



**TÀI LIỆU THUỘC KHÓA HỌC  
“LIVE VIP 9+ TOÁN ”**

**ĐĂNG KÝ HỌC EM INBOX THẦY TU  
VẤN NHÉ!**

## **TOÁN THỰC TẾ LỚP 12 SGK MỚI** **( Chuẩn cấu trúc SGK form mới)**

### **CHUYÊN ĐỀ 6. THỐNG KÊ**

#### **A. KIẾN THỨC CƠ BẢN CẦN NẮM**

##### **I. KHOẢNG BIẾN THIÊN**

Cho mẫu số liệu ghép nhóm:

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

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

Khoảng biến thiên của mẫu 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.

**Lưu ý:** Khoảng biến thiên của MSL ghép nhóm luôn  $\geq$  khoảng biến thiên của MSL gốc.

##### **II. KHOẢNG TỨ PHÂN VỊ**

Tứ phân vị thứ  $r$  là

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

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

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

**Ý nghĩa.** Khoảng tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ cho khoảng tứ phân vị của mẫu số liệu gốc. Khoảng tứ phân vị cũng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm. Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

##### **Nhận xét.**

- Do khoảng tứ phân vị của mẫu số liệu ghép nhóm chỉ phụ thuộc vào nửa giữa của mẫu số liệu, nên không bị ảnh hưởng bởi các giá trị bất thường và có thể dùng đại lượng này để loại giá trị bất thường.
- Khoảng tứ phân vị được dùng để xác định giá trị ngoại lệ trong mẫu số liệu. Giá trị  $x$  trong mẫu số liệu là giá trị ngoại lệ nếu  $x > Q_3 + 1,5\Delta_Q$  hoặc  $x < Q_1 - 1,5\Delta_Q$

#### **III. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN**

##### **1. Phương sai**

Phương sai của mẫu số liệu ghép nhóm, kí hiệu  $S^2$ , được tính bởi công thức:

$$S^2 = \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right]$$

Trong đó:  $n = n_1 + n_2 + \dots + n_k$  là cỡ mẫu;

$\bar{x} = \frac{1}{n} (n_1 c_1 + n_2 c_2 + \dots + n_k c_k)$  là số trung bình

## 2. Độ lệch chuẩn

Độ lệch chuẩn của mẫu số liệu ghép nhóm kí hiệu  $S$ ,

Độ lệch chuẩn là căn bậc hai số học của phương sai:

$$\text{Độ lệch chuẩn} = \sqrt{S^2}$$

### Chú ý:

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

$$S^2 = \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + \dots + n_k c_k^2) - \bar{x}^2$$

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

$$S^2 = \frac{1}{n-1} \left[ n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2 \right].$$

### Ý nghĩa:

- Phương sai** của MSL ghép nhóm là giá trị xấp xỉ cho phương sai của MSL gốc.

**Độ lệch chuẩn** của MSL ghép nhóm cũng là giá trị xấp xỉ cho độ lệch chuẩn của MSL gốc.

→ Được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.

- Phương sai và độ lệch chuẩn càng lớn thì dữ liệu càng phân tán.
- Độ lệch chuẩn có cùng đơn vị với đơn vị của mẫu số liệu.

## B. BÀI TẬP VẬN DỤNG

**Câu 1:** Dữ liệu về tốc độ của 100 xe ô tô lưu thông trên một đoạn đường cao tốc vào giờ cao điểm, được trích xuất từ camera của cơ quan cảnh sát giao thông. Hãy tìm khoảng biến thiên của mẫu số liệu (bảng số liệu hình bên dưới).

| Tốc độ (km/h) | $[60; 70)$ | $[70; 80)$ | $[80; 90)$ | $[90; 100)$ | $[100; 110)$ |
|---------------|------------|------------|------------|-------------|--------------|
| Số xe         | 10         | 20         | 20         | 35          | 15           |

### Lời giải

Ta có  $u_{k+1} = 110$  và  $u_1 = 60$

Khoảng biến thiên  $R = u_{k+1} - u_1 = 110 - 60 = 50$  km/h.

**Câu 2:** Thời gian hoàn thành bài kiểm tra của các bạn trong lớp 12A được cho bằng bảng sau:

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

- Tính khoảng biến thiên  $R$  cho mẫu số liệu ghép nhóm trên.
- Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 27 phút và muộn nhất mất 43 phút. Hãy so sánh khoảng biến thiên của mẫu số liệu ghép nhóm và mẫu số liệu gốc.

### Lời giải

- Tính khoảng biến thiên  $R$  cho mẫu số liệu ghép nhóm trên.

Khoảng biến thiên  $R = u_{k+1} - u_1 = 45 - 25 = 20$  phút.

- Hãy so sánh khoảng biến thiên của mẫu số liệu ghép nhóm và mẫu số liệu gốc.

Khoảng biến thiên của mẫu số liệu gốc là  $43 - 27 = 16$  phút.

Vậy khoảng biến thiên của mẫu số liệu ghép nhóm nhỏ hơn khoảng biến thiên của mẫu số liệu gốc.

**Câu 3:** Để chuẩn bị mở một trung tâm thể dục thể thao, anh Tiến đã tiến hành điều tra tuổi thọ của máy chạy bộ (đơn vị: năm) do hai hãng X, Y sản xuất. Bảng biểu thị hai mẫu số liệu mà anh thu thập được qua Internet.

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

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

### Lời giải

Khoảng biến thiên của tuổi thọ máy chạy bộ do hãng X:  $R_X = 12 - 2 = 10$

Khoảng biến thiên của tuổi thọ máy chạy bộ do hãng Y:  $R_Y = 12 - 4 = 8$

Vì  $R_X > R_Y$  nên có thể nói là máy do hãng X sản xuất có tuổi thọ phân tán hơn so với máy của hãng Y.

**Câu 4:** Bảng dưới biểu thị kết quả điều tra thời gian sử dụng Internet hằng ngày của một số người.

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

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

### Lời giải

Đầu mút phải của nhóm ghép cuối cùng là 180,

Đầu mút trái của nhóm ghép đầu tiên là 30.

Vậy khoảng biến thiên của mẫu số liệu là  $R = 180 - 30 = 150$ .

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

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

Bảng. Điểm đánh giá của khán giả

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

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

### Lời giải

$$R_A = 100 - 50 = 50$$

$$R_B = 90 - 60 = 30$$

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

**Câu 6:** Một người ghi lại thời gian đàm thoại của một số cuộc gọi cho kết quả như bảng sau:

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

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

### Lời giải

| 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

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

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

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

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

**Câu 7:** Thầy Cư rất thích chạy bộ. Thời gian chạy bộ mỗi ngày trong thời gian gần đây của Thầy Cư 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          |

Hãy tính khoảng tứ phân vị của mẫu số liệu ghép nhóm trong bảng trên.

**Lời giải**

Cỡ mẫu  $n = 18$ .

Gọi  $x_1; x_2; \dots; x_{18}$  là mẫu số liệu gốc gồm thời gian của 18 ngày chạy bộ của Thầy Cư được sắp 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 ghép nhóm là:

$$Q_1 = 20 + \frac{\frac{18}{4} - 0}{6} \cdot (25 - 20) = 23,75.$$

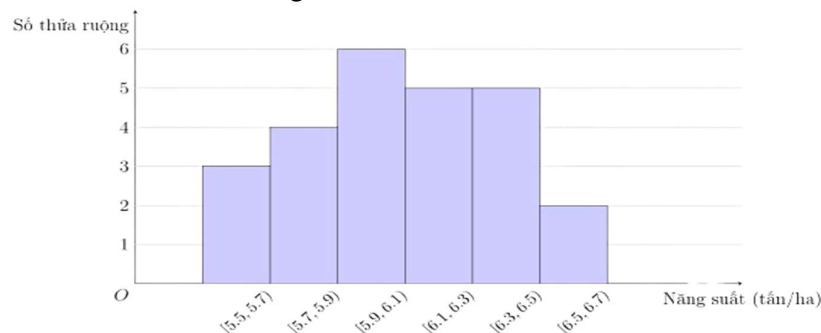
Tứ phân vị thứ ba của mẫu số liệu gốc là  $x_{14} \in [30; 35)$ .

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

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

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là:  $\Delta_Q = 31,875 - 23,75 = 8,125$ .

**Câu 8:** Khảo sát năng suất của một số thửa ruộng được minh họa ở biểu đồ sau:



a) Có bao nhiêu thửa ruộng được khảo sát?

b) Lập bảng tần số ghép nhóm

c) Hãy xác định khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu trên.

**Lời giải**

a) Có bao nhiêu thửa ruộng được khảo sát?

Có tất cả  $n = 3 + 4 + 6 + 5 + 5 + 2 = 25$  thửa ruộng được khảo sát.

b) Lập bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên.

Bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên.

| Năng suất (tấn/ha) | $[5,5;5,7)$ | $[5,7;5,9)$ | $[5,9;6,1)$ | $[6,1;6,3)$ | $[6,3;6,5)$ | $[6,5;6,7)$ |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Số thửa ruộng      | 3           | 4           | 6           | 5           | 5           | 2           |

c) Hãy xác định khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu trên.

+ Khoảng biến thiên của mẫu số liệu trên là:  $6,7 - 5,5 = 1,2$ .

+ Cỡ mẫu  $n = 25$ .

Gọi  $x_1; x_2; \dots; x_{25}$  là mẫu số liệu gốc gồm năng suất của 25 thửa ruộng được sắp 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  $\frac{1}{2}(x_6 + x_7) \in [5,7;5,9).$

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

$$Q_1 = 5,7 + \frac{\frac{25}{4} - 3}{4}(5,9 - 5,7) = 5,8625.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là  $\frac{1}{2}(x_{19} + x_{20}) \in [6,3;6,5).$

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

$$Q_3 = 6,3 + \frac{\frac{3 \cdot 25}{4} - (3 + 4 + 6 + 5)}{5}(6,5 - 6,3) = 6,33.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là:  $\Delta_Q = 6,33 - 5,8625 = 0,4675$ .

**Câu 9:** cho kết quả như sau:

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

(Theo premierleague.com)

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

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

**Lời giải**

a) Bảng số liệu ghép nhóm:

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

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

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

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

Vì  $n = 20$  nên tứ phân vị thứ nhất là trung vị của dãy số liệu: 42; 47; 50; 55; 55; 57; 59; 60; 61; 63. Do đó,

$$Q_1 = \frac{55 + 57}{2} = 56$$

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

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

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

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

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

Tứ phân vị thứ nhất của mẫu số liệu gốc là  $\frac{x_5 + x_6}{2}$  nên nhóm chứa tứ phân vị thứ nhất là nhóm [50; 60)

$$\text{và ta có: } Q'_1 = 50 + \frac{\frac{20}{4} - 2}{5} \cdot 10 = 56$$

Tứ phân vị thứ ba của mẫu số liệu gốc là  $\frac{x_{15} + x_{16}}{2}$  nên nhóm chứa tứ phân vị thứ ba là nhóm [70; 80)

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

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

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

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

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

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

**Lời giải**

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

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

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

$$\frac{6,5 \cdot 20 + 9,5 \cdot 35 + 12,5 \cdot 45 + 15,5 \cdot 35 + 18,5 \cdot 20}{20 + 35 + 45 + 35 + 20} = \frac{25}{2} \text{ (triệu đồng)}$$

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

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

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

$$\text{thứ nhất là: } Q'_1 = 8 + \frac{\frac{110}{4} - 17}{23} \cdot 3 = \frac{431}{46}$$

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

$$[14;17) \text{ và tứ phân vị thứ ba là: } Q'_3 = 14 + \frac{\frac{3 \cdot 110}{4} - (17 + 23 + 30)}{23} \cdot 3 = \frac{719}{46}$$

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

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

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

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

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

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

**Lời giải**

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

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

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

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

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

$$[165;170) \text{ và tứ phân vị thứ ba là: } Q_3 = 165 + \frac{\frac{3.43}{4} - (1+15+12)}{10} \cdot 5 = 167,125$$

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

$$\text{phân vị thứ ba là } Q_3' = 165 + \frac{\frac{3.42}{4} - (17+10)}{9} \cdot 5 = \frac{335}{2}$$

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

**Bảng a. Điểm của lớp A**

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

**Bảng b. Điểm của lớp B**

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

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

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

**c)** Mẫu số liệu nào có độ phân tán lớn hơn? Minh 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.

**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  $\Delta_Q^A$  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  $\Delta_Q^B$  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             |

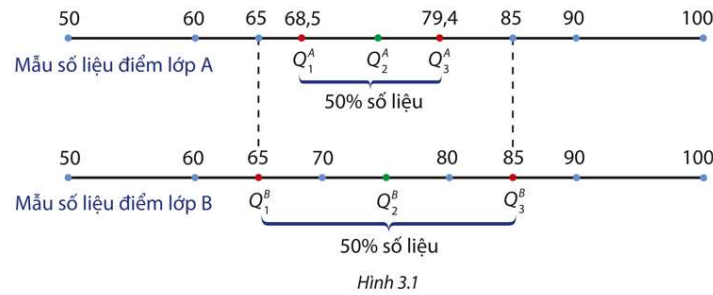
Nhóm chứa  $Q_1^B$  là [60; 70).  $Q_1^B = 60 + \frac{25-15}{20} \cdot 10 = 65$ .

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 13:** Kết quả đo chiều cao của 100 cây dừa trồng sau 10 năm tại một vườn trái cây ở Bến Tre 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, khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.  
b) Trong 100 cây dừa trên có 1 cây cao 8,4 m. Hỏi chiều cao của cây dừa này có phải là giá trị ngoại lệ không?

**Lời giải**

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

Khoảng biến thiên của mẫu số liệu là  $R = 9,4 - 8,4 = 1$ .

Ta có cỡ mẫu là  $n = 100$ .

Gọi  $x_1, x_2, \dots, x_{100}$  là mẫu số liệu gồm chiều cao của 100 cây dừa.

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 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

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

Tứ phân vị thứ ba của mẫu số liệu 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

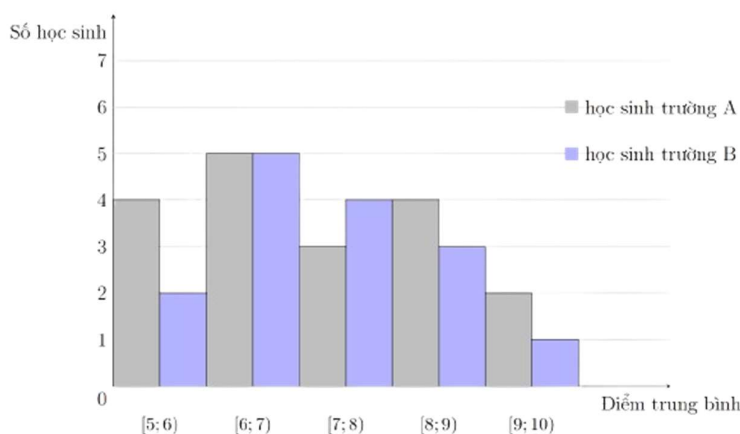
$$Q_3 = 9,0 + \frac{\frac{3 \cdot 100}{4} - (5 + 12 + 25)}{44} \cdot (9,2 - 9,0) = 9,15.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là  $\Delta_Q = Q_3 - Q_1 = 0,286$ .

- b) Hỏi chiều cao của cây dừa này có phải là giá trị ngoại lệ không?

Vì  $Q_1 - 1,5\Delta_Q = 8,435$  và  $Q_3 + 1,5\Delta_Q = 9,579$  nên cây dừa có chiều cao 8,4 m là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

**Câu 14:** Biểu đồ sau mô tả kết quả điều tra về điểm trung bình năm học của học sinh hai trường A và B.



- a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu trên.  
b) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đều hơn?

**Lời giải**

- a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu trên.  
Ta có:

| 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) Nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đều hơn?

▪ **Xét mẫu số liệu trường A**

Cỡ mẫu  $n = 18$ .

Gọi  $x_1, x_2, \dots, x_{18}$  là mẫu số liệu gốc gồm điểm trung bình của 18 học sinh trường A được sắp xếp theo thứ tự không giảm.

Ta có:  $x_1, \dots, x_4 \in [5;6)$ ;  $x_5, \dots, x_9 \in [6;7)$ ;  $x_{10}, x_{11}, x_{12} \in [7;8)$ ;  $x_{13}, \dots, x_{16} \in [8;9)$ ;  $x_{17}, x_{18} \in [9;10)$ .

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

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

$$Q_1 = 6 + \frac{\frac{18}{4} - 4}{5} (7 - 6) = 6,1.$$

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

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

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

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là:  $\Delta_{Q(A)} = 8,375 - 6,1 = 2,275$ .

▪ **Xét mẫu số liệu ở trường B**

Cỡ mẫu  $n = 15$ .

Gọi  $x_1, x_2, \dots, x_{15}$  là mẫu số liệu gốc gồm điểm trung bình của 15 học sinh trường B được sắp xếp theo thứ tự không giảm.

Ta có:  $x_1, x_2 \in [5;6)$ ;  $x_3, \dots, x_7 \in [6;7)$ ;  $x_8, \dots, x_{11} \in [7;8)$ ;  $x_{12}, x_{13}, x_{14} \in [8;9)$ ;  $x_{15} \in [9;10)$ .

Tứ phân vị thứ nhất của mẫu số liệu gốc là  $x_4 \in [6; 7)$ .

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

$$Q_1 = 6 + \frac{\frac{15}{4} - 2}{5} (7 - 6) = 6,35.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là  $x_{12} \in [8; 9)$ .

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

$$Q_3 = 8 + \frac{\frac{3 \cdot 15}{4} - (2 + 5 + 4)}{3} (9 - 8) = \frac{97}{12}.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là:  $\Delta_{Q(B)} = \frac{97}{12} - 6,35 = \frac{26}{15} \approx 1,73$ .

Nếu so sánh theo khoảng tứ phân vị thì học sinh trường B có điểm trung bình đồng đều hơn.

**Câu 15:** Thầy giáo thống kê lại điểm trung bình cuối năm của các học sinh lớp 11 A và 11B ở bảng sau:

| Điểm trung bình | $[5; 6)$ | $[6; 7)$ | $[7; 8)$ | $[8; 9)$ | $[9; 10)$ |
|-----------------|----------|----------|----------|----------|-----------|
| 11A             | 1        | 0        | 11       | 22       | 6         |
| 11B             | 0        | 6        | 8        | 14       | 12        |

- a) Nếu so sánh theo khoảng biến thiên thì học sinh lớp nào có điểm trung bình ít phân tán hơn?  
 b) Nếu so sánh theo độ lệch chuẩn thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

**Lời giải**

- a) Nếu so sánh theo khoảng biến thiên thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

Khoảng biến thiên của điểm số học sinh lớp 11A là:  $10 - 5 = 5$ .

Khoảng biến thiên của điểm số học sinh lớp 11B là:  $10 - 6 = 4$ .

Nếu so sánh theo **khoảng biến thiên** thì điểm trung bình của các học sinh lớp 11B ít phân tán hơn điểm trung bình của các học sinh lớp 11A

- b) Nếu so sánh theo độ lệch chuẩn thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

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 |
|------------------|-----|-----|-----|-----|-----|
| 11A              | 1   | 0   | 11  | 22  | 6   |
| 11B              | 0   | 6   | 8   | 14  | 12  |

**Xét mẫu số liệu của lớp 11A:**

Cỡ mẫu là  $n_1 = 1 + 11 + 22 + 6 = 40$ .

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{1 \cdot 5,5 + 11 \cdot 7,5 + 22 \cdot 8,5 + 6 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{40} (1 \cdot 5,5^2 + 11 \cdot 7,5^2 + 22 \cdot 8,5^2 + 6 \cdot 9,5^2) - 8,3^2 = 0,61.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là  $S_1 = \sqrt{0,61}$ .

**Xét mẫu số liệu của lớp 11B:**

Cỡ mẫu là  $n_2 = 6 + 8 + 14 + 12 = 40$ .

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{6 \cdot 6,5 + 8 \cdot 7,5 + 14 \cdot 8,5 + 12 \cdot 9,5}{40} = 8,3.$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{40} (6 \cdot 6,5^2 + 8 \cdot 7,5^2 + 14 \cdot 8,5^2 + 12 \cdot 9,5^2) - 8,3^2 = 1,06.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là  $S_2 = \sqrt{1,06}$ .

Do  $S_1 < S_2$  nên nếu so sánh theo **độ lệch chuẩn** thì học sinh lớp 11A có điểm trung bình ít phân tán hơn học sinh lớp 11B

**Nhận xét:**

- Trong bài tập trên, kết quả so sánh độ phân tán theo giá trị trung bình và độ lệch chuẩn có sự khác biệt. Điều này là do mẫu số liệu của học sinh lớp 11A có một giá trị ngoại lệ.
- Với các mẫu số liệu ghép nhóm có cùng số trung bình (hoặc xấp xỉ nhau), ta thường sử dụng phương sai và độ lệch chuẩn để so sánh mức độ phân tán của các mẫu số liệu đó.

**Câu 16:** Cân nặng (kg) 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 cây giống  | 6     | 12    | 19     | 9       | 4       |

Hãy tính phương sai của mẫu số liệu ghép nhóm trên.

**Lời giải**

Ta có giá trị đại diện được thể hiện trong bảng sau:

| Cân nặng (kg)    | [4;6) | [6;8) | [8;10) | [10;12) | [12;14) |
|------------------|-------|-------|--------|---------|---------|
| Giá trị đại diện | 5     | 7     | 9      | 11      | 13      |
| Số quả mít       | 6     | 12    | 19     | 9       | 4       |

Cỡ mẫu:  $n = 50$ .

$$\text{Số trung bình } \bar{x} = \frac{m_1 \cdot x_1 + m_2 \cdot x_2 + \dots + m_k \cdot x_k}{n} = \frac{6 \cdot 5 + 12 \cdot 7 + 19 \cdot 9 + 9 \cdot 11 + 4 \cdot 13}{50} = 8,72.$$

$$\begin{aligned} \text{Phương sai: } s^2 &= \frac{1}{n} (m_1 \cdot x_1^2 + m_2 \cdot x_2^2 + \dots + m_k \cdot x_k^2) - (\bar{x})^2 \\ &= \frac{1}{50} (6 \cdot 5^2 + 12 \cdot 7^2 + 19 \cdot 9^2 + 9 \cdot 11^2 + 4 \cdot 13^2) - (8,72)^2 = 4,8016. \end{aligned}$$

**Câu 17:** Thống kê điểm trung bình của hai lớp 12A và 12B được cho ở bảng sau

| Điểm trung bình     | [5;6) | [6;7) | [7;8) | [8;9) | [9;10) |
|---------------------|-------|-------|-------|-------|--------|
| Số học sinh lớp 12A | 1     | 0     | 11    | 22    | 6      |
| Số học sinh lớp 12B | 0     | 6     | 8     | 14    | 12     |

Nếu so sánh phương sai thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

**Lời giải**

Ta có bảng thống kê điểm trung bình theo giá trị đại diện như sau:

| Điểm trung bình     | [5;6) | [6;7) | [7;8) | [8;9) | [9;10) |
|---------------------|-------|-------|-------|-------|--------|
| Giá trị đại diện    | 5,5   | 6,5   | 7,5   | 8,5   | 9,5    |
| Số học sinh lớp 12A | 1     | 0     | 11    | 22    | 6      |
| Số học sinh lớp 12B | 0     | 6     | 8     | 14    | 12     |

**Xét mẫu số liệu lớp 12A:**

Cỡ mẫu:  $n_1 = 40$ .

$$\text{Số trung bình: } \bar{x} = \frac{m_1 \cdot x_1 + m_2 \cdot x_2 + \dots + m_k \cdot x_k}{n_1} = \frac{1 \cdot 5,5 + 0 \cdot 6,5 + 11 \cdot 7,5 + 22 \cdot 8,5 + 6 \cdot 9,5}{40} = 8,3.$$

Phương sai:  $s_1^2 = \frac{1}{n_1} (m_1 \cdot x_1^2 + m_2 \cdot x_2^2 + \dots + m_k \cdot x_k^2) - (\bar{x})^2$

$$= \frac{1}{40} (1.5 \cdot 5^2 + 0.6 \cdot 5^2 + 11.7 \cdot 5^2 + 22.8 \cdot 5^2 + 6.9 \cdot 5^2) - (8,3)^2 = 0,61.$$

**Xét mẫu số liệu lớp 12B:**

Cỡ mẫu:  $n_2 = 40$ .

Số trung bình:  $\bar{y} = \frac{m_1 \cdot y_1 + m_2 \cdot y_2 + \dots + m_k \cdot y_k}{n_2} = \frac{0.5 \cdot 5 + 6.6 \cdot 5 + 8.7 \cdot 5 + 14.8 \cdot 5 + 12.9 \cdot 5}{40} = 8,3$ .

Phương sai:  $s_2^2 = \frac{1}{n_2} (m_1 \cdot y_1^2 + m_2 \cdot y_2^2 + \dots + m_k \cdot y_k^2) - \bar{y}^2$

$$= \frac{1}{40} (0.5 \cdot 5^2 + 6.6 \cdot 5^2 + 8.7 \cdot 5^2 + 14.8 \cdot 5^2 + 12.9 \cdot 5^2) - (8,3)^2 = 1,06.$$

→ Vì  $s_1^2 < s_2^2$  nên điểm trung bình của học sinh lớp 12B phân tán ít hơn điểm trung bình của học sinh lớp 12A

**Câu 18:** Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi hai phân xưởng được cho như sau:

| Tuổi thọ (năm)                | $[1,5;2)$ | $[2;2,5)$ | $[2,5;3)$ | $[3;3,5)$ | $[3,5;4)$ |
|-------------------------------|-----------|-----------|-----------|-----------|-----------|
| Số linh kiện của phân xưởng 1 | 4         | 9         | 13        | 8         | 6         |
| Số linh kiện của phân xưởng 2 | 2         | 8         | 20        | 7         | 3         |

Tính phương sai của mỗi mẫu số liệu ghép nhóm và nhận xét về độ phân tán của tuổi thọ các linh kiện điện tử được sản xuất bởi mỗi phân xưởng.

**Lời giải**

Ta có bảng thống kê điểm trung bình theo giá trị đại diện như sau:

| 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         |

**Xét mẫu số liệu Phân xưởng 1:**

Tổng số linh kiện:  $4 + 9 + 13 + 8 + 6 = 40$

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:  $s_1^2 = \frac{1}{40} (1,75^2 \cdot 4 + 2,25^2 \cdot 9 + 2,75^2 \cdot 13 + 3,25^2 \cdot 8 + 3,75^2 \cdot 6) - \left( \frac{223}{80} \right)^2 = \frac{2271}{6400} \approx 0,354$

**Xét mẫu số liệu Phân xưởng 2:**

Tổng số linh kiện:  $2 + 8 + 20 + 7 + 3 = 40$

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:  $s_2^2 = \frac{1}{40} (1,75^2 \cdot 2 + 2,25^2 \cdot 8 + 2,75^2 \cdot 20 + 3,25^2 \cdot 7 + 3,75^2 \cdot 3) - \left( \frac{221}{80} \right)^2 = \frac{1399}{6400} \approx 0,218$

→ Vì  $s_1^2 > s_2^2$  nên số linh kiện phân xưởng 2 phân tán nhiều hơn số linh kiện phân xưởng 1.

**Câu 19:** Cho mẫu số liệu ghép nhóm về độ tuổi của cư dân trong một khu phố.

| Độ tuổi          | [20;30) | [30;40) | [40;50) | [50;60) | [60;70) | [70;80) |
|------------------|---------|---------|---------|---------|---------|---------|
| Giá trị đại diện | 25      | 35      | 45      | 55      | 65      | 75      |
| Số cư dân        | 25      | 20      | 20      | 15      | 14      | 6       |

Tính phương sai và độ lệch chuẩn của mẫu số liệu đó. (Kết quả làm tròn đến hàng phần mười)

**Lời giải**

$$\begin{aligned} \text{Số trung bình cộng là } \bar{x} &= \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} \\ &= \frac{25 \cdot 25 + 20 \cdot 35 + 20 \cdot 45 + 15 \cdot 55 + 14 \cdot 65 + 6 \cdot 75}{100} = 44,1 \text{ (tuổi)}. \end{aligned}$$

$$\begin{aligned} \text{Phương sai là } S^2 &= \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} \\ S^2 &= \frac{25(25 - \bar{x})^2 + 20(35 - \bar{x})^2 + 20(45 - \bar{x})^2 + 15(55 - \bar{x})^2 + 14(65 - \bar{x})^2 + 6(75 - \bar{x})^2}{100} = 244,2. \end{aligned}$$

$$\text{Độ lệch chuẩn là } S = \sqrt{S^2} \approx 15,6 \text{ (tuổi)}.$$

**Câu 20:** Cho hai mẫu số liệu ghép nhóm thống kê mức lương của hai công ty A, B (đơn vị là triệu đồng)

| Bảng thống kê mức lương công ty A<br>(triệu đồng) |                    | Bảng thống kê mức lương công ty B<br>(triệu đồng) |                    |
|---------------------------------------------------|--------------------|---------------------------------------------------|--------------------|
| Mức lương                                         | Số lượng nhân viên | Mức lương                                         | Số lượng nhân viên |
| [10;15)                                           | 15                 | [10;15)                                           | 25                 |
| [15;20)                                           | 18                 | [15;20)                                           | 15                 |
| [20;25)                                           | 10                 | [20;25)                                           | 7                  |
| [25;30)                                           | 10                 | [25;30)                                           | 5                  |
| [30;35)                                           | 5                  | [30;35)                                           | 5                  |
| [35;40)                                           | 2                  | [35;40)                                           | 3                  |

a) Tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm lần lượt biểu diễn mức lương của hai công ty A, B. (Kết quả làm tròn đến hàng phần mười)

b) Công ty nào có mức lương đồng đều hơn?

**Lời giải**

a) Tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm lần lượt biểu diễn mức lương của hai công ty A, B. (Kết quả làm tròn đến hàng phần mười)

**Xét Mức lương tại công ty A**

$$\text{Số trung bình của công ty A là } \bar{x}_A = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = \frac{62}{3} \approx 20,7 \text{ (triệu đồng)}.$$

$$\text{Phương sai của công ty A là } S_A^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = \frac{1769}{36}.$$

$$\text{Độ lệch chuẩn của công ty A là } S_A = \sqrt{S_A^2} \approx 7 \text{ (triệu đồng)}.$$

**Xét Mức lương tại công ty B**

$$\text{Số trung bình của công ty B là } \bar{x}_B = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = \frac{229}{12} \approx 19,1 \text{ (triệu đồng)}.$$

Phương sai của công ty B là  $S_B^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = \frac{8339}{144}$ .

Độ lệch chuẩn của công ty B là  $S_B = \sqrt{S_B^2} \approx 7,6$  (triệu đồng).

b) Công ty nào có mức lương đồng đều hơn?

Vì độ lệch chuẩn của công ty A nhỏ hơn độ lệch chuẩn của công ty B và lương bình quân của cả hai công ty xấp xỉ nhau nên công ty A có mức lương đồng đều hơn.

**Câu 21:** 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 là 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   | 9         | 8         | 11        | 13        | 9         |
| Cổ phiếu B   | 14        | 6         | 5         | 20        | 5         |

a) Tính phương sai và độ lệch chuẩn của hai mã cổ phiếu A, B. (Kết quả làm tròn đến hàng phần mười)

b) So sánh độ rủi ro của cổ phiếu A và B?

**Lời giải**

a) Tính phương sai và độ lệch chuẩn của hai mã cổ phiếu A, B.

Số trung bình của cổ phiếu A là  $\bar{x}_A = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = 125,2$ .

Phương sai của cổ phiếu A là  $S_A^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = 7,4$ .

Độ lệch chuẩn của cổ phiếu A là  $S_A = \sqrt{S_A^2} \approx 2,7$ .

Số trung bình của cổ phiếu B là  $\bar{x}_B = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = 124,8$ .

Phương sai của cổ phiếu B là  $S_B^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = 8,1$ .

Độ lệch chuẩn của cổ phiếu B là  $S_B = \sqrt{S_B^2} \approx 2,9$ .

b) So sánh độ rủi ro của cổ phiếu A và B?

Vì độ lệch chuẩn về giá của cổ phiếu A nhỏ hơn độ lệch chuẩn về giá của cổ phiếu B nên cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B.

**Câu 22:** Chỉ số P/B (viết tắt của từ Price to book ratio) là một chỉ số tài chính quan trọng, được sử dụng để so sánh giữa giá của cổ phiếu với giá trị ghi sổ của cổ phiếu đó. Khi chỉ số P/B cao, nghĩa là thị trường đang có nhiều kỳ vọng đối với cổ phiếu này, doanh nghiệp có tiềm năng phát triển tốt trong tương lai. Bảng sau thống kê chỉ số P/B của hai mã cổ phiếu A và B trong 10 năm giao dịch liên tiếp.

| Giá đóng cửa | [0;2) | [2;4) | [4;6) | [6;8) | [8;10) | [10;12) |
|--------------|-------|-------|-------|-------|--------|---------|
| Cổ phiếu A   | 1     | 0     | 4     | 2     | 2      | 1       |
| Cổ phiếu B   | 1     | 2     | 1     | 2     | 3      | 1       |

a) Tính phương sai và độ lệch chuẩn của hai mã cổ phiếu A, B. (Kết quả làm tròn đến hàng phần mười)

b) So sánh độ rủi ro của cổ phiếu A và B? Biết rằng các chỉ số tài chính còn lại của hai cổ phiếu đó là xấp xỉ nhau.

**Lời giải**

a) Tính phương sai và độ lệch chuẩn của hai mã cổ phiếu A, B.

Số trung bình của cổ phiếu A là  $\bar{x}_A = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = 6,4$ .

Phương sai của cổ phiếu A là  $S_A^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = 7,2$ .

Độ lệch chuẩn của cổ phiếu A là  $S_A = \sqrt{S_A^2} \approx 2,7$ .

Số trung bình của cổ phiếu B là  $\bar{x}_B = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = 6,4$ .

Phương sai của cổ phiếu B là  $S_B^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N} = 9,6$ .

Độ lệch chuẩn của cổ phiếu B là  $S_B = \sqrt{S_B^2} \approx 3,1$ .

b) So sánh độ rủi ro của cổ phiếu A và B?

Vì độ lệch chuẩn về chỉ số P/B của cổ phiếu A nhỏ hơn độ lệch chuẩn của cổ phiếu B nên cổ phiếu A có độ rủi ro thấp hơn cổ phiếu B.

**Câu 23:** Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

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

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

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

c) Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

**Lời giải**

| Giá trị đại diện  | 6,5 | 7,5 | 8,5 | 9,5 | 10,5 |
|-------------------|-----|-----|-----|-----|------|
| Học sinh trường X | 8   | 10  | 13  | 10  | 9    |
| Học sinh trường Y | 4   | 12  | 17  | 14  | 3    |

a) Cỡ mẫu:  $n = 50$

Xét số liệu của trường X:

Số trung bình:  $\bar{x}_X = \frac{8.6,5 + 10.7,5 + 13.8,5 + 10.9,5 + 9.10,5}{50} = 8,54$

Xét số liệu của trường Y:

Số trung bình:  $\bar{x}_Y = \frac{4.6,5 + 12.7,5 + 17.8,5 + 14.9,5 + 3.10,5}{50} = 8,5$

Vậy nếu so sánh theo số trung bình thì học sinh trường Y viết nhanh hơn

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

Ta có:  $x_1; \dots; x_8 \in [6; 7); x_9; \dots; x_{18} \in [7; 8); x_{19}; \dots; x_{31} \in [8; 9); x_{32}; \dots; x_{41} \in [9; 10); x_{42}; \dots; x_{50} \in [10; 11)$

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

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

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

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

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 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

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

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 Y có tốc độ viết đồng đều hơn

**c)** Xét số liệu của trường X :

$$\text{Độ lệch chuẩn: } \sigma_X = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,54^2} \approx 1,33$$

Xét số liệu của trường Y :

$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{4.6,5^2 + 12.7,5^2 + 17.8,5^2 + 14.9,5^2 + 3.10,5^2}{50} - 8,5^2} \approx 1,04$$

Vậy nếu so sánh theo độ lệch chuẩn thì học sinh trường Y có tốc độ viết đồng đều hơn

**Câu 24:** 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 ( $\mu m$ ) | [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 ( $\mu m$ ) | [4,5;5) | [5;5,5) | [5,5;6) | [6;6,5) |
| Giá trị đại diện       | 4,75    | 5,25    | 5,75    | 6,25    |
| Số học sinh            | 3       | 8       | 7       | 2       |

$$\text{Số trung bình: } \bar{x} = \frac{4,75.3 + 5,25.8 + 5,75.7 + 6,25.2}{20} = \frac{109}{20} (\mu m)$$

$$\text{Phương sai: } s^2 = \frac{1}{20} (4,75^2.3 + 5,25^2.8 + 5,75^2.7 + 6,25^2.2) - \left( \frac{109}{20} \right)^2 = \frac{37}{200}$$

$$\text{Độ lệch chuẩn: } s = \sqrt{\frac{37}{200}} \approx 0,43$$

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 25:** Thời gian chạy tập luyện cự li 100 m của hai vận động viên được cho trong bảng sau:

| Thời gian (giây)  | [10;10,3) | [10,3;10,6) | [10,6;10,9) | [10,9;11,2) |
|-------------------|-----------|-------------|-------------|-------------|
| Số lần chạy của A | 2         | 10          | 5           | 3           |
| Số lần chạy của B | 3         | 7           | 9           | 6           |

Dựa trên độ lệch chuẩn của các mẫu số liệu ghép nhóm, hãy cho biết vận động viên nào có thành tích luyện tập ổn định hơn.

**Lời giải**

Ta có bảng số liệu với giá trị đại diện:

| Thời gian (giây)  | [10;10,3) | [10,3;10,6) | [10,6;10,9) | [10,9;11,2) |
|-------------------|-----------|-------------|-------------|-------------|
| Giá trị đại diện  | 10,15     | 10,45       | 10,75       | 11,05       |
| Số lần chạy của A | 2         | 10          | 5           | 3           |
| Số lần chạy của B | 3         | 7           | 9           | 6           |

Vận động viên A:

$$\text{Giá trị trung bình } \bar{x}_A = \frac{10,15 \cdot 2 + 10,45 \cdot 10 + 10,75 \cdot 5 + 11,05 \cdot 3}{2 + 10 + 5 + 3} = \frac{2117}{20}$$

$$\text{Phương sai: } s_A^2 = \frac{1}{20} (10,15^2 \cdot 2 + 10,45^2 \cdot 10 + 10,75^2 \cdot 5 + 11,05^2 \cdot 3) - \left( \frac{2117}{20} \right)^2 = \frac{2691}{40000}$$

$$\text{Độ lệch chuẩn: } s_A = \sqrt{\frac{2691}{40000}} \approx 0,26$$

Vận động viên B:

$$\text{Giá trị trung bình } \bar{x}_B = \frac{10,15 \cdot 3 + 10,45 \cdot 7 + 10,75 \cdot 9 + 11,05 \cdot 6}{3 + 7 + 9 + 6} = \frac{5333}{500}$$

$$\text{Phương sai: } s_B^2 = \frac{1}{25} (10,15^2 \cdot 3 + 10,45^2 \cdot 7 + 10,75^2 \cdot 9 + 11,05^2 \cdot 6) - \left( \frac{5333}{500} \right)^2 = \frac{1296}{15625}$$

$$\text{Độ lệch chuẩn: } s_B = \sqrt{\frac{1296}{15625}} = 0,288$$

Vì  $s_A < s_B$  nên vận động viên A có thành tích luyện tập ổn định hơn.

**Câu 26:** Mai và Ngọc cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

| Số bước (đơn vị: nghìn) | [3;5) | [5;7) | [7;9) | [9;11) | [11;13) |
|-------------------------|-------|-------|-------|--------|---------|
| Mai                     | 6     | 7     | 6     | 6      | 5       |
| Ngọc                    | 2     | 5     | 13    | 8      | 2       |

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

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

**Lời giải**

| Giá trị đại diện | 4 | 6 | 8  | 10 | 12 |
|------------------|---|---|----|----|----|
| Mai              | 6 | 7 | 6  | 5  |    |
| Ngọc             | 2 | 5 | 13 | 8  | 2  |

a) Cỡ mẫu:  $n = 30$

- Xét mẫu số liệu của Mai:

$$\text{Số trung bình: } \bar{x}_1 = \frac{6.4 + 7.6 + 6.8 + 6.10 + 5.12}{30} = 7,8$$

$$\text{Phương sai: } S_1^2 = \frac{(6.4^2 + 7.6^2 + 6.8^2 + 6.10^2 + 5.12^2)}{30} - 7,8^2 = 7,56$$

$$\text{Độ lệch chuẩn: } \sigma_1 = \sqrt{7,56} \approx 2,75$$

- Xét mẫu số liệu của Ngọc:

$$\text{Số trung bình: } \bar{x}_2 = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$$

$$\text{Phương sai: } S_2^2 = \frac{(2.4^2 + 5.6^2 + 13.8^2 + 8.10^2 + 2.12^2)}{30} - 8,2^2 \approx 3,83$$

$$\text{Độ lệch chuẩn: } \sigma_2 = \sqrt{3,83} \approx 1,96$$

**b)** Nếu so sánh theo độ lệch chuẩn thì bạn Ngọc có số lượng bước chân đi mỗi ngày đều đặn hơn  
Kết quả các phép tính làm tròn đến hàng phần nghìn.

**Câu 27:** Tốc độ của 20 xe hơi khi đi qua một trạm kiểm tra tốc độ (đơn vị: km/h) được thống kê lại như sau:

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

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

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

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

#### Lời giải

**a)** Khoảng biến thiên của mẫu số liệu:  $61,1 - 42 = 19,1 (km/h)$

Cỡ mẫu:  $n = 20$

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

$$\text{Trung vị } Q_2 = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(48,4 + 50,8) = 49,6$$

$$\text{Tứ phân vị thứ nhất là trung vị của nửa số liệu bên trái } Q_1 = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(46,7 + 46,8) = 46,75$$

$$\text{Tứ phân vị thứ ba là trung vị của nửa số liệu bên phải } Q_3 = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(54,8 + 55,6) = 55,2$$

$$\text{Khoảng tứ phân vị của mẫu số liệu là: } \Delta_Q = Q_3 - Q_1 = 8,45$$

$$\text{Số trung bình: } \bar{x} = \frac{42 + 43,4 + \dots + 61,1}{20} = 50,945$$

$$\text{Phương sai: } S^2 = \frac{42^2 + 43,4^2 + \dots + 61,1^2}{20} - 50,945^2 \approx 32,2$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{32,2} \approx 5,67$$

**b)**

| Tốc độ | $[42; 46)$ | $[46; 50)$ | $[50; 54)$ | $[54; 58)$ | $[58; 62)$ |
|--------|------------|------------|------------|------------|------------|
| Số xe  | 3          | 7          | 4          | 3          | 3          |

**c)** Ta có:  $x_1; \dots; x_3 \in [42; 46); x_4; \dots; x_{10} \in [46; 50); x_{11}; \dots; x_{14} \in [50; 54); x_{15}; \dots; x_{17} \in [54; 58);$

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

Tứ phân vị thứ nhất của mẫu số liệu gốc là  $\frac{1}{2}(x_5 + x_6) \in [46; 50)$ . Do đó, tứ phân vị thứ nhất của mẫu

$$\text{số liệu ghép nhóm là: } Q_1' = 46 + \frac{\frac{20}{4} - 3}{7} (50 - 46) = \frac{330}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là  $\frac{1}{2}(x_{15} + x_{16}) \in [54; 58)$ . Do đó, tứ phân vị thứ ba của mẫu số

$$\text{liệu ghép nhóm là: } Q_3' = 54 + \frac{\frac{3.20}{4} - (3 + 7 + 4)}{3} (58 - 54) = \frac{166}{3}$$

$$\text{Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: } \Delta_Q' = Q_3' - Q_1' = \frac{172}{21}$$

|                  |    |    |    |    |    |
|------------------|----|----|----|----|----|
| Giá trị đại diện | 44 | 48 | 52 | 56 | 60 |
| Số xe            | 3  |    | 4  | 3  | 3  |

$$\text{Số trung bình: } \bar{x} = \frac{3.44 + 7.48 + 4.52 + 3.56 + 3.60}{20} = 41,8$$

$$\text{Phương sai: } S^2 = \frac{3.44^2 + 7.48^2 + 4.52^2 + 3.56^2 + 3.60^2}{20} - 41,8^2 = 364,96$$

$$\text{Độ lệch chuẩn: } \sigma = \sqrt{364,96} = 19,1$$

**Câu 28:** 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 |    |    |

- 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  $[5; 7)$ .
- 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).
- So sánh và nêu ý nghĩa các kết quả tìm được ở câu b với các kết quả tương ứng của mẫu số liệu gốc.

#### Lời giải

a) Căn cứ vào giá trị nhỏ nhất và giá trị lớn nhất của mẫu số liệu, ta thấy khoảng biến thiên là  $R = 9$  nên nếu muốn mỗi nhóm có độ dài bằng 2 và nhóm đầu tiên là  $[5; 7)$  thì phải chia số liệu thành 5 nhóm, với 4 nhóm ghép tiếp theo là  $[7; 9), [9; 11), [11; 13), [13; 15)$ . Đếm số lần xuất hiện các giá trị trong mỗi nhóm, ta lập được bảng sau:

|        |          |          |           |            |            |
|--------|----------|----------|-----------|------------|------------|
| Nhóm   | $[5; 7)$ | $[7; 9)$ | $[9; 11)$ | $[11; 13)$ | $[13; 15)$ |
| Tần số | 2        | 4        | 9         | 6          | 4          |

b) Bổ sung thêm giá trị đại diện của mỗi nhóm vào bảng lập ở trên, ta có

|                  |          |          |           |            |            |          |
|------------------|----------|----------|-----------|------------|------------|----------|
| Nhóm             | $[5; 7)$ | $[7; 9)$ | $[9; 11)$ | $[11; 13)$ | $[13; 15)$ |          |
| Giá trị đại diện | 6        | 8        | 10        | 12         | 14         |          |
| Tần số           | 2        | 4        | 9         | 6          | 4          | $N = 25$ |

Sử dụng Bảng 3.15b để tính số trung bình  $\bar{x}_N$  của mẫu số liệu ghép nhóm:

$$\bar{x}_N = \frac{2.6 + 4.8 + 9.10 + 6.12 + 4.14}{25} = \frac{262}{25} = 10,48 \text{ (giờ)}$$

Suy ra phương sai  $s_N^2$  của mẫu số liệu ghép nhóm là:

$$s_N^2 = \frac{2(6-10,48)^2 + 4(8-10,48)^2 + 9(10-10,48)^2 + 6(12-10,48)^2 + 4(14-10,48)^2}{25}$$

$$= \frac{130,24}{25} = 5,2096 \approx 5,21.$$

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  | $N = 25$ |

Sử dụng các công thức tính trung bình, phương sai, độ lệch chuẩn của mẫu số liệu không ghép nhóm, ta có:

$$\bar{x}_G = \frac{5 + 6 + 7 + 3 \cdot 8 + 5 \cdot 9 + 4 \cdot 10 + 3 \cdot 11 + 3 \cdot 12 + 2 \cdot 13 + 2 \cdot 14}{25} = 10.$$

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.

**Quý phụ huynh và các em học sinh đky khóa học**

**1. Zalo thầy Thuận Toán: 0869998668**

[zalo.me/84869998668](https://www.zalo.me/84869998668)

**2. Facebook thầy:**

<https://www.facebook.com/Thaygiaothuan.99/>

**3. Fanpage thầy:**

<https://www.facebook.com/thaythuantoan>

**4. Tiktok thầy:**

[https://www.tiktok.com/@thay\\_hothucthuan](https://www.tiktok.com/@thay_hothucthuan)

**5. Kênh Youtube thầy:**

<https://www.youtube.com/@HoThucThuanOfficial>