



CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU KHÔNG GHÉP NHÓM

BÀI 14. CÁC SỐ ĐẶC TRƯNG ĐO ĐỘ PHÂN TÁN



LÝ THUYẾT.

1. KHOẢNG BIẾN THIÊN VÀ KHOẢNG TỬ PHÂN VỊ

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 1. Điểm kiểm tra học kì môn Toán của các bạn Tổ 1, Tổ 2 lớp 10A được cho như sau:

Tổ 1: 7 8 8 9 8 8 8
Tổ 2: 10 6 8 9 9 7 8 7 8.

- Điểm kiểm tra trung bình của hai tổ có như nhau không?
- Tính các khoảng biến thiên của hai mẫu số liệu. Căn cứ trên chỉ số này, các bạn tổ nào học đồng đều hơn?

Giải

- Điểm kiểm tra trung bình của hai tổ đều bằng 8.
- Đối với Tổ 1: Điểm kiểm tra thấp nhất, cao nhất tương ứng là 7;9. Do đó, khoảng biến thiên là: $R_1 = 9 - 7 = 2$.

Đối với Tổ 2: Điểm kiểm tra thấp nhất, cao nhất tương ứng là 6;10. Do đó, khoảng biến thiên là: $R_2 = 10 - 6 = 4$.

Do $R_2 > R_1$ nên ta nói các bạn Tổ 1 học đều hơn các bạn Tổ 2.

Luyện 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.

Giải

Chiều cao thấp nhất, cao nhất tương ứng là 159; 172. Do đó, khoảng biến thiên là:
 $R = 172 - 159 = 13$.

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.



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

$$\Delta_Q = Q_3 - Q_1$$

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 2. Mẫu số liệu sau cho biết số ghế trống tại một rạp chiếu phim trong 9 ngày:

7 8 22 20 15 18 19 13 11.

Tìm khoảng tứ phân vị cho mẫu số liệu này.

Giải

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

7 8 11 13 15 18 19 20 22.

Mẫu số liệu gồm 9 giá trị nên trung vị là số ở vị trí chính giữa $Q_2 = 15$.

Nửa số liệu bên trái là 7, 8, 11, 13 gồm 4 giá trị, hai phần tử chính giữa là 8, 11.

Do đó, $Q_1 = (8 + 11) : 2 = 9,5$.

Nửa số liệu bên phải là 18, 19, 20, 22 gồm 4 giá trị, hai phần tử chính giữa là 19, 20.

Do đó, $Q_3 = (19 + 20) : 2 = 19,5$.

Vậy khoảng tứ phân vị cho mẫu số liệu là: $\Delta_Q = 19,5 - 9,5 = 10$.

2. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và nhỏ nhất của mẫu số liệu (bỏ qua thông tin của tất cả các giá trị khác), còn 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 là 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:

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

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

Ví dụ 3. Mẫu số liệu sau đây cho biết sĩ số của 5 lớp khối 10 tại một trường Trung

43 45 46 41 40

Tìm phương sai và độ lệch chuẩn cho mẫu số liệu này.



Giải

Số trung bình của mẫu số liệu là $\bar{x} = \frac{43 + 45 + 46 + 41 + 40}{5} = 43$.

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
43	$43 - 43 = 0$	0
45	$45 - 43 = 2$	4
46	$46 - 43 = 3$	9
41	$41 - 43 = -2$	4
40	$40 - 43 = -3$	9
Tổng		26

Bạn có thể sử dụng máy tính cầm tay, phần mềm bảng tính hay phần mềm thống kê để tính các số đặc trưng.

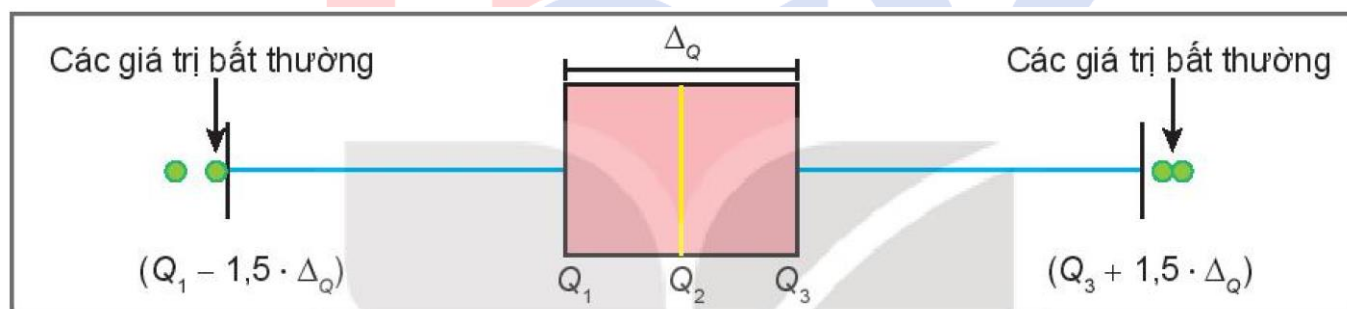


Mẫu số liệu gồm 5 giá trị nên $n = 5$. Do đó phương sai là $s^2 = \frac{26}{5} = 5,2$.

Độ lệch chuẩn là: $s = \sqrt{5,2} \approx 2,28$.

3. PHÁT HIỆN SỐ LIỆU BẤT THƯỜNG HOẶC KHÔNG CHÍNH XÁC BẰNG BIỂU ĐỒ HỘP

Trong mẫu số liệu thống kê, có khi gặp 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.



Hình 5.5. Biểu đồ hộp

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 Natri (đơn vị mg) 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.

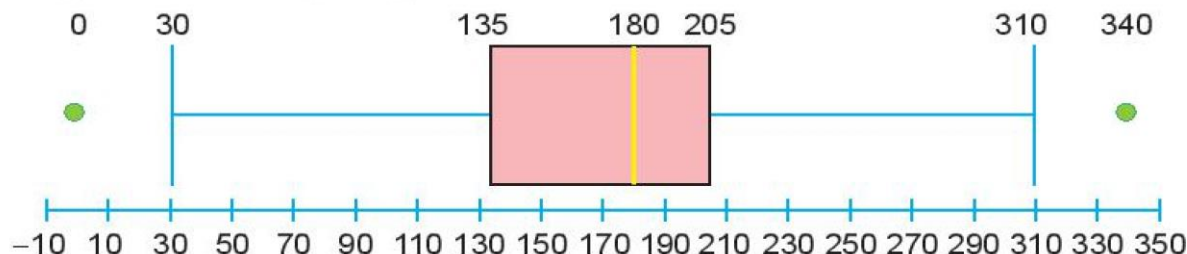
Tìm giá trị bất thường trong mẫu số liệu trên bằng cách sử dụng biểu đồ hộp.

Giải

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

$$\Delta_Q = 205 - 135 = 70$$

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 (lớn hơn 310 mg) và 0 mg (bé hơn 30 mg).



BÀI TẬP SÁCH GIÁO KHOA.

5.11. Mỗi khẳng định sau đúng hay sai?

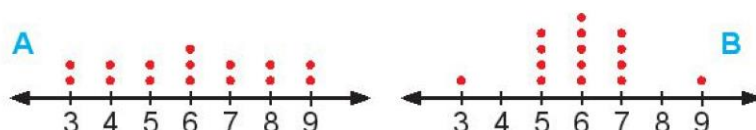
- (1) Nếu các giá trị của mẫu số liệu càng tập trung quanh giá trị trung bình thì độ lệch chuẩn càng lớn.
- (2) Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và bé nhất, bỏ qua thông tin của các giá trị còn lại.
- (3) Khoảng tứ phân vị có sử dụng thông tin của giá trị lớn nhất, giá trị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

Giải

Các khẳng định đúng: (2), (5).

Các khẳng định sai: (1), (3), (4).

5.12. Cho hai biểu đồ chấm biểu diễn hai mẫu số liệu A, B như sau:



Không tính toán, hãy cho biết:

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

Giải

- a) Khoảng biến thiên của hai mẫu số liệu bằng nhau.

Số trung bình của hai mẫu số liệu bằng nhau.

- b) Mẫu số liệu A có phương sai lớn hơn mẫu số liệu B.

5.13. 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.

Số chấm trên mỗi giá trị biểu diễn cho tần số của giá trị đó.





Giải

a) Nhân mỗi giá trị của mẫu số liệu với 2 thì:

Khoảng biến thiên tăng gấp 2 lần.

Khoảng tứ phân vị tăng gấp 2 lần.

Độ lệch chuẩn tăng gấp 4 lần.

b) Cộng mỗi giá trị của mẫu số liệu với 2 thì:

Khoảng biến thiên giữ nguyên.

Khoảng tứ phân vị giữ nguyên.

Độ lệch chuẩn giữ nguyên.

5.14. 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.

Giải

a) Từ mẫu số liệu về thuế thuốc lá của 51 thành phố tại một quốc gia, người ta tính được $Q_1 = 36$ nên có 12 thành phố có thuế thuốc lá lớn hơn 36.

Vì vậy, tỷ lệ thành phố có thuế thuốc lá lớn hơn 36 là: $\frac{12}{51} = \frac{4}{7}$.

b) Hai giá trị có 50% giá trị của mẫu số liệu nằm giữa là 36 và 100.

c) Khoảng tứ phân vị của mẫu số liệu là $\Delta_Q = Q_3 - Q_1 = 100 - 36 = 64$.

5.15. 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.

Giải

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

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

Khoảng biến thiên là $R = 4,236 - 2,593 = 1,643$.

Ta có: $Q_2 = 3,3995$; $Q_1 = 3,155$; $Q_3 = 3,920$

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

Độ lệch chuẩn là $s \approx 0,52$.

5.16. Tỷ lệ thất nghiệp ở một số 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.

Giải

Từ mẫu số liệu ta tính được $Q_1 = 4,5$ và $Q_3 = 7,8$. Do đó, khoảng tứ phân vị là:

$$\Delta_Q = 7,8 - 4,5 = 3,3$$

Ta có $Q_1 - 1,5\Delta_Q = -0,45$ và $Q_3 + 1,5\Delta_Q = 12,75$ nên trong mẫu số liệu trên không có giá trị bất thường.



HỆ THỐNG BÀI TẬP TỰ LUẬN.

I. KHOẢNG BIẾN THIÊN VÀ KHOẢNG TỨ PHÂN VỊ

Câu 1: Mẫu số liệu sau cho biết điểm kiểm tra giữa kì môn Toán của 10 bạn trong tổ 1 của lớp 10A:

8 9 7 10 8 9 10 9 8 7

Tính khoảng biến thiên của mẫu số liệu này.

Lời giải

Điểm thấp nhất, cao nhất của mẫu số liệu tương ứng là: 7; 10. Do đó, khoảng biến thiên là:

$$R = 10 - 7 = 3$$

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.

Câu 2: Mẫu số liệu sau cho biết số học sinh tham gia ủng hộ sách cũ và quần áo cũ cho học sinh vùng cao trong 9 ngày phát động chiến dịch “Món quà năm học mới” của Câu lạc bộ thiện nguyện trường THPT X:

37 87 30 56 120 78 96 88 92

Tìm khoảng tứ phân vị cho mẫu số liệu này.

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

$$\text{Do đó } Q_1 = (37 + 56) : 2 = 46,5$$

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

$$\text{Do đó } Q_3 = (92 + 96) : 2 = 94$$

Vậy khoảng tứ phân vị cho mẫu số liệu là: $\Delta_Q = Q_3 - Q_1 = 94 - 46,5 = 47,5$.



Câu 3: 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 4: 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 5: 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

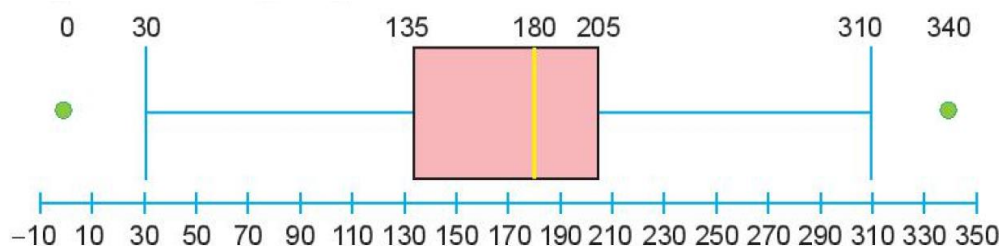
Tìm giá trị bất thường trong mẫu số liệu trên bằng cách sử dụng biểu đồ hộp.

Lời giải

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

$$\Delta_Q = 205 - 135 = 70$$

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 6: 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

Con đường B :

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 đường 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 .

II. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

PHƯƠNG PHÁP

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:

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

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 .

Chú ý: Có thể biến đổi công thức tính phương sai

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

Câu 7: 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



$$S_x^2 = \frac{(\bar{x}-9)^2 + (\bar{x}-10)^2 + 3(\bar{x}-11)^2 + 5(\bar{x}-12)^2 + \dots + 2(\bar{x}-19)^2}{100} \approx 3,957.$$

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

Câu 8: Sản lượng lúa các năm từ 2014 đến 2018 của hai tỉnh Thái Bình và Hậu Giang được cho ở bảng sau

Tỉnh \ Năm	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

a) Hãy tính độ lệch chuẩn của sản lượng lúa từng tỉnh.

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

Lời giải

a) Độ lệch chuẩn của sản lượng lúa từng tỉnh.

+/- Tỉnh Thái Bình:

$$\text{Số trung bình: } \bar{x}_1 = \frac{942,6 + 1030,4 + 1053,6 + 1061,9 + 1061,9}{5} = 1030,08.$$

Giá trị	Độ lệch	Bình phương độ lệch
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
Tổng		10231,06

Mẫu số liệu gồm 5 giá trị nên $n = 5$. Do đó phương sai là: $s_1^2 = \frac{10231,06}{5} \approx 2046,21$.

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

+/- Tỉnh Hậu Giang

$$\text{Số trung bình: } \bar{x}_2 = \frac{1204,6 + 1293,1 + 1231,0 + 1261,0 + 1246,1}{5} = 1247,16.$$

Giá trị	Độ lệch	Bình phương độ lệch
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
Tổng		4375,65

Mẫu số liệu gồm 5 giá trị nên $n = 5$. Do đó phương sai là: $s_2^2 = \frac{4375,65}{5} = 875,13$.

Độ lệch chuẩn là: $s_2 = \sqrt{875,13} \approx 29,58$.



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

Do $45,24 > 29,58$ nên độ lệch chuẩn của sản lượng lúa tỉnh Thái Bình lớn hơn tỉnh Hậu Giang, điều đó có nghĩa là sản lượng lúa của tỉnh Thái Bình trong các năm từ 2014 đến 2018 có độ phân tán cao hơn tỉnh Hậu Giang.

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

Câu 9: Số quần jeans bán được trong một quý ở một cửa hàng thời trang được thống kê ở bảng sau:

Size	26	27	28	29	30	31	32	Cộng
Tần số	128	105	119	16	8	15	9	400

Tìm phương sai và độ lệch chuẩn cho mẫu số liệu này.

Lời giải

Số trung bình của mẫu số liệu là

$$\bar{x} = \frac{26.128 + 27.105 + 28.119 + 29.16 + 30.8 + 31.15 + 32.9}{400} = 27,38.$$

Phương sai của mẫu số liệu là

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

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

Câu 10: Tiền thưởng của 50 cán bộ và nhân viên trong một công ty được cho ở bảng dưới đây:

Tiền thưởng	1	2	3	4	5	Cộng
Tần số	10	8	20	10	2	50

Tìm phương sai và độ lệch chuẩn cho mẫu số liệu này.

Lời giải

$$\bar{x} = \frac{1.10 + 2.8 + 3.20 + 4.10 + 5.2}{50} = 2,72.$$

Phương sai của mẫu số liệu là

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

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

Câu 11: Cho mẫu số liệu thống kê: $\{2, 4, 6, 8, 10\}$. Tính phương sai và độ lệch chuẩn của mẫu số liệu trên.

Lời giải

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10}{5} = 6.$$



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

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

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

Câu 12: 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

- Ta có:

$$\begin{aligned} \text{Số trung bình là } \bar{x} &= \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_k x_k) \\ &= \frac{1}{25} (1.30 + 1.35 + 1.39 + 1.41 + 1.45 + 1.48 + 1.50 + 1.51 + 1.54 + 1.58 + 1.60 + 3.61 + 2.65 + \\ &\quad 1.68 + 3.72 + 2.75 + 1.80 + 1.83 + 1.84) = 60,2. \end{aligned}$$

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

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

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

Câu 13: 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

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

Lời giải

- 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 14: 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 15: 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:



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 16: 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

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

III. PHÁT HIỆN SỐ LIỆU BẤT THƯỜNG HOẶC KHÔNG CHÍNH XÁC BẰNG BIỂU ĐỒ HỘP PHƯƠNG PHÁP

Trong mẫu số liệu, những giá trị quá lớn hay quá nhỏ so với các giá trị khác được gọi là **giá trị bất thường**.

Cách tìm giá trị bất thường trong mẫu số liệu:

- ✓ Xác định các tứ phân vị Q_1, Q_2, Q_3 của mẫu số liệu.
- ✓ Xác định khoảng tứ phân vị: $\Delta_Q = Q_3 - Q_1$.
- ✓ Tính $Q_1 - 1,5 \cdot \Delta_Q, Q_3 + 1,5 \cdot \Delta_Q$, khi đó: Các giá trị bé hơn $Q_1 - 1,5 \cdot \Delta_Q$ hay lớn hơn $Q_3 + 1,5 \cdot \Delta_Q$ được xem là các giá trị bất thường.

Câu 17: 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 18: 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 $Q_3 = 164$



Vậy $\Delta_Q = Q_3 - Q_1 = 164 - 154 = 10$

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

Câu 20: 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 $Q_1 - 1,5 \cdot \Delta_Q = 2$ và $Q_3 + 1,5 \cdot \Delta_Q = 26$ nên trong mẫu số liệu không có giá trị bất thường.

Câu 21: 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:

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

+) 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 22: 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

+) Từ mẫu số liệu ta tính được $Q_1 = 19$ và $Q_3 = 26$. Do đó khoảng tứ phân vị là:

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

+) Ta có $Q_1 - 1,5\Delta_Q = 8,5$ và $Q_3 + 1,5\Delta_Q = 36,5$ nên trong mẫu số liệu không có giá trị bất thường.

Câu 23: 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

Có hay không giá trị bất thường của mẫu số liệu?

Lời giải

Khi sắp xếp mẫu số liệu theo giá trị không giảm ta có giá trị thứ 15,16 lần lượt là 43 và 45.

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

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

$$\Rightarrow Q_3 = 48$$

$$\text{Ta có: } \Delta_Q = 48 - 43 = 5$$

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



$$Q_3 + 1,5.\Delta_Q = 48 + 1,5.5 = 55,5.$$

Vậy mẫu có giá trị bất thường là 56.

