

Chủ đề 3. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM

Bài 6. MẪU SỐ LIỆU GHÉP NHÓM

PHẦN A. KIẾN THỨC CẦN NẮM

1. KHOẢNG BIẾN THIÊN CỦA MẪU SỐ LIỆU GHÉP NHÓM

a) Định nghĩa

Với mẫu số liệu ghép nhóm cho bởi Bảng 1.

Gọi u_1 và u_{k+1} lần lượt là đầu mút trái của nhóm đầu tiên và đầu mút phải của nhóm cuối cùng.

Hiệu $R = u_{k+1} - u_1$ được gọi là khoảng biến thiên của mẫu số liệu ghép nhóm đó.

Bảng 1

Nhóm	$[u_1; u_2)$	$[u_2; u_3)$	$[u_3; u_4)$	$\dots$	$[u_k; u_{k+1})$
Tần số	n_1	n_2	n_3	$\dots$	n_k

b) Ý nghĩa của khoảng biến thiên

- Khoảng biến thiên của mẫu số liệu ghép nhóm đo mức độ phân tán của mẫu số liệu đó.
- Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.
- Khoảng biến thiên chỉ sử dụng hai giá trị u_1 và u_{k+1} của mẫu số liệu nên đại lượng đó dễ bị ảnh hưởng bởi các giá trị bất thường.
- Khoảng biến thiên của mẫu số liệu ghép nhóm xấp xỉ cho khoảng biến thiên của mẫu số liệu gốc.

Lưu ý: Đối với mẫu số liệu ghép nhóm mà ta biết mẫu số liệu không ghép nhóm sinh ra nó thì ta cũng có thể chọn khoảng biến thiên của mẫu số liệu không ghép nhóm chính là khoảng biến thiên của mẫu số liệu ghép nhóm.

2. KHOẢNG TỨ PHÂN VỊ CỦA MẪU SỐ LIỆU GHÉP NHÓM

a) Định nghĩa

Với mẫu số liệu ghép nhóm cho bởi Bảng 1 .

Gọi Q_1, Q_2, Q_3 là tứ phân vị của mẫu số liệu đó. Ta gọi hiệu $\Delta_Q = Q_3 - Q_1$ là khoảng tứ phân vị của mẫu số liệu đó.

Lưu ý: Tứ phân vị thứ i , kí hiệu là Q_i , với $i=1,2,3$ của mẫu số liệu ghép nhóm (Bảng 1) được xác định như sau:

$$Q_i = u_m + \frac{\frac{i \cdot n}{4} - C}{n_m} (u_{m+1} - u_m)$$

Trong đó:

- $n = n_1 + n_2 + \dots + n_k$ là cỡ mẫu;

- $[u_m; u_{m+1})$ là nhóm chứa tứ phân vị thứ i ;
- n_m là tần số của nhóm chứa tứ phân vị thứ i ;
- $C = n_1 + n_2 + \dots + n_{m-1}$.

b) Ý nghĩa khoảng tứ phân vị của mẫu số liệu ghép nhóm

- Khoảng tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ khoảng tứ phân vị của mẫu số liệu gốc và là một đại lượng cho biết mức độ phân tán của nửa giữa mẫu số liệu.

- Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

- Khoảng tứ phân vị của mẫu số liệu ghép nhóm giúp xác định các giá trị bất thường của mẫu đó. Khoảng tứ phân vị thường được sử dụng thay cho khoảng biến thiên vì nó loại trừ hầu hết giá trị bất thường của mẫu số liệu và nó không bị ảnh hưởng bởi các giá trị bất thường đó.

3. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MẪU SỐ LIỆU GHEP NHÓM

a) Định nghĩa

Cho mẫu số liệu ghép nhóm như ở Bảng 2.

Bảng 2

Nhóm	Giá trị đại diện	Tần số
$[a_1; a_2)$	c_1	n_1
$[a_2; a_3)$	c_2	n_2
...	...	...
$[a_k; a_{k+1})$	c_m	n_m
		$n = n_1 + n_2 + \dots + n_m$

- Gọi $\bar{x}$ là số trung bình cộng của mẫu số liệu đó. Ta có:
$$\bar{x} = \frac{1}{n}(n_1 c_1 + n_2 c_2 + \dots + n_k c_k)$$
.
- 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.

- Căn bậc hai số học của phương sai, kí hiệu là S , được gọi là độ lệch chuẩn của mẫu số liệu đó, nghĩa là $S = \sqrt{S^2}$.

Lời bình: Chúng ta có thể tính phương sai thông qua số trung bình bình phương.

$$\overline{x^2} = \frac{1}{n}(n_1c_1^2 + n_2c_2^2 + \dots + n_kc_k^2)$$

$$S^2 = \overline{x^2} - (\overline{x})^2$$

b) Ý nghĩa của phương sai và độ lệch chuẩn

- Phương sai (độ lệch chuẩn) của mẫu số liệu ghép nhóm được dùng để đo mức độ 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.

4. MỘT CỦA SỐ LIỆU GHEP NHÓM

Nhóm chứa một của mẫu số liệu ghép nhóm là nhóm có tần số lớn nhất.

Giả sử nhóm chứa một là $[u_m; u_{m+1})$, khi đó một của mẫu số liệu ghép nhóm, kí hiệu là M_o , được xác

$$M_o = u_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \cdot (u_{m+1} - u_m)$$

định bởi công thức

*Chú ý: Nếu không có nhóm kề trước của nhóm chứa một thì $n_{m-1} = 0$. Nếu không có nhóm kề sau của nhóm chứa một thì $n_{m+1} = 0$.

* Ý nghĩa của một của mẫu số liệu ghép nhóm

- Một của mẫu số liệu không ghép nhóm là giá trị có khả năng xuất hiện cao nhất khi lấy mẫu.
Một của mẫu số liệu sau khi ghép nhóm M_o xấp xỉ với một của mẫu số liệu không ghép nhóm.
Các giá trị nằm xung quanh M_o thường có khả năng xuất hiện cao hơn các giá trị khác.
- Một mẫu số liệu ghép nhóm có thể có nhiều nhóm chứa một và nhiều một.

PHẦN B. CÂU TẬP ÁP DỤNG

Câu 1. Kết quả thăm dò ngẫu nhiên thời gian chờ xe buýt mỗi ngày của một số sinh viên được biểu thị ở bảng sau:

Thời gian (phút)	$[5;10)$	$[10;15)$	$[15;20)$	$[20;25)$	$[25;30)$
Số người	4	5	6	3	2

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 30, đầu mút trái của nhóm ghép đầu tiên là 5.

Vậy khoảng biến thiên của mẫu số liệu là $R = 30 - 5 = 25$.

Kết quả này cho biết thời gian chờ xe buýt mỗi ngày của các sinh viên thuộc nhóm người được điều tra chênh lệch nhau nhiều nhất là 25 phút.

Câu 2. Trong một phòng thí nghiệm, nghiên cứu thời gian sạc pin của hai loại điện thoại X và Y . Bảng dưới đây biểu thị thời gian sạc pin (phút) của hai loại điện thoại:

Thời gian sạc (phút)	$[30; 60)$	$[60; 90)$	$[90; 120)$	$[120; 150)$	$[150; 180)$
Số điện thoại loại X	0	8	25	32	35
Số điện thoại loại Y	20	12	15	23	30

Tính khoảng biến thiên và sánh thời gian sạc pin của hai loại điện thoại trên?

Lời giải

Khoảng biến thiên của điện thoại loại X là: $R_x = 180 - 60 = 120$ phút.

Khoảng biến thiên của điện thoại loại Y là: $R_y = 180 - 30 = 150$ phút.

Vì $R_x < R_y$ nên có thể nói là thời gian sạc pin điện thoại Y có sự phân tán hơn so điện thoại loại X .

Điều này có thể cho thấy rằng thời gian sạc pin của điện thoại loại Y không đồng đều và ổn định như điện thoại loại X .

Câu 3. Mẫu số liệu sau cho biết số lượng sách bán được tại một hiệu sách trong 10 ngày:

15	22	20	30	28	25	40	37	35	32
----	----	----	----	----	----	----	----	----	----

Tìm khoảng tứ phân vị cho mẫu số liệu trên.

Lời giải

Sắp xếp mẫu số liệu theo thứ tự không giảm: 15 20 22 25 28 30 32 35 37 40

Mẫu số liệu gồm 10 giá trị nên trung vị là giá trị trung bình của hai giá trị ở vị trí chính giữa, tức là giá

$$Q_2 = \frac{28+30}{2} = 29$$

trị thứ 5 và thứ 6 trong dãy đã sắp xếp:

Nửa số liệu bên trái là: 15, 20, 22, 25, 28 gồm 5 giá trị. Do đó tứ phân vị thứ nhất giá trị thứ ba nên ta có $Q_1 = 22$.

Nửa số liệu bên phải là: 30, 32, 35, 37, 40 gồm 5 giá trị. Do đó tứ phân vị thứ ba giá trị thứ ba nên ta có $Q_3 = 35$.

Vậy khoảng tứ phân vị của mẫu số liệu trên là: $\Delta_Q = Q_3 - Q_1 = 35 - 22 = 13$.

Câu 4. Một nghiên cứu theo dõi thời gian làm Câu tập về nhà của học sinh lớp 12C4 tại trường Trung học Phổ thông X. Kết quả được mô tả ở bảng sau:

Thời gian (phút)	$[30; 40)$	$[40; 50)$	$[50; 60)$	$[60; 70)$
------------------	------------	------------	------------	------------

Số học sinh	5	10	8	3
-------------	---	----	---	---

a) Tính khoảng biến thiên R cho mẫu số liệu ghép nhóm.

b) Nếu biết học sinh làm Câu tập ít nhất là 32 phút và nhiều nhất là 68 phút, thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

Lời giải

a) Khoảng biến thiên R của mẫu số liệu ghép nhóm là: $R = 70 - 30 = 40$ phút.

b) Nếu biết học sinh làm Câu tập ít nhất là 32 phút và nhiều nhất là 68 phút thì khoảng biến thiên của mẫu số liệu gốc là sự chênh lệch giữa thời gian hoàn thành sớm nhất và muộn nhất. Do đó, khoảng biến thiên của mẫu số liệu gốc là: $R = 68 - 32 = 36$ phút.

Câu 5. Cân nặng của học sinh lớp 10 A và 10 B ở một trường trung học phổ thông được thống kê ở bảng sau:

Cân nặng (kg)	[40; 45)	[45; 50)	[50; 55)	[55; 60)	[60; 65)	[65; 70)
Số học sinh lớp 10A	3	5	10	7	4	1
Số học sinh lớp 10B	2	8	7	6	3	2

So sánh theo khoảng biến thiên, cân nặng của học sinh lớp nào có độ phân tán lớn hơn?

Lời giải

Khoảng biến thiên của cân nặng của học sinh lớp 10 A là: $R_{10A} = 70 - 40 = 30$ kg.

Khoảng biến thiên của cân nặng của học sinh lớp 10 B là: $R_{10B} = 70 - 40 = 30$ kg.

Khoảng biến thiên của cả hai lớp 10 A và 10 B đều là 30 kg. Do đó, theo khoảng biến thiên, cân nặng của học sinh hai lớp này có độ phân tán như nhau.

Câu 6. Nhà đầu tư theo dõi doanh thu bán một loại sản phẩm của hai cửa hàng bán lẻ trong 5 năm. Số liệu được thống kê ở bảng sau:

Số tiền doanh thu (triệu đồng)	[10; 15)	[15; 20)	[20; 25)	[25; 30)	[30; 35)
Số tháng của cửa hàng A	5	15	25	10	5
Số tháng của cửa hàng B	10	10	20	10	10

So sánh giá trị trung bình và độ lệch chuẩn của doanh thu hàng tháng của hai cửa hàng A và B. Cửa hàng nào có doanh thu biến động lớn hơn? Đầu tư vào cửa hàng nào "rủi ro" hơn?

Lời giải

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

Giá trị đại diện	12,5	17,5	22,5	27,5	32,5
Số tháng của cửa hàng A	5	15	25	10	5
Số tháng của cửa hàng B	10	10	20	10	10

Doanh thu trung bình thu được của cửa hàng A và B tương ứng là:

$$x_A = \frac{1}{60} (12,5 \times 5 + 17,5 \times 15 + 22,5 \times 25 + 27,5 \times 10 + 32,5 \times 5) = \frac{265}{12} \text{ triệu đồng.}$$

$$x_B = \frac{1}{60} (12,5 \times 10 + 17,5 \times 10 + 22,5 \times 20 + 27,5 \times 10 + 32,5 \times 10) = 22,5 \text{ triệu đồng.}$$

Như vậy, về trung bình doanh thu thì cửa hàng B cao hơn cửa hàng A .

Độ lệch chuẩn của số tiền doanh thu hàng tháng tương ứng là:

Cửa hàng A

$$S_A^2 = \frac{1}{60} \left[\left(12,5 - \frac{265}{12} \right)^2 \cdot 5 + \left(17,5 - \frac{265}{12} \right)^2 \cdot 15 + \left(22,5 - \frac{265}{12} \right)^2 \cdot 25 + \left(27,5 - \frac{265}{12} \right)^2 \cdot 10 + \left(32,5 - \frac{265}{12} \right)^2 \cdot 5 \right] = \frac{3875}{144}$$

$$\Rightarrow s_A = \sqrt{S_A^2} = 5,19$$

Cửa hàng B

$$S_B^2 = \frac{(12,5 - 22,5)^2 \cdot 10 + (17,5 - 22,5)^2 \cdot 10 + (22,5 - 22,5)^2 \cdot 20 + (27,5 - 22,5)^2 \cdot 10 + (32,5 - 22,5)^2 \cdot 10}{60} = \frac{125}{3}$$

$$\Rightarrow S_B = \sqrt{S_B^2} = 6,45.$$

Như vậy, độ lệch chuẩn của doanh thu hàng tháng của cửa hàng B cao hơn cửa hàng A , điều này cho thấy doanh thu hàng tháng của cửa hàng B biến động lớn hơn. Do đó, đầu vào cửa hàng B được xem là "rủi ro" hơn so với cửa hàng A .

Cách 2: Tương tự như cách 1 nhưng cách tính phương sai khác.

Tính phương sai theo số trung bình bình phương

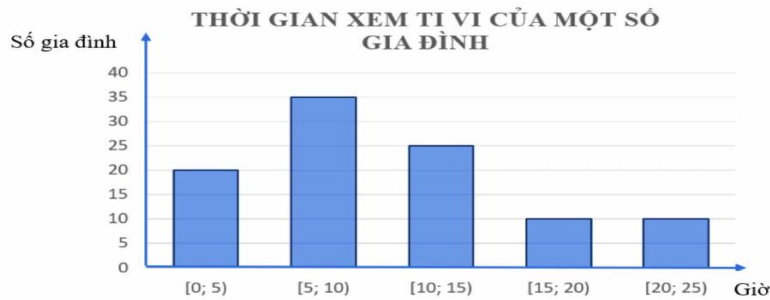
$$\overline{x_A^2} = \frac{1}{60} (12,5^2 \cdot 5 + 17,5^2 \cdot 15 + 22,5^2 \cdot 25 + 27,5^2 \cdot 10 + 32,5^2 \cdot 5) = \frac{6175}{12}$$

$$S_A^2 = \overline{x_A^2} - (\bar{x})^2 = \frac{6175}{12} - \left(\frac{265}{12}\right)^2 = \frac{3875}{144} \Rightarrow s_A = \sqrt{S_A^2} = 5,19$$

$$\overline{x_B^2} = \frac{1}{60} (12.5^2 \times 10 + 17.5^2 \times 10 + 22.5^2 \times 20 + 27.5^2 \times 10 + 32.5^2 \times 10) = \frac{6575}{12}$$

$$S_B^2 = \overline{x_B^2} - (\bar{x}_B)^2 = \frac{125}{3} \Rightarrow S_B = \sqrt{S_B^2} = 6,45$$

Câu 7. Kết quả điều tra về số giờ xem ti vi của một số hộ gia đình được biểu diễn ở biểu đồ dưới y:



Hãy lập bảng tần số ghép nhóm, tìm khoảng biến thiên, khoảng tứ phân vị và giá trị trung bình của mẫu số liệu.

Lời giải

a) Bảng tần số ghép nhóm:

Nhóm giờ xem Ti vi (Giờ)	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)
Số gia đình	20	35	25	10	10

b) Tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn của mẫu số liệu.

Khoảng biến thiên: $R = 25 - 0 = 25$ (giờ).

Xác định các tứ phân vị

Tứ phân vị thứ nhất Q_1 :

Nhóm chứa tứ phân vị thứ nhất là nhóm $[5;10)$.

$$Q_1 = 5 + \frac{\frac{100}{4} - 20}{35} \cdot (10 - 5) = \frac{40}{7}$$

Ta có:

Nhóm chứa tứ phân vị thứ ba là nhóm $[10;15)$.

$$Q_3 = 10 + \frac{\frac{3 \cdot 100}{4} - (20 + 35)}{25} \cdot (15 - 10) = 14$$

Tương tự:

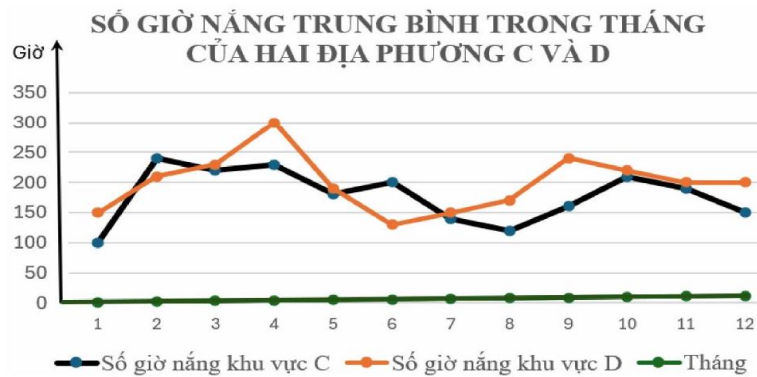
$$\Delta_Q = 14 - \frac{40}{7} = \frac{58}{7}$$

Vậy khoảng tứ phân vị

Số trung bình:

$$\bar{x}_A = \frac{1}{100} (2.5 \times 20 + 7.5 \times 35 + 12.5 \times 25 + 17.5 \times 10 + 22.5 \times 10) = 10,25 \text{ (giờ)}.$$

Câu 8. Hình sau là biểu đồ biểu diễn giờ nắng trung bình hàng tháng của hai địa phương C và D .



- a) Lập bảng số liệu ghép nhóm với dài các nhóm là 40 và đầu mút phải của nhóm cuối cùng là 300 .
 b) Tìm khoảng tứ phân vị của số giờ nắng mỗi khu vực và cho biết số giờ nắng của khu vực nào ít biến động hơn.

Lời giải

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

Khu vực C

Nhóm giờ nắng (giờ)	[100;140)	[140;180)	[180;220)	[220;260)	[260;300)
Số tháng	3	4	4	1	0

Khu vực D

Mức tiêu thụ nước (m ³)	[100;140)	[140;180)	[180;220)	[220;260)	[260;300)
Số tháng	1	5	3	2	1

b) Tìm các khoảng tứ phân vị

Khu vực C: Với $n = 12$.

Tứ phân vị thứ nhất Q_1^C :

Mẫu là 12 nên tứ phân vị thứ nhất là số thứ 3 ứng với khoảng $[100;140)$ và số thứ 4 ứng với khoảng $[140;180)$ nên: $Q_1^C = 140$.

$$Q_1^C = 140 + \frac{1 \cdot \frac{12}{4} - 1}{5} \cdot (180 - 140) = 156$$

Mẫu là 12 nên tứ phân vị thứ ba là số thứ 9 và số thứ 10 đều ứng với khoảng $[180;220)$ nên:

$$Q_3^C = 180 + \frac{3 \cdot \frac{12}{4} - (3+4)}{4} \cdot (220 - 180) = 200$$

Vậy khoảng tứ phân vị $\Delta_Q^C = 200 - 140 = 60$.

Khu dân cư D: Với $n = 12$.

Tứ phân vị thứ nhất Q_1^D :

Nhóm chứa tứ phân vị thứ nhất là nhóm $[140;180)$.

$$Q_1^D = 140 + \frac{1 \cdot \frac{12}{4} - 1}{5} \cdot (180 - 140) = 156$$

Ta có:

Nhóm chứa tứ phân vị thứ ba là nhóm $[220;260)$.

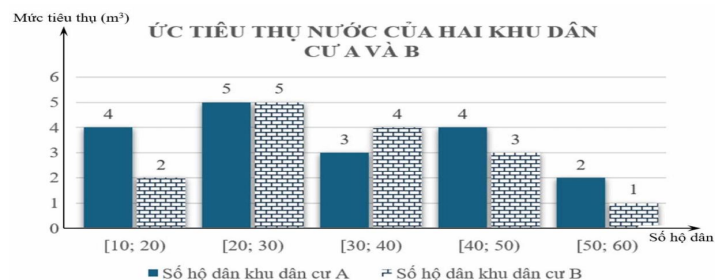
$$Q_3^D = 220 + \frac{3 \cdot \frac{12}{4} - (1+5+3)}{2} \cdot (260 - 220) = 220$$

Ta có:

Vậy khoảng tứ phân vị $\Delta_Q^D = 220 - 156 = 64$.

Vậy khu vực C có khoảng tứ phân vị nhỏ hơn khu vực D, cho thấy số giờ nắng của khu vực C ít biến động hơn so với khu vực D.

Câu 9. Biểu đồ sau mô tả kết quả điều tra về mức tiêu thụ nước hàng tháng của hai khu dân cư 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 khu dân cư nào có mức tiêu thụ nước đồng đều hơn?

c) Nếu so sánh theo độ lệch chuẩn của mẫu số liệu ghép nhóm thì khu dân cư nào có mức tiêu thụ nước đồng đều hơn?

Lời giải

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

Khu dân cư A

Mức tiêu thụ nước (m ³)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)	Tổng
Giá trị đại diện	15	25	35	45	55	
Số hộ dân	4	5	3	4	2	18

Khu dân cư B

Mức tiêu thụ nước (m ³)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)	Tổng
Giá trị đại diện	15	25	35	45	55	
Số hộ dân	2	5	4	3	1	15

b) So sánh theo khoảng tứ phân vị ghép nhóm

Khu dân cư A: Với $n = 18$.

Tứ phân vị thứ nhất Q_1^A :

Nhóm chứa tứ phân vị thứ nhất là nhóm [20; 30) .

$$Q_1^A = 20 + \frac{1 \cdot \frac{18}{4} - 4}{5} \cdot (30 - 20) = 21$$

Ta có:

Nhóm chứa tứ phân vị thứ ba là nhóm [40; 50) .

$$Q_3^A = 40 + \frac{3 \cdot \frac{18}{4} - (4 + 5 + 3)}{4} \cdot (50 - 40) = 43,75$$

Ta có:

Vậy khoảng tứ phân vị $\Delta_Q^A = 43,75 - 21 = 22,75$.

Khu dân cư B: với $n = 15$.

Tứ phân vị thứ nhất Q_1^B :

Nhóm chứa tứ phân vị thứ nhất là nhóm $[20; 30)$.

$$Q_1^B = 20 + \frac{1 \cdot \frac{15}{4} - 2}{5} \cdot (30 - 20) = 23,5$$

Ta có:

Nhóm chứa tứ phân vị thứ ba là nhóm $[40; 50)$.

$$Q_3^B = 40 + \frac{3 \cdot \frac{15}{4} - (2 + 5 + 4)}{3} \cdot (50 - 40) \approx 40,83$$

Ta có:

$$\Delta_Q^B = 40,83 - 23,5 = 17,33$$

Vậy khoảng tứ phân vị

Khu dân cư A có khoảng tứ phân vị lớn hơn khoảng tứ phân vị của khu dân cư B, cho thấy mức tiêu thụ nước hàng tháng của khu dân cư B đồng đều hơn so với khu dân cư A.

c) So sánh độ lệch chuẩn của mẫu số liệu ghép nhóm.

Khu dân cư A

$$\bar{x}_A = \frac{4 \times 15 + 5 \times 25 + 3 \times 35 + 4 \times 45 + 2 \times 55}{18} \approx 32,22$$

Số trung bình:

Độ lệch chuẩn:

$$S_A^2 = \frac{4 \times (15 - 32,22)^2 + 5 \times (25 - 32,22)^2 + 3 \times (35 - 32,22)^2 + 4 \times (45 - 32,22)^2 + 2 \times (55 - 32,22)^2}{18}$$

$$\Rightarrow s_A = \sqrt{S_A^2} \approx 11,79$$

Khu dân cư B

$$\bar{x}_B = \frac{2 \times 15 + 5 \times 25 + 4 \times 35 + 3 \times 45 + 1 \times 55}{15} \approx 32,33$$

Số trung bình:

Độ lệch chuẩn:

$$S_B^2 = \frac{2 \times (15 - 32,33)^2 + 5 \times (25 - 32,33)^2 + 4 \times (35 - 32,33)^2 + 3 \times (45 - 32,33)^2 + 1 \times (55 - 32,33)^2}{15}$$

$$\Rightarrow S_B = \sqrt{S_B^2} \approx 11,31$$

Vậy khu dân cư B có độ lệch chuẩn nhỏ hơn so với độ lệch chuẩn của khu dân cư A, cho thấy mức tiêu thụ nước hàng tháng của khu dân cư B đồng đều hơn khu dân cư A.

Câu 10. Điểm thi môn Toán của 25 học sinh được chọn ngẫu nhiên trong kỳ thi Tốt nghiệp Trung học Phổ thông năm học 2024 - 2025 được thống kê lại như sau:

4,4	8,6	4,8	5,0	5,4	5,4	5,6
5,2	6,0	6,2	6,2	8,0	7,8	7,6
7,2	7,4	7,2	6,0	5,8	5,8	8,2

7,0	6,8	6,4	6,6			
-----	-----	-----	-----	--	--	--

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 $[4;5)$ và độ dài mỗi nhóm bằng 1

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: $R = 8.6 - 4.4 = 4.2$ điểm.

Tứ phân vị:

Sắp xếp mẫu số liệu theo thứ tự không giảm:

4,4 4,8 5,0 5,2 5,4 5,4 5,6 5,8 5,8 6,0 6,0 6,2 6,2 6,4 6,6 6,8 7,0
7,2 7,2 7,4 7,6 7,8 8,0 8,2 8,6

Mẫu là 25 nên số trung vị là số thứ 13 nên $Q_2 = 6,2$.

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

$$Q_1 = \frac{1}{2}(x_6 + x_7) = \frac{5.4 + 5.6}{2} = 5.5$$

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

$$Q_3 = \frac{1}{2}(x_{18} + x_{19}) = \frac{7.2 + 7.4}{2} = 7.3$$

Khoảng tứ phân vị: $\Delta Q = Q_3 - Q_1 = 7.3 - 5.5 = 1.8$.

$$\bar{x} = \frac{4,4 + 4,8 + 5,0 + 5,4 + 5,4 + 5,6 + \dots + 8,0 + 8,2 + 8,6}{25} = 6,296$$

Số trung bình:

$$S^2 = \frac{1}{25} [(4,4 - 6,291)^2 + (4,8 - 6,296)^2 + \dots + (8,6 - 6,296)^2] \approx 1.32$$

Phương sai:

Độ lệch chuẩn: $s = \sqrt{S} = \sqrt{1.32} \approx 1.15$.

b) Lập bảng tần số ghép nhóm với nhóm đầu tiên là $[4;5)$ và độ dài mỗi nhóm bằng 1

Điểm	$[4;5)$	$[5;6)$	$[6;7)$	$[7;8)$	$[8;9)$
Số học sinh (Tần số)	2	7	9	5	2

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

Khoảng biến thiên: $R' = 9 - 4 = 5$.

Tứ phân vị của mẫu số liệu ghép nhóm:

Giá trị đại diện cho mỗi nhóm

Điểm	[4;5)	[5;6)	[6;7)	[7;8)	[8;9)
Giá trị đại diện	4,5	5,5	6,5	7,5	8,5
Số học sinh (Tần số)	2	7	9	5	2

$$[5;6): Q'_1 = 5 + \frac{\frac{25}{4} - 2}{7} \cdot (6 - 5) \approx 5,61$$

Nhóm chứa tứ phân vị thứ nhất là nhóm

$$[6;7): Q'_2 = 6 + \frac{2 \cdot \frac{25}{4} - (2+7)}{9} \cdot (7 - 6) \approx 6,39$$

Nhóm chứa tứ phân vị thứ hai là nhóm

$$[7;8): Q'_3 = 7 + \frac{3 \cdot \frac{25}{4} - (2+7+9)}{5} \cdot (8 - 7) = 7,15$$

Nhóm chứa tứ phân vị thứ ba là nhóm

Khoảng tứ phân vị là: $\Delta Q' = 7,15 - 5,61 = 1,54$

$$\bar{x} = \frac{2 \cdot 4,5 + 7 \cdot 5,5 + 9 \cdot 6,5 + 5 \cdot 7,5 + 2 \cdot 8,5}{25} = 6,42$$

Số trung bình của mẫu số liệu ghép nhóm:

Phương sai của mẫu số liệu ghép nhóm:

$$S^2 = \frac{1}{25} (2 \cdot 4,5^2 + 7 \cdot 5,5^2 + 9 \cdot 6,5^2 + 5 \cdot 7,5^2 + 2 \cdot 8,5^2) - 6,42^2 \approx 1,22$$

Độ lệch chuẩn của mẫu số liệu ghép nhóm là: $S' = \sqrt{S^2} = \sqrt{1,22} \approx 1,10$

Câu 11. Người ta thống kê chiều cao của 40 học sinh lớp 12A ta được bảng sau:

Chiều cao(cm)	[150;154]	[155;159]	[160;164]	[165;169]	[170;174]	[175;179]
Số học sinh	5	7	8	9	6	5

Tìm tứ phân vị thứ 2 .

Lời giải

Do mẫu $n = 40$ nên tứ phân vị thứ 2 là số thứ 20 ứng với khoảng [160;164] và số thứ 21 ứng

với khoảng $[165;169]$ nên tứ phân vị thứ hai $Q_2 = \frac{164 + 165}{2} = 164,5$

Câu 12. Người ta thống kê tuổi của 20 cá thể hổ tại 1 vườn bách thú ta được bảng sau:

Chiều cao (cm)	[10;11)	[11;12)	[12;13)	[13;14)	[14;15)	[15;16)
Số học sinh	2	3	6	4	4	1

Tìm tứ phân vị thứ 1 và 3 .

Lời giải

Do mẫu $n = 20$ nên tứ phân vị thứ 1 là số thứ 5 ứng với khoảng $[11;12)$ và số thứ 6 ứng với khoảng $[12;13)$ nên tứ phân vị thứ nhất $Q_2 = 12$.

Do mẫu $n = 20$ nên tứ phân vị thứ 3 là số thứ 15 ứng với khoảng $[13;14)$ và số thứ 16 ứng với khoảng $[14;15)$ nên tứ phân vị thứ nhất $Q_2 = 14$.

Câu 13. Cân nặng của một số lợn con mới sinh thuộc hai giống A và B được cho ở bảng đây (đơn vị: kg)

Cân nặng (kg)	$[1,0;1,1)$	$[1,1;1,2)$	$[1,2;1,3)$	$[1,3;1,4)$
Số con giống A	8	28	32	17
Số con giống B	13	14	24	14

- Hãy so sánh cân nặng của lợn con mới sinh giống A và giống B theo số trung bình và trung vị.
- Hãy ước lượng tứ phân vị thứ nhất và thứ ba của cân nặng lợn con mới sinh giống A và của cân nặng lợn con mới sinh giống B?
- Tính phương sai và độ lệch chuẩn của mẫu số liệu.

Lời giải

Cân nặng trung bình của giống A là: $\frac{1,05.8 + 1,15.28 + 1,25.32 + 1,35.17}{85} = 1,22$.

Cân nặng trung bình của giống B là: $\frac{1,05.13 + 1,15.14 + 1,25.24 + 1,35.14}{65} = 1,21$.

Vậy theo số trung bình thì cân nặng của lợn con giống A lớn hơn giống B .

Cân nặng	$[1,1;1,1)$	$[1,1;1,2)$	$[1,2;1,3)$	$[1,3;1,4)$
Giá trị đại diện	1,05	1,15	1,25	1,35
Số con giống A	8	28	32	17
Tần số lũy tiến nhóm A	8	36	68	85
Số con giống B	13	14	24	14
Tần số lũy tiến nhóm B	13	27	51	65

b) $Q_{1,A} = 1,1 + \frac{\frac{85}{4} - 8}{28} (1,2 - 1,1) = 1,15$.

$$Q_{3A} = 1,2 + \frac{\frac{3 \cdot 85}{4} - 36}{32} (1,3 - 1,2) = 1,29$$

$$Q_{1B} = 1,1 + \frac{\frac{65}{4} - 13}{14} (1,2 - 1,1) = 1,12$$

$$Q_{3B} = 1,2 + \frac{\frac{3 \cdot 65}{4} - 27}{24} (1,3 - 1,2) = 1,29$$

c)

$$s_A^2 = \frac{8 \cdot (1,05 - 1,22)^2 + 28 \cdot (1,15 - 1,22)^2 + 32 \cdot (1,25 - 1,22)^2 + 17 \cdot (1,35 - 1,22)^2}{85} = 8,05 \cdot 10^{-3}$$

$$\Rightarrow s_A = \sqrt{s_A^2} = 0,09$$

$$s_B^2 = \frac{13 \cdot (1,05 - 1,21)^2 + 14 \cdot (1,15 - 1,21)^2 + 24 \cdot (1,25 - 1,21)^2 + 14 \cdot (1,35 - 1,21)^2}{65} = 0,0107$$

$$\Rightarrow s_B = \sqrt{s_B^2} = 0,1$$

Câu 14. 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 bao nhiêu?

Lời giải

Mẫu là 25 nên tứ phân vị thứ nhất là số thứ 6 và số thứ 7 ứng với khoảng [10;12)

$$Q_1 = 10 + \frac{\frac{25}{4} - 4}{6} (12 - 10) = 10,75$$

Mẫu là 25 nên tứ phân vị thứ ba là số thứ 18 ứng với [12;14) và số thứ 19 ứng với khoảng [14;16)

$$Q_3 = 14$$

$$\text{Khoảng tứ phân vị } \Delta Q = Q_3 - Q_1 = 14 - 10,75 = 3,25$$

Câu 15 .

a) Bảng tần số ghép nhóm:

Nhóm	Giá trị đại diện	Tần số	Tần số tích lũy
------	------------------	--------	-----------------

[100,120)	110	1	1
[120,140)	130	4	5
[140,160)	150	7	12
[160,180)	170	11	23
[180,200)	190	7	30
		$n=30$	

b) Số trung bình cộng: $\bar{x} = \frac{1 \cdot 110 + 4 \cdot 130 + 7 \cdot 150 + 11 \cdot 170 + 7 \cdot 190}{30} \approx 163$ (lượt)

Số phần tử của mẫu là $n=30 \Rightarrow \frac{n}{2} = 15$.

Nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 15.

Ta có số trung vị của mẫu số liệu là $M_e = 160 + \frac{15-12}{11} \cdot 20 \approx 165$ (lượt)

Số phần tử của mẫu là $n=40$.

Ta có: $\frac{n}{4} = 7,5$. Nhóm 3 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 7,5.

Ta có tứ phân vị thứ nhất là $Q_1 = 140 + \frac{7,5-5}{7} \cdot 20 \approx 147$ (lượt).

Ta có: $\frac{3n}{4} = 22,5$. Nhóm 4 là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng 22,5

Ta có tứ phân vị thứ ba là $Q_3 = 160 + \frac{15-12}{11} \cdot 20 \approx 165$ (lượt).

Vậy tứ phân vị của mẫu số liệu trên là $Q_1 = 147; Q_2 = 165; Q_3 = 165$.

Câu 15. Mẫu số liệu dưới đây ghi số lượt xe ô tô khi đi qua trạm thu phí A trong 30 ngày (đơn vị: lượt).

100 134 124 150 138 140 155 176 180 198
156 173 185 163 165 185 190 143 179 123
176 150 167 168 198 178 182 164 175 153

a) Lập bảng tần số ghép nhóm cho mẫu số liệu trên có năm nhóm ứng với năm nửa khoảng:

$[100,120), [120,140), [140,160), [160,180), [180,200)$

b) Xác định số trung bình cộng, trung vị, tứ phân vị của mẫu số liệu ghép nhóm trên.

c) Một của mẫu số liệu ghép nhóm trên là bao nhiêu?

d) Tính phương sai và độ lệch chuẩn của mẫu số liệu

Câu 16. Một công ty xây dựng khảo sát khách hàng xem họ có nhu cầu mua nhà ở mức giá nào. Kết quả khảo sát được ghi lại ở bảng sau:

Mức giá (triệu đồng/ m^2)	$[10;14)$	$[14;18)$	$[18;22)$	$[22;26)$	$[26;30)$
Số khách hàng	54	78	120	45	12

a) Tìm một của mẫu số liệu ghép nhóm trên.

b) Công ty nên xây nhà ở mức giá nào để nhiều người có nhu cầu mua nhất?

Lời giải

Bảng đã hiệu chỉnh.

a) Nhóm chứa một của mẫu số liệu trên là nhóm $[18;22)$.

Do đó $u_m = 18, n_{m-1} = 78, n_m = 120, n_{m+1} = 45, u_{m+1} - u_m = 22 - 18 = 4$.

$$M_o = 18 + \frac{120 - 78}{(120 - 78) + (120 - 45)} \cdot 4 = \frac{758}{39} \approx 19,4$$

Một của mẫu số liệu ghép nhóm là

b) Dựa vào kết quả trên ta có thể dự đoán rằng nếu công ty xây nhà ở mức giá 19,4 triệu đồng/ m^2 thì sẽ có nhiều người có nhu cầu mua nhất.

Câu 17. Số cuộc gọi đi n thoại anh An thực hiện mỗi ngày trong 30 ngày được lựa chọn ngẫu nhiên được thống kê trong bảng sau:

Số cuộc gọi	$[3;5]$	$[6;8]$	$[9;11]$	$[12;14]$	$[15;17]$
Số ngày	5	10	3	10	2

Hỏi anh An gọi một ngày bao nhiêu cuộc gọi là nhiều nhất?

Lời giải

Hiệu chỉnh bảng lại như sau:

Số cuộc gọi	$[2,5;5,5)$	$[5,5;8,5)$	$[8,5;11,5)$	$[11,5;14,5)$	$[14,5;17,5)$
Số ngày	5	10	3	10	2

Nhóm chứa một của mẫu số liệu trên là nhóm $[5,5;8,5)$ và $[11,5;14,5)$.

$$[5,5;8,5) M_o^A = 5,5 + \frac{10-5}{(10-5)+(10-3)} \cdot 3 = 6,75$$

Nhóm

$$[11,5;14,5) M_o^B = 11,5 + \frac{10-3}{(10-3)+(10-2)} \cdot 3 = 12,9$$

Nhóm

Vậy anh An thực hiện cuộc gọi khoảng 7 cuộc gọi hoặc 13 cuộc gọi là nhiều nhất.

Câu 18. Anh Võ ghi lại cự li 40 lần ném lao của mình ở bảng sau (đơn vị: mét):

Cự li (m)	[69,2;70)	[70;70,8)	[70,8;71,6)	[71,6;72,4)	[72,4;73,2)
Số lần	5	6	10	10	9

Hỏi Anh Võ ném khoảng bao nhiêu mét là nhiều nhất?

Lời giải

Bảng đã hiệu chỉnh nên

Nhóm chứa một của mẫu số liệu trên là nhóm $[70,8;71,6)$ và $[71,6;72,4)$.

$$[70,8;71,6): M_o^A = 70,8 + \frac{10-6}{(10-6)+(10-10)} \cdot 0,8 = 71,6$$

Nhóm

$$M_o^B = 71,6 + \frac{10-10}{(10-10)+(10-9)} \cdot 0,8 = 71,6$$

Nhóm $[71,6;72,4)$:

Vậy anh Võ ném khoảng 71,6 m là nhiều nhất

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

Thời gian (phút)	[20;25)	[25;30)	[30;35)	[35;40)	[40;45)
Số ngày	6	5	4	2	1

Hỏi Cô Thúy Kiều nhảy thời gian bao nhiêu phút là nhiều nhất?

Lời giải

Bảng đã hiệu chỉnh nên

Nhóm chứa một của mẫu số liệu trên là nhóm $[20;25)$ nên một là

$$M_o = 20 + \frac{6-0}{(6-0)+(6-5)} \cdot (25-20) = \frac{170}{7}$$

Vậy Cô Thúy Kiều nhảy khoảng 24,3 phút là nhiều nhất.

Câu 20. Khi thống kê chiều cao của học sinh khối lớp 12 trong một trường trung học, ta thu được mẫu số liệu ghép nhóm sau:

Chiều cao (cm)	[150;156)	[156;162)	[162;168)	[168;174)	[174;180)	[180;186)
Số học sinh	5	18	40	26	8	3

Hãy tính giá trị trung bình, mốt, trung vị; tứ phân vị, phương sai và độ lệch chuẩn của mẫu số liệu trên.

Lời giải

Ta có bảng thống kê chiều cao học sinh lớp 12 theo giá trị đại diện:

Chiều cao đại diện	153	159	165	171	177	183
Số học sinh	5	18	40	26	8	3

- Chiều cao trung bình của học sinh lớp 11:

$$\bar{x} = \frac{153.5 + 159.18 + 165.40 + 171.26 + 177.8 + 183.3}{100} = 166,38$$

- Ta có nhóm chứa một của mẫu số liệu trên là $[162; 168)$, khi đó $u_m = 162, u_{m+1} = 168$ và $n_m = 40, n_{m+1} = 26, n_{m-1} = 18$

Một của mẫu số liệu ghép nhóm là:

$$\begin{aligned} M_0 &= u_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \cdot (u_{m+1} - u_m) \\ &= 162 + \frac{40 - 18}{(40 - 18) + (40 - 26)} \cdot (168 - 162) \approx 165,67 \end{aligned}$$

- Ta có $n = 100$ nên trung vị của mẫu số liệu thuộc nhóm $[162; 168)$, khi đó $u_m = 162, u_{m+1} = 168, n_m = 40, C = 5 + 18 = 23$

Trung vị của mẫu số liệu ghép nhóm là:

$$M_c = u_m + \frac{\frac{n}{2} - C}{n_m} \cdot (u_{m+1} - u_m) = 162 + \frac{\frac{100}{2} - 23}{40} \cdot (168 - 162) = 166,05$$

- Xác định tứ phân vị:

Tứ phân vị thứ hai: $Q_2 = M_c = 166,05$

Tứ phân vị thứ nhất của dãy số liệu $x_1, x_2, x_3, \dots, x_{100}$ là $x_{25} \in [162; 168)$. Do đó tứ phân vị thứ nhất của mẫu ghép nhóm là:

$$Q_1 = u_m + \frac{\frac{n}{4} - C}{n_m} \cdot (u_{m+1} - u_m) = 162 + \frac{\frac{100}{4} - 23}{40} \cdot (168 - 162) = 162,3$$

Tứ phân vị thứ ba của dãy số liệu $x_1, x_2, x_3, \dots, x_{100}$ là $x_{76} \in [168; 174)$, với $C = 5 + 18 + 40 = 63$. Do đó tứ phân vị thứ ba của mẫu ghép nhóm là

$$Q_3 = u_m + \frac{\frac{3n}{4} - C}{n_m} \cdot (u_{m+1} - u_m) = 168 + \frac{\frac{3 \cdot 100}{4} - 63}{26} \cdot (174 - 168) \approx 170,77$$

- Xác định phương sai và độ lệch chuẩn:

$$s^2 = \frac{5 \times (153 - 166,38)^2 + 18 \times (159 - 166,38)^2 + 40 \times (165 - 166,38)^2 + 26 \times (171 - 166,38)^2 + 8 \times (177 - 166,38)^2 + 3 \times (183 - 166,38)^2}{100}$$

$$= 42,8$$

$$\Rightarrow s = \sqrt{s^2} = 6,54$$