

Bài 2. Các số đặc trưng đo xu thế trung tâm cho mẫu số liệu không ghép nhóm

A. Lý thuyết

I. Số trung bình cộng (Số trung bình)

1. Định nghĩa

Số trung bình cộng của một mẫu n số liệu thống kê bằng tổng của các số liệu chia cho số các số liệu đó. Số trung bình cộng \bar{x} của mẫu số liệu x_1, x_2, \dots, x_n là:

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

Ví dụ: Tìm số trung bình cộng của các số 13, 15, 17, 20.

Hướng dẫn giải

Mẫu trên có 4 số liệu.

Khi đó, số trung bình cộng là $\bar{x} = \frac{13 + 15 + 17 + 20}{4} = 16,25$.

Vậy trung bình cộng của các số đã cho là 16,25.

Nhận xét:

- Đối với bảng tần số:

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Số trung bình cộng \bar{x} của mẫu số liệu thống kê trong bảng phân bố tần số là:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n_1 + n_2 + \dots + n_k}.$$

- Đối với bảng tần số tương đối:

Giá trị	x_1	x_2	...	x_k
Tần số tương đối	f_1	f_2	...	f_k

Số trung bình cộng \bar{x} của mẫu số liệu thống kê trong bảng phân bố tần số tương đối là:

$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_k x_k,$$

trong đó $f_1 = \frac{n_1}{n}$, $f_2 = \frac{n_2}{n}$, ..., $f_k = \frac{n_k}{n}$, với $n = n_1 + n_2 + \dots + n_k$.

Ví dụ:

a) Thời gian giải một bài toán (đơn vị: phút) của 30 học sinh được ghi lại trong bảng tần số sau:

Thời gian (phút)	5	6	7	8	9	10	12	13	15
Số học sinh	5	4	2	3	4	1	3	5	3

Tính thời gian trung bình để giải bài toán trên.

b) Số cân nặng (đơn vị: kg) của 20 học sinh được ghi lại trong bảng tần số tương đối sau:

Cân nặng (kg)	28	29	30	35	37	42
Tần số tương đối	$\frac{1}{10}$	$\frac{3}{20}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{20}$

Hãy tính cân nặng trung bình của 20 học sinh.

Hướng dẫn giải

a) Thời gian trung bình để giải bài toán trên là

$$\bar{x} = \frac{5.5 + 4.6 + 2.7 + 3.8 + 4.9 + 1.10 + 3.12 + 5.13 + 3.15}{5 + 4 + 2 + 3 + 4 + 1 + 3 + 5 + 3} = 9,3.$$

Vậy thời gian trung bình để giải bài toán trên là 9,3 phút.

b) Ta có cân nặng trung bình của 20 học sinh là:

$$\bar{x} = \frac{1}{10}.28 + \frac{3}{20}.29 + \frac{1}{5}.30 + \frac{3}{10}.35 + \frac{1}{5}.37 + \frac{1}{20}.42 = 33,15.$$

Vậy cân nặng trung bình của 20 học sinh là 33,15 kg.

2. Ý nghĩa

Khi các số liệu trong mẫu ít sai lệch với số trung bình cộng, ta có thể giải quyết được vấn đề trên bằng cách lấy số trung bình cộng làm đại diện cho mẫu số liệu.

Ví dụ: Để dự báo lượng mưa trong tháng 8 tại Hà Nội người ta tiến hành đo lượng mưa của từng ngày của tháng 8 gồm 31 số liệu. Số trung bình cộng của mẫu số liệu đó được xem như lượng mưa trung bình tháng 8 tại Hà Nội. Thống kê lượng mưa trung bình tháng 8 tại Hà Nội trong nhiều năm liên tiếp sẽ cho ta những dự báo lượng mưa trung bình tháng 8 tại Hà Nội trong những năm sắp tới.

II. Trung vị

1. Định nghĩa

Sắp thứ tự mẫu số liệu gồm n số liệu thành một dãy không giảm (hoặc không tăng).

- Nếu n là số lẻ thì số liệu đứng ở vị trí thứ $\frac{n+1}{2}$ (số đứng chính giữa) gọi là *trung vị*.

- Nếu n là số chẵn thì số trung bình cộng của hai số liệu đứng ở vị trí thứ $\frac{n}{2}$ và $\frac{n}{2} + 1$ gọi là *trung vị*.

Trung vị kí hiệu là M_e .

Nhận xét:

- Trung vị không nhất thiết là một số trong mẫu số liệu và dễ tính toán.
- Khi các số liệu trong mẫu không có sự chênh lệch lớn thì số trung bình cộng và trung vị xấp xỉ nhau.

Ví dụ: Điểm kiểm tra Toán của 7 bạn học sinh tổ 1 lớp 10B như sau: 9; 5; 4; 5; 8; 7; 9. Tìm trung vị M_e của mẫu số liệu trên.

Hướng dẫn giải

- Sắp xếp số liệu của mẫu theo thứ tự không giảm:

4 5 5 7 8 9 9

- Xác định xem số các số liệu là chẵn hay lẻ để tìm số trung vị:

Mẫu có 7 số liệu. Giá trị chính giữa là 7. Vì thế, trung vị của mẫu là 7.

Vậy $M_e = 7$.

2. Ý nghĩa

Nếu những số liệu trong mẫu có sự chênh lệch lớn thì ta nên chọn thêm trung vị làm đại diện cho mẫu số liệu đó nhằm điều chỉnh một số hạn chế khi sử dụng số trung bình cộng. Những kết luận về đối tượng thống kê rút ra khi đó sẽ tin cậy hơn.

Ví dụ: Thời gian giải một bài tập (đơn vị: phút) của nhóm học sinh như sau:

20 3 2 5 6 1

Tính trung vị của mẫu và số trung bình cộng của mẫu.

Ta nên chọn trung vị hay số trung bình cộng để đại diện cho mẫu thì kết luận về thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn?

Hướng dẫn giải

Sắp xếp các số liệu theo thứ tự không giảm:

1 2 3 5 6 20

Mẫu có 6 số liệu, khi đó trung vị của mẫu là trung bình cộng của 3 và 5.

Ta có $M_e = \frac{3+5}{2} = 4$.

Trung bình cộng của mẫu số liệu: $\bar{x} = \frac{1+2+3+5+6+20}{6} \approx 6,2$.

Ta thấy nên lựa chọn trung vị $M_e = 4$ đại diện cho mẫu thì kết luận thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn.

Vậy trung vị của mẫu là $M_e = 4$; số trung bình cộng là 6,2 và nên lựa chọn trung vị $M_e = 4$ đại diện cho mẫu thì kết luận thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn.

III. Tứ phân vị

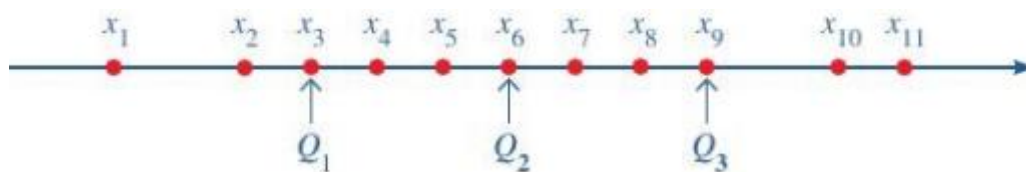
1. Định nghĩa

Sắp thứ tự mẫu số liệu gồm n số liệu thành một dãy không giảm.

Tứ phân vị của mẫu số liệu trên là bộ ba giá trị: tứ phân vị thứ nhất, tứ phân vị thứ hai và tứ phân vị thứ ba; ba giá trị này chia mẫu số liệu thành bốn phần có số lượng phần tử bằng nhau.

- Tứ phân vị thứ hai Q_2 bằng trung vị.
- Nếu n là số chẵn thì tứ phân vị thứ nhất Q_1 bằng trung vị của nửa dãy phía dưới và tứ phân vị thứ ba Q_3 bằng trung vị của nửa dãy phía trên.
- Nếu n là số lẻ thì tứ phân vị thứ nhất Q_1 bằng trung vị của nửa dãy phía dưới (không bao gồm Q_2) và tứ phân vị thứ ba Q_3 bằng trung vị của nửa dãy phía trên (không bao gồm Q_2).

Ta minh họa tứ phân vị của mẫu số liệu gồm 11 số liệu trên trục số như sau:



Ví dụ: Tìm tứ phân vị của mẫu số liệu sau:

21 32 10 45 11 35 24 8

Hướng dẫn giải

Mẫu số liệu trên được sắp xếp theo thứ tự tăng dần như sau:

8 10 11 21 24 32 35 45

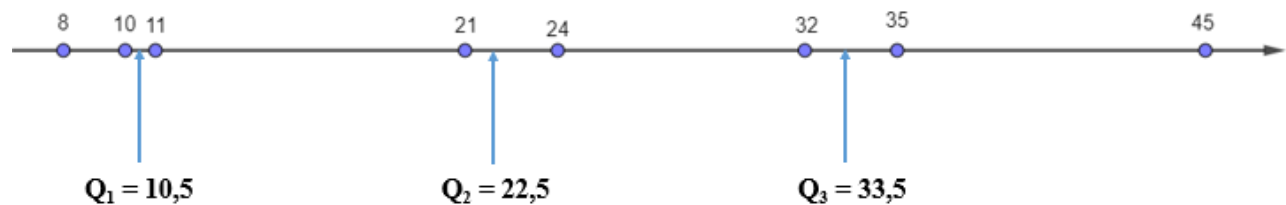
Dãy số liệu trên gồm 8 số liệu, là số chẵn.

Do đó ta có:

- Trung vị của mẫu số liệu trên là: $Q_2 = \frac{21+24}{2} = 22,5$.
- Trung vị của dãy 8, 10, 11, 21 là $Q_1 = \frac{10+11}{2} = 10,5$.
- Trung vị của dãy 24, 32, 35, 45 là $Q_3 = \frac{32+35}{2} = 33,5$.

Vậy $Q_1 = 10,5$, $Q_2 = 22,5$, $Q_3 = 33,5$.

Tứ phân vị đó được biểu diễn trên trục số như sau:



2. Ý nghĩa

- Trong thực tiễn, có những mẫu số liệu mà nhiều số liệu trong mẫu đó vẫn còn sự chênh lệch lớn so với trung vị. Ta nên chọn thêm những số khác cùng làm đại diện cho mẫu đó. Bằng cách lấy thêm trung vị của từng dãy số liệu tách ra bởi trung vị của mẫu nói trên, ta nhận được tứ phân vị đại diện cho mẫu số liệu đó.

- Bộ ba giá trị Q_1 , Q_2 , Q_3 trong tứ phân vị phản ánh độ phân tán của mẫu số liệu. Nhưng mỗi giá trị Q_1 , Q_2 , Q_3 lại đo xu thế trung tâm của phần số liệu tương ứng của mẫu đó.

IV. Mốt

1. Định nghĩa

Mốt của mẫu số liệu là giá trị có tần số lớn nhất trong bảng phân bố tần số và kí hiệu là M_o .

Chú ý: Một mẫu số liệu có thể có nhiều mốt.

Ví dụ: Cho bảng tần số sau:

Giá trị	1	2	3	4	5	6	7
Tần số	12	14	6	25	25	7	6

Tìm mốt của mẫu số liệu trên.

Hướng dẫn giải

Quan sát bảng tần số ta thấy giá trị 4 và 5 có tần số lớn nhất bằng 25.

Suy ra mốt của dấu hiệu là $M_o = 4$ và $M_o = 5$.

Vậy mốt của dấu hiệu là $M_o = 4$ và $M_o = 5$.

2. Ý nghĩa

Mốt của một mẫu số liệu đặc trưng cho số lần lặp đi lặp lại nhiều nhất tại một vị trí của mẫu số liệu đó. Dựa vào mốt, ta có thể đưa ra những kết luận (có ích) về đối tượng thống kê.

Ví dụ: Một cửa hàng bán 5 loại quạt với giá tiền là 150; 200; 350; 400; 500 (nghìn đồng). Số quạt bán ra trong mùa hè vừa qua được thống kê trong bảng sau:

Giá tiền	150	200	350	400	500
Số quạt bán được	25	80	100	123	75

Năm nay cửa hàng nên nhập nhiều số lượng loại quạt có giá tiền bao nhiêu để bán?

Hướng dẫn giải

Quan sát bảng thống kê trên ta thấy quạt có giá 400 nghìn đồng có số lượng bán được nhiều nhất, nghĩa là quạt giá 400 nghìn có tần số lớn nhất.

Suy ra $M_o = 400$.

Vậy năm nay của hàng nên nhập nhiều quạt có giá tiền 400 nghìn đồng về để bán.

V. Tính hợp lí của số liệu thống kê

Sau khi thu thập, tổ chức, phân loại và biểu diễn số liệu bằng bảng hoặc biểu đồ, ta cần phân tích và xử lí các số liệu đó để xem xét tính hợp lí của số liệu thống kê, đặc biệt chỉ ra được những số liệu bất thường (hay còn gọi là dị biệt, trong tiếng Anh là Outliers). Ta có thể sử dụng các số liệu đặc trưng đo xu thế trung tâm cho mẫu số liệu không ghép nhóm để thực hiện điều đó.

Ví dụ: Chiều cao của một nhóm học sinh nữ 6 tuổi (đơn vị cm) được ghi lại như sau:

100	112	113	115	115
116	118	117	120	122
122	125	123	162	97

a) Tìm tứ phân vị của mẫu trên.

b) Từ kết quả câu a) bước đầu xác định những giá trị bất thường của mẫu số liệu trên.

Hướng dẫn giải

a) Ta sắp xếp các số liệu theo thứ tự tăng dần như sau:

97	100	112	113	115
115	116	117	118	120
122	122	123	125	145

Mẫu trên có 15 số liệu.

Trung vị của mẫu số liệu trên là $Q_2 = 117$.

Trung vị nửa phía dưới 97, 100, 112, 113, 115, 115, 116 là $Q_1 = 113$.

Trung vị nửa phía trên 118, 120, 122, 122, 123, 125, 145 là $Q_3 = 122$.

Vậy tứ phân vị của mẫu là $Q_1 = 113$; $Q_2 = 117$; $Q_3 = 122$.

b) Dựa vào trung vị và tứ phân vị của mẫu số liệu, bước đầu ta thấy những số liệu bất thường trong mẫu là 97 và 145.

B. Bài tập tự luyện

B.1 Bài tập tự luận

Bài 1. Điểm số bài kiểm tra Toán của các bạn trong 2 tổ của lớp 10A được ghi lại trong bảng sau:

Tổ 1	8	7	8	8	7	9
Tổ 2	7	8	9	7	6	9

Tổ nào có kết quả tốt hơn? Tại sao?

Hướng dẫn giải

Ta có số bạn trong mỗi tổ 1 và 2 là 6 (người).

Điểm trung bình của tổ 1 là : $\overline{x_1} = \frac{8+7+8+8+7+9}{6} \approx 7,83$

Điểm trung bình của tổ 2 là : $\overline{x_2} = \frac{7+8+9+7+6+9}{6} \approx 7,67$

Vì $\overline{x_2} < \overline{x_1}$ ($7,67 < 7,83$) nên điểm trung bình của tổ 2 nhỏ hơn của tổ 1.

Vậy tổ 1 có kết quả bài kiểm tra tốt hơn tổ 2.

Bài 2. Cân nặng của 12 học sinh (đơn vị: kg) của một lớp 5 được cho trong bảng sau:

25	34	36	41	19	35
27	57	36	37	29	30

- a) Xác định trung vị, tứ phân vị của mẫu số liệu trên.
- b) Từ kết quả của câu a), bước đầu xác định những số liệu bất thường trong mẫu số liệu trên.

Hướng dẫn giải

- a) Ta sắp xếp các số liệu theo thứ tự tăng dần như sau:

19	25	27	29	30	34
35	36	36	37	41	51

Mẫu trên có 12 số liệu, là số chẵn

Trung vị của mẫu số liệu trên là $Q_2 = \frac{34 + 35}{2} = 34,5$.

Trung vị nửa phía dưới 19, 25, 27, 29, 30, 34 là $Q_1 = \frac{27 + 29}{2} = 28$.

Trung vị nửa phía trên 35, 36, 36, 37, 41, 51 là $Q_3 = \frac{36 + 37}{2} = 36,5$.

Vậy tứ phân vị của mẫu là $Q_1 = 28$; $Q_2 = 34,5$; $Q_3 = 36,5$.

- b) Dựa vào trung vị và tứ phân vị của mẫu số liệu, bước đầu ta thấy số liệu bất thường trong mẫu là 51.

Bài 3. Một cửa hàng thống kê số lượng một loại giày bán được trong 1 tháng trong bảng sau:

Size giày	35	36	37	38	39
Số giày bán được	12	50	31	32	22

a) Tìm một của mẫu số liệu trên?

b) Cửa hàng nên nhập nhiều hơn size giày nào để bán?

Hướng dẫn giải

a) Ta thấy số giày bán được của size giày 36 là lớn nhất bằng 50 nên một của mẫu số liệu là $M_o = 36$.

Vậy một của mẫu số liệu là $M_o = 36$.

b) Vì $M_o = 36$ nên số lượng bán được của size 36 là lớn nhất.

Do đó, cửa hàng nên nhập nhiều số lượng giày size 36.

Vậy cửa hàng nên nhập nhiều số lượng giày size 36.

B.2 Bài tập trắc nghiệm

Câu 1. Người ta thống kê cân nặng của 10 học sinh theo thứ tự tăng dần. Số trung vị của mẫu số liệu trên là:

A. Khối lượng của học sinh thứ 5;

B. Khối lượng của học sinh thứ 6;

C. Không tìm được trung vị;

D. Số trung bình cộng khối lượng của học sinh thứ 5 và thứ 6.

Hướng dẫn giải

Đáp án đúng là: D

Mẫu số liệu trên có 10 số (là số chẵn) nên trung vị là số trung bình cộng của hai số

liệu đứng ở vị trí thứ $\frac{10}{2} = 5$ và $\frac{10}{2} + 1 = 6$.

Nghĩa là trung vị của dãy số liệu trên là số trung bình cộng khối lượng của học sinh thứ 5 và thứ 6.

Câu 2. 100 học sinh tham dự kì thi học sinh giỏi toán (thang điểm là 20). Kết quả cho trong bảng sau:

Điểm (x)	9	10	11	12	13	14	15	16	17	18	19
Số học sinh (n)	1	1	3	5	8	13	19	24	14	10	2

Điểm trung bình của các học sinh dự thi môn toán là bao nhiêu?

- A. 15;
- B. 15,23;
- C. 15,50;
- D. 16.

Hướng dẫn giải

Đáp án đúng là: B

Điểm trung bình của các học sinh dự thi môn toán là:

$$\frac{9.1 + 10.1 + 11.3 + 12.5 + 13.8 + 14.13 + 15.19 + 16.24 + 17.14 + 18.10 + 19.2}{100}$$

$$= 15,23.$$

Câu 3. Cho mẫu số liệu thống kê: 6 5 5 2 9 10 8. Một và trung vị của mẫu số

liệu lần lượt là.

A. 5 và 6;

B. 8 và 6;

C. 6 và 5;

D. 2 và 10.

Hướng dẫn giải

Đáp án đúng là: A

Số 5 xuất hiện nhiều lần nhất trong mẫu số liệu (2 lần) nên $M_0 = 5$.

Sắp xếp mẫu số liệu theo thứ tự tăng dần 2 5 5 6 8 9 10

Mẫu số liệu có 7 số nên trung vị $M_e = 6$.