

IIT Madras ONLINE DEGREE

Statistics for Data Science -1

Lecture 3.3: Describing Numerical Data- Median and Mode

Usha Mohan

Indian Institute of Technology Madras

Median

Another frequently used measure of center is the median.

Essentially, the median of a data set is the number that divides the bottom 50% of the data from the top 50%.

Definition

The median of a data set is the middle value in its ordered list.

Steps to obtain median

Arrange the data in increasing order. Let n be the total number of observations in the dataset.

- 1. If the number of observations is odd, then the median is the observation exactly in the middle of the ordered list, i.e. $\frac{n+1}{2}$ observation
- 2. If the number of obsevations is even, then the median is the mean of the two middle observations in the ordered list, i.e. mean of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n = 7 odd,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n = 7 odd, median is the $\frac{n+1}{2} =$

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order
 - 2, 3, 5, 6, 7, 7, 12
 - 1.2 n = 7 odd, median is the $\frac{n+1}{2} = \frac{8}{2} = 4^{th}$ observation,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105
 - $2.2 \ n = 7 \text{ odd},$

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5, 6, 7, 7, 105
 - 2.2 n = 7 odd, median is the $\frac{n+1}{2} = \frac{8}{2} = 4^{th}$ observation,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 105

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2. 3. 5. 6. 7. 7. 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 105
 - 3.2 n = 6 even,

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 105
 - 3.2 n=6 even, median is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 105
 - 3.2 n=6 even, median is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation $=\frac{5+6}{2}=$

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 12
 - 1.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 2. 2, 105, 5, 7, 6, 7, 3
 - 2.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 7, 105
 - 2.2 n=7 odd, median is the $\frac{n+1}{2}=\frac{8}{2}=4^{th}$ observation, "6".
- 3. 2, 105, 5, 7, 6, 3
 - 3.1 Arrange the data in increasing order 2, 3, 5, 6, 7, 105
 - 3.2 n=6 even, median is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation $=\frac{5+6}{2}=5.5$.

1. 2, 12, 5, 7, 6, 7, 3 1.1 Sample mean= $\frac{2+3+5+6+7+7+12}{7} = 6$ 1.2 Sample median = 6

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Sample mean= $\frac{2+3+5+6+7+7+12}{7} = 6$
 - 1.2 Sample median = 6
- 2. 2, 117, 5, 7, 6, 7, 3
 - 2.1 Sample mean= $\frac{2+3+5+6+7+7+117}{7} = 21$
 - 2.2 Sample median = 6

- 1. 2, 12, 5, 7, 6, 7, 3
 - 1.1 Sample mean= $\frac{2+3+5+6+7+7+12}{7} = 6$
 - 1.2 Sample median = 6
- 2. 2, 117, 5, 7, 6, 7, 3
 - 2.1 Sample mean= $\frac{2+3+5+6+7+7+117}{7} = 21$
 - 2.2 Sample median = 6

The sample mean is sensitive to outliers, whereas the sample median is not sensitive to outliers.

Let $y_i = x_i + c$ where c is a constant then new median = old median + c

- Let $y_i = x_i + c$ where c is a constant then new median = old median + c
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66. Arranging in ascending order 35,38,47,58,61,66,68,68,70,79 The median for this data is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation which is $\frac{61+66}{2}=\frac{127}{2}=63.5$

- Let $y_i = x_i + c$ where c is a constant then new median = old median + c
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66. Arranging in ascending order 35,38,47,58,61,66,68,68,70,79 The median for this data is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation which is $\frac{61+66}{2}=\frac{127}{2}=63.5$
- Suppose the teacher has decided to add 5 marks to each student.
- ► Then the data in ascending order is 40,43,52,63,66,71,73,73,75,84

- Let $y_i = x_i + c$ where c is a constant then new median = old median + c
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66. Arranging in ascending order 35,38,47,58,61,66,68,68,70,79 The median for this data is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ observation which is $\frac{61+66}{2}=\frac{127}{2}=63.5$
- Suppose the teacher has decided to add 5 marks to each student.
- Then the data in ascending order is 40,43,52,63,66,71,73,73,75,84
- ► The median of the new dataset is $\frac{66+71}{2} = \frac{137}{2} = 68.5$
- ► Note 68.5=63.5+5

▶ Let $y_i = x_i c$ where c is a constant then

 $new median = old median \times c$

▶ Let $y_i = x_i c$ where c is a constant then

 $new median = old median \times c$

Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66. We already know median for this data is 63.5

Let $y_i = x_i c$ where c is a constant then

 $new median = old median \times c$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.We already know median for this data is 63.5
- Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.

▶ Let $y_i = x_i c$ where c is a constant then

$$new median = old median \times c$$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.We already know median for this data is 63.5
- Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
- Then the data becomes 27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4 The ascending order is 14, 15.2, 18.8, 23.2, 24.4, 26.4, 27.2, 28, 31.6

The median of new dataset is $\frac{24.4+26.4}{2} = \frac{50.8}{2} = 25.4$

Let $y_i = x_i c$ where c is a constant then

$$new median = old median \times c$$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.
 - We already know median for this data is 63.5
- Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
- Then the data becomes 27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4

The ascending order is 14, 15.2, 18.8, 23.2, 24.4, 26.4, 27.2, 28, 31.6

The median of new dataset is $\frac{24.4+26.4}{2} = \frac{50.8}{2} = 25.4$

Note $25.4 = 0.4 \times 63.5$

Mode

Another measure of central tendency is the sample mode.

Definition

The mode of a data set is its most frequently occurring value.

Steps to obtain mode

- 1. If no value occurs more than once, then the data set has no mode.
- 2. Else, the value that occurs with the greatest frequency is a mode of the data set.

1. 2, 12, 5, 7, 6, 7, 3;

- 1. 2, 12, 5, 7, 6, 7, 3; 7 occurs twice, hence 7 is mode
- 2. 2, 105, 5, 7, 6, 7, 3

- 1. 2, 12, 5, 7, 6, 7, 3; 7 occurs twice, hence 7 is mode
- 2. 2, 105, 5, 7, 6, 7, 3 7 is mode

- 1. 2, 12, 5, 7, 6, 7, 3; 7 occurs twice, hence 7 is mode
- 2. 2, 105, 5, 7, 6, 7, 3 7 is mode
- 3. 2, 105, 5, 7, 6, 3

- 1. 2, 12, 5, 7, 6, 7, 3; 7 occurs twice, hence 7 is mode
- 2. 2, 105, 5, 7, 6, 7, 3 7 is mode
- 3. 2, 105, 5, 7, 6, 3 no mode

Let $y_i = x_i + c$ where c is a constant then new $mode = old \ mode + c$

- Let $y_i = x_i + c$ where c is a constant then new $mode = old \ mode + c$
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.
 The mode for this data is 68

- Let $y_i = x_i + c$ where c is a constant then new $mode = old \ mode + c$
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.
 The mode for this data is 68
- Suppose the teacher has decided to add 5 marks to each student.
- ► Then the data in ascending order is 40,43,52,63,66,71,73,73,75,84

- Let $y_i = x_i + c$ where c is a constant then new $mode = old \ mode + c$
- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.
 The mode for this data is 68
- Suppose the teacher has decided to add 5 marks to each student.
- Then the data in ascending order is 40,43,52,63,66,71,73,73,75,84
- ▶ The mode of the new dataset is 73
- Note 73 = 68 + 5

Let $y_i = x_i c$ where c is a constant then

 $new \ mode = old \ mode \times c$

Let $y_i = x_i c$ where c is a constant then

$$new \ mode = old \ mode \times c$$

Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66. We already know mode for this data is 68

Let $y_i = x_i c$ where c is a constant then

$$new \ mode = old \ mode \times c$$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.We already know mode for this data is 68
- Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.

Let $y_i = x_i c$ where c is a constant then

$$new \ mode = old \ mode \times c$$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.We already know mode for this data is 68
- ➤ Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
- ► Then the data becomes 27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4 The mode of new dataset is 27.2

Let $y_i = x_i c$ where c is a constant then

$$new \ mode = old \ mode \times c$$

- Example: Recall the marks of students 68,79,38,68,35,70,61,47,58,66.We already know mode for this data is 68
- Suppose the teacher has decided to scale down each mark by 40%, in other words each mark is multiplied by 0.4.
- ► Then the data becomes 27.2, 31.6, 15.2, 27.2, 14, 28, 24.4, 18.8, 23.2, 26.4 The mode of new dataset is 27.2
- Note $27.2 = 0.4 \times 68$

Section summary

- ► Measures of central tendency
 - 1. Mean
 - 2. Median
 - 3. Mode
- Impact of adding a constant or multiplying with a constant on the measures.