Measures of Center

Objectives

1. Calculate the mean, median, and mode of a dataset
2. Calculate the weighted mean of a dataset
3. Approximate the mean for a grouped dataset
4. Determine if a dataset appears skewed left, skewed right, or normal

Measures of Center

In this section, we will look at finding a central location for a dataset.

Measures of Center

In this section, we will look at finding a central location for a dataset.

In some instances, this can give us a good value to expect from that dataset.

The Mean

Mean

The **mean** of a dataset is found by adding all of the data values together and then dividing by the total number of values.

The Mean

Mean

The **mean** of a dataset is found by adding all of the data values together and then dividing by the total number of values.

When most people use the term *average*, they are referring to the mean.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.

- The mean of a data set uses every value, unless the mean is a *trimmed mean*.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.
- The mean of a data set uses every value, unless the mean is a *trimmed mean*.
- One extreme value (called an **outlier**) can change the value of the mean drastically.

Mean Formula

Sample mean: $\overline{x}$ Population mean: μ

Mean Formula

Sample mean: $\overline{x}$ Population mean: μ

$$\overline{x}, \text{ or } \mu, = \frac{\sum x_i}{n} = \frac{1}{n} \sum x_i$$

The Median

Median

The **median** of a dataset is found by first arranging the data values from least to greatest, then selecting the data value in the middle.

The Median

Median

The **median** of a dataset is found by first arranging the data values from least to greatest, then selecting the data value in the middle.

If there are 2 dava values in the middle, the median will be the mean of these two values.

Properties of the Median

The median separates the top 50% of the data from the bottom 50%.

Properties of the Median

The median separates the top 50% of the data from the bottom 50%.

Unlike, the mean, the median typically does not change by large amounts when we include just a few extreme values; in other words, the median is **resistant**.

Properties of the Median

The median separates the top 50% of the data from the bottom 50%.

Unlike, the mean, the median typically does not change by large amounts when we include just a few extreme values; in other words, the median is **resistant**.

The median is denoted by $\tilde{x}$, however, some technologies uses `Med` to denote it.

The Mode

Mode

The **mode** of a dataset is the value(s) that occur the most.

The Mode

Mode

The **mode** of a dataset is the value(s) that occur the most.

There may be one mode, no mode, or many modes (2 modes is called **bimodal**).

The Mode

Mode

The **mode** of a dataset is the value(s) that occur the most.

There may be one mode, no mode, or many modes (2 modes is called **bimodal**).

The mode is the only measure of center that is applicable to qualitative data.

Example 1

The dataset below represents the number of complaints I receive each week about my teaching.

8, 2, 3, 7, 4, 4, 1, 9, 7, 5

Calculate the mean, median, and mode of the number of complaints.

Example 1

The total number of complaints is 50. There are 10 complaints listed.

Example 1

The total number of complaints is 50. There are 10 complaints listed.

$$\text{Mean} = 50/10 = 5 \text{ complaints per week.}$$

Example 1

The total number of complaints is 50. There are 10 complaints listed.

$$\text{Mean} = 50/10 = 5 \text{ complaints per week.}$$

Sorted: 1, 2, 3, 4, 4, 5, 7, 7, 8, 9

Example 1

The total number of complaints is 50. There are 10 complaints listed.

$$\text{Mean} = 50/10 = 5 \text{ complaints per week.}$$

Sorted: 1, 2, 3, 4, 4, 5, 7, 7, 8, 9

Median number of complaints per week is 4.5

Example 1

The total number of complaints is 50. There are 10 complaints listed.

$$\text{Mean} = 50/10 = 5 \text{ complaints per week.}$$

Sorted: 1, 2, 3, 4, 4, 5, 7, 7, 8, 9

Median number of complaints per week is 4.5

Mode number of complaints per week are 4 and 7.

Example 2

The next week, I received 400 complaints. Re-calculate the mean, median, and mode now.

Example 2

The next week, I received 400 complaints. Re-calculate the mean, median, and mode now.

1, 2, 3, 4, 4, 5, 7, 7, 8, 9, 400

Example 2

The next week, I received 400 complaints. Re-calculate the mean, median, and mode now.

$$1, 2, 3, 4, 4, 5, 7, 7, 8, 9, 400$$

Mean: $450 / 11 = 40.9$ complaints per week.

Example 2

The next week, I received 400 complaints. Re-calculate the mean, median, and mode now.

$$1, 2, 3, 4, 4, 5, 7, 7, 8, 9, 400$$

Mean: $450 / 11 = 40.9$ complaints per week.

Median: 5 complaints per week.

Example 2

The next week, I received 400 complaints. Re-calculate the mean, median, and mode now.

$$1, 2, 3, 4, 4, 5, 7, 7, 8, 9, 400$$

Mean: $450 / 11 = 40.9$ complaints per week.

Median: 5 complaints per week.

Mode: 4 and 7 complaints per week.

Objectives

1. Calculate the mean, median, and mode of a dataset
2. Calculate the weighted mean of a dataset
3. Approximate the mean for a grouped dataset
4. Determine if a dataset appears skewed left, skewed right, or normal

Example 3

California has a mean class size of 20.9 students per teacher and Alaska has a mean of 16.8 students per teacher.

If we combine the two states, we might find the mean number of students per teacher to be 18.85 (0.5 * (20.9 + 16.8)) but is this result correct? Why or why not?

Example 3

California has a mean class size of 20.9 students per teacher and Alaska has a mean of 16.8 students per teacher.

If we combine the two states, we might find the mean number of students per teacher to be 18.85 (0.5 * (20.9 + 16.8)) but is this result correct? Why or why not?

It is not correct because California has a much higher population that Alaska. We would have to find what is known as the **weighted mean**.

Weighted Mean

Sometimes it is necessary to take into account how large each class of a data set is (such as the population of each US state).

Weighted Mean

Sometimes it is necessary to take into account how large each class of a data set is (such as the population of each US state).

If w_i represents the **weight** of each class, then the **weighted mean** can be found via

$$\frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

Weighted Mean

Sometimes it is necessary to take into account how large each class of a data set is (such as the population of each US state).

If w_i represents the **weight** of each class, then the **weighted mean** can be found via

$$\frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

In other words,

1. Multiply each data value by its corresponding weight.

Weighted Mean

Sometimes it is necessary to take into account how large each class of a data set is (such as the population of each US state).

If w_i represents the **weight** of each class, then the **weighted mean** can be found via

$$\frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

In other words,

1. Multiply each data value by its corresponding weight.
2. Add those results.

Weighted Mean

Sometimes it is necessary to take into account how large each class of a data set is (such as the population of each US state).

If w_i represents the **weight** of each class, then the **weighted mean** can be found via

$$\frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

In other words,

1. Multiply each data value by its corresponding weight.
2. Add those results.
3. Then divide that by the total of the weights.

Example 4

You've recently completed a semester. Determine the semester's GPA (A = 4pts, B = 3pts, etc).

Course	**Grade**	**Credit Hours**
Statistics	A	4
Advanced Chris Farley	A	3
Airplane! Quotes	B	5
Obnoxious Examples	C	3

Example 4

Grade	Credit Hours	TOTALS
4	4	16
4	3	12
3	5	15
2	3	6
	15	49

Example 4

Grade	Credit Hours	TOTALS
4	4	16
4	3	12
3	5	15
2	3	6
	15	49

Weighted Mean: $49/15 = 3.267$

Example 5

In a statistics course, tests count for 60% of the final grade, homework for 20% and midterm and final exams are 10% each. Suppose you've earned an 87% average on tests, 94% average on homeworks and a 77% average on the exams.

What is your overall percentage?

Example 5

In a statistics course, tests count for 60% of the final grade, homework for 20% and midterm and final exams are 10% each. Suppose you've earned an 87% average on tests, 94% average on homeworks and a 77% average on the exams.

What is your overall percentage?

Assessment	**Your Scores**	**Grade Weights**	**TOTALS**
Tests	0.87	0.60	0.522
Homework	0.94	0.20	0.188
Exams	0.77	0.20	0.154

Example 5

Grade Weights	TOTALS
0.60	0.522
0.20	0.188
0.20	0.154

Example 5

Grade Weights	TOTALS
0.60	0.522
0.20	0.188
0.20	0.154
1	0.864

Example 5

Grade Weights	TOTALS
0.60	0.522
0.20	0.188
0.20	0.154
1	0.864

Grade: $0.864/1 = 86.4\%$

Objectives

1. Calculate the mean, median, and mode of a dataset
2. Calculate the weighted mean of a dataset
3. Approximate the mean for a grouped dataset
4. Determine if a dataset appears skewed left, skewed right, or normal

Grouped Data

When data is presented in its grouped form, we lose out on knowing the individual elements within the dataset.

Grouped Data

When data is presented in its grouped form, we lose out on knowing the individual elements within the dataset.

This means our measures of center, such as mean, can only be our best educated guess.

Grouped Data

When data is presented in its grouped form, we lose out on knowing the individual elements within the dataset.

This means our measures of center, such as mean, can only be our best educated guess.

We will use similar techniques like we did in finding the weighted mean, however we will have to use the class midpoints as our observed values.

Example 6

(I swear this is a true story) On one very hungry day of mine, I ordered and consumed 19 sushi rolls from an all-you-can-eat sushi restaurant. This, as you might guess, is not typical (at least for me). The table below indicates the frequencies of sushi rolls I typically eat.

Number of Rolls	**Frequency**
1 – 5	4
6 – 10	17
11 – 15	12

Estimate the mean number of rolls consumed.

Example 6

Class Midpoint	Frequency	TOTALS
3	4	12
8	17	136
13	12	156
	33	304

Example 6

Class Midpoint	Frequency	TOTALS
3	4	12
8	17	136
13	12	156
	33	304

Mean: $304/33 \approx 9.2$ rolls per visit.

Objectives

1. Calculate the mean, median, and mode of a dataset
2. Calculate the weighted mean of a dataset
3. Approximate the mean for a grouped dataset
4. Determine if a dataset appears skewed left, skewed right, or normal

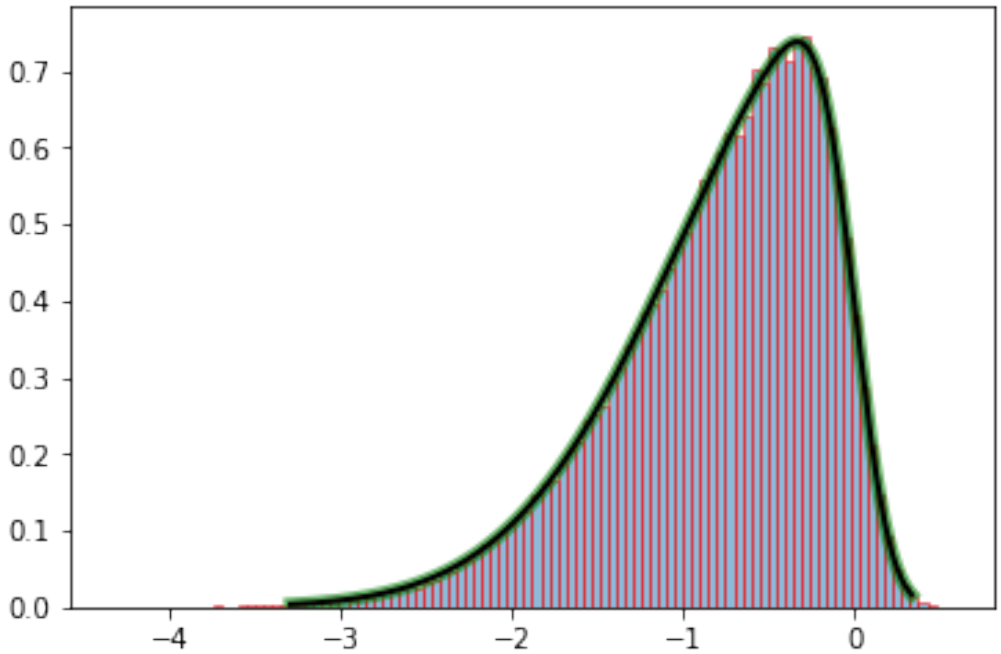
Skewness

When normally distributed (bell-shaped), the mean, median, and mode are (roughly) equal. However, some data sets my be *skewed* (remember, skewness refers to the tail).

Skewness

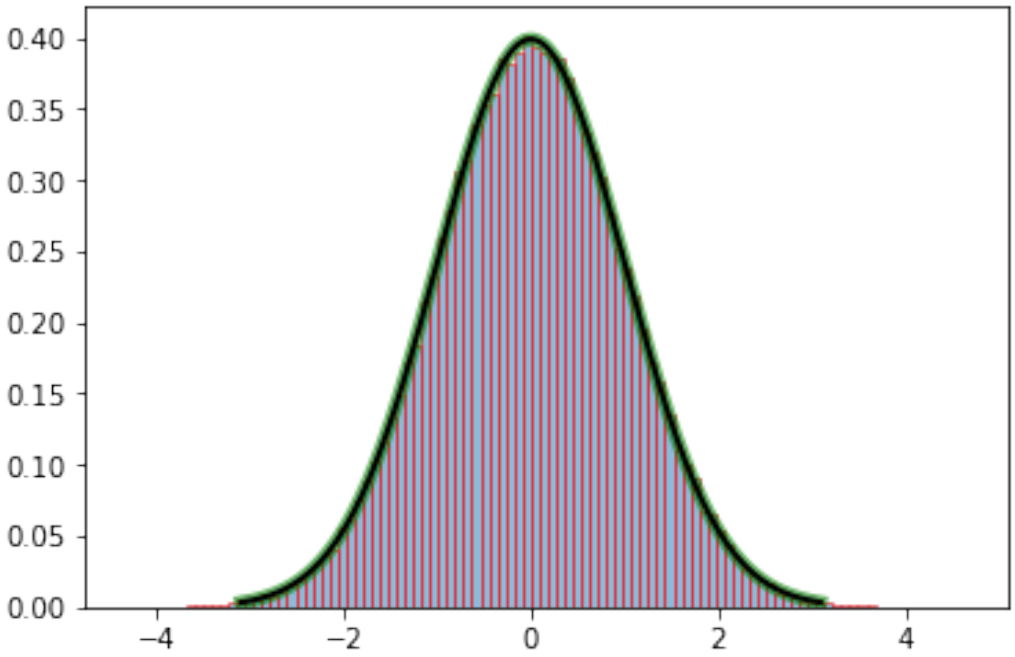
When normally distributed (bell-shaped), the mean, median, and mode are (roughly) equal. However, some data sets my be *skewed* (remember, skewness refers to the tail).

If the mean is (significantly) less than the median, the data is **skewed left** (or **negatively skewed**).



Skewness

If the mean is (approximately) equal to the median, the data is **normal** (no skewness)



Skewness

If the mean is (significantly) greater than the median, the data is **skewed right** (or **positively skewed**).