

## 2.7: Skewness and the Mean, Median, and Mode

Consider the following data set.

4; 5; 6; 6; 6; 7; 7; 7; 7; 7; 8; 8; 8; 9; 10

This data set can be represented by following histogram. Each interval has width one, and each value is located in the middle of an interval.

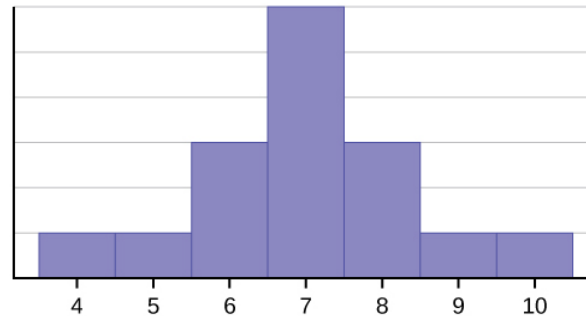


Figure 2.7.1

The histogram displays a **symmetrical** distribution of data. A distribution is symmetrical if a vertical line can be drawn at some point in the histogram such that the shape to the left and the right of the vertical line are mirror images of each other. The mean, the median, and the mode are each seven for these data. **In a perfectly symmetrical distribution, the mean and the median are the same.** This example has one mode (unimodal), and the mode is the same as the mean and median. In a symmetrical distribution that has two modes (bimodal), the two modes would be different from the mean and median.

The histogram for the data: 4; 5; 6; 6; 6; 7; 7; 7; 7; 8 is not symmetrical. The right-hand side seems "chopped off" compared to the left side. A distribution of this type is called **skewed to the left** because it is pulled out to the left.

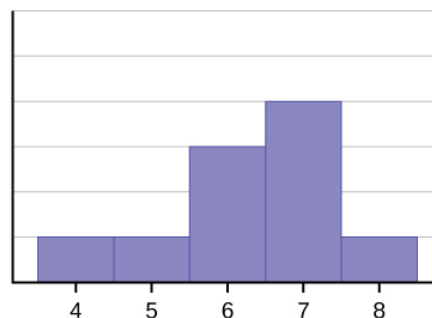


Figure 2.7.2

The mean is 6.3, the median is 6.5, and the mode is seven. **Notice that the mean is less than the median, and they are both less than the mode.** The mean and the median both reflect the skewing, but the mean reflects it more so.

The histogram for the data: 6; 7; 7; 7; 7; 8; 8; 8; 9; 10, is also not symmetrical. It is **skewed to the right**.

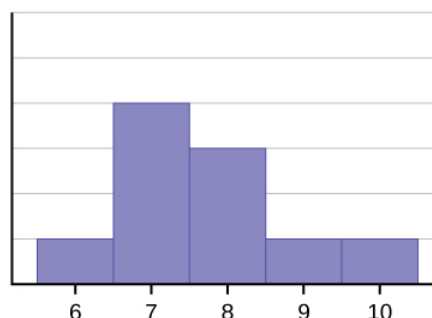


Figure 2.7.3

The mean is 7.7, the median is 7.5, and the mode is seven. Of the three statistics, **the mean is the largest, while the mode is the smallest**. Again, the mean reflects the skewing the most.

*Generally, if the distribution of data is skewed to the left, the mean is less than the median, which is often less than the mode. If the distribution of data is skewed to the right, the mode is often less than the median, which is less than the mean.*

Skewness and symmetry become important when we discuss probability distributions in later chapters.

### Example 2.7.1

Statistics are used to compare and sometimes identify authors. The following lists shows a simple random sample that compares the letter counts for three authors.

- Terry: 7; 9; 3; 3; 3; 4; 1; 3; 2; 2
- Davis: 3; 3; 3; 4; 1; 4; 3; 2; 3; 1
- Maris: 2; 3; 4; 4; 4; 6; 6; 6; 8; 3

- Make a dot plot for the three authors and compare the shapes.
- Calculate the mean for each.
- Calculate the median for each.
- Describe any pattern you notice between the shape and the measures of center.

#### Solution

a.


 This dot plot matches the supplied data for Terry. The plot uses a number line from 1 to 10. It shows one x over 1, two x's over 2, four x's over 3, one x over 4, one x over 7, and one x over 9. There are no x's over the numbers 5, 6, 8, and 10.

Figure 2.7.4: This dot plot matches the supplied data for Terry. The plot uses a number line from 1 to 10. It shows one x over 1, two x's over 2, four x's over 3, one x over 4, one x over 7, and one x over 9. There are no x's over the numbers 5, 6, 8, and 10.


 This dot plot matches the supplied data for Davi. The plot uses a number line from 1 to 10. It shows two x's over 1, one x over 2, five x's over 3, and two x's over 4. There are no x's over the numbers 5, 6, 7, 8, 9, and 10.

Figure 2.7.5: Copy and Paste Caption here. (Copyright; author via source)


 This dot plot matches the supplied data for Mari. The plot uses a number line from 1 to 10. It shows one x over 2, two x's over 3, three x's over 4, three x's over 6, and one x over 8. There are no x's over the numbers 1, 5, 7, 9, and 10.

Figure 2.7.6: Copy and Paste Caption here. (Copyright; author via source)

- Terry's mean is 3.7, Davis' mean is 2.7, Maris' mean is 4.6.
- Terry's median is three, Davis' median is three. Maris' median is four.
- It appears that the median is always closest to the high point (the mode), while the mean tends to be farther out on the tail. In a symmetrical distribution, the mean and the median are both centrally located close to the high point of the distribution.

### Exercise 2.7.1

Discuss the mean, median, and mode for each of the following problems. Is there a pattern between the shape and measure of the center?

a.


 This dot plot matches the supplied data. The plot uses a number line from 0 to 14. It shows two x's over 0, four x's over 1, three x's over 2, one x over 3, two x's over the number 4, 5, 6, and 9, and 1 x each over 10 and 14. There are no x's over the numbers 7, 8, 11, 12, and 13.

Figure 2.7.7: This dot plot matches the supplied data. The plot uses a number line from 0 to 14. It shows two x's over 0, four x's over 1, three x's over 2, one x over 3, two x's over the number 4, 5, 6, and 9, and 1 x each over 10 and 14. There are no x's over the numbers 7, 8, 11, 12, and 13.

b.

#### The Ages Former U.S Presidents Died

4	6 9
---	-----

### The Ages Former U.S Presidents Died

5	3 6 7 7 7 8
6	0 0 3 3 4 4 5 6 7 7 7 8
7	0 1 1 2 3 4 7 8 8 9
8	0 1 3 5 8
9	0 0 3 3

Key: 8|0 means 80.

c.


 This is a histogram titled Hours Spent Playing Video Games on Weekends. The x-axis shows the number of hours spent playing video games with bars showing values at intervals of 5. The y-axis shows the number of students. The first bar for 0 - 4.99 hours has a height of 2. The second bar from 5 - 9.99 has a height of 3. The third bar from 10 - 14.99 has a height of 4. The fourth bar from 15 - 19.99 has a height of 7. The fifth bar from 20 - 24.99 has a height of 9.

Figure 2.7.8: This is a histogram titled Hours Spent Playing Video Games on Weekends. The x-axis shows the number of hours spent playing video games with bars showing values at intervals of 5. The y-axis shows the number of students. The first bar for 0 - 4.99 hours has a height of 2. The second bar from 5 - 9.99 has a height of 3. The third bar from 10 - 14.99 has a height of 4. The fourth bar from 15 - 19.99 has a height of 7. The fifth bar from 20 - 24.99 has a height of 9.

## Chapter Review

Looking at the distribution of data can reveal a lot about the relationship between the mean, the median, and the mode. There are three types of distributions:

- A **right (or positive) skewed** distribution has a shape like Figure 2.7.3.
- A **left (or negative) skewed** distribution has a shape like Figure 2.7.2.
- A **symmetrical** distribution looks like Figure 2.7.1.

Use the following information to answer the next three exercises: State whether the data are symmetrical, skewed to the left, or skewed to the right.

### Exercise 2.7.2

1; 1; 1; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 5; 5

#### Answer

The data are symmetrical. The median is 3 and the mean is 2.85. They are close, and the mode lies close to the middle of the data, so the data are symmetrical.

### Exercise 2.7.3

16; 17; 19; 22; 22; 22; 22; 22; 23

### Exercise 2.7.4

87; 87; 87; 87; 87; 88; 89; 89; 90; 91

#### Answer

The data are skewed right. The median is 87.5 and the mean is 88.2. Even though they are close, the mode lies to the left of the middle of the data, and there are many more instances of 87 than any other number, so the data are skewed right.

### Exercise 2.7.5

When the data are skewed left, what is the typical relationship between the mean and median?

### Exercise 2.7.6

When the data are symmetrical, what is the typical relationship between the mean and median?

**Answer**

When the data are symmetrical, the mean and median are close or the same.

Exercise 2.7.7

What word describes a distribution that has two modes?

Exercise 2.7.8

Describe the shape of this distribution.

Figure 2.7.9: This is a histogram which consists of 5 adjacent bars with the x-axis split into intervals of 1 from 3 to 7. The bar heights peak at the first bar and taper lower to the right.

**Answer**

The distribution is skewed right because it looks pulled out to the right.

Exercise 2.7.9

Describe the relationship between the mode and the median of this distribution.

Figure 2.7.10: This is a histogram which consists of 5 adjacent bars with the x-axis split into intervals of 1 from 3 to 7. The bar heights peak at the first bar and taper lower to the right. The bar heights from left to right are: 8, 4, 2, 2, 1.

Exercise 2.7.10

Describe the relationship between the mean and the median of this distribution.

Figure 2.7.11: This is a histogram which consists of 5 adjacent bars with the x-axis split into intervals of 1 from 3 to 7. The bar heights peak at the first bar and taper lower to the right. The bar heights from left to right are: 8, 4, 2, 2, 1.

**Answer**

The mean is 4.1 and is slightly greater than the median, which is four.

Exercise 2.7.11

Describe the shape of this distribution.

Figure 2.7.12

Exercise 2.7.12

Describe the relationship between the mode and the median of this distribution.

Figure 2.7.13

**Answer**

The mode and the median are the same. In this case, they are both five.

Exercise 2.7.13

Are the mean and the median the exact same in this distribution? Why or why not?

Figure 2.7.14

Exercise 2.7.14

Describe the shape of this distribution.

Figure 2.7.15

**Answer**

The distribution is skewed left because it looks pulled out to the left.

Exercise 2.7.15

Describe the relationship between the mode and the median of this distribution.

Figure 2.7.16: Copy and Paste Caption here. (Copyright; author via source)

### Exercise 2.7.16

Describe the relationship between the mean and the median of this distribution.

Figure 2.7.17

#### Answer

The mean and the median are both six.

### Exercise 2.7.17

The mean and median for the data are the same.

3; 4; 5; 5; 6; 6; 6; 6; 7; 7; 7; 7; 7; 7

Is the data perfectly symmetrical? Why or why not?

### Exercise 2.7.18

Which is the greatest, the mean, the mode, or the median of the data set?

11; 11; 12; 12; 12; 12; 13; 15; 17; 22; 22; 22

#### Answer

The mode is 12, the median is 12.5, and the mean is 15.1. The mean is the largest.

### Exercise 2.7.19

Which is the least, the mean, the mode, and the median of the data set?

56; 56; 56; 58; 59; 60; 62; 64; 64; 65; 67

### Exercise 2.7.20

Of the three measures, which tends to reflect skewing the most, the mean, the mode, or the median? Why?

#### Answer

The mean tends to reflect skewing the most because it is affected the most by outliers.

### Exercise 2.7.21

In a perfectly symmetrical distribution, when would the mode be different from the mean and median?

## Contributors

- Barbara Illowsky and Susan Dean (De Anza College) with many other contributing authors. Content produced by OpenStax College is licensed under a Creative Commons Attribution License 4.0 license. Download for free at <http://cnx.org/contents/30189442-699...b91b9de@18.114>.

This page titled [2.7: Skewness and the Mean, Median, and Mode](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.