



CHƯƠNG



CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU

BÀI. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN



I. LÝ THUYẾT

I. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Phương sai của mẫu số liệu ghép nhóm, kí hiệu là s^2 , là một số được tính theo công thức sau:

Câu 1:
$$s^2 = \frac{m_1(x_1 - \bar{x})^2 + \dots + m_k(x_k - \bar{x})^2}{n}$$

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

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

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

Ý nghĩa. Phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm là các xấp xỉ cho 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 đó. Phương sai, độ lệch chuẩn càng lớn thì mẫu số liệu càng phân tán.

Chú ý: Người ta còn sử dụng các đại lượng sau để đo mức độ phân tán của mẫu số liệu ghép nhóm:

Câu 2:
$$s^2 = \frac{m_1(x_1 - \bar{x})^2 + \dots + m_k(x_k - \bar{x})^2}{n-1}, s = \sqrt{s^2}.$$

Ví dụ 1: Người ta theo dõi sự thay đổi cân nặng, 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

Chọn giá trị đại diện cho các nhóm số liệu, ta có:

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

Tổng số người nam là: $n_1 = 2 + 3 + 5 + 3 + 2 = 15$.

Tổng số người nữ là: $n_2 = 2 + 7 + 12 + 7 + 2 = 30$.

Thay đổi cân nặng trung bình của người nam là:

Câu 3: $\bar{x}_1 = \frac{1}{15} [2 \cdot (-0,5) + 3 \cdot 0,5 + 5 \cdot 1,5 + 3 \cdot 2,5 + 2 \cdot 3,5] = 1,5 \text{ (kg)}$

Thay đổi cân nặng trung bình của người nữ là:

Câu 4: $\bar{x}_2 = \frac{1}{30} [2 \cdot (-0,5) + 7 \cdot 0,5 + 12 \cdot 1,5 + 7 \cdot 2,5 + 2 \cdot 3,5] = 1,5 \text{ (kg)}$

Phương sai và độ lệch chuẩn của mẫu số liệu về thay đổi cân nặng của người nam là:

Câu 5: $s_1^2 = \frac{1}{15} [2 \cdot (-0,5)^2 + 3 \cdot 0,5^2 + 5 \cdot 1,5^2 + 3 \cdot 2,5^2 + 2 \cdot 3,5^2] - 1,5^2 \approx 1,21^2; s_1 \approx 1,21$

Phương sai và độ lệch chuẩn của mẫu số liệu về thay đổi cân nặng của người nữ là:

Câu 6: $s_2^2 = \frac{1}{30} [2 \cdot (-0,5)^2 + 7 \cdot 0,5^2 + 12 \cdot 1,5^2 + 7 \cdot 2,5^2 + 2 \cdot 3,5^2] - 1,5^2 \approx 2,06^2; s_2 \approx 2,06$.

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.

II. SỬ DỤNG PHƯƠNG SAI, ĐỘ LỆCH CHUẨN ĐO ĐỘ RỦI RO

Để so sánh độ phân tán của hai mẫu số liệu khi đơn vị đo trên hai mẫu số liệu khác nhau hoặc giá trị trung bình của hai mẫu số liệu này khác nhau rất nhiều người ta dùng hệ số biến thiên CV (Coefficient of Variation). Hệ số biến thiên được tính theo công thức:

Câu 7: $CV = \frac{s}{\bar{x}}$

trong đó s là độ lệch chuẩn và $\bar{x}$ là số trung bình của mẫu số liệu.



Ví dụ 2: Anh An đầu tư số tiền bằng nhau vào hai lĩnh vực kinh doanh A, B . Anh An 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 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:

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 đầu tư vào các lĩnh vực A, B tương ứng là:

Câu 8: $\bar{x}_A = \frac{1}{60}(5 \cdot 7,5 + \dots + 5 \cdot 27,5) = 17,5$ (triệu đồng);

Câu 9: $\bar{x}_B = \frac{1}{60}(20 \cdot 7,5 + \dots + 20 \cdot 27,5) = 17,5$ (triệu đồng).

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 đầu tư vào các lĩnh vực A, B tương ứng là:

Câu 10: $s_A = \sqrt{\frac{1}{60}(5 \cdot 7,5^2 + \dots + 5 \cdot 27,5^2) - (17,5)^2} = 5$

Câu 11: $s_B = \sqrt{\frac{1}{60}(20 \cdot 7,5^2 + \dots + 20 \cdot 27,5^2) - (17,5)^2} \approx 8,42$.

Như vậy, độ lệch chuẩn của mẫu số liệu về số tiền thu được hàng tháng khi 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.

Ví dụ sau cho thấy không phải lúc nào ta cũng có thể dùng độ lệch chuẩn của lợi nhuận thu được để so sánh độ rủi ro của các phương án đầu tư.

Ví dụ 3: Thống kê lợi nhuận hàng tháng (đơn vị: triệu đồng) trong 20 tháng của hai nhà đầu tư được cho như sau:

Lợi nhuận	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
Số tháng	2	4	8	4	2

Bảng 3.2. Lợi nhuận theo tháng của nhà đầu tư nhỏ



Lợi nhuận	[510; 520)	[520; 530)	[530; 540)	[540; 550)	[550; 560)
Số tháng	4	3	6	3	4

Bảng 3.3. Lợi nhuận theo tháng của nhà đầu tư lớn

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

Lời giải

Chọn điểm đại diện cho các nhóm số liệu ta tính được các số đặc trưng như sau:

Lợi nhuận trung bình một tháng của các nhà đầu tư tương ứng là:

$$\bar{x}_A = \frac{1}{20}(2 \cdot 15 + \dots + 2 \cdot 55) = 35 \text{ (triệu đồng)}; \bar{x}_B = \frac{1}{20}(4 \cdot 515 + \dots + 4 \cdot 555) = 535 \text{ (triệu đồng)}.$$

Độ lệch chuẩn của lợi nhuận hàng tháng của hai nhà đầu tư tương ứng là:

Câu 12: $s_A = \sqrt{\frac{1}{20}(2 \cdot 15^2 + \dots + 2 \cdot 55^2) - (35)^2} \approx 10,95$

Câu 13: $s_B = \sqrt{\frac{1}{20}(4 \cdot 515^2 + \dots + 4 \cdot 555^2) - (535)^2} \approx 13,78$

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.



HỆ THỐNG BÀI TẬP TỰ LUẬN.

Câu 1: Cân nặng của một số quả mít trong một khu vườn được thống kê ở bảng sau:

Cân nặng (kg)	[4; 6)	[6; 8)	[8; 10)	[10; 12)	[12; 14)
Số quả mít	6	12	19	9	4

Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm trên. (Kết quả các phép tính làm tròn đến hàng phần trăm.)

Lời giải

Ta có bảng thống kê cân nặng của các quả mít theo giá trị đại diện:

Cân nặng đại diện (kg)	5	7	9	11	13
Tần số	6	12	19	9	4



Cỡ mẫu $n = 6 + 12 + 19 + 9 + 4 = 50$.

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x} = \frac{6.5 + 12.7 + 19.9 + 9.11 + 4.13}{50} = 8,72$

Phương sai của mẫu số liệu ghép nhóm là

$$S^2 = \frac{1}{50} (6.5^2 + 12.7^2 + 19.9^2 + 9.11^2 + 4.13^2) - 8,72^2 \approx 4,80.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S \approx \sqrt{4,80} \approx 2,19$

Câu 2: Thống kê tổng số giờ nắng trong tháng 9 tại một trạm quan trắc đặt ở Cà Mau trong các năm từ 2002 đến 2021 được thống kê như sau:

111,6 134,9 130,3 134,2 140,9 109,3 154,4 156,3 116,1 96,7

105,2 80,8 80,8 110 109 139 145 161 126 114

(Nguồn: Tổng cục Thống kê)

- Hãy tính phương sai và độ lệch chuẩn của mẫu số liệu trên.
- Hãy lập bảng tần số ghép nhóm với nhóm đầu tiên là $[80; 98)$ và độ dài mỗi nhóm bằng 18. Tính phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm.
- Hãy tính sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc. (Kết quả các phép tính làm tròn đến hàng phần nghìn.)

Lời giải

a) Cỡ mẫu là $n = 20$.

Số trung bình của mẫu số liệu trên là: $\bar{x}_1 = \frac{111,6 + 134,9 + \dots + 114}{20} = 122,755$

Phương sai của mẫu số liệu trên là: $S_1^2 = \frac{1}{20} (111,6^2 + 134,9^2 + \dots + 114^2) - 122,755^2 \approx 515,453$.

Độ lệch chuẩn của mẫu số liệu trên là $S_1 \approx \sqrt{515,453} \approx 22,704$

b) Ta có bảng sau:

Số giờ nắng	$[80; 98)$	$[98; 116)$	$[116; 134)$	$[134; 152)$	$[152; 170)$
Giá trị đại diện	89	107	125	143	161
Số năm	3	6	3	5	3

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_2 = \frac{3.89 + 6.107 + 3.125 + 5.143 + 3.161}{20} = 124,1$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{20} (3.89^2 + 6.107^2 + 3.125^2 + 5.143^2 + 3.161^2) - 124,1^2 = 566,19.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 \approx \sqrt{566,19} \approx 23,795$

c) Sai số tương đối của độ lệch chuẩn của mẫu số liệu ghép nhóm so với độ lệch chuẩn của mẫu số liệu gốc là

$$\frac{|S_2 - S_1|}{S_1} = \frac{|23,795 - 22,704|}{22,704} \cdot 100\% \approx 4,805\%$$



Câu 3: 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) Nếu so sánh theo khoảng biến thiên thì học sinh lớp nào có điểm trung bình ít phân tán hơn?

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

Lời giải

a) Khoảng biến thiên của điểm số học sinh lớp 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) Ta có bảng thống kê điểm trung bình theo giá trị đại diện:

Giá trị đại diện \ Lớp	5,5	6,5	7,5	8,5	9,5
11A	1	0	11	22	6
11B	0	6	8	14	12

• Xét mẫu số liệu của lớp 11A: Cỡ mẫu là $n_1 = 1 + 11 + 22 + 6 = 40$.

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_1 = \frac{1.5,5 + 11.7,5 + 22.8,5 + 6.9,5}{40} = 8,3$

Phương sai của mẫu số liệu ghép nhóm: $S_1^2 = \frac{1}{40} (1. 5,5^2 + 11. 7,5^2 + 22. 8,5^2 + 6. 9,5^2) - 8,3^2 = 0,61$.

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

• 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: $\bar{x}_2 = \frac{6.6,5 + 8.7,5 + 14.8,5 + 12.9,5}{40} = 8,3$

Phương sai của mẫu số liệu ghép nhóm: $S_2^2 = \frac{1}{40} (6. 6,5^2 + 8. 7,5^2 + 14. 8,5^2 + 12. 9,5^2) - 8,3^2 = 1,06$.

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

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.

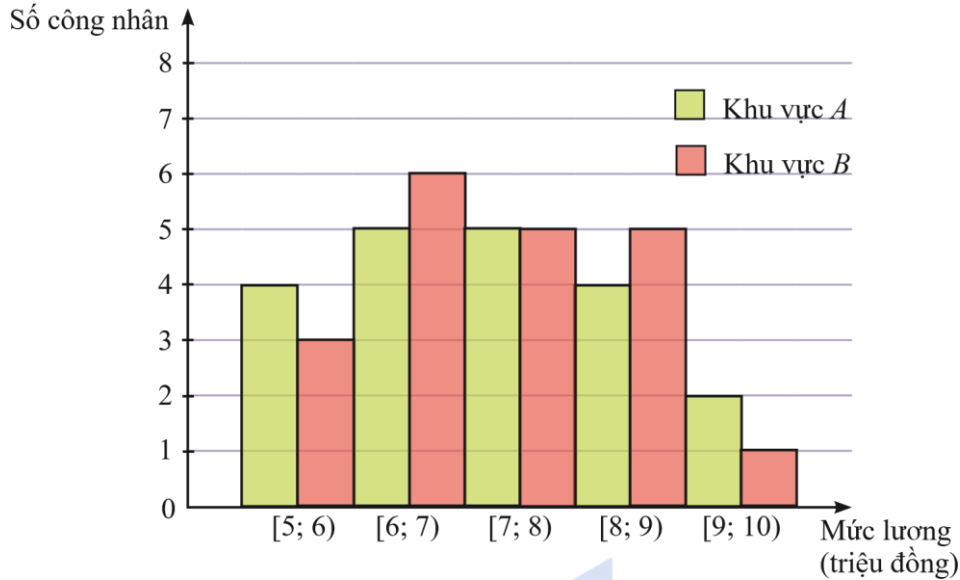
Chú ý: Trong ví dụ trên, kết quả so sánh độ phân tán theo giá trị trung bình và độ lệch chuẩn có sự khác biệt. Điều này là do mẫu số liệu của học sinh lớp 11A có một giá trị ngoại lệ.

Câu 4: Biểu đồ dưới đây mô tả kết quả điều tra về mức lương khởi điểm (đơn vị: triệu đồng) của một số công nhân ở hai khu vực A và B.



- a) Hãy xác định giá trị đại diện cho mỗi nhóm và lập bảng tần số ghép nhóm cho mẫu số liệu đó.
- b) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì công nhân ở khu vực nào có mức lương khởi điểm đồng đều hơn?

Mức lương khởi điểm của công nhân ở hai khu vực A và B



Lời giải

a) Ta có bảng sau:

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

b) Xét mẫu số liệu của khu vực A: Cỡ mẫu là $n_A = 4 + 5 + 5 + 4 + 2 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là: $\bar{x}_A = \frac{4.5,5 + 5.6,5 + 5.7,5 + 4.8,5 + 2.9,5}{20} = 7,25$

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 $S_A \approx \sqrt{1,5875}$

Xét mẫu số liệu của khu vực B: Cỡ mẫu là $n_B = 3 + 6 + 5 + 5 + 1 = 20$.

Số trung bình của mẫu số liệu ghép nhóm là $\bar{x}_B = \frac{3.5,5 + 6.6,5 + 5.7,5 + 5.8,5 + 1.9,5}{20} = 7,25$

Phương sai của mẫu số liệu ghép nhóm là

$$S_B^2 = \frac{1}{20} (3.5,5^2 + 6.6,5^2 + 5.7,5^2 + 5.8,5^2 + 1.9,5^2) - (7,25)^2 = 1,2875.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_B \approx \sqrt{1,2875}$

Do $S_A > S_B$ nên nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì mức lương khởi điểm của công nhân khu vực B đồng đều hơn của công nhân khu vực **A**.



Câu 5: Giá đóng cửa của một cổ phiếu là giá của cổ phiếu đó cuối một phiên giao dịch. Bảng sau thống kê giá đóng cửa (đơn vị: nghìn đồng) của hai mã cổ phiếu A và B trong 50 ngày giao dịch liên tiếp.

Giá đóng cửa	[120; 122)	[122; 124)	[124; 126)	[126; 128)	[128; 130)
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Người ta có thể dùng phương sai và độ lệch chuẩn để so sánh mức độ rủi ro của các loại cổ phiếu có giá trị trung bình gần bằng nhau. Cổ phiếu nào có phương sai, độ lệch chuẩn cao hơn thì được coi là có độ rủi ro lớn hơn. Theo quan điểm trên, hãy so sánh độ rủi ro của cổ phiếu A và cổ phiếu B.

Lời giải

Ta có bảng thống kê giá đóng cửa theo giá trị đại diện:

Giá đóng cửa	121	123	125	127	129
Cổ phiếu A	8	9	12	10	11
Cổ phiếu B	16	4	3	6	21

Xét mẫu số liệu của cổ phiếu A:

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_1 = \frac{8.121 + 9.123 + 12.125 + 10.127 + 11.129}{50} = 125,28$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_1^2 = \frac{1}{50} (8.121^2 + 9.123^2 + 12.125^2 + 10.127^2 + 11.129^2) - (125,28)^2 = 7,5216.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_1 = \sqrt{S_1^2} = \sqrt{7,5216}$

Xét mẫu số liệu của cổ phiếu B:

Số trung bình của mẫu số liệu ghép nhóm là

$$\bar{x}_2 = \frac{16.121 + 4.123 + 3.125 + 6.127 + 21.129}{50} = 125,28$$

Phương sai của mẫu số liệu ghép nhóm là

$$S_2^2 = \frac{1}{50} (16.121^2 + 4.123^2 + 3.125^2 + 6.127^2 + 21.129^2) - (125,28)^2 = 12,4096.$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là $S_2 = \sqrt{S_2^2} = \sqrt{12,4096}$

Vậy nếu đánh giá độ rủi ro theo phương sai và độ lệch chuẩn thì cổ phiếu A có độ rủi ro thấp hơn cổ phiếu **B**.

Câu 6: 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 7: 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 8: 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:

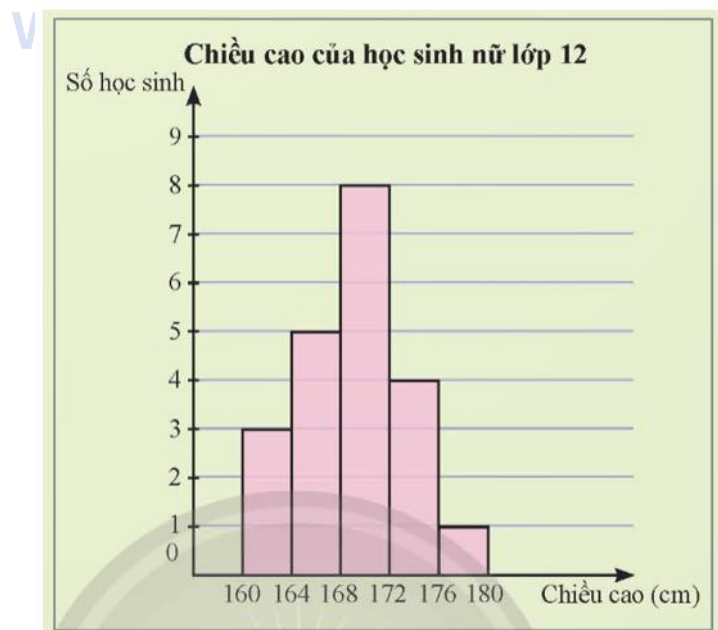
$$S_B^2 = \frac{1}{100} (22.31^2 + 27.33^2 + 19.35^2 + 18.37^2 + 14.39^2) - (34,5)^2 \approx 7,31.$$

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

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

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 9: 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: $\bar{x} = \frac{3550}{21}$

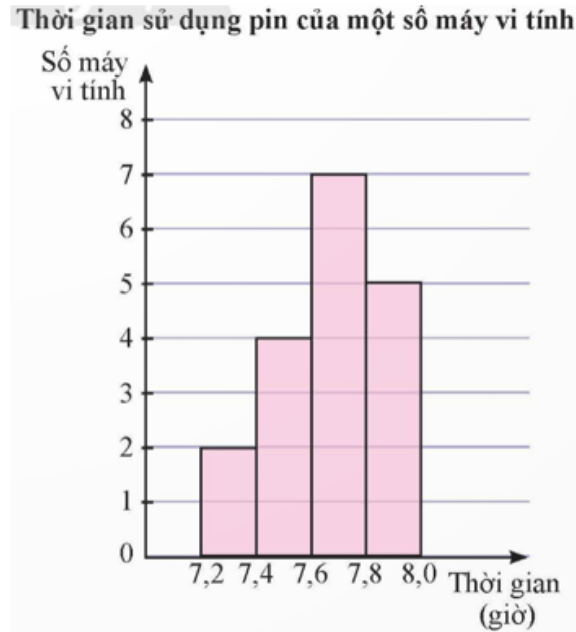
Phương sai của mẫu số liệu ghép nhóm là:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm là:

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

Câu 10: 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 11: 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
----	------	------	------	------	------	------	------	------	------

$$\text{Do đó, } Q_1 = \frac{46,7 + 46,8}{2} = 46,75$$

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
------	------	------	------	------	------	------	------	------	------

$$\text{Do đó, } Q_3 = \frac{54,8 + 55,6}{2} = 55,2$$

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:

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