

Behavioral Statistics



By Robert G. Kelley, Ph.D.

Table of Contents

CHAPTER 5: MEASURES OF CENTRAL TENDENCY	3
Reporting Options: Sharing What Is ‘Typical’	3
The Reason They are Called Measures of Central Tendency	5
The Normal Distribution	7
Be Careful What You Report: When the Distribution is Not Normal	10
Be Careful What You Report: Outliers Distort the Mean	13
Symbols & Formulas: The Mean	16
Chapter Resources & Assessment	20

CHAPTER 5: MEASURES OF CENTRAL TENDENCY

Imagine being asked to describe the ‘typical’ age of students in your class. After recording each person’s age, what method would you choose to summarize the data? In this chapter, we’ll look at how to best determine what is ‘typical’ in a data set.

Reporting Options: Sharing What Is ‘Typical’

Mean: The arithmetic average for a set of scores. To calculate the mean, add up all the values, and divide by the total number of values. The mean is the most common way of describing what is “typical.”

Median: The “Middle Value” in a ranked data set. It is also known as the 50th percentile. To determine the median, first list the scores from lowest to highest. The median score is the score that occurs in the exact middle (e.g., 1, 3, 7, 9, 842; the median in this case is seven - the middle score). Note that the median is unaffected by outliers (i.e., the extreme score of 842). If two scores occur in the exact middle (e.g., 1, 3, 7, 8, 9, 842), then add up the two middle scores and divide by two (e.g., 7.5).

Mode: The mode is the most frequently occurring score (e.g., 4, 5, 5, 8, 9, 10, 11, 12; the mode in this case is 5, since it occurred most often). A group of scores can be bi-modal (having two modes), or multi-modal (having more than two modes). The mode is often used with categorical data (e.g., if asked what is the most common fruit purchased, would respond with the mode).

Test Your Understanding: Determining the Mean, Median, and Mode

For each problem set, determine the values for the mean, median, and mode.

1. Problem Set A: 2, 2, 6, 8, 12

Mean: _____

Median: _____

Mode: _____

2. Problem Set B: 1, 1, 2, 4, 8, 8

Mean: _____

Median: _____

Mode: _____

Answers

1. Problem Set A

Mean: 6, Median: 6, Mode: 2

2. Problem Set B

Mean: 4, Median: 3, Mode: 1, 8

Online Resources: Determining the Mean, Median, and Mode

Mean, Median, and Mode Song



A basic overview of how to determine the mean, median and mode. Includes music and animation. Posted to YouTube by LearningUpGrade.com. View at tinyurl.com/vfsnmh9

Statistics: The Average



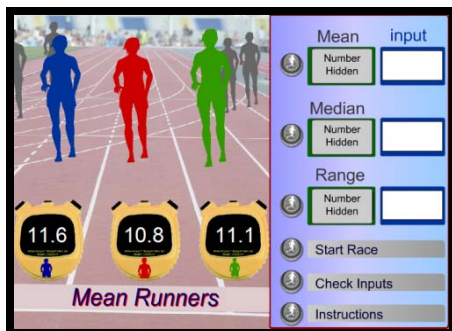
A more in depth, college level, introduction to the mean, median, and mode. Note – starts with a blank screen, which is then written upon. Posted to YouTube by khanacademy. View at tinyurl.com/ykbbvmj

It's Not Hard (Averages Song)



This is just for fun... Covers mean, median, and mode in a way that may leave you laughing aloud. Posted to YouTube by jalapenojane. View at tinyurl.com/yhc885w

Mean Runners – An Interactive Demonstration



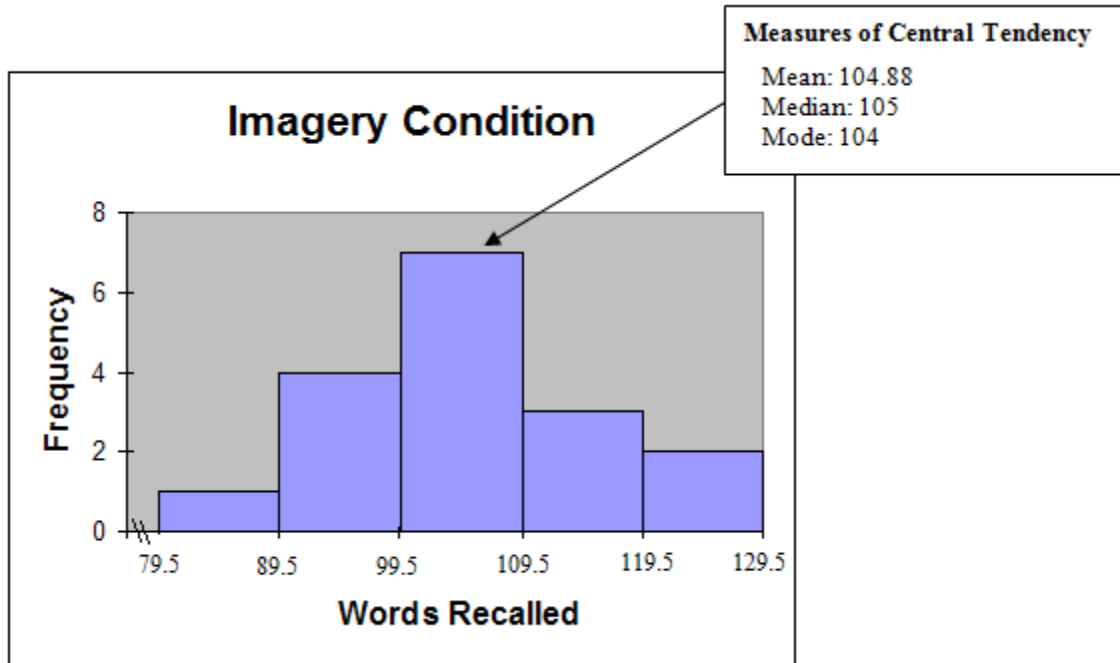
After observing a race, you are asked to determine the mean, median, and range. Directions and an example are provided. Posted at Crickweb.co.uk – primary interactive resources. Participate at tinyurl.com/yf7kv54

The Reason They are Called Measures of Central Tendency

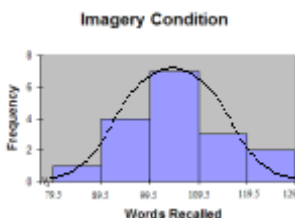
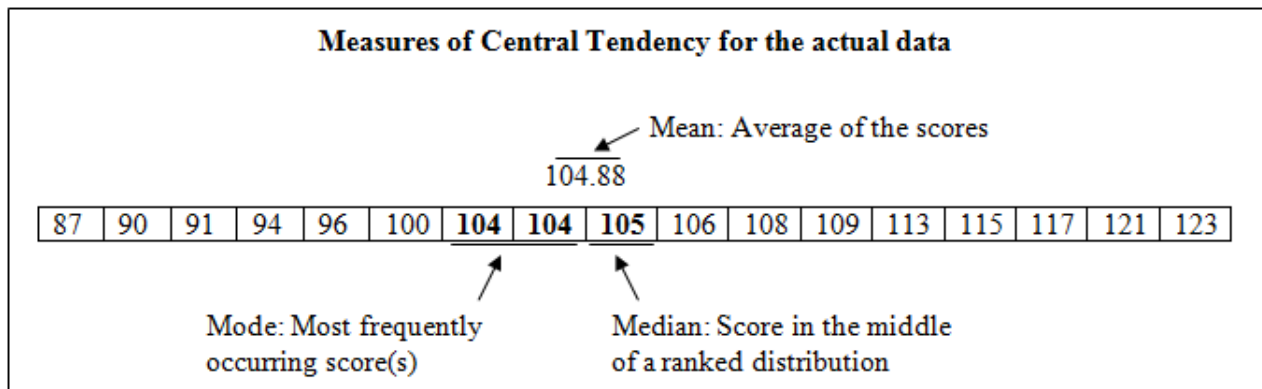
Notice in the graph below that most of the scores are bunched together in the middle of the graph (i.e., the greatest frequency is in the center of the graph). The measures that we have discussed (mean, median, and mode) all occur in the center of the graph. Because these measures typically occur in the center of the graph, they are referred to as measures of central tendency.

The measures of central tendency (mean, median, mode) help us to describe an entire data set with a single well chosen number.

An Example. Imagine that we are interested in whether the use of visual imagery improves memory. We give the participants a list of 130 words, and ask them to create a visual image for each word. We later ask them to recall as many words as possible. How many words did people typically recall?



Below you can see the ranked distribution of scores for the imagery study... Note that the measures of central tendency (mean, median, mode) occur in the middle of the data set, helping us to get an idea of what is 'typical.'



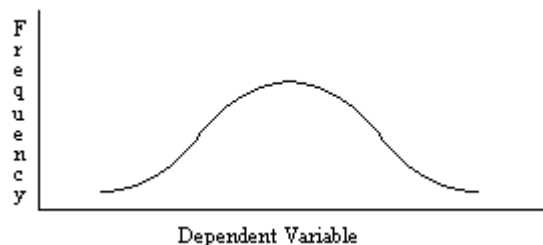
The shape of the histogram for the imagery scores (i.e., bell shaped curve) is fairly common. In fact, this bell shaped curve occurs so often, that it is now simply called the 'Normal Distribution.' With the normal distribution, people will report the mean as the measure of what is typical.

The Normal Distribution

Almost anything measurable (e.g., height, weight, IQ, g.p.a., # of keys) is also normally distributed. That is to say, most people will appear to be 'typical,' with a few above and a few below the mean. For example, in the case of height most men are typically around 5' 7'' (five foot, seven inches). There are some (though fewer) men who are taller than 5'7''. There are also some (though fewer) men who are smaller than 5'7''. But most men are around 5'7''.

Important Characteristics of the Normal Distribution

1. The normal distribution is bell shaped & symmetrical.
2. The mean, median, and mode are all equal.
3. The highest frequency is in the middle of the curve.
4. The frequency gradually tapers off as the scores approach the ends of the curve.
5. The curve approaches but never meets the x-axis at both the high and low ends.



Test Your Understanding: The Mean, Median, Mode and the Normal Distribution

For questions 1 and 2, determine the values for the mean, median, and mode. Note that the raw data is provided (i.e., the data is not ranked).

1. Problem Set A: 95, 80, 99, 54, 93, 95

Mean: _____

Median: _____

Mode: _____

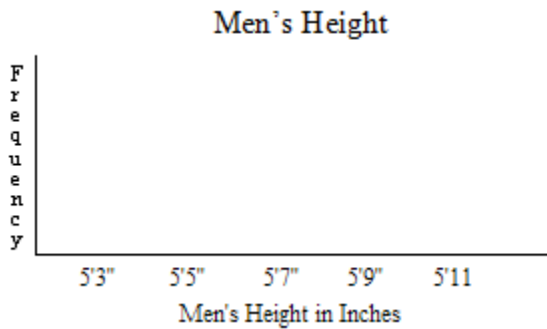
2. Problem Set B: 1, 2, 3, 4, 5

Mean: _____

Median: _____

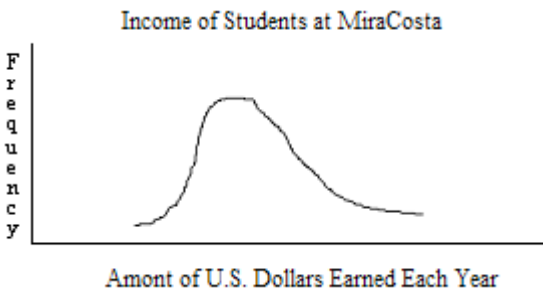
Mode: _____

3. Draw an example of the normal distribution. The middle of the distribution should be centered at 5'7''.



4. The reason that the mean, median, and mode are called measures of central tendency is because they tend to occur in the _____ of the frequency histogram.

5. Why is the graph below NOT an example of a normal distribution?



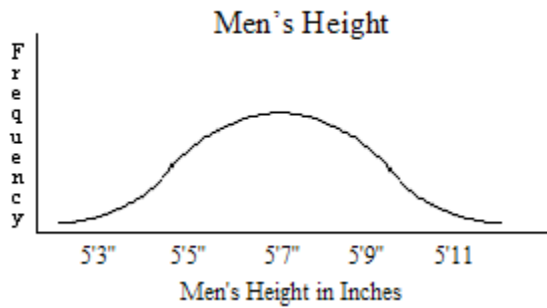
- | |
|--|
| <p>Characteristics of the Normal Distribution</p> <ol style="list-style-type: none"> 1. The normal distribution is bell shaped & symmetrical. 2. The mean, median, and mode are all equal. 3. The highest frequency is in the middle of the curve. 4. The frequency gradually tapers off as the scores approach the ends of the curve. 5. The curve approaches but never meets the x-axis at both the high and low ends. |
|--|

Answer: _____

Answers

1. Mean: 86; Median: 94; Mode: 95
2. Mean: 3; Median: 3; Mode: 1, 2, 3, 4, 5 (since all of the scores tied as occurring the most frequently)

3.



4. center

5. This is not a normal distribution. It is not bell shaped. It is not symmetrical (the left and the right side are not mirror opposites of each other). The highest frequency is not in the exact middle (it is slightly to the left of what would be the middle). The values do not gradually taper off on the left side, but rather decreased sharply. The mean, median, and mode will not be the same.

Online Resources: The Shape of the Normal Distribution

The Amazing Normal Distribution Function



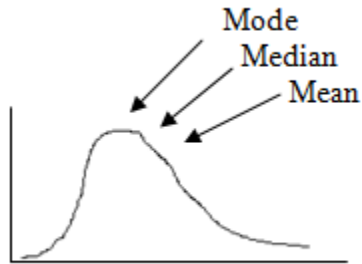
Video: As balls roll between the pins into the bins below, they form a normal distribution. The bean machine was invented by Sir Francis Galton. Posted by K0x on YouTube. View at tinyurl.com/yzpxbza.

Be Careful What You Report: When the Distribution is Not Normal

When the distribution (i.e., set of scores) is normal, then the mean, median, and mode will all be fairly similar. When the distribution is skewed to one side or the other, then we want to

- [1] Let the reader know the shape of the distribution, and
- [2] Report the median as the measure of central tendency.

Positively Skewed Distribution



The tail of the graph extends off to the right.

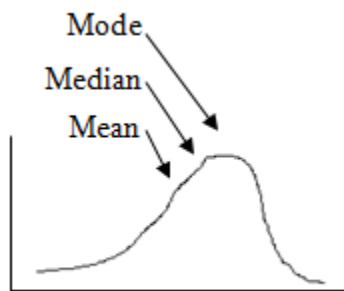
What it is..... Rather than being symmetrical, the right side of the graph extends far out to the right.

What it means... While the majority of the values are smaller, there are some larger values.

An example..... Income tends to be positively skewed. Most people work for around minimum wage, with some people earning large salaries.

What to do..... Tell your reader that the distribution is positively skewed. Report the median, which for this situation is a better measure of what is typical.

Negatively Skewed Distribution



The tail of the graph extends off to the left

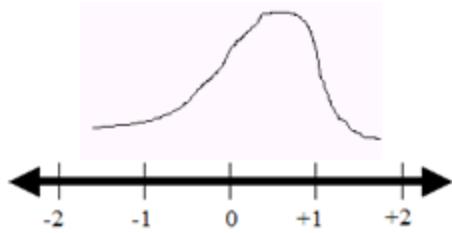
What it is..... Rather than being symmetrical, the left side of the graph extends far out to the left.

What it means... While some of the values are smaller, most of the values are larger.

An example..... The negatively skewed graph on the left indicates that the majority of people watch a lot of TV, with the minority watching only a little TV.

What to do..... Tell your reader that the distribution is negatively skewed. Be sure to report the median (rather than just the mean, which can be misleading).

The Reason These Distributions Are Referred to as Negatively and Positively Skewed



Think of a number line, with the negative numbers on the left and the positive numbers on the right. If the tail of the distribution extends far off to the left, then we say it is a negatively skewed distribution. This does not mean that the values are negative – just that the tail extends off in the negative direction.

Test Your Understanding: Positively and Negatively Skewed Distributions

1. Draw the frequency histogram that corresponds to the following grouped frequency distribution. Use of real limits along the x-axis. The frequency histogram will communicate the number of positive statements used by the clients when describing themselves to their therapist.

Clients' Self-Descriptions

Number of Positive Statements	Frequency
30 – 34	1
25 – 29	2
20 – 24	4
15 – 19	8
10 – 14	20
5 – 9	16
0 – 4	2

2. What is the shape of this distribution? _____

3. Which of the following measures of central tendency (mean, median, mode) is most affected by extreme scores? _____

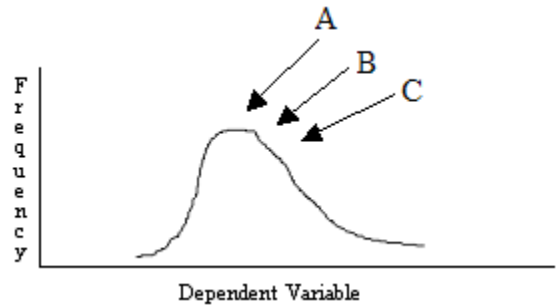
4. When the distribution is skewed or there are extreme scores, which measure of central tendency should we report? _____

5. For the frequency histogram below, identify the mean, median, and mode. Remember that the mode is the most frequently occurring score. The mean is most influenced by extreme scores. The median is the value in the middle.

A: _____

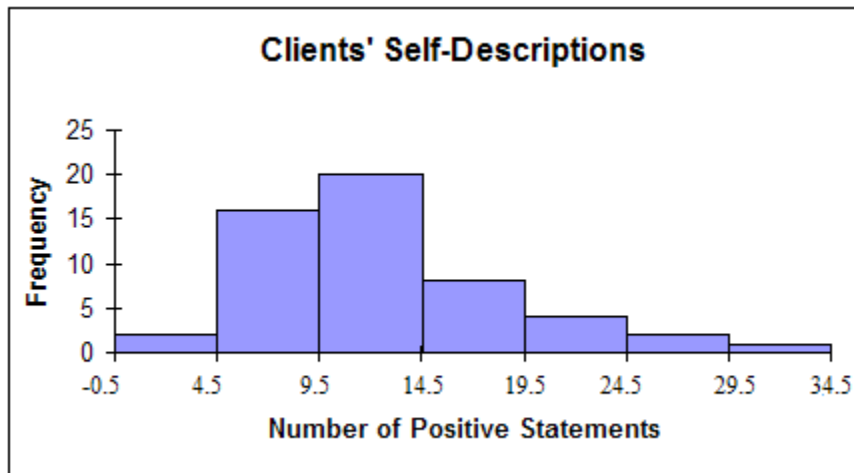
B: _____

C: _____



Answers

1.



2. What is the shape of this distribution? Positively Skewed

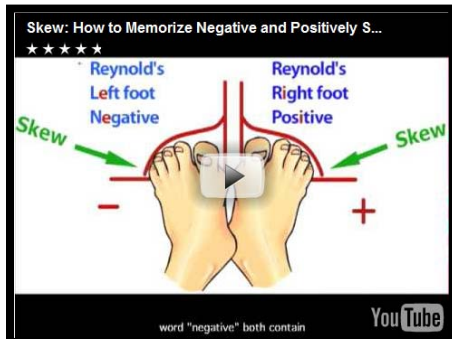
3. Which of the following measures of central tendency (mean, median, mode) is most affected by extreme scores? Mean

4. When the distribution is skewed or there are extreme scores, which measure of central tendency should we report? Median

5. For the frequency histogram below, identify the mean, median, and mode. Remember that the mode is the most frequently occurring score. The mean is most influenced by extreme scores. The median is the value in the middle.

A: Mode, B: Median, C: Mean

Skew: How to Memorize Negative and Positively Skewed Distributions

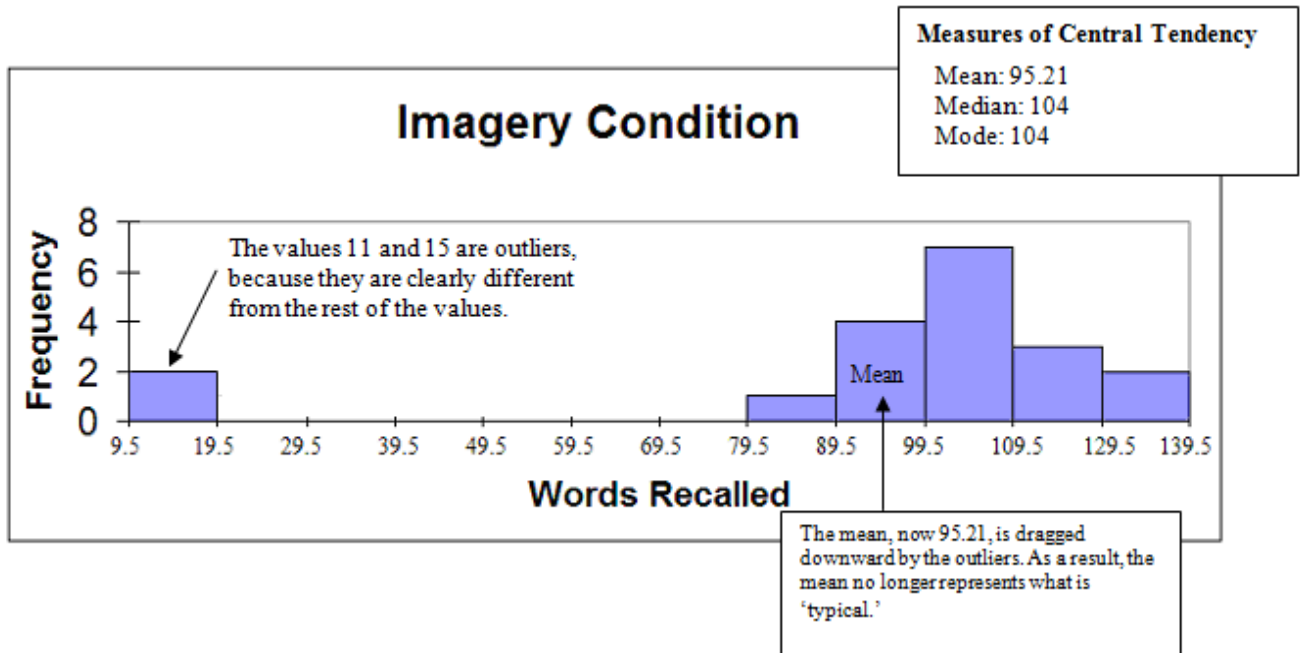


Breaking news – Man has skewed feet! A short video covering how to recognize and remember the names for positively and negatively skewed distributions. Posted by rodolfo1114 on YouTube. View at tinyurl.com/ygukovyn.

Be Careful What You Report: Outliers Distort the Mean

Outliers are *extreme values* that do not ‘fit’ with the rest of the data. The distribution can be normal, but have just one extreme score. When you notice an outlier, the first thing to do is to make sure it is not an error (i.e., a data entry error). If the outlier is an actual value, then you have to be careful. This is because the outlier will overly influence the mean. As a result, the mean will no longer be a good indicator of what is ‘typical’ for your data.

An Example: Consider the scenario that we discussed earlier, regarding the number of words recalled by participants who used visual imagery. But this time, two outliers are also included. They are ‘11’ and ‘15’ words recalled (much lower values than all the other values).



Notice that in the above example, that the two outliers dragged the mean downward – so much so that the mean no longer represents the ‘typical’ value. On the positive side, the median was not greatly affected by the outliers.

When there are outliers, it is a good idea to:

- [1] Inform the reader that there are outliers (report the outlier values)
- [2] Report the median

If you report only the mean when there is an outlier you end up misleading your reader (and perhaps also yourself).

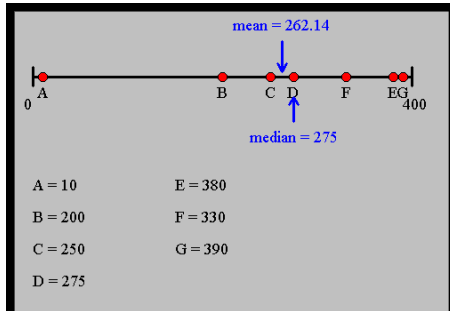
Test Your Understanding: Outliers

1. An extreme value is referred to as an _____.
2. Which measure of central tendency will an outlier affect? _____
3. When there are outliers, which measure of central tendency should be included? _____

Answers

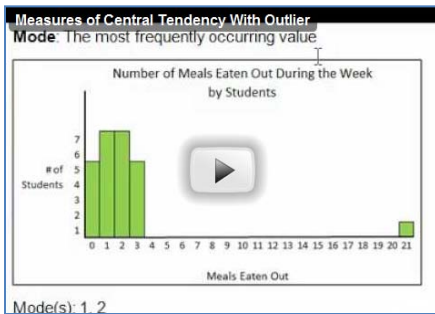
1. An extreme value is referred to as an outlier.
2. An outlier will affect the mean of a distribution.
3. When there are outliers, the median of a distribution should also be reported.

Comparing the Properties of the Mean and the Median



This is an interactive demonstration. Move the numbers around on the number line, and see the corresponding effect on the mean and median. How do outliers affect the mean and median? Participate at tinyurl.com/28blmz2.

What to Report When There is an Outlier



Briefly review the measures of central tendency (mean, median, and mode), learn how each measure is affected by outliers (i.e., an extreme score), and find out what to report when there is an outlier in the data set. View at tinyurl.com/23rbhvy [closed captioned video].

Symbols & Formulas: The Mean

Symbols

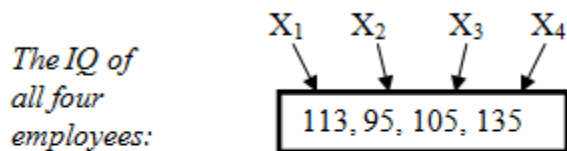
	<i>Population Parameters</i>	<i>Estimates (Based on a Sample)</i>
Mean	μ	\bar{X}
Size	N	n
Finally, the symbol “ Σ ” (pronounced ‘Sigma’) means to add up all the values that follow.		

The symbol μ is pronounced “mu.” The symbol \bar{X} is pronounced “X bar.”

Mathematical symbols and formulas are great – when you know what they mean. By using symbols, people can save time (e.g., rather than writing out “Population Mean,” the person can instead use the symbol μ). Furthermore, rather than writing out a set of instructions in English on how to calculate the mean, mathematicians prefer to write it out as a formula – it takes less space, and states exactly what is needed to determine the answer. A formulas is essentially a set of instructions.

Consider the population mean. In plain English, if we wanted to tell someone how to calculate the population mean, we might say, “The population mean is calculated by adding up all of the values, and then dividing that sum by the size of the population.”

Remember that a population represents the entire group that you wish to describe. For the sake of discussion, we’ll say that our population consists of just four values.

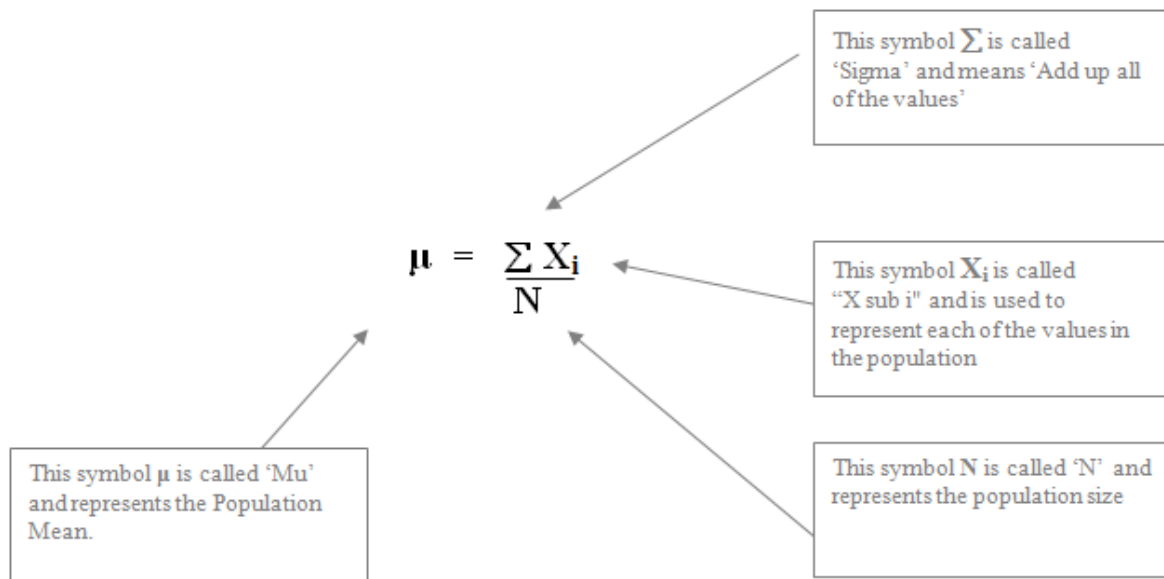


Each value in our population is represented by the variable X_i . Recall that a variable is something that can vary in value. In this case, the variable X represents IQ score.

The symbol X_3 (read it as ‘X sub 3’) refers to the IQ score of the third employee, which is 105..

Anytime we have a population or sample, statisticians like to use the variable ‘X’ to refer to the values... with X_1 being the first value, and X_N being the last value (note that N represents the size of the entire population).

If we wanted to tell someone how to calculate the population mean using mathematics, we could write out the formula:



Notice that our mathematical formula can be thought of as a simple way to communicate the steps needed to obtain the answer. The mathematical formula for the population mean specifies the steps needed to calculate the correct answer.

The formula for the *population mean* (μ) says that we need to *add up all* (\sum) of the *values* (X_i), and divide by the *population size* (N).

$$\mu = \frac{\sum X_i}{N}$$

As you learn the basic statistical symbols, you will improve in your ability to use the mathematical language (i.e., formulas) to communicate and understand basic statistical concepts.

Remember the difference between populations and samples? A population means that every member of the group in which you are interested has been interviewed. A sample is a subset of the population. Statisticians like to keep track of whether you are discussing a population or a sample. To do this, they use different symbols, depending upon whether it is the population or the sample that is being discussed.

When we interview an entire population, then our descriptions (e.g., mean) are referred to as *population parameters*. When a sample is interviewed, (i.e., not the entire population, due to

time or money constraints), then we are using an *estimate* of the population mean. The symbol we use to report our results (e.g., the mean) informs our readers whether we interviewed the entire population or used a sample.

Measure of Central Tendency – The Sample Mean. The formula for the *sample mean* (\bar{X}) says that we need to *add up all* (Σ) of the *values* (X_i), and then divide by the *sample size* (n).

$$\bar{X} = \frac{\sum X_i}{n}$$

Notice that the sample mean is calculated in exactly the same way as a population mean. Only the symbols have been changed (as we are dealing with a sample). When we report a sample mean (referred to as x-bar; an x with a horizontal bar drawn over it) we are sharing with others, based on our sample, our best *estimate* of the population mean.

Test Your Understanding: Symbols and Formulas

1. What is the symbol for the population mean? _____
2. What symbol is used to say ‘Add them all up’? _____
3. What does the symbol x-bar (an x with a horizontal bar drawn over it) represent?

4. What is the symbol for the population size? _____
5. What does the symbol ‘n’ represent? _____
6. What is the formula for the sample mean?

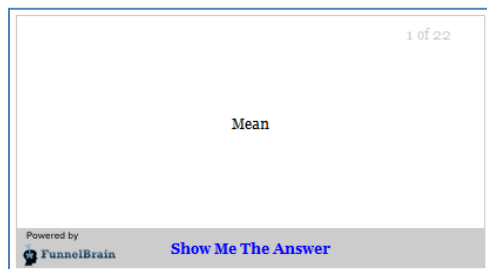
Answers

1. μ
2. Σ
3. Sample mean
4. N
5. Sample size
- 6.

$$\bar{X} = \frac{\sum X_i}{n}$$

Chapter Resources & Assessment

Chapter Flash Cards



Use these online flash card to review what you have learned and to quiz yourself. The online flash cards cover the topic of Measures of Central Tendency (Mean, Median, & Mode)! Review the important terms and concepts covered. View cards at tinyurl.com/2892yq.

Online Statistics Ebooks

Students are responsible for all material covered in this workbook and presented as part of the course. With that said, there is more than one way to learn a topic. Below are listed several different online resources that cover the topic of Measures of Central Tendency. Consider skimming through each of the resources below, to see if there is one that closely matches your learning style.

Introductory Statistics: Concepts, Models, and Applications. This is an online ebook covering statistics, which includes a chapter on measures of central tendency. View at <http://www.psychstat.missouristate.edu/introbook/sbk13.htm>

Online Statistics: An Interactive Multimedia Course of Study. This is an online ebook covering statistics, which includes a chapter on measures of central tendency (including an interactive quiz). View at <http://onlinestatbook.com/chapter3/measures.html>

Seeing Statistics. This is an online ebook covering statistics, which includes a chapter on measures of central tendency. It includes several interactive demonstrations. View at <http://www.seeingstatistics.com/seeing1999/resources/opening.html>