Bài 3. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

A. Lý thuyết

1. Số trung bình

1.1. Công thức tính số trung bình

• Giả sử ta có một mẫu số liệu là $x_1, x_2, \ldots, x_n$.

Số trung bình (hay ***số trung bình cộng)*** của mẫu số liệu này, kí hiệu là $\overline{x}$, được tính bởi công thức

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

• Giả sử mẫu số liệu được cho dưới dạng bảng tần số

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

Khi đó, công thức tính số trung bình trở thành

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{n}.$$

Trong đó $n = n_1 + n_2 + \ldots + n_k$. Ta gọi n là ***cỡ mẫu***.

Chú ý: Nếu kí hiệu $f_k = \frac{n_k}{n}$ là tần số tương đối (hay còn gọi là tần suất) của x_k trong mẫu số liệu thì số trung bình còn có thể biểu diễn là: $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_kx_k$.

Ví dụ: Điểm số bài thực hành môn Toán của các bạn học sinh trong nhóm A là 10; 5; 7; 9; 8; 6, còn của các bạn nhóm B là 9; 9; 8; 7; 6; 8. Tính điểm trung bình của mỗi nhóm.

Hướng dẫn giải

Điểm trung bình của nhóm A là: $\frac{1}{6}(10+5+7+9+8+6)=7,5$.

Điểm trung bình của nhóm B là: $\frac{1}{6}(9+9+8+7+6+8)\approx 7,83$.

1.2.Ý nghĩa của số trung bình

Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đo xu thế trung tâm của mẫu đó.

Ví dụ: Ở trong Ví dụ thuộc phần 1.1. trên, ta thấy điểm số trung bình của nhóm B cao hơn nhóm A (7,83 > 7,5), ta có thể nói rằng thành tích thực hành của nhóm B tốt hơn nhóm A.

2. Trung vị và tứ phân vị

2.1. Trung vị

2.1.1 Định nghĩa và cách tính số trung vị

Khi các số liệu trong mẫu số liệu chênh lệch nhau quá lớn, ta dùng một đặc trưng khác của mẫu số liệu, gọi là **trung vị** để so sánh các mẫu số liệu với nhau.

Trung vị được định nghĩa như sau:

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được:

$$x_1 \le x_2 \le \ldots \le x_n.$$

Trung vị của mẫu, kí hiệu là M_e, là giá trị ở chính giữa dãy $x_1, x_2, \ldots, x_n$. Cụ thể:

- Nếu n = 2k + 1, $k \in \mathbb{N}$ (tức n là số tự nhiên lẻ), thì trung vị của mẫu $M_e = x_{k+1}$.

- Nếu n = 2k, $k \in \mathbb{N}$ (tức n là số tự nhiên chẵn), thì trung vị của mẫu M_e = $\frac{1}{2}\left(x_k + x_{k+1}\right)$.

Ví dụ: Tính các trung vị của điểm thực hành môn Toán của các bạn học sinh trong nhóm A và nhóm B trong Ví dụ thuộc phần 1.1.

Hướng dẫn giải

+ Sắp xếp điểm số của mỗi bạn trong nhóm A theo thứ tự không giảm, ta được:

5; 6; 7; 8; 9; 10

Vì cỡ mẫu bằng 6 nên trung vị của nhóm A là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là M_e = $\frac{1}{2}\left(7+8\right) = 7,5$.

+ Sắp xếp điểm số của mỗi bạn trong nhóm B theo thứ tự không giảm, ta được:

6; 7; 8; 8; 9; 9

Vì cỡ mẫu bằng 6 nên trung vị của nhóm B là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là M_e = $\frac{1}{2}\left(8+8\right) = 8$.

2.1.2 Ý nghĩa của số trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có ít nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

Ví dụ: Bảng sau thống kê số sách mỗi bạn học sinh Tổ 1 và Tổ 2 đã đọc ở thư viện trường trong một tháng:

Tổ 1	3	1	2	1	2	2	3	25	1
Tổ 2	4	5	4	3	3	4	5	4	

a) Trung bình mỗi bạn Tổ 1 và mỗi bạn Tổ 2 đọc bao nhiêu quyển sách ở thư viện trường trong tháng đó?

b) Em hãy thảo luận với các bạn trong nhóm xem tổ nào chăm đọc sách ở thư viện hơn.

Hướng dẫn giải

a) Trung bình mỗi bạn Tổ 1 đọc số quyển sách ở thư viện trong tháng trên là:

$$\frac{3+1+2+1+2+2+3+25+1}{9}=\frac{40}{9}\approx 4,4.$$

Trung bình mỗi bạn Tổ 2 đọc số quyển số ở thư viện trong tháng trên là:

$$\frac{4+5+4+3+3+4+5+4}{8}=\frac{32}{8}=4.$$

b) Vì 4,4 > 4 nên theo số trung bình, các bạn Tổ 1 đọc sách chăm hơn.

Nếu dựa vào số trung bình để đánh giá xem tổ nào chăm đọc sách hơn trong bài này thì không phù hợp, do có một số liệu trong mẫu số liệu của Tổ 1 quá lớn so với các số liệu còn lại. Ta sử dụng trung vị để so sánh độ chăm học giữa hai tổ.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 1:

1; 1; 1; 2; 2; 2; 3; 3; 25

Vì cỡ mẫu $n_1 = 9$ là số lẻ, nên trung vị của mẫu số liệu Tổ 1 là $M_{e1} = 2$.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 2:

$$3; 3; 4; 4; 4; 4; 5; 5$$

Vì cỡ mẫu $n_2 = 8$ là số chẵn, nên trung vị của mẫu số liệu Tổ 2 là $M_{e2} = \frac{1}{2}(4+4) = 4$

.

Do đó ta có: $M_{e2} > M_{e1}$.

Vậy theo trung vị, các bạn Tổ 2 chăm đọc sách ở thư viện hơn Tổ 1.

2.2. Tứ phân vị

• Trung vị chia mẫu thành hai phần. Trong thực tế người ta cũng quan tâm đến trung vị của mỗi phần đó. Ba trung vị này được gọi là **tứ phân vị** của mẫu.

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được:

$$x_1 \le x_2 \le \ldots \le x_n.$$

Tứ phân vị của một mẫu số liệu gồm ba giá trị, gọi là **tứ phân vị thứ nhất, thứ hai và thứ ba** (lần lượt kí hiệu là Q_1, Q_2, Q_3). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

- Giá trị tứ phân vị thứ hai, Q_2, chính là số trung vị của mẫu.

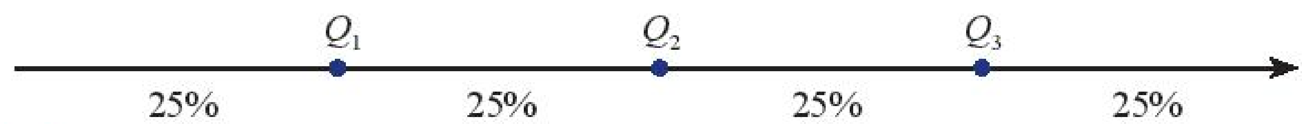
- Giá trị tứ phân vị thứ nhất, Q_1, là trung vị của nửa số liệu đã sắp xếp bên trái Q_2 (không bao gồm Q_2 nếu n lẻ).

- Giá trị tứ phân vị thứ ba, Q_3, là trung vị của nửa số liệu đã sắp xếp bên phải Q_2 (không bao gồm Q_2 nếu n lẻ).

• Ý nghĩa của tứ phân vị

Các điểm tứ phân vị 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 chia khoảng 25% tổng số liệu đã thu thập được.

Tứ phân vị thứ nhất Q_1 còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liệu phía dưới. Tứ phân vị thứ ba Q_3, còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu ở phía trên.



Ví dụ: Tìm tứ phân vị của mẫu số liệu sau: 3; 5; 6; 13; 25; 17; 19.

Hướng dẫn giải

Sắp xếp các số liệu theo thứ tự không giảm ta được:

3; 5; 6; 13; 17; 19; 25.

Vì cỡ mẫu n = 7, là số lẻ, nên giá trị tứ phân vị thứ hai là $Q_2 = 13$.

Tứ phân vị thứ nhất là trung vị của mẫu: 3; 5; 6. Do đó $Q_1 = 5$.

Tứ phân vị thứ ba là trung vị của mẫu: 17; 19; 25. Do đó $Q_3 = 19$.

3. Mốt

Cho mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu và kí hiệu là M_o.

Ý nghĩa của mốt: Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

Chú ý: Một mẫu số liệu có thể có rất nhiều mốt. Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

Ví dụ: Cho mẫu số liệu:

Giá trị	35	38	40	45
Tần số	10	5	6	3

Ta thấy giá trị 35 có tần số lớn nhất, do đó, mốt của mẫu số liệu trên là $M_o = 35$.

B. Bài tập tự luyện

Bài 1. Hãy tìm số trung bình, tứ phân vị và mốt của mẫu số liệu sau:

56; 45; 65; 45; 56; 78; 100; 78; 78.

Hướng dẫn giải

Cỡ mẫu: n = 9.

Số trung bình: $\overline{x} = \frac{1}{9}(56 + 45 + 65 + 45 + 56 + 78 + 100 + 78 + 78) \approx 66,78$.

Sắp xếp các số liệu theo thứ tự không giảm, ta được:

45; 45; 56; 56; 65; 78; 78; 78; 100.

Vì cỡ mẫu là 9, là số lẻ nên tứ phân vị thứ hai là $Q_2 = 65$.

Tứ phân vị thứ nhất là trung vị của mẫu: 45; 45; 56; 56. Do đó $Q_1 = \frac{1}{2}(45 + 56) =$ 50,5.

Tứ phân vị thứ ba là trung vị của mẫu: 78; 78; 78; 100. Do đó $Q_3 = \frac{1}{2}(78 + 78) =$ 78.

Giá trị 78 có tần số lớn nhất nên mốt của mẫu số liệu là $M_o = 78$.

Bài 2. Hãy tìm số trung bình, trung vị và mốt của mẫu số liệu sau:

Giá trị	20	25	30	35
Tần số	2	3	5	7

Hướng dẫn giải

Cỡ mẫu n = 2 + 3 + 5 + 7 = 17.

Số trung bình: $\overline{x} = \frac{1}{17}\left(2.20 + 3.25 + 5.30 + 7.35\right) = 30$.

Sắp xếp các số liệu đã cho theo thứ tự không giảm, ta được:

20; 20; 25; 25; 25; 30; 30; 30; 30; 30; 35; 35; 35; 35; 35; 35; 35.

Vì cỡ mẫu là 17 là số lẻ nên trung vị là $M_e = 30$.

Giá trị 35 có tần số lớn nhất nên mốt của mẫu số liệu là $M_o = 35$.

Bài 3. Trong một cuộc thi nghề, người ta ghi lại thời gian hoàn thành một sản phẩm của một số thí sinh ở bảng sau:

Thời gian (đơn vị: phút)	5	6	7	8	35
Số thí sinh	1	3	5	2	1

a) Hãy tìm số trung bình, tứ phân vị và mốt của thời gian thi nghề của các thí sinh trên.

b) Năm ngoái, thời gian thi của các thí sinh có số trung bình và trung vị đều bằng 7. Bạn hãy so sánh thời gian thi nói chung của các thí sinh trong hai năm.

Hướng dẫn giải

a) Cỡ mẫu là n = 1 + 3 + 5 + 2 + 1 = 12.

Số trung bình là: $\overline{x} = \frac{1.5 + 3.6 + 5.7 + 2.8 + 1.35}{12} \approx 9,08$.

Số thí sinh là trong thời gian 7 phút là nhiều nhất nên mốt của mẫu là $M_o = 7$.

Sắp xếp các giá trị của mẫu theo thứ tự không giảm, ta được:

5; 6; 6; 6; 7; 7; 7; 7; 7; 8; 8; 35.

Vì cỡ mẫu là số chẵn nên tứ phân vị thứ hai là $Q_2 = \frac{1}{2}(7+7) = 7$.

Tứ phân vị thứ nhất là trung vị của mẫu: 5; 6; 6; 6; 7; 7. Do đó $Q_1 = 6$.

Tứ phân vị thứ ba là trung vị của mẫu: 7; 7; 7; 8; 8; 35. Do đó $Q_3 = 7,5$.

b) Dựa theo số trung bình, vì 9,08 > 7 nên thời gian thi của các thí sinh năm nay nhiều hơn năm ngoái.

Dựa theo trung vị, thì cả hai năm trung vị đều bằng nhau và bằng 7 nên thời gian của các thí sinh trong hai năm là ngang nhau.

Vì trong mẫu số liệu của năm nay có số liệu 35 lớn hơn so với các số liệu còn lại rất nhiều, do đó ta dùng trung vị để so sánh sẽ phù hợp hơn.

Vậy thời gian thi nói chung của các thí sinh trong hai năm là ngang nhau.

Bài 4. Người ta đã tiến hành thăm dò ý kiến của khách hàng về các mẫu 1, 2, 3, 4 của một loại sản phẩm mới được sản xuất ở nhà máy. Dưới đây là bảng phân bố tần số theo số phiếu bình chọn tín nhiệm cho các mẫu kể trên.

Mẫu	1	2	3	4	Cộng
Tần số	195	300	356	149	1000

a) Tìm mốt của mẫu số liệu.

b) Trong sản xuất, nhà máy nên ưu tiên cho mẫu nào?

Hướng dẫn giải

a) Quan sát bảng tần số ta thấy mẫu 3 có tần số lớn nhất nên mốt của mẫu số liệu đã cho là $M_o = 3$.

b) Vì $M_o = 3$ nên trong sản xuất, nhà máy nên ưu tiên cho mẫu 3.