

CHƯƠNG 3

CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHEP NHÓM

BÀI 1

KHOẢNG BIẾN THIÊN, KHOẢNG TƯ PHÂN VỊ CỦA MẪU SỐ LIỆU GHEP NHÓM

1. Khoảng biến thiên

a. Định nghĩa

Nhóm	Tần số
$[a_1; a_2)$	n_1
$[a_2; a_3)$	n_2
...	...
$[a_m; a_{m+1})$	n_m
	n

Xét mẫu số liệu ghép nhóm cho bởi bảng trên.

Gọi a_1, a_{m+1} lần lượt là các đầu mút trái của nhóm 1, đầu mút phải của nhóm m .

Hiệu $R = a_{m+1} - a_1$ được gọi là khoảng biến thiên của mẫu số liệu ghép nhóm đó.

Chú 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.

b. Ý nghĩa

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.

Trong các đại lượng đo mức độ phân tán của của mẫu số liệu ghép nhóm, khoảng biến thiên là đại lượng dễ hiểu, dễ tính toán. Tuy nhiên, do khoảng biến thiên chỉ sử dụng hai giá trị a_1 và a_{m+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.

2. Khoảng tứ phân vị

a. Định nghĩa

Nhóm	Tần số	Tần số tích lũy
$[a_1; a_2)$	n_1	$cf_1 = n_1$
$[a_2; a_3)$	n_2	$cf_2 = n_1 + n_2$
...	...	...
$[a_m; a_{m+1})$	n_m	$cf_m = n_1 + n_2 + \dots + n_m$
	n	

Xét mẫu số liệu ghép nhóm cho bởi bảng trên.

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.

b) Ý nghĩa

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

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. Nhắc lại kiến thức 11.

a. Trung vị

Cho mẫu số liệu ghép nhóm như sau:

Nhóm	Tần số	Tần số tích lũy
$[a_1; a_2)$	n_1	$cf_1 = n_1$
$[a_2; a_3)$	n_2	$cf_2 = n_1 + n_2$
...	...	...
$[a_m; a_{m+1})$	n_m	$cf_m = n_1 + n_2 + \dots + n_m$
	n	

Giả sử nhóm k là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{n}{2}$, tức là $cf_{k-1} < \frac{n}{2}$ nhưng $cf_k \geq \frac{n}{2}$. Ta gọi r, d, n_k lần lượt là đầu mút trái, độ dài, tần số của nhóm k ; cf_{k-1} là tần số tích lũy của nhóm $k - 1$.

Trung vị của mẫu số liệu ghép nhóm, kí hiệu M_e , được tính theo công thức sau:

$$M_e = r + \left(\frac{\frac{n}{2} - cf_{k-1}}{n_k} \right) \cdot d$$

Trung vị của mẫu số liệu sau khi ghép nhóm xấp xỉ với trung vị của mẫu số liệu không ghép nhóm ban đầu và có thể dùng để đại diện cho mẫu số liệu đã cho.

b. Tứ phân vị

Cho mẫu số liệu ghép nhóm như sau:

Nhóm	Tần số	Tần số tích lũy
$[a_1; a_2)$	n_1	$cf_1 = n_1$
$[a_2; a_3)$	n_2	$cf_2 = n_1 + n_2$
...	...	...
$[a_m; a_{m+1})$	n_m	$cf_m = n_1 + n_2 + \dots + n_m$
	n	

- **Tứ phân vị thứ hai** Q_2 bằng trung vị M_e .

• Giả sử nhóm P là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{n}{4}$, tức là $cf_{k-1} < \frac{n}{4}$ nhưng $cf_k \geq \frac{n}{4}$. Ta gọi s, h, n_p lần lượt là đầu mút trái, độ dài, tần số của nhóm P ; cf_{p-1} là tần số tích lũy của nhóm $p-1$. Khi đó **tứ phân vị thứ nhất** Q_1 được tính theo công thức sau:

$$Q_1 = s + \left(\frac{\frac{n}{4} - cf_{p-1}}{n_p} \right) \cdot h$$

• Giả sử nhóm q là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\frac{3n}{4}$, tức là $cf_{q-1} < \frac{3n}{4}$ nhưng $cf_q \geq \frac{3n}{4}$. Ta gọi t, l, n_q lần lượt là đầu mút trái, độ dài, tần số của nhóm q ; cf_{q-1} là tần số tích lũy của nhóm $q-1$. Khi đó **tứ phân vị thứ ba** Q_3 được tính theo công thức sau:

$$Q_3 = t + \left(\frac{\frac{3n}{4} - cf_{q-1}}{n_q} \right) \cdot l$$

Ý nghĩa:

- Như ta đã biết, đối với mẫu số liệu không ghép nhóm đã sắp xếp theo thứ tự từ nhỏ đến lớn, các điểm Q_1, Q_2, Q_3 chia mẫu số liệu đó thành bốn phần, mỗi phần chứa 25% giá trị.
- Bằng cách ghép nhóm mẫu số liệu và tính toán tứ phân vị của mẫu số liệu ghép nhóm, ta nhận được ba giá trị mới cũng có thể dùng để đại diện cho mẫu số liệu đã cho.
- Các giá trị Q_1, Q_2, Q_3 trong tứ phân vị của mẫu số liệu sau khi ghép nhóm xấp xỉ với các tứ phân vị của mẫu số liệu không ghép nhóm ban đầu.

Câu 1. Bảng sau thống kê cân nặng của 50 quả xoài được lựa chọn ngẫu nhiên sau khi thu hoạch ở một nông trường

Cân nặng (g)	[250; 290)	[290; 330)	[330; 370)	[370; 410)	[410; 450)
Số quả xoài	3	13	18	11	5

Có ý kiến cho rằng: “Trong 50 quả xoài trên, hiệu số cân nặng của hai quả bất kì không vượt quá 200 g”. Ý kiến đó đúng hay sai? Giải thích.

Lời giải

Ý kiến nêu trên là đúng.

Giải thích: Quan sát bảng thống kê đã cho, ta thấy cân nặng lớn nhất quả xoài có thể đạt được là dưới 450 g, cân nặng nhỏ nhất quả xoài có thể đạt được là 250 g. Mà ta có $450 - 350 = 200$. Do đó, hai quả bất kì nào cũng có hiệu số cân nặng không vượt quá 200 g.

Câu 2. Kết quả điều tra tổng thu nhập trong năm 2024 của một số hộ gia đình ở thành phố Nha Trang được ghi lại ở bảng sau:

Tổng thu nhập (triệu đồng)	[200; 250)	[250; 300)	[300; 350)	[350; 400)	[400; 450)
Số hộ gia đình	24	62	34	21	9

a) Hãy tìm các tứ phân vị Q_1 và Q_3 .

b) Một doanh nghiệp địa phương muốn hướng dịch vụ của mình đến các gia đình có mức thu nhập ở tầm trung, tức là 50% các hộ gia đình có mức thu nhập ở chính giữa so với mức thu nhập của tất cả các hộ gia đình của địa phương. Hỏi doanh nghiệp cần hướng đến các gia đình có mức thu nhập trong khoảng nào?

Lời giải

a) Số hộ gia đình được khảo sát (cỡ mẫu) là $n = 24 + 62 + 34 + 21 + 9 = 150$.

Gọi $x_1; x_2; \dots; x_{150}$ là tổng thu nhập trong năm 2024 của 150 hộ gia đình được xếp theo thứ tự không giảm.

Ta có:

$$x_1; \dots; x_{24} \in [200; 250)$$

$$x_{25}; \dots; x_{86} \in [300; 350)$$

$$x_{87}; \dots; x_{120} \in [300; 350)$$

$$x_{121}; \dots; x_{141} \in [350; 400)$$

$$x_{142}; \dots; x_{150} \in [400; 450)$$

Do đó, đối với dãy số liệu $x_1; x_2; \dots; x_{150}$ thì

Tứ phân vị thứ nhất Q_1 là $x_{38} \in [250; 300)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 250 + \frac{\frac{150}{4} - 24}{62} (300 - 250) = \frac{16175}{62}$$

Tứ phân vị thứ ba Q_3 là $x_{113} \in [300; 350)$. Do đó, tứ phân thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 300 + \frac{\frac{3 \cdot 150}{4} - (24 + 62)}{34} (350 - 300) = \frac{11525}{34}$$

b) Doanh nghiệp cần hướng đến các gia đình có mức thu nhập trong khoảng:

$$[Q_1; Q_3) = \left[\frac{16175}{62}; \frac{11525}{34} \right) = [260,89; 338,97) \quad (\text{triệu đồng}).$$

Câu 3. Kết quả đo chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được cho ở bảng sau:

Chiều cao (m)	[8,4; 8,6)	[8,6; 8,8)	[8,8; 9,0)	[9,0; 9,2)	[9,2; 9,4)
Số cây	5	12	25	44	14

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

b) Trong 100 cây keo trên có 1 cây cao 8,4 m. Hỏi chiều cao của cây keo này có phải là giá trị ngoại lệ không?

Lời giải

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là:

$$R = 9,4 - 8,4 = 1 \text{ (m)}.$$

Cỡ mẫu $n = 100$.

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc về chiều cao của 100 cây keo 3 năm tuổi tại một nông trường được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_5 \in [8,4; 8,6),$$

$$x_6; \dots; x_{17} \in [8,6; 8,8),$$

$$x_{18}; \dots; x_{42} \in [8,8; 9,0),$$

$$x_{43}; \dots; x_{86} \in [9,0; 9,2),$$

$$x_{87}; \dots; x_{100} \in [9,2; 9,4).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{25} + x_{26}}{2} \in [8,8; 9,0)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 8,8 + \frac{\frac{100}{4} - (5 + 12)}{25} (9,0 - 8,8) = 8,864$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{75} + x_{76}}{2} \in [9,0; 9,2)$.

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

$$Q_3 = 20 + \frac{\frac{3 \cdot 100}{4} - (5 + 12 + 25)}{44} (9,2 - 9,0) = 9,15$$

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

$$\Delta_Q = Q_3 - Q_1 = 9,15 - 8,864 = 0,286$$

b) Trong 100 cây keo trên có 1 cây cao 8,4 m thuộc nhóm $[8,4; 8,6)$.

Vì $Q_1 - 1,5\Delta_Q = 8,864 - 1,5 \cdot 0,286 = 8,435 > 8,4$ nên chiều cao của cây keo cao 8,4 m là giá trị ngoại lệ của mẫu số liệu ghép nhóm.

Câu 4. Bạn Trang thống kê chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau:

Chiều cao (cm)	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
Số học sinh nữ lớp 12C	2	7	12	3	0	1
Số học sinh nữ lớp 12D	5	9	8	2	1	0

a) Sử dụng khoảng biến thiên, hãy cho biết chiều cao của học sinh nữ lớp nào có độ phân tán lớn hơn.

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của học sinh nữ lớp 12C và 12D.

Lời giải

a) Khoảng biến thiên của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12C là: $185 - 155 = 30$ (cm).

Trong mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D, khoảng đầu tiên chứa dữ liệu là $[155; 160)$ và khoảng cuối cùng chứa dữ liệu là $[175; 180)$.

Khoảng biến thiên của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D là: $180 - 155 = 25$ (cm).

Vậy nếu căn cứ theo khoảng biến thiên thì chiều cao của học sinh nữ lớp 12C có độ phân tán lớn hơn lớp 12D.

b)

• Lớp 12C:

Cỡ mẫu $n = 2 + 7 + 12 + 3 + 0 + 1 = 25$.

Gọi $x_1; x_2; \dots; x_{25}$ là mẫu số liệu gốc về chiều cao của 25 học sinh nữ lớp 12C được xếp theo thứ tự không giảm.

Ta có:

$$x_1; x_2 \in [155; 160)$$

$$x_3; \dots; x_9 \in [160; 165)$$

$$x_{10}; \dots; x_{21} \in [165; 170)$$

$$x_{22}; x_{23}; x_{24} \in [170; 175)$$

$$x_{25} \in [180; 185)$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_6 + x_7}{2} \in [160; 165)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 160 + \frac{\frac{25}{4} - 2}{7} (165 - 160) = \frac{4565}{28}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{19} + x_{20}}{2} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 165 + \frac{\frac{3 \cdot 25}{4} - (2 + 7)}{12} (170 - 165) = \frac{2705}{16}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12C là:

$$\Delta_Q = Q_3 - Q_1 = \frac{2705}{16} - \frac{4565}{28} \approx 6,03$$

• Lớp 12D:

ỡ mẫu $n' = 5 + 9 + 8 + 2 + 1 = 25$.

Gọi $y_1; y_2; \dots; y_{25}$ là mẫu số liệu gốc về chiều cao của 25 học sinh nữ lớp 12D được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_5 \in [155; 160),$$

$$y_6; \dots; y_{14} \in [160; 165),$$

$$y_{15}; \dots; y_{22} \in [165; 170),$$

$$y_{23}; y_{24} \in [170; 175),$$

$$y_{25} \in [175; 180).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{y_6 + y_7}{2} \in [160; 165)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là

$$Q'_1 = 160 + \frac{\frac{25}{4} - 5}{9} (165 - 160) = \frac{5785}{36}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{y_{19} + y_{20}}{2} \in [165; 170)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q'_3 = 165 + \frac{\frac{3.25}{4} - (5+9)}{8} (170 - 165) = \frac{5395}{32}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về chiều cao của các bạn học sinh nữ lớp 12D là:

$$\Delta'_o = Q'_3 - Q'_1 = \frac{5375}{32} - \frac{5785}{36} \approx 7,27$$

Ta có $\Delta_o \approx 6,03 < \Delta'_o \approx 7,27$

Câu 5. Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau:

Tuổi kết hôn	[19; 22)	[22; 25)	[25; 28)	[28; 31)	[31; 34)
Số phụ nữ khu vực A	10	27	31	25	7
Số phụ nữ khu vực B	47	40	11	2	0

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.

b) Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?

Lời giải

a)

• Khu vực A:

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực A là:

$$R = 34 - 19 = 15.$$

$$\text{Cỡ mẫu } n = 10 + 27 + 31 + 25 + 7 = 100.$$

Gọi $x_1; x_2; \dots; x_{100}$ là mẫu số liệu gốc về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực A được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_{10} \in [19; 22),$$

$$x_{11}; \dots; x_{37} \in [22; 25),$$

$$x_{38}; \dots; x_{68} \in [25; 28),$$

$$x_{69}; \dots; x_{93} \in [28; 31),$$

$$x_{94}; \dots; x_{100} \in [31; 34).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{25} + x_{26}}{2} \in [22; 25)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 22 + \frac{\frac{100}{4} - 10}{27} (25 - 22) = \frac{71}{3}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{75} + x_{76}}{2} \in [28; 31)$. Do đó, tứ phân thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 28 + \frac{\frac{3 \cdot 100}{4} - (10 + 27 + 31)}{25} (31 - 28) = \frac{721}{25}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực A là:

$$\Delta_Q = Q_3 - Q_1 = \frac{721}{25} - \frac{71}{3} \approx 5,17$$

• Khu vực B:

Khoảng biến thiên của mẫu số liệu ghép nhóm ứng với khu vực B là:

$$R' = 31 - 19 = 12.$$

Cỡ mẫu $n' = 47 + 40 + 11 + 2 = 100$.

Gọi $y_1; y_2; \dots; y_{100}$ là mẫu số liệu gốc về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực B được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_{47} \in [19; 22),$$

$$y_{48}; \dots; y_{87} \in [22; 25),$$

$$y_{88}; \dots; y_{98} \in [25; 28),$$

$$y_{99}; y_{100} \in [28; 31).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{y_{25} + y_{26}}{2} \in [19; 22)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là:

$$Q'_1 = 19 + \frac{\frac{100}{4}}{47} (22 - 19) = \frac{968}{47}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{y_{75} + y_{76}}{2} \in [22; 25)$. Do đó, tứ phân thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 22 + \frac{\frac{3 \cdot 100}{4} - 47}{40} (25 - 22) = \frac{241}{10}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình ở khu vực B là:

$$\Delta'_o = Q_3 - Q_1 = \frac{241}{10} - \frac{968}{47} \approx 3,5$$

Vì $\Delta_o \approx 5,17 > \Delta'_o \approx 3,5$ nên phụ nữ ở khu vực B có độ tuổi kết hôn đồng đều hơn.

Câu 6. Bảng sau thống kê tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau.

341,4	187,1	242,2	522,9	251,4	432,2	200,7	388,6	258,4	288,5
298,1	413,5	413,5	332	421	475	400	305	520	147

- Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu trên.
- Hãy chia mẫu số liệu trên thành 4 nhóm với nhóm đầu tiên là $[140; 240)$ và lập bảng tần số ghép nhóm.
- Hãy tìm khoảng biến thiên, khoảng tứ phân vị của mẫu số liệu ghép nhóm và so sánh với kết quả tương ứng thu được ở câu a).

Lời giải

a) Sắp xếp lại mẫu số liệu trên theo thứ tự không giảm, ta được:

147	187,1	200,7	242,2	251,4	258,4	288,5
298,1	305	332	341,4	388,6	400	413,5
413,5	421	432,2	475	520	522,9	

Khoảng biến thiên của mẫu số liệu trên là:

$$R = 522,9 - 147 = 375,9 \text{ (mm)}.$$

Cỡ mẫu $n = 20$.

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

$$147; 187,1; 200,7; 242,2; 251,4; 258,4; 288,5; 298,1; 305; 332.$$

$$Q_1 = \frac{251,4 + 258,4}{2} = 254,9$$

Do đó,

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

$$341,4; 388,6; 400; 413,5; 413,5; 421; 432,2; 475; 520; 522,9.$$

$$Q_3 = \frac{413,5 + 421}{2} = 417,25$$

Do đó,

Khoảng tứ phân vị của mẫu số liệu đã cho là:

$$\Delta_Q = Q_3 - Q_1 = 417,25 - 254,9 = 162,35 \quad \Delta_Q = Q_3 - Q_1 = 417,25 - 254,9 = 162,35.$$

b) Nhóm đầu tiên là $[140; 240)$, ta chọn 3 nhóm còn lại là

$[240; 340)$, $[340; 440)$, $[440; 540)$.

Từ bảng thống kê ban đầu, ta lập được bảng tần số ghép nhóm như sau:

Lượng mưa (mm)	$[140; 240)$	$[240; 340)$	$[340; 440)$	$[440; 540)$
Số tháng	3	7	7	3

c) Cỡ mẫu $n = 20$.

Khoảng biến thiên của mẫu số liệu ghép nhóm là

$$R' = 540 - 140 = 400 \text{ (mm)}.$$

Gọi $x_1; x_2; \dots; x_{20}$ là mẫu số liệu gốc về lượng mưa đo được vào tháng 7 từ năm 2002 đến 2021 tại một trạm quan trắc đặt ở Cà Mau được xếp theo thứ tự không giảm.

Ta có:

$$x_1; \dots; x_3 \in [140; 240),$$

$$x_4; \dots; x_{10} \in [240; 340),$$

$$x_{11}; \dots; x_{17} \in [340; 440),$$

$$x_{18}; \dots; x_{20} \in [440; 540).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_5 + x_6}{2} \in [240; 340)$.

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

$$Q'_1 = 240 + \frac{\frac{20}{4} - 3}{7} (340 - 240) = \frac{1880}{7}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{15} + x_{16}}{2} \in [340; 440)$.

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

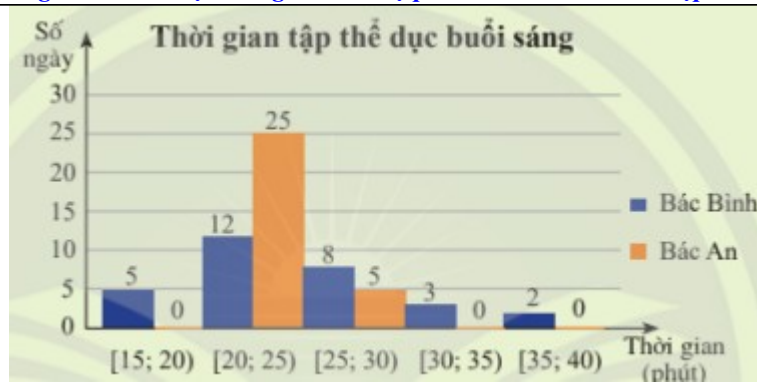
$$Q'_3 = 340 + \frac{\frac{3 \cdot 20}{4} - (3 + 7)}{7} (440 - 340) = \frac{2880}{7}$$

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

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{2880}{7} - \frac{1880}{7} = 142,86$$

Ta thấy khoảng biến thiên của mẫu số liệu ghép nhóm lớn hơn mẫu số liệu đã cho; khoảng tứ phân vị của mẫu số liệu ghép nhóm nhỏ hơn mẫu số liệu đã cho.

Câu 7. Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng của Bác Bình và Bác An



a) Ai là người có thời gian tập đều hơn?

b) Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.

Lời giải

Ta có bảng thống kê sau:

Thời gian (phút)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
Số ngày tập của Bác Bình	5	12	8	3	2
Số ngày tập của Bác An	0	25	5	0	0

a)

Khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác Bình là $40 - 15 = 25$ (phút).

Tuy nhiên, trong mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An, khoảng đầu tiên chứa dữ liệu là $[20; 25)$ và khoảng cuối cùng chứa dữ liệu là $[25; 30)$.

Do đó, khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng của bác An là $30 - 20 = 10$ (phút).

Nếu căn cứ theo khoảng biến thiên thì bác Bình có thời gian tập thể dục phân tán hơn bác An, vậy bác An là người có thời gian tập đều hơn.

b) Cỡ mẫu $n = 30$.

• Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình:

Gọi $x_1; x_2; \dots; x_{30}$ là mẫu số liệu gốc về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_5 \in [15; 20),$$

$$x_6; \dots; x_{17} \in [20; 25),$$

$$x_{18}; \dots; x_{25} \in [25; 30),$$

$$x_{26}; \dots; x_{28} \in [30; 35),$$

$$x_{29}; x_{30} \in [35; 40).$$

Tứ phân vị thứ nhất Q_1 của mẫu số liệu gốc là $x_8 \in [20; 25)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là

$$Q_1 = 20 + \frac{\frac{30}{4} - 5}{12} (25 - 20) = \frac{505}{24}$$

Tứ phân vị thứ ba Q_3 của mẫu số liệu gốc là $x_{23} \in [25; 30)$. Do đó, tứ phân thứ ba của mẫu số liệu ghép nhóm là

$$Q_3 = 25 + \frac{\frac{3.30}{4} - (5+12)}{8} (30 - 25) = \frac{455}{16}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình là

$$\Delta_Q = Q_3 - Q_1 = \frac{455}{16} - \frac{505}{24} = \frac{355}{48} \approx 7,4$$

- Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An:

Gọi $y_1; y_2; \dots; y_{30}$ là mẫu số liệu gốc về thời gian tập thể dục buổi sáng mỗi ngày của bác An được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_{25} \in [20; 25),$$

$$y_{26}; \dots; y_{30} \in [25; 30).$$

Tứ phân vị thứ nhất Q'_1 của mẫu số liệu gốc là $y_8 \in [20; 25)$. Do đó, tứ phân thứ nhất của mẫu số liệu ghép nhóm là

$$Q'_1 = 20 + \frac{\frac{30}{4}}{25} (25 - 20) = \frac{43}{2}$$

Tứ phân vị thứ ba Q'_3 của mẫu số liệu gốc là $y_{23} \in [20; 25)$. Do đó, tứ phân thứ ba của mẫu số liệu ghép nhóm là

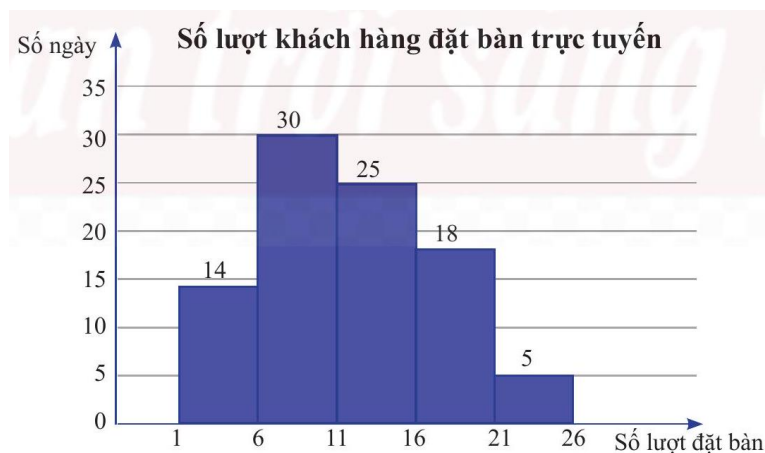
$$Q'_3 = 20 + \frac{\frac{3.30}{4}}{25} (25 - 20) = \frac{49}{2}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác An là

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{49}{2} - \frac{43}{2} = 3$$

Vì $\Delta_Q \approx 7,4 > \Delta'_Q = 3$ nên khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình lớn hơn bác An.

Câu 8. Biểu đồ dưới đây biểu diễn số lượt khách hàng đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2024 của một nhà hàng. Cột thứ nhất biểu diễn số ngày có từ 1 đến dưới 6 lượt đặt bàn; cột thứ hai biểu diễn số ngày có từ 6 đến dưới 11 lượt đặt bàn; ...



Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm cho bởi biểu đồ trên.

Lời giải

Từ biểu đồ đã cho, ta có có bảng thống kê sau:

Số lượt đặt bàn	[1; 6)	[6; 11)	[11; 16)	[16; 21)	[21; 26)
Số ngày	14	30	25	18	5

Cỡ mẫu $n = 14 + 30 + 25 + 18 + 5 = 92$.

Gọi $x_1; x_2; \dots; x_{92}$ là mẫu số liệu gốc về số lượt khách đặt bàn qua hình thức trực tuyến mỗi ngày trong quý III năm 2022 của một nhà hàng được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_{14} \in [1; 6),$$

$$x_{15}; \dots; x_{44} \in [6; 11),$$

$$x_{45}; \dots; x_{69} \in [11; 16),$$

$$x_{70}; \dots; x_{87} \in [16; 21),$$

$$x_{88}; \dots; x_{92} \in [21; 26).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $\frac{x_{23} + x_{24}}{2} \in [6; 11)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 6 + \frac{\frac{92}{4} - 14}{30} (11 - 6) = 7,5$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $\frac{x_{69} + x_{70}}{2}$.

Mà $x_{69} \in [11; 16)$ và $x_{70} \in [16; 21)$

Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là $Q_3 = 16$.

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

$$\Delta_Q = Q_3 - Q_1 = 16 - 7,5 = 8,5$$

Câu 9. Hai bảng tần số ghép nhóm dưới đây thống kê theo độ tuổi số lượng thành viên nam và thành viên nữ đang sinh hoạt trong một câu lạc bộ dưỡng sinh.

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nam	4	7	4	6	15	12	2	0

Khoảng tuổi	[50; 55)	[55; 60)	[60; 65)	[65; 70)	[70; 75)	[75; 80)	[80; 85)	[85; 90)
Số thành viên nữ	3	4	5	3	7	14	13	1

a) Hãy tính các khoảng tứ phân vị của tuổi nam giới và nữ giới trong mỗi bảng số liệu ghép nhóm trên.

b) Hãy cho biết trong câu lạc bộ trên, nam giới hay nữ giới có tuổi đồng đều hơn.

Lời giải

a)

• Nam giới:

Cỡ mẫu $n = 4 + 7 + 4 + 6 + 15 + 12 + 2 = 50$.

Gọi $x_1; x_2; \dots; x_{50}$ là mẫu số liệu gốc về tuổi của nam giới đang sinh hoạt trong câu lạc bộ dưỡng sinh được xếp theo thứ tự không giảm.

Ta có

$$x_1; \dots; x_4 \in [50; 55),$$

$$x_5; \dots; x_{11} \in [55; 60),$$

$$x_{12}; \dots; x_{15} \in [60; 65),$$

$$x_{16}; \dots; x_{21} \in [65; 70),$$

$$x_{22}; \dots; x_{36} \in [70; 75),$$

$$x_{37}; \dots; x_{48} \in [75; 80),$$

$$x_{49}; x_{50} \in [80; 85).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $x_{13} \in [60; 65)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q_1 = 60 + \frac{\frac{50}{4} - (4+7)}{4} (65 - 60) = 61,875$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $x_{38} \in [75; 80)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là:

$$Q_3 = 75 + \frac{\frac{3.50}{4} - (4 + 7 + 4 + 6 + 15)}{12} (80 - 75) = 75,625$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về tuổi của nam giới đang sinh hoạt trong câu lạc bộ dưỡng sinh là:

$$\Delta_Q = Q_3 - Q_1 = 75,625 - 61,875 = 13,75$$

• Nữ giới:

Cỡ mẫu $n' = 3 + 4 + 5 + 3 + 7 + 14 + 13 + 1 = 50$.

Gọi $y_1; \dots; y_{50}$ là mẫu số liệu gốc về tuổi của nữ giới đang sinh hoạt trong câu lạc bộ dưỡng sinh được xếp theo thứ tự không giảm.

Ta có

$$y_1; \dots; y_4 \in [50; 55),$$

$$y_4; \dots; y_7 \in [55; 60),$$

$$y_8; \dots; y_{12} \in [60; 65),$$

$$y_{13}; \dots; y_{15} \in [65; 70),$$

$$y_{16}; \dots; y_{22} \in [70; 75),$$

$$y_{23}; \dots; y_{36} \in [75; 80),$$

$$y_{37}; \dots; y_{49} \in [80; 85),$$

$$y_{50} \in [85; 90).$$

Tứ phân vị thứ nhất của mẫu số liệu gốc là $y_{13} \in [65; 70)$. Do đó, tứ phân vị thứ nhất của mẫu số liệu ghép nhóm là:

$$Q'_1 = 65 + \frac{\frac{50}{4} - (3 + 4 + 5)}{3} (70 - 65) = \frac{395}{6}$$

Tứ phân vị thứ ba của mẫu số liệu gốc là $y_{38} \in [80; 85)$. Do đó, tứ phân vị thứ ba của mẫu số liệu ghép nhóm là

$$Q'_3 = 80 + \frac{\frac{3.50}{4} - (3 + 4 + 5 + 3 + 7 + 14)}{13} (85 - 80) = \frac{2095}{26}$$

Khoảng tứ phân vị của mẫu số liệu ghép nhóm về tuổi của nữ giới đang sinh hoạt trong câu lạc bộ dưỡng sinh là:

$$\Delta'_Q = Q'_3 - Q'_1 = \frac{2095}{26} - \frac{395}{6} \approx 14,74$$

b) Ta có $\Delta'_Q \approx 14,74 > \Delta_Q = 13,75$ nên trong câu lạc bộ dưỡng sinh, nam giới có tuổi đồng đều hơn.