

# VII

## THỐNG KÊ

### 1. Số trung bình

- Giả sử ta có một mẫu số liệu là  $x_1, x_2, \dots, x_n$ .

**Số trung bình** (hay **số trung bình cộng**) của mẫu số liệu này, kí hiệu là  $\bar{x}$ , được tính bởi công thức  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ .

- Giả sử mẫu số liệu được cho dưới dạng bảng tần số

|         |       |       |     |       |
|---------|-------|-------|-----|-------|
| Giá trị | $x_1$ | $x_2$ | ... | $x_k$ |
| Tần số  | $n_1$ | $n_2$ | ... | $n_k$ |

Khi đó, công thức tính số trung bình trở thành  $\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n}$ , trong đó

$n = n_1 + n_2 + \dots + n_k$ . Ta gọi  $n$  là cỡ mẫu.

**Chú ý:** Kí hiệu  $f_k = \frac{n_k}{n}$  là tần số tương đối (hay còn gọi là tần suất) của  $x_k$  trong mẫu số liệu thì số trung bình còn có thể biểu diễn là  $\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_k x_k$ .

**- Ý nghĩa của số trung bình:** Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đo xu thế trung tâm của mẫu đó.

### 2. Trung vị và tứ phân vị

Sắp xếp lại mẫu số liệu theo thứ tự không giảm, ta được  $x_1 \leq x_2 \leq \dots \leq x_n$ .

- **Trung vị** của mẫu, kí hiệu là  $M_e$ , là giá trị ở chính giữa dãy  $x_1, x_2, \dots, x_n$ . Cụ thể:

+ Nếu  $n = 2k + 1, k \in \mathbb{N}$  thì trung vị mẫu là  $M_e = x_{k+1}$ .

+ Nếu  $n = 2k, k \in \mathbb{N}$  thì trung vị mẫu là  $M_e = \frac{1}{2}(x_k + x_{k+1})$ .

**- Ý nghĩa của trung vị:** Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có ít nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

- **Tứ phân vị** của một mẫu ngẫu nhiên gồm 3 giá trị, đó là **tứ phân vị thứ nhất, thứ hai và thứ ba** (lần lượt kí hiệu là  $Q_1, Q_2, Q_3$ ). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

+ Giá trị tứ phân vị thứ hai,  $Q_2$ , chính là trung vị của mẫu.

+ Giá trị tứ phân vị thứ nhất,  $Q_1$ , là trung vị của nửa số liệu đã sắp xếp bên trái  $Q_2$  (không bao gồm  $Q_2$  nếu  $n$  lẻ).

+ Giá trị tứ phân vị thứ ba,  $Q_3$ , là trung vị của nửa số liệu đã sắp xếp bên phải  $Q_2$  (không bao gồm  $Q_2$  nếu  $n$  lẻ).

- **Ý nghĩa của tứ phân vị:** Các điểm tứ phân vị  $Q_1, Q_2, Q_3$  chia mẫu số liệu đã sắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần chứa khoảng 25% tổng số số liệu đã thu thập được. Tứ phân vị thứ nhất  $Q_1$  còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liệu phía dưới. Tứ phân vị thứ ba  $Q_3$  còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu phía trên.

### 3. Một

- Cho một mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu, kí hiệu là  $M_o$ .

- **Ý nghĩa của mốt:** Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

*Chú ý:* Một mẫu số liệu có thể có nhiều mốt. Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

### 4. Khoảng biến thiên và khoảng tứ phân vị

Sắp xếp lại mẫu số liệu theo thứ tự không giảm, ta được  $x_1 \leq x_2 \leq \dots \leq x_n$ .

- **Khoảng biến thiên** của một mẫu số liệu, kí hiệu là  $R$ , là hiệu giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó, tức là:  $R = x_n - x_1$ .

- **Khoảng tứ phân vị**, kí hiệu là  $\Delta_Q$ , là hiệu giữa  $Q_3$  và  $Q_1$ , tức là  $\Delta_Q = Q_3 - Q_1$ .

- **Ý nghĩa của khoảng biến thiên và khoảng tứ phân vị**

+ Khoảng biến thiên đặc trưng cho độ phân tán của toàn bộ mẫu số liệu.

+ Khoảng tứ phân vị đặc trưng cho độ phân tán của một nửa các số liệu, có giá trị thuộc đoạn từ  $Q_1$  đến  $Q_3$  trong mẫu.

+ Khoảng tứ phân vị không bị ảnh hưởng bởi các giá trị rất lớn hoặc rất bé trong mẫu.

- **Giá trị ngoại lệ**

+ Khoảng tứ phân vị được dùng để xác định các **giá trị ngoại lệ** trong mẫu, đó là các giá trị quá nhỏ hay quá lớn so với đa số các giá trị của mẫu. Cụ thể, phần tử  $x$  trong mẫu là giá trị ngoại lệ nếu  $x > Q_3 + 1,5\Delta_Q$  hoặc  $x < Q_1 - 1,5\Delta_Q$ .

+ Khi mẫu có giá trị ngoại lệ, người ta thường sử dụng trung vị và khoảng tứ phân vị để đo mức độ tập trung và mức độ phân tán của đa số các phần tử trong mẫu số liệu.

## 5. Phương sai và độ lệch chuẩn

Giả sử ta có một mẫu số liệu là  $x_1, x_2, \dots, x_n$ .

- Phương sai của mẫu số liệu này, kí hiệu là  $S^2$ , được tính bởi công thức

$$S^2 = \frac{1}{n} \left[ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right] = \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - \bar{x}^2.$$

trong đó  $\bar{x}$  là số trung bình của mẫu số liệu.

- Căn bậc hai của phương sai được gọi là **độ lệch chuẩn**, kí hiệu là  $S$ .

Giả sử mẫu số liệu được cho dưới dạng bảng tần số

|         |       |       |         |       |
|---------|-------|-------|---------|-------|
| Giá trị | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
| Tần số  | $n_1$ | $n_2$ | $\dots$ | $n_k$ |

- Khi đó, công thức tính phương sai trở thành

$$S^2 = \frac{1}{n} \left[ n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_k (x_k - \bar{x})^2 \right]$$

trong đó  $n = n_1 + n_2 + \dots + n_k$ .

- Có thể biến đổi công thức tính phương sai trên thành  $S^2 = \frac{1}{n} (n_1 x_1^2 + n_2 x_2^2 + \dots + n_k x_k^2) - \bar{x}^2$

- **Ý nghĩa của phương sai và độ lệch chuẩn**

+ Phương sai là trung bình cộng của các bình phương độ lệch từ mỗi giá trị của mẫu số liệu đến số trung bình.

+ Phương sai và độ lệch chuẩn được dùng để đo mức độ phân tán của các số liệu trong mẫu quanh số trung bình. Phương sai và độ lệch chuẩn càng lớn thì các giá trị của mẫu càng cách xa nhau (có độ phân tán lớn).

## TÍNH SAI SỐ TUYỆT ĐỐI, ĐỘ CHÍNH XÁC CỦA MỘT SỐ GẦN ĐÚNG.

♦ Nếu  $a$  là số gần đúng của  $\bar{a}$  thì  $\Delta_a = |\bar{a} - a|$  được gọi là sai số tuyệt đối của số gần đúng  $a$

♦ Nếu  $\Delta_a \leq d$  thì  $a - d \leq \bar{a} \leq a + d$  khi đó ta viết  $\bar{a} = a \pm d$

$d$  gọi là độ chính xác của số gần đúng.

**Câu 1:** Cho số  $x = \frac{2}{7}$  và các giá trị gần đúng của  $x$  là 0,28; 0,29; 0,286; 0,3. Hãy xác định sai số tuyệt đối trong từng trường hợp và cho biết giá trị gần đúng nào là tốt nhất.

**Lời giải**

Ta có các sai số tuyệt đối là:

$$\Delta_a = \left| \frac{2}{7} - 0,28 \right| = \frac{1}{175}, \Delta_b = \left| \frac{2}{7} - 0,29 \right| = \frac{3}{700}, \Delta_c = \left| \frac{2}{7} - 0,286 \right| = \frac{1}{3500}, \Delta_d = \left| \frac{2}{7} - 0,3 \right| = \frac{1}{70}$$

Vì  $\Delta_c < \Delta_b < \Delta_a < \Delta_d$  nên  $c = 0,286$  là số gần đúng tốt nhất.

**Câu 2:** Hình chữ nhật có các cạnh  $x = 10 \pm 0,01$  m,  $y = 5 \pm 0,02$  m. Tính chu vi hình chữ nhật và sai số tuyệt đối của giá trị đó.

**Lời giải**

Ta có chu vi hình chữ nhật là  $P = 2(5 + 10) = 30$  m.

Giả sử  $x = 10 + a$ ,  $y = 5 + b$  với  $-0,01 \leq a \leq 0,01$ ;  $-0,02 \leq b \leq 0,02$ .

Do đó  $P = 2(x + y) = 2(15 + a + b) = 30 + 2(a + b)$ .

Vì  $-0,01 \leq a \leq 0,01$ ;  $-0,02 \leq b \leq 0,02$  nên  $-0,06 \leq 2(a + b) \leq 0,06$ .

Do đó  $|P - 30| = |2(a + b)| \leq 0,06$ .

Vậy  $P = 30\text{m} \pm 0,06\text{m}$ , sai số tuyệt đối  $\Delta_p \leq 0,06$ .

**Câu 3:** Chiều dài của một chi tiết máy được ghi  $l = 20\text{cm} \pm 0,01\text{cm}$ . Chiều dài đúng của cây thước là bao nhiêu?

**Lời giải**

Chiều dài đúng của cây thước là một số nằm trong khoảng từ  $20\text{cm} - 0,01\text{cm}$  đến  $20\text{cm} + 0,01\text{cm}$ , hay từ  $19,99\text{cm}$  đến  $20,01\text{cm}$ .

**Câu 4:** Số lượng que diêm của một bao diêm được ghi trên bao bì là  $45 \pm 5$  que. Số lượng que diêm của một bao diêm đúng quy định là bao nhiêu?

**Lời giải**

Số lượng que diêm của bao diêm đúng yêu cầu quy định là từ  $45 - 5$  đến  $45 + 5$ , hay từ 40 que đến 50 que diêm.

**Câu 5:** Cho  $\bar{a} = \frac{12}{7} = 1,71428517...$  Hãy xác định số gần đúng của  $\bar{a}$  với độ chính xác  $d = 0,002$ .

**Lời giải**

Hàng của chữ số khác 0 đầu tiên bên trái của  $d = 0,002$  là hàng phần nghìn. Quy tròn  $\bar{a}$  đến hàng phần trăm ta được số gần đúng của  $\bar{a}$  là  $a = 1,71$ .

**Câu 6:** Cho  $\bar{b} = \frac{1 - \sqrt{5}}{2} = -0,61803398...$  Hãy xác định số gần đúng của  $\bar{b}$  với độ chính xác  $d = 0,0005$ .

**Lời giải**

Hàng của chữ số khác 0 đầu tiên bên trái của  $d = 0,0005$  là hàng phần chục nghìn. Quy tròn  $\bar{b}$  đến hàng phần nghìn ta được số gần đúng của  $\bar{b}$  là  $b = -0,618$ .

**Câu 7:** An tính diện tích hình tròn có bán kính  $r = 4\text{cm}$  bằng công thức  $S = 3,145.4^2 = 50,32\text{cm}^2$ . Biết rằng  $3,14 < \pi < 3,15$ , hãy ước lượng độ chính xác của  $S$ .

**Lời giải**

Diện tích đúng kí hiệu là  $\bar{S}$  của hình tròn trên thỏa mãn

$$3,14.4^2 < \bar{S} < 3,15.4^2 \Leftrightarrow 50,24 < \bar{S} < 50,4$$

$$\text{Do đó } 50,24 - 50,32 < \bar{S} - S < 50,4 - 50,32 \Leftrightarrow |\bar{S} - S| < 0,08.$$

Vậy kết quả của An có độ chính xác là 0,08.

Nói cách khác, diện tích của hình tròn là  $50,32 \pm 0,08\text{cm}^2$

**Câu 8:** Độ dài đường chéo của một hình vuông có cạnh bằng 10cm được tính bởi  $d = 10.1,415 = 14,15$ . Cho biết  $1,41 < \sqrt{2} < 1,42$ . Hãy xác định độ chính xác của kết quả tìm được.

**Lời giải**

Gọi  $\bar{d}$  là độ dài đúng của đường chéo của hình vuông.

$$\text{Khi đó } \bar{d} = 10.\sqrt{2}\text{cm}$$

$$\text{Ta có } 1,41.10 < \bar{d} < 1,42.10 \Leftrightarrow 14,1 < \bar{d} < 14,2$$

$$\text{Do đó } 14,1 - 14,15 < \bar{d} - d < 14,2 - 14,15 \Leftrightarrow |\bar{d} - d| < 0,05.$$

Vậy độ chính xác của kết quả là 0,05.

**Câu 9:** Cho tam giác  $ABC$  có độ dài ba cạnh đo được như sau  $a = 12\text{cm} \pm 0,2\text{cm}$ ;  $b = 10,2\text{cm} \pm 0,2\text{cm}$ ;  $c = 8\text{cm} \pm 0,1\text{cm}$ . Tính chu vi  $P$  của tam giác và đánh giá sai số tuyệt đối

**Lời giải**

$$\text{Ta có } P = 12 + 10,2 + 8 = 30,2\text{cm}.$$

$$\text{Giả sử } a = 12 + d_1, b = 10,2 + d_2, c = 8 + d_3.$$

$$\text{Khi đó } P = a + b + c = 30,2 + d_1 + d_2 + d_3.$$

$$\text{Theo giả thiết, ta có } -0,2 \leq d_1 \leq 0,2; -0,2 \leq d_2 \leq 0,2; -0,1 \leq d_3 \leq 0,1.$$

$$\text{Suy ra } -0,5 \leq d_1 + d_2 + d_3 \leq 0,5.$$

$$\text{Do đó } P = 30,2\text{cm} \pm 0,5\text{cm}.$$

$$\text{Sai số tuyệt đối } \Delta_P \leq 0,5.$$

**Câu 10:** Hình vuông có cạnh  $x = 2 \pm 0,02\text{m}$ . Tính diện tích hình vuông và sai số tuyệt đối của giá trị đó.

**Lời giải**

$$\text{Diện tích hình vuông } S = x^2 = 4\text{m}^2.$$

Ta có  $x = 2 \pm 0,02 \text{ m} \Rightarrow 1,98 \text{ m} \leq x \leq 2,02 \text{ m}$ .

Do đó  $3,9204 \leq x^2 \leq 4,0804$ .

Sai số tuyệt đối  $\Delta_s \leq |4 - 4,0804| \Rightarrow \Delta_s \leq 0,0804$ .

### SAI SỐ TƯƠNG ĐỐI CỦA SỐ GẦN ĐÚNG

Sai số tương đối của số gần đúng  $a$ , kí hiệu là  $\delta_a$  là tỉ số giữa sai số tuyệt đối và  $|a|$ .

$$\circ \delta_a = \frac{\Delta_a}{|a|}.$$

$$\circ \delta_a \leq \frac{d}{|a|}.$$

Do đó  $\frac{d}{|a|}$  càng nhỏ thì chất lượng của phép đo đạc hay tính toán càng cao.

**Câu 11:** Trong một cuộc điều tra dân số, người ta viết dân số của một huyện là  $3574625 \text{ người} \pm 25000 \text{ người}$ . Hãy ước lượng sai số tương đối của số gần đúng này.

**Lời giải**

Số gần đúng  $a = 3574625$  người.

Độ chính xác  $d = 25000$  người.

Do đó sai số tương đối  $\delta_a \leq \frac{25000}{3574625} \approx 0,7\%$ .

**Câu 12:** Một thửa ruộng hình chữ nhật có chiều dài là  $x = 23 \text{ m} \pm 0,01 \text{ m}$  và chiều rộng là  $y = 15 \text{ m} \pm 0,01 \text{ m}$ . Tính chu vi của thửa ruộng và ước lượng sai số tương đối của số gần đúng của chu vi qua phép đo đó.

**Lời giải**

Giả sử  $x = 23 + a$ ,  $y = 15 + b$  với  $-0,01 \leq a, b \leq 0,01$ .

Ta có chu vi ruộng là  $P = 2(x + y) = 2(38 + a + b) = 76 + 2(a + b)$ .

Vì  $-0,01 \leq a, b \leq 0,01$  nên  $-0,04 \leq 2(a + b) \leq 0,04$ .

Do đó  $\bar{P} = 76 \text{ m} \pm 0,04 \text{ m}$ .

Sai số tương đối  $\delta_p \leq \frac{d}{P} \approx 0,0526\%$ .

**Câu 13:** Độ cao của một ngọn núi được ghi lại như sau  $\bar{h} = 1372 \text{ m} \pm 0,2 \text{ m}$ . Hãy ước lượng sai số tương đối của số gần đúng của phép đo đã cho.

**Lời giải**

Sai số tương đối  $\delta_h \leq \frac{0,2}{1372} = 0,00014577259 \approx 0,0146\%$

**Câu 14:** Đo chiều dài của một cây thước, ta được kết quả  $\bar{a} = 45 \pm 0,3(\text{cm})$ . Hãy ước lượng sai số tương đối của số gần đúng trong phép đo trên.

**Lời giải**

$$\text{Sai số tương đối } \delta_a \leq \frac{0,3}{45} = 0,00(6) \approx 0,67\%.$$

**Câu 15:** Cho tam giác  $ABC$  có độ dài ba cạnh đo được như sau  $a = 12\text{cm} \pm 0,2\text{cm}$ ;  $b = 10,2\text{cm} \pm 0,2\text{cm}$ ;  $c = 8\text{cm} \pm 0,1\text{cm}$ . Tính chu vi  $P$  của tam giác và ước lượng sai số tuyệt đối, sai số tương đối của số gần đúng của chu vi qua phép đo.

**Lời giải**

$$\text{Giả sử } a = 12 + d_1, b = 10,2 + d_2, c = 8 + d_3.$$

$$\text{Ta có } P = a + b + c + d_1 + d_2 + d_3 = 30,2 + d_1 + d_2 + d_3.$$

$$\text{Theo giả thiết, ta có } -0,2 \leq d_1 \leq 0,2; -0,2 \leq d_2 \leq 0,2; -0,1 \leq d_3 \leq 0,1.$$

$$\text{Suy ra } -0,5 \leq d_1 + d_2 + d_3 \leq 0,5.$$

$$\text{Do đó } \bar{P} = 30,2 \text{ cm} \pm 0,5 \text{ cm}.$$

$$\text{Sai số tuyệt đối } \Delta_p \leq 0,5.$$

$$\text{Sai số tương đối } \delta_p \leq \frac{d}{|P|} \approx 1,66\%.$$

**Câu 16:** Bạn Nam đo chiều dài của một cây cầu được kết quả là  $150 \pm 0,2\text{m}$ . Bạn Bình đo chiều cao của một dây nhả của trường mà hai bạn học được kết quả  $14 \pm 0,1\text{m}$ . Hãy đánh giá độ chính xác về kết quả đo của hai bạn Nam và Bình theo sai số tương đối.

**Lời giải**

$$\text{Phép đo của bạn Nam có sai số tương đối } \delta_1 \leq \frac{0,2}{150} \approx 0,0013333 \approx 0,133\%.$$

$$\text{Phép đo của bạn Bình có sai số tương đối } \delta_2 \leq \frac{0,1}{14} \approx 0,00714286 \approx 0,714\%.$$

$$\text{Ta có } \delta_1 < \delta_2.$$

Như vậy phép đo của bạn Nam có độ chính xác cao hơn phép đo của bạn Bình.

## CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM

Để xác định các số đặc trưng đo xu thế trung tâm ta làm như sau:

- ♦ Liệt kê số liệu theo thứ tự không giảm.
- ♦ Tính số trung bình, trung vị, mốt, tứ phân vị theo yêu cầu của bài toán.
- ♦ Kết luận về mẫu số liệu

**Câu 17:** Bảng số liệu sau cho biết số lượng học sinh đạt học lực giỏi năm học 2022 – 2023 của một số tỉnh

| Tỉnh        | Thanh Hóa | Hà Nội | Hải Phòng | Bạc Liêu | Nam Định | Thái Bình | Quảng Ninh | Bình Phước | Bình Dương |
|-------------|-----------|--------|-----------|----------|----------|-----------|------------|------------|------------|
| Số lượng HS | 2324      | 4305   | 3524      | 2324     | 4322     | 3846      | 4305       | 2542       | 4305       |

Giá trị về số trung bình, trung vị thay đổi thế nào nếu bỏ đi số lượng học sinh giỏi của tỉnh Bạc Liêu và Thái Bình

### Lời giải

$$\text{Số trung bình} = \frac{2324 + 4305 + 3524 + 2324 + 4322 + 3846 + 4305 + 2542 + 4305}{9} = 3533$$

Sắp xếp theo thứ tự không giảm:

2324 2324 2542 3524 3846 4305 4305 4305 4322

Trung vị là 3846

Khi bỏ đi số lượng học sinh giỏi của hai tỉnh Bạc Liêu và Thái Bình thì:

$$\text{Số trung bình} = \frac{2324 + 4305 + 3524 + 4322 + 4305 + 2542 + 4305}{7} = \frac{21305}{7}$$

Sắp xếp theo thứ tự không giảm:

2324 2542 3524 4305 4305 4305 4322

Trung vị là 4305

Vậy khi bỏ đi số lượng học sinh giỏi của hai tỉnh Bạc Liêu và Thái Bình thì cả số trung bình và trung vị đều thay đổi.

**Câu 18:** Hãy chọn số đặc trưng đo xu thế trung tâm của mỗi mẫu số liệu sau giải thích và tính giá trị của số đặc trưng đó với bảng số liệu sau cho biết số lượng học sinh đạt điểm 10 ở các môn học của một trường THPT.

| Môn         | Toán | Lý | Hóa | Sinh | Anh | Địa | Sử |
|-------------|------|----|-----|------|-----|-----|----|
| Số lượng HS | 80   | 70 | 30  | 25   | 5   | 7   | 2  |

### Lời giải

Chọn số đặc trưng là tứ phân vị, vì các số liệu không đều nhau, nhiều số liệu trong mẫu chênh lệch lớn so với trung vị

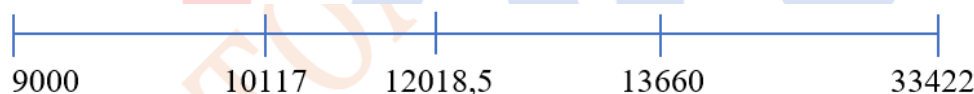
Tứ phân vị  $Q_1 = 5, Q_2 = 25, Q_3 = 70$ .

## Lời giải

$$Q_2 = \frac{(1200 + 12037)}{2} = 12018,5$$

Và tìm được  $Q_1 = \frac{(9000 + 11234)}{2} = 10117$

và tìm được  $Q_3 = \frac{(13200 + 14120)}{2} = 13660$



Các tứ phân vị cho ta hình ảnh phân bố của mẫu số liệu. Khoảng cách từ  $Q_1$  đến  $Q_2$  là 1901,5 trong khi khoảng cách từ  $Q_2$  đến  $Q_3$  là 1641,5 điều này cho thấy mẫu số liệu tập chung cao hơn ở bên trái  $Q_2$  và thấp hơn ở bên phải  $Q_2$ .

## Lời giải

$$R = 10 - 7 = 3$$

**Nhận xét.** Sử dụng khoảng biến thiên có ưu điểm là đơn giản, dễ tính toán song khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và giá trị nhỏ nhất mà bỏ qua thông tin từ tất cả các giá trị khác. Do đó, khoảng biến thiên rất dễ bị ảnh hưởng bởi các giá trị bất thường.

### Lời giải

Trước hết, ta sắp xếp mẫu số liệu theo thứ tự không giảm:

30 37 56 78 87 88 92 96 120

Mẫu số liệu gồm 9 giá trị nên trung vị là số ở vị trí chính giữa:  $Q_2 = 87$

Nửa số liệu bên trái là 30; 37; 56; 78 gồm 4 giá trị, hai phần tử chính giữa là: 37; 56

Do đó  $Q_1 = (37 + 56) : 2 = 46.5$

Nửa số liệu bên phải là: 88; 92; 96; 120 gồm 4 giá trị, hai phần tử chính giữa là: 92; 96

Do đó  $Q_3 = (92 + 96) : 2 = 94$

Vậy khoảng tứ phân vị cho mẫu số liệu là:  $\Delta_Q = Q_3 - Q_1 = 94 - 46.5 = 47.5$ .

**Câu 22:** Cho bảng số liệu ghi lại điểm của 40 học sinh bài kiểm tra 1 tiết môn Toán:

| Điểm        | 3 | 4 | 5 | 6  | 7 | 8 | 9 | 10 | Cộng |
|-------------|---|---|---|----|---|---|---|----|------|
| Số học sinh | 2 | 3 | 7 | 18 | 3 | 2 | 4 | 1  | 40   |

Tính khoảng biến thiên của mẫu số liệu này.

**Lời giải**

Điểm cao thấp nhất, cao nhất tương ứng là 10; 3. Do đó, khoảng biến thiên là:  $R = 10 - 3 = 7$ .

**Câu 23:** Số máy tính bán được trong 7 tháng liên tiếp của một cửa hàng được ghi lại trong bảng sau:

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| 83 | 79 | 92 | 71 | 69 | 83 | 74 |
|----|----|----|----|----|----|----|

Tính số trung bình, phương sai và độ lệch chuẩn. Nhận xét về số máy tính bán được trong 7 tháng liên tiếp của một cửa hàng có đều không?

**Lời giải**

Số trung bình là  $\bar{x} = \frac{83 + 79 + 92 + 71 + 69 + 83 + 74}{7} \approx 78,71$ .

Ta có  $\frac{1}{N} \sum_{i=1}^N x_i^2 = 6251,57$  và  $\frac{1}{N^2} \left( \sum_{i=1}^N x_i \right)^2 = 6195,94$ .

Suy ra  $s^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^N x_i \right)^2 = 6251,57 - 6195,94 = 55,63$ .

nên  $s = \sqrt{55,63} \approx 7,46$ .

Nhận xét độ lệch chuẩn lớn nên độ phân tán của các số liệu thống kê lớn suy ra số máy tính bán ra không được đều trong 7 tháng liên tiếp của một cửa hàng.

**Câu 24:** Hàm lượng Natri trong 100 g một số loại ngũ cốc được cho như sau:

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0   | 340 | 70  | 140 | 200 | 180 | 210 | 150 | 100 | 130 |
| 140 | 180 | 190 | 160 | 290 | 50  | 220 | 180 | 200 | 210 |

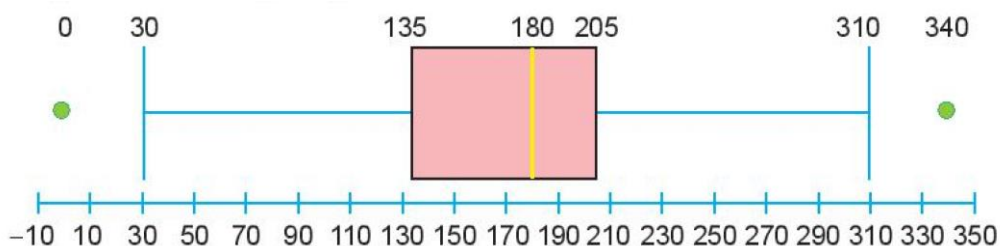
Tìm giá trị bất thường trong mẫu số liệu trên bằng cách sử dụng biểu đồ hộp.

**Lời giải**

Từ mẫu số liệu ta tính được  $Q_1 = 135$  và  $Q_3 = 205$ . Do đó, khoảng tứ phân vị là:

$$\Delta_o = 205 - 135 = 70$$

Biểu đồ hộp cho mẫu số liệu này là:



Ta có  $Q_1 - 1,5.\Delta_Q = 30$  và  $Q_3 + 1,5.\Delta_Q = 310$  nên trong mẫu số liệu có hai giá trị được xem là bất thường là 340 mg và 0 mg.

**Câu 25:** Trên hai con đường  $A$  và  $B$ , trạm kiểm soát đã ghi lại tốc độ (km/h) của 30 chiếc ô tô trên mỗi con đường như sau:

Con đường A: 

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 60 | 65 | 70 | 68 | 62 | 75 | 80 | 83 | 82 | 69 | 73 | 75 | 85 | 72 | 67 |
| 88 | 90 | 85 | 72 | 63 | 75 | 76 | 85 | 84 | 70 | 61 | 60 | 65 | 73 | 76 |

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 76 | 64 | 58 | 82 | 72 | 70 | 68 | 75 | 63 | 67 | 74 | 70 | 79 | 74 | 60 |
| 80 | 73 | 75 | 71 | 68 | 72 | 73 | 79 | 80 | 63 | 62 | 71 | 70 | 69 | 63 |

a) Tính số trung bình, phương sai và độ lệch chuẩn của tốc độ ô tô trên mỗi con đường A, B.

b) Theo em thì chạy xe trên con đường nào an toàn hơn?

## Lời giải

a) Trên con đường **A.** Ta có:  $\bar{x}_A \approx 73,63 \text{ km/h}$ ,  $s_A^2 \approx 74,77$ ,  $s_A \approx 8.65 \text{ km/h}$ .

Trên con đườn  $B$ . Ta có:  $\bar{x}_B \approx 70,7 \text{ km/h}$ ,  $s_B^2 \approx 38,21$ ,  $s_B \approx 6,18 \text{ km/h}$ .

b) Nhận xét: Trên con đường  $B$ , có tốc độ trung bình và độ lệch chuẩn đều nhỏ hơn trên con đường  $A$ . Do đó chạy xe trên con đường  $B$  sẽ an toàn hơn trên con đường  $A$ .

**Câu 26:** Thống kê 100 học sinh tham gia kì thi học sinh giỏi toán. Kết quả được thống kê trong bảng sau:

|        |   |    |    |    |    |    |    |    |    |    |    |           |
|--------|---|----|----|----|----|----|----|----|----|----|----|-----------|
| Điểm   | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |           |
| Tần số | 1 | 1  | 3  | 5  | 8  | 13 | 19 | 24 | 14 | 10 | 2  | $N = 100$ |

Tính độ lệch chuẩn của bảng số liệu thống kê.

### Lời giải

Điểm số trung bình của các học sinh tham gia thi học sinh giỏi là

$$\bar{x} = \frac{1.9+1.10+3.11+5.12+8.13+13.14+19.15+24.16+14.17+10.18+2.19}{100} = 15,23.$$

Phương sai của số liệu thống kê là

Suy ra độ lệch chuẩn của bảng số liệu thống kê là  $S_x = \sqrt{S_x^2} \approx 1,99$ .

| Năm<br>Tỉnh | 2014   | 2015   | 2016   | 2017   | 2018   |
|-------------|--------|--------|--------|--------|--------|
| Thái Bình   | 1061,9 | 1061,9 | 1053,6 | 942,6  | 1030,4 |
| Hậu Giang   | 1204,6 | 1293,1 | 1231,0 | 1261,0 | 1246,1 |

- ### Lời giải

- +/
- Tỉnh Thái Bình
- :

| <b>Giá trị</b> | <b>Độ lệch</b> | <b>Bình phương độ lệch</b> |
|----------------|----------------|----------------------------|
| 942,6          | -87,48         | 7652,75                    |
| 1030,4         | 0,32           | 0,10                       |
| 1053,6         | 23,52          | 553,19                     |
| 1061,9         | 31,82          | 1012,51                    |
| 1061,9         | 31,82          | 1012,51                    |
| <b>Tổng</b>    |                | <b>10231,06</b>            |

Độ lệch chuẩn là:  $s_1 = \sqrt{2046,21} \approx 45,24$ .

+ / Tỉnh Hậu Giang

| <b>Giá trị</b> | <b>Độ lệch</b> | <b>Bình phương độ lệch</b> |
|----------------|----------------|----------------------------|
| 1204,6         | -42,56         | 1811,35                    |
| 1231,0         | -16,16         | 261,15                     |
| 1246,1         | -1,06          | 1,12                       |
| 1261,0         | 13,84          | 191,55                     |
| 1293,1         | 45,94          | 2110,48                    |
| <b>Tổng</b>    |                | <b>4375,65</b>             |

12 | Ths. Lê Việt – THPT Thăng Long, Hà Nội – Phone: 0835221183 - Ôn thi cam kết điểm số

**b)** Tỉnh nào có sản lượng lúa ổn định hơn? Tại sao?

Vây tỉnh Hậu Giang có sản lượng lúa ổn định hơn.

|  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

### Lời giải

$$\bar{x} = \frac{26.128 + 27.105 + 28.119 + 29.16 + 30.8 + 31.15 + 32.9}{400} = 27,38.$$
$$s^2 = \frac{(26 - \bar{x})^2 \cdot 128 + (27 - \bar{x})^2 \cdot 105 + (28 - \bar{x})^2 \cdot 119 + (29 - \bar{x})^2 \cdot 16 + (30 - \bar{x})^2 \cdot 8 + (31 - \bar{x})^2 \cdot 15 + (32 - \bar{x})^2 \cdot 9}{400} = 1,9756.$$

Độ lệch chuẩn của mẫu số liệu là  $s = \sqrt{s^2} = \sqrt{1,9756} \approx 1,41$ .

|       |  |  |  |  |  |
|-------|--|--|--|--|--|
| hưởng |  |  |  |  |  |
| số    |  |  |  |  |  |

### Lời giải

Số trung bình của mẫu số liệu là  $\bar{x} = \frac{1.10 + 2.8 + 3.20 + 4.10 + 5.2}{50} = 2,72.$

$$s^2 = \frac{(1-\bar{x})^2 \cdot 10 + (2-\bar{x})^2 \cdot 8 + (3-\bar{x})^2 \cdot 20 + (4-\bar{x})^2 \cdot 10 + (5-\bar{x})^2 \cdot 2}{50} = 1,2416.$$

Độ lệch chuẩn của mẫu số liệu là  $s = \sqrt{s^2} = \sqrt{1,2416} \approx 1,114$ .

## Lời giải

Ta có số trung bình là  $\bar{x} = \frac{2+4+6+8+10}{5} = 6$ .

Áp dụng tính phương sai của mẫu số liệu là:  $s^2 = \frac{1}{5} \sum_{i=1}^5 (x_i - \bar{x})^2$ , ta có

$$s^2 = \frac{1}{5}[(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2] = 8.$$

Độ lệch chuẩn của mẫu số liệu trên là  $s = \sqrt{s^2} = \sqrt{8} = 2\sqrt{2}$ .

**Câu 31:** Người ta tiến hành phỏng vấn một số người về chất lượng của một loại sản phẩm mới. Người điều tra yêu cầu cho điểm sản phẩm kết quả như sau:

80 65 51 48 45 61 30 35 84 83 60 50 75  
72 68 39 41 54 61 72 75 72 61 58 65

- Tìm phương sai và độ lệch chuẩn.
- Nhận xét gì về các kết quả nhận được.

## Lời giải

Ta lập bảng phân bố tần số như sau

|        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Điểm   | 30 | 35 | 39 | 41 | 45 | 48 | 50 | 51 | 54 | 58 | 60 | 61 | 65 | 68 | 72 | 75 | 80 | 83 | 84 |
| Tần số | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 3  | 2  | 1  | 3  | 2  | 1  | 1  | 1  |

- a) Ta có:

Số trung bình là  $\bar{x} = \frac{1}{n}(n_1x_1 + n_2x_2 + \dots + n_kx_k)$

$$= \frac{1}{25}(1.30+1.35+1.39+1.41+1.45+1.48+1.50+1.51+1.54+1.58+1.60+3.61+2.65+1.68+3.72+2.75+1.80+1.83+1.84) = 60,2.$$

Phương sai:  $s_x^2 = \frac{1}{n} [n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2] = 216,8.$

Độ lệch chuẩn  $s_x = \sqrt{s_x^2} = \sqrt{216,8} = 14,724$ .

- b) Nhân xét: mức độ chênh lệch điểm giữa các giá trị là khá lớn.

**Câu 32:** Để khảo sát kết quả thi tuyển sinh môn Toán trong kì thi tuyển sinh đại học năm vừa qua của trường THPT Hai Bà Trưng, người điều tra chọn một mẫu gồm 100 học sinh tham gia kì thi tuyển sinh đó. Điểm môn Toán của các học sinh này được cho ở bảng phân bố tần số sau đây

|        |   |   |   |   |   |    |    |    |    |    |    |       |
|--------|---|---|---|---|---|----|----|----|----|----|----|-------|
| Điểm   | 0 | 1 | 2 | 3 | 4 | 5  | 6  | 7  | 8  | 9  | 10 |       |
| Tần số | 1 | 1 | 3 | 5 | 8 | 13 | 19 | 24 | 14 | 10 | 2  | N=100 |

- a) Tìm một, số trung vị.
- b) Tìm số trung bình, phương sai và độ lệch chuẩn.

**Lời giải**

a) Ta có giá trị có tần số lớn nhất  $M_o = 7$

Kích thước mẫu là số chẵn nên số trung vị là trung bình cộng của hai số đứng giữa

Vậy  $M_e = \frac{6+7}{2} = 6,5$

b) Số trung bình của mẫu số liệu là

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = \frac{0.1+1.1+2.3+\dots+10.2}{100} = 6,23$$

Ta có bảng sau:

| Giá trị | Độ lệch            | Bình phương độ lệch |
|---------|--------------------|---------------------|
| 0       | $0 - 6.23 = -6.23$ | 38.8129             |
| 1       | $1 - 6.23 = -5.23$ | 27.3529             |
| 2       | $2 - 6.23 = -4.23$ | 17.8929             |
| 3       | $3 - 6.23 = -3.23$ | 10.4329             |
| 4       | $4 - 6.23 = -2.23$ | 4.7929              |
| 5       | $5 - 6.23 = -1.23$ | 1.5129              |
| 6       | $6 - 6.23 = -0.23$ | 0.0529              |
| 7       | $7 - 6.23 = 0.77$  | 0.5929              |
| 8       | $8 - 6.23 = 1.77$  | 3.1329              |
| 9       | $9 - 6.23 = 2.77$  | 7.6729              |
| 10      | $10 - 6.23 = 3.77$ | 14.2129             |

Mẫu số liệu gồm 100 giá trị nên  $n = 100$ .

Do đó phương sai là  $s^2 = 3.9571$ .

Độ lệch chuẩn là  $s = \sqrt{3.9571} \approx 1,99$ .

**Câu 33:** Nhiệt độ trung bình các tháng trong năm 2021 tại Vĩnh Phúc và TP Thủ Đức được cho trong bảng sau

| Tháng      | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Vĩnh Phúc  | 16,1 | 17,0 | 20,2 | 23,7 | 27,3 | 28,8 | 29,9 | 28,2 | 27,2 | 24,6 | 21,4 | 18,2 |
| TP Thủ Đức | 25,8 | 26,7 | 27,9 | 28,9 | 28,3 | 27,5 | 27,1 | 27,1 | 26,8 | 26,7 | 26,4 | 25,7 |

a) Tính khoảng biến thiên, khoảng tứ phân và độ lệch chuẩn cho mỗi dãy số liệu trên.

b) Có nhận xét gì về sự biến động của nhiệt độ trung bình các tháng trong năm của tỉnh Vĩnh Phúc và thành phố Thủ Đức?

**Lời giải**

a) Với dãy số liệu về nhiệt độ trung bình các tháng của Vĩnh Phúc:

+ Giá trị nhỏ nhất là 16,1

+ Giá trị lớn nhất là 29,9

+ Khoảng biến thiên là  $R = 29,9 - 16,1 = 13,8$

Dãy số liệu sắp xếp thứ tự không giảm

16,1 17,0 18,2 20,2 21,4 23,7 24,6 27,2 27,3 28,2 28,8 29,9

+ Trung vị  $Q_2 = (23,7 + 24,6) : 2 = 24,15$

Nửa dữ liệu bên trái  $Q_2$  là: 16,1 17,0 18,2 20,2 21,4 23,7

Do đó  $Q_1 = (18,2 + 20,2) : 2 = 19,2$

Nửa dữ liệu bên phải  $Q_2$  là: 24,6 27,2 27,3 28,2 28,8 29,9

Do đó  $Q_3 = (27,3 + 28,2) : 2 = 27,75$

+ Khoảng tứ phân vị cho mẫu số liệu  $\Delta Q = Q_3 - Q_1 = 27,75 - 19,2 = 8,55$

+ Số trung bình của mẫu số liệu là

$$\bar{x} = \frac{16,1 + 17,0 + 18,2 + 20,2 + 21,4 + 23,7 + 24,6 + 27,2 + 27,3 + 28,2 + 28,8 + 29,9}{12} = 23,55.$$

+ Độ lệch chuẩn là

$$S_1 = \sqrt{\frac{(16,1 - 23,55)^2 + (17,0 - 23,55)^2 + \dots + (28,8 - 23,55)^2 + (29,9 - 23,55)^2}{12}} \approx 4,66.$$

Với dãy số liệu về nhiệt độ trung bình các tháng của thành phố Thủ Đức

+ Giá trị nhỏ nhất là 25,7

+ Giá trị lớn nhất là 28,9

+ Khoảng biến thiên là  $R = 28,9 - 25,7 = 3,2$

Dãy số liệu sắp xếp thứ tự không giảm

25,7 25,8 26,4 26,7 26,7 26,8 27,1 27,1 27,5 27,9 28,3 28,9

+ Trung vị  $Q_2 = (26,8 + 27,1) : 2 = 26,95$

Nửa dữ liệu bên trái  $Q_2$  là:

25,7 25,8 26,4 26,7 26,7 26,8

Do đó  $Q_1 = (26,4 + 26,7) : 2 = 26,55$

Nửa dữ liệu bên phải  $Q_2$  là:

27,1 27,1 27,5 27,9 28,3 28,9

Do đó  $Q_3 = (27,5 + 27,9) : 2 = 27,7$

+ Khoảng tứ phân vị cho mẫu số liệu  $\Delta Q = Q_3 - Q_1 = 27,7 - 26,55 = 1,15$

+ Số trung bình của mẫu số liệu là

$$\bar{x} = \frac{25,7 + 25,8 + 26,4 + 26,7 + 26,7 + 26,8 + 27,1 + 27,1 + 27,5 + 27,9 + 28,3 + 28,9}{12} \approx 27,08.$$

+ Độ lệch chuẩn là

$$S_2 = \sqrt{\frac{(25,7 - 27,08)^2 + (25,8 - 27,08)^2 + \dots + (28,3 - 27,08)^2 + (28,9 - 27,08)^2}{12}} \approx 0,91.$$

b) Khoảng biến thiên của Vĩnh Phúc cao hơn nhiều so với TP Thủ Đức

Khoảng tứ phân vị của Vĩnh Phúc cao hơn nhiều so với TP Thủ Đức

Độ lệch chuẩn của Vĩnh Phúc cao hơn nhiều so với TP Thủ Đức

Vậy nhiệt độ trung bình của năm 2021 của tỉnh Vĩnh Phúc khắc nghiệt hơn so với thành phố Thủ Đức.

**Câu 34:** Bảng dưới đây thống kê tổng số giờ nắng trong năm 2019 theo từng tháng được đo bởi hai trạm quan sát khí tượng đặt ở Tuyên Quang và Cà Mau.

| Tháng       | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Tuyên Quang | 25  | 89  | 72  | 117 | 106 | 177 | 156 | 203 | 227 | 146 | 117 | 145 |
| Cà Mau      | 180 | 223 | 257 | 245 | 191 | 111 | 141 | 134 | 130 | 122 | 157 | 173 |

a) Hãy tính phương sai và độ lệch chuẩn của dữ liệu từng tỉnh.

b) Nêu nhận xét về sự thay đổi tổng số giờ nắng theo từng tháng ở mỗi tỉnh.

**Lời giải**

a) Hãy tính phương sai và độ lệch chuẩn của dữ liệu từng tỉnh.

+/- Tỉnh Tuyên Quang:

$$\text{Số trung bình: } \bar{x}_1 = \frac{25 + 89 + 72 + 117 + 106 + 177 + 156 + 203 + 227 + 146 + 117 + 145}{12} \approx 131,67.$$

| Giá trị | Độ lệch | Bình phương độ lệch |
|---------|---------|---------------------|
| 25      | -106,67 | 11378,49            |
| 72      | -59,67  | 3560,51             |
| 89      | -42,67  | 1820,73             |
| 106     | -25,67  | 658,95              |
| 117     | -14,67  | 215,21              |
| 117     | -14,67  | 215,21              |
| 145     | 13,33   | 177,69              |
| 146     | 14,33   | 205,35              |
| 156     | 24,33   | 591,95              |
| 177     | 45,33   | 2054,81             |
| 203     | 71,33   | 5087,97             |
| 227     | 95,33   | 9087,81             |
| Tổng    |         | 35054,68            |

Mẫu số liệu gồm 12 giá trị nên  $n = 12$ . Do đó phương sai là:  $s_1^2 = \frac{35054,68}{12} \approx 2921,22$ .

Độ lệch chuẩn là:  $s_1 = \sqrt{2921,22} \approx 54,05$ .

+/ Tỉnh Cà Mau:

$$\text{Số trung bình: } \bar{x}_2 = \frac{180 + 223 + 257 + 245 + 191 + 111 + 141 + 134 + 130 + 122 + 157 + 173}{12} = 172.$$

| Giá trị | Độ lệch | Bình phương độ lệch |
|---------|---------|---------------------|
| 111     | -61     | 3721                |
| 122     | -50     | 2500                |
| 130     | -42     | 1764                |
| 134     | -38     | 1444                |
| 141     | -31     | 961                 |
| 157     | -15     | 225                 |
| 173     | 1       | 1                   |
| 180     | 8       | 64                  |
| 191     | 19      | 361                 |
| 223     | 51      | 2601                |
| 245     | 73      | 5329                |
| 257     | 85      | 7225                |
| Tổng    |         | 26196               |

Mẫu số liệu gồm 12 giá trị nên  $n=12$ . Do đó phương sai là:  $s_2^2 = \frac{26196}{12} = 2183$ .

Độ lệch chuẩn là:  $s_2 = \sqrt{2183} \approx 46,72$ .

**b)** Nêu nhận xét về sự thay đổi tổng số giờ nắng theo từng tháng ở mỗi tỉnh.

Phương sai mẫu và độ lệch chuẩn mẫu số liệu ở tỉnh Tuyên Quang cao hơn tỉnh Cà Mau nên tổng số giờ nắng trong năm 2019 theo từng tháng ở tỉnh Tuyên Quang có độ phân tán cao hơn ở tỉnh Cà Mau. Do đó, sự thay đổi tổng số giờ nắng theo từng tháng ở tỉnh Cà Mau ổn định hơn so với tỉnh Tuyên Quang.

**Câu 35:** Mẫu số liệu sau cho biết điểm kiểm tra thường xuyên của 20 học sinh lớp 10A.

|   |   |   |   |   |    |   |   |   |    |
|---|---|---|---|---|----|---|---|---|----|
| 4 | 6 | 8 | 6 | 9 | 10 | 7 | 7 | 6 | 5  |
| 5 | 2 | 5 | 9 | 9 | 8  | 3 | 8 | 8 | 10 |

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

**Lời giải**

Sắp xếp mẫu số liệu theo thứ tự không giảm:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |    |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|
| 2 | 3 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|

Khoảng biến thiên:  $R = 10 - 2 = 8$ .

Khoảng tứ phân vị:  $Q_2 = 7, Q_3 = 8,5, Q_1 = 5 \Rightarrow \Delta_Q = Q_3 - Q_1 = 8,5 - 5 = 3,5$ .

Số trung bình của mẫu số liệu là

$$\bar{x} = \frac{2.1 + 3.1 + 4.1 + 5.3 + 6.3 + 7.2 + 8.4 + 9.3 + 10.2}{20} = 6,75.$$

Phương sai của mẫu số liệu là

$$s^2 = \frac{(2-\bar{x})^2 \cdot 1 + (3-\bar{x})^2 \cdot 1 + (4-\bar{x})^2 \cdot 1 + (5-\bar{x})^2 \cdot 3 + (6-\bar{x})^2 \cdot 3 + (7-\bar{x})^2 \cdot 2 + (8-\bar{x})^2 \cdot 4 + (9-\bar{x})^2 \cdot 3 + (10-\bar{x})^2 \cdot 2}{20} = 4,8875$$

Độ lệch chuẩn của mẫu số liệu là

$$s = \sqrt{s^2} = \sqrt{4,8875} \approx 2,21.$$

**Câu 36:** Một mẫu số liệu có tứ phân vị thứ nhất là 54, tứ phân vị thứ ba là 76. Có bao nhiêu giá trị bất thường trong các giá trị sau: 19, 26, 58, 72, 105, 115?

**Lời giải**

Ta có  $Q_1 = 54; Q_3 = 76$

$$\Delta_Q = 76 - 54 = 22.$$

$$Q_1 - 1,5 \cdot \Delta_Q = 54 - 1,5 \cdot 22 = 21; Q_3 + 1,5 \cdot \Delta_Q = 76 + 1,5 \cdot 22 = 109$$

Ta thấy  $19 < Q_1 - 1,5 \cdot \Delta_Q$  và  $115 > Q_3 + 1,5 \cdot \Delta_Q$  nên 19, 115 là các giá trị bất thường.

Vậy có tất cả 2 giá trị bất thường trong các giá trị trên.

**Câu 37:** Đo chiều cao (cm) của 11 học sinh tổ 2 của lớp 10A ta có mẫu số liệu: 145, 164, 164, 155, 154, 174, 158, 146, 180, 160, 157. Xác định các giá trị bất thường của mẫu số liệu.

**Lời giải**

Ta sắp xếp mẫu số liệu theo thứ tự không giảm như sau: 145, 146, 154, 155, 157, 158, 160, 164, 164, 174, 180.

Vì  $n = 11$ , nên ta có  $Q_2 = 158$

Ta tìm  $Q_1$  nửa bên trái  $Q_2$  là 145, 146, 154, 155, 157 gồm 5 giá trị và ta tìm được  $Q_1 = 154$

Ta tìm  $Q_3$  nửa bên phải  $Q_2$  là 160, 164, 164, 174, 180 gồm 5 giá trị và ta tìm được  $Q_3 = 164$

$$\text{Vậy } \Delta_Q = Q_3 - Q_1 = 164 - 154 = 10$$

$$\text{Ta có } Q_1 - 1,5 \cdot \Delta_Q = 154 - 1,5 \cdot 10 = 139; Q_3 + 1,5 \cdot \Delta_Q = 164 + 1,5 \cdot 10 = 179$$

Ta thấy  $180 > Q_3 + 1,5 \cdot \Delta_Q$  nên 180 là giá trị bất thường của mẫu số liệu.

**Câu 38:** Điểm khảo sát đầu năm môn Toán của 30 học sinh trong một lớp của khối 10 được thống kê như sau:

|   |   |   |   |    |   |   |    |   |   |   |   |   |    |   |
|---|---|---|---|----|---|---|----|---|---|---|---|---|----|---|
| 3 | 5 | 7 | 9 | 10 | 6 | 8 | 3  | 4 | 6 | 5 | 7 | 8 | 10 | 7 |
| 9 | 3 | 6 | 4 | 7  | 8 | 9 | 10 | 6 | 9 | 7 | 4 | 5 | 3  | 8 |

Tìm giá trị bất thường trong mẫu số liệu trên.

**Lời giải**

+) Sắp xếp mẫu số liệu theo thứ tự không giảm:

3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10

+) Từ mẫu số liệu ta tính được:  $Q_2 = 7$ ;  $Q_1 = 5$  và  $Q_3 = 8$ . Do đó khoảng tứ phân vị là:

$$\Delta_Q = 8 - 5 = 3$$

+) Ta có  $Q_1 - 1,5\Delta_Q = 0,5$  và  $Q_3 + 1,5\Delta_Q = 12,5$  nên trong mẫu số liệu không có giá trị bất thường.

**Câu 39:** Trong một kì thi chọn học sinh giỏi môn Toán dành cho khối 12 kết quả cho trong bảng sau:

| Điểm   | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|--------|---|----|----|----|----|----|----|----|----|----|----|
| Tần số | 1 | 1  | 3  | 5  | 8  | 13 | 19 | 24 | 14 | 10 | 2  |

Tìm giá trị bất thường trong mẫu số liệu trên.

**Lời giải**

+) Từ mẫu số liệu ta tính được  $Q_1 = 11$  và  $Q_3 = 17$ . Do đó khoảng tứ phân vị là:

$$\Delta_Q = 17 - 11 = 6$$

+) Ta có  $Q_1 - 1,5\Delta_Q = 2$  và  $Q_3 + 1,5\Delta_Q = 26$  nên trong mẫu số liệu không có giá trị bất thường.

**Câu 40:** Để điều tra số con trong một gia đình trong một tổ dân phố người ta chọn ra 20 gia đình thu được mẫu số liệu sau đây:

2, 4, 2, 1, 3, 5, 1, 1, 2, 3, 1, 2, 2, 3, 4, 1, 1, 2, 3, 4.

Tìm giá trị bất thường trong mẫu số liệu trên.

**Lời giải**

+) Sắp xếp mẫu số liệu theo thứ tự không giảm:

1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5.

+) Từ mẫu số liệu ta tính được  $Q_1 = 1$  và  $Q_3 = 3$ . Do đó khoảng tứ phân vị là:  $\Delta_Q = 3 - 1 = 2$ .

+) Ta có  $Q_1 - 1,5\Delta_Q = -2$  và  $Q_3 + 1,5\Delta_Q = 6$  nên trong mẫu số liệu không có giá trị bất thường.

**Câu 41:** Trong một vườn trồng các cây chôm chôm người ta chọn ngẫu nhiên 100 quả và kiểm tra trọng lượng của chúng. Kết quả thu được như sau:

| Cân nặng | 8 | 19 | 23 | 24 | 25 | 26 | 28 |
|----------|---|----|----|----|----|----|----|
| Số quả   | 3 | 24 | 26 | 30 | 10 | 5  | 2  |

Tìm giá trị bất thường trong mẫu số liệu trên.

**Lời giải**

$$\Delta_O = 26 - 19 = 7.$$

**Câu 42:** Tiến hành một cuộc thăm dò về cân nặng ( $kg$ ) của mỗi học sinh nữ lớp 10A; biết lớp có 45 học sinh và 15 học sinh nam. Kết quả thu được ghi lại như sau:

|                   |    |    |    |    |    |    |    |    |
|-------------------|----|----|----|----|----|----|----|----|
| Cân nặng ( $kg$ ) | 36 | 38 | 40 | 43 | 45 | 48 | 50 | 56 |
| Tần số ( $n$ )    | 1  | 2  | 4  | 8  | 6  | 4  | 3  | 2  |

### Lời giải

$$\text{Vây } Q_2 = (43 + 45) : 2 = 44.$$

15 giá trị bên trái của  $Q_2$  là:

|    |    |    |    |
|----|----|----|----|
| 36 | 38 | 40 | 43 |
| 1  | 2  | 4  | 8  |

$$\Rightarrow Q_1 = 43.$$

15 giá trị bên phải của  $Q_2$  là:

|    |    |    |    |
|----|----|----|----|
| 45 | 48 | 50 | 56 |
| 6  | 4  | 3  | 2  |

$$\Rightarrow Q_3 = 48$$

Ta có:  $\Delta_o = 48 - 43 = 5$

$$Q_1 - 1,5 \cdot \Delta_0 = 43 - 1,5 \cdot 5 = 35,5.$$

$$Q_3 + 1,5 \cdot \Delta_Q = 48 + 1,5 \cdot 5 = 55,5.$$

Vậy mẫu có giá trị bất thường là 56.

**Câu 43:** Chiều cao (đơn vị: xăng-ti-mét) của các bạn tổ lữ lớp 10A lần lượt là:  
165 155 171 167 159 155 165 160 158

Đối với mẫu số liệu trên, hãy tìm:

- a. Số trung bình cộng;  
b. Trung vị:

d. Tứ phân vi.

155 155 158 159 160 165 165 167 171

d.  $M_o = 155$  và  $M_o = 165$

$$\bar{x} = \frac{16,4 + 17 + 20,2 + 23,7 + 27,3 + 28,8 + 28,9 + 28,2 + 27,2 + 24,6 + 21,4 + 18,2}{12} = 23,49$$

b. Nhiệt độ trung bình của tháng có giá trị thấp nhất là 16,4 độ C. Cao nhất là 28,9 độ C.

**Câu 46:** Bảng 3 cho biết tổng diện tích rừng từ năm 2008 đến năm 2019 ở nước ta.

| Năm                            | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|--------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Tổng diện tích rừng (triệu ha) | 13,1 | 13,2 | 13,4 | 13,5 | 13,9 | 14,0 | 13,8 | 14,1 | 14,4 | 14,4 | 14,5 | 14,6 |

- Diện tích rừng trung bình của nước ta từ năm 2008 đến năm 2019 là bao nhiêu?
- Từ năm 2008 đến năm 2019, diện tích rừng của năm có giá trị thấp nhất là bao nhiêu triệu héc-ta? Cao nhất là bao nhiêu triệu héc-ta?
- So với năm 2008, tỉ lệ tổng diện tích rừng của nước ta năm 2019 tăng lên được bao nhiêu phần trăm? Theo em, tỉ lệ tăng đó là cao hay thấp?
- Hãy tìm hiểu số liệu về tổng diện tích rừng của tỉnh em đang sống trong một số năm gần đây.

**Lời giải**

- Diện tích rừng trung bình của nước ta từ năm 2008 đến năm 2019 là:

$$\bar{x} = \frac{13,1 + 13,2 + 13,4 + 13,5 + 13,9 + 14 + 13,8 + 14,1 + 14,4 + 14,4 + 14,5 + 14,6}{12} \approx 13,9$$

- Từ năm 2008 đến năm 2019:

- Diện tích rừng của năm có giá trị thấp nhất là 13,1 triệu héc-ta.
- Diện tích rừng của năm có giá trị cao nhất là 14,6 triệu héc-ta.

- So với năm 2008, tỉ lệ tổng diện tích rừng của nước ta năm 2019 tăng lên:

$$\frac{14,6 - 13,1}{13,1} = 11,45$$

- HS tự thực hiện.

**Câu 47:** Bốn bạn Bình, Cường, Hoa, Kiên cùng thi vào trường phổ thông chất lượng cao Bình Minh. Kết quả thi được cho bởi bảng thống kê sau:

| Học sinh | Điểm Toán | Điểm Ngữ Văn | Điểm Tiếng Anh |
|----------|-----------|--------------|----------------|
| Bình     | 10        | 8            | 9              |
| Cường    | 6         | 7            | 5              |
| Hoa      | 10        | 10           | 4              |
| Kiên     | 9         | 5            | 10             |

Tính điểm trung bình kết quả thi 3 môn Toán, Ngữ Văn, Tiếng Anh của mỗi bạn và cho biết bạn nào trúng tuyển. Biết rằng, nếu muốn trúng tuyển, điểm trung bình các môn thi ở trên phải lớn hơn hoặc bằng 8 và không môn nào dưới 5 điểm.

**Lời giải**

$$\begin{aligned}\bar{x}_B &= \frac{10+8+9}{3} = 9, & \bar{x}_C &= \frac{6+7+5}{3} = 6, \\ \bar{x}_H &= \frac{10+10+4}{3} = 8, & \bar{x}_K &= \frac{9+5+10}{3} = 8.\end{aligned}$$

## Lời giải

102 109 115 118 121 127 130 132 138

Trung vị của dãy 102, 109, 115, 118 là:  $\frac{109+115}{2}=112$ .

Trung vị của dãy 127, 130, 132, 138 là:  $\frac{130+132}{2} = 131$ .

**Câu 52:** Một cửa hàng bán giày thống kê số đôi giày bán được trong Quý III năm 2020 như sau:

|                              |    |    |    |    |    |    |    |    |
|------------------------------|----|----|----|----|----|----|----|----|
| Cờ già                       | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 |
| Số đôi già bán được (Tần số) | 41 | 49 | 50 | 71 | 53 | 46 | 27 | 5  |

b) Cửa hàng đó nên nhập về nhiều hơn cỡ giày nào để bán tiếp?

### Lời giải

b) Cửa hàng nên nhập về nhiều hơn cỡ giày 40 để bán tiếp.

a) Tính số trung bình, trung vị của dãy số liệu trên.

b) Giải thích ý nghĩa các giá trị thu được.

### Lời giải

a) Có mẫu  $n=10$ . Số trung bình là:  $\bar{x} = \frac{-42+(-58)+\dots+(-54)}{10} = -48,8$ .

Sắp xếp các giá trị trên theo thứ tự không giảm:

-58 -56 -54 -53 -52 -50 -45 -42 -41 -37.

Vì  $n=10$  là số chẵn nên trung vị là trung bình công của hai giá trị ở các vị trí thứ 5 và thứ 6:

$$Me = \frac{-52 + (-50)}{2} = -51$$

b) Về trung bình, mức đóng góp của ngành du lịch cho GDP năm 2021 giảm khoảng 48,8% so với mức đóng góp của ngành du lịch cho GDP năm 2020.

Trung vị  $Me = -51\%$  tức là có 50% số khu vực (5 khu vực) có mức giảm dưới 51% và có 50% số khu vực (5 khu vực) có mức giảm trên 51%.

**Câu 54:** Tính các tứ phân vị cho dữ liệu về diện tích đất (đơn vị:  $km^2$ ) của 266 quốc gia và vùng lãnh thổ cho số liệu như sau:

$$Q_1 = 20574,1; \quad Q_2 = 194690; \quad Q_3 = 1249825.$$

(Theo World Bank)

a) Có bao nhiêu quốc gia, vùng lãnh thổ có diện tích đất lớn hơn  $194690 km^2$ ?

b) Diện tích đất của Việt Nam khoảng  $310070 km^2$  có thuộc nhóm 25% quốc gia và vùng lãnh thổ có diện tích đất lớn nhất không?

#### Lời giải

a) Vì  $Q_2 = 194690$  nên có 133 số quốc gia, vùng lãnh thổ (50%) có diện tích đất lớn hơn  $194690 km^2$ .

b) Do diện tích đất của Việt Nam nhỏ hơn  $Q_3 = 1249825$  nên Việt Nam không thuộc nhóm 25% quốc gia và vùng lãnh thổ có diện tích đất lớn nhất.

**Câu 55:** Tìm số trung bình, trung vị, một và tứ phân vị của mỗi mẫu số liệu sau đây:

a) Số điểm mà năm vận động viên bóng rổ ghi được trong một trận đấu:

9 8 15 8 20

b) Giá của một số loại giày (đơn vị nghìn đồng): 350 300 650 300 450 500 300 250

c) Số kênh được chiếu của một số hãng truyền hình cáp: 36 38 33 34 32 30 34 35

#### Lời giải

a) Số điểm mà năm vận động viên bóng rổ ghi được trong một trận đấu:

9 8 15 8 20

$$\text{Số trung bình: } \bar{X} = \frac{9+8+15+8+20}{5} = 12$$

Trung vị:

Sắp xếp theo thứ tự không giảm:

8 8 9 15 20

Ta có  $n=5$  là số lẻ nên trung vị là 9.

Mốt: Ta thấy số 8 là số có tần số cao nhất (xuất hiện 2 lần)

Tứ phân vị:

+ Tìm  $Q_2$

Ta có trung vị là  $9 \Rightarrow Q_2 = 9$ .

+ Tìm  $Q_1$

Nửa số liệu bên trái là: 8 8

Trung vị của mẫu này là  $\frac{8+8}{2} = 8 \Rightarrow Q_1 = 8$

+ Tìm  $Q_3$

Nửa số liệu bên phải là: 15 20

Trung vị của mẫu này là  $\frac{15+20}{2} = 17,5 \Rightarrow Q_3 = 17,5$

Vậy số trung bình là 12, trung vị là 9 và một là 8,  $Q_1 = 8, Q_3 = 17,5$

b) Giá của một số loại giày (đơn vị nghìn đồng):

350 300 650 300 450 500 300 250

Số trung bình:  $\bar{X} = \frac{350+300.3+650+450+500+250}{8} = 387,5$

Trung vị:

Sắp xếp theo thứ tự không giảm:

250 300 300 **300 350** 450 500 650

Ta có  $n=8$  là số chẵn nên trung vị là trung bình cộng của hai số chính giữa.

Hai số chính giữa là 300 và 350

$\Rightarrow$  Trung vị là  $\frac{300+350}{2} = 325$

Mốt: Ta thấy số 300 là số có tần số cao nhất (xuất hiện 3 lần)

Tứ phân vị:

+ Tìm  $Q_2$

Ta có trung vị là  $325 \Rightarrow Q_2 = 325$ .

+ Tìm  $Q_1$

Vì  $n$  chẵn nên nửa số liệu bên trái là: 250 **300 300** 300

Trung vị của mẫu này là  $\frac{300+300}{2} = 300 \Rightarrow Q_1 = 300$

+ Tìm  $Q_3$

Vì  $n$  chẵn nên nửa số liệu bên phải là: 350 **450 500** 650

Trung vị của mẫu này là  $\frac{450+500}{2} = 475 \Rightarrow Q_3 = 475$

Vậy số trung bình là 387,5, trung vị là 325 và một là 300,  $Q_1 = 300$ ,  $Q_3 = 475$

c) Số kênh được chiếu của một số hãng truyền hình cáp: 36 38 33 34 32 30 34

Số trung bình:  $\bar{X} = \frac{36+38+33+34.2+32+30+35}{8} = 34$

Trung vị:

Sắp xếp theo thứ tự không giảm: 30 32 33 **34 34** 35 36 38

Ta có  $n=8$  là số chẵn nên trung vị là trung bình công của hai số chính giữa.

Hai số chính giữa là 34 và 34

=> Trung vị là 34

Một: Ta thấy số 34 là số có tần số cao nhất (xuất hiện 2 lần)

Tứ phân vị:

 $+ \text{Tim } Q_2$ 

Ta có trung vị là  $34 \Rightarrow Q_2 = 34$ .

 $+ \text{Tim } Q_1$ 

Vì n chẵn nên nửa số liệu bên trái là: 30 **32 33 34**

Trung vị của mẫu này là  $\frac{32+33}{2} = 32,5 \Rightarrow Q_1 = 32,5$

 $+ \text{Tim } Q_3$ 

Vì n chẵn nên nửa số liệu bên phải là: 34 35 36 38

Trung vị của mẫu này là  $\frac{35+36}{2} = 35,5 \Rightarrow Q_3 = 35,5$

Vậy số trung bình là 34, trung vị là 34 và mode là 34,  $Q_1 = 32,5$ ,  $Q_3 = 35,5$

**Câu 56:** Hãy chọn số đặc trưng đo xu thế trung tâm của mỗi mẫu số liệu sau. Giải thích và tính giá trị của số đặc trưng đó.

a) Số mắt trắng đã biết của các hành tinh:

| Hành<br>tinh    | Thủy<br>tinh | Kim<br>tinh | Trái<br>Đất | Hoả<br>tinh | Mộc<br>tinh | Thổ<br>tinh | Thiên<br>Vương<br>tinh | Hải<br>Vương<br>tinh |
|-----------------|--------------|-------------|-------------|-------------|-------------|-------------|------------------------|----------------------|
| Số mặt<br>trắng | 0            | 0           | 1           | 2           | 63          | 34          | 27                     | 13                   |

b) Số đường chuyền thành công trong một trận đấu của một số cầu thủ bóng đá:

32 24 20 14 23.

c) Chỉ số IQ của một nhóm học sinh:

80 10 83 10 10 94 11 10 10 100  
2 3 8 0 6 4

d) Các sai số trong một phép đo: 10 15 18 15 14 13 42 15 12 14 42.

### Lời giải

a) Sắp xếp lại số liệu:

0 0 1 2 13 27 34 63

Trung vị là  $\frac{(2+13)}{2} = 7,5$ .

Ta không chọn số trung bình vì số trung bình là 17,5 chênh lệch với 63 lớn. Một cũng thế.

b) Các số liệu bài cho không chênh lệch quá lớn với số trung bình nên ta chọn số trung bình.

Số đường truyền trung bình là:  $\frac{32+24+20+14+23}{5} = 22,6$

c) Các số liệu bài cho không chênh lệch quá lớn với số trung bình nên ta chọn số trung bình.

IQ trung bình là  $\frac{80+102+83+103+108+94+110+106+104+100}{10} = 99$

d) Ta thấy có hai giá trị 42 chênh lệch lớn với các số còn lại nên ta chọn Một để đo xu thế trung tâm.

Mốt là 15 (tần số là 3).

**Câu 57:** Số lượng học sinh giỏi Quốc gia năm học 2018-2019 của 10 trường Trung học phổ thông được cho như sau:

0 0 4 0 0 0 10 0 6 0

a) Tìm số trung bình, mốt, các tứ phân vị của mẫu số liệu trên.

b) Giải thích tại sao tứ phân vị thứ nhất và trung vị trùng nhau.

### Lời giải

a) Sắp xếp theo thứ tự không giảm:

0 0 0 0 0 0 0 4 6

Số trung bình:  $\bar{X} = \frac{0.7+4+6+10}{10} = 2$

Trung vị:  $Q_2 = 0$

Tứ phân vị:

+ Nửa bên trái của  $Q_2$ : 0 0 0 0 0

$$\Rightarrow Q_1 = 0$$

+ Nửa bên phải của  $Q_2$ : 0 0 4 6 10

$$\Rightarrow Q_3 = 4$$

b) Tứ phân vị thứ nhất và trung vị trùng nhau vì mật độ của mẫu số liệu tập trung hết ở nửa trái của trung vị, mẫu số liệu bên trái có số liệu bằng 0 hết.

**Câu 58:** Bảng sau đây cho biết số chỗ ngồi của một số sân vận động được sử dụng trong Giải Bóng đá Vô địch Quốc gia Việt Nam năm 2018 (số liệu gần đúng).

| Sân vận động | phả | trường | Đầy | Hóa | ình |
|--------------|-----|--------|-----|-----|-----|
| Số ngồi      |     |        |     |     |     |

(Theo vov, vn)

Các giá trị số trung bình, trung vị, một bị ảnh hưởng thế nào nếu bỏ đi số liệu chỗ ngồi của Sân vận động Quốc gia Mỹ Đình?

**Lời giải**

Sắp xếp lại mẫu số liệu:

20120 20120 21315 23405 37546

Số trung bình:

$$\frac{20120.2 + 21315 + 23405 + 37546}{5} = 24501,2$$

Trung vị: 21315

Mốt: 20120

Nếu bỏ đi số liệu chỗ ngồi của Sân vận động Quốc gia Mỹ Đình thì số trung bình giảm, trung vị giảm và Mốt thì vẫn giữ nguyên.

Cụ thể: số trung bình là 21 240; trung vị là 20 717,5 và Mốt vẫn là 20120.

**Câu 59:** Tổng số ca mắc Covid-19 tính đến ngày 26-8-2021 tại Thành phố Hồ Chí Minh và một số tỉnh lân cận được thống kê như sau:

190174 81182 19728 19048 8155 6103 5807  
4544 3760 3297 2541 2000 1934 1602 1195.

(Theo Bộ Y tế?)

a) Tính số trung bình và trung vị cho dãy số liệu trên.

b) Giải thích tại sao số trung bình và trung vị lại khác nhau nhiều?

**Lời giải**

a) Số trung bình:  $\bar{x} = \frac{190174 + \dots + 1195}{15} \approx 23404,67.$

Sắp xếp dãy số liệu theo thứ tự không giảm:

1195 1602 1934 2000 2541 3297 3760 4544  
5807 6103 8155 19048 19728 81182 190174.

Vì  $n=15$  nên vị trí chính giữa là vị trí số 8. Do đó, trung vị là 4544.

b) Số trung bình lớn hơn trung vị nhiều là do trong dãy số có một giá trị rất lớn là 190174. Trung vị không bị ảnh hưởng bởi giá trị "bất thường" này.

**Câu 60:** Lan thống kê số anh, chị, em ruột của các bạn trong lớp thu được bảng số liệu sau:

|                      |   |    |   |   |
|----------------------|---|----|---|---|
| Số anh, chị, em ruột | 0 | 1  | 2 | 3 |
| Số bạn               | 4 | 25 | 5 | 1 |

Xác định một cho mẫu số liệu trên và giải thích ý nghĩa.

### Lời giải

Mốt = 1; Ý nghĩa: Số gia đình có hai con là nhiều nhất.

**Câu 61:** Thống kê GDP năm 2020 (đơn vị: tỉ đô la Mỹ) của 10 nước tại khu vực Đông Nam Á được kết quả như sau:

|         |             |           |          |          |
|---------|-------------|-----------|----------|----------|
| Brunei  | Campuchia   | Indonesia | Lào      | Malaysia |
| 12,02   | 25,95       | 1059,64   | 19,08    | 338,28   |
| Myanmar | Philippines | Singapore | Thái Lan | Việt Nam |
| 81,26   | 362,24      | 339,98    | 501,89   | 340,82   |

(Theo statista.com)

a) Tìm các tứ phân vị cho dãy số liệu trên.

b) Giải thích ý nghĩa của các tứ phân vị này. Việt Nam có thuộc nhóm 25% quốc gia có GDP năm 2020 cao nhất trong khu vực Đông Nam Á không?

### Lời giải

a) Sắp xếp dãy số liệu theo thứ tự không giảm:

12,02 19,08 25,95 81,26 338,28  
339,98 340,82 362,24 501,89 1059,64.

Vì  $n=10$  nên trung vị là trung bình cộng của 2 giá trị chính giữa (vị trí 5 và 6):

$$Q_2 = \frac{338,28 + 339,98}{2} = 339,13.$$

Nửa dữ liệu bên trái  $Q_2$  là: 12,02 19,08 25,95 81,26 338,28; gồm 5 số do đó số chính giữa là số ở vị trí thứ 3 nên  $Q_1 = 25,95$ .

Nửa dữ liệu bên phải  $Q_2$  là: 339,98 340,82 362,24 501,89 1059,64; gồm 5 số do đó trung vị là số chính giữa ở vị trí thứ 3 nên  $Q_3 = 362,24$ .

b) GDP của Việt Nam năm 2020 là 340,82 tỉ đô la Mỹ (nhỏ hơn  $Q_3$ ) nên Việt Nam không thuộc nhóm 25% quốc gia trong khu vực Đông Nam Á có GDP cao nhất.

**Câu 62:** Diện tích của các tỉnh đồng bằng sông Cửu Long năm 2022 (đơn vị: nghìn  $km^2$ ) là:

- Tính số trung bình, trung vị cho dãy số liệu trên.
- Giải thích ý nghĩa của mỗi số thu được ở câu a.

**Lời giải**

a) Số trung bình  $\bar{X} \approx 3,14$ . Trung vị  $Me = 2,67$ .

b) Diện tích trung bình của các tỉnh đồng bằng sông Cửu Long là 3,14 nghìn  $km^2$ .

Trung vị  $Me = 2,67$  (nghìn  $km^2$ ) nghĩa là số tỉnh có diện tích nhỏ hơn 2,67 nghìn  $km^2$  bằng số tỉnh có diện tích lớn hơn 2,67 nghìn  $km^2$ .

**Câu 63:** Trong một đợt khảo sát về tốc độ viết của học sinh lớp 3, người ta cho hai nhóm học sinh chép một đoạn văn trong 15 phút. Bảng dưới đây thống kê số chữ mỗi bạn viết được.

|        |    |    |    |    |    |    |    |    |
|--------|----|----|----|----|----|----|----|----|
| Nhóm 1 | 72 | 79 | 77 | 75 | 74 | 77 | 71 |    |
| Nhóm 2 | 70 | 65 | 68 | 90 | 73 | 78 | 72 | 84 |

- Có bao nhiêu học sinh tham gia đợt khảo sát?
- Sử dụng số trung bình để so sánh tốc độ viết của học sinh hai nhóm.
- Sử dụng trung vị để so sánh tốc độ viết của học sinh hai nhóm.

**Lời giải**

a) Số học sinh tham gia khảo sát là:  $7+8=15$ .

b) Số chữ trung bình mỗi học sinh nhóm 1 viết được là:

$$\frac{1}{7}(72+79+77+75+74+77+71)=75.$$

Số chữ trung bình mỗi học sinh nhóm 2 viết được là:

$$\frac{1}{8}(70+65+68+90+73+78+72+84)=75.$$

Vậy nếu so sánh theo số trung bình thì tốc độ viết của học sinh hai nhóm là bằng nhau.

c) Sắp xếp số chữ học sinh nhóm 1 viết được theo thứ tự không giảm, ta được dãy:

71 ; 72 ; 74 ; 75 ; 77 ; 77 ; 79 .

Trung vị của nhóm 1 là 75.

Sắp xếp số chữ học sinh nhóm 2 viết được theo thứ tự không giảm, ta được dãy:

65 ; 68 ; 70 ; 72 ; 73 ; 78 ; 84 ; 90 .

Trung vị của nhóm 2 là  $(72+73):2=72,5$ .

Vậy nếu so sánh theo trung vị thì tốc độ viết của học sinh nhóm 1 cao hơn của học sinh nhóm 2.

**Câu 64:** Khối lượng cơ thể lúc trưởng thành của 10 con chim được ghi lại ở bảng sau (đơn vị: gam).

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 155 | 165 | 150 | 155 | 165 | 170 | 165 | 150 | 155 | 160 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Hãy tìm các tứ phân vị và một của mẫu số liệu trên.



- c) Sử dụng trung vị để so sánh số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố ở hai khu vực.
- d) Hãy giải thích tại sao lại có sự khác biệt khi so sánh bằng số trung bình và trung vị.
- e) Hãy tìm một của hai khu vực.

### Lời giải

a) Khu vực Đông Nam Bộ có 6 tỉnh/thành phố. Khu vực Đồng bằng sông Cửu Long có 13 tỉnh/thành phố.

b) Trung bình số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đông Nam Bộ là  $(10+8+8+9+6+24):6 \approx 10,83$ .

Trung bình số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đồng bằng sông Cửu Long là  $(14+10+8+8+7+10+9+13+9+7+10+6+8):13 \approx 9,15$ .

Vậy nếu so sánh theo số trung bình thì số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đông Nam Bộ nhiều hơn của các tỉnh Đồng bằng sông Cửu Long.

c) Trung vị số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đông Nam Bộ là 8,5 và của các tỉnh Đồng bằng sông Cửu Long là 9.

Vậy nếu so sánh theo trung vị thì số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đông Nam Bộ ít hơn của các tỉnh Đồng bằng sông Cửu Long.

d) Sự khác biệt khi so sánh bằng số trung bình và bằng trung vị là do có 1 tỉnh/ thành phố khu vực Đông Nam Bộ có quá nhiều đơn vị hành chính cấp quận/ huyện/thị xã so với các tỉnh/thành phố khác.

e) Một của số liệu khu vực Đông Nam Bộ là 8. Một của số liệu khu vực Đồng bằng sông Cửu Long là 8 và 10.

**Câu 67:** An lấy ra ngẫu nhiên 3 quả bóng từ một hộp có chứa nhiều bóng xanh và bóng đỏ. An đếm xem có bao nhiêu bóng đỏ trong 3 bóng lấy ra rồi trả bóng lại hộp. An lặp lại phép thử trên 100 lần và ghi lại kết quả ở bảng sau:

|            |    |    |    |    |
|------------|----|----|----|----|
| Số bóng đỏ | 0  | 1  | 2  | 3  |
| Số lần     | 10 | 30 | 40 | 20 |

Hãy tìm số trung bình, tứ phân vị và một của bảng kết quả trên.

### Lời giải

+) Số trung bình:  $\bar{x} = \frac{0.10 + 1.30 + 2.40 + 3.20}{100} = 1,7$

+) Tứ phân vị:  $Q_1, Q_2, Q_3$

Bước 1: Sắp xếp mẫu số liệu theo thứ tự không giảm,

$$\underbrace{0, \dots, 0}_{10}, \underbrace{1, \dots, 1}_{30}, \underbrace{2, \dots, 2}_{40}, \underbrace{3, \dots, 3}_{20}.$$

Bước 2: Vì  $n=100$ , là số chẵn nên  $Q_2 = \frac{1}{2}(2+2) = 2$

$Q_1$  là trung vị của nửa số liệu:  $\underbrace{0, \dots, 0}_{10}, \underbrace{1, \dots, 1}_{30}, \underbrace{2, \dots, 2}_{10}$ . Do đó  $Q_1 = \frac{1}{2}(1+1) = 1$

$Q_3$  là trung vị của nửa số liệu  $\underbrace{2, \dots, 2}_{30}, \underbrace{3, \dots, 3}_{20}$ . Do đó  $Q_3 = \frac{1}{2}(2+2) = 2$

+) Một  $M_o = 2$

**Câu 68:** Trong một cuộc thi nghề, người ta ghi lại thời gian hoàn thành một sản phẩm của một số thí nghiệm ở bảng sau:

|                             |   |   |   |   |    |
|-----------------------------|---|---|---|---|----|
| Thời gian<br>(đơn vị: phút) | 5 | 6 | 7 | 8 | 35 |
| Số thí sinh                 | 1 | 3 | 5 | 2 | 1  |

a) Hãy tìm số trung bình, tứ phân vị và một của thời gian thi nghề của các thí sinh trên.

b) Năm ngoái, thời gian thi của các thí sinh có số trung bình và trung vị đều bằng 7. Bạn hãy so sánh thời gian thi nói chung của các thí sinh trong hai năm.

### Lời giải

a) Hãy tìm số trung bình, tứ phân vị và một của thời gian thi nghề của các thí sinh trên.

+) Số trung bình:  $\bar{x} = \frac{1.5 + 3.6 + 5.7 + 2.8 + 1.35}{1 + 3 + 5 + 2 + 1} = 9,08$

+) Tứ phân vị:  $Q_1, Q_2, Q_3$

Bước 1: Sắp xếp mẫu số liệu theo thứ tự không giảm,

5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 35

Bước 2: Vì  $n = 12$ , là số chẵn nên  $Q_2 = \frac{1}{2}(7+7) = 7$

$Q_1$  là trung vị của nửa số liệu: 5, 6, 6, 6, 7, 7 Do đó  $Q_1 = \frac{1}{2}(6+6) = 6$

$Q_3$  là trung vị của nửa số liệu 7, 7, 7, 8, 8, 35 Do đó  $Q_3 = \frac{1}{2}(7+8) = 7,5$

+) Một  $M_o = 7$

b) Năm ngoái, thời gian thi của các thí sinh có số trung bình và trung vị đều bằng 7. Bạn hãy so sánh thời gian thi nói chung của các thí sinh trong hai năm.

+) Nếu so sánh số trung bình:  $9,08 > 7$  do đó thời gian thi nói chung của các thí sinh trong năm nay là lớn hơn so với năm trước.

+) Nếu so sánh trung vị: Trung vị của hai năm đều bằng 7 do đó thời gian thi nói chung của các thí sinh trong hai năm là như nhau.

Do có 1 thí sinh có thời gian thi lớn hơn hẳn so với các thí sinh khác  $\Rightarrow$  nên so sánh theo trung vị.

|          |   |   |   |   |   |   |   |   |    |   |
|----------|---|---|---|---|---|---|---|---|----|---|
| Bác Dũng | 2 | 7 | 3 | 6 | 1 | 4 | 1 | 4 | 5  | 1 |
| Bác Thu  | 1 | 3 | 1 | 2 | 3 | 4 | 1 | 2 | 20 | 2 |

- ## Lời giải

b) Do  $3,9 > 3,4$  nên theo số trung bình thì bác Thu có nhiều cuộc điện thoại hơn.

c) Do  $3,5 > 2$  nên theo số trung vị thì bác Dũng có nhiều cuộc điện thoại hơn.

d) Vì trong mẫu số liệu có một ngày bác Thu có tới 20 cuộc điện thoại, lớn hơn nhiều so với các ngày khác, do đó ta nên so sánh theo số trung vị.

**Câu 70:** Tổng số điểm mà các thành viên đội tuyển Olympic Toán quốc tế (IMO) của Việt Nam đạt được trong 20 kì thi được cho ở bảng sau:

| Năm  | Tổng điểm | Năm  | Tổng điểm | Năm  | Tổng điểm | Năm  | Tổng điểm |
|------|-----------|------|-----------|------|-----------|------|-----------|
| 2020 | 150       | 2015 | 151       | 2010 | 133       | 2005 | 143       |
| 2019 | 177       | 2014 | 157       | 2009 | 161       | 2004 | 196       |
| 2018 | 148       | 2013 | 180       | 2008 | 159       | 2003 | 172       |
| 2017 | 155       | 2012 | 148       | 2007 | 168       | 2002 | 166       |
| 2016 | 151       | 2011 | 113       | 2006 | 131       | 2001 | 139       |

(Nguồn: <https://imo-offial.org>)

Có ý kiến cho rằng điểm thi của đội tuyển giai đoạn 2001 - 2010 cao hơn giai đoạn 2011 - 2020. Hãy sử dụng số trung bình và trung vị để kiểm nghiệm xem ý kiến trên có đúng không.

### Lời giải

+) Giai đoạn 2001 - 2010

Số trung bình

$$\bar{x} = \frac{139+166+172+196+143+131+168+159+161+133}{10} = 156,8$$

Sắp xếp số liệu theo thứ tự không giảm, ta được: 131,133,139,143,159,161,166,168,172,196

Do  $n = 10$ , là số chẵn nên trung vị là:  $M_e = \frac{1}{2}(159+161) = 160$

+) Giai đoạn 2011–2020 Số trung bình

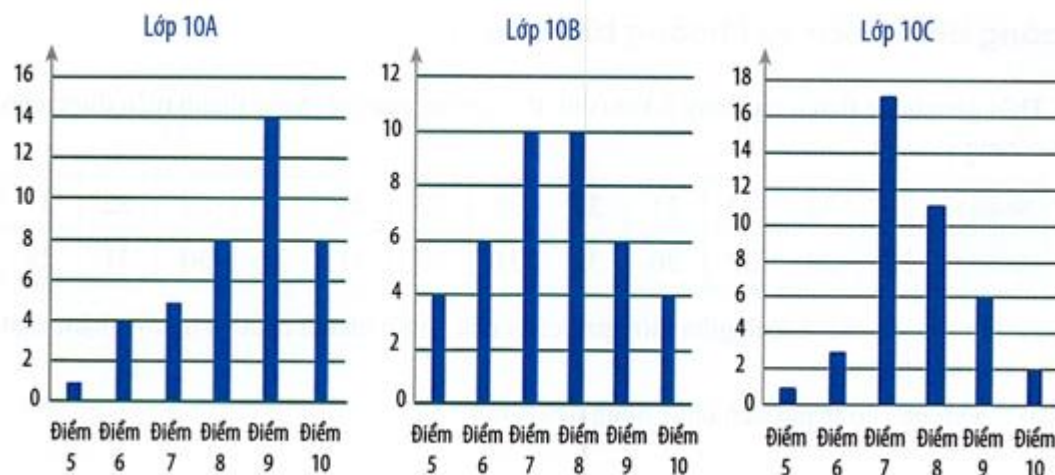
$$\bar{x} = \frac{150+177+148+155+151+151+157+180+148+113}{10} = 153$$

Sắp xếp số liệu theo thứ tự không giảm, ta được: 113,148,148,150,151,151,155,157,177,180 Do  $n = 10$ , là số chẵn nên

trung vị là:  $M_e = \frac{1}{2}(151+151) = 151$

+) So sánh theo số trung bình hay số trung vị ta đều thấy điểm thi của đội tuyển giai đoạn 2001 - 2010 cao hơn giai đoạn 2011 - 2020. Vậy ý kiến trên là đúng.

**Câu 71:** Kết quả bài kiểm tra giữa kì cả các bạn học sinh lớp 10A, 10B, 10C được thống kê ở các biểu đồ dưới đây.



- a) Hãy lập thống kê số lượng học sinh theo điểm số ở mỗi lớp.  
 b) Hãy so sánh điểm số của học sinh các lớp đó theo số trung bình, trung vị và môđ.

### Lời giải

a)

|         |       |   |   |    |    |    |    |
|---------|-------|---|---|----|----|----|----|
| Lớp 10A | Điểm  | 5 | 6 | 7  | 8  | 9  | 10 |
|         | Số HS | 1 | 4 | 5  | 8  | 14 | 8  |
| Lớp 10B | Điểm  | 5 | 6 | 7  | 8  | 9  | 10 |
|         | Số HS | 4 | 6 | 10 | 10 | 6  | 4  |
| Lớp 10C | Điểm  | 5 | 6 | 7  | 8  | 9  | 10 |
|         | Số HS | 1 | 3 | 17 | 11 | 6  | 2  |

b)

+) Lớp 10A

$$\text{Số trung bình } \bar{x} = \frac{5 \cdot 1 + 6 \cdot 4 + 7 \cdot 5 + 8 \cdot 8 + 9 \cdot 14 + 10 \cdot 8}{1 + 4 + 5 + 8 + 14 + 8} = 8,35$$

Sắp xếp số liệu theo thứ tự không giảm, ta được:

$$5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, \dots, 8, 9, \dots, 9, 10, \dots, 10$$

Do  $n = 40$ , là số chẵn nên trung vị là:  $M_e = \frac{1}{2}(9 + 9) = 9$  Môđ  $M_e = 9$

+) Lớp 10B

$$\text{Số trung bình } \bar{x} = \frac{5 \cdot 4 + 6 \cdot 6 + 7 \cdot 10 + 8 \cdot 10 + 9 \cdot 6 + 10 \cdot 4}{4 + 6 + 10 + 10 + 6 + 4} = 7,5$$

Sắp xếp số liệu theo thứ tự không giảm, ta được:

$$5, 5, 5, 5, 6, \dots, 6, 7, \dots, 7, 8, \dots, 8, 9, \dots, 9, 10, 10, 10, 10$$

Do  $n = 40$ , là số chẵn nên trung vị là:  $M_e = \frac{1}{2}(7 + 8) = 7,5$  Môđ  $M_e = 7; M_e = 8$ .

+) Lớp 10C

$$\text{Số trung bình } \bar{x} = \frac{5.1 + 6.3 + 7.17 + 8.11 + 9.6 + 10.2}{1 + 3 + 17 + 11 + 6 + 2} = 7,6$$

Sắp xếp số liệu theo thứ tự không giảm, ta được: 5, 6, 6, 6,  $\underbrace{7, \dots, 7}_{17}$ ,  $\underbrace{8, \dots, 8}_{11}$ ,  $\underbrace{9, \dots, 9}_6$ , 10, 10

Do  $n = 40$ , là số chẵn nên trung vị là:  $M_e = \frac{1}{2}(7 + 7) = 7$

Mốt  $M_o = 7$

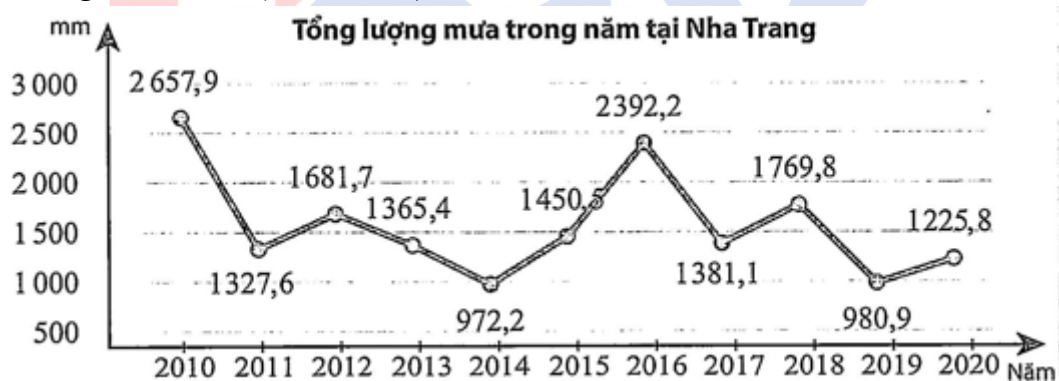
+) So sánh:

Số trung bình:  $8,35 > 7,6 > 7,5 \Rightarrow$  Điểm số của HS các lớp theo thứ tự giảm dần là 10A, 10C, 10B.

Số trung vị:  $9 > 7,5 > 7 \Rightarrow$  Điểm số của HS các lớp theo thứ tự giảm dần là 10A, 10B, 10C.

Mốt: Lớp 10A có 14 điểm 9, Lớp 10B có 10 điểm 7 và 10 điểm 8, Lớp 10C có 17 điểm 7. Do đó so sánh theo một thì điểm số các lớp giảm dần theo thứ tự là: 10A, 10B, 10C.

**Câu 72:** Tổng lượng mưa trong năm tại một trạm quan trắc đặt tại Nha Trang từ năm 2010 đến 2020 được thể hiện trong biểu đồ sau (đơn vị: mm).



a) Hãy tính lượng mưa trung bình tại trạm quan trắc trên từ năm 2010 đến 2020.

b) Hãy tìm các tứ phân vị của mẫu số liệu đó.

**Lời giải**

a)  $\bar{x} = 1564,1$ .

b)  $Q_1 = 1225,8; Q_2 = 1381,1; Q_3 = 1769,8$ .

**Câu 73:** Số huy chương vàng và bạc trong các giải thể thao quốc tế mà đoàn thể thao Việt Nam đạt được tại các giải đấu ở châu Á trong các năm từ năm 2010 đến 2019 được thống kê ở bảng sau:

| Năm             | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|-----------------|------|------|------|------|------|------|------|------|------|------|
| Huy chương vàng | 39   | 43   | 115  | 52   | 56   | 62   | 130  | 82   | 74   | 120  |
| Huy chương bạc  | 61   | 63   | 121  | 47   | 58   | 73   | 134  | 87   | 74   | 105  |

(Nguồn: Tổng cục Thống kê)

- a) Tìm số trung bình và trung vị huy chương vàng và huy chương bạc mà đoàn thể thao Việt Nam đạt được trong 10 năm trên.
- b) Hãy so sánh số huy chương vàng đoàn thể thao Việt Nam đạt được trong giai đoạn 2010 - 2014 với giai đoạn 2015 - 2019.

### Lời giải

a) Trung bình số huy chương vàng và trung bình số huy chương bạc đạt được trong 10 năm là 77,3 và 82,3.

Trung vị số huy chương vàng và trung vị số huy chương bạc đạt được trong 10 năm là 68 và 73,5.

b) Trung bình và trung vị số huy chương vàng đạt được trong giai đoạn 2010-2014 là 61 và 52.

Trung bình và trung vị số huy chương vàng đạt được trong giai đoạn 2015 - 2019 là 93,6 và 82.

Vậy nếu so sánh theo số trung bình và số trung vị thì Việt Nam đều giành được nhiều huy chương vàng hơn trong giai đoạn 2015 - 2019 so với giai đoạn 2010-2014.

**Câu 74:** Bảng sau ghi lại độ tuổi của hai nhóm vận động viên tham gia một cuộc thi.

|        |    |    |    |    |    |    |    |    |    |    |    |    |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|
| Nhóm 1 | 20 | 32 | 27 | 31 | 32 | 30 | 32 | 29 | 17 | 29 | 22 | 31 |
| Nhóm 2 | 22 | 29 | 22 | 30 | 22 | 31 | 29 | 21 | 32 | 20 | 31 | 29 |

a) Hãy so sánh độ tuổi của hai nhóm vận động viên theo số trung bình và trung vị.

b) Tìm tứ phân vị của độ tuổi vận động viên cả hai nhóm gộp lại.

### Lời giải

a) Nhóm 1:  $\bar{x} = 27,67; M_e = 29,5$ .

Nhóm 2:  $\bar{x} = 26,5; M_e = 29$ .

Vậy độ tuổi của các vận động viên nhóm 1 cao hơn nhóm 2.

b)  $Q_1 = 22; Q_2 = 29; Q_3 = 31$ .

**Câu 75:** Minh và Thuỷ ghi lại số thư điện tử mà mỗi người nhận được mỗi ngày trong 10 ngày được lựa chọn ngẫu nhiên từ tháng 01/ 2021 ở bảng sau:

|      |   |   |   |   |   |   |   |   |    |   |
|------|---|---|---|---|---|---|---|---|----|---|
| Minh | 6 | 7 | 3 | 6 | 1 | 4 | 1 | 4 | 5  | 1 |
| Thuỷ | 2 | 3 | 1 | 2 | 3 | 4 | 1 | 2 | 20 | 2 |

a) Hãy tìm số trung bình, trung vị và mốt của số thư điện tử mà mỗi bạn nhận được theo số liệu trên.

b) Nếu so sánh theo số trung bình thì ai nhận được nhiều thư điện tử hơn?

c) Nếu so sánh theo trung vị thì ai nhận được nhiều thư điện tử hơn?

d) Nên dùng số trung bình hay số trung vị để so sánh xem ai nhận được nhiều thư điện tử hơn mỗi ngày?

### Lời giải

a) Minh:  $\bar{x} = 3,8; M_e = 4; M_o = 1$ . Thuỷ:  $\bar{x} = 4; M_e = 2; M_o = 2$ .

- b) Nếu so sánh theo số trung bình thì Thuý nhận được nhiều thư điện tử hơn Minh.  
c) Nếu so sánh theo số trung vị thì Minh nhận được nhiều thư điện tử hơn Thuý.  
d) Nên dùng số trung vị để so sánh.

**Câu 76:** Bạn Út ghi lại khối lượng của một số quả xoài Keo và xoài Thanh Ca ở bảng sau (đơn vị: gam).

|               |     |     |     |     |     |     |     |     |     |     |     |     |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Xoài Keo      | 370 | 320 | 350 | 290 | 300 | 350 | 310 | 330 | 340 | 370 | 390 |     |
| Xoài Thanh Ca | 350 | 310 | 410 | 390 | 380 | 370 | 320 | 350 | 330 | 340 | 370 | 400 |

- a) Sử dụng số trung bình, hãy so sánh khối lượng của hai loại xoài.  
b) Sử dụng trung vị, hãy so sánh khối lượng của hai loại xoài.  
c) Hãy tính các tứ phân vị của hai mẫu số liệu trên.  
d) Nếu bạn Út mua 5 kg xoài Keo thì sẽ được khoảng bao nhiêu quả?



Nếu bạn Út mua 5 kg xoài Thanh Ca thì sẽ được khoảng bao nhiêu quả?

#### Lời giải

- a) Khối lượng trung bình xoài Keo là  $338,18g$ . Khối lượng trung bình xoài Thanh Ca là  $360g$ . Vậy nếu so sánh theo số trung bình thì khối lượng xoài Thanh Ca cao hơn khối lượng xoài Keo.  
b) Trung vị của khối lượng xoài Keo là  $340g$ . Trung vị khối lượng xoài Thanh Ca là  $360g$ . Vậy nếu so sánh theo trung vị thì khối lượng xoài Thanh Ca cao hơn khối lượng xoài Keo.  
c) Xoài Keo:  $Q_1 = 310; Q_2 = 340; Q_3 = 370$ .

Xoài Thanh Ca:  $Q_1 = 335; Q_2 = 360; Q_3 = 385$ .

d) Do  $5000 : 338,18 \approx 14,79$  nên nếu bạn Út mua 5 kg xoài Keo thì sẽ được khoảng 14 đến 15 quả.

Do  $5000 : 360 \approx 13,89$  nên nếu bạn Út mua 5kg xoài Thanh Ca thì sẽ được khoảng 13 đến 14 quả.

**Câu 77:** Số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực Đồng bằng sông Hồng và khu vực Trung du và miền núi phía Bắc vào năm 2019 được cho như sau:

Đồng bằng sông Hồng: 30; 7; 7; 10; 10; 15; 9; 7; 5; 9; 6.

Trung du và miền núi phía Bắc: 10; 12; 7; 6; 8; 8; 7; 10; 9; 12; 9; 7; 11; 10.

(Nguồn: Tổng cục Thống kê)

- a) Mỗi khu vực nêu trên có bao nhiêu tỉnh/thành phố?

- b) Sử dụng số trung bình, hãy so sánh số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố ở hai khu vực.
- c) Sử dụng trung vị, hãy so sánh số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố ở hai khu vực.
- d) Hãy giải thích tại sao lại có sự khác biệt khi so sánh bằng số trung bình và trung vị.
- e) Hãy tìm tứ phân vị và một của hai khu vực.

### Lời giải

a) Khu vực Đồng bằng sông Hồng ( OBSH) có 11 tỉnh/thành phố; Khu vực Trung du và miền núi phía Bắc (TDMNPB) có 14 tỉnh/thành phố.

b) Trung bình số đơn vị hành chính cấp quận/huyện/thị xã của khu vực ĐBSH là 10,45 và của khu vực TDMNPB là 9. Do đó theo số trung bình thì các tỉnh/thành phố khu vực ĐBSH có nhiều đơn vị hành chính cấp quận/huyện/thị xã hơn khu vực TDMNPB.

c) Trung vị số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực ĐBSH và TDMNPB đều là 9.

Vậy nếu so sánh theo trung vị thì số đơn vị hành chính cấp quận/huyện/thị xã của các tỉnh/thành phố khu vực ĐBSH và TDMNPB là bằng nhau.

d) Có sự khác biệt khi so sánh bằng số trung bình và số trung vị là do có một tỉnh/ thành phố khu vực ĐBSH có quá nhiều đơn vị hành chính cấp quận/huyện/thị xã so với các tỉnh/thành phố khác.

e) ĐBSH  $Q_1 = 7; Q_2 = 9; Q_3 = 10; M_o = 7$ .

TDMNPB  $Q_1 = 7; Q_2 = 9; Q_3 = 10; M_o \in \{7; 10\}$ .

WE COACH. WE INSPIRE