BÀI GIẢNG: CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU RỜI RẠC
CHUYÊN ĐỀ: CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU
MÔN: TOÁN 12
GIÁO VIÊN: ĐOÀN NGỌC DUY TRÌNH

MỤC TIÊU

- ✓ *Hiểu được khái niệm mẫu số liệu rời rạc và vai trò của các số đặc trưng trong mô tả dữ liệu.*
- ✓ *Tính được các số đặc trưng cho xu thế trung tâm và mức độ phân tán.*
- ✓ *Vận dụng được các số đặc trưng để nhận xét, so sánh và phân tích dữ liệu thực tế.*

I. KHÁI NIỆM MẪU SỐ LIỆU RỜI RẠC

Mẫu số liệu rời rạc (hay còn gọi là mẫu số liệu không ghép nhóm) là mẫu số liệu thường được trình bày dưới dạng bảng tần số có dạng như sau:

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

Chẳng hạn, điểm số bài kiểm tra môn toán của các bạn học sinh tổ 1 là 5; 6; 7; 7; 8; 8; 9; 9; 9; 10. Đây là mẫu số liệu rời rạc, được trình bày dưới dạng bảng tần số như sau:

Giá trị	5	6	7	8	9	10
Tần số	1	1	2	2	3	1

II. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU RỜI RẠC

1 Số trung bình

Cho một mẫu số liệu rời rạc, được trình bày dưới dạng bảng tần số như sau:

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

Khi đó, số trung bình của mẫu số liệu đã cho được tính bởi công thức:

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

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

Ví dụ 1: Sau khi thống kê số sách mà mỗi bạn trong lớp đã đọc trong năm 2023, An thu được kết quả như bảng bên dưới.

Số cuốn sách	1	2	3	4	5
Số bạn	3	6	15	10	6

Tính số trung bình của mẫu số liệu có trong bảng đã cho.

A. 3. **B.** $\frac{13}{4}$. **C.** 4. **D.** $\frac{11}{4}$.

Cách giải:

Số trung bình của mẫu số liệu có trong bảng đã cho là: $\overline{x}=\frac{1.3+2.6+3.15+4.10+5.6}{3+6+15+10+6}=\frac{13}{4}$

Chọn B.

Ví dụ 2: Cho biểu đồ kết quả kiểm tra môn Toán của một lớp học như hình vẽ bên. Số trung bình của mẫu số liệu ở biểu đồ này là



A. 7. **B.** 6,65.

C. 6. **D.** 7,35.

Cách giải:

Số trung bình của mẫu số liệu ở biểu đồ này là

$\overline{x}=\frac{4\cdot 2+5\cdot 3+6\cdot 5+7\cdot 11+8\cdot 10+9\cdot 6+10\cdot 3}{2+3+5+11+10+6+3}=7,35$

Chọn D.

❖ Ý nghĩa của số trung bình

Số trung bình (hay số trung bình cộng) 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.

2 Tiêu đề

a. Trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị, kí hiệu là M_e, 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.

Để xác định trung vị của mẫu số liệu không ghép nhóm, ta làm như sau:

Bước 1: Sắp xếp các giá trị trong mẫu số liệu theo thứ tự không giảm.

Bước 2: Nếu số giá trị của mẫu số liệu là:

- Số lẻ thì trung vị là giá trị chính giữa của mẫu.
- Số chẵn thì trung vị là trung bình cộng của hai giá trị chính giữa của mẫu.

Ví dụ 3: Khảo sát số ghế trống tại một rạp chiếu phim trong một khoảng thời gian, thu được mẫu số liệu sau:

Số ghế trống	7	8	9	10	11	12	15
Số ngày	4	5	6	8	4	1	1

Trung vị của mẫu số liệu nêu trên là

A. 15. **B.** 9. **C.** 10. **D.** 9,5.

Cách giải:

Cỡ mẫu $n = 29$

Trung vị của mẫu số liệu nêu trên là $M_e = x_{\frac{29+1}{2}} = x_{15}$

Chọn A.

Ví dụ 4: Dưới đây là bảng phân bố tần số về sản lượng chè thu được trong một năm của một số hộ gia đình.

Sản lượng (kg/sào)	111	112	113	114	115	116	117
Số hộ đạt được	1	4	5	4	3	2	1

Trung vị của mẫu số liệu nêu trên là

A. 113,5. **B.** 113. **C.** 114. **D.** 114,5.

Cách giải:

Cỡ mẫu: $n = 20$

Trung vị của mẫu số liệu nêu trên là: $Me = \frac{x_{20+1}}{2} = \frac{1}{2}\left(x_{10} + x_{11}\right) = \frac{1}{2}(113+114) = 113,5$

Chọn A.

b. Tứ vị phân

Trung vị chia mẫu số liệu thành hai phần. Trong thực tế, người ta cũng quan tâm đến trung vị của mỗi phần đó. Ba trung vị này chia tập hợp dữ liệu của mẫu khi sắp xếp theo thứ tự không giảm thành bốn phần đều nhau, được gọi là tứ phân vị của mẫu. Tứ phân vị của mẫu gồm có tứ phân vị thứ nhất, thứ hai và thứ ba (lần lượt kí hiệu là Q_1, Q_2, Q_3). Các điểm Q_1, Q_2, Q_3 chia mẫu số liệu đã sắp xếp theo thứ tự không giảm thành bốn phần, mỗi phần chứa khoảng 25% số lượng giá trị của mẫu.

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

Để tìm các tứ phân vị của mẫu số liệu rời rạc, ta làm như sau:

Bước 1: Sắp xếp mẫu số liệu theo thứ tự không giảm.

Bước 2: Tìm trung vị M_e của mẫu. Giá trị này cũng chính là tứ phân vị thứ hai Q_2.

Bước 3: Tìm tứ phân vị thứ nhất và tứ phân vị thứ ba.

- Tìm trung vị của nửa số liệu bên trái Q_2 (không bao gồm Q_2). Giá trị này là tứ phân vị thứ nhất Q_1
- Tìm trung vị của nửa số liệu bên phải Q_2 (không bao gồm Q_2). Giá trị này là tứ phân vị thứ ba Q_3

Ví dụ 5: Số nhân khẩu trong các hộ gia đình ở một xóm được thống kê ở bảng sau:

Số nhân khẩu	1	2	3	4	5	6
Số hộ gia đình	1	4	7	11	5	3

Tìm tứ phân vị thứ nhất của mẫu số liệu nêu trên.

A. 3. **B.** 2. **C.** 2,5. **D.** 3,5.

Cách giải:

Cỡ mẫu $n=31$

Ta có $M_2=Q_2=x_{\frac{31+1}{2}}=x_{16}$

Tứ phân vị thứ nhất của mẫu số liệu nêu trên là: $Q_1=x_{\frac{15+1}{2}}=x_8=3$

Chọn A.

Ví dụ 6: Một cơ sở chăn nuôi gia cầm tiến hành nuôi thử nghiệm giống gà đẻ trứng mới. Khi gà đã cho trứng, họ tiến hành khảo sát với một số quả trứng thì được cân nặng (gam) như sau:

40	42	36	38	40	42	29	48	43	43
41	41	39	44	45	41	40	39	42	41

Tứ phân vị thứ ba của mẫu số liệu nêu trên là

A. 43. **B.** 42. **C.** 42,5. **D.** 43,5.

Cách giải:

Bảng tần số:

Giá trị	29	36	38	39	40	41	42	43	44	45	48
Tần số	1	1	1	2	3	4	3	2	1	1	1

Cỡ mẫu: $n=20$

Tứ phân vị thứ 2: $Q_2=x_{\frac{20+1}{2}}=x_{10,5}$

$\Rightarrow Q_1=x_{\frac{10+1}{2}}=x_{5,5}$

$\Rightarrow Q_3=x_{20+1-5,5}=x_{15,5}$

$Q_3=\frac{1}{2}\left(x_{15}+x_{16}\right)=\frac{1}{2}(42+43)=42,5$

Chọn C.

3 Mốt

Cho một mẫu số liệu dưới dạng bảng tần số. Khi đó, mốt của mẫu số liệu, kí hiệu M_o, là giá trị có tần số lớn nhất. Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

Ví dụ 7: Để khảo sát kết quả thi tuyển sinh môn Toán trong kì thi tuyển sinh đại học năm vừa qua của trường A, người điều tra chọn một mẫu gồm 100 học sinh tham gia kì thi tuyển sinh đó. Điểm môn Toán (thang điểm 10) của các học sinh này được cho ở bảng phân bố tần số sau đây:

Điểm	0	1	2	3	4	5	6	7	8	9	10
Tần số	1	1	3	5	8	13	19	24	14	10	2

Tìm mốt của mẫu số liệu nêu trên.

A. 6. **B.** 24. **C.** 6,5. **D.** 7.

Cách giải:

Mốt của mẫu số liệu nêu trên là : $M_o = 7$

Chọn D.

Ví dụ 8: Thống kê cỡ giày trong tổng số 30 đôi bán được của tuần lễ vừa qua tại một cửa hàng X như sau:

37	39	38	39	40	40	42	41	38	39
39	40	43	41	37	39	38	39	42	39
37	39	43	38	39	40	39	39	38	42

Mốt của mầu số liệu trên là

A. 42. **B.** 39. **C.** 37. **D.** 11.

Cách giải:

Mốt của mầu số liệu trên là 39.

Chọn B.

III. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU RỜI RẠC

1 Khoảng biến thiên và khoảng tứ phân vị

Khoảng biến thiên của mẫu số liệu, kí hiệu R, là hiệu giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó, tức là $R = x_n - x_1$.

Khoảng tứ phân vị, kí hiệu Δ_Q, là hiệu giữa Q_3 và Q_1, tức là $\Delta_Q = Q_3 - Q_1$.

Phần tử x trong mẫu số liệu được gọi là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.

Ví dụ 9: Thống kê số bàn thắng ghi được bởi 20 cầu thủ có thành tích tốt hàng đầu của một giải đấu bóng đá trong một mùa giải nọ như sau:

Số bàn thắng	27	21	19	18	17	16	15	14	12	11
Số cầu thủ	1	2	3	1	1	3	2	2	3	2

Tính khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu trên.

A. 15 và 6. **B.** 16 và 5. **C.** 16 và 6. **D.** 15 và 5.

Cách giải:

Khoảng biến thiên: $R = x_n - x_1 = 27 - 11 = 16$

Ta có $Q_2 = x_{\frac{20+1}{2}} = x_{10,5}$

$$Q_1 = x_{\frac{10+1}{2}} = x_{5,5} = \frac{1}{2}\left(x_5 + x_6\right) = 19$$

$$Q_3 = x_{20+1} - 5,5 = x_{15,5} = \frac{1}{2}\left(x_{15} + x_{16}\right) = \frac{1}{2}(14+12) = 13$$

Khoảng tứ phân vị: $\Delta_Q = Q_3 - Q_1 = 19 - 13 = 6$

Chọn C.

Ví dụ 10: Thống kê điểm thi giữa kì môn Giải tích của các sinh viên năm nhất ở một lớp nọ theo bảng sau:

Điểm	0	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Số sinh viên	2	1	1	1	2	10	12	13	10	7	2

Số giá trị ngoại lệ của mẫu số liệu trên là

A. 5. **B.** 0. **C.** 3. **D.** 2.

Cách giải:

Cỡ mẫu: $n = 61$

Gọi x là giá trị ngoại lệ với $\left[\begin{array}{l} x > Q_3 + 1,5\Delta_Q \\ x < Q_1 - 1,5\Delta_Q \end{array}\right.$

Ta có $Q_2 = x_{\frac{6+1+1}{2}} = x_{31}$

$$Q_1 = x_{\frac{30+1}{2}} = x_{15,5} = \frac{1}{2}\left(x_{15} + x_{16}\right) = \frac{1}{2}(7,5+7,5) = 7,5$$

$$Q_3 = x_{61+1-15,5} = x_{46,5} = \frac{1}{2}(9+9) = 9$$

Khoảng tứ phân vị: $\Delta_Q = Q_3 - Q_1 = 9 - 7,5 = 1,5$

Với $\left[\begin{array}{l} x > Q_3 + 1,5\Delta_Q = 11,25 \text{ (L)} \\ x < Q_1 - 1,5\Delta_Q = 5,25 \end{array}\right.$

$\Rightarrow x < 5,25$. Có 2 giá trị ngoại lệ của mẫu số liệu trên

Chọn D.

2 Phương sai và độ lệch chuẩn

Giả sử ta có mẫu số liệu được cho dưới dạng bảng tần số như sau:

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

Khi đó:

- Phương sai của mẫu số liệu này, kí hiệu là S^2, được tính bởi công thức:

$$S^2 = \frac{1}{n}\left[n_1\left(x_1 - \overline{x}\right)^2 + n_2\left(x_2 - \overline{x}\right)^2 + \ldots + n_k\left(x_k - \overline{x}\right)^2\right] \text{ hoặc } S^2 = \frac{1}{n}\left(n_1x_1^2 + n_2x_2^2 + \ldots + n_kx_k^2\right) - \overline{x}^2$$

trong đó $\overline{x}$ là số trung bình của mẫu số liệu đã Cho.

- Độ lệch chuẩn, kí hiệu là S, bằng căn bậc hai của phương sai, tức là $S = \sqrt{S^2}$

Chú ý:

Trong thống kê, người ta cũng quan tâm đến phương sai hiệu chỉnh, kí hiệu là $\hat{s}^2$ và độ lệch chuẩn hiệu chỉnh, kí hiệu là $\hat{s}$, trong đó:

- Phương sai hiệu chỉnh được tính bởi công thức :

$$\hat{s}^2=\frac{1}{n-1}\left[n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+\ldots+n_k\left(x_k-\overline{x}\right)^2\right].$$

- Độ lệch chuẩn hiệu chỉnh được tính bởi công thức $\Delta\hat{s}=\sqrt{\hat{s}^2}$.

Ví dụ 11: Thống kê điểm thi giữa học kì I môn Toán của các bạn học sinh trong một lớp nọ theo bảng sau:

Điểm	5	6	7	8	9	10
Số học sinh	1	3	10	15	7	4

Phương sai của mẫu số liệu trên là

A. 1,34. **B.** 1,16. **C.** 1,37. **D.** 1,17

Cách giải:

Cỡ mẫu: $n=40$

Ta có $\overline{x}=\frac{5.1+6.3+7.10+8.15+9.7+10.4}{40}=7,9$

Phương sai của mẫu số liệu trên là:

$$s^2=\frac{1}{40}\cdot\left(1\cdot(5-7,9)^2+3\cdot(6-7,9)^2+10\cdot(7-7,9)^2+15\cdot(8-7,9)^2+7\cdot(9-7,9)^2+4\cdot(10-7,9)^2\right)=1,34$$

Chọn A.

Ví dụ 12: Bảng số liệu sau cho biết thời gian chạy cự li 100 m của các bạn trong lớp 12 A.

Thời gian	12	13	14	15	16	17
Số bạn	5	6	10	6	7	8

Độ lệch chuẩn hiệu chỉnh của mẫu số liệu trên (kết quả tính làm tròn đến hàng phần trăm) là

A. 2,70. **B.** 1,64. **C.** 2,77. **D.** 1,66.

Cách giải:

Cỡ mẫu: $n=42$

$$\overline{x}=\frac{12\cdot5+13\cdot6+14\cdot10+15\cdot6+16\cdot7+17\cdot8}{42}=\frac{44}{3}$$

Phương sai:

$$\hat{s}^2=\frac{1}{42-1}\cdot\left[5\cdot\left(12-\frac{44}{3}\right)^2+6\cdot\left(13-\frac{44}{3}\right)^2+10\cdot\left(14-\frac{44}{3}\right)^2+6\cdot\left(15-\frac{44}{3}\right)^2+7\cdot\left(16-\frac{44}{3}\right)^2+8\cdot\left(17-\frac{44}{3}\right)^2\right]=\frac{340}{123}$$

Độ lệch chuẩn: $\hat{s}=\sqrt{\hat{s}^2}=\sqrt{\frac{340}{123}}\approx1,66$

Chọn D.