

BÀI GIẢNG GIÁO KHOA

Thầy giáo : Nguyễn Quốc Tùng

Toán 11

TẬP HAI

§1

**CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM
CHO MẪU SỐ LIỆU GHÉP NHÓM**

I. MẪU SỐ LIỆU GHÉP NHÓM

1. Khái niệm và ví dụ mẫu số liệu ghép nhóm:

Ví dụ 1:

Số lượng khách mua bảo hiểm nhân thọ trong một ngày

Khoảng tuổi	[20;30)	[30;40)	[40;50)	[50;60)	[60;70)
Số khách hàng	1	7	12	4	0

Nhận xét: Khách hàng mua bảo hiểm nhân thọ được phân loại vào các khoảng tuổi.

* Định nghĩa

Mẫu số liệu ghép nhóm thường được trình bày dưới dạng bảng thống kê như sau:

Bảng 1: Bảng tần số ghép nhóm

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Tần số	n_1	n_2	...	n_k

Chú ý:

- Bảng trên gồm k nhóm $[u_j; u_{j+1})$ với $1 \leq j \leq k$, mỗi nhóm gồm một số giá trị được ghép theo một tiêu chí xác định.
- Cỡ mẫu $n = n_1 + n_2 + \dots + n_k$.
- Giá trị chính giữa mỗi nhóm được dùng làm giá trị đại diện cho nhóm ấy. Ví dụ nhóm $[u_1; u_2)$ có giá trị đại diện là $\frac{1}{2}(u_1 + u_2)$.
- Hiệu $u_{j+1} - u_j$ được gọi là độ dài của nhóm $[u_j; u_{j+1})$.

2. Ghép nhóm mẫu số liệu:

* Quy tắc thường dùng:

Mỗi mẫu số liệu có thể được ghép nhóm theo nhiều cách khác nhau nhưng thường tuân theo một số quy tắc sau:

- Tính độ lớn của khoảng biến thiên R .
- Chọn từ $k=5$ đến $k=20$ nhóm. Cỡ mẫu càng lớn thì cần càng nhiều nhóm số liệu. Các nhóm có cùng độ dài bằng L . Cách chọn: $L > R/k$.
- Giá trị nhỏ nhất của mẫu thuộc vào nhóm $[u_1; u_2)$ và càng gần u_1 càng tốt. Giá trị lớn nhất của mẫu thuộc nhóm $[u_k; u_{k+1})$ và càng gần u_{k+1} càng tốt.

Ví dụ 2: Cân nặng của 28 học sinh nam lớp 11 được cho như sau:

55,4 62,6 54,2 56,8 58,8 59,4 60,7 58 59,5 63,6 61,8 52,3 63,4 57,9
49,7 45,1 56,2 63,2 46,1 49,6 59,1 55,3 55,8 45,5 46,8 54 49,2 52,6

Hãy chia mẫu dữ liệu trên thành 5 nhóm, lập bảng tần số ghép nhóm và xác định giá trị đại diện cho mỗi nhóm.

Giải

Khoảng biến thiên của mẫu số liệu trên là $R=63,6-45,1=18,5$.

$$\text{Độ dài mỗi nhóm } L > \frac{R}{k} = \frac{18,5}{5} = 3,7$$

Ta chọn $L = 4$ và chia dữ liệu thành các nhóm $[45; 49), [49; 53), [53; 57), [57; 61), [61; 65)$.

Khi đó, ta có bảng tần số ghép nhóm sau:

Cân nặng	$[45; 49)$	$[49; 53)$	$[53; 57)$	$[57; 61)$	$[61; 65)$
Giá trị đại diện	47	51	55	59	63
Số học sinh	4	5	7	7	5

II. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU GHEP NHÓM

Chú ý:

- Số trung bình (số trung vị, tứ phân vị, mốt) của mẫu số liệu ghép nhóm đều là các ước lượng của số trung bình (số trung vị, tứ phân vị, mốt) của mẫu số liệu gốc (trước khi ghép nhóm).
- Đối với mẫu số liệu rời rạc (mẫu số liệu thuộc tập số tự nhiên), thường được cho các nhóm dưới dạng $[k_1; k_2]$ hay $k_1 - k_2$, trong đó $k_1, k_2 \in \mathbb{N}$, để tiện cho việc tính toán, ta có thể hiệu chỉnh nhóm $[k_1; k_2]$ thành $[k_1 - 0,5; k_2 + 0,5)$.

Ví dụ:

Mẫu số liệu trước hiệu chỉnh:

Điểm thi	1 – 4	5 – 7	8 – 10
Số học sinh	5	20	10

Mẫu số liệu sau hiệu chỉnh:

Điểm thi	[0,5; 4,5)	[4,5; 7,5)	[7,5; 10)
Số học sinh	5	20	10

1. Số trung bình

Giả sử mẫu số liệu được cho dưới dạng bảng tần số ghép nhóm:

Nhóm	Nhóm 1	Nhóm 2	...	Nhóm k
Giá trị đại diện	c_1	c_2	...	c_k
Tần số	n_1	n_2	...	n_k

Số trung bình của mẫu số liệu ghép nhóm, kí hiệu $\bar{x}$, được tính như sau:

$$\bar{x} = \frac{n_1 \cdot c_1 + n_2 \cdot c_2 + \dots + n_k \cdot c_k}{n}$$

trong đó $n = n_1 + n_2 + \dots + n_k$.

Ví dụ 1:

Các bạn học sinh lớp 11A1 trả lời 40 câu hỏi trong một bài kiểm tra. Kết quả được thống kê ở bảng sau:

Số câu trả lời đúng	[16; 21)	[21; 26)	[26; 31)	[31; 36)	[36; 41)
Số học sinh	4	6	8	18	4

Hãy ước lượng trung bình số câu trả lời đúng của các học sinh lớp 11A1.

Giải

Ta có bảng thống kê số câu trả lời đúng theo giá trị đại diện:

Số câu trả lời đúng	18,5	23,5	28,5	33,5	38,5
Số học sinh	4	6	8	18	4

Áp dụng công thức tính giá trị trung bình, ta có:

$$\bar{x} = \frac{18,5.4 + 23,5.6 + 28,5.8 + 33,5.18 + 38,5.4}{4 + 6 + 8 + 18 + 4} = 30$$

Ý nghĩa: Số trung bình của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho số trung bình của mẫu số liệu gốc. Nó thường dùng để đo xu thế trung tâm của mẫu số liệu.

2. Mốt

Nhóm chứa mốt của mẫu số liệu ghép nhóm là nhóm có tần số lớn nhất.

Giả sử nhóm chứa mốt là $[u_m; u_{m+1})$, khi đó **mốt của mẫu số liệu ghép nhóm**, kí hiệu là M_o , được xác định bởi công thức

$$M_o = u_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \cdot (u_{m+1} - u_m)$$

Chú ý: Nếu không có nhóm kế trước của nhóm chứa mốt thì $n_{m-1} = 0$. Nếu không có nhóm kế sau của nhóm chứa mốt thì $n_{m+1} = 0$.

Ví dụ 2:

Một công ty xây dựng khảo sát khách hàng xem họ có nhu cầu mua nhà ở mức giá nào. Kết quả khảo sát được ghi lại ở bảng sau:

Mức giá (triệu đồng / m ²)	[10; 14)	[14; 18)	[18; 22)	[22; 26)	[26; 30)
Số khách hàng	54	78	120	45	12

- Tìm mốt của mẫu số liệu ghép nhóm trên.
- Công ty nên xây nhà ở mức giá nào để nhiều người có nhu cầu mua nhất?

Giải

- Nhóm chứa mốt (gốc) của mẫu số liệu trên là nhóm $[u_m; u_{m+1}) = [u_3; u_4) = [18; 22)$.

Vậy $m = 3$. Thay số vào công thức, ta có:

$$u_m = u_3 = 18, n_{m-1} = n_2 = 78, n_m = n_3 = 120, n_{m+1} = n_4 = 45,$$

$$u_{m+1} - u_m = u_4 - u_3 = 22 - 18 = 4.$$

Mốt của mẫu số liệu ghép nhóm là

$$M_o = 18 + \frac{120 - 78}{(120 - 78) + (120 - 45)} \cdot 4 = \frac{758}{39} \approx 19,4$$

b) Dựa vào kết quả trên ta có thể dự đoán rằng nếu công ty xây nhà ở mức giá 19,4 triệu đồng / m² thì sẽ có nhiều người có nhu cầu mua nhất.

Ý nghĩa:

- Mốt của mẫu số liệu không ghép nhóm là giá trị có khả năng xuất hiện cao nhất khi lấy mẫu. Mốt của mẫu số liệu sau khi ghép nhóm M_o xấp xỉ với mốt của mẫu số liệu không ghép nhóm.
- Các giá trị nằm xung quanh M_o thường có khả năng xuất hiện cao hơn các giá trị khác.
- Một mẫu số liệu ghép nhóm có thể có nhiều nhóm chứa mốt và nhiều mốt.

III. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU GHEP NHÓM (TIẾP)

3. Trung vị

Công thức xác định trung vị của mẫu số liệu ghép nhóm:

- Gọi n là cỡ mẫu.
- Giả sử nhóm $[u_m; u_{m+1})$ chứa trung vị của mẫu số liệu gốc;
- n_m là tần số của nhóm chứa trung vị;
- $C = n_1 + n_2 + \dots + n_{m-1}$.

$$M_e = u_m + \frac{\frac{n}{2} - C}{n_m} \cdot (u_{m+1} - u_m)$$

Ví dụ 1: Kết quả khảo sát cân nặng của 25 quả bơ ở một lô hàng cho trong bảng sau:

Cân nặng (g)	[150; 155)	[155; 160)	[160; 165)	[165; 170)	[170; 175)
Số quả bơ	1	7	12	3	2

Hãy tìm trung vị của mẫu số liệu ghép nhóm trên.

Gọi $x_1; x_2; \dots; x_{25}$ là cân nặng của 25 quả bơ xếp theo thứ tự không giảm.

Ta xác định được số trung vị là x_{13} và từ bảng, ta có $x_{13} \in [160; 165)$.

Áp dụng công thức với $n = 25, n_m = 12, C = 1 + 7 = 8, u_m = 160, u_{m+1} = 165$, trung vị của mẫu số liệu ghép nhóm là

$$M_e = 160 + \frac{\frac{25}{2} - 8}{12} \cdot (165 - 160) = 161,875.$$

Ý nghĩa: Từ dữ liệu ghép nhóm nói chung không thể xác định chính xác trung vị của mẫu số liệu gốc. Trung vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho mẫu số liệu gốc và có thể lấy làm giá trị đại diện cho mẫu số liệu.

4. Tứ phân vị

Cách xác định tứ phân vị của mẫu số liệu ghép nhóm:

Tứ phân vị thứ hai, Q_2 , cũng chính là trung vị của mẫu số liệu ghép nhóm.

Cách tìm tứ phân vị thứ nhất Q_1 :

- Giả sử nhóm $[u_i; u_{i+1})$ chứa tứ phân vị thứ nhất (mẫu gốc);
- n_m là tần số của nhóm chứa tứ phân vị thứ nhất (mẫu gốc);
- $C = n_1 + n_2 + \dots + n_{i-1}$.

Khi đó:

$$Q_1 = u_i + \frac{\frac{n}{4} - C}{n_i}(u_{i+1} - u_i)$$

Cách tìm tứ phân vị thứ nhất Q_3 :

- Giả sử nhóm $[u_j; u_{j+1})$ chứa tứ phân vị thứ ba (mẫu gốc);
- n_j là tần số của nhóm chứa tứ phân vị thứ ba (mẫu gốc);
- $C = n_1 + n_2 + \dots + n_{j-1}$.

Khi đó:

$$Q_3 = u_j + \frac{\frac{3n}{4} - C}{n_j}(u_{j+1} - u_j)$$

