

PHẦN B. BÀI TẬP TỰ LUẬN

Câu 1. Thời gian hoàn thành một bài viết chính tả của một số học sinh lớp 4 hai trường X và Y được ghi lại ở bảng sau:

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

- Nếu so sánh theo số trung bình thì học sinh trường nào viết nhanh hơn?
- Nếu so sánh theo khoảng tứ phân vị thì học sinh trường nào có tốc độ viết đồng đều hơn?
- Nếu so sánh theo độ lệch chuẩn thì học sinh trường nào có tốc độ viết đồng đều hơn?

Lời giải

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

a) Cỡ mẫu: $n = 50$

Xét số liệu của trường X :

$$\text{Số trung bình: } \bar{x}_x = \frac{8 \cdot 6,5 + 10 \cdot 7,5 + 13 \cdot 8,5 + 10 \cdot 9,5 + 9 \cdot 10,5}{50} = 8,54$$

Xét số liệu của trường Y :

$$\text{Số trung bình: } \bar{x}_y = \frac{4 \cdot 6,5 + 12 \cdot 7,5 + 17 \cdot 8,5 + 14 \cdot 9,5 + 3 \cdot 10,5}{50} = 8,5$$

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

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

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

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

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

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

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

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

$$Q_1 = 7 + \frac{\frac{50}{4} - 4}{12} (8 - 7) = \frac{185}{24}$$

ghép nhóm là:

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

$$Q_3 = 9 + \frac{\frac{3.50}{4} - (4+12+17)}{14} (10 - 9) = \frac{261}{28}$$

nhóm là:

$$\Delta_Q = Q_3 - Q_1 = \frac{271}{168}$$

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

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 :

$$\sigma_x = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,54^2} \approx 1,33$$

Độ lệch chuẩn:

Xét số liệu của trường Y :

$$\sigma_y = \sqrt{\frac{4.6,5^2 + 12.7,5^2 + 17.8,5^2 + 14.9,5^2 + 3.10,5^2}{50} - 8,5^2} \approx 1,04$$

Độ lệch chuẩn:

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

Câu 2. Tuổi thọ của một số linh kiện điện tử (đơn vị: năm) được sản xuất bởi hai phân xưởng được cho như sau:

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

Số linh kiện của phân xưởng 2	2	8	20	7	3
-------------------------------	---	---	----	---	---

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

Lời giải

Phân xưởng 1: Tổng số linh kiện: $4 + 9 + 13 + 8 + 6 = 40$

$$\bar{x}_1 = \frac{1,75 \cdot 4 + 2,25 \cdot 9 + 2,75 \cdot 13 + 3,25 \cdot 8 + 3,75 \cdot 6}{4 + 9 + 13 + 8 + 6} = \frac{223}{40}$$

$$s_1^2 = \frac{1}{40} (1,75^2 \cdot 4 + 2,25^2 \cdot 9 + 2,75^2 \cdot 13 + 3,25^2 \cdot 8 + 3,75^2 \cdot 6) - \left(\frac{223}{40} \right)^2 = \frac{2271}{1600}$$

$$s_1 = \sqrt{\frac{2271}{1600}} \approx 1,2$$

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

$$\bar{x}_2 = \frac{1,75 \cdot 2 + 2,25 \cdot 8 + 2,75 \cdot 20 + 3,25 \cdot 7 + 3,75 \cdot 3}{2 + 8 + 20 + 7 + 3} = \frac{221}{40}$$

$$s_2^2 = \frac{1}{40} (1,75^2 \cdot 2 + 2,25^2 \cdot 8 + 2,75^2 \cdot 20 + 3,25^2 \cdot 7 + 3,75^2 \cdot 3) - \left(\frac{221}{40} \right)^2 = \frac{1399}{1600}$$

$$s_2 = \sqrt{\frac{1399}{1600}} \approx 0,94$$

Vì $1,2 > 0,94$ nên độ phân tán của phân xưởng 1 lớn hơn độ phân tán của phân xưởng 2.

Câu 3. Một nhóm 20 học sinh dùng một thiết bị đo đường kính của một nhân tế bào cho kết quả như sau:

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

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

b) Số trung bình và độ lệch chuẩn cho biết thông tin gì?

Lời giải

a) Mẫu số liệu ghép nhóm với giá trị đại diện:

Kết quả đo (μm)	[4,5;5)	[5;5,5)	[5,5;6)	[6;6,5)
Giá trị đại diện	4,75	5,25	5,75	6,25
Số học sinh	3	8	7	2

$$\bar{x} = \frac{4,75 \cdot 3 + 5,25 \cdot 8 + 5,75 \cdot 7 + 6,25 \cdot 2}{20} = \frac{109}{20} (\mu m)$$

Số trung bình:

$$\text{Phương sai: } s^2 = \frac{1}{20} (4,75^2 \cdot 3 + 5,25^2 \cdot 8 + 5,75^2 \cdot 7 + 6,25^2 \cdot 2) - \left(\frac{109}{20} \right)^2 = \frac{37}{200}$$

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

b) Số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm xấp xỉ với số trung bình và độ lệch chuẩn của mẫu số liệu gốc. Do đó, mẫu số liệu gốc có số trung bình xấp xỉ $\frac{109}{20} \mu m$ và độ lệch chuẩn xấp xỉ $0,43 \mu m$.

Câu 4. Thời gian chạy tập luyện cự li 100 m của hai vận động viên được cho trong bảng sau:

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

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

Lời giải

Ta có bảng số liệu với giá trị đại diện:

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

Vận động viên A:

$$\text{Giá trị trung bình } \bar{x}_A = \frac{10,15 \cdot 2 + 10,45 \cdot 10 + 10,75 \cdot 5 + 11,05 \cdot 3}{2 + 10 + 5 + 3} = \frac{2117}{200}$$

$$\text{Phương sai: } s_A^2 = \frac{1}{20} (10,15^2 \cdot 2 + 10,45^2 \cdot 10 + 10,75^2 \cdot 5 + 11,05^2 \cdot 3) - \left(\frac{2117}{200} \right)^2 = \frac{2691}{40000}$$

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

Vận động viên B:

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

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

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 5. Có nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán của hai mẫu số liệu ghép nhóm trong mỗi trường hợp sau không? Tại sao?

- a) Các mẫu số liệu ghép nhóm về điểm thi tốt nghiệp môn Toán của học sinh hai trường trung học phổ thông có chất lượng tương đương.
- b) Các mẫu số liệu ghép nhóm về doanh thu của 100 cửa hàng bán lẻ và doanh thu của 100 siêu thị.

Lời giải

a) Có nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán của học sinh hai trường vì 2 trường này có chất lượng tương đương nhau

b) Không nên dùng phương sai (hoặc độ lệch chuẩn) để so sánh độ phân tán về doanh thu của 100 cửa hàng bán lẻ và 100 siêu thị vì doanh thu của 100 cửa hàng bán lẻ và 100 siêu thị là không gần tương đương nhau.

Câu 6. Mai và Ngọc cùng sử dụng vòng đeo tay thông minh để ghi lại số bước chân hai bạn đi mỗi ngày trong một tháng. Kết quả được ghi lại ở bảng sau:

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

- a) Hãy tính số trung bình và độ lệch chuẩn của mẫu số liệu ghép nhóm trên.
- b) Nếu so sánh theo độ lệch chuẩn thì bạn nào có số lượng bước chân đi mỗi ngày đều đặn hơn?

Lời giải

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

a) Cỡ mẫu: $n = 30$

- Xét mẫu số liệu của Mai:

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: $\sigma_1 = \sqrt{7,56} \approx 2,75$

- Xét mẫu số liệu của Ngọc:

$$\bar{x}_2 = \frac{2.4 + 5.6 + 13.8 + 8.10 + 2.12}{30} = 8,2$$

Số trung bình:

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

Phương sai:

$$\sigma_2 = \sqrt{3,83} \approx 1,96$$

Độ lệch chuẩn:

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

Câu 7. Người ta ghi lại tiền lãi (đơn vị: triệu đồng) của một số nhà đầu tư (với số tiền đầu tư như nhau), khi đầu tư vào hai lĩnh vực A, B cho kết quả như sau:

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

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

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

Lời giải

Ta có mẫu số liệu ghép nhóm với giá trị đại diện là:

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

Tiền lãi trung bình khi đầu tư vào lĩnh vực A là:

$$\bar{x}_A = \frac{7,5.2 + 12,5.5 + 17,5.8 + 22,5.6 + 27,5.4}{2 + 5 + 8 + 6 + 4} = 18,5 \quad (\text{triệu đồng})$$

Tiền lãi trung bình khi đầu tư vào lĩnh vực B là:

$$\bar{x}_B = \frac{7,5.8 + 12,5.4 + 17,5.2 + 22,5.5 + 27,5.6}{8 + 4 + 2 + 5 + 6} = 16,9 \quad (\text{triệu đồng})$$

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:

$$s_A^2 = \frac{1}{25} (7,5^2.2 + 12,5^2.5 + 17,5^2.8 + 22,5^2.6 + 27,5^2.4) - 18,5^2 = 34$$

Độ lệch chuẩn của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực A: $s_A = \sqrt{34}$ (triệu đồng)

Phương sai của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực B:

$$s_B^2 = \frac{1}{25}(7,5^2 \cdot 8 + 12,5^2 \cdot 4 + 17,5^2 \cdot 2 + 22,5^2 \cdot 5 + 27,5^2 \cdot 6) - 16,9^2 = 64,64$$

Độ lệch chuẩn của mẫu số liệu về tiền lãi khi đầu tư vào lĩnh vực B: $s_B = \sqrt{64,64}$ (triệu đồng)

Như vậy, độ lệch chuẩn của mẫu số liệu về tiền lãi đầu tư vào lĩnh vực B lớn hơn độ lệch chuẩn của mẫu số liệu về tiền lãi đầu tư vào lĩnh vực A nên đầu tư vào lĩnh vực B là rủi ro hơn.

Câu 8. Tốc độ của 20 xe hơi khi đi qua một trạm kiểm tra tốc độ (đơn vị: km/h) được thống kê lại như sau:

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

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

Lời giải

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

Cỡ mẫu: $n = 20$

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

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

Tứ phân vị thứ nhất là trung vị của nửa số liệu bên trái Q_1 :

$$Q_1 = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(46,7 + 46,8) = 46,75$$

Tứ phân vị thứ ba là trung vị của nửa số liệu bên phải Q_3 :

$$Q_3 = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(54,8 + 55,6) = 55,2$$

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

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

$$\text{liệu ghép nhóm là: } Q_3' = 54 + \frac{\frac{3 \cdot 20}{4} - (3 + 7 + 4)}{3}(58 - 54) = \frac{166}{3}$$

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

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

$$\text{Số trung bình: } \bar{x} = \frac{3 \cdot 44 + 7 \cdot 48 + 4 \cdot 52 + 3 \cdot 56 + 3 \cdot 60}{20} = 41,8$$

$$\text{Phương sai: } S^2 = \frac{3 \cdot 44^2 + 7 \cdot 48^2 + 4 \cdot 52^2 + 3 \cdot 56^2 + 3 \cdot 60^2}{20} - 41,8^2 = 364,96$$

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

Câu 9. Trong bài thực hành đo hiệu điện thế của mạch điện, An và Bình đã dùng hai vôn kế khác nhau để đo, mỗi bạn tiến hành đo 10 lần cho kết quả như sau:

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

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

Lời giải

Mẫu số liệu với giá trị đại diện

Hiệu điện thế đo được (Vôn)	[3,85; 3,90)	[3,90; 3,95)	[3,95; 4,00)	[4,00; 4,05)
Giá trị đại diện	3,875	3,925	3,975	4,025
Số lần An đo	1	6	2	1
Số lần Bình đo	1	3	4	2

Kết quả đo của An:

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 Bình:

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: $s_2 = \sqrt{0,002025} \approx 0,045$

Vì $s_2 > s_1$ nên Vôn kế của bạn An cho kết quả ổn định hơn.

Câu 10. Điều tra thời gian phải làm thêm trung bình hàng 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.

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:

$$\bar{x}_N = \frac{2 \cdot 6 + 4 \cdot 8 + 9 \cdot 10 + 6 \cdot 12 + 4 \cdot 14}{25} = \frac{262}{25} = 10,48 \quad (\text{giờ})$$

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

$$\begin{aligned} s_N^2 &= \frac{2(6 - 10,48)^2 + 4(8 - 10,48)^2 + 9(10 - 10,48)^2 + 6(12 - 10,48)^2 + 4(14 - 10,48)^2}{25} \\ &= \frac{130,24}{25} = 5,2096 \approx 5,21. \end{aligned}$$

Độ lệch chuẩn s_N của mẫu số liệu ghép nhóm là $s_N = \sqrt{s_N^2} \approx 2,28$.

c) Gọi $\bar{x}_G, s_G^2, s_G$ tương ứng là trung bình, phương sai, độ lệch chuẩn của mẫu số liệu gốc. Để tính giá trị của các số đặc trưng này, ta lập bảng tần số:

Thời gian (giờ)	5	6	7	8	9	10	11	12	13	14	
Tần số	1	1	1	3	5	4	3	3	2	2	$N = 25$

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

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

Phương sai:

$$\begin{aligned} s_G^2 &= \frac{1(5 - 10)^2 + 1(6 - 10)^2 + 1(7 - 10)^2 + \dots + 3(12 - 10)^2 + 2(13 - 10)^2 + 2(14 - 10)^2}{25} \\ &= \frac{132}{25} = 5,28. \end{aligned}$$

Độ lệch chuẩn: $s_G = \sqrt{s_G^2} = \sqrt{5,28} \approx 2,30$.

Quan sát kết quả tìm được ở hai câu b và c, ta thấy $\bar{x}_N \approx \bar{x}_G; s_N^2 \approx s_G^2; s_N \approx s_G$. Như vậy số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm tương ứng xấp xỉ với số trung bình, phương sai, độ lệch chuẩn của mẫu số liệu ban đầu, khi chưa ghép nhóm. Nếu chỉ có mẫu số liệu ghép nhóm thì ta có thể kết luận là thời gian làm thêm trung bình hằng tuần của 25 bác sĩ được điều tra là 10,48 giờ và so với số này thì thời gian làm thêm của các bác sĩ chênh lệch trung bình khoảng 2,28 giờ.

Câu 11. Một giống cây xoan đào được trồng tại hai địa điểm A và B. Người ta thống kê đường kính thân của một số cây xoan đào 5 năm tuổi ở bảng sau:

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

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

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

Lời giải

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

- Xét mẫu số liệu của địa điểm A:

$$\text{Số trung bình: } \bar{x}_A = \frac{25 \cdot 31 + 38 \cdot 33 + 20 \cdot 35 + 10 \cdot 37 + 7 \cdot 39}{100} = 33,72$$

- Xét mẫu số liệu của địa điểm B:

$$\bar{x}_B = \frac{22 \cdot 31 + 27 \cdot 33 + 19 \cdot 35 + 18 \cdot 37 + 14 \cdot 39}{100} = 34,5$$

Đường kính trung bình của thân cây xoan đào trồng tại địa điểm A nhỏ hơn tại địa điểm B

b) - Xét mẫu số liệu của địa điểm A:

$$\sigma_A = \sqrt{\frac{25 \cdot 31^2 + 38 \cdot 33^2 + 20 \cdot 35^2 + 10 \cdot 37^2 + 7 \cdot 39^2}{100} - 33,72^2} = 2,32$$

- Xét mẫu số liệu của địa điểm B:

$$\sigma_B = \sqrt{\frac{22 \cdot 31^2 + 27 \cdot 33^2 + 19 \cdot 35^2 + 18 \cdot 37^2 + 14 \cdot 39^2}{100}} = 2,7$$

Vậy cây trồng tại địa điểm A có đường kính đồng đều hơn

