**Bài 3. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu**

****Từ khoá: Số trung bình cộng; Trung vị; Tứ phân vị; Mốt.

**Đề nhanh chóng nắm bắt được những đặc điểm, những thông tin quan trọng trong một mẫu số liệu người ta thường hay tính các số đặc trưng của mẫu số liệu đó.**

**Trong bài này chúng ta sẽ tìm hiểu các số đặc trưng đo xu thế trung tâm của mẫu số liệu là: số trung bình; trung vị; tứ phân vị và mốt.**

**1. Số trung bình**

Điểm số bài kiềm tra môn Toán của các bạn trong Tổ 1 là 6; 10; 6, 6; 7; 10, còn của các bạn Tổ 2 là 10; 6; 9; 9; 8; 9, Theo em, tổ nào có kết quả kiểm tra tốt hơn? Tại sao?

* Giá sử ta có một mẫu số liệu là x1,x2,….xn

**Sô trung bình** (hay **số trung bình cộng**) của mẫu số liệu này, kí hiệu là $\overbar{X}$, được tính bởi

công thức

$$\overbar{x}=\frac{x\_{1}+x\_{2}+...x\_{n}}{n}$$

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Giá trị | x1 | x2 | …. | Xk |
| Tần số | n1 | n2 | … | nk |

Khi đó, công thức tính số trung bình trở thành

$$\overbar{x}=\frac{n\_{1}x\_{1}+n\_{2}x\_{2}+...n\_{k}x\_{n}}{n}$$

trong đó n =$n\_{1}$ +n2 +... + nk. Ta gọi n là **cỡ mẫu**.

**Chú ý:** Nếu kí hiệu $f\_{k}=\frac{nk}{n} $là tần số tương đối (hay còn gọi là tần suất) của x, trong

mẫu số liệu thì số trung bình còn có thể biểu diễn là: $\overbar{x}$ =$f\_{1}x\_{1}+f\_{2}x\_{2}+...f\_{k}x\_{k}$

***Ví dụ 1***

Một của hàng bán xe đạp thống kê số xe bán được hằng tháng trong năm 2021 ở bảng sau:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tháng | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Số xe | 10 | 8 | 7 | 5 | 8 | 22 | 28 | 25 | 20 | 10 | 9 | 7 |

a) Hãy tính số xe trung bình của hàng bán được mỗi tháng trong năm 2021.

b) Hãy so sánh hiệu quả kinh doanh trong quý III của của hàng với 6 tháng đầu năm 2021.

**Giải**

a) Số xe trung bình của hàng bán được mỗi tháng trong năm 2021 là:

$\frac{1}{12}$(10+8 + 7 +5 + 8 + 22 + 28 + 25 + 20 + 10+9+7) = 13,25 (xe).

b) Số xe trung binh bán được trong 6 tháng đầu năm là:

$\frac{1}{6}$ (10 +8+7 +5 + 8 + 22) = 10 (xc).

Số xe trung bình bán được trong quý III của năm là:

$\frac{1}{3}$ (28 + 25 + 20) = $\frac{73}{3}$ =$≈ $24,33 (xe).

Như vậy hiệu quả kinh doanh của của hàng trong quý III cao hơn trong 6 tháng đầu năm.

**Ý nghĩa của số trung bình**

*Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu, Nó là một số đo xu thể trung tâm của mẫu đó.*

Ví dụ như trong , vì điểm trung bình của các bạn Tổ 1 là 7,83 và của các bạn Tổ 2 là 8,5 nên ta có thể cho rằng nới chung các bạn Tố 2 học Toán tốt hơn các bạn Tổ 1.

Thời gian chạy 100 mét (đơn vị: giây) của các bạn học sinh ở hai nhóm A và B được ghí lại ở bảng sau:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nhóm A | 12,2 | 13,5 | 12,7 | 13,1 | 12,5 | 12,9 | 13,2 | 12,8 |
| Nhóm B | 12,1 | 14,3 | 13,2 | 12,9 | 13,7 |  |  |  |

Nhóm nào có thành tích chạy tốt hơn?

Số bàn thắng mà một đội bóng ghi được ở mỗi trận đấu trong một mùa giải được thống kê lại ở bảng sau:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Số bàn thắng | 0 | 1 | 2 | 3 | 4 | 6 |
| Số trận | 5 | 10 | 5 | 3 | 2 | 1 |

Hãy xác định số bàn thăng trung bình đội đó ghi được trong một trận đầu của mùa giải.

**2. Trung vị và tứ phân vị**

Bảng sau thống kê số sách mỗi bạn học sinh Tổ 1 và Tổ 2 đã đọc ở thư viện trường trong một tháng:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tổ 1 | 3 | 1 | 2 | 1 | 2 | 2 | 3 | 25 | 1 |
| Tổ 2 | 4 | 5 | 4 | 3 | 3 | 4 | 5 | 4 |  |

a) Trung bình mỗi bạn Tổ 1 và mỗi bạn Tổ 2 đọc bao nhiêu quyển sách ở thư viện trường trong tháng đó?

b) Em hãy thảo luận với các bạn trong nhóm xem tô nào chăm đọc sách ở thư viện hơn.

Trong tình huống trên ta thấy việc sử dụng số trung bình để so sánh độ chăm đọc sách thư chất của viện của hai tổ là không phù hợp, Điều này là do Tổ 1 có một số liệu quá lớn so với các số liệu còn lại. Trong tình huống như vậy, để so sánh độ chăm đọc sách giữa hai tổ, người ta thường dùng một số đặc trưng khác của mẫu số liệu, gọi là **trung vị**, được định nghĩa như sau:

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

X1$\leq $X2 $\leq $…$\leq $xn

**Trung vị** của mẫu, kí hiệu là Me, là giá trị ở chính giữa dãy X1, X2,….Xn . Cụ thể:

* Nếu n = 2k + 1, k $\in $ N, thì trung vị của mẫu Me = xk+1.
* Nếu n = 2k, k $\in $ N, thì trung vị của mẫu Me = $\frac{1}{2}$(xk + xk+1.).

Ý nghĩa của trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu, Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có it nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

***Ví dụ 2***

a) Tính các trung vị của số sách các bạn ở Tổ 1 và số sách các bạn ở Tổ 2 đã đọc.

b) Sử dụng trung vị, hãy so sánh xem các bạn ở tổ nào đọc nhiều sách ở thư viện hơn.

**Giải**

a) Sắp xếp số sách mỗi bạn Tố 1 đã đọc theo thứ tự không giảm, ta được dãy.

1; 1; 1; 2; 2; 2; 3; 3; 25.

Vi cỡ mẫu bằng 9 nên trung vị của Tổ 1 là số liệu thứ 5 của dãy trên, tức là Me = 2.

Sắp xếp số sách mỗi bạn Tố 2 đã đọc theo thứ tự không giảm, ta được dãy:

3, 3, 4, 4, 4, 5; 5;

Vì cỡ mẫu bằng 8 nên trung vị của Tổ 2 là trung bình cộng của số liệu thứ 4 và thứ 5 của dãy trên, tức là Me = $\frac{1 }{2 } $(4 + 4) = 4.

b) Nếu so sánh theo trung vị thi các bạn Tổ 2 đọc nhiều sách ở thư viện hơn các bạn Tổ 1.

***Ví dụ 3***

Khi kiểm tra ngẫu nhiên một số công nhân trong một xí nghiệp, người ta thống kê lại độ tuổi của họ ở bảng sau:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tuổi** | 25 | 26 | 27 | 29 | 31 | 34 |
| **Số công nhân** | 4 | 9 | 8 | 3 | 1 | 1 |

Tìm trung vị và trung bình cộng của mẫu số liệu trên.

**Giải**

Cỡ mẫu là n = 26. Khi sắp xếp độ tuổi các công nhân theo thứ tự không giảm thì số liệu thứ 13 và 14 lần lượt là 26 và 27. Vậy

Me = $\frac{1 }{2 }$(26+27)=26,5

Số trung bình cộng của mẫu là

 $\overbar{X}$= $\frac{1 }{26 } $(25.4 + 26.9 + 27,8 + 29.3 + 31 + 34) = 27.

Hãy tìm trung vị của các số liệu

Cân nặng của 20 vận động viên môn vật của một câu lạc bộ được ghi lại ở bảng sau:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 50 | 56 | 57 | 62 | 58 | 52 | 66 | 61 | 54 | 61 |
| 64 | 69 | 52 | 65 | 58 | 68 | 67 | 56 | 59 | 54 |

Đề thuận tiện cho việc luyện tập, ban huấn luyện muốn xếp 20 vận động viên trên thành 4 nhóm, mỗi nhóm gồm 25% số vận động viên có cân nặng gần nhau. Bạn hãy giúp ban huấn luyện xác định các ngưỡng cân nặng để phân nhóm mỗi vận động viên.

Trung vị chia mẫu thành hai phần, Trong thực tề người ta cũng quan tâm đền trung vị của mỗi phần đó. Ba trung vị này được gọi là **tứ phân vị** của mẫu.

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

X1$\leq $X2 $\leq $…$\leq $xn

**Tứ phân vị** của một mẫu số liệu gồm ba giá trị, gọi là tứ phân vị thứ nhất, thứ hai và thứ ba (lần lượt kí hiệu là Q1, Q2,Q3). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

* Giá trị tứ phân vị thứ hai Q2 , chính là số trung vị của mẫu.
* Giá trị tử phân vị thứ nhất Q1, là trung vị của nửa số liệu đã sắp xếp bên trái Q2 (không bao gồm Q2, nếu n lẻ).
* Giá trị tử phân vị thứ ba Q3, là trung vị của nửa số liệu đã sắp xếp bên phải Q2,(không bao gồm Q2 nếu n lẻ).

Ý nghĩa của tứ phân vị

*Các điểm tứ phân vị Q1, Q2,Q3, chia mẫu số liệu đâ xắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần chứa khoảng 25% tổng số liệu đã thu thập được.*

*Tứ phân vị thứ nhất Q1, còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liện phía trên. Tứ phân vị thứ ba Q3 còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu phía trên*

(đéo save được)

***Ví dụ 4***

*Tìm tứ phân vị của các mẫu số liệu sau:*

*a) 5; 13; 5; 7; 10; 2; 3. b) 2; 3; 10; 13; 5; 15; 5; 5; 7.*

**Giải**

*a) Sắp xếp lại mẫu số liệu theo thứ tự không giảm, ta được: 2; 3; 5; 5; 7; 10; 13*.

* Vì cỡ mẫu là n = 7, là số lẻ, nên giá trị tứ phân vị thứ hai là Q2= 5.
* Tứ phân vị thứ nhất là trung vị của mẫu: 2; 3; 5. Do đó Q1 = 3.
* Tứ phân vị thứ ba là trung vị của mẫu: 7; 10; 13. Do đó Q3 = 10.

b) Sắp xếp lại mẫu số liệu theo thứ tự không giảm, ta được: 2; 3; 5; 5; 7; 10; 13;15 ; là

* Vì cỡ mẫu là n = 8, là số chẵn, nên giá trị tứ phân vị thứ hai là

Q2 =$ \frac{1}{2} $(5+7)=6

* Tứ phân vị thứ nhất là trung vị của mẫu: 2; 3; 5; 5. Do đó Q1 = 4.
* Tứ phân vị thứ ba là trung vị của mẫu: 7; 10; 13; 15. Do đó Q3 = 11,5.

Trong ví dụ trên, khoảng cách giữa Q1, và Q3, nhỏ hơn khoảng cách giữa O2 và O3. Ta nói mật độ số liệu ở bên trái Q2 cao hơn ở bên phải Q2.

Hãy tìm 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;1;2;5;15

**3. Mốt**

Một cửa hàng kinh doanh hoa thống kê số hoa hồng bán được trong ngày 14 tháng 2 theo loại hoa và thu được bảng tần số sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Loại hoa | Hồng bạch | Hồng nhung | Hồng vàng | Hồng vàng |
| Số bông bán được | 120 | 230 | 180 | 150 |

Của hàng nên nhập loại hoa hồng nào nhiều nhất để bán trong ngày 14 tháng 2 năm tiếp

theo? Tại sao?

Cho một mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu và kí hiệu là Mo.

Ví dụ 5

Số vụ va chạm giao thông mỗi ngày tại một ngã tư được ghi lại trong bảng tần số sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Số vụ va chạm | 0 | 1 | 2 | 3 | 4 |
| Số ngày | 12 | 17 | 6 | 4 | 1 |

Tìm mốt của mẫu số liệu trên.

**Giải**

Số ngày có 1 vụ va chạm là 17, lớn hơn số ngày có 0, 2, 3, 4 vụ va chạm. Do đó mẫu số liệu trên có Mo = 1.

**Ý nghĩa của mốt**

*Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.*

**Chú ý:** Một mẫu số liệu có thể có nhiều một, Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

Hãy tìm mốt của số liệu điểm kiểm tra của các bạn Tổ 1

**BÀI TẬP**

1. Hãy tìm số trung bình, tứ phân vị và mốt của các mẫu số liệu sau:

a) 23; 41; 71; 29; 48; 45; 72; 41. b) 12; 32; 93; 78; 24; 12; 54; 66; 78.

2. Hãy tìm số trung bình, tứ phân vị và mốt của các mẫu số liệu sau:

a)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Giá trị | 23 | 25 | 28 | 31 | 33 | 37 |
| Tần số | 6 | 8 | 10 | 6 | 4 | 3 |

b)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Giá trị | 0 | 2 | 4 | 5 |
| Tần số tương đối | 0,6 | 0,2 | 0,1 | 0,1 |

3. An lấy ra ngẫu nhiên 3 quả bóng từ một hộp có chứa nhiều bóng xanh và bóng đỏ. An đếm xem có bao nhiêu bóng đỏ trong 3 bóng lấy ra đó rồi trả bóng lại hộp. An lặp lại phép thử trên 100 lần và ghi lại kết quả ở bảng sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Số bóng đỏ | 0 | 1 | 2 | 3 |
| Số lần | 10 | 30 | 40 | 20 |

Hãy tìm số trung bình, tứ phân vị và mốt của bảng kết quả trên.

4. Trong một cuộc thi nghề, người ta ghi lại thời gian hoàn thành một sản phẩm của một số thí sinh ở bảng sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Thời gian (đơn vị :phút) | 5 | 6 | 7 | 8 | 35 |
| Số thí sinh | 1 | 3 | 5 | 2 | 1 |

a) Hãy tìm số trung bình, tứ phân vị và mốt của thời gian thi nghề của các thí sinh trên.

b) Năm ngoái, thời gian thi của các thí sinh có số trung bình và trung vị đều bằng 7. Bạn hãy so sánh thời gian thi nói chung của các thí sinh trong hai năm.

5. Bác Dũng và bác Thu ghi lại số cuộc điện thoại mà mỗi người gọi mỗi ngày trong 10 ngày được lựa chọn ngẫu nhiên từ tháng 01/2021 ở bảng sau:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bác Dũng | 2 | 7 | 3 | 6 | 1 | 4 | 1 | 4 | 5 | 1 |
| Bác Thu | 1 | 3 | 1 | 2 | 3 | 4 | 1 | 2 | 20 | 2 |

a) Hãy tìm số trung bình, tứ phân vị và mốt của số cuộc điện thoại mà mỗi bác gọi theo số liệu trên.

b) Nếu so sánh theo số trung bình thì ai có nhiều cuộc điện thoại hơn?

c) Nếu so sánh theo số trung vị thì ai có nhiều cuộc điện thoại hơn?

d) theo bạn, dùng số trung bình hay số trung vị để so sánh xem ai có nhiều cuộc gọi điện thoại hơn mỗi ngày?

6. Tổng số điểm mà các thành viên đội tuyển Olympic Toán quốc tế (IMO) của Việt Nam đạt được trong 20 kì thi được cho ở bảng sau:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Năm | Tổng điểm | Năm | Tổng điểm | Năm | Tổng điểm | Năm | Tổng điểm |
| 2020 | 150 | 2015 | 151 | 2010 | 133 | 2005 | 143 |
| 2019 | 177 | 2014 | 157 | 2009 | 161 | 2004 | 196 |
| 2018 | 148 | 2013 | 180 | 2008 | 159 | 2003 | 172 |
| 2017 | 155 | 2012 | 148 | 2007 | 168 | 2002 | 166 |
| 2016 | 151 | 2011 | 113 | 2006 | 131 | 2001 | 139 |

 (Nguồn: https: Vimo-official.org)

Có ý kiến cho rằng điểm thi của đội tuyển giai đoạn 2001 - 2010 cao hơn giai đoạn 2011 - 2020. Hãy sử dụng số trung bình và trung vị để kiểm nghiệm xem ý kiến trên có đúng không.

**7.** Kết quả bài kiểm tra giữa kì của các bạn học sinh lớp 10A, 10B, 10C được thống kê các

biểu đồ dưới đây.

a) Hãy lập bảng thống kê số lượng học sinh theo điểm số ở mỗi lớp.

b) Hãy so sánh điểm số của học sinh các lớp đó theo số trung bình, trung vị và mốt.

**Bạn có biết?**

Các số đặc trưng cho ta thông tin về quy luật phân phối của mẫu số liệu, Mẫu số liệu được gọi là **đối xứng** nếu trung bình của mẫu và trung vị bằng nhau, lệch trái nếu trung bình của mẫu nhỏ hơn trung vị; lệch phải nêu trung bình của mẫu lớn hơn trung vị.

Ví dụ trong Bài tập 7, điểm của lớp 10A là lệch trái, điểm của lớp 10B là đối xứng và điểm của lớp 10C là lệch phải.