

3.1: Measures of Central Tendency

Let's start this section with an example and a multiple choice question:

Example: Pizza delivery



Anthony's Pizza, a Detroit based company, offers pizza delivery to its customers. A driver for Anthony's Pizza will often make several deliveries on a single delivery run. A sample of 5 delivery runs by a driver showed the total number of pizzas delivered on each run:²³

2 2 5 9 12

What is the "average" number of pizzas sent out on a delivery run?

- a. 2 pizzas
- b. 5 pizzas
- c. 6 pizzas

Pick what you think is the answer and we will return to this example and discuss the answer at the end of this section.

Sample Mean

The **sample mean** is the arithmetic average of the data values. You simply add up all the numbers and divide by the sample size. The symbol \bar{X} (pronounced X-bar) refers to the sample mean.

Definition: Sample Mean

If X_1, X_2, \dots, X_n represents a sample of size n , then the **sample mean** is:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum X_i}{n}$$

For the [Example - Pizza delivery](#) data, the sample mean is $\bar{X} = \frac{2 + 2 + 5 + 9 + 12}{5} = 6$ pizzas (the middle value).

Sample Median

The **sample median** is the value that represents the exact middle of data, when the values are sorted from lowest to highest.

Procedure for finding the sample median

1. Sort the data values from lowest to highest.
2. If there is an odd number of values, the sample median is the middle value.

The median of $\{1, 3, 8, 13, 14\}$ is 8

3. If there is an even number of values, the sample median is the mean of the 2 middle values

The median of $\{1, 3, 8, 10, 13, 14\}$ is $\frac{8 + 10}{2} = 9$

Example: Pizza delivery

For the pizza delivery data {2, 2, 5, 9, 12}, the sample median is 5 pizzas (the middle value).

Example: Home prices in a single neighborhood



Here are the selling prices of 6 homes in the same neighborhood in Antioch, California²⁴:

\$500,000 \$550,000 \$600,000 \$700,000 \$700,000 \$1,950,000

The sample mean is \$1,000,000 (add up the values and divide by 5).

The sample median is \$650,000 (\$600,000 plus \$700,000 divided by 2).

Which of the two values is a better measure of the “average” home in this neighborhood?

Here the sample median is a better measure of center, because \$650,000 better represents a typical home in this neighborhood. The mean is not a good measure of center here because the **value** of the outlier home, which costs \$1,950,000. The median will never be affected by outliers because it is only **location** that matters when calculating the median.

Unlike the mean, the median (which is based on ranking instead of values), can be calculated for ordinal categorical data, but not for nominal data.

Example: Grades in a math class

In a community college algebra class, an instructor gave out the following grades to 40 students. Determine the median grade for the course.

A	C	C	B	B	B-	A	A+	B-	A-	C+	B-	C	A	B-	D	F	B+	B	C+
C	A-	A-	B	A-	B	B+	B+	C+	F	B-	F	A+	B+	F	C	B	A-	D	B

The first step is to sort the grades from lowest to highest:

F	F	F	F	D	D	C	C	C	C	C	C+	C+	C+	B-	B-	B-	B-	B-	B
B	B	B	B	B	B	B+	B+	B+	B+	A-	A-	A-	A-	A-	A	A	A	A+	A+

The middle values are both B's, so the median grade is B.

Sample mode

The **sample mode** is the most frequently occurring value in the data. If there are multiple values that occur most frequently, then there are multiple modes in the data.

Example: Pizza delivery

For the pizza delivery data {2, 2, 5, 9, 12}, the sample mode is 2 pizzas because 2 occurs most frequently in the data.

Let's now return to the original question at the beginning of this section.

What is the "average" number of pizzas sent out on a delivery run?

- a. 2 pizzas
- b. 5 pizzas
- c. 6 pizzas

Since 2 is the mode, 5 is the median and 6 is the mean, practically speaking **all 3 answers are examples of "averages"**. Lightbulb Books humorously calls these statistics "The Average Bears."²⁵



Many (including some Statistics texts) will automatically assume that average is the same as mean. In general life, people will use the terms mean and average interchangeably. But in Statistics, when we use the word "average", we mean a value that represents the center of the data. There are many statistics that represent the center of the data, including the mean, median and mode.

The mode can also be used for both nominal and ordinal categorical data.

Example: Nominal data - Marital status

Let's return to the sample of 500 adults (aged 18 and over) from Santa Clara County taken from the year 2000 United States Census.

Marital Status	Frequency
Married	270
Widowed	22
Divorced - not remarried	42
Separated	10
Single - never married	156
Total	500

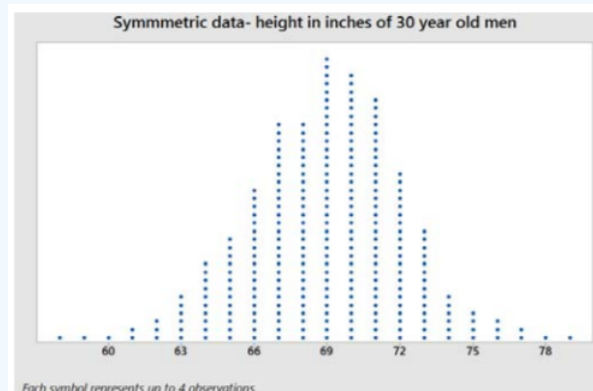
The mode for this data is value with the highest frequency, "Married."

Using the mean and median to determine skewness

Skewness is a measure of how asymmetric the data values are. Data can be positively skewed (stretched to the right), negatively skewed (stretched to the left) or symmetric (no skewness). Let's now explore what effect skewness has on measures of center with several examples.

Example: Symmetric data – Heights of men

Here is a dot plot and summary statistics of the heights in inches of 1000 men, aged 30 years



Sample mean = 68.98 inches

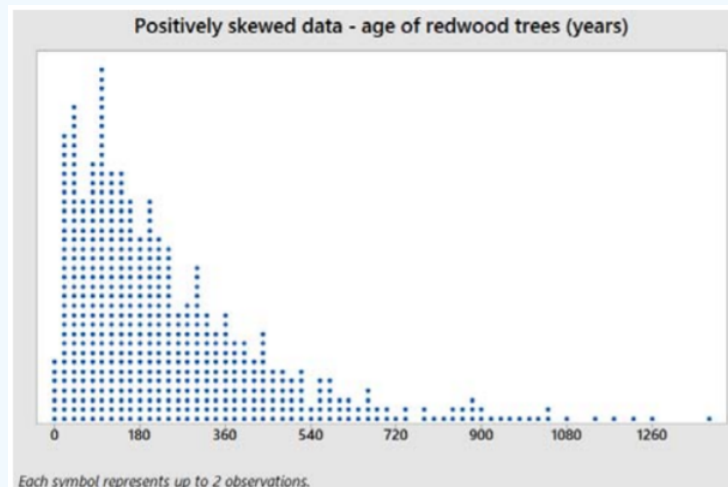
Sample median = 69 inches

Sample mode = 69 inches

The data values are evenly spread on the right and left of the peak. When data are symmetric, the mean, median and mode are about the same.

Example: Positively skewed data – Redwood trees

Here is a dot plot and summary statistics of the age of 1000 redwood trees sampled in California parks.



Sample mean = 237.48 years

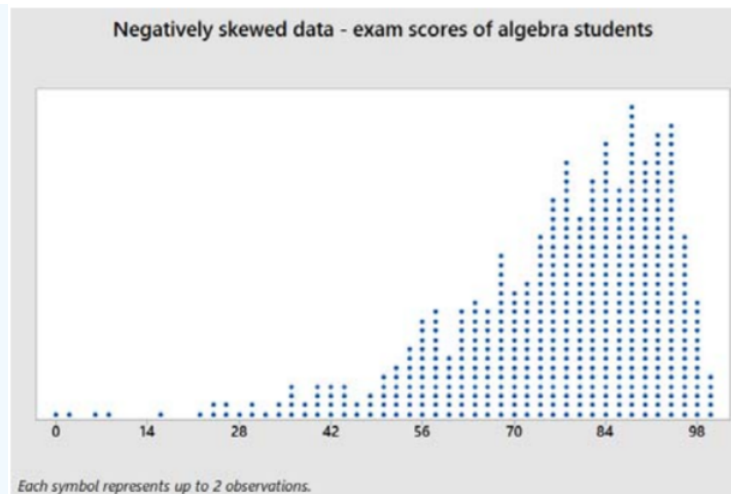
Sample median = 180 years

Sample mode = 100 years

The data values are stretched to the right of center, causing the mean to be greater than the median. Also, the median will usually be greater than the mode for positively skewed data.

Example: Negatively skewed data – Exam grades

Here is a dot plot and summary statistics of the percentage grade of 1000 midterm exams given by a math instructor to algebra students.



Sample mean = 76.21

Sample median = 80

Sample mode = 91

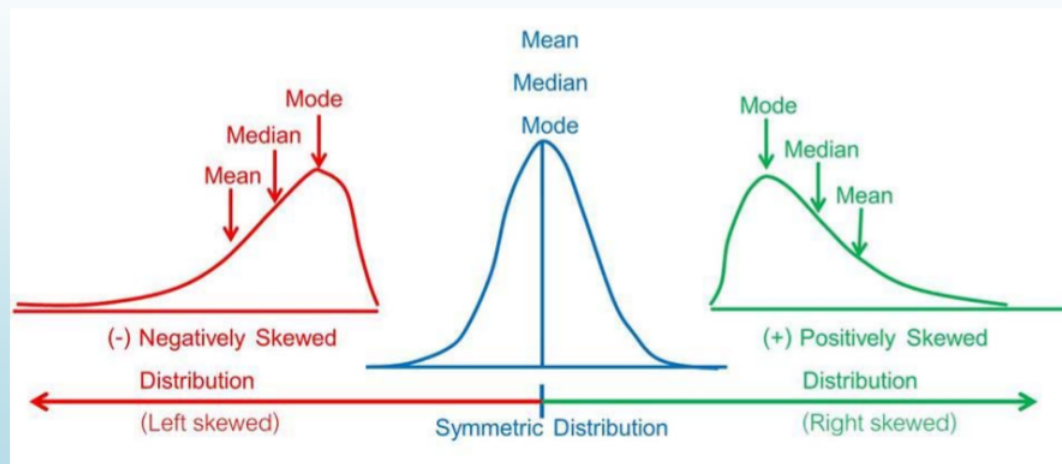
The data values are stretched to the left of center, causing the mean to be less than the median. Also, the median will usually be less than the mode for negatively skewed data.

Using the mean and median to find skewness in data²⁶

For **negatively skewed data**, the mean is less than the median

For **positively skewed data**, the mean is greater than the median

For **symmetric data**, the mean and median are about the same



Example: Students browsing the web

From a prior example, this stem and leaf graph represents how much time 30 students spent on a web browser (on the Internet) in a 24 hour period. Data is rounded to the nearest minute.

6	7
7	18
8	25677
9	25799
10	01233455789
11	268
12	245

The sample median is 101.5 minutes, since the 15th observation is 101 and the 16th observation is 102.

Since the data is skewed negative, we would expect the sample mean to be less than the sample median.

Adding up the values and dividing by 30, we calculate that the sample mean is 96.6 minutes, consistent with data values that are negatively skewed.

Note that the mode is not helpful in this example since the sample size is small.

This page titled [3.1: Measures of Central Tendency](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Maurice A. Geraghty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.