

**CÁC SỐ ĐẶC TRƯNG.**

**ĐO ĐỘ PHÂN TÁN**

**Đ**

**14**

❶. Giáo viên Soạn: Văn Quý Vênh

❷. Tiết 43 – 45 . Ngày soạn 9/12/2024

**THUẬT NGỮ KIẾN THỨC, KĨ NĂNG**

• Khoảng biến thiên • Tính các số đặc trưng đo độ phân tán.

• Khoảng tứ phân vị • Biết ý nghĩa của các số đặc trưng đo độ phân tán.

• Phương sai • Phát hiện các giá trị bất thường sử dụng các công cụ toán học.

• Độ lệch chuẩn

Dưới đây là điểm trung bình môn học kỳ I của hai bạn An và Bình:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Toán | Vật lí | Hóa 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”?

Bài này sẽ giới thiệu một vài số đặc trưng như vậy.

**1. KHOẢNG BIẾN THIÊN VÀ KHOẢNG TỨ PHÂN VỊ**

Một cổ động viên của câu lạc bộ Everton, Anh đã thống kê điểm số mà hai 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:

**HĐ1:**

Leicester City: 41 81 44 47 52

Everton: 47 47 61 49 54.

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

**Giải**

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

Câu lạc bộ Everton có điểm cao nhất là 61 và điểm thấp nhất là 47 nên khoảng cách giữa điểm cao nhất và thấp nhất là 14.

Ta thấy  nên câu lạc bộ Everton thi đấu ổn định hơn.

Trong 5 mùa giải, điểm thấp nhất, cao nhất của Leicester City lần lượt là 41; 81 trong khi của Everton là 47; 61. Về trực quan, thành tích của Everton ổn định hơn Leicester City. Người ta có nhiều cách để đo sự ổn định này. Cách đơn giải nhất là dung hiệu số (Điểm cao nhất – Điểm thấp nhất). Giá trị này được gọi là 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 dung 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.

a) Điểm kiểm tra trung bình của hai tổ có như nhau không?

b) 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**

a) Điểm kiểm tra trung bình của hai tổ đều bằng 8.

b) Đối với Tổ 1: Điểm kiểm tra thấp nhất, cao nhất tương ứng là 7;9. Do đó, khoảng biến thiên là: .

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

Do  nên ta nói các bạn Tổ 1 học đều hơn các bạn Tổ 2.

Mẫu số liệu sau cho biết chiều cao (đơn vị cm) của các bạn trong tổ:

**Luyện tập 1.**

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: .

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.

Trong một tuần, nhiệt độ cao nhất trong ngày (đơn vị ) tại hai thành phố Hà Nội và Điện Biên được cho như sau:

**HĐ2:**

Hà Nội: 23 25 28 28 32 33 35

Điện Biên: 16 24 26 26 26 27 28

a) Tính các 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  cho mỗi mẫu số liệu. Có thể dùng hiệu này để đo độ phân tán của mẫu số liệu không?

**Giải**

a) Ở Hà Nội, nhiệt độ thấp nhất, cao nhất trong ngày tương ứng là: 23; 35. Do đó, khoảng biến thiên là .

Ở Điện Biên, nhiệt độ thấp nhất, cao nhất trong ngày tương ứng là 16; 28. Do đó, khoảng biến thiên là 

b) Số 16 làm cho khoảng biến thiên về nhiệt độ tại Điện Biên lớn hơn.

c) Ở Hà Nội: 23 25 28 28 32 33 35.

Mẫu số liệu gồm 7 giá trị nên số trung vị là .

Nửa số liệu bên trái gồm 23; 25; 28 gồm 3 giá trị nên .

Nửa số liệu bên phải gồm 32; 33; 35 gồm 3 giá trị nên .

Khi đó, .

Ở Điện Biên: 16 24 26 26 26 27 28.

Mẫu số liệu gồm 7 giá trị nên số trung vị là .

Nửa số liệu bên trái gồm 16; 24; 26 gồm 3 giá trị nên .

Nửa số liệu bên phải gồm 26; 27; 28 gồm 3 giá trị nên .

Khi đó, .

Ta có thể dùng hiệu này để đo độ phân tán của mẫu số liệu.

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



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 .

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

Do đó, .

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

Do đó, .

Vậy khoảng tứ phân vị cho mẫu số liệu là: .

Mẫu số liệu sau đây cho biết số bài hát ở mỗi album trong bộ sưu tập của An:

**Luyện tập 2.**

12 7 10 9 12 9 10 11 10 14.

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

**Giải**

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

7 9 9 10 10 10 11 12 12 14.

Mẫu số liệu gồm 10 giá trị nên trung vị là .

Nửa số liệu bên trái là 7; 9; 9 ; 10 gồm 4 giá trị, hai phần tử chính giữa là 9; 9.

Do đó, .

Nửa số liệu bên phải là 11; 12; 12; 14 gồm 4 giá trị, hai phần tử chính giữa là 12; 12.

Do đó, .

Vậy khoảng tứ phân vị cho mẫu số liệu là: .

**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 , nếu gọi số trung bình là  thì với mỗi giá trị , độ lệch của nó so với giá trị trung bình là .

* Phương sai là giá trị .
* Căn bậc hai của phương sai, , được gọi là độ lệch chuẩn.

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

.

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 học:

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 

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 |
| 10 | 40 – 43 = - 3 | 9 |
| Tổng | | 26 |

Mẫu số liệu gồm 5 giá trị nên . Do đó phương sai là 

Độ lệch chuẩn là: 

**Luyện tập 3.** 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  đến điểm . Kết quả đo như sau:



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

Hãy tính 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?

**Giải**

Số trung bình của mẫu số liệu là 

Ta có bẳng sau:

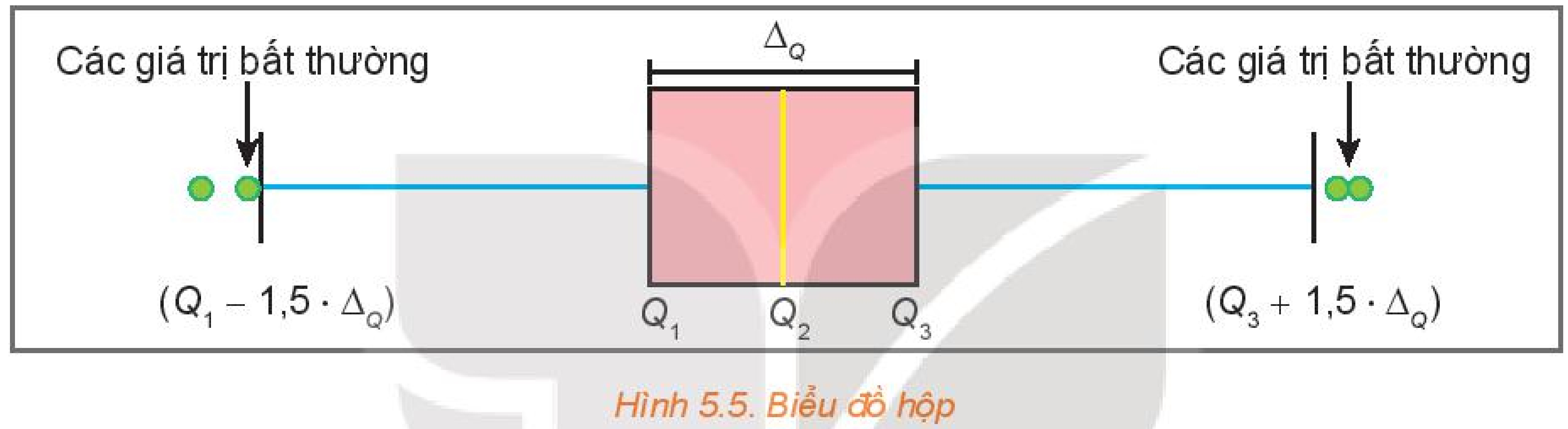
|  |  |  |
| --- | --- | --- |
| Giá trị | Độ lệch | Bình phương độ lệch |
| 0,398 |  |  |
| 0,399 |  |  |
| 0,408 |  |  |
| 0,410 |  |  |
| 0,406 |  |  |
| 0,405 |  |  |
| 0,402 |  |  |
| Tổng | |  |

Mẫu số liệu gồm 7 giá trị nên . Do đó phương sai là 

Độ lệch chuẩn là: .

**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.



Các giá trị lớn hơn  hoặc bé hơn  được xem là giá trị bất thường.

**Ví dụ 4.**  Hàm lượng Natri (đơn vị mg) trong 100 g một số loại ngũ cốc được cho như sau:



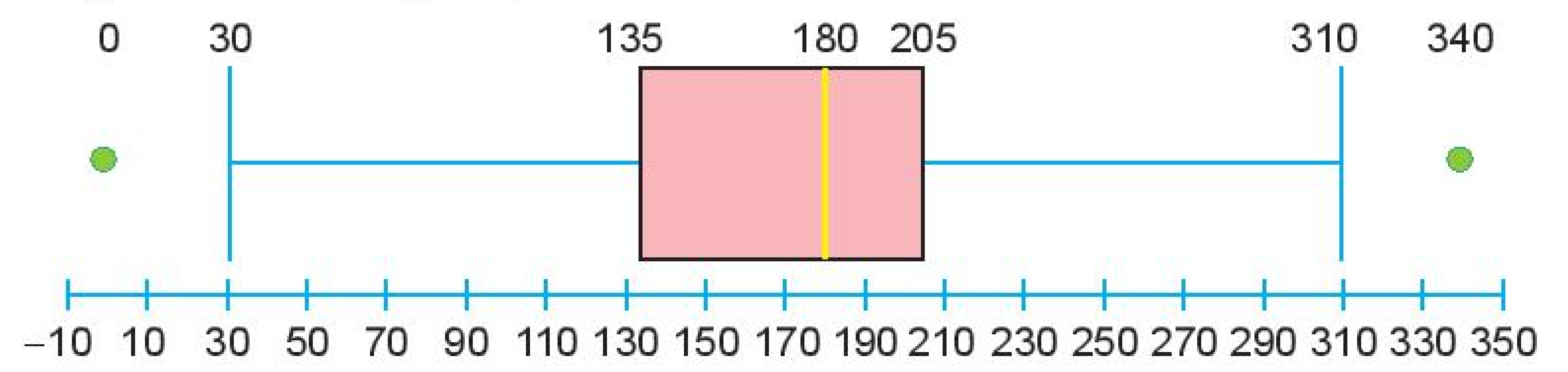
Tìm giá trị bất thường trong mẫu số liệu trên bằng cách sử dụng biểu đồ hộp.

**Giải**

Từ mẫu số liệu ta tính được  và . Do đó, khoảng tứ phân vị là:



Biểu đồ hộp cho mẫu số liệu này là:



Ta có  và  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).

**Luyện tập 4.** Một mẫu số liệu có tứ phân vị thứ nhất là 56 và từ 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.

**Giải**

Theo đề bài ta có  và , do đó, khoảng tứ phân vị là:



Ta có  và  nên cả hai giá trị 10 và 100 đều không phải hai giá trị bất thường.

**BÀI TẬP**

**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.

**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; ; giá trị lớn nhất bằng 205.

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

b) Chỉ ra hai giá trị sao cho có  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  nên có 12 thành phố có thuế thuốc lá lớn hơn 36.

Vì vậy, tỉ lệ thành phố có thuế thuốc lá lớn hơn 36 là: 

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

c) Khoảng tứ phân vị của mẫu số liệu là .

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



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:



Khoảng biến thiên là .

Ta có: ; ; 

Khoảng tứ phân vị là .

Độ lệch chuẩn là .

**5.16.** Tỉ lệ thất nghiệp ở một số quốc gia vào năm 2007 (đơn vị %) được cho như sau:



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

**Giải**

Từ mẫu số liệu ta tính được  và . Do đó, khoảng tứ phân vị là:



Ta có  và  nên trong mẫu số liệu trên không có giá trị bất thường.