**Bài 4. Các số đặc trưng đo mức độ phân tán của mẫu số liệu chất**

Từ khoá: **Khoảng biến thiên; Khoảng tứ phân vị; Phương sai: Độ lệch chuẩn.**

****

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

Thời gian hoàn thành bài chạy 5 km (tính theo phút) của hai nhóm thanh niên được cho ở bảng sau:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nhóm 1 | 30 | 32 | 47 | 31 | 32 | 30 | 32 | 29 | 17 | 29 | 32 | 31 |
| Nhóm 2 | 32 | 29 | 32 | 30 | 32 | 31 | 29 | 31 | 32 | 30 | 31 | 29 |

a) Hãy tính độ chênh lệch giữa thời gian chạy của người nhanh nhất và người chậm chất trong từng nhóm.

b) Nhóm nào có thành tích chạy đồng đều hơn?



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

*X1*$ \leq $ *X2* $\leq $ *……..* $\leq $ *Xn*

* **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 = Xn- X1*

* **Khoảng tứ phân vị,** kí hiệu là $∆$Q, là hiệu giữa Q3 và Q1 , tức là:

$∆$*Q = Q3-Q1*

Trong Độ chênh lệch giữa kết quả cao nhất và kết quả thấp nhất chính là khoảng biển thiên của kết quả các lần chạy của từng nhóm.

***Ví dụ 1***

Hãy tính khoảng biền thiên và khoảng tứ phân vị của mẫu số liệu: 10; 20; 3; 1; 3; 4; 7 ; 4; 9

**Giải**

Xét mẫu số liệu đã sắp xếp là: 1; 3; 3; 4; 4; 7; 9; 10; 20.

* Khoảng biến thiên của mẫu số liệu là: *R* = 20 - 1 = 19.
* Cỡ mẫu là n = 9 là số lẻ nên giá trị tứ phân vị thứ hai là: Q2 = 4.
* Tứ phân vị thứ nhất là trung vị của mấu: 1; 3; 3; 4. Do đó Q1 = 3.
* Tứ phân vị thứ ba là trung vị của mẫu: 7; 9; 10; 20. Do đó Q3= 9,5.
* Khoảng tứ phân vị của mẫu là: $∆$Q = 9,5 - 3 = 6,5.

**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 trung cho độ phân tán của một nứa các số liệu, có giá trị thuộc đoạn*

*từ Q1 đến Q3 trong mẫu.*

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

Trong Có sự khác biệt lớn nếu sử dụng khoảng biến thiên để so sánh độ chênh lệch kết

quả giữa hai nhóm. Nhưng nều sử dụng khoảng tứ phân vị thì thấy sự chênh lệch thời gian

chạy của đa số các thanh niên ở hai nhóm là như nhau.

Hãy tìm khoảng biên thiên và khoảng tứ phân vị của các mẫu số liệu sau:

a) 10; 13; 15; 2; 10; 19; 2; 5; 7. b) 15; 19; 10; 5; 9; 10; 5; 9; 10; 1; 2; 5; 15.

Dưới đây là bảng số liệu thống kê của Biểu đồ nhiệt độ trung bình các tháng trong năm 2019 của hai tỉnh Lai Châu và Lâm Đồng (được đề cập đến ở hoạt động khởi độngcủa bài học).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tháng | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Lai châu | 14,8 | 18,8 | 20,3 | 23,5 | 24,7 | 24,2 | 23,6 | 24,6 | 22,7 | 21,0 | 18,6 | 14,2 |
| Lâm đồng | 16,3 | 17,4 | 18,7 | 19,8 | 20,2 | 20,3 | 19,5 | 19,3 | 18,6 | 18,5 | 17,5 | 16,0 |

a) Hãy tìm khoảng biến thiên và khoảng tứ phân vị của nhiệt độ trung bình mỗi tháng của tỉnh Lai Châu và Lâm Đồng.

b) Hãy cho biết trong một năm, nhiệt độ ở địa phương nào ít thay đổi hơn.

**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 thề, phần tử x trong mẫu là giá trị

ngoại lệ nếu x > Q3 + 1,5$∆$Q, hoặc x <$∆$1 - 1,5$∆$Q

Trong Ví dụ 1, Q3+ 1,5$∆$Q = 9,5 + 1,5. 6,5 = 19,25 và Q1= 1,5$∆$Q = 3- 1,5.6,5 = -6,75

nên mẫu có một giá trị ngoại lệ là 20.

Sự xuất hiện của các giá trị ngoại lệ làm cho số trung bình và phạm vi của mẫu thay đổi lớn. Do

đó, khi mẫu có giá trị ngoại lệ, người ta thường sử dụng trung vị và khoảng tứ phân vị đo

mức độ tập trung và mức độ phân tán của đa số các phần tử trong mẫu số liệu.

Hãy tìm giá trị ngoại lệ của mẫu số liệu: 37; 12; 3; 9; 10; 9; 12; 3; 10.

**2. Phương sai và độ lệch chuẩn**

Hai cung thủ *A* và *B* đã ghi lại kết quả từng lần bắn của mình ở bảng sau:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cung thủ A | 8 | 9 | 10 | 7 | 6 | 10 | 6 | 7 | 9 | 8 |
| Cung thủ B | 10 | 6 | 8 | 7 | 9 | 9 | 8 | 7 | 8 | 8 |

a) Tính kết quả trung bình của mỗi cung thủ trên.

b) Cung thủ nào có kết quả các lần bắn ổn định hơn?

Ngoài khoảng biền thiên và khoảng tứ phân vị, người ta cũng sử dụng phương sai và độ lệch chuẩn để đo độ phân tán của mẫu số liệu.

 Giả sử ta có một mẫu số liệu là X1,X2 ,…..,Xn

* **Phương sai** của mẫu số liệu này, kí hiệu là S2, được tính bởi công thức:



trong đó $\overbar{X}$ 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 ở trên thành:

Trong thống kê, người ta cũng quan tâm đến phương sai hiệu chỉnh, kí hiệu là $\hat{S^{2}}$, được tính

bởi công thức:



**Ví dụ 2**

Hãy tính phương sai và độ lệch chuẩn của mỗi mẫu số liệu ghi kết quả các lần bắn của

từng cung thủ trong

**Giải**

Số trung bình của kết quả các lần bắn của cung thủ A là:

(8 + 9 + 10 + 7+ 6 + 10 + 6 + 7+ 9 + 8): 10 = 8.

Số trung bình của kết quả các lần bắn của cung thủ B là:

(10 +6 + 8 + 7+9 +9 + 8+ 7 +8 +8): 10 = 8.

Phương sai mãu số liệu của cung thủ A là:

SA2 = $\frac{1}{10}$( 82+ 92 + 102 + 72 + 62 + 102 + 62 + 72 + 92 + 82 ) - 82 = 2.

Độ lệch chuẩn mẫu số liệu của cung thủ A là: SA = $\sqrt{S\_{A}^{2}}$ =$\sqrt{2}≈$1,41

Phương sai mẫu số liệu của cung thủ B là

SB2 = $\frac{1}{10}$( 102+ 62 + 82 + 72 + 92 + 92 + 82 + 72 + 82 + 82 ) - 82 =1,2

Độ lệch chuẩn mẫu số liệu của cung thủ B là: :$ S\_{B} \sqrt{S\_{B}^{2}}$ =$\sqrt{1,2}≈$1,10

**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 cùng lớn thì các giá trị của mẫu càng cách xa nhau (có độ phân tán lớn).*

Trong Kết quả các lần bắn của hai cung thủ có cùng khoảng biến thiên và khoảng tứ phân vị. Tuy nhiên, nếu so sánh bằng phương sai hoặc độ lệch chuẩn thì kết quả của cung thủ A có độ phân tán cao hơn cung thủ B. Do đó, cung thủ 3 bắn ôn định hơn cung thủ A.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Giá trị | X1 | X2 | …. | XN |
| Tần số | n1 | n2 | ….. | Nn |

Khi đó, công thức tính phương sai trở thành:



trong đó n = n1, + n2 +.. + nn

Có thể biến đổi công thức tính phương sai trên thành:



***Ví dụ 3***

Điều tra một số học sinh về số cái bánh chưng mà gia đình mỗi bạn tiêu thụ trong dịp Tết Nguyên đán, kết quả được ghi lại ở bảng sau:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Số cái bánh chưng | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Số gia đình | 5 | 7 | 10 | 8 | 5 | 4 | 1 |

Số trung bình của mẫu số liệu trên là:

$\overbar{x}$=$\frac{1}{40}$(5.6+ 7.7 + 8.10 + 9.8 + 10.5 + 11.4 + 15.1) =8.5

Phương sai của mẫu số liệu trên là:

S2 = $\frac{1}{40}$(5.62+ 7.72 + 8.102 + 9.82 + 10.52 + 11.42 + 15.12) - 8,52 = 3,25

Độ lệch chuẩn của mẫu số liệu là:

S = $\sqrt{S^{2}}$= $\sqrt{3,25}$s $≈ $1,80.

Bảng dưới đây thông kê tổng số giờ năng 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 |

(Nguồn: Tổng cục thống kê)

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

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

BÀI TẬP

1. Hãy chọn ngẫu nhiên trong lớp ra 5 bạn nam và 5 bạn nữ rồi đo chiều cao các bạn đó, So sánh xem chiều cao của các bạn nam hay các bạn nữ đồng đều hơn.

2. Hãy tim độ lệch chuẩn, khoảng biến thiên, khoảng tứ phân vị và các giá trị ngoại lệ của các mẫu số liệu sau:

a) 6, 8; 3; 4; 5; 6; 7; 2; 4. b) 13; 37; 64; 12; 26; 43; 29; 23.

3. Hãy tìm độ lệch chuẩn, khoảng biến thiên và khoảng tứ phân vị của các mẫu số liệu sau:

a)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Giá trị | -2 | -1 | 0 | 1 | 2 |
| Tần số | 10 | 20 | 30 | 20 | 10 |

b)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Giá trị | 0 | 1 | 2 | 3 | 4 |
| Tần số | 0,1 | 0,2 | 0,4 | 0,2 | 0,1 |

4. Hãy so sánh số trung bình, phương sai và độ lệch chuẩn của ba mẫu số liệu sau:

Mẫu 1: 0,1; 0,3, 0,5; 0,5; 0,3; 0,7.

Mẫu 2: 1,1; 1,3; 1,5; 1,5; 1,3; 1,7.

Mãu 3: 1 ; 3 ; 5; 5; 3; 7.

5. Sản lượng lúa các năm từ 2014 đến 20l8 của hai tinh Thái Bình và Hậu Giang được cho ở bảng sau (đơn vị: nghìn tấn):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TỉnhNăm text here | **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 |

 (Nguồn: Tổng cục thống kê)

a) Hãy tính độ lệch chuẩn và khoảng biến thiên của sản lượng lúa từng tinh.

b) Tinh nào có sản lượng lúa ốn định hơn? Tại sao?

6. Kết quả điều tra mức lương hằng tháng của một số công nhân của hai nhà máy A và B được

cho ở bảng sau (đơn vị: triệu đồng):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Công nhân nhà máy A | 4 | 5 | 5 | 47 | 5 | 6 | 4 | 4 |  |
| Công nhân nhà máy B | 2 | 9 | 9 | 8 | 10 | 9 | 9 | 11 | 9 |

a) Hãy tìm số trung binh, mốt, tứ phân vị và độ lệch chuẩn của hai mẫu số liệu lấy từ nhà máy A và nhà máy B.

b) Hãy tìm các giá trị ngoại lệ trong mỗi mấu số liệu trên. Công nhân nhà máy nào có mức lương cao hơn? Tại sao?