Averages and Variation

3

Section 3.1

Measures of Central Tendency: Mode, Median, and Mean

Focus Points

- Compute mean, median, and mode from raw data.
- Interpret what mean, median, and mode tell you.
- Explain how mean, median, and mode can be affected by extreme data values.
- Define and compute a trimmed mean.
- Compute a weighted average.

Measures of Central Tendency: Mode, Median, and Mean

The average price of an ounce of gold is $1500. The Zippy car averages 39 miles per gallon on the highway. A survey showed the average shoe size for women is size 9.

In each of the preceding statements, *one* number is used to describe the entire sample or population. Such a number is called an *average*.

There are many ways to compute averages, but we will study only three of the major ones. The easiest average to compute is the *mode*.

The **mode** of a data set is the value that occurs most frequently.

Example 1 – *Mode*

Count the letters in each word of this sentence and give the mode. The numbers of letters in the words of the sentence are

5 3 7 2 4 4 2 4 8 3 4 3 4

Scanning the data, we see that 4 is the mode because more words have 4 letters than any other number.

For larger data sets, it is useful to order—or sort—the data before scanning them for the mode.

Measures of Central Tendency: Mode, Median, and Mean

Not every data set has a mode. For example, if Professor Fair gives equal numbers of A’s, B’s, C’s, D’s, and F’s, then there is no modal grade.

In addition, the mode is not very stable. Changing just one number in a data set can change the mode dramatically.

However, the mode is a useful average when we want to know the most frequently occurring data value, such as the most frequently requested shoe size.

Measures of Central Tendency: Mode, Median, and Mean

Another average that is useful is the *median*, or central value, of an ordered distribution.

When you are given the median, you know there are an equal number of data values in the ordered distribution that are above it and below it.

Measures of Central Tendency: Mode, Median, and Mean

Procedure:

How to find the median

The **median** is the central value of an ordered distribution. To find it,

1. Order the data from smallest to largest.
2. For a distribution with an *odd* number of data values,

$$\text{Median} = \text{Middle data value}$$

3. For a distribution with an *even* number of data values,

$$\text{Median} = \frac{\text{Sum of middle two values}}{2}$$

Example 2 – *Median*

What do barbecue-flavored potato chips cost? According to *Consumer Reports*, Vol. 66, No. 5, the prices per ounce in cents of the rated chips are

19 19 27 28 18 35

(a) To find the median, we first order the data, and then note that there are an even number of entries.

So the median is constructed using the two middle values.

18 19 19 27 28 35

middle values

Example 2 – *Median*

cont'd

$$\text{Median} = \frac{19 + 27}{2} = 23 \text{ cents}$$

(b) According to *Consumer Reports*, the brand with the lowest overall taste rating costs 35 cents per ounce.

Eliminate that brand, and find the median price per ounce for the remaining barbecue-flavored chips.

Example 2 – *Median*

cont'd

Again order the data. Note that there are an odd number of entries, so the median is simply the middle value.

18   19   19   27   28

middle values

Median = middle value

= 19 cents

Example 2 – *Median*

cont'd

(c) One ounce of potato chips is considered a small serving. Is it reasonable to budget about $10.45 to serve the barbecue-flavored chips to 55 people?

Yes, since the median price of the chips is 19 cents per small serving. This budget for chips assumes that there is plenty of other food!

Measures of Central Tendency: Mode, Median, and Mean

The median uses the *position* rather than the specific value of each data entry. If the extreme values of a data set change, the median usually does not change.

This is why the median is often used as the average for house prices.

If one mansion costing several million dollars sells in a community of much lower-priced homes, the median selling price for houses in the community would be affected very little, if at all.

Note:

For small ordered data sets, we can easily scan the set to find the *location* of the median.

However, for large ordered data sets of size *n*, it is convenient to have a formula to find the middle of the data set.

For an ordered data set of size n,

$$\text{Postition of the middle value} = \frac{n+1}{2}$$

Measures of Central Tendency: Mode, Median, and Mean

For instance, if n = 99 then the middle value is the (99 +1)/2 or 50th data value in the ordered data.

If n = 100, then (100 + 1)/2 = 50.5 tells us that the two middle values are in the 50th and 51st positions.

An average that uses the exact value of each entry is the *mean* (sometimes called the *arithmetic mean*).

Measures of Central Tendency: Mode, Median, and Mean

To compute the mean, we add the values of all the entries and then divide by the number of entries.

$$\text{Mean} = \frac{\text{Sum of all entries}}{\text{Number of entries}}$$

The mean is the average usually used to compute a test average.

Example 3 – *Mean*

To graduate, Linda needs at least a B in biology. She did not do very well on her first three tests; however, she did well on the last four. Here are her scores:

58 67 60 84 93 98 100

Compute the mean and determine if Linda’s grade will be a B (80 to 89 average) or a C (70 to 79 average).

Example 3 – *Solution*

$$\text{Mean} = \frac{\text{Sum of scores}}{\text{Number of scores}}$$

$$= \frac{58 + 67 + 60 + 84 + 93 + 98 + 100}{7}$$

$$= \frac{560}{7}$$

$$= 80$$

Since the average is 80, Linda will get the needed B.

Measures of Central Tendency: Mode, Median, and Mean

Comment:

When we compute the mean, we sum the given data. There is a convenient notation to indicate the sum.

Let x represent any value in the data set. Then the notation

Σx (read "the sum of all given x values")

means that we are to sum all the data values. In other words, we are to sum all the entries in the distribution.

Measures of Central Tendency: Mode, Median, and Mean

The *summation symbol* Σ means *sum the following* and is capital sigma, the *S* of the Greek alphabet.

The symbol for the mean of a *sample* distribution of *x* values is $\bar{x}$ (read "*x* bar").

If your data comprise the entire *population*, we use the symbol μ (lowercase Greek letter mu, pronounced "mew") to represent the mean.

Measures of Central Tendency: Mode, Median, and Mean

Procedure:

How to find the mean

1. Compute Σx; that is, find the sum of all the data values.
2. Divide the sum total by the number of data values.

Sample statistic $\overline{x}$ Population parameter μ

$$\overline{x} = \frac{\Sigma x}{n} \qquad \mu = \frac{\Sigma x}{N}$$

where n = number of data values in the sample

N = number of data values in the population

Measures of Central Tendency: Mode, Median, and Mean

We have seen three averages: the mode, the median, and the mean. For later work, the mean is the most important.

A disadvantage of the mean, however, is that it can be affected by exceptional values. A *resistant measure* is one that is not influenced by extremely high or low data values.

The mean is not a resistant measure of center because we can make the mean as large as we want by changing the size of only one data value.

Measures of Central Tendency: Mode, Median, and Mean

The median, on the other hand, is more resistant. However, a disadvantage of the median is that it is not sensitive to the specific size of a data value.

A measure of center that is more resistant than the mean but still sensitive to specific data values is the *trimmed mean*.

A trimmed mean is the mean of the data values left after "trimming" a specified percentage of the smallest and largest data values from the data set.

Measures of Central Tendency: Mode, Median, and Mean

Usually a 5% trimmed mean is used. This implies that we trim the lowest 5% of the data as well as the highest 5% of the data. A similar procedure is used for a 10% trimmed mean.

Procedure:

How to compute a 5% trimmed mean

1. Order the data from smallest to largest.
2. Delete the bottom 5% of the data and the top 5% of the data. *Note:* If the calculation of 5% of the number of data values does not produce a whole number, *round* to the nearest integer.
3. Compute the mean of the remaining 90% of the data.

Distribution shapes and averages

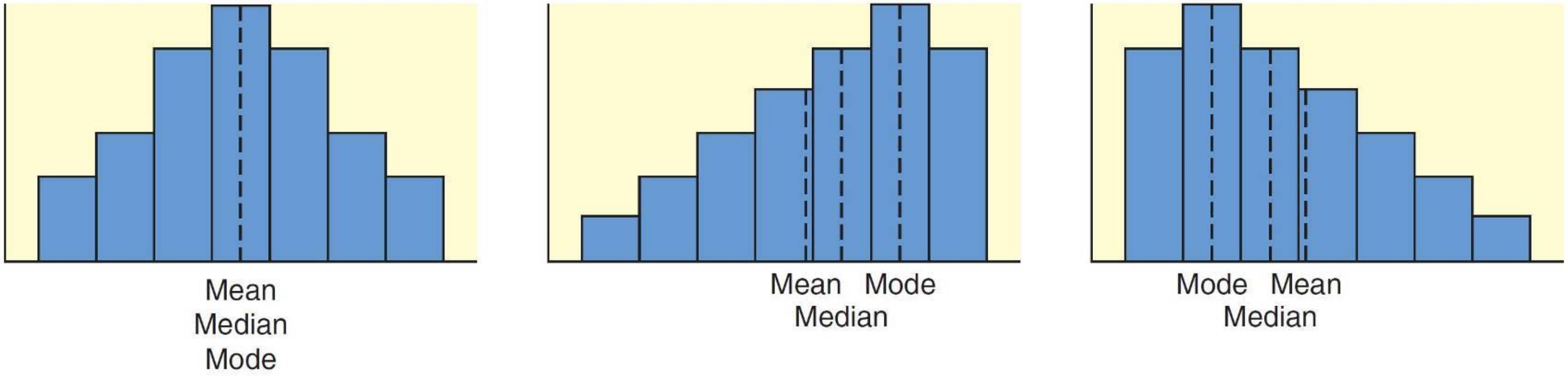
In general, when a data distribution is mound-shaped symmetrical, the values for the mean, median, and mode are the same or almost the same.

For skewed-left distributions, the mean is less than the median and the median is less than the mode.

For skewed-right distributions, the mode is the smallest value, the median is the next largest, and the mean is the largest.

Measures of Central Tendency: Mode, Median, and Mean

Figure 3-1, shows the general relationships among the mean, median, and mode for different types of distributions.



(a) Mound-shaped symmetrical

(b) Skewed left

(c) Skewed right

Figure 3-1

Weighted Average

Weighted Average

Sometimes we wish to average numbers, but we want to assign more importance, or weight, to some of the numbers.

For instance, suppose your professor tells you that your grade will be based on a midterm and a final exam, each of which is based on 100 possible points.

However, the final exam will be worth 60% of the grade and the midterm only 40%. How could you determine an average score that would reflect these different weights?

Weighted Average

The average you need is the *weighted average*.

$$\text{Weighted average} = \frac{\Sigma xw}{\Sigma w}$$

where x is a data value and w is the weight assigned to that data value. The sum is taken over all data values.

Example 4 – *Weighted Average*

Suppose your midterm test score is 83 and your final exam score is 95.

Using weights of 40% for the midterm and 60% for the final exam, compute the weighted average of your scores.

If the minimum average for an A is 90, will you earn an A?

Solution:

By the formula, we multiply each score by its weight and add the results together.

Example 4 – *Solution*

cont'd

Then we divide by the sum of all the weights. Converting the percentages to decimal notation, we get

$$\text{Weighted average} = \frac{83(0.40) + 95(0.60)}{0.40 + 0.60}$$

$$= \frac{33.2 + 57}{1}$$

$$= 90.2$$

Your average is high enough to earn an A.