

## A. Lý thuyết

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

#### a) Khoảng biến thiên

Khoảng biến thiên, kí hiệu là  $R$ , là hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất trong mẫu số liệu.

**Ý nghĩa:** Khoảng biến thiên dùng để đo độ phân tán của mẫu số liệu. Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

**Ví dụ:** Hai xạ thủ A và B cùng bắn 10 phát đạn, kết quả được ghi lại như bảng sau:

A	7	9	8	9	9	10	8	7	9	10
B	8	9	10	7	6	9	10	7	10	10

a) Điểm số trung bình của hai xạ thủ A và B có như nhau không?

b) Tính các khoảng biến thiên của hai mẫu số liệu. Căn cứ trên chỉ số này, xạ thủ nào bắn đều hơn?

#### Hướng dẫn giải

a) Điểm số trung bình của xạ thủ A là:  $\frac{7.2 + 8.2 + 9.4 + 10.2}{10} = 8,6$  (điểm).

Điểm số trung bình của xạ thủ B là:  $\frac{6 + 7.2 + 8 + 9.2 + 10.4}{10} = 8,6$  (điểm)

Vậy điểm kiểm tra trung bình của hai xạ thủ A và B đều bằng 8,6.

b) Đối với xạ thủ A: Điểm số thấp nhất và cao nhất tương ứng là 7 và 10. Do đó khoảng biến thiên là  $R_A = 10 - 7 = 3$ .

Đối với xạ thủ B: Điểm số thấp nhất và cao nhất tương ứng là 6 và 10. Do đó khoảng biến thiên là  $R_B = 10 - 6 = 4$ .

Do  $R_B > R_A$  nên ta nói xạ thủ A bắn đều hơn xạ thủ B.

**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.

### b) Khoảng tứ phân vị

Khoảng tứ phân vị, kí hiệu là  $\Delta_Q$ , là hiệu số giữa tứ phân vị thứ ba và tứ phân vị thứ nhất, tức là:

$$\Delta_Q = Q_3 - Q_1.$$

Về bản chất, khoảng tứ phân vị là khoảng biến thiên của 50% số liệu chính giữa của mẫu số liệu đã sắp xếp.

**Ý nghĩa:** Khoảng tứ phân vị cũng là một số đo độ phân tán của mẫu số liệu. Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

**Chú ý:** Một số tài liệu gọi khoảng biến thiên là *biên độ* và khoảng tứ phân vị là *độ trải giữa*.

**Ví dụ:** Mẫu số liệu sau cho biết số ghế trống tại một rạp xiếc trong 9 ngày:

0    7    3    9    20    11    5    16    19

Tìm khoảng tứ phân vị cho mẫu số liệu trên.

### Hướng dẫn giải

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

0    3    5    7    9    11    16    19    20

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

Nửa số liệu bên trái là 0; 3; 5; 7 gồm 4 giá trị, hai phần tử chính giữa là 3; 5.

Do đó,  $Q_1 = (3 + 5) : 2 = 4$ .

Nửa số liệu bên phải là 11; 16; 19; 20 gồm 4 giá trị, hai phần tử chính giữa là 16; 19.

Do đó,  $Q_3 = (16 + 19) : 2 = 17,5$ .

Vậy khoảng tứ phân vị cho mẫu số liệu là  $\Delta_Q = Q_3 - Q_1 = 17,5 - 4 = 13,5$ .

## 2. Phương sai và độ lệch chuẩn

Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu (bỏ qua thông tin của tất cả các giá trị khác). Khoảng tứ phân vị chỉ sử dụng thông tin của 50% số liệu chính giữa. Có một vài số đặc trưng khác đo độ phân tán sử dụng thông tin của tất cả các giá trị trong mẫu số liệu. Hai trong số đó là *phương sai* và *độ lệch chuẩn*.

Cụ thể với mẫu số liệu  $x_1, x_2, \dots, x_n$ , nếu gọi số trung bình là  $\bar{x}$  thì với mỗi giá trị  $x_i$ , độ lệch của nó so với giá trị trung bình là  $x_i - \bar{x}$ .

• **Phương sai** là giá trị  $s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ .

• Căn bậc hai của phương sai,  $s = \sqrt{s^2}$ , được gọi là **độ lệch chuẩn**.

**Chú ý:** Người ta còn sử dụng đại lượng để đo độ phân tán của mẫu số liệu:

$$\hat{s}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}.$$

**Ý nghĩa:** Nếu số liệu càng phân tán thì phương sai và độ lệch chuẩn càng lớn.

**Ví dụ:** Mẫu số liệu sau đây cho biết số học sinh được lên lớp của 7 lớp khối 10 tại một trường Trung học phổ thông:

45   42   47   40   41   44   42

Tìm phương sai và độ lệch chuẩn cho mẫu số liệu trên. Qua các đại lượng này, em có nhận xét gì về độ phân tán của mẫu số liệu?

### Hướng dẫn giải

Số trung bình của mẫu số liệu là:  $\bar{x} = \frac{45 + 42 + 47 + 40 + 41 + 44 + 42}{7} = 43$ .

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
45	$45 - 43 = 2$	4
42	$42 - 43 = -1$	1
47	$47 - 43 = 4$	16
40	$40 - 43 = -3$	9
41	$41 - 43 = -2$	4
44	$44 - 43 = 1$	1
42	$42 - 43 = -1$	1
Tổng		36

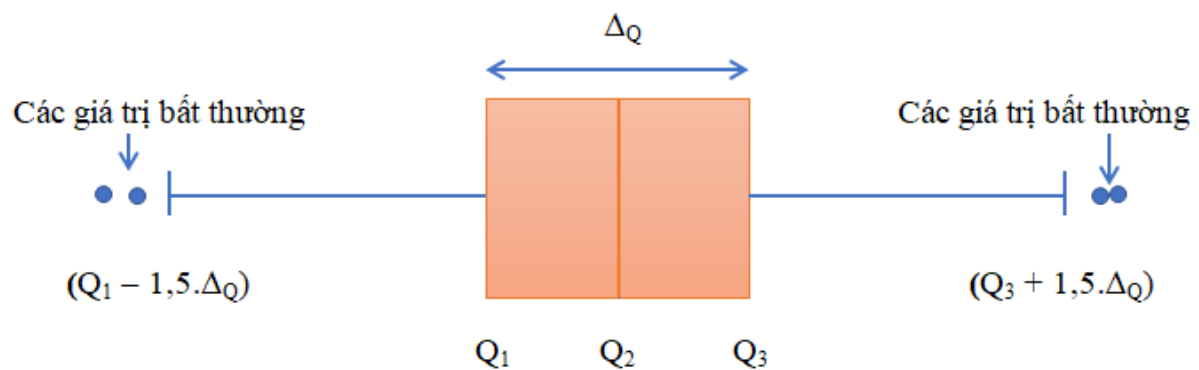
Mẫu số liệu gồm 7 giá trị nên  $n = 7$ . Do đó phương sai là:  $s^2 = \frac{36}{7} \approx 5,14$ .

Độ lệch chuẩn là:  $s = \sqrt{5,14} \approx 2,27$ .

Qua các đại lượng này, ta thấy phương sai và độ lệch chuẩn không lớn nên số liệu không quá phân tán.

### 3. Phát hiện số liệu bất thường hoặc không chính xác bằng biểu đồ hộp

Trong mẫu số liệu thống kê, có khi ta sẽ gặp phải những giá trị quá lớn hoặc quá nhỏ so với đa số các giá trị khác. Những giá trị này được gọi là **giá trị bất thường**. Chúng xuất hiện trong mẫu số liệu có thể do nhầm lẫn hay sai sót nào đó. Ta có thể dùng biểu đồ hộp để phát hiện những giá trị bất thường này.



Các giá trị lớn hơn  $Q_3 + 1,5 \cdot \Delta_Q$  hoặc bé hơn  $Q_1 - 1,5 \cdot \Delta_Q$  được xem là **giá trị bất thường**.

**Ví dụ:** Hàm lượng Canxi (đơn vị mg) trong 100 g một số loại thực phẩm được cho như trong bảng sau:

22	20	20	19	20	0	29	16	13	21
18	34	16	18	10	15	18	14	4	8

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

#### Hướng dẫn giải

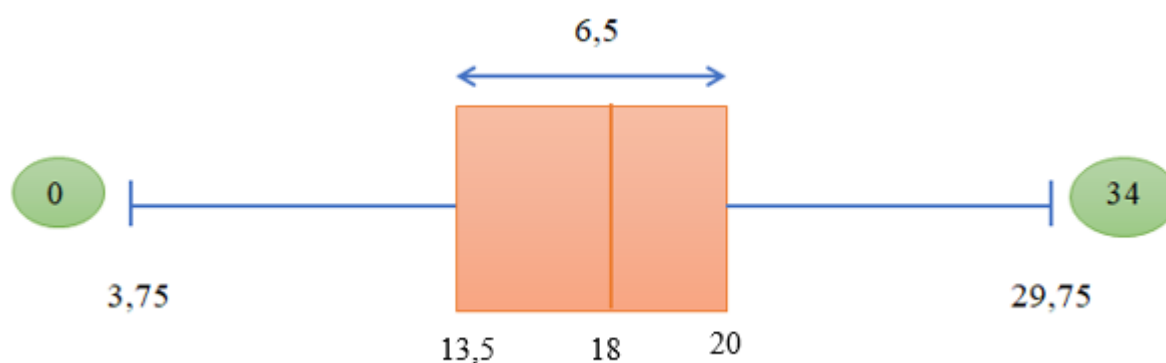
Sắp xếp các số liệu theo thứ tự không giảm, ta được:

0; 4; 8; 10; 13; 14; 15; 16; 16; 18; 18; 18; 19; 20; 20; 20; 21; 22; 29; 34.

Từ mẫu số liệu trên, ta tính được  $Q_2 = 18$ ;  $Q_1 = 13,5$  và  $Q_3 = 20$ . Do đó khoảng tứ phân vị là:

$$\Delta_Q = Q_3 - Q_1 = 20 - 13,5 = 6,5.$$

Biểu đồ hộp cho mẫu số liệu này là:



Ta có  $Q_1 - 1,5.\Delta_Q = 3,75$  và  $Q_3 + 1,5.\Delta_Q = 29,75$  nên trong mẫu số liệu có hai giá trị được xem là bất thường là 0 mg (bé hơn 3,75 mg) và 34 mg (lớn hơn 29,75 mg).

## B. Bài tập tự luyện

### B1. Bài tập trắc nghiệm

**Câu 1.** Điểm kiểm tra của 11 học sinh cho bởi bảng số liệu sau

Điểm	7	7,5	8	8,5	9	9,5
Tần số	1	2	3	2	2	1

Tìm phương sai của bảng số liệu

- A. 0,34;
- B. 0,50;
- C. 0,65;
- D. 5,54.

### Hướng dẫn giải

**Đáp án đúng là: B**

Giá trị trung bình của mẫu số liệu là

$$\bar{x} = \frac{1.7 + 2.7,5 + 3.8 + 2.8,5 + 2.9 + 1.9,5}{11} \approx 8,23.$$

Ta có bảng sau

Giá trị	Độ lệch	Bình phương độ lệch
7	$7 - 8,23 = -1,23$	1,5129
7,5	$7,5 - 8,23 = -0,73$	0,5329
7,5	$7,5 - 8,23 = -0,73$	0,5329
8	$8 - 8,23 = -0,23$	0,0529
8	$8 - 8,23 = -0,23$	0,0529
8	$8 - 8,23 = -0,23$	0,0529
8,5	$8,25 - 8,23 = 0,02$	0,0004
8,5	$8,25 - 8,23 = 0,02$	0,0004
9	$9 - 8,23 = 0,77$	0,5929
9	$9 - 8,23 = 0,77$	0,5929
9,5	$9,5 - 8,23 = 1,27$	1,6129
Tổng		5,5369

Vì có 11 giá trị nên  $n = 11$  do đó  $s^2 = \frac{5,5369}{11} \approx 0,5$ .

**Câu 2.** Điểm kiểm tra học kỳ của 10 học sinh được thống kê như sau: 6; 7; 7; 5; 8; 6; 9; 9; 8; 6. Khoảng biến thiên của dãy số là

A. 5;

B. 4;

C. 3;

D. 2.

### Hướng dẫn giải

**Đáp án đúng là: B**

Ta có giá trị lớn nhất của số liệu là 9 và giá trị nhỏ nhất của số liệu là 5.

Khoảng biến thiên là hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất trong mẫu số liệu.

Vậy  $R = 9 - 5 = 4$ .

**Câu 3.** Điều tra chiều cao của 10 hs lớp 10A cho kết quả như sau: 154; 160; 155; 162; 165; 162; 155; 160; 165; 162 (đơn vị cm). Khoảng tứ phân vị là

A. 5;

B. 6;

C. 7;

D. 8.

**Hướng dẫn giải**

**Đáp án đúng là: C**

Ta sắp xếp số liệu theo thứ tự không giảm như sau: 154; 155; 155; 160; 160; 162; 162; 162; 165; 165.

Vì  $n = 10$  là số chẵn nên  $Q_2$  là trung bình cộng của hai số chính giữa

$$Q_2 = (160 + 162) : 2 = 161$$

Ta tìm  $Q_1$  là trung vị nửa số liệu bên trái  $Q_2$  là 154; 155; 155; 160; 160 gồm 5 giá trị và tìm được  $Q_1 = 155$

Ta tìm  $Q_3$  là trung vị nửa số liệu bên phải  $Q_2$  là 162; 162; 162; 165; 165 gồm 5 giá trị và tìm được  $Q_3 = 162$

Vậy khoảng tứ phân vị  $\Delta_Q = Q_3 - Q_1 = 162 - 155 = 7$ .

## **B2. Bài tập tự luận**

**Câu 1.** Mỗi khẳng định sau đúng hay sai?

a) Nếu các giá trị của mẫu số liệu càng tập trung quanh giá trị trung bình thì độ lệch chuẩn càng lớn;

b) Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và bé nhất, bỏ qua thông tin của các giá trị còn lại;



c) Khoảng tứ phân vị có sử dụng thông tin của các giá trị lớn nhất và giá trị bé nhất;

d) Các số đo độ phân tán đều không âm.

### **Hướng dẫn giải**

a) Sai.

Nếu các giá trị của mẫu số liệu đều tập trung quanh giá trị trung bình thì độ lệch của mỗi giá trị  $x_i - \bar{x}$  càng nhỏ.

Ta suy ra  $(x_i - \bar{x})^2$  cũng càng nhỏ.

Do đó tổng các bình phương độ lệch cũng càng nhỏ.

Mà độ lệch chuẩn tỉ lệ thuận với tổng các bình phương độ lệch.

Vậy trong trường hợp này, độ lệch chuẩn càng nhỏ.

b) Đúng.

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.

c) Sai.

Khoảng tứ phân vị chỉ sử dụng 50% số liệu chính giữa của mẫu số liệu đã sắp xếp. Vậy khoảng biến thiên không sử dụng giá trị lớn nhất và giá trị bé nhất.

d) Đúng.

Khoảng biến thiên là hiệu của giá trị lớn nhất và giá trị bé nhất. Do đó khoảng biến thiên luôn không âm.

Khoảng tứ phân vị là hiệu của tứ phân vị thứ ba và tứ phân vị thứ nhất. Do đó khoảng tứ phân vị luôn không âm.

Phương sai là tổng các bình phương các độ lệch nên luôn không âm.

Độ lệch chuẩn là căn bậc hai của phương sai nên độ lệch chuẩn luôn không âm.

**Câu 2.** Mẫu số liệu sau đây cho biết cân nặng của 10 trẻ sơ sinh (đơn vị kg):

2,8   2,5   3,1   3,3   2,9   3,5   4,1   2,8   3,0   3,3

Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn cho mẫu số liệu trên.

### **Hướng dẫn giải**

+ Khoảng biến thiên:

Số cân nặng cao nhất, thấp nhất lần lượt là 4,1 (kg) và 2,5 (kg). Do đó khoảng biến thiên là  $R = 4,1 - 2,5 = 1,6$  (kg).

+ Khoảng tứ phân vị:

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

2,5   2,8   2,8   2,9   3,0   3,1   3,3   3,3   3,5   4,1

Mẫu số liệu trên gồm 10 giá trị, hai phần tử chính giữa là 3,0 và 3,1.

Do đó trung vị là  $Q_2 = (3,0 + 3,1) : 2 = 3,05$ .

Nửa số liệu bên trái là 2,5; 2,8; 2,8; 2,9; 3,0 gồm 5 giá trị. Do đó trung vị là số ở vị trí chính giữa  $Q_1 = 2,8$ .

Nửa số liệu bên phải là 3,1; 3,3; 3,3; 3,5; 4,1 gồm 5 giá trị. Do đó trung vị là số ở vị trí chính giữa  $Q_3 = 3,3$ .

Vậy khoảng tứ phân vị cho mẫu số liệu là  $\Delta_Q = Q_3 - Q_1 = 3,3 - 2,8 = 0,5$ .

+ Độ lệch chuẩn:

Số trung bình của mẫu số liệu là:

$$\bar{x} = \frac{2,5 + 2,8.2 + 2,9 + 3,0 + 3,1 + 3,3.2 + 3,5 + 4,1}{10} = 3,13$$

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
2,8	$2,8 - 3,13 = -0,33$	0,1089
2,5	$2,5 - 3,13 = -0,63$	0,3969
3,1	$3,1 - 3,13 = -0,03$	0,0009
3,3	$3,3 - 3,13 = 0,17$	0,0289
2,9	$2,9 - 3,13 = -0,23$	0,0529
3,5	$3,5 - 3,13 = 0,37$	0,1369
4,1	$4,1 - 3,13 = 0,97$	0,9409
2,8	$2,8 - 3,13 = -0,33$	0,1089
3,0	$3,0 - 3,13 = -0,13$	0,0169
3,3	$3,3 - 3,13 = 0,17$	0,0289
Tổng		1,821

Mẫu số liệu gồm 10 giá trị nên  $n = 10$ . Do đó phương sai là:  $s^2 = \frac{1,821}{10} = 0,1821$ .

Độ lệch chuẩn là:  $s = \sqrt{0,1821} \approx 0,4267$ .

**Câu 3.** Tỷ lệ thất nghiệp từ năm 2006 – 2020 tại Việt Nam (đơn vị %) được cho như sau:

0,00      2,03      0,00      1,74      1,11      1,00      1,03      1,32  
1,26      1,85      1,85      1,87      1,16      2,04      2,39

Hãy tìm các giá trị bất thường (nếu có) của mẫu số liệu trên.

### Hướng dẫn giải

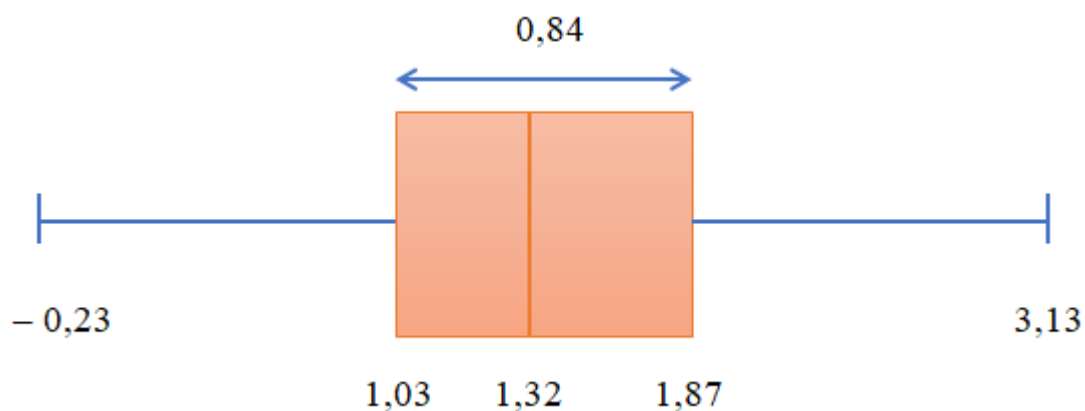
Sắp xếp các số liệu theo thứ tự không giảm ta có:

0,00; 0,00; 1,00; 1,03; 1,11; 1,16; 1,26; 1,32; 1,74; 1,85; 1,85; 1,87; 2,03; 2,04; 2,39.

Từ mẫu số liệu trên, ta tính được  $Q_2 = 1,32$ ;  $Q_1 = 1,03$  và  $Q_3 = 1,87$ . Do đó khoảng tứ phân vị là:

$$\Delta_Q = Q_3 - Q_1 = 1,87 - 1,03 = 0,84.$$

Biểu đồ hộp cho mẫu số liệu này là:



Ta có  $Q_1 - 1,5 \cdot \Delta_Q = -0,23$  và  $Q_3 + 1,5 \cdot \Delta_Q = 3,13$  nên trong mẫu số liệu trên không có giá trị nào được xem là bất thường.