

CHƯƠNG 3**CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHEP NHÓM****BÀI 1****KHOẢNG BIẾN THIÊN VÀ KHOẢNG TƯ PHÂN VỊ****1. Khoảng biến thiên**

Cho mẫu số liệu ghép nhóm:

Nhóm	$[a_1; a_2)$	...	$[a_p; a_{p+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

Trong đó các tần số $m_1 > 0, m_k > 0$ và $n = m_1 + \dots + m_k$ là cỡ mẫu.

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là $R = a_{k+1} - a_1$

Ý nghĩa

Khoảng biến thiên của mẫu số liệu ghép nhóm xấp xỉ cho khoảng biến thiên của mẫu số liệu gốc.

Khoảng biến thiên được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.

Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

2. Khoảng tứ phân vị

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[a_1; a_2)$	...	$[a_p; a_{p+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

Khoảng tứ phân vị của mẫu số liệu ghép nhóm, kí hiệu Δ_Q , là hiệu giữa tứ phân vị thứ ba Q_3 và tứ phân vị thứ nhất Q_1 của mẫu số liệu ghép nhóm đó, tức là: $\Delta_Q = Q_3 - Q_1$

Chú ý:

Tứ phân vị thứ r là: $Q_r = a_p + \frac{\frac{rn}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$

Trong đó: $[a_p; a_{p+1})$ là nhóm chứa tứ phân vị thứ $r = 1, 2, 3$.

Ý nghĩa

Khoảng tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ khoảng tứ phân vị của mẫu số liệu gốc.

Khoảng tứ phân vị cũng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.

Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

Nhận xét: Do khoảng tứ phân vị của mẫu số liệu ghép nhóm chỉ phụ thuộc vào nửa giữa của mẫu số liệu, nên không bị ảnh hưởng bởi các giá trị bất thường và có thể dùng đại lượng này để loại giá trị bất thường.

Câu 1. Bảng sau thống kê cân nặng của 50 quả xoài được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường

Cân nặng (g)	[250; 290)	[290; 330)	[330; 370)	[370; 410)	[410; 450)
Số quả xoài	3	13	18	11	5

Có ý kiến cho rằng: “Trong 50 quả xoài trên, hiệu số cân nặng của hai quả bất kì không vượt quá 200 g”. Ý kiến đó đúng hay sai? Giải thích.

Câu 2. Kết quả điều tra tổng thu nhập trong năm 2024 của một số hộ gia đình ở thành phố Nha Trang được ghi lại ở bảng sau:

Tổng thu nhập (triệu đồng)	[200; 250)	[250; 300)	[300; 350)	[350; 400)	[400; 450)
Số hộ gia đình	24	62	34	21	9

- a) Hãy tìm các tứ phân vị Q_1 và Q_3 .
- b) Một doanh nghiệp địa phương muốn hướng dịch vụ của mình đến các gia đình có mức thu nhập ở tầm trung, tức là 50% các hộ gia đình có mức thu nhập ở chính giữa so với mức thu nhập của tất cả các hộ gia đình của địa phương. Hỏi doanh nghiệp cần hướng đến các gia đình có mức thu nhập trong khoảng nào?

Câu 3. Kết quả đo chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được cho ở bảng sau:

Chiều cao (m)	[8,4; 8,6)	[8,6; 8,8)	[8,8; 9,0)	[9,0; 9,2)	[9,2; 9,4)
Số cây	5	12	25	44	14

- a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.
- b) Trong 100 cây keo trên có 1 cây cao 8,4 m. Hỏi chiều cao của cây keo này có phải là giá trị ngoại lệ không?

Câu 4. Bạn Trang thống kê chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau:

Chiều cao (cm)	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
Số học sinh nữ lớp 12C	2	7	12	3	0	1
Số học sinh nữ lớp 12D	5	9	8	2	1	0

- a) Sử dụng khoảng biến thiên, hãy cho biết chiều cao của học sinh nữ lớp nào có độ phân tán lớn hơn.

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của học sinh nữ lớp 12C và 12D.

Câu 5. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

Tuổi kết hôn	[19; 22)	[22; 25)	[25; 28)	[28; 31)	[31; 34)
Số phụ nữ khu vực A	10	27	31	25	7
Số phụ nữ khu vực B	47	40	11	2	0

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.

b) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?

Câu 6. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

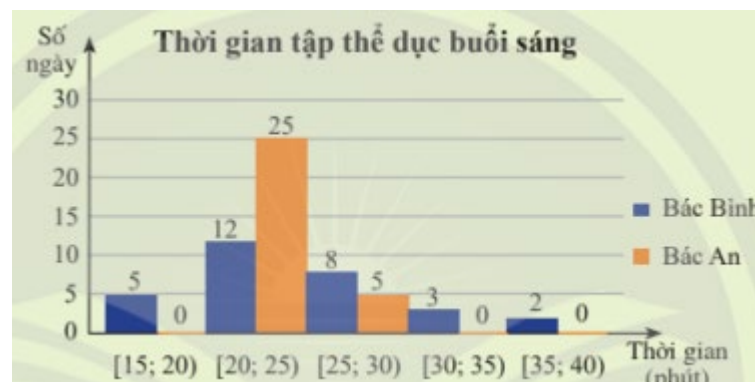
341,4	187,1	242,2	522,9	251,4	432,2	200,7	388,6	258,4	288,5
298,1	413,5	413,5	332	421	475	400	305	520	147

a) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu trên.

b) Hãy chia mẫu số liệu trên thành 4 nhóm với nhóm đầu tiên là [140; 240) và lập bảng tần số ghép nhóm.

c) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu ghép nhóm và so sánh với kết quả tương ứng thu được ở câu a).

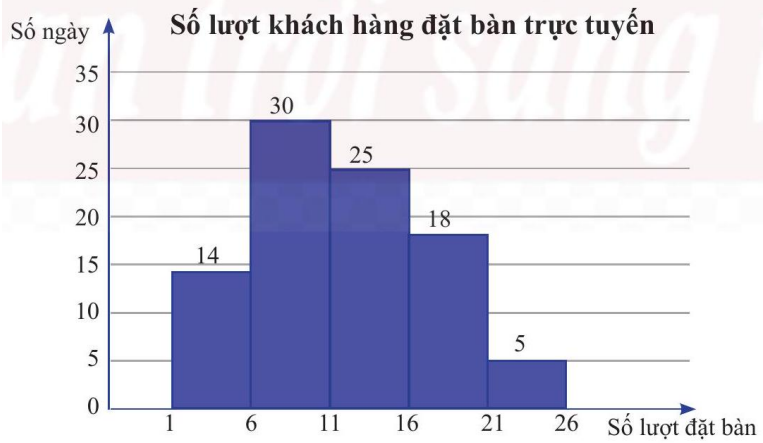
Câu 7. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng của Bác Bình và Bác An



a) Ai là người có thời gian tập đều hơn?

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

Câu 8. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2024 của một nhà hàng. Cột thứ nhất biểu diễn số ngày có từ 1 đến dưới 6 lượt đặt bàn; cột thứ hai biểu diễn số ngày có từ 6 đến dưới 11 lượt đặt bàn; ...



Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi biểu đồ trên.

Câu 9. Hai bảng tần số ghép nhóm dưới đây thống kê theo độ tuổi số lượng thành viên nam và thành viên nữ đang sinh hoạt trong một câu lạc bộ dưỡng sinh.

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nam	4	7	4	6	15	12	2	0

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nữ	3	4	5	3	7	14	13	1

- a) Hãy tính các khoảng tứ phân vị của tuổi nam giới và nữ giới trong mỗi bảng số liệu ghép nhóm trên.
- b) Hãy cho biết trong câu lạc bộ trên, nam giới hay nữ giới có tuổi đồng đều hơn.

CHƯƠNG 3**CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM****BÀI 1****KHOẢNG BIẾN THIÊN VÀ KHOẢNG TƯ PHÂN VỊ****1. Khoảng biến thiên**

Cho mẫu số liệu ghép nhóm:

Nhóm	$[a_1; a_2)$	...	$[a_p; a_{p+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

Trong đó các tần số $m_1 > 0, m_k > 0$ và $n = m_1 + \dots + m_k$ là cỡ mẫu.

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là $R = a_{k+1} - a_1$

Ý nghĩa

Khoảng biến thiên của mẫu số liệu ghép nhóm xấp xỉ cho khoảng biến thiên của mẫu số liệu gốc.

Khoảng biến thiên được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.

Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

2. Khoảng tứ phân vị

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[a_1; a_2)$	...	$[a_p; a_{p+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

Khoảng tứ phân vị của mẫu số liệu ghép nhóm, kí hiệu Δ_Q , là hiệu giữa tứ phân vị thứ ba Q_3 và tứ phân vị thứ nhất Q_1 của mẫu số liệu ghép nhóm đó, tức là: $\Delta_Q = Q_3 - Q_1$

Chú ý:

Tứ phân vị thứ r là: $Q_r = a_p + \frac{\frac{rn}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$

Trong đó: $[a_p; a_{p+1})$ là nhóm chứa tứ phân vị thứ $r = 1, 2, 3$.

Ý nghĩa

Khoảng tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ khoảng tứ phân vị của mẫu số liệu gốc.

Khoảng tứ phân vị cũng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.

Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

Nhận xét: Do khoảng tứ phân vị của mẫu số liệu ghép nhóm chỉ phụ thuộc vào nửa giữa của mẫu số liệu, nên không bị ảnh hưởng bởi các giá trị bất thường và có thể dùng đại lượng này để loại giá trị bất thường.

Câu 1. Bảng sau thống kê cân nặng của 50 quả xoài được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường

Cân nặng (g)	[250; 290)	[290; 330)	[330; 370)	[370; 410)	[410; 450)
Số quả xoài	3	13	18	11	5

Có ý kiến cho rằng: “Trong 50 quả xoài trên, hiệu số cân nặng của hai quả bất kì không vượt quá 200 g”. Ý kiến đó đúng hay sai? Giải thích.

Lời giải

Ý kiến nêu trên là đúng.

Giải thích: Quan sát bảng thống kê đã cho, ta thấy cân nặng lớn nhất quả xoài có thể đạt được là dưới 450 g, cân nặng nhỏ nhất quả xoài có thể đạt được là 250 g. Mà ta có $450 - 350 = 200$. Do đó, hai quả bất kì nào cũng có hiệu số cân nặng không vượt quá 200 g.

Câu 2. Kết quả điều tra tổng thu nhập trong năm 2024 của một số hộ gia đình ở thành phố Nha Trang được ghi lại ở bảng sau:

Tổng thu nhập (triệu đồng)	[200; 250)	[250; 300)	[300; 350)	[350; 400)	[400; 450)
Số hộ gia đình	24	62	34	21	9

- a) Hãy tìm các tứ phân vị Q_1 và Q_3 .
- b) Một doanh nghiệp địa phương muốn hướng dịch vụ của mình đến các gia đình có mức thu nhập ở tầm trung, tức là 50% các hộ gia đình có mức thu nhập ở chính giữa so với mức thu nhập của tất cả các hộ gia đình của địa phương. Hỏi doanh nghiệp cần hướng đến các gia đình có mức thu nhập trong khoảng nào?

Lời giải

a) Số hộ gia đình được khảo sát (cỡ mẫu) là $n = 24 + 62 + 34 + 21 + 9 = 150$.

Gọi $x_1; x_2; \dots; x_{150}$ là tổng thu nhập trong năm 2024 của 150 hộ gia đình được xếp theo thứ tự không giảm.

Ta có:

$$x_1; \dots; x_{24} \in [200; 250)$$

$$x_{25}; \dots; x_{86} \in [300; 350)$$

$$x_{87}; \dots; x_{120} \in [300; 350)$$

$$x_{121}; \dots; x_{141} \in [350; 400)$$

$$x_{142}; \dots; x_{150} \in [400; 450)$$

Do đó, đối với dãy số liệu $x_1; x_2; \dots; x_{150}$ thì

Tứ phân vị thứ nhất Q_1 là $x_{38} \in [250; 300)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 250 + \frac{\frac{150}{4} - 24}{62} (300 - 250) = \frac{16175}{62}$$

Tứ phân vị thứ ba Q_3 là $x_{113} \in [300; 350)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 300 + \frac{\frac{3.150}{4} - (24 + 62)}{34} (350 - 300) = \frac{11525}{34}$$

b) Doanh nghiệp cần hướng đến các gia đình có mức thu nhập trong khoảng:

$$[Q_1; Q_3] = \left[\frac{16175}{62}; \frac{11525}{34} \right] = [260,89; 338,97] \text{ (triệu đồng)}.$$

Câu 3. Kết quả đo chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được cho ở bảng sau:

Chiều cao (m)	[8,4; 8,6)	[8,6; 8,8)	[8,8; 9,0)	[9,0; 9,2)	[9,2; 9,4)
Số cây	5	12	25	44	14

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

b) Trong 100 cây keo trên có 1 cây cao 8,4 m. Hỏi chiều cao của cây keo này có phải là giá trị ngoại lệ không?

Lời giải

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là:

$$R = 9,4 - 8,4 = 1 \text{ (m)}.$$

Cỡ mẫu $n = 100$.

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc về chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_5 \in [8,4; 8,6),$$

$$x_6; \dots; x_{17} \in [8,6; 8,8),$$

$$x_{18}; \dots; x_{42} \in [8,8; 9,0),$$

$$x_{43}; \dots; x_{86} \in [9,0; 9,2),$$

$$x_{87}; \dots; x_{100} \in [9,2; 9,4).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{25} + x_{26}}{2} \in [8,8; 9,0)$. Do đó, tứ phân vị thứ nhất của mẫu số

liệu ghép nhóm là:

$$Q_1 = 8,8 + \frac{\frac{100}{4} - (5 + 12)}{25} (9,0 - 8,8) = 8,864$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{75} + x_{76}}{2} \in [9,0; 9,2)$.

Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 20 + \frac{\frac{3 \cdot 100}{4} - (5 + 12 + 25)}{44} (9,2 - 9,0) = 9,15$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là:

$$\Delta_Q = Q_3 - Q_1 = 9,15 - 8,864 = 0,286$$

b) Trong 100 cây keo trên có 1 cây cao 8,4 m thuộc nhóm $[8,4; 8,6)$.

Vì $Q_1 - 1,5\Delta_Q = 8,864 - 1,5 \cdot 0,286 = 8,435 > 8,4$ nên chiều cao của cây keo cao 8,4 m là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Câu 4. Bạn Trang thống kê chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau:

Chiều cao (cm)	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
Số học sinh nữ lớp 12C	2	7	12	3	0	1
Số học sinh nữ lớp 12D	5	9	8	2	1	0

a) Sử dụng khoảng biến thiên, hãy cho biết chiều cao của học sinh nữ lớp nào có độ phân tán lớn hơn.

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của học sinh nữ lớp 12C và 12D.

Lời giải

a) Khoảng biến thiên của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12C là: $185 - 155 = 30$ (cm).

Trong mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D, khoảng đầu tiên chứa dữ liệu là $[155; 160)$ và khoảng cuối cùng chứa dữ liệu là $[175; 180)$.

Khoảng biến thiên của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D là: $180 - 155 = 25$ (cm).

Vậy nếu căn cứ theo khoảng biến thiên thì chiều cao của học sinh nữ lớp 12C có độ phân tán lớn hơn lớp 12D.

b)

• Lớp 12C:

Cỡ mẫu $n = 2 + 7 + 12 + 3 + 0 + 1 = 25$.

Gọi $x_1; x_2; \dots; x_{25}$ là mẫu số liệu gốc về chiều cao của 25 học sinh nữ lớp 12C được xếp theo thứ tự không giảm.

Ta có:

$$x_1; x_2 \in [155; 160)$$

$$x_3; \dots; x_9 \in [160; 165)$$

$$x_{10}; \dots; x_{21} \in [165; 170)$$

$$x_{22}; x_{23}; x_{24} \in [170; 175)$$

$$x_{25} \in [180; 185)$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_6 + x_7}{2} \in [160; 165)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 160 + \frac{\frac{25}{4} - 2}{7} (165 - 160) = \frac{4565}{28}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{19} + x_{20}}{2} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 165 + \frac{\frac{3 \cdot 25}{4} - (2 + 7)}{12} (170 - 165) = \frac{2705}{16}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12C là:

$$\Delta_Q = Q_3 - Q_1 = \frac{2705}{16} - \frac{4565}{28} \approx 6,03$$

• Lớp 12D:

ỡ mẫu $n' = 5 + 9 + 8 + 2 + 1 = 25$.

Gọi $y_1; y_2; \dots; y_{25}$ là mẫu số liệu gốc về chiều cao của 25 học sinh nữ lớp 12D được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_5 \in [155; 160),$$

$$y_6; \dots; y_{14} \in [160; 165),$$

$$y_{15}; \dots; y_{22} \in [165; 170),$$

$$y_{23}; y_{24} \in [170; 175),$$

$$y_{25} \in [175; 180).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{y_6 + y_7}{2} \in [160; 165)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q'_1 = 160 + \frac{\frac{25}{4} - 5}{9} (165 - 160) = \frac{5785}{36}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{y_{19} + y_{20}}{2} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q'_3 = 165 + \frac{\frac{3.25}{4} - (5+9)}{8} (170 - 165) = \frac{5395}{32}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D là:

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{5375}{32} - \frac{5785}{36} \approx 7,27$$

Ta có $\Delta_Q \approx 6,03 < \Delta'_Q \approx 7,27$

Câu 5. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

Tuổi kết hôn	[19; 22)	[22; 25)	[25; 28)	[28; 31)	[31; 34)
Số phụ nữ khu vực A	10	27	31	25	7
Số phụ nữ khu vực B	47	40	11	2	0

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.

b) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?

Lời giải

a)

• Khu vực A:

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là:

$$R = 34 - 19 = 15.$$

Cỡ mẫu $n = 10 + 27 + 31 + 25 + 7 = 100$.

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực A được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_{10} \in [19; 22),$$

$$x_{11}; \dots; x_{37} \in [22; 25),$$

$$x_{38}; \dots; x_{68} \in [25; 28),$$

$$x_{69}; \dots; x_{93} \in [28; 31),$$

$$x_{94}; \dots; x_{100} \in [31; 34).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{25} + x_{26}}{2} \in [22; 25)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu

ghép nhóm là:

$$Q_1 = 22 + \frac{\frac{100}{4} - 10}{27} (25 - 22) = \frac{71}{3}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{75} + x_{76}}{2} \in [28; 31)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 28 + \frac{\frac{3.100}{4} - (10 + 27 + 31)}{25} (31 - 28) = \frac{721}{25}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực A là:

$$\Delta_Q = Q_3 - Q_1 = \frac{721}{25} - \frac{71}{3} \approx 5,17$$

• Khu vực B:

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là:

$$R' = 31 - 19 = 12.$$

Cỡ mẫu $n' = 47 + 40 + 11 + 2 = 100$.

Gọi $y_1; y_2; \dots; y_{100}$ là mẫu số liệu gốc về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực B được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_{47} \in [19; 22),$$

$$y_{48}; \dots; y_{87} \in [22; 25),$$

$$y_{88}; \dots; y_{98} \in [25; 28),$$

$$y_{99}; y_{100} \in [28; 31).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{y_{25} + y_{26}}{2} \in [19; 22)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q'_1 = 19 + \frac{\frac{100}{4} - 47}{47} (22 - 19) = \frac{968}{47}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{y_{75} + y_{76}}{2} \in [22; 25)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q'_3 = 22 + \frac{\frac{3.100}{4} - 47}{40} (25 - 22) = \frac{241}{10}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực B là:

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{241}{10} - \frac{968}{47} \approx 3,5$$

Vì $\Delta_Q \approx 5,17 > \Delta'_Q \approx 3,5$ nên phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn.

Câu 6. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

341,4	187,1	242,2	522,9	251,4	432,2	200,7	388,6	258,4	288,5
298,1	413,5	413,5	332	421	475	400	305	520	147

- a) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu trên.
- b) Hãy chia mẫu số liệu trên thành 4 nhóm với nhóm đầu tiên là $[140; 240)$ và lập bảng tần số ghép nhóm.
- c) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu ghép nhóm và so sánh với kết quả tương ứng thu được ở câu a).

Lời giải

a) Sắp xếp lại mẫu số liệu trên theo thứ tự không giảm, ta được:

147	187,1	200,7	242,2	251,4	258,4	288,5
298,1	305	332	341,4	388,6	400	413,5
413,5	421	432,2	475	520	522,9	

Khoảng biến thiên của mẫu số liệu trên là:

$$R = 522,9 - 147 = 375,9 \text{ (mm)}.$$

Cỡ mẫu $n = 20$.

Tứ phân vị thứ nhất là trung vị của mẫu số liệu:

$$147; 187,1; 200,7; 242,2; 251,4; 258,4; 288,5; 298,1; 305; 332.$$

$$\text{Do đó, } Q_1 = \frac{251,4 + 258,4}{2} = 254,9$$

Tứ phân vị thứ ba là trung vị của mẫu số liệu:

$$341,4; 388,6; 400; 413,5; 413,5; 421; 432,2; 475; 520; 522,9.$$

$$\text{Do đó, } Q_3 = \frac{413,5 + 421}{2} = 417,25$$

Khoảng tứ phân vị của mẫu số liệu đã cho là:

$$\Delta_Q = Q_3 - Q_1 = 417,25 - 254,9 = 162,35 \quad \Delta_Q = Q_3 - Q_1 = 417,25 - 254,9 = 162,35.$$

b) Nhóm đầu tiên là $[140; 240)$, ta chọn 3 nhóm còn lại là

$$[240; 340), [340; 440), [440; 540).$$

Từ bảng thống kê ban đầu, ta lập được bảng tần số ghép nhóm như sau:

Lượng mưa (mm)	$[140; 240)$	$[240; 340)$	$[340; 440)$	$[440; 540)$
Số tháng	3	7	7	3

c) Cỡ mẫu $n = 20$.

Khoảng biến thiên của mẫu số liệu ghép nhóm là

$$R' = 540 - 140 = 400 \text{ (mm)}.$$

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về lượng mưa đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau được xếp theo thứ tự không giảm.

Ta có:

$$x_1; \dots; x_3 \in [140; 240),$$

$$x_4; \dots; x_{10} \in [240; 340),$$

$$x_{11}; \dots; x_{17} \in [340; 440),$$

$$x_{18}; \dots; x_{20} \in [440; 540).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [240; 340)$.

Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q'_1 = 240 + \frac{\frac{20}{4} - 3}{7} (340 - 240) = \frac{1880}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [340; 440)$.

Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

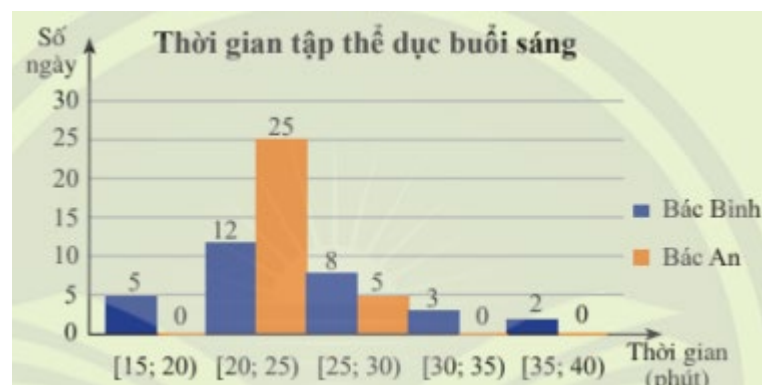
$$Q'_3 = 340 + \frac{\frac{3 \cdot 20}{4} - (3 + 7)}{7} (440 - 340) = \frac{2880}{7}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là:

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{2880}{7} - \frac{1880}{7} = 142,86$$

Ta thấy khoảng biến thiên của mẫu số liệu ghép nhóm lớn hơn mẫu số liệu đã cho; khoảng tứ phân vị của mẫu số liệu ghép nhóm nhỏ hơn mẫu số liệu đã cho.

Câu 7. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng của Bác Bình và Bác An



a) Ai là người có thời gian tập đều hơn?

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

Lời giải

Ta có bảng thống kê sau:

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Số ngày tập của Bác Bình	5	12	8	3	2
Số ngày tập của Bác An	0	25	5	0	0

a)

Khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác Bình là $40 - 15 = 25$ (phút).

Tuy nhiên, trong mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An, khoảng đầu tiên chứa dữ liệu là $[20; 25)$ và khoảng cuối cùng chứa dữ liệu là $[25; 30)$.

Do đó, khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An là $30 - 20 = 10$ (phút).

Nếu căn cứ theo khoảng biến thiên thì bác Bình có thời gian tập thể dục phân tán hơn bác An, vậy bác An là người có thời gian tập đều hơn.

b) Cỡ mẫu $n = 30$.

- Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình:

Gọi $x_1; x_2; \dots; x_{30}$ là mẫu số liệu gốc về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_5 \in [15; 20),$$

$$x_6; \dots; x_{17} \in [20; 25),$$

$$x_{18}; \dots; x_{25} \in [25; 30),$$

$$x_{26}; \dots; x_{28} \in [30; 35),$$

$$x_{29}; x_{30} \in [35; 40).$$

Tứ phân vị thứ nhất Q_1 của mẫu số liệu gốc là $x_8 \in [20; 25)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 20 + \frac{\frac{30}{4} - 5}{12} (25 - 20) = \frac{505}{24}$$

Tứ phân vị thứ ba Q_3 của mẫu số liệu gốc là $x_{23} \in [25; 30)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 25 + \frac{\frac{3.30}{4} - (5+12)}{8} (30-25) = \frac{455}{16}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình là

$$\Delta_Q = Q_3 - Q_1 = \frac{445}{16} - \frac{505}{24} = \frac{355}{48} \approx 7,4 \square$$

• Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An:

Gọi $y_1; y_2; \dots; y_{30}$ là mẫu số liệu gốc về thời gian tập thể dục buổi sáng mỗi ngày của bác An được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_{25} \in [20; 25),$$

$$y_{26}; \dots; y_{30} \in [25; 30).$$

Tứ phân vị thứ nhất Q'_1 của mẫu số liệu gốc là $y_8 \in [20; 25)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q'_1 = 20 + \frac{\frac{30}{4} - (25-20)}{25} (25-20) = \frac{43}{2}$$

Tứ phân vị thứ ba Q'_3 của mẫu số liệu gốc là $y_{23} \in [20; 25)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

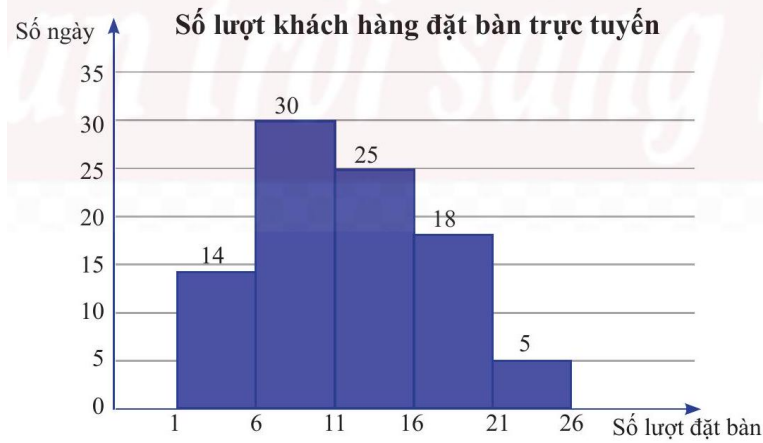
$$Q'_3 = 20 + \frac{\frac{3.30}{4} - (25-20)}{25} (25-20) = \frac{49}{2}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An là

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{49}{2} - \frac{43}{2} = 3$$

$\square$ Vì $\Delta_Q \approx 7,4 > \Delta'_Q = 3$ nên khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình lớn hơn bác An.

Câu 8. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2024 của một nhà hàng. Cột thứ nhất biểu diễn số ngày có từ 1 đến dưới 6 lượt đặt bàn; cột thứ hai biểu diễn số ngày có từ 6 đến dưới 11 lượt đặt bàn; ...



Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi biểu đồ trên.

Lời giải

Từ biểu đồ đã cho, ta có có bảng thống kê sau:

Số lượt đặt bàn	[1; 6)	[6; 11)	[11; 16)	[16; 21)	[21; 26)
Số ngày	14	30	25	18	5

Cỡ mẫu $n = 14 + 30 + 25 + 18 + 5 = 92$.

Gọi $x_1; x_2; \dots; x_{92}$ là mẫu số liệu gốc về số lượt khách đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 của một nhà hàng được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_{14} \in [1; 6),$$

$$x_{15}; \dots; x_{44} \in [6; 11),$$

$$x_{45}; \dots; x_{69} \in [11; 16),$$

$$x_{70}; \dots; x_{87} \in [16; 21),$$

$$x_{88}; \dots; x_{92} \in [21; 26).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{23} + x_{24}}{2} \in [6; 11)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu

ghép nhóm là:

$$Q_1 = 6 + \frac{\frac{92}{4} - 14}{30} (11 - 6) = 7,5$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{69} + x_{70}}{2}$.

Mà $x_{69} \in [11; 16)$ và $x_{70} \in [16; 21)$

Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là $Q_3 = 16$.

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là:

$$\Delta_Q = Q_3 - Q_1 = 16 - 7,5 = 8,5$$

Câu 9. Hai bảng tần số ghép nhóm dưới đây thống kê theo độ tuổi số lượng thành viên nam và thành viên nữ đang sinh hoạt trong một câu lạc bộ dưỡng sinh.

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nam	4	7	4	6	15	12	2	0

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nữ	3	4	5	3	7	14	13	1

a) Hãy tính các khoảng tứ phân vị của tuổi nam giới và nữ giới trong mỗi bảng số liệu ghép nhóm trên.

b) Hãy cho biết trong câu lạc bộ trên, nam giới hay nữ giới có tuổi đồng đều hơn.

Lời giải

a)

• Nam giới:

Cỡ mẫu $n = 4 + 7 + 4 + 6 + 15 + 12 + 2 = 50$.

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về tuổi của nam giới đang sinh hoạt trong câu lạc bộ đường sinh được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_4 \in [50; 55),$$

$$x_5; \dots; x_{11} \in [55; 60),$$

$$x_{12}; \dots; x_{15} \in [60; 65),$$

$$x_{16}; \dots; x_{21} \in [65; 70),$$

$$x_{22}; \dots; x_{36} \in [70; 75),$$

$$x_{37}; \dots; x_{48} \in [75; 80),$$

$$x_{49}; x_{50} \in [80; 85).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_{13} \in [60; 65)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 60 + \frac{\frac{50}{4} - (4 + 7)}{4} (65 - 60) = 61,875$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{38} \in [75; 80)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 75 + \frac{\frac{3 \cdot 50}{4} - (4 + 7 + 4 + 6 + 15)}{12} (80 - 75) = 75,625.$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về tuổi của nam giới đang sinh hoạt trong câu lạc bộ đường sinh là:

$$\Delta_Q = Q_3 - Q_1 = 75,625 - 61,875 = 13,75$$

• Nữ giới:

$$\text{Cỡ mẫu } n' = 3 + 4 + 5 + 3 + 7 + 14 + 13 + 1 = 50.$$

Gọi $y_1; \dots; y_{50}$ là mẫu số liệu gốc về tuổi của nữ giới đang sinh hoạt trong câu lạc bộ dưỡng sinh được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_4 \in [50; 55),$$

$$y_4; \dots; y_7 \in [55; 60),$$

$$y_8; \dots; y_{12} \in [60; 65),$$

$$y_{13}; \dots; y_{15} \in [65; 70),$$

$$y_{16}; \dots; y_{22} \in [70; 75),$$

$$y_{23}; \dots; y_{36} \in [75; 80),$$

$$y_{37}; \dots; y_{49} \in [80; 85),$$

$$y_{50} \in [85; 90).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $y_{13} \in [65; 70)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q'_1 = 65 + \frac{\frac{50}{4} - (3 + 4 + 5)}{3} (70 - 65) = \frac{395}{6}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $y_{38} \in [80; 85)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q'_3 = 80 + \frac{\frac{3 \cdot 50}{4} - (3 + 4 + 5 + 3 + 7 + 14)}{13} (85 - 80) = \frac{2095}{26}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về tuổi của nữ giới đang sinh hoạt trong câu lạc bộ dưỡng sinh là:

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{2095}{26} - \frac{395}{6} \approx 14,74$$

b) Ta có $\Delta'_Q \approx 14,74 > \Delta_Q = 13,75$ nên trong câu lạc bộ dưỡng sinh, nam giới có tuổi đồng đều hơn.

BÀI 2**PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN****1. Phương sai và độ lệch chuẩn**

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

- **Phương sai** của mẫu số liệu ghép nhóm, kí hiệu S^2 , được tính theo công thức sau:

$$S^2 = \frac{1}{n} \left[m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2 \right]$$

Trong đó:

$$n = m_1 + m_2 + \dots + m_k; \quad x_i = \frac{a_i + a_{i+1}}{2} \text{ với } i = 1, 2, \dots, k \text{ là giá trị đại diện cho nhóm } [a_i; a_{i+1}).$$

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_k x_k}{n} \text{ là số trung bình của mẫu số liệu ghép nhóm.}$$

- **Độ lệch chuẩn** của mẫu số liệu ghép nhóm, kí hiệu S , là căn bậc hai số học của phương sai, nghĩa là $S = \sqrt{S^2}$.

Chú ý:

- Phương sai của mẫu số liệu ghép nhóm có thể được tính theo công thức sau:

$$S^2 = \frac{1}{n} \left[m_1 x_1^2 + m_2 x_2^2 + \dots + m_k x_k^2 \right] - \bar{x}^2$$

- Trong thống kê, người ta còn dùng đại lượng sau để đo mức độ phân tán của mẫu số liệu ghép nhóm

$$\hat{s}^2 = \frac{1}{n-1} \left[m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2 \right]; \quad \hat{s} = \sqrt{\hat{s}^2}$$

Ý nghĩa

- Phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm xấp xỉ phương sai (độ lệch chuẩn) của mẫu số liệu gốc. Chúng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm xung quanh số trung bình của mẫu số liệu.

- Phương sai, độ lệch chuẩn càng lớn thì mẫu số liệu càng phân tán.

2. Sử dụng phương sai, độ lệch chuẩn đo độ rủi ro

Câu 1. Bảng dưới đây thống kê cự li ném tạ của một vận động viên.

Cự li (m)	[19; 19,5)	[19,5; 20)	[20; 20,5)	[20,5; 21)	[21; 21,5)
Tần số	13	45	24	12	6

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

Câu 2. Minh Hiền và Minh Nhân cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

Số bước (đơn vị: nghìn)	[3; 5)	[5; 7)	[7; 9)	[9; 11)	[11; 13)
Số ngày của Minh Hiền	6	7	6	6	5
Số ngày của Minh Nhân	2	5	13	8	2

a) Hãy tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

b) Nếu so sánh theo độ lệch chuẩn thì bạn nào có số lượng bước chân đi mỗi ngày đều đặn hơn?

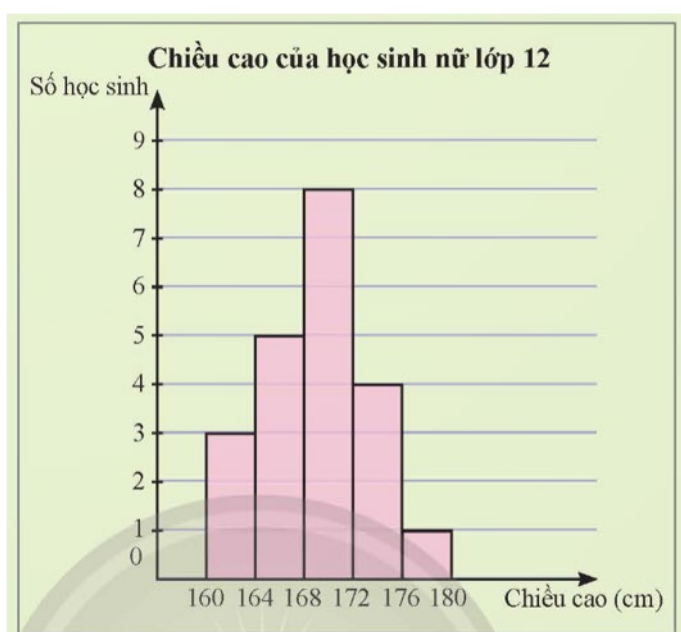
Câu 3. Một giống cây xoan đào được trồng tại hai địa điểm A và B. Người ta thống kê đường kính thân của một số cây xoan đào 5 năm tuổi ở bảng sau:

Đường kính (cm)	[30; 32)	[32; 34)	[34; 36)	[36; 38)	[38; 40)
Số cây trồng ở địa điểm A	25	38	20	10	7
Số cây trồng ở địa điểm B	22	27	19	18	14

a) Hãy so sánh đường kính trung bình của thân cây xoan đào trồng tại địa điểm A và địa điểm B.

b) Nếu so sánh theo độ lệch chuẩn thì cây trồng tại địa điểm nào có đường kính đồng đều hơn?

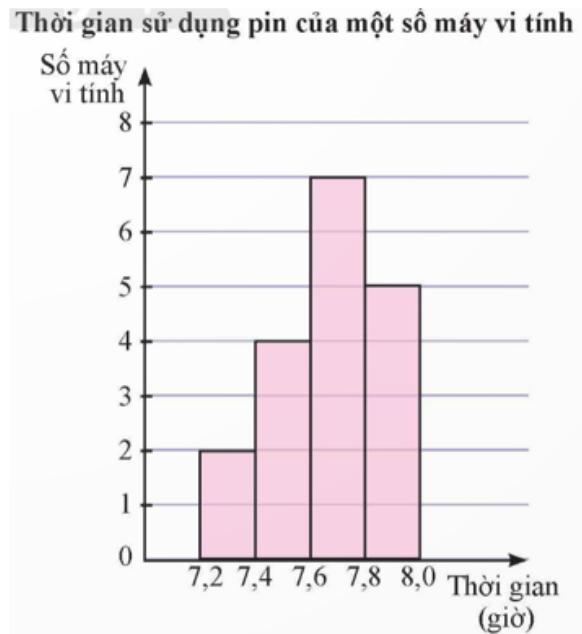
Câu 4. Biểu đồ sau biểu diễn chiều cao của học sinh nữ lớp 12.



a) Hãy lập bảng tần số ghép nhóm cho mẫu số liệu ở biểu đồ trên và xác định giá trị đại diện của mỗi nhóm và tính số trung bình của mẫu số liệu ghép nhóm.

b) Xét mẫu số liệu mới gồm các giá trị đại diện của các nhóm, tần số của mỗi giá trị đại diện bằng tần số của nhóm tương ứng. Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu mới.

Câu 5. Kết quả khảo sát thời gian sử dụng liên tục (đơn vị: giờ) từ lúc sạc đầy cho đến khi hết của pin một số máy vi tính cùng loại được mô tả bằng biểu đồ bên.



a) Hãy cho biết có bao nhiêu máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ?

b) Hãy xác định số trung bình và độ lệch chuẩn của thời gian sử dụng pin.

Câu 6. Tốc độ của 20 xe hơi khi đi qua một trạm kiểm tra tốc độ (đơn vị: km/h) được thống kê lại như sau:

42	43,4	43,4	46,5	46,7	46,8	47,5	47,7	48,1	48,4
50,8	52,1	52,7	53,9	54,8	55,6	57,5	59,6	60,3	61,1

a) Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên.

b) Hãy lập bảng tần số ghép nhóm với nhóm đầu tiên là $[42; 46)$ và độ dài mỗi nhóm bằng 4.

c) Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu ghép nhóm.

BÀI 2**PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN****1. Phương sai và độ lệch chuẩn**

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
Tần số	m_1	...	m_i	...	m_k

- **Phương sai** của mẫu số liệu ghép nhóm, kí hiệu S^2 , được tính theo công thức sau:

$$S^2 = \frac{1}{n} \left[m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2 \right]$$

Trong đó:

$$n = m_1 + m_2 + \dots + m_k; \quad x_i = \frac{a_i + a_{i+1}}{2} \text{ với } i = 1, 2, \dots, k \text{ là giá trị đại diện cho nhóm } [a_i; a_{i+1}).$$

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_k x_k}{n} \text{ là số trung bình của mẫu số liệu ghép nhóm.}$$

- **Độ lệch chuẩn** của mẫu số liệu ghép nhóm, kí hiệu S , là căn bậc hai số học của phương sai, nghĩa là $S = \sqrt{S^2}$.

Chú ý:

- Phương sai của mẫu số liệu ghép nhóm có thể được tính theo công thức sau:

$$S^2 = \frac{1}{n} \left[m_1 x_1^2 + m_2 x_2^2 + \dots + m_k x_k^2 \right] - \bar{x}^2$$

- Trong thống kê, người ta còn dùng đại lượng sau để đo mức độ phân tán của mẫu số liệu ghép nhóm

$$\hat{s}^2 = \frac{1}{n-1} \left[m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2 \right]; \quad \hat{s} = \sqrt{\hat{s}^2}$$

Ý nghĩa

- Phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm xấp xỉ phương sai (độ lệch chuẩn) của mẫu số liệu gốc. Chúng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm xung quanh số trung bình của mẫu số liệu.

- Phương sai, độ lệch chuẩn càng lớn thì mẫu số liệu càng phân tán.

2. Sử dụng phương sai, độ lệch chuẩn đo độ rủi ro

Câu 1. Bảng dưới đây thống kê cự li ném tạ của một vận động viên.

Cự li (m)	[19; 19,5)	[19,5; 20)	[20; 20,5)	[20,5; 21)	[21; 21,5)
Tần số	13	45	24	12	6

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

Lời giải

Ta có bảng sau:

Cự li (m)	[19; 19,5)	[19,5; 20)	[20; 20,5)	[20,5; 21)	[21; 21,5)
Giá trị đại diện	19,25	19,75	20,25	20,75	21,25
Tần số	13	45	24	12	6

Cỡ mẫu là $n = 13 + 45 + 24 + 12 + 6 = 100$.

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{13 \cdot 19,25 + 45 \cdot 19,75 + 24 \cdot 20,25 + 12 \cdot 20,75 + 6 \cdot 21,25}{100} = 20,015$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{100} \left[13 \cdot (19,25)^2 + 45 \cdot (19,75)^2 + 24 \cdot (20,25)^2 + 12 \cdot (20,75)^2 + 6 \cdot (21,25)^2 \right] - (20,015)^2 \approx 0,277$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} \approx \sqrt{0,277} \approx 0,526$

Câu 2. Minh Hiền và Minh Nhân cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

Số bước (đơn vị: nghìn)	[3; 5)	[5; 7)	[7; 9)	[9; 11)	[11; 13)
Số ngày của Minh Hiền	6	7	6	6	5
Số ngày của Minh Nhân	2	5	13	8	2

a) Hãy tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

b) Nếu so sánh theo độ lệch chuẩn thì bạn nào có số lượng bước chân đi mỗi ngày đều đặn hơn?

Lời giải

a) Ta có bảng sau:

Số bước (đơn vị: nghìn)	[3; 5)	[5; 7)	[7; 9)	[9; 11)	[11; 13)
Số bước đại diện	4	6	8	10	12
Số ngày của Minh Hiền	6	7	6	6	5
Số ngày của Minh Nhân	2	5	13	8	2

• Xét mẫu số liệu của Minh Hiền:

Cỡ mẫu là $n_H = 6 + 7 + 6 + 6 + 5 = 30$

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_H = \frac{6.4 + 7.6 + 6.8 + 6.10 + 5.12}{30} = 7,8$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S_H^2 = \frac{1}{30} (6.4^2 + 7.6^2 + 6.8^2 + 6.10^2 + 5.12^2) - (7,8)^2 = 7,56$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S_H = \sqrt{S_H^2} = \sqrt{7,56} \approx 2,75$

• Xét mẫu số liệu của Minh Nhân:

Cỡ mẫu là $n_N = 2 + 5 + 13 + 8 + 2 = 30$

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_N = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S_N^2 = \frac{1}{30} (2.4^2 + 5.6^2 + 13.8^2 + 8.10^2 + 2.12^2) - (8,2)^2 = 3,83$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S_N = \sqrt{S_N^2} \approx \sqrt{3,83} \approx 1,96$

b) Ta thấy $S_H \approx 2,75 > S_N \approx 1,96$

Do đó, nếu so sánh theo độ lệch chuẩn thì bạn Minh Nhân có số lượng bước chân đi mỗi ngày đều đặn hơn Minh Hiền.

Câu 3. Một giống cây xoan đào được trồng tại hai địa điểm A và B. Người ta thống kê đường kính thân của một số cây xoan đào 5 năm tuổi ở bảng sau:

Đường kính (cm)	[30; 32)	[32; 34)	[34; 36)	[36; 38)	[38; 40)
Số cây trồng ở địa điểm A	25	38	20	10	7
Số cây trồng ở địa điểm B	22	27	19	18	14

a) Hãy so sánh đường kính trung bình của thân cây xoan đào trồng tại địa điểm A và địa điểm B.

b) Nếu so sánh theo độ lệch chuẩn thì cây trồng tại địa điểm nào có đường kính đồng đều hơn?

Lời giải

a) Ta có bảng sau:

Đường kính (cm)	[30; 32)	[32; 34)	[34; 36)	[36; 38)	[38; 40)
Giá trị đại diện	31	33	35	37	39
Số cây trồng ở địa điểm A	25	38	20	10	7
Số cây trồng ở địa điểm B	22	27	19	18	14

Cỡ mẫu: $n_A = 25 + 38 + 20 + 10 + 7 = 100$; $n_B = 22 + 27 + 19 + 18 + 14 = 100$.

Đường kính trung bình của thân cây xoan đào trồng tại địa điểm A là:

$$\bar{x}_A = \frac{25.31 + 38.33 + 20.35 + 10.37 + 7.39}{100} = 33,72$$

Đường kính trung bình của thân cây xoan đào trồng tại địa điểm B là:

$$\bar{x}_B = \frac{22.31 + 27.33 + 19.35 + 18.37 + 14.39}{100} = 34,5$$

Vì $\bar{x}_A = 33,72 < \bar{x}_B = 34,5$ nên đường kính trung bình của thân cây xoan đào trồng tại địa điểm A nhỏ hơn tại địa điểm B.

b) Phương sai của mẫu số liệu ghép nhóm về đường kính của thân cây xoan đào trồng tại địa điểm A là:

$$S_A^2 = \frac{1}{100} (25.31^2 + 38.33^2 + 20.35^2 + 10.37^2 + 7.39^2) - (33,72)^2 \approx 5,402$$

Độ lệch chuẩn mẫu số liệu ghép nhóm về đường kính của thân cây xoan đào trồng tại địa điểm A là:

$$S_A = \sqrt{S_A^2} \approx \sqrt{5,402} \approx 2,324$$

Phương sai của mẫu số liệu ghép nhóm về đường kính của thân cây xoan đào trồng tại địa điểm B là:

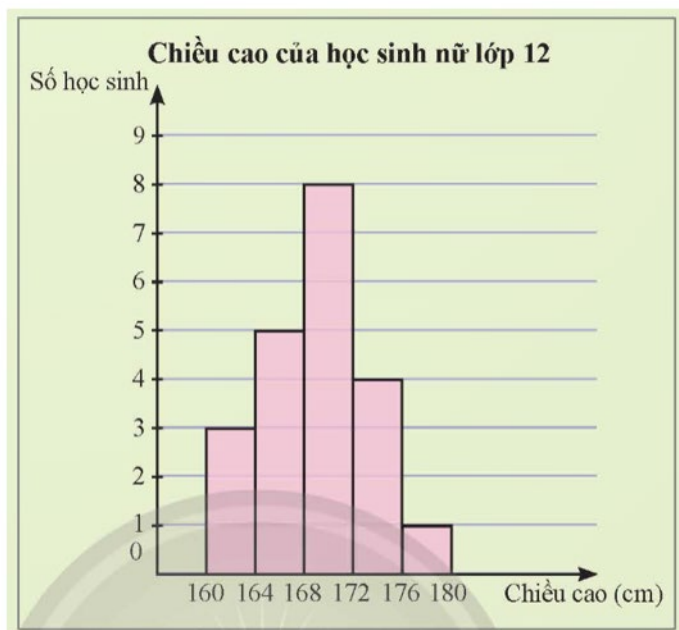
$$S_B^2 = \frac{1}{100} (22.31^2 + 27.33^2 + 19.35^2 + 18.37^2 + 14.39^2) - (34,5)^2 \approx 7,31.$$

Độ lệch chuẩn mẫu số liệu ghép nhóm về đường kính của thân cây xoan đào trồng tại địa điểm B là:

$$S_B = \sqrt{S_B^2} = \sqrt{7,31} \approx 2,704$$

Vì $S_A \approx 2,324 < S_B \approx 2,704$ nên nếu so sánh theo độ lệch chuẩn thì cây trồng tại địa điểm A có đường kính đồng đều hơn.

Câu 4. Biểu đồ sau biểu diễn chiều cao của học sinh nữ lớp 12.



a) Hãy lập bảng tần số ghép nhóm cho mẫu số liệu ở biểu đồ trên và xác định giá trị đại diện của mỗi nhóm và tính số trung bình của mẫu số liệu ghép nhóm.

b) Xét mẫu số liệu mới gồm các giá trị đại diện của các nhóm, tần số của mỗi giá trị đại diện bằng tần số của nhóm tương ứng. Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu mới.

Lời giải

a) Từ biểu đồ, ta lập được bảng tần số ghép nhóm và tính được giá trị đại diện của mỗi nhóm như sau:

Chiều cao	[160; 164)	[164; 168)	[168; 172)	[172; 176)	[176; 180)
-----------	------------	------------	------------	------------	------------

(cm)					
Số học sinh	3	5	8	4	1
Giá trị đại diện	162	166	170	174	178

Số học sinh nữ lớp 12 tham gia khảo sát là $n = 3 + 5 + 8 + 4 + 1 = 21$.

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{3.162 + 5.166 + 8.170 + 4.174 + 1.178}{21} = \frac{3550}{21}$$

b) Ta có bảng thống kê mẫu số liệu mới:

Giá trị đại diện	162	166	170	174	178
Số học sinh	3	5	8	4	1

Cỡ mẫu $n = 21$.

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{3550}{21}$

Phương sai của mẫu số liệu ghép nhóm là:

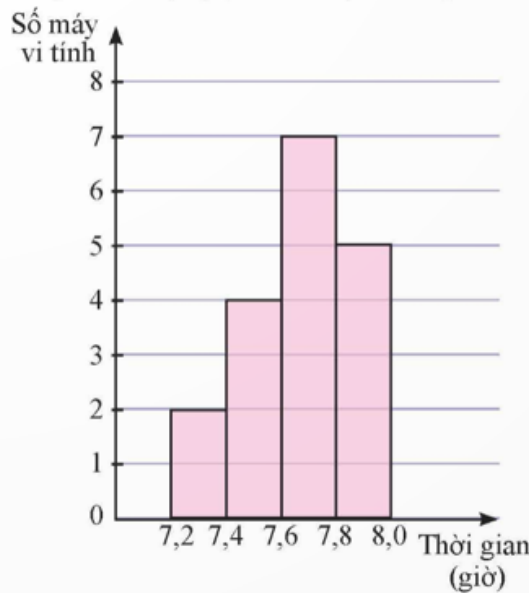
$$s^2 = \frac{3\left(162 - \frac{3550}{21}\right)^2 + 5\left(166 - \frac{3550}{21}\right)^2 + 8\left(170 - \frac{3550}{21}\right)^2 + 4\left(174 - \frac{3550}{21}\right)^2 + 1\left(178 - \frac{3550}{21}\right)^2}{21} = \frac{8000}{441} \approx 18,14$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là:

$$s = \sqrt{s^2} = \sqrt{\frac{8000}{441}} \approx 4,26$$

Câu 5. Kết quả khảo sát thời gian sử dụng liên tục (đơn vị: giờ) từ lúc sạc đầy cho đến khi hết của pin một số máy vi tính cùng loại được mô tả bằng biểu đồ bên.

Thời gian sử dụng pin của một số máy vi tính



- a) Hãy cho biết có bao nhiêu máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ?
 b) Hãy xác định số trung bình và độ lệch chuẩn của thời gian sử dụng pin.

Lời giải

- a) Từ biểu đồ ta thấy có 2 máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ.
 b) Từ biểu đồ, ta có bảng thống kê sau:

Thời gian (giờ)	[7,2; 7,4)	[7,4; 7,6)	[7,6; 7,8)	[7,8; 8,0)
Giá trị đại diện	7,3	7,5	7,7	7,9
Số máy vi tính	2	4	7	5

Cỡ mẫu là $n = 2 + 4 + 7 + 5 = 18$.

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{2 \cdot 7,3 + 4 \cdot 7,5 + 7 \cdot 7,7 + 5 \cdot 7,9}{18} = \frac{23}{3}$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{18} \left[2 \cdot (7,3)^2 + 4 \cdot (7,5)^2 + 7 \cdot (7,7)^2 + 5 \cdot (7,9)^2 \right] - \left(\frac{23}{3} \right)^2 \approx 0,032$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} \approx \sqrt{0,032} \approx 0,179$

Câu 6. Tốc độ của 20 xe hơi khi đi qua một trạm kiểm tra tốc độ (đơn vị: km/h) được thống kê lại như sau:

42 43,4 43,4 46,5 46,7 46,8 47,5 47,7 48,1 48,4
 50,8 52,1 52,7 53,9 54,8 55,6 57,5 59,6 60,3 61,1

- a) Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên.
 b) Hãy lập bảng tần số ghép nhóm với nhóm đầu tiên là [42; 46) và độ dài mỗi nhóm bằng 4.
 c) Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu ghép nhóm.

Lời giải

a) Mẫu số liệu đã cho đã được xếp theo thứ tự không giảm.

Khoảng biến thiên của mẫu số liệu là:

$$R = 61,1 - 42 = 19,1 \text{ (km/h)}.$$

Cỡ mẫu $n = 20$.

Tứ phân vị thứ nhất là trung vị của mẫu số liệu:

42 43,4 43,4 46,5 46,7 46,8 47,5 47,7 48,1 48,4

$$\text{Do đó, } Q_1 = \frac{46,7 + 46,8}{2} = 46,75$$

Tứ phân vị thứ ba là trung vị của mẫu số liệu:

50,8 52,1 52,7 53,9 54,8 55,6 57,5 59,6 60,3 61,1

$$\text{Do đó, } Q_3 = \frac{54,8 + 55,6}{2} = 55,2$$

Khoảng tứ phân vị của mẫu số liệu là: $\Delta Q = Q_3 - Q_1 = 55,2 - 46,75 = 8,45$.

Số trung bình của mẫu số liệu là:

$$\bar{x} = \frac{42 + 43,4 + \dots + 60,3 + 61,1}{20} = 50,945$$

Phương sai của mẫu số liệu là:

$$S^2 = \frac{1}{20} \left[42^2 + (43,4)^2 + \dots + (60,3)^2 + (61,1)^2 \right] - (50,945)^2 \approx 32,2$$

Độ lệch chuẩn của mẫu số liệu là: $S = \sqrt{S^2} \approx \sqrt{32,2} \approx 5,675$

b) Ta lập được bảng tần số ghép nhóm như sau:

Tốc độ (km/h)	[42; 46)	[46; 50)	[50; 54)	[54; 58)	[58; 62)
Số xe	3	7	4	3	3

c) Khoảng biến thiên của mẫu số liệu ghép nhóm là:

$$R' = 62 - 42 = 20 \text{ (km/h)}.$$

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về tốc độ của 20 xe hơi đi qua một trạm kiểm tra tốc độ được xếp theo thứ tự không giảm.

Ta có

$$x_1; x_2; x_3 \in [42; 46),$$

$$x_4; \dots; x_{10} \in [46; 50),$$

$$x_{11}; \dots; x_{14} \in [50; 54),$$

$$x_{15}; \dots; x_{17} \in [54; 58),$$

$$x_{18}; x_{19}; x_{20} \in [58; 62).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [46; 50)$.

Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1' = 46 + \frac{\frac{20}{4} - 3}{7} (50 - 46) = \frac{330}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [54; 58)$.

Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3' = 54 + \frac{\frac{3 \cdot 20}{4} - (3 + 7 + 4)}{3} (58 - 54) = \frac{166}{3}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là:

$$\Delta_Q' = Q_3' - Q_1' = \frac{166}{3} - \frac{330}{7} \approx 8,19$$

Từ bảng tần số ghép nhóm, ta có bảng sau:

Tốc độ (km/h)	[42; 46)	[46; 50)	[50; 54)	[54; 58)	[58; 62)
Giá trị đại diện	44	48	52	56	60
Số xe	3	7	4	3	3

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}' = \frac{3 \cdot 44 + 7 \cdot 48 + 4 \cdot 52 + 3 \cdot 56 + 3 \cdot 60}{20} = 51,2$$

Phương sai của mẫu số liệu ghép nhóm là:

$$(S')^2 = \frac{1}{20} [3 \cdot 44^2 + 7 \cdot 48^2 + 4 \cdot 52^2 + 3 \cdot 56^2 + 3 \cdot 60^2] - (51,2)^2 = 26,56$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là:

$$S' = \sqrt{(S')^2} = \sqrt{26,56} \approx 5,154$$

BÀI 3**ÔN TẬP CHƯƠNG 3****CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM**

PHẦN I. Câu trắc nghiệm nhiều phương án lựa chọn. Mỗi câu hỏi thí sinh chỉ chọn một phương án.

Câu 1. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

Tuổi thọ	[14; 15)	[15; 16)	[16; 17)	[17; 18)	[18; 19)
Số con hổ	1	3	8	6	2

a) Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 4. **B. 5.** C. 6. D. 7.

b) Nhóm chứa tứ phân vị thứ nhất là

- A. [14; 15). B. [15; 16). **C. [16; 17).** D. [17; 18).

c) Nhóm chứa tứ phân vị thứ ba là

- A. [18; 19). B. [15; 16). C. [16; 17). **D. [17; 18).**

d) Số đặc trưng nào không sử dụng thông tin của nhóm số liệu đầu tiên và nhóm số liệu cuối cùng.

- A. Khoảng biến thiên. **B. khoảng tứ phân vị.**
C. Phương sai. D. Độ lệch chuẩn.

e) Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì số đặc trưng nào sau đây không thay đổi?

- A. Khoảng biến thiên.** B. khoảng tứ phân vị.
C. Phương sai. D. Độ lệch chuẩn.

Câu 2. Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

Quãng đường (km)	[2,7; 3,0)	[3,0; 3,3)	[3,3; 3,6)	[3,6; 3,9)	[3,9; 4,2)
Số ngày	3	6	5	4	2

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là

- A. 1,5.** B. 0,9. C. 0,6. D. 0,3.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A. 0,9. B. 0,975. C. 0,5. **D. 0,575.**

c) Phương sai của mẫu số liệu ghép nhóm là

- A. 3,39. B. 11,62. **C. 0,1314.** D. 0,36.

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

- A. 3,41. B. 11,62. C. 0,017. **D. 0,36.**

Câu 3. Cô Minh Hiền rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

Thời gian (phút)	[20; 25)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Số ngày	6	6	4	1	1

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm là
A. 25. **B. 20.** **C. 15.** **D. 30.**
- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là
A. 23,75. **B. 27,5.** **C. 31,88.** **D. 8,125.**
- c) Phương sai của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?
A. 31,77. **B. 32.** **C. 31.** **D. 31,44.**

Câu 4. Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

Thời gian giải rubik (giây)	[8; 10)	[10; 12)	[12; 14)	[14; 16)	[16; 18)
Số lần	4	6	8	4	3

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm nhận giá trị nào trong các giá trị dưới đây?
A. 6. **B. 8.** **C. 10.** **D. 12.**
- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là
A. 10,75. **B. 1,75.** **C. 3,63.** **D. 14,38.**
- c) Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?
A. 5,98. **B. 6.** **C. 2,44.** **D. 2,5.**

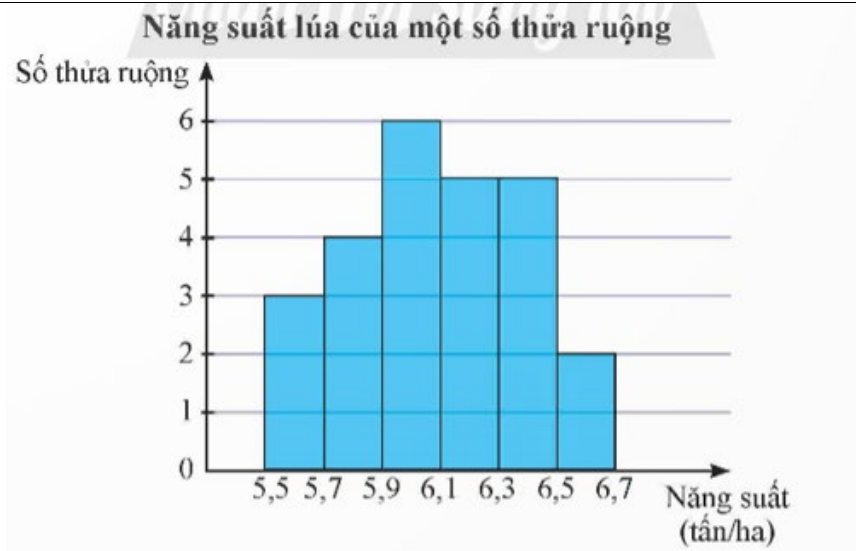
PHẦN II. Câu trắc nghiệm đúng sai. Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai.

Câu 5. Một bác tài xế thống kê lại độ dài quãng đường (đơn vị: km) bác đã lái xe mỗi ngày trong một tháng ở bảng sau:

Độ dài quãng đường (km)	[50; 100)	[100; 150)	[150; 200)	[200; 250)	[250; 300)
Số ngày	5	10	9	4	2

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 250 (km).
- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm gần bằng 79,17 .
- c) Số trung bình của mẫu số liệu ghép nhóm là 145 .
- d) Độ lệch chuẩn của mẫu số liệu ghép nhóm gần bằng 55,68 .

Câu 6. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau:



- a) Có 25 thửa ruộng đã được khảo sát.
- b) Khoảng biến thiên của mẫu số liệu trên là 1,2 (tấn/ha).
- c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là 0,4675 .
- d) Phương sai của mẫu số liệu ghép nhóm trên là 0,086656 .

Câu 7. Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

Thời gian (phút)	[6; 7)	[7; 8)	[8; 9)	[9; 10)	[10; 11)
Số học sinh trường X	8	10	13	10	9
Số học sinh trường Y	4	12	17	14	3

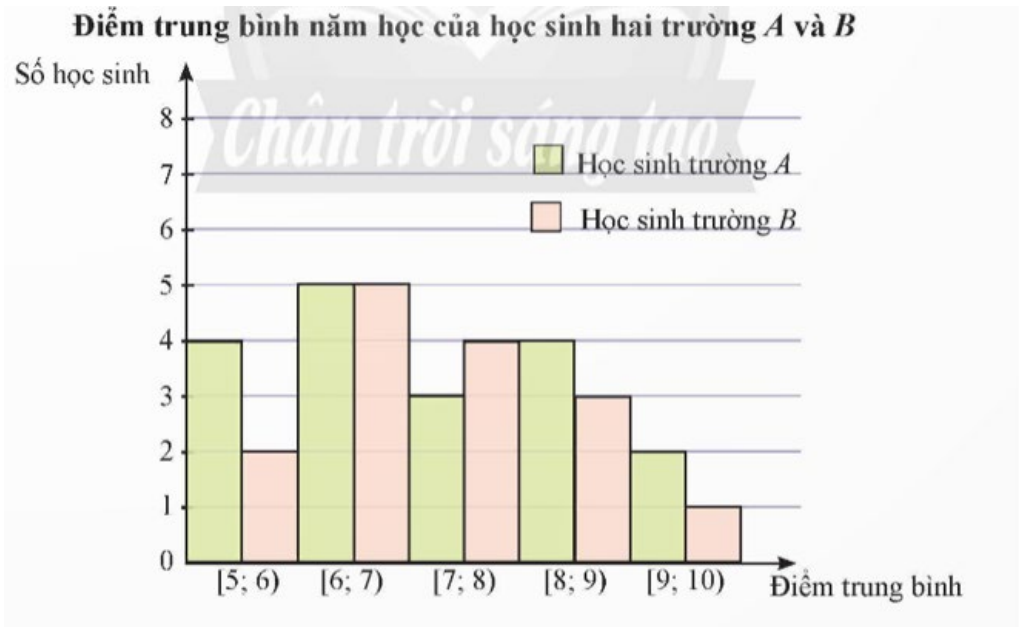
- a) Nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn.
- b) Nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn.
- c) Phương sai của mẫu số liệu ghép nhóm của trường X là 1,08 và Phương sai của mẫu số liệu ghép nhóm của trường Y là 1,7584 .
- d) Nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn.

Câu 8. Bảng sau thống kê lại tổng số giờ nắng trong tháng 6 của các năm từ 2002 đến 2021 tại hai trạm quan trắc đặt ở Nha Trang và Quy Nhơn.

Số giờ nắng	[130; 160)	[160; 190)	[190; 220)	[220; 250)	[250; 280)	[280; 310)
Số năm ở Nha Trang	1	1	1	8	7	2
Số năm ở Quy Nhơn	0	1	2	4	10	3

- a) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trạm quan trắc ở Nha Trang bằng 45 .
- b) Nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn Nha Trang.
- c) Số trung bình của mẫu số liệu ghép nhóm trạm quan trắc ở Quy Nhơn bằng 242,5 .
- d) Nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn Nha Trang..

Câu 9. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



a) Giá trị đại diện cho mỗi nhóm và bảng tần số ghép nhóm của mẫu số liệu trên là:

Điểm trung bình	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Giá trị đại diện	5,5	6,5	7,5	8,5	9,5
Số học sinh trường A	4	5	3	4	2
Số học sinh trường B	2	5	4	3	1

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm của trường A là 2,275.

c) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường A có điểm trung bình đồng đều hơn trường B.

d) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn.

PHẦN III. Câu trắc nghiệm trả lời ngắn. Mỗi câu hỏi thí sinh chỉ trả lời đáp án.

Câu 10. Thời gian hoàn thành bài kiểm tra môn Toán của các bạn trong lớp 12C được cho trong bảng sau:

Thời gian (phút)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Số học sinh	8	16	4	2

a) Tính khoảng biến thiên R cho mẫu số liệu ghép nhóm trên.

b) Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 27 phút và muộn nhất mất 43 phút thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

Trả lời:

Câu 11. Để đánh giá chất lượng một loại pin điện thoại mới, người ta ghi lại thời gian nghe nhạc liên tục của điện thoại được sạc đầy pin cho đến khi hết pin cho kết quả sau:

Thời gian (giờ)	[5; 5,5)	[5,5; 6)	[6; 6,5)	[6,5; 7)	[7; 7,5)
Số chiếc điện thoại (tần số)	2	8	15	10	5

Tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

Trả lời:

Câu 12. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư (với số tiền đầu tư như nhau), khi đầu tư vào hai lĩnh vực A, B cho kết quả như sau:

Tiền lãi	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
Số nhà đầu tư vào lĩnh vực A	2	5	8	6	4
Số nhà đầu tư vào lĩnh vực B	8	4	2	5	6

a) Về trung bình, đầu tư vào lĩnh vực nào đem lại tiền lãi cao hơn?

b) Tính độ lệch chuẩn cho các mẫu số liệu về tiền lãi của các nhà đầu tư ở hai lĩnh vực này và giải thích ý nghĩa của các số thu được.

Trả lời:

Câu 13. Thành tích môn nhảy cao của các vận động viên tại một giải điền kinh dành cho học sinh trung học phổ thông như sau:

Mức xà (cm)	[170; 172)	[172; 174)	[174; 176)	[176; 180)
Số vận động viên	3	10	6	1

a) Tính các số đặc trưng đo mức độ phân tán của mẫu số liệu ghép nhóm trên.

b) Độ phân tán của mẫu số liệu cho biết điều gì?

Trả lời:

Câu 14. Trong thực hành đo hiệu điện thế của mạch điện, An và Bình đã dùng hai vôn kế khác nhau để đo, mỗi bạn tiến hành đo 10 lần và cho kết quả như sau:

Hiệu điện thế đo được (Vôn)	[3,85; 3,90)	[3,90; 3,95)	[3,95; 4,00)	[4,00; 4,05)
Số lần An đo	1	6	2	1
Số lần Bình đo	1	3	4	2

Tính độ lệch chuẩn của các mẫu số liệu ghép nhóm cho kết quả đo của An và Bình. Từ đó kết luận xem vôn kế của bạn nào cho kết quả đo ổn định hơn.

Trả lời:

Câu 15. Một người ghi lại thời gian đàm thoại của một số cuộc gọi cho kết quả như bảng sau:

Thời gian t (phút)	Số cuộc gọi
$0 \leq t < 1$	8
$1 \leq t < 2$	17
$2 \leq t < 3$	25
$3 \leq t < 4$	20
$4 \leq t < 5$	10

Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

Trả lời:

BÀI 3**ÔN TẬP CHƯƠNG 3****CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM**

PHẦN I. Câu trắc nghiệm nhiều phương án lựa chọn. Mỗi câu hỏi thí sinh chỉ chọn một phương án.

Câu 1. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

Tuổi thọ	[14; 15)	[15; 16)	[16; 17)	[17; 18)	[18; 19)
Số con hổ	1	3	8	6	2

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm này là
 A. 4. **B. 5.** C. 6. D. 7.
- b) Nhóm chứa tứ phân vị thứ nhất là
 A. [14; 15). **B. [15; 16).** **C. [16; 17).** D. [17; 18).
- c) Nhóm chứa tứ phân vị thứ ba là
 A. [18; 19). **B. [15; 16).** C. [16; 17). **D. [17; 18).**
- d) Số đặc trưng nào không sử dụng thông tin của nhóm số liệu đầu tiên và nhóm số liệu cuối cùng.
 A. Khoảng biến thiên. **B. khoảng tứ phân vị.**
 C. Phương sai. D. Độ lệch chuẩn.
- e) Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì số đặc trưng nào sau đây không thay đổi?
A. Khoảng biến thiên. B. khoảng tứ phân vị.
 C. Phương sai. D. Độ lệch chuẩn.

Lời giải

a) **Chọn B**

Khoảng biến thiên $R = 19 - 14 = 5$

b) **Chọn C**

Cỡ mẫu là: $1 + 3 + 8 + 6 + 2 = 20$.

Gọi $x_1; x_2; \dots; x_{20}$ là tuổi thọ của 20 con hổ được sắp xếp theo thứ tự tăng dần.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [16; 17)$ nên nhóm chứa tứ phân vị thứ nhất là [16; 17).

c) **Chọn D**

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [17; 18)$. Do đó nhóm chứa tứ phân vị thứ ba là [17; 18).

d) **Chọn B**

Số đặc trưng không sử dụng thông tin của nhóm số liệu đầu tiên và nhóm số liệu cuối cùng là khoảng tứ phân vị.

e) **Chọn A**

Khoảng biến thiên sẽ không thay đổi nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4.

Câu 2. Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

Quãng đường (km)	[2,7; 3,0)	[3,0; 3,3)	[3,3; 3,6)	[3,6; 3,9)	[3,9; 4,2)
Số ngày	3	6	5	4	2

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là

A. 1,5.

B. 0,9.

C. 0,6.

D. 0,3.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 0,9.

B. 0,975.

C. 0,5.

D. 0,575.

c) Phương sai của mẫu số liệu ghép nhóm là

A. 3,39.

B. 11,62.

C. 0,1314.

D. 0,36.

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 3,41.

B. 11,62.

C. 0,017.

D. 0,36.

Lời giải

a) **Chọn A**

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 4,2 - 2,7 = 1,5$ (km).

b) **Chọn D**

Cỡ mẫu $n = 20$.

Gọi $x_1; \dots; x_{20}$ là mẫu số liệu gốc về quãng đường đi bộ mỗi ngày của bác Hương trong 20 ngày được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_3 \in [2,7; 3,0),$$

$$x_4; \dots; x_9 \in [3,0; 3,3),$$

$$x_{10}; \dots; x_{14} \in [3,3; 3,6),$$

$$x_{15}; \dots; x_{18} \in [3,6; 3,9),$$

$$x_{19}; x_{20} \in [3,9; 4,2).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [3,0; 3,3)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1 = 3,0 + \frac{\frac{20}{4} - 3}{6}(3,3 - 3,0) = 3,1$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [3,6; 3,9)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3 = 3,6 + \frac{\frac{3 \cdot 20}{4} - (3 + 6 + 5)}{4} (3,9 - 3,6) = 3,675$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 3,675 - 3,1 = 0,575$

c) **Chọn C**

Ta có bảng sau:

Quãng đường (km)	[2,7; 3,0)	[3,0; 3,3)	[3,3; 3,6)	[3,6; 3,9)	[3,9; 4,2)
Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
Số ngày	3	6	5	4	2

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{3 \cdot 2,85 + 6 \cdot 3,15 + 5 \cdot 3,45 + 4 \cdot 3,75 + 2 \cdot 4,05}{20} = 3,39$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{20} [3 \cdot (2,85)^2 + 6 \cdot (3,15)^2 + 5 \cdot (3,45)^2 + 4 \cdot (3,75)^2 + 2 \cdot (4,05)^2] - (3,39)^2 = 0,1314$$

d) **Chọn D**

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} = \sqrt{0,1314} \approx 0,36$

Câu 3. Cô Minh Hiền rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

Thời gian (phút)	[20; 25)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Số ngày	6	6	4	1	1

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là

A. 25.

B. 20.

C. 15.

D. 30.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 23,75.

B. 27,5.

C. 31,88.

D. 8,125.

c) Phương sai của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 31,77.

B. 32.

C. 31.

D. 31,44.

Lời giải

a) **Chọn A**

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 45 - 20 = 25$ (phút).

b) **Chọn D**

Cỡ mẫu $n = 6 + 6 + 4 + 1 + 1 = 18$.

Gọi $x_1; \dots; x_{18}$ là mẫu số liệu gốc về thời gian tập nhảy mỗi ngày của Cô Minh Hiền được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_6 \in [20; 25),$$

$$x_7; \dots; x_{12} \in [25; 30),$$

$$x_{13}; \dots; x_{16} \in [30; 35),$$

$$x_{17} \in [35; 40),$$

$$x_{18} \in [40; 45).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_5 \in [20; 25)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu

$$\text{ghép nhóm là: } Q_1 = 20 + \frac{18}{4}(25 - 20) = 23,75.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{14} \in [30; 35)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép

$$\text{nhóm là: } Q_3 = 30 + \frac{3 \cdot 18 - (6 + 6)}{4}(35 - 30) = 31,875$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 31,875 - 23,75 = 8,125$

c) Chọn D

Ta có bảng sau:

Thời gian (phút)	[20; 25)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Giá trị đại diện	22,5	27,5	32,5	37,5	42,5
Số ngày	6	6	4	1	1

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{6 \cdot 22,5 + 6 \cdot 27,5 + 4 \cdot 32,5 + 1 \cdot 37,5 + 1 \cdot 42,5}{18} = \frac{85}{3}.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{18} \left[6 \cdot (22,5)^2 + 6 \cdot (27,5)^2 + 4 \cdot (32,5)^2 + 1 \cdot (37,5)^2 + 1 \cdot (42,5)^2 \right] - \left(\frac{85}{3} \right)^2 = 31,25$$

Do đó, phương sai của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị 31,44.

Câu 4. Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

Thời gian giải rubik (giây)	[8; 10)	[10; 12)	[12; 14)	[14; 16)	[16; 18)
Số lần	4	6	8	4	3

a) Khoảng biến thiên của mẫu số liệu ghép nhóm nhận giá trị nào trong các giá trị dưới đây?

A. 6.

B. 8.

C. 10.

D. 12.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 10,75.

B. 1,75.

C. 3,63.

D. 14,38.

c) Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 5,98.

B. 6.

C. 2,44.

D. 2,5.

Lời giảia) **Chọn C**Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 18 - 8 = 10$ (giây).b) **Chọn C**Cỡ mẫu $n = 25$.Gọi $x_1; \dots; x_{25}$ là mẫu số liệu gốc về thời gian giải rubik trong 25 lần giải liên tiếp được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_4 \in [8; 10),$$

$$x_5; \dots; x_{10} \in [10; 12),$$

$$x_{11}; \dots; x_{18} \in [12; 14),$$

$$x_{19}; \dots; x_{22} \in [14; 16),$$

$$x_{23}; \dots; x_{25} \in [16; 18).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_6 + x_7}{2} \in [10; 12)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu

$$\text{ghép nhóm là: } Q_1 = 10 + \frac{\frac{25}{4} - 4}{6}(12 - 10) = 10,75$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{19} + x_{20}}{2} \in [14; 16)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3 = 10 + \frac{\frac{3 \cdot 25}{4} - (4 + 6 + 8)}{4}(16 - 14) = 14,375$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 14,375 - 10,75 = 3,625$ c) **Chọn C**

Ta có bảng sau:

Thời gian giải rubik (giây)	[8; 10)	[10; 12)	[12; 14)	[14; 16)	[16; 18)
Giá trị đại diện	9	11	13	15	17
Số lần	4	6	8	4	3

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{4 \cdot 9 + 6 \cdot 11 + 8 \cdot 13 + 4 \cdot 15 + 3 \cdot 17}{25} = 12,68$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{25} \left[4.9^2 + 6.11^2 + 8.13^2 + 4.15^2 + 3.17^2 \right] - (12,68)^2 = 5,9776$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} = \sqrt{5,9776} \approx 2,44$

PHẦN II. Câu trắc nghiệm đúng sai. Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai.

Câu 5. Một bác tài xế thống kê lại độ dài quãng đường (đơn vị: km) bác đã lái xe mỗi ngày trong một tháng ở bảng sau:

Độ dài quãng đường (km)	[50; 100)	[100; 150)	[150; 200)	[200; 250)	[250; 300)
Số ngày	5	10	9	4	2

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 250 (km).
- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm gần bằng 79,17 .
- c) Số trung bình của mẫu số liệu ghép nhóm là 145 .
- d) Độ lệch chuẩn của mẫu số liệu ghép nhóm gần bằng 55,68 .

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 300 - 50 = 250$ (km).

Cỡ mẫu $n = 5 + 10 + 9 + 4 + 2 = 30$.

Gọi $x_1;...;x_{30}$ là mẫu số liệu gốc về độ dài quãng đường bác tài xế đã lái xe mỗi ngày trong một tháng được xếp theo thứ tự không giảm.

Ta có

$$x_1;...;x_5 \in [50; 100),$$

$$x_6;...;x_{15} \in [100; 150),$$

$$x_{16};...;x_{24} \in [150; 200),$$

$$x_{25};...;x_{28} \in [200; 250),$$

$$x_{29};x_{30} \in 250; 300).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_8 \in [100; 150)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu

ghép nhóm là: $Q_1 = 100 + \frac{\frac{30}{4} - 5}{10} (150 - 100) = 112,5$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{23} \in [150; 200)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép

$$\text{nhóm là: } Q_3 = 100 + \frac{\frac{3 \cdot 30}{4} - (5+10)}{9} (200 - 150) = \frac{575}{3}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = \frac{575}{3} - 112,5 \approx 79,17$$

Ta có bảng sau:

Độ dài quãng đường (km)	[50; 100)	[100; 150)	[150; 200)	[200; 250)	[250; 300)
Giá trị đại diện	75	125	175	225	275
Số ngày	5	10	9	4	2

Số trung bình của mẫu số liệu ghép nhóm là:

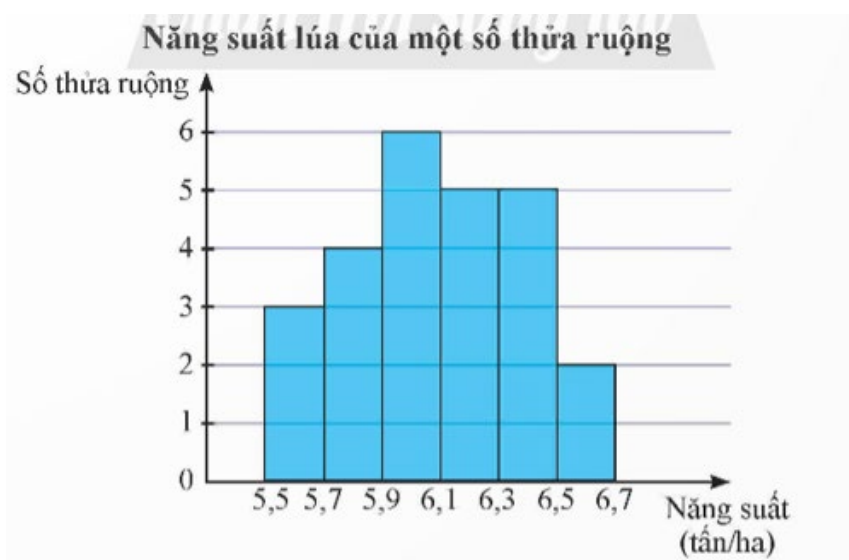
$$\bar{x} = \frac{5 \cdot 75 + 10 \cdot 125 + 9 \cdot 175 + 4 \cdot 225 + 2 \cdot 275}{30} = 155.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{30} [5 \cdot 75^2 + 10 \cdot 125^2 + 9 \cdot 175^2 + 4 \cdot 225^2 + 2 \cdot 275^2] - 155^2 = 3100$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} = \sqrt{3100} \approx 55,68$.

Câu 6. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau:



- Có 25 thửa ruộng đã được khảo sát.
- Khoảng biến thiên của mẫu số liệu trên là 1,2 (tấn/ha).
- Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là 0,4675.
- Phương sai của mẫu số liệu ghép nhóm trên là 0,086656.

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	ĐÚNG	ĐÚNG

a) Số thửa ruộng được khảo sát là: $n = 3 + 4 + 6 + 5 + 5 + 2 = 25$.

b) Từ biểu đồ, ta có bảng tần số ghép nhóm của mẫu số liệu như sau:

Năng suất (tấn/ha)	[5,5; 5,7)	[5,7; 5,9)	[5,9; 6,1)	[6,1; 6,3)	[6,3; 6,5)	[6,5; 6,7)
Giá trị đại diện (tấn/ha)	5,6	5,8	6,0	6,2	6,4	6,6
Tần số tương đối	3	4	6	5	5	2

Khoảng biến thiên của mẫu số liệu đã cho là: $R = 6,7 - 5,5 = 1,2$ (tấn/ha).

c)

Cỡ mẫu $n = 25$.

Gọi $x_1; \dots; x_{25}$ là mẫu số liệu gốc về năng suất của một số thửa ruộng được khảo sát được xếp theo thứ tự không giảm.

Ta có

$$x_1; x_2; x_3 \in [5,5; 5,7),$$

$$x_4; \dots; x_7 \in [5,7; 5,9),$$

$$x_8; \dots; x_{13} \in [5,9; 6,1),$$

$$x_{14}; \dots; x_{18} \in [6,1; 6,3),$$

$$x_{19}; \dots; x_{23} \in [6,3; 6,5),$$

$$x_{24}; x_{25} \in [6,5; 6,7).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_6 + x_7}{2} \in [5,7; 5,9)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1 = 5,7 + \frac{\frac{25}{4} - 3}{4} (5,9 - 5,7) = 5,8625$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{19} + x_{20}}{2} \in [6,3; 6,5)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3 = 6,3 + \frac{\frac{3 \cdot 25}{4} - (3 + 4 + 6 + 5)}{5} (6,5 - 6,3) = 6,33$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 6,33 - 5,8625 = 0,4675$

d) Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{3 \cdot 5,6 + 4 \cdot 5,8 + 6 \cdot 6,2 + 5 \cdot 6,4 + 2 \cdot 6,6}{25} = 6,088.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{25} \left[3 \cdot (5,6)^2 + 4 \cdot (5,8)^2 + 6 \cdot (6,0)^2 + 5 \cdot (6,2)^2 + 5 \cdot (6,4)^2 + 2 \cdot (6,6)^2 \right] - (6,088)^2 = 0,086656$$

Câu 7. Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

Thời gian (phút)	[6; 7)	[7; 8)	[8; 9)	[9; 10)	[10; 11)
Số học sinh trường X	8	10	13	10	9
Số học sinh trường Y	4	12	17	14	3

a) Nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn.

b) Nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn.

c) Phương sai của mẫu số liệu ghép nhóm của trường X là 1,08 và Phương sai của mẫu số liệu ghép nhóm của trường Y là 1,7584.

d) Nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn.

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Ta có bảng sau:

Thời gian (phút)	[6; 7)	[7; 8)	[8; 9)	[9; 10)	[10; 11)
Giá trị đại diện	6,5	7,5	8,5	9,5	10,5
Số học sinh trường X	8	10	13	10	9
Số học sinh trường Y	4	12	17	14	3

Cỡ mẫu $n_X = 8 + 10 + 13 + 10 + 9 = 50$, $n_Y = 4 + 12 + 17 + 14 + 3 = 50$.

Thời gian trung bình hoàn thành một bài viết chính tả của học sinh trường X là:

$$\bar{x}_X = \frac{6 \cdot 6,5 + 10 \cdot 7,5 + 13 \cdot 8,5 + 10 \cdot 9,5 + 9 \cdot 10,5}{50} = 8,54$$

Thời gian trung bình hoàn thành một bài viết chính tả của học sinh trường Y là:

$$\bar{x}_Y = \frac{4 \cdot 6,5 + 12 \cdot 7,5 + 17 \cdot 8,5 + 14 \cdot 9,5 + 3 \cdot 10,5}{50} = 8,5$$

Vì $\bar{x}_X > \bar{x}_Y$ nên nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn.

b)

• Xét mẫu số liệu của học sinh trường X:

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 trường X được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_8 \in [6; 7),$$

$$x_9; \dots; x_{18} \in [7; 8),$$

$$x_{19}; \dots; x_{31} \in [8; 9),$$

$$x_{32}; \dots; x_{41} \in [9; 10),$$

$$x_{42}; \dots; x_{50} \in [10; 11).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_{13} \in [7; 8)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm của trường X là:

$$Q_1 = 7 + \frac{\frac{50}{4} - 8}{10}(8 - 7) = 7,45$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{38} \in [9; 10)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm của trường X là:

$$Q_3 = 9 + \frac{\frac{3.50}{4} - (8 + 10 + 13)}{10}(10 - 9) = 9,65$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm của trường X là: $\Delta_Q = Q_3 - Q_1 = 9,65 - 7,45 = 2,2$

• Xét mẫu số liệu của học sinh trường Y:

Gọi $y_1; \dots; y_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 trường Y được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_4 \in [6; 7),$$

$$y_5; \dots; y_{16} \in [7; 8),$$

$$y_{17}; \dots; y_{33} \in [8; 9),$$

$$y_{34}; \dots; y_{47} \in [9; 10),$$

$$y_{48}; \dots; y_{50} \in [10; 11).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $y_{13} \in [7; 8)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm của trường Y là:

$$Q'_1 = 7 + \frac{\frac{50}{4} - 4}{12}(8 - 7) = \frac{185}{24}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $y_{38} \in [9; 10)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm của trường Y là:

$$Q'_3 = 9 + \frac{\frac{3.50}{4} - (4 + 12 + 17)}{14}(10 - 9) = \frac{261}{28}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm của trường Y là:

$$\Delta_Q = Q'_3 - Q'_1 = \frac{261}{28} - \frac{185}{24} \approx 1,61$$

Vì $\Delta_Q = 2,2 > \Delta'_Q \approx 1,61$ nên nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn.

c)

Phương sai của mẫu số liệu ghép nhóm của trường X là:

$$S_X^2 = \frac{1}{50} \left[8.(6,5)^2 + 10.(7,5)^2 + 13.(8,5)^2 + 10.(9,5)^2 + 9.(10,5)^2 \right] - (8,54)^2 = 1,7584$$

Phương sai của mẫu số liệu ghép nhóm của trường Y là:

$$S_Y^2 = \frac{1}{50} \left[4.(6,5)^2 + 12.(7,5)^2 + 17.(8,5)^2 + 14.(9,5)^2 + 3.(10,5)^2 \right] - (8,5)^2 = 1,08$$

d)

Độ lệch chuẩn của mẫu số liệu ghép nhóm của trường X là: $S_X = \sqrt{S_X^2} = \sqrt{1,7584} \approx 1,33$

Độ lệch chuẩn của mẫu số liệu ghép nhóm của trường Y là: $S_Y = \sqrt{S_Y^2} = \sqrt{1,08} \approx 1,04$

Vì $S_X \approx 1,33 > S_Y \approx 1,04$ nên nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn.

Câu 8. Bảng sau thống kê lại tổng số giờ nắng trong tháng 6 của các năm từ 2002 đến 2021 tại hai trạm quan trắc đặt ở Nha Trang và Quy Nhơn.

Số giờ nắng	[130; 160)	[160; 190)	[190; 220)	[220; 250)	[250; 280)	[280; 310)
Số năm ở Nha Trang	1	1	1	8	7	2
Số năm ở Quy Nhơn	0	1	2	4	10	3

a) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trạm quan trắc ở Nha Trang bằng 45 .

b) Nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn Nha Trang.

c) Số trung bình của mẫu số liệu ghép nhóm trạm quan trắc ở Quy Nhơn bằng 242,5 .

d) Nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn Nha Trang..

Lời giải

A.	B.	C.	D.
SAI	ĐÚNG	SAI	ĐÚNG

a) Cỡ mẫu $n = 20$.

• Xét mẫu số liệu của trạm quan trắc ở Nha Trang:

Gọi $x_1; \dots; x_{20}$ là mẫu số liệu gốc về tổng số giờ nắng trong tháng 6 của các năm 2002 đến 2021 tại trạm quan trắc đặt ở Nha Trang được xếp theo thứ tự không giảm.

Ta có

$$x_1 \in [130; 160),$$

$$x_2 \in [160; 190),$$

$$x_3 \in [190; 220),$$

$$x_4; \dots; x_{11} \in [220; 250),$$

$$x_{12}; \dots; x_{18} \in [250; 280),$$

$$x_{19}; x_{20} \in [280; 310).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [220; 250)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1 = 220 + \frac{\frac{20}{4} - (1+1+1)}{8} (250 - 220) = 227,5$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [250; 280)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3 = 250 + \frac{\frac{3.20}{4} - (1+1+1+8)}{7} (280 - 250) = \frac{1870}{7}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = \frac{1870}{7} - 227,5 \approx 39,64$$

- Xét mẫu số liệu của trạm quan trắc ở Quy Nhơn:

Gọi $y_1; \dots; y_{20}$ là mẫu số liệu gốc về tổng số giờ nắng trong tháng 6 của các năm 2022 đến 2021 tại trạm quan trắc đặt ở Quy Nhơn được xếp theo thứ tự không giảm.

Ta có

$$y_1 \in [160; 190),$$

$$y_2; y_3 \in [190; 220),$$

$$y_4; \dots; y_7 \in [220; 250),$$

$$y_8; \dots; y_{17} \in [250; 280),$$

$$y_{18}; \dots; y_{20} \in [280; 310).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{y_5 + y_6}{2} \in [220; 250)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q'_1 = 200 + \frac{\frac{20}{4} - (1+2)}{4} (250 - 200) = 235$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{y_{15} + y_{16}}{2} \in [250; 280)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q'_3 = 250 + \frac{\frac{3.20}{4} - (1+2+4)}{10} (280 - 250) = 274$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta'_Q = Q'_3 - Q'_1 = 274 - 235 = 39.$$

b) Vì $\Delta_Q \approx 39,64 > \Delta'_Q = 39$ nên nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn Nha Trang.

c) Ta có bảng sau:

Số giờ nắng	[130; 160)	[160; 190)	[190; 220)	[220; 250)	[250; 280)	[280; 310)
Giá trị đại diện	145	175	205	235	265	295
Số năm ở Nha Trang	1	1	1	8	7	2
Số năm ở Quy Nhơn	0	1	2	4	10	3

- Xét mẫu số liệu của trạm quan trắc ở Nha Trang:

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_N = \frac{1.145 + 1.175 + 1.205 + 8.235 + 7.265 + 2.295}{20} = 242,5.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S_N^2 = \frac{1}{20} [1.145^2 + 1.175^2 + 1.205^2 + 8.235^2 + 7.265^2 + 2.295^2] - (242,5)^2 = 1248,75$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S_N = \sqrt{S_N^2} = \sqrt{1248,75} \approx 35,34.$

- Xét mẫu số liệu của trạm quan trắc ở Quy Nhơn:

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_Q = \frac{0.145 + 2.175 + 4.205 + 4.235 + 10.265 + 3.295}{20} = 253.$$

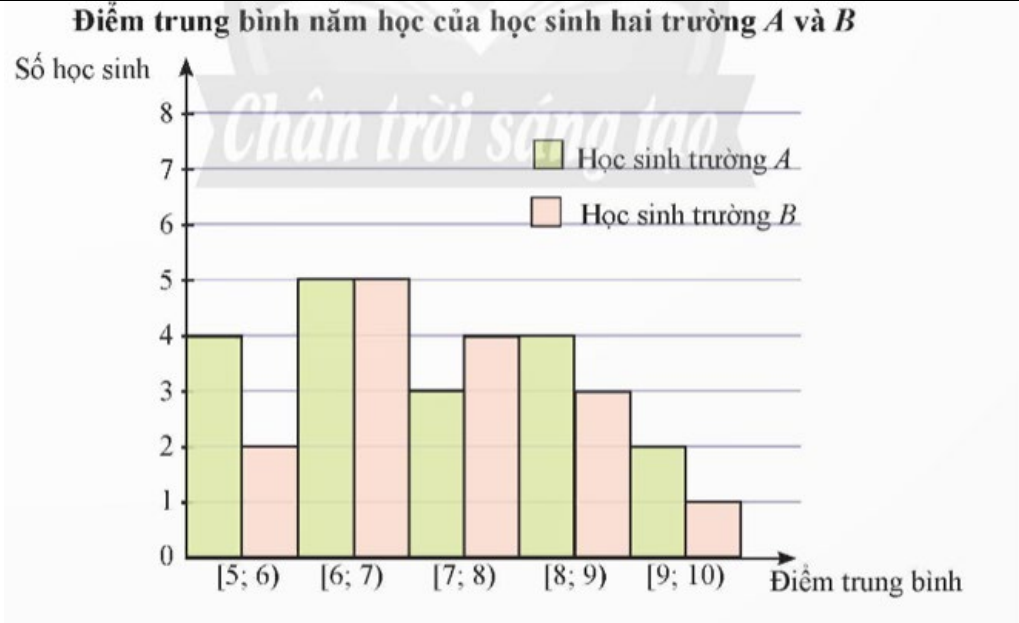
Phương sai của mẫu số liệu ghép nhóm là:

$$S_Q^2 = \frac{1}{20} [0.145^2 + 2.175^2 + 4.205^2 + 4.235^2 + 10.265^2 + 3.295^2] - (253)^2 = 936$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S_Q = \sqrt{S_Q^2} = \sqrt{936} \approx 30,59.$

d) Vì $S_N \approx 35,34 > S_Q \approx 30,59$ nên nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn.

Câu 9. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



a) Giá trị đại diện cho mỗi nhóm và bảng tần số ghép nhóm của mẫu số liệu trên là:

Điểm trung bình	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Giá trị đại diện	5,5	6,5	7,5	8,5	9,5
Số học sinh trường A	4	5	3	4	2
Số học sinh trường B	2	5	4	3	1

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm của trường A là 2,275.

c) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường A có điểm trung bình đồng đều hơn trường B.

d) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn.

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Giá trị đại diện của nhóm [5; 6) là 5,5.

Giá trị đại diện của nhóm [6; 7) là 6,5.

Giá trị đại diện của nhóm [7; 8) là 7,5.

Giá trị đại diện của nhóm [8; 9) là 8,5.

Giá trị đại diện của nhóm [9; 10) là 9,5.

Từ biểu đồ, ta có bảng tần số ghép nhóm sau:

Điểm trung bình	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Giá trị đại diện	5,5	6,5	7,5	8,5	9,5
Số học sinh trường A	4	5	3	4	2
Số học sinh trường B	2	5	4	3	1

b)

- Xét mẫu số liệu của trường A:

$$\text{Cỡ mẫu } n_A = 4 + 5 + 3 + 4 + 2 = 18.$$

Gọi $x_1; \dots; x_{18}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh trường A được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_4 \in [5; 6),$$

$$x_5; \dots; x_9 \in [6; 7),$$

$$x_{10}; \dots; x_{12} \in [7; 8),$$

$$x_{13}; \dots; x_{16} \in [8; 9),$$

$$x_{17}; x_{18} \in [9; 10).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_5 \in [6; 7)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép

$$\text{nhóm là: } Q_1 = 6 + \frac{\frac{18}{4} - 4}{5} (7 - 6) = 6,1$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{14} \in [8; 9)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép

$$\text{nhóm là: } Q_3 = 8 + \frac{\frac{3 \cdot 18}{4} - (4 + 5 + 3)}{4} (9 - 8) = 8,375$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 8,375 - 6,1 = 2,275$.

- Xét mẫu số liệu của trường B:

$$\text{Cỡ mẫu } n_B = 2 + 5 + 4 + 3 + 1 = 15.$$

Gọi $y_1; \dots; y_{20}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh trường B được xếp theo thứ tự không giảm.

Ta có

$$y_1; y_2 \in [5; 6),$$

$$y_3; \dots; y_7 \in [6; 7),$$

$$y_8; \dots; y_{11} \in [7; 8),$$

$$y_{12}; \dots; y_{14} \in [8; 9),$$

$$y_{15} \in [9; 10).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $y_4 \in [6; 7)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép

$$\text{nhóm là: } Q'_1 = 6 + \frac{\frac{15}{4} - 2}{5} (7 - 6) = 6,35$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $y_{12} \in [8; 9)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép

$$\text{nhóm là: } Q'_3 = 8 + \frac{\frac{3.15}{4} - (2+5+4)}{3} (9-8) = \frac{97}{12}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta'_Q = Q'_3 - Q'_1 = \frac{97}{12} - 6,35 \approx 1,73$$

c) Vì $\Delta_Q = 2,275 > \Delta'_Q \approx 1,73$ nên nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn.

d)

•□ Xét mẫu số liệu của trường A:

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_A = \frac{4.5,5 + 5.6,5 + 3.7,5 + 4.8,5 + 2.9,5}{18} = \frac{65}{9}.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S_A^2 = \frac{1}{18} \left[4.(5,5)^2 + 5.(6,5)^2 + 3.(7,5)^2 + 4.(8,5)^2 + 2.(9,5)^2 \right] - \left(\frac{65}{9} \right)^2 = \frac{569}{324}$$

$$\text{Độ lệch chuẩn của mẫu số liệu ghép nhóm là: } S_A = \sqrt{S_A^2} = \sqrt{\frac{569}{324}} \approx 1,33.$$

•□ Xét mẫu số liệu của trường B:

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x}_B = \frac{2.5,5 + 5.6,5 + 4.7,5 + 3.8,5 + 1.9,5}{15} = \frac{217}{30}.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S_B^2 = \frac{1}{15} \left[2.(5,5)^2 + 5.(6,5)^2 + 4.(7,5)^2 + 3.(8,5)^2 + 1.(9,5)^2 \right] - \left(\frac{217}{30} \right)^2 = \frac{284}{225}$$

$$\text{Độ lệch chuẩn của mẫu số liệu ghép nhóm là: } S_B = \sqrt{S_B^2} = \sqrt{\frac{284}{225}} \approx 1,12.$$

Vì $S_A \approx 1,33 > S_B \approx 1,12$ nên nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn.

PHẦN III. Câu trắc nghiệm trả lời ngắn. Mỗi câu hỏi thí sinh chỉ trả lời đáp án.

Câu 10. Thời gian hoàn thành bài kiểm tra môn Toán của các bạn trong lớp 12C được cho trong bảng sau:

Thời gian (phút)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Số học sinh	8	16	4	2

- a) Tính khoảng biến thiên R cho mẫu số liệu ghép nhóm trên.
 b) Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 27 phút và muộn nhất mất 43 phút thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

Trả lời:

Lời giải**Đáp án:**

- a) 20
 b) 16

- a) Khoảng biến thiên R cho mẫu số liệu ghép nhóm là $R = 45 - 25 = 20$.
 b) Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 27 phút và muộn nhất mất 43 phút thì khoảng biến thiên của mẫu số liệu gốc là $43 - 27 = 16$.

Câu 11. Để đánh giá chất lượng một loại pin điện thoại mới, người ta ghi lại thời gian nghe nhạc liên tục của điện thoại được sạc đầy pin cho đến khi hết pin cho kết quả sau:

Thời gian (giờ)	[5; 5,5)	[5,5; 6)	[6; 6,5)	[6,5; 7)	[7; 7,5)
Số chiếc điện thoại (tần số)	2	8	15	10	5

Tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.

Trả lời:

Lời giải**Đáp án:**

- Khoảng biến thiên: $R = 7,5 - 5 = 2,5$.
- Khoảng tứ phân vị $\Delta_Q = 0,75$
- Độ lệch chuẩn là $S \approx 0,53$

Khoảng biến thiên: $R = 7,5 - 5 = 2,5$.

Cỡ mẫu là $n = 2 + 8 + 15 + 10 + 5 = 40$.

Gọi $x_1; x_2; \dots; x_{40}$ thời gian nghe nhạc liên tục của điện thoại được sạc đầy pin cho đến khi hết pin và được sắp xếp theo thứ tự tăng dần.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{10} + x_{11}}{2}$.

Mà $x_{10} \in [5,5; 6)$; $x_{11} \in [6; 6,5)$. Do đó $Q_1 = 6$.

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{30} + x_{31}}{2} \in [6,5; 7)$ nên nhóm chứa tứ phân vị thứ ba là $[6,5; 7)$.

$$Q_3 = 6,5 + \frac{\frac{3.40}{4} - 25}{10} (7 - 6,5) = 6,75$$

Khoảng tứ phân vị $\Delta_Q = Q_3 - Q_1 = 6,75 - 6 = 0,75$.

Chọn giá trị đại diện cho mẫu số liệu ta có

Thời gian (giờ)	[5; 5,5)	[5,5; 6)	[6; 6,5)	[6,5; 7)	[7; 7,5)
Giá trị đại diện	5,25	5,75	6,25	6,75	7,25
Số chiếc điện thoại (tần số)	2	8	15	10	5

Thời gian trung bình là $\bar{x} = 6,35$

Phương sai là: $S^2 = 0,2775$

Độ lệch chuẩn là $S = \sqrt{S^2} = \sqrt{0,2775} \approx 0,53$

Câu 12. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư (với số tiền đầu tư như nhau), khi đầu tư vào hai lĩnh vực A, B cho kết quả như sau:

Tiền lãi	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
Số nhà đầu tư vào lĩnh vực A	2	5	8	6	4
Số nhà đầu tư vào lĩnh vực B	8	4	2	5	6

- a) Về trung bình, đầu tư vào lĩnh vực nào đem lại tiền lãi cao hơn?
- b) Tính độ lệch chuẩn cho các mẫu số liệu về tiền lãi của các nhà đầu tư ở hai lĩnh vực này và giải thích ý nghĩa của các số thu được.

Trả lời:

Lời giải

Đáp án:

a) Về trung bình, nên đầu tư vào lĩnh vực A thì đem lại lãi cao hơn lĩnh vực B.

b) Độ lệch chuẩn mẫu số liệu lĩnh vực A: $S_A \approx 5,83$

Độ lệch chuẩn là mẫu số liệu lĩnh vực B: $S_B \approx 8,04$

Giải thích : Dựa vào độ lệch chuẩn, ta thấy rằng tiền lãi của các nhà đầu tư trong lĩnh vực B có sự biến động lớn hơn và có xu hướng phân tán rộng hơn so với tiền lãi của các nhà đầu tư trong lĩnh vực A.

a) Chọn giá trị đại diện cho mẫu số liệu ta có:

Tiền lãi	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
Giá trị đại diện	7,5	12,5	17,5	22,5	27,5
Số nhà đầu tư vào lĩnh vực A	2	5	8	6	4
Số nhà đầu tư vào lĩnh vực B	8	4	2	5	6

Trung bình tiền lãi đầu tư vào lĩnh vực A là: $\bar{x}_A = 18,5$

Trung bình tiền lãi đầu tư vào lĩnh vực B là: $\bar{x}_B = 16,9$

Vì $\bar{x}_A = 18,5 > \bar{x}_B = 16,9$ nên đầu tư vào lĩnh vực A thì đem lại lãi cao hơn lĩnh vực B.

b)

Phương sai và độ lệch chuẩn của tiền lãi của nhà đầu tư vào lĩnh vực A

$$\text{Phương sai là: } S_A^2 = 34$$

$$\text{Độ lệch chuẩn là } S_A = \sqrt{S_A^2} = \sqrt{34} \approx 5,83$$

Phương sai và độ lệch chuẩn của tiền lãi của nhà đầu tư vào lĩnh vực B

$$\text{Phương sai là: } S_B^2 = 64,64$$

$$\text{Độ lệch chuẩn là } S_B = \sqrt{S_B^2} = \sqrt{64,64} \approx 8,04$$

Dựa vào độ lệch chuẩn, ta thấy rằng tiền lãi của các nhà đầu tư trong lĩnh vực B có sự biến động lớn hơn và có xu hướng phân tán rộng hơn so với tiền lãi của các nhà đầu tư trong lĩnh vực A.

Câu 13. Thành tích môn nhảy cao của các vận động viên tại một giải điền kinh dành cho học sinh trung học phổ thông như sau:

Mức xà (cm)	[170; 172)	[172; 174)	[174; 176)	[176; 180)
Số vận động viên	3	10	6	1

a) Tính các số đặc trưng đo mức độ phân tán của mẫu số liệu ghép nhóm trên.

b) Độ phân tán của mẫu số liệu cho biết điều gì?

Trả lời:

Lời giải

Đáp án:

a)

khoảng tứ phân vị là $\Delta_Q \approx 2,3$

Phương sai $S^2 \approx 2,75$ và độ lệch chuẩn $S \approx 1,66$

b) Dựa vào các số liệu ở câu a, ta thấy mẫu dữ liệu có sự biến động lớn, các giá trị phân tán rộng và không đồng đều. Có sự chênh lệch đáng kể giữa các kết quả của các vận động viên.

a) Khoảng biến thiên của mẫu số liệu là: $R = 180 - 170 = 10$.

Cỡ mẫu là: $n = 3 + 10 + 6 + 1 = 20$.

Gọi $x_1; x_2; \dots; x_{20}$ là mức xà của 20 vận động viên được sắp xếp theo thứ tự tăng dần.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [172; 174)$. Do đó, tứ phân vị thứ nhất của mẫu số

liệu ghép nhóm là: $Q_1 = 172 + \frac{\frac{20}{4} - 3}{10}(174 - 172) = 172,4$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [174; 176)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

ghép nhóm là: $Q_3 = 174 + \frac{\frac{3 \cdot 20}{4} - 13}{6}(176 - 174) \approx 174,7$

Do đó khoảng tứ phân vị là $\Delta_Q = Q_3 - Q_1 \approx 174,7 - 172,4 \approx 2,3$

Chọn giá trị đại diện cho mẫu số liệu ta có

Mức xà (cm)	[170; 172)	[172; 174)	[174; 176)	[176; 180)
Giá trị đại diện	171	173	175	178
Số vận động viên	3	10	6	1

Mức xà trung bình là: $\bar{x} = 173,55$

Phương sai và độ lệch chuẩn

$$S^2 \approx 2,75 \Rightarrow S = \sqrt{S^2} = \sqrt{2,75} \approx 1,66$$

b) Dựa vào các số liệu ở câu a, ta thấy mẫu dữ liệu có sự biến động lớn, các giá trị phân tán rộng và không đồng đều. Có sự chênh lệch đáng kể giữa các kết quả của các vận động viên.

Câu 14. Trong thực hành đo hiệu điện thế của mạch điện, An và Bình đã dùng hai vôn kế khác nhau để đo, mỗi bạn tiến hành đo 10 lần và cho kết quả như sau:

Hiệu điện thế đo được (Vôn)	[3,85; 3,90)	[3,90; 3,95)	[3,95; 4,00)	[4,00; 4,05)
Số lần An đo	1	6	2	1
Số lần Bình đo	1	3	4	2

Tính độ lệch chuẩn của các mẫu số liệu ghép nhóm cho kết quả đo của An và Bình. Từ đó kết luận xem vôn kế của bạn nào cho kết quả đo ổn định hơn.

Trả lời:

Lời giải

Đáp án:

Độ lệch chuẩn về mẫu số liệu ghép nhóm của An đo là: $S_A \approx 0,039$

Độ lệch chuẩn về mẫu số liệu ghép nhóm của Bình đo là $S_B = 0,045$

Dựa vào kết quả tính được của độ lệch chuẩn, ta thấy vốn kế của An cho kết quả ổn định hơn vốn kế của Bình.

Chọn giá trị đại diện cho mẫu số liệu ta có:

Hiệu điện thế đo được (Vôn)	[3,85; 3,90)	[3,90; 3,95)	[3,95; 4,00)	[4,00; 4,05)
Giá trị đại diện	3,875	3,925	3,975	4,025
Số lần An đo	1	6	2	1
Số lần Bình đo	1	3	4	2

Hiệu điện thế trung bình của An đo là: $\bar{x}_A = 3,94$

Hiệu điện thế trung bình của Bình đo là: $\bar{x}_B = 3,96$

Phương sai và độ lệch chuẩn về mẫu số liệu ghép nhóm của An đo là:

$$S_A^2 = 1,525 \cdot 10^{-3} \Rightarrow S_A = \sqrt{S_A^2} = \sqrt{1,525 \cdot 10^{-3}} \approx 0,039$$

Phương sai và độ lệch chuẩn về mẫu số liệu ghép nhóm của Bình đo là

$$S_B^2 = 2,025 \cdot 10^{-3} \Rightarrow S_B = \sqrt{S_B^2} = \sqrt{2,025 \cdot 10^{-3}} = 0,045$$

Dựa vào kết quả tính được của độ lệch chuẩn, ta thấy vốn kế của An cho kết quả ổn định hơn vốn kế của Bình.

Câu 15. Một người ghi lại thời gian đàm thoại của một số cuộc gọi cho kết quả như bảng sau:

Thời gian t (phút)	Số cuộc gọi
$0 \leq t < 1$	8
$1 \leq t < 2$	17
$2 \leq t < 3$	25
$3 \leq t < 4$	20
$4 \leq t < 5$	10

Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

Trả lời:

Lời giải

Đáp án: Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 1,8$

Ta có bảng mẫu số liệu ghép nhóm được viết lại như sau

Thời gian t (phút)	[0;1)	[1; 2)	[2; 3)	[3; 4)	[4; 5)
----------------------	-------	--------	--------	--------	--------

Số cuộc gọi	8	17	25	20	10
-------------	---	----	----	----	----

Có cỡ mẫu $n = 8 + 17 + 25 + 20 + 10 = 80$.

Giả sử $x_1; x_2; \dots; x_{80}$ là thời gian đàm thoại của 80 cuộc gọi được sắp xếp theo thứ tự tăng dần.

Ta có tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{20} + x_{21}}{2} \in [1; 2)$ nên nhóm chứa tứ phân vị thứ nhất là $[1; 2)$.

$$Q_1 = 1 + \frac{\frac{80}{4} - 8}{17}(2 - 1) \approx 1,7$$

Tứ phân vị thứ ba của mẫu số liệu là $\frac{x_{60} + x_{61}}{2} \in [3; 4)$ nên nhóm chứa tứ phân vị thứ ba là $[3; 4)$.

$$Q_3 = 3 + \frac{\frac{3 \cdot 80}{4} - 50}{20}(4 - 3) = 3,5$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 1,8$