

Introduction to Statistics & Measures of Central Tendency

In our day to day activities, we often deal with many problems that involve related items of numerical information called **data**. **Statistics** is the study of sets of such numerical data.

There are three steps in a statistical study

- 1 - Collection of data
- 2 - Organization of this data into tables, charts, and graphs (material we will soon study)
- 3 - Drawing conclusions from a statistical analysis of this data

<p>To Find the Mean</p> <p>Find sum of all values, and divide by total number of values. (average)</p>	<p>To Find the Median</p> <p>List the values in order from least to greatest. The value in the middle is the median. If there are two numbers in the middle then average them to find the median.</p>	<p>To Find the Mode</p> <p>The mode is the value that is repeated the most. If each value is used once then there is no mode. There can be more than one mode.</p>
---	--	---

Exercise #1: Students in Mr. Ramirez’s statistics class were trying to determine if people speed along a certain section of roadway. They collected speeds of 20 vehicles, as displayed in the table below.

- (a) List the values in order from least to greatest. You should have 20 values listed.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

- (b) Find the mean, median, and mode for this data set. These are called the **Measures of Center**.

Mean:

Median:

Mode:

Speed (mph)	Number of Cars
29	1
33	2
34	4
35	5
36	3
38	2
39	2
54	1

- (c) The speed limit along this part of the highway is 35 mph. Based on your results from part (a), is it a fair to make the conclusion that the average driver does speed on this roadway? Answer with a sentence that uses the measures of center to support your conclusion.

When conducting a statistical study, it is not always possible to obtain information about every person (census) or situation to which the study applies. Unlike a census, in which every person is counted, some studies use only a **sample** or portion of the items being investigated. Whenever a sample is taken, it is vital that it be **fair**, and **representative of the population** being studied.

Exercise #2: Samuel is trying to determine the average height of high school male students. Because he is on the basketball team, he uses the heights of the 14 players on the team, which are given below in inches.

69, 70, 72, 72, 74, 74, 74, 75, 76, 76, 76, 77, 77, 82

(a) Calculate the mean, median, and mode (measures of center) for this data set. Round answers to the nearest tenth.

Mean: _____

Median: _____

Mode: _____

(b) Is the data set above a representative sample to use to determine the average height of high school male students? Explain your answer.

(c) If the value 82 was changed to 84, which measure(s) of center would stay the same? Which measure(s) of center would change? Find the new measures of center. Round answers to the nearest tenth.

Mean:

Median

Mode:

(d) Explain why the mean value changed but the median and mode did not change.

Mean:

Median

Mode:

Data sets can have more than one mode, as the previous example illustrates. When a data set has two distinct modes, it is called **bimodal**. A data set can also have no mode if there isn't any data value that occurs more than any other.

Exercise #3: In Mr. Smith's Advanced Calculus Course, ten students recently took a test. Their grades were as follows:

45, 78, 82, 85, 87, 88, 89, 90, 93, 95

(a) Calculate the mean and median of this data set. (This data set has no mode)

Mean:

Median:

(b) What score is an **extreme or influential value** in this data set? What makes it influential?

(c) Which value, the mean or the median, do you predict would be a better measure of how well the average student did on Mr. Smith's quiz? **Write 1-2 sentences** about why you think so.

(d) If the value 45 was changed to 70, predict which measure of center(s) would change and how would it change (that is, would it get bigger or smaller?) Do not calculate the centers- just use what you already know. Which measure(s) of center would stay the same?

Introduction to Statistics & Measures of Central Tendency- HW

1. For the following data set, calculate the mean, median and mode. Any non-integer answers should be rounded to the nearest tenth.

3, 5, 8, 8, 12, 16, 17, 20, 24

2. For the following data set, calculate the mean, median and mode. Any non-integer answers should be rounded to the nearest tenth.

5, 5, 9, 10, 13, 16, 18, 20, 22, 22

3. Which of the following is true about the data set {3, 5, 5, 7, 9}? (multiple choice)

(A) mode > median (B) mean > median

(C) median > mode (D) median > mean

4. Which of the following data sets has no mode? (multiple choice)

(A) {2, 3, 4, 4, 5} (B) {7, 7, 8, 9, 9}

(C) {3, 4, 5, 6, 7} (D) {1, 1, 4, 5, 7}

5. Which of the following data sets has a median of 7.5? (multiple choice)

(A) {6, 7, 8, 9, 10} (B) {3, 5, 7, 8, 10, 14}

(C) {1, 3, 7, 10, 14} (D) {2, 7, 9, 11, 14, 17}

Applications

6. Here are some annual salaries for employees of Subway in a local franchise (in thousands of dollars):

35, 35, 35, 35, 35, 37, 37, 38, 40, 40, 65, 70, 120

(a) Why do you think some people might make so much more than others?

(b) Find the mean, median, and mode of this set. (Round answers to nearest hundredth)

(d) If you were a manager in charge of hiring at this franchise, which measure of salary would you report to your applicants- mean, median, or mode? Why? Write 1-2 sentences.

(d) If you were a union representative for Subway employees in charge of bargaining for higher salaries, which measure of salary would you report- mean, median, or mode? Why? Write 1-2 sentences.

7. A survey is taken by an insurance company to determine how many car accidents the average New York City resident has gotten into in the past 10 years. The company surveyed 20 people who are getting off a train at a subway station. The following table gives the results of the survey.

(a) Calculate the mean, median, and mode number of accidents of this data set.

Number of Accidents	Number of People
0	6
1	8
2	4
3	1
11	1

(b) Are there any extreme or influential values in this data set? If so, what data value?

(c) Which number, the mean or the median, better represents the number of accidents an average person in this survey had over this 10 year period? Explain your answer.

(d) Explain why this sample does not fairly represent the average number of accidents a typical New York City resident would get into over a 10 year period?