

BÀI 1. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CHO MẪU SỐ LIỆU GHEP NHÓM

A. KIẾN THỨC CẦN NHỚ

1. Mẫu số liệu ghép nhóm

a) Bảng tần số ghép nhóm

- Mẫu số liệu ghép nhóm là mẫu số liệu cho dưới dạng bảng tần số ghép nhóm.
- Mỗi nhóm số liệu gồm một số giá trị của mẫu số liệu được ghép nhóm theo một tiêu chí xác định có dạng $[a; b)$, trong đó a là đầu mút trái, b là đầu mút phải. Độ dài nhóm là $b - a$.
- Tần số của một nhóm là số số liệu trong mẫu số liệu thuộc vào nhóm đó. Tần số của nhóm 1, nhóm 2, ..., nhóm m kí hiệu lần lượt là $n_1, n_2, \dots, n_m$.
- Bảng tần số ghép nhóm được lập như ở Bảng 1, trong đó mẫu số liệu gồm n số liệu được chia thành m nhóm ứng với m nửa khoảng $[a_1; a_2); [a_2; a_3); \dots; [a_m; a_{m+1})$, ở đó $a_1 < a_2 < \dots < a_m < a_{m+1}$ và $n = n_1 + n_2 + \dots + n_m$.

Nhóm	Tần số
$[a_1; a_2)$	n_1
$[a_2; a_3)$	n_2
...	...
$[a_m; a_{m+1})$	n_m
	n

Bảng 1

b) Ghép nhóm mẫu số liệu. Tần số tích lũy

Để chuyển mẫu số liệu không ghép nhóm thành mẫu số liệu ghép nhóm, ta thực hiện như sau:

- Chia miền giá trị của mẫu số liệu thành một số nhóm theo tiêu chí cho trước;
- Đếm số giá trị của mẫu số liệu thuộc mỗi nhóm (tần số) và lập bảng tần số ghép nhóm.

Chú ý: Khi ghép nhóm số liệu, ta thường phân chia các nhóm có độ dài bằng nhau và đầu mút của các nhóm có thể không phải là giá trị của mẫu số liệu. Nhóm cuối cùng có thể là $[a_m; a_{m+1}]$.

- Tần số tích lũy của một nhóm là số số liệu trong mẫu số liệu có giá trị nhỏ hơn giá trị đầu mút phải của nhóm đó. Tần số tích lũy của nhóm 1, nhóm 2, ..., nhóm m kí hiệu lần lượt là $cf_1, cf_2, \dots, cf_m$.
- Bảng tần số ghép nhóm bao gồm cả tần số tích lũy được lập như ở Bảng 2.

Nhóm	Tần số	Tần số tích lũy
$[a_1; a_2)$	n_1	$cf_1 = n_1$
$[a_2; a_3)$	n_2	$cf_2 = n_1 + n_2$
...	...	...
$[a_m; a_{m+1})$	n_m	$cf_m = n_1 + n_2 + \dots + n_m$

	n	
--	-----	--

Bảng 2

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

a) Số trung bình cộng (số trung bình)

Cho mẫu số liệu ghép nhóm như ở Bảng 3, trong đó giá trị đại diện của nhóm là trung điểm x_i của nửa khoảng (tính bằng trung bình cộng của hai đầu mút) ứng với nhóm i .

Số trung bình cộng của mẫu số liệu ghép nhóm, kí hiệu $\bar{x}$, được tính theo công thức:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_m x_m}{n}.$$

Nhóm	Giá trị đại diện	Tần số
$[a_1; a_2)$	x_1	n_1
$[a_2; a_3)$	x_2	n_2
...	...	...
$[a_m; a_{m+1})$	x_m	n_m
		$n = n_1 + n_2 + \dots + n_m$

Bảng 3

b) Trung vị

Cho mẫu số liệu ghép nhóm bao gồm cả tần số tích lũy như ở Bảng 2.

Giả sử nhóm k là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{n}{2}$, tức là $cf_{k-1} < \frac{n}{2}$ nhưng $cf_k \geq \frac{n}{2}$. Ta gọi r, d, n_k lần lượt là đầu mút trái, độ dài, tần số của nhóm k ; cf_{k-1} là tần số tích lũy của nhóm $k-1$.

Trung vị của mẫu số liệu ghép nhóm, kí hiệu M_e , được tính theo công thức sau:

$$M_e = r + \left(\frac{\frac{n}{2} - cf_{k-1}}{n_k} \right) \cdot d.$$

Quy ước: $cf_0 = 0$.

c) Tứ phân vị

Cho mẫu số liệu ghép nhóm bao gồm cả tần số tích lũy như ở Bảng 2.

- Tứ phân vị thứ hai, kí hiệu Q_2 , bằng trung vị M_e .

- Giả sử nhóm p là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{n}{4}$, tức là $cf_{p-1} < \frac{n}{4}$ nhưng $cf_p \geq \frac{n}{4}$. Ta gọi s, h, n_p lần lượt là đầu mút trái, độ dài, tần số của nhóm p ; cf_{p-1} là tần số tích lũy của nhóm $p-1$.

Tứ phân vị thứ nhất, kí hiệu Q_1 , được tính bằng công thức sau:

$$Q_1 = s + \left(\frac{\frac{n}{4} - cf_{p-1}}{n_p} \right) \cdot h.$$

- Giả sử nhóm q là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{3n}{4}$, tức là $cf_{q-1} < \frac{3n}{4}$ nhưng $cf_q \geq \frac{3n}{4}$. Ta gọi t, l, n_q lần lượt là đầu mút trái, độ dài, tần số của nhóm q ; cf_{q-1} là tần số tích lũy của nhóm $q-1$.

Tứ phân vị thứ ba, kí hiệu Q_3 , được tính bằng công thức sau:

$$Q_3 = t + \left(\frac{\frac{3n}{4} - cf_{q-1}}{n_q} \right) \cdot l.$$

d) Mốt

Cho mẫu số liệu ghép nhóm như ở Bảng 1.

Giả sử nhóm i là nhóm có tần số lớn nhất. Ta gọi u, g, n_i lần lượt là đầu mút trái, độ dài, tần số của nhóm i ; n_{i-1}, n_{i+1} lần lượt là tần số của nhóm $i-1$, nhóm $i+1$.

Mốt của mẫu số liệu ghép nhóm, kí hiệu M_o , được tính theo công thức sau:

$$M_o = u + \left(\frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}} \right) \cdot g.$$

Quy ước: $n_0 = 0; n_{m+1} = 0$.

B. VÍ DỤ

Vấn đề 1. Mẫu số liệu ghép nhóm. Lập bảng tần số ghép nhóm của mẫu số liệu

Ví dụ 1. Bảng 4 biểu diễn mẫu số liệu ghép nhóm được cho dưới dạng bảng tần số ghép nhóm. Hãy cho biết:

- Mẫu số liệu đó có bao nhiêu số liệu; bao nhiêu nhóm;
- Tần số của mỗi nhóm.

Nhóm	Tần số
[0;10)	8
[10; 20)	7
[20; 30)	9
[30; 40)	6
	$n = 30$

Bảng 4

Giải

Từ Bảng 4, ta thấy:

a) Mẫu số liệu đó gồm 30 số liệu và 4 nhóm.

b) Tần số của các nhóm 1,2, 3,4 lần lượt là $8, 7, 9, 6$.

Ví dụ 2. Mẫu số liệu dưới đây ghi lại tốc độ của 42 ô tô khi đi qua một trạm đo tốc độ (đơn vị: km / h):

47,5 49,5 46 51 52,5 45 61
 42 67 48 63 65 62,5 49,5
 43,5 41 57,5 63,5 56,5 53 48
 61,5 46 57 69 44,5 52 50
 45 55 47 60 67,5 62 58
 56 51,5 57,5 59 52 43 56

Lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy cho mẫu số liệu trên có sáu nhóm ứng với sáu nửa khoảng:

$[40; 45), [45; 50), [50; 55), [55; 60), [60; 65), [65; 70)$.

Giải

Bảng tần số ghép nhóm bao gồm cả tần số tích lũy như ở Bảng 5:

Nhóm	Tần số	Tần số tích lũy
[40; 45)	5	5
[45; 50)	10	15
[50; 55)	7	22
[55; 60)	9	31
[60; 65)	7	38
[65; 70)	4	42
	$n = 42$	

Bảng 5

Vấn đề 2. Xác định số trung bình cộng của mẫu số liệu ghép nhóm

Ví dụ 3. Tính số trung bình cộng của mẫu số liệu ghép nhóm được cho ở Bảng 5 (làm tròn kết quả đến hàng phần mười).

Giải

Từ mẫu số liệu ghép nhóm được cho ở Bảng 5, ta có bảng tần số ghép nhóm như ở Bảng 6:

Nhóm	Giá trị đại diện	Tần số
[40; 45)	42,5	5
[45; 50)	47,5	10
[50; 55)	52,5	7
[55; 60)	57,5	9
[60; 65)	62,5	7
[65; 70)	67,5	4
		$n = 42$

Bảng 6

Số trung bình cộng của mẫu số liệu trên là:

$$\bar{x} = \frac{5 \cdot 42,5 + 10 \cdot 47,5 + 7 \cdot 52,5 + 9 \cdot 57,5 + 7 \cdot 62,5 + 4 \cdot 67,5}{42} \approx 54,3 (km/h).$$

Vấn đề 3. Xác định trung vị, tứ phân vị của mẫu số liệu ghép nhóm

Ví dụ 4. Xác định trung vị và tứ phân vị của mẫu số liệu ghép nhóm được cho ở Bảng 5 (làm tròn các kết quả đến hàng phần mười).

Giải

Số phần tử của mẫu là $n = 42$.

- Ta có: $\frac{n}{2} = \frac{42}{2} = 21$ mà $15 < 21 < 22$. Suy ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 21.

Xét nhóm 3 là nhóm [50; 55) có $r = 50, d = 5, n_3 = 7$ và nhóm 2 là nhóm [45; 50) có $cf_2 = 15$.

Áp dụng công thức, ta có trung vị của mẫu số liệu là:

$$M_e = 50 + \left(\frac{21 - 15}{7} \right) \cdot 5 \approx 54,3 (km/h).$$

Tứ phân vị thứ hai của mẫu số liệu là: $Q_2 = M_e \approx 54,3 (km/h)$.

- Ta có: $\frac{n}{4} = \frac{42}{4} = 10,5$ mà $5 < 10,5 < 15$ nên nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 10,5.

Xét nhóm 2 là nhóm [45; 50) có $s = 45, h = 5, n_2 = 10$ và nhóm 1 là nhóm [40; 45) có $cf_1 = 5$.

Áp dụng công thức, ta có tứ phân vị thứ nhất của mẫu số liệu là:

$$Q_1 = 45 + \left(\frac{10,5 - 5}{10} \right) \cdot 5 \approx 47,8 (km/h).$$

- Ta có: $\frac{3n}{4} = \frac{3 \cdot 42}{4} = 31,5$ mà $31 < 31,5 < 38$ nên nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 31,5.

Xét nhóm 5 là nhóm [60; 65) có $t = 60, l = 5, n_5 = 7$ và nhóm 4 là nhóm [55; 60) có $cf_4 = 31$.

Áp dụng công thức, ta có tứ phân vị thứ ba của mẫu số liệu là:

$$Q_3 = 60 + \left(\frac{31,5 - 31}{7} \right) \cdot 5 \approx 60,4 (km / h).$$

Vậy tứ phân vị của mẫu số liệu trên là: $Q_1 \approx 47,8 (km / h); Q_2 \approx 54,3 (km / h); Q_3 \approx 60,4 (km / h).$

Vấn đề 5. Xác định một của mẫu số liệu ghép nhóm

Ví dụ 5. Xác định một của mẫu số liệu ghép nhóm được cho ở Bảng 5 (làm tròn kết quả đến hàng phần mười).

Giải

Ta thấy: Nhóm 2 ứng với nửa khoảng $[45; 50)$ là nhóm có tần số lớn nhất với $u = 45; g = 5, n_2 = 10$.
Nhóm 1 có tần số $n_1 = 5$, nhóm 3 có tần số $n_3 = 7$.

Áp dụng công thức, ta có một của mẫu số liệu là:

$$M_o = 45 + \left(\frac{10 - 5}{2 \cdot 10 - 5 - 7} \right) \cdot 5 \approx 48,1 (km / h).$$

C. BÀI TẬP

Câu 1. Khi thống kê chiều cao của 40 bạn lớp 11^A , ta thu được mẫu số liệu ghép nhóm được cho ở Bảng 7 (đơn vị: centimét).

Nhóm	Tần số
[155;160)	5
[160;165)	12
[165;170)	16
[170;175)	7
	$n = 40$

Bảng 7

a) Độ dài của mỗi nhóm bằng:

- A. 155 .
- B. 5 .
- C. 175 .
- D. 20 .

Lời giải

Chọn B

b) Tần số của nhóm $[160;165)$ là bao nhiêu?

- A. 5 .
- B. 16 .
- C. 12 .
- D. 7 .

Lời giải

Chọn C

c) Nhóm có tần số lớn nhất là:

A. $[155;160)$.

B. $[160;165)$.

C. $[165;170)$.

D. $[170;175)$.

Lời giải

Chọn C

d) Giá trị cf_3 bằng:

A. 16 .

B. 17 .

C. 23 .

D. 33 .

Lời giải

Chọn D

e) Giá trị đại diện của nhóm $[155; 160)$ bằng:

A. 157,5.

B. 155 .

C. 160 .

D. 5 .

Lời giải

Chọn A

g) Nhóm có giá trị đại diện bằng 162,5 là:

A. $[155;160)$.

B. $[160;165)$.

C. $[165;170)$.

D. $[170;175)$.

Lời giải

Chọn B

Câu 2. Xác định các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm như ở Bảng 7 (làm tròn các kết quả đến hàng phần mười).

Lời giải

Số trung bình cộng là:

$$\bar{x} = \frac{5 \cdot 157,5 + 12 \cdot 162,5 + 16 \cdot 167,5 + 7 \cdot 172,5}{40} \approx 165,6.$$

Bảng tần số ghép nhóm bao gồm cả tần số tích lũy được cho như ở Bảng 9:

Ta có: $\frac{n}{2} = 20, \frac{n}{4} = 10, \frac{3n}{4} = 30$.

Vì $17 < 20 < 33$ nên nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 20.

Suy ra trung vị là:

$$M_e = 165 + \left(\frac{20 - 17}{16} \right) \cdot 5 \approx 165,9.$$

Nhóm	Tần số	Tần số tích lũy
[155;160)	5	5
[160;165)	12	17
[165;170)	16	33
[170;175)	7	40
	$n = 40$	

Bảng 9

Tứ phân vị thứ hai là: $Q_2 = M_e \approx 165,9$.

Vì $5 < 10 < 17$ nên nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 10. Suy ra tứ phân vị thứ nhất là:

$$Q_1 = 160 + \left(\frac{10 - 5}{12} \right) \cdot 5 \approx 162,1.$$

Vì $17 < 30 < 33$ nên nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn

hơn hoặc bằng 30. Suy ra tứ phân vị thứ ba là:

$$Q_3 = 165 + \left(\frac{30 - 17}{16} \right) \cdot 5 \approx 169,1$$

Trong các nhóm, nhóm 3 có tần số lớn nhất. Suy ra mốt là:

$$M_o = 165 + \left(\frac{16 - 12}{2 \cdot 16 - 12 - 7} \right) \cdot 5 \approx 166,5.$$

Câu 3. Cho mẫu số liệu ghép nhóm thống kê thời gian sử dụng điện thoại trước khi ngủ (đơn vị: phút) của một người trong 120 ngày như ở Bảng 8. Xác định các số đặc trưng đo xu thế trung tâm cho mẫu số liệu đó (làm tròn các kết quả đến hàng phần mười).

Nhóm	Tần số
[0;4)	13
[4;8)	29

[8;12)	48
[12;16)	22
[16; 20)	8
	$n = 120$

Bảng 8

Lời giải

Số trung bình cộng là:

$$\bar{x} = \frac{13 \cdot 2 + 29 \cdot 6 + 48 \cdot 10 + 22 \cdot 14 + 8 \cdot 18}{120} \approx 9,4.$$

Bảng tần số ghép nhóm bao gồm cả tần số tích lũy được cho như ở Bảng 10:

Nhóm	Tần số	Tần số tích lũy
[0; 4)	13	13
[4; 8)	29	42
[8; 12)	48	90
[12; 16)	22	112
[16; 20)	8	120
	$n = 120$	

Bảng 10

Ta có: $\frac{n}{2} = 60, \frac{n}{4} = 30, \frac{3n}{4} = 90$.

Vì $42 < 60 < 90$ nên nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 60 .

Suy ra trung vị là: $M_e = 8 + \left(\frac{60 - 42}{48} \right) \cdot 4 = 9,5.$

Tứ phân vị thứ hai là: $Q_2 = M_e = 9,5$.

Vì $13 < 30 < 42$ nên nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 30 . Suy ra tứ

phân vị thứ nhất là: $Q_1 = 4 + \left(\frac{30 - 13}{29} \right) \cdot 4 \approx 6,3$

Vì $42 < 90 \leq 90$ nên nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 90 . Suy ra tứ

phân vị thứ ba là: $Q_3 = 8 + \left(\frac{90 - 42}{48} \right) \cdot 4 = 12$

Trong các nhóm, nhóm 3 có tần số lớn nhất. Suy ra một là: $M_0 = 8 + \left(\frac{48 - 29}{2 \cdot 48 - 29 - 22} \right) \cdot 4 \approx 9,7.$

Câu 4. Khi thống kê chỉ số đường huyết (đơn vị: $mmol / L$) của 28 người cao tuổi trong một lần đo, ta được kết quả sau:

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

a) Lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy có năm nhóm ứng với năm nửa khoảng: $[7,0;7,2), [7,2;7,4), [7,4;7,6), [7,6;7,8), [7,8;8,0]$

Lời giải

a) Bảng tần số ghép nhóm bao gồm cả tần số tích lũy được cho như ở Bảng 11 :

Nhóm	Tần số	Tần số tích lũy
$[7,0;7,2)$	7	7
$[7,2;7,4)$	6	13
$[7,4;7,6)$	7	20
$[7,6;7,8)$	5	25
$[7,8;8,0]$	3	28
	$n = 28$	

Bảng 11

b) Độ dài của mỗi nhóm bằng:

- A. 7 .
- B. 8 .
- C. 1 .
- D. 0,2 .

Lời giải

Chọn D

c) Tần số của nhóm $[7,8;8,0]$ bằng:

- A. 3 .
- B. 5 .
- C. 6 .
- D. 7.

Lời giải

Chọn A

d) Giá trị cf_3 bằng:

- A. 7.
- B. 13.
- C. 20 .

D. 25 .

Lời giải

Chọn C

e) Giá trị đại diện của nhóm $[7, 4; 7, 6)$ bằng:

A. 7,4 .

B. 7,6.

C. 7,5.

D. 2 .

Lời giải

Chọn C

g) Nhóm có giá trị đại diện bằng 7,7 là:

A. $[7, 0; 7, 2)$.

B. $[7, 2; 7, 4)$.

C. $[7, 4; 7, 6)$.

D. $[7, 6; 7, 8)$.

Lời giải

Chọn D

Câu 5. Với mẫu số liệu ghép nhóm thu được ở Bài 4 , xác định các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm đó (làm tròn các kết quả đến hàng phần mười).

Lời giải

Số trung bình cộng là:

$$\bar{x} = \frac{7 \cdot 7,1 + 6 \cdot 7,3 + 7 \cdot 7,5 + 5 \cdot 7,7 + 3 \cdot 7,9}{28} \approx 7,4.$$

Ta có: $\frac{n}{2} = 14, \frac{n}{4} = 7, \frac{3n}{4} = 21$.

Vì $13 < 14 < 20$ nên nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 14.

Suy ra trung vị là: $M_e = 7,4 + \left(\frac{14 - 13}{7} \right) \cdot 0,2 \approx 7,4$.

Tứ phân vị thứ hai là: $Q_2 = M_e \approx 7,4$.

Vì $0 < 7 \leq 7$ nên nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 7 . Suy ra tứ phân

$$Q_1 = 7,0 + \left(\frac{7 - 0}{7} \right) \cdot 0,2 = 7,2$$

vị thứ nhất là:

Vì $20 < 21 < 25$ nên nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 21 . Suy ra tứ phân vị thứ ba là:

$$Q_3 = 7,6 + \left(\frac{21 - 20}{5} \right) \cdot 0,2 \approx 7,6$$

Trong các nhóm, nhóm 1 và nhóm 3 có tần số lớn nhất nên ta có hai mốt là:

$$M_0 = 7,0 + \left(\frac{7 - 0}{2 \cdot 7 - 0 - 6} \right) \cdot 0,2 \approx 7,2$$

$$M_o' = 7,4 + \left(\frac{7 - 6}{2 \cdot 7 - 6 - 5} \right) \cdot 0,2 \approx 7,5$$