

## Bài 14. Các số đặc trưng. Đo độ phân tán

**Mở đầu trang 84 SGK Toán 10 tập 1:** Dưới đây là điểm trung bình môn học kì I của hai bạn An và Bình:

	Toán	Vật lí	Hoá học	Ngữ văn	Lịch sử	Địa lí	Tin học	Tiếng Anh
An	9,2	8,7	9,5	6,8	8,0	8,0	7,3	6,5
Bình	8,2	8,1	8,0	7,8	8,3	7,9	7,6	8,1

Điểm trung bình môn học kì của An và Bình đều là 8,0 nhưng rõ ràng Bình “học đều” hơn An. Có thể dùng những số đặc trưng nào để đo mức độ “học đều”?

### Lời giải

Bài học này sẽ giới thiệu một vài số đặc trưng như: khoảng biến thiên, khoảng tứ phân vị, độ lệch chuẩn và phương sai.

Ở đây ta sẽ sử dụng độ lệch chuẩn để so sánh

Điểm trung bình môn học kì I của An là:  $\overline{X}_1 = 8,0$

$$\Rightarrow s_1^2 = \frac{9,2 - 8,0^2 + 8,7 - 8,0^2 + \dots + 7,3 - 8,0^2 + 6,5 - 8,0^2}{8} = 1,045$$

$$\Rightarrow s_1 = \sqrt{s_1^2} = \sqrt{1,045} \approx 1,02.$$

Điểm trung bình môn học kì I của Bình là  $\overline{X}_2 = 8,0$

$$\Rightarrow s_2^2 = \frac{(8,2 - 8,0)^2 + (8,1 - 8,0)^2 + \dots + (7,6 - 8,0)^2 + (8,1 - 8,0)^2}{8} = 0,045$$

$$\Rightarrow s_2 = \sqrt{s_2^2} = \sqrt{0,045} \approx 0,21.$$

Vì  $s_2 < s_1$  nên độ phân tán của số liệu 2 nhỏ hơn độ phân tán của số liệu 1 hay bạn Bình học đều hơn bạn An.

**Hoạt động 1 trang 84 SGK Toán 10 tập 1:** Một cổ động viên đã thống kê điểm số mà hai câu lạc bộ Leicester City và Everton đạt được trong năm mùa giải Ngoại hạng Anh gần đây, từ mùa giải 2014 – 2015 đến mùa giải 2018 – 2019 như sau:

Leicester City: 41 81 44 47 52.

Everton: 47 47 61 49 54.

Cổ động viên cho rằng, Everton thi đấu ổn hơn Leicester City. Em có đồng ý với nhận định này không? Vì sao?

**Lời giải:**

Ta có câu lạc bộ Leicester City có điểm lớn nhất là 81 và nhỏ nhất là 41 nên khoảng cách giữa điểm cao nhất và thấp nhất là 40.

Câu lạc bộ Everton có điểm lớn nhất là 61 và nhỏ nhất là 41 nên khoảng cách giữa điểm cao nhất và thấp nhất là 20.

Ta thấy  $20 < 40$  nên câu lạc bộ Everton thi đấu ổn định hơn.

**Luyện tập 1 trang 85 SGK Toán 10 tập 1:** Mẫu số liệu sau cho biết chiều cao (đơn vị cm) của các bạn trong tổ:

163 159 172 167 165 168 170 161.

Tính khoảng biến thiên của mẫu số liệu này.

**Lời giải:**

Trong mẫu số liệu trên, chiều cao cao nhất là 172 cm và thấp nhất là 159 cm.

Do đó, khoảng biến thiên là  $R = 172 - 159 = 13$  (cm).

Vậy khoảng biến thiên  $R = 13$ cm.

**Hoạt động 2 trang 85 SGK Toán 10 tập 1:** Trong một tuần, nhiệt độ cao nhất trong ngày (đơn vị  $^{\circ}\text{C}$ ) tại hai thành phố Hà Nội và Điện Biên được cho như sau:

Hà Nội: 23 25 28 28 32 33 35.

Điện Biên: 16 24 26 26 26 27 28.

- a) Tính khoảng biến thiên của mỗi mẫu số liệu và so sánh.
- b) Em có nhận xét gì về sự ảnh hưởng của giá trị 16 đến khoảng biến thiên của mẫu số liệu về nhiệt độ cao nhất trong ngày tại Điện Biên?
- c) Tính các tứ phân vị và hiệu  $Q_3 - Q_1$  cho mỗi mẫu số liệu. Có thể dùng hiệu này để đo độ phân tán của mẫu số liệu không?

**Lời giải:**

a)

· Hà Nội: Nhiệt độ cao nhất là 35, nhiệt độ thấp nhất là 23.

Khi đó, khoảng biến thiên là:  $R_1 = 35 - 23 = 12$ .

· Điện Biên: Nhiệt độ cao nhất là 28, nhiệt độ thấp nhất là 16.

Khi đó, khoảng biến thiên là:  $R_2 = 28 - 16 = 12$ .

Ta thấy  $R_1 = R_2 = 12$ .

Vậy khoảng biến thiên về nhiệt độ của Hà Nội và Điện Biên bằng nhau.

b) Về trực quan nhiệt độ tại Điện Biên thay đổi khá ít, riêng một ngày có nhiệt độ thấp hần là  $16^{\circ}\text{C}$ , giá trị 16 này đã ảnh hưởng rất nhiều đến khoảng biến thiên.

c)

· Hà Nội: 23 25 28 28 32 33 35.

Vì  $n = 7$  là số lẻ nên số trung vị là số chính giữa là  $Q_2 = 28$ .

Ta tìm  $Q_1$  là trung vị của nửa số liệu bên trái  $Q_2$ :

23; 25; 28.

Do đó  $Q_1 = 25$ .

Ta tìm  $Q_3$  là trung vị của nửa số liệu bên phải  $Q_2$ :

32; 33; 35.

Do đó  $Q_3 = 33$ .

Tứ phân vị cho mẫu số liệu này là:  $Q_1 = 25$ ;  $Q_2 = 28$ ,  $Q_3 = 33$ .

Suy ra  $\Delta_Q = Q_3 - Q_1 = 33 - 25 = 8$ .

· Điện Biên: 16 24 26 26 26 27 28.

Vì  $n = 7$  là số lẻ nên số trung vị là số chính giữa là  $Q'_2 = 26$ .

Ta tìm  $Q'_1$  là trung vị của nửa số liệu bên trái  $Q'_2$ :

16; 24; 26.

Do đó  $Q'_1 = 24$ .

Ta tìm  $Q'_3$  là trung vị của nửa số liệu bên phải  $Q'_2$ :

26; 27; 28.

Do đó  $Q'_3 = 27$ .

Tứ phân vị cho mẫu số liệu này là  $Q'_1 = 24$ ;  $Q'_2 = 26$ ,  $Q'_3 = 27$ .

Suy ra  $\Delta'_Q = Q'_3 - Q'_1 = 27 - 24 = 3$ .

Có thể dùng số liệu này để đo độ phân tán của mẫu số liệu.

**Luyện tập 2 trang 86 SGK Toán 10 tập 1:** Mẫu số liệu sau đây cho biết số bài hát ở mỗi album trong bộ sưu tập của An:

12 7 10 9 12 9 10 11 10 14.

Hãy tìm khoảng tứ phân vị cho mẫu số liệu này.

### Lời giải

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

7; 9; 9; 10; 10; 10; 11; 12; 12; 14.

Ta có  $n = 10$  nên trung vị bằng trung bình cộng hai giá trị chính giữa:

$$Q_2 = (10 + 10) : 2 = 10.$$

Nửa số liệu bên trái  $Q_2$ : 7 9 9 10 10

Do đó  $Q_1 = 9$ .

Nửa số liệu bên phải  $Q_2$ : 10 11 12 12 14

Do đó  $Q_3 = 12$ .

Vậy khoảng tứ phân vị cho mẫu số liệu là:  $\Delta_Q = Q_3 - Q_1 = 12 - 9 = 3$ .

**Luyện tập 3 trang 87 SGK Toán 10 tập 1:** Dùng đồng hồ đo thời gian có độ chia nhỏ nhất đến 0,001 giây để đo 7 lần thời gian rơi tự do của một vật bắt đầu từ điểm A ( $V_A = 0$ ) đến điểm B. Kết quả đo như sau:

0,398      0,399      0,408      0,410      0,406      0,405      0,402.

(Theo Bài tập Vật lí 10, Nhà xuất bản Giáo dục Việt Nam, 2018)

Hãy tìm phương sai và độ lệch chuẩn cho mẫu số liệu này. Qua các đại lượng này, em có nhận xét gì về độ chính xác của phép đo trên?

### Lời giải

Số trung bình của mẫu số liệu là:

$$\bar{X} = \frac{0,398 + 0,399 + 0,408 + 0,410 + 0,406 + 0,405 + 0,402}{7} = 0,404.$$

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
0,398	$0,398 - 0,404 = -0,006$	0,000036
0,399	$0,399 - 0,404 = -0,005$	0,000025
0,408	$0,408 - 0,404 = 0,004$	0,000016
0,410	$0,410 - 0,404 = 0,006$	0,000036
0,406	$0,406 - 0,404 = 0,002$	0,000004
0,405	$0,405 - 0,404 = 0,001$	0,000001
0,402	$0,402 - 0,404 = -0,002$	0,000004
Tổng		0,000122

Mẫu số liệu gồm 7 giá trị nên  $n = 7$ .

Do đó phương sai là:  $s^2 = \frac{0,000122}{7} \approx 0,000017$ .

Độ lệch chuẩn là:  $s = \sqrt{\frac{0,000122}{7}} \approx 0,0042$ .

Đối với số liệu này phương sai và độ lệch chuẩn nhỏ nên độ phân tán của số liệu thấp. Do đó phép đo trên khá chính xác.

**Luyện tập 4 trang 87 SGK Toán 10 tập 1:** Một mẫu số liệu có tứ phân vị thứ nhất là 56 và tứ phân vị thứ ba là 84. Hãy kiểm tra xem trong hai giá trị 10 và 100 giá trị nào được xem là giá trị bất thường.

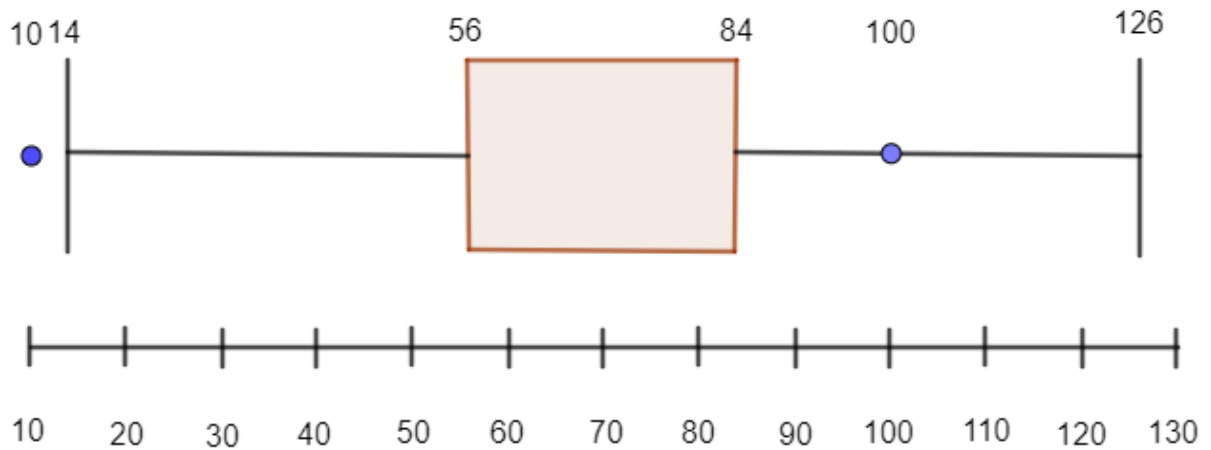
**Lời giải**

Ta có  $Q_1 = 56$  và  $Q_3 = 84$ .

Do đó, khoảng tứ phân vị là:

$$\Delta_Q = Q_3 - Q_1 = 84 - 56 = 28.$$

Biểu đồ hộp cho mẫu số liệu này là:



Ta có:

$$Q_1 - 1,5 \cdot \Delta_Q = 56 - 1,5 \cdot 28 = 14;$$

$$Q_3 + 1,5 \cdot \Delta_Q = 84 + 1,5 \cdot 28 = 126.$$

Ta thấy  $10 < 14$  nên 10 là giá trị bất thường

$14 < 100 < 126$  nên 100 không là giá trị bất thường.

**Bài 5.11 trang 88 SGK Toán 10 tập 1:** Mỗi khẳng định sau đúng hay sai?

- (1) Nếu các giá trị của mẫu số liệu càng tập trung quanh giá trị trung bình thì độ lệch chuẩn càng lớn.
- (2) 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.
- (3) Khoảng tứ phân vị có sử dụng thông tin của giá trị lớn nhất, giá trị bé nhất.

(4) Khoảng tứ phân vị chính là khoảng biến thiên của nửa dưới mẫu số liệu đã sắp xếp.

(5) Các số đo độ phân tán đều không âm.

**Lời giải:**

· Khẳng định (1): Nếu các giá trị của mẫu số liệu càng tập trung quanh giá trị trung bình thì độ lệch của mỗi giá trị so với giá trị trung bình càng nhỏ (tức là  $x_i - \bar{x}$  càng nhỏ, với  $i = 1; 2; \dots; n$ ), dẫn đến độ lệch chuẩn càng nhỏ.

Do đó, khẳng định (1) sai.

· Khẳng định (2): Khoảng biến thiên R bằng hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất nên chỉ sử dụng thông tin của giá trị lớn nhất và bé nhất.

Do đó, khẳng định (2) đúng.

· Khẳng định (3): Khoảng tứ phân vị  $\Delta_Q = Q_3 - Q_1$ , các giá trị  $Q_1, Q_3$  không bị ảnh hưởng bởi giá trị của giá trị lớn nhất và giá trị nhỏ nhất (với  $n > 4$ ).

Do đó, khẳng định (3) sai.

· Khẳng định (4): Khoảng tứ phân vị chính là khoảng biến thiên của 50% số liệu chính giữa của mẫu số liệu đã sắp xếp

Do đó, khẳng định (4) sai.

· Khẳng định (5): Các số đo độ phân tán là:

Khoảng biến thiên  $R = \text{Số lớn nhất} - \text{Số nhỏ nhất} > 0$

Trước khi tính khoảng tứ phân vị thì mẫu số liệu được sắp xếp theo thứ tự không giảm

$$\Rightarrow Q_3 > Q_1 \Rightarrow \Delta_Q = Q_3 - Q_1 > 0$$



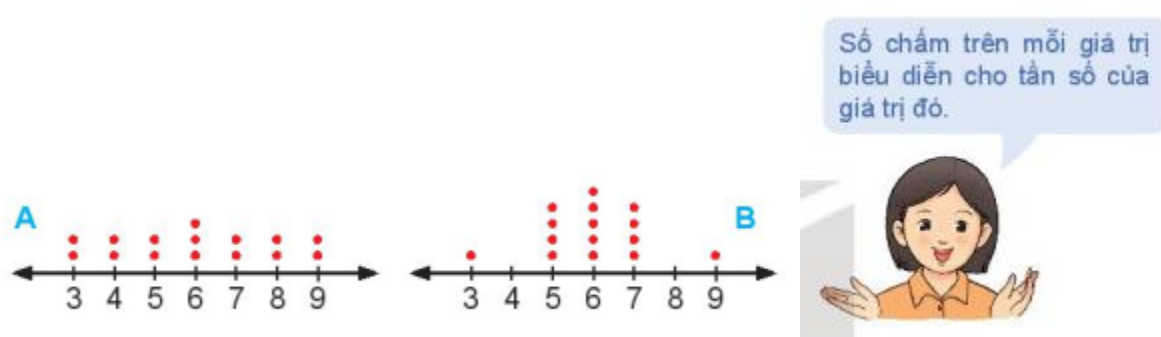
Phương sai: 
$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} > 0$$

Độ lệch chuẩn:  $s = \sqrt{s^2} > 0$

Do đó, các số đo độ phân tán đều không âm

Do đó, khẳng định (5) đúng.

**Bài 5.12 trang 88 SGK Toán 10 tập 1:** Cho hai biểu đồ chấm điểm biểu diễn hai mẫu số liệu A, B như sau:



Không tính toán, hãy cho biết:

- Hai mẫu số liệu này có cùng khoảng biến thiên và số trung bình không?
- Mẫu số liệu nào có phương sai lớn hơn?

**Lời giải:**

a) Ta có cả hai mẫu số liệu đều có giá trị nhỏ nhất là 3, giá trị lớn nhất là 9.

Do đó cả hai mẫu số liệu có cùng khoảng biến thiên.

Hai biểu đồ này có dạng đối xứng qua điểm 6 nên giá trị trung bình của hai mẫu số liệu A và B bằng nhau và bằng 6.

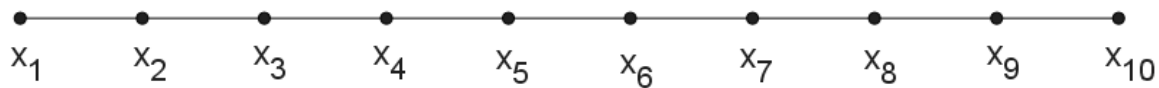
b) Từ biểu đồ, ta thấy các giá trị của dãy số liệu B tập trung nhiều hơn quanh giá trị trung bình nên mẫu số liệu B có phương sai nhỏ hơn. Vậy mẫu số liệu A có phương sai lớn hơn.

**Bài 5.13 trang 88 SGK Toán 10 tập 1:** Cho mẫu số liệu gồm 10 số dương không hoàn toàn giống nhau. Các số đo độ phân tán (khoảng biến thiên, khoảng tứ phân vị, độ lệch chuẩn) sẽ thay đổi như thế nào nếu:

- a) Nhân mỗi giá trị của mẫu số liệu với 2.
- b) Cộng mỗi giá trị của mẫu số liệu với 2.

**Lời giải:**

a) Gọi các giá trị dương của mẫu số liệu ban đầu theo thứ tự không giảm là:



Ta có  $n = 10$  là số chẵn nên trung vị là giá trị trung bình của số thứ 5 và thứ 6.

Do đó  $Q_1$  là số thứ 3 và  $Q_3$  là số thứ 8.

a) Khi nhân mỗi giá trị của mẫu số liệu với 2 thì

+ Số lớn nhất tăng 2 lần và số nhỏ nhất tăng 2 lần nên khoảng biến thiên  $R$  tăng 2 lần.

+  $Q_1$  và  $Q_3$  tăng 2 lần nên khoảng tứ phân vị  $\Delta_Q = Q_3 - Q_1$  tăng 2 lần.

+ Giá trị trung bình tăng 2 lần.

Nên độ lệch của mỗi giá trị so với giá trị trung bình  $|x_i - \bar{x}|$  cũng tăng 2 lần.

Suy ra  $(x_i - \bar{x})^2$  tăng 4 lần.

Khi đó, phương sai tăng 4 lần.

Do đó độ lệch chuẩn tăng 2 lần.

Vậy các khoảng biến thiên, độ lệch chuẩn, khoảng tứ phân vị của dãy số liệu mới bằng hai lần các khoảng biến thiên, độ lệch chuẩn, khoảng tứ phân vị ban đầu.

b) Khi cộng mỗi giá trị của mẫu số liệu với 2 thì

+ Số lớn nhất tăng 2 đơn vị và số nhỏ nhất tăng 2 đơn vị.

Suy ra khoảng biến thiên R không đổi vì phần tăng thêm bị triệt tiêu cho nhau.

+  $Q_1$  và  $Q_3$  tăng 2 đơn vị nên khoảng tứ phân vị  $\Delta_Q = Q_3 - Q_1$  không đổi vì phần tăng thêm bị triệt tiêu cho nhau.

+ Giá trị trung bình tăng 2 đơn vị

Nên độ lệch của mỗi giá trị so với giá trị trung bình  $|x_i - \bar{x}|$  không đổi vì phần tăng thêm bị triệt tiêu cho nhau.

Suy ra  $(x_i - \bar{x})^2$  không đổi

Khi đó, phương sai không đổi.

Do đó độ lệch chuẩn không đổi.

Vậy khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn đều không đổi.

**Bài 5.14 trang 88 SGK Toán 10 tập 1:** Từ mẫu số liệu về thuế thuốc lá của 51 thành phố tại một quốc gia, người ta tính được:

Giá trị nhỏ nhất bằng 2,5;  $Q_1 = 36$ ;  $Q_2 = 60$ ;  $Q_3 = 100$ ; giá trị lớn nhất bằng 205.

a) Tỷ lệ thành phố có thuế thuốc lá lớn hơn 36 là bao nhiêu?

b) Chỉ ra hai giá trị sao cho có 50% giá trị của mẫu số liệu nằm giữa hai giá trị này.

c) Tìm khoảng tứ phân vị của mẫu số liệu.

**Lời giải:**

a) Vì số các giá trị của số liệu  $n = 51$  là số lẻ nên trung vị của số liệu là giá trị thứ 26.

Nửa bên trái số trung vị gồm 25 số liệu là số lẻ nên tứ phân vị thứ nhất là giá trị thứ 13 có giá trị là 36.

Do đó có  $51 - 13 = 38$  thành phố có thuế thuốc lá lớn hơn 36.

Suy ra tỉ lệ các thành phố có thuế thuốc lá lớn hơn 36 là:  $\frac{38}{51} \approx 74,51\%$ .

Vậy tỉ lệ các thành phố có thuế thuốc là lớn hơn 36 khoảng 74,51%.

b)

Có nhiều phương án để lựa chọn trong bài này.

Chẳng hạn ta chọn hai giá trị là  $Q_1$  và  $Q_3$ , vì khoảng giữa hai giá trị này là khoảng tứ phân vị và khoảng này là khoảng biến thiên của 50% số liệu chính giữa của mẫu số liệu đã sắp xếp.

Vậy giữa hai giá trị  $Q_1 = 36$  và  $Q_3 = 100$  có 50% giá trị của mẫu số liệu nằm giữa hai giá trị này.

c) Khoảng tứ phân vị của mẫu số liệu này là:

$$\Delta_Q = Q_3 - Q_1 = 100 - 36 = 64.$$

**Bài 5.15 trang 88 SGK Toán 10 tập 1:** Mẫu số liệu sau đây cho biết cân nặng của 10 trẻ sơ sinh (đơn vị kg):

2,977	3,155	3,920	3,412	4,236
2,593	3,270	3,813	4,042	3,387.

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

**Lời giải**

Sắp xếp các giá trị của số liệu trên theo thứ tự từ không giảm là:

2,593; 2,977; 3,155; 3,270; 3,387; 3,412; 3,813; 3,920; 4,042; 4,236. .

Ta có giá trị lớn nhất là 4,236 kg và giá trị nhỏ nhất là 2,593 kg.

Khi đó, khoảng biến thiên là:  $R = 4,236 - 2,593 = 1,643$ .

Vì  $n = 10$  là số chẵn nên trung vị là trung bình cộng của hai giá trị chính giữa:

$$Q_2 = (3,387 + 3,412) : 2 = 3,3995.$$

Nửa số liệu bên trái gồm 5 số liệu là một số lẻ nên tứ phân vị thứ nhất là:  $Q_1 = 3,155$ .

Nửa số liệu bên phải gồm 5 số liệu là một số lẻ nên tứ phân vị thứ ba là:  $Q_3 = 3,920$ .

Khoảng tứ phân vị là:  $\Delta_Q = Q_3 - Q_1 = 3,920 - 3,155 = 0,765$ .

Số trung bình cộng của mẫu số liệu là:

$$\begin{aligned}\bar{X} &= \frac{2,593 + 2,977 + 3,155 + 3,270 + 3,387 + 3,412 + 3,813 + 3,920 + 4,042 + 4,236}{10} \\ &= 3,4805.\end{aligned}$$

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
2,593	0,888	0,789
2,977	0,504	0,254
3,155	0,326	0,106
3,270	0,211	0,045
3,387	0,094	0,009
3,412	0,069	0,005

3,813	0,332	0,110
3,920	0,439	0,193
4,042	0,561	0,315
4,236	0,755	0,570
Tổng		2,396

Phương sai:  $s^2 \approx \frac{2,396}{10} = 0,2396$

Độ lệch chuẩn:  $s = \sqrt{s^2} \approx \sqrt{0,2396} \approx 0,489$ .

Vậy khoảng biến thiên  $R = 1,643$ , khoảng tứ phân vị  $\Delta_Q = 0,765$ ; độ lệch chuẩn  $s \approx 0,489$ .

**Bài 5.16 trang 88 SGK Toán 10 tập 1:** Tỷ lệ thất nghiệp ở một quốc gia vào năm 2007 (đơn vị %) được cho như sau:

7,8    3,2    7,7    8,7    8,6    8,4    7,2    3,6

5,0    4,4    6,7    7,0    4,5    6,0    5,4.

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

**Lời giải:**

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

3,2; 3,6; 4,4; 4,5; 5,0; 5,4; 6,0; 6,7; 7,0; 7,2; 7,7; 7,8; 8,4; 8,6; 8,7.

Vì  $n = 15$  là số lẻ nên số trung vị là giá trị chính giữa  $Q_2 = 6,7$ .

Nửa số liệu bên trái có 7 số liệu nên có tứ phân vị thứ nhất là  $Q_1 = 4,5$ .

Nửa số liệu bên phải có 7 số liệu nên có tứ phân vị thứ ba là  $Q_3 = 7,8$ .

Khoảng tứ phân vị là:  $\Delta_Q = Q_3 - Q_1 = 7,8 - 4,5 = 3,3$ .

Ta có:

$$Q_1 - 1,5\Delta_Q = 4,5 - 4,95 = -0,45;$$

$$Q_3 + 1,5\Delta_Q = 7,8 + 4,95 = 12,75.$$

Ta thấy không có giá trị nào dưới  $-0,45$  và trên  $12,75$  nên không có giá trị bất thường.

Vậy mẫu số liệu đã cho không có giá trị nào bất thường.