

CHƯƠNG 3

CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM

BÀI 1,2

KHOẢNG BIẾN THIÊN VÀ KHOẢNG TƯ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHÉP NHÓM

I. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM

1. Khoảng biến thiên của mẫu số liệu ghép nhóm

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 số liệu ghép nhóm trên là $R = a_{k+1} - a_1$

Ý nghĩa

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

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

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

2. Khoảng tứ phân vị của mẫu số liệu ghép nhóm

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

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

Chú ý:

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

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

Ý nghĩa

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

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

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

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 (còn gọi giá trị bất thường) nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$

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

3. Phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

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

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

Trong đó:

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

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + \dots + m_kx_k}{n} \text{ là số trung bình của mẫu số liệu ghép nhóm.}$$

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

Chú ý:

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

$$S^2 = \frac{1}{n} [m_1x_1^2 + m_2x_2^2 + \dots + m_kx_k^2] - \bar{x}^2$$

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

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

Ý nghĩa

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

Chú ý: Sử dụng phương sai, độ lệch chuẩn đo độ rủi ro:

Trong tài chính, người ta có nhiều cách để đo độ rủi ro của một phương án đầu tư. Một trong các cách đó là sử dụng độ lệch chuẩn của lợi nhuận thu được theo phương án đầu tư. Độ lệch chuẩn càng lớn thì phương án đầu tư càng rủi ro.

II. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CHO MẪU SỐ LIỆU GHEP NHÓM

(Nhắc lại kiến thức lớp 11)

1. Số trung bình của mẫu số liệu ghép nhóm

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

Số trung bình của mẫu số liệu ghép nhóm, kí hiệu $\bar{x}$, được tính như sau:

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + \dots + m_kx_k}{n}$$

Trong đó:

- $n = m_1 + m_2 + \dots + m_k$ là cỡ mẫu
- $x_i = \frac{a_i + a_{i+1}}{2}$ (với $i = 1, 2, \dots, k$) là giá trị đại diện của nhóm $[a_i; a_{i+1})$

Chú ý: Đối với số liệu rời rạc, người ta thường cho các nhóm dưới dạng $k_1 - k_2$ trong đó $k_1, k_2 \in \mathbb{N}$.

Nhóm $k_1 - k_2$ được hiểu là nhóm gồm các giá trị $k_1, k_1 + 1, \dots, k_2$. Khi đó, ta cần hiệu chỉnh mẫu số liệu ghép nhóm đưa về dạng bảng sau, trước khi thực hiện tính toán các số đặc trưng bằng cách hiệu chỉnh nhóm $k_1 - k_2$ với $k_1, k_2 \in \mathbb{N}$ thành nhóm $[k_1 - 0,5; k_2 + 0,5)$.

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

Ý nghĩa: Số trung bình của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho số trung bình của mẫu gốc, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

2. Trung vị của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tính trung vị của mẫu số liệu ghép nhóm, ta làm các bước sau:

Bước 1: Xác định nhóm chứa trung vị. Giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$

Bước 2: Trung vị là: $M_e = a_p + \frac{\frac{n}{2} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

Ý nghĩa: Trung vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho mẫu số liệu gốc, nó chia mẫu số liệu thành hai phần, mỗi phần chứa 50% giá trị.

3. Tứ phân vị của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tìm **tứ phân vị thứ nhất** Q_1 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_1 , giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó:

$$Q_1 = a_p + \frac{\frac{n}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

Để tìm **tứ phân vị thứ ba** Q_3 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_3 , giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó:

$$Q_3 = a_p + \frac{\frac{3n}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

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

Nhận xét: Ta có thể xác định nhóm chứa tứ phân vị thứ r nhờ tính chất: có khoảng $\left(\frac{r.n}{4}\right)$ giá trị nhỏ hơn tứ phân vị này.

Ý nghĩa: Các tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho các tứ phân vị của mẫu số liệu gốc, chúng chia mẫu số liệu thành 4 phần, mỗi phần chứa 25% giá trị.

4. Một của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tìm **mốt** của mẫu số liệu ghép nhóm, ta làm các bước sau:

Bước 1: Xác định nhóm có chứa tần số lớn nhất (gọi là nhóm chứa mốt), giả sử là nhóm $j: [a_j; a_{j+1})$

Bước 2: Mốt được xác định là: $M_o = a_j + \frac{m_j - m_{j-1}}{(m_j - m_{j-1}) + (m_j - m_{j+1})} \cdot h$

Trong đó:

- m_j là tần số của nhóm j (quy ước $m_0 = m_{k+1} = 0$).
- h là độ dài của nhóm.

Chú ý:

- Người ta chỉ định nghĩa mốt cho mẫu ghép nhóm có độ dài các nhóm bằng nhau. Một mẫu có thể không có mốt hoặc có nhiều hơn một mốt
- Khi tần số của các nhóm số liệu bằng nhau thì mẫu số liệu ghép nhóm không có mốt.

Ý nghĩa: Mốt của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho mốt của mẫu gốc, nó được dùng để đo xu thế trung tâm của mẫu số liệu.

III. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU KHÔNG GHEP NHÓM (Nhắc lại kiến thức lớp 10)

1. Số trung bình

- Giả sử ta có một mẫu số liệu là $x_1 + x_2 + \dots + x_n$.

Số trung bình (số trung bình cộng) của mẫu số liệu $x_1 + x_2 + \dots + x_n$, kí hiệu là $\bar{x}$, được tính bằng công thức: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

Ví dụ 1: Điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II được ghi ở bảng sau:

Nhóm I	10	5	7	9	8	6
Nhóm II	9	9	8	7	6	8

Nhóm nào có thành tích điểm tốt hơn?

Lời giải

Điểm trung bình của nhóm I là: $\frac{10+5+7+9+8+6}{6} = 7,5$

Điểm trung bình của nhóm II là: $\frac{9+9+8+7+6+8}{6} \approx 7,83$.

Vậy nhóm II có thành tích điểm tốt hơn nhóm I.

- Giả sử mẫu số liệu được cho dưới dạng bảng tần số

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Khi đó, công thức tính số trung bình trở thành: $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n}$

Trong đó $n = n_1 + n_2 + \dots + n_k$. Ta gọi n là **cỡ mẫu**.

Chú ý: Nếu kí hiệu $f_k = \frac{n_k}{n}$ là tần số tương đối (hay còn gọi là tần suất) của x_k trong mẫu số liệu thì số

trung bình còn có thể biểu diễn là: $\bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k$

Ví dụ 2: Kết quả điều tra về số con của một số hộ gia đình trong một tổ dân phố được ghi lại trong bảng sau:

Số con	0	1	3	4	5
Số hộ gia đình	4	4	8	3	1

Hỏi trung bình mỗi hộ gia đình trong tổ dân phố có bao nhiêu con?

Lời giải

Tổng số hộ gia đình là: $n = 4 + 4 + 8 + 3 + 1 = 20$ (hộ gia đình).

Trung bình mỗi hộ gia đình trong tổ dân phố có số con là: $\frac{0.4 + 1.4 + 3.8 + 4.3 + 5.1}{20} = 1,65$ (con)

Ý nghĩa của số trung bình

Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đo xu thế trung tâm của mẫu đó.

Chẳng hạn, ở trong ví dụ 1 trên, ta thấy điểm số trung bình của nhóm II cao hơn nhóm I ($7,83 > 7,5$), ta có thể nói rằng thành tích thực hành của nhóm II tốt hơn nhóm I.

2. Trung vị và tứ phân vị

a. Trung vị

- **Định nghĩa và cách tính số trung vị**

Khi các số liệu trong mẫu số liệu chênh lệch nhau quá lớn, ta dùng một đặc trưng khác của mẫu số liệu, gọi là **trung vị** để so sánh các mẫu số liệu với nhau.

Trung vị được định nghĩa như sau:

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

Trung vị của mẫu, kí hiệu là M_e , là giá trị ở chính giữa dãy $x_1, x_2, \dots, x_n$. Cụ thể:

- Nếu $n = 2k + 1, k \in \mathbb{N}$ (tức n là số tự nhiên lẻ), thì trung vị của mẫu $M_e = x_k + 1$.

- Nếu $n = 2k, k \in \mathbb{N}$, (tức n là số tự nhiên chẵn), thì trung vị của mẫu $M_e = \frac{x_k + x_{k+1}}{2}$.

Ví dụ 3: Điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II được ghi ở bảng sau:

Nhóm I	10	5	7	9	8	6
Nhóm II	9	9	8	7	6	8

Tính các trung vị của điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II.

Lời giải

+ Sắp xếp điểm số của mỗi bạn trong nhóm I theo thứ tự không giảm, ta được: 5; 6; 7; 8; 9; 10

Vì cỡ mẫu bằng 6 nên trung vị của nhóm I là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là $M_e = \frac{7+8}{2} = 7,5$.

+ Sắp xếp điểm số của mỗi bạn trong nhóm II theo thứ tự không giảm, ta được: 6; 7; 8; 8; 9; 9

Vì cỡ mẫu bằng 6 nên trung vị của nhóm II là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là $M_e = \frac{8+8}{2} = 8$.

• Ý nghĩa của số trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có ít nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

Ví dụ 4: Bảng sau thống kê số sách mỗi bạn học sinh Tổ 1 và Tổ 2 đã đọc ở thư viện trường trong một tháng:

Tổ 1	3	1	2	1	2	2	3	25	1
Tổ 2	4	5	4	3	3	4	5	4	

a) Trung bình mỗi bạn Tổ 1 và mỗi bạn Tổ 2 đọc bao nhiêu quyển sách ở thư viện trường trong tháng đó?

b) Em hãy thảo luận với các bạn trong nhóm xem tổ nào chăm đọc sách ở thư viện hơn.

Lời giải

a) Trung bình mỗi bạn Tổ 1 đọc số quyển sách ở thư viện trong tháng trên là:

$$\frac{3+1+2+1+2+2+3+25+1}{9} = \frac{40}{9} \approx 4,4$$

Trung bình mỗi bạn Tổ 2 đọc số quyển sách ở thư viện trong tháng trên là:

$$\frac{4+5+4+3+3+4+5+4}{8} = \frac{32}{8} = 4$$

b) Vì $4,4 > 4$ nên theo số trung bình, các bạn Tổ 1 đọc sách chăm hơn.

Nếu dựa vào số trung bình để đánh giá xem tổ nào chăm đọc sách hơn trong bài này thì không phù hợp, do có một số liệu trong mẫu số liệu của Tổ 1 quá lớn so với các số liệu còn lại. Ta sử dụng trung vị để so sánh độ chăm học giữa hai tổ.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 1: 1; 1; 1; 2; 2; 2; 3; 3; 25

Vì cỡ mẫu $n_1 = 9$ là số lẻ, nên trung vị của mẫu số liệu Tổ 1 là $M_{e_1} = 2$.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 2: 3; 3; 4; 4; 4; 4; 5; 5

Vì cỡ mẫu $n_2 = 8$ là số chẵn, nên trung vị của mẫu số liệu Tổ 2 là $M_{e_2} = \frac{4+4}{2} = 4$.

Do đó ta có: $M_{e_2} > M_{e_1}$.

Vậy theo trung vị, các bạn Tổ 2 chăm đọc sách ở thư viện hơn Tổ 1.

b. Tứ phân vị

• Trung vị chia mẫu thành hai phần. Trong thực tế người ta cũng quan tâm đến trung vị của mỗi phần đó. Ba trung vị này được gọi là **tứ phân vị** của mẫu.

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

Tứ phân vị của một mẫu số liệu gồm ba giá trị, gọi là **tứ phân vị thứ nhất, thứ hai và thứ ba** (lần lượt kí hiệu là Q_1, Q_2, Q_3). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

- Giá trị tứ phân vị thứ hai, Q_2 , chính là số trung vị của mẫu.

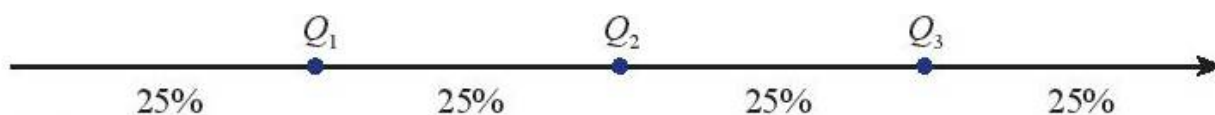
- Giá trị tứ phân vị thứ nhất, Q_1 , là trung vị của nửa số liệu đã sắp xếp bên trái Q_2 (không bao gồm Q_2 nếu n lẻ).

- Giá trị tứ phân vị thứ ba, Q_3 , là trung vị của nửa số liệu đã sắp xếp bên phải Q_2 (không bao gồm Q_2 nếu n lẻ).

• Ý nghĩa của tứ phân vị

Các điểm tứ phân vị Q_1, Q_2, Q_3 chia mẫu số liệu đã sắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần chia khoảng 25% tổng số liệu đã thu thập được.

Tứ phân vị thứ nhất Q_1 còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liệu phía dưới. Tứ phân vị thứ ba Q_3 , còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu ở phía trên.



Ví dụ 5: Tìm tứ phân vị của mẫu số liệu sau: 3; 5; 6; 13; 25; 17; 19.

Lời giải

Sắp xếp các số liệu theo thứ tự không giảm ta được: 3; 5; 6; 13; 17; 19; 25.

Vì cỡ mẫu $n = 7$, là số lẻ, nên giá trị tứ phân vị thứ hai là $Q_2 = 13$.

Tứ phân vị thứ nhất là trung vị của mẫu: 3; 5; 6. Do đó $Q_1 = 5$.

Tứ phân vị thứ ba là trung vị của mẫu: 17; 19; 25. Do đó $Q_3 = 19$.

3. Một

Cho mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu và kí hiệu là M_o .

Ý nghĩa của mốt: Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

Chú ý: Một mẫu số liệu có thể có rất nhiều mốt. Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

Ví dụ 6: Cho mẫu số liệu:

Giá trị	35	38	40	45
Tần số	10	5	6	3

Tìm mốt của mẫu số liệu trên.

Lời giải

Ta thấy giá trị 35 có tần số 10 là lớn nhất, do đó, mốt của mẫu số liệu trên là $M_o = 35$.

IV. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU KHÔNG GHÉP NHÓM (Nhắc lại kiến thức lớp 10)

1. Khoảng biến thiên và khoảng tứ phân vị

a. Khoảng biến thiên và khoảng tứ phân vị

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

• **Khoảng biến thiên** của một mẫu số liệu, kí hiệu là R , là hiệu giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó, tức là: $R = x_n - x_1$

• **Khoảng tứ phân vị**, kí hiệu là Δ_Q , là hiệu giữa Q_3 và Q_1 , tức là: $\Delta_Q = Q_3 - Q_1$

Ví dụ 1: Hãy tính khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu: 10; 3; 5; 7; 20; 1; 4; 9.

Lời giải

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: 1; 3; 4; 5; 7; 9; 10; 20.

- Khoảng biến thiên của mẫu số liệu là $R = 20 - 1 = 19$.
- Cỡ mẫu là $n = 8$, là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = 6$.
- Tứ phân vị thứ nhất là trung vị của mẫu: 10; 3; 5; 7. Do đó $Q_1 = 4$.
- Tứ phân vị thứ 3 là trung vị của mẫu: 7; 9; 10; 20. Do đó $Q_3 = 9,5$.
- Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 9,5 - 4 = 5,5$.

b. Ý nghĩa của khoảng biến thiên và khoảng tứ phân vị

- Khoảng biến thiên đặc trưng cho độ phân tán của toàn bộ mẫu số liệu.
- Khoảng tứ phân vị đặc trưng cho độ phân tán của một nửa các số liệu, có giá trị thuộc đoạn từ Q_1 đến Q_3 trong mẫu.
- Khoảng tứ phân vị không bị ảnh hưởng bởi các giá trị rất lớn hoặc rất bé trong mẫu.

Ví dụ 2: Dưới đây là bảng số liệu thống kê của Biểu đồ nhiệt độ trung bình các tháng trong năm 2019 của hai tỉnh Lai Châu và Lâm Đồng .

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Lai Châu	14,8	18,8	20,3	23,5	24,7	24,2	23,6	24,6	22,7	21,0	18,6	14,2
Lâm Đồng	16,3	17,4	18,7	19,8	20,2	20,3	19,5	19,3	18,6	18,5	17,5	16,0

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của nhiệt độ trung bình mỗi tháng của tỉnh Lai Châu và Lâm Đồng.

b) Hãy cho biết trong một năm, nhiệt độ ở địa phương nào ít thay đổi hơn.

Lời giải

a)

* Tỉnh Lai Châu:

Sắp xếp các số liệu theo thứ tự không giảm, ta được:

14,2; 14,8; 18,6; 18,8; 20,3; 21,0; 22,7; 23,5; 23,6; 24,2; 24,6; 24,7.

+ Khoảng biến thiên của mẫu số liệu là: $R = 24,7 - 14,2 = 10,5$.

+ Cỡ mẫu là $n = 12$ là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = \frac{21,0 + 22,7}{2} = 21,85$

+ Tứ phân vị thứ nhất là trung vị của mẫu: 14,2; 14,8; 18,6; 18,8; 20,3; 21,0.

Do đó $Q_1 = \frac{18,6 + 18,8}{2} = 18,7$.

+ Tứ phân vị thứ ba là trung vị của mẫu: 22,7; 23,5; 23,6; 24,2; 24,6; 24,7.

Do đó $Q_3 = \frac{23,6 + 24,2}{2} = 23,9$.

+ Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 23,9 - 18,7 = 5,2$

* Tỉnh Lâm Đồng:

Sắp xếp các số liệu theo thứ tự không giảm, ta được:

16,0; 16,3; 17,4; 17,5; 18,5; 18,6; 18,7; 19,3; 19,5; 19,8; 20,2; 20,3.

+ Khoảng biến thiên của mẫu số liệu là: $R' = 20,3 - 16,0 = 4,3$.

+ Cỡ mẫu là $n = 12$ là số chẵn nên giá trị tứ phân vị thứ hai là: $Q'_2 = \frac{18,6 + 18,7}{2} = 18,65$

+ Tứ phân vị thứ nhất là trung vị của mẫu: 16,0; 16,3; 17,4; 17,5; 18,5; 18,6.

$$\text{Do đó } Q'_1 = \frac{17,4+17,5}{2} = 17,45.$$

+ Tứ phân vị thứ ba là trung vị của mẫu: 18,7; 19,3; 19,5; 19,8; 20,2; 20,3.

$$\text{Do đó } Q'_3 = \frac{19,5+19,8}{2} = 19,65.$$

+ Khoảng tứ phân vị của mẫu là: $\Delta'_Q = Q'_3 - Q'_1 = 19,65 - 17,45 = 2,2$.

b) Xét về cả khoảng biến thiên và khoảng tứ phân vị của nhiệt độ trung bình mỗi tháng của cả hai tỉnh, ta thấy: $10,5 > 4,3$ hay $R > R'$ và $5,2 > 2,2$ hay $\Delta_Q > \Delta'_Q$.

Điều đó có nghĩa là trong một năm, nhiệt độ ở Lâm Đồng ít thay đổi hơn.

c. Giá trị ngoại lệ

Khoảng tứ phân vị được dùng để xác định các **giá trị ngoại lệ** trong mẫu, đó là các giá trị quá nhỏ hay quá lớn so với đa số các giá trị của mẫu. Cụ thể, phần tử x trong mẫu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.

Sự xuất hiện của các giá trị ngoại lệ làm cho số trung bình và phạm vi của mẫu thay đổi lớn. Do đó, khi mẫu có giá trị ngoại lệ, người ta thường sử dụng trung vị và khoảng tứ phân vị để đo mức độ tập trung và mức độ phân tán của đa số các phần tử trong mẫu số liệu.

Ví dụ 3: Hãy tính giá trị ngoại lệ (nếu có) của mẫu số liệu: 10; 3; 5; 7; 20; 1; 4; 9.

Lời giải

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: 1; 3; 4; 5; 7; 9; 10; 20.

Khoảng biến thiên của mẫu số liệu là $R = 20 - 1 = 19$.

Cỡ mẫu là $n = 8$, là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = 6$.

Tứ phân vị thứ nhất là trung vị của mẫu: 10; 3; 5; 7. Do đó $Q_1 = 4$.

Tứ phân vị thứ 3 là trung vị của mẫu: 7; 9; 10; 20. Do đó $Q_3 = 9,5$.

Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 9,5 - 4 = 5,5$.

$$Q_1 - 1,5\Delta_Q = 4 - 1,5 \cdot 5,5 = -4,25$$

$$Q_3 + 1,5\Delta_Q = 9,5 + 1,5 \cdot 5,5 = 17,75$$

Do đó, mẫu có một giá trị ngoại lệ là 20.

2. Phương sai và độ lệch chuẩn

a. Công thức tính phương sai và độ lệch chuẩn

• Giả sử ta có một mẫu số liệu là $x_1, x_2, \dots, x_n$.

Phương sai của mẫu số liệu này, kí hiệu là S^2 , được tính bởi công thức:

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

trong đó $\bar{x}$ là số trung bình của mẫu số liệu.

Căn bậc hai của phương sai được gọi là **độ lệch chuẩn**, kí hiệu là S .

Chú ý:

+ Có thể biến đổi công thức tính phương sai ở trên thành: $S^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$

+ Trong thống kê, người ta cũng quan tâm đến **phương sai hiệu chỉnh**, kí hiệu là $\hat{S}^2$, được tính bởi công thức: $\hat{S}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$

Ví dụ 4: Tính phương sai và độ lệch chuẩn của mẫu số liệu sau: 8; 10; 9; 7; 6; 10; 6; 7; 8; 9.

Lời giải

Cỡ mẫu $n = 10$.

Số trung bình: $\bar{x} = \frac{8+10+9+7+6+7+8+9}{10} = 8$

Phương sai mẫu số liệu là: $S^2 = \frac{8^2+10^2+9^2+7^2+6^2+7^2+8^2+9^2}{10} - 8^2 = 2$

Độ lệch chuẩn mẫu số liệu là $S = \sqrt{S^2} = \sqrt{2} \approx 1,41$.

• Giả sử mẫu số liệu được cho dưới dạng bảng tần số:

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Khi đó, công thức tính phương sai trở thành: $S^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{n}$

trong đó $n = n_1 + n_2 + \dots + n_k$.

Có thể biến đổi công thức tính phương sai trên thành: $S^2 = \frac{n_1x_1^2 + n_2x_2^2 + \dots + n_kx_k^2}{n} - \bar{x}^2$

Ví dụ 5: Điều tra số con của mỗi hộ gia đình trong tổ dân cư xóm 2, kết quả được ghi lại ở bảng sau:

Số con	0	1	2	3	4
Số hộ gia đình	4	4	8	3	1

Tính phương sai và độ lệch chuẩn của mẫu số liệu.

Lời giải

Tổng số hộ gia đình là: $n = 4 + 4 + 8 + 3 + 1 = 20$ (hộ gia đình).

Số trung bình của mẫu số liệu trên là: $\bar{x} = \frac{4.0+4.1+8.2+3.3+1.4}{20} = 1,65$

Phương sai của mẫu số liệu trên là: $S^2 = \frac{4.0^2+4.1^2+8.2^2+3.3^2+1.4^2}{20} - 1,65^2 = 1,2275$

Độ lệch chuẩn của mẫu số liệu trên là: $S = \sqrt{S^2} = \sqrt{1,2275} \approx 1,11$

b. Ý nghĩa của phương sai và độ lệch chuẩn

- Phương sai là trung bình cộng của các bình phương độ lệch từ mỗi giá trị của mẫu số liệu đến số trung bình.
- Phương sai và độ lệch chuẩn được dùng để đo mức độ phân tán của các số liệu trong mẫu quanh số trung bình. Phương sai và độ lệch chuẩn càng lớn thì các giá trị của mẫu càng cách xa nhau (có độ phân tán lớn).

Ví dụ 6: Bảng dưới đây thống kê tổng số giờ nắng trong năm 2019 theo từng tháng được đo bởi hai trạm quan sát khí tượng đặt ở Tuyên Quang và Cà Mau.

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Tuyên Quang	25	89	72	117	106	177	156	203	227	146	117	145
Cà Mau	180	223	257	245	191	111	141	134	130	122	157	173

(Nguồn: Tổng cục Thống kê)

- Hãy tính phương sai và độ lệch chuẩn của dữ liệu từng tỉnh.
- Nêu nhận xét về sự thay đổi tổng số giờ nắng theo từng tháng ở mỗi tỉnh.

Lời giải

a)* Tỉnh Tuyên Quang:

$$+ \text{Số trung bình: } \bar{x}_1 = \frac{25 + 89 + 72 + 117 + 106 + 177 + 156 + 203 + 227 + 146 + 117 + 145}{12} = 131,67$$

+ Phương sai mẫu số liệu ở tỉnh Tuyên Quang là:

$$S_1^2 = \frac{25^2 + 89^2 + 72^2 + 117^2 + 106^2 + 177^2 + 156^2 + 203^2 + 227^2 + 146^2 + 117^2 + 145^2}{12} - 131,67^2 = 2920,34$$

$$+ \text{Độ lệch chuẩn mẫu số liệu ở tỉnh Tuyên Quang là: } S_1 = \sqrt{S_1^2} = \sqrt{2920,34} \approx 54,04$$

* Tỉnh Cà Mau:

$$+ \text{Số trung bình: } \bar{x}_2 = \frac{180 + 223 + 257 + 245 + 191 + 111 + 141 + 134 + 130 + 122 + 157 + 173}{12} = 172$$

+ Phương sai mẫu số liệu ở tỉnh Cà Mau là:

$$S_2^2 = \frac{180^2 + 223^2 + 257^2 + 245^2 + 191^2 + 111^2 + 141^2 + 134^2 + 130^2 + 122^2 + 157^2 + 173^2}{12} - 172^2 = 2183$$

$$+ \text{Độ lệch chuẩn mẫu số liệu ở tỉnh Cà Mau là: } S_2 = \sqrt{S_2^2} = \sqrt{2183} \approx 46,72$$

b) Phương sai mẫu và độ lệch chuẩn mẫu số liệu ở tỉnh Tuyên Quang cao hơn tỉnh Cà Mau nên tổng giờ nắng trong năm 2019 theo từng tháng ở tỉnh Tuyên Quang có độ phân tán cao hơn ở tỉnh Cà Mau. Do đó, sự thay đổi tổng số giờ nắng theo từng tháng ở tỉnh Cà Mau ổn định (có ít sự thay đổi) hơn so với tỉnh Tuyên Quang.

PHẦN A
TỰ LUẬN PHÂN DẠNG TOÁN

DẠNG 1
TÌM KHOẢNG BIẾN THIÊN CHO MẪU SỐ LIỆU GHÉP NHÓM

Bài toán: Tính khoảng biến thiên của mẫu số liệu ghép nhóm sau:

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.

Phương pháp:

Xác định a_1 là giá trị đầu mút trái của nhóm đầu tiên và a_{k+1} là giá trị đầu mút phải của nhóm cuối cùng có chứa dữ liệu (tần số khác 0).

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

Bài 1. Bảng dưới thống kê thành tích nhảy xa của 58 học sinh lớp 12A.

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

a) Tìm khoảng biến thiên thành tích nhảy xa của học sinh lớp 12A.

b) Bạn Khánh Linh ý kiến cho rằng: “Trong 58 học sinh lớp 12A trên, hiệu số thành tích nhảy xa của hai học sinh bất kì không vượt quá 160 cm”. Ý kiến đó đúng hay sai?

Bài 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

a) Tìm khoảng biến thiên của mẫu số liệu đã cho.

b) Kết quả câu a) cho biết điều gì?

Bài 3. Bảng sau thống kê cân nặng của 50 quả xoài được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường

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

Có ý kiến cho rằng: “Trong 50 quả xoài trên, hiệu số cân nặng của hai quả bất kì không vượt quá 200 g”.

Ý kiến đó đúng hay sai? Giải thích.

Bài 4. Thời gian hoàn thành bài kiểm tra môn Toán của các bạn trong lớp 12A 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 cho mẫu số liệu ghép nhóm trên.

b) Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 25 phút và muộn nhất mất 44 phút thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

Bài 5. Để chuẩn bị mở một cơ sở thể dục thể thao, Bác Minh đã tiến hành điều tra tuổi thọ của máy chạy bộ do hai hãng A, B sản xuất. Bảng dưới biểu thị hai mẫu số liệu mà Bác Minh thu thập được qua Internet, với tuổi thọ của máy chạy bộ tính theo đơn vị năm.

Tuổi thọ (năm)	[2;4)	[4;6)	[6;8)	[8;10)	[10;12)
Số máy của hãng A	7	20	36	20	17
Số máy của hãng B	0	20	35	35	10

a) Tính khoảng biến thiên của mẫu số liệu hai hãng A, B .

b) Hỏi máy chạy bộ do hãng nào sản xuất có tuổi thọ phân tán hơn?

Bài 6. Thống kê thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12A và 12B được kết quả như bảng sau:

Thời gian sử dụng (phút)	[0;10)	[10;30)	[30;60)	[60;90)
Số học sinh lớp 12A	10	15	15	5
Số học sinh lớp 12B	25	5	15	0

Tìm khoảng biến thiên cho thời gian sử dụng mạng xã hội của học sinh mỗi lớp và giải thích ý nghĩa.

Bài 7. 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:

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?

DẠNG 2

TÌM TỨ PHÂN VỊ CHO MẪU SỐ LIỆU GHÉP NHÓM

Bài toán: Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi bảng sau:

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

Phương pháp:

Tìm tứ phân vị Q_1 và Q_3 theo công thức:

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

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

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

Bài 1. Kết quả điều tra tổng thu nhập trong năm 2024 của một số hộ gia đình ở thành phố Nha Trang được ghi lại ở bảng sau:

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

a) Hãy tìm các tứ phân vị Q_1 và Q_3 .

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

Bài 2. 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 3. Bạn Trang thống kê chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau:

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

- a) Sử dụng khoảng biến thiên, hãy cho biết chiều cao của học sinh nữ lớp nào có độ phân tán lớn hơn.
 b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của học sinh nữ lớp 12C và 12D.

Bài 4. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

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

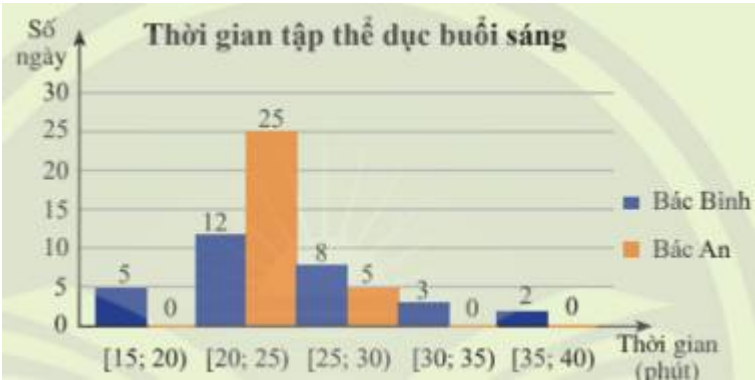
- a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.
 b) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?

Bài 5. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

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

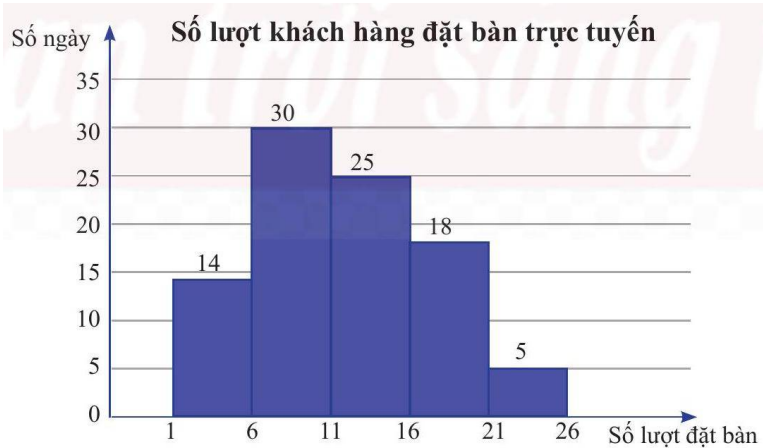
- a) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu trên.
 b) Hãy chia mẫu số liệu trên thành 4 nhóm với nhóm đầu tiên là [140; 240) và lập bảng tần số ghép nhóm.
 c) Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu ghép nhóm và so sánh với kết quả tương ứng thu được ở câu a).

Bài 6. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng của Bác Bình và Bác An



- a) Ai là người có thời gian tập đều hơn?
- b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

Bài 7. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2024 của một nhà hàng. Cột thứ nhất biểu diễn số ngày có từ 1 đến dưới 6 lượt đặt bàn; cột thứ hai biểu diễn số ngày có từ 6 đến dưới 11 lượt đặt bàn; ...



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

Bài 8. Hai bảng tần số ghép nhóm dưới đây thống kê theo độ tuổi số lượng thành viên nam và thành viên nữ đang sinh hoạt trong một câu lạc bộ dưỡng sinh.

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

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

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

DẠNG 3

TÍNH PHƯƠNG SAI, ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHEP NHÓM

Bài toán: Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

Tính phương sai và độ lệch chuẩn của mẫu số liệu trên.

Phương pháp:

Bước 1: Tính giá trị đại diện của từng nhóm:

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

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
Giá trị đại diện	x_1	...	x_i	...	x_k
Tần số	m_1	...	m_i	...	m_k

Bước 2: Tính số trung bình cộng của mẫu số liệu ghép nhóm trên được tính bằng công thức:

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_k x_k}{n}$$

Trong đó: $n = m_1 + m_2 + \dots + m_k$ là cỡ mẫu.

Bước 3:

Tính **phương sai** của mẫu số liệu ghép nhóm theo công thức sau:

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

Hoặc công thức sau:
$$S^2 = \frac{1}{n} [m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2]$$

Tính **độ lệch chuẩn** của mẫu số liệu ghép nhóm theo công thức sau: $S = \sqrt{S^2}$.

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. Người ta theo dõi sự thay đổi cân nặng ở người nam và người nữ, được tính bằng hiệu cân nặng trước và sau ba tháng áp dụng chế độ ăn kiêng của một số người cho kết quả như sau:

Thay đổi cân nặng (kg)	$[-1; 0)$	$[0; 1)$	$[1; 2)$	$[2; 3)$	$[3; 4)$
Số người nam	2	3	5	3	2
Số người nữ	2	7	12	7	2

Tính số trung bình, phương sai, độ lệch chuẩn và nhận xét về sự thay đổi cân nặng của người nam, người nữ sau ba tháng áp dụng chế độ ăn kiêng.

Bài 3. Minh Hiền và Minh Nhân cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

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

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

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

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?

Bài 5. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của hai nhà đầu tư được cho như sau:

Bảng lợi nhuận theo tháng của nhà đầu tư nhỏ

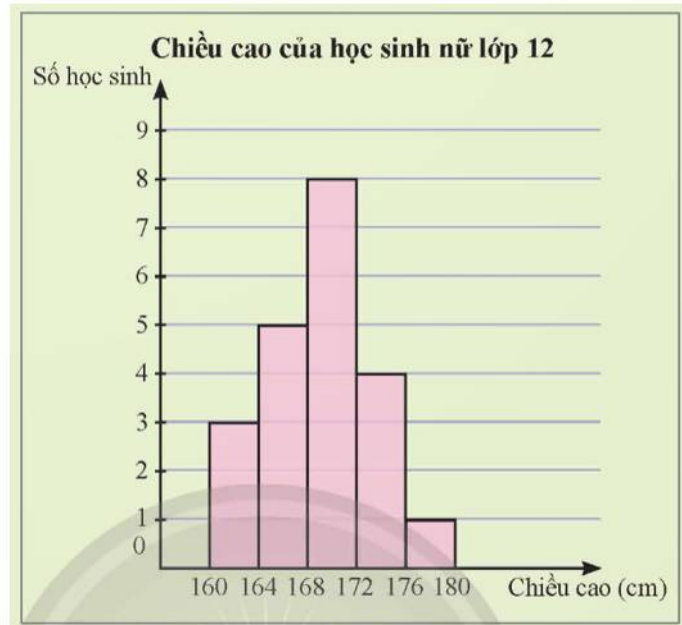
Lợi nhuận (đơn vị: triệu đồng)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
Số tháng	2	4	8	4	2

Bảng lợi nhuận theo tháng của nhà đầu tư lớn

Lợi nhuận (đơn vị: triệu đồng)	[510;520)	[520;530)	[530;540)	[540;550)	[550;560)
Số tháng	4	3	6	3	4

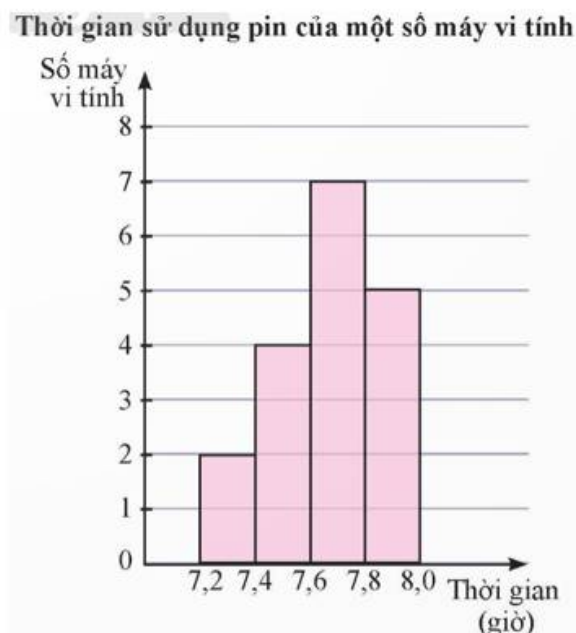
- a) Tính độ lệch chuẩn của hai mẫu số liệu ghép nhóm trên.
b) Có nên dựa vào độ lệch chuẩn để so sánh độ rủi ro của hai nhà đầu tư này không?

Bài 6. Biểu đồ sau biểu diễn chiều cao của học sinh nữ lớp 12.



- a) Hãy lập bảng tần số ghép nhóm cho mẫu số liệu ở biểu đồ trên và xác định giá trị đại diện của mỗi nhóm và tính số trung bình của mẫu số liệu ghép nhóm.
b) Xét mẫu số liệu mới gồm các giá trị đại diện của các nhóm, tần số của mỗi giá trị đại diện bằng tần số của nhóm tương ứng. Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu mới.

Bài 7. 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.

Bài 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.

PHẦN B

TRẮC NGHIỆM VÀ TỰ LUẬN TỔNG HỢP GỒM BỐN PHẦN

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

Câu 1. Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì số đặc trưng nào sau đây không thay đổi?

A. Khoảng biến thiên.

B. Khoảng tứ phân vị.

C. Phương sai.

D. Độ lệch chuẩn.

Câu 2. 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 3. Trong các khẳng định sau, khẳng định nào sai?

A. Phương sai là bình phương của độ lệch chuẩn.

B. Phương sai luôn luôn là số không âm.

C. Phương sai luôn luôn lớn hơn độ lệch chuẩn.

D. Phương sai càng lớn thì độ phân tác của các giá trị quanh số trung bình càng lớn.

Câu 4. Thời gian hoàn thành bài kiểm tra môn Toán của các học sinh trong lớp 12A được cho trong bảng sau:

Thời gian (phút)	[25;30)	[30;35)	[35;40)	[40;45)
Số học sinh	8	16	12	2

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 20.

B. 30.

C. 27.

D. 35.

Câu 5. Thời gian hoàn thành giải chạy của các vận động viên được cho như bảng sau:

Thời gian (phút)	[15;20)	[20;25)	[25;30)	[30;35)
Số vận động viên	5	10	8	2

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 10.

B. 25.

C. 30.

D. 20.

Câu 6. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 120 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	[40;50)	[50;60)	[60;70)	[70;80)	[80;90)
Tần số	6	12	38	46	18

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

- A. 30. B. 40. **C. 50.** D. 60.

Câu 7. Mức thưởng tết (triệu đồng) cho các nhân viên của một công ty được thống kê trong bảng sau:

Mức thưởng tết	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số nhân viên	10	5	6	4	5

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 20. **B. 25.** C. 47. D. 23

Câu 8. Chiều cao của 80 học sinh ở một trường Đại Học Thể Dục Thể Thao được thống kê trong bảng sau:

Chiều cao (cm)	[160;163,5)	[164;167,5)	[168;171,5)	[172;175,5)
Tần số	18	40	14	8

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 31. **B. 15,5.** C. 74. D. 32.

Câu 9. Thời gian dùng điện thoại 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	0	24	30	48	52

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 26. B. 15. C. 20. **D. 12.**

Câu 10. Một Bác nông dân thống kê lại đường kính thân gỗ của một số cây xoan đào 7 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	9	18	18	10	0

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

- A. 25. B. 30. **C. 20.** D. 35.

Câu 11. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

Tuổi thọ (năm)	[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 12. 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 13. Khảo sát thời gian tập thể dục trong ngày của một số học sinh khối 12 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 14. 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 15. 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 16. 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

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

A. 20,67.

B. 17,24.

C. 18,24.

D. 22,67.

Câu 17. 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

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

A. 18, 24.

B. 17, 24.

C. 22, 67.

D. 20, 67.

Câu 18. 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 19. 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

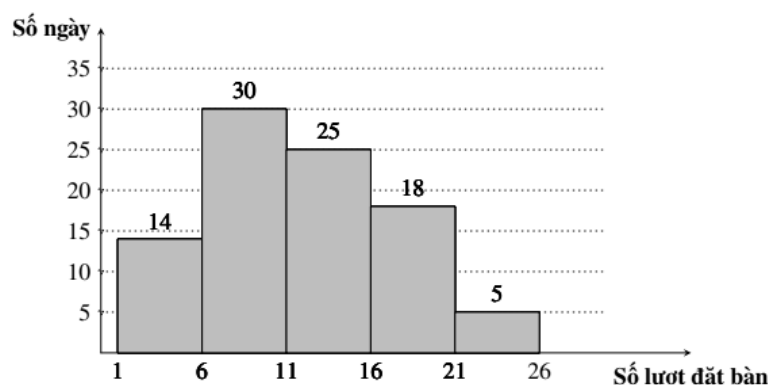
A. 12,6.

B. 11,5.

C. 14,3.

D. 16,8.

Câu 20. Biểu đồ dưới đây biểu diễn số lượng 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 6 lượt đặt bàn, cột thứ hai biểu diễn số ngày có từ 6 đến 11 lượt đặt bàn;...



Khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi biểu đồ trên

A. 9,5.

B. 8,5.

C. 10,5.

D. 7,5.

Câu 21. 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 22. Trong buổi thực hành, 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

Tính số trung bình của mẫu số liệu trên.

A. 4,44.

B. 5,45.

C. 4,45.

D. 6,44.

Câu 23. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư vào hai lĩnh vực A cho kết quả như sau:

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

Tính số trung bình của mẫu số liệu trên.

A. 17,9 (triệu đồng).

B. 19,5 (triệu đồng).

C. 16,9 (triệu đồng).

D. 18,5 (triệu đồng).

Câu 24. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư vào lĩnh vực B cho kết quả như sau:

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

Tính số trung bình của mẫu số liệu trên.

A. 17,9 (triệu đồng).

B. 19,5 (triệu đồng).

C. 16,9 (triệu đồng).

D. 18,5 (triệu đồng).

Câu 25. Anh Bảo Toàn đầu tư số tiền vào lĩnh vực kinh doanh A. Anh Bảo Toàn thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực A	5	10	30	10	5

Độ lệch chuẩn của số tiền thu được hàng tháng khi anh Bảo Toàn đầu tư vào lĩnh vực A là:

A. 4,5.

B. 5.

C. 5,5.

D. 6.

Câu 26. Cô Thúy đầu tư số tiền vào lĩnh vực kinh doanh B. Cô Thúy thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực B	20	5	10	5	20

Độ lệch chuẩn của số tiền thu được hàng tháng khi Cô Thúy đầu tư vào lĩnh vực B là:

A. 8,42 .

B. 9,22 .

C. 7,42 .

D. 7,22 .

Câu 27. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, bạn Minh Hiền đã dùng hai vôn kế khác nhau để đo 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 Minh Hiền đo	1	6	2	1

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm cho kết quả đo của bạn Minh Hiền.

A. 0,039 .

B. 0,0015.

C. 0,024 .

D. 0,0025.

Câu 28. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, bạn Trúc Linh đã dùng hai vôn kế khác nhau để đo 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 Trúc Linh đo	1	3	4	2

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm cho kết quả đo của bạn Trúc Linh

A. 0,002 .

B. 0,012 .

C. 0,045 .

D. 0,055.

Câu 29. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi phân xưởng A 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 A	4	9	13	8	6

Tính phương sai của mẫu số liệu ghép nhóm trên

A. 0,145 .

B. 0,245 .

C. 0,445 .

D. 0,355 .

Câu 30. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi phân xưởng B 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 B	2	8	20	7	3

Tính phương sai của mẫu số liệu ghép nhóm trên

A. 0,219 .

B. 0,329 .

C. 0,119 .

D. 0,129 .

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

Câu 31. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm này là 5

b) Nhóm chứa tứ phân vị thứ nhất là nhóm [15;16)

c) Nhóm chứa tứ phân vị thứ ba là nhóm [17;18)

d) Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì khoảng biến thiên không thay đổi.

Câu 32. 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

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

b) Mốt của mẫu số liệu là $M_o = 8,0625$.

c) Trung vị của mẫu số liệu là $M_e \approx 9,95$

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 $14,79m^3$ nước trở lên.

Câu 33. Hằng ngày bạn Minh Hiền đều đi xe buýt từ nhà đến trường học. Dưới đây là bảng thống kê thời gian của 100 lần bạn Minh Hiền đi xe buýt từ nhà đến trường.

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

a) Cỡ mẫu $n = 100$.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = \frac{983}{37}$.

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta_Q = \frac{615}{113}$

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

Câu 34. 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) Khoảng biến 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.

Câu 35. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

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

a) 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: 15 (tuổi)

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm ứng với khu vực A là: $\frac{721}{25}$ (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 36. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 10 từ năm 2005 đến 2024 tại một trạm quan trắc đặt ở Nha Trang.

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

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

b) Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = \frac{17221}{100}$

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

Khi đó, khoảng biến thiên của mẫu số liệu ghép nhóm là $400(mm)$

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

Khi đó, khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\frac{1100}{7}$

Câu 37. Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 0,6 .

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là 0,575 .

c) Số trung bình của mẫu số liệu ghép nhóm là 3,62 .

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm 0,36 (kết quả làm tròn đến phần trăm).

Câu 38. Cô Minh Hiền rất thích luyện tập yoga. Thời gian tập mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 25.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là 8,125 .

c) Số trung bình của mẫu số liệu ghép nhóm là $\frac{85}{3}$.

d) Phương sai của mẫu số liệu ghép nhóm là 31,25 .

Câu 39. Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 12.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = 10,75$

c) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm là: $Q_3 = 14,375$

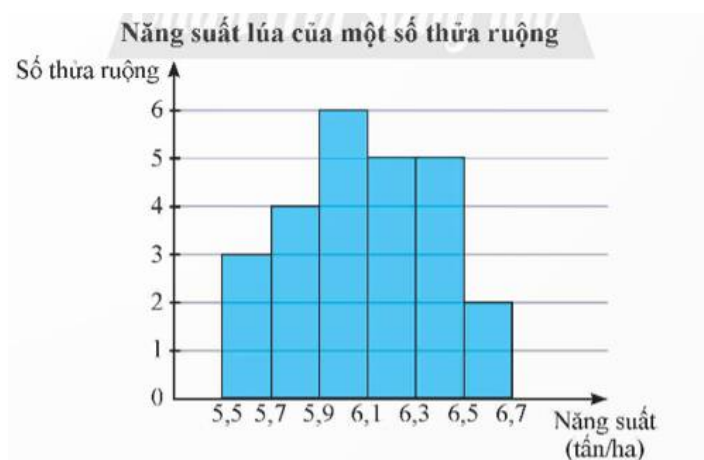
d) Độ lệch chuẩn của mẫu số liệu ghép nhóm là 2,74 (kết quả làm tròn đến phần trăm).

Câu 40. Một bác tài xế thống kê lại độ dài quãng đường (đơn vị: km) bác đã lái xe mỗi ngày trong một tháng ở bảng sau:

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

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

Câu 41. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau:



- a) Có 18 thửa ruộng đã được khảo sát.
 b) Khoảng biến thiên của mẫu số liệu trên là 1,4(tấn/ha).
 c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là 0,4675 .
 d) Phương sai của mẫu số liệu ghép nhóm trên là 0,086656 .

Câu 42. Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

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

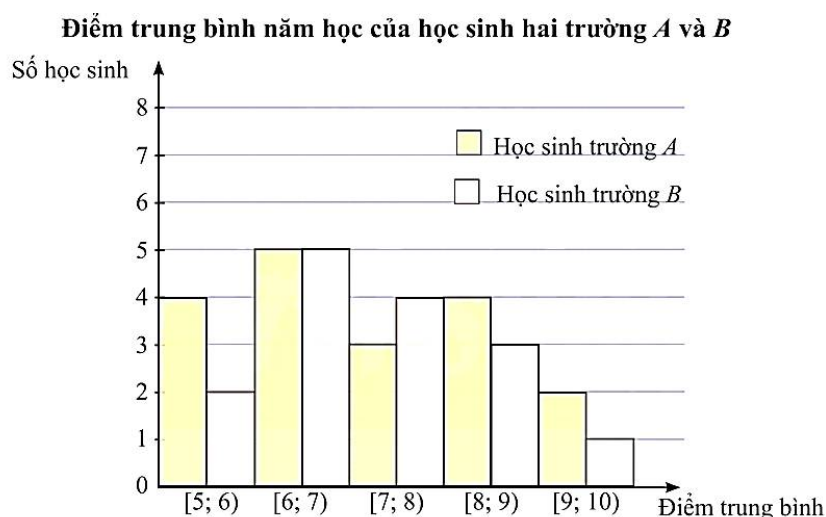
- a) Nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn.
 b) Nếu so sánh theo khoảng tứ phân vị thì học sinh trường Y có tốc độ viết đồng đều hơn.
 c) Phương sai của mẫu số liệu ghép nhóm của trường X là 1,08 và phương sai của mẫu số liệu ghép nhóm của trường Y là 1,7584 .
 d) Nếu so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

Câu 43. Bảng sau thống kê lại tổng số giờ nắng trong tháng 6 của các năm từ 2002 đến 2021 tại hai trạm quan trắc đặt ở Nha Trang và Quy Nhơn.

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

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

Câu 44. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.

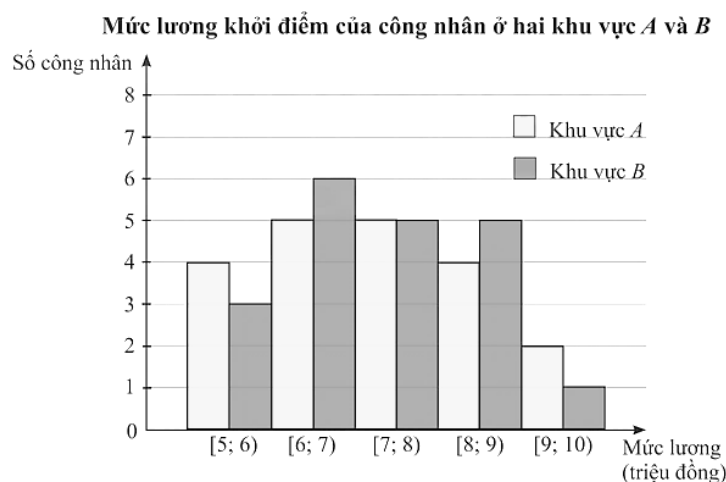


- a) Giá trị đại diện cho mỗi nhóm và bảng tần số ghép nhóm của mẫu số liệu trên là:

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

- b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm của trường A là 2,275.
- c) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường A có điểm trung bình đồng đều hơn trường B.
- d) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn.

Câu 45. 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) Bảng tần số ghép nhóm và giá trị đại diện cho mẫu số liệu trên như sau:

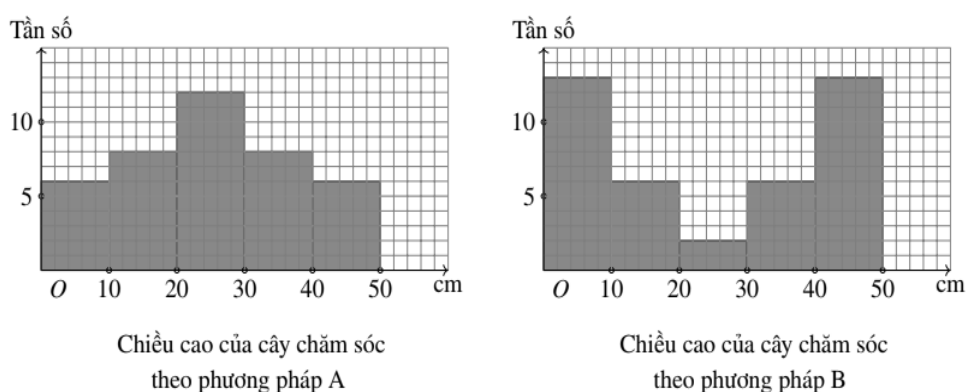
Mức lương (triệu đồng)	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
Mức lương đại diện (triệu đồng)	5,5	6,5	7,5	8,5	9,5
Khu vực A	4	5	5	4	2
Khu vực B	3	6	5	5	1

b) Xét mẫu số liệu của khu vực A 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 46. Một công ty giống cây trồng đã thử nghiệm hai phương pháp chăm sóc khác nhau cho cây hướng dương. Sau hai tuần, người ta thấy cây được chăm sóc theo cả hai phương pháp đều thấp hơn 50 cm.



Từ biểu đồ, ta có bảng tần số ghép nhóm về chiều cao của cây được chăm sóc theo phương pháp A và B là:

Chiều cao (cm)	[0;10)	[10; 20)	[20; 30)	[30; 40)	[40; 50)
Mức lương đại diện (cm)	5	15	25	35	45

Chăm sóc theo phương pháp A	6	8	12	8	6
Chăm sóc theo phương pháp B	13	6	2	6	13

- a) Khoảng biến thiên của chiều cao các cây được chăm sóc theo mỗi phương pháp A và B bằng nhau.
- b) Trung bình của chiều cao các cây được chăm sóc theo mỗi phương pháp A và B bằng nhau.
- c) Độ lệch chuẩn của chiều cao các cây được chăm sóc theo phương án A là 12,65(cm).
- d) Dựa vào độ lệch chuẩn thì chiều cao của các loại cây được chăm sóc theo phương án B ít bị chênh lệch hơn so với phương án A .

Câu 47. Kết quả 40 lần nhảy xa của hai vận động viên Dũng và Huy được lần lượt thống kê trong bảng bên dưới (đơn vị: mét)

Nhóm	Dũng	Huy
[6, 22; 6, 46)	3	2
[6, 46; 6, 70)	7	5
[6, 70; 6, 94)	5	8
[6, 94; 7, 18)	20	19
[7, 18; 7, 42)	5	6
	$n = 40$	$n = 40$

- a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng là 6,92(m) (kết quả làm tròn đến phần trăm của mét).
- b) Độ lệch chuẩn của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho là 0,26(m) (kết quả làm tròn đến phần trăm của mét).
- c) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho là 0,16 (kết quả làm tròn đến phần trăm).
- d) Kết quả nhảy xa của vận động viên Dũng đồng đều hơn kết quả nhảy xa của vận động viên Huy.

Câu 48. Mẫu số liệu ghép nhóm thống kê mức lương của hai công ty A, B (đơn vị: triệu đồng) được thể hiện như bảng dưới đây

Nhóm	Công ty A	Công ty B
[10;15)	15	25
[15;20)	18	15
[20;25)	10	7
[25;30)	10	5
[30;35)	5	5
[35;40)	2	3
	$n = 60$	$n = 60$

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 là 52,91 (làm tròn kết quả đến hàng phần trăm)

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 49. Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11A 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) 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 $\sqrt{0,59}$.

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.

Câu 50. 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

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.

Câu 51. 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:

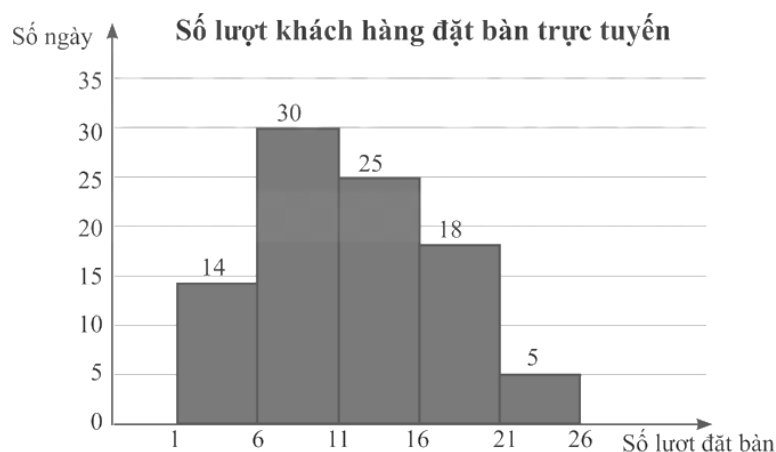
Câu 52. Bảng sau thống kê cân nặng của 50 quả xoài Úc được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường.

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

Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm đó (kết quả làm tròn đến phần mười).

Trả lời:

Câu 53. 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:

Câu 54. Thời gian luyện tập trong một ngày (tính theo giờ) của một số vận động viên được ghi lại ở bảng sau:

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

Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên (kết quả làm tròn đến phần trăm).

Trả lời:

Câu 55. 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 (kết quả làm tròn đến hàng đơn vị).

Trả lời:

Câu 56. 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:

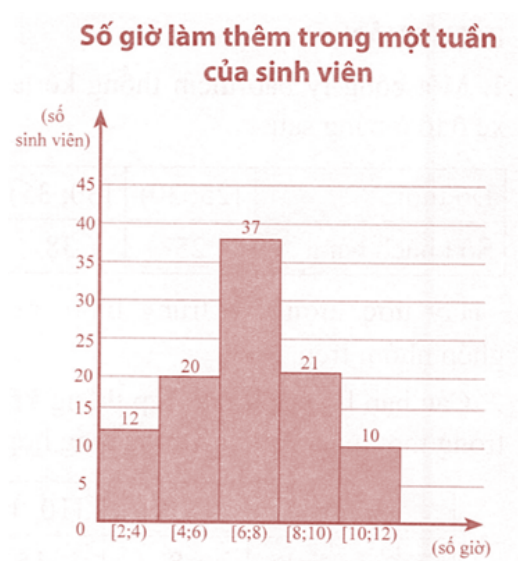
Câu 57. Thời gian sử dụng điện thoại trong một ngày của 30 sinh viên được ghi lại ở bảng sau (đơn vị: phút).

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 (kết quả làm tròn đến hàng đơn vị).

Trả lời:

Câu 58. 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:

Câu 59. 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 (kết quả làm tròn đến phần mười).

Trả lời:

Câu 60. 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:

Câu 61. 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:

Câu 62. 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 mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

Trả lời:

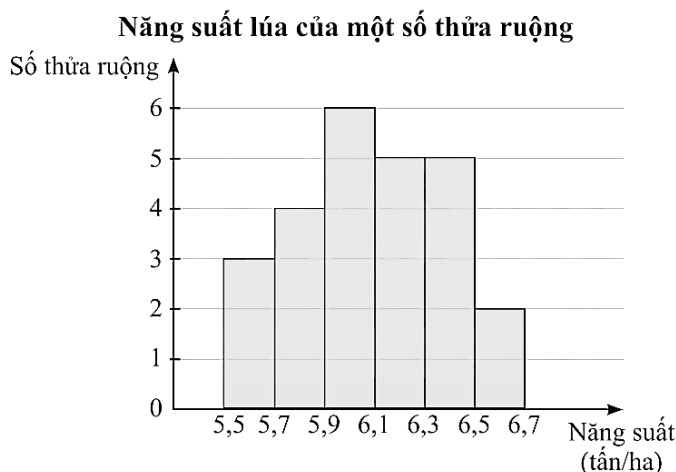
Câu 63. Chiều dài của 40 bé trai sơ sinh 12 ngày tuổi chọn ngẫu nhiên ở một bệnh viện được nhà nghiên cứu thống kê trong Bảng dưới đây:

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:

Câu 64. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



Hãy xác định độ lệch chuẩn của mẫu số liệu trên (làm tròn kết quả đến hàng phần trăm).

Trả lời:

Câu 65. 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 (làm tròn kết quả đến hàng phần trăm).

Trả lời:

Câu 66. 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:

Câu 67. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của một nhà đầu tư được cho như sau:

Lợi nhuận (đơn vị: triệu đồng)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
Số tháng	2	4	8	4	2

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần mười).

Trả lời:

Câu 68. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của một nhà đầu tư được cho như sau:

Lợi nhuận (đơn vị: triệu đồng)	[510;520)	[520;530)	[530;540)	[540;550)	[550;560)
Số tháng	4	3	6	3	4

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần mười).

Trả lời:

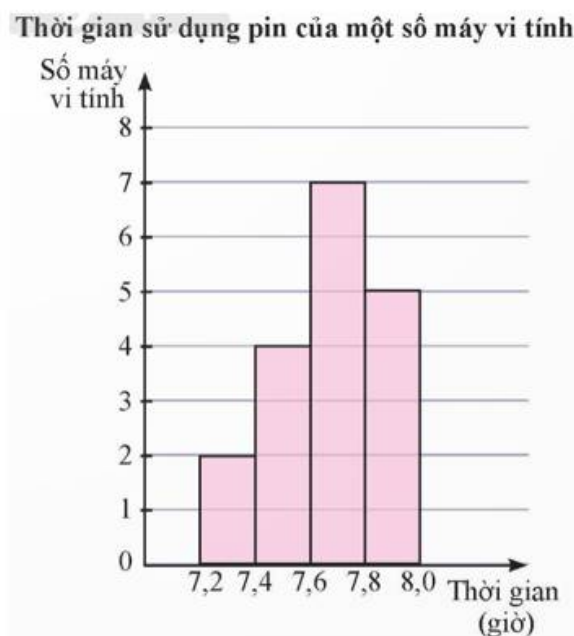
Câu 69. Trong buổi thực hành, một nhóm 20 học sinh dùng một thiết bị đo đường kính của một nhân tế bào cho kết quả như sau:

Kết quả đo (μm)	[4,5;5)	[5,5;5)	[5,5;6)	[6,6;5)
Số học sinh	3	8	7	2

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần trăm).

Trả lời:

Câu 70. 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.



Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần trăm).

Trả lời:

PHẦN IV. Câu tự luận. Mỗi câu hỏi thí sinh trình bày cách giải tự luận.

Câu 71. Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của cô Minh Nhân và cô Minh Hiền.

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Cô Minh Nhân	5	12	8	3	2
Cô Minh Hiền	0	25	5	0	0

Hỏi thời gian tập thể dục buổi sáng mỗi ngày của ai phân tán hơn?

Câu 72. Thời gian chờ khám bệnh của các bệnh nhân tại phòng khám X được cho trong bảng sau:

Thời gian (phút)	[0; 5)	[5; 10)	[10; 15)	[15; 20)
Số bệnh nhân	3	12	15	8

- a) Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm này.
- b) Từ một mẫu số liệu về thời gian chờ khám bệnh của các bệnh nhân tại phòng khám Y người ta tính được khoảng tứ phân vị bằng 9,23. Hỏi thời gian chờ của bệnh nhân tại phòng khám nào phân tán hơn?

Câu 73. Thống kê số ngày trong tháng 11 năm 2023 và năm 2024 theo nhiệt độ cao nhất trong ngày tại Nha Trang, 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 11/2023	0	2	8	5	6	9
Số ngày trong tháng 11/2024	2	3	4	11	8	2

Hỏi tháng sáu năm nào ở Nha Trang nhiệt độ cao nhất trong ngày biến đổi nhiều hơn?

Câu 74. 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 75. Bảng sau đây cho biết chiều cao của các học sinh lớp 12A và 12B.

Thu nhập	[145; 150)	[150; 155)	[155; 160)	[160; 165)	[165; 170)	[170; 175)
Số học sinh lớp 12A	1	0	15	12	10	5
Số học sinh 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 76. 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 77. 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 họa câu trả lời bằng cách biểu diễn các tứ phân vị và khoảng tứ phân vị của mỗi mẫu số liệu trên trục số.

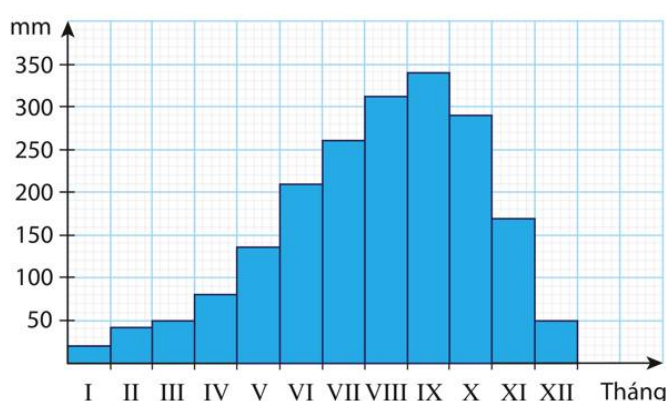
Câu 78. Thống kê số thẻ vàng của mỗi câu lạc bộ trong giải ngoại hạng Anh mùa giải 2021 - 2022 cho kết quả như sau:

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

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 79. 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.

Câu 80. Hằng ngày Bác Toàn đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bản thống kê thời gian của 100 lần Bác Toàn đi xe buýt từ nhà đến cơ quan:

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

a) Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm phía trên (Làm tròn kết quả đến hàng phần trăm)

b) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần Bác Toàn đi hết 32 phút. Thời gian của lần đi đó có phải giá trị ngoại lệ không?

Câu 81. Trong buổi thực hành, 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 82. Hai vận động viên điền kinh A và B ghi lại thời gian chạy tập luyện chạy cự li 100 m của mình 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 83. Để rèn luyện kỹ năng đi bộ, Cô Minh Nhân và Cô Minh Hiền cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân đi bộ của hai cô 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)
Cô Minh Nhân	6	7	6	6	5
Cô Minh Hiền	2	5	13	8	2

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

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

Câu 84. Một cuộc khảo sát thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 5 hai trường A và B được ghi lại ở bảng sau:

Thời gian (phút)	[6;7)	[7;8)	[8;9)	[9;10)	[10;11)
Học sinh trường A	8	10	13	10	9
Học sinh trường B	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 85. 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

a) Tính phương sai và độ lệch chuẩn của mỗi mẫu số liệu ghép nhóm.

b) Nhận xét về độ phân tán của tuổi thọ các linh kiện điện tử được sản xuất bởi mỗi phân xưởng.

Câu 86. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, hai bạn Việt và Nam đã dùng hai vôn kế khác nhau để đo 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 Việt đo	1	6	2	1
Số lần Nam đo	1	3	4	2

a) Tính độ lệch chuẩn của các mẫu số liệu ghép nhóm cho kết quả đo của Việt và Nam.

b) Từ đó kết luận xem vôn kế của bạn nào cho kết quả đo ổn định hơn.

Câu 87. Thầy Tuấn đầu tư số tiền bằng nhau vào hai lĩnh vực kinh doanh A, B . Thầy Tuấn thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực A	5	10	30	10	5
Số tháng đầu tư vào lĩnh vực B	20	5	10	5	20

So sánh giá trị trung bình và độ lệch chuẩn của số tiền thu được mỗi tháng khi Thầy Tuấn đầu tư vào mỗi lĩnh vực A, B . Đầu tư vào lĩnh vực nào "rủi ro" hơn?

Câu 88. 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 (triệu đồng)	[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.

Câu 89. 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 90. 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.

CHƯƠNG 3

CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM

BÀI 1,2

KHOẢNG BIẾN THIÊN VÀ KHOẢNG TƯ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHÉP NHÓM

I. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM

1. Khoảng biến thiên của mẫu số liệu ghép nhóm

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 số liệu ghép nhóm trên là $R = a_{k+1} - a_1$

Ý nghĩa

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

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

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

2. Khoảng tứ phân vị của mẫu số liệu ghép nhóm

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

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

Chú ý:

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

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

Ý nghĩa

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

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

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

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 (còn gọi giá trị bất thường) nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$

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

3. Phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm

Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

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

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

Trong đó:

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

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + \dots + m_kx_k}{n} \text{ là số trung bình của mẫu số liệu ghép nhóm.}$$

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

Chú ý:

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

$$S^2 = \frac{1}{n} [m_1x_1^2 + m_2x_2^2 + \dots + m_kx_k^2] - \bar{x}^2$$

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

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

Ý nghĩa

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

Chú ý: Sử dụng phương sai, độ lệch chuẩn đo độ rủi ro:

Trong tài chính, người ta có nhiều cách để đo độ rủi ro của một phương án đầu tư. Một trong các cách đó là sử dụng độ lệch chuẩn của lợi nhuận thu được theo phương án đầu tư. Độ lệch chuẩn càng lớn thì phương án đầu tư càng rủi ro.

II. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CHO MẪU SỐ LIỆU GHEP NHÓM

(Nhắc lại kiến thức lớp 11)

1. Số trung bình của mẫu số liệu ghép nhóm

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

Số trung bình của mẫu số liệu ghép nhóm, kí hiệu $\bar{x}$, được tính như sau:

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + \dots + m_kx_k}{n}$$

Trong đó:

- $n = m_1 + m_2 + \dots + m_k$ là cỡ mẫu
- $x_i = \frac{a_i + a_{i+1}}{2}$ (với $i = 1, 2, \dots, k$) là giá trị đại diện của nhóm $[a_i; a_{i+1})$

Chú ý: Đối với số liệu rời rạc, người ta thường cho các nhóm dưới dạng $k_1 - k_2$ trong đó $k_1, k_2 \in \mathbb{N}$.

Nhóm $k_1 - k_2$ được hiểu là nhóm gồm các giá trị $k_1, k_1 + 1, \dots, k_2$. Khi đó, ta cần hiệu chỉnh mẫu số liệu ghép nhóm đưa về dạng bảng sau, trước khi thực hiện tính toán các số đặc trưng bằng cách hiệu chỉnh nhóm $k_1 - k_2$ với $k_1, k_2 \in \mathbb{N}$ thành nhóm $[k_1 - 0,5; k_2 + 0,5)$.

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

Ý nghĩa: Số trung bình của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho số trung bình của mẫu gốc, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

2. Trung vị của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tính trung vị của mẫu số liệu ghép nhóm, ta làm các bước sau:

Bước 1: Xác định nhóm chứa trung vị. Giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$

Bước 2: Trung vị là: $M_e = a_p + \frac{\frac{n}{2} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

Ý nghĩa: Trung vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho mẫu số liệu gốc, nó chia mẫu số liệu thành hai phần, mỗi phần chứa 50% giá trị.

3. Tứ phân vị của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tìm **tứ phân vị thứ nhất** Q_1 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_1 , giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó:

$$Q_1 = a_p + \frac{\frac{n}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

Để tìm **tứ phân vị thứ ba** Q_3 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_3 , giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó:

$$Q_3 = a_p + \frac{\frac{3n}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$$

Trong đó:

- n là cỡ mẫu.
- m_p là tần số của nhóm p , với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$

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

Nhận xét: Ta có thể xác định nhóm chứa tứ phân vị thứ r nhờ tính chất: có khoảng $\left(\frac{r.n}{4}\right)$ giá trị nhỏ hơn tứ phân vị này.

Ý nghĩa: Các tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho các tứ phân vị của mẫu số liệu gốc, chúng chia mẫu số liệu thành 4 phần, mỗi phần chứa 25% giá trị.

4. Một của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm như sau:

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

Để tìm **mốt** của mẫu số liệu ghép nhóm, ta làm các bước sau:

Bước 1: Xác định nhóm có chứa tần số lớn nhất (gọi là nhóm chứa mốt), giả sử là nhóm $j: [a_j; a_{j+1})$

Bước 2: Mốt được xác định là: $M_o = a_j + \frac{m_j - m_{j-1}}{(m_j - m_{j-1}) + (m_j - m_{j+1})} \cdot h$

Trong đó:

- m_j là tần số của nhóm j (quy ước $m_0 = m_{k+1} = 0$).
- h là độ dài của nhóm.

Chú ý:

• Người ta chỉ định nghĩa mốt cho mẫu ghép nhóm có độ dài các nhóm bằng nhau. Một mẫu có thể không có mốt hoặc có nhiều hơn một mốt

- Khi tần số của các nhóm số liệu bằng nhau thì mẫu số liệu ghép nhóm không có mốt.

Ý nghĩa: Mốt của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho mốt của mẫu gốc, nó được dùng để đo xu thế trung tâm của mẫu số liệu.

III. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU KHÔNG GHÉP NHÓM (Nhắc lại kiến thức lớp 10)

1. Số trung bình

- Giả sử ta có một mẫu số liệu là $x_1 + x_2 + \dots + x_n$.

Số trung bình (số trung bình cộng) của mẫu số liệu $x_1 + x_2 + \dots + x_n$, kí hiệu là $\bar{x}$, được tính bằng

công thức: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

Ví dụ 1: Điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II được ghi ở bảng sau:

Nhóm I	10	5	7	9	8	6
Nhóm II	9	9	8	7	6	8

Nhóm nào có thành tích điểm tốt hơn?

Lời giải

Điểm trung bình của nhóm I là: $\frac{10+5+7+9+8+6}{6} = 7,5$

Điểm trung bình của nhóm II là: $\frac{9+9+8+7+6+8}{6} \approx 7,83$.

Vậy nhóm II có thành tích điểm tốt hơn nhóm I.

- Giả sử mẫu số liệu được cho dưới dạng bảng tần số

Giá trị	x_1	x_2	$\dots$	x_k
Tần số	n_1	n_2	$\dots$	n_k

Khi đó, công thức tính số trung bình trở thành: $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n}$

Trong đó $n = n_1 + n_2 + \dots + n_k$. Ta gọi n là **cỡ mẫu**.

Chú ý: Nếu kí hiệu $f_k = \frac{n_k}{n}$ là tần số tương đối (hay còn gọi là tần suất) của x_k trong mẫu số liệu thì số

trung bình còn có thể biểu diễn là: $\bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k$

Ví dụ 2: Kết quả điều tra về số con của một số hộ gia đình trong một tổ dân phố được ghi lại trong bảng sau:

Số con	0	1	3	4	5
Số hộ gia đình	4	4	8	3	1

Hỏi trung bình mỗi hộ gia đình trong tổ dân phố có bao nhiêu con?

Lời giải

Tổng số hộ gia đình là: $n = 4 + 4 + 8 + 3 + 1 = 20$ (hộ gia đình).

Trung bình mỗi hộ gia đình trong tổ dân phố có số con là: $\frac{0.4 + 1.4 + 3.8 + 4.3 + 5.1}{20} = 1,65$ (con)

Ý nghĩa của số trung bình

Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đo xu thế trung tâm của mẫu đó.

Chẳng hạn, ở trong ví dụ 1 trên, ta thấy điểm số trung bình của nhóm II cao hơn nhóm I ($7,83 > 7,5$), ta có thể nói rằng thành tích thực hành của nhóm II tốt hơn nhóm I.

2. Trung vị và tứ phân vị

a. Trung vị

- **Định nghĩa và cách tính số trung vị**

Khi các số liệu trong mẫu số liệu chênh lệch nhau quá lớn, ta dùng một đặc trưng khác của mẫu số liệu, gọi là **trung vị** để so sánh các mẫu số liệu với nhau.

Trung vị được định nghĩa như sau:

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

Trung vị của mẫu, kí hiệu là M_e , là giá trị ở chính giữa dãy $x_1, x_2, \dots, x_n$. Cụ thể:

- Nếu $n = 2k + 1, k \in \mathbb{N}$ (tức n là số tự nhiên lẻ), thì trung vị của mẫu $M_e = x_k + 1$.

- Nếu $n = 2k, k \in \mathbb{N}$, (tức n là số tự nhiên chẵn), thì trung vị của mẫu $M_e = \frac{x_k + x_{k+1}}{2}$.

Ví dụ 3: Điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II được ghi ở bảng sau:

Nhóm I	10	5	7	9	8	6
Nhóm II	9	9	8	7	6	8

Tính các trung vị của điểm số thi học kỳ 1 môn Toán 10 của các bạn học sinh trong nhóm I và nhóm II.

Lời giải

+ Sắp xếp điểm số của mỗi bạn trong nhóm I theo thứ tự không giảm, ta được: 5; 6; 7; 8; 9; 10

Vì cỡ mẫu bằng 6 nên trung vị của nhóm I là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là $M_e = \frac{7+8}{2} = 7,5$.

+ Sắp xếp điểm số của mỗi bạn trong nhóm II theo thứ tự không giảm, ta được: 6; 7; 8; 8; 9; 9

Vì cỡ mẫu bằng 6 nên trung vị của nhóm II là trung bình cộng của số liệu thứ 3 và thứ 4 của dãy trên, tức là $M_e = \frac{8+8}{2} = 8$.

• Ý nghĩa của số trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có ít nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

Ví dụ 4: Bảng sau thống kê số sách mỗi bạn học sinh Tổ 1 và Tổ 2 đã đọc ở thư viện trường trong một tháng:

Tổ 1	3	1	2	1	2	2	3	25	1
Tổ 2	4	5	4	3	3	4	5	4	

a) Trung bình mỗi bạn Tổ 1 và mỗi bạn Tổ 2 đọc bao nhiêu quyển sách ở thư viện trường trong tháng đó?

b) Em hãy thảo luận với các bạn trong nhóm xem tổ nào chăm đọc sách ở thư viện hơn.

Lời giải

a) Trung bình mỗi bạn Tổ 1 đọc số quyển sách ở thư viện trong tháng trên là:

$$\frac{3+1+2+1+2+2+3+25+1}{9} = \frac{40}{9} \approx 4,4$$

Trung bình mỗi bạn Tổ 2 đọc số quyển sách ở thư viện trong tháng trên là:

$$\frac{4+5+4+3+3+4+5+4}{8} = \frac{32}{8} = 4$$

b) Vì $4,4 > 4$ nên theo số trung bình, các bạn Tổ 1 đọc sách chăm hơn.

Nếu dựa vào số trung bình để đánh giá xem tổ nào chăm đọc sách hơn trong bài này thì không phù hợp, do có một số liệu trong mẫu số liệu của Tổ 1 quá lớn so với các số liệu còn lại. Ta sử dụng trung vị để so sánh độ chăm học giữa hai tổ.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 1: 1; 1; 1; 2; 2; 2; 3; 3; 25

Vì cỡ mẫu $n_1 = 9$ là số lẻ, nên trung vị của mẫu số liệu Tổ 1 là $M_{e_1} = 2$.

+ Sắp xếp mẫu số liệu theo thứ tự không giảm của Tổ 2: 3; 3; 4; 4; 4; 4; 5; 5

Vì cỡ mẫu $n_2 = 8$ là số chẵn, nên trung vị của mẫu số liệu Tổ 2 là $M_{e_2} = \frac{4+4}{2} = 4$.

Do đó ta có: $M_{e_2} > M_{e_1}$.

Vậy theo trung vị, các bạn Tổ 2 chăm đọc sách ở thư viện hơn Tổ 1.

b. Tứ phân vị

• Trung vị chia mẫu thành hai phần. Trong thực tế người ta cũng quan tâm đến trung vị của mỗi phần đó. Ba trung vị này được gọi là **tứ phân vị** của mẫu.

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

Tứ phân vị của một mẫu số liệu gồm ba giá trị, gọi là **tứ phân vị thứ nhất, thứ hai và thứ ba** (lần lượt kí hiệu là Q_1, Q_2, Q_3). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

- Giá trị tứ phân vị thứ hai, Q_2 , chính là số trung vị của mẫu.

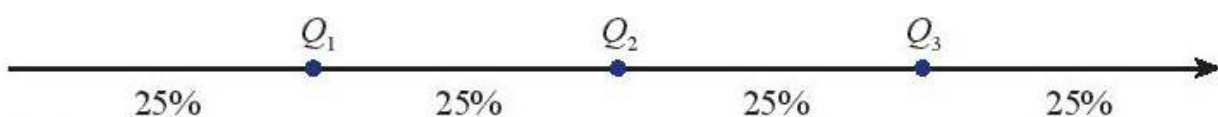
- Giá trị tứ phân vị thứ nhất, Q_1 , là trung vị của nửa số liệu đã sắp xếp bên trái Q_2 (không bao gồm Q_2 nếu n lẻ).

- Giá trị tứ phân vị thứ ba, Q_3 , là trung vị của nửa số liệu đã sắp xếp bên phải Q_2 (không bao gồm Q_2 nếu n lẻ).

• Ý nghĩa của tứ phân vị

Các điểm tứ phân vị Q_1, Q_2, Q_3 chia mẫu số liệu đã sắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần chia khoảng 25% tổng số liệu đã thu thập được.

Tứ phân vị thứ nhất Q_1 còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liệu phía dưới. Tứ phân vị thứ ba Q_3 , còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu ở phía trên.



Ví dụ 5: Tìm tứ phân vị của mẫu số liệu sau: 3; 5; 6; 13; 25; 17; 19.

Lời giải

Sắp xếp các số liệu theo thứ tự không giảm ta được: 3; 5; 6; 13; 17; 19; 25.

Vì cỡ mẫu $n = 7$, là số lẻ, nên giá trị tứ phân vị thứ hai là $Q_2 = 13$.

Tứ phân vị thứ nhất là trung vị của mẫu: 3; 5; 6. Do đó $Q_1 = 5$.

Tứ phân vị thứ ba là trung vị của mẫu: 17; 19; 25. Do đó $Q_3 = 19$.

3. Một

Cho mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu và kí hiệu là M_o .

Ý nghĩa của mốt: Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

Chú ý: Một mẫu số liệu có thể có rất nhiều mốt. Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

Ví dụ 6: Cho mẫu số liệu:

Giá trị	35	38	40	45
Tần số	10	5	6	3

Tìm mốt của mẫu số liệu trên.

Lời giải

Ta thấy giá trị 35 có tần số 10 là lớn nhất, do đó, mốt của mẫu số liệu trên là $M_o = 35$.

IV. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU KHÔNG GHÉP NHÓM (Nhắc lại kiến thức lớp 10)

1. Khoảng biến thiên và khoảng tứ phân vị

a. Khoảng biến thiên và khoảng tứ phân vị

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$

• **Khoảng biến thiên** của một mẫu số liệu, kí hiệu là R , là hiệu giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó, tức là: $R = x_n - x_1$

• **Khoảng tứ phân vị**, kí hiệu là Δ_Q , là hiệu giữa Q_3 và Q_1 , tức là: $\Delta_Q = Q_3 - Q_1$

Ví dụ 1: Hãy tính khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu: 10; 3; 5; 7; 20; 1; 4; 9.

Lời giải

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: 1; 3; 4; 5; 7; 9; 10; 20.

- Khoảng biến thiên của mẫu số liệu là $R = 20 - 1 = 19$.
- Cỡ mẫu là $n = 8$, là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = 6$.
- Tứ phân vị thứ nhất là trung vị của mẫu: 10; 3; 5; 7. Do đó $Q_1 = 4$.
- Tứ phân vị thứ 3 là trung vị của mẫu: 7; 9; 10; 20. Do đó $Q_3 = 9,5$.
- Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 9,5 - 4 = 5,5$.

b. Ý nghĩa của khoảng biến thiên và khoảng tứ phân vị

- Khoảng biến thiên đặc trưng cho độ phân tán của toàn bộ mẫu số liệu.
- Khoảng tứ phân vị đặc trưng cho độ phân tán của một nửa các số liệu, có giá trị thuộc đoạn từ Q_1 đến Q_3 trong mẫu.
- Khoảng tứ phân vị không bị ảnh hưởng bởi các giá trị rất lớn hoặc rất bé trong mẫu.

Ví dụ 2: Dưới đây là bảng số liệu thống kê của Biểu đồ nhiệt độ trung bình các tháng trong năm 2019 của hai tỉnh Lai Châu và Lâm Đồng .

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Lai Châu	14,8	18,8	20,3	23,5	24,7	24,2	23,6	24,6	22,7	21,0	18,6	14,2
Lâm Đồng	16,3	17,4	18,7	19,8	20,2	20,3	19,5	19,3	18,6	18,5	17,5	16,0

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của nhiệt độ trung bình mỗi tháng của tỉnh Lai Châu và Lâm Đồng.

b) Hãy cho biết trong một năm, nhiệt độ ở địa phương nào ít thay đổi hơn.

Lời giải

a)

* Tỉnh Lai Châu:

Sắp xếp các số liệu theo thứ tự không giảm, ta được:

14,2; 14,8; 18,6; 18,8; 20,3; 21,0; 22,7; 23,5; 23,6; 24,2; 24,6; 24,7.

+ Khoảng biến thiên của mẫu số liệu là: $R = 24,7 - 14,2 = 10,5$.

+ Cỡ mẫu là $n = 12$ là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = \frac{21,0 + 22,7}{2} = 21,85$

+ Tứ phân vị thứ nhất là trung vị của mẫu: 14,2; 14,8; 18,6; 18,8; 20,3; 21,0.

Do đó $Q_1 = \frac{18,6 + 18,8}{2} = 18,7$.

+ Tứ phân vị thứ ba là trung vị của mẫu: 22,7; 23,5; 23,6; 24,2; 24,6; 24,7.

Do đó $Q_3 = \frac{23,6 + 24,2}{2} = 23,9$.

+ Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 23,9 - 18,7 = 5,2$

* Tỉnh Lâm Đồng:

Sắp xếp các số liệu theo thứ tự không giảm, ta được:

16,0; 16,3; 17,4; 17,5; 18,5; 18,6; 18,7; 19,3; 19,5; 19,8; 20,2; 20,3.

+ Khoảng biến thiên của mẫu số liệu là: $R' = 20,3 - 16,0 = 4,3$.

+ Cỡ mẫu là $n = 12$ là số chẵn nên giá trị tứ phân vị thứ hai là: $Q'_2 = \frac{18,6 + 18,7}{2} = 18,65$

+ Tứ phân vị thứ nhất là trung vị của mẫu: 16,0; 16,3; 17,4; 17,5; 18,5; 18,6.

$$\text{Do đó } Q'_1 = \frac{17,4+17,5}{2} = 17,45.$$

+ Tứ phân vị thứ ba là trung vị của mẫu: 18,7; 19,3; 19,5; 19,8; 20,2; 20,3.

$$\text{Do đó } Q'_3 = \frac{19,5+19,8}{2} = 19,65.$$

+ Khoảng tứ phân vị của mẫu là: $\Delta'_Q = Q'_3 - Q'_1 = 19,65 - 17,45 = 2,2$.

b) Xét về cả khoảng biến thiên và khoảng tứ phân vị của nhiệt độ trung bình mỗi tháng của cả hai tỉnh, ta thấy: $10,5 > 4,3$ hay $R > R'$ và $5,2 > 2,2$ hay $\Delta_Q > \Delta'_Q$.

Điều đó có nghĩa là trong một năm, nhiệt độ ở Lâm Đồng ít thay đổi hơn.

c. Giá trị ngoại lệ

Khoảng tứ phân vị được dùng để xác định các **giá trị ngoại lệ** trong mẫu, đó là các giá trị quá nhỏ hay quá lớn so với đa số các giá trị của mẫu. Cụ thể, phần tử x trong mẫu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.

Sự xuất hiện của các giá trị ngoại lệ làm cho số trung bình và phạm vi của mẫu thay đổi lớn. Do đó, khi mẫu có giá trị ngoại lệ, người ta thường sử dụng trung vị và khoảng tứ phân vị để đo mức độ tập trung và mức độ phân tán của đa số các phần tử trong mẫu số liệu.

Ví dụ 3: Hãy tính giá trị ngoại lệ (nếu có) của mẫu số liệu: 10; 3; 5; 7; 20; 1; 4; 9.

Lời giải

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: 1; 3; 4; 5; 7; 9; 10; 20.

Khoảng biến thiên của mẫu số liệu là $R = 20 - 1 = 19$.

Cỡ mẫu là $n = 8$, là số chẵn nên giá trị tứ phân vị thứ hai là: $Q_2 = 6$.

Tứ phân vị thứ nhất là trung vị của mẫu: 10; 3; 5; 7. Do đó $Q_1 = 4$.

Tứ phân vị thứ 3 là trung vị của mẫu: 7; 9; 10; 20. Do đó $Q_3 = 9,5$.

Khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 9,5 - 4 = 5,5$.

$$Q_1 - 1,5\Delta_Q = 4 - 1,5 \cdot 5,5 = -4,25$$

$$Q_3 + 1,5\Delta_Q = 9,5 + 1,5 \cdot 5,5 = 17,75$$

Do đó, mẫu có một giá trị ngoại lệ là 20.

2. Phương sai và độ lệch chuẩn

a. Công thức tính phương sai và độ lệch chuẩn

• Giả sử ta có một mẫu số liệu là $x_1, x_2, \dots, x_n$.

Phương sai của mẫu số liệu này, kí hiệu là S^2 , được tính bởi công thức:

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

trong đó $\bar{x}$ là số trung bình của mẫu số liệu.

Căn bậc hai của phương sai được gọi là **độ lệch chuẩn**, kí hiệu là S .

Chú ý:

+ Có thể biến đổi công thức tính phương sai ở trên thành: $S^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$

+ Trong thống kê, người ta cũng quan tâm đến **phương sai hiệu chỉnh**, kí hiệu là $\hat{S}^2$, được tính bởi công thức: $\hat{S}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$

Ví dụ 4: Tính phương sai và độ lệch chuẩn của mẫu số liệu sau: 8; 10; 9; 7; 6; 10; 6; 7; 8; 9.

Lời giải

Cỡ mẫu $n = 10$.

Số trung bình: $\bar{x} = \frac{8+10+9+7+6+7+8+9}{10} = 8$

Phương sai mẫu số liệu là: $S^2 = \frac{8^2+10^2+9^2+7^2+6^2+7^2+8^2+9^2}{10} - 8^2 = 2$

Độ lệch chuẩn mẫu số liệu là $S = \sqrt{S^2} = \sqrt{2} \approx 1,41$.

• Giả sử mẫu số liệu được cho dưới dạng bảng tần số:

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Khi đó, công thức tính phương sai trở thành: $S^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{n}$

trong đó $n = n_1 + n_2 + \dots + n_k$.

Có thể biến đổi công thức tính phương sai trên thành: $S^2 = \frac{n_1x_1^2 + n_2x_2^2 + \dots + n_kx_k^2}{n} - \bar{x}^2$

Ví dụ 5: Điều tra số con của mỗi hộ gia đình trong tổ dân cư xóm 2, kết quả được ghi lại ở bảng sau:

Số con	0	1	2	3	4
Số hộ gia đình	4	4	8	3	1

Tính phương sai và độ lệch chuẩn của mẫu số liệu.

Lời giải

Tổng số hộ gia đình là: $n = 4 + 4 + 8 + 3 + 1 = 20$ (hộ gia đình).

Số trung bình của mẫu số liệu trên là: $\bar{x} = \frac{4.0+4.1+8.2+3.3+1.4}{20} = 1,65$

Phương sai của mẫu số liệu trên là: $S^2 = \frac{4.0^2+4.1^2+8.2^2+3.3^2+1.4^2}{20} - 1,65^2 = 1,2275$

Độ lệch chuẩn của mẫu số liệu trên là: $S = \sqrt{S^2} = \sqrt{1,2275} \approx 1,11$

b. Ý nghĩa của phương sai và độ lệch chuẩn

- Phương sai là trung bình cộng của các bình phương độ lệch từ mỗi giá trị của mẫu số liệu đến số trung bình.
- Phương sai và độ lệch chuẩn được dùng để đo mức độ phân tán của các số liệu trong mẫu quanh số trung bình. Phương sai và độ lệch chuẩn càng lớn thì các giá trị của mẫu càng cách xa nhau (có độ phân tán lớn).

Ví dụ 6: Bảng dưới đây thống kê tổng số giờ nắng trong năm 2019 theo từng tháng được đo bởi hai trạm quan sát khí tượng đặt ở Tuyên Quang và Cà Mau.

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Tuyên Quang	25	89	72	117	106	177	156	203	227	146	117	145
Cà Mau	180	223	257	245	191	111	141	134	130	122	157	173

(Nguồn: Tổng cục Thống kê)

- Hãy tính phương sai và độ lệch chuẩn của dữ liệu từng tỉnh.
- Nêu nhận xét về sự thay đổi tổng số giờ nắng theo từng tháng ở mỗi tỉnh.

Lời giải

a)* Tỉnh Tuyên Quang:

$$+ \text{Số trung bình: } \bar{x}_1 = \frac{25 + 89 + 72 + 117 + 106 + 177 + 156 + 203 + 227 + 146 + 117 + 145}{12} = 131,67$$

+ Phương sai mẫu số liệu ở tỉnh Tuyên Quang là:

$$S_1^2 = \frac{25^2 + 89^2 + 72^2 + 117^2 + 106^2 + 177^2 + 156^2 + 203^2 + 227^2 + 146^2 + 117^2 + 145^2}{12} - 131,67^2 = 2920,34$$

$$+ \text{Độ lệch chuẩn mẫu số liệu ở tỉnh Tuyên Quang là: } S_1 = \sqrt{S_1^2} = \sqrt{2920,34} \approx 54,04$$

* Tỉnh Cà Mau:

$$+ \text{Số trung bình: } \bar{x}_2 = \frac{180 + 223 + 257 + 245 + 191 + 111 + 141 + 134 + 130 + 122 + 157 + 173}{12} = 172$$

+ Phương sai mẫu số liệu ở tỉnh Cà Mau là:

$$S_2^2 = \frac{180^2 + 223^2 + 257^2 + 245^2 + 191^2 + 111^2 + 141^2 + 134^2 + 130^2 + 122^2 + 157^2 + 173^2}{12} - 172^2 = 2183$$

$$+ \text{Độ lệch chuẩn mẫu số liệu ở tỉnh Cà Mau là: } S_2 = \sqrt{S_2^2} = \sqrt{2183} \approx 46,72$$

b) Phương sai mẫu và độ lệch chuẩn mẫu số liệu ở tỉnh Tuyên Quang cao hơn tỉnh Cà Mau nên tổng giờ nắng trong năm 2019 theo từng tháng ở tỉnh Tuyên Quang có độ phân tán cao hơn ở tỉnh Cà Mau. Do đó, sự thay đổi tổng số giờ nắng theo từng tháng ở tỉnh Cà Mau ổn định (có ít sự thay đổi) hơn so với tỉnh Tuyên Quang.

PHẦN A
TỰ LUẬN PHÂN DẠNG TOÁN

DẠNG 1
TÌM KHOẢNG BIẾN THIÊN CHO MẪU SỐ LIỆU GHÉP NHÓM

Bài toán: Tính khoảng biến thiên của mẫu số liệu ghép nhóm sau:

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.

Phương pháp:

Xác định a_1 là giá trị đầu mút trái của nhóm đầu tiên và a_{k+1} là giá trị đầu mút phải của nhóm cuối cùng có chứa dữ liệu (tần số khác 0).

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

Bài 1. Bảng dưới thống kê thành tích nhảy xa của 58 học sinh lớp 12A.

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

- a) Tìm khoảng biến thiên thành tích nhảy xa của học sinh lớp 12A.
b) Bạn Khánh Linh ý kiến cho rằng: “Trong 58 học sinh lớp 12A trên, hiệu số thành tích nhảy xa của hai học sinh bất kì không vượt quá 160 cm”. Ý kiến đó đúng hay sai?

Lời giải

- a) Đầu mút phải của nhóm ghép cuối cùng là 300, đầu mút trái của nhóm ghép đầu tiên là 150.
Vậy khoảng biến thiên của mẫu số liệu là $R = 300 - 150 = 150$
b) Ta có khoảng biến thiên của mẫu số liệu là 150 nên hiệu số thành tích nhảy xa của hai học sinh bất kì không vượt quá 150 cm, do đó ý kiến của bạn Khánh Linh sai.

Bài 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

- a) Tìm khoảng biến thiên của mẫu số liệu đã cho.
b) Kết quả câu a) cho biết điều gì?

Lời giải

a) 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 $R = 180 - 30 = 150$.

b) Kết quả câu a) 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.

Bài 3. Bảng sau thống kê cân nặng của 50 quả xoài được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường

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

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

Lời giải

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

Do đó ý kiến nêu trên là đúng.

Bài 4. Thời gian hoàn thành bài kiểm tra môn Toán của các bạn trong lớp 12A 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 cho mẫu số liệu ghép nhóm trên.

b) Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 25 phút và muộn nhất mất 44 phút thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

Lời giải

a) Đầu mút phải của nhóm ghép cuối cùng là 45, đầu mút trái của nhóm ghép đầu tiên là 25.

Vậy khoảng biến thiên của mẫu số liệu là $R = 45 - 25 = 20$.

b) Khoảng biến thiên của mẫu số liệu gốc là: $R = 44 - 25 = 19$

Bài 5. Để chuẩn bị mở một cơ sở thể dục thể thao, Bác Minh đã tiến hành điều tra tuổi thọ của máy chạy bộ do hai hãng A, B sản xuất. Bảng dưới biểu thị hai mẫu số liệu mà Bác Minh thu thập được qua Internet, với tuổi thọ của máy chạy bộ tính theo đơn vị năm.

Tuổi thọ (năm)	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
Số máy của hãng A	7	20	36	20	17
Số máy của hãng B	0	20	35	35	10

a) Tính khoảng biến thiên của mẫu số liệu hai hãng A, B.

b) Hỏi máy chạy bộ do hãng nào sản xuất có tuổi thọ phân tán hơn?

Lời giải

a) Khoảng biến thiên của tuổi thọ máy chạy bộ do hãng A sản xuất là: $R_A = 12 - 2 = 10$ (năm)

Do trong mẫu số liệu ghép nhóm về của tuổi thọ máy chạy bộ do hãng B sản xuất, khoảng đầu tiên chứa dữ liệu là $[4; 6)$ và khoảng cuối cùng chứa dữ liệu là $[10; 12)$. Nên khoảng biến thiên của tuổi thọ máy chạy bộ do hãng B sản xuất là: $R_B = 12 - 4 = 8$ (năm)

b) Vì $R_A > R_B$ nên máy do hãng A sản xuất có tuổi thọ phân tán hơn so với máy của hãng B.

Bài 6. Thống kê thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12A và 12B được kết quả như bảng sau:

Thời gian sử dụng (phút)	$[0; 10)$	$[10; 30)$	$[30; 60)$	$[60; 90)$
Số học sinh lớp 12A	10	15	15	5
Số học sinh lớp 12B	25	5	15	0

Tìm khoảng biến thiên cho thời gian sử dụng mạng xã hội của học sinh mỗi lớp và giải thích ý nghĩa.

Lời giải

Khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12A là: $R_A = 90 - 0 = 90$

Do trong mẫu số liệu ghép nhóm về thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12B, khoảng đầu tiên chứa dữ liệu là $[0; 10)$ và khoảng cuối cùng chứa dữ liệu là $[30; 60)$. Nên khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12B là: $R_B = 60 - 0 = 60$

Do $R_A > R_B$ nên ta có thể kết luận rằng thời gian sử dụng mạng xã hội trong ngày của các bạn lớp 12A phân tán hơn thời gian sử dụng mạng xã hội của các bạn lớp 12B.

Bài 7. 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:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm A là: $R_A = 100 - 50 = 50$

Do trong mẫu số liệu ghép nhóm B khoảng đầu tiên chứa dữ liệu là $[60; 70)$ và khoảng cuối cùng chứa dữ liệu là $[80; 90)$. Nên khoảng biến thiên của mẫu số liệu ghép nhóm B là: $R_B = 90 - 60 = 30$

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

DẠNG 2

TÌM TỨ PHÂN VỊ CHO MẪU SỐ LIỆU GHÉP NHÓM

Bài toán: Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi bảng sau:

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

Phương pháp:

Tìm tứ phân vị Q_1 và Q_3 theo công thức:

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

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

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

Bài 1. Kết quả điều tra tổng thu nhập trong năm 2024 của một số hộ gia đình ở thành phố Nha Trang được ghi lại ở bảng sau:

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

a) Hãy tìm các tứ phân vị Q_1 và Q_3 .

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

Lời giải

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

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

Ta có:

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

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

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

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

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

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

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

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

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

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

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

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

Bài 2. Kết quả đo chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được cho ở bảng sau:

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

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

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

Lời giải

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

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

Ta có

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

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

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

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

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

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

liệu ghép nhóm là:

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

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

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

$$Q_3 = 20 + \frac{\frac{3.100}{4} - (5+12+25)}{44} (9,2 - 9,0) = 9,15$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 9,15 - 8,864 = 0,286$

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

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

Bài 3. Bạn Trang thống kê chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau:

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

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

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

Lời giải

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

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

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

$$R = 180 - 155 = 25 \text{ (cm)}.$$

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

b)

• Lớp 12C:

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

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

Ta có:

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

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

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

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

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

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

$$\text{liệu ghép nhóm là: } Q_1 = 160 + \frac{\frac{25}{4} - 2}{7} (165 - 160) = \frac{4565}{28}$$

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

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

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

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

• Lớp 12D:

ỡ mẫu $n' = 5 + 9 + 8 + 2 + 1 = 25$.

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

Ta có

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

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

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

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

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

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

liệu ghép nhóm là

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

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

ghép nhóm là

$$Q_3' = 165 + \frac{\frac{3.25}{4} - (5+9)}{8} (170 - 165) = \frac{5395}{32}$$

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

$$\Delta_Q' = Q_3' - Q_1' = \frac{5375}{32} - \frac{5785}{36} \approx 7,27$$

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

Bài 4. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

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

- a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.
- b) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?

Lời giải

a)

- Khu vực A:

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

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

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

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

Ta có

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

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

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

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

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

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

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

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

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

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

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

• Khu vực B:

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

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

Cỡ mẫu $n' = 47 + 40 + 11 + 2 = 100$.

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

Ta có

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

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

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

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

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

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

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

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

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

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

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

Bài 5. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

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

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

Lời giải

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

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

Khoảng biến thiên của mẫu số liệu trên là: $R = 522,9 - 147 = 375,9$ (mm).

Cỡ mẫu $n = 20$.

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

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

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

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

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

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

Khoảng tứ phân vị của mẫu số liệu đã cho là: $\Delta_Q = Q_3 - Q_1 = 417,25 - 254,9 = 162,35$

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

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

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

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

c) Cỡ mẫu $n = 20$.

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R' = 540 - 140 = 400$ (mm).

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

Ta có:

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

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

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

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

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

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

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

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

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

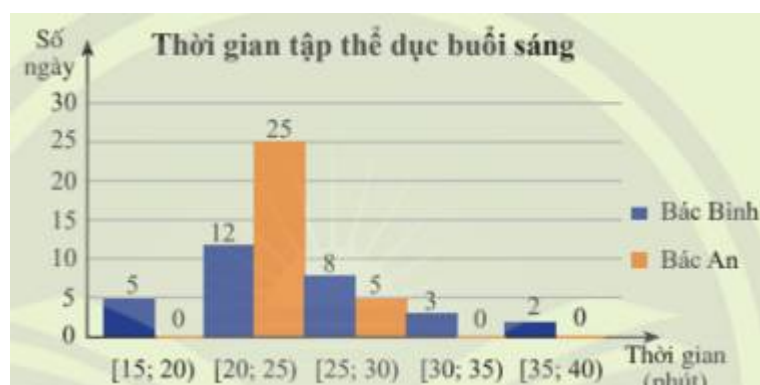
$$Q'_3 = 340 + \frac{\frac{3 \cdot 20}{4} - (3 + 7)}{7} (440 - 340) = \frac{2880}{7}$$

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

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

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

Bài 6. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng của Bác Bình và Bác An



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

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

Lời giải

Ta có bảng thống kê sau:

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

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

$$40 - 15 = 25 \text{ (phút)}.$$

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:

$$30 - 20 = 10 \text{ (phút)}.$$

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

b) Cỡ mẫu $n = 30$.

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

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

Ta có

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

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

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

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

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

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

$$\text{ghép nhóm là: } Q_1 = 20 + \frac{\frac{30}{4} - 5}{12} (25 - 20) = \frac{505}{24}$$

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

$$\text{nhóm là: } Q_3 = 25 + \frac{\frac{3 \cdot 30}{4} - (5 + 12)}{8} (30 - 25) = \frac{455}{16}$$

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

$$\text{là: } \Delta_Q = Q_3 - Q_1 = \frac{455}{16} - \frac{505}{24} = \frac{355}{48} \approx 7,4$$

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

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

Ta có

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

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

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

$$\text{ghép nhóm là: } Q'_1 = 20 + \frac{\frac{30}{4}}{25} (25 - 20) = \frac{43}{2}$$

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

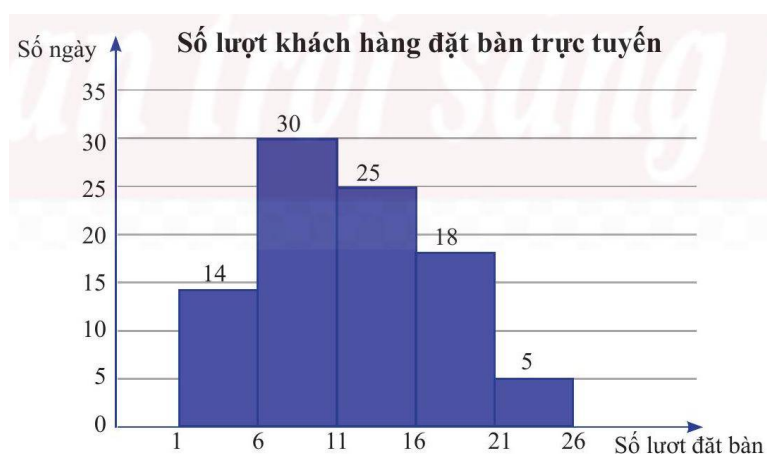
$$\text{nhóm là: } Q'_3 = 20 + \frac{\frac{3.30}{4}}{25} (25 - 20) = \frac{49}{2}$$

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

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

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

Bài 7. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2024 của một nhà hàng. Cột thứ nhất biểu diễn số ngày có từ 1 đến dưới 6 lượt đặt bàn; cột thứ hai biểu diễn số ngày có từ 6 đến dưới 11 lượt đặt bàn; ...



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

Lời giải

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

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

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

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

Ta có

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

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

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

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

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

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

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

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

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 16 - 7,5 = 8,5$

Bài 8. Hai bảng tần số ghép nhóm dưới đây thống kê theo độ tuổi số lượng thành viên nam và thành viên nữ đang sinh hoạt trong một câu lạc bộ dưỡng sinh.

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

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

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

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

Lời giải

a)

• Nam giới:

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

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

Ta có

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

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

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

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

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

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

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

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

$$\text{ghép nhóm là: } Q_1 = 60 + \frac{\frac{50}{4} - (4+7)}{4} (65 - 60) = 61,875$$

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về tuổi của nam giới đang sinh hoạt trong câu lạc bộ dưỡng sinh là: $\Delta_Q = Q_3 - Q_1 = 75,625 - 61,875 = 13,75$

• Nữ giới:

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

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

Ta có

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

$$y_5; \dots; y_7 \in [55; 60),$$

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

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

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

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

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

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

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

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

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

$$\text{nhóm là: } Q'_3 = 80 + \frac{\frac{3.50}{4} - (3+4+5+3+7+14)}{13} (85-80) = \frac{2095}{26}$$

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

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

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

DẠNG 3

TÍNH PHƯƠNG SAI, ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHEP NHÓM

Bài toán: Xét mẫu số liệu ghép nhóm cho bởi bảng sau:

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

Tính phương sai và độ lệch chuẩn của mẫu số liệu trên.

Phương pháp:

Bước 1: Tính giá trị đại diện của từng nhóm:

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

Nhóm	$[a_1; a_2)$	...	$[a_i; a_{i+1})$	...	$[a_k; a_{k+1})$
Giá trị đại diện	x_1	...	x_i	...	x_k
Tần số	m_1	...	m_i	...	m_k

Bước 2: Tính số trung bình cộng của mẫu số liệu ghép nhóm trên được tính bằng công thức:

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_k x_k}{n}$$

Trong đó: $n = m_1 + m_2 + \dots + m_k$ là cỡ mẫu.

Bước 3:

Tính **phương sai** của mẫu số liệu ghép nhóm theo công thức sau:

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

Hoặc công thức sau:
$$S^2 = \frac{1}{n} [m_1 (x_1 - \bar{x})^2 + m_2 (x_2 - \bar{x})^2 + \dots + m_k (x_k - \bar{x})^2]$$

Tính **độ lệch chuẩn** của mẫu số liệu ghép nhóm theo công thức sau: $S = \sqrt{S^2}$.

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

Ta có bảng sau:

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

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

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

Bài 2. Người ta theo dõi sự thay đổi cân nặng ở người nam và người nữ, được tính bằng hiệu cân nặng trước và sau ba tháng áp dụng chế độ ăn kiêng của một số người cho kết quả như sau:

Thay đổi cân nặng (kg)	[-1; 0)	[0; 1)	[1; 2)	[2; 3)	[3; 4)
Số người nam	2	3	5	3	2
Số người nữ	2	7	12	7	2

Tính số trung bình, phương sai, độ lệch chuẩn và nhận xét về sự thay đổi cân nặng của người nam, người nữ sau ba tháng áp dụng chế độ ăn kiêng.

Lời giải

Ta có bảng giá trị đại diện cho các nhóm số liệu trên

Thay đổi cân nặng (kg)	[-1; 0)	[0; 1)	[1; 2)	[2; 3)	[3; 4)
Giá trị đại diện	-0,5	0,5	1,5	2,5	3,5
Số người nam	2	3	5	3	2
Số người nữ	2	7	12	7	2

• Người nam

Tổng số người nam là: $n_1 = 2 + 3 + 5 + 3 + 2 = 15$.

Thay đổi cân nặng trung bình của người nam là:

$$\bar{x}_1 = \frac{1}{15} [2 \cdot (-0,5) + 3 \cdot 0,5 + 5 \cdot 1,5 + 3 \cdot 2,5 + 2 \cdot 3,5] = 1,5(kg).$$

Phương sai của mẫu số liệu về thay đổi cân nặng của người nam là:

$$s_1^2 = \frac{1}{15} [2 \cdot (-0,5)^2 + 3 \cdot 0,5^2 + 5 \cdot 1,5^2 + 3 \cdot 2,5^2 + 2 \cdot 3,5^2] - 1,5^2 \approx 1,4641$$

Độ lệch chuẩn của mẫu số liệu về thay đổi cân nặng của người nam là: $s_1 = \sqrt{s_1^2} \approx 1,21$

• Người nữ

Tổng số người nữ là: $n_2 = 2 + 7 + 12 + 7 + 2 = 30$.

Thay đổi cân nặng trung bình của người nữ là:

$$\bar{x}_2 = \frac{1}{30} [2 \cdot (-0,5) + 7 \cdot 0,5 + 12 \cdot 1,5 + 7 \cdot 2,5 + 2 \cdot 3,5] = 1,5(kg).$$

Phương sai chuẩn của mẫu số liệu về thay đổi cân nặng của người nữ là:

$$s_2^2 = \frac{1}{30} [2 \cdot (-0,5)^2 + 7 \cdot 0,5^2 + 12 \cdot 1,5^2 + 7 \cdot 2,5^2 + 2 \cdot 3,5^2] - 1,5^2 \approx 4,2436$$

Độ lệch chuẩn của mẫu số liệu về thay đổi cân nặng của người nữ là: $s_2 \approx 2,06$

Như vậy, sau ba tháng áp dụng chế độ ăn kiêng này, về trung bình sự thay đổi cân nặng của nam và nữ là như nhau. Tuy nhiên, sự biến động về thay đổi cân nặng của nữ nhiều hơn so với của nam.

Bài 3. Minh Hiền và Minh Nhân cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

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

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

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

Lời giải

a) Ta có bảng sau:

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

• Xét mẫu số liệu của Minh Hiền:

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

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_H = \frac{6.4 + 7.6 + 6.8 + 6.10 + 5.12}{30} = 7,8$

Phương sai của mẫu số liệu ghép nhóm là: $S_H^2 = \frac{1}{30} (6.4^2 + 7.6^2 + 6.8^2 + 6.10^2 + 5.12^2) - (7,8)^2 = 7,56$

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

• Xét mẫu số liệu của Minh Nhân:

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

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_N = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$

Phương sai của mẫu số liệu ghép nhóm là: $S_N^2 = \frac{1}{30} (2.4^2 + 5.6^2 + 13.8^2 + 8.10^2 + 2.12^2) - (8,2)^2 = 3,83$

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

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

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

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

a) Ta có bảng sau:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Bài 5. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của hai nhà đầu tư được cho như sau:

Bảng lợi nhuận theo tháng của nhà đầu tư nhỏ

Lợi nhuận (đơn vị: triệu đồng)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
Số tháng	2	4	8	4	2

Bảng lợi nhuận theo tháng của nhà đầu tư lớn

Lợi nhuận (đơn vị: triệu đồng)	[510; 520)	[520; 530)	[530; 540)	[540; 550)	[550; 560)
Số tháng	4	3	6	3	4

a) Tính độ lệch chuẩn của hai mẫu số liệu ghép nhóm trên.

b) Có nên dựa vào độ lệch chuẩn để so sánh độ rủi ro của hai nhà đầu tư này không?

Lời giải

• Giá trị đại diện lợi nhuận theo tháng của nhà đầu tư nhỏ

Lợi nhuận (đơn vị: triệu đồng)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
Giá trị đại diện	15	25	35	45	55
Số tháng	2	4	8	4	2

Lợi nhuận trung bình một tháng của các nhà đầu tư nhỏ là: $\bar{x}_A = \frac{1}{20} (2 \cdot 15 + \dots + 2 \cdot 55) = 35$ (triệu đồng)

Độ lệch chuẩn của lợi nhuận hàng tháng của nhà đầu tư nhỏ là:

$$s_A = \sqrt{\frac{1}{20}(2 \cdot 15^2 + \dots + 2 \cdot 55^2) - (35)^2} \approx 10,95$$

- Giá trị đại diện lợi nhuận theo tháng của nhà đầu tư lớn

Lợi nhuận (đơn vị: triệu đồng)	[510;520)	[520;530)	[530;540)	[540;550)	[550;560)
Giá trị đại diện	515	525	535	545	555
Số tháng	4	3	6	3	4

Lợi nhuận trung bình một tháng của các nhà đầu tư lớn là:

$$\bar{x}_B = \frac{1}{20}(4 \cdot 515 + \dots + 4 \cdot 555) = 535 \text{ (triệu đồng)}.$$

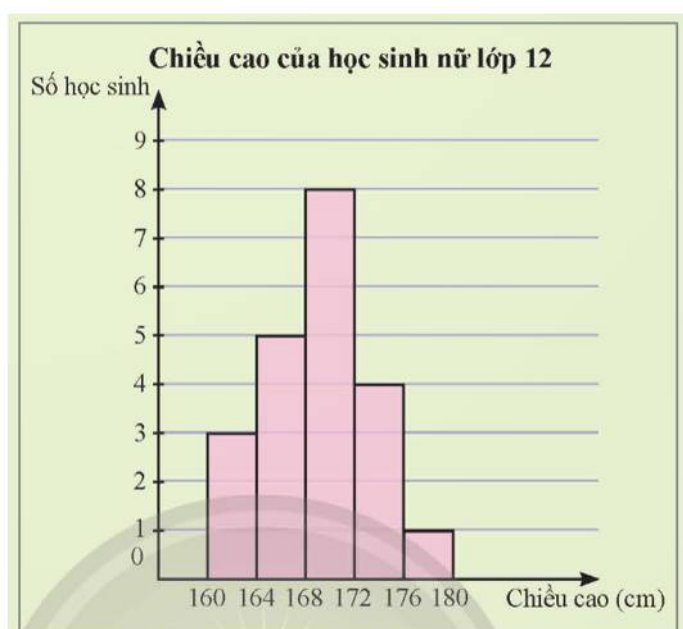
Độ lệch chuẩn của lợi nhuận hàng tháng của nhà đầu tư lớn là:

$$s_B = \sqrt{\frac{1}{20}(4 \cdot 515^2 + \dots + 4 \cdot 555^2) - (535)^2} \approx 13,78$$

Độ lệch chuẩn cho lợi nhuận hàng tháng của nhà đầu tư lớn cao hơn của nhà đầu tư nhỏ. Lợi nhuận trung bình của hai nhà đầu tư khác nhau rất nhiều, do đó ta không nên dùng độ lệch chuẩn để so sánh mức độ rủi ro của hai nhà đầu tư này.

Nhận xét. Ta không nên dùng phương sai hay độ lệch chuẩn để so sánh độ rủi ro của hai phương án đầu tư khi lợi nhuận trung bình của hai phương án đầu tư này khác nhau rất nhiều.

Bài 6. Biểu đồ sau biểu diễn chiều cao của học sinh nữ lớp 12.



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

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

Lời giải

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

Chiều cao (cm)	[160; 164)	[164; 168)	[168; 172)	[172; 176)	[176; 180)
Giá trị đại diện	162	166	170	174	178
Số học sinh	3	5	8	4	1

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

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{3.162 + 5.166 + 8.170 + 4.174 + 1.178}{21} = \frac{3550}{21}$

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

Chiều cao (cm)	[160; 164)	[164; 168)	[168; 172)	[172; 176)	[176; 180)
Giá trị đại diện	162	166	170	174	178
Số học sinh	3	5	8	4	1

Cỡ mẫu $n = 21$

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

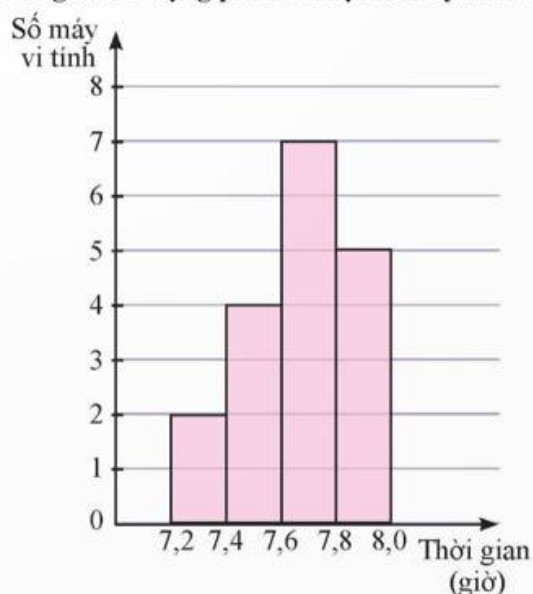
Phương sai của mẫu số liệu ghép nhóm là:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $s = \sqrt{s^2} = \sqrt{\frac{8000}{441}} \approx 4,26$

Bài 7. Kết quả khảo sát thời gian sử dụng liên tục (đơn vị: giờ) từ lúc sạc đầy cho đến khi hết của pin một số máy vi tính cùng loại được mô tả bằng biểu đồ bên.

Thời gian sử dụng pin của một số máy vi tính



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

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

Lời giải

a) Từ biểu đồ ta thấy có 2 máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ.

b) Từ biểu đồ, ta có bảng thống kê sau:

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

Cỡ mẫu là $n = 2 + 4 + 7 + 5 = 18$

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{2 \cdot 7,3 + 4 \cdot 7,5 + 7 \cdot 7,7 + 5 \cdot 7,9}{18} = \frac{23}{3}$

Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

Khoảng biến thiên của mẫu số liệu là: $R = 61,1 - 42 = 19,1$ (km/h).

Cỡ mẫu $n = 20$

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

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

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

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

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

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

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

Số trung bình của mẫu số liệu là: $\bar{x} = \frac{42 + 43,4 + \dots + 60,3 + 61,1}{20} = 50,945$

Phương sai của mẫu số liệu là: $S^2 = \frac{1}{20} \left[42^2 + (43,4)^2 + \dots + (60,3)^2 + (61,1)^2 \right] - (50,945)^2 \approx 32,2$

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

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

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

c) Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R' = 62 - 42 = 20$ (km/h).

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

Ta có

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

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

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

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

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

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

$$\text{ghép nhóm là: } Q'_1 = 46 + \frac{\frac{20}{4} - 3}{7} (50 - 46) = \frac{330}{7}$$

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

$$\text{ghép nhóm là: } Q'_3 = 54 + \frac{\frac{3 \cdot 20}{4} - (3 + 7 + 4)}{3} (58 - 54) = \frac{166}{3}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta'_Q = Q'_3 - Q'_1 = \frac{166}{3} - \frac{330}{7} \approx 8,19$

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

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

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}' = \frac{3.44 + 7.48 + 4.52 + 3.56 + 3.60}{20} = 51,2$

Phương sai của mẫu số liệu ghép nhóm là:

$$(S')^2 = \frac{1}{20} [3.44^2 + 7.48^2 + 4.52^2 + 3.56^2 + 3.60^2] - (51,2)^2 = 26,56$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S' = \sqrt{(S')^2} = \sqrt{26,56} \approx 5,154$

PHẦN B

TRẮC NGHIỆM VÀ TỰ LUẬN TỔNG HỢP GỒM BỐN PHẦN

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

Câu 1. 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 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.

Câu 3. Trong các khẳng định sau, khẳng định nào sai?

A. Phương sai là bình phương của độ lệch chuẩn.

B. Phương sai luôn luôn là số không âm.

C. Phương sai luôn luôn lớn hơn độ lệch chuẩn.

D. Phương sai càng lớn thì độ phân tán của các giá trị quanh số trung bình càng lớn.

Lời giải

Chọn C.

Khi $x \in (0;1)$ thì $s^2 < s$ nên khẳng định phương sai luôn lớn hơn độ lệch chuẩn là sai.

Câu 4. Thời gian hoàn thành bài kiểm tra môn Toán của các học sinh trong lớp 12A được cho trong bảng sau:

Thời gian (phút)	[25;30)	[30;35)	[35;40)	[40;45)
Số học sinh	8	16	12	2

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 20.

B. 30.

C. 27.

D. 35.

Lời giải

Chọn A.

Khoảng biến thiên: $R = 45 - 25 = 20$.

Câu 5. Thời gian hoàn thành giải chạy của các vận động viên được cho như bảng sau:

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)
Số vận động viên	5	10	8	2

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 10.

B. 25.

C. 30.

D. 20.

Lời giải

Chọn D.

Khoảng biến thiên: $R = 35 - 15 = 20$.

Câu 6. Bảng dưới biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 120 khách hàng mua sách ở một cửa hàng trong một ngày.

Nhóm	[40; 50)	[50; 60)	[60; 70)	[70; 80)	[80; 90)
Tần số	6	12	38	46	18

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

A. 30.

B. 40.

C. 50.

D. 60.

Lời giải

Chọn C.

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

Câu 7. Mức thưởng tết (triệu đồng) cho các nhân viên của một công ty được thống kê trong bảng sau:

Mức thưởng tết	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
Số nhân viên	10	5	6	4	5

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

A. 20.

B. 25.

C. 47.

D. 23

Lời giải

Chọn B.

Khoảng biến thiên: $R = 30 - 5 = 25$

Câu 8. Chiều cao của 80 học sinh ở một trường Đại Học Thủ Đức Thủ Đức được thống kê trong bảng sau:

Chiều cao (cm)	[160; 163,5)	[164; 167,5)	[168; 171,5)	[172; 175,5)
Tần số	18	40	14	8

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 31. B. 15,5. C. 74. D. 32.

Lời giải

Chọn B.

Khoảng biến thiên: $R = 175,5 - 160 = 15,5$

Câu 9. Thời gian dùng điện thoại 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	0	24	30	48	52

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A. 26. B. 15. C. 20. D. 12.

Lời giải

Chọn D.

Khoảng biến thiên: $R = 24,5 - 12,5 = 12.$

Câu 10. Một Bác nông dân thống kê lại đường kính thân gỗ của một số cây xoan đào 7 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	9	18	18	10	0

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

- A. 25. B. 30. C. 20. D. 35.

Lời giải

Chọn C.

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là $R = 60 - 40 = 20(cm).$

Câu 11. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

Tuổi thọ (năm)	[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: $R = 19 - 14 = 5$

Câu 12. 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. $Q_3 + Q_1 - Q_2.$

Chọn C.

Lý thuyết: $Q_3 - Q_1$

Câu 13. Khảo sát thời gian tập thể dục trong ngày của một số học sinh khối 12 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 12 và 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 14. 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

Ta có cỡ mẫu là $n = 20$.

Gọi $x_1, x_2, \dots, x_{20}$ là doanh thu bán hàng trong 20 ngày của một cửa hàng và 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 10 số liệu đầu nên $Q_1 = \frac{1}{2}(x_4 + x_5)$.

Do 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

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

Câu 15. 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 của mẫu số

$$\text{liệu là } Q_3 = 9 + \frac{\frac{3 \cdot 20}{4} - 9}{7} (11 - 9) = 10,7$$

Câu 16. 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

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

A. 20,67 .

B. 17,24 .

C. 18,24 .

D. 22,67 .

Lời giải

Chọn C.

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:

$$\begin{aligned} x_1, \dots, x_{22} &\in [15; 18); \\ x_{23}, \dots, x_{60} &\in [18; 21); \\ x_{61}, \dots, x_{87} &\in [21; 24); \\ x_{88}, \dots, x_{95} &\in [24; 27); \\ x_{96}, \dots, x_{99} &\in [27; 30); \\ x_{100} &\in [30; 33) \end{aligned}$$

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

$$\text{liệu ghép nhóm là: } Q_1 = 18 + \frac{\frac{100}{4} - 22}{38} \cdot (21 - 18) = \frac{693}{38} \approx 18,24$$

Câu 17. 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

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

- A. 18,24. B. 17,24. **C. 22,67.** D. 20,67.

Lời giải

Chọn C.

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:

$$\begin{aligned} x_1, \dots, x_{22} &\in [15;18); \\ x_{23}, \dots, x_{60} &\in [18;21); \\ x_{61}, \dots, x_{87} &\in [21;24); \\ x_{88}, \dots, x_{95} &\in [24;27); \\ x_{96}, \dots, x_{99} &\in [27;30); \\ x_{100} &\in [30;33) \end{aligned}$$

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

$$\text{ghép nhóm là: } Q_3 = 21 + \frac{\frac{3 \cdot 100}{4} - (22 + 38)}{27} \cdot (24 - 21) = \frac{68}{3} \approx 22,67$$

Câu 18. 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 ta có

$$Q_3 = 18,5 + \frac{\frac{3.56}{4} - 30}{24} \cdot 3 = 20.$$

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

Câu 19. 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, $p = 3; a_3 = 160; m_3 = 35; m_1 + m_2 = 18 + 28 = 46; a_4 - a_3 = 165 - 160 = 5$ và ta có

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

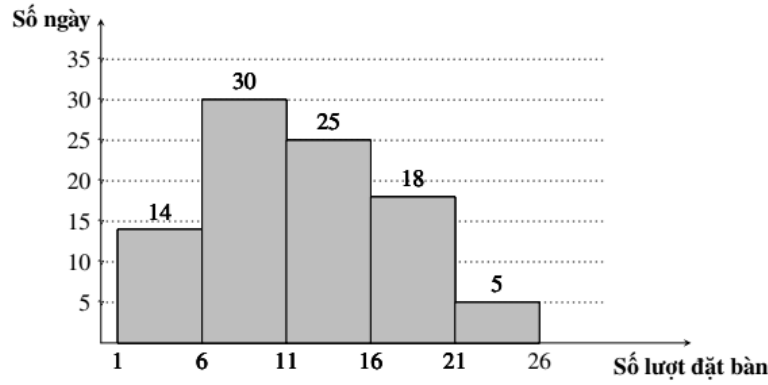
Tứ phân vị thứ ba Q_3 là $\frac{x_{150} + x_{151}}{2}$. Do x_{150}, x_{151} đều thuộc nhóm $[170; 175)$ nên nhóm này chứa Q_3 . Do

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

$$Q_3 = 170 + \frac{\frac{3.200}{4} - 124}{41} \cdot 5 \approx 173,17$$

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

Câu 20. Biểu đồ dưới đây biểu diễn số lượng 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 6 lượt đặt bàn, cột thứ hai biểu diễn số ngày có từ 6 đến 11 lượt đặt bàn;...



Khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi biểu đồ trên

A. 9,5.

B. 8,5.

C. 10,5.

D. 7,5.

Lời giải

Chọn B.

Dựa vào biểu đồ ta lập được bảng ghép nhóm như sau:

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

Cỡ mẫu $n = 92$ và gọi $x_1, x_2, \dots, x_{92}$ là mẫu số liệu đã cho

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

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

Tứ phân vị thứ ba của mẫu số liệu là $\frac{x_{69} + x_{70}}{2} \in [11;16)$ và $x_{70} \in [16;21)$. Do đó tứ phân vị thứ ba của mẫu số liệu là $Q_3 = 16$.

Vậy khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = 16 - 7,5 = 8,5$.

Câu 21. 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 22. Trong buổi thực hành, 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

Tính số trung bình của mẫu số liệu trên.

- A. 4,44 . B. 5,45 . C. 4,45 . D. 6,44 .

Lời giải

Chọn B.

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

Cỡ mẫu: $n = 3 + 8 + 7 + 2 = 20$

Số trung bình: $\bar{x} = \frac{4,75.3 + 5,25.8 + 5,75.7 + 6,25.2}{20} = \frac{109}{20} = 5,45(\mu m)$

Câu 23. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư vào hai lĩnh vực A cho kết quả như sau:

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

Tính số trung bình của mẫu số liệu trên.

- A. 17,9 (triệu đồng). B. 19,5 (triệu đồng). C. 16,9 (triệu đồng). D. 18,5 (triệu đồng).

Lời giải

Chọn D.

Ta có mẫu số liệu ghép nhóm với giá trị đại diện là:

Tiền lãi (triệu đồng)	[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

Tiền lãi trung bình khi đầu tư vào lĩnh vực A là:

$$\bar{x}_A = \frac{7,5 \cdot 2 + 12,5 \cdot 5 + 17,5 \cdot 8 + 22,5 \cdot 6 + 27,5 \cdot 4}{2 + 5 + 8 + 6 + 4} = 18,5 \text{ (triệu đồng)}$$

Câu 24. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư vào lĩnh vực B cho kết quả như sau:

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

Tính số trung bình của mẫu số liệu trên.

- A. 17,9 (triệu đồng). B. 19,5 (triệu đồng). C. 16,9 (triệu đồng). D. 18,5 (triệu đồng).

Lời giải

Chọn C.

Ta có mẫu số liệu ghép nhóm với giá trị đại diện là:

Tiền lãi (triệu đồng)	[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 B	8	4	2	5	6

Tiền lãi trung bình khi đầu tư vào lĩnh vực B là:

$$\bar{x}_B = \frac{7,5 \cdot 8 + 12,5 \cdot 4 + 17,5 \cdot 2 + 22,5 \cdot 5 + 27,5 \cdot 6}{8 + 4 + 2 + 5 + 6} = 16,9 \text{ (triệu đồng)}$$

Câu 25. Anh Bảo Toàn đầu tư số tiền vào lĩnh vực kinh doanh A. Anh Bảo Toàn thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực A	5	10	30	10	5

Độ lệch chuẩn của số tiền thu được hàng tháng khi anh Bảo Toàn đầu tư vào lĩnh vực A là:

- A. 4,5. B. 5. C. 5,5. D. 6.

Lời giải

Chọn B.

Chọn giá trị đại diện cho các nhóm số liệu ta có:

Số tiền (triệu đồng)	[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 tháng đầu tư vào lĩnh vực A	5	10	30	10	5

Số tiền trung bình thu được khi Anh Bảo Toàn đầu tư vào các lĩnh vực A là:

$$\bar{x}_A = \frac{1}{60}(5 \cdot 7,5 + \dots + 5 \cdot 27,5) = 17,5$$

Độ lệch chuẩn của số tiền thu được hàng tháng khi Anh Bảo Toàn đầu tư vào các lĩnh vực A là:

$$s_A = \sqrt{\frac{1}{60}(5 \cdot 7,5^2 + \dots + 5 \cdot 27,5^2) - (17,5)^2} = 5$$

Câu 26. Cô Thúy đầu tư số tiền vào lĩnh vực kinh doanh B. Cô Thúy thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực B	20	5	10	5	20

Độ lệch chuẩn của số tiền thu được hàng tháng khi Cô Thúy đầu tư vào lĩnh vực B là:

- A.** 8,42. **B.** 9,22. **C.** 7,42. **D.** 7,22.

Lời giải

Chọn A.

Chọn giá trị đại diện cho các nhóm số liệu ta có:

Số tiền (triệu đồng)	[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 tháng đầu tư vào lĩnh vực B	20	5	10	5	20

Số tiền trung bình thu được khi Cô Thúy đầu tư vào các lĩnh vực B là:

$$\bar{x}_B = \frac{1}{60}(20 \cdot 7,5 + \dots + 20 \cdot 27,5) = 17,5$$

Độ lệch chuẩn của số tiền thu được hàng tháng khi Cô Thúy đầu tư vào các lĩnh vực B là:

$$s_B = \sqrt{\frac{1}{60}(20 \cdot 7,5^2 + \dots + 20 \cdot 27,5^2) - (17,5)^2} \approx 8,42.$$

Câu 27. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, bạn Minh Hiền đã dùng hai vôn kế khác nhau để đo 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 Minh Hiền đo	1	6	2	1

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm cho kết quả đo của bạn Minh Hiền.

A. 0,039. B. 0,0015. C. 0,024. D. 0,0025.

Lời giải

Chọn A.

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 Minh Hiền đo	1	6	2	1

Giá trị trung bình: $\bar{x}_1 = \frac{1}{10}(3,875.1 + 3,925.6 + 3,975.2 + 4,025.1) = 3,94$

Phương sai của mẫu số liệu: $s_1^2 = \frac{1}{10}(3,875^2.1 + 3,925^2.6 + 3,975^2.2 + 4,025^2.1) - 3,94^2 = 0,001525$

Độ lệch chuẩn của mẫu số liệu: $s_1 = \sqrt{0,001525} \approx 0,039$

Câu 28. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, bạn Trúc Linh đã dùng hai vôn kế khác nhau để đo 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 Trúc Linh đo	1	3	4	2

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm cho kết quả đo của bạn Trúc Linh

A. 0,002. B. 0,012. C. 0,045. D. 0,055.

Lời giải

Chọn C.

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 Trúc Linh đo	1	3	4	2

Giá trị trung bình: $\bar{x}_1 = \frac{1}{10}(3,875.1 + 3,925.3 + 3,975.4 + 4,025.2) = 3,96$

Phương sai của mẫu số liệu: $s_2^2 = \frac{1}{10}(3,875^2.1 + 3,925^2.3 + 3,975^2.4 + 4,025^2.2) - 3,96^2 = 0,002025$

Độ lệch chuẩn của mẫu số liệu: $s_2 = \sqrt{0,002025} \approx 0,045$

Câu 29. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi phân xưởng A 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 A	4	9	13	8	6

Tính phương sai của mẫu số liệu ghép nhóm trên

A. 0,145.

B. 0,245.

C. 0,445.

D. 0,355.

Lời giải

Chọn D.

Ta có bảng số liệu với giá trị đại diện:

Tuổi thọ (năm)	[1,5; 2)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)
Giá trị đại diện	1,75	2,25	2,75	3,25	3,75
Số linh kiện của phân xưởng A	4	9	13	8	6

Tổng số linh kiện: $n_1 = 4 + 9 + 13 + 8 + 6 = 40$

Giá trị trung bình $\bar{x}_1 = \frac{1,75 \cdot 4 + 2,25 \cdot 9 + 2,75 \cdot 13 + 3,25 \cdot 8 + 3,75 \cdot 6}{4 + 9 + 13 + 8 + 6} = \frac{223}{80}$

Phương sai: $s_1^2 = \frac{1}{40} (1,75^2 \cdot 4 + 2,25^2 \cdot 9 + 2,75^2 \cdot 13 + 3,25^2 \cdot 8 + 3,75^2 \cdot 6) - \left(\frac{223}{80} \right)^2 = \frac{2271}{6400} \approx 0,355$

Câu 30. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi phân xưởng B 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 B	2	8	20	7	3

Tính phương sai của mẫu số liệu ghép nhóm trên

A. 0,219.

B. 0,329.

C. 0,119.

D. 0,129.

Lời giải

Chọn A.

Ta có bảng số liệu với giá trị đại diện:

Tuổi thọ (năm)	[1,5; 2)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)
Giá trị đại diện	1,75	2,25	2,75	3,25	3,75
Số linh kiện của phân xưởng B	2	8	20	7	3

Tổng số linh kiện: $n_2 = 2 + 8 + 20 + 7 + 3 = 40$

Giá trị trung bình $\bar{x}_2 = \frac{1,75 \cdot 2 + 2,25 \cdot 8 + 2,75 \cdot 20 + 3,25 \cdot 7 + 3,75 \cdot 3}{2 + 8 + 20 + 7 + 3} = \frac{221}{80}$

Phương sai: $s_2^2 = \frac{1}{40} (1,75^2 \cdot 2 + 2,25^2 \cdot 8 + 2,75^2 \cdot 20 + 3,25^2 \cdot 7 + 3,75^2 \cdot 3) - \left(\frac{221}{80} \right)^2 = \frac{1399}{6400} \approx 0,219$

sai.

Câu 31. Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm này là 5

b) Nhóm chứa tứ phân vị thứ nhất là nhóm [15;16)

c) Nhóm chứa tứ phân vị thứ ba là nhóm [17;18)

d) Nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4 thì khoảng biến thiên không thay đổi.

Lời giải

a)	b)	c)	d)
ĐÚNG	SAI	ĐÚNG	ĐÚNG

a) Khoảng biến thiên $R = 19 - 14 = 5$

b) Cỡ mẫu là: $1 + 3 + 8 + 6 + 2 = 20$.

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

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

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

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

Câu 32. 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

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

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

c) Trung vị của mẫu số liệu là $M_e \approx 9,95$

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 $14,79m^3$ nước trở lên.

Lời giải

a)	b)	c)	d)
ĐÚNG	ĐÚNG	SAI	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 $11,79m^3$ nước trở lên.

Câu 33. Hằng ngày bạn Minh Hiền đều đi xe buýt từ nhà đến trường học. Dưới đây là bảng thống kê thời gian của 100 lần bạn Minh Hiền đi xe buýt từ nhà đến trường.

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

a) Cỡ mẫu $n = 100$.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = \frac{983}{37}$.

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta_Q = \frac{615}{113}$

d) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần bạn Minh Hiền 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)	b)	c)	d)
ĐÚNG	SAI	SAI	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 bạn Minh Hiền.

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

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

c) 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

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

$$\text{Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là } \Delta_Q = \frac{68}{3} - \frac{693}{38} = \frac{505}{114}$$

d) Trong lần duy nhất bạn Minh Hiền đi hết hơn 29 phút, thời gian đi của ông thuộc nhóm $[30; 33)$. Vì

$$Q_3 + 1,5\Delta_Q = \frac{6683}{228} < 30 \text{ nên thời gian của lần bạn Minh Hiền đi hết hơn 29 phút là giá trị ngoại lệ của}$$

mẫu số liệu ghép nhóm.

Câu 34. 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) 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)	b)	c)	d)
ĐÚNG	ĐÚNG	SAI	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 nhất của mẫu

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

c) 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 mẫu số

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

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

d) 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 $x > 10,5 + 1,5.1,06 = 12,09$ hoặc $x < 9,44 - 1,5.1,06 = 7,85$

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

Câu 35. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

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

a) 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: 15 (tuổi)

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm ứng với khu vực A là: $\frac{721}{25}$ (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)	b)	c)	d)
ĐÚNG	SAI	SAI	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 nhất của mẫu số

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

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm A 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 nhất của mẫu số

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

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

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

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

Câu 36. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 10 từ năm 2005 đến 2024 tại một trạm quan trắc đặt ở Nha Trang.

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

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

b) Khoảng tứ phân vị của mẫu số liệu là: $\Delta_Q = \frac{17221}{100}$

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

Khi đó, khoảng biến thiên của mẫu số liệu ghép nhóm là 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

Khi đó, khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\frac{1100}{7}$

Lời giải

a)	b)	c)	d)
ĐÚNG	SAI	ĐÚNG	SAI

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

b) 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: $\Delta_Q = Q_3 - Q_1 = \frac{18177}{100}$

c)

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: $540 - 140 = 400(mm)$

d)

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

Cỡ mẫu $n = 20$;

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về lượng mưa đo được vào tháng 10 từ năm 2005 đến 2024 tại một trạm quan trắc đặt ở Nha Trang 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

$$\text{liệu ghép nhóm là: } Q_3' = 340 + \frac{\frac{3.20}{4} - (3+7)}{7} (440 - 340) = \frac{2880}{7}$$

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

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

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 0,6.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là 0,575.

c) Số trung bình của mẫu số liệu ghép nhóm là 3,62.

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm 0,36 (kết quả làm tròn đến phần trăm).

Lời giải

a)	b)	c)	d)
SAI	ĐÚNG	SAI	ĐÚNG

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

b) Cỡ mẫu $n = 20$.

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

Ta có

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

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

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

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

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

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

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

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

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

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

c) Ta có bảng giá trị đại diện sau:

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

d) Phương sai của mẫu số liệu ghép nhóm là:

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

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

Câu 38. Cô Minh Hiền rất thích luyện tập yoga. Thời gian tập mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 25.

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là 8,125.

c) Số trung bình của mẫu số liệu ghép nhóm là $\frac{85}{3}$.

d) Phương sai của mẫu số liệu ghép nhóm là 31,25.

Lời giải

a)	b)	c)	d)
ĐÚNG	ĐÚNG	ĐÚNG	ĐÚNG

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

b) Cỡ mẫu $n = 6 + 6 + 4 + 1 + 1 = 18$.

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

Ta có

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

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

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

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

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

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

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

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 số liệu ghép

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

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

c) Ta có bảng giá trị đại diện sau:

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

d) Phương sai của mẫu số liệu ghép nhóm là:

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

Câu 39. Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

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

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là 12.

b) Tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là $Q_1 = 10,75$

c) Tứ phân vị thứ ba của mẫu số liệu ghép nhóm là: $Q_3 = 14,375$

d) Độ lệch chuẩn của mẫu số liệu ghép nhóm là 2,74 (kết quả làm tròn đến phần trăm).

Lời giải

a)	b)	c)	d)
SAI	ĐÚNG	ĐÚNG	SAI

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 18 - 8 = 10$ (giây).

b) Cỡ mẫu $n = 25$.

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

Ta có

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

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

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

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

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

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

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

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

$$\text{ghép nhóm là: } Q_3 = 10 + \frac{\frac{3.25}{4} - (4 + 6 + 8)}{4} (16 - 14) = 14,375$$

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

d) Ta có bảng giá trị đại diện sau:

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

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{4.9 + 6.11 + 8.13 + 4.15 + 3.17}{25} = 12,68$$

Phương sai của mẫu số liệu ghép nhóm là:

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

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

Câu 40. Một bác tài xế thống kê lại độ dài quãng đường (đơn vị: km) bác đã lái xe mỗi ngày trong một tháng ở bảng sau:

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

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

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

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

b) Cỡ mẫu $n = 5 + 10 + 9 + 4 + 2 = 30$.

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

Ta có

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

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

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

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

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

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 ghép

$$\text{nhóm là: } Q_3 = 100 + \frac{\frac{3.30}{4} - (5+10)}{9} (200 - 150) = \frac{575}{3}$$

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

c) Ta có bảng giá trị đại diện sau:

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

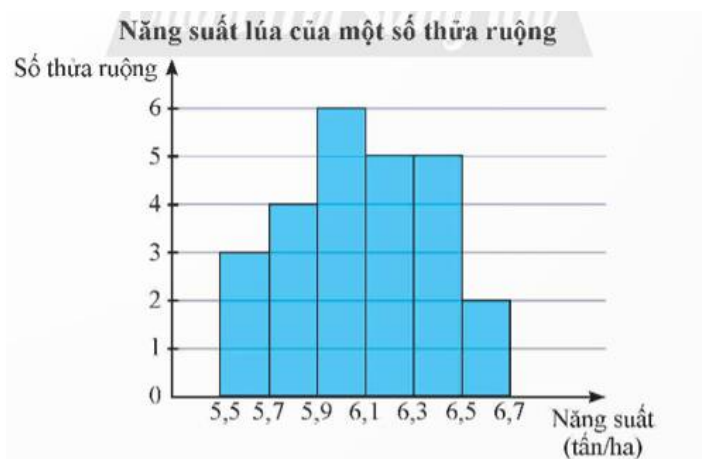
Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{5.75 + 10.125 + 9.175 + 4.225 + 2.275}{30} = 155$.

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{30} [5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2] - 155^2 = 3100$$

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

Câu 41. Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau:



a) Có 18 thửa ruộng đã được khảo sát.

b) Khoảng biến thiên của mẫu số liệu trên là 1,4(tấn/ha).

c) Khoảng tứ phân vị của mẫu số liệu ghép nhóm trên là 0,4675.

d) Phương sai của mẫu số liệu ghép nhóm trên là 0,086656.

Lời giải

A.	B.	C.	D.
SAI	SAI	ĐÚNG	ĐÚNG

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

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

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

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

c) Cỡ mẫu $n = 25$.

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

Ta có

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

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

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

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

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

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

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

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

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

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

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

d) Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{3 \cdot 5,6 + 4 \cdot 5,8 + 6 \cdot 6,0 + 5 \cdot 6,2 + 5 \cdot 6,4 + 2 \cdot 6,6}{25} = 6,088$.

Phương sai của mẫu số liệu ghép nhóm là:

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

Câu 42. Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

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

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

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

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

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

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Ta có bảng sau:

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

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

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

$$\bar{x}_X = \frac{6.6,5 + 10.7,5 + 13.8,5 + 10.9,5 + 9.10,5}{50} = 8,54$$

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

$$\bar{x}_Y = \frac{4.6,5 + 12.7,5 + 17.8,5 + 14.9,5 + 3.10,5}{50} = 8,5$$

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

b)

• Xét mẫu số liệu của học sinh trường X:

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

Ta có

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

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

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

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

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

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

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

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

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

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

• Xét mẫu số liệu của học sinh trường Y:

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

Ta có

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

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

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

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

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

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

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

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

$$Q'_3 = 7 + \frac{\frac{3.50}{4} - (4+12+17)}{14} (10-9) = \frac{261}{28}$$

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

$$\Delta_Q = Q_3 - Q_1 = \frac{261}{28} - \frac{185}{24} \approx 1,61$$

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

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

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

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

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

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

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

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

Câu 43. Bảng sau thống kê lại tổng số giờ nắng trong tháng 6 của các năm từ 2002 đến 2021 tại hai trạm quan trắc đặt ở Nha Trang và Quy Nhơn.

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

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

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

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

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

Lời giải

A.	B.	C.	D.
SAI	ĐÚNG	SAI	ĐÚNG

a) Cỡ mẫu $n = 20$.

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

Gọi $x_1; \dots; x_{20}$ là mẫu số liệu gốc về tổng số giờ nắng trong tháng 6 của các năm 2022 đến 2021 tại trạm quan trắc đặt ở Nha Trang được xếp theo thứ tự không giảm.

Ta có

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

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

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

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

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

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

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

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

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

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

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

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

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

Ta có

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

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

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

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

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

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

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

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

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

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

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

c) Ta có bảng giá trị đại diện sau:

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

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

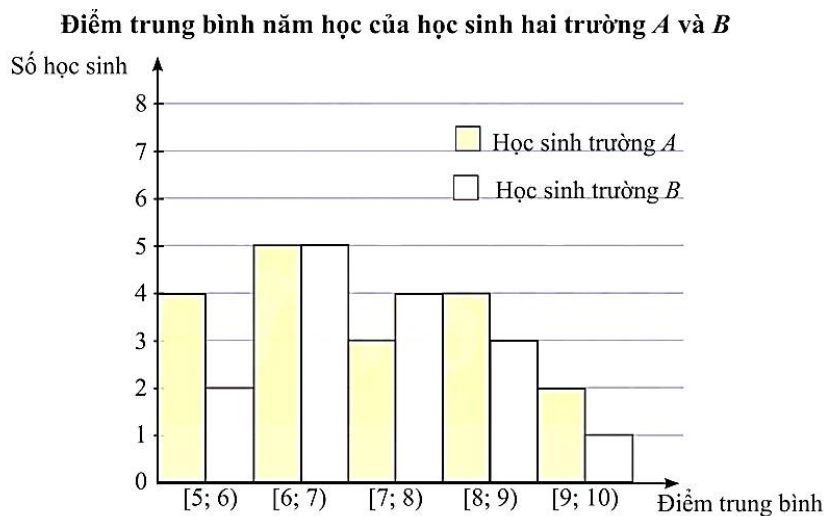
Phương sai của mẫu số liệu ghép nhóm là:

$$S_Q^2 = \frac{1}{20} [0.145^2 + 1.175^2 + 2.205^2 + 4.235^2 + 10.265^2 + 3.295^2] - (253)^2 = 936$$

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

d) Vì $S_N \approx 35,54 > S_N \approx 30,59$ nên nếu so sánh theo độ lệch chuẩn thì số giờ nắng trong tháng 6 của Quy Nhơn đồng đều hơn.

Câu 44. Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



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

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

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

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

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

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

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

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

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

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

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

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

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

b)

• Xét mẫu số liệu của trường A:

Cỡ mẫu $n_A = 4 + 5 + 3 + 4 + 2 = 18$.

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

Ta có

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

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

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

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

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

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

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

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

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

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

- Xét mẫu số liệu của trường B:

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

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

Ta có

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

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

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

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

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

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

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

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

$$\text{nhóm là: } Q'_3 = 8 + \frac{\frac{3 \cdot 15}{4} - (2 + 5 + 4)}{3} (9 - 8) = \frac{97}{12}$$

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

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

d)

- Xét mẫu số liệu của trường A:

$$\text{Số trung bình của mẫu số liệu ghép nhóm là: } \bar{x}_A = \frac{4 \cdot 5,5 + 5 \cdot 6,5 + 3 \cdot 7,5 + 4 \cdot 8,5 + 2 \cdot 9,5}{18} = \frac{65}{9}.$$

Phương sai của mẫu số liệu ghép nhóm là:

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

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

- Xét mẫu số liệu của trường B:

$$\text{Số trung bình của mẫu số liệu ghép nhóm là: } \bar{x}_B = \frac{2 \cdot 5,5 + 5 \cdot 6,5 + 4 \cdot 7,5 + 3 \cdot 8,5 + 1 \cdot 9,5}{15} = \frac{217}{30}.$$

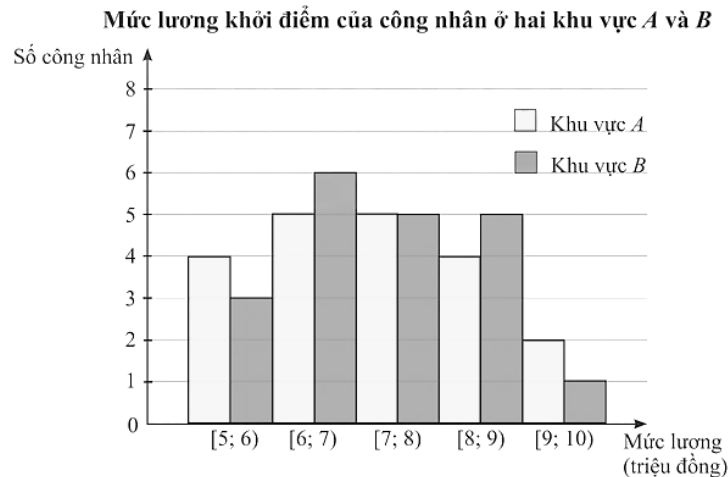
Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

Câu 45. 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) Bảng tần số ghép nhóm và giá trị đại diện cho mẫu số liệu trên như sau:

Mức lương (triệu đồng)	[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 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.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Từ biểu đồ, ta lập được bảng tần số ghép nhóm cho mẫu số liệu như sau:

Mức lương (triệu đồng)	[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 $\bar{x}_A = \frac{4.5,5 + 5.6,5 + 5.7,5 + 4.8,5 + 2.9,5}{20} = 7,25$.

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: $S_A = \sqrt{1,5875}$

c) 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: $\bar{x}_B = \frac{3.5,5 + 6.6,5 + 5.7,5 + 5.8,5 + 1.9,5}{20} = 7,25$.

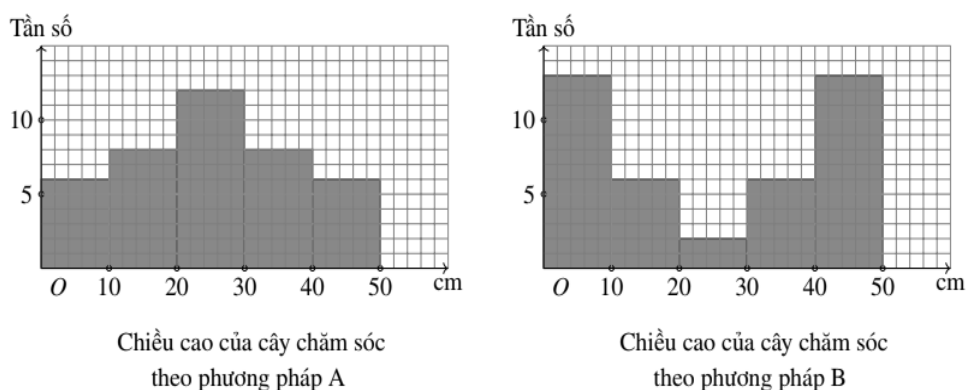
Phương sai của mẫu số liệu ghép nhóm là

$$S_B^2 = \frac{1}{20} (3 \cdot 5,5^2 + 6 \cdot 6,5^2 + 5 \cdot 7,5^2 + 5 \cdot 8,5^2 + 1 \cdot 9,5^2) - (7,25)^2 = 1,2875$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S_B = \sqrt{1,2875}$

d) Đúng: 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.

Câu 46. Một công ty giống cây trồng đã thử nghiệm hai phương pháp chăm sóc khác nhau cho cây hướng dương. Sau hai tuần, người ta thấy cây được chăm sóc theo cả hai phương pháp đều thấp hơn 50 cm.



Từ biểu đồ, ta có bảng tần số ghép nhóm về chiều cao của cây được chăm sóc theo phương pháp A và B là:

Chiều cao (cm)	[0;10)	[10;20)	[20;30)	[30;40)	[40;50)
Mức lương đại diện (cm)	5	15	25	35	45
Chăm sóc theo phương pháp A	6	8	12	8	6
Chăm sóc theo phương pháp B	13	6	2	6	13

a) Khoảng biến thiên của chiều cao các cây được chăm sóc theo mỗi phương pháp A và B bằng nhau.

b) Trung bình của chiều cao các cây được chăm sóc theo mỗi phương pháp A và B bằng nhau.

c) Độ lệch chuẩn của chiều cao các cây được chăm sóc theo phương án A là 12,65(cm).

d) Dựa vào độ lệch chuẩn thì chiều cao của các loại cây được chăm sóc theo phương án B ít bị chênh lệch hơn so với phương án A .

Lời giải

A.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Khoảng biến thiên của chiều cao các cây được chăm sóc theo mỗi phương pháp A và B bằng nhau và cùng bằng 50.

b) Ước tính số trung bình và độ lệch chuẩn của chiều cao các cây được chăm sóc theo mỗi phương pháp. Cỡ mẫu của hai mẫu số liệu thống kê là $N = 40$.

Chiều cao trung bình của các cây được chăm sóc theo phương án A là:

$$\bar{x}_A = \frac{5.6 + 18.5 + 25.12 + 35.8 + 45.6}{40} = 25$$

Chiều cao trung bình của các cây được chăm sóc theo phương án B là:

$$\bar{x}_B = \frac{5.13 + 15.6 + 25.2 + 35.6 + 45.13}{40} = 25 \text{ cm.}$$

c) Độ lệch chuẩn của chiều cao các cây được chăm sóc theo phương án A là:

$$s_A = \sqrt{\frac{5^2.6 + 15^2.8 + 25^2.12 + 35^2.8 + 45^2.6}{40} - 25^2} \approx 12,65$$

d) Độ lệch chuẩn của chiều cao các cây được chăm sóc theo phương án B là:

$$s_B = \sqrt{\frac{5^2.13 + 15^2.6 + 25^2.2 + 35^2.6 + 45^2.13}{40} - 25^2} \approx 17,03 \text{ cm}^2$$

Câu 47. Kết quả 40 lần nhảy xa của hai vận động viên Dũng và Huy được lần lượt thống kê trong bảng bên dưới (đơn vị: mét)

Nhóm	Dũng	Huy
[6, 22; 6, 46)	3	2
[6, 46; 6, 70)	7	5
[6, 70; 6, 94)	5	8
[6, 94; 7, 18)	20	19
[7, 18; 7, 42)	5	6
	$n = 40$	$n = 40$

a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng là $6,92(m)$ (kết quả làm tròn đến phần trăm của mét).

b) Độ lệch chuẩn của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng cho là $0,26(m)$ (kết quả làm tròn đến phần trăm của mét).

c) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy cho là $0,16$ (kết quả làm tròn đến phần trăm).

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)	b)	c)	d)
ĐÚNG	ĐÚNG	SAI	SAI

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
		$n = 40$

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
		$n = 40$

a) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng là:

$$\bar{x}_D = \frac{3.6,34 + 7.6,58 + 5.6,82 + 20.7,06 + 5.7,30}{40} = \frac{276,88}{40} \approx 6,92 \text{ (m)}$$

b) Phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Dũng (làm tròn kết quả đến hàng phần trăm) là:

$$s_D^2 = \frac{1}{40} \left[3.(6,34 - 6,92)^2 + 7.(6,58 - 6,92)^2 + 5.(6,82 - 6,92)^2 + 20.(7,06 - 6,92)^2 + 5.(7,30 - 6,92)^2 \right] = \frac{2,9824}{40} \approx 0,07.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s_D \approx \sqrt{0,07} \approx 0,26 \text{ (m)}$

c) Số trung bình cộng của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy là:

$$\bar{x}_H = \frac{2.6,34 + 5.6,58 + 8.6,82 + 19.7,06 + 6.7,30}{40} = \frac{278,08}{40} \approx 6,95 \text{ (m)}$$

Vậy phương sai của mẫu số liệu ghép nhóm biểu diễn kết quả 40 lần nhảy xa của vận động viên Huy (làm tròn kết quả đến hàng phần trăm) là:

$$s_H^2 = \frac{1}{40} \left[2.(6,34 - 6,95)^2 + 5.(6,58 - 6,95)^2 + 8.(6,82 - 6,95)^2 + 19.(7,06 - 6,95)^2 + 6.(7,30 - 6,95)^2 \right] = \frac{2,5288}{40} \approx 0,06$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm trên là: $s_H \approx \sqrt{0,06} \approx 0,24 \text{ (m)}$

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.

Câu 48. Mẫu số liệu ghép nhóm thống kê mức lương của hai công ty A, B (đơn vị: triệu đồng) được thể hiện như bảng dưới đây

Nhóm	Công ty A	Công ty B
[10;15)	15	25
[15;20)	18	15
[20;25)	10	7
[25;30)	10	5
[30;35)	5	5
[35;40)	2	3
	$n = 60$	$n = 60$

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 là 52,91 (làm tròn kết quả đến hàng phần trăm)

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.	B.	C.	D.
ĐÚNG	SAI	SAI	ĐÚNG

Nhóm	Giá trị đại diện	Tần số
[10; 15)	12,5	15
[15; 20)	17,5	18
[20; 25)	22,5	10
[25; 30)	27,5	10
[30; 35)	32,5	5
[35; 40)	37,5	2
		$n = 60$

Nhóm	Giá trị đại diện	Tần số
[10; 15)	12,5	25
[15; 20)	17,5	15
[20; 25)	22,5	7
[25; 30)	27,5	5
[30; 35)	32,5	5
[35; 40)	37,5	3
		$n = 60$

a) Số trung bình cộng của mẫu số liệu ghép nhóm của công ty A là:

$$\bar{x}_A = \frac{15.12,5 + 18.17,5 + 10.22,5 + 10.27,5 + 5.32,5 + 2.37,5}{60} = \frac{62}{3}$$

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: $s_A = \sqrt{s_A^2} = \sqrt{49,14} \approx 7$

c) Sai: 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:

$$s_B^2 = \frac{25 \cdot \left(12,5 - \frac{229}{12}\right)^2 + 15 \cdot \left(17,5 - \frac{229}{12}\right)^2 + 7 \cdot \left(22,5 - \frac{229}{12}\right)^2 + 5 \cdot \left(27,5 - \frac{229}{12}\right)^2}{60} + \frac{5 \cdot \left(32,5 - \frac{229}{12}\right)^2 + 3 \cdot \left(37,5 - \frac{229}{12}\right)^2}{60} \approx 57,91$$

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.

Câu 49. Thầy Tuấn thống kê lại điểm trung bình cuối năm của các học sinh lớp 11A 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) 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 $\sqrt{0,59}$.

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.	B.	C.	D.
ĐÚNG	ĐÚNG	SAI	ĐÚNG

a) Khoảng biến thiên của điểm số học sinh lớp 11A là: $10 - 5 = 5$.

b) 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.

c) 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 $S_1 = \sqrt{0,61}$.

d) 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.

Câu 50. 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

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.	B.	C.	D.
ĐÚNG	SAI	SAI	ĐÚNG

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

a) 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 $S_1 = \sqrt{S_1^2} = \sqrt{7,5216}$.

b) Xét mẫu số liệu của cổ phiếu B:

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_2 = \frac{16.121 + 4.123 + 3.125 + 6.127 + 21.129}{50} = 125,28.$

c) Sai: Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{50} (16.121^2 + 4.123^2 + 3.125^2 + 6.127^2 + 21.129^2) - (125,28)^2 = 12,4096.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{S_2^2} = \sqrt{12,4096}$.

d) 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.

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

Câu 51. 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:

Lời giải

Đáp án: 15

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 52. Bảng sau thống kê cân nặng của 50 quả xoài Úc 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

Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm đó (kết quả làm tròn đến phần mười).

Trả lời:

Lời giải

Đáp án: 63,5

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

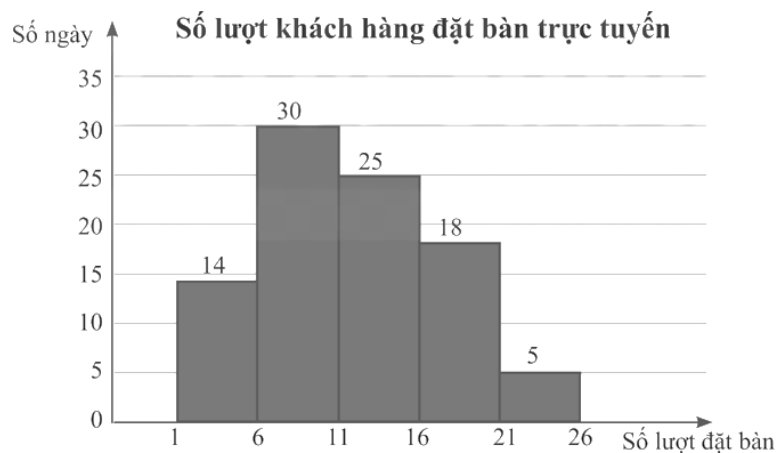
$$\text{ghép nhóm là: } Q_1 = 290 + \frac{\frac{50}{4} - 3}{13} \cdot (330 - 290) = \frac{4150}{13}.$$

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

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

$$\text{Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q = \frac{4210}{11} - \frac{4150}{13} = \frac{9080}{143} \approx 63,5$$

Câu 53. 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:

Lời giải

Đáp án: 8,5

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

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

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: $\Delta_Q = Q_3 - Q_1 = 8,5$

Câu 54. Thời gian luyện tập trong một ngày (tính theo giờ) của một số vận động viên được ghi lại ở bảng sau:

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

Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên (kết quả làm tròn đến phần trăm).

Trả lời:

Lời giải

Đáp án: 1,63

Gọi $x_1; x_2; \dots; x_{39}$ là thời gian luyện tập của 39 vận động viên.

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

$$\text{ghép nhóm là } Q_2 = 4 + \frac{\frac{2 \cdot 39}{4} - (3 + 8)}{12} \cdot (6 - 4) = \frac{65}{12} \approx 5,417.$$

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

$$\text{ghép nhóm là } Q_1 = 2 + \frac{\frac{1 \cdot 39}{4} - 3}{8} \cdot (4 - 2) = \frac{59}{16} = 3,6875.$$

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q \approx 1,63$

Câu 55. 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 (kết quả làm tròn đến hàng đơn vị).

Trả lời:

Lời giải

Đáp án: 116

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:

$$x_1, \dots, x_8 \in [0; 60)$$

$$x_9, \dots, x_{18} \in [60; 120)$$

$$x_{19}, \dots, x_{25} \in [120; 180)$$

$$x_{26}, \dots, x_{30} \in [180; 240)$$

$$x_{31}, x_{32} \in [240; 300)$$

$$x_{33} \in [300; 360)$$

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 hai của mẫu

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

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 nhất của mẫu

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

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 ba của mẫu số

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

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

Câu 56. 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:

Lời giải

Đáp án: 9,7

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 thứ hai của mẫu

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

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 thứ nhất của

$$\text{mẫu số liệu là } Q_1 = 10,5 + \frac{\frac{20}{4} - 3}{9}(15,5 - 10,5) = 11,6$$

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 thứ ba của mẫu

$$\text{số liệu là } Q_3 = 20,5 + \frac{\frac{3 \cdot 20}{4} - 14}{6}(25,5 - 20,5) = 21,3$$

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

Câu 57. 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 (kết quả làm tròn đến hàng đơn vị).

Trả lời:

Lời giải

Đáp án: 112

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 nhóm là

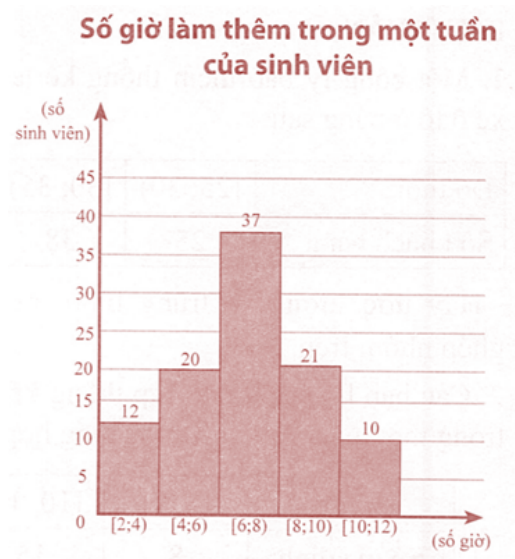
$Q_1 = 60 + \frac{\frac{30}{4} - 2}{7} \cdot (120 - 60) = \frac{750}{7} \approx 107,14.$

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 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: $\Delta_Q \approx 112$

Câu 58. 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:

Lời giải

Đáp án: 3,27

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 $\Delta Q = 3,27$

Câu 59. 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 (kết quả làm tròn đến phần mười).

Trả lời:

Lời giải

Đáp án: 23,7

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, $p = 5$; $a_5 = 39,5$; $m = m_5 = 15$; $m_1 + m_2 + m_3 + m_4 = 13$; $a_6 - a_5 = 10$ 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: $\Delta_Q \approx 23,7$

Câu 60. 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:

Lời giải

Đáp án: 0,53

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:

$$s^2 = \frac{1}{40} (5,25^2 \cdot 2 + 5,75^2 \cdot 8 + 6,25^2 \cdot 15 + 6,75^2 \cdot 10 + 7,25^2 \cdot 5) - 6,35^2 = 0,2775$$

Độ lệch chuẩn của mẫu số liệu là: $\sqrt{0,2775} = \frac{\sqrt{111}}{20} \approx 0,53$

Câu 61. 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:

Lời giải

Đáp án: 0,53

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

Cỡ mẫu: $n = 100$

$$\text{Số trung bình: } \bar{x} = \frac{13.19,25 + 45.19,75 + 24.20,25 + 12.20,75 + 6.21,25}{100} = 20,015$$

$$\text{Phương sai: } S^2 = \frac{13.19,25^2 + 45.19,75^2 + 24.20,25^2 + 12.20,75^2 + 6.21,25^2}{100} - 20,015^2 \approx 0,28$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{0,28} \approx 0,53$$

Câu 62. 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 mẫu số liệu ghép nhóm trên (kết quả được làm tròn đến hàng phần trăm)

Trả lời:

Lời giải

Đáp án: 1,53

Mẫu số liệu với giá trị đại diện

Mức xà (cm)	[170;172)	[172;174)	[174;176)	[176;180)
Giá trị đại diện	171	173	175	177
Số vận động viên	3	10	6	1

$$\text{Giá trị trung bình: } \bar{x} = \frac{1}{20} (171.3 + 173.10 + 175.6 + 177.1) = 173,5(\text{cm})$$

$$\text{Phương sai của mẫu số liệu: } s^2 = \frac{1}{20} (171^2.3 + 173^2.10 + 175^2.6 + 177^2.1) - 173,5^2 = 2,35$$

$$\text{Độ lệch chuẩn của mẫu số liệu: } s = \sqrt{2,35} = \frac{\sqrt{235}}{10} \approx 1,53$$

Câu 63. 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:

Lời giải

Đáp án: 2,43

Mẫu số liệu với giá trị đại diện

Chiều dài (cm)	[44; 46)	[46; 48)	[48; 50)	[50; 52)	[52; 54)	[54; 56)
Giá trị đại diện	45	47	49	51	53	55
Số trẻ	3	3	10	15	7	2

Từ mẫu số liệu đã cho, ta tính được số trung bình là:

$$\bar{x} = \frac{3.45 + 3.47 + 10.49 + 15.51 + 7.53 + 2.55}{40} = \frac{2012}{40} = 50,3.$$

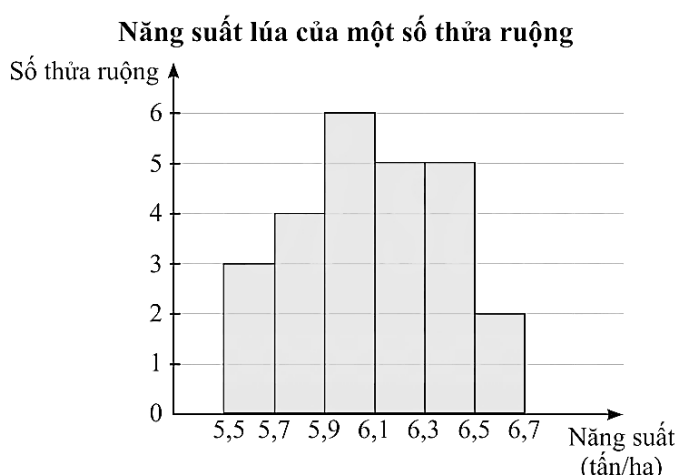
$\bar{x}$ không phải là số nguyên nên để tính phương sai ta tính:

$$\overline{x^2} = \frac{3.45^2 + 3.47^2 + 10.49^2 + 15.51^2 + 7.53^2 + 2.55^2}{40} = 2536.$$

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



Hãy xác định độ lệch chuẩn của mẫu số liệu trên (làm tròn kết quả đến hàng phần trăm).

Trả lời:

Lời giải

Đáp án: 0,29

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

Số trung bình: $\bar{x} = \frac{3.5,6 + 4.5,8 + 6.6,0 + 5.6,2 + 5.6,4 + 2.6,6}{25} = 6,088$

Độ lệch chuẩn: $s = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} \approx 0,29$

Câu 65. 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 (làm tròn kết quả đến hàng phần trăm).

Trả lời:

Lời giải

Đáp án: 0,03

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: $\bar{x} = \frac{1}{20}(10,3.3 + 10,5.7 + 10,7.8 + 10,9.2) = 10,59$

(giây)

Phương sai của mẫu số liệu là: $s^2 = \frac{1}{20}(10,3^2.3 + 10,5^2.7 + 10,7^2.8 + 10,9^2.2) - 10,59^2 \approx 0,03$

Câu 66. 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:

Lời giải

Đáp án: 4,8

Ta có bảng thống kê cân nặng của các quả mít theo giá trị đại diện:

Cân nặng (kg)	[4;6)	[6;8)	[8;10)	[10;12)	[12;14)
Cân nặng đại diện (kg)	5	7	9	11	13
Số quả mít	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: $S^2 = \frac{1}{50} (6 \cdot 5^2 + 12 \cdot 7^2 + 19 \cdot 9^2 + 9 \cdot 11^2 + 4 \cdot 13^2) - 8,72^2 \approx 4,8$

Câu 67. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của một nhà đầu tư được cho như sau:

Lợi nhuận (đơn vị: triệu đồng)	[10;20)	[20;30)	[30;40)	[40;50)	[50;60)
Số tháng	2	4	8	4	2

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần mười).

Trả lời:

Lời giải

Đáp án: 11

Giá trị đại diện lợi nhuận theo tháng của nhà đầu tư

Lợi nhuận (đơn vị: triệu đồng)	[10;20)	[20;30)	[30;40)	[40;50)	[50;60)
Giá trị đại diện	15	25	35	45	55
Số tháng	2	4	8	4	2

Lợi nhuận trung bình một tháng của nhà đầu tư là: $\bar{x}_A = \frac{1}{20} (2 \cdot 15 + \dots + 2 \cdot 55) = 35$ (triệu đồng)

Độ lệch chuẩn của lợi nhuận hàng tháng của nhà đầu tư là:

$$s_A = \sqrt{\frac{1}{20} (2 \cdot 15^2 + \dots + 2 \cdot 55^2) - (35)^2} \approx 11$$

Câu 68. Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của một nhà đầu tư được cho như sau:

Lợi nhuận (đơn vị: triệu đồng)	[510;520)	[520;530)	[530;540)	[540;550)	[550;560)
Số tháng	4	3	6	3	4

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần mười).

Trả lời:

Lời giải

Đáp án: 13,8

Giá trị đại diện lợi nhuận theo tháng của nhà đầu tư

Lợi nhuận (đơn vị: triệu đồng)	[510;520)	[520;530)	[530;540)	[540;550)	[550;560)
Giá trị đại diện	515	525	535	545	555
Số tháng	4	3	6	3	4

Lợi nhuận trung bình một tháng của các nhà đầu tư là:

$$\bar{x}_B = \frac{1}{20}(4 \cdot 515 + \dots + 4 \cdot 555) = 535 \text{ (triệu đồng)}.$$

Độ lệch chuẩn của lợi nhuận hàng tháng của nhà đầu tư là:

$$s_B = \sqrt{\frac{1}{20}(4 \cdot 515^2 + \dots + 4 \cdot 555^2) - (535)^2} \approx 13,8$$

Câu 69. Trong buổi thực hành, 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

Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần trăm).

Trả lời:

Lời giải

Đáp án: 0,43

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

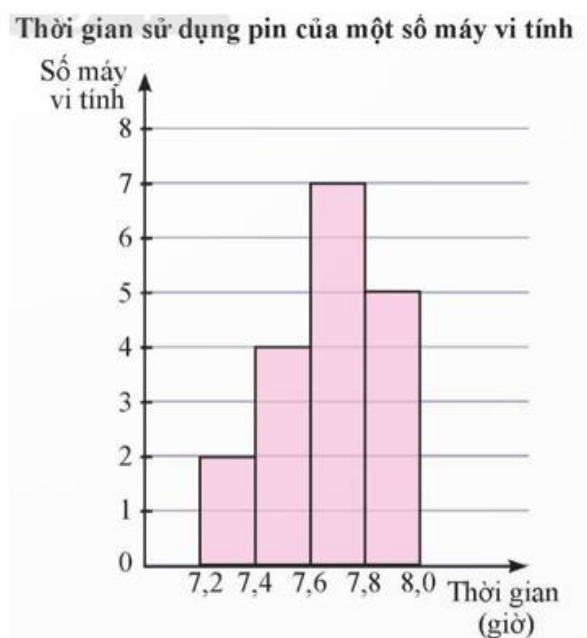
Cỡ mẫu: $n = 3 + 8 + 7 + 2 = 20$

$$\text{Số trung bình: } \bar{x} = \frac{4,75 \cdot 3 + 5,25 \cdot 8 + 5,75 \cdot 7 + 6,25 \cdot 2}{20} = \frac{109}{20} (\mu m)$$

Phương sai: $s^2 = \frac{1}{20} \left(4,75^2 \cdot 3 + 5,25^2 \cdot 8 + 5,75^2 \cdot 7 + 6,25^2 \cdot 2 \right) - \left(\frac{109}{20} \right)^2 = \frac{37}{200}$

Độ lệch chuẩn: $s = \sqrt{\frac{37}{200}} \approx 0,43$

Câu 70. 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.



Tính độ lệch chuẩn của mẫu số liệu ghép nhóm trên (kết quả các phép tính làm tròn đến hàng phần trăm).

Trả lời:

Lời giải

Đáp án: 0,18

Từ biểu đồ, ta có bảng thống kê sau:

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

Cỡ mẫu là $n = 2 + 4 + 7 + 5 = 18$

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{2 \cdot 7,3 + 4 \cdot 7,5 + 7 \cdot 7,7 + 5 \cdot 7,9}{18} = \frac{23}{3}$

Phương sai của mẫu số liệu ghép nhóm là:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} \approx \sqrt{0,032} \approx 0,18$

Câu 71. Hãy tìm khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của cô Minh Nhân và cô Minh Hiền.

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Cô Minh Nhân	5	12	8	3	2
Cô Minh Hiền	0	25	5	0	0

Hỏi thời gian tập thể dục buổi sáng mỗi ngày của ai phân tán hơn?

Lời giải

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

Do trong mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của cô Minh Hiền, khoảng đầu tiên chứa dữ liệu là $[20; 25)$ và khoảng cuối cùng chứa dữ liệu là $[25; 30)$. Nên khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của cô Minh Hiền là $R_H = 30 - 20 = 10$ (phút).

Ta thấy $R_N > R_H$ nên thời gian tập thể dục buổi sáng mỗi ngày của cô Minh Nhân phân tán hơn cô Minh Hiền

Câu 72. Thời gian chờ khám bệnh của các bệnh nhân tại phòng khám X được cho trong bảng sau:

Thời gian (phút)	[0; 5)	[5; 10)	[10; 15)	[15; 20)
Số bệnh nhân	3	12	15	8

a) Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm này.

b) Từ một mẫu số liệu về thời gian chờ khám bệnh của các bệnh nhân tại phòng khám Y người ta tính được khoảng tứ phân vị bằng 9,23. Hỏi thời gian chờ của bệnh nhân tại phòng khám nào phân tán hơn?

Lời giải

a) Cỡ mẫu là $n = 3 + 12 + 15 + 8 = 38$.

Gọi $x_1, \dots, x_{38}$ là thời gian chờ khám bệnh của 38 bệnh nhân này và giả sử rằng dãy số liệu gốc này đã được sắp xếp theo thứ tự tăng dần. Tứ phân vị thứ nhất của mẫu số liệu gốc là x_{10} nên nhóm chứa tứ phân vị thứ nhất là nhóm $[5; 10)$ và ta có:

$$Q_1 = 5 + \left[\frac{\frac{38}{4} - 3}{12} \right] \cdot 5 \approx 7,71.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là x_{29} nên nhóm chứa tứ phân vị thứ ba là nhóm $[10;15)$ và ta có:

$$Q_3 = 10 + \left[\frac{\frac{3 \cdot 38}{4} - 15}{15} \right] \cdot 5 = 14,5.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 \approx 14,5 - 7,71 = 6,79$.

b) Do $\Delta_Q = 6,79 < 9,23$ nên thời gian chờ của bệnh nhân tại phòng khám Y phân tán hơn thời gian chờ của bệnh nhân tại phòng khám X .

Câu 73. Thống kê số ngày trong tháng 11 năm 2023 và năm 2024 theo nhiệt độ cao nhất trong ngày tại Nha Trang, 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 11/2023	0	2	8	5	6	9
Số ngày trong tháng 11/2024	2	3	4	11	8	2

Hỏi tháng sáu năm nào ở Nha Trang nhiệt độ cao nhất trong ngày biến đổi nhiều hơn?

Lời giải

• Năm 2023:

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 11 năm 2023 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 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 nhóm $[38;40)$

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

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

• Năm 2024:

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 11 năm 2024 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 $[32; 34)$ 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 11 năm 2024 biến đổi nhiều hơn nhiệt độ cao nhất trong ngày vào tháng 11 năm 2023.

Theo khoảng tứ phân vị: vì $\Delta_{Q_1} > \Delta_{Q_2}$ nên nhiệt độ cao nhất trong ngày vào tháng 11 năm 2023 biến đổi nhiều hơn nhiệt độ cao nhất trong ngày vào tháng 11 năm 2024.

Câu 74. 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:

$$\frac{6,5 \cdot 20 + 9,5 \cdot 35 + 12,5 \cdot 45 + 15,5 \cdot 35 + 18,5 \cdot 20}{20 + 35 + 45 + 35 + 20} = \frac{25}{2} \text{ (triệu đồng)}$$

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

$$\frac{6,5 \cdot 17 + 9,5 \cdot 23 + 12,5 \cdot 30 + 15,5 \cdot 23 + 18,5 \cdot 17}{17 + 23 + 30 + 23 + 17} = \frac{25}{2} \text{ (triệu đồng)}$$

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) và tứ phân vị

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

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

$$[14;17) \text{ và tứ phân vị thứ ba là: } Q_3 = 14 + \frac{\frac{3 \cdot 155}{4} - (20 + 35 + 45)}{35} \cdot 3 = \frac{431}{28}$$

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

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 tứ phân vị thứ

$$\text{nhất là: } Q'_1 = 8 + \frac{\frac{110}{4} - 17}{23} \cdot 3 = \frac{431}{46}$$

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)

$$\text{và tứ phân vị thứ ba là: } Q'_3 = 14 + \frac{\frac{3 \cdot 110}{4} - (17 + 23 + 30)}{23} \cdot 3 = \frac{719}{46}$$

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

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

Câu 75. Bảng sau đây cho biết chiều cao của các học sinh lớp 12A và 12B.

Thu nhập	[145;150)	[150;155)	[155;160)	[160;165)	[165;170)	[170;175)
Số học sinh lớp 12A	1	0	15	12	10	5
Số học sinh 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

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

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

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

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 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) và tứ

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

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 76. 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 vị thứ nhất là:

$$Q_1 = 1 + \frac{\frac{80}{4} - 8}{17} \cdot 1 = \frac{29}{17}$$

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

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

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

Câu 77. 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

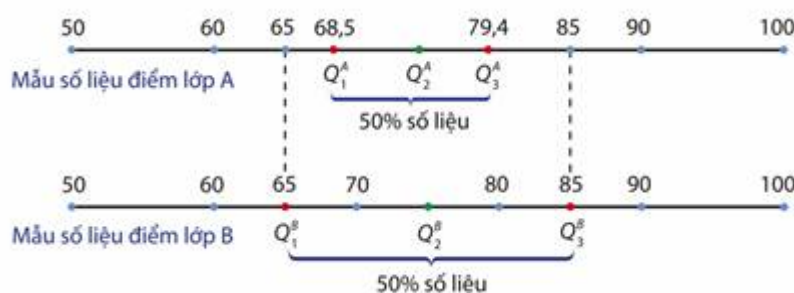
Nhóm chứa Q_1^B là $[60; 70)$. $Q_1^B = 60 + \frac{25-15}{20} \cdot 10 = 65$.

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 họa 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.



Câu 78. 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

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.

$$\text{Do đó, } Q_3 = \frac{73 + 75}{2} = 74.$$

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)

$$\text{và ta có: } Q'_1 = 50 + \frac{\frac{20}{5} - 2}{5} \cdot 10 = 56$$

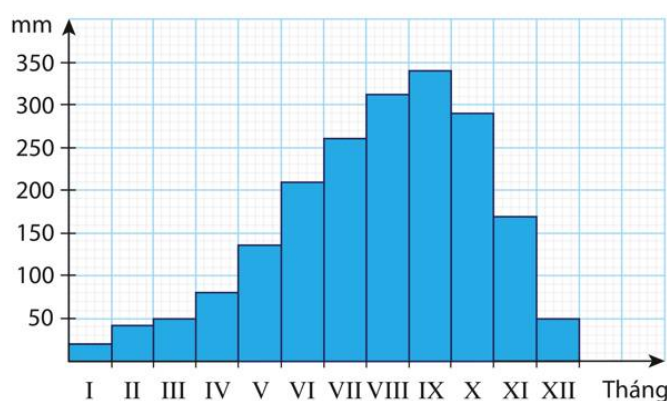
Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2}$ nên nhóm chứa tứ phân vị thứ ba là nhóm [70; 80) và

$$\text{ta có: } Q'_3 = 70 + \frac{\frac{3 \cdot 20}{5} - (2 + 5 + 7)}{5} \cdot 10 = 72$$

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 79. 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). Nếu 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.

Câu 80. Hằng ngày Bác Toàn đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bản thống kê thời gian của 100 lần Bác Toàn đi xe buýt từ nhà đến cơ quan:

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

a) Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm phía trên (Làm tròn kết quả đến hàng phần trăm)

b) Biết rằng trong 100 lần đi trên, chỉ có đúng một lần Bác Toàn đi hết 32 phút. Thời gian của lần đi đó có phải giá trị ngoại lệ không?

Lời giải

a) Cỡ mẫu $n = 100$. Gọi $x_1, x_2, x_3, \dots, x_{100}$ là mẫu số liệu gốc gồm 100 lần đi xe buýt của Bác Toàn. Khi đó ta có:

$$x_1, \dots, x_{22} \in [15; 18)$$

$$x_{23}, \dots, x_{60} \in [18; 21)$$

$$x_{61}, \dots, x_{87} \in [21; 24)$$

$$x_{88}, \dots, x_{95} \in [21; 27)$$

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

$$x_{100} \in [30; 33)$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{1}{2}(x_{75} + x_{76}) \in [18; 21)$

Do đó tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là: $Q_1 = 18 + \frac{\frac{100}{4} - 22}{38}(21 - 18) = \frac{693}{38}$

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: $Q_3 = 18 + \frac{\frac{3 \cdot 100}{4} - (22 + 38)}{27}(24 - 21) = \frac{68}{3}$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta_Q = \frac{68}{3} - \frac{693}{38} = \frac{505}{114} \approx 4,43$.

b) Trong lần duy nhất Bác Toàn đi hết 32 phút, thời gian đi của Bác thuộc nhóm $[30; 33)$

Vì $Q_3 + 1,5\Delta_Q \approx 29,31 < 30$ nên đây là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Câu 81. Trong buổi thực hành, 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

Cỡ mẫu: $n = 3 + 8 + 7 + 2 = 20$

$$\text{Số trung bình: } \bar{x} = \frac{4,75 \cdot 3 + 5,25 \cdot 8 + 5,75 \cdot 7 + 6,25 \cdot 2}{20} = \frac{109}{20} (\mu m)$$

$$\text{Phương sai: } s^2 = \frac{1}{20} (4,75^2 \cdot 3 + 5,25^2 \cdot 8 + 5,75^2 \cdot 7 + 6,25^2 \cdot 2) - \left(\frac{109}{20} \right)^2 = \frac{37}{200}$$

$$\text{Độ lệch chuẩn: } s = \sqrt{\frac{37}{200}} \approx 0,43$$

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 82. Hai vận động viên điền kinh A và B ghi lại thời gian chạy tập luyện chạy cự li 100 m của mình 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:

$$\text{Giá trị trung bình } \bar{x}_B = \frac{10,15 \cdot 3 + 10,45 \cdot 7 + 10,75 \cdot 9 + 11,05 \cdot 6}{3 + 7 + 9 + 6} = \frac{5333}{500}$$

$$\text{Phương sai: } s_B^2 = \frac{1}{25} (10,15^2 \cdot 3 + 10,45^2 \cdot 7 + 10,75^2 \cdot 9 + 11,05^2 \cdot 6) - \left(\frac{5333}{500} \right)^2 = \frac{1296}{15625}$$

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 83. Để rèn luyện kỹ năng đi bộ, Cô Minh Nhân và Cô Minh Hiền cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân đi bộ của hai cô 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)
Cô Minh Nhân	6	7	6	6	5
Cô Minh Hiền	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

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

a) Cỡ mẫu: $n = 30$

- Xét mẫu số liệu của Cô Minh Nhân:

Số trung bình: $\bar{x}_1 = \frac{6.4 + 7.6 + 6.8 + 6.10 + 5.12}{30} = 7,8$

Phương sai: $S_1^2 = \frac{(6.4^2 + 7.6^2 + 6.8^2 + 6.10^2 + 5.12^2)}{30} - 7,8^2 = 7,56$

Độ lệch chuẩn: $S_1 = \sqrt{7,56} \approx 2,75$

- Xét mẫu số liệu của Cô Minh Hiền:

Số trung bình: $\bar{x}_2 = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$

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: $S_2 = \sqrt{3,83} \approx 1,96$

b) Nếu so sánh theo độ lệch chuẩn thì Cô Minh Hiền có số lượng bước chân đi mỗi ngày đều đặn hơn Cô Minh Nhân

Kết quả các phép tính làm tròn đến hàng phần nghìn.

Câu 84. Một cuộc khảo sát thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 5 hai trường A và B được ghi lại ở bảng sau:

Thời gian (phút)	[6;7)	[7;8)	[8;9)	[9;10)	[10;11)
Học sinh trường A	8	10	13	10	9
Học sinh trường B	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

Ta có bảng số liệu với giá trị đại diện:

Thời gian (phút)	[6;7)	[7;8)	[8;9)	[9;10)	[10;11)
Giá trị đại diện	6,5	7,5	8,5	9,5	10,5
Học sinh trường A	8	10	13	10	9
Học sinh trường B	4	12	17	14	3

a) Cỡ mẫu: $n = 50$

Xét số liệu của trường A:

$$\text{Số trung bình: } \bar{x}_A = \frac{8 \cdot 6,5 + 10 \cdot 7,5 + 13 \cdot 8,5 + 10 \cdot 9,5 + 9 \cdot 10,5}{50} = 8,54$$

Xét số liệu của trường B:

$$\text{Số trung bình: } \bar{x}_B = \frac{4 \cdot 6,5 + 12 \cdot 7,5 + 17 \cdot 8,5 + 14 \cdot 9,5 + 3 \cdot 10,5}{50} = 8,5$$

Vậy nếu so sánh theo số trung bình thì học sinh trường B 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 5 trường A 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

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

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

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

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 5 trường B 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

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

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

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

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 B có tốc độ viết đồng đều hơn

c) Xét số liệu của trường A:

$$\text{Độ lệch chuẩn: } S_A = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,54^2} \approx 1,33$$

Xét số liệu của trường Y:

$$\text{Độ lệch chuẩn: } S_B = \sqrt{\frac{4.6,5^2 + 12.7,5^2 + 17.8,5^2 + 14.9,5^2 + 3.10,5^2}{50} - 8,5^2} \approx 1,04$$

Vậy nếu so sánh theo độ lệch chuẩn thì học sinh trường B có tốc độ viết đồng đều hơn

Câu 85. 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

a) Tính phương sai và độ lệch chuẩn của mỗi mẫu số liệu ghép nhóm.

b) Nhận xét về độ phân tán của tuổi thọ các linh kiện điện tử được sản xuất bởi mỗi phân xưởng.

Lời giải

a) Ta có bảng số liệu với giá trị đại diện:

Tuổi thọ (năm)	[1,5;2)	[2;2,5)	[2,5;3)	[3;3,5)	[3,5;4)
Giá trị đại diện	1,75	2,25	2,75	3,25	3,75
Số linh kiện của phân xưởng 1	4	9	13	8	6
Số linh kiện của phân xưởng 2	2	8	20	7	3

• Phân xưởng 1: Tổng số linh kiện: $n_1 = 4 + 9 + 13 + 8 + 6 = 40$

$$\text{Giá trị trung bình } \bar{x}_1 = \frac{1,75 \cdot 4 + 2,25 \cdot 9 + 2,75 \cdot 13 + 3,25 \cdot 8 + 3,75 \cdot 6}{4 + 9 + 13 + 8 + 6} = \frac{223}{40}$$

$$\text{Phương sai: } s_1^2 = \frac{1}{40} (1,75^2 \cdot 4 + 2,25^2 \cdot 9 + 2,75^2 \cdot 13 + 3,25^2 \cdot 8 + 3,75^2 \cdot 6) - \left(\frac{223}{40} \right)^2 = \frac{2271}{1600}$$

$$\text{Độ lệch chuẩn: } s_1 = \sqrt{\frac{2271}{1600}} \approx 0,6$$

• Phân xưởng 2: Tổng số linh kiện: $n_2 = 2 + 8 + 20 + 7 + 3 = 40$

$$\text{Giá trị trung bình } \bar{x}_2 = \frac{1,75 \cdot 2 + 2,25 \cdot 8 + 2,75 \cdot 20 + 3,25 \cdot 7 + 3,75 \cdot 3}{2 + 8 + 20 + 7 + 3} = \frac{221}{40}$$

$$\text{Phương sai: } s_2^2 = \frac{1}{40} (1,75^2 \cdot 2 + 2,25^2 \cdot 8 + 2,75^2 \cdot 20 + 3,25^2 \cdot 7 + 3,75^2 \cdot 3) - \left(\frac{221}{40} \right)^2 = \frac{1399}{1600}$$

$$\text{Độ lệch chuẩn: } s_2 = \sqrt{\frac{1399}{1600}} \approx 0,47$$

b) Vì $0,6 > 0,47$ nên độ phân tán của phân xưởng 1 lớn hơn độ phân tán của phân xưởng 2.

Câu 86. Trong bài thực hành Vật lí đo hiệu điện thế của mạch điện, hai bạn Việt và Nam đã dùng hai vôn kế khác nhau để đo 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 Việt đo	1	6	2	1
Số lần Nam đo	1	3	4	2

a) Tính độ lệch chuẩn của các mẫu số liệu ghép nhóm cho kết quả đo của Việt và Nam.

b) Từ đó kết luận xem vôn kế của bạn nào cho kết quả đo ổn định hơn.

Lời giải

a) 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 Việt đo	1	6	2	1
Số lần Nam đo	1	3	4	2

- Kết quả đo của Việt:

Giá trị trung bình: $\bar{x}_1 = \frac{1}{10}(3,875.1 + 3,925.6 + 3,975.2 + 4,025.1) = 3,94$

Phương sai của mẫu số liệu: $s_1^2 = \frac{1}{10}(3,875^2.1 + 3,925^2.6 + 3,975^2.2 + 4,025^2.1) - 3,94^2 = 0,001525$

Độ lệch chuẩn của mẫu số liệu: $s_1 = \sqrt{0,001525} \approx 0,039$

- Kết quả đo của Nam:

Giá trị trung bình: $\bar{x}_1 = \frac{1}{10}(3,875.1 + 3,925.3 + 3,975.4 + 4,025.4) = 3,96$

Phương sai của mẫu số liệu: $s_2^2 = \frac{1}{10}(3,875^2.1 + 3,925^2.3 + 3,975^2.4 + 4,025^2.2) - 3,96^2 = 0,002025$

Độ lệch chuẩn của mẫu số liệu: $s_2 = \sqrt{0,002025} \approx 0,045$

b) Vì $s_2 > s_1$ nên Vân kể của bạn Việt cho kết quả ổn định hơn.

Câu 87. Thầy Tuấn đầu tư số tiền bằng nhau vào hai lĩnh vực kinh doanh A, B . Thầy Tuấn thống kê số tiền thu được mỗi tháng trong vòng 60 tháng theo mỗi lĩnh vực cho kết quả như sau:

Số tiền (triệu đồng)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)
Số tháng đầu tư vào lĩnh vực A	5	10	30	10	5
Số tháng đầu tư vào lĩnh vực B	20	5	10	5	20

So sánh giá trị trung bình và độ lệch chuẩn của số tiền thu được mỗi tháng khi Thầy Tuấn đầu tư vào mỗi lĩnh vực A, B . Đầu tư vào lĩnh vực nào "rủi ro" hơn?

Lời giải

Chọn giá trị đại diện cho các nhóm số liệu ta có:

Số tiền (triệu đồng)	[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 tháng đầu tư vào lĩnh vực A	5	10	30	10	5
Số tháng đầu tư vào lĩnh vực B	20	5	10	5	20

Số tiền trung bình thu được khi Thầy Tuấn đầu tư vào các lĩnh vực A, B tương ứng là:

$$\bar{x}_A = \frac{1}{60}(5 \cdot 7,5 + \dots + 5 \cdot 27,5) = 17,5$$

$$\bar{x}_B = \frac{1}{60}(20 \cdot 7,5 + \dots + 20 \cdot 27,5) = 17,5$$

Như vậy, về trung bình đầu tư vào các lĩnh vực A, B số tiền thu được hàng tháng như nhau.

Độ lệch chuẩn của số tiền thu được hàng tháng khi Thầy Tuấn đầu tư vào các lĩnh vực A, B tương ứng là:

$$s_A = \sqrt{\frac{1}{60}(5 \cdot 7,5^2 + \dots + 5 \cdot 27,5^2) - (17,5)^2} = 5$$

$$s_B = \sqrt{\frac{1}{60}(20 \cdot 7,5^2 + \dots + 20 \cdot 27,5^2) - (17,5)^2} \approx 8,42.$$

Như vậy, độ lệch chuẩn của mẫu số liệu về số tiền thu được hàng tháng khi Thầy Tuấn đầu tư vào lĩnh vực B cao hơn khi đầu tư vào lĩnh vực A . Người ta nói rằng, đầu tư vào lĩnh vực B là "rủi ro" hơn.

Câu 88. 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 (triệu đồng)	[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 (triệu đồng)	[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

a) Tiền lãi trung bình khi đầu tư vào lĩnh vực A là:

$$\bar{x}_A = \frac{7,5 \cdot 2 + 12,5 \cdot 5 + 17,5 \cdot 8 + 22,5 \cdot 6 + 27,5 \cdot 4}{2 + 5 + 8 + 6 + 4} = 18,5 \text{ (triệu đồng)}$$

Tiền lãi trung bình khi đầu tư vào lĩnh vực B là:

$$\bar{x}_B = \frac{7,5 \cdot 8 + 12,5 \cdot 4 + 17,5 \cdot 2 + 22,5 \cdot 5 + 27,5 \cdot 6}{8 + 4 + 2 + 5 + 6} = 16,9 \text{ (triệu đồng)}$$

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 89. 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 $R = 14 - 5 = 9$ 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) 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

Cơ mẫu: $n = 2 + 4 + 9 + 6 + 4 = 25$

Số trung bình $\bar{x}_N$ của mẫu số liệu ghép nhóm:

$$\bar{x}_N = \frac{2.6 + 4.8 + 9.10 + 6.12 + 4.14}{25} = \frac{262}{25} = 10,48 \text{ (giờ)}$$

Phương sai s_N^2 của mẫu số liệu ghép nhóm là:

$$S_N^2 = \frac{2(6-10,48)^2 + 4(8-10,48)^2 + 9(10-10,48)^2 + 6(12-10,48)^2 + 4(14-10,48)^2}{25}$$

$$= \frac{130,24}{25} = 5,2096 \approx 5,21$$

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

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:

$$S_G^2 = \frac{1(5-10)^2 + 1(6-10)^2 + 1(7-10)^2 + \dots + 3(12-10)^2 + 2(13-10)^2 + 2(14-10)^2}{25} = \frac{132}{25} = 5,28$$

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 90. 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: $R = 61,1 - 42 = 19,1(km/h)$

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.

$$\text{Trung vị } Q_2 = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(48,4 + 50,8) = 49,6$$

$$\text{Tứ phân vị thứ nhất là trung vị của nửa số liệu bên trái } Q_1 = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(46,7 + 46,8) = 46,75$$

$$\text{Tứ phân vị thứ ba là trung vị của nửa số liệu bên phải } Q_3 = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(54,8 + 55,6) = 55,2$$

$$\text{Khoảng tứ phân vị của mẫu số liệu là: } \Delta_Q = Q_3 - Q_1 = 8,45$$

$$\text{Số trung bình: } \bar{x} = \frac{42 + 43,4 + \dots + 61,1}{20} = 50,945$$

$$\text{Phương sai: } S^2 = \frac{42^2 + 43,4^2 + \dots + 61,1^2}{20} - 50,945^2 \approx 32,2$$

$$\text{Độ lệch chuẩn: } S = \sqrt{32,2} \approx 5,67$$

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

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 mẫu số

$$\text{liệu ghép nhóm là: } Q_1' = 46 + \frac{\frac{20}{4} - 3}{7}(50 - 46) = \frac{330}{7}$$

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 số liệu

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

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q' = Q_3' - Q_1' = \frac{172}{21}$$

• Bảng giá trị đại diện của mỗi nhóm vào bảng lập ở trên, ta có:

Tốc độ	$[42; 46)$	$[46; 50)$	$[50; 54)$	$[54; 58)$	$[58; 62)$
--------	------------	------------	------------	------------	------------

Giá trị đại diện	44	48	52	56	60
Số xe	3	7	4	3	3

Số trung bình: $\bar{x} = \frac{3.44 + 7.48 + 4.52 + 3.56 + 3.60}{20} = 41,8$

Phương sai: $S^2 = \frac{3.44^2 + 7.48^2 + 4.52^2 + 3.56^2 + 3.60^2}{20} - 41,8^2 = 364,96$

Độ lệch chuẩn: $S = \sqrt{364,96} = 19,1$