SECTION 3.1 – Measures of Central Tendency

The Three Measures of Central Tendency

There are three measures of central tendency:

Mean, Median, & Mode.

Population Mean, μ

A *parameter* that is computed by adding all the values of the variable in the population data set and dividing by the number of observations or population size, N. The population mean is denoted by the symbol μ.

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{\sum x_i}{N}$$

Sample Mean, $\bar{x}$

A *statistic* that is computed by adding all the values of the variable in the sample data set and dividing by the number of observations or sample size, n. The sample mean is denoted by the symbol $\bar{x}$.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Median, M

The median of a variable is the value that lies in the middle of the data when arranged in ascending order. We use M to represent the median.

Steps in Finding the Median of a Data Set:

Step 1 Arrange the data in ascending order.
Step 2 Determine the number of observations, n.
Step 3 Determine the observation in the middle of the data set.

- If the number of observations is odd, then the median is the data value exactly in the middle of the ordered data set. That is, the median is the observation that lies in the $\frac{n+1}{2}$ position.
- If the number of observations is even, then the median is the mean of the two middle observations in the ordered data set. That is, the median is the mean of the observations that lie in the $\frac{n}{2}$ position and the $\frac{n}{2}+1$ position.

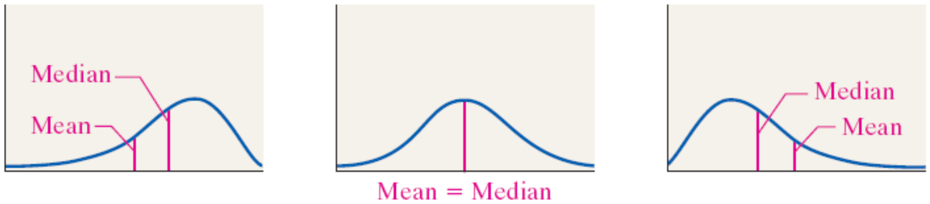
Note: The median is a useful number in cases where the distribution has very extreme values (very large or small) which would otherwise skew the data.

Resistant

The value of a descriptive statistic is said to be **resistant** if extreme values (very large or small) relative to the data do not affect its value substantially. Thus, in this case, the median is resistant and the mean is not resistant.

SECTION 3.1 – Measures of Central Tendency

Relative positions of the mean and median for left-skewed, symmetric, and right-skewed



(a) Skewed Left
Mean < Median

(b) Symmetric
Mean = Median

(c) Skewed Right
Mean > Median

Mode
The mode of a variable is the most frequent observation of the variable that occurs in a data set.

- If no value occurs more than once, then the data set has no mode.
- Otherwise, any value that occurs with the greatest frequency is a mode of the data set.

Bimodal
When a data set has exactly two modes.

Multimodal
When a data set has three or more modes.

Summary: Measures of Central Tendency

Measure of Central Tendency	Computation	Interpretation	When to Use
Mean	Population mean: $\mu = \frac{\sum x_i}{N}$ Sample mean: $\bar{x} = \frac{\sum x_i}{n}$	Center of gravity	When data are quantitative and the frequency distribution is roughly symmetric
Median	Arrange data in ascending order and divide the data set in half	Divides the bottom 50% of the data from the top 50%	When the data are quantitative and the frequency distribution is skewed left or right
Mode	Tally data to determine most frequent observation	Most frequent observation	When the most frequent observation is the desired measure of central tendency or the data are qualitative

SECTION 3.1 – Measures of Central Tendency

☺ **Exercises**:

1) The time, in minutes, it took 9 students to complete their first quiz in a Biology class was:

5, 7, 5, 2, 9, 4, 9, 9, 10

Find the sample mean, median, and mode for the above data set of times.

2) In Exercise #1, suppose the time of 10 was actually a time of 30. How does this affect the mean? How does this affect the median? What property does this illustrate?

3) Find the median and mode of the following data set: 4, 5, 6, 5, 6, 7, 7, 4.

4) Find the median and mode of the following data set: 14, 17, 19, 23, 24, 37, 11.

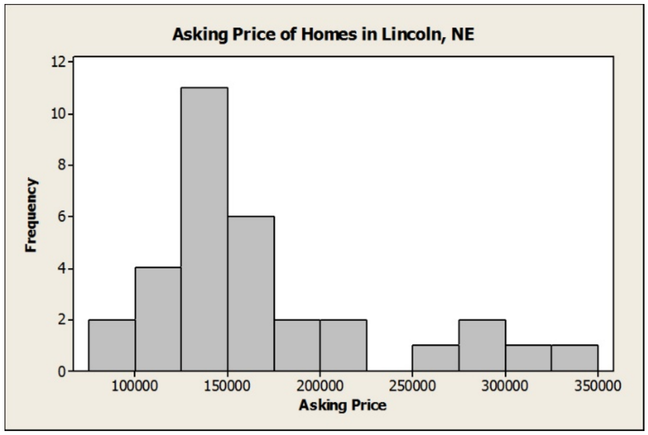
5) Refer to the frequency distribution to the right. Find the median.

6) Refer to the frequency distribution to the right. Find the mode.

x_i	f
0	3
1	0
2	4
3	4
4	9
5	11

☺ **Exercises**:

7) The following frequency histogram represents the asking price of homes for sale in Lincoln, Nebraska. What is the relationship between the mean and median based of the shape of the distribution?



8) The following relative frequency histogram represents the onboard charges for singles on a 7-day cruise sailing to the Mexican Riviera from Los Angeles. What is the relationship between the mean and median based of the shape of the distribution?