



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



HỆ THỐNG BÀI TẬP TỰ LUẬN.

Câu 1: Một bác tài xế thống kê lại độ dài quãng đường (đơn vị: km) bác đã lái xe mỗi ngày trong một tháng ở bảng sau:

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

Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên

Lời giải

Khoảng biến thiên của mẫu số liệu là: $300 - 50 = 250$ (km)

Cỡ mẫu $n = 30$

Gọi $x_1; x_2; \dots; x_{30}$ là mẫu số liệu gốc về độ dài quãng đường bác tài xế lái mỗi ngày trong một tháng được xếp theo thứ tự không giảm.

Ta có:

$$x_1; \dots; x_5 \in [50; 100); x_6; \dots; x_{15} \in [100; 150); x_{16}; \dots; x_{24} \in [150; 200); x_{25}; \dots; x_{28} \in [200; 250); \\ x_{29}; x_{30} \in [250; 300)$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_8 \in [100; 150)$. Do đó, tứ phân vị thứ nhất của mẫu

$$\text{số liệu ghép nhóm là: } Q_1 = 100 + \frac{\frac{30}{4} - 5}{10} (150 - 100) = 112,5$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{23} \in [150; 200)$. Do đó, tứ phân vị thứ ba của mẫu số

$$\text{liệu ghép nhóm là: } Q_3 = 150 + \frac{\frac{3.30}{4} - (5+10)}{9} (200 - 150) = \frac{575}{3}$$

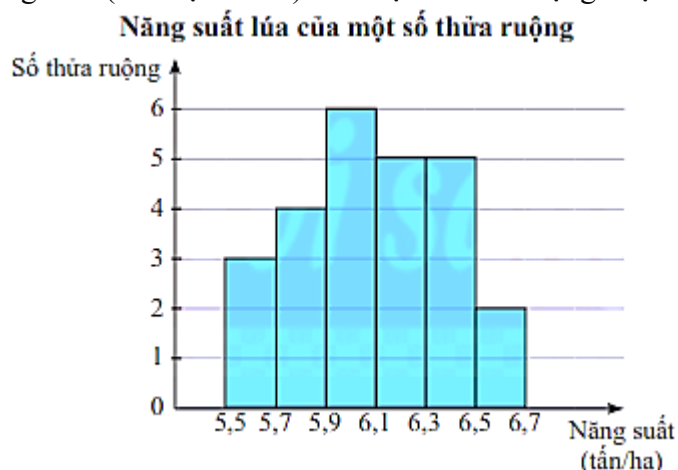
Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 79,17$

Giá trị đại diện	75	125	175	225	275
Số ngày	5	10	9	4	2

Số trung bình: $\bar{x} = \frac{5.75 + 10.125 + 9.175 + 4.225 + 2.275}{30} = 155$

Độ lệch chuẩn: $\sigma = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} = 3100$

Câu 2: Kết quả khảo sát năng suất (đơn vị: tấn/ha) của một số thửa ruộng được minh họa ở biểu đồ sau.



- Có bao nhiêu thửa ruộng đã được khảo sát?
- Lập bảng tần số ghép nhóm và tần số tương đối ghép nhóm tương ứng của mẫu số liệu trên.
- Hãy xác định khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu trên.

Lời giải

a) Có $3 + 4 + 6 + 5 + 5 + 2 = 25$ thửa ruộng đã được khảo sát

b)

Năng suất	[5,5;5,7)	[5,7;5,9)	[5,9;6,1)	[6,1;6,3)	[6,3;6,5)	[6,5;6,7)
Số thửa ruộng	3	4	6	5	5	2

Giá trị đại diện	5,6	5,8	6,0	6,2	6,4	6,6
Tần số tương đối	3	4	6	5	5	2

c) Khoảng biến thiên của mẫu số liệu là: $6,7 - 5,5 = 1,2$ (tấn/ha)

Cỡ mẫu $n = 25$

Gọi $x_1; x_2; \dots; x_{25}$ là mẫu số liệu gốc về năng suất của 25 thửa ruộng được xếp theo thứ tự không giảm.

Ta có: $x_1; \dots; x_3 \in [5, 5; 5, 7); x_4; \dots; x_7 \in [5, 7; 5, 9); x_8; \dots; x_{13} \in [5, 9; 6, 1); x_{14}; \dots; x_{18} \in [6, 1; 6, 3); x_{19}; \dots; x_{23} \in [6, 3; 6, 5); x_{24}; x_{25} \in [6, 5; 6, 7)$

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 $x_{19} \in [6, 3; 6, 5)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm

là: $Q_3 = 6,3 + \frac{\frac{3.25}{4} - (3 + 4 + 6 + 5)}{5}(6,5 - 6,3) = 6,33$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 0,4675$

Số trung bình: $\bar{x} = \frac{3.5,6 + 4.5,8 + 6.6,0 + 5.6,2 + 5.6,4 + 2.6,6}{25} = 6,088$

Độ lệch chuẩn: $\sigma = \sqrt{\frac{5.75^2 + 10.125^2 + 9.175^2 + 4.225^2 + 2.275^2}{30} - 155^2} \approx 0,29$

Câu 3: 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:

- 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 :

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

Xét số liệu của trường Y:

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

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

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

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 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

$$\text{liệu ghép nhóm là: } Q_1' = 7 + \frac{\frac{50}{4} - 4}{12} (8 - 7) = \frac{185}{24}$$

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 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:

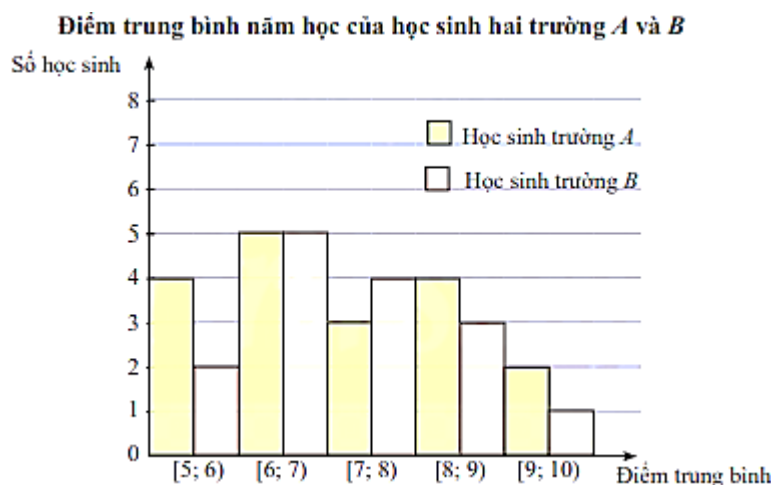
$$\text{Độ lệch chuẩn: } \sigma_Y = \sqrt{\frac{8.6,5^2 + 10.7,5^2 + 13.8,5^2 + 10.9,5^2 + 9.10,5^2}{50} - 8,54^2} \approx 1,33$$

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 4: 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 đồng đều hơn?

c) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường nào có điểm trung bình đồng đều hơn?

Lời giải

a)

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) Cỡ mẫu: $n_A = 18$

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

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

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

$$\text{liệu ghép nhóm là: } Q_1 = 6 + \frac{\frac{18}{4} - 4}{5} (7 - 6) = 6,1$$

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

$$\text{ghép nhóm là: } Q_3 = 8 + \frac{\frac{3 \cdot 18}{4} - (4 + 5 + 3)}{4} (9 - 8) = 8,375$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 2,275$

Cỡ mẫu: $n_B = 15$

Gọi $y_1; y_2; \dots; y_{15}$ là mẫu số liệu gốc về điểm trung bình năm học của học sinh hai trường B được xếp theo thứ tự không giảm.

Ta có: $y_1; y_2 \in [5;6); y_3; \dots; y_7 \in [6;7); y_8; \dots; y_{11} \in [7;8); y_{12}; \dots; y_{14} \in [8;9); y_{15} \in [9;10)$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $y_4 \in [6;7)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1' = 6 + \frac{\frac{15}{4} - 2}{5} (7 - 6) = 6,35$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $y_{12} \in [8;9)$. Do đó, tứ phân vị thứ ba của mẫu số liệu

$$\text{ghép nhóm là: } Q_3' = 8 + \frac{\frac{3 \cdot 15}{4} - (2 + 5 + 4)}{3} (9 - 8) = 8,08$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q' = Q_3' - Q_1' = 1,73$

Vậy nếu so sánh theo khoảng tứ phân vị của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn

c) Xét số liệu của trường A:

$$\text{Số trung bình: } \bar{x}_x = \frac{4.5,5 + 5.6,5 + 3.7,5 + 4.8,5 + 2.9,5}{18} = 7,22$$

$$\text{Độ lệch chuẩn: } \sigma_x = \sqrt{\frac{4.5,5^2 + 5.6,5^2 + 3.7,5^2 + 4.8,5^2 + 2.9,5^2}{18} - 7,22^2} \approx 1,79$$

Xét số liệu của trường B:

$$\text{Số trung bình: } \bar{x}_y = \frac{2.5,5 + 5.6,5 + 4.7,5 + 3.8,5 + 1.9,5}{15} = 7,23$$

$$\text{Độ lệch chuẩn: } \sigma_y = \sqrt{\frac{2.5,5^2 + 5.6,5^2 + 4.7,5^2 + 3.8,5^2 + 1.9,5^2}{15} - 7,23^2} \approx 1,31$$

Vậy nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì học sinh trường B có điểm trung bình đồng đều hơn

Câu 5: Bảng 22, Bảng 23 lần lượt biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội và Huế (Đơn vị: $^{\circ}\text{C}$)

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	2
[19,8; 22,8)	21,3	3
[22,8; 25,8)	24,3	2
[25,8; 28,8)	27,3	1
[28,8; 31,8)	30,3	4
		$n = 12$

Bảng 22

Nhóm	Giá trị đại diện	Tần số
[16,8; 19,8)	18,3	1
[19,8; 22,8)	21,3	2
[22,8; 25,8)	24,3	3
[25,8; 28,8)	27,3	2
[28,8; 31,8)	30,3	4
		$n = 12$

Bảng 23

- Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Hà Nội và Huế.
- Trong hai thành phố Hà Nội và Huế, thành phố nào có nhiệt độ không khí trung bình tháng đồng đều hơn?

Lời giải

a) **Mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội (Bảng 22)**

+ Khoảng biến thiên là $R = 31,8 - 16,8 = 15 (^{\circ}\text{C})$

+ Khoảng tứ phân vị

Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 2, cf_2 = 5, cf_3 = 7, cf_4 = 8, cf_5 = 12$

Ta có $\frac{n}{4} = \frac{12}{4} = 3$ mà $2 < 3 < 5$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3. Xét nhóm 2 là nhóm $[19,8; 22,8)$ có $s = 19,8; h = 3; n_2 = 3$ và nhóm 1 là nhóm $[16,8; 19,8)$ có $cf_1 = 2$. Ta có tứ phân vị thứ nhất là

$$Q_1 = s + \frac{\frac{3n}{4} - cf_1}{n_2} h = 19,8 + \frac{3 - 2}{3} 3 = 20,8(^{\circ}C)$$

Ta có $\frac{3n}{4} = \frac{3.12}{4} = 9$ mà $8 < 9 < 12$. Suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 5 là nhóm $[28,8; 31,8)$ có $t = 28,8; l = 3; n_5 = 4$ và nhóm 4 là nhóm $[25,8; 28,8)$ có $cf_4 = 8$. Ta có tứ phân vị thứ ba là

$$Q_3 = t + \frac{\frac{3n}{4} - cf_4}{n_5} l = 28,8 + \frac{9 - 8}{4} 3 = 29,55(^{\circ}C)$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Hà Nội (Bảng 22) là $D_Q = Q_3 - Q_1 = 29,55 - 20,8 = 8,75(^{\circ}C)$

+ Phương sai và độ lệch chuẩn

$$\text{Ta có } \bar{x}_1 = \frac{2.18,3 + 3.21,3 + 2.24,3 + 27,3 + 4.30,3}{12} = 24,8(^{\circ}C)$$

Phương sai

$$s_1^2 = \frac{2(18,3 - 24,8)^2 + 3(21,3 - 24,8)^2 + 2(24,3 - 24,8)^2 + (27,3 - 24,8)^2 + 4(30,3 - 24,8)^2}{12} = 20,75$$

Độ lệch chuẩn

$$s_1 = \sqrt{s_1^2} = \sqrt{20,75} \approx 4,56(^{\circ}C)$$

Mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Huế (Bảng 23)

+ Khoảng biến thiên là $R = 31,8 - 16,8 = 15(^{\circ}C)$.

+ Khoảng tứ phân vị

Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 1, cf_2 = 3, cf_3 = 6, cf_4 = 8, cf_5 = 12$

Ta có $\frac{n}{4} = \frac{12}{4} = 3$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3. Xét nhóm 2 là nhóm $[19,8; 22,8)$ có $s = 19,8; h = 3; n_2 = 2$ và nhóm 1 là nhóm $[16,8; 19,8)$ có $cf_1 = 1$. Ta có tứ phân vị thứ nhất là

$$Q_1 = s + \frac{\frac{cf_1}{4} - \frac{\ddot{o}}{h}}{n_2} h = 19,8 + \frac{\frac{1}{4} - \frac{1}{3}}{2} 3 = 22,8(^{\circ}C)$$

Ta có $\frac{3n}{4} = \frac{3.12}{4} = 9$ mà $8 < 9 < 12$. Suy ra nhóm 5 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 5 là nhóm $[28,8; 31,8)$ có $t = 28,8; l = 3; n_5 = 4$ và nhóm 4 là nhóm $[25,8; 28,8)$ có $cf_4 = 8$. Ta có tứ phân vị thứ nhất là

$$Q_3 = t + \frac{\frac{cf_4}{4} - \frac{\ddot{o}}{l}}{n_5} l = 28,8 + \frac{\frac{8}{4} - \frac{8}{3}}{4} 3 = 29,55(^{\circ}C)$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng năm 2021 tại Huế (Bảng 23) là $D_Q = Q_3 - Q_1 = 29,55 - 22,8 = 6,75$

+ Phương sai và độ lệch chuẩn

$$\text{Ta có } \bar{x}_2 = \frac{18,3 + 2.21,3 + 3.24,3 + 2.27,3 + 4.30,3}{12} = 25,8(^{\circ}C)$$

Phương sai

$$s_2^2 = \frac{(18,3 - 25,8)^2 + 2(21,3 - 25,8)^2 + 3(24,3 - 25,8)^2 + 2(27,3 - 24,8)^2 + 4(30,3 - 24,8)^2}{12} = 15,75$$

Độ lệch chuẩn

$$s_2 = \sqrt{s_2^2} = \sqrt{15,75} \approx 3,97(^{\circ}C)$$

b) Vì $s_2 < s_1$ nên Huế có nhiệt độ không khí trung bình tháng đồng đều hơn Hà Nội.

Câu 6: Bảng sau thống kê độ ẩm không khí trung bình các tháng 2021 tại Đà Lạt và Vũng Tàu (đơn vị: %)

Tháng Độ ẩm	1	2	3	4	5	6	7	8	9	10	11	12
Đà Lạt	83	79	79	87	87	87	88	89	90	91	88	86
Vũng Tàu	75	77	78	77	79	79	81	79	81	83	80	77

a) Hãy lần lượt ghép lớp các số liệu của Đà Lạt, Vũng Tàu thành năm nhóm sau:

$$[75; 78,3), [78,3; 81,6), [81,6; 84,9), [84,9; 88,2), [88,2; 91,5).$$

b) Tính khoảng biến thiên, khoảng tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm của Đà Lạt và Vũng Tàu.

c) Trong hai thành phố Đà Lạt và Vũng Tàu, thành phố nào có độ ẩm không khí trung bình tháng đồng đều hơn?

Lời giải

a)

Đà Lạt		
Nhóm	Giá trị đại diện	Tần số
[75 ; 78,3)	76,65	0
[78,3; 81,6)	79,95	2
[81,6; 84,9)	83,25	1
[84,9; 88,2)	86,55	6
[88,2; 91,5)	89,85	3
		$n = 12$

Vũng Tàu		
Nhóm	Giá trị đại diện	Tần số
[75 ; 78,3)	76,65	5
[78,3; 81,6)	79,95	6
[81,6; 84,9)	83,25	1
[84,9; 88,2)	86,55	0
[88,2; 91,5)	89,85	0
		$n = 12$

b)

Mẫu số liệu ghép nhóm về độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt

+ Khoảng biến thiên là $R = 91,5 - 75 = 16,5(\%)$.

+ Khoảng tứ phân vị

Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 0, cf_2 = 2, cf_3 = 3, cf_4 = 9, cf_5 = 12$

Ta có $\frac{n}{4} = \frac{12}{4} = 3$. Suy ra nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3. Xét nhóm 3 là nhóm $[81,6; 84,9)$ có $s = 81,6; h = 3,3; n_3 = 1$ và nhóm 2 là nhóm $[78,3; 81,6)$ có $cf_2 = 2$. Ta có tứ phân vị thứ nhất là

$$Q_1 = s + \frac{\frac{n}{4} - cf_2}{n_3} \cdot h = 81,6 + \frac{3 - 2}{1} \cdot 3,3 = 84,9(\%)$$

Ta có $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$. Suy ra nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9.

Xét nhóm 4 là nhóm $[84,9; 88,2)$ có $t = 84,9; l = 3,3; n_4 = 6$ và nhóm 3 là nhóm $[81,6; 84,9)$ có $cf_3 = 3$. Ta có tứ phân vị thứ ba là

$$Q_3 = t + \frac{\frac{3n}{4} - cf_3}{n_4} \cdot l = 84,9 + \frac{9 - 3}{6} \cdot 3,3 = 88,2(\%)$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ ẩm không khí trung bình các tháng năm 2021 tại Đà Lạt là $D_Q = Q_3 - Q_1 = 88,2 - 84,9 = 3,3(\%)$

+ Phương sai và độ lệch chuẩn

$$\text{Ta có } \bar{x}_1 = \frac{0,76,65 + 2,79,95 + 83,25 + 6,86,55 + 3,89,85}{12} = 86(\%)$$

Phương sai

$$s_1^2 = \frac{0(76,65 - 86)^2 + 2(79,95 - 86)^2 + (83,25 - 86)^2 + 6(86,55 - 86)^2 + 3(89,95 - 86)^2}{12} = 10,78$$

Độ lệch chuẩn

$$s_1 = \sqrt{s_1^2} = \sqrt{10,78} \approx 3,28(\%)$$

Mẫu số liệu ghép nhóm về độ ẩm không khí trung bình các tháng năm 2021 tại Vũng Tàu

+ Khoảng biến thiên là $R = 91,5 - 75 = 16,5(\%)$.

+ Khoảng tứ phân vị

Số phần tử của mẫu là $n = 12$

Tần số tích lũy của các nhóm lần lượt là $cf_1 = 5, cf_2 = 11, cf_3 = 12, cf_4 = 12, cf_5 = 12$

Ta có $\frac{n}{4} = \frac{12}{4} = 3$. Suy ra nhóm 1 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 3. Xét nhóm 1 là nhóm $[75; 78,3)$ có $s = 75; h = 3,3; n_1 = 5$. Ta có tứ phân vị thứ nhất là

$$Q_1 = s + \frac{\frac{n}{4} - cf_0}{n_1} \cdot h = 75 + \frac{3 - 0}{5} \cdot 3,3 = 76,98(\%)$$

Ta có $\frac{3n}{4} = \frac{3 \cdot 12}{4} = 9$ mà $5 < 9 < 11$. Suy ra nhóm 2 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 9. Xét nhóm 2 là nhóm $[78,3); 81,6)$ có $t = 78,3; l = 3,3; n_2 = 6$ và nhóm 1 là nhóm $[75; 78,3)$ có $cf_1 = 5$. Ta có tứ phân vị thứ ba là

$$Q_3 = t + \frac{\frac{3n}{4} - cf_4}{n_5} \cdot l = 75 + \frac{9 - 5}{6} \cdot 3,3 = 77,2(\%)$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ ẩm không khí trung bình các tháng năm 2021 tại Vũng Tàu là $D_Q = Q_3 - Q_1 = 77,2 - 76,98 = 0,22(\%)$

+ Phương sai và độ lệch chuẩn

$$\text{Ta có } \bar{x}_2 = \frac{5.76,65 + 6.79,95 + 83,25 + 0.86,55 + 0.89,85}{12} = 78,85(\%)$$

Phương sai

$$s_2^2 = \frac{5(76,65 - 78,85)^2 + 6(79,95 - 78,85)^2 + (83,25 - 78,85)^2}{12} = 4,24$$

Độ lệch chuẩn

$$s_2 = \sqrt{s_2^2} = \sqrt{4,24} \approx 2,06(\%)$$

b) Vì $s_2 < s_1$ nên Vũng Tàu có độ ẩm không khí trung bình tháng đồng đều hơn Đà Lạt.



HỆ THỐNG BÀI TẬP TRẮC NGHIỆM.

Câu 7: Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm này là

- A.** 4. **B.** 5. **C.** 6. **D.** 7.

Lời giải

Khoảng biến thiên $R = 19 - 14 = 5$

Câu 8: Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

- A.** 1,5. **B.** 0,9. **C.** 0,6. **D.** 0,3.

Lời giải

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 4,2 - 2,7 = 1,5$ (km).

Câu 9: Cô Minh Hiền rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm là

- A.** 25. **B.** 20. **C.** 15. **D.** 30.

Lời giải

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 45 - 20 = 25$ (phút).

Câu 10: Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

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

Khoảng biến thiên của mẫu số liệu ghép nhóm nhận giá trị nào trong các giá trị dưới đây?

- A.** 6. **B.** 8. **C.** 10. **D.** 12.

Lời giải

Khoảng biến thiên của mẫu số liệu ghép nhóm là: $R = 18 - 8 = 10$ (giây).

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

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

Nhóm chứa tứ phân vị thứ nhất là

- A.** [14; 15). **B.** [15; 16). **C.** [16; 17). **D.** [17; 18).

Lời giải

Cỡ mẫu là: $1 + 3 + 8 + 6 + 2 = 20$.

Gọi $x_1; x_2; \dots; x_{20}$ là tuổi thọ của 20 con hổ được sắp xếp theo thứ tự tăng dần.

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [16; 17)$ nên nhóm chứa tứ phân vị thứ nhất là [16; 17).

Câu 12: Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

Nhóm chứa tứ phân vị thứ ba là

- A.** [18; 19). **B.** [15; 16). **C.** [16; 17). **D.** [17; 18).

Lời giải

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [17; 18)$. Do đó nhóm chứa tứ phân vị thứ ba là [17; 18).

Câu 13: Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

Số đặc trưng nào không sử dụng thông tin của nhóm số liệu đầu tiên và nhóm số liệu cuối cùng.

- A.** Khoảng biến thiên. **B.** khoảng tứ phân vị.
C. Phương sai. **D.** Độ lệch chuẩn.

Lời giải

Số đặc trưng không sử dụng thông tin của nhóm số liệu đầu tiên và nhóm số liệu cuối cùng là khoảng tứ phân vị.

Câu 14: Một vườn thú ghi lại tuổi thọ (đơn vị: năm) của 20 con hổ và thu được kết quả như sau:

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

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

- A.** Khoảng biến thiên. **B.** khoảng tứ phân vị.
C. Phương sai. **D.** Độ lệch chuẩn.

Lời giải

Khoảng biến thiên sẽ không thay đổi nếu thay tất cả các tần số trong mẫu số liệu ghép nhóm trên bằng 4.

Câu 15: Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A.** 0,9. **B.** 0,975. **C.** 0,5. **D.** 0,575.

Lời giải

Cỡ mẫu $n = 20$.

Gọi $x_1; \dots; x_{20}$ là mẫu số liệu gốc về quãng đường đi bộ mỗi ngày của bác Hương trong 20 ngày được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_3 \in [2,7; 3,0),$$

$$x_4; \dots; x_9 \in [3,0; 3,3),$$

$$x_{10}; \dots; x_{14} \in [3,3; 3,6),$$

$$x_{15}; \dots; x_{18} \in [3,6; 3,9),$$

$$x_{19}; x_{20} \in [3,9; 4,2).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [3,0; 3,3)$. Do đó, tứ phân vị thứ nhất của

$$\text{mẫu số liệu ghép nhóm là: } Q_1 = 3,0 + \frac{\frac{20}{4} - 3}{6}(3,3 - 3,0) = 3,1$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [3,6; 3,9)$. Do đó, tứ phân vị thứ ba của mẫu

$$\text{số liệu ghép nhóm là: } Q_3 = 3,6 + \frac{\frac{3 \cdot 20}{4} - (3 + 6 + 5)}{4}(3,9 - 3,6) = 3,675$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 3,675 - 3,1 = 0,575$

Câu 16: Cô Minh Hiền rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

A. 23,75.

B. 27,5.

C. 31,88.

D. 8,125.

Lời giải

Cỡ mẫu $n = 6 + 6 + 4 + 1 + 1 = 18$.

Gọi $x_1; \dots; x_{18}$ là mẫu số liệu gốc về thời gian tập nhảy mỗi ngày của Cô Minh Hiền được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_6 \in [20; 25),$$

$$x_7; \dots; x_{12} \in [25; 30),$$

$$x_{13}; \dots; x_{16} \in [30; 35),$$

$$x_{17} \in [35; 40),$$

$$x_{18} \in [40; 45).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_5 \in [20; 25)$. Do đó, tứ phân vị thứ nhất của mẫu số

$$\text{liệu ghép nhóm là: } Q_1 = 20 + \frac{\frac{18}{4} - 3}{6}(25 - 20) = 23,75.$$

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

$$\text{ghép nhóm là: } Q_3 = 30 + \frac{\frac{3 \cdot 18}{4} - (6 + 6)}{4}(35 - 30) = 31,875$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 31,875 - 23,75 = 8,125$

Câu 17: Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là

- A. 10,75. B. 1,75. C. 3,63. D. 14,38.

Lời giải

Cỡ mẫu $n = 25$.

Gọi $x_1; \dots; x_{25}$ là mẫu số liệu gốc về thời gian giải rubik trong 25 lần giải liên tiếp được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_4 \in [8; 10),$$

$$x_5; \dots; x_{10} \in [10; 12),$$

$$x_{11}; \dots; x_{18} \in [12; 14),$$

$$x_{19}; \dots; x_{22} \in [14; 16),$$

$$x_{23}; \dots; x_{25} \in [16; 18).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_6 + x_7}{2} \in [10; 12)$. Do đó, tứ phân vị thứ nhất của

$$\text{mẫu số liệu ghép nhóm là: } Q_1 = 10 + \frac{\frac{25}{4} - 4}{6} (12 - 10) = 10,75$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{19} + x_{20}}{2} \in [14; 16)$. Do đó, tứ phân vị thứ ba của mẫu

$$\text{số liệu ghép nhóm là: } Q_3 = 10 + \frac{\frac{3 \cdot 25}{4} - (4 + 6 + 8)}{4} (16 - 14) = 14,375$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 = 14,375 - 10,75 = 3,625 \approx 3,63$

Câu 18: Mỗi ngày bác Hương đều đi bộ để rèn luyện sức khỏe. Quãng đường đi bộ mỗi ngày (đơn vị: km) của bác Hương trong 20 ngày được thống kê lại ở bảng sau:

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

Phương sai của mẫu số liệu ghép nhóm là

- A. 3,39. B. 11,62. C. 0,1314. D. 0,36.

Lời giải

Ta có bảng sau:

Quãng đường (km)	[2,7; 3,0)	[3,0; 3,3)	[3,3; 3,6)	[3,6; 3,9)	[3,9; 4,2)
Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
Số ngày	3	6	5	4	2

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{3.2,85 + 6.3,15 + 5.3,45 + 4.3,75 + 2.4,05}{20} = 3,39$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{20} \left[3.(2,85)^2 + 6.(3,15)^2 + 5.(3,45)^2 + 4.(3,75)^2 + 2.(4,05)^2 \right] - (3,39)^2 = 0,1314$$

Câu 19: Cô Minh Hiền rất thích nhảy hiện đại. Thời gian tập nhảy mỗi ngày trong thời gian gần đây của Cô Minh Hiền được thống kê lại ở bảng sau:

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

Phương sai của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 31,77.

B. 32.

C. 31.

D. 31,25.

Lời giải

Ta có bảng sau:

Thời gian (phút)	[20; 25)	[25; 30)	[30; 35)	[35; 40)	[40; 45)
Giá trị đại diện	22,5	27,5	32,5	37,5	42,5
Số ngày	6	6	4	1	1

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{6.22,5 + 6.27,5 + 4.32,5 + 1.37,5 + 1.42,5}{18} = \frac{85}{3}.$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{18} \left[6.(22,5)^2 + 6.(27,5)^2 + 4.(32,5)^2 + 1.(37,5)^2 + 1.(42,5)^2 \right] - \left(\frac{85}{3} \right)^2 = 31,25$$

Do đó, phương sai của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị 31,25.

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

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 3,41.

B. 11,62.

C. 0,017.

D. 0,36.

Lời giải

Ta có bảng sau:

Quãng đường (km)	[2,7; 3,0)	[3,0; 3,3)	[3,3; 3,6)	[3,6; 3,9)	[3,9; 4,2)
Giá trị đại diện	2,85	3,15	3,45	3,75	4,05
Số ngày	3	6	5	4	2

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{3.2,85 + 6.3,15 + 5.3,45 + 4.3,75 + 2.4,05}{20} = 3,39$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{20} [3.(2,85)^2 + 6.(3,15)^2 + 5.(3,45)^2 + 4.(3,75)^2 + 2.(4,05)^2] - (3,39)^2 = 0,1314$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} = \sqrt{0,1314} \approx 0,36$

Câu 21: Dũng là học sinh rất giỏi chơi rubik, bạn có thể giải nhiều loại khối rubik khác nhau. Trong một lần tập luyện giải khối rubik 3×3 , bạn Dũng đã tự thống kê lại thời gian giải rubik trong 25 lần giải liên tiếp ở bảng sau:

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

Độ lệch chuẩn của mẫu số liệu ghép nhóm có giá trị gần nhất với giá trị nào dưới đây?

A. 5,98.

B. 6.

C. 2,44.

D. 2,5.

Lời giải

Ta có bảng sau:

Thời gian giải rubik (giây)	[8; 10)	[10; 12)	[12; 14)	[14; 16)	[16; 18)
Giá trị đại diện	9	11	13	15	17
Số lần	4	6	8	4	3

Số trung bình của mẫu số liệu ghép nhóm là:

$$\bar{x} = \frac{4.9 + 6.11 + 8.13 + 4.15 + 3.17}{25} = 12,68$$

Phương sai của mẫu số liệu ghép nhóm là:

$$S^2 = \frac{1}{25} [4.9^2 + 6.11^2 + 8.13^2 + 4.15^2 + 3.17^2] - (12,68)^2 = 5,9776$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S = \sqrt{S^2} = \sqrt{5,9776} \approx 2,44$