↓  
Giá trị chính giữa

Giá trị chính giữa lúc này là 26. Ta suy ra  $Q_3 = 26$ .

Vậy các tứ phân vị của mẫu số liệu là  $Q_1 = 19$ ;  $Q_2 = 22$ ;  $Q_3 = 26$ .

b) Số cân nặng trung bình của 9 học sinh lớp 10D là:

$$\frac{37.2 + 40.2 + 42.3 + 45 + 55}{6} \approx 63,33 \text{ (kg)}$$

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

37 37 40 40 42 42 42 45 55  
↓  
Giá trị chính giữa

Trung vị của mẫu số liệu là: 42.

+ Ta thấy giá trị 42 xuất hiện với tần số cao nhất (3 lần) nên một của mẫu số liệu này là 42.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị  $Q_2$  chính là trung vị của mẫu số liệu. Do đó ta có  $Q_2 = 42$ .

Ta tìm  $Q_1$  là trung vị của nửa số liệu bên trái  $Q_2$ :

37    37 40    40

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa là 37 và 40. Do đó  $Q_1 = (37 + 40) : 2 = 38,5$ .

Ta tìm  $Q_3$  là trung vị của nửa số liệu bên phải  $Q_2$ :

42    42 45    55

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa là 42 và 45. Do đó  $Q_3 = (42 + 45) : 2 = 43,5$ .

Vậy các tứ phân vị của mẫu số liệu là  $Q_1 = 38,5$ ;  $Q_2 = 42$ ;  $Q_3 = 43,5$ .

**Câu 2.** Số lượng học sinh giỏi Quốc gia năm học 2019 – 2020 của 10 trường Trung học phổ thông được cho như sau:

0 0 9 0 0 4 0 0 0 6

a) Tìm số trung bình, một và các tứ phân vị của mẫu số liệu trên.

b) Giải thích tại sao tứ phân vị thứ nhất và trung vị trùng nhau.

### Hướng dẫn giải

a) Số lượng học sinh trung bình là:  $\frac{0.7 + 9 + 4 + 6}{10} = 1,9$  (học sinh)

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

0 0 0 0 0 0 0 4 6 9

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa cùng bằng 0.

Vậy trung vị của mẫu số liệu cũng bằng 0.

+ Ta thấy giá trị 0 xuất hiện với tần số cao nhất (7 lần) nên một của mẫu số liệu này là 0.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị  $Q_2$  chính là trung vị của mẫu số liệu. Do đó ta có  $Q_2 = 0$ .

Ta tìm  $Q_1$  là trung vị của nửa số liệu bên trái  $Q_2$ :

0 0 0 0 0  
          ↓

Giá trị chính giữa

Giá trị chính giữa lúc này là 0. Ta suy ra  $Q_1 = 0$ .

Ta tìm  $Q_3$  là trung vị của nửa số liệu bên phải  $Q_2$ :

0   0   4   6   9  
          ↓  
Giá trị chính giữa

Giá trị chính giữa lúc này là 4. Ta suy ra  $Q_3 = 4$ .

Vậy các tứ phân vị của mẫu số liệu là  $Q_1 = 0$ ;  $Q_2 = 0$ ;  $Q_3 = 4$ .

b) Vì mẫu số liệu tập trung với mật độ cao ở bên trái  $Q_2$ , tức là trung vị, nên tứ phân vị thứ nhất  $Q_1$  và trung vị  $Q_2$  trùng nhau.

**Câu 3.** Sức chứa của một số sân vận động tại Việt Nam đã được thống kê và ghi lại như trong bảng sau (số liệu gần đúng):

Sân vận động	Mỹ Đình	Cần Thơ	Thiên Trường	Việt Trì
Sức chỗ ngồi	40 000	50 000	30 000	20 000

Các giá trị số trung bình, trung vị và một bị ảnh hưởng như thế nào nếu bỏ đi số liệu sức chứa của sân vận động Cần Thơ?

### Hướng dẫn giải

+ Trước khi bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta có:

Sức chỗ ngồi trung bình là:

$$\frac{40\,000 + 50\,000 + 30\,000 + 20\,000}{4} = 35\,000.$$

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

$$20\ 000 \quad \underbrace{30\ 000 \quad 40\ 000} \quad 50\ 000$$

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa lần lượt là 30 000 và 40 000.

Do đó trung vị của mẫu số liệu là:  $\frac{30\ 000 + 40\ 000}{2} = 35\ 000$ .

Trong mẫu dữ liệu trên, sức chứa ở sân vận động Cần Thơ là lớn nhất.

Do đó mốt của mẫu dữ liệu trên là “Cần Thơ”.

+ Sau khi bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta có:

Sức chỗ ngồi trung bình là:

$$\frac{40\ 000 + 30\ 000 + 20\ 000}{3} = 30\ 000.$$

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

$$20\ 000 \quad 30\ 000 \quad 40\ 000$$

↓

Giá trị chính giữa

Dãy trên có giá trị chính giữa lần lượt là 30 000.

Do đó trung vị của mẫu số liệu là: 30 000.

Trong mẫu dữ liệu trên, sức chứa ở sân vận động Mỹ Đình là lớn nhất.

Do đó mốt của mẫu dữ liệu trên là “Mỹ Đình”.

+ Nếu bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta thấy:

Giá trị số trung bình và trung vị đều giảm 5000 (giảm từ 35 000 xuống còn 30 000).

Mốt của dữ liệu cũng bị thay đổi từ “Cần Thơ” sang “Mỹ Đình”.