

Trong thực tế ta thường gặp những tập số liệu rất lớn, ví dụ như điểm thi tốt nghiệp Trung học phổ thông của học sinh cả nước hay số tiền thuế thu nhập cá nhân mỗi cá nhân phải đóng trong năm. Để thuận lợi cho việc lưu trữ, trình bày và biểu diễn các số liệu này, người ta thường sử dụng bảng số liệu ghép nhóm, tổng đó các số liệu riêng lẻ có giá trị gần nhau được sắp xếp vào từng nhóm.

Trong chương này, chúng ta sẽ tìm hiểu cách xác định, ý nghĩa và vai trò của các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm.

**Học xong chương này, bạn có thể:**

- Tính được các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm; số trung bình cộng (hay số trung bình), trung vị, tứ phân vị, mốt.
- Hiểu được ý nghĩa và vai trò của các số đặc trưng nói trên của mẫu số liệu trong thực tiễn.
- Rút ra được kết luận nhờ ý nghĩa của các số đặc trưng nói trên của mẫu số liệu trong trường hợp đơn giản.
- Nhận biết được mối liên hệ giữa thống kê với những kiến thức của các môn học khác trong chương trình lớp 11 và trong thực tiễn.

BÀI 1. SỐ TRUNG BÌNH VÀ MỐT CỦA MẪU SỐ LIỆU GHÉP NHÓM.

Từ khoá : Số trung bình, mốt.

► Một đại lý bảo hiểm đã thống kê số lượng khách mua bảo hiểm nhân thọ trong một ngày ở biểu đồ bên.

Hãy so sánh độ tuổi trung bình của khách hàng nam nữ.



1. Số liệu ghép nhóm.

• 1. Sử dụng dữ liệu ở biểu đồ trong ►, hoàn thiện bảng thống kê về số lượng khách hàng nữ theo tuổi sau :

Khoảng tuổi	[20; 30)	[30; 40)	[40; 50)	[50; 60)	[60; 70)
Số khách hàng nữ	3	?	?	?	?

Một số loại số liệu điều tra có thể nhận rất nhiều những giá trị khác nhau, hoặc khó xác định được giá trị chính xác, ví dụ như chiều cao, cân nặng, tuổi thọ, ... Để thuận tiện cho việc lưu trữ và xử lý các loại số liệu này, người ta thường ghép các số liệu gần nhau lại thành nhóm.

► 1. Mẫu số liệu ghép nhóm thường được trình bày dưới dạng bảng thống kê có dạng 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 của 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 $[u_j; u_{j+1})$.

Ví dụ 1. Tính giá trị đại diện và độ dài của mỗi nhóm trong mẫu số liệu ở •1.

Giải

Khoảng tuổi	[20; 30)	[30; 40)	[40; 50)	[50; 60)	[60; 70)
Giá trị đại diện	25	35	45	55	65
Độ dài của nhóm	10	10	10	10	10

Một số quy tắc ghép nhóm của mẫu số liệu

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

- Sử dụng 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 thoả mãn $R < k.L$, trong đó R là khoảng biến thiên, k là số nhóm.

- Giá trị nhỏ nhất của mẫu số 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 $R = 63,6 - 45,1 = 18,5$.

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 số ghép nhóm như 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

Chú ý:

- Các đầu mút của các nhóm có thể không là giá trị của mẫu số liệu.
- Ta hay gặp các bảng số liệu ghép nhóm là số nguyên, chẳng hạn như bảng thống kê số lỗi chính tả trong bài kiểm tra giữa học kì I môn Ngữ Văn của học sinh khối 11 như sau :

Số lỗi	[1; 2]	[3; 4]	[5; 6]	[7; 8]	[9; 10]
Số bài	122	75	14	5	2

Bảng số liệu này không có dạng như bảng 1. Để thuận lợi cho việc tính các số đặc trưng cho bảng số liệu này, người ta hiệu chỉnh về dạng như Bảng 1 bằng cách thêm và bớt $0,5$ đơn vị vào đầu mút bên phải và bên trái của mỗi nhóm số liệu như sau :

Số lỗi	[0,5; 2,5)	[2,5; 4,5)	[4,5; 6,5)	[6,5; 8,5)	[8,5; 10,5)
Số bài	122	75	14	5	2

•1. Một cửa hàng đã thống kê số ba lô bán được mỗi ngày trong tháng 9 với kết quả cho như sau :

12 29 12 19 15 21 19 29 28 12 15 25 16 20 29
21 12 24 14 10 12 10 23 27 28 18 16 10 20 21

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

2. Số trung bình

2. 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

a) Tính giá trị đại diện $c_i, 1 \leq i \leq 5$, của từng nhóm số liệu.

b) Tính $n_1c_1 + n_2c_2 + n_3c_3 + n_4c_4 + n_5c_5$.

$$\bar{x} = \frac{n_1c_1 + n_2c_2 + n_3c_3 + n_4c_4 + n_5c_5}{40}$$

c) Tính $\bar{x}$.

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 :

$$\bar{x} = \frac{n_1c_1 + n_2c_2 + \dots + n_kc_k}{n}, \text{ trong đó } n = n_1 + n_2 + \dots + n_k.$$

Ví dụ 3. Kết quả khảo sát cân nặng của 25 quả cam ở mỗi lô hàng A và B được cho ở bảng sau:

Cân nặng (g)	[150; 155)	[155; 160)	[160; 165)	[165; 170)	[170; 175)
Số quả cam ở lô hàng A	2	6	12	4	1
Số quả cam ở lô hàng B	1	3	7	10	4

a) Hãy ước lượng cân nặng trung bình của mỗi quả cam ở lô hàng A và lô hàng B.

b) Nếu so sánh theo số trung bình thì cam ở lô hàng nào nặng hơn ?

Giải

Ta có bảng thống kê số lượng cam theo giá trị đại diện :

Cân nặng đại diện (g)	152,5	157,5	162,5	167,5	172,5
Số quả cam ở lô hàng A	2	6	12	4	1
Số quả cam ở lô hàng B	1	3	7	10	4

a) Cân nặng trung bình của mỗi quả cam ở lô hàng A xấp xỉ bằng

$$(2.152,5 + 6.157,5 + 12.162,5 + 4.167,5 + 1.172,5) : 25 = 161,7(\text{g})$$

Cân nặng trung bình của mỗi quả cam ở lô hàng B xấp xỉ bằng

$$(1.152,5 + 3.157,5 + 7.162,5 + 10.167,5 + 4.172,5) : 25 = 165,1(\text{g})$$

b) Nếu so sánh theo số trung bình thì cam ở lô hàng B nặng hơn cam ở lô hàng A.

Ý nghĩa của số trung bình của mẫu số liệu ghép nhóm

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.

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

Hãy ước lượng cân nặng trung bình của học sinh trong Ví dụ 2 sau khi ghép nhóm và so sánh kết quả tìm được với cân nặng trung bình của mẫu số liệu gốc.

3. Một

Từ mẫu số liệu ở , hãy cho biết khách hàng nam và khách hàng nữ ở khoảng tuổi nào mua bảo hiểm nhân thọ nhiều nhất. Ta có thể biết một của mẫu số liệu đó không?

Xét mẫu số liệu cho ở dạng Bảng 1.

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

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

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

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

Ví dụ 4. 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

a) Tìm một của mẫu số liệu ghép nhóm trên.

b) Công ty nên xây nhà ở mức giá nào để nhiều người có nhu cầu mua nhất?

Giải

a) Nhóm chứa một của mẫu số liệu trên là nhóm $[18;22)$.

Do đó $u_m = 18, n_{m-1} = 78, n_m = 120, n_{m+1} = 45, u_{m+1} - u_m = 22 - 18 = 4$.

Một của mẫu số liệu ghép nhóm là

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

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.

Ví dụ 5. Số cuộc gọi điện thoại một người thực hiện mỗi ngày trong 30 ngày được lựa chọn ngẫu nhiên được thống kê trong bảng sau:

Số cuộc gọi	[3;5]	[6;8]	[9;11]	[12;14]	[15;17]
Số ngày	5	13	7	3	2

a) Tìm một của mẫu số liệu ghép nhóm trên.

b) Hãy dự đoán xem khả năng người đó thực hiện bao nhiêu cuộc gọi mỗi ngày là cao nhất.

Giải

Do số cuộc gọi là số nguyên nên ta hiệu chỉnh lại như sau:

Số cuộc gọi	[2,5;5,5)	[5,5;8,5)	[8,5;11,5)	[11,5;14,5)	[14,5;17,5)
-------------	-----------	-----------	------------	-------------	-------------

Số ngày	5	13	7	3	2
---------	---	----	---	---	---

a) Nhóm chứa một của mẫu số liệu trên là nhóm $[5,5;8,5)$.

Do đó $u_m = 5,5; n_{m-1} = 5; n_m = 13; n_{m+1} = 7; u_{m+1} - u_m = 8,5 - 5,5 = 3$.

Mốt của mẫu số liệu ghép nhóm là

$$M_o = 5,5 + \frac{13 - 5}{(13 - 5) + (13 - 7)} \cdot 3 \approx 7,2$$

b) Dựa vào kết quả trên ta có thể dự đoán rằng khả năng người đó thực hiện 7 cuộc gọi mỗi ngày là cao nhất.

Ý nghĩa của một của mẫu số liệu ghép nhóm

- 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 một và nhiều mốt.

Hãy sử dụng dữ liệu ở để tư vấn cho đại lý bảo hiểm xác định khách hàng nam và nữ ở tuổi nào hay mua bảo hiểm nhất.

BÀI TẬP

1. Anh Văn ghi lại cự li 30 lần ném lao của mình ở bảng sau (đơn vị: mét):

72,1	72,9	70,2	70,9	72,2	71,5	72,5	69,3	72,3	69,7
72,3	71,5	71,2	69,8	72,3	71,1	69,5	72,2	71,9	73,1
71,6	71,3	72,2	71,8	70,8	72,2	72,2	72,9	72,7	70,7

a) Tính cự li trung bình của mỗi lần ném.

b) Tổng hợp lại kết quả ném của anh Văn vào bảng tần số ghép nhóm theo mẫu sau:

Cự li (m)	[69,2;70)	[70;70,8)	[70,8;71,6)	[71,6;72,4)	[72,4;73,2)
Số lần	?	?	?	?	?

c) Hãy ước lượng cự li trung bình mỗi lần ném từ bảng tần số ghép nhóm trên.

d) Khả năng anh Văn ném được khoảng bao nhiêu mét là cao nhất?

2. Người ta đếm số xe ô tô đi qua một trạm thu phí mỗi phút trong khoảng thời gian từ 9 giờ đến 9 giờ 30 phút sáng. Kết quả được ghi lại ở bảng sau:

15	16	13	21	17	23	15	21	6	11	12	23	19	25	11
25	7	29	10	28	29	24	6	11	23	11	21	9	27	15

a) Tính số xe trung bình đi qua trạm thu phí trong mỗi phút.

b) Tổng hợp lại số liệu trên vào bảng tần số ghép nhóm theo mẫu sau:

Số xe	[6;10]	[11;15]	[16;20]	[21;25]	[26;30]
Số lần	?	?	?	?	?

c) Hãy ước lượng trung bình số xe đi qua trạm thu phí trong mỗi phút từ bảng tần số ghép nhóm trên.

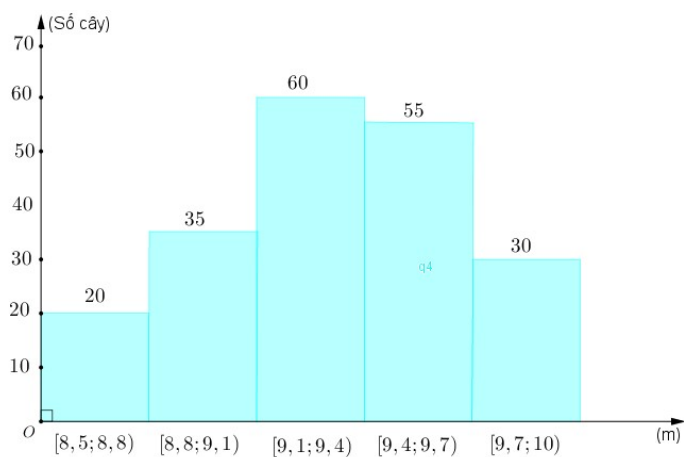
3. Một thư viện thống kê số lượng sách được mượn mỗi ngày trong ba tháng ở bảng sau:

Số sách	[16;20]	[21;25]	[26;30]	[31;35]	[36;40]	[41;45]	[46;50]
Số ngày	3	6	15	27	22	14	5

Hãy ước lượng số trung bình và một của mẫu số liệu ghép nhóm trên.

4. Kết quả đo chiều cao của 200 cây keo 3 năm tuổi ở một nông trường được biểu diễn ở biểu đồ dưới đây.

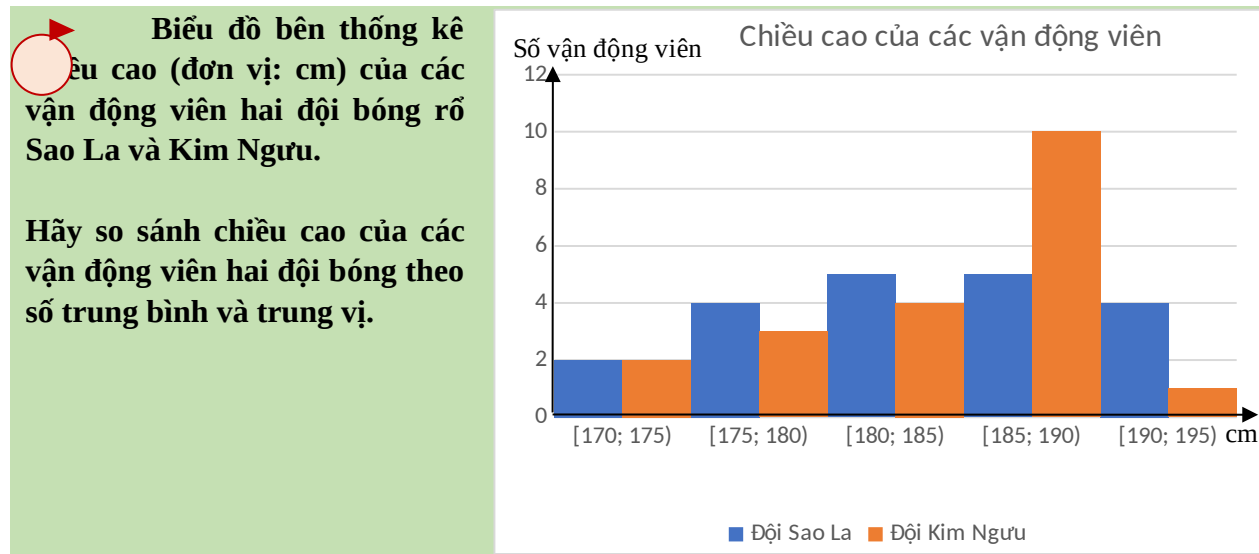
Chiều cao 200 cây keo 3 năm tuổi



Hãy ước lượng số trung bình và mốt của mẫu số liệu ghép nhóm trên.

Bài 2. TRUNG VỊ VÀ TỬ PHÂN VỊ CỦA MẪU SỐ LIỆU GHEP NHÓM

Từ khóa: Trung vị; Tử phân vị



1. Trung vị



a) Sử dụng biểu đồ ở , hoàn thiện bảng thống kê sau:

Chiều cao (cm)	[170; 175)	[175; 180)	[180; 185)	[185; 190)	[190; 195)
Đội Sao La	2	4	5	5	4
Đội Kim Ngưu	?	?	?	?	?

b) Tìm các nhóm chứa giá trị trung vị chiều cao thành viên mỗi đội.

Gọi $x_1; x_2; \dots; x_{20}$ là chiều cao của 20 thành viên đội Sao La xếp theo thứ tự không giảm. Từ bảng số liệu trên ta thấy $x_1; x_2 \in [170; 175)$; $x_3; \dots; x_6 \in [175; 180)$; $x_7; \dots; x_{11} \in [180; 185)$ nên

trung vị của mẫu số liệu $x_1; x_2; \dots; x_{20}$ là $\frac{1}{2}(x_{10} + x_{11})$ sẽ thuộc nhóm $[180; 185)$.

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;
- n_m là tần số của nhóm chứa trung vị;
- $C = n_1 + n_2 + \dots + n_{m-1}$.

Khi đó

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

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.

Giải

Gọi $x_1; x_2; \dots; x_{25}$ là cân nặng của 25 quả bơ xếp theo thứ tự không giảm. Do $x_1 \in [150; 155)$; $x_2; \dots; x_8 \in [155; 160)$; $x_9; \dots; x_{20} \in [160; 165)$ nên trung vị của mẫu số liệu $x_1; x_2; \dots; x_{25}$ là $x_{13} \in [160; 165)$.

Ta xác định được $n = 15, n_m = 12, C = 1 + 7 = 8, u_m = 160, u_{m+1} = 165$.

Vậy trung vị của mẫu số liệu ghép nhóm là

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

Ví dụ 2. Trong tuần lễ bảo vệ môi trường, các học sinh khối 11 tiến hành thu nhặt vỏ chai nhựa để tái chế. Nhà trường thống kê kết quả thu nhặt vỏ chai của học sinh khối 11 ở bảng sau:

Số vỏ chai nhựa	[11; 15]	[16; 20]	[21; 25]	[26; 30]	[31; 35]
Số học sinh	53	82	48	39	18

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

Giải

Do số vỏ chai nhựa là số nguyên nên ta hiệu chỉnh lại như sau:

Số vỏ chai nhựa	[10,5; 15,5)	[15,5; 20,5)	[20,5; 25,5)	[25,5; 30,5)	[30,5; 35,5)
Số học sinh	53	82	48	39	18

Số học sinh tham gia thu nhặt vỏ chai nhựa là $n = 53 + 82 + 48 + 39 + 18 = 240$.

Gọi $x_1; x_2; \dots; x_{240}$ lần lượt là số vỏ chai nhựa 240 học sinh khối 11 thu nhặt được xếp theo thứ tự không giảm. Do $x_1; \dots; x_{53} \in [10,5; 15,5)$; $x_{54}; \dots; x_{135} \in [15,5; 20,5)$ nên trung vị của mẫu số

liệu $x_1; x_2; \dots; x_{240}$ là $\frac{1}{2}(x_{120} + x_{121}) \in [15,5; 20,5)$.

Ta xác định được $n = 240, n_m = 82, C = 53, u_m = 15,5; u_{m+1} = 20,5$.

Vậy trung vị của mẫu số liệu ghép nhóm là

$$M_e = 15,5 + \frac{\frac{240}{2} - 53}{82} \cdot (20,5 - 15,5) = \frac{803}{41} \approx 19,59$$

Ý nghĩa của trung vị của mẫu số liệu ghép nhóm

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.



Hãy trả lời câu hỏi ở .



Trong một hội thao, thời gian chạy 200 m của một nhóm các vận động viên được ghi lại ở bảng sau:

Thời gian (giây)	[21; 21,5)	[21,5; 22)	[22; 22,5)	[22,5; 23)	[23; 23,5)
Số vận động viên	5	12	32	45	30

Dựa vào bảng số liệu trên, ban tổ chức muốn chọn ra khoảng 50% số vận động viên chạy nhanh nhất để tiếp tục thi vòng 2. Ban tổ chức nên chọn các vận động viên có thời gian chạy không quá bao nhiêu giây?

2. Tứ phân vị



Thời gian luyện tập trong một ngày (tính theo giờ) của một số vận động viên được ghi lại ở bảng sau:

Thời gian luyện tập (giờ)	[0; 2)	[2; 4)	[4; 6)	[6; 8)	[8; 10)
Số vận động viên	3	8	12	12	4

Huấn luyện viên muốn xác định nhóm gồm 25% các vận động viên có số giờ luyện tập cao nhất. Hỏi huấn luyện viên nên chọn các vận động viên có thời gian luyện tập từ bao nhiêu giờ trở lên vào nhóm này?



Trong , số vận động viên được khảo sát là $n = 3 + 8 + 12 + 12 + 4 = 39$.

Gọi $x_1; x_2; \dots; x_{39}$ là thời gian luyện tập của 39 vận động viên được xếp theo thứ tự không giảm.

Ta có $x_1; x_2; x_3 \in [0; 2)$; $x_4; \dots; x_{11} \in [2; 4)$; $x_{12}; \dots; x_{23} \in [4; 6)$; $x_{24}; \dots; x_{35} \in [6; 8)$; $x_{36}; \dots; x_{39} \in [8; 10)$. Do đó đối với dãy số liệu $x_1; x_2; \dots; x_{39}$ thì

- Tứ phân vị thứ nhất là x_{10} thuộc nhóm $[2; 4)$;
- Tứ phân vị thứ hai là x_{20} thuộc nhóm $[4; 6)$;
- Tứ phân vị thứ ba là x_{30} thuộc nhóm $[6; 8)$;

Ta nói nhóm $[2; 4)$ là nhóm chứa tứ phân vị thứ nhất; nhóm $[4; 6)$ là nhóm chứa tứ phân vị thứ hai; nhóm $[6; 8)$ là nhóm chứa tứ phân vị thứ ba.

Công thức xác định tứ phân vị của mẫu số liệu ghép nhóm

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

Để tìm tứ phân vị thứ nhất của mẫu số liệu ghép nhóm, kí hiệu Q_1 , ta thực hiện như sau:

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

Khi đó

$$Q_1 = u_m + \frac{\frac{n}{4} - C}{n_m} \cdot (u_{m+1} - u_m)$$

Tương tự, để tìm tứ phân vị thứ ba của mẫu số liệu ghép nhóm, kí hiệu Q_3 , ta thực hiện như sau:

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

Khi đó

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

Ví dụ 3. Hãy xác định các tứ phân vị của mẫu số liệu trong Bài 2.

Lời giải

Nhắc lại: $x_1 \leq x_2 \leq \dots \leq x_{39}$ là thời gian luyện tập của 39 vận động viên.

Tứ phân vị thứ hai của dãy số liệu $x_1; x_2; \dots; x_{39}$ là $x_{20} \in [4; 6)$. Do đó tứ phân vị thứ hai của mẫu số liệu ghép nhóm là

$$Q_2 = 4 + \frac{\frac{2.39}{4} - (3+8)}{12} \cdot (6 - 4) = \frac{65}{12} \approx 5,417$$

Tứ phân vị thứ nhất của dãy số liệu $x_1; x_2; \dots; x_{39}$ là $x_{10} \in [2; 4)$. Do đó tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 2 + \frac{\frac{1.39}{4} - 3}{8} \cdot (4 - 2) = \frac{59}{16} = 3,6875$$

Tứ phân vị thứ ba của dãy số liệu $x_1; x_2; \dots; x_{39}$ là $x_{30} \in [6; 8)$. Do đó tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 6 + \frac{\frac{3.39}{4} - (3+8+12)}{12} \cdot (8 - 6) = \frac{169}{24} \approx 7,042$$

Chú ý: Nếu tứ phân vị thứ k là $\frac{1}{2}(x_m + x_{m+1})$, trong đó x_m và x_{m+1} thuộc hai nhóm liên tiếp, ví dụ như $x_m \in [u_{j-1}; u_j)$ và $x_{m+1} \in [u_j; u_{j+1})$ thì ta lấy $Q_k = u_j$.

Ví dụ 4. Một hãng ô tô thống kê lại số lần gặp sự cố về động cơ của 100 chiếc xe cùng loại sau 2 năm sử dụng đầu tiên ở bảng sau:

Số lần gặp sự cố	[1; 2]	[3; 4]	[5; 6]	[9; 10]	[9; 10]
Số xe	17	33	25	20	5

a) Hãy ước lượng các tứ phân vị của mẫu số liệu ghép số trên.

b) Một người cho rằng có trên 25% xe của hãng gặp không ít hơn 4 sự cố về động cơ trong 2 năm sử dụng đầu tiên. Nhận định trên có hợp lí không?

Lời giải

a) Do số lần gặp sự cố là số nguyên nên ta hiệu chỉnh lại như sau:

Số lần gặp sự cố	[0,5; 2,5)	[2,5; 4,5)	[4,5; 6,5)	[6,5; 8,5)	[8,5; 10,5)
Số xe	17	33	25	20	5

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu được xếp theo thứ tự không giảm.

Ta có $x_1, \dots, x_{17} \in [0,5; 2,5); x_{18}, \dots, x_{50} \in [2,5; 4,5); x_{51}, \dots, x_{75} \in [4,5; 6,5);$

$x_{76}, \dots, x_{95} \in [6,5; 8,5); x_{96}, \dots, x_{100} \in [8,5; 10,5)$.

Tứ phân vị thứ hai của dãy số liệu $x_1; x_2; \dots; x_{100}$ là $\frac{1}{2}(x_{50} + x_{51})$. Do $x_{50} \in [2,5; 4,5)$ và $x_{51} \in [4,5; 6,5)$ nên tứ phân vị thứ hai của mẫu số liệu ghép nhóm là $Q_2 = 4,5$.

Tứ phân vị thứ nhất của dãy số liệu $x_1; x_2; \dots; x_{100}$ là $\frac{1}{2}(x_{25} + x_{26})$. Do x_{25} và x_{26} thuộc nhóm $[2, 5; 4, 5)$ nên tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 2,5 + \frac{\frac{1.100}{4} - 17}{33} (4,5 - 2,5) = \frac{197}{66} \approx 2,98.$$

Tứ phân vị thứ ba của dãy số liệu $x_1; x_2; \dots; x_{100}$ là $\frac{1}{2}(x_{75} + x_{76})$. Do $x_{75} \in [4, 5; 6, 5)$ và $x_{76} \in [6, 5; 8, 5)$ nên tứ phân vị thứ ba của mẫu số liệu ghép nhóm là $Q_3 = 6,5$.

b) Do tứ phân vị thứ nhất $Q_1 \approx 2,98$ nên nhận định trên là hợp lí.

Ý nghĩa của tứ phân vị của mẫu số liệu ghép nhóm

Ba điểm tứ phân vị chia mẫu số liệu đã sắp xếp theo thứ tự không giảm thành bốn phần đều nhau. Giống như với trung vị, nói chung không thể xác định chính xác các điểm tứ phân vị của mẫu số liệu ghép nhóm.

Bộ ba tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho tứ phân vị của mẫu số liệu gốc và được sử dụng làm giá trị đo xu thế trung tâm của mẫu số liệu.

Tứ phân vị thứ nhất và thứ ba đo xu thế trung tâm của nửa dưới (các dữ liệu nhỏ hơn Q_2) và nửa trên (các dữ liệu lớn hơn Q_2) của mẫu số liệu.

- Một người thống kê lại thời gian thực hiện các cuộc gọi điện thoại của người đó trong một tuần ở bảng sau:

Thời gian (đơn vị : giây)	$[0; 60)$	$[60; 120)$	$[120; 180)$	$[180; 240)$	$[240; 300)$	$[300; 360)$
Số cuộc gọi	8	10	7	5	2	1

Hãy ước lượng các tứ phân vị của mẫu số liệu ghép nhóm trên.

- Một phòng khám thống kê số bệnh nhân đến khám bệnh mỗi ngày trong tháng 4 năm 2022 ở bảng sau:

Số bệnh nhân	$[1; 10]$	$[11; 20]$	$[21; 30]$	$[31; 40]$	$[41; 50]$
Số ngày	7	8	7	6	2

a) Hãy ước lượng các tứ phân vị của mẫu số liệu ghép nhóm trên.

b) Quản lí phòng khám cho rằng có khoảng 25% số ngày khám có nhiều hơn 35 bệnh nhân đến khám. Nhận định trên có hợp lí không?

BÀI TẬP

1. Lương tháng của một số nhân viên một văn phòng được ghi lại như sau (đơn vị: triệu đồng):

12,5	9,6	11,7	12,7	10,0	10,0	12,2	9,8	10,9	6,7	13,6	9,2
13,1	6,5	10,7	8,9	11,2	13,2	8,3	11,1	11,9	8,4	6,7	13,8

a) Tìm tứ phân vị của dãy số liệu trên.

b) Tổng hợp lại dãy số liệu trên vào bảng tần số ghép nhóm theo mẫu sau:

Lương tháng (triệu đồng)	$[6; 8)$	$[8; 10)$	$[10; 12)$	$[12; 14)$
Số nhân viên	?	?	?	?

c) Hãy ước lượng tứ phân vị của số liệu ở bảng tần số ghép trên.

2. Số điểm một cầu thủ bóng rổ ghi được trong 20 trận đấu được cho ở bảng sau:

25	23	21	13	8	14	15	18	22	11
24	12	14	14	18	6	8	25	10	11

a) Tìm tứ phân vị của dãy số liệu trên.

b) Tổng hợp lại dãy số liệu trên vào bảng tần số ghép nhóm theo mẫu sau:

Điểm số	[6;10]	[11;15]	[16;20]	[21;25]
Số trận	?	?	?	?

c) Hãy ước lượng tứ phân vị của số liệu ở bảng tần số ghép trên.

3. Kiểm tra điện lượng của một số viên pin tiêu do một hãng sản xuất thu được kết quả sau:

Điện lượng (nghìn mAh)	[0,9;0,95)	[0,95;1;0)	[1,0;1,05)	[1,05;1,1)	[1,1;1,15)
Số viên pin	10	20	35	15	5

Hãy ước lượng số trung bình, một và tứ phân vị của mẫu số liệu ghép nhóm trên.

4. Cân nặng của một số lợn con mới sinh thuộc hai giống A và B được cho ở biểu đồ dưới đây (đơn vị: kg).



- a) Hãy so sánh cân nặng của lợn con mới sinh giống A và giống B theo số trung bình và trung vị.
- b) Hãy ước lượng tứ phân vị thứ nhất và thứ ba của cân nặng lợn con mới sinh giống A và của cân nặng lợn con mới sinh giống B .

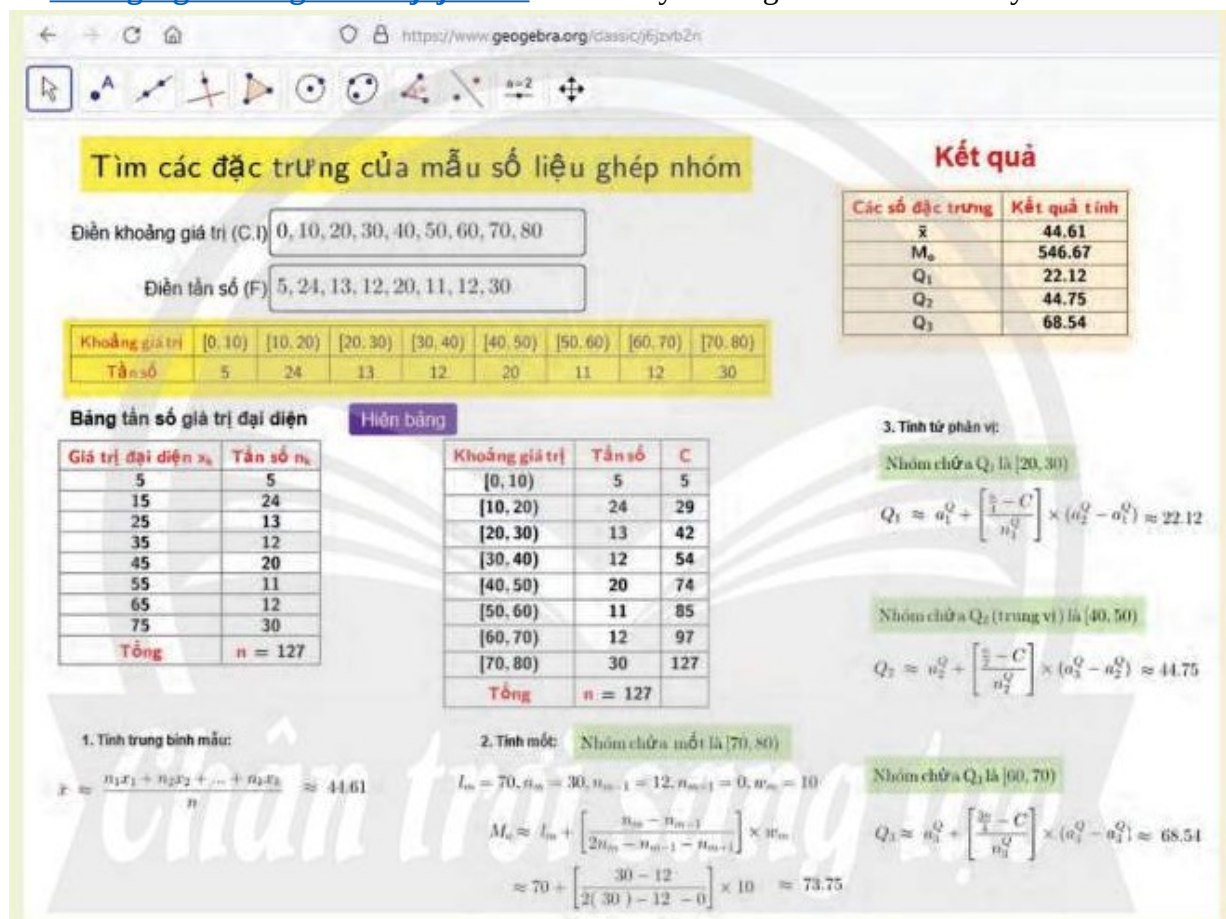
BÀI TẬP CUỐI CHƯƠNG V

Bạn có biết

Ở lớp 10 chúng ta đã được làm quen với phần mềm GeoGebra để vẽ đồ thị hàm số bậc hai và vẽ ba đường conic. Phần mềm GeoGebra còn có thể được dùng để tính các số đặc trưng của mẫu số liệu ghép nhóm.

Ví dụ, để tính các số đặc trưng của mẫu số liệu ghép nhóm, hãy vào địa chỉ:

<https://www.geogebra.org/classic/j6jzvb2n> em sẽ thấy chương trình như dưới đây.



Hướng dẫn sử dụng: Nhập các đầu mút của mỗi nhóm lần lượt từ nhỏ đến lớn vào ô “Điền khoảng giá trị (C.I)”; Nhập tần số tương ứng vào ô “Điền tần số (F)”. Sau khi nhập xong dữ liệu trên, ấn vào nút **Hiện bảng** để xem các số đặc trưng.

CÂU HỎI TRẮC NGHIỆM

Chọn phương án đúng.

Doanh thu bán hàng trong 20 ngày được lựa chọn ngẫu nhiên của một cửa hàng được ghi lại ở bảng sau (đơn vị: triệu đồng).

Doanh thu	[5; 7)	[7; 9)	[9; 11)	[11; 13)	[13; 15)
Số ngày	2	7	7	3	1

1. Số trung bình của mẫu số liệu trên thuộc khoảng nào trong các khoảng dưới đây?

- A. [7; 9) B. [9; 11) C. [11; 13) D. [13; 15)

2. Trung vị của mẫu số liệu trên thuộc khoảng nào trong các khoảng dưới đây?

- A. [7; 9) B. [9; 11) C. [11; 13) D. [13; 15)

3. Mốt của mẫu số liệu trên thuộc khoảng nào trong các khoảng dưới đây?

- A. $[7;9)$. B. $[9;11)$. C. $[11;13)$. D. $[13;15)$.
4. Tứ phân vị thứ nhất của mẫu số liệu trên gần nhất với giá trị nào trong các giá trị sau?
A. 7 . B. $7,6$. C. 8 . D. $8,6$.
5. Tứ phân vị thứ ba của mẫu số liệu trên gần nhất với giá trị nào trong các giá trị sau?
A. 10 . B. 11 . C. 12 . D. 13 .

BÀI TẬP TỰ LUẬN

6. Thống kê điểm trung bình môn Toán của một số học sinh lớp 11 được cho ở bảng sau:

Khoảng điểm	$[6,5;7)$	$[7;7,5)$	$[7,5;8)$	$[8;8,5)$	$[8,5;9)$	$[9;9,5)$	$[9,5;10)$
Tần số	8	10	16	24	13	7	4

Hãy ước lượng số trung bình, tứ phân vị và mốt của mẫu số liệu ghép nhóm trên.

7. Để kiểm tra thời gian sử dụng pin của chiếc điện thoại mới, chị An thống kê thời gian sử dụng điện thoại của mình từ lúc sạc đầy pin cho đến khi hết pin ở bảng sau:

Thời gian sử dụng (giờ)	$[7;9)$	$[9;11)$	$[11;13)$	$[13;15)$	$[15;17)$
Số lần	2	5	7	6	3

- a. Hãy ước lượng thời gian sử dụng trung bình từ lúc chị An sạc đầy pin điện thoại cho tới khi hết pin.
b. Chị An cho rằng có khoảng 25% số lần sạc điện thoại chỉ dùng được dưới 10 giờ. Nhận định của chị An có hợp lí không?
8. Tổng lượng mưa trong tháng 8 đo được tại một trạm quan trắc đặt tại Vũng Tàu từ năm 2002 đến năm 2020 được ghi lại như dưới đây (đơn vị: mm):

121,8 158,3 334,9 200,9 165,6 161,5 194,3 220,7 189,8 234,2
165,9 165,9 134 173 169 189 254 168 255

(Nguồn: Tổng cục Thống kê)

- a. Xác định số trung bình, tứ phân vị và mốt của mẫu số liệu trên.
b. Hoàn thiện bảng tần số ghép nhóm theo mẫu sau:

Tổng lượng mưa trong tháng 8 (mm)	$[120;175)$	$[175;230)$	$[230;285)$	$[285;340)$
Số năm	?	?	?	?

c. Hãy ước lượng số trung bình, tứ phân vị và mốt của mẫu số liệu ở bảng tần số ghép nhóm trên.

9. Bảng sau thống kê số ca nhiễm mới SARS-CoV-2 mỗi ngày trong tháng 12/2021 tại Việt Nam.

Ngày	Số ca	Ngày	Số ca	Ngày	Số ca	Ngày	Số ca
1	15 139	9	15 965	17	15 685	25	16 046
2	14 295	10	15 474	18	16 363	26	15 667
3	14 254	11	16 830	19	16 586	27	15 310
4	14 598	12	15 264	20	15 420	28	14 866
5	14 927	13	16 035	21	16 806	29	14 299
6	15 215	14	15 871	22	17 044	30	20 454
7	14 433	15	16 192	23	16 860	31	17 004
8	15 223	16	15 720	24	16 633		

(Nguồn: worldometers.info)

- a. Xác định số trung bình và tứ phân vị của mẫu số liệu trên. Mẫu số liệu có bao nhiêu giá trị ngoại lệ?
b. Hoàn thiện bảng tần số ghép nhóm theo mẫu sau:

Số ca (nghìn)	$[14; 15,5)$	$[15,5; 17)$	$[17; 18,5)$	$[18,5; 20)$	$[20; 21,5)$
Số ngày	?	?	?	?	?

- c. Hãy ước lượng số trung bình và tứ phân vị của mẫu số liệu ở bảng tần số ghép nhóm trên.