

CHƯƠNG III. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM**BÀI 1. KHOẢNG BIẾN THIÊN KHOẢNG TƯ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM. 2**

A. KIẾN THỨC CƠ BẢN CẦN NẮM.....	2
B. GIẢI BÀI TẬP SÁCH GIÁO KHOA	5
C. CÁC DẠNG TOÁN	6
Dạng 1: Tìm khoảng biến thiên	6
Dạng 2. Xác định khoảng tứ phân vị của mẫu số liệu ghép nhóm.....	7
D. TRẮC NGHIỆM 4 PHƯƠNG ÁN	9
E. TRẮC NGHIỆM ĐÚNG SAI.....	13
F. TRẢ LỜI NGẮN	17

CHƯƠNG III. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM

BÀI 1. KHOẢNG BIẾN THIÊN KHOẢNG TỬ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM

A. KIẾN THỨC CƠ BẢN CẦN NẮM

1. KHOẢNG BIẾN THIÊN

Khoảng biến thiên, kí hiệu R , của mẫu số liệu ghép nhóm là hiệu số giữa đầu mút phải của nhóm cuối cùng và đầu mút trái của nhóm đầu tiên có chứa dữ liệu của mẫu số liệu.

Chú ý:

Xét mẫu số liệu ghép nhóm được cho ở bảng sau:

Bảng 1

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Tần số	n_1	n_2	...	n_k

Nếu n_1 và n_{k+1} cùng khác 0 thì $R = u_{k+1} - u_1$.

Khoảng biến thiên của mẫu số liệu ghép nhóm luôn lớn hơn hoặc bằng khoảng biến thiên của mẫu số liệu gốc.

Khoảng biến thiên của mẫu số liệu ghép nhóm là giá trị xấp xỉ khoảng biến thiên của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của mẫu số liệu.

Khoảng biến thiên $R = u_{k+1} - u_1$ chưa phản ánh được đầy đủ mức độ phân tán của phần lớn các số liệu.

Hơn nữa, giá trị của R thường tăng vọt khi xuất hiện giá trị ngoại lệ trong mẫu số liệu. Do đó, để phản ánh mức độ phân tán của số liệu, người ta còn dùng các số đặc trưng khác.

Ví dụ 1: Cô Hà thống kê lại đường kính thân gỗ của một số cây xoan đào 6 năm tuổi được trồng ở một lâm trường ở bảng sau.

Đường kính (cm)	[40; 45)	[45; 50)	[50; 55)	[55; 60)	[60; 65)
Tần số	5	20	18	7	3

Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm trên.

Ý nghĩa của khoảng biến thiên của mẫu số liệu ghép nhóm

- Khoảng biến thiên của mẫu số liệu ghép nhóm là giá trị xấp xỉ khoảng biến thiên của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của mẫu số liệu.

- Khoảng biến thiên $R = u_{k+1} - u_1$ chưa phản ánh được đầy đủ mức độ phân tán của phần lớn các số liệu.

Hơn nữa, giá trị của R thường tăng vọt khi xuất hiện giá trị ngoại lệ trong mẫu số liệu. Do đó, để phản ánh mức độ phân tán của số liệu, người ta còn dùng các số đặc trưng khác.

Ví dụ 2: Sử dụng dữ liệu ở biểu đồ trong , chọn số thích hợp thay vào các vị trí được đánh dấu? ở bảng sau:

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Bác Bình	?	12	8	3	2
Bác An	?	?	?	?	?

a) Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

b) Sử dụng khoảng biến thiên, hãy cho biết bác nào có thời gian tập phân tán hơn.

2.KHOẢNG TỨ PHÂN VỊ

Bảng 1

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Tần số	n_1	n_2	...	n_k

Tứ phân vị thứ i , kí hiệu là Q_i , với $i = 1, 2, 3$ của mẫu số liệu ghép nhóm (Bảng 1) được xác định như sau:

$$Q_i = u_m + \frac{\frac{in}{4} - C}{n_m} (u_{m+1} - u_m),$$

Trong đó:

- $n = n_1 + n_2 + \dots + n_k$ là cỡ mẫu;
- $[u_m; u_{m+1})$ là nhóm chứa tứ phân vị thứ i ;
- n_m là tần số của nhóm chứa tứ phân vị thứ i ;
- $C = n_1 + n_2 + \dots + n_{m-1}$.

Khoảng tứ phân vị của mẫu số liệu ghép nhóm cũng được xác định dựa trên tứ phân vị thứ nhất và tứ phân vị thứ ba như đối với mẫu số liệu không ghép nhóm.

Khoảng tứ phân vị của mẫu số liệu ghép nhóm cho ở Bảng 1, 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:

$$\Delta_Q = Q_3 - Q_1.$$

Ví dụ 3: Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trong hoạt động 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

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

- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho khoảng tứ phân vị của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của nửa giữa của mẫu số liệu (tập hợp gồm 50% số liệu nằm chính giữa mẫu số liệu).
- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm càng nhỏ thì dữ liệu càng tập trung xung quanh trung vị.
- ✓ Khoảng tứ phân vị được dùng để xác định giá trị ngoại lệ trong mẫu số liệu. Giá trị x trong mẫu số liệu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.
- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm không bị ảnh hưởng nhiều bởi các giá trị ngoại lệ trong mẫu số liệu.

Ví dụ 4: Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

- a) Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.
- b) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần ông Thắng đi hết hơn 29 phút. Thời gian của lần đi đó có phải là giá trị ngoại lệ không?

Trả lời câu hỏi Thực hành 3 trang 72 SGK Toán 12 Chân trời sáng tạo

- a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm ở Ví dụ 4 sau khi đã loại bỏ các giá trị ngoại lệ. Em có nhận xét gì về khoảng biến thiên, khoảng tứ phân vị vừa tìm được và khoảng biến thiên, khoảng tứ phân vị ban đầu?

Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

- b) Hãy so sánh độ phân tán của nửa giữa hai mẫu số liệu chiều cao của các học sinh nữ lớp 12C và 12D ở Thực hành 1.

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

Trả lời câu hỏi Vận dụng trang 73 SGK Toán 12 Chân trời sáng tạo

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:

- 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?

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

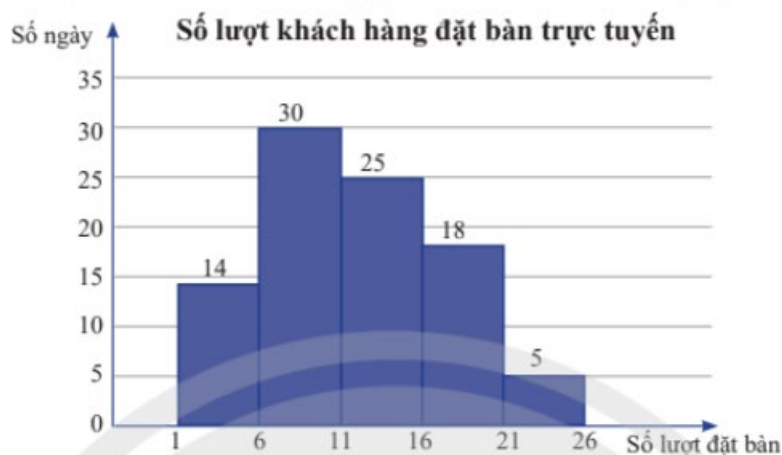
Bài 1: Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

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

(Nguồn: Tổng cục Thống kê)

- 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).

Bài 2: Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 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.

Bài 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?

Bài 4: Bảng tần số ghép nhóm dưới đây thể hiện kết quả điều tra về tuổi thọ trung bình của nam giới và nữ giới ở 50 quốc gia.

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 thọ trung bình của nam giới và nữ giới trong mẫu số liệu ghép nhóm trên.
- b) Hãy cho biết tuổi thọ trung bình của nam giới hay nữ giới trong mẫu số liệu ghép nhóm trên đồng đều hơn.

C. CÁC DẠNG TOÁN

Dạng 1: Tìm khoảng biến thiên

1. Phương pháp

Cho mẫu số liệu ghép nhóm:

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
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 ghép nhóm trên là $R = a_{k+1} - a_1$.

2. Ví dụ

Câu 1. 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?

Câu 2. Bảng dưới biểu thị kết quả điều tra thời gian sử dụng Internet hằng ngày của một số người.

Thời gian (phút)	[30;60)	[60;90)	[90;120)	[120;150)	[150;180)
Số người	2	4	10	5	3

Tìm khoảng biến thiên của mẫu số liệu đã cho. Kết quả cho biết điều gì?

Câu 3. Bảng dưới thống kê thành tích nhảy xa của một số học sinh lớp 12. Tìm khoảng biến thiên thành tích nhảy xa của số học sinh này.

Thành tích (cm)	[150;180)	[180;210)	[210;240)	[240;270)	[270;300)
Số học sinh	3	5	28	14	8

Câu 4. Để chuẩn bị mở một trung tâm thể dục thể thao, anh Dũng đã tiến hành điều tra tuổi thọ của máy chạy bộ do hai hãng X, Y sản xuất. Bảng dưới biểu thị hai mẫu số liệu mà anh thu thập được qua Internet.

Bảng. Tuổi thọ của máy chạy bộ (đơn vị: năm)

Tuổi thọ	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
Số máy của hãng X	7	20	36	20	17
Số máy của hãng Y	0	20	35	35	10

Khoảng biến thiên của mẫu số liệu nào lớn hơn? Từ đó có thể nói là máy chạy bộ do hãng nào sản xuất có tuổi thọ phân tán hơn?

Câu 5. Người ta tiến hành phỏng vấn hai nhóm khán giả về một bộ phim mới công chiếu. Nhóm A gồm những khán giả thuộc lứa tuổi 20 - 30, nhóm B thuộc lứa tuổi trên 30. Người được hỏi ý kiến phải đánh giá bộ phim bằng cách cho điểm theo một số tiêu chí nêu trong phiếu điều tra và sau đó lấy tổng số điểm (thang điểm 100). Bảng dưới đây trình bày kết quả điều tra hai nhóm khán giả:

Bảng. Điểm đánh giá của khán giả

Điểm	[50; 60)	[60; 70)	[70; 80)	[80; 90)	[90; 100)
Số người của nhóm A	6	10	14	12	8
Số người của nhóm B	0	8	14	28	0

Ý kiến đánh giá của nhóm khán giả nào phân tán hơn?

Câu 6. Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

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

Dạng 2. Xác định khoảng tứ phân vị của mẫu số liệu ghép nhóm

1. Phương pháp

Tứ phân vị thứ r là

$$Q_r = a_p + \frac{\frac{r \cdot n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p),$$

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

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

2. Ví dụ

Câu 1. 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

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

Câu 2. Thống kê số ngày trong tháng Sáu năm 2021 và năm 2022 theo nhiệt độ cao nhất trong ngày tại Hà Nội, người ta thu được bảng sau:

Nhiệt độ ($^{\circ}\text{C}$)	[28;30)	[30;32)	[32;34)	[34;36)	[36;38)	[38;40)
Số ngày trong tháng 6/2021	0	2	8	5	6	9
Số ngày trong tháng 6/2022	2	3	4	11	8	2

Hỏi tháng Sáu năm nào ở Hà Nội nhiệt độ cao nhất trong ngày biến đổi nhiều hơn?

Câu 3. Thống kê số thẻ vàng của mỗi câu lạc bộ trong giải ngoại hạng Anh mùa giải 2021 - 2022 cho kết quả như sau:

101 79 79 78 75 73 68 67 67 63
63 61 60 59 57 55 55 50 47 42.

(Theo premierleague.com)

a) Hãy ghép nhóm dãy số liệu trên thành các nhóm có độ dài bằng nhau với nhóm đầu tiên là [40;50).

b) Tính khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu gốc và mẫu số liệu ghép nhóm thu được ở câu a. Giá trị nào là giá trị chính xác? Giá trị nào là giá trị xấp xỉ?

Câu 4. Thu nhập theo tháng (đơn vị: triệu đồng) của người lao động ở hai nhà máy như sau:

Thu nhập	[5;8)	[8;11)	[11;14)	[14;17)	[17;20)
Số người của nhà máy A	20	35	45	35	20
Số người của nhà máy B	17	23	30	23	17

Tính mức thu nhập trung bình của người lao động ở hai nhà máy trên. Dựa vào khoảng tứ phân vị, hãy xác định xem mức thu nhập của người lao động ở nhà máy nào biến động nhiều hơn.

Câu 5. Bảng sau đây cho biết chiều cao của các học sinh lớp 12A và 12B.

Chiều cao (cm)	[145; 150)	[150; 155)	[155; 160)	[160; 165)	[165; 170)	[170; 175)
Số học sinh của lớp 12A	1	0	15	12	10	5
Số học sinh của lớp 12B	0	0	17	10	9	6

a) Tìm khoảng biến thiên, khoảng tứ phân vị cho các mẫu số liệu ghép nhóm về chiều cao của học sinh lớp 12A, 12B.

b) Để so sánh độ phân tán về chiều cao của học sinh hai lớp này ta nên dùng khoảng biến thiên hay khoảng tứ phân vị? Vì sao?

Câu 6. Điểm kiểm tra cuối khoá môn Tiếng Anh của hai lớp ở một trung tâm ngoại ngữ được thống kê trong các Bảng a và b.

Bảng a. Điểm của lớp A

Điểm	Số học viên (tần số)
[50; 60)	8
[60; 70)	20
[70; 80)	50
[80; 90)	17
[90; 100)	5

Bảng b. Điểm của lớp B

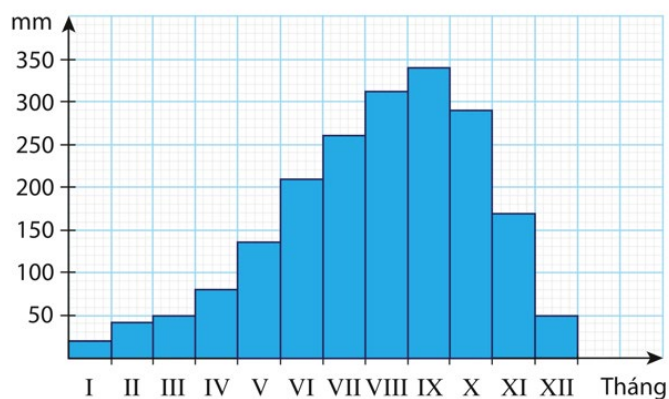
Điểm	Số học viên (tần số)
[50; 60)	15
[60; 70)	20
[70; 80)	30
[80; 90)	20
[90; 100)	15

a) Tìm khoảng biến thiên của mỗi mẫu số liệu. Có thể dùng khoảng biến thiên để biết điểm của lớp nào đồng đều hơn không?

b) Tìm các tứ phân vị và khoảng tứ phân vị của mỗi mẫu số liệu.

c) Mẫu số liệu nào có độ phân tán lớn hơn? Minh hoạ câu trả lời bằng cách biểu diễn các tứ phân vị và khoảng tứ phân vị của mỗi mẫu số liệu trên trục số.

Câu 7. Hình dưới là biểu đồ biểu diễn lượng mưa trung bình của các tháng trong năm ở thành phố #A.



Hình 3.2. Biểu đồ lượng mưa ở thành phố A

a) Lập bảng số liệu ghép nhóm về lượng mưa của thành phố A, với độ dài các nhóm là 50 và đầu mút phải của nhóm cuối cùng là 350.

b) Xác định khoảng tứ phân vị của mẫu số liệu (làm tròn kết quả đến hàng đơn vị). Nêu ý nghĩa của kết quả tìm được.

D. TRẮC NGHIỆM 4 PHƯƠNG ÁN

Câu 1. Cô Hà thống kê lại đường kính thân gỗ của một số cây xoan đào 6 năm tuổi được trồng ở một lâm trường ở bảng sau.

Đường kính (cm)	[40; 45)	[45; 50)	[50; 55)	[55; 60)	[60; 65)
Tần số	5	20	18	7	3

Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm trên.

A. 25.

B. 30.

C. 6.

D. 69,8.

Câu 2. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 60 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	Tần số
[40 ; 50)	3
[50 ; 60)	6
[60 ; 70)	19
[70 ; 80)	23
[80 ; 90)	9
	$n = 60$

Bảng 8

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là:

A. 50.

B. 30.

C. 6.

D. 69,8.

Câu 3. Cho mẫu số liệu ghép nhóm có tứ phân vị thứ nhất, thứ hai, thứ ba lần lượt là Q_1, Q_2, Q_3 . Khoảng tứ phân vị của mẫu số liệu ghép nhóm đó bằng:

A. $2Q_2$.B. $Q_1 - Q_3$.C. $Q_3 - Q_1$.D. $Q_3 + Q_1 - Q_2$.

Câu 4. 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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

A. 1,5.

B. 0,9.

C. 0,6.

D. 0,3.

Câu 5. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

A. 25.

B. 20.

C. 15.

D. 30.

Câu 6. 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

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.

Câu 7. 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

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 3.

B. 4.

C. 5.

D. 6.

Câu 8. 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

Nhóm chứa tứ phân vị thứ nhất là

A. [14;15).

B. [15;16).

C. [16;17).

D. [17;18).

Câu 9. 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

Nhóm chứa tứ phân vị thứ ba là

A. [15;16).

B. [16;17).

C. [17;18).

D. [18;19).

Câu 10. Khảo sát thời gian tập thể dục trong ngày của một số học sinh khối 11 thu được mẫu số liệu ghép nhóm sau:

Thời gian (phút)	[0;20)	[20;40)	[40;60)	[60;80)	[80;100)
Số học sinh	5	9	12	10	6

Nhóm chứa tứ phân vị thứ nhất là

A. [0;20).

B. [20;40).

C. [40;60).

D. [60;80).

Câu 11. 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

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.

Câu 12. 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

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.

Câu 13. Cho mẫu số liệu ghép nhóm về tuổi thọ (đơn vị tính là năm) của một loại bóng đèn mới như sau.

Tuổi thọ	[2;3,5)	[3,5;5)	[5;6,5)	[6,5;8)
Số bóng đèn	8	22	35	15

Nhóm chứa tứ phân vị thứ nhất của mẫu số liệu là

A. [2;3,5).

B. [3,5;5).

C. [5;6,5).

D. [6,5;8).

Câu 14. Cho mẫu số liệu ghép nhóm về tuổi thọ (đơn vị tính là năm) của một loại bóng đèn mới như sau.

Tuổi thọ	[2;3,5)	[3,5;5)	[5;6,5)	[6,5;8)
----------	---------	---------	---------	---------

Số bóng đèn	8	22	35	15
-------------	---	----	----	----

Nhóm chứa tứ phân vị thứ ba của mẫu số liệu là

- A. $[2; 3; 5)$. B. $[3; 5; 5)$. C. $[5; 6; 5)$. D. $[6; 5; 8)$.

Câu 15. Cho bảng số liệu về chiều cao của 100 học sinh một trường trung học phổ thông dưới đây.

Nhóm	Chiều cao (cm)	Số học sinh
1	$[150; 153)$	7
2	$[153; 156)$	13
3	$[156; 159)$	40
4	$[159; 162)$	21
5	$[162; 165)$	13
6	$[165; 168)$	6

Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng phần trăm) là

- A. 156,25. B. 157,5. C. 156,38. D. 157,54.

Câu 16. Cho bảng số liệu về chiều cao của 100 học sinh một trường trung học phổ thông dưới đây.

Nhóm	Chiều cao (cm)	Số học sinh
1	$[150; 153)$	7
2	$[153; 156)$	13
3	$[156; 159)$	40
4	$[159; 162)$	21
5	$[162; 165)$	13
6	$[165; 168)$	6

Tứ phân vị thứ ba của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng phần trăm) là:

- A. 160,52. B. 161,52. C. 161,14. D. 162,25.

Câu 17. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 60 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	Tần số
$[40 ; 50)$	3
$[50 ; 60)$	6
$[60 ; 70)$	19
$[70 ; 80)$	23
$[80 ; 90)$	9
	$n = 60$

Bảng 8

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

- A. 50. B. 40. C. 14,23. D. 70,87.

Câu 18. 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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

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

Câu 19. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A.** 23,75. **B.** 27,5. **C.** 31,88. **D.** 8,125.

Câu 20. 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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A.** 10,75. **B.** 1,75. **C.** 3,63. **D.** 14,38.

Câu 21. Thời gian (phút) truy cập Internet mỗi buổi tối của một số học sinh được cho trong bảng sau:

Thời gian (phút)	[9,5;12,5)	[12,5;15,5)	[15,5;18,5)	[18,5;21,5)	[21,5;24,5)
Số học sinh	3	12	15	24	2

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A.** 10,75. **B.** 4,75. **C.** 4,63. **D.** 4,38.

Câu 22. Ghi lại tốc độ bóng trong 200 lần giao bóng của một vận động viên môn quần vợt cho kết quả như bảng bên.

Tốc độ $v(km/h)$	Số lần
$150 \leq v < 155$	18
$155 \leq v < 160$	28
$160 \leq v < 165$	35
$165 \leq v < 170$	43
$170 \leq v < 175$	41
$175 \leq v < 180$	35

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A.** 12,6. **B.** 11,5. **C.** 14,3. **D.** 16,8.

E. TRẮC NGHIỆM ĐÚNG SAI

Câu 1. Bạn Trang thống kê lại chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau.

Chiều cao (cm) \ Lớp	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
12C	2	7	12	3	0	1
12D	5	9	8	2	1	0

Các mệnh đề sau đúng hay sai?

- Chiều cao cao nhất của các bạn học sinh trong lớp 12D là 185 (cm)
- Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12C là: 30 (cm)
- Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12D là: 25 (cm)
- Chiều cao của học sinh lớp 12C có độ phân tán bé hơn

Câu 2. Bảng sau biểu diễn mẫu số liệu ghép nhóm thống kê mức lương của một công ty (đơn vị: triệu đồng).

Nhóm	Tần số
[10;15)	15
[15;20)	18
[20;25)	10
[25;30)	10
[30;35)	5
[35;40)	2
	$n = 60$

Các mệnh đề sau đúng hay sai?

- Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 30$
- Số phần tử của mẫu là $n = 60$
- Tứ phân vị thứ nhất là: $Q_1 = 15$
- Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $\Delta_Q = 3$

Câu 3. Bảng sau biểu diễn mẫu số liệu ghép nhóm về độ tuổi của cư dân trong một khu phố.

Nhóm	Tần số
[20;30)	25
[30;40)	20
[40;50)	20
[50;60)	15
[60;70)	14
[70;80)	6
	$n = 100$

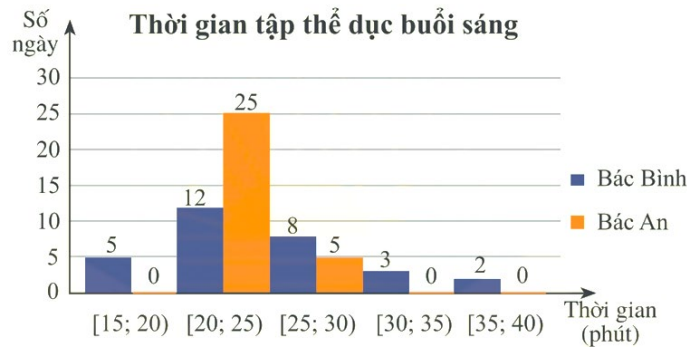
Các mệnh đề sau đúng hay sai?

- Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 60$
- Tứ phân vị thứ nhất là: $Q_1 = 35$

c) Tứ phân vị thứ ba là: $Q_3 = \frac{160}{3}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $\Delta_Q = \frac{65}{3}$

Câu 4. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng 9/2022 của bác Bình và bác An.



Các mệnh đề sau đúng hay sai?

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

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An là: $\Delta_Q = 2$

c) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác Bình là: $Q_3' = \frac{455}{16}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An lớn hơn bác Bình

Câu 5. Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

Các mệnh đề sau đúng hay sai?

a) Cỡ mẫu $n = 100$.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = \frac{683}{38}$.

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta_Q = \frac{515}{114}$

d) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần ông Thắng đi hết hơn 29 phút. Thời gian của lần đi đó là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Câu 6. 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

Các mệnh đề sau đúng hay sai?

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là: 15 (tuổi)
- b) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là: 12 (tuổi)
- c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm ứng với khu vực A là: $\frac{61}{3}$ (tuổi)
- d) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn

Câu 7. 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

Các mệnh đề sau đúng hay sai?

- a) Khoảng biến thiên của mẫu số liệu: 375,9(mm)
- b) Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = \frac{1827}{100}$
- c) Chia mẫu số liệu trên thành 4 nhóm như bảng:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm: 400(mm)

- d) Chia mẫu số liệu trên thành 4 nhóm như bảng:

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = \frac{1000}{7}$

Câu 8. 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

Các mệnh đề sau đúng hay sai?

- a) Khoảng biên thiên của mẫu số liệu ghép nhóm: 1(m)
- b) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm là: $Q_3 = 10,5$
- c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 2,06$
- d) Trong 100 cây keo trên có 1 cây cao 8,4 m . Thì chiều cao của cây keo này là giá trị ngoại lệ

Câu 9. 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
--------	---	----	----	----	----	---	---

Các mệnh đề sau đúng hay sai?

- Khoảng biến thiên của mẫu số liệu là $R = 4$
- Số trung bình của mẫu số liệu xấp xỉ bằng: 8,12
- Mốt của mẫu số liệu là: 6, 21
- Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 2,05$

Câu 10. Một công ty cung cấp nước sạch thống kê lượng nước các hộ gia đình trong một khu vực tiêu thụ trong một tháng ở bảng sau:

Lượng nước tiêu thụ (m^3)	[3;6)	[6;9)	[9;12)	[12;15)	[15;18)
Số hộ gia đình	24	57	42	29	8

Các mệnh đề sau đúng hay sai?

- Số trung bình của mẫu số liệu ghép nhóm đã cho là 9,375.
- Mốt của mẫu số liệu là $M_o = 8,0625$.
- Khoảng biến thiên của mẫu số liệu ghép nhóm đã cho là 15
- Công ty muốn gửi một thông báo khuyến nghị tiết kiệm nước đến 25% các hộ gia đình có lượng nước tiêu thụ cao nhất. Khi đó công ty nên gửi thông báo tiết kiệm nước đến các hộ gia đình có lượng nước tiêu thụ từ $14,79m^3$ nước trở lên.

F. TRẢ LỜI NGẮN

Câu 1. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về chiều cao (đơn vị: centimét) của 36 học sinh nam lớp 12 ở một trường trung học phổ thông. Tìm khoảng biến thiên của mẫu số liệu ghép nhóm đó.

Nhóm	Tần số
[160;163)	6
[163;166)	11
[166;169)	9
[169;172)	7
[172;175)	3
	$n = 36$

Trả lời: 15

Câu 2. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về chiều cao của 42 mẫu cây ở một vườn thực vật (đơn vị: centimét). Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm đó (làm tròn kết quả đến hàng phần mười nếu cần).

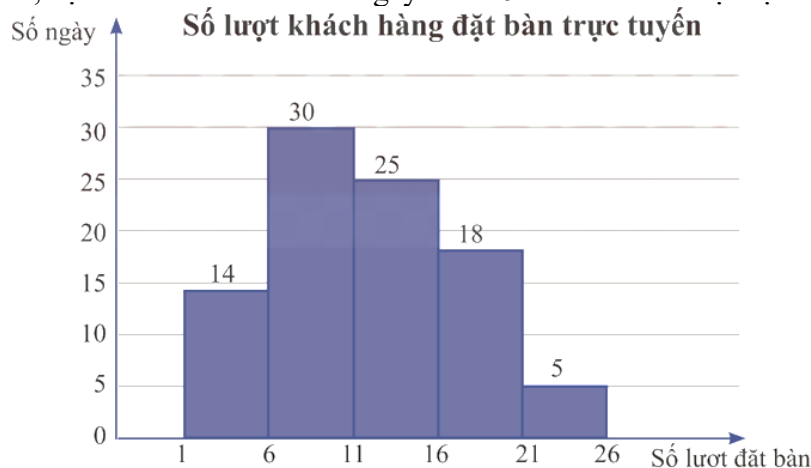
Nhóm	Tần số	Tần số tích lũy
[40 ; 45)	5	5
[45 ; 50)	10	15
[50 ; 55)	7	22
[55 ; 60)	9	31
[60 ; 65)	7	38
[65 ; 70)	4	42
	$n = 42$	

Bảng 7

Câu 3. Bảng sau thống kê cân nặng của 50 quả xoài Thanh Ca được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường. Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm đó

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

Câu 4. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 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 5. Thời gian luyện tập trong một ngày (tính theo giờ) của một số vận động viên được ghi lại ở bảng sau:

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

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

Câu 6. 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 tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

Câu 7. Số điểm một cầu thủ bóng rổ ghi được trong 20 trận đấu được cho ở bảng sau:

Điểm số	[5,5;10,5)	[10,5;15,5)	[15,5;20,5)	[20,5;25,5)
Số trận	3	9	2	6

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

Câu 8. Thời gian sử dụng điện thoại trong một ngày của 30 sinh viên được ghi lại ở bảng sau (đơn vị: phút).

Thời gian (phút)	[0;60)	[60;120)	[120;180)	[180;240)	[240;300)
Số sinh viên	2	7	7	10	4

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

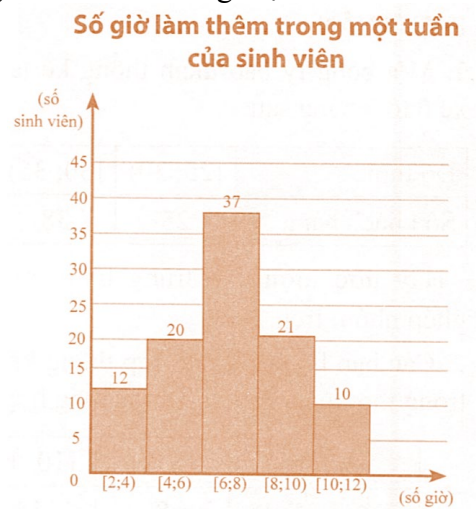
Câu 9. Điểm thi môn Toán (thang điểm 100, điểm được làm tròn đến 1) của 60 thí sinh được cho mẫu số liệu ghép nhóm như sau:

Điểm	[0;9,5)	[9,5;19,5)	[19,5;29,5)	[29,5;39,5)	[39,5;49,5)
------	---------	------------	-------------	-------------	-------------

Số thí sinh	1	2	4	6	15
Điểm	[49,5;59,5)	[59,5;69,5)	[69,5;79,5)	[79,5;89,5)	[89,5;99,5)
Số thí sinh	12	10	6	3	1

Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm

Câu 10. Kết quả điều tra về số giờ làm thêm trong một tuần của 100 sinh viên được cho ở biểu đồ bên.



Tìm khoảng tứ phân vị của số liệu đó.

CHƯƠNG III. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM

BÀI 1. KHOẢNG BIẾN THIÊN KHOẢNG TỬ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM

A. KIẾN THỨC CƠ BẢN CẦN NẮM

1. KHOẢNG BIẾN THIÊN

Khoảng biến thiên, kí hiệu R , của mẫu số liệu ghép nhóm là hiệu số giữa đầu mút phải của nhóm cuối cùng và đầu mút trái của nhóm đầu tiên có chứa dữ liệu của mẫu số liệu.

Chú ý:

Xét mẫu số liệu ghép nhóm được cho ở bảng sau:

Bảng 1

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Tần số	n_1	n_2	...	n_k

Nếu n_1 và n_{k+1} cùng khác 0 thì $R = u_{k+1} - u_1$.

Khoảng biến thiên của mẫu số liệu ghép nhóm luôn lớn hơn hoặc bằng khoảng biến thiên của mẫu số liệu gốc.

Khoảng biến thiên của mẫu số liệu ghép nhóm là giá trị xấp xỉ khoảng biến thiên của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của mẫu số liệu.

Khoảng biến thiên $R = u_{k+1} - u_1$ chưa phản ánh được đầy đủ mức độ phân tán của phần lớn các số liệu. Hơn nữa, giá trị của R thường tăng vọt khi xuất hiện giá trị ngoại lệ trong mẫu số liệu. Do đó, để phản ánh mức độ phân tán của số liệu, người ta còn dùng các số đặc trưng khác.

Ví dụ 1: Cô Hà thống kê lại đường kính thân gỗ của một số cây xoan đào 6 năm tuổi được trồng ở một lâm trường ở bảng sau.

Đường kính (cm)	[40; 45)	[45; 50)	[50; 55)	[55; 60)	[60; 65)
Tần số	5	20	18	7	3

Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm trên.

Lời giải

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là

$$65 - 40 = 25 \text{ (cm)}.$$

Ý nghĩa của khoảng biến thiên của mẫu số liệu ghép nhóm

- Khoảng biến thiên của mẫu số liệu ghép nhóm là giá trị xấp xỉ khoảng biến thiên của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của mẫu số liệu.

- Khoảng biến thiên $R = u_{k+1} - u_1$ chưa phản ánh được đầy đủ mức độ phân tán của phần lớn các số liệu. Hơn nữa, giá trị của R thường tăng vọt khi xuất hiện giá trị ngoại lệ trong mẫu số liệu. Do đó, để phản ánh mức độ phân tán của số liệu, người ta còn dùng các số đặc trưng khác.

Ví dụ 2: Sử dụng dữ liệu ở biểu đồ trong , chọn số thích hợp thay vào các vị trí được đánh dấu? ở bảng sau:

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Bác Bình	?	12	8	3	2
Bác An	?	?	?	?	?

- a) Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An .
- b) Sử dụng khoảng biến thiên, hãy cho biết bác nào có thời gian tập phân tán hơn.

Lời giải

- a) Ta có bảng sau:

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

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).

- b) Nếu căn cứ theo khoảng biến thiên thì bác Bình có thời gian tập phân tán hơn bác An.

2.. KHOẢNG TỨ PHÂN VỊ

Bảng 1

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Tần số	n_1	n_2	...	n_k

Tứ phân vị thứ i , kí hiệu là Q_i , với $i = 1, 2, 3$ của mẫu số liệu ghép nhóm (Bảng 1) được xác định như sau:

$$Q_i = u_m + \frac{\frac{in}{4} - C}{n_m} (u_{m+1} - u_m),$$

Trong đó:

- $n = n_1 + n_2 + \dots + n_k$ là cỡ mẫu;
- $[u_m; u_{m+1})$ là nhóm chứa tứ phân vị thứ i ;
- n_m là tần số của nhóm chứa tứ phân vị thứ i ;
- $C = n_1 + n_2 + \dots + n_{m-1}$.

Khoảng tứ phân vị của mẫu số liệu ghép nhóm cũng được xác định dựa trên tứ phân vị thứ nhất và tứ phân vị thứ ba như đối với mẫu số liệu không ghép nhóm.

Khoảng tứ phân vị của mẫu số liệu ghép nhóm cho ở Bảng 1, 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:

$$\Delta_Q = Q_3 - Q_1.$$

Ví dụ 3: Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trong hoạt động 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

Lời giải

Cỡ mẫu $n = 50$.

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc gồm cân nặng của 50 quả xoài được xếp theo thứ tự không giảm.

Ta có: $x_1, x_2, x_3 \in [250; 290); x_4, \dots, x_{16} \in [290; 330); x_{17}, \dots, x_{34} \in [330; 370);$
 $x_{35}, \dots, x_{45} \in [370; 410); x_{46}, \dots, x_{50} \in [410; 450).$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_{13} \in [290; 330)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 290 + \frac{\frac{50}{4} - 3}{13} \cdot (330 - 290) = \frac{4150}{13}.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{38} \in [370; 410)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 370 + \frac{\frac{3 \cdot 50}{4} - (3 + 13 + 18)}{11} \cdot (410 - 370) = \frac{4210}{11}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là

$$\Delta_Q = \frac{4210}{11} - \frac{4150}{13} = \frac{9080}{143}$$

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

- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho khoảng tứ phân vị của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của nửa giữa của mẫu số liệu (tập hợp gồm 50% số liệu nằm chính giữa mẫu số liệu).
- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm càng nhỏ thì dữ liệu càng tập trung xung quanh trung vị.
- ✓ Khoảng tứ phân vị được dùng để xác định giá trị ngoại lệ trong mẫu số liệu. Giá trị x trong mẫu số liệu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.
- ✓ Khoảng tứ phân vị của mẫu số liệu ghép nhóm không bị ảnh hưởng nhiều bởi các giá trị ngoại lệ trong mẫu số liệu.

Ví dụ 4: Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

- Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.
- Biết rằng trong 100 lần đi trên, chỉ có đúng một lần ông Thắng đi hết hơn 29 phút. Thời gian của lần đi đó có phải là giá trị ngoại lệ không?

Lời giải

a) Cỡ mẫu $n = 100$.

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc gồm thời gian 100 lần đi xe buýt của ông Thắng.

Ta có: $x_1, \dots, x_{22} \in [15; 18); x_{23}, \dots, x_{60} \in [18; 21); x_{61}, \dots, x_{87} \in [21; 24); x_{88}, \dots, x_{95} \in [24; 27);$

$x_{96}, \dots, x_{99} \in [27; 30); x_{100} \in [30; 33).$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{25} + x_{26}) \in [18; 21)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 18 + \frac{\frac{100}{4} - 22}{38} \cdot (21 - 18) = \frac{693}{38}.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [21; 24)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 21 + \frac{\frac{3 \cdot 100}{4} - (22 + 38)}{27} \cdot (24 - 21) = \frac{68}{3}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là

$$\Delta_Q = \frac{68}{3} - \frac{693}{38} = \frac{505}{114}.$$

b) Trong lần duy nhất ông Thắng đi hết hơn 29 phút, thời gian đi của ông thuộc nhóm $[30; 33)$.

Vì $Q_3 + 1,5\Delta_Q = \frac{6683}{228} < 30$ nên thời gian của lần ông Thắng đi hết hơn 29 phút là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Trả lời câu hỏi Thực hành 3 trang 72 SGK Toán 12 Chân trời sáng tạo

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm ở Ví dụ 4 sau khi đã loại bỏ các giá trị ngoại lệ. Em có nhận xét gì về khoảng biến thiên, khoảng tứ phân vị vừa tìm được và khoảng biến thiên, khoảng tứ phân vị ban đầu?

Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

b) Hãy so sánh độ phân tán của nửa giữa hai mẫu số liệu chiều cao của các học sinh nữ lớp 12C và 12D ở Thực hành 1.

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

Lời giải

a) Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc gồm thời gian 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan được xếp theo thứ tự không giảm.

Khoảng biến thiên $R = 33 - 15 = 18$ (phút)

ta có: $x_1; x_2; \dots; x_{22} \in [15; 18); x_{23}; \dots; x_{60} \in [18; 21); x_{61}; \dots; x_{87} \in [21; 24);$

$x_{88}; \dots; x_{95} \in [24; 27); x_{96}; \dots; x_{99} \in [27; 30); x_{100} \in [30; 33)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{25} + x_{26}) \in [18; 21)$. Do đó, tứ phân vị thứ nhất

$$\text{của mẫu số liệu ghép nhóm là: } Q_1 = 18 + \frac{\frac{100}{4} - 22}{38}(21 - 18) = \frac{693}{38}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [21; 24)$. Do đó, tứ phân vị thứ ba của

$$\text{mẫu số liệu ghép nhóm là: } Q_3 = 21 + \frac{\frac{3 \cdot 100}{4} - (22 + 38)}{27}(24 - 21) = \frac{68}{3}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = \frac{505}{114}$$

Giá trị x trong mẫu số liệu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$

$$\text{Hay } x > \frac{39}{2} + 1,5 \cdot \frac{24}{19} = 21,39 \text{ hoặc } x < \frac{693}{38} - 1,5 \cdot \frac{24}{19} = 16,34$$

Vậy các giá trị ngoại lệ thuộc khoảng $[15; 18)$; $[24; 27)$; $[27; 30)$; $[30; 33)$

Khoảng biến thiên của mẫu số liệu ghép nhóm sau khi loại bỏ giá trị ngoại lệ: $24 - 18 = 6$ (phút)

Gọi $z_1; z_2; \dots; z_{65}$ là mẫu số liệu gốc gồm thời gian 65 lần ông Thắng đi xe buýt từ nhà đến cơ quan được xếp theo thứ tự không giảm, sau khi đã loại bỏ các giá trị ngoại lệ

Ta có: $z_1; z_2; \dots; z_{38} \in [18; 21)$; $z_{39}; \dots; z_{65} \in [21; 24)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $z_{17} \in [18; 21)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1'' = 18 + \frac{\frac{65}{4}}{38}(21 - 18) = \frac{2931}{152}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $z_{50} \in [21; 24)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3'' = 21 + \frac{\frac{3 \cdot 65}{4} - 38}{27}(24 - 21) = \frac{799}{36}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q'' = Q_3'' - Q_1'' = 2,91$$

Nhận xét: Sau khi loại bỏ giá trị ngoại lệ, khoảng biến thiên mới giảm mạnh còn khoảng tứ phân vị mới không bị ảnh hưởng nhiều

b) Cỡ mẫu $n = 25$

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

Ta có: $x_1; x_2 \in [155; 160); x_3; \dots; x_9 \in [160; 165); x_{10}; \dots; x_{21} \in [165; 170); x_{22}; \dots; x_{24} \in [170; 175); x_{25} \in [180; 185)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_6 + x_7) \in [160; 165)$. 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 $x_{19} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số

liệu ghép nhóm là: $Q_3 = 165 + \frac{\frac{3.25}{4} - (2 + 7)}{12}(170 - 165) = \frac{2705}{16}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = \frac{675}{12}$

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

ta có: $y_1; y_2; \dots; y_5 \in [155; 160); y_6; \dots; y_{14} \in [160; 165); y_{15}; \dots; y_{22} \in [165; 170); y_{23}; y_{24} \in [170; 175); y_{25} \in [175; 180)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(y_6 + y_7) \in [160; 165)$. 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 $y_{19} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số

liệu ghép nhóm là: $Q_3' = 165 + \frac{\frac{3.25}{4} - (5 + 9)}{8}(170 - 165) = \frac{5375}{32}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = \frac{2095}{288}$

Có $\Delta_Q' > \Delta_Q$ nên chiều cao của các bạn học sinh nữ lớp 12D có độ phân tán lớn hơn lớp 12C

Trả lời câu hỏi Vận dụng trang 73 SGK Toán 12 Chân trời sáng tạo

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:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là: $34 - 19 = 15(\text{tuổi})$

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là: $31 - 19 = 12(\text{tuổi})$

Cỡ mẫu $n = 100$

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

Ta có: $x_1; x_2; \dots; x_{10} \in [19; 22); x_{11}; \dots; x_{37} \in [22; 25); x_{38}; \dots; x_{68} \in [25; 28); x_{69}; \dots; x_{93} \in [28; 31); x_{94}; \dots; x_{100} \in [31; 34)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{25} + x_{26}) \in [22; 25)$. 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 $\frac{1}{2}(x_{75} + x_{76}) \in [28; 31)$. Do đó, tứ phân vị thứ ba của

mẫu số liệu ghép nhóm là: $Q_3 = 28 + \frac{\frac{3 \cdot 100}{4} - (10 + 27 + 31)}{25}(31 - 28) = \frac{721}{25}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = \frac{388}{75}$

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

Ta có: $y_1; y_2; \dots; y_{47} \in [19; 22); y_{48}; \dots; y_{87} \in [22; 25); y_{88}; \dots; y_{98} \in [25; 30); y_{99}; y_{100} \in [28; 31)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(y_{25} + y_{26}) \in [19; 22)$. Do đó, tứ phân vị thứ nhất

của mẫu số liệu ghép nhóm là: $Q_1' = 19 + \frac{\frac{100}{4}}{47}(22 - 19) = \frac{968}{47}$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(y_{75} + y_{76}) \in [22; 25)$. Do đó, tứ phân vị thứ ba của

mẫu số liệu ghép nhóm là: $Q_3' = 22 + \frac{\frac{3 \cdot 100}{4} - 47}{40}(25 - 22) = \frac{241}{10}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = \frac{1647}{470}$

b) Có $\Delta_Q' < \Delta_Q$ nên phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 1: 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

(Nguồn: Tổng cục Thống kê)

- 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 theo thứ tự tăng dần:

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

Tứ phân vị thứ nhất của mẫu số liệu là trung vị của 147; 187,1; 200,7; 242,2; 251,4; 258,4; 288,5; 298,1; 305; 332 nên: $Q_1 = \frac{6276}{25}$

Tứ phân vị thứ nhất của mẫu số liệu là trung vị của 341,4; 388,6; 400; 413,5; 413,5; 421; 432,2; 475; 520; 522,9 nên: $Q_3 = \frac{43281}{100}$

Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = \frac{18177}{100}$

- b)

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

- c) Khoảng biến thiên của mẫu số liệu ghép nhóm: $540 - 140 = 400(\text{ mm})$

Cỡ mẫu $n = 20$;

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 $\frac{1}{2}(x_5 + x_6) \in [240; 340)$. Do đó, tứ phân vị thứ nhất của

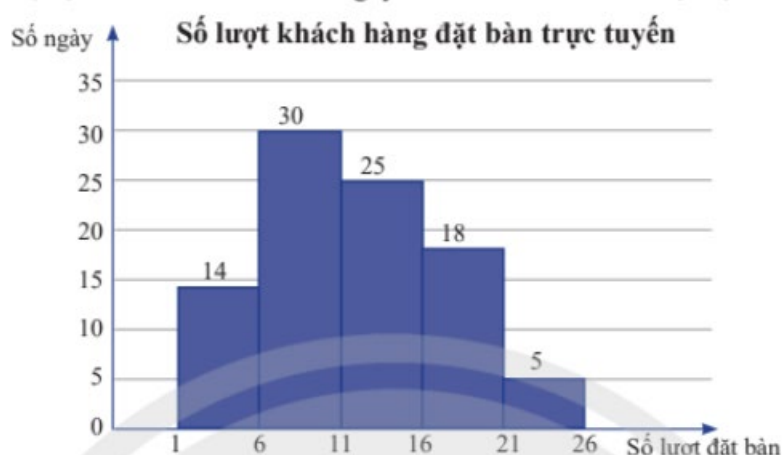
$$\text{mẫu số liệu ghép nhóm là: } Q_1' = 240 + \frac{\frac{20}{4} - 3}{7}(340 - 240) = \frac{1880}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{15} + x_{16}) \in [340; 440)$. Do đó, tứ phân vị thứ ba của

$$\text{mẫu số liệu ghép nhóm là: } Q_3' = 340 + \frac{\frac{3 \cdot 20}{4} - (3 + 7)}{7}(440 - 340) = \frac{2880}{7}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = \frac{1000}{7}$ Khoảng biến thiên của mẫu số liệu ghép nhóm lớn hơn mẫu số liệu; khoảng tứ phân vị của mẫu số liệu ghép nhóm nhỏ hơn mẫu số liệu

Bài 2: Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 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

Cỡ mẫu $n = 92$;

Gọi $x_1; x_2; \dots; x_{92}$ là mẫu số liệu gốc về số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 của 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 $\frac{1}{2}(x_{23} + x_{24}) \in [6; 11)$. Do đó, tứ phân vị thứ nhất của

$$\text{mẫu số liệu ghép nhóm là: } Q_1 = 6 + \frac{\frac{92}{4} - 14}{30}(11 - 6) = 7,5$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{69} + x_{70}) \in [11; 16)$ và $[16; 21)$. 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: $\Delta_Q = Q_3 - Q_1 = 8,5$

Bài 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: $9,4 - 8,4 = 1(\text{m})$

Cỡ mẫu $n = 100$;

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc về số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 của nhà hàng 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 $\frac{1}{2}(x_{25} + x_{26}) \in [8,8; 9,0)$. Do đó, tứ phân vị thứ nhất

$$\text{của mẫu số liệu ghép nhóm là: } Q_1 = 8,8 + \frac{\frac{100}{4} - (5+12)}{25}(9,0 - 8,8) = 9,44$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [9,0; 9,2)$. Do đó, tứ phân vị thứ ba của

$$\text{mẫu số liệu ghép nhóm là: } Q_3 = 9,0 + \frac{\frac{3 \cdot 100}{4} - (5+12+25)}{44}(9,2 - 9,0) = 10,5$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 1,06$

b) Giá trị x trong mẫu số liệu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$

Hay $x > 10,5 + 1,5 \cdot 1,06 = 12,09$ hoặc $x < 9,44 - 1,5 \cdot 1,06 = 7,85$

Vậy cây cao 8,4m không phải là giá trị ngoại lệ

Bài 4: Bảng tần số ghép nhóm dưới đây thể hiện kết quả điều tra về tuổi thọ trung bình của nam giới và nữ giới ở 50 quốc gia.

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 thọ trung bình của nam giới và nữ giới trong mẫu số liệu ghép nhóm trên.

b) Hãy cho biết tuổi thọ trung bình của nam giới hay nữ giới trong mẫu số liệu ghép nhóm trên đồng đều hơn.

Lời giải

Cỡ mẫu $n = 50$

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về tuổi thọ trung bình của nam giới ở 50 quốc gia được xếp theo thứ tự không giảm.

Ta có: $x_1; x_2; \dots; x_4 \in [50; 55); x_5; \dots; x_{11} \in [55; 60); x_{12}; \dots; x_{15} \in [60; 65); x_{16}; \dots; x_{21} \in [65; 70); x_{22}; \dots; x_{36} \in [70; 75); x_{37}; \dots; x_{48} \in [75; 80); x_{49}; x_{50} \in [80; 85)$

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: $Q_1 = 60 + \frac{\frac{50}{4} - (4 + 7)}{4} (65 - 60) = 71,875$

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 là: $\Delta_Q = Q_3 - Q_1 = 3,75$

Gọi $y_1; y_2; \dots; y_{50}$ là mẫu số liệu gốc về tuổi thọ trung bình của nữ giới ở 50 quốc gia được xếp theo thứ tự không giảm.

ta có: $y_1; \dots; y_3 \in [50; 55); y_4; \dots; y_7 \in [55; 60); y_8; \dots; y_{12} \in [60; 65); y_{13}; \dots; y_{15} \in [65; 70);$
 $y_{16}; \dots; y_{22} \in [70; 75); y_{23}; \dots; y_{36} \in [75; 80); y_{37}; \dots; y_{49} \in [80; 85); y_{50} \in [85; 90)$

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

$$\text{liệu ghép nhóm là: } Q_1' = 65 + \frac{\frac{50}{4} - (3 + 4 + 5)}{3} (70 - 65) = \frac{395}{6}$$

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

$$\text{liệu ghép nhóm là: } Q_3' = 80 + \frac{\frac{3 \cdot 50}{4} - (3 + 4 + 5 + 3 + 7 + 14)}{13} (85 - 80) = \frac{2095}{26}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta Q' = Q_3' - Q_1' = \frac{575}{39}$$

b) Có $\Delta_Q' > \Delta_Q$ nên độ tuổi trung bình của nam giới đồng đều hơn

C. CÁC DẠNG TOÁN

Dạng 1: Tìm khoảng biến thiên

1. Phương pháp

Cho mẫu số liệu ghép nhóm:

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
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 ghép nhóm trên là $R = a_{k+1} - a_1$.

2. Ví dụ

Câu 1. 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?

Lời giải

a) Khoảng biến thiên R cho mẫu số liệu ghép nhóm trên là: $45 - 25 = 20$

b) Khoảng biến thiên của mẫu số liệu gốc là: $43 - 27 = 16$

Câu 2. Bảng dưới biểu thị kết quả điều tra thời gian sử dụng Internet hằng ngày của một số người.

Thời gian (phút)	[30;60)	[60;90)	[90;120)	[120;150)	[150;180)
Số người	2	4	10	5	3

Tìm khoảng biến thiên của mẫu số liệu đã cho. Kết quả cho biết điều gì?

Lời giải

Đầu mút phải của nhóm ghép cuối cùng là 180, đầu mút trái của nhóm ghép đầu tiên là 30. Vậy khoảng biến thiên của mẫu số liệu là $R = 180 - 30 = 150$.

Kết quả này cho biết thời gian sử dụng Internet hằng ngày của các thành viên thuộc nhóm người được điều tra chênh lệch nhau nhiều nhất là 150 phút.

Câu 3. Bảng dưới thống kê thành tích nhảy xa của một số học sinh lớp 12. Tìm khoảng biến thiên thành tích nhảy xa của số học sinh này.

Thành tích (cm)	[150;180)	[180;210)	[210;240)	[240;270)	[270;300)
Số học sinh	3	5	28	14	8

Lời giải

Khoảng biến thiên của mẫu số liệu là $R = 300 - 150 = 150$

Câu 4. Để chuẩn bị mở một trung tâm thể dục thể thao, anh Dũng đã tiến hành điều tra tuổi thọ của máy chạy bộ do hai hãng X, Y sản xuất. Bảng dưới biểu thị hai mẫu số liệu mà anh thu thập được qua Internet.

Bảng. Tuổi thọ của máy chạy bộ (đơn vị: năm)

Tuổi thọ	[2;4)	[4;6)	[6;8)	[8;10)	[10;12)
Số máy của hãng X	7	20	36	20	17
Số máy của hãng Y	0	20	35	35	10

Khoảng biến thiên của mẫu số liệu nào lớn hơn? Từ đó có thể nói là máy chạy bộ do hãng nào sản xuất có tuổi thọ phân tán hơn?

Lời giải

Khoảng biến thiên của tuổi thọ máy chạy bộ do hãng X và hãng Y sản xuất tương ứng là $R_X = 12 - 2 = 10$ và $R_Y = 12 - 4 = 8$. Vì $R_X > R_Y$ nên có thể nói là máy do hãng X sản xuất có tuổi thọ phân tán hơn so với máy của hãng Y .

Câu 5. Người ta tiến hành phỏng vấn hai nhóm khán giả về một bộ phim mới công chiếu. Nhóm A gồm những khán giả thuộc lứa tuổi 20 - 30, nhóm B thuộc lứa tuổi trên 30. Người được hỏi ý kiến phải đánh giá bộ phim bằng cách cho điểm theo một số tiêu chí nêu trong phiếu điều tra và sau đó lấy tổng số điểm (thang điểm 100). Bảng dưới đây trình bày kết quả điều tra hai nhóm khán giả:

Bảng. Điểm đánh giá của khán giả

Điểm	[50;60)	[60;70)	[70;80)	[80;90)	[90;100)
Số người của nhóm A	6	10	14	12	8

Số người của nhóm B	0	8	14	28	0
---------------------	---	---	----	----	---

Ý kiến đánh giá của nhóm khán giả nào phân tán hơn?

Lời giải

$$R_A = 100 - 50 = 50$$

$$R_B = 90 - 60 = 30$$

Ta thấy $R_A > R_B$ nên nhóm khán giả A phân tán hơn

Câu 6. Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

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

Lời giải

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).

Dạng 2. Xác định khoảng tứ phân vị của mẫu số liệu ghép nhóm

1. Phương pháp

Tứ phân vị thứ r là

$$Q_r = a_p + \frac{\frac{r \cdot n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p),$$

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

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

2. Ví dụ

Câu 7. 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

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

Lời giải

Hiệu chỉnh lại bảng số liệu ta có:

Thời gian t (phút)	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)
Số cuộc gọi	8	17	25	20	10

Cỡ mẫu $n = 80$. Giả sử $x_1, x_2, \dots, x_{80}$ là thời gian đàm thoại của 80 cuộc gọi và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự tăng dần.

Vì $\frac{n}{4} = 20$ và $8 < 20 < 8 + 17$ nên nhóm chứa tứ phân vị thứ nhất là nhóm [1;2) và tứ phân

$$\text{vị thứ nhất là: } Q_1 = 1 + \frac{\frac{80}{4} - 8}{17} \cdot 1 = \frac{29}{17}$$

vì $\frac{3n}{4} = 60$ và $8 + 17 + 25 < 60 < 8 + 17 + 25 + 20$ nên nhóm chứa tứ phân vị thứ ba là nhóm

$$[3;4) \text{ và tứ phân vị thứ ba là: } Q_3 = 3 + \frac{\frac{3 \cdot 80}{4} - (8 + 17 + 25)}{20} \cdot 1 = 3,5$$

$$\text{Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } 3,5 - \frac{29}{17} = \frac{61}{34}$$

Câu 8. Thống kê số ngày trong tháng Sáu năm 2021 và năm 2022 theo nhiệt độ cao nhất trong ngày tại Hà Nội, người ta thu được bảng sau:

Nhiệt độ ($^{\circ}\text{C}$)	[28;30)	[30;32)	[32;34)	[34;36)	[36;38)	[38;40)
Số ngày trong tháng 6/2021	0	2	8	5	6	9
Số ngày trong tháng 6/2022	2	3	4	11	8	2

Hỏi tháng Sáu năm nào ở Hà Nội nhiệt độ cao nhất trong ngày biến đổi nhiều hơn?

Lời giải

Năm 2021: Khoảng biến thiên của nhiệt độ là: $R_1 = 40 - 30 = 10$

Cỡ mẫu $n = 30$. Giả sử $y_1, y_2, \dots, y_{30}$ là nhiệt độ cao nhất trong ngày của 30 ngày tháng Sáu năm 2021 và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự không giảm.

Vì $\frac{n}{4} = \frac{30}{4} = 7,5$ và $2 < 7,5 < 2 + 8$ nên tứ phân vị thứ nhất thuộc nhóm $[32; 34)$ và tứ phân

$$\text{vị thứ nhất là: } Q_1 = 32 + \frac{\frac{30}{4} - 2}{8} \cdot 2 = 33,375$$

vì $\frac{3n}{4} = \frac{3 \cdot 30}{4} = 22,5$ và $2 + 8 + 5 + 6 < 22,5 < 2 + 8 + 5 + 6 + 9$ nên tứ phân vị thứ ba thuộc

$$\text{nhóm } [38; 40) \text{ và tứ phân vị thứ ba là: } Q_3 = 38 + \frac{\frac{3 \cdot 30}{4} - (2 + 8 + 5 + 6)}{9} \cdot 2 = \frac{115}{3}$$

$$\text{Khoảng biến thiên của mẫu số liệu ghép nhóm là: } \Delta_{Q_1} = \frac{115}{3} - 33,375 = \frac{119}{24}$$

Năm 2022: Khoảng biến thiên của nhiệt độ là: $R_2 = 40 - 28 = 12$

Cỡ mẫu $n = 30$. Giả sử $z_1, z_2, \dots, z_{30}$ là nhiệt độ cao nhất trong ngày của 30 ngày tháng Sáu năm 2022 và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự tăng dần.

vì $\frac{n}{4} = \frac{30}{4} = 7,5$ và $2 + 3 < 7,5 < 2 + 3 + 4$ nên nhóm chứa tứ phân vị thứ nhất là nhóm

$$[32; 34) \text{ và tứ phân vị thứ nhất là: } Q'_1 = 32 + \frac{\frac{30}{4} - (2 + 3)}{4} \cdot (34 - 32) = 33,25$$

vì $\frac{3n}{4} = \frac{3 \cdot 30}{4} = 22,5$ và $2 + 3 + 4 + 11 < 22,5 < 2 + 3 + 4 + 11 + 8$ nên nhóm chứa tứ phân vị thứ ba là nhóm $[36; 38)$ và tứ phân vị thứ ba là:

$$Q'_3 = 36 + \frac{\frac{3 \cdot 30}{4} - (2 + 3 + 4 + 11)}{8} \cdot (38 - 36) = 36,625$$

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $\Delta_{Q_2} = 36,625 - 33,25 = 3,375$

Theo khoảng biến thiên: vì $R_2 > R_1$ nên nhiệt độ cao nhất trong ngày vào tháng 6 năm 2022 biến đổi nhiều hơn nhiệt độ cao nhất trong ngày vào tháng 6 năm 2021.

Theo khoảng tứ phân vị: vì $\Delta_{Q_1} > \Delta_{Q_2}$ nên nhiệt độ cao nhất trong ngày vào tháng 6 năm 2021 biến đổi nhiều hơn nhiệt độ cao nhất trong ngày vào tháng 6 năm 2022.

Câu 9. Thống kê số thẻ vàng của mỗi câu lạc bộ trong giải ngoại hạng Anh mùa giải 2021 - 2022 cho kết quả như sau:

101	79	79	78	75	73	68	67	67	63
63	61	60	59	57	55	55	50	47	42.

(Theo premierleague.com)

a) Hãy ghép nhóm dãy số liệu trên thành các nhóm có độ dài bằng nhau với nhóm đầu tiên là $[40; 50)$.

b) Tính khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu gốc và mẫu số liệu ghép nhóm thu được ở câu a. Giá trị nào là giá trị chính xác? Giá trị nào là giá trị xấp xỉ?

Lời giải

a) Bảng số liệu ghép nhóm:

Số thẻ	[40;50)	[50;60)	[60;70)	[70;80)	[80;90)	[90;100)	[100;110)
Tần số	2	5	7	5	0	0	1

b) Với mẫu số liệu gốc: Khoảng biến thiên là: $R_1 = 101 - 42 = 59$

Sắp xếp mẫu số liệu theo thứ tự không giảm là:

42; 47; 50; 55; 55; 57; 59; 60; 61; 63; 63; 67; 67; 68; 73; 75; 78; 79; 79; 101

Vì $n = 20$ nên tứ phân vị thứ nhất là trung vị của dãy số liệu:

42; 47; 50; 55; 55; 57; 59; 60; 61; 63. Do đó, $Q_1 = \frac{55+57}{2} = 56$

Tứ phân vị thứ ba là trung vị của dãy số liệu: 63; 67; 67; 68; 73; 75; 78; 79; 79; 101. Do đó,

$Q_3 = \frac{73+75}{2} = 74$.

Khoảng tứ phân vị là: $\Delta_{Q_1} = 74 - 56 = 18$

Với mẫu số liệu ghép nhóm: Khoảng biến thiên là: $R_2 = 110 - 40 = 70$

Cỡ mẫu $n = 20$. Giả sử $x_1, x_2, \dots, x_{20}$ là số thẻ vàng mà mỗi câu lạc bộ ngoại hạng Anh nhận được mùa giải 2021-2022, các giá trị này đã được sắp xếp theo thứ tự không giảm.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2}$ nên nhóm chứa tứ phân vị thứ nhất là

nhóm [50;60) và ta có: $Q'_1 = 50 + \frac{\frac{20}{4} - 2}{5} \cdot 10 = 56$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2}$ nên nhóm chứa tứ phân vị thứ ba là

nhóm [70;80) và ta có: $Q'_3 = 70 + \frac{\frac{3 \cdot 20}{4} - (2+5+7)}{5} \cdot 10 = 72$

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $\Delta_{Q_2} = 72 - 56 = 16$

Giá trị chính xác là $R_1; \Delta Q_1$, giá trị xấp xỉ là $R_2; \Delta Q_2$

Câu 10. Thu nhập theo tháng (đơn vị: triệu đồng) của người lao động ở hai nhà máy như sau:

Thu nhập	[5;8)	[8;11)	[11;14)	[14;17)	[17;20)
Số người của nhà máy A	20	35	45	35	20
Số người của nhà máy B	17	23	30	23	17

Tính mức thu nhập trung bình của người lao động ở hai nhà máy trên. Dựa vào khoảng tứ phân vị, hãy xác định xem mức thu nhập của người lao động ở nhà máy nào biến động nhiều hơn.

Lời giải

Ta có bảng số liệu với giá trị đại diện của nhóm là:

Thu nhập	[5;8)	[8;11)	[11;14)	[14;17)	[17;20)
Giá trị đại diện	6,5	9,5	12,5	15,5	18,5
Số người của nhà máy A	20	35	45	35	20
Số người của nhà máy B	17	23	30	23	17

Mức thu nhập trung bình của người lao động nhà máy A là:

$$\frac{6,5.20 + 9,5.35 + 12,5.45 + 15,5.35 + 18,5.20}{20 + 35 + 45 + 35 + 20} = \frac{25}{2} \text{ (triệu đồng)}$$

Mức thu nhập trung bình của người lao động nhà máy B là:

$$\frac{6,5.17 + 9,5.23 + 12,5.30 + 15,5.23 + 18,5.17}{17 + 23 + 30 + 23 + 17} = \frac{25}{2} \text{ (triệu đồng)}$$

Nhà máy A: Ta có cỡ mẫu $n = 155$. Giả sử $x_1, x_2, \dots, x_{155}$ là mức thu nhập của người lao động nhà máy A và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự không giảm.

Vì $\frac{n}{4} = 38,75$ và $20 < 38,75 < 20 + 35$ nên nhóm chứa tứ phân vị thứ nhất là nhóm [8;11)

$$\text{và tứ phân vị thứ nhất là: } Q_1 = 8 + \frac{\frac{155}{4} - 20}{35} \cdot 3 = \frac{269}{28}$$

Vì $\frac{3n}{4} = 116,25$ và $20 + 35 + 45 < 116,25 < 20 + 35 + 45 + 35$ nên nhóm chứa tứ phân vị thứ

$$\text{ba là nhóm [14;17) và tứ phân vị thứ ba là: } Q_3 = 14 + \frac{\frac{3.155}{4} - (20 + 35 + 45)}{35} \cdot 3 = \frac{431}{28}$$

$$\text{Khoảng biến thiên của mẫu số liệu ghép nhóm là: } \Delta_{Q_1} = \frac{431}{28} - \frac{269}{28} = \frac{81}{14}$$

Nhà máy B: Ta có cỡ mẫu $n = 110$. Giả sử $x_1, x_2, \dots, x_{110}$ là mức thu nhập của người lao động nhà máy B và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự không giảm.

Vì $\frac{n}{4} = 27,5$ và $17 < 27,5 < 17 + 23$ nên nhóm chứa tứ phân vị thứ nhất là nhóm [8;11) và

$$\text{tứ phân vị thứ nhất là: } Q_1' = 8 + \frac{\frac{110}{4} - 17}{23} \cdot 3 = \frac{431}{46}$$

Vì $\frac{3n}{4} = 82,5$ và $17 + 23 + 30 < 82,5 < 17 + 23 + 30 + 23$ nên nhóm chứa tứ phân vị thứ ba

là nhóm $[14; 17)$ và tứ phân vị thứ ba là: $Q_3' = 14 + \frac{\frac{3 \cdot 110}{4} - (17 + 23 + 30)}{23} \cdot 3 = \frac{719}{46}$

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $\Delta_{Q_2} = \frac{719}{46} - \frac{431}{46} = \frac{144}{23}$

Vì $\Delta_{Q_1} < \Delta_{Q_2}$ nên mức thu nhập của người lao động nhà máy B biến động nhiều hơn.

Câu 11. Bảng sau đây cho biết chiều cao của các học sinh lớp 12A và 12B.

Chiều cao (cm)	[145; 150)	[150; 155)	[155; 160)	[160; 165)	[165; 170)	[170; 175)
Số học sinh của lớp 12A	1	0	15	12	10	5
Số học sinh của lớp 12B	0	0	17	10	9	6

a) Tìm khoảng biến thiên, khoảng tứ phân vị cho các mẫu số liệu ghép nhóm về chiều cao của học sinh lớp 12A, 12B.

b) Để so sánh độ phân tán về chiều cao của học sinh hai lớp này ta nên dùng khoảng biến thiên hay khoảng tứ phân vị? Vì sao?

Lời giải

a) Lớp 12A: Khoảng biến thiên: $R = 175 - 145 = 30$

Ta có cỡ mẫu $n = 43$. Giả sử $x_1, x_2, \dots, x_{43}$ là chiều cao của các học sinh lớp 12A và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự không giảm.

vì $\frac{n}{4} = 10,75$ và $1 < 10,75 < 1 + 15$ nên nhóm chứa tứ phân vị thứ nhất là nhóm $[150; 160)$

và tứ phân vị thứ nhất là: $Q_1 = 155 + \frac{\frac{43}{4} - 1}{15} \cdot 5 = 158,25$

vì $\frac{3n}{4} = 32,25$ và $1 + 15 + 12 < 32,25 < 1 + 15 + 12 + 10$ nên nhóm chứa tứ phân vị thứ ba là

nhóm $[165; 170)$ và tứ phân vị thứ ba là: $Q_3 = 165 + \frac{\frac{3 \cdot 43}{4} - (1 + 15 + 12)}{10} \cdot 5 = 167,125$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_{Q_1} = 167,125 - 158,25 = 8,875$

Lớp 12B: Khoảng biến thiên: $R = 175 - 155 = 20$

Ta có cỡ mẫu $n = 42$. Giả sử $x_1, x_2, \dots, x_{42}$ là chiều cao của các học sinh lớp 12B và giả sử dãy số liệu gốc này đã được sắp xếp theo thứ tự không giảm.

Vì $\frac{n}{4} = 10,5$ và $0 < 10,5 < 17$ nên nhóm chứa tứ phân vị thứ nhất là nhóm $[155;160)$ và ta

$$\text{có: } Q'_1 = 155 + \frac{\frac{42}{4} - 0}{17} \cdot 5 = \frac{5375}{34}$$

vì $\frac{3n}{4} = 31,5$ và $17 + 10 < 31,5 < 17 + 10 + 9$ nên nhóm chứa tứ phân vị thứ ba là nhóm

$$[165;170) \text{ và tứ phân vị thứ ba là } Q'_3 = 165 + \frac{\frac{3 \cdot 42}{4} - (17 + 10)}{9} \cdot 5 = \frac{335}{2}$$

$$\text{Khoảng biến thiên của mẫu số liệu ghép nhóm là: } \Delta_{Q_2} = \frac{335}{2} - \frac{5375}{34} = \frac{160}{17}$$

b) Để so sánh độ phân tán về chiều cao của học sinh hai lớp này, ta nên dùng khoảng tứ phân vị vì khoảng tứ phân vị chỉ phụ thuộc vào nửa giữa của mẫu số liệu, không bị ảnh hưởng bởi các giá trị bất thường.

Câu 12. Điểm kiểm tra cuối khoá môn Tiếng Anh của hai lớp ở một trung tâm ngoại ngữ được thống kê trong các Bảng a và **b**.

Bảng a. Điểm của lớp A

Điểm	Số học viên (tần số)
[50;60)	8
[60;70)	20
[70;80)	50
[80;90)	17
[90;100)	5

Bảng b. Điểm của lớp B

Điểm	Số học viên (tần số)
[50;60)	15
[60;70)	20
[70;80)	30
[80;90)	20
[90;100)	15

a) Tìm khoảng biến thiên của mỗi mẫu số liệu. Có thể dùng khoảng biến thiên để biết điểm của lớp nào đồng đều hơn không?

b) Tìm các tứ phân vị và khoảng tứ phân vị của mỗi mẫu số liệu.

c) Mẫu số liệu nào có độ phân tán lớn hơn? Minh hoạ câu trả lời bằng cách biểu diễn các tứ phân vị và khoảng tứ phân vị của mỗi mẫu số liệu trên trục số.

Lời giải

a) Hai mẫu số liệu đều có khoảng biến thiên là $R = 100 - 50 = 50$ nên không thể căn cứ vào đó để nói điểm của lớp nào đồng đều hơn.

b) Kích thước của hai mẫu số liệu đều là $N = 100$. Ta có $\frac{N}{4} = 25$; $\frac{N}{2} = 50$; $\frac{3N}{4} = 75$.

- Đối với mẫu số liệu về điểm của lớp A , ta tìm các tứ phân vị Q_1^A, Q_2^A, Q_3^A và khoảng tứ phân vị Δ_Q^A qua bảng tần số tích lũy dưới đây:

Điểm	Tần số	Tần số tích lũy
[50; 60)	8	8
[60; 70)	20	28
[70; 80)	50	78
[80; 90)	17	95
[90; 100)	5	100

Nhóm chứa Q_1^A là [60; 70). $Q_1^A = 60 + \frac{25-8}{20} \cdot 10 = 68,5$.

Nhóm chứa Q_2^A là [70; 80). $Q_2^A = 70 + \frac{50-28}{50} \cdot 10 = 74,4$.

Nhóm chứa Q_3^A là [70; 80). $Q_3^A = 70 + \frac{75-28}{50} \cdot 10 = 79,4$.

Vậy $\Delta_Q^A = 79,4 - 68,5 = 10,9$.

- Gọi Q_1^B, Q_2^B, Q_3^B là các tứ phân vị và Δ_Q^B là khoảng tứ phân vị của mẫu số liệu về điểm của lớp B . Ta lập bảng tần số tích lũy và tính được:

Điểm	Tần số	Tần số tích lũy
[50; 60)	15	15
[60; 70)	20	35
[70; 80)	30	65
[80; 90)	20	85
[90; 100)	15	100

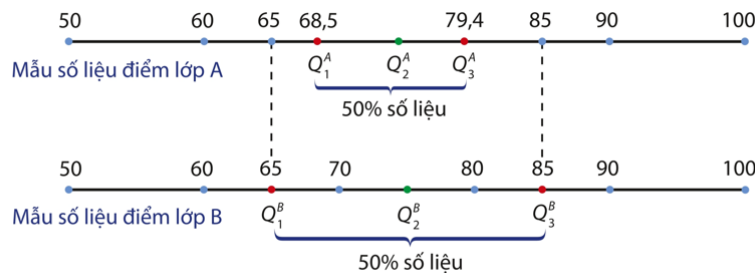
Nhóm chứa Q_1^B là [60; 70). $Q_1^B = 60 + \frac{25-15}{20} \cdot 10 = 65$.

Nhóm chứa Q_2^B là [70; 80). $Q_2^B = 70 + \frac{50-35}{30} \cdot 10 = 75$.

Nhóm chứa Q_3^B là $[80; 90)$. $Q_3^B = 80 + \frac{75-65}{20} \cdot 10 = 85$.

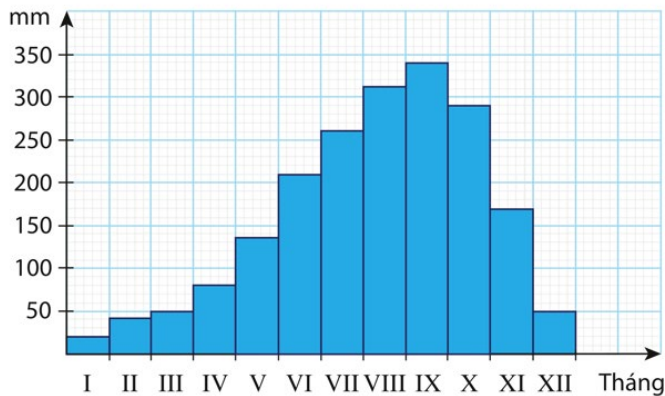
Vậy $\Delta_Q^B = 85 - 65 = 20$.

c) $\Delta_Q^B > \Delta_Q^A$ nên điểm của lớp B phân tán hơn điểm của lớp A . Minh hoạ trên trục số: Mỗi mẫu đều có 100 số liệu thuộc khoảng $[50; 100)$. Có 50% số liệu ở giữa của bảng điểm lớp B thuộc khoảng (Q_1^B, Q_3^B) . Bảng điểm lớp A cũng có 50% số liệu ở giữa thuộc khoảng (Q_1^A, Q_3^A) . Vì $(Q_1^A, Q_3^A) \subset (Q_1^B, Q_3^B)$ nên ta có thể nói là điểm của lớp B phân tán hơn so với điểm lớp A .



Hình 3.1

Câu 13. Hình dưới là biểu đồ biểu diễn lượng mưa trung bình của các tháng trong năm ở thành phố #A.



Hình 3.2. Biểu đồ lượng mưa ở thành phố A

a) Lập bảng số liệu ghép nhóm về lượng mưa của thành phố A , với độ dài các nhóm là 50 và đầu mút phải của nhóm cuối cùng là 350.

b) Xác định khoảng tứ phân vị của mẫu số liệu (làm tròn kết quả đến hàng đơn vị). Nêu ý nghĩa của kết quả tìm được.

Lời giải

a) Bảng số liệu về lượng mưa của thành phố A :

Lượng mưa (mm)	Tần số	Tần số tích lũy
$[0; 50)$	2	2
$[50; 100)$	3	5
$[100; 150)$	1	6
$[150; 200)$	1	7

[200;250)	1	8
[250;300)	2	10
[300;350)	2	12

b) $N = 12$ nên $\frac{N}{4} = 3; \frac{3N}{4} = 9$.

Nhóm chứa Q_1 là $[50;100)$. $Q_1 = 50 + \frac{3-2}{3} \cdot 50 \approx 67$.

Nhóm chứa Q_3 là $[250;300)$. $Q_3 = 250 + \frac{9-8}{2} \cdot 50 = 275$.

Vậy $\Delta_Q = 275 - 67 = 208$.

Kết quả tìm được cho thấy: Hằng năm, ở thành phố A có 3 tháng có lượng mưa trung bình không vượt quá 67 mm và 3 tháng có lượng mưa trung bình ít nhất là 275 mm . Trong 6 tháng còn lại, lượng mưa trung bình đạt từ 67 mm đến 275 mm và như vậy là lượng mưa của 6 tháng này có thể chênh lệch nhau đến 208 mm .

D. TRẮC NGHIỆM 4 PHƯƠNG ÁN

Câu 1. Cô Hà thống kê lại đường kính thân gỗ của một số cây xoan đào 6 năm tuổi được trồng ở một lâm trường ở bảng sau.

Đường kính (cm)	[40;45)	[45;50)	[50;55)	[55;60)	[60;65)
Tần số	5	20	18	7	3

Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm trên.

A. 25.

B. 30.

C. 6.

D. 69,8.

Lời giải

Chọn A

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là $65 - 40 = 25(\text{cm})$.

Câu 2. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 60 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	Tần số
[40 ; 50)	3
[50 ; 60)	6
[60 ; 70)	19
[70 ; 80)	23
[80 ; 90)	9
	$n = 60$

Bảng 8

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là:

- A.** 50. **B.** 30. **C.** 6. **D.** 69,8.

Lời giải

Chọn A

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 90 - 40 = 50$

Câu 3. Cho mẫu số liệu ghép nhóm có tứ phân vị thứ nhất, thứ hai, thứ ba lần lượt là Q_1, Q_2, Q_3 . Khoảng tứ phân vị của mẫu số liệu ghép nhóm đó bằng:

- A.** $2Q_2$. **B.** $Q_1 - Q_3$.
C. $Q_3 - Q_1$. **D.** $Q_3 + Q_1 - Q_2$.

Lời giải

Chọn C

Chọn phương án đúng.

Câu 4. Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khoẻ. 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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

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

Lời giải

Chọn A

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

Câu 5. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

- A.** 25. **B.** 20. **C.** 15. **D.** 30.

Lời giải

Chọn A

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

Câu 6. 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 (giờ)	[8;10)	[10;12)	[12;14)	[14;16)	[16;18)
Số lần	4	6	8	4	3

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.

Lời giải

Chọn C

Khoảng biến thiên của mẫu số liệu là: $18 - 8 = 10$ (giờ)

Câu 7. 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

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A.** 3. **B.** 4. **C.** 5. **D.** 6.

Lời giải

Chọn C

Khoảng biến thiên: $19 - 14 = 5$

Câu 8. 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

Nhóm chứa tứ phân vị thứ nhất là

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

Lời giải

Chọn C

Ta có: $\frac{n}{4} = \frac{20}{4} = 5$ và $1 + 3 < 5 < 1 + 3 + 8$ nên tứ phân vị thứ nhất của mẫu số liệu thuộc nhóm [16;17)

Câu 9. 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

Nhóm chứa tứ phân vị thứ ba là

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

Lời giải

Chọn C

Ta có: $\frac{3 \cdot 20}{4} = 15$ và $1 + 3 + 8 < 15 < 1 + 3 + 8 + 6$ tứ phân vị thứ ba thuộc nhóm [17;18)

Câu 10. Khảo sát thời gian tập thể dục trong ngày của một số học sinh khối 11 thu được mẫu số liệu ghép nhóm sau:

Thời gian (phút)	[0; 20)	[20; 40)	[40; 60)	[60; 80)	[80; 100)
Số học sinh	5	9	12	10	6

Nhóm chứa tứ phân vị thứ nhất là

- A.** [0; 20). **B.** [20; 40). **C.** [40; 60). **D.** [60; 80).

Lời giải

Chọn B

Ta có cỡ mẫu là $n = 5 + 9 + 12 + 10 + 6 = 42$.

Gọi $x_1, x_2, \dots, x_{42}$ là thời gian tập thể dục trong ngày của 42 học sinh khối 11 và giả sử dãy này đã sắp xếp theo thứ tự tăng dần.

Khi đó tứ phân vị thứ nhất Q_1 là trung vị của dãy gồm 21 số liệu đầu nên $Q_1 = x_{11}$. Do x_{11} thuộc nhóm [20; 40) nên nhóm này chứa Q_1 .

Câu 11. 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

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.

Lời giải

Chọn C

Tứ phân vị thứ nhất của dãy số liệu là $\frac{1}{2}(x_4 + x_5)$ thuộc nhóm [7; 9) nên tứ phân vị thứ

nhất của mẫu số liệu là $Q_1 = 7 + \frac{\frac{20}{2} - 2}{7}(9 - 7) = 7,86$

Câu 12. 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

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.

Lời giải

Chọn B

Tứ phân vị thứ ba của dãy số liệu là $\frac{1}{2}(x_{15} + x_{16})$ thuộc nhóm [9;11) nên tứ phân vị thứ ba

$$\text{của mẫu số liệu là } Q_3 = 9 + \frac{\frac{3.20}{4} - 9}{7}(11 - 9) = 10,7$$

Câu 13. Cho mẫu số liệu ghép nhóm về tuổi thọ (đơn vị tính là năm) của một loại bóng đèn mới như sau.

Tuổi thọ	[2;3,5)	[3,5;5)	[5;6,5)	[6,5;8)
Số bóng đèn	8	22	35	15

Nhóm chứa tứ phân vị thứ nhất của mẫu số liệu là

- A.** [2;3,5) . **B.** [3,5;5) . **C.** [5;6,5) . **D.** [6,5;8) .

Lời giải

Chọn B

Câu 14. Cho mẫu số liệu ghép nhóm về tuổi thọ (đơn vị tính là năm) của một loại bóng đèn mới như sau.

Tuổi thọ	[2;3,5)	[3,5;5)	[5;6,5)	[6,5;8)
Số bóng đèn	8	22	35	15

Nhóm chứa tứ phân vị thứ ba của mẫu số liệu là

- A.** [2;3,5) . **B.** [3,5;5) . **C.** [5;6,5) . **D.** [6,5;8) .

Lời giải

Chọn C

Câu 15. Cho bảng số liệu về chiều cao của 100 học sinh một trường trung học phổ thông dưới đây.

Nhóm	Chiều cao (cm)	Số học sinh
1	[150;153)	7
2	[153;156)	13
3	[156;159)	40
4	[159;162)	21
5	[162;165)	13
6	[165;168)	6

Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng phần trăm) là

- A.** 156,25. **B.** 157,5. **C.** 156,38. **D.** 157,54.

Lời giải

Chọn C

Câu 16. Cho bảng số liệu về chiều cao của 100 học sinh một trường trung học phổ thông dưới đây.

Nhóm	Chiều cao (cm)	Số học sinh
1	[150;153)	7
2	[153;156)	13
3	[156;159)	40
4	[159;162)	21

5	[162;165)	13
6	[165;168)	6

Tứ phân vị thứ ba của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng phần trăm) là:

- A.** 160,52. **B.** 161,52. **C.** 161,14. **D.** 162,25.

Lời giải

Chọn C

Câu 17. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 60 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	Tần số
[40 ; 50)	3
[50 ; 60)	6
[60 ; 70)	19
[70 ; 80)	23
[80 ; 90)	9
	$n = 60$

Bảng 8

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

- A.** 50. **B.** 40. **C.** 14,23. **D.** 70,87.

Lời giải

Chọn C

Số phần tử của mẫu là $n = 60$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 3, cf_2 = 9, cf_3 = 28, cf_4 = 51, cf_5 = 60$

Ta có: $\frac{n}{4} = \frac{60}{4} = 15$ mà $9 < 15 < 28$ suy ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 15. Xét nhóm 3 là nhóm $[60; 70]$ có $s = 60, h = 10, n_3 = 19$ và nhóm 2 là nhóm $[50; 60]$ có $cf_2 = 9$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{15 - cf_2}{n_3} \right) \cdot h = 60 + \left(\frac{15 - 9}{19} \right) \cdot 10 = \frac{1200}{19}$

Ta có: $\frac{3n}{4} = \frac{3.60}{4} = 45$ mà $28 < 45 < 51$ suy ra nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 45. Xét nhóm 4 là nhóm $[70; 80]$ có $t = 70, \neq 10, n_4 = 23$ và nhóm 3 là nhóm $[60; 70]$ có $cf_3 = 28$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{45 - cf_3}{n_4} \right) . l = 70 + \left(\frac{45 - 28}{23} \right) . 10 = \frac{1780}{23}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = \frac{1780}{23} - \frac{1200}{19} \approx 14,23$

Câu 18. 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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 0,9.

B. 0,975.

C. 0,5.

D. 0,575.

Lời giải

Chọn D

Cỡ mẫu

$n = 20$

Gọi $x_1; x_2; \dots; x_{20}$ 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 $\frac{1}{2}(x_5 + x_6) \in [3, 0; 3, 3)$.

Do đó, tứ phân vị thứ nhất của mẫu số 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 $\frac{1}{2}(x_{15} + x_{16}) \in [3, 6; 3, 9)$.

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

$Q_3 = 3,6 + \frac{\frac{3.20}{4} - (3 + 6 + 5)}{4}(3,9 - 3,6) = 3,675$

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

$\Delta_Q = Q_3 - Q_1 = 0,575$

Câu 19. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 23,75.

B. 27,5.

C. 31,88.

D. 8,125.

Lời giải

Chọn D

Cỡ mẫu $n = 18$

Gọi $x_1; x_2; \dots; x_{18}$ là mẫu số liệu gốc về thời gian tập nhảy mỗi ngày của bạn Chi 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

$$\text{mẫu số liệu ghép nhóm là: } Q_1 = 20 + \frac{\frac{18}{4}}{6} (25 - 20) = 23,75$$

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

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

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = 8,125$$

Câu 20. 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 (giây)	[8;10)	[10;12)	[12;14)	[14;16)	[16;18)
Số lần	4	6	8	4	3

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 10,75.

B. 1,75.

C. 3,63.

D. 14,38.

Lời giải

Chọn C

Cỡ mẫu $n = 25$

Gọi $x_1; x_2; \dots; x_{25}$ là mẫu số liệu gốc về thời gian giải rubik trong 25 lần của bạn Dũng được xếp theo thứ tự không giảm.

$$\text{Ta có: } x_1; \dots; x_4 \in [8; 10); x_5; \dots; x_{10} \in [10; 12); x_{11}; \dots; x_{18} \in [12; 14); x_{19}; \dots; x_{22} \in [14; 16);$$

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

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_6 + x_7) \in [10; 12)$. Do đó, tứ phân vị thứ nhất

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

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

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

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = 3,63$$

Câu 21. Thời gian (phút) truy cập Internet mỗi buổi tối của một số học sinh được cho trong bảng sau:

Thời gian (phút)	[9,5;12,5)	[12,5;15,5)	[15,5;18,5)	[18,5;21,5)	[21,5;24,5)
Số học sinh	3	12	15	24	2

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 10,75.

B. 4,75.

C. 4,63.

D. 4,38.

Lời giải

Chọn B

Cỡ mẫu là $n = 56$.

Tứ phân vị thứ nhất Q_1 là $\frac{x_{14} + x_{15}}{2}$. Do x_{14}, x_{15} đều thuộc nhóm $[12,5; 15,5)$ nên nhóm này chứa Q_1 . Do đó, $p = 2; a_2 = 12,5; m_2 = 12; m_1 = 3; a_3 - a_2 = 3$ và ta có

$$Q_1 = 12,5 + \frac{\frac{56}{4} - 3}{12} \cdot 3 = 15,25$$

Với tứ phân vị thứ ba Q_3 là $\frac{x_{42} + x_{43}}{2}$. Do x_{42}, x_{43} đều thuộc nhóm $[18,5; 21,5)$ nên nhóm này chứa Q_3 . Do đó, $p = 4; a_4 = 18,5; m_4 = 24; m_1 + m_2 + m_3 = 3 + 12 + 15 = 30; a_5 - a_4 = 3$ và

$$\text{ta có } Q_3 = 18,5 + \frac{\frac{3 \cdot 56}{4} - 30}{24} \cdot 3 = 20.$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = 4,75$$

Câu 22. Ghi lại tốc độ bóng trong 200 lần giao bóng của một vận động viên môn quần vợt cho kết quả như bảng bên.

Tốc độ $v(km/h)$	Số lần
$150 \leq v < 155$	18
$155 \leq v < 160$	28

$160 \leq v < 165$	35
$165 \leq v < 170$	43
$170 \leq v < 175$	41
$175 \leq v < 180$	35

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 12,6.

B. 11,5.

C. 14,3.

D. 16,8.

Lời giải

Chọn A

Cỡ mẫu là $n = 200$.

Tứ phân vị thứ nhất Q_1 là $\frac{x_{50} + x_{51}}{2}$. Do x_{50}, x_{51} đều thuộc nhóm $[160; 165)$ nên nhóm này chứa Q_1 . Do đó, $p = 3$; $a_3 = 160$; $m_3 = 35$; $m_1 + m_2 = 18 + 28 = 46$; $a_4 - a_3 = 165 - 160 = 5$

$$\text{và ta có } Q_1 = 160 + \frac{\frac{200}{4} - 46}{35} \cdot 5 \approx 160,57$$

Tứ phân vị thứ ba Q_3 là $\frac{x_{150} + x_{151}}{2}$. Do x_{150}, x_{151} đều thuộc nhóm $[170; 175)$ nên nhóm này chứa Q_3 . Do đó, $p = 5$; $a_5 = 170$;

$m_5 = 41$; $m_1 + m_2 + m_3 + m_4 = 18 + 28 + 35 + 43 = 124$; $a_6 - a_5 = 175 - 170 = 5$ và ta có

$$Q_3 = 170 + \frac{\frac{3 \cdot 200}{4} - 124}{41} \cdot 5 \approx 173,17$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 \approx 12,6$

E. TRẮC NGHIỆM ĐÚNG SAI

Câu 1. Bạn Trang thống kê lại chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau.

Chiều cao (cm) \ Lớp	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
12C	2	7	12	3	0	1
12D	5	9	8	2	1	0

Các mệnh đề sau đúng hay sai?

a) Chiều cao cao nhất của các bạn học sinh trong lớp 12D là 185 (cm)

b) Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12C là: 30 (cm)

c) Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12D là: 25 (cm)

d) Chiều cao của học sinh lớp 12C có độ phân tán bé hơn

Lời giải

a) Sai	b) Đúng	c) Đúng	c) Sai
---------------	----------------	----------------	---------------

- a) Sai vì chú ý 180 là nửa đầu mút phải
- b) Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12C là: $185 - 155 = 30$ (cm)
- c) Khoảng biến thiên của chiều cao của các bạn học sinh nữ lớp 12D là: $180 - 155 = 25$ (cm)
- d) Vậy nếu so sánh theo khoảng biến thiên thì chiều cao của học sinh lớp 12C có độ phân tán lớn hơn

Câu 2. Bảng sau biểu diễn mẫu số liệu ghép nhóm thống kê mức lương của một công ty (đơn vị: triệu đồng).

Nhóm	Tần số
[10;15)	15
[15;20)	18
[20;25)	10
[25;30)	10
[30;35)	5
[35;40)	2
	$n = 60$

Các mệnh đề sau đúng hay sai?

- a)** Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 30$
- b)** Số phần tử của mẫu là $n = 60$
- c)** Tứ phân vị thứ nhất là: $Q_1 = 15$
- d)** Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $\Delta_Q = 3$

Lời giải

a) Đúng	b) Đúng	c) Đúng	c) Sai
----------------	----------------	----------------	---------------

- a)** Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 40 - 10 = 30$
- b) Số phần tử của mẫu là $n = 60$
- c) Tần số tích lũy của các nhóm lần lượt là $cf_1 = 15, cf_2 = 33, cf_3 = 43, cf_4 = 53, cf_5 = 58, cf_6 = 60$

Ta có: $\frac{n}{4} = \frac{60}{4} = 15$ suy ra nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

15. Xét nhóm 1 là nhóm [10;15] có $s = 10, h = 5, n_1 = 15$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{15 - cf_0}{n_1} \right) \cdot h = 10 + \left(\frac{15 - 0}{15} \right) \cdot 5 = 15$

d) Ta có: $\frac{3n}{4} = \frac{3 \cdot 60}{4} = 45$ mà $43 < 45 < 53$ suy ra nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 45. Xét nhóm 4 là nhóm $[25;30]$ có $t = 25, \neq 5, n_4 = 10$ và nhóm 3 là nhóm $[20;25]$ có $cf_3 = 43$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{45 - cf_3}{n_4} \right) \cdot h = 25 + \left(\frac{45 - 43}{10} \right) \cdot 5 = 26$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = 26 - 15 = 9$

Câu 3. Bảng sau biểu diễn mẫu số liệu ghép nhóm về độ tuổi của cư dân trong một khu phố.

Nhóm	Tần số
[20;30)	25
[30;40)	20
[40;50)	20
[50;60)	15
[60;70)	14
[70;80)	6
	$n = 100$

Các mệnh đề sau đúng hay sai?

a) Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 60$

b) Tứ phân vị thứ nhất là: $Q_1 = 35$

c) Tứ phân vị thứ ba là: $Q_3 = \frac{160}{3}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $\Delta_Q = \frac{65}{3}$

Lời giải

a) Đúng	b) Đúng	c) Sai	d) Đúng
---------	---------	--------	---------

a) Khoảng biến thiên của mẫu số liệu ghép nhóm trên là: $R = 80 - 20 = 60$

b) Số phần tử của mẫu là $n = 100$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 25, cf_2 = 45, cf_3 = 65, cf_4 = 80, cf_5 = 94, cf_6 = 100$

Ta có: $\frac{n}{4} = \frac{100}{4} = 25$ suy ra nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 25. Xét nhóm 1 là nhóm $[20;30]$ có $s = 20, h = 10, n_1 = 25$

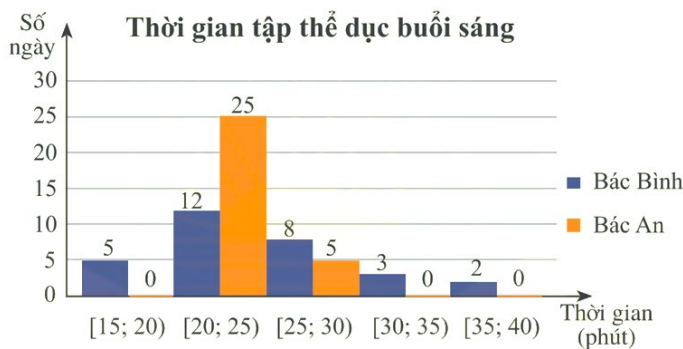
Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{25 - cf_0}{n_1} \right) \cdot h = 25 + \left(\frac{25 - 0}{25} \right) \cdot 10 = 35$

c) Ta có: $\frac{3n}{4} = \frac{3.100}{4} = 75$ mà $65 < 75 < 80$ suy ra nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 80. Xét nhóm 4 là nhóm $[50; 60]$ có $t = 50, l = 10, n_4 = 15$ và nhóm 3 là nhóm $[40; 50]$ có $cf_3 = 65$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{75 - cf_3}{n_4} \right) \cdot l = 50 + \left(\frac{75 - 65}{15} \right) \cdot 10 = \frac{170}{3}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = \frac{170}{3} - 35 = \frac{65}{3}$

Câu 4. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng 9/2022 của bác Bình và bác An.



Các mệnh đề sau đúng hay sai?

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

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An là: $\Delta_Q = 2$

c) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác Bình là: $Q_3' = \frac{455}{16}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An lớn hơn bác Bình

Lời giải

a) Đúng	b) Sai	c) Đúng	d) Sai
----------------	---------------	----------------	---------------

Ta có bảng sau

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

Cỡ mẫu $n = 30$;

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

b) 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 An được xếp theo thứ tự không giảm.

Ta có: $x_1; x_2; \dots; x_{25} \in [20; 25); x_{26}; \dots; x_{30} \in [25; 30)$;

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

$$\text{mẫu số liệu ghép nhóm là: } Q_1 = 20 + \frac{\frac{30}{4}}{25} (25 - 20) = \frac{43}{2}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{23} \in [20; 25)$. Do đó, tứ phân vị thứ ba của mẫu

$$\text{số liệu ghép nhóm là: } Q_3 = 20 + \frac{\frac{3.30}{4}}{25} (25 - 20) = \frac{49}{2}$$

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

c) 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 Bình được xếp theo thứ tự không giảm.

Ta có:

$y_1; y_2; \dots; y_5 \in [15; 20); y_6; \dots; y_{17} \in [20; 25); y_{18}; \dots; y_{25} \in [25; 30); y_{26}; y_{27}; y_{28} \in [30; 35);$

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

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

$$\text{mẫu số liệu ghép nhóm là: } Q'_1 = 20 + \frac{\frac{30}{4}}{12} (25 - 20) = \frac{185}{8}$$

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

$$\text{số liệu ghép nhóm là: } Q'_3 = 25 + \frac{\frac{3.30}{4} - (5 + 12)}{8} (30 - 25) = \frac{455}{16}$$

d) Vậy 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 5. Hằng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

Các mệnh đề sau đúng hay sai?

a) Cỡ mẫu $n = 100$.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = \frac{683}{38}$.

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta_Q = \frac{515}{114}$

d) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần ông Thắng đi hết hơn 29 phút. Thời gian của lần đi đó là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Lời giải

a) Đúng	b) Sai	c) Sai	c) Đúng
---------	--------	--------	---------

a) Cỡ mẫu $n = 100$.

b) Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc gồm thời gian 100 lần đi xe buýt của ông Thắng.

Ta có: $x_1, \dots, x_{22} \in [15; 18); x_{23}, \dots, x_{60} \in [18; 21); x_{61}, \dots, x_{87} \in [21; 24); x_{88}, \dots, x_{95} \in [24; 27); x_{96}, \dots, x_{99} \in [27; 30); x_{100} \in [30; 33)$.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{25} + x_{26}) \in [18; 21)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 18 + \frac{\frac{100}{4} - 22}{38} \cdot (21 - 18) = \frac{693}{38}.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [21; 24)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 21 + \frac{\frac{3 \cdot 100}{4} - (22 + 38)}{27} \cdot (24 - 21) = \frac{68}{3}.$$

c) Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là

$$\Delta_Q = \frac{68}{3} - \frac{693}{38} = \frac{505}{114}$$

d) Trong lần duy nhất ông Thắng đi hết hơn 29 phút, thời gian đi của ông thuộc nhóm $[30; 33)$. Vì $Q_3 + 1,5\Delta_Q = \frac{6683}{228} < 30$ nên thời gian của lần ông Thắng đi hết hơn 29 phút là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Câu 6. 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

Các mệnh đề sau đúng hay sai?

- a) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là: 15 (tuổi)
- b) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là: 12 (tuổi)
- c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm ứng với khu vực A là: $\frac{61}{3}$ (tuổi)
- d) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn

Lời giải

a) Đúng	b) Đúng	c) Sai	c) Đúng
---------	---------	--------	---------

a) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là: $34 - 19 = 15$ (tuổi)

b) Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là: $31 - 19 = 12$ (tuổi)

c) Cỡ mẫu $n = 100$

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

Ta có:

$$x_1; x_2; \dots; x_{10} \in [19; 22); x_{11}; \dots; x_{37} \in [22; 25); x_{38}; \dots; x_{68} \in [25; 28); x_{69}; \dots; x_{93} \in [28; 31); \\ x_{94}; \dots; x_{100} \in [31; 34)$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{25} + x_{26}) \in [22; 25)$. Do đó, tứ phân vị thứ

$$\text{nhất của mẫu số liệu ghép nhóm là: } Q_1 = 22 + \frac{\frac{100}{4} - 10}{27} (25 - 22) = \frac{71}{3}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [28; 31)$. Do đó, tứ phân vị thứ ba

$$\text{của mẫu số liệu ghép nhóm là: } Q_3 = 28 + \frac{\frac{3 \cdot 100}{4} - (10 + 27 + 31)}{25} (31 - 28) = \frac{721}{25}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = \frac{388}{75}$$

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

Ta có: $y_1; y_2; \dots; y_{47} \in [19; 22); y_{48}; \dots; y_{87} \in [22; 25); y_{88}; \dots; y_{98} \in [25; 30); y_{99}; y_{100} \in [28; 31)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(y_{25} + y_{26}) \in [19; 22)$. Do đó, tứ phân vị thứ

$$\text{nhất của mẫu số liệu ghép nhóm là: } Q_1' = 19 + \frac{\frac{100}{4}}{47} (22 - 19) = \frac{968}{47}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(y_{75} + y_{76}) \in [22; 25)$. Do đó, tứ phân vị thứ ba

$$\text{của mẫu số liệu ghép nhóm là: } Q_3' = 22 + \frac{\frac{3 \cdot 100}{40} - 47}{40} (25 - 22) = \frac{241}{10}$$

Có $\Delta_Q' < \Delta_Q$ nên phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn

Câu 7. 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

Các mệnh đề sau đúng hay sai?

a) Khoảng biến thiên của mẫu số liệu: $375,9(mm)$

b) Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = \frac{1827}{100}$

c) Chia mẫu số liệu trên thành 4 nhóm như bảng:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm: $400(mm)$

d) Chia mẫu số liệu trên thành 4 nhóm như bảng:

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = \frac{1000}{7}$

Lời giải

a) Đúng	b) Sai	c) Đúng	d) Đúng
----------------	---------------	----------------	----------------

a) Sắp xếp lại mẫu số liệu theo thứ tự tăng dần:

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: $522,9 - 147 = 375,9(mm)$

Tứ phân vị thứ nhất của mẫu số liệu là trung vị của 147; 187,1; 200,7; 242,2; 251,4; 258,4; 288,5; 298,1; 305; 332 nên: $Q_1 = \frac{6276}{25}$

Tứ phân vị thứ nhất của mẫu số liệu là trung vị của 341,4; 388,6; 400; 413,5; 413,5; 421; 432,2; 475; 520; 522,9 nên: $Q_3 = \frac{43281}{100}$

Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = \frac{18177}{100}$

b)

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

c) Khoảng biến thiên của mẫu số liệu ghép nhóm: $540 - 140 = 400(mm)$

Cỡ mẫu $n = 20$;

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 $\frac{1}{2}(x_5 + x_6) \in [240; 340)$. 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 $\frac{1}{2}(x_{15} + x_{16}) \in [340; 440)$. Do đó, tứ phân vị thứ

ba của mẫu số liệu ghép nhóm là: $Q_3' = 340 + \frac{\frac{3 \cdot 20}{4} - (3 + 7)}{7}(440 - 340) = \frac{2880}{7}$

d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = \frac{1000}{7}$

Khoảng biến thiên của mẫu số liệu ghép nhóm lớn hơn mẫu số liệu; khoảng tứ phân vị của mẫu số liệu ghép nhóm nhỏ hơn mẫu số liệu

Câu 8. 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

Các mệnh đề sau đúng hay sai?

a) Khoảng biên thiên của mẫu số liệu ghép nhóm: $1(m)$

b) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm là: $Q_3 = 10,5$

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 2,06$

d) Trong 100 cây keo trên có 1 cây cao $8,4m$. Thì chiều cao của cây keo này là giá trị ngoại lệ

Lời giải

a) Đúng	b) Đúng	c) Sai	c) Sai
----------------	----------------	---------------	---------------

a) Khoảng biên thiên của mẫu số liệu ghép nhóm: $9,4 - 8,4 = 1(m)$

b) Cỡ mẫu $n = 100$;

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc 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 $\frac{1}{2}(x_{25} + x_{26}) \in [8, 8; 9, 0)$. Do đó, tứ phân vị thứ

$$\text{nhất của mẫu số liệu ghép nhóm là: } Q_1 = 8,8 + \frac{\frac{100}{4} - (5 + 12)}{25} (9,0 - 8,8) = 9,44$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [9, 0; 9, 2)$. Do đó, tứ phân vị thứ

$$\text{ba của mẫu số liệu ghép nhóm là: } Q_3 = 9,0 + \frac{\frac{3 \cdot 100}{4} - (5 + 12 + 25)}{44} (9,2 - 9,0) = 10,5$$

$$\text{c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = Q_3 - Q_1 = 1,06$$

$$\text{d) Giá trị } x \text{ trong mẫu số liệu là giá trị ngoại lệ nếu } x > Q_3 + 1,5\Delta_Q \text{ hoặc } x < Q_1 - 1,5\Delta_Q$$

$$\text{Hay } x > 10,5 + 1,5 \cdot 1,06 = 12,09 \text{ hoặc } x < 9,44 - 1,5 \cdot 1,06 = 7,85$$

Vậy cây cao 8,4m không phải là giá trị ngoại lệ

Câu 9. 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

Các mệnh đề sau đúng hay sai?

$$\text{a) Khoảng biến thiên của mẫu số liệu là } R = 4$$

$$\text{b) Số trung bình của mẫu số liệu xấp xỉ bằng: } 8,12$$

$$\text{c) Một của mẫu số liệu là: } 6,21$$

$$\text{d) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = 2,05$$

Lời giải

a) Đúng	b) Đúng	c) Sai	d) Sai
---------	---------	--------	--------

$$\text{a) Khoảng biến thiên của mẫu số liệu là } R = 10 - 6 = 4$$

b)

Khoảng điểm	[6,5;7)	[7;7,5)	[7,5;8)	[8;8,5)	[8,5;9)	[9;9,5)	[9,5;10)
Giá trị đại diện	6,75	7,25	7,75	8,25	8,75	9,25	9,75
Tần số	8	10	16	24	13	7	4

Số trung bình của mẫu số liệu xấp xỉ bằng:

$$(6,75.8 + 7,25.10 + 7,75.16 + 8,25.24 + 8,75.13 + 9,25.7 + 9,75.4) : 82 = 8,12$$

Nhóm chứa một của mẫu số liệu là: $[8;8,5)$

c) Một của mẫu số liệu là:

$$M_0 = 8 + \frac{24-16}{(24-16)+(24-13)} \cdot (8,5-8) = 8,21$$

d) Gọi $x_1; x_2; x_3; \dots; x_{85}$ lần lượt là tần số theo thứ tự không gian

Do $x_1, \dots, x_8 \in [6,5;7); x_9, \dots, x_{18} \in [7;7,5); x_{19}, \dots, x_{34} \in [7,5;8);$

$x_{35}, \dots, x_{58} \in [8;8,5); x_{59}, \dots, x_{71} \in [8,5;9); \dots$

Tứ phân vị thứ hai của dãy số liệu là $\frac{1}{2}(x_{41} + x_{42})$ thuộc nhóm $[8;8,5)$ nên tứ phân vị thứ

$$\text{hai của mẫu số liệu là } Q_2 = 8 + \frac{\frac{82}{2} - 34}{24} (8,5 - 8) = 8,15$$

Tứ phân vị thứ nhất của dãy số liệu là $\frac{1}{2}(x_{20} + x_{21})$ thuộc nhóm $[7,5;8)$ nên tứ phân vị thứ

$$\text{nhất của mẫu số liệu là } Q_1 = 7,5 + \frac{\frac{82}{2} - 18}{16} (8 - 7,5) = 7,58$$

Tứ phân vị thứ ba của dãy số liệu là $\frac{1}{2}(x_{61} + x_{62})$ thuộc nhóm $[8,5;9)$ nên tứ phân vị thứ

$$\text{ba của mẫu số liệu là } Q_3 = 8,5 + \frac{\frac{3.82}{2} - 58}{13} (9 - 8,5) = 8,63$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 1,05$

Câu 10. Một công ty cung cấp nước sạch thống kê lượng nước các hộ gia đình trong một khu vực tiêu thụ trong một tháng ở bảng sau:

Lượng nước tiêu thụ (m^3)	$[3;6)$	$[6;9)$	$[9;12)$	$[12;15)$	$[15;18)$
Số hộ gia đình	24	57	42	29	8

Các mệnh đề sau đúng hay sai?

a) Số trung bình của mẫu số liệu ghép nhóm đã cho là 9,375.

b) Một của mẫu số liệu là $M_0 = 8,0625$.

c) Khoảng biến thiên của mẫu số liệu ghép nhóm đã cho là 15

d) Công ty muốn gửi một thông báo khuyến nghị tiết kiệm nước đến 25% các hộ gia đình có lượng nước tiêu thụ cao nhất. Khi đó công ty nên gửi thông báo tiết kiệm nước đến các hộ gia đình có lượng nước tiêu thụ từ $14,79 m^3$ nước trở lên.

Lời giải

a) Đúng	b) Đúng	c) Đúng	c) Sai
----------------	----------------	----------------	---------------

a)

Cỡ mẫu $n = 160$.

Bảng tần số ghép nhóm của mẫu số liệu trên như sau:

Lượng nước tiêu thụ (m^3)	[3;6)	[6;9)	[9;12)	[12;15)	[15;18)
Giá trị đại diện	4,5	7,5	10,5	13,5	16,5
Số hộ gia đình	24	57	42	29	8

Số trung bình của mẫu số liệu ghép nhóm đã cho là

$$\bar{x} = \frac{1}{160} (24 \cdot 4,5 + 57 \cdot 7,5 + 42 \cdot 10,5 + 29 \cdot 13,5 + 8 \cdot 16,5) = 9,375.$$

b) Nhóm chứa một của mẫu số liệu ghép nhóm trên là nhóm [6;9).

Do đó: $u_m = 6; n_m = 57; n_{m-1} = 24; n_{m+1} = 42; u_{m+1} = 9$.

Mốt của mẫu số liệu là $M_o = 6 + \frac{(57-24)}{(57-24) + (57-42)} \cdot (9-6) = 8,0625$.

c) Gọi $x_1; x_2; \dots; x_{160}$ là mẫu số liệu được xếp theo thứ tự không giảm.

Ta có: $x_1, \dots, x_{24} \in [3;6); x_{25}, \dots, x_{81} \in [6;9); x_{82}, \dots, x_{123} \in [9;12);$

$x_{124}, \dots, x_{152} \in [12;15); x_{153}, \dots, x_{160} \in [15;18).$

Cỡ mẫu $n = 160$ là số chẵn nên trung vị là $M_e = \frac{1}{2}(x_{80} + x_{81})$.

Do x_{80} và x_{81} thuộc nhóm [6;9) nên trung vị của mẫu số liệu là $M_e = 6 + \frac{\frac{160}{2} - 24}{57} \cdot (9-6) \approx 8,95$.

d) 25% các hộ gia đình có lượng nước tiêu thụ cao nhất có lượng nước tiêu thụ không nhỏ hơn Q_3 , với Q_3 là tứ phân vị thứ ba của mẫu số liệu.

Tứ phân vị thứ ba của dãy số liệu $x_1; x_2; \dots; x_{160}$ là $\frac{1}{2}(x_{120} + x_{121})$. Do x_{120} và x_{121} thuộc nhóm [9;12)

nên tứ phân vị thứ ba của mẫu số liệu là $Q_3 = 9 + \frac{\frac{160 \cdot 3}{4} - (24 + 57)}{42} \cdot (12-9) \approx 11,79$.

Vậy công ty nên gửi thông báo tiết kiệm nước đến các hộ gia đình có lượng nước tiêu thụ từ $11,79 m^3$ nước trở lên.

F. TRẢ LỜI NGẮN

Câu 1. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về chiều cao (đơn vị: centimét) của 36 học sinh nam lớp 12 ở một trường trung học phổ thông. Tìm khoảng biến thiên của mẫu số liệu ghép nhóm đó.

Nhóm	Tần số
[160;163)	6
[163;166)	11
[166;169)	9
[169;172)	7
[172;175)	3
	$n = 36$

Trả lời: 15

Lời giải

Trong mẫu số liệu ghép nhóm đó, ta có: đầu mút trái của nhóm 1 là $a_1 = 160$, đầu mút phải của nhóm 5 là $a_6 = 175$. Vậy khoảng biến thiên của mẫu số liệu ghép nhóm đó là:

$$R = a_6 - a_1 = 175 - 160 = 15(cm).$$

Câu 2. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về chiều cao của 42 mẫu cây ở một vườn thực vật (đơn vị: centimét). Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm đó (làm tròn kết quả đến hàng phần mười nếu cần).

Nhóm	Tần số	Tần số tích lũy
[40 ; 45)	5	5
[45 ; 50)	10	15
[50 ; 55)	7	22
[55 ; 60)	9	31
[60 ; 65)	7	38
[65 ; 70)	4	42
	$n = 42$	

Bảng 7

Trả lời: 12,65

Lời giải

Số phần tử của mẫu là $n = 42$.

- Ta có: $\frac{n}{4} = \frac{42}{4} = 10,5$ mà $5 < 10,5 < 15$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 10,5. Xét nhóm 2 là nhóm $[45; 50)$ có $s = 45; h = 5; n_2 = 10$ và nhóm 1 là nhóm $[40; 45)$ có $cf_1 = 5$.

Áp dụng công thức, ta có tứ phân vị thứ nhất là:

$$Q_1 = 45 + \left(\frac{10,5 - 5}{10} \right) \cdot 5 = 47,75(cm).$$

- Ta có: $\frac{3n}{4} = \frac{3.42}{4} = 31,5$ mà $31 < 31,5 < 38$. Suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 31,5. Xét nhóm 5 là nhóm $[60; 65)$ có $t = 60; l = 5; n_5 = 7$ và nhóm 4 là nhóm $[55; 60)$ có $cf_4 = 31$.

Áp dụng công thức, ta có tứ phân vị thứ ba là:

$$Q_3 = 60 + \left(\frac{31,5 - 31}{7} \right) \cdot 5 \approx 60,4 (cm).$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm đã cho là:

$$\Delta_Q = Q_3 - Q_1 \approx 60,4 - 47,75 = 12,65 (cm).$$

Câu 3. Bảng sau thống kê cân nặng của 50 quả xoài Thanh Ca được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường. Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm đó

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

Trả lời: 63,5

Lời giải

Cỡ mẫu $n = 50$.

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc gồm cân nặng của 50 quả xoài được xếp theo thứ tự không giảm.

Ta có: $x_1, x_2, x_3 \in [250; 290); x_4, \dots, x_{16} \in [290; 330); x_{17}, \dots, x_{34} \in [330; 370); x_{35}, \dots, x_{45} \in [370; 410); x_{46}, \dots, x_{50} \in [410; 450)$.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_{13} \in [290; 330)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 290 + \frac{\frac{50}{4} - 3}{13} \cdot (330 - 290) = \frac{4150}{13}.$$

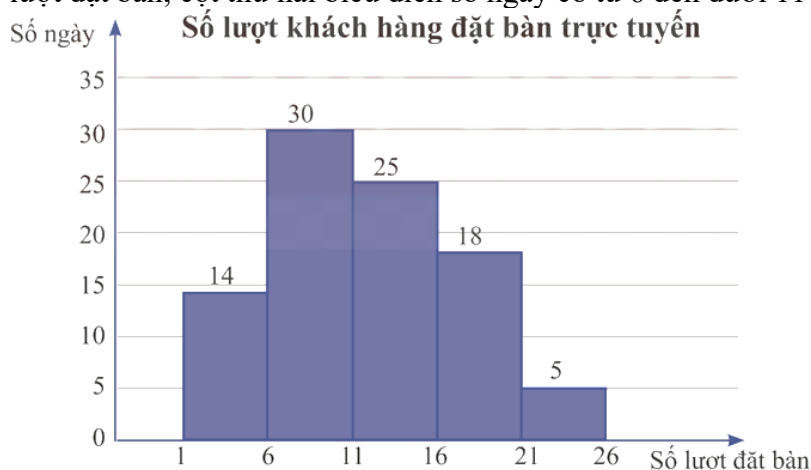
Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{38} \in [370; 410)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 370 + \frac{\frac{3.50}{4} - (3 + 13 + 18)}{11} \cdot (410 - 370) = \frac{4210}{11}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là

$$\Delta_Q = \frac{4210}{11} - \frac{4150}{13} = \frac{9080}{143} \approx 63,5$$

Câu 4. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 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.

Trả lời: 8,5

Lời giải

Cỡ mẫu $n = 92$;

Gọi $x_1; x_2; \dots; x_{92}$ là mẫu số liệu gốc về số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 của 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 $\frac{1}{2}(x_{23} + x_{24}) \in [6; 11)$. Do đó, tứ phân vị thứ

nhất của mẫu số liệu ghép nhóm là: $Q_1 = 6 + \frac{\frac{92}{2} - 14}{30}(11 - 6) = 7,5$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{69} + x_{70}) \in [11; 16)$ và $[16; 21)$. 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: $\Delta_Q = Q_3 - Q_1 = 8,5$

Câu 5. Thời gian luyện tập trong một ngày (tính theo giờ) của một số vận động viên được ghi lại ở bảng sau:

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

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

Trả lời: 1,63

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

$$\text{của mẫu số liệu ghép nhóm là } Q_2 = 4 + \frac{\frac{2.39}{4} - (3+8)}{12} \cdot (6-4) = \frac{65}{12} \approx 5,417.$$

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

$$\text{của mẫu số liệu ghép nhóm là } Q_1 = 2 + \frac{\frac{1.39}{4} - 3}{8} \cdot (4-2) = \frac{59}{16} = 3,6875.$$

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

$$\text{mẫu số liệu ghép nhóm là } Q_3 = 6 + \frac{\frac{3.39}{4} - (3+8+12)}{12} \cdot (8-6) = \frac{169}{24} \approx 7,042.$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 1,625$

Câu 6. 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 tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

Trả lời: 116,3

Lời giải

Số lần thực hiện cuộc gọi là: $n = 8 + 10 + 7 + 5 + 2 + 1 = 33$

Gọi $x_1; x_2; x_3; \dots; x_{33}$ lần lượt là thời gian thực hiện cuộc gọi theo thứ tự không gian

Do $x_1, \dots, x_8 \in [0; 60); x_9, \dots, x_{18} \in [60; 120); x_{19}, \dots, x_{25} \in [120; 180);$
 $x_{26}, \dots, x_{30} \in [180; 240); \dots$

Tứ phân vị thứ hai của dãy số liệu là $\frac{1}{2}(x_{16} + x_{17})$ thuộc nhóm $[60; 120)$ nên tứ phân vị thứ

$$\text{hai của mẫu số liệu là } Q_2 = 60 + \frac{\frac{33}{2} - 8}{10} (120 - 60) = 111$$

Tứ phân vị thứ nhất của dãy số liệu là $\frac{1}{2}(x_8 + x_9)$ thuộc nhóm $[60;120)$ nên tứ phân vị thứ

$$\text{nhất của mẫu số liệu là } Q_1 = 60 + \frac{\frac{33}{2} - 8}{10}(120 - 60) = 61,5$$

Tứ phân vị thứ ba của dãy số liệu là $\frac{1}{2}(x_{24} + x_{25})$ thuộc nhóm $[120;180)$ nên tứ phân vị thứ

$$\text{ba của mẫu số liệu là } Q_3 = 120 + \frac{\frac{3.33}{2} - 18}{7}(180 - 120) = 177,8$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 116,3$

Câu 7. Số điểm một cầu thủ bóng rổ ghi được trong 20 trận đấu được cho ở bảng sau:

Điểm số	$[5,5;10,5)$	$[10,5;15,5)$	$[15,5;20,5)$	$[20,5;25,5)$
Số trận	3	9	2	6

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

Trả lời: 9,7

Lời giải

Gọi $x_1; x_2; x_3; \dots; x_{20}$ lần lượt là số trận theo thứ tự không gian

Do $x_1, \dots, x_3 \in [5,5;10,5); x_4, \dots, x_{12} \in [10,5;15,5); x_{13}, x_{14} \in [15,5;20,5);$
 $x_{15}, \dots, x_{20} \in [20,5;25,5)$

Tứ phân vị thứ hai của dãy số liệu là $\frac{1}{2}(x_{10} + x_{11})$ thuộc nhóm $[10,5;15,5)$ nên tứ phân vị

$$\text{thứ hai của mẫu số liệu là } Q_2 = 10,5 + \frac{\frac{20}{2} - 3}{9}(15,5 - 10,5) = 14,4$$

Tứ phân vị thứ nhất của dãy số liệu là $\frac{1}{2}(x_5 + x_6)$ thuộc nhóm $[10,5;15,5)$ nên tứ phân vị

$$\text{thứ nhất của mẫu số liệu là } Q_1 = 10,5 + \frac{\frac{20}{2} - 3}{9}(15,5 - 10,5) = 11,6$$

Tứ phân vị thứ ba của dãy số liệu là $\frac{1}{2}(x_{15} + x_{16})$ thuộc nhóm $[20,5;25,5)$ nên tứ phân vị

$$\text{thứ ba của mẫu số liệu là } Q_3 = 20,5 + \frac{\frac{3.20}{2} - 14}{6}(25,5 - 20,5) = 21,3$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 9,7$

Câu 8. Thời gian sử dụng điện thoại trong một ngày của 30 sinh viên được ghi lại ở bảng sau (đơn vị: phút).

Thời gian (phút)	[0; 60)	[60; 120)	[120; 180)	[180; 240)	[240; 300)
Số sinh viên	2	7	7	10	4

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

Trả lời: 111,86

Lời giải

Gọi $x_1; x_2; \dots; x_{30}$ là mẫu số liệu được xếp theo thứ tự không giảm.

Ta có: $x_1, x_2 \in [0; 60); x_3, \dots, x_9 \in [60; 120); x_{10}, \dots, x_{16} \in [120; 180);$

$x_{17}, \dots, x_{26} \in [180; 240); x_{27}, \dots, x_{30} \in [240; 300).$

Tứ phân vị thứ hai của mẫu số liệu ghép nhóm là $\frac{1}{2}(x_{15} + x_{16})$. Do x_{15} và x_{16} thuộc nhóm [120; 180) nên tứ phân vị thứ hai của mẫu số liệu ghép nhóm là

$$Q_2 = 120 + \frac{\frac{30}{2} - (2 + 7)}{7} \cdot (180 - 120) = \frac{1200}{7} \approx 171,43.$$

Tứ phân vị thứ nhất của mẫu là $x_8 \in [60; 120)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép

$$\text{nhóm là } Q_1 = 60 + \frac{\frac{30}{4} - 2}{7} \cdot (120 - 60) = \frac{750}{7} \approx 107,14.$$

Tứ phân vị thứ ba của mẫu số liệu là $x_{23} \in [180; 240)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là } Q_3 = 180 + \frac{\frac{3 \cdot 30}{4} - (2 + 7 + 7)}{10} \cdot (240 - 180) = 219.$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 111,86$

Câu 9. Điểm thi môn Toán (thang điểm 100, điểm được làm tròn đến 1) của 60 thí sinh được cho mẫu số liệu ghép nhóm như sau:

Điểm	[0; 9,5)	[9,5; 19,5)	[19,5; 29,5)	[29,5; 39,5)	[39,5; 49,5)
Số thí sinh	1	2	4	6	15
Điểm	[49,5; 59,5)	[59,5; 69,5)	[69,5; 79,5)	[79,5; 89,5)	[89,5; 99,5)
Số thí sinh	12	10	6	3	1

Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm

Trả lời: 23,67

Lời giải

Cỡ mẫu là $n = 60$.

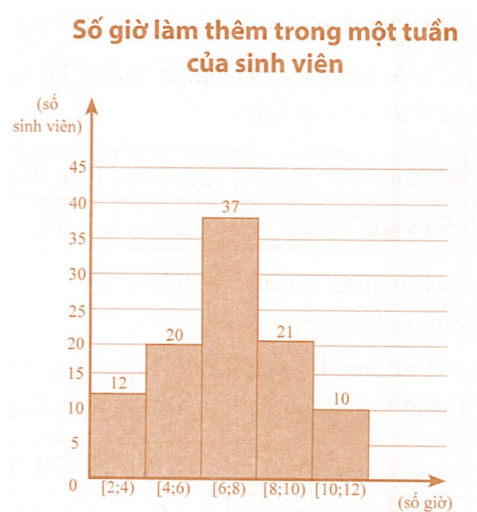
Gọi $x_1, x_2, \dots, x_{60}$ là điểm thi môn Toán của 60 thí sinh và giả sử dãy này đã được sắp xếp theo thứ tự tăng dần.

Tứ phân vị thứ nhất Q_1 là $\frac{x_{15} + x_{16}}{2}$. Do x_{15} và x_{16} đều thuộc nhóm $[39,5; 49,5)$ nên nhóm này chứa Q_1 . Do đó, $p = 5$; $a_5 = 39,5$; $m_5 = 15$; $m_1 + m_2 + m_3 + m_4 = 13$; $a_6 - a_5 = 10$ và ta có $Q_1 = 39,5 + \frac{\frac{60}{4} - 13}{15} \cdot 10 \approx 40,83$

Tứ phân vị thứ ba Q_3 là $\frac{x_{45} + x_{46}}{2}$. Do x_{45} và x_{46} đều thuộc nhóm $[59,5; 69,5)$ nên nhóm này chứa Q_3 . Do đó, $p = 7$; $a_7 = 59,5$; $m_7 = 10$; $m_1 + m_2 + m_3 + m_4 + m_5 + m_6 = 40$; $a_6 - a_5 = 10$ và ta có $Q_3 = 59,5 + \frac{\frac{3 \cdot 60}{4} - 40}{10} \cdot 10 = 64,5$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = 23,67$

Câu 10. Kết quả điều tra về số giờ làm thêm trong một tuần của 100 sinh viên được cho ở biểu đồ bên.



Tìm khoảng tứ phân vị của số liệu đó.

Trả lời: 3,27

Lời giải

Từ mẫu số liệu ghép nhóm, ta có bảng thống kê số giờ làm thêm trong một tuần của 100 sinh viên như sau:

Số giờ làm thêm	[2;4)	[4;6)	[6;8)	[8;10)	[10;12)
Số giờ làm thêm đại diện	3	5	7	9	11
Số sinh viên	12	20	37	21	10

Cỡ mẫu $n = 100$.

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

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} và x_{51} thuộc nhóm $[6; 8)$ nên $Q_2 = 6 + \frac{\frac{2 \cdot 100}{4} - 32}{37} \cdot (8 - 6) \approx 6,97$.

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 $[4; 6)$

nên $Q_1 = 4 + \frac{\frac{1 \cdot 100}{4} - 12}{20} \cdot (6 - 4) = 5,3$.

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} và x_{76} thuộc nhóm $[8; 10)$

nên $Q_3 = 8 + \frac{\frac{3 \cdot 100}{4} - 69}{21} \cdot (10 - 8) \approx 8,57$.

Khoảng tứ phân vị là $\Delta Q = 3,27$

CHƯƠNG 3: CÁC SỐ ĐẶC TRƯNG ĐO ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM

BÀI 2. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHÉP NHÓM.....	2
A. KIẾN THỨC CƠ BẢN CẦN NẮM.....	2
B. GIẢI BÀI TẬP SÁCH GIÁO KHOA	4
C. CÁC DẠNG TOÁN.....	5
D. TRẮC NGHIỆM 4 PHƯƠNG ÁN	7
E. TRẮC NGHIỆM ĐÚNG SAI.....	9
F. TRẢ LỜI NGẮN	14

BÀI 2. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHEP NHÓM

A. KIẾN THỨC CƠ BẢN CẦN NẮM

1. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Ở lớp 11, ta đã biết giá trị đại diện của nhóm $[a; b)$ là $c = \frac{1}{2}(a + b)$.

Tương tự như cách tính số trung bình, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm có thể được xác định thông qua giá trị đại diện và tần số của mỗi nhóm.

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Giá trị đại diện	c_1	c_2	...	c_k
Tần số	n_1	n_2	...	n_k

Phương sai của mẫu số liệu ghép nhóm, kí hiệu S^2 , được tính bởi công thức:

$$S^2 = \frac{1}{n} \left[n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right]$$

trong đó: $n = n_1 + n_2 + \dots + n_k$ là cỡ mẫu:

$$\bar{x} = \frac{1}{n} (n_1 c_1 + n_2 c_2 + \dots + n_k c_k) \text{ là } \text{so trung bình.}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm, kí hiệu S , là căn bậc hai số học của phương sai.

Chú ý:

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

$$S^2 = \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + \dots + n_k c_k^2) - \bar{x}^2$$

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

$$S^2 = \frac{1}{n-1} \left[n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right].$$

Ví dụ 1: Cân nặng của một số quả mít trong một khu vườn được thống kê ở bảng sau:

Cân nặng (kg)	[4; 6)	[6; 8)	[8; 10)	[10; 12)	[12; 14)
Số quả mít	6	12	19	9	4

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Ví dụ 2: Thống kê tổng số giờ nắng trong tháng 9 tại một trạm quan trắc đặt ở Cà Mau trong các năm từ 2002 đến 2021 được thống kê như sau:

111,6 134,9 130,3 134,2 140,9 109,3 154,4 156,3 116,1 96,7
105,2 80,8 80,8 110 109 139 145 161 126 114

a) Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu trên.

b) Hãy lập bảng tần số ghép nhóm với nhóm đầu tiên là $[80; 98)$ và độ dài mỗi nhóm bằng 18. Tính phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm.

c) Hãy tính sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc.

(Kết quả các phép tính làm tròn đến hàng phần nghìn.)

Ý nghĩa của phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm

- Phương sai của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho phương sai của mẫu số liệu gốc. Độ lệch chuẩn của mẫu số liệu ghép nhóm cũng là giá trị xấp xỉ cho độ lệch chuẩn của mẫu số liệu gốc. Chúng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm. Phương sai và độ lệch chuẩn càng lớn thì dữ liệu càng phân tán.
- Độ lệch chuẩn có cùng đơn vị với đơn vị của mẫu số liệu.

Ví dụ 3: Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11 A và 11B ở bảng sau:

Điểm trung bình \ Lớp	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
11A	1	0	11	22	6
11B	0	6	8	14	12

a) Nếu so sánh theo khoảng biến thiên thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

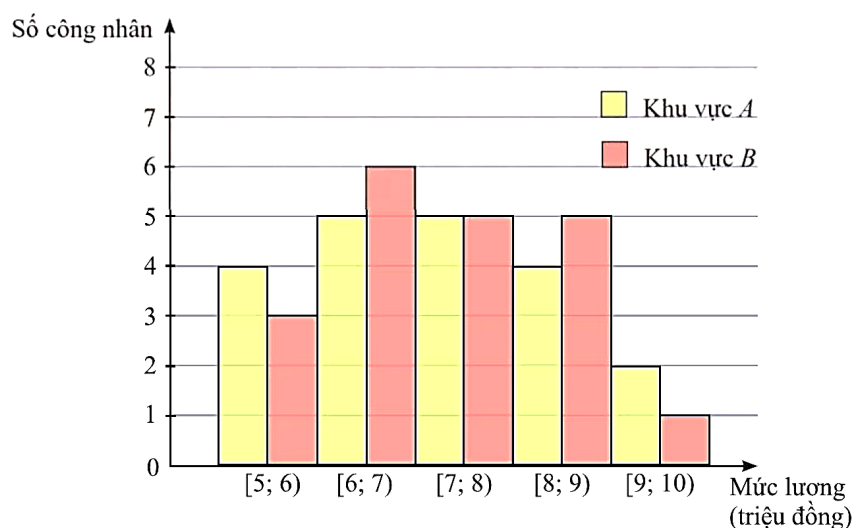
b) Nếu so sánh theo độ lệch chuẩn thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

Ví dụ 4: Biểu đồ dưới đây mô tả kết quả điều tra về mức lương khởi điểm (đơn vị: triệu đồng) của một số công nhân ở hai khu vực A và B.

a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu đó.

b) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì công nhân ở khu vực nào có mức lương khởi điểm đồng đều hơn?

Mức lương khởi điểm của công nhân ở hai khu vực A và B



Ví dụ 5: Giá đóng cửa của một cổ phiếu là giá của cổ phiếu đó cuối một phiên giao dịch. Bảng sau thống kê giá đóng cửa (đơn vị: nghìn đồng) của hai mã cổ phiếu A và B trong 50 ngày giao dịch liên tiếp.

Giá đóng cửa	[120; 122)	[122; 124)	[124; 126)	[126; 128)	[128; 130)
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Người ta có thể dùng phương sai và độ lệch chuẩn để so sánh mức độ rủi ro của các loại cổ phiếu có giá trị trung bình gần bằng nhau. Cổ phiếu nào có phương sai, độ lệch chuẩn cao hơn thì được coi là có độ rủi ro lớn hơn.

Theo quan điểm trên, hãy so sánh độ rủi ro của cổ phiếu A và cổ phiếu B .

Trả lời câu hỏi Thực hành 1 trang 82 SGK Toán 12 Chân trời sáng tạo

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm ở Khởi động

Chiều cao (cm)	[160;164)	[164;168)	[168;172)	[172;176)	[176;180)
Số học sinh	3	5	8	4	1

Trả lời câu hỏi Thực hành 2 trang 82 SGK Toán 12 Chân trời sáng tạo

Mai và Ngọc 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:

- Hãy tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.
- Nếu so sánh theo độ lệch chuẩn thì bạn nào có số lượng bước chân đi mỗi ngày đều đặn hơn?

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

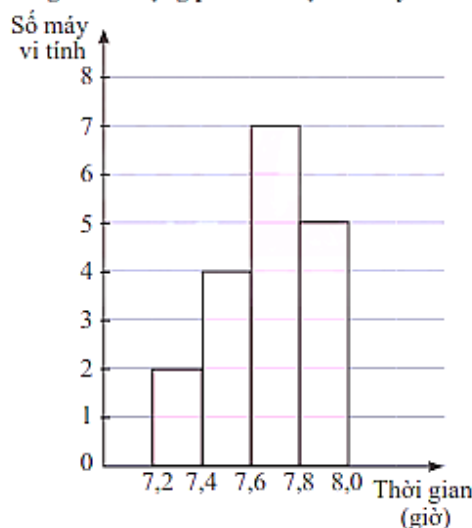
Bài 1: Bảng dưới đây thống kê cự li ném tạ của một vận động viên.

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.

Bài 2:

Thời gian sử dụng pin của một số máy vi tính



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.

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

Bài 3: 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 4: 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?

C. CÁC DẠNG TOÁN

Câu 1. 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)	$[6; 7)$	$[7; 8)$	$[8; 9)$	$[9; 10)$	$[10; 11)$
Học sinh trường X	8	10	13	10	9
Học sinh trường Y	4	12	17	14	3

- a) Nếu so sánh theo số trung bình thì học sinh trường nào viết nhanh hơn?
b) Nếu so sánh theo khoảng tứ phân vị thì học sinh trường nào có tốc độ viết đồng đều hơn?
c) Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

Câu 2. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi hai phân xưởng được cho như sau:

Tuổi thọ (năm)	$[1, 5; 2)$	$[2; 2, 5)$	$[2, 5; 3)$	$[3; 3, 5)$	$[3, 5; 4)$
Số linh kiện của phân xưởng 1	4	9	13	8	6
Số linh kiện của phân xưởng 2	2	8	20	7	3

Tính phương sai và độ lệch chuẩn của mỗi mẫu số liệu ghép nhóm và nhận xét về độ phân tán của tuổi thọ các linh kiện điện tử được sản xuất bởi mỗi phân xưởng.

Câu 3. Một nhóm 20 học sinh dùng một thiết bị đo đường kính của một nhân tế bào cho kết quả như sau:

Kết quả đo (μm)	$[4, 5; 5)$	$[5; 5, 5)$	$[5, 5; 6)$	$[6; 6, 5)$
Số học sinh	3	8	7	2

- a) Tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.
b) Số trung bình và độ lệch chuẩn cho biết thông tin gì?

Câu 4. Thời gian chạy tập luyện cự li 100 m của hai vận động viên được cho trong bảng sau:

Thời gian (giây)	[10;10,3)	[10,3;10,6)	[10,6;10,9)	[10,9;11,2)
Số lần chạy của A	2	10	5	3
Số lần chạy của B	3	7	9	6

Dựa trên độ lệch chuẩn của các mẫu số liệu ghép nhóm, hãy cho biết vận động viên nào có thành tích luyện tập ổn định hơn.

Câu 5. Có nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán của hai mẫu số liệu ghép nhóm trong mỗi trường hợp sau không? Tại sao?

- Các mẫu số liệu ghép nhóm về điểm thi tốt nghiệp môn Toán của học sinh hai trường trung học phổ thông có chất lượng tương đương.
- Các mẫu số liệu ghép nhóm về doanh thu của 100 cửa hàng bán lẻ và doanh thu của 100 siêu thị.

Câu 6. Mai và Ngọc 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)
Mai	6	7	6	6	5
Ngọc	2	5	13	8	2

- Hãy tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.
- Nếu so sánh theo độ lệch chuẩn thì bạn nào có số lượng bước chân đi mỗi ngày đều đặn hơn?

Câu 7. 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

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

Câu 8. 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

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

Câu 9. Trong bài 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 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 vốn kế của bạn nào cho kết quả đo ổn định hơn.

Câu 10. Điều tra thời gian phải làm thêm trung bình hằng tuần của các bác sĩ ở một bệnh viện, người ta thu được số liệu sau:

5	6	7	8	8	8	9	9	9
9	9	10	10	10	10	11	11	11
12	12	12	13	13	14	14		

a) Chuyển mẫu số liệu đã cho về mẫu số liệu ghép nhóm với độ dài các nhóm ghép bằng 2 và nhóm đầu tiên là $[5; 7)$.

b) Tính số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm lập ở câu a (làm tròn kết quả đến hàng phần trăm).

c) So sánh và nêu ý nghĩa các kết quả tìm được ở câu b với các kết quả tương ứng của mẫu số liệu gốc.

Câu 11. 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?

D. TRẮC NGHIỆM 4 PHƯƠNG ÁN

Câu 1. 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.

Câu 2. 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 9. 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
---------	---	---	---	---	---

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. $[13;15)$.

Câu 4. Một siêu thị thống kê số tiền (đơn vị: chục nghìn đồng) mà 44 khách hàng mua hàng ở siêu thị đó trong một ngày. Số liệu được ghi lại trong Bảng 18.

Nhóm	Giá trị đại diện	Tần số
$[40 ; 45)$	42,5	4
$[45 ; 50)$	47,5	14
$[50 ; 55)$	52,5	8
$[55 ; 60)$	57,5	10
$[60 ; 75)$	62,5	6
$[65 ; 70)$	67,5	2
		$n = 44$

Bảng 18

Phương sai của mẫu số liệu ghép nhóm trên là:

- A. 53,2. B. 46,1. C. 30. D. 11.

Câu 5. Một siêu thị thống kê số tiền (đơn vị: chục nghìn đồng) mà 44 khách hàng mua hàng ở siêu thị đó trong một ngày. Số liệu được ghi lại trong Bảng 18.

Nhóm	Giá trị đại diện	Tần số
$[40 ; 45)$	42,5	4
$[45 ; 50)$	47,5	14
$[50 ; 55)$	52,5	8
$[55 ; 60)$	57,5	10
$[60 ; 75)$	62,5	6
$[65 ; 70)$	67,5	2
		$n = 44$

Bảng 18

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng đơn vị) là:

- A. 6,8. B. 7,3. C. 3,3. D. 46,1.

Câu 6. 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

Phương sai của mẫu số liệu ghép nhóm là

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

Câu 7. 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

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 8. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

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 9. 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

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.

E. TRẮC NGHIỆM ĐÚNG SAI

Câu 1. Ta có các bảng thống kê sau:

Nhóm	Giá trị đại diện	Tần số
[6,22 ; 6,46)	6,34	3
[6,46 ; 6,70)	6,58	7
[6,70 ; 6,94)	6,82	5
[6,94 ; 7,18)	7,06	20
[7,18 ; 7,42)	7,30	5
		$n = 40$

Bảng 15

Nhóm	Giá trị đại diện	Tần số
[6,22 ; 6,46)	6,34	2
[6,46 ; 6,70)	6,58	5
[6,70 ; 6,94)	6,82	8
[6,94 ; 7,18)	7,06	19
[7,18 ; 7,42)	7,30	6
		$n = 40$

Bảng 16

Các mệnh đề sau đúng hay sai?

- a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 (làm tròn kết quả đến hàng phần trăm) là: $\approx 6,92(m)$.
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 (làm tròn kết quả đến hàng phần trăm) là: $0,26(m)$.
- c) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho bởi Bảng 16 (làm tròn kết quả đến hàng phần trăm) là: $0,16$.
- d) Kết quả nhảy xa của vận động viên Dũng đồng đều hơn kết quả nhảy xa của vận động viên Huy.

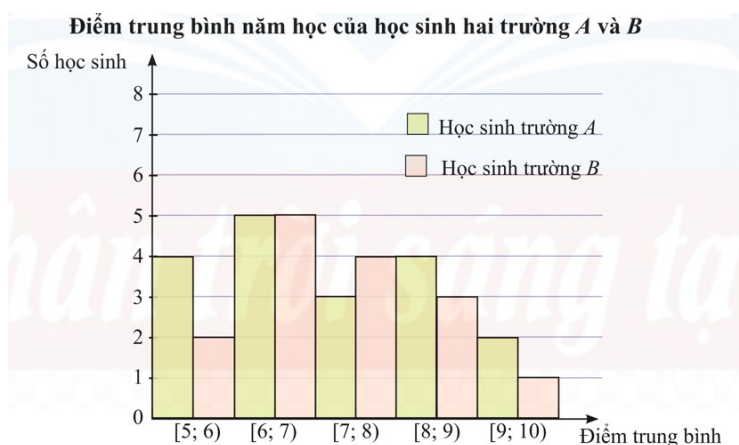
Câu 2. Bảng 1, Bảng 2 lần lượt biểu diễn mẫu số liệu ghép nhóm thống kê mức lương của hai công ty A, B (đơn vị: triệu đồng)

Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[10; 15)	12,5	15	[10; 15)	12,5	25
[15; 20)	17,5	18	[15; 20)	17,5	15
[20; 25)	22,5	10	[20; 25)	22,5	7
[25; 30)	27,5	10	[25; 30)	27,5	5
[30; 35)	32,5	5	[30; 35)	32,5	5
[35; 40)	37,5	2	[35; 40)	37,5	3
		$n = 60$			$n = 60$
Bảng 1			Bảng 2		

Các mệnh đề sau đúng hay sai?

- a) Số trung bình cộng của mẫu số liệu ghép nhóm của công ty A là: $\frac{62}{3}$
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm của công ty A là: 5
- c) Phương sai của mẫu số liệu ghép nhóm của công ty B (làm tròn kết quả đến hàng phần trăm) là: 52,91
- b) Nhận thấy độ lệch chuẩn của công ty A nhỏ hơn công ty B nên mức lương của công ty A đồng đều hơn

Câu 3. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B



Người ta lập được bảng tần số ghép nhóm cho mẫu số liệu trên như 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

Học sinh trường A	4	5	3	4	2
Học sinh trường B	2	5	4	3	1

Các mệnh đề sau đúng hay sai?

- Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm của học sinh trường A là: 6,1
- Khoảng tứ phân vị của mẫu số liệu ghép nhóm của học sinh trường B là: 1,73
- 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
- Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường A có điểm trung bình đồng đều hơn

Câu 4. Bảng 1, Bảng 2 lần lượt biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội và Huế (đơn vị: độ C).

Nhóm	Giá trị đại diện	Tần số
[16, 8; 19, 8)	18,3	2
[19, 8; 22, 8)	21,3	3
[22, 8; 25, 8)	24,3	2
[25, 8; 28, 8)	27,3	1
[28, 8; 31, 8)	30,3	4
		$n = 12$

Nhóm	Giá trị đại diện	Tần số
[16, 8; 19, 8)	18,3	1
[19, 8; 22, 8)	21,3	2
[22, 8; 25, 8)	24,3	3
[25, 8; 28, 8)	27,3	2
[28, 8; 31, 8)	30,3	4
		$n = 12$

Bảng 1

Bảng 2

Các mệnh đề sau đúng hay sai?

- Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên số liệu ở Hà Nội là: 8,75
- Độ lệch chuẩn của mẫu số liệu ghép nhóm trên số liệu ở Hà Nội (làm tròn kết quả đến hàng phần trăm) là: 3,56
- Phương sai của mẫu số liệu ghép nhóm trên số liệu ở Huế là: 12,75
- Huế có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn

Câu 5. Bảng 24 thống kê độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt và Vũng Tàu (đơn vị: %).

Tháng Độ ẩm	1	2	3	4	5	6	7	8	9	10	11	12
Đà Lạt	83	79	79	87	87	87	88	89	90	91	88	86
Vũng Tàu	75	77	78	77	79	79	81	79	81	83	80	77

Bảng 24

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

Người ta lần lượt ghép các số liệu của Đà Lạt, Vũng Tàu thành mẫu số liệu ghép nhóm như sau:

Đà Lạt			Vũng Tàu		
Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[75;78,3)	76,65	0	[75;78,3)	76,65	5
[78,3;81,6)	79,95	2	[78,3;81,6)	79,95	6
[81,6;84,9)	83,25	1	[81,6;84,9)	83,25	1
[84,9;88,2)	86,55	6	[84,9;88,2)	86,55	0
[88,2;91,5)	89,85	3	[88,2;91,5)	89,85	0

Các mệnh đề sau đúng hay sai?

- a) Xét số liệu ở Đà Lạt ta có khoảng biến thiên là: 16,5
- b) Xét số liệu ở Đà Lạt thì ta có độ lệch chuẩn của mẫu số liệu ghép nhóm (làm tròn kết quả đến hàng phần trăm) là: 3,28
- c) Xét số liệu ở Vũng Tàu thì phương sai của mẫu số liệu ghép nhóm là: 4,235
- d) Đà Lạt có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn

Câu 6. 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

(Nguồn: Tổng cục Thống kê)

Các mệnh đề sau đúng hay sai?

- a) Xét số liệu ở Nha Trang thì khoảng tứ phân vị của mẫu số liệu ghép nhóm là: 32,64
- b) Nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn
- c) Xét số liệu của Quy Nhơn ta có độ lệch chuẩn của mẫu số liệu ghép nhóm (làm tròn kết quả đến hàng phần trăm) là: 30,59
- d) Nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Nha Trang đồng đều hơn

Câu 7. Thống kê tổng số giờ nắng trong tháng 9 tại một trạm quan trắc đặt ở Cà Mau trong các năm từ 2002 đến 2021 được thống kê như sau:

111,6 134,9 130,3 134,2 140,9 109,3 154,4 156,3 116,1 96,7
 105,2 80,8 80,8 110 109 139 145 161 126 114

(Nguồn: Tổng cục Thống kê)

Người ta lập được bảng tần số ghép nhóm như sau:

Số giờ nắng	[80;98)	[98;116)	[116;134)	[134;152)	[152;170)
Giá trị đại diện	89	107	125	143	161
Số năm	3	6	3	5	3

Các mệnh đề sau đúng hay sai?

- Số trung bình của mẫu số liệu ghép nhóm là 124,1.
- Phương sai của mẫu số liệu ghép nhóm là 566,19.
- Độ lệch chuẩn của mẫu số liệu ghép nhóm (kết quả các phép tính làm tròn đến hàng phần nghìn) là 23,795.
- Sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc (kết quả các phép tính làm tròn đến hàng phần nghìn) là 4,805%.

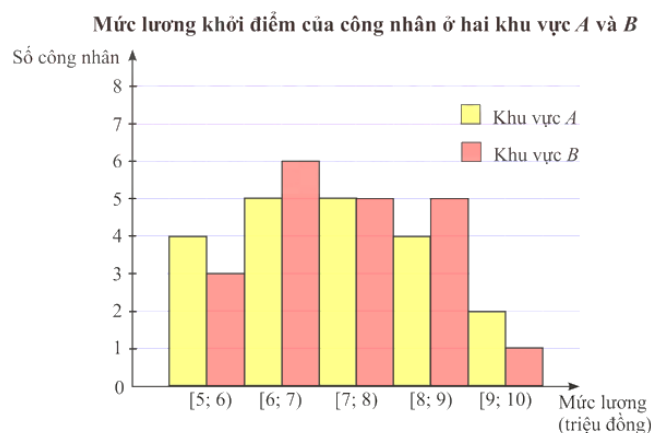
Câu 8. Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11A và 11B ở bảng sau:

Lớp \ Điểm trung bình					
	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
11A	1	0	11	22	6
11B	0	6	8	14	12

Các mệnh đề sau đúng hay sai?

- Khoảng biến thiên của điểm số học sinh lớp 11A là: 5.
- Nếu so sánh theo khoảng biến thiên thì điểm trung bình của các học sinh lớp 11B ít phân tán hơn điểm trung bình của các học sinh lớp 11A.
- Xét mẫu số liệu của lớp 11A ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{0,51}$.
- Nếu so sánh theo độ lệch chuẩn thì học sinh lớp 11A có điểm trung bình ít phân tán hơn học sinh lớp 11B.

Câu 9. Biểu đồ dưới đây mô tả kết quả điều tra về mức lương khởi điểm (đơn vị: triệu đồng) của một số công nhân ở hai khu vực A và B.



Người ta lập được bảng tần số ghép nhóm cho mẫu số liệu như sau:

Mức lương	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Mức lương đại diện (triệu đồng)	5,5	6,5	7,5	8,5	9,5
Khu vực A	4	5	5	4	2
Khu vực B	3	6	5	5	1

Các mệnh đề sau đúng hay sai?

- a) Xét mẫu số liệu của khu vực A ta có số trung bình của mẫu số liệu ghép nhóm là 6,25.
- b) Xét mẫu số liệu của khu vực A ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{1,5875}$
- c) Xét mẫu số liệu của khu vực B ta có phương sai của mẫu số liệu ghép nhóm là 1,3875.
- d) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì mức lương khởi điểm của công nhân khu vực B đồng đều hơn của công nhân khu vực A.

Câu 10. Giá đóng cửa của một cổ phiếu là giá của cổ phiếu đó cuối một phiên giao dịch. Bảng sau thống kê giá đóng cửa (đơn vị: nghìn đồng) của hai mã cổ phiếu A và B trong 50 ngày giao dịch liên tiếp.

Giá đóng cửa	[120; 122)	[122; 124)	[124; 126)	[126; 128)	[128; 130)
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Các mệnh đề sau đúng hay sai?

- a) Xét mẫu số liệu của cổ phiếu A ta có phương sai của mẫu số liệu ghép nhóm là 7,5216
- b) Xét mẫu số liệu của cổ phiếu B ta có số trung bình của mẫu số liệu ghép nhóm là 115,28.
- c) Xét mẫu số liệu của cổ phiếu B ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{15,4096}$
- d) Người ta có thể dùng phương sai và độ lệch chuẩn để so sánh mức độ rủi ro của các loại cổ phiếu có giá trị trung bình gần bằng nhau. Cổ phiếu nào có phương sai, độ lệch chuẩn cao hơn thì được coi là có độ rủi ro lớn hơn. Theo quan điểm trên, thì cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B.

F. TRẢ LỜI NGẮN

Câu 1. Hãy tính độ lệch chuẩn của mẫu số liệu ghép nhóm (kết quả được làm tròn đến hàng phần trăm)

Chiều cao (cm)	[160;164)	[164;168)	[168;172)	[172;176)	[176;180)
Số học sinh	3	5	8	4	1

Câu 2. 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 như 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 độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

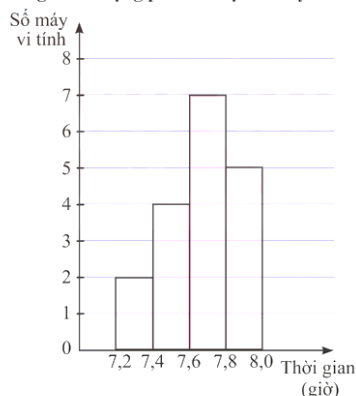
Câu 3. Bảng dưới đây thống kê cự li ném tạ của một vận động viên.

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 độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

Câu 4. 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



Hãy xác định độ lệch chuẩn của thời gian sử dụng pin (kết quả được làm tròn đến hàng phần trăm)

Câu 5. 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

Hãy xác định độ lệch chuẩn của thời gian sử dụng pin (kết quả được làm tròn đến hàng phần trăm)

Câu 6. Chiều dài của 40 bé trai sơ sinh 12 ngày tuổi chọn ngẫu nhiên ở một bệnh viện được nhà nghiên cứu thống kê trong Bảng dưới đây:

Chiều dài (cm)	[44;46)	[46;48)	[48;50)	[50;52)	[52;54)	[54;56)
Số trẻ	3	3	10	15	7	2

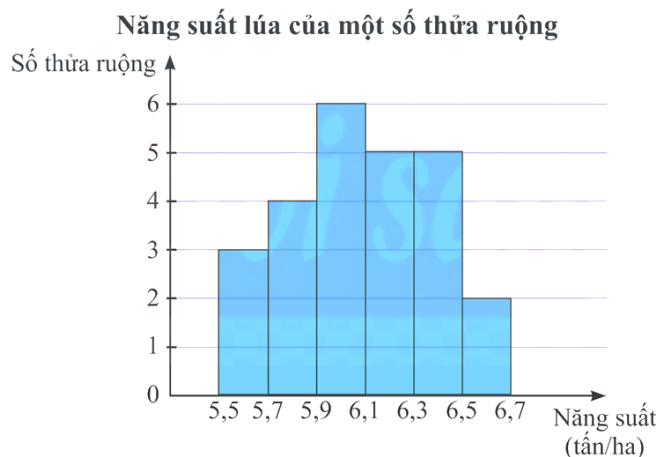
Tính độ lệch chuẩn của chiều dài nhóm 40 bé trai sơ sinh (làm tròn kết quả đến hàng phần nghìn).

Câu 7. 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

Hãy xác định độ lệch chuẩn của mẫu số liệu trên

Câu 8. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



Người ta lập được bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên như sau:

Năng suất	$[5,5;5,7)$	$[5,7;5,9)$	$[5,9;6,1)$	$[6,1;6,3)$	$[6,3;6,5)$	$[6,5;6,7)$
Số thửa ruộng	3	4	6	5	5	2
Giá trị đại diện	5,6	5,8	6,0	6,2	6,4	6,6
Tần số tương đối	3	4	6	5	5	2

Hãy xác định độ lệch chuẩn của mẫu số liệu trên (làm tròn kết quả đến hàng phần nghìn).

Câu 9. Một vận động viên luyện tập chạy cự li 100 m đã ghi lại kết quả luyện tập như sau:

Thời gian (giây)	$[10,2;10,4)$	$[10,4;10,6)$	$[10,6;10,8)$	$[10,8;11)$
Số vận động viên	3	7	8	2

Hãy xác định phương sai của mẫu số liệu trên

Câu 10. Cân nặng của một số quả mít trong một khu vườn được thống kê ở bảng sau:

Cân nặng (kg)	$[4;6)$	$[6;8)$	$[8;10)$	$[10;12)$	$[12;14)$
Số quả mít	6	12	19	9	4

Hãy tính phương sai của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Câu 11. Bảng 21 biểu diễn mẫu số liệu ghép nhóm về độ tuổi của cư dân trong một khu phố.

Nhóm	Giá trị đại diện	Tần số
[20 ; 30)	25	25
[30 ; 40)	35	20
[40 ; 50)	45	20
[50 ; 60)	55	15
[60 ; 70)	65	14
[70 ; 80)	75	6
		$n = 100$

Bảng 21

Hãy tính độ lệch của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

BÀI 2. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHÉP NHÓM

A. KIẾN THỨC CƠ BẢN CẦN NẮM

1. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Ở lớp 11, ta đã biết giá trị đại diện của nhóm $[a; b)$ là $c = \frac{1}{2}(a + b)$.

Tương tự như cách tính số trung bình, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm có thể được xác định thông qua giá trị đại diện và tần số của mỗi nhóm.

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	...	$[u_k; u_{k+1})$
Giá trị đại diện	c_1	c_2	...	c_k
Tần số	n_1	n_2	...	n_k

Phương sai của mẫu số liệu ghép nhóm, kí hiệu S^2 , được tính bởi công thức:

$$S^2 = \frac{1}{n} \left[n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right]$$

trong đó: $n = n_1 + n_2 + \dots + n_k$ là cỡ mẫu:

$$\bar{x} = \frac{1}{n} (n_1 c_1 + n_2 c_2 + \dots + n_k c_k) \text{ là } so \text{ trung bình.}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm, kí hiệu S , là căn bậc hai số học của phương sai.

Chú ý:

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

$$S^2 = \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + \dots + n_k c_k^2) - \bar{x}^2$$

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

$$S^2 = \frac{1}{n-1} \left[n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right].$$

Ví dụ 1: Cân nặng của một số quả mít trong một khu vườn được thống kê ở bảng sau:

Cân nặng (kg)	[4; 6)	[6; 8)	[8; 10)	[10; 12)	[12; 14)
Số quả mít	6	12	19	9	4

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Lời giải

Ta có bảng thống kê cân nặng của các quả mít theo giá trị đại diện:

Cân nặng đại diện (kg)	5	7	9	11	13
Tần số	6	12	19	9	4

Cỡ mẫu $n = 6 + 12 + 19 + 9 + 4 = 50$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x} = \frac{6.5 + 12.7 + 19.9 + 9.11 + 4.13}{50} = 8,72.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S^2 = \frac{1}{50} (6 \cdot 5^2 + 12 \cdot 7^2 + 19 \cdot 9^2 + 9 \cdot 11^2 + 4 \cdot 13^2) - 8,72^2 \approx 4,80$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S \approx \sqrt{4,80} \approx 2,19.$$

Ví dụ 2: Thống kê tổng số giờ nắng trong tháng 9 tại một trạm quan trắc đặt ở Cà Mau trong các năm từ 2002 đến 2021 được thống kê như sau:

111,6 134,9 130,3 134,2 140,9 109,3 154,4 156,3 116,1 96,7
105,2 80,8 80,8 110 109 139 145 161 126 114

- a) Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu trên.
b) Hãy lập bảng tần số ghép nhóm với nhóm đầu tiên là $[80; 98)$ và độ dài mỗi nhóm bằng 18. Tính phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm.
c) Hãy tính sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc.

(Kết quả các phép tính làm tròn đến hàng phần nghìn.)

Lời giải

a) Cỡ mẫu là $n = 20$.

Số trung bình của mẫu số liệu trên là

$$\bar{x}_1 = \frac{111,6 + 134,9 + \dots + 114}{20} = 122,755.$$

Phương sai của mẫu số liệu trên là

$$S_1^2 = \frac{1}{20} (111,6^2 + 134,9^2 + \dots + 114^2) - 122,755^2 \approx 515,453.$$

Độ lệch chuẩn của mẫu số liệu trên là

$$S_1 \approx \sqrt{515,453} \approx 22,704.$$

b) Ta có bảng sau:

Số giờ nắng	$[80; 98)$	$[98; 116)$	$[116; 134)$	$[134; 152)$	$[152; 170)$
Giá trị đại diện	89	107	125	143	161
Số năm	3	6	3	5	3

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{3.89 + 6.107 + 3.125 + 5.143 + 3.161}{20} = 124,1.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{20} (3.89^2 + 6.107^2 + 3.125^2 + 5.143^2 + 3.161^2) - 124,1^2 = 566,19.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_2 = \sqrt{566,19} \approx 23,795.$$

c) Sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc là

$$\frac{|S_2 - S_1|}{S_1} = \frac{|23,795 - 22,704|}{22,704} \cdot 100\% \approx 4,805\%.$$

Ý nghĩa của phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm

- Phương sai của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho phương sai của mẫu số liệu gốc. Độ lệch chuẩn của mẫu số liệu ghép nhóm cũng là giá trị xấp xỉ cho độ lệch chuẩn của mẫu số liệu gốc. Chúng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm. Phương sai và độ lệch chuẩn càng lớn thì dữ liệu càng phân tán.
- Độ lệch chuẩn có cùng đơn vị với đơn vị của mẫu số liệu.

Ví dụ 3: Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11 A và 11B ở bảng sau:

Điểm trung bình \ Lớp	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
11A	1	0	11	22	6
11B	0	6	8	14	12

- a) Nếu so sánh theo khoảng biến thiên thì học sinh lớp nào có điểm trung bình ít phân tán hơn?
b) Nếu so sánh theo độ lệch chuẩn thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

Lời giải

a) Khoảng biến thiên của điểm số học sinh lớp 11 A là: $10 - 5 = 5$.

Khoảng biến thiên của điểm số học sinh lớp 11B là: $10 - 6 = 4$.

Nếu so sánh theo khoảng biến thiên thì điểm trung bình của các học sinh lớp 11B ít phân tán hơn điểm trung bình của các học sinh lớp 11A.

b) Ta có bảng thống kê điểm trung bình theo giá trị đại diện:

Giá trị đại diện \ Lớp	5,5	6,5	7,5	8,5	9,5
11A	1	0	11	22	6
11B	0	6	8	14	12

Xét mẫu số liệu của lớp 11A:

Cỡ mẫu là $n_1 = 1 + 11 + 22 + 6 = 40$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{1 \cdot 5,5 + 11 \cdot 7,5 + 22 \cdot 8,5 + 6 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{40} (1 \cdot 5,5^2 + 11 \cdot 7,5^2 + 22 \cdot 8,5^2 + 6 \cdot 9,5^2) - 8,3^2 = 0,61.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_1 = \sqrt{0,61}$.

Xét mẫu số liệu của lớp 11B:

Cỡ mẫu là $n_2 = 6 + 8 + 14 + 12 = 40$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{6 \cdot 6,5 + 8 \cdot 7,5 + 14 \cdot 8,5 + 12 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{40} (6 \cdot 6,5^2 + 8 \cdot 7,5^2 + 14 \cdot 8,5^2 + 12 \cdot 9,5^2) - 8,3^2 = 1,06.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{1,06}$.

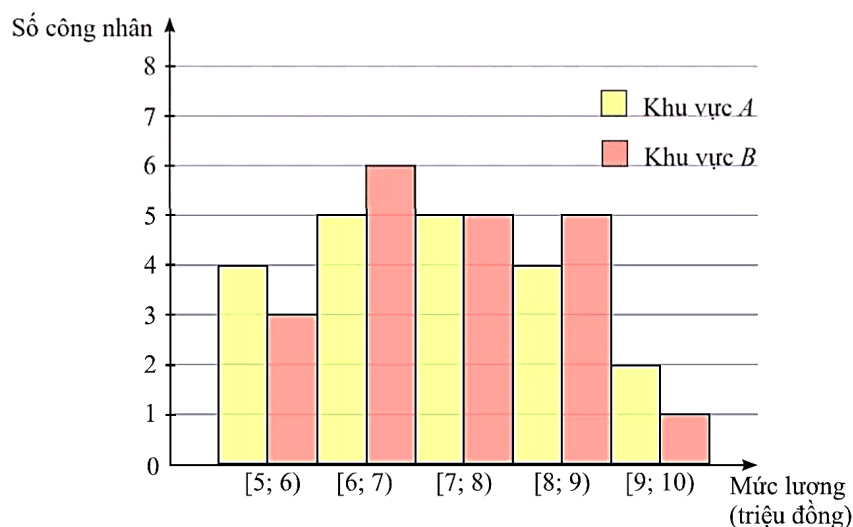
Do $S_1 < S_2$ nên nếu so sánh theo độ lệch chuẩn thì học sinh lớp 11 A có điểm trung bình ít phân tán hơn học sinh lớp 11B.

Ví dụ 4: Biểu đồ dưới đây mô tả kết quả điều tra về mức lương khởi điểm (đơn vị: triệu đồng) của một số công nhân ở hai khu vực A và B.

a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu đó.

b) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì công nhân ở khu vực nào có mức lương khởi điểm đồng đều hơn?

Mức lương khởi điểm của công nhân ở hai khu vực A và B



Lời giải

a) Ta có bảng sau:

Mức lương	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Mức lương đại diện (triệu đồng)	5,5	6,5	7,5	8,5	9,5
Khu vực A	4	5	5	4	2
Khu vực B	3	6	5	5	1

b)

Xét mẫu số liệu của khu vực A :

Cỡ mẫu là $n_A = 4 + 5 + 5 + 4 + 2 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_A = \frac{4 \cdot 5,5 + 5 \cdot 6,5 + 5 \cdot 7,5 + 4 \cdot 8,5 + 2 \cdot 9,5}{20} = 7,25.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_A^2 = \frac{1}{20} (4 \cdot 5,5^2 + 5 \cdot 6,5^2 + 5 \cdot 7,5^2 + 4 \cdot 8,5^2 + 2 \cdot 9,5^2) - (7,25)^2 = 1,5875.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_A = \sqrt{1,5875}.$$

Xét mẫu số liệu của khu vực B :

Cỡ mẫu là $n_B = 3 + 6 + 5 + 5 + 1 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_B = \frac{3 \cdot 5,5 + 6 \cdot 6,5 + 5 \cdot 7,5 + 5 \cdot 8,5 + 1 \cdot 9,5}{20} = 7,25.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_B^2 = \frac{1}{20} (3 \cdot 5,5^2 + 6 \cdot 6,5^2 + 5 \cdot 7,5^2 + 5 \cdot 8,5^2 + 1 \cdot 9,5^2) - (7,25)^2 = 1,2875.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_B = \sqrt{1,2875}$$

Do $S_A > S_B$ nên nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì mức lương khởi điểm của công nhân khu vực B đồng đều hơn của công nhân khu vực A.

Chú ý: Với các mẫu số liệu ghép nhóm có cùng số trung bình (hoặc xấp xỉ nhau), ta thường sử dụng phương sai và độ lệch chuẩn để so sánh mức độ phân tán của các mẫu số liệu đó.

Ví dụ 5: Giá đóng cửa của một cổ phiếu là giá của cổ phiếu đó cuối một phiên giao dịch. Bảng sau thống kê giá đóng cửa (đơn vị: nghìn đồng) của hai mã cổ phiếu A và B trong 50 ngày giao dịch liên tiếp.

Giá đóng cửa	[120; 122)	[122; 124)	[124; 126)	[126; 128)	[128; 130)
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Người ta có thể dùng phương sai và độ lệch chuẩn để so sánh mức độ rủi ro của các loại cổ phiếu có giá trị trung bình gần bằng nhau. Cổ phiếu nào có phương sai, độ lệch chuẩn cao hơn thì được coi là có độ rủi ro lớn hơn.

Theo quan điểm trên, hãy so sánh độ rủi ro của cổ phiếu A và cổ phiếu B.

Lời giải

Ta có bảng thống kê giá đóng cửa theo giá trị đại diện:

Giá đóng cửa	121	123	125	127	129
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Xét mẫu số liệu của cổ phiếu A :

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{8.121 + 9.123 + 12.125 + 10.127 + 11.129}{50} = 125,28.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{50} (8.121^2 + 9.123^2 + 12.125^2 + 10.127^2 + 11.129^2) - (125,28)^2 = 7,5216$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_1 = \sqrt{S_1^2} = \sqrt{7,5216}$.

Xét mẫu số liệu của cổ phiếu B :

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{16.121 + 4.123 + 3.125 + 6.127 + 21.129}{50} = 125,28.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{50} (16.121^2 + 4.123^2 + 3.125^2 + 6.127^2 + 21.129^2) - (125,28)^2 = 12,4096.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{S_2^2} = \sqrt{12,4096}$.

Vậy nếu đánh giá độ rủi ro theo phương sai và độ lệch chuẩn thì cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B.

Trả lời câu hỏi Thực hành 1 trang 82 SGK Toán 12 Chân trời sáng tạo

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm ở Khởi động

Chiều cao (cm)	[160;164)	[164;168)	[168;172)	[172;176)	[176;180)
Số học sinh	3	5	8	4	1

Lời giải

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

Cỡ mẫu: $n = 21$

Giá trị trung bình của mẫu số liệu mới:

$$\bar{x} = \frac{1}{n}(n_1c_1 + n_2c_2 + \dots + n_kc_k) = \frac{1}{21}(3 \cdot 162 + 5 \cdot 166 + 8 \cdot 170 + 4 \cdot 174 + 1 \cdot 178) = \frac{3550}{21}$$

Phương sai của mẫu số liệu mới:

$$\begin{aligned} S^2 &= \frac{1}{n} \left[n_1(c_1 - \bar{x})^2 + n_2(c_2 - \bar{x})^2 + \dots + n_k(c_k - \bar{x})^2 \right] \\ &= \frac{1}{21} \left[3 \left(162 - \frac{3550}{21} \right)^2 + 5 \left(166 - \frac{3550}{21} \right)^2 + \dots \right. \\ &\quad \left. + 1 \left(178 - \frac{3550}{21} \right)^2 \right] = \frac{8000}{441} \end{aligned}$$

$$\text{Độ lệch chuẩn của mẫu số liệu mới: } \sigma = \sqrt{S^2} = \sqrt{\frac{8000}{441}} = \frac{40\sqrt{5}}{21}$$

Trả lời câu hỏi Thực hành 2 trang 82 SGK Toán 12 Chân trời sáng tạo

Mai và Ngọc 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:

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

Giá trị đại diện	4	6	8	10	12
Mai	6	7	6	6	5
Ngọc	2	5	13	8	2

a) Cỡ mẫu: $n = 30$

Xét mẫu số liệu của Mai:

$$\text{Số trung bình: } \bar{x}_1 = \frac{6 \cdot 4 + 7 \cdot 6 + 6 \cdot 8 + 6 \cdot 10 + 5 \cdot 12}{30} = 7,8$$

$$\text{Phương sai: } S_1^2 = \frac{(6 \cdot 4^2 + 7 \cdot 6^2 + 6 \cdot 8^2 + 6 \cdot 10^2 + 5 \cdot 12^2)}{30} - 7,8^2 = 7,56$$

$$\text{Độ lệch chuẩn: } \sigma_1 = \sqrt{7,56} \approx 2,75$$

Xét mẫu số liệu của Ngọc:

$$\text{Số trung bình: } \bar{x}_2 = \frac{2 \cdot 4 + 5 \cdot 6 + 13 \cdot 8 + 8 \cdot 10 + 2 \cdot 12}{30} = 8,2$$

Phương sai: $S_2^2 = \frac{(2.4^2 + 5.6^2 + 13.8^2 + 8.10^2 + 2.12^2)}{30} - 8,2^2 \approx 3,83$

Độ lệch chuẩn: $\sigma_2 = \sqrt{3,83} \approx 1,96$

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

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 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

Giá trị đại diện	19,25	19,75	20,25	20,75	21,25
Tần số	13	45	24	12	6

Cỡ mẫu: $n = 100$

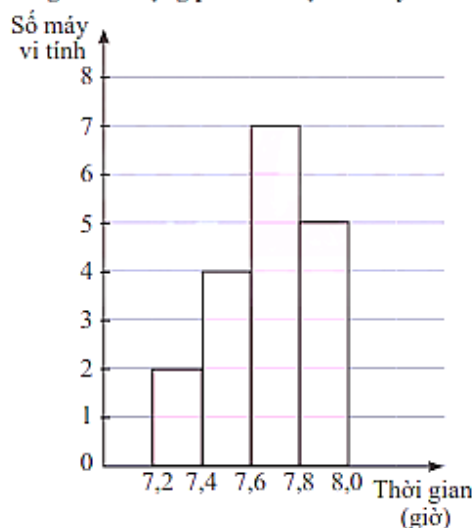
Số trung bình: $\bar{x} = \frac{13.19,25 + 45.19,75 + 24.20,25 + 12.20,75 + 6.21,25}{100} = 20,015$

Phương sai: $S^2 = \frac{13.19,25^2 + 45.19,75^2 + 24.20,25^2 + 12.20,75^2 + 6.21,25^2}{100} - 20,015^2 \approx 0,28$

Độ lệch chuẩn: $\sigma = \sqrt{0,28} \approx 0,53$

Bài 2:

Thời gian sử dụng pin của một số máy vi tính



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.

Lời giải

a) Có 2 máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ

b)

Giá trị đại diện	7,3	7,5	7,7	7,9
Số máy	2	4	7	5

Cỡ mẫu: $n = 18$

$$\text{Số trung bình: } \bar{x} = \frac{2 \cdot 7,3 + 4 \cdot 7,5 + 7 \cdot 7,7 + 5 \cdot 7,9}{18} \approx 7,67$$

$$\text{Phương sai: } S^2 = \frac{2 \cdot 7,3^2 + 4 \cdot 7,5^2 + 7 \cdot 7,7^2 + 5 \cdot 7,9^2}{18} - 7,67^2 \approx 0,04$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{0,04} \approx 0,19$$

Bài 3: 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) Khoảng biến thiên của mẫu số liệu: $61,1 - 42 = 19,1$ (km/h)

Cỡ mẫu: $n = 20$

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

$$\text{Trung vị } Q_2 = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(48,4 + 50,8) = 49,6$$

Tứ phân vị thứ nhất là trung vị của nửa số liệu bên trái

$$Q_2 Q_1 = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(46,7 + 46,8) = 46,75$$

Tứ phân vị thứ ba là trung vị của nửa số liệu bên phải

$$Q_2 : Q_3 = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(54,8 + 55,6) = 55,2$$

Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = 8,45$

$$\text{Số trung bình: } \bar{x} = \frac{42 + 43,4 + \dots + 61,1}{20} = 50,945$$

$$\text{Phương sai: } S^2 = \frac{42^2 + 43,4^2 + \dots + 61,1^2}{20} - 50,945^2 \approx 32,2$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{32,2} \approx 5,67$$

b)

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

c) Ta có: $x_1; \dots; x_3 \in [42; 46)$; $x_4; \dots; x_{10} \in [46; 50)$; $x_{11}; \dots; x_{14} \in [50; 54)$; $x_{15}; \dots; x_{17} \in [54; 58)$; $x_{18}; \dots; x_{20} \in [58; 62)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_5 + x_6) \in [46; 50)$. Do đó, tứ phân vị thứ nhất của

$$\text{mẫu số liệu ghép nhóm là: } Q_1' = 46 + \frac{\frac{20}{4} - 3}{7}(50 - 46) = \frac{330}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{15} + x_{16}) \in [54; 58)$. Do đó, tứ phân vị thứ ba của mẫu

$$\text{số liệu ghép nhóm là: } Q_3' = 54 + \frac{\frac{3 \cdot 20}{4} - (3 + 7 + 4)}{3}(58 - 54) = \frac{166}{3} \text{ Khoảng tứ phân vị của mẫu số}$$

$$\text{liệu ghép nhóm là: } \Delta_Q' = Q_3' - Q_1' = \frac{172}{21}$$

Giá trị đại diện	44	48	52	56	60
Số xe	3	7	4	3	3

$$\text{Số trung bình: } \bar{x} = \frac{3 \cdot 44 + 7 \cdot 48 + 4 \cdot 52 + 3 \cdot 56 + 3 \cdot 60}{20} = 41,8$$

$$\text{Phương sai: } S^2 = \frac{3 \cdot 44^2 + 7 \cdot 48^2 + 4 \cdot 52^2 + 3 \cdot 56^2 + 3 \cdot 60^2}{20} - 41,8^2 = 364,96$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{364,96} = 19,1$$

Bài 4: 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

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

Xét mẫu số liệu của địa điểm A:

$$\text{Số trung bình: } \bar{x}_A = \frac{25.31 + 38.33 + 20.35 + 10.37 + 7.39}{100} = 33,72$$

Xét mẫu số liệu của địa điểm B:

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

Đường kính trung bình của thân cây xoan đào trồng tại địa điểm A nhỏ hơn tại địa điểm B.

b) - Xét mẫu số liệu của địa điểm A:

$$\sigma_A = \sqrt{\frac{25.31^2 + 38.33^2 + 20.35^2 + 10.37^2 + 7.39^2}{100} - 33,72^2} = 2,32$$

Xét mẫu số liệu của địa điểm B:

$$\sigma_B = \sqrt{\frac{22.31^2 + 27.33^2 + 19.35^2 + 18.37^2 + 14.39^2}{100}} = 2,7$$

Vậy cây trồng tại địa điểm A có đường kính đồng đều hơn

C. CÁC DẠNG TOÁN

Câu 1. 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)
Học sinh trường X	8	10	13	10	9
Học sinh trường Y	4	12	17	14	3

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

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

c) Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

Lời giải

Giá trị đại diện	6,5	7,5	8,5	9,5	10,5
Học sinh trường X	8	10	13	10	9
Học sinh trường Y	4	12	17	14	3

a) Cỡ mẫu: $n = 50$

Xét số liệu của trường X :

$$\text{Số trung bình: } \bar{x}_X = \frac{8.6,5 + 10.7,5 + 13.8,5 + 10.9,5 + 9.10,5}{50} = 8,54$$

Xét số liệu của trường Y :

Số trung bình: $\bar{x}_Y = \frac{4.6,5 + 12.7,5 + 17.8,5 + 14.9,5 + 3.10,5}{50} = 8,5$

Vậy nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn

b) Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của 50 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 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 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 là: $\Delta_Q = Q_3 - Q_1 = 2,2$

Gọi $y_1; y_2; \dots; y_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của 50 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 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 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 là: $\Delta_{Q'} = Q'_3 - Q'_1 = \frac{271}{168}$

Vậy nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn

c) Xét số liệu của trường X :

Độ lệch chuẩn: $\sigma_X = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,5^2} \approx 1,33$

Xét số liệu của trường Y :

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{4.6,5^2 + 12.7,5^2 + 17.8,5^2 + 14.9,5^2 + 3.10,5^2}{50} - 8,5^2} \approx 1,04$$

Vậy nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn

Câu 2. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi hai phân xưởng được cho như sau:

Tuổi thọ (năm)	[1,5;2)	[2,2,5)	[2,5;3)	[3;3,5)	[3,5;4)
Số linh kiện của phân xưởng 1	4	9	13	8	6
Số linh kiện của phân xưởng 2	2	8	20	7	3

Tính phương sai và độ lệch chuẩn của mỗi mẫu số liệu ghép nhóm và nhận xét về độ phân tán của tuổi thọ các linh kiện điện tử được sản xuất bởi mỗi phân xưởng.

Lời giải

Phân xưởng 1: Tổng số linh kiện: $4 + 9 + 13 + 8 + 6 = 40$

$$\text{Giá trị trung bình } \bar{x}_1 = \frac{1,75.4 + 2,25.9 + 2,75.13 + 3,25.8 + 3,75.6}{4 + 9 + 13 + 8 + 6} = \frac{223}{80}$$

$$\text{Phương sai: } s_1^2 = \frac{1}{40} (1,75^2.4 + 2,25^2.9 + 2,75^2.13 + 3,25^2.8 + 3,75^2.6) - \left(\frac{223}{80} \right)^2 = \frac{2271}{6400}$$

$$\text{Độ lệch chuẩn: } s_1 = \sqrt{\frac{2271}{6400}} \approx 0,6$$

Phân xưởng 2: Tổng số linh kiện: $2 + 8 + 20 + 7 + 3 = 40$

$$\text{Giá trị trung bình } \bar{x}_2 = \frac{1,75.2 + 2,25.8 + 2,75.20 + 3,25.7 + 3,75.3}{2 + 8 + 20 + 7 + 3} = \frac{221}{80}$$

$$\text{Phương sai: } s_2^2 = \frac{1}{40} (1,75^2.2 + 2,25^2.8 + 2,75^2.20 + 3,25^2.7 + 3,75^2.3) - \left(\frac{221}{80} \right)^2 = \frac{1399}{6400}$$

$$\text{Độ lệch chuẩn: } s_2 = \sqrt{\frac{1399}{6400}} \approx 0,47$$

Vì $0,6 > 0,47$ nên độ phân tán của phân xưởng 1 lớn hơn độ phân tán của phân xưởng 2.

Câu 3. Một nhóm 20 học sinh dùng một thiết bị đo đường kính của một nhân tế bào cho kết quả như sau:

Kết quả đo (μm)	[4,5;5)	[5;5,5)	[5,5;6)	[6;6,5)
Số học sinh	3	8	7	2

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

b) Số trung bình và độ lệch chuẩn cho biết thông tin gì?

Lời giải

a) Mẫu số liệu ghép nhóm với giá trị đại diện:

Kết quả đo (μm)	[4,5;5)	[5;5,5)	[5,5;6)	[6;6,5)
Giá trị đại diện	4,75	5,25	5,75	6,25
Số học sinh	3	8	7	2

$$\text{Số trung bình: } \bar{x} = \frac{4,75 \cdot 3 + 5,25 \cdot 8 + 5,75 \cdot 7 + 6,25 \cdot 2}{20} = \frac{109}{20} (\mu m)$$

$$\text{Phương sai: } s^2 = \frac{1}{20} (4,75^2 \cdot 3 + 5,25^2 \cdot 8 + 5,75^2 \cdot 7 + 6,25^2 \cdot 2) - \left(\frac{109}{20} \right)^2 = \frac{37}{200}$$

$$\text{Độ lệch chuẩn: } s = \sqrt{\frac{37}{200}} \approx 0,43$$

b) Số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm xấp xỉ với số trung bình và độ lệch chuẩn của mẫu số liệu gốc. Do đó, mẫu số liệu gốc có số trung bình xấp xỉ $\frac{109}{20} \mu m$ và độ lệch chuẩn xấp xỉ $0,43 \mu m$.

Câu 4. Thời gian chạy tập luyện cự li 100 m của hai vận động viên được cho trong bảng sau:

Thời gian (giây)	[10;10,3)	[10,3;10,6)	[10,6;10,9)	[10,9;11,2)
Số lần chạy của A	2	10	5	3
Số lần chạy của B	3	7	9	6

Dựa trên độ lệch chuẩn của các mẫu số liệu ghép nhóm, hãy cho biết vận động viên nào có thành tích luyện tập ổn định hơn.

Lời giải

Ta có bảng số liệu với giá trị đại diện:

Thời gian (giây)	[10;10,3)	[10,3;10,6)	[10,6;10,9)	[10,9;11,2)
Giá trị đại diện	10,15	10,45	10,75	11,05
Số lần chạy của A	2	10	5	3
Số lần chạy của B	3	7	9	6

Vận động viên A:

$$\text{Giá trị trung bình } \bar{x}_A = \frac{10,15 \cdot 2 + 10,45 \cdot 10 + 10,75 \cdot 5 + 11,05 \cdot 3}{2 + 10 + 5 + 3} = \frac{2117}{200}$$

$$\text{Phương sai: } s_A^2 = \frac{1}{20} (10,15^2 \cdot 2 + 10,45^2 \cdot 10 + 10,75^2 \cdot 5 + 11,05^2 \cdot 3) - \left(\frac{2117}{200} \right)^2 = \frac{2691}{40000}$$

$$\text{Độ lệch chuẩn: } s_A = \sqrt{\frac{2691}{40000}} \approx 0,26$$

Vận động viên B:

$$\text{Giá trị trung bình } \bar{x}_B = \frac{10,15.3 + 10,45.7 + 10,75.9 + 11,05.6}{3 + 7 + 9 + 6} = \frac{5333}{500}$$

$$\text{Phương sai: } s_B^2 = \frac{1}{25} (10,15^2.3 + 10,45^2.7 + 10,75^2.9 + 11,05^2.6) - \left(\frac{5333}{500} \right)^2 = \frac{1296}{15625}$$

$$\text{Độ lệch chuẩn: } s_B = \sqrt{\frac{1296}{15625}} = 0,288$$

Vì $s_A < s_B$ nên vận động viên A có thành tích luyện tập ổn định hơn.

Câu 5. Có nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán của hai mẫu số liệu ghép nhóm trong mỗi trường hợp sau không? Tại sao?

a) Các mẫu số liệu ghép nhóm về điểm thi tốt nghiệp môn Toán của học sinh hai trường trung học phổ thông có chất lượng tương đương.

b) Các mẫu số liệu ghép nhóm về doanh thu của 100 cửa hàng bán lẻ và doanh thu của 100 siêu thị.

Lời giải

a) Có nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán của học sinh hai trường vì 2 trường này có chất lượng tương đương nhau

b) Không nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán về doanh thu của 100 cửa hàng bán lẻ và 100 siêu thị vì doanh thu của 100 cửa hàng bán lẻ và 100 siêu thị là không gần tương đương nhau.

Câu 6. Mai và Ngọc 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)
Mai	6	7	6	6	5
Ngọc	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

Giá trị đại diện	4	6	8	10	12
Mai	6	7	6	5	
Ngọc	2	5	13	8	2

a) Cỡ mẫu: $n = 30$

- Xét mẫu số liệu của Mai:

$$\text{Số trung bình: } \bar{x}_1 = \frac{6.4 + 7.6 + 6.8 + 6.10 + 5.12}{30} = 7,8$$

$$\text{Phương sai: } S_1^2 = \frac{(6.4^2 + 7.6^2 + 6.8^2 + 6.10^2 + 5.12^2)}{30} - 7,8^2 = 7,56$$

$$\text{Độ lệch chuẩn: } \sigma_1 = \sqrt{7,56} \approx 2,75$$

- Xét mẫu số liệu của Ngọc:

$$\text{Số trung bình: } \bar{x}_2 = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$$

$$\text{Phương sai: } S_2^2 = \frac{(2.4^2 + 5.6^2 + 13.8^2 + 8.10^2 + 2.12^2)}{30} - 8,2^2 \approx 3,83$$

$$\text{Độ lệch chuẩn: } \sigma_2 = \sqrt{3,83} \approx 1,96$$

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

Kết quả các phép tính làm tròn đến hàng phần nghìn.

Câu 7. 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.

Lời giải

Ta có mẫu số liệu ghép nhóm với giá trị đại diện là:

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

Tiền lãi trung bình khi đầu tư vào lĩnh vực A là:

$$\bar{x}_A = \frac{7,5.2 + 12,5.5 + 17,5.8 + 22,5.6 + 27,5.4}{2 + 5 + 8 + 6 + 4} = 18,5 \text{ (triệu đồng)}$$

Tiền lãi trung bình khi đầu tư vào lĩnh vực B là:

$$\bar{x}_B = \frac{7,5.8 + 12,5.4 + 17,5.2 + 22,5.5 + 27,5.6}{8 + 4 + 2 + 5 + 6} = 16,9 \text{ (triệu đồng)}$$

Do đó, về trung bình, đầu tư vào lĩnh vực A đem lại tiền lãi cao hơn.

b) Phương sai của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực A:

$$s_A^2 = \frac{1}{25} (7,5^2 \cdot 2 + 12,5^2 \cdot 5 + 17,5^2 \cdot 8 + 22,5^2 \cdot 6 + 27,5^2 \cdot 4) - 18,5^2 = 34$$

Độ lệch chuẩn của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực A: $s_A = \sqrt{34}$ (triệu đồng)

Phương sai của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực B:

$$s_B^2 = \frac{1}{25} (7,5^2 \cdot 8 + 12,5^2 \cdot 4 + 17,5^2 \cdot 2 + 22,5^2 \cdot 5 + 27,5^2 \cdot 6) - 16,9^2 = 64,64$$

Độ lệch chuẩn của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực B: $s_B = \sqrt{64,64}$ (triệu đồng)

Như vậy, độ lệch chuẩn của mẫu số liệu về tiền lãi đầu tư vào lĩnh vực B lớn hơn độ lệch chuẩn của mẫu số liệu về tiền lãi đầu tư vào lĩnh vực A nên đầu tư vào lĩnh vực B là rủi ro hơn.

Câu 8. 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) Khoảng biến thiên của mẫu số liệu: $61,1 - 42 = 19,1 (km / h)$

Cỡ mẫu: $n = 20$

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

$$\text{Trung vị } Q_2 = \frac{1}{2} (x_{10} + x_{11}) = \frac{1}{2} (48,4 + 50,8) = 49,6$$

Tứ phân vị thứ nhất là trung vị của nửa số liệu bên trái Q_1 :

$$Q_1 = \frac{1}{2} (x_5 + x_6) = \frac{1}{2} (46,7 + 46,8) = 46,75$$

Tứ phân vị thứ ba là trung vị của nửa số liệu bên phải Q_3 :

$$Q_3 = \frac{1}{2} (x_{15} + x_{16}) = \frac{1}{2} (54,8 + 55,6) = 55,2$$

Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = 8,45$

$$\text{Số trung bình: } \bar{x} = \frac{42 + 43,4 + \dots + 61,1}{20} = 50,945$$

$$\text{Phương sai: } S^2 = \frac{42^2 + 43,4^2 + \dots + 61,1^2}{20} - 50,945^2 \approx 32,2$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{32,2} \approx 5,67$$

b)

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

c) Ta có: $x_1; \dots; x_3 \in [42; 46); x_4; \dots; x_{10} \in [46; 50); x_{11}; \dots; x_{14} \in [50; 54); x_{15}; \dots; x_{17} \in [54; 58); x_{18}; \dots; x_{20} \in [58; 62)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_5 + x_6) \in [46; 50)$. Do đó, tứ phân vị thứ

$$\text{nhat của mẫu số liệu ghép nhóm là: } Q'_1 = 46 + \frac{\frac{20}{4} - 3}{7}(50 - 46) = \frac{330}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(x_{15} + x_{16}) \in [54; 58)$. Do đó, tứ phân vị thứ ba

$$\text{của mẫu số liệu ghép nhóm là: } Q'_3 = 54 + \frac{\frac{3 \cdot 20}{4} - (3 + 7 + 4)}{3}(58 - 54) = \frac{166}{3}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta'_Q = Q'_3 - Q'_1 = \frac{172}{21}$$

Giá trị đại diện	44	48	52	56	60
Số xe	3		4	3	3

$$\text{Số trung bình: } \bar{x} = \frac{3 \cdot 44 + 7 \cdot 48 + 4 \cdot 52 + 3 \cdot 56 + 3 \cdot 60}{20} = 41,8$$

$$\text{Phương sai: } S^2 = \frac{3 \cdot 44^2 + 7 \cdot 48^2 + 4 \cdot 52^2 + 3 \cdot 56^2 + 3 \cdot 60^2}{20} - 41,8^2 = 364,96$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{364,96} = 19,1$$

Câu 9. Trong bài 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 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.

Lời giải

Mẫu số liệu với giá trị đại diện

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

Kết quả đo của An:

$$\text{Giá trị trung bình: } \bar{x}_1 = \frac{1}{10}(3,875 \cdot 1 + 3,925 \cdot 6 + 3,975 \cdot 2 + 4,025 \cdot 1) = 3,94$$

Phương sai của mẫu số liệu:

$$s_1^2 = \frac{1}{10}(3,875^2 \cdot 1 + 3,925^2 \cdot 6 + 3,975^2 \cdot 2 + 4,025^2 \cdot 1) - 3,94^2 = 0,001525$$

$$\text{Độ lệch chuẩn của mẫu số liệu: } s_1 = \sqrt{0,001525} \approx 0,039$$

Kết quả đo của Bình:

$$\text{Giá trị trung bình: } \bar{x}_1 = \frac{1}{10}(3,875 \cdot 1 + 3,925 \cdot 3 + 3,975 \cdot 4 + 4,025 \cdot 2) = 3,96$$

Phương sai của mẫu số liệu:

$$s_2^2 = \frac{1}{10}(3,875^2 \cdot 1 + 3,925^2 \cdot 3 + 3,975^2 \cdot 4 + 4,025^2 \cdot 2) - 3,96^2 = 0,002025$$

$$\text{Độ lệch chuẩn của mẫu số liệu: } s_2 = \sqrt{0,002025} \approx 0,045$$

Vì $s_2 > s_1$ nên Vôn kế của bạn An cho kết quả ổn định hơn.

Câu 10. Điều tra thời gian phải làm thêm trung bình hằng tuần của các bác sĩ ở một bệnh viện, người ta thu được số liệu sau:

5	6	7	8	8	8	9	9	9
9	9	10	10	10	10	11	11	11
12	12	12	13	13	14	14		

a) Chuyển mẫu số liệu đã cho về mẫu số liệu ghép nhóm với độ dài các nhóm ghép bằng 2 và nhóm đầu tiên là $[5; 7)$.

b) Tính số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm lập ở câu a (làm tròn kết quả đến hàng phần trăm).

c) So sánh và nêu ý nghĩa các kết quả tìm được ở câu b với các kết quả tương ứng của mẫu số liệu gốc.

Lời giải

a) Căn cứ vào giá trị nhỏ nhất và giá trị lớn nhất của mẫu số liệu, ta thấy khoảng biến thiên là $R = 9$ nên nếu muốn mỗi nhóm có độ dài bằng 2 và nhóm đầu tiên là $[5; 7)$ thì phải chia số liệu thành 5 nhóm, với 4 nhóm ghép tiếp theo là $[7; 9), [9; 11), [11; 13), [13; 15)$. Đếm số lần xuất hiện các giá trị trong mỗi nhóm, ta lập được bảng sau:

Nhóm	$[5; 7)$	$[7; 9)$	$[9; 11)$	$[11; 13)$	$[13; 15)$
Tần số	2	4	9	6	4

b) Bổ sung thêm giá trị đại diện của mỗi nhóm vào bảng lập ở trên, ta có

Nhóm	$[5; 7)$	$[7; 9)$	$[9; 11)$	$[11; 13)$	$[13; 15)$	
Giá trị đại diện	6	8	10	12	14	
Tần số	2	4	9	6	4	$N = 25$

Sử dụng Bảng 3.15b để tính số trung bình $\bar{x}_N$ của mẫu số liệu ghép nhóm:

$$\bar{x}_N = \frac{2.6 + 4.8 + 9.10 + 6.12 + 4.14}{25} = \frac{262}{25} = 10,48 \text{ (giờ)}$$

Suy ra phương sai s_N^2 của mẫu số liệu ghép nhóm là:

$$\begin{aligned} S_N^2 &= \frac{2(6-10,48)^2 + 4(8-10,48)^2 + 9(10-10,48)^2 + 6(12-10,48)^2 + 4(14-10,48)^2}{25} \\ &= \frac{130,24}{25} = 5,2096 \approx 5,21. \end{aligned}$$

Độ lệch chuẩn s_N của mẫu số liệu ghép nhóm là $s_N = \sqrt{s_N^2} \approx 2,28$.

c) Gọi $\bar{x}_G, s_G^2, s_G$ tương ứng là trung bình, phương sai, độ lệch chuẩn của mẫu số liệu gốc. Để tính giá trị của các số đặc trưng này, ta lập bảng tần số:

Thời gian (giờ)	5	6	7	8	9	10	11	12	13	14	
Tần số	1	1	1	3	5	4	3	3	2	2	$N = 25$

Sử dụng các công thức tính trung bình, phương sai, độ lệch chuẩn của mẫu số liệu không ghép nhóm, ta có:

$$\bar{x}_G = \frac{5+6+7+3.8+5.9+4.10+3.11+3.12+2.13+2.14}{25} = 10.$$

Phương sai:

$$\begin{aligned} S_G^2 &= \frac{1(5-10)^2 + 1(6-10)^2 + 1(7-10)^2 + \dots + 3(12-10)^2 + 2(13-10)^2 + 2(14-10)^2}{25} \\ &= \frac{132}{25} = 5,28. \end{aligned}$$

Độ lệch chuẩn: $s_G = \sqrt{s_G^2} = \sqrt{5,28} \approx 2,30$.

Quan sát kết quả tìm được ở hai câu b và c, ta thấy $\bar{x}_N \approx \bar{x}_G; s_N^2 \approx s_G^2; s_N \approx s_G$. Như vậy số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm tương ứng xấp xỉ với số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ban đầu, khi chưa ghép nhóm. Nếu chỉ có mẫu số liệu ghép nhóm thì ta có thể kết luận là thời gian làm thêm trung bình hằng tuần của 25 bác sĩ được điều tra là 10,48 giờ và so với số này thì thời gian làm thêm của các bác sĩ chênh lệch trung bình khoảng 2,28 giờ.

Câu 11. 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

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

- Xét mẫu số liệu của địa điểm A:

$$\text{Số trung bình: } \bar{x}_A = \frac{25 \cdot 31 + 38 \cdot 33 + 20 \cdot 35 + 10 \cdot 37 + 7 \cdot 39}{100} = 33,72$$

- Xét mẫu số liệu của địa điểm B:

$$\bar{x}_B = \frac{22 \cdot 31 + 27 \cdot 33 + 19 \cdot 35 + 18 \cdot 37 + 14 \cdot 39}{100} = 34,5$$

Đường kính trung bình của thân cây xoan đào trồng tại địa điểm A nhỏ hơn tại địa điểm B

b) - Xét mẫu số liệu của địa điểm A:

$$\sigma_A = \sqrt{\frac{25 \cdot 31^2 + 38 \cdot 33^2 + 20 \cdot 35^2 + 10 \cdot 37^2 + 7 \cdot 39^2}{100} - 33,72^2} = 2,32$$

- Xét mẫu số liệu của địa điểm B:

$$\sigma_B = \sqrt{\frac{22 \cdot 31^2 + 27 \cdot 33^2 + 19 \cdot 35^2 + 18 \cdot 37^2 + 14 \cdot 39^2}{100}} = 2,7$$

Vậy cây trồng tại địa điểm A có đường kính đồng đều hơn

D. TRẮC NGHIỆM 4 PHƯƠNG ÁN

Câu 1. 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.

Lời giải

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.

Câu 2. 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

Chọn A

Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì số đặc trưng không đổi là khoảng biến thiên.

Câu 9. 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

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) .

Lời giải

Chọn B

Bảng tần số ghép nhóm theo giá trị đại diện là

Doanh thu	[5; 7)	[7; 9)	[9; 11)	[11; 13)	[13; 15)
Giá trị đại diện	6	8	10	12	14
Số ngày	2	7	7	3	1

Số trung bình: $\bar{x} = \frac{2.6 + 7.8 + 7.10 + 3.12 + 1.14}{20} = 9,4$.

Câu 4. Một siêu thị thống kê số tiền (đơn vị: chục nghìn đồng) mà 44 khách hàng mua hàng ở siêu thị đó trong một ngày. Số liệu được ghi lại trong Bảng 18.

Nhóm	Giá trị đại diện	Tần số
[40 ; 45)	42,5	4
[45 ; 50)	47,5	14
[50 ; 55)	52,5	8
[55 ; 60)	57,5	10
[60 ; 75)	62,5	6
[65 ; 70)	67,5	2
		$n = 44$

Bảng 18

Phương sai của mẫu số liệu ghép nhóm trên là:

A. 53,2.

B. 46,1.

C. 30.

D. 11.

Lời giải

Chọn D

Số trung bình cộng của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{4.42,5 + 14.47,5 + 8.52,5 + 10.57,5 + 6.62,5 + 2.67,5}{44} = \frac{585}{11}$$

Phương sai của mẫu số liệu ghép nhóm là:

$$\begin{aligned}
 s^2 &= \frac{4\left(42,5 - \frac{585}{11}\right)^2 + 14\left(47,5 - \frac{585}{11}\right)^2 + 8\left(52,5 - \frac{585}{11}\right)^2 + 10\left(57,5 - \frac{585}{11}\right)^2}{44} \\
 &+ \frac{6\left(62,5 - \frac{585}{11}\right)^2 + 2\left(67,5 - \frac{585}{11}\right)^2}{44} \\
 &\approx 46,12
 \end{aligned}$$

Câu 5. Một siêu thị thống kê số tiền (đơn vị: chục nghìn đồng) mà 44 khách hàng mua hàng ở siêu thị đó trong một ngày. Số liệu được ghi lại trong Bảng 18.

Nhóm	Giá trị đại diện	Tần số
[40 ; 45)	42,5	4
[45 ; 50)	47,5	14
[50 ; 55)	52,5	8
[55 ; 60)	57,5	10
[60 ; 75)	62,5	6
[65 ; 70)	67,5	2
		$n = 44$

Bảng 18

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên (làm tròn kết quả đến hàng đơn vị) là:

A. 6,8.

B. 7,3.

C. 3,3.

D. 46,1.

Lời giải

Chọn A

Số trung bình cộng của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{4.42,5 + 14.47,5 + 8.52,5 + 10.57,5 + 6.62,5 + 2.67,5}{44} = \frac{585}{11}$$

Phương sai của mẫu số liệu ghép nhóm là:

$$s^2 = \frac{4\left(42,5 - \frac{585}{11}\right)^2 + 14\left(47,5 - \frac{585}{11}\right)^2 + 8\left(52,5 - \frac{585}{11}\right)^2 + 10\left(57,5 - \frac{585}{11}\right)^2 + 6\left(62,5 - \frac{585}{11}\right)^2 + 2\left(67,5 - \frac{585}{11}\right)^2}{44} \approx 46,12$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s = \sqrt{s^2} = \sqrt{46,12} \approx 6,8$

Câu 6. 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

Phương sai của mẫu số liệu ghép nhóm là

A. 3,39.

B. 11,62.

C. 0,1314.

D. 0,36.

Lời giải

Chọn C

Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
------------------	------	------	------	------	------

Số ngày	3	6	5	4	2
---------	---	---	---	---	---

Số trung bình:

$$\bar{x} = \frac{3.2,85 + 6.3,15 + 5.3,45 + 4.3,75 + 2.4,05}{20} = 3,39$$

Phương sai:

$$S^2 = \frac{3.2,85^2 + 6.3,15^2 + 5.3,45^2 + 4.3,75^2 + 2.4,05^2}{20} - 3,39^2 = 0,1314$$

Câu 7. 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 (<i>km</i>)	[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

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

Chọn D

Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
Số ngày	3	6	5	4	2

Số trung bình:

$$\bar{x} = \frac{3.2,85 + 6.3,15 + 5.3,45 + 4.3,75 + 2.4,05}{20} = 3,39$$

Phương sai:

$$S^2 = \frac{3.2,85^2 + 6.3,15^2 + 5.3,45^2 + 4.3,75^2 + 2.4,05^2}{20} - 3,39^2 = 0,1314$$

Độ lệch chuẩn:

$$\sigma = \sqrt{0,1314} \approx 0,36$$

Câu 8. Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi 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

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

Chọn D

$$\text{Số trung bình: } \bar{x} = \frac{6.22,5 + 6.27,5 + 4.32,5 + 37,5 + 42,5}{18} \approx 28,33$$

$$\text{Phương sai: } S^2 = \frac{6.22,5^2 + 6.27,5^2 + 4.32,5^2 + 37,5^2 + 42,5^2}{18} - 28,33^2 = 31,25$$

Câu 9. 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 (giây)	[8;10)	[10;12)	[12;14)	[14;16)	[16;18)
Số lần	4	6	8	4	3

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 giải

Chọn A

Giá trị đại diện	9	11	13	15	17
Số lần	4	6	8	4	3

$$\text{Số trung bình: } \bar{x} = \frac{4.9 + 6.11 + 8.13 + 4.15 + 3.17}{25} = 12,68$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{\frac{4.9^2 + 6.11^2 + 8.13^2 + 4.15^2 + 3.17^2}{25} - 12,68^2} \approx 5,98$$

E. TRẮC NGHIỆM ĐÚNG SAI

Câu 1. Ta có các bảng thống kê sau:

Nhóm	Giá trị đại diện	Tần số
[6,22 ; 6,46)	6,34	3
[6,46 ; 6,70)	6,58	7
[6,70 ; 6,94)	6,82	5
[6,94 ; 7,18)	7,06	20
[7,18 ; 7,42)	7,30	5
		$n = 40$

Bảng 15

Nhóm	Giá trị đại diện	Tần số
[6,22 ; 6,46)	6,34	2
[6,46 ; 6,70)	6,58	5
[6,70 ; 6,94)	6,82	8
[6,94 ; 7,18)	7,06	19
[7,18 ; 7,42)	7,30	6
		$n = 40$

Bảng 16

Các mệnh đề sau đúng hay sai?

- a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 (làm tròn kết quả đến hàng phần trăm) là: $\approx 6,92(m)$.
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 (làm tròn kết quả đến hàng phần trăm) là: $0,26(m)$.
- c) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho bởi Bảng 16 (làm tròn kết quả đến hàng phần trăm) là: $0,16$.
- d) Kết quả nhảy xa của vận động viên Dũng đồng đều hơn kết quả nhảy xa của vận động viên Huy.

Lời giải

a) Đúng	b) Đúng	c) Sai	d) Sai
----------------	----------------	---------------	---------------

a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 là:

$$\bar{x}_D = \frac{3 \cdot 6,34 + 7 \cdot 6,58 + 5 \cdot 6,82 + 20 \cdot 7,06 + 5 \cdot 7,30}{40} = \frac{276,88}{40} \approx 6,92(m). \text{ nên } \mathbf{a \text{ đúng}}$$

b) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho bởi Bảng 15 (làm tròn kết quả đến hàng phần trăm) là:

$$s_D^2 = \frac{1}{40} \left[3 \cdot (6,34 - 6,92)^2 + 7 \cdot (6,58 - 6,92)^2 + 5 \cdot (6,82 - 6,92)^2 + 20 \cdot (7,06 - 6,92)^2 + 5 \cdot (7,30 - 6,92)^2 \right] = \frac{2,9824}{40} \approx 0,07.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s_D \approx \sqrt{0,07} \approx 0,26(m)$ nên **b đúng**

c) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho bởi Bảng 16 là:

$$\bar{x}_H = \frac{2 \cdot 6,34 + 5 \cdot 6,58 + 8 \cdot 6,82 + 19 \cdot 7,06 + 6 \cdot 7,30}{40} = \frac{278,08}{40} \approx 6,95(m).$$

Vậy phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho bởi Bảng 16 (làm tròn kết quả đến hàng phần trăm) là:

$$s_H^2 = \frac{1}{40} \left[2 \cdot (6,34 - 6,95)^2 + 5 \cdot (6,58 - 6,95)^2 + 8 \cdot (6,82 - 6,95)^2 + 19 \cdot (7,06 - 6,95)^2 + 6 \cdot (7,30 - 6,95)^2 \right] = \frac{2,5288}{40} \approx 0,06.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s_H \approx \sqrt{0,06} \approx 0,24(m)$ nên **c sai**

d) Do $s_H \approx 0,24 < s_D \approx 0,26$ nên kết quả nhảy xa của vận động viên Huy đồng đều hơn kết quả nhảy xa của vận động viên Dũng nên **d sai**

Câu 2. Bảng 1, Bảng 2 lần lượt biểu diễn mẫu số liệu ghép nhóm thống kê mức lương của hai công ty A, B (đơn vị: triệu đồng)

Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[10; 15)	12,5	15	[10; 15)	12,5	25
[15; 20)	17,5	18	[15; 20)	17,5	15
[20; 25)	22,5	10	[20; 25)	22,5	7
[25; 30)	27,5	10	[25; 30)	27,5	5
[30; 35)	32,5	5	[30; 35)	32,5	5
[35; 40)	37,5	2	[35; 40)	37,5	3
		$n = 60$			$n = 60$
Bảng 1			Bảng 2		

Các mệnh đề sau đúng hay sai?

- a) Số trung bình cộng của mẫu số liệu ghép nhóm của công ty A là: $\frac{62}{3}$
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm của công ty A là: 5
- c) Phương sai của mẫu số liệu ghép nhóm của công ty B (làm tròn kết quả đến hàng phần trăm) là: 52,91
- b) Nhận thấy độ lệch chuẩn của công ty A nhỏ hơn công ty B nên mức lương của công ty A đồng đều hơn

Lời giải

a) Đúng	b) Sai	c) Sai	d) Đúng
---------	--------	--------	---------

- a) Số trung bình cộng của mẫu số liệu ghép nhóm của công ty A là:
 $\bar{x}_A = \frac{15.12,5 + 18.17,5 + 10.22,5 + 10.27,5 + 5.32,5 + 2.37,5}{60} = \frac{62}{3}$ nên **a đúng**

- b) Phương sai của mẫu số liệu ghép nhóm của công ty A là:

$$s_A^2 = \frac{15 \cdot \left(12,5 - \frac{62}{3}\right)^2 + 18 \cdot \left(17,5 - \frac{62}{3}\right)^2 + 10 \cdot \left(22,5 - \frac{62}{3}\right)^2 + 10 \cdot \left(27,5 - \frac{62}{3}\right)^2}{60} + \frac{5 \cdot \left(32,5 - \frac{62}{3}\right)^2 + 2 \cdot \left(37,5 - \frac{62}{3}\right)^2}{60} \approx 49,14$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm của công ty A là: $s_A = \sqrt{s_A^2} = \sqrt{49,13} \approx 7$ nên **b sai**

c) Số trung bình cộng của mẫu số liệu ghép nhóm của công ty B là:
 $\bar{x}_B = \frac{25 \cdot 12,5 + 15 \cdot 17,5 + 7 \cdot 22,5 + 5 \cdot 27,5 + 5 \cdot 32,5 + 3 \cdot 37,5}{60} = \frac{229}{12}$

Phương sai của mẫu số liệu ghép nhóm của công ty B là:

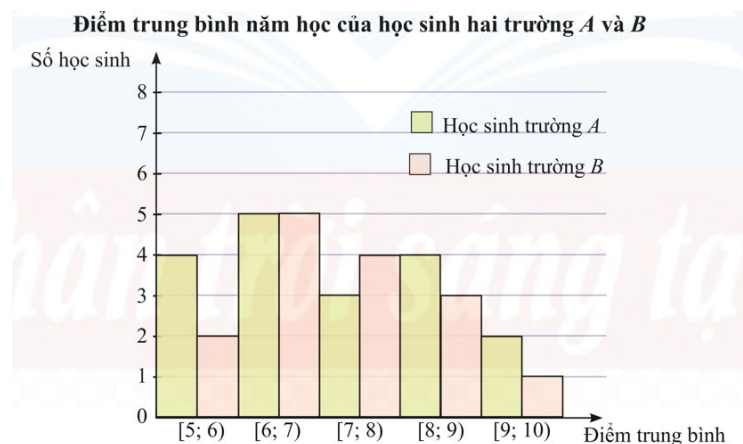
$$s_B^2 = \frac{25 \cdot \left(12,5 - \frac{229}{12}\right)^2 + 15 \cdot \left(17,5 - \frac{229}{12}\right)^2 + 7 \cdot \left(22,5 - \frac{229}{12}\right)^2 + 5 \cdot \left(27,5 - \frac{229}{12}\right)^2}{60} + \frac{5 \cdot \left(32,5 - \frac{229}{12}\right)^2 + 3 \cdot \left(37,5 - \frac{229}{12}\right)^2}{60} \approx 57,91$$

nên **c sai**

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm của công ty B là: $s_B = \sqrt{s_B^2} = \sqrt{57,91} \approx 7,61$

Nhận thấy độ lệch chuẩn của công ty A nhỏ hơn công ty B nên mức lương của công ty A đồng đều hơn **nên d đúng**

Câu 3. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



Người ta lập được bảng tần số ghép nhóm cho mẫu số liệu trên như 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
Học sinh trường A	4	5	3	4	2
Học sinh trường B	2	5	4	3	1

Các mệnh đề sau đúng hay sai?

- a) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm của học sinh trường A là: 6,1
- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm của học sinh trường B là: 1,73
- c) 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) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường A có điểm trung bình đồng đều hơn

Lời giải

a) Đúng	b) Đúng	c) Đúng	d) Sai
----------------	----------------	----------------	---------------

a) Cỡ mẫu: $n_A = 18$

Gọi $x_1; x_2; \dots; x_{18}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh hai 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

$$\text{số liệu ghép nhóm là: } Q_1 = 6 + \frac{\frac{18}{4} - 4}{5} (7 - 6) = 6,1 \text{ nên } \mathbf{a \text{ đúng}}$$

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 2,275$ nên **b sai**

b) Cỡ mẫu: $n_B = 15$

Gọi $y_1; y_2; \dots; y_{15}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh hai 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

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

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

$$\text{liệu ghép nhóm là: } Q'_3 = 8 + \frac{\frac{3.15}{4} - (2 + 5 + 4)}{3} (9 - 8) = 8,08$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q'_3 - Q'_1 = 1,73$

c) Vậy 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 nên **c đúng**

d) Xét số liệu của trường A:

$$\text{Số trung bình: } \bar{x}_A = \frac{4.5,5 + 5.6,5 + 3.7,5 + 4.8,5 + 2.9,5}{18} = 7,22$$

$$\text{Độ lệch chuẩn: } \sigma_A = \sqrt{\frac{4.5,5^2 + 5.6,5^2 + 3.7,5^2 + 4.8,5^2 + 2.9,5^2}{18} - 7,22^2} \approx 1,79$$

Xét số liệu của trường B:

$$\text{Số trung bình: } \bar{x}_B = \frac{2.5,5 + 5.6,5 + 4.7,5 + 3.8,5 + 1.9,5}{15} = 7,23$$

$$\text{Độ lệch chuẩn: } \sigma_B = \sqrt{\frac{2.5,5^2 + 5.6,5^2 + 4.7,5^2 + 3.8,5^2 + 1.9,5^2}{15} - 7,23^2} \approx 1,31$$

Vậy 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 nên **d sai**

Câu 4. Bảng 1, Bảng 2 lần lượt biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội và Huế (đơn vị: độ C).

Bảng 1			Bảng 2		
Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[16, 8; 19, 8)	18,3	2	[16, 8; 19, 8)	18,3	1
[19, 8; 22, 8)	21,3	3	[19, 8; 22, 8)	21,3	2
[22, 8; 25, 8)	24,3	2	[22, 8; 25, 8)	24,3	3
[25, 8; 28, 8)	27,3	1	[25, 8; 28, 8)	27,3	2
[28, 8; 31, 8)	30,3	4	[28, 8; 31, 8)	30,3	4
		$n = 12$			$n = 12$

Các mệnh đề sau đúng hay sai?

- a) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên số liệu ở Hà Nội là: 8,75
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm trên số liệu ở Hà Nội (làm tròn kết quả đến hàng phần trăm) là: 3,56
- c) Phương sai của mẫu số liệu ghép nhóm trên số liệu ở Huế là: 12,75
- d) Huế có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn

Lời giải

a) Đúng	b) Sai	c) Sai	d) Đúng
----------------	---------------	---------------	----------------

a) - Xét số liệu ở Hà Nội:

+ Khoảng biến thiên: $R = 31,8 - 16,8 = 15$

+ Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 2, cf_2 = 5, cf_3 = 7, cf_4 = 8, cf_5 = 12$

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ mà $2 < 3 < 5$ suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3. Xét nhóm 2 là nhóm [19,8;22,8) có $s = 19,8, h = 3, n_2 = 3$ và nhóm 1 là nhóm [16,8;19,8) có $cf_1 = 2$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{3 - cf_1}{n_2} \right) \cdot h = 19,8 + \left(\frac{3 - 2}{3} \right) \cdot 3 = 20,8$

Ta có: $\frac{3n}{4} = \frac{3.12}{4} = 9$ mà $8 < 9 < 12$ suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 5 là nhóm $[28,8;31,8)$ có $t = 28,8, l = 3, n_5 = 4$ và nhóm 4 là nhóm $[25,8;28,8)$ có $cf_4 = 8$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{9 - cf_4}{n_5} \right) \cdot l = 28,8 + \left(\frac{9 - 8}{4} \right) \cdot 3 = 29,55$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = 29,55 - 20,8 = 8,75$ nên **a đúng**

b) + Số trung bình cộng của mẫu số liệu ghép nhóm là:
 $\bar{x}_1 = \frac{2.18,3 + 3.21,3 + 2.24,3 + 27,3 + 4.30,3}{12} = 24,8$

Phương sai của mẫu số liệu ghép nhóm là:

$$s_1^2 = \frac{2(18,3 - 24,8)^2 + 3(21,3 - 24,8)^2 + 2(24,3 - 24,8)^2 + (27,3 - 24,8)^2 + 4(30,3 - 24,8)^2}{12} = 20,75$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s_1 = \sqrt{s_1^2} = \sqrt{20,75} \approx 4,56$ nên **b sai**

c) - Xét số liệu ở Huế:

+ Khoảng biến thiên: $R = 31,8 - 16,8 = 15$

+ Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 1, cf_2 = 3, cf_3 = 6, cf_4 = 8, cf_5 = 12$

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 2 là nhóm $[19,8;22,8)$ có $s = 19,8, h = 3, n_2 = 2$ và nhóm 1 là nhóm $[16,8;19,8)$ có $cf_1 = 1$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{3 - cf_1}{n_2} \right) \cdot h = 19,8 + \left(\frac{3 - 1}{2} \right) \cdot 3 = 22,8$

Ta có: $\frac{3n}{4} = \frac{3.12}{4} = 9$ mà $8 < 9 < 12$ suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 5 là nhóm $[28,8;31,8)$ có $t = 28,8, l = 3, n_5 = 4$ và nhóm 4 là nhóm $[25,8;28,8)$ có $cf_4 = 8$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{9 - cf_4}{n_5} \right) \cdot l = 28,8 + \left(\frac{9 - 8}{4} \right) \cdot 3 = 29,55$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = 29,55 - 22,8 = 6,75$

+ Số trung bình cộng của mẫu số liệu ghép nhóm là:

$$\bar{x}_2 = \frac{18,3 + 2.21,3 + 3.24,3 + 2.27,3 + 4.30,3}{12} = 25,8$$

Phương sai của mẫu số liệu ghép nhóm là:

$$s_2^2 = \frac{(18,3 - 25,8)^2 + 3(21,3 - 25,8)^2 + 3(24,3 - 25,8)^2 + 2(27,3 - 25,8)^2 + 4(30,3 - 25,8)^2}{12} = 15,75$$

Nên c sai

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s_2 = \sqrt{s_2^2} = \sqrt{15,75} \approx 3,97$

Huế có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn **nên d đúng**

Câu 5. Bảng 24 thống kê độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt và Vũng Tàu (đơn vị: %).

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Độ ẩm												
Đà Lạt	83	79	79	87	87	87	88	89	90	91	88	86
Vũng Tàu	75	77	78	77	79	79	81	79	81	83	80	77

Bảng 24

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

Người ta lần lượt ghép các số liệu của Đà Lạt, Vũng Tàu thành mẫu số liệu ghép nhóm như sau:

Đà Lạt			Vũng Tàu		
Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[75; 78,3)	76,65	0	[75; 78,3)	76,65	5
[78,3; 81,6)	79,95	2	[78,3; 81,6)	79,95	6
[81,6; 84,9)	83,25	1	[81,6; 84,9)	83,25	1
[84,9; 88,2)	86,55	6	[84,9; 88,2)	86,55	0
[88,2; 91,5)	89,85	3	[88,2; 91,5)	89,85	0

Các mệnh đề sau đúng hay sai?

- a) Xét số liệu ở Đà Lạt ta có khoảng biến thiên là: 16,5
- b) Xét số liệu ở Đà Lạt thì ta có độ lệch chuẩn của mẫu số liệu ghép nhóm (làm tròn kết quả đến hàng phần trăm) là: 3,28
- c) Xét số liệu ở Vũng Tàu thì phương sai của mẫu số liệu ghép nhóm là: 4,235
- d) Đà Lạt có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn

Lời giải

a) Đúng	b) Đúng	c) Đúng	d) Sai
----------------	----------------	----------------	---------------

a) - Xét số liệu ở Đà Lạt:

+ Khoảng biến thiên: $R = 91,5 - 75 = 16,5$ nên **a đúng**

b) + Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 0, cf_2 = 2, cf_3 = 3, cf_4 = 9, cf_5 = 12$

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ suy ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 3 là nhóm $[81,6;84,9)$ có $s = 81,6, h = 3,3, n_3 = 1$ và nhóm 2 là nhóm $[78,3;81,6)$ có $cf_2 = 2$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{3 - cf_2}{n_3} \right) \cdot h = 81,6 + \left(\frac{3 - 2}{1} \right) \cdot 3,3 = 84,9$

Ta có: $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$ suy ra nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9.

Xét nhóm 4 là nhóm $[84,9;88,2)$ có $t = 84,9, l = 3,3, n_4 = 6$ và nhóm 3 là nhóm $[81,6;84,9)$ có $cf_3 = 3$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{9 - cf_3}{n_4} \right) \cdot l = 84,9 + \left(\frac{9 - 3}{6} \right) \cdot 3,3 = 88,2$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = 88,2 - 84,9 = 3,3$

+ Số trung bình cộng của mẫu số liệu ghép nhóm là:
 $\bar{x}_1 = \frac{0,76,65 + 2,79,95 + 83,25 + 6,86,55 + 3,89,85}{12} = 86$

Phương sai của mẫu số liệu ghép nhóm là:

$$s_1^2 = \frac{0(76,65 - 86)^2 + 2(79,95 - 86)^2 + (83,25 - 86)^2 + 6(86,55 - 86)^2 + 3(89,95 - 86)^2}{12} = 10,7825$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s_1 = \sqrt{s_1^2} = \sqrt{10,7825} \approx 3,28$ nên **b đúng**

c) - Xét số liệu ở Vũng Tàu:

+ Khoảng biến thiên: $R = 91,5 - 75 = 16,5$

+ Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 5, cf_2 = 11, cf_3 = 12, cf_4 = 12, cf_5 = 12$

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ suy ra nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 1 là nhóm $[75;78,3)$ có $s = 75, h = 3,3, n_1 = 5$

Ta có tứ phân vị thứ nhất là: $Q_1 = s + \left(\frac{3 - cf_0}{n_1} \right) \cdot h = 75 + \left(\frac{3 - 0}{5} \right) \cdot 3,3 = 76,98$

Ta có: $\frac{3n}{4} = \frac{3.12}{4} = 9$ suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

9. Xét nhóm 2 là nhóm $[78,3;81,6)$ có $t = 78,3, l = 3,3, n_2 = 6$ và nhóm 1 là nhóm $[75;78,3)$ có $cf_1 = 5$

Ta có tứ phân vị thứ ba là: $Q_3 = t + \left(\frac{9 - cf_1}{n_2} \right) . l = 75 + \left(\frac{9 - 5}{6} \right) . 3,3 = 77,2$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là: $Q_3 - Q_1 = 77,2 - 76,98 = 0,22$

+ Số trung bình cộng của mẫu số liệu ghép nhóm là:
 $\bar{x}_2 = \frac{5.76,65 + 6.79,95 + 83,25}{12} = 78,85$

Phương sai của mẫu số liệu ghép nhóm là:

$$s_2^2 = \frac{5(76,65 - 78,85)^2 + 6(79,95 - 78,85)^2 + (83,25 - 78,85)^2}{12} = 4,235 \text{ nên c đúng}$$

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s_2 = \sqrt{s_2^2} = \sqrt{4,235} \approx 2,06$

Vũng Tàu có nhiệt độ không khí trung bình tháng đồng đều hơn vì độ lệch chuẩn nhỏ hơn
 nên d sai

Câu 6. 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

(Nguồn: Tổng cục Thống kê)

Các mệnh đề sau đúng hay sai?

a) Xét số liệu ở Nha Trang thì khoảng tứ phân vị của mẫu số liệu ghép nhóm là: 32,64

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

c) Xét số liệu của Quy Nhơn ta có độ lệch chuẩn của mẫu số liệu ghép nhóm (làm tròn kết quả đến hàng phần trăm) là: 30,59

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

Lời giải

a) Sai	b) Đúng	c) Đúng	d) Sai
---------------	----------------	----------------	---------------

a) Cỡ mẫu: $n = 20$

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về số giờ nắng trong tháng 6 trong 20 năm của 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 $\frac{1}{2}(x_5 + x_6) \in [220; 250)$. Do đó, tứ phân vị thứ

nhất của mẫu số 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 $\frac{1}{2}(x_{15} + x_{16}) \in [250; 280)$. Do đó, tứ phân vị thứ

ba của mẫu số liệu ghép nhóm là: $Q_3 = 250 + \frac{\frac{3.20}{4} - (1+1+1+8)}{7}(280 - 250) = \frac{1870}{7}$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 39,64$ nên **a sai**

b) Gọi $y_1; y_2; \dots; y_{30}$ là mẫu số liệu gốc về số giờ nắng trong tháng 6 trong 20 năm của Quy Nhơn được xếp theo thứ tự không giảm.

Ta có:

$y_1; \dots; y_7 \in [160; 190); y_8; \dots; y_{14} \in [190; 220); y_{15}; \dots; y_{21} \in [220; 250); y_{22}; \dots; y_{28} \in [250; 280); y_{29}; y_{30} \in [280; 310)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(y_5 + y_6) \in [220; 250)$. Do đó, tứ phân vị thứ

nhất của mẫu số liệu ghép nhóm là: $Q'_1 = 220 + \frac{\frac{20}{4} - (1+2)}{4}(250 - 220) = 235$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(y_{15} + y_{16}) \in [250; 280)$. Do đó, tứ phân vị thứ

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_{Q'} = Q'_3 - Q'_1 = 39$

Vậy nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn nên **b đúng**

Xét số liệu của Nha Trang:

Số trung bình: $\bar{x}_X = \frac{1.145 + 1.175 + 1.205 + 8.235 + 7.265 + 2.295}{20} = 242,5$

Độ lệch chuẩn:

$\sigma_X = \sqrt{\frac{1.145^2 + 1.175^2 + 1.205^2 + 8.235^2 + 7.265^2 + 2.295^2}{20} - 242,5^2} \approx 35,34$

c) Xét số liệu của Quy Nhơn:

$$\text{Số trung bình: } \bar{x}_Y = \frac{1.175 + 2.205 + 4.235 + 10.265 + 3.295}{20} = 253$$

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{1.175^2 + 2.205^2 + 4.235^2 + 10.265^2 + 3.295^2}{20} - 253^2} \approx 30,59 \text{ nên c đúng}$$

d) Vậy nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn **nên d sai**

Câu 7. Thống kê tổng số giờ nắng trong tháng 9 tại một trạm quan trắc đặt ở Cà Mau trong các năm từ 2002 đến 2021 được thống kê như sau:

111,6 134,9 130,3 134,2 140,9 109,3 154,4 156,3 116,1 96,7
105,2 80,8 80,8 110 109 139 145 161 126 114

(Nguồn: Tổng cục Thống kê)

Người ta lập được bảng tần số ghép nhóm như sau:

Số giờ nắng	[80;98)	[98;116)	[116;134)	[134;152)	[152;170)
Giá trị đại diện	89	107	125	143	161
Số năm	3	6	3	5	3

Các mệnh đề sau đúng hay sai?

a) Số trung bình của mẫu số liệu ghép nhóm là 124,1.

b) Phương sai của mẫu số liệu ghép nhóm là 566,19.

c) Độ lệch chuẩn của mẫu số liệu ghép nhóm (kết quả các phép tính làm tròn đến hàng phần nghìn) là 23,795.

d) Sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc (kết quả các phép tính làm tròn đến hàng phần nghìn) là 4,805%.

Lời giải

a) Đúng	b) Đúng	c) Đúng	d) Đúng
----------------	----------------	----------------	----------------

-Xét theo mẫu số liệu gốc

Cỡ mẫu là $n = 20$.

Số trung bình của mẫu số liệu trên là

$$\bar{x}_1 = \frac{111,6 + 134,9 + \dots + 114}{20} = 122,755.$$

Phương sai của mẫu số liệu trên là

$$S_1^2 = \frac{1}{20} (111,6^2 + 134,9^2 + \dots + 114^2) - 122,755^2 \approx 515,453.$$

Độ lệch chuẩn của mẫu số liệu trên là

$$S_1 \approx \sqrt{515,453} \approx 22,704.$$

-Xét theo mẫu số liệu ghép nhóm

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{3.89 + 6.107 + 3.125 + 5.143 + 3.161}{20} = 124,1. \text{ nên } \mathbf{a \text{ đúng}}$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{20} (3.89^2 + 6.107^2 + 3.125^2 + 5.143^2 + 3.161^2) - 124,1^2 = 566,19. \text{ nên } \mathbf{b \text{ đúng}}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_2 = \sqrt{566,19} \approx 23,795.77 \text{ nên } \mathbf{c \text{ đúng}}$$

Sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc là

$$\frac{|S_2 - S_1|}{S_1} = \frac{|23,795 - 22,704|}{22,704} \cdot 100\% \approx 4,805\%. \text{ nên } \mathbf{d \text{ đúng}}$$

Câu 8. Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11A và 11B ở bảng sau:

Lớp \ Điểm trung bình					
	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
11A	1	0	11	22	6
11B	0	6	8	14	12

Các mệnh đề sau đúng hay sai?

- a)** Khoảng biến thiên của điểm số học sinh lớp 11A là: 5.
- b)** Nếu so sánh theo khoảng biến thiên thì điểm trung bình của các học sinh lớp 11B ít phân tán hơn điểm trung bình của các học sinh lớp 11A.
- c)** Xét mẫu số liệu của lớp 11A ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{0,51}$.
- d)** Nếu so sánh theo độ lệch chuẩn thì học sinh lớp 11A có điểm trung bình ít phân tán hơn học sinh lớp 11B.

Lời giải

a) Đúng	b) Đúng	c) Sai	d) Đúng
----------------	----------------	---------------	----------------

-Khoảng biến thiên của điểm số học sinh lớp 11A là: $10 - 5 = 5$. Nên **a đúng**

Khoảng biến thiên của điểm số học sinh lớp 11B là: $10 - 6 = 4$.

Nếu so sánh theo khoảng biến thiên thì điểm trung bình của các học sinh lớp 11B ít phân tán hơn điểm trung bình của các học sinh lớp 11A nên **b đúng**

-Ta có bảng thống kê điểm trung bình theo giá trị đại diện:

Giá trị đại diện	5,5	6,5	7,5	8,5	9,5
Lớp					
11A	1	0	11	22	6
11B	0	6	8	14	12

- Xét mẫu số liệu của lớp 11A:

Cỡ mẫu là $n_1 = 1 + 11 + 22 + 6 = 40$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{1 \cdot 5,5 + 11 \cdot 7,5 + 22 \cdot 8,5 + 6 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{40} (1 \cdot 5,5^2 + 11 \cdot 7,5^2 + 22 \cdot 8,5^2 + 6 \cdot 9,5^2) - 8,3^2 = 0,61.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_1 = \sqrt{0,61}$ **nên c sai**

- Xét mẫu số liệu của lớp 11B:

Cỡ mẫu là $n_2 = 6 + 8 + 14 + 12 = 40$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{6 \cdot 6,5 + 8 \cdot 7,5 + 14 \cdot 8,5 + 12 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

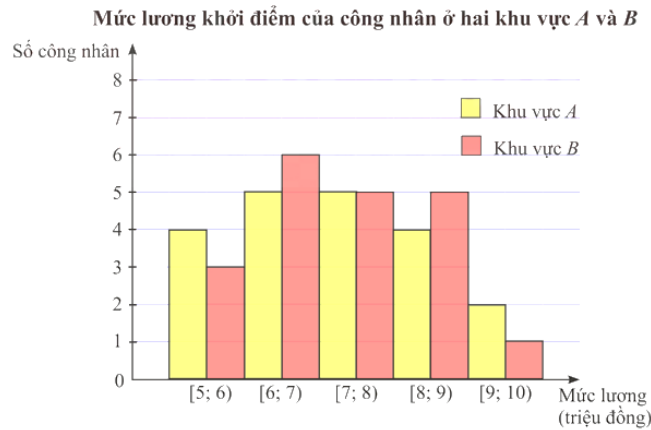
$$S_2^2 = \frac{1}{40} (6 \cdot 6,5^2 + 8 \cdot 7,5^2 + 14 \cdot 8,5^2 + 12 \cdot 9,5^2) - 8,3^2 = 1,06.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{1,06}$.

Do $S_1 < S_2$ nên nếu so sánh theo độ lệch chuẩn thì học sinh lớp 11A có điểm trung bình ít phân tán hơn học sinh lớp 11B nên **d đúng**

Chú ý: Trong bài toán trên, kết quả so sánh độ phân tán theo giá trị trung bình và độ lệch chuẩn có sự khác biệt. Điều này là do mẫu số liệu của học sinh lớp 11A có một giá trị ngoại lệ.

Câu 9. Biểu đồ dưới đây mô tả kết quả điều tra về mức lương khởi điểm (đơn vị: triệu đồng) của một số công nhân ở hai khu vực A và B.



Người ta lập được bảng tần số ghép nhóm cho mẫu số liệu như sau:

Mức lương	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Mức lương đại diện (triệu đồng)	5,5	6,5	7,5	8,5	9,5
Khu vực A	4	5	5	4	2
Khu vực B	3	6	5	5	1

Các mệnh đề sau đúng hay sai?

- a)** Xét mẫu số liệu của khu vực A ta có số trung bình của mẫu số liệu ghép nhóm là 6,25.
- b)** Xét mẫu số liệu của khu vực A ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{1,5875}$
- c)** Xét mẫu số liệu của khu vực B ta có phương sai của mẫu số liệu ghép nhóm là 1,3875.
- d)** Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì mức lương khởi điểm của công nhân khu vực B đồng đều hơn của công nhân khu vực A.

Lời giải

a) Sai	b) Đúng	c) Sai	d) Đúng
---------------	----------------	---------------	----------------

- Xét mẫu số liệu của khu vực A :

Cỡ mẫu là $n_A = 4 + 5 + 5 + 4 + 2 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_A = \frac{4.5,5 + 5.6,5 + 5.7,5 + 4.8,5 + 2.9,5}{20} = 7,25. \text{ nên a sai}$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_A^2 = \frac{1}{20} (4.5,5^2 + 5.6,5^2 + 5.7,5^2 + 4.8,5^2 + 2.9,5^2) - (7,25)^2 = 1,5875$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_A = \sqrt{1,5875} \text{ nên b đúng}$$

- Xét mẫu số liệu của khu vực B :

Cỡ mẫu là $n_B = 3 + 6 + 5 + 5 + 1 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_B = \frac{3 \cdot 5,5 + 6 \cdot 6,5 + 5 \cdot 7,5 + 5 \cdot 8,5 + 1 \cdot 9,5}{20} = 7,25.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_B^2 = \frac{1}{20} (3 \cdot 5,5^2 + 6 \cdot 6,5^2 + 5 \cdot 7,5^2 + 5 \cdot 8,5^2 + 1 \cdot 9,5^2) - (7,25)^2 = 1,2875. \text{ nên c sai}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là

$$S_B = \sqrt{1,2875}$$

Do $S_A > S_B$ nên nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì mức lương khởi điểm của công nhân khu vực B đồng đều hơn của công nhân khu vực A nên **d đúng**

Chú ý: Với các mẫu số liệu ghép nhóm có cùng số trung bình (hoặc xấp xỉ nhau), ta thường sử dụng phương sai và độ lệch chuẩn để so sánh mức độ phân tán của các mẫu số liệu đó.

Câu 10. Giá đóng cửa của một cổ phiếu là giá của cổ phiếu đó cuối một phiên giao dịch. Bảng sau thống kê giá đóng cửa (đơn vị: nghìn đồng) của hai mã cổ phiếu A và B trong 50 ngày giao dịch liên tiếp.

Giá đóng cửa	[120; 122)	[122; 124)	[124; 126)	[126; 128)	[128; 130)
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Các mệnh đề sau đúng hay sai?

a) Xét mẫu số liệu của cổ phiếu A ta có phương sai của mẫu số liệu ghép nhóm là 7,5216

b) Xét mẫu số liệu của cổ phiếu B ta có số trung bình của mẫu số liệu ghép nhóm là 115,28.

c) Xét mẫu số liệu của cổ phiếu B ta có độ lệch chuẩn của mẫu số liệu ghép nhóm là $\sqrt{15,4096}$

d) Người ta có thể dùng phương sai và độ lệch chuẩn để so sánh mức độ rủi ro của các loại cổ phiếu có giá trị trung bình gần bằng nhau. Cổ phiếu nào có phương sai, độ lệch chuẩn cao hơn thì được coi là có độ rủi ro lớn hơn. Theo quan điểm trên, thì cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B .

Lời giải

a) Đúng	b) Sai	c) Sai	d) Đúng
----------------	---------------	---------------	----------------

Ta có bảng thống kê giá đóng cửa theo giá trị đại diện:

Giá đóng cửa	121	123	125	127	129
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

- Xét mẫu số liệu của cổ phiếu A :

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{8.121 + 9.123 + 12.125 + 10.127 + 11.129}{50} = 125,28.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{50} (8.121^2 + 9.123^2 + 12.125^2 + 10.127^2 + 11.129^2) - (125,28)^2 = 7,5216 \text{ nên a đúng}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_1 = \sqrt{S_1^2} = \sqrt{7,5216}$.

- Xét mẫu số liệu của cổ phiếu B:

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{16.121 + 4.123 + 3.125 + 6.127 + 21.129}{50} = 125,28. \text{ nên b sai}$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{50} (16.121^2 + 4.123^2 + 3.125^2 + 6.127^2 + 21.129^2) - (125,48)^2 = 12,4096.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{S_2^2} = \sqrt{12,4096}$. **Nên c sai**

Vậy nếu đánh giá độ rủi ro theo phương sai và độ lệch chuẩn thì cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B. **Nên d đúng**

F. TRẢ LỜI NGẮN

PHẦN E. TRẢ LỜI NGẮN

Câu 1. Hãy tính độ lệch chuẩn của mẫu số liệu ghép nhóm (kết quả được làm tròn đến hàng phần trăm)

Chiều cao (cm)	[160;164)	[164;168)	[168;172)	[172;176)	[176;180)
Số học sinh	3	5	8	4	1

Trả lời: 4,26

Lời giải

Cỡ mẫu: $n = 21$

Giá trị trung bình của mẫu số liệu mới:

$$\bar{x} = \frac{1}{n} (n_1 c_1 + n_2 c_2 + \dots + n_k c_k) = \frac{1}{21} (3.162 + 5.166 + 8.170 + 4.174 + 1.178) = \frac{3550}{21}$$

Phương sai của mẫu số liệu mới:

$$\begin{aligned} S^2 &= \frac{1}{n} [n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2] \\ &= \frac{1}{21} \left[3 \left(162 - \frac{3550}{21} \right)^2 + 5 \left(166 - \frac{3550}{21} \right)^2 + \dots + 1 \left(178 - \frac{3550}{21} \right)^2 \right] = \frac{8000}{441} \end{aligned}$$

$$\text{Độ lệch chuẩn của mẫu số liệu mới: } \sigma = \sqrt{S^2} = \sqrt{\frac{8000}{441}} = \frac{40\sqrt{5}}{21} \approx 4,26$$

Câu 2. 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 như 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 độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

Trả lời: 0,53

Lời giải

Chọn giá trị đại diện cho các nhóm 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 nghe nhạc liên tục của điện thoại là:

$$\bar{x} = \frac{1}{40}(5,25 \cdot 2 + 5,75 \cdot 8 + 6,25 \cdot 15 + 6,75 \cdot 10 + 7,25 \cdot 5) = 6,35$$

Phương sai của mẫu số liệu là:

$$s^2 = \frac{1}{40}(5,25^2 \cdot 2 + 5,75^2 \cdot 8 + 6,25^2 \cdot 15 + 6,75^2 \cdot 10 + 7,25^2 \cdot 5) - 6,35^2 = 0,2775$$

$$\text{Độ lệch chuẩn của mẫu số liệu là: } \sqrt{0,2775} = \frac{\sqrt{111}}{20} \approx 0,53$$

Câu 3. 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 độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

Trả lời: 0,53

Lời giải

Giá trị đại diện	19,25	19,75	20,25	20,75	21,25
Tần số	13	45	24	12	6

Cỡ mẫu: $n = 100$

$$\text{Số trung bình: } \bar{x} = \frac{13.19,25 + 45.19,75 + 24.20,25 + 12.20,75 + 6.21,25}{100} = 20,015$$

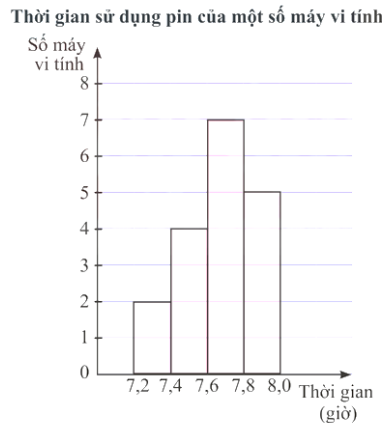
Phương

sai:

$$S^2 = \frac{13.19,25^2 + 45.19,75^2 + 24.20,25^2 + 12.20,75^2 + 6.21,25^2}{100} - 20,015^2 \approx 0,28$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{0,28} \approx 0,53$$

Câu 4. 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.



Hãy xác định độ lệch chuẩn của thời gian sử dụng pin (kết quả được làm tròn đến hàng phần trăm)

Trả lời: 0,19

Lời giải

Giá trị đại diện	7,3	7,5	7,7	7,9
Số máy	2	4	7	5

Cỡ mẫu: $n = 18$

$$\text{Số trung bình: } \bar{x} = \frac{2.7,3 + 4.7,5 + 7.7,7 + 5.7,9}{18} \approx 7,67$$

$$\text{Phương sai: } S^2 = \frac{2.7,3^2 + 4.7,5^2 + 7.7,7^2 + 5.7,9^2}{18} - 7,67^2 \approx 0,04$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{0,04} \approx 0,19$$

Câu 5. 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

Hãy xác định độ lệch chuẩn của thời gian sử dụng pin (kết quả được làm tròn đến hàng phần trăm)

Trả lời: 1,53

Lời giải

Mẫu số liệu với giá trị đại diện

Mức xà (cm)	[170;172)	[172;174)	[174;176)	[176;178)
Giá trị đại diện	171	173	175	177
Số vận động viên	3	10	6	1

$$\text{Giá trị trung bình: } \bar{x} = \frac{1}{20}(171 \cdot 3 + 173 \cdot 10 + 175 \cdot 6 + 177 \cdot 1) = 173,5(\text{cm})$$

Phương sai của mẫu số liệu:

$$s^2 = \frac{1}{20}(171^2 \cdot 3 + 173^2 \cdot 10 + 175^2 \cdot 6 + 177^2 \cdot 1) - 173,5^2 = 2,35$$

$$\text{Độ lệch chuẩn của mẫu số liệu: } s = \sqrt{2,35} = \frac{\sqrt{235}}{10} \approx 1,53(\text{cm})$$

Câu 6. Chiều dài của 40 bé trai sơ sinh 12 ngày tuổi chọn ngẫu nhiên ở một bệnh viện được nhà nghiên cứu thống kê trong Bảng dưới đây:

Chiều dài (cm)	[44;46)	[46;48)	[48;50)	[50;52)	[52;54)	[54;56)
Số trẻ	3	3	10	15	7	2

Tính độ lệch chuẩn của chiều dài nhóm 40 bé trai sơ sinh (làm tròn kết quả đến hàng phần nghìn).

Trả lời: 2,43

Lời giải

Bổ sung thêm các giá trị đại diện vào Bảng 3.17, ta lập được bảng sau:

Chiều dài của 40 bé trai sơ sinh (đơn vị: cm)

Nhóm	c_i	n_i
[44;46)	45	3
[46;48)	47	3
[48;50)	49	10
[50;52)	51	15
[52;54)	53	7
[54;56)	55	2
		$N = 40$

Từ mẫu số liệu đã cho, ta tính được số trung bình là:

$$\bar{x} = \frac{3 \cdot 45 + 3 \cdot 47 + 10 \cdot 49 + 15 \cdot 51 + 7 \cdot 53 + 2 \cdot 55}{40} = \frac{2012}{40} = 50,3.$$

$\bar{x}$ không phải là số nguyên nên để tính phương sai ta tính:

$$\overline{x^2} = \frac{3 \cdot 45^2 + 3 \cdot 47^2 + 10 \cdot 49^2 + 15 \cdot 51^2 + 7 \cdot 53^2 + 2 \cdot 55^2}{40} = 2536.$$

Do đó $s^2 = \overline{x^2} - (\bar{x})^2 = 2536 - 50,3^2 = 2536 - 2530,09 = 5,91$.

Vậy mẫu số liệu về chiều dài của 40 trẻ sơ sinh có độ lệch chuẩn là $s = \sqrt{5,91} \approx 2,43$.

Câu 7. 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

Hãy xác định độ lệch chuẩn của mẫu số liệu trên

Trả lời: 3100

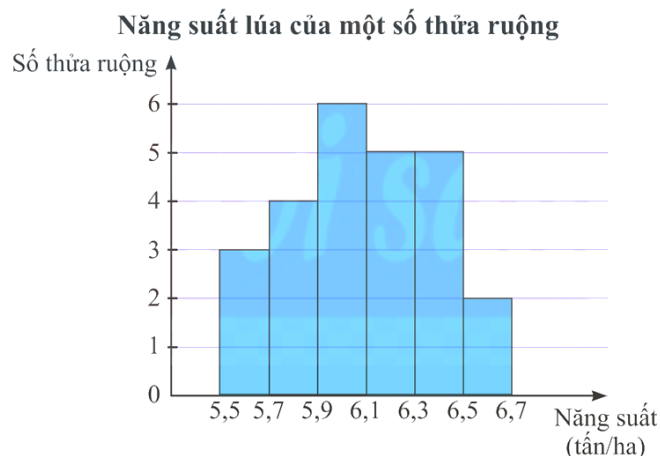
Lời giải

Giá trị đại diện	75	125	175	225	275
Số ngày	5	10	9	9	2

Số trung bình: $\bar{x} = \frac{5.75 + 10.125 + 9.175 + 4.225 + 2.275}{30} = 155$

Độ lệch chuẩn: $\sigma = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} = 3100$

Câu 8. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



Người ta lập được bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên như sau:

Năng suất	[5,5;5,7)	[5,7;5,9)	[5,9;6,1)	[6,1;6,3)	[6,3;6,5)	[6,5;6,7)
Số thửa ruộng	3	4	6	5	5	2
Giá trị đại diện	5,6	5,8	6,0	6,2	6,4	6,6

Tần số tương đối	3	4	6	5	5	2
------------------	---	---	---	---	---	---

Hãy xác định độ lệch chuẩn của mẫu số liệu trên (làm tròn kết quả đến hàng phần nghìn).

Trả lời: 0,29

Lời giải

$$\text{Số trung bình: } \bar{x} = \frac{3.5,6 + 4.5,8 + 6.6,0 + 5.6,2 + 5.6,4 + 2.6,6}{25} = 6,088$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} \approx 0,29$$

Câu 9. Một vận động viên luyện tập chạy cự li 100 m đã ghi lại kết quả luyện tập như sau:

Thời gian (giây)	[10,2;10,4)	[10,4;10,6)	[10,6;10,8)	[10,8;11)
Số vận động viên	3	7	8	2

Hãy xác định phương sai của mẫu số liệu trên

Trả lời: 0,0299

Lời giải

Chọn giá trị đại diện cho các nhóm số liệu, ta có:

Thời gian (giây)	[10,2;10,4)	[10,4;10,6)	[10,6;10,8)	[10,8;11)
Giá trị đại diện	10,3	10,5	10,7	10,9
Số vận động viên	3	7	8	2

Tổng số vận động viên là: $3 + 7 + 8 + 2 = 20$

Thời gian chạy trung bình của các vận động viên là:

$$\bar{x} = \frac{1}{20}(10,3 \cdot 3 + 10,5 \cdot 7 + 10,7 \cdot 8 + 10,9 \cdot 2) = 10,59 \text{ (giây)}$$

Phương sai của mẫu số liệu là:

$$s^2 = \frac{1}{20}(10,3^2 \cdot 3 + 10,5^2 \cdot 7 + 10,7^2 \cdot 8 + 10,9^2 \cdot 2) - 10,59^2 = 0,0299$$

Câu 10. Cân nặng của một số quả mít trong một khu vườn được thống kê ở bảng sau:

Cân nặng (kg)	[4;6)	[6;8)	[8;10)	[10;12)	[12;14)
Số quả mít	6	12	19	9	4

Hãy tính phương sai của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Trả lời: 4,8

Lời giải

Ta có bảng thống kê cân nặng của các quả mít theo giá trị đại diện:

Cân nặng đại diện (kg)	5	7	9	11	13
Tần số	6	12	19	9	4

Cỡ mẫu $n = 6 + 12 + 19 + 9 + 4 = 50$.

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x} = \frac{6 \cdot 5 + 12 \cdot 7 + 19 \cdot 9 + 9 \cdot 11 + 4 \cdot 13}{50} = 8,72.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S^2 = \frac{1}{50} (6 \cdot 5^2 + 12 \cdot 7^2 + 19 \cdot 9^2 + 9 \cdot 11^2 + 4 \cdot 13^2) - 8,72^2 \approx 4,80.$$

Câu 11. Bảng 21 biểu diễn mẫu số liệu ghép nhóm về độ tuổi của cư dân trong một khu phố.

Nhóm	Giá trị đại diện	Tần số
[20 ; 30)	25	25
[30 ; 40)	35	20
[40 ; 50)	45	20
[50 ; 60)	55	15
[60 ; 70)	65	14
[70 ; 80)	75	6
		$n = 100$

Bảng 21

Hãy tính độ lệch của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Trả lời: 15,63

Lời giải

Số trung bình cộng của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{25 \cdot 25 + 20 \cdot 35 + 20 \cdot 45 + 15 \cdot 55 + 14 \cdot 65 + 6 \cdot 75}{100} = 44,1$$

Phương sai của mẫu số liệu ghép nhóm là:

$$\begin{aligned}
s^2 &= \frac{25.(25-44,1)^2 + 20.(35-44,1)^2 + 20.(45-44,1)^2 + 15.(55-44,1)^2}{100} \\
&+ \frac{14.(65-44,1)^2 + 6.(75-44,1)^2}{100} \\
&= 244,19
\end{aligned}$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s = \sqrt{s^2} = \sqrt{244,19} \approx 15,63$

ÔN TẬP CHƯƠNG III

PHẦN 1. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 1: 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:

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.

Bài 2: Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi được thống kê lại ở bảng sau:

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.

Bài 3: 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:

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?

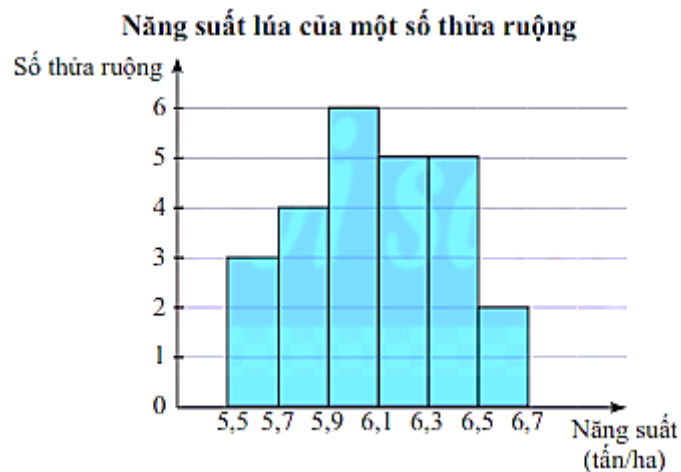
- A.** 5,98. **B.** 6. **C.** 2,44. **D.** 2,5.

Bài 4: 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

Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên

Bài 5: Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



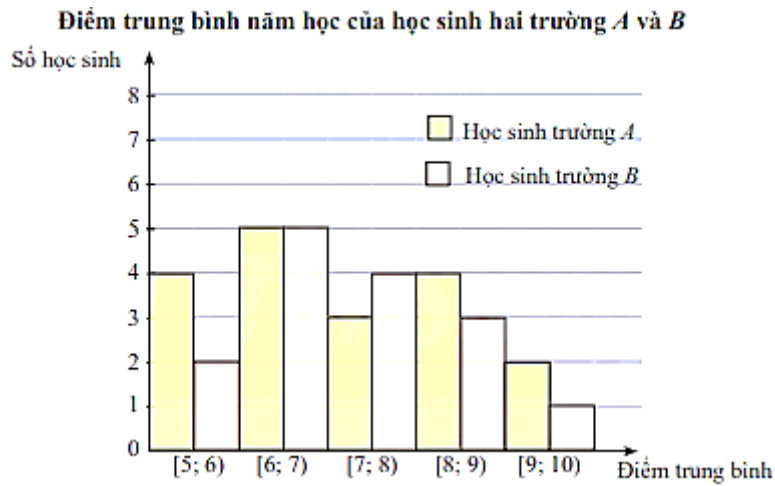
- Có bao nhiêu thửa ruộng đã được khảo sát?
- Lập bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên.
- Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên.

Bài 6: 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:

- Nếu so sánh theo số trung bình thì học sinh trường nào viết nhanh hơn?
- Nếu so sánh theo khoảng tứ phân vị thì học sinh trường nào có tốc độ viết đồng đều hơn?
- Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

Bài 7: Khoảng tứ phân vị nhỏ hơn thì số giờ nắng trong tháng 6 của địa phương đó đồng đều hơn Độ lệch chuẩn nhỏ hơn thì số giờ nắng trong tháng 6 của địa phương đó đồng đều hơn

Bài 8: Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



- a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu trên.
- b) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đồng đều hơn?
- c) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đồng đều hơn?

PHẦN 2. BÀI TẬP THÊM

Bài 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

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A.** 3. **B.** 4. **C.** 5. **D.** 6.

Bài 2: 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

Nhóm chứa tứ phân vị thứ nhất là

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

Bài 3: 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

Nhóm chứa tứ phân vị thứ ba là

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

Bài 4: 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

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.

Bài 5: 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

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.

Bài 6: 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.

Bài 7: 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:

- 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.

Bài 8: 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?

Bài 9: 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.

Bài 10: Cho mẫu số liệu ghép nhóm có tứ phân vị thứ nhất, thứ hai, thứ ba lần lượt là Q_1 , Q_2 , Q_3 . Khoảng tứ phân vị của mẫu số liệu ghép nhóm đó bằng:

A. $2Q_2$.

B. $Q_1 - Q_3$.

C. $Q_3 - Q_1$.

D. $Q_3 + Q_1 - Q_2$.

Bài 11: Bảng 22, Bảng 23 lần lượt biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội và Huế (đơn vị: độ C).

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	2
[19,8; 22,8)	21,3	3
[22,8; 25,8)	24,3	2
[25,8; 28,8)	27,3	1
[28,8; 31,8)	30,3	4
		n = 12

Bảng 22

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	1
[19,8; 22,8)	21,3	2
[22,8; 25,8)	24,3	3
[25,8; 28,8)	27,3	2
[28,8; 31,8)	30,3	4
		n = 12

Bảng 23

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

a) Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Hà Nội và Huế.

b) Trong hai thành phố Hà Nội và Huế, thành phố nào có nhiệt độ không khí trung bình tháng đồng đều hơn?

Bài 12: Bảng 24 thống kê độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt và Vũng Tàu (đơn vị: %).

Tháng Độ ẩm	1	2	3	4	5	6	7	8	9	10	11	12
Đà Lạt	83	79	79	87	87	87	88	89	90	91	88	86
Vũng Tàu	75	77	78	77	79	79	81	79	81	83	80	77

Bảng 24

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

a) Hãy lần lượt ghép các số liệu của Đà Lạt, Vũng Tàu thành năm nhóm sau:

$[75; 78,3)$, $[78,3; 81,6)$, $[81,6; 84,9)$, $[84,9; 88,2)$, $[88,2; 91,5)$.

b) Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Đà Lạt và Vũng Tàu.

c) Trong hai thành phố Đà Lạt và Vũng Tàu, thành phố nào có độ ẩm không khí trung bình tháng đồng đều hơn?

ÔN TẬP CHƯƠNG III

PHẦN 1. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 1: 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:

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: $4,2 - 2,7 = 1,5(\text{km})$

b) **Chọn D**

Cỡ mẫu

$$n = 20$$

Gọi

$$x_1; x_2; \dots; x_{20}$$

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

$$\frac{1}{2}(x_5 + x_6) \in [3,0; 3,3)$$

Do đó, tứ phân vị thứ nhất của mẫu số 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

$$\frac{1}{2}(x_{15} + x_{16}) \in [3,6; 3,9)$$

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

$$\Delta_Q = Q_3 - Q_1 = 0,575$$

c) **Chọn C**

Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
Số ngày	3	6	5	4	2

Số trung bình:

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

Phương sai:

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

d) **Chọn D**

Độ lệch chuẩn:

$$\sigma = \sqrt{0,1314} \approx 0,36$$

Bài 2: Bạn Chi rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của bạn Chi được thống kê lại ở bảng sau:

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: $45 - 20 = 25$ (phút)

b) **Chọn D**

Cỡ mẫu $n = 18$

Gọi $x_1; x_2; \dots; x_{18}$ là mẫu số liệu gốc về thời gian tập nhảy mỗi ngày của bạn Chi được xếp theo thứ tự không giảm.

Ta có: $x_1; \dots; x_6 \in [20; 25); x_7; \dots; x_{12} \in [25; 30); x_{13}; \dots; x_{16} \in [30; 35); x_{17}; \dots; x_{18} \in [35; 40); x_{19} \in [40; 45)$

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

ghép nhóm là: $Q_1 = 20 + \frac{18}{6}(25 - 20) = 23,75$

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

nhóm là: $Q_3 = 30 + \frac{\frac{3.18}{4} - (6+6)}{4} (35 - 30) = 31,875$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_o = Q_3 - Q_1 = 8,125$

c) **Chọn D**

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

nhóm là: $Q_3 = 14 + \frac{\frac{3.25}{4} - (4 + 6 + 8)}{4}(16 - 14) = 14,375$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 3,63$

c) **Chọn A**

Giá trị đại diện	9	11	13	15	17
Số lần	4	6	8	4	3

Số trung bình: $\bar{x} = \frac{4.9 + 6.11 + 8.13 + 4.15 + 3.17}{25} = 12,68$

Độ lệch chuẩn: $\sigma = \sqrt{\frac{4.9^2 + 6.11^2 + 8.13^2 + 4.15^2 + 3.17^2}{25} - 12,68^2} \approx 5,98$

Bài 4: 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

Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên

Lời giải

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

Cỡ mẫu $n = 30$

Gọi $x_1; x_2; \dots; x_{30}$ là mẫu số liệu gốc về độ dài quãng đường bác tài xế lái mỗi ngày trong một tháng được xếp theo thứ tự không giảm.

Ta có: $x_1; \dots; x_5 \in [50;100); x_6; \dots; x_{15} \in [100;150); x_{16}; \dots; x_{24} \in [150;200); x_{25}; \dots; x_{28} \in [200;250); x_{29}; x_{30} \in [250;300)$

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

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

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

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

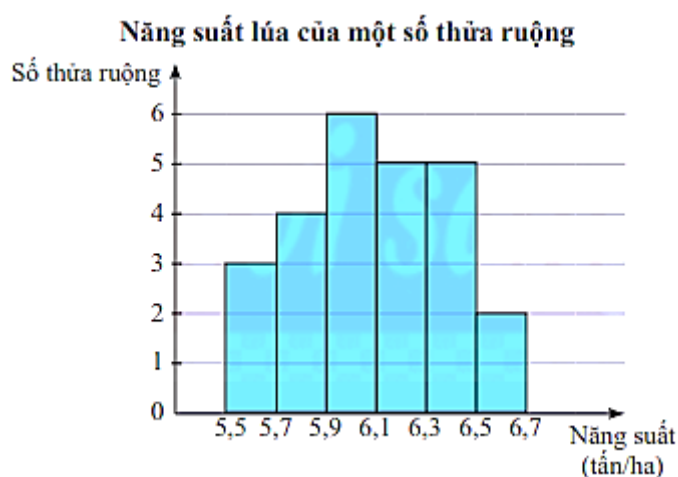
Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 79,17$

Giá trị đại diện	75	125	175	225	275
Số ngày	5	10	9	4	2

$$\text{Số trung bình: } \bar{x} = \frac{5 \cdot 75 + 10 \cdot 125 + 9 \cdot 175 + 4 \cdot 225 + 2 \cdot 275}{30} = 155$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{\frac{5 \cdot 75^2 + 10 \cdot 125^2 + 9 \cdot 175^2 + 4 \cdot 225^2 + 2 \cdot 275^2}{30} - 155^2} = 3100$$

Bài 5: Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



- Có bao nhiêu thửa ruộng đã được khảo sát?
- Lập bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên.
- Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên.

Lời giải

a) Có $3 + 4 + 6 + 5 + 5 + 2 = 25$ thửa ruộng đã được khảo sát

b)

Năng suất	[5,5;5,7)	[5,7;5,9)	[5,9;6,1)	[6,1;6,3)	[6,3;6,5)	[6,5;6,7)
Số thửa ruộng	3	4	6	5	5	2

Giá trị đại diện	5,6	5,8	6,0	6,2	6,4	6,6
Tần số tương đối	3	4	6	5	5	2

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

Cỡ mẫu $n = 25$

Gọi $x_1; x_2; \dots; x_{25}$ là mẫu số liệu gốc về năng suất của 25 thửa ruộng được xếp theo thứ tự không giảm.

Ta có: $x_1; \dots; x_3 \in [5,5;5,7); x_4; \dots; x_7 \in [5,7;5,9); x_8; \dots; x_{13} \in [5,9;6,1); x_{14}; \dots; x_{18} \in [6,1;6,3);$

$x_{19}; \dots; x_{23} \in [6,3;6,5); x_{24}; x_{25} \in [6,5;6,7)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_6; x_7) \in [5,7;5,9)$. Do đó, tứ phân vị thứ nhất của mẫu

số 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 $x_{19} \in [6,3;6,5)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm

là: $Q_3 = 6,3 + \frac{\frac{3.25}{4} - (3+4+6+5)}{5}(6,5 - 6,3) = 6,33$

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

Số trung bình: $\bar{x} = \frac{3.5,6 + 4.5,8 + 6.6,0 + 5.6,2 + 5.6,4 + 2.6,6}{25} = 6,088$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} \approx 0,29$$

Bài 6: 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:

- Nếu so sánh theo số trung bình thì học sinh trường nào viết nhanh hơn?
- Nếu so sánh theo khoảng tứ phân vị thì học sinh trường nào có tốc độ viết đồng đều hơn?
- Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

Lời giải

Giá trị đại diện	6,5	7,5	8,5	9,5	10,5
Học sinh trường X	8	10	13	10	9
Học sinh trường Y	4	12	17	14	3

a) Cỡ mẫu: $n = 50$

Xét số liệu của trường X :

$$\text{Số trung bình: } \bar{x}_X = \frac{8.6,5 + 10.7,5 + 13.8,5 + 10.9,5 + 9.10,5}{50} = 8,54$$

Xét số liệu của trường Y:

$$\text{Số trung bình: } \bar{x}_Y = \frac{4.6,5 + 12.7,5 + 17.8,5 + 14.9,5 + 3.10,5}{50} = 8,5$$

Vậy nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn

b) Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của 50 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

$$\text{ghép nhóm là: } Q_1 = 7 + \frac{\frac{50}{4} - 8}{10}(8 - 7) = 7,45$$

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 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 là: $\Delta_Q = Q_3 - Q_1 = 2,2$

Gọi $y_1; y_2; \dots; y_{50}$ là mẫu số liệu gốc về thời gian hoàn thành một bài viết chính tả của 50 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

$$\text{ghép nhóm là: } Q_1' = 7 + \frac{\frac{50}{4} - 4}{12}(8 - 7) = \frac{185}{24}$$

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 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 là: $\Delta_Q' = Q_3' - Q_1' = \frac{271}{168}$

Vậy nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn

c) Xét số liệu của trường X:

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,54^2} \approx 1,33$$

Xét số liệu của trường Y:

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{4.6,5^2 + 12.7,5^2 + 17.8,5^2 + 14.9,5^2 + 3.10,5^2}{50} - 8,5^2} \approx 1,04$$

Vậy nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn

Bài 7: Khoảng tứ phân vị nhỏ hơn thì số giờ nắng trong tháng 6 của địa phương đó đồng đều hơn Độ lệch chuẩn nhỏ hơn thì số giờ nắng trong tháng 6 của địa phương đó đồng đều hơn

Lời giải

a) Cỡ mẫu: $n = 20$

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về số giờ nắng trong tháng 6 trong 20 năm của 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 $\frac{1}{2}(x_5 + x_6) \in [220; 250)$. Do đó, tứ phân vị thứ nhất của

mẫu số 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 $\frac{1}{2}(x_{15} + x_{16}) \in [250; 280)$. Do đó, tứ phân vị thứ ba của mẫu

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 39,64$

Gọi $y_1; y_2; \dots; y_{50}$ là mẫu số liệu gốc về số giờ nắng trong tháng 6 trong 20 năm của Quy Nhơn được xếp theo thứ tự không giảm.

Ta có: $y_1; \dots; y_{10} \in [160; 190); y_{11}; y_{12} \in [190; 220); y_{13}; \dots; y_{17} \in [220; 250); y_{18}; \dots; y_{19} \in [250; 280);$

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

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(y_5 + y_6) \in [220; 250)$. Do đó, tứ phân vị thứ nhất của

$$\text{mẫu số liệu ghép nhóm là: } Q_1' = 220 + \frac{\frac{20}{4} - (1+2)}{4} (250 - 220) = 235$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{1}{2}(y_{15} + y_{16}) \in [250; 280)$. Do đó, tứ phân vị thứ ba của mẫu

$$\text{số liệu ghép nhóm là: } Q_3' = 250 + \frac{\frac{3.20}{4} - (1+2+4)}{10} (280 - 250) = 274$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q' = Q_3' - Q_1' = 39$$

Vậy nếu so sánh theo khoảng tứ phân vị thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn

b)

Xét số liệu của Nha Trang:

$$\text{Số trung bình: } \bar{x}_X = \frac{1.145 + 1.175 + 1.205 + 8.235 + 7.265 + 2.295}{20} = 242,5$$

$$\text{Độ lệch chuẩn: } \sigma_X = \sqrt{\frac{1.145^2 + 1.175^2 + 1.205^2 + 8.235^2 + 7.265^2 + 2.295^2}{20} - 242,5^2} \approx 35,34$$

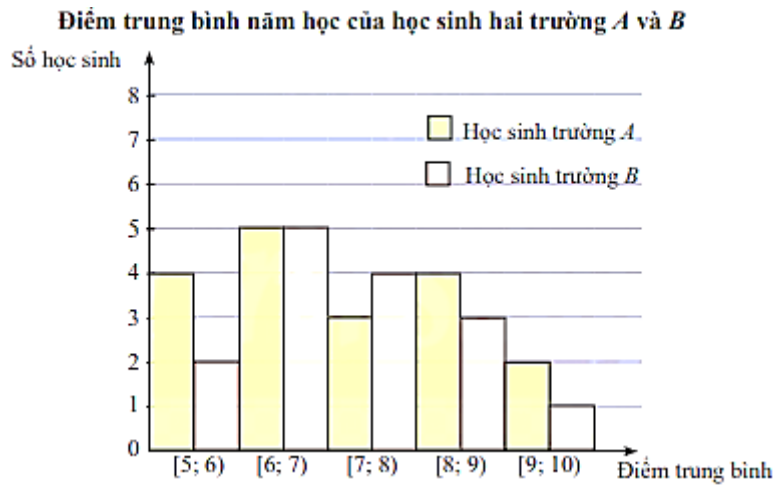
Xét số liệu của Quy Nhơn:

$$\text{Số trung bình: } \bar{x}_Y = \frac{1.175 + 2.205 + 4.235 + 10.265 + 3.295}{20} = 253$$

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{1.175^2 + 2.205^2 + 4.235^2 + 10.265^2 + 3.295^2}{20} - 253^2} \approx 30,59$$

Vậy nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn

Bài 8: Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



- a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu trên.
- b) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đồng đều hơn?
- c) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đồng đều hơn?

Lời giải

a)

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
Học sinh trường A	4	5	3	4	2
Học sinh trường B	2	5	4	3	1

b) Cỡ mẫu: $n_A = 18$

Gọi $x_1; x_2; \dots; x_{18}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh hai 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 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

$$\text{nhóm là: } Q_3 = 8 + \frac{\frac{3.18}{4} - (4 + 5 + 3)}{4} (9 - 8) = 8,375$$

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

Cỡ mẫu: $n_B = 15$

Gọi $y_1; y_2; \dots; y_{15}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh hai 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

$$\text{ghép nhóm là: } Q_1' = 6 + \frac{\frac{15}{4} - 2}{5} (7 - 6) = 6,35$$

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

$$\text{nhóm là: } Q_3' = 8 + \frac{\frac{3.15}{4} - (2 + 5 + 4)}{3} (9 - 8) = 8,08$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = 1,73$

Vậy 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

c) Xét số liệu của trường A:

$$\text{Số trung bình: } \bar{x}_X = \frac{4.5,5 + 5.6,5 + 3.7,5 + 4.8,5 + 2.9,5}{18} = 7,22$$

$$\text{Độ lệch chuẩn: } \sigma_X = \sqrt{\frac{4.5,5^2 + 5.6,5^2 + 3.7,5^2 + 4.8,5^2 + 2.9,5^2}{18} - 7,22^2} \approx 1,79$$

Xét số liệu của trường B :

$$\text{Số trung bình: } \bar{x}_Y = \frac{2.5,5 + 5.6,5 + 4.7,5 + 3.8,5 + 1.9,5}{15} = 7,23$$

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

Nhóm chứa tứ phân vị thứ ba là

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

Lời giải

Chọn C

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2}$.

Mà $x_{15}; x_{16}$ đều thuộc nhóm [17; 18). Do đó nhóm chứa tứ phân vị thứ ba là [17; 18).

Bài 4: 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

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.

Lời giải

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.

Bài 5: 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

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

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.

Bài 6: 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.

Lời giải

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 $\frac{x_{30} + x_{31}}{2}$.

Mà $x_{30}; x_{31} \in [6, 5; 7)$ nên nhóm chứa tứ phân vị thứ ba là $[6, 5; 7)$.

Ta có $Q_3 = 6,5 + \frac{\frac{3 \cdot 40}{4} - 25}{10} \cdot (7 - 6,5) = 6,75$.

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

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

$$\bar{x} = \frac{5,25 \cdot 2 + 5,75 \cdot 8 + 6,25 \cdot 15 + 6,75 \cdot 10 + 7,25 \cdot 5}{40} = 6,35.$$

Phương sai và độ lệch chuẩn là:

$$s^2 = \frac{5,25^2 \cdot 2 + 5,75^2 \cdot 8 + 6,25^2 \cdot 15 + 6,75^2 \cdot 10 + 7,25^2 \cdot 5}{40} - 6,35^2 = 0,2775.$$

$$\text{Suy ra } s = \sqrt{0,2775} \approx 0,53.$$

Bài 7: 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:

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

Lời giải

- 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:

$$\overline{x_A} = \frac{2.7,5 + 5.12,5 + 8.17,5 + 6.22,5 + 4.27,5}{25} = 18,5.$$

Trung bình tiền lãi đầu tư vào lĩnh vực B là:

$$\overline{x_B} = \frac{8.7,5 + 4.12,5 + 2.17,5 + 5.22,5 + 6.27,5}{25} = 16,9.$$

Vì $\overline{x_A} > \overline{x_B}$ nên đầu tư vào lĩnh vực A thì đem lại lãi cao hơn.

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

$$s_A^2 = \frac{2.7,5^2 + 5.12,5^2 + 8.17,5^2 + 6.22,5^2 + 4.27,5^2}{25} - 18,5^2 = 34.$$

$$\text{Suy ra } s_A = \sqrt{34} \approx 5,83.$$

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

$$s_B^2 = \frac{8.7,5^2 + 4.12,5^2 + 2.17,5^2 + 5.22,5^2 + 6.27,5^2}{25} - 16,9^2 = 64,64.$$

$$\text{Suy ra } s_B = \sqrt{64,64} \approx 8,04.$$

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.

Bài 8: 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?

Lời giải

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 là $\frac{x_5 + x_6}{2}$ mà $x_5; x_6$ thuộc nhóm $[172; 174)$.

$$\text{Ta có } Q_1 = 172 + \frac{\frac{20}{4} - 3}{10} \cdot (174 - 172) = 172,4.$$

Tứ phân vị thứ ba của mẫu số liệu là $\frac{x_{15} + x_{16}}{2}$ mà $x_{15}; x_{16}$ thuộc nhóm $[174; 176)$.

$$\text{Ta có } Q_3 = 174 + \frac{\frac{3 \cdot 20}{4} - 13}{6} \cdot (176 - 174) \approx 174,7.$$

Do đó khoảng tứ phân vị là $\Delta_Q = 174,7 - 172,4 = 2,3$.

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:

$$\bar{x} = \frac{3 \cdot 171 + 10 \cdot 173 + 6 \cdot 175 + 1 \cdot 178}{20} = 173,55.$$

Phương sai và độ lệch chuẩn

$$s^2 = \frac{3 \cdot 171^2 + 10 \cdot 173^2 + 6 \cdot 175^2 + 1 \cdot 178^2}{20} - 173,55^2 \approx 2,75.$$

$$\text{Suy ra } s = \sqrt{2,75} \approx 1,66.$$

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.

Bài 9: 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 vốn kế của bạn nào cho kết quả đo ổn định hơn.

Lời giải

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 <u>An</u> đo	1	6	2	1
Số lần Bình đo	1	3	4	2

Hiệu điện thế trung bình của An đo là:

$$\bar{x}_1 = \frac{3,875 \cdot 1 + 3,925 \cdot 6 + 3,975 \cdot 2 + 4,025 \cdot 1}{10} = 3,94.$$

Hiệu điện thế trung bình của Bình đo là:

$$\bar{x}_2 = \frac{3,875 \cdot 1 + 3,925 \cdot 3 + 3,975 \cdot 4 + 4,025 \cdot 2}{10} = 3,96.$$

Phương sai và độ lệch chuẩn về mẫu số liệu ghép nhóm của An đo là:

$$s_1^2 = \frac{3,875^2 \cdot 1 + 3,925^2 \cdot 6 + 3,975^2 \cdot 2 + 4,025^2 \cdot 1}{10} - 3,94^2 = 1,525 \cdot 10^{-3}.$$

$$\text{Suy ra } s_1 = \sqrt{1,525 \cdot 10^{-3}} \approx 0,039.$$

Phương sai và độ lệch chuẩn về mẫu số liệu ghép nhóm của Bình đo là:

$$s_2^2 = \frac{3,875^2 \cdot 1 + 3,925^2 \cdot 3 + 3,975^2 \cdot 4 + 4,025^2 \cdot 2}{10} - 3,96^2 = 2,025 \cdot 10^{-3}.$$

$$\text{Suy ra } s_2 = \sqrt{2,025 \cdot 10^{-3}} = 0,045.$$

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.

Bài 10: Cho mẫu số liệu ghép nhóm có tứ phân vị thứ nhất, thứ hai, thứ ba lần lượt là Q_1 , Q_2 , Q_3 . Khoảng tứ phân vị của mẫu số liệu ghép nhóm đó bằng:

- A. $2Q_2$.
- B. $Q_1 - Q_3$.
- C. $Q_3 - Q_1$.
- D. $Q_3 + Q_1 - Q_2$.

Lời giải

Chọn C

Khoảng tứ phân vị của mẫu số liệu ghép nhóm đó là $\Delta_Q = Q_3 - Q_1$.

Bài 11: Bảng 22, Bảng 23 lần lượt biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội và Huế (đơn vị: độ C).

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	2
[19,8; 22,8)	21,3	3
[22,8; 25,8)	24,3	2
[25,8; 28,8)	27,3	1
[28,8; 31,8)	30,3	4
		n = 12

Bảng 22

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	1
[19,8; 22,8)	21,3	2
[22,8; 25,8)	24,3	3
[25,8; 28,8)	27,3	2
[28,8; 31,8)	30,3	4
		n = 12

Bảng 23

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

- a) Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Hà Nội và Huế.
- b) Trong hai thành phố Hà Nội và Huế, thành phố nào có nhiệt độ không khí trung bình tháng đồng đều hơn?

Lời giải

a)

Hà Nội

Trong mẫu số liệu ghép nhóm ở Bảng 22, ta có: đầu mút trái của nhóm 1 là $a_1 = 16,8$; đầu mút phải của nhóm 5 là $a_6 = 31,8$.

Vậy khoảng biến thiên của mẫu số liệu ghép nhóm được cho bởi *Bảng 22* là:

$$R = a_6 - a_1 = 31,8 - 16,8 = 15 \text{ (độ C)}.$$

Từ *Bảng 22* ta có bảng thống kê sau:

Nhóm	Tần số	Tần số tích lũy
[16,8; 19,8)	2	2
[19,8; 22,8)	3	5
[22,8; 25,8)	2	7
[25,8; 28,8)	1	8
[28,8; 31,8)	4	12
	n = 12	

Số phần tử của mẫu là $n = 12$.

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ mà $2 < 3 < 5$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng

3. Xét nhóm 2 là nhóm $[19,8; 22,8)$ có $s = 19,8; h = 3; n_2 = 3$ và nhóm 1 là nhóm $[16,8; 19,8)$ có $cf_1 = 2$.

Áp dụng công thức, ta có tứ phân vị thứ nhất là:

$$Q_1 = 19,8 + \left(\frac{3-2}{3} \right) \cdot 3 = 20,8 \text{ (độ C)}.$$

Ta có: $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9 > 8$ mà $8 < 9 < 12$. Suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc

bằng 9. Xét nhóm 5 là nhóm $[28,8; 31,8)$ có $t = 28,8; l = 3; n_5 = 4$ và nhóm 4 là nhóm $[25,8; 28,8)$ có $cf_4 = 8$.

Áp dụng công thức, ta có tứ phân vị thứ ba là:

$$Q_3 = 28,8 + \left(\frac{9-8}{4} \right) \cdot 3 = 29,55 \text{ (độ C)}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm được cho bởi *Bảng 22* là:

$$\Delta_Q = Q_3 - Q_1 = 29,55 - 20,8 = 8,75 \text{ (độ C)}.$$

Số trung bình cộng của mẫu số liệu ghép nhóm được cho bởi *Bảng 22* là:

$$\bar{x} = \frac{2 \cdot 18,3 + 3 \cdot 21,3 + 2 \cdot 24,3 + 1 \cdot 27,3 + 4 \cdot 30,3}{12} = \frac{297,6}{12} = 24,8 \text{ (độ C)}.$$

Vậy phương sai của của mẫu số liệu ghép nhóm được cho bởi Bảng 22 là:

$$s^2 = \frac{1}{12} \cdot \left[2 \cdot (18,3 - 24,8)^2 + 3 \cdot (21,3 - 24,8)^2 + 2 \cdot (24,3 - 24,8)^2 \right. \\ \left. + 1 \cdot (27,3 - 24,8)^2 + 4 \cdot (30,3 - 24,8)^2 \right] = \frac{249}{12} = 20,75.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s = \sqrt{20,75} \approx 4,56$ (độ C).

Huế

Trong mẫu số liệu ghép nhóm ở Bảng 23, ta có: đầu mút trái của nhóm 1 là $a_1 = 16,8$; đầu mút phải của nhóm 5 là $a_6 = 31,8$.

Vậy khoảng biến thiên của mẫu số liệu ghép nhóm được cho bởi Bảng 23 là:

$$R' = a_6 - a_1 = 31,8 - 16,8 = 15 \text{ (đ} \spadesuit \text{ C)}.$$

Từ Bảng 23 ta có bảng thống kê sau:

Nhóm	Tần số	Tần số tích lũy
[16,8; 19,8)	1	1
[19,8; 22,8)	2	3
[22,8; 25,8)	3	6
[25,8; 28,8)	2	8
[28,8; 31,8)	4	12
	n = 12	

Số phần tử của mẫu là $n = 12$.

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ mà $1 < 3$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 2 là nhóm $[19,8; 22,8)$ có $s = 19,8; h = 3; n_2 = 2$ và nhóm 1 là nhóm $[16,8; 19,8)$ có $cf_1 = 1$.

Áp dụng công thức, ta có tứ phân vị thứ nhất là:

$$Q_1' = 19,8 + \left(\frac{3-1}{2} \right) \cdot 3 = 22,8 \text{ (đ} \spadesuit \text{ C)}.$$

Ta có: $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$ mà $8 < 9 < 12$. Suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 5 là nhóm $[28,8; 31,8)$ có $t = 28,8; I = 3; n_5 = 4$ và nhóm 4 là nhóm $[25,8; 28,8)$ có $cf_4 = 8$.

Áp dụng công thức, ta có tứ phân vị thứ ba là:

$$Q'_3 = 28,8 + \left(\frac{9-8}{4} \right) \cdot 3 = 29,55 \text{ (đ}\blacklozenge\text{ C)}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm được cho bởi Bảng 23 là:

$$\Delta'_Q = Q'_3 - Q'_1 = 29,55 - 22,8 = 6,75 \text{ (đ}\blacklozenge\text{ C)}.$$

Số trung bình cộng của mẫu số liệu ghép nhóm được cho bởi Bảng 23 là:

$$\overline{x'} = \frac{1 \cdot 18,3 + 2 \cdot 21,3 + 3 \cdot 24,3 + 2 \cdot 27,3 + 4 \cdot 30,3}{12} = \frac{309,6}{12} = 25,8 \text{ (đ}\blacklozenge\text{ C)}.$$

Vậy phương sai của của mẫu số liệu ghép nhóm được cho bởi Bảng 23 là:

$$s'^2 = \frac{1}{12} \cdot [1 \cdot (18,3 - 25,8)^2 + 2 \cdot (21,3 - 25,8)^2 + 3 \cdot (24,3 - 25,8)^2 + 2 \cdot (27,3 - 25,8)^2 + 4 \cdot (30,3 - 25,8)^2] = \frac{189}{12} = 15,75.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s' = \sqrt{15,75} \approx 3,97$ (độ C).

b) Vì $s' \approx 3,97 < s \approx 4,56$ nên thành phố Huế có nhiệt độ không khí trung bình tháng đồng đều hơn thành phố Hà Nội.

Bài 12: Bảng 24 thống kê độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt và Vũng Tàu (đơn vị: %).

Tháng Độ ẩm	1	2	3	4	5	6	7	8	9	10	11	12
Đà Lạt	83	79	79	87	87	87	88	89	90	91	88	86
Vũng Tàu	75	77	78	77	79	79	81	79	81	83	80	77

Bảng 24

(Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022)

a) Hãy lần lượt ghép các số liệu của Đà Lạt, Vũng Tàu thành năm nhóm sau:

$[75; 78,3)$, $[78,3; 81,6)$, $[81,6; 84,9)$, $[84,9; 88,2)$, $[88,2; 91,5)$.

b) Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Đà Lạt và Vũng Tàu.

c) Trong hai thành phố Đà Lạt và Vũng Tàu, thành phố nào có độ ẩm không khí trung bình tháng đồng đều hơn?

Lời giải

a) Từ *Bảng 24*, ta có các bảng thống kê sau:

Nhóm	Giá trị đại diện	Tần số	Nhóm	Giá trị đại diện	Tần số
[75; 78,3)	76,65	0	[75; 78,3)	76,65	5
[78,3; 81,6)	79,95	2	[78,3; 81,6)	79,95	6
[81,6; 84,9)	83,25	1	[81,6; 84,9)	83,25	1
[84,9; 88,2)	86,55	7	[84,9; 88,2)	86,55	0
[88,2; 91,5)	89,85	2	[88,2; 91,5)	89,85	0
		n = 12			n = 12

Đà Lạt

Vũng Tàu

b)

Đà Lạt

Khoảng biến thiên của mẫu số liệu ghép nhóm của Đà Lạt là:

$R = 91,5 - 78,3 = 13,2 \text{ (\%)}$.

Từ bảng thống kê trên, ta có bảng thống kê của mẫu số liệu ghép nhóm của Đà Lạt:

Nhóm	Tần số	Tần số tích lũy
[78,3; 81,6)	2	2
[81,6; 84,9)	1	3
[84,9; 88,2)	7	10
[88,2; 91,5)	2	12
	n = 12	

Số phần tử của mẫu là $n = 12$.

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ mà $2 < 3$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 2 là nhóm $[81,6;84,9)$ có $s = 81,6; h = 3,3; n_2 = 1$ và nhóm 1 là nhóm $[78,3;81,6)$ có $cf_1 = 2$.

Áp dụng công thức, ta có tứ phân vị thứ nhất là:

$$Q_1 = 81,6 + \left(\frac{3-2}{1} \right) \cdot 3,3 = 84,9(\%).$$

Ta có: $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$ mà $3 < 9 < 10$. Suy ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 3 là nhóm $[84,9;88,2)$ có $t = 84,9; 1 = 3,3; n_3 = 7$ và nhóm 2 là nhóm $[81,6; 84,9)$ có $cf_2 = 3$.

Áp dụng công thức, ta có tứ phân vị thứ ba là:

$$Q_3 = 84,9 + \left(\frac{9-3}{7} \right) \cdot 3,3 \approx 87,7(\%).$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm của Đà Lạt là:

$$\Delta_Q = Q_3 - Q_1 = 87,7 - 84,9 = 2,8(\%).$$

Số trung bình cộng của mẫu số liệu ghép nhóm của Đà Lạt là:

$$\bar{x} = \frac{2 \cdot 79,95 + 1 \cdot 83,25 + 7 \cdot 86,55 + 2 \cdot 89,85}{12} = \frac{1028,7}{12} = 85,725(\%).$$

Vậy phương sai của của mẫu số liệu ghép nhóm của Đà Lạt là:

$$s^2 = \frac{1}{12} \cdot$$

$$\left[2 \cdot (79,95 - 85,725)^2 + 1 \cdot (83,25 - 85,725)^2 + 7 \cdot (86,55 - 85,725)^2 + 2 \cdot (89,85 - 85,725)^2 \right]$$

$$= \frac{111,6225}{12} \approx 9,3.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm của Đà Lạt là: $s \approx \sqrt{9,3} \approx 3,05 (\%).$

Vũng Tàu

Khoảng biến thiên của mẫu số liệu ghép nhóm của Vũng Tàu là:

$$R' = 84,9 - 75 = 9,9 (\%).$$

Từ bảng thống kê trên, ta có bảng thống kê của mẫu số liệu ghép nhóm của Vũng Tàu:

Nhóm	Tần số	Tần số tích lũy
[75; 78,3)	5	5
[78,3; 81,6)	6	11
[81,6; 84,9)	1	12
	n = 12	

Số phần tử của mẫu là $n = 12$.

Ta có: $\frac{n}{4} = \frac{12}{4} = 3$ mà $5 > 3$. Suy ra nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3.

Xét nhóm 1 là nhóm [75; 78,3) có $s = 75; h = 3,3; n_1 = 5$.

Áp dụng công thức, ta có tứ phân vị thứ nhất là:

$$Q_1' = 75 + \frac{3}{5} \cdot 3,3 = 76,98(\%).$$

Ta có: $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$ mà $5 < 9 < 11$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 2 là nhóm [78,3; 81,6) có $t = 78,3; 1 = 3,3; n_2 = 6$ và nhóm 1 là nhóm [75; 78,3) có $cf_1 = 5$.

Áp dụng công thức, ta có tứ phân vị thứ ba là:

$$Q_3' = 78,3 + \left(\frac{9-5}{6} \right) \cdot 3,3 = 80,5(\%).$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm của Vũng Tàu là:

$$\Delta_Q' = Q_3' - Q_1' = 80,5 - 76,98 = 3,52(\%).$$

Số trung bình cộng của mẫu số liệu ghép nhóm của Vũng Tàu là:

$$\overline{x'} = \frac{5 \cdot 76,65 + 6 \cdot 79,95 + 1 \cdot 83,25}{12} = \frac{946,2}{12} = 78,85(\%).$$

Vậy phương sai của của mẫu số liệu ghép nhóm của Vũng Tàu là:

$$s'^2 = \frac{1}{12} \cdot \left[5 \cdot (76,65 - 78,85)^2 + 6 \cdot (79,95 - 78,85)^2 + 1 \cdot (83,25 - 78,85)^2 \right]$$

$$= \frac{50,82}{12} = 4,235.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm của Vũng Tàu là: $s' = \sqrt{4,235} \approx 2,06$ (%).

c) Vì $s' \approx 2,06 < s \approx 3,05$ nên thành phố Vũng Tàu có độ ẩm không khí trung bình tháng đồng đều hơn thành phố Đà Lạt.