

Bài 1. Các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm

Chương 5 | Cánh Diều | Toán 11

ZO | 2025-03-20

Nếu không có yêu cầu làm tròn đặc biệt, các con số sẽ được làm tròn đến 4 chữ số thập phân để tiện đối chiếu khi bạn tự tìm đáp án.

1 Bài học

Hoạt động 1

Trong *Bảng 1*, ta thấy:

- Có 13 ô-tô có độ tuổi dưới 4;
- Có 29 ô-tô có độ tuổi từ 4 đến dưới 8.

Hãy xác định số ô-tô có độ tuổi:

- Từ 8 đến dưới 12;
- Từ 12 đến dưới 16;
- Từ 16 đến dưới 20.

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

Lời giải

Mẹo

Bạn có biết? Khi chưa quen với cách lập luận, bắt chước sách giáo khoa là một cách tốt để làm quen với cách trình bày. Khi đã hiểu rõ, bạn có thể tự phát triển cách lập luận của riêng mình.

Từ *Bảng 1*, quan sát theo hàng để đếm số ô-tô trong từng nhóm độ tuổi.

- Có 48 ô-tô có độ tuổi từ 8 đến dưới 12.
- Có 22 ô-tô có độ tuổi từ 12 đến dưới 16.
- Có 8 ô-tô có độ tuổi từ 16 đến dưới 20.

Nhóm & tần số

Việc xác định có 13 ô-tô có độ tuổi dưới 4 cho mình biết rằng nhóm $[0; 4)$ có tần số 13.

- **Nhóm:** $[a; b)$, chứa một số giá trị của mẫu số liệu, a là *đầu mút trái*, b là *đầu mút phải*.
Độ dài nhóm là $b - a$.
- **Tần số:** số giá trị của mẫu số liệu có trong nhóm $[a; b)$.

Biết có 13 ô-tô có độ tuổi dưới 4, nhưng không biết tuổi cụ thể của từng ô-tô.

Khám phá

Nhìn vào bảng số liệu, bạn có nhận xét gì về số lượng ô-tô theo từng nhóm độ tuổi?

- Số lượng ô-tô cũ (trên 12 năm) có ít hơn xe mới không?
- Điều này có ý nghĩa gì?

Luyện tập 1

Mẫu số liệu ghép nhóm ở *Bảng 1* có bao nhiêu số liệu? Bao nhiêu nhóm? Tìm tần số của mỗi nhóm.

Nhóm	Tần số
$[0; 4)$	13
$[4; 8)$	29
$[8; 12)$	48
$[12; 16)$	22

(continued)

Nhóm	Tần số
[16; 20)	8
$n = 120$	

Lời giải

- Mẫu số liệu ghép nhóm ở *Bảng 1* có 120 số liệu, chính là tổng của cột Tần số.
- Mẫu số liệu này có 5 nhóm, bao gồm $[0; 4)$, $[4; 8)$, $[8; 12)$, $[12; 16)$ và $[16; 20)$. Để ngắn gọn, gọi các nhóm này lần lượt là 1, 2, 3, 4 và 5.
- Tần số của các nhóm 1, 2, 3, 4, 5 lần lượt là 13, 29, 48, 22, 8.

Hoạt động 2

Một trường trung học phổ thông chọn 36 học sinh nam của khối lớp 11, đo chiều cao của các bạn học sinh đó và thu được mẫu số liệu sau (đơn vị: cm):

160	161	161	162	162	162	163	163	163	164	164	164
164	165	165	165	165	165	166	166	166	166	167	167
168	168	168	168	169	169	170	171	171	172	172	174

Từ mẫu số liệu không ghép nhóm trên, hãy ghép các số liệu thành năm nhóm theo các nửa khoảng có độ dài bằng nhau.

Lời giải

Nhận thức quan trọng

Ở Hoạt động 1, chúng mình làm việc với một bảng tần số ghép nhóm đã có sẵn, nhưng không biết chính xác các số liệu gốc.

Sang Hoạt động 2, tình huống đảo ngược: số liệu cụ thể được cho, nhưng bảng tần số thì chưa có. Nhiệm vụ của chúng mình là ghép các số liệu thành năm nhóm có độ dài bằng nhau.

Điều thú vị là, với hai tiêu chí “số lượng nhóm” và “độ dài nhóm”, mình chưa thể xác lập một cách ghép nhóm duy nhất. Vẫn còn nhiều cách khác nhau để chia nhóm, và đây chính là cơ hội để khám phá sâu hơn về cách tổ chức và sắp xếp số liệu.

- Hoạt động 1: Có bảng tần số nhưng thiếu số liệu.
- Hoạt động 2: Có số liệu nhưng cần tự ghép nhóm, và việc ghép nhóm không có duy nhất một cách làm.

Số liệu nhỏ nhất là 160, số liệu lớn nhất là 174, nên miền chứa toàn bộ số liệu là [160; 174]. Cần chia miền này thành 5 nhóm có độ dài bằng nhau, nên độ dài mỗi nhóm là

$$\frac{174 - 160}{5} = 2,8.$$

Nhưng nếu chia nhóm theo độ dài 2,8 cm, sẽ gặp các ranh giới thập phân không đẹp:

[160; 162,8), [162,8; 165,6), [165,6; 168,4), [168,4; 171,2), [171,2; 174].

Rõ ràng, các số thập phân như 162,8 rất khó đọc, khó vẽ biểu đồ và không phù hợp với cách làm thống kê trong thực tế.

Số 3 là số nguyên nhỏ nhất gần với 2,8. Nếu chọn độ dài nhóm là 3, miền dữ liệu cần mở rộng ra một chút để bao phủ toàn bộ dữ liệu cũ. Lúc này, nếu vẫn giữ đầu dưới là 160, thì đầu trên là

$$160 + 5 \cdot 3 = 175.$$

Khi đó, các nhóm mới là

[160; 163), [163; 166), [166; 169), [169; 172), [172; 175).

Nhóm cuối, [172; 175), ghi theo dạng nửa khoảng là hợp lý, vì 175 không có trong số liệu gốc.

Khám phá

Có điều gì ngăn cản mình chọn ghép các số liệu thành năm nhóm sau đây không:

(159; 162], (162; 165], (165; 168], (168; 171], (171; 174];

hoặc

[159; 162), [162; 165), [165; 168), [168; 171), [171; 174)?

Nếu không, câu hỏi kế tiếp là có bao nhiêu cách ghép nhóm? Cách nào là tốt nhất và dựa trên tiêu chí nào để khẳng định?

Luyện tập 2

Một thư viện thống kê số người đến đọc sách vào buổi tối trong 30 ngày của tháng vừa qua như sau:

85	81	65	58	47	30	51	92	85	42
55	37	31	82	63	33	44	93	77	57
44	74	63	67	46	73	52	53	47	35

Lập bảng tần số ghép nhóm có tám nhóm ứng với tám nửa khoảng sau: [25; 34), [34; 43), [43; 52), [52; 61), [61; 70), [70; 79), [79; 88), [88; 97).

Lời giải

Mẹo

Sắp xếp số liệu theo thứ tự *tăng dần* giúp mình dễ phân nhóm hơn.

Để đảm bảo phân nhóm chính xác, số liệu nên được sắp xếp theo thứ tự tăng dần.

30	31	33	35	37	42	44	44	46	47
47	51	52	53	55	57	58	63	63	65
67	73	74	77	81	82	85	85	92	93

Nhìn vào dãy có thứ tự tăng dần, dễ dàng thấy rằng thuộc vào nhóm $[25; 34)$ có ba số liệu bao gồm 30; 31; 33. Bảng tần số ghép nhóm được thành lập dựa trên các quan sát như thế cho đến khi duyệt hết số liệu.

Nhóm	Tần số
$[25; 34)$	3
$[34; 43)$	3
$[43; 52)$	6
$[52; 61)$	5
$[61; 70)$	4
$[70; 79)$	3
$[79; 88)$	4
$[88; 97)$	2
$n = 30$	

Hoạt động 3

Trong *Bảng 4*, có bao nhiêu số liệu với giá trị không vượt quá giá trị đầu mút phải:

- a. 163 của nhóm 1?
- b. 166 của nhóm 2?
- c. 169 của nhóm 3?
- d. 172 của nhóm 4?
- e. 175 của nhóm 5?

Nhóm	Tần số
[160; 163)	6
[163; 166)	12
[166; 169)	10
[169; 172)	5
[172; 175)	3
$n = 36$	

Lời giải

- a. Nhóm 1, [160; 163), có tần số là 6. Vậy có 6 số liệu không vượt quá 163.
b. Nhóm 2, [163; 166), có tần số là 12. Các số liệu trong nhóm 1 cũng không vượt quá 166. Vậy tổng số là

$$6 + 12 = 18.$$

- c. Tương tự, số số liệu không vượt quá đầu mút phải 169 của nhóm 3 là

$$6 + 12 + 10 = 28.$$

- d. Số số liệu không vượt quá đầu mút phải 172 của nhóm 4 là

$$6 + 12 + 10 + 5 = 33.$$

- e. Số số liệu không vượt quá đầu mút phải 175 của nhóm 5 là

$$6 + 12 + 10 + 5 + 3 = 36.$$

Tần số tích lũy

Đáp số câu a cho biết tần số tích lũy của nhóm 1 là $cf_1 = 6$. Đáp số câu b cho biết tần số tích lũy của nhóm 2 là $cf_2 = 18$. Đáp số câu e cho biết tần số tích lũy của nhóm 5 là $cf_5 = 36$.
Tần số tích lũy của một nhóm là số số liệu có giá trị nhỏ hơn đầu mút phải của nhóm đó.

Luyện tập 3

Trong bài toán ở Luyện tập 2, lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy có tám nhóm ứng với tám nửa khoảng: [25; 34), [34; 43), [43; 52), [52; 61), [61; 70), [70; 79), [79; 88), [88; 97).

Lời giải

Mẹo

Có hai cách để tính tần số tích lũy cf_3 của nhóm 3.

- Cách 1: $cf_3 = n_1 + n_2 + n_3$, tính tần số tích lũy của nhóm 3 bằng cách cộng tần số của nhóm 1, nhóm 2 và nhóm 3.
- Cách 2: $cf_3 = cf_2 + n_3$, tính tần số tích lũy của nhóm 3 bằng cách cộng tần số tích lũy của nhóm 2 và tần số của nhóm 3.

Theo bạn, khi nào cách 2 nhanh hơn?

Nhóm	Tần số	Tần số tích lũy
[25; 34)	3	3
[34; 43)	3	6
[43; 52)	6	12
[52; 61)	5	17
[61; 70)	4	21
[70; 79)	3	24
[79; 88)	4	28
[88; 97)	2	30
$n = 30$		

Hoạt động 4

Xét mẫu số liệu được cho trong *Bảng 4*.

Nhóm	Tần số
[160; 163)	6
[163; 166)	12
[166; 169)	10
[169; 172)	5
[172; 175)	3
$n = 36$	

- Tìm trung điểm x_1 của nửa khoảng (tính bằng trung bình cộng của hai đầu mút) ứng với nhóm 1. Ta gọi trung điểm x_1 là *giá trị đại diện* của nhóm 1.
- Bằng cách tương tự, hãy tìm giá trị đại diện của bốn nhóm còn lại. Từ đó, hãy hoàn thiện các số liệu trong *Bảng 7*.

c. Tính giá trị \bar{x} cho bởi công thức sau:

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_5x_5}{n}.$$

Giá trị \bar{x} được gọi là *số trung bình cộng* của mẫu số liệu đã cho.

Nhóm	Giá trị đại diện	Tần số
[160; 163)	$x_1 = ?$	$n_1 = ?$
[163; 166)	$x_2 = ?$	$n_2 = ?$
[166; 169)	$x_3 = ?$	$n_3 = ?$
[169; 172)	$x_4 = ?$	$n_4 = ?$
[172; 175)	$x_5 = ?$	$n_5 = ?$
		$n = 36$

Lời giải

a. Giá trị đại diện của nhóm 1, [160; 163), là

$$x_1 = \frac{160 + 163}{2} = 161,5 \text{ (cm)}.$$

b. Tương tự, giá trị đại diện cho các nhóm còn lại được tính như dưới đây.

- Nhóm 2, [163; 166): $x_2 = \frac{163+166}{2} = 164,5 \text{ (cm)}.$
- Nhóm 3, [166; 169): $x_3 = \frac{166+169}{2} = 167,5 \text{ (cm)}.$
- Nhóm 4, [169; 172): $x_4 = \frac{169+172}{2} = 170,5 \text{ (cm)}.$
- Nhóm 5, [172; 175): $x_5 = \frac{172+175}{2} = 173,5 \text{ (cm)}.$

c. Từ *Bảng 4* và các giá trị đại diện vừa tính, *Bảng 7* được hoàn thiện như dưới đây.

Nhóm	Giá trị đại diện	Tần số
[160; 163)	$x_1 = 161,5$	$n_1 = 6$
[163; 166)	$x_2 = 164,5$	$n_2 = 12$
[166; 169)	$x_3 = 167,5$	$n_3 = 10$
[169; 172)	$x_4 = 170,5$	$n_4 = 5$
[172; 175)	$x_5 = 173,5$	$n_5 = 3$
		$n = 36$

Số trung bình cộng của mẫu số liệu là

$$\begin{aligned}\bar{x} &= \frac{1}{36}(6 \cdot 161,5 + 12 \cdot 164,5 + 10 \cdot 167,5 + 5 \cdot 170,5 + 3 \cdot 173,5) \\ &= \frac{1997}{12} \\ &\approx 166,4167 \text{ (cm)}.\end{aligned}$$

Luyện tập 4

Xác định số trung bình cộng của mẫu số liệu ghép nhóm trong bài toán ở Luyện tập 2.

Lời giải

Bảng tần số ghép nhóm và giá trị đại diện của mẫu số liệu trong bài toán ở Luyện tập 2 được cho trong bảng dưới đây.

Nhóm	Giá trị đại diện	Tần số
[25; 34)	29,5	3
[34; 43)	38,5	3
[43; 52)	47,5	6
[52; 61)	56,5	5
[61; 70)	65,5	4
[70; 79)	74,5	3
[79; 88)	83,5	4
[88; 97)	92,5	2
		$n = 30$

Trung bình cộng là

$$\begin{aligned}\bar{x} &= \frac{1}{30}(3 \cdot 29,5 + 3 \cdot 38,5 + 6 \cdot 47,5 + 5 \cdot 56,5 + 4 \cdot 65,5 + 3 \cdot 74,5 + 4 \cdot 83,5 + 2 \cdot 92,5) \\ &= \frac{296}{5} \\ &= 59,2 \text{ (người)}.\end{aligned}$$

Hoạt động 5

Trong phòng thí nghiệm, người ta chia 99 mẫu vật thành năm nhóm căn cứ trên khối lượng của chúng (đơn vị: g) và lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy như *Bảng 10*.

- a. Nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

$$\frac{n}{2} = \frac{99}{2} = 49,5 \text{ có đúng không?}$$

- b. Tìm đầu mút trái r , độ dài d , tần số n_3 của nhóm 3 và tần số tích lũy cf_2 của nhóm 2.
c. Tính giá trị M_e theo công thức sau:

$$M_e = r + \left(\frac{49,5 - cf_2}{n_3} \right) \cdot d.$$

Giá trị M_e được gọi là *trung vị* của mẫu số liệu ghép nhóm đã cho.

Nhóm	Tần số	Tần số tích lũy
[27, 5; 32, 5)	16	16
[32, 5; 37, 5)	24	40
[37, 5; 42, 5)	20	60
[42, 5; 47, 5)	30	90
[47, 5; 52, 5)	9	99
$n = 99$		

Lời giải

- a. Vì $cf_2 = 40 < \frac{n}{2} = \frac{99}{2} = 49,5 < cf_3 = 60$, nên khẳng định nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 49,5 là đúng.
b. Nhóm 3, [37, 5; 42, 5), có đầu mút trái $r = 37,5$, độ dài $d = 42,5 - 37,5 = 5$, tần số $n_3 = 20$. Tần số tích lũy của nhóm 2 là $cf_2 = 40$.
c. Giá trị M_e theo công thức là

$$\begin{aligned} M_e &= r + \left(\frac{49,5 - cf_2}{n_3} \right) \cdot d \\ &= 37,5 + \left(\frac{49,5 - 40}{20} \right) \cdot 5 \\ &= 39,875 \text{ (g)}. \end{aligned}$$

Luyện tập 5

Xác định trung vị của mẫu số liệu ghép nhóm ở *Bảng 1*.

Lời giải

Thêm vào *Bảng 1* cột tần số tích lũy để thu được bảng dưới đây.

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
$n = 120$		

Vì

$$cf_2 = 42 < \frac{n}{2} = \frac{120}{2} = 60 < cf_3 = 90,$$

nên nhóm 3, [8; 12), là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 60. Nhóm này có đầu mút trái $r = 8$, độ dài $d = 4$, tần số $n_3 = 48$ và nhóm 2 có tần số tích lũy $cf_2 = 42$.

Trung vị của mẫu số liệu là

$$\begin{aligned} M_e &= r + \left(\frac{\frac{n}{2} - cf_2}{n_3} \right) \cdot d \\ &= 8 + \left(\frac{60 - 42}{48} \right) \cdot 4 \\ &= \frac{19}{2} \\ &= 9,5 \text{ (năm)}. \end{aligned}$$

Hoạt động 6

Giáo viên chủ nhiệm chia thời gian sử dụng Internet trong một ngày của 40 học sinh thành năm nhóm (đơn vị: phút) và lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy như *Bảng 12*.

- a. Tìm trung vị M_e của mẫu số liệu ghép nhóm đó. Trung vị M_e còn gọi là tứ phân vị thứ hai Q_2 của mẫu số liệu trên.
- b. Nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

$$\frac{n}{4} = \frac{40}{4} = 10 \text{ có đúng không?}$$

Tìm đầu mút trái s , độ dài h , tần số n_2 của nhóm 2 và tần số tích lũy cf_1 của nhóm 1. Sau đó, hãy tính giá trị Q_1 theo công thức sau:

$$Q_1 = s + \left(\frac{10 - cf_1}{n_2} \right) \cdot h.$$

Giá trị nói trên được gọi là *tứ phân vị thứ nhất* Q_1 của mẫu số liệu đã cho.

- c. Nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

$$\frac{3n}{4} = \frac{3 \cdot 40}{4} = 30 \text{ có đúng không?}$$

Tìm đầu mút trái t , độ dài l , tần số n_3 của nhóm 3 và tần số tích lũy cf_2 của nhóm 2. Sau đó, hãy tính giá trị Q_3 theo công thức sau:

$$Q_3 = t + \left(\frac{30 - cf_2}{n_3} \right) \cdot l.$$

Giá trị nói trên được gọi là *tứ phân vị thứ ba* Q_3 của mẫu số liệu đã cho.

Nhóm	Tần số	Tần số tích lũy
[0; 60)	6	6
[60; 120)	13	19
[120; 180)	13	32
[180; 240)	6	38
[240; 300)	2	40
$n = 40$		

Lời giải

- a. Quan sát từ cột Tần số tích lũy,

$$cf_2 = 19 < \frac{n}{2} = \frac{40}{2} = 20 < cf_3 = 32,$$

suy ra nhóm 3, [120; 180), là nhóm đầu tiên có tần số tích lũy lớn hơn 20. Nhóm này có đầu mút trái $r = 120$, độ dài $d = 60$, tần số $n_3 = 13$ và nhóm 2 có tần số tích lũy $cf_2 = 19$. Trung vị M_e hay tứ phân vị thứ hai Q_2 là

$$\begin{aligned}
M_e = Q_2 &= r + \left(\frac{\frac{n}{2} - cf_2}{n_3} \right) \cdot d \\
&= 120 + \left(\frac{20 - 19}{13} \right) \cdot 60 \\
&= \frac{1620}{13} \\
&= 124,6154 \text{ (phút)}.
\end{aligned}$$

b. Quan sát từ cột Tần số tích lũy,

$$cf_1 = 6 < \frac{n}{4} = \frac{40}{4} = 10 < cf_2 = 19,$$

nên khẳng định nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 10 là khẳng định đúng. Nhóm 2, [60; 120), có đầu mút trái $s = 60$, độ dài $h = 60$, tần số $n_2 = 13$ và nhóm 1 có tần số tích lũy $cf_1 = 6$. Tứ phân vị thứ nhất là

$$\begin{aligned}
Q_1 &= s + \left(\frac{10 - cf_1}{n_2} \right) \cdot h \\
&= 60 + \left(\frac{10 - 6}{13} \right) \cdot 60 \\
&= \frac{1020}{13} \\
&= 78,4615 \text{ (phút)}.
\end{aligned}$$

c. Quan sát từ cột Tần số tích lũy,

$$cf_2 = 19 < \frac{n}{4} = \frac{3 \cdot 40}{4} = 30 < cf_3 = 32,$$

nên khẳng định nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 30 là khẳng định đúng. Nhóm 3, [120; 180), có đầu mút trái $t = 120$, độ dài $l = 60$, tần số $n_3 = 13$ và nhóm 2 có tần số tích lũy $cf_2 = 19$. Tứ phân vị thứ ba là

$$\begin{aligned}
Q_3 &= t + \left(\frac{30 - cf_2}{n_3} \right) \cdot l \\
&= 120 + \left(\frac{30 - 19}{13} \right) \cdot 60 \\
&= \frac{2220}{13} \\
&= 170,7692 \text{ (phút)}.
\end{aligned}$$

Luyện tập 6

Tìm tứ phân vị của mẫu số liệu trong *Bảng 1* (làm tròn các kết quả đến hàng phần mười).

Lời giải

Bảng 1 thêm vào cột Tần số tích lũy được trình bày như dưới đây.

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
$n = 120$		

- Quan sát từ cột Tần số tích lũy,

$$cf_1 = 13 < \frac{n}{4} = \frac{120}{4} = 30 < cf_2 = 42,$$

nên nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn 30. Nhóm 2, [4; 8), có đầu mút trái $s = 4$, độ dài $h = 4$, tần số $n_2 = 29$ và nhóm 1 có tần số tích lũy $cf_1 = 13$. Tứ phân vị thứ nhất là

$$\begin{aligned} Q_1 &= 4 + \left(\frac{30 - 13}{29} \right) \cdot 4 \\ &= \frac{184}{29} \\ &\approx 6 \text{ (năm)}. \end{aligned}$$

- Quan sát từ cột Tần số tích lũy,

$$cf_2 = 42 < \frac{n}{2} = \frac{120}{2} = 60 < cf_3 = 90,$$

rút ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn 60. Nhóm 3, [8; 12), có đầu mút trái $r = 8$, độ dài $d = 4$, tần số $n_3 = 48$ và nhóm 2 có tần số tích lũy $cf_2 = 42$. Tứ phân vị thứ hai là

$$Q_2 = 8 + \left(\frac{60 - 42}{48} \right) \cdot 4$$

$$= \frac{19}{2}$$

$$\approx 10 \text{ (năm)}.$$

- Quan sát từ cột Tần số tích lũy,

$$cf_2 = 42 < \frac{n}{4} = \frac{3 \cdot 120}{4} = 90 = cf_3,$$

nên nhóm 3, [8; 12), là nhóm đầu tiên có tần số tích lũy bằng 90. Tứ phân vị thứ ba là

$$Q_3 = 8 + \left(\frac{90 - 42}{48} \right) \cdot 4$$

$$= 12 \text{ (năm)}.$$

Hoạt động 7

Quan sát bảng tần số ghép nhóm bao gồm cả tần số tích lũy ở *Ví dụ 6* và cho biết:

- Nhóm nào có tần số lớn nhất;
- Đầu mút trái và độ dài của nhóm có tần số lớn nhất bằng bao nhiêu.

Lời giải

Bảng tần số ghép nhóm bao gồm cả tần số tích lũy ở *Ví dụ 6* được trình bày như dưới đây.

Nhóm	Tần số	Tần số tích lũy
[30; 40)	2	2
[40; 50)	10	12
[50; 60)	16	28
[60; 70)	8	36
[70; 80)	2	38
[80; 90)	2	40
$n = 40$		

- Nhóm 3, [50; 60), có tần số lớn nhất 16.
- Nhóm 3 có đầu mút trái 50 và độ dài 10.

Luyện tập 7

Tìm mốt của mẫu số liệu trong *Ví dụ 6* (làm tròn các kết quả đến hàng phần mười).

Lời giải

Bảng tần số ghép nhóm của mẫu số liệu *Ví dụ 6* được trình bày dưới đây.

Nhóm	Tần số
[30; 40)	2
[40; 50)	10
[50; 60)	16
[60; 70)	8
[70; 80)	2
[80; 90)	2
$n = 40$	

Quan sát cột Tần số, nhóm 3, [50; 60), là nhóm có tần số lớn nhất $n_3 = 16$, có đầu mút trái $u = 50$ và độ dài $g = 10$. Nhóm 2 có tần số $n_2 = 10$ và nhóm 4 có tần số $n_4 = 8$. Mốt là

$$\begin{aligned}M_o &= u + \left(\frac{n_3 - n_2}{2n_3 - n_2 - n_4} \right) \cdot g \\&= 50 + \left(\frac{16 - 10}{2 \cdot 16 - 10 - 8} \right) \cdot 10 \\&= \frac{380}{7} \\&= 54,3 \text{ (kg)}.\end{aligned}$$

2 Bài tập

Xem Đáp án để tra nhanh kết quả.

48,5	43,0	50,0	55,0	45,0	60,0	53,0	55,5	44,0	65,0
51,0	62,5	41,0	44,5	57,0	57,0	68,0	49,0	46,5	53,5
61,0	49,5	54,0	62,0	59,0	56,0	47,0	50,0	60,0	61,0
49,5	52,5	57,0	47,0	60,0	55,0	45,0	47,5	48,0	61,5

Bài tập 1

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

- a. Lập bảng tần số ghép nhóm cho mẫu số liệu trên có sáu nhóm ứng với sáu nửa khoảng:

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

- b. Xác định trung bình cộng, trung vị, tứ phân vị của mẫu số liệu ghép nhóm trên.

- c. Mốt của mẫu số liệu ghép nhóm trên là bao nhiêu?

Lời giải

- a. Để đảm bảo phân nhóm chính xác, số liệu nên được sắp xếp theo thứ tự tăng dần.

41,0	43,0	44,0	44,5	45,0	45,0	46,5	47,0	47,0	47,5
48,0	48,5	49,0	49,5	49,5	50,0	50,0	51,0	52,5	53,0
53,5	54,0	55,0	55,0	55,5	56,0	57,0	57,0	57,0	59,0
60,0	60,0	60,0	61,0	61,0	61,5	62,0	62,5	65,0	68,0

Nhìn vào dãy có thứ tự tăng dần, dễ dàng thấy rằng thuộc vào nhóm $[40; 45)$ có bốn số liệu bao gồm 41; 43; 44; 44,5. Bảng tần số ghép nhóm được thành lập dựa trên các quan sát như thế cho đến khi duyệt hết số liệu.

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

- b. Thêm vào bảng tần số ghép nhóm ở trên cột số thứ tự để dễ quan sát, cột Giá trị đại diện để tính trung bình, và cột Tần số tích lũy để tính tứ phân vị.

	Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
1	[40; 45)	42,5	4	4
2	[45; 50)	47,5	11	15
3	[50; 55)	52,5	7	22
4	[55; 60)	57,5	8	30
5	[60; 65)	62,5	8	38
6	[65; 70)	67,5	2	40
			$n = 40$	

- Mẫu có $n = 40$ số liệu. Nhóm 1, [40; 45), có trung điểm $x_1 = 42,5$ làm giá trị đại diện và có tần số $n_1 = 4$. Ký hiệu tương tự cho các nhóm còn lại. Trung bình cộng là

$$\begin{aligned}
 \bar{x} &= \frac{1}{n}(n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + n_4 \cdot x_4 + n_5 \cdot x_5 + n_6 \cdot x_6) \\
 &= \frac{1}{40}(4 \cdot 42,5 + 11 \cdot 47,5 + 7 \cdot 52,5 + 8 \cdot 57,5 + 8 \cdot 62,5 + 2 \cdot 67,5) \\
 &= \frac{431}{8} \\
 &= 53,875 \text{ (km/h)}.
 \end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_2 = 15 < \frac{n}{2} = \frac{40}{2} = 20 < cf_3 = 22.$$

Suy ra nhóm 3, [50; 55), là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 20. Nhóm này có đầu mút trái $r = 50$, độ dài $d = 5$ và tần số $n_3 = 7$. Trung vị là

$$\begin{aligned}
 M_e &= r + \left(\frac{\frac{n}{2} - cf_2}{n_3} \right) \cdot d \\
 &= 50 + \left(\frac{20 - 15}{7} \right) \cdot 5 \\
 &= \frac{375}{7} \\
 &\approx 53,5714 \text{ (km/h)}.
 \end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_1 = 4 < \frac{n}{4} = \frac{40}{4} = 10 < cf_2 = 15.$$

Suy ra nhóm 2, [45; 50), là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 10. Nhóm này có đầu mút trái $s = 45$, độ dài $h = 5$ và tần số $n_2 = 11$. Tứ phân vị thứ nhất là

$$\begin{aligned} Q_1 &= s + \left(\frac{\frac{n}{4} - cf_1}{n_2} \right) \cdot h \\ &= 45 + \left(\frac{10 - 4}{11} \right) \cdot 5 \\ &= \frac{525}{11} \\ &\approx 47,7273 \text{ (km/h)}. \end{aligned}$$

- Tứ phân vị thứ hai Q_2 chính là trung vị M_e , hay $Q_2 = M_e = 53,5714$ (km/h).
- Quan sát cột Tần số tích lũy, ta thấy

$$cf_3 = 22 < \frac{3n}{4} = \frac{3 \cdot 40}{4} = 30 = cf_4.$$

Suy ra nhóm 4, [55; 60), là nhóm đầu tiên có tần số tích lũy bằng với 30. Nhóm này có đầu mút trái $t = 55$, độ dài $l = 5$ và tần số $n_4 = 8$. Tứ phân vị thứ ba là

$$\begin{aligned} Q_3 &= t + \left(\frac{\frac{3n}{4} - cf_3}{n_4} \right) \cdot l \\ &= 55 + \left(\frac{30 - 22}{8} \right) \cdot 5 \\ &= 60 \text{ (km/h)}. \end{aligned}$$

- c. Quan sát cột Tần số, nhóm 2, [45; 50), là nhóm có tần số lớn nhất $n_2 = 11$. Nó có đầu mút trái $u = 45$ và độ dài $g = 5$. Nhóm 1 có tần số $n_1 = 4$ và nhóm 3 có tần số $n_3 = 7$. Mốt là

$$\begin{aligned} M_o &= u + \left(\frac{n_2 - n_1}{2n_2 - n_1 - n_3} \right) \cdot g \\ &= 45 + \left(\frac{11 - 4}{2 \cdot 11 - 4 - 7} \right) \cdot 5 \\ &= \frac{530}{11} \\ &\approx 48,1818 \text{ (km/h)}. \end{aligned}$$

Tóm lại, bảng tần số ghép nhóm là

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

Các số đặc trưng đo xu thế trung tâm là

\bar{x}	Q_1	$Q_2 (M_e)$	Q_3	M_o
53,875	47,7273	53,5714	60	48,1818.

Bài tập 2

Mẫu số liệu sau ghi lại cân nặng của 30 bạn học sinh (đơn vị: kg):

17,0	40,0	39,0	40,5	42,0	51,0	41,5	39,0	41,0	30,0
40,0	42,0	40,5	39,5	41,0	40,5	37,0	39,5	40,0	41,0
38,5	39,5	40,0	41,0	39,0	40,5	40,0	38,5	39,5	41,5

a. Lập bảng tần số ghép nhóm cho mẫu số liệu trên có tám nhóm ứng với tám nửa khoảng:

[15; 20), [20; 25), [25; 30), [30; 35), [35; 40), [40; 45), [45; 50), [50; 55).

b. Xác định trung bình cộng, trung vị, tứ phân vị của mẫu số liệu ghép nhóm trên.

c. Mốt của mẫu số liệu ghép nhóm trên là bao nhiêu?

Lời giải

a. Để đảm bảo phân nhóm chính xác, số liệu nên được sắp xếp theo thứ tự tăng dần.

17,0	30,0	37,0	38,5	38,5	39,0	39,0	39,0	39,5	39,5
39,5	39,5	40,0	40,0	40,0	40,0	40,0	40,5	40,5	40,5
40,5	41,0	41,0	41,0	41,0	41,5	41,5	42,0	42,0	51,0

Nhìn vào dãy có thứ tự tăng dần, dễ dàng thấy rằng thuộc vào nhóm [15; 20) chỉ có một số liệu là 17. Bảng tần số ghép nhóm được thành lập dựa trên các quan sát như thế cho đến khi duyệt hết số liệu.

Nhóm	Tần số
[15; 20)	1
[20; 25)	0
[25; 30)	0
[30; 35)	1
[35; 40)	10
[40; 45)	17
[45; 50)	0
[50; 55)	1
$n = 30$	

- b. Thêm vào bảng tần số ghép nhóm ở trên cột số thứ tự để dễ quan sát, cột Giá trị đại diện để tính trung bình, và cột Tần số tích lũy để tính tứ phân vị.

	Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
1	[15; 20)	17,5	1	1
2	[20; 25)	22,5	0	1
3	[25; 30)	27,5	0	1
4	[30; 35)	32,5	1	2
5	[35; 40)	37,5	10	12
6	[40; 45)	42,5	17	29
7	[45; 50)	47,5	0	29
8	[50; 55)	52,5	1	30
$n = 30$				

- Mẫu có $n = 30$ số liệu. Nhóm 1, [15; 20), có trung điểm $x_1 = 17,5$ làm giá trị đại diện và có tần số $n_1 = 1$. Ký hiệu tương tự với các nhóm còn lại. Trung bình là

$$\begin{aligned}\bar{x} &= \frac{1}{n}(n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + n_4 \cdot x_4 + n_5 \cdot x_5 + n_6 \cdot x_6 + n_7 \cdot x_7 + n_8 \cdot x_8) \\ &= \frac{1}{30}(1 \cdot 17,5 + 0 \cdot 22,5 + 0 \cdot 27,5 + 1 \cdot 32,5 + 10 \cdot 37,5 + 17 \cdot 42,5 + 0 \cdot 47,5 + 52,5 \cdot 1) \\ &= 40 \text{ (kg)}.\end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_5 = 12 < \frac{n}{2} = 15 < cf_6 = 29.$$

Suy ra nhóm 6, [40; 45), là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 15. Nhóm này có đầu mút trái $r = 40$, độ dài $d = 5$ và tần số $n_6 = 17$. Trung vị là

$$\begin{aligned}M_e &= r + \left(\frac{\frac{n}{2} - cf_5}{n_6} \right) \cdot d \\ &= 40 + \left(\frac{15 - 12}{17} \right) \cdot 5 \\ &= \frac{695}{17} \\ &\approx 40,8824 \text{ (kg)}.\end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_4 = 2 < \frac{n}{4} = \frac{30}{4} = 7,5 < cf_5 = 12.$$

Suy ra nhóm 5, $[35; 40)$, là nhóm đầu tiên có tần số tích lũy, không nhỏ hơn 7,5. Nhóm này có đầu mút trái $s = 35$, độ dài $h = 5$ và tần số $n_5 = 10$. Tứ phân vị thứ nhất là

$$\begin{aligned} Q_1 &= s + \left(\frac{\frac{n}{4} - cf_4}{n_5} \right) \cdot h \\ &= 35 + \left(\frac{7,5 - 2}{10} \right) \cdot 5 \\ &= \frac{151}{4} \\ &= 37,75 \text{ (kg)}. \end{aligned}$$

- Tứ phân vị thứ hai Q_2 chính là trung vị M_e , hay $Q_2 = M_e \approx 40,8824$ (kg).
- Quan sát cột Tần số tích lũy, ta có

$$cf_5 = 12 < \frac{3n}{4} = \frac{3 \cdot 30}{4} = 22,5 < cf_6 = 29.$$

Suy ra nhóm 6, $[40; 45)$ là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 22,5. Nhóm này có đầu mút trái $t = 40$, độ dài $l = 5$ và tần số $n_6 = 17$. Tứ phân vị thứ ba là

$$\begin{aligned} Q_3 &= t + \left(\frac{\frac{3n}{4} - cf_5}{n_6} \right) \cdot l \\ &= 40 + \left(\frac{22,5 - 12}{17} \right) \cdot 5 \\ &= \frac{1465}{34} \\ &\approx 43,0882 \text{ (kg)}. \end{aligned}$$

- c. Quan sát cột Tần số, nhóm 6, $[40; 45)$ là nhóm có tần số lớn nhất $n_6 = 17$. Nhóm này có đầu mút trái $u = 40$ và độ dài $g = 5$. Nhóm 5 liền trước nó có tần số $n_5 = 10$ và nhóm 7 liền sau nó có tần số $n_7 = 0$. Mốt là

$$\begin{aligned} M_o &= u + \left(\frac{n_6 - n_5}{2n_6 - n_5 - n_7} \right) \cdot g \\ &= 40 + \left(\frac{17 - 10}{2 \cdot 17 - 10 - 0} \right) \cdot 5 \\ &= \frac{995}{24} \\ &\approx 41,5 \text{ (kg)}. \end{aligned}$$

Tóm lại, bảng tần số ghép nhóm là

$[15; 20)$	$[20; 25)$	$[25; 30)$	$[30; 35)$	$[35; 40)$	$[40; 45)$	$[45; 50)$	$[50; 55)$
1	0	0	1	10	17	0	1.

Các số đặc trưng đo xu thế trung tâm là

\bar{x}	Q_1	$Q_2 (M_e)$	Q_3	M_o
40	37,75	40,8824	43,0882	41,4583.

Bài tập 3

Bảng 15 cho ta bảng tần số ghép nhóm số liệu thống kê chiều cao của 40 mẫu cây ở một vườn thực vật (đơn vị: cm).

- Xác định trung bình cộng, trung vị, tứ phân vị của mẫu số liệu ghép nhóm trên.
- Mốt của mẫu số liệu ghép nhóm trên là bao nhiêu?

Nhóm	Tần số	Tần số tích lũy
[30; 40)	4	4
[40; 50)	10	14
[50; 60)	14	28
[60; 70)	6	34
[70; 80)	4	38
[80; 90)	2	40
$n = 40$		

Lời giải

- Thêm vào bảng đã cho cột Giá trị đại diện để tính trung bình, và cột số thứ tự để dễ quan sát.

	Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
1	[30; 40)	35	4	4
2	[40; 50)	45	10	14
3	[50; 60)	55	14	28
4	[60; 70)	65	6	34
5	[70; 80)	75	4	38
6	[80; 90)	85	2	40
$n = 40$				

- Mẫu có $n = 40$ số liệu. Nhóm 1 có giá trị đại diện $x_1 = 35$ và tần số $n_1 = 4$, vậy vậy cho đến nhóm 6 có giá trị đại diện $x_6 = 85$ và tần số $n_6 = 2$. Trung bình là

$$\begin{aligned}
\bar{x} &= \frac{1}{n}(n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + n_4 \cdot x_4 + n_5 \cdot x_5 + n_6 \cdot x_6) \\
&= \frac{1}{40}(4 \cdot 35 + 10 \cdot 45 + 14 \cdot 55 + 6 \cdot 65 + 4 \cdot 75 + 2 \cdot 85) \\
&= \frac{111}{2} \\
&= 55,5 \text{ (cm)}.
\end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_2 = 14 < \frac{n}{2} = \frac{40}{2} = 20 < cf_3 = 28.$$

Suy ra nhóm 3, [50; 55), là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 20. Nhóm này có đầu mút trái $r = 50$, độ dài $d = 10$ và tần số $n_3 = 14$. Trung vị là

$$\begin{aligned}
M_e &= r + \left(\frac{\frac{n}{2} - cf_2}{n_3} \right) \cdot d \\
&= 50 + \left(\frac{20 - 14}{14} \right) \cdot 10 \\
&= \frac{380}{7} \\
&\approx 54,2857 \text{ (cm)}.
\end{aligned}$$

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_1 = 4 < \frac{n}{4} = \frac{40}{4} = 10 < cf_2 = 14.$$

Suy ra nhóm 2, [40; 50), là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 10. Nhóm này có đầu mút trái $s = 40$, độ dài $h = 10$ và tần số $n_2 = 10$. Tứ phân vị thứ nhất là

$$\begin{aligned}
Q_1 &= s + \left(\frac{\frac{n}{4} - cf_1}{n_2} \right) \cdot h \\
&= 40 + \left(\frac{10 - 4}{10} \right) \cdot 10 \\
&= 46 \text{ (cm)}.
\end{aligned}$$

- Tứ phân vị thứ hai Q_2 chính là trung vị M_e , hay $Q_2 = M_e = 54,2857 \text{ (cm)}$.

- Quan sát cột Tần số tích lũy, ta thấy

$$cf_3 = 28 < \frac{3n}{4} = \frac{3 \cdot 40}{4} = 30 < cf_4 = 34.$$

Suy ra nhóm 4, $[60; 70)$, là nhóm đầu tiên có tần số tích lũy không nhỏ hơn 30. Nhóm này có đầu mút trái $t = 60$, độ dài $l = 10$ và tần số $n_4 = 6$. Tứ phân vị thứ ba là

$$\begin{aligned} Q_3 &= t + \left(\frac{\frac{3n}{4} - cf_3}{n_4} \right) \cdot l \\ &= 60 + \left(\frac{30 - 28}{6} \right) \cdot 10 \\ &= \frac{190}{3} \\ &= 63,3333 \text{ (cm)}. \end{aligned}$$

- b. Quan sát cột Tần số, nhóm 3, $[50; 60)$, là nhóm có tần số lớn nhất $n_3 = 14$. Nó có đầu mút trái $u = 50$ và độ dài $g = 10$. Nhóm 2 có tần số $n_2 = 10$ và nhóm 4 có tần số $n_4 = 6$. Mốt là

$$\begin{aligned} M_o &= u + \left(\frac{n_3 - n_2}{2n_3 - n_2 - n_4} \right) \cdot g \\ &= 50 + \left(\frac{14 - 10}{2 \cdot 14 - 10 - 6} \right) \cdot 10 \\ &= \frac{160}{3} \\ &\approx 53,3333 \text{ (cm)}. \end{aligned}$$

Tóm lại, bảng tần số ghép nhóm là

$[30; 40)$	$[40; 50)$	$[50; 60)$	$[60; 70)$	$[70; 80)$	$[80; 90)$
1	0	0	1	10	17.

Các số đặc trưng đo xu thế trung tâm là

\bar{x}	Q_1	$Q_2 (M_e)$	Q_3	M_o
55,5	46	54,2857	63,3333	53,3333.