BÀI GIẢNG: CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CHO MẪU SỐ LIỆU GHÉP NHÓM

CHUYÊN ĐỀ: THỐNG KÊ VÀ XÁC SUẤT

MÔN: TOÁN 11 CƠ BẢN & NÂNG CAO

GIÁO VIÊN: NGÔ VĂN TOẢN

MỤC TIÊU

- ✓ *Biết được các số đặc trưng đo xu thế trung tâm (số trung bình cộng, trung vị, tứ phân vị, mốt) cho mẫu số liệu ghép nhóm.*
- ✓ *Tính được các số đặc trưng đo xu thế trung tâm (số trung bình cộng, trung vị, tứ phân vị, mốt) cho mẫu số liệu ghép nhóm…*

Ví dụ: Một nhà thực vật học đo chiều dài của 74 lá cây (đơn vị: milimet) và thu được tần số như sau. Tính chiều dài trung bình của 74 lá cây trên theo đơn vi milimet (làm tròn kết quả đến hàng phần trăm).

Nhóm	Giá trị đại diện	Tần số
[5,45;5,85)	5,65	5
[5,85;6,25)	6,05	9
[6,25;6,65)	6,45	15
[6,65;7,05)	6,85	19
[7,05;7,45)	7,25	16
[7,45;7,85)	7,65	8
[7,85;8,25)	8,05	2
		$n = 74$

1 MẪU SỐ LIỆU GHÉP NHÓM

1. 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 liệu trong mẫu số liệu thuộc vào nhóm đó. Tần số của nhóm 1, nhóm $2,\ldots$, nhóm m kí hiệu lần lượt là $n_1, n_2, \ldots, n_m$.

Bảng tần số ghép nhóm được lập ở bảng bên, trong đó mẫu số liệu n số liệu được chia thành m nhóm ứng với m nữa khoảng $[a_1;a_2)$; $[a_2;a_3);\ldots;[a_m;a_{m+1})$, ở đó

$a_1 < a_2 < \ldots < a_m < a_{m+1}$ và $n = n_1 + n_2 + \ldots + n_m$.

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

2. Ghép nhóm mẫu số liệu

a) Ghép nhóm mẫu số liệu

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

+) Bước 1:Chia miền giá trị của mẫu số liệu thành một số nhóm theo tiêu chí cho trước.

+) Bước 2: 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 đài 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 $[a_m;a_{m+1}]$

Ví dụ 1: Mẫu số liệu đây ghi lại tốc độ của 40 ô tô khi đi qua một trạm đo tốc dộ (đơn vị: km / h)

48,5	43	50	55	45	60	53	55,5	44	65
52	62,5	41	44,5	57	57	68	49	46,5	53,5
61	49,5	54	62	59	56	47	50	60	61
49,5	52,5	57	47	60	55	45	47,5	48	61,5

Lập bảng tần số ghép nhóm 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)

Cách giải:

Nhóm	Tần số
[40;45)	4
[45; 50)	11
[50; 55)	7
[55;60)	8
[60;65)	8
[65; 70)	2
	$n = 40$

b) Giá trị đại diện của nhóm

Trung điểm x_i của nửa khoảng (tính bằng trung bình cộng của hai đầu mút) ứng với nhóm i là giá trị đại diện của nhóm đó.

Nhóm	Tần số	Giá trị đại diện
$[a_1;a_2)$	n_1	$x_1=(a_1+a_2):2$
$[a_2;a_3)$	n_2	$x_2=(a_2+a_3):2$
…	…	
$[a_m;a_{m+1})$	n_m	$x_m=(a_m+a_{m+1}):2$
	n	

Ví dụ 2: Tính các giá trị đại diện của các nhóm ở bảng tần số ghép nhóm sau:

Nhóm	Giá trị đại diện
[40;45)	42,5
[45; 50)	47,5
[50; 55)	52,5
[55;60)	57,5
[60;65)	62,5
[65; 70)	67,5

c) Tần số tích lũy của nhóm

Tần số tích luỹ 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 luỹ của nhóm 1 , nhóm $2,\ldots$, nhóm m kí hiệu lần lượt là $cf_1,cf_2,\ldots,cf_m$.

Bảng tần số ghép nhóm bao gồm cả tần số tích luỹ được lập như ở bảng:

Nhóm	Tần số	Giá trị đại diện	Tần số tích luỹ
$[a_1;a_2)$	n_1	$x_1=(a_1+a_2):2$	$cf_1=n_1$
$[a_2;a_3)$	n_2	$x_2=(a_2+a_3):2$	$cf_2=n_1+n_2$
…	…		…
$[a_m;a_{m+1})$	n_m	$x_m=(a_m+a_{m+1}):2$	$cf_m=n_1+n_2+...+n_m$
	n		

Ví dụ 3. Tính các giá trị tần số tích luỹ của các nhóm ở bảng tần số ghép nhóm sau:

	Tần số	Tần số tích luỹ
[40;45)	4	4
[45; 50)	11	15
[50; 55)	7	22
[55;60)	8	30
[60;65)	8	38
[65; 70)	2	40
	$n=40$	

II SỐ TRUNG BÌNH CỘNG (SỐ TRUNG BÌNH)

Công thức:

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

$$\overline{X}=\frac{n_1x_1+n_2x_2+\cdots+n_mx_m}{n}$$

Ý nghĩa: Như ta đã biết, số trung bình cộng của mẫu số liệu không ghép nhóm là giá trị trung bình cộng của các số trong mẫu số liệu đó, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu khi các số liệu trong mẫu ít sai lệch với số trung bình cộng. Số trung bình cộng của mẫu số liệu sau khi ghép nhóm xấp xỉ với số trung bình cộng của mẫu số liệu không ghép nhóm ban đầu và có thể làm đại diện cho vị tri trung tâm của mẫu số liệu.

Ví dụ: Một nhà thực vật học đo chiều dài của 74 lá cây (đơn vị: milimet) và thu được tần số như sau. Tính chiều dài trung bình của 74 lá cây trên theo đơn vi milimet (làm tròn kết quả đến hàng phần trăm).

Nhóm	Giá trị đại diện	Tần số
[5,45;5,85)	5,65	5
[5,85;6,25)	6,05	9
[6,25;6,65)	6,45	15
[6,65;7,05)	6,85	19
[7,05;7,45)	7,25	16
[7,45;7,85)	7,65	8
[7,85;8,25)	8,05	2
		$n = 74$

Cách giải:

Nhóm	Giá trị đại diện	Tần số
[5,45;5,85)	5,65	5
[5,85;6,25)	6,05	9
[6,25;6,65)	6,45	15
[6,65;7,05)	6,85	19
[7,05;7,45)	7,25	16
[7,45;7,85)	7,65	8
[7,85;8,25)	8,05	2
		$n = 74$

$$\overline{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 x_3 + n_4 x_4 + n_5 x_5 + n_6 \cdot x_6 + n_7 x_7}{n} \approx 6,8(cm)$$

III TRUNG VỊ

1. Cách tìm trung vị của mẫu số liệu ghép nhóm có n giá trị

Bước 1: Tính $\frac{n}{2}$; tính tần số tích lũy của các nhóm (lập bảng tần số + cột tần số tích luỹ)

Bước 2: Tìm nhóm đầu tiên theo thứ tự có tần số tích luỹ lớn hơn hoặc bằng $\frac{n}{2}$ (Giả sử là nhóm thứ $k : [a_k ; a_{k+1})$

Bước 3 : Ta tìm các thông số :

$r = a_k$ là giá trị đầu mút trái cùa nhóm **k**

$d = a_{k+1} - a_k$ là độ dài của nhóm k

n_k : tần số của nhóm k

cf_{k-1} : tần số tích lũy của nhóm $k-1$

(nhóm ngay trước nhóm k)

Bước 4. Thay vào công thức : $M_e = r + \left(\dfrac{\dfrac{n}{2} - cf_{k-1}}{n_k} \right) \cdot d$

2. Ý nghĩa

Trung vị của mẫu số liệu sau khi ghép nhóm xấp xỉ với trung vị của mẫu số liệu không ghép nhóm ban đầu và có thể dùng để đại diện cho mẫu số liệu đã cho.

Ví dụ 4. Tìm trung vị của mẫu số liệu được mô tả trong bảng tần số ghép nhóm dưới đây:

	Tần số	Tần số tích luỹ
[40;45)	4	4
[45; 50)	11	15
[50; 55)	7	22
[55;60)	8	30
[60;65)	8	38
[65; 70)	2	40
	$n = 40$	

Cách giải:

Ta có $\dfrac{n}{2} = \dfrac{40}{2} = 20$.

Nhóm 3 là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng 20.

$$M_e = r + \left(\frac{\frac{n}{2} + c.f_2}{n_3} \right) d = 50 + \left(\frac{20-15}{7} \right).5 \approx 53,57$$

Ví dụ 5: Bảng dưới đây cho ta bảng tần số ghép nhóm số liệu thông kê chiều cao của 40 mẫu cây ở một vườn thực vật (đơn vị: centimét). Tìm trung vị của mẫu số liệu được mô tả trong bảng tần số ghép nhóm dưới đây:

	Tần số
[40;45)	4
[45; 50)	11
[50; 55)	7
[55;60)	8
[60;65)	8
[65; 70)	2

Cách giải:

	Tần số	Tần số tích luỹ
[40;45)	4	4
[45; 50)	11	15
[50; 55)	7	22
[55;60)	8	30
[60;65)	8	38
[65; 70)	2	40
	$n = 40$	

$\frac{n}{2} = \frac{40}{2} = 20$

Nhóm 3

$$\Rightarrow M_e = r + \left(\frac{\frac{n}{2} + c.f_2}{n_3} \right) d = 50 + \left(\frac{20-14}{14} \right).10 \approx 54,29$$

IV TỨ PHÂN VỊ

1. Cách tìm tứ phân vị $Q_1; Q_2; Q_3$

Tứ phân vị thứ hai $Q_2 = M_e = r + \left(\frac{\frac{n}{2} cf_{k-1}}{n_k} \right) \cdot d$

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 luỹ của nhóm $p-1$.

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

+) Tứ phân vị thứ ba Q_3

Giả sử nhơm q là nhóm đầu tiên có tần số tích luỹ lóm hơn hoặc bằng $\frac{3n}{4}$, tức là $cf_{q-1} < \frac{3n}{4}$

nhưng $cf_q \geq \frac{3n}{4}$.

Ta gọi ℓ, L, n_q lần lượt là đầu mút trái, độ dài, tần số của nhóm $q; cf_{q-1}$ là tần số tích luỹ của nhóm $q-1$.

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

2. Ý nghĩa

Như ta đã biết, đối với mẫu số liệu không ghép nhóm đã sắp xếp theo thứ tự từ nhỏ đến lớn, các điểm Q_1, Q_2, Q_3 chia mẫu số liệu đó thành bốn phần, mỗi phần đều chứa 25% giá trị.

Bằng cách ghép nhóm mẫu số liệu và tính toán tứ phân vị của mẫu số liệu ghép nhóm, ta nhận được ba giá trị mới cũng có thể dùng để đại diện cho mẫu số liệu đã cho.

Lưu ý rằng bộ ba giá trị Q_1, Q_2, Q_3 trong tứ phân vị của mẫu số liệu sau khi ghép nhóm xấp xỉ với bộ ba giá trị trong tứ phân vị của mẫu số liệu không ghép nhóm ban đầu.

Ví dụ 6: Tìm tứ phân vị của mẫu số liệu được mo tả bằng bảng tần số ghép nhóm dưới đây:

	Tần số	Tần số tích luỹ
[40;45)	4	4
[45; 50)	11	15
[50; 55)	7	22
[55;60)	8	30
[60;65)	8	38
[65; 70)	2	40
	$n = 40$	

Cách giải:

Ta có $\frac{n}{20} = 20$

$$Q_2 = M_2 = 50 + \left(\frac{20 - 13}{7} \right) \cdot 5 \approx 53,53$$

+) $\frac{n}{4} = \frac{40}{4} = 10$.

$$Q_1 = 45 + \left(\frac{10 - 4}{11} \right) \cdot 5 \approx 47,73.$$

+) $\frac{3n}{4} = \frac{3.40}{4} = 30$.

$$Q_3 = 55 + \left(\frac{30 - 22}{8} \right) \cdot 5 = 60.$$

Ví dụ 7. Tìm tứ phân vị của mẫu số liệu dưới đây

Nhóm	Tần số
[30;40)	4
[40; 50)	10
[50; 60)	14

[60;70)	6
[70;80)	4
[80; 90)	2
	$n = 40$

Cách giải:

Nhóm	Tần số	Tần số tích luỹ
[30;40)	4	$cf_1 = 4$
[40; 50)	10	$cf_2 = 14$
[50; 60)	14	$cf_3 = 28$
[60;70)	6	$cf_4 = 34$
[70;80)	4	$cf_5 = 38$
[80; 90)	2	$cf_6 = 40$
	$n = 40$	

Ta có $\frac{n}{20} = 20$

$$Q_2 = M_l = 50 + \left(\frac{20-14}{14}\right)\cdot 10 = 54,29.$$

$$Q_1 = 40 + \left(\frac{10-4}{10}\right)\cdot 10 = 46.$$

$$Q_3 = 60 + \left(\frac{30-28}{6}\right)\cdot 10 \approx 63,33.$$

V MỐT

1. Cách tìm Mốt của mẫu số liệu ghép nhóm

Bước 1. Xác định nhóm có tần số lớn nhất (giả sử là nhóm thứ i)

Bước 2. Ta xác định các thông số sau:

$u = a_i$ là đầu mút trái của nhóm i

$g = a_{i+1} - a_i$ là độ dài của nhóm i

$n_i; n_{i-1}; n_{i+1}$ lần lượt là tần số của nhóm thứ i; thứ $i-1$

(liền trước nhóm i); thứ $i+1$ (liền sau nhóm i);

Bước 3. Thay vào công thức $M_o = u + \left(\frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}\right)\cdot g.$

Chú ý: • Khi $i = 0$ thì $n_0 = 0$; • Khi $i = m$ thì $n_{m+1} = 0$.

2. Ý nghĩa

Như ta đã biết, mốt của một mẫu số liệu không ghép nhóm đặc trưng cho số lần lặp đi lặp lại nhiều nhất tại một giá trị của mẩu số liệu đó. Vì thế, có thể dùng mốt để đo xu thế trung tâm của mẫu số liệu khi mẫu số liệu có nhiều giá trị trùng nhau.

Bằng cách ghép nhóm mẫu số liệu và tính toán mốt của mẫu số liệu ghép nhóm, ta nhận được giá trị mới cũng có thể dùng để đo xu thế trung tâm của mẫu số liệu đã cho.

Mốt của mẫu số liệu sau khi ghép nhóm xấp xỉ với mốt của mẫu số liệu không ghép nhóm ban đầu. Một mẫu số liệu ghép nhóm có thể có nhiều mốt.

Ví dụ 8. Cho mẫu số liệu ghép nhóm thống kê thời gian sự dụng điện thoại trước khi ngủ (đơn vị: phút) của một người trong 120 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
	$n=120$

Cách giải:

Nhóm 3 là nhóm có tần số lớn nhất: $u=8; n_3=48; n_2=29; n_4=22$, $g=4$.

$$M_0=8+\left(\frac{n_3-n_2}{2n_3-n_2-n_4}\right)g=8+\left(\frac{48-29}{2.48-29-22}\right)\cdot 4\approx 9,69.$$

Ví dụ 9. 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 lưy có năm nhóm ứng với năm nửa khoảng: $[7,0;7,2),[7,2;7,4),[7,4;7,6),[7,6;7,8),[7,8;8,0]$

b) Xác dịnhh 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).

Cách giải:

Nhóm	Tần số	Tần số tích luỹ	Giá trị đại diện
$[7,0;7,2)$	7	7	7,1
$[7,2;7,4)$	6	13	7,3
$[7,4;7,6)$	7	20	7,5
$[7,6;7,8)$	5	25	7,7
$[7,8;8,0]$	3	28	7,9

Có hai nhóm có tần số lớn nhất là nhóm thứ nhất và nhóm thứ ba.

$$M_0 = u + \left(\frac{n_1 - n_0}{2n_1 - n_0 - n_2} \right).g$$

$$\Rightarrow M_0 = 7 + \left(\frac{7-0}{2.7-0-6} \right) \cdot 0,2 \approx 7,2$$

Xét nhóm thứ ba:

$$M_0 = u + \left(\frac{n_3 - n_2}{2n_3 - n_2 - n_4} \right) g = 7,4 + \left(\frac{7-6}{2.7-6-5} \right) \cdot 0,2 \approx 7,5$$