

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

Chương 5 | Cánh Diều | Toán 11

ZO | 2025-03-15

Đây là Bài 1: Các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm, Chương V: Một số yếu tố thống kê và xác suất, bộ Cánh Diều, Toán 11 (trang 14, tập 2).

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.

Nội dung *Bảng 1*:

- **Nhóm:** Mỗi nhóm là một khoảng tuổi ô-tô.
- **Tần số:** Số lượng ô-tô thuộc nhóm đó.

Câu trả lời bắt chước theo sách giáo khoa :

- 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.

Suy nghĩ sâu hơn

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 - vận dụng 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
[16; 20)	8
$n = 120$	

Lời giải

Mẫu số liệu ghép nhóm ở *Bảng 1* có 120 số liệu. Có 5 nhóm. Tần số của các nhóm 1, 2, 3, 4, 5 lần lượt là 13, 29, 48, 22, 8.

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.



Mẹo

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

Số liệu nhỏ nhất là 160, số liệu lớn nhất là 174. Mình cần chia 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, mình sẽ gặp các ranh giới 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 2,8; 162,8; 165,6 vâng vâng, 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 mình 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; 3), [163; 166), [166; 169), [169; 172), [172; 175).

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

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):

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

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

[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, trước tiên mình sắp xếp số liệu 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, bạn dễ dàng thấy rằng thuộc vào nhóm [40; 45) có bốn số liệu bao gồm 41; 43; 44; 44,5. Hãy tiếp tục quan sát để thu được bảng tần số ghép nhóm.

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

- b. Mình 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

(continued)

	Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
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 một, [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

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

- Nhóm ba, [50; 55), là nhóm đầu tiên có tần số tích lũy $cf_3 = 22$ không nhỏ hơn $\frac{n}{2} = 20$. Nhóm này có đầu mút trái $r = 50$, độ dài $d = 5$ và tần số $n_3 = 7$. Nhóm hai có tần số tích lũy $cf_2 = 15$. Trung vị là

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

- Nhóm hai, [45; 50), là nhóm đầu tiên có tần số tích lũy $cf_2 = 15$ không nhỏ hơn $\frac{n}{4} = 10$. Nhóm này có đầu mút trái $s = 45$, độ dài $h = 5$ và tần số $n_2 = 11$. Nhóm một có tần số tích lũy là $cf_1 = 4$. 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).
- Nhóm bốn, $[55; 60)$, là nhóm đầu tiên có tần số tích lũy $cf_4 = 30$ bằng với $\frac{3n}{4} = \frac{3 \cdot 40}{4} = 30$. Nhóm này có đầu mút trái $t = 55$, độ dài $l = 5$ và tần số $n_4 = 8$. Nhóm thứ ba có tần số tích lũy là $cf_3 = 22$. 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. Nhóm hai, $[45; 50)$, là nhóm có tần số lớn nhất $n_2 = 11$. Nó có đầu mút trái $u = 45$ và độ dài $g = 5$. Nhóm một có tần số $n_1 = 4$ và nhóm ba 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, trước tiên mình sắp xếp số liệu 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, bạn dễ dàng thấy rằng thuộc vào nhóm $[15; 20)$ chỉ có một số liệu là 17. Hãy tiếp tục quan sát để thu được bảng tần số ghép nhóm.

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
$n = 40$	

- b. Mình 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

(continued)

	Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
6	[40; 45)	42,5	17	29
7	[45; 50)	47,5	0	29
8	[50; 55)	52,5	1	30
$n = 40$				

- Mẫu có $n = 30$ số liệu. Nhóm một, $[15; 20)$, 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

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

- Nhóm sáu, $[40; 45)$, là nhóm đầu tiên có tần số tích lũy $cf_6 = 29$ không nhỏ hơn $\frac{n}{2} = 15$. Nhóm này có đầu mút trái $r = 40$, độ dài $d = 5$ và tần số $n_6 = 17$. Nhóm năm liền trước nó có tần số tích lũy $cf_5 = 12$. Trung vị là

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

- Nhóm năm $[35; 40)$ là nhóm đầu tiên có tần số tích lũy $cf_5 = 12$ không nhỏ hơn $\frac{n}{4} = 7,5$. Nhóm này có đầu mút trái $s = 35$, độ dài $h = 5$ và tần số $n_5 = 10$. Nhóm bốn liền trước nó có tần số tích lũy $cf_4 = 2$. 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 $Q_2 = M_e \approx 40,8824 \text{ (kg)}$.

- Nhóm sáu [40; 45) là nhóm đầu tiên có tần số tích lũy $cf_6 = 29$ không nhỏ hơn $\frac{3n}{4} = 22,5$. Nhóm này có đầu mút trái $t = 40$, độ dài $l = 5$ và tần số $n_6 = 17$. Nhóm năm liền trước nó có tần số tích lũy là $cf_5 = 12$. 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. Nhóm sáu [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 năm liền trước nó có tần số $n_5 = 10$ và nhóm bảy 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).

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

b. 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

a. Mình sẽ 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 một có giá trị đại diện $x_1 = 35$ và tần số $n_1 = 4$, vậy vậy cho đến nhóm sáu có giá trị đại diện $x_6 = 85$ và tần số $n_6 = 2$. Trung bình là

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

- Nhóm ba, [50; 55), là nhóm đầu tiên có tần số tích lũy $cf_3 = 28$ không nhỏ hơn $\frac{n}{2} = 20$. Nhóm này có đầu mút trái $r = 50$, độ dài $d = 10$ và tần số $n_3 = 14$. Nhóm hai liền trước nó có tần số tích lũy $cf_2 = 14$. Trung vị là

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

- Nhóm hai, $[40; 50)$, là nhóm đầu tiên có tần số tích lũy $cf_2 = 14$ không nhỏ hơn $\frac{n}{4} = 10$. Nhóm này có đầu mút trái $s = 40$, độ dài $h = 10$ và tần số $n_2 = 10$. Nhóm một có tần số tích lũy là $cf_1 = 4$. 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 $Q_2 = M_e = 54,2857 \text{ (cm)}$.
- Nhóm bốn, $[60; 70)$, là nhóm đầu tiên có tần số tích lũy $cf_4 = 34$ không nhỏ hơn $\frac{3n}{4} = 30$. Nhóm này có đầu mút trái $t = 60$, độ dài $l = 10$ và tần số $n_4 = 6$. Nhóm thứ ba có tần số tích lũy là $cf_3 = 28$. 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. Nhóm ba, $[50; 60)$, là nhóm có tần số lớn nhất $n_3 = 14$. Nó có đầu mút trái $u = 50$ và độ dài

$g = 10$. Nhóm hai có tần số $n_2 = 10$ và nhóm bốn 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.