

BÀI 2**PHƯƠNG SAI, ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHEP NHÓM****1. Định nghĩa**

Nhóm	Giá trị đại diện	Tần số
$[a_1; a_2)$	x_1	n_1
$[a_2; a_3)$	x_2	n_2
...	...	...
$[a_m; a_{m+1})$	x_m	n_m
		$n = n_1 + n_2 + \dots + n_m$

Xét mẫu số liệu ghép nhóm cho bởi bảng trên.

Gọi $\bar{x}$ số trung bình cộng mẫu số liệu trên.

$$s^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_m(x_m - \bar{x})^2}{n}$$

được gọi là *phương sai* của mẫu số liệu trên.

Căn bậc hai (số học) của phương sai gọi là *độ lệch chuẩn* của mẫu số liệu ghép nhóm, kí hiệu là s , nghĩa là $s = \sqrt{s^2}$.

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 và được dùng để đo phân tán của mẫu số liệu ghép nhóm đó.
- Độ lệch chuẩn có cùng đơn vị với đơn vị của mẫu số liệu.
- Khi hai mẫu số liệu ghép nhóm có cùng đơn vị đo và có số trung bình cộng bằng nhau (hoặc xấp xỉ nhau), mẫu số liệu nào có độ lệch chuẩn nhỏ hơn thì mức độ phân tán (so với số trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn.

3. Nhắc lại kiến thức lớp 11: số trung bình cộng của mẫu số liệu ghép nhóm

Cho mẫu số liệu ghép nhóm:

Nhóm	Giá trị đại diện	Tần số
$[a_1; a_2)$	x_1	n_1
$[a_2; a_3)$	x_2	n_2
...	...	...
$[a_m; a_{m+1})$	x_m	n_m
		$n = n_1 + n_2 + \dots + n_m$

- Trung điểm của nửa khoảng (tính bằng trung bình cộng của hai đầu mút) ứng với nhóm là giá trị đại diện của nhóm đó.

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

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_mx_m}{n}$$

Ý nghĩa:

- Như ta đã biết, số trung bình cộng của mẫu số liệu không ghép nhóm là giá trị trung bình cộng của các số trong mẫu số liệu đó, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu trong mẫu ít sai lệch với số trung bình cộng

- Số trung bình cộng của mẫu số liệu sau khi ghép nhóm là giá trị xấp xỉ cho số trung bình của mẫu số liệu không ghép nhóm ban đầu và có thể làm đại diện cho vị trí trung tâm của mẫu số liệu.

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

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

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

Lời giải

Ta có bảng sau:

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

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

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{13.19,25 + 45.19,75 + 24.20,25 + 12.20,75 + 6.21,25}{100} = 20,015$$

Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

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

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

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

Lời giải

a) Ta có bảng sau:

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

• Xét mẫu số liệu của Minh Hiền:

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

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

• Xét mẫu số liệu của Minh Nhân:

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

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

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

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

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

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

Lời giải

a) Ta có bảng sau:

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

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

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

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

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

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

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

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

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

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

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

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

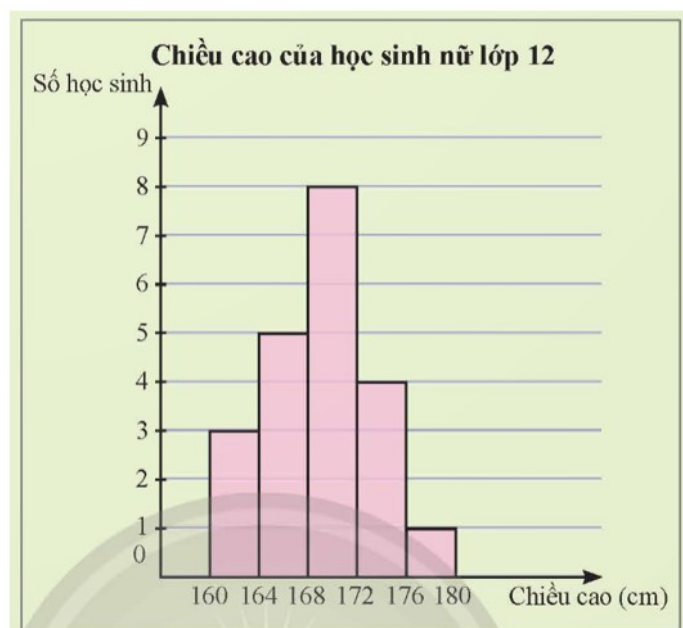
$$S_B^2 = \frac{1}{100} (22.31^2 + 27.33^2 + 19.35^2 + 18.37^2 + 14.39^2) - (34,5)^2 \approx 7,31$$

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

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

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

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



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

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

Lời giải

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

Chiều cao (cm)	[160; 164)	[164; 168)	[168; 172)	[172; 176)	[176; 180)
Số học sinh	3	5	8	4	1
Giá trị đại diện	162	166	170	174	178

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

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

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

Cỡ mẫu $n = 21$.

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

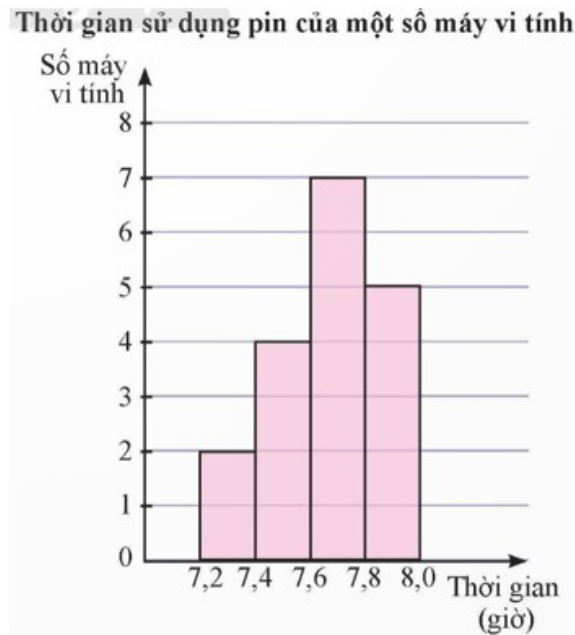
Phương sai của mẫu số liệu ghép nhóm là:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm là:

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

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



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

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

Lời giải

a) Từ biểu đồ ta thấy có 2 máy vi tính có thời gian sử dụng pin từ 7,2 đến dưới 7,4 giờ.

b) Từ biểu đồ, ta có bảng thống kê sau:

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

Cỡ mẫu là $n = 2 + 4 + 7 + 5 = 18$.

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

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

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

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

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

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

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

Lời giải

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

Khoảng biến thiên của mẫu số liệu là:

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

Cỡ mẫu $n = 20$.

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

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

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:

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

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

Ta có

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Số trung bình của mẫu số liệu ghép nhóm là:

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

Phương sai của mẫu số liệu ghép nhóm là:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm là:

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