

Statistical Reasoning
Measures of Central Tendency

Name: _____ Date: _____ Class: _____

Mean vs Median vs Mode

What is the best measure of central tendency?

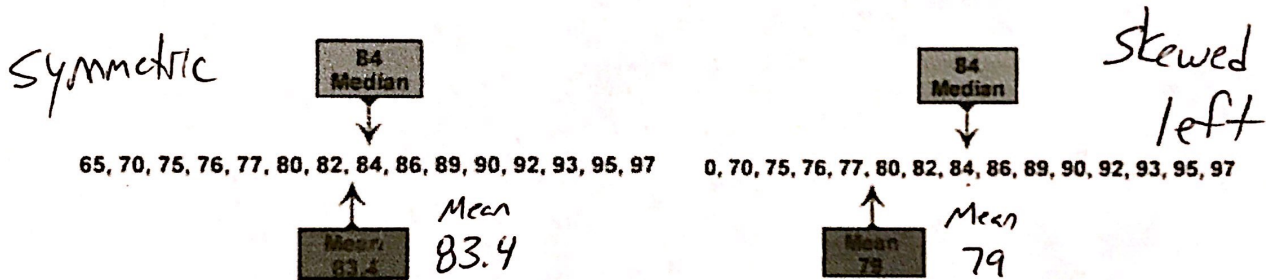
There is no one "best" measure of central tendency, but is often better than another. The most representative measure of central tendency will depend on the type of data and outliers from your data.

and what you are showing from your data.

When is the mean the best measure of central tendency?

The Mean is usually the best measure of central tendency because it analyzes all the data in your set. However, because of this, the mean has one main disadvantage: it is particularly susceptible to the influence of outliers. These are values that are unusually large or small.

For example, consider the two data sets below:



The set on the right shows what would happen if one of the scores was WAY out of range in regard to the other scores. Such a term is called an outlier.

Without the outlier, no changes
With the outlier, mean moved (left), median same

What is the most appropriate measure of central tendency when the data has outliers?

The median is usually preferred in situations where your data is skewed from outliers because the value of the mean can be changed by outliers. However, it will depend on how influential the outliers are. If they do not significantly distort the mean, using the mean as the measure of central tendency will usually be preferred.

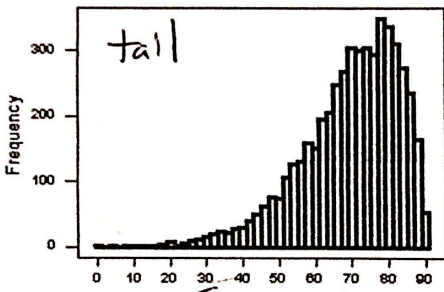
When is the mode the best measure of central tendency?

The mode is the least used of the measures of central tendency. The mode is often used when dealing with categorical data. The mean and/or median are usually preferred when dealing with all other types of data, but this does not mean it is never used with these data types.

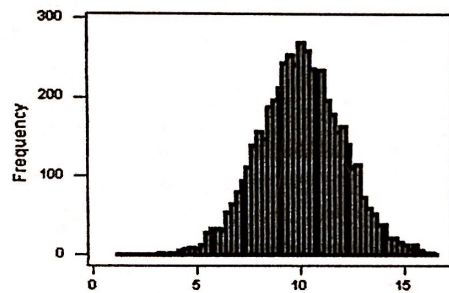
Graphs like histograms, boxplots, and dotplots are useful in visualizing data's central tendency and can assist in deciding which central tendency statistic is most appropriate for a given data set.

In a data set that has few outliers, the data is not skewed and the mean is the best measure of center.

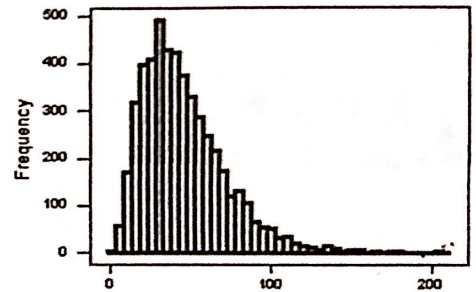
Skewed-Left Distribution



Symmetric, Single-peaked (Unimodal) Distribution



Skewed-Right Distribution

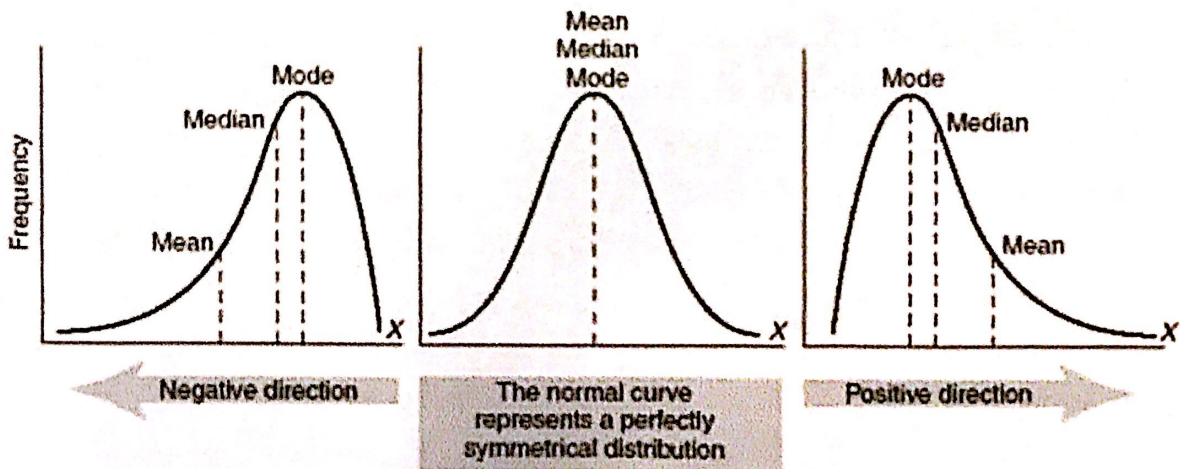


As extreme data is introduced to the sample, the mean begins to be influenced and pulled in positive and negative directions. In this example, the reference lines (from left to right) represent the median, trimmed mean, and mean. In this case, the median is most appropriate measure of center.

(a) Negatively skewed

(b) Normal (no skew)

(c) Positively skewed



MEDIAN

Use the median to describe the middle of a set of data that does have an outlier.

Advantages:

- outliers do not affect the median
- useful when comparing sets of data
- unique - only one median

Disadvantages:

- not as popular as using the mean

MEAN

Use the mean to describe the middle of a set of data that does not have an outlier.

Advantages:

- most popular
- unique - only one mean
- useful when comparing data sets

Disadvantages:

- affected by outliers

MODE

Use the mode when the data is non-numeric or when asked to choose the most popular item.

Advantages:

- outliers do not affect the mode

Disadvantages:

- not as popular as median / ~~mean~~ mean.
- not always unique
- useless when all values repeat the same amount
- when there is more than one mode, difficult to interpret