Section 3.1

Measures of Central Tendency

Objectives

1. Determine the arithmetic mean of a variable from raw data
2. Determine the median of a variable from raw data
3. Explain what it means for a statistic to be resistant
4. Determine the mode of a variable from raw data

The **arithmetic mean** of a variable is computed by adding all the values of the variable in the data set and dividing by the number of observations.

$$\bar{x} = \frac{(x_1 + x_2 + \ldots\ldots + x_n)}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Compute the population mean of this data.

$\bar{x} = 24.857$

Teacher - can't read the 10th score 3 10 Avg. 81

Nine others Avg. 84

Find that missing score.

$x + 756 = 810$

$x = 54$

The **median** of a variable is the value that lies in the middle of the data when arranged in ascending order.

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.

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Determine the median of this data.

5 18 23 23 26 36 43

Median

5 18 23 24 26 36

Median 23.5

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43, 130

Suppose a new employee is hired who has a 130 minute commute. How does this impact the value of the mean and median?

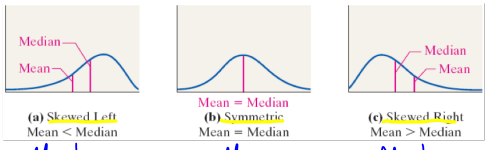
A numerical summary of data is said to be **resistant** if extreme values (very large or small) relative to the data do not affect its value substantially.

Mean before new hire: 24.9 minutes
Median before new hire: 23 minutes

Mean after new hire: 38 minutes
Median after new hire: 24.5 minutes

Relation Between the Mean, Median, and Distribution Shape

Distribution Shape	Mean versus Median
Skewed left	Mean substantially smaller than median
Symmetric	Mean roughly equal to median
Skewed right	Mean substantially larger than median



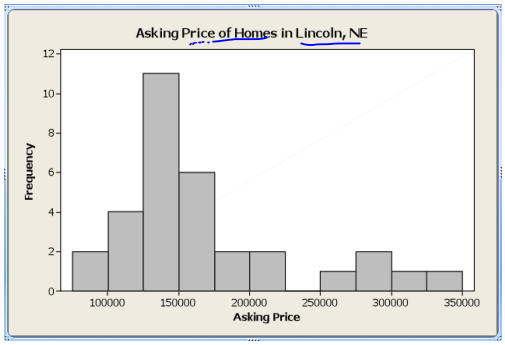
Median Mean Median

The following data represent the asking price of homes for sale in Lincoln, NE.

79,995	128,950	149,900	189,900
99,899	130,950	151,350	203,950
105,200	131,800	154,900	217,500
111,000	132,300	159,900	260,000
120,000	134,950	163,300	284,900
121,700	135,500	165,000	299,900
125,950	138,500	174,850	309,900
126,900	147,500	180,000	349,900

Find the mean and median. Use the mean and median to identify the shape of the distribution. Verify your result by drawing a histogram of the data.

Mean 168,320.13 Mean > Median
Median 148,700.00 Skewed Right



The **mode** of a variable is the most frequent observation of the variable that occurs in the data set.
A set of data can have no mode, one mode, or more than one mode.
If no observation occurs more than once, we say the data have **no mode**.

5, 6, 7, 2 3, 9

Centers:

Mean -- Symmetric

Median - Skewed

Mode - Qualitative or Categorical

Section 3.2

Measures of Dispersion

Objectives

- Determine the range of a variable from raw data
- Determine the standard deviation of a variable from raw data
- Determine the variance of a variable from raw data
- Use the Empirical Rule to describe data that are bell shaped

The **range, *R*,** of a variable is the difference between the largest data value and the smallest data values.

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Find the range.

Range = 43 - 5 = 38

The **standard deviation** of a variable

$$\sigma = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2}{N}} = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$$

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Compute the population standard deviation of this data.

s_x - Sample standard dev.

σ_x - Population " "

$\sigma_x = 36.359$

The **variance** of a variable is the square of the standard deviation. The **population variance** is σ^2 and the **sample variance** is s^2.

$\sigma_x = 36.359 \leftarrow$

$\sigma_x^2 = 1321.98$

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

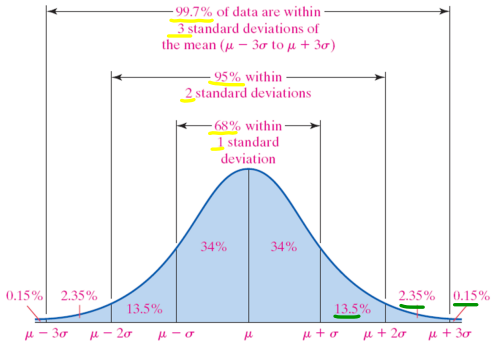
Compute the population and sample variance of this data.

Spread:

Range - Skewed

Standard deviation - Symmetric

The Empirical Rule

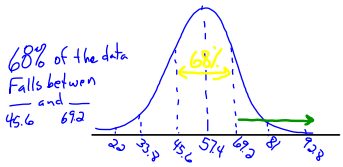


The following data represent the serum HDL cholesterol of the 54 female patients of a family doctor.

41	48	43	38	35	37	44	44	44
62	75	77	58	82	39	85	55	54
67	69	69	70	65	72	74	74	74
60	60	60	61	62	63	64	64	64
54	54	55	56	56	56	57	58	59
45	47	47	48	48	50	52	52	53

9/54

Create the Normal curve for the data. $\bar{X} = 57.4$ $S = 11.8$



use Emp. Rule

Find the % above 69.2 16%

Find the Actual % above 69.2 .166

16.7%