Chapter 5:
Statistical Reasoning

Section 5.1
Analyzing Data Sets

Learning targets:

1. Demonstrate understanding of new terminology.
2. Calculate the mean, median, mode and range for a set of data.
3. Draw and interpret line plots and bar graphs that represent sets of data.

Measures of Central Tendency:

- **Mean:** the sum of the data divided by how many pieces of data are in the set. Commonly referred to as the "average" of the numbers.
- **Median:** the middle value of an ordered set of data. If the data set has more than one middle value, the median is the mean of the two in the middle.
- **Mode:** the value that occurs most often in a data set. A set of data may have one mode, more than one mode, or it may have no mode.

The range of a set of data:

- **Definition:** a measure of how spread out the data is (a measure of "*dispersion*")

- **How to calculate range:** highest value minus lowest value (*max* – *min*)

The measures of central tendency and the range are useful for comparing sets of data.

Other Key Terms:

- **Outlier:** a value in a data set that is very different from other values in the data set.

- **Line Plot:** a graph that records each data value in a data set as a point above the number line.

- **Histogram:** a bar graph where the height of each bar represents the frequency of the occurrence of a data value

Analyzing Sets of Data

1. Organize the data in a systematic way.
   - *list the data in ascending or descending order*

2. Determine or calculate the measures of central tendency.
   - *mean, median and mode*

3. Calculate the range of the data.

4. If comparing two or more sets of data, look for both **similarities** as well as **differences**.

Sets of data may be presented as lists of data, or in some sort of graphical representation (line plot, bar graph or histogram).

Example 1:

Brand A and **Brand B** fluorescent light bulbs were examined to see how long a life span they had. Eight bulbs of each brand were tested and the results are summarized as follows (in thousands of hours):

Brand A:	6.0	6.5	7.0	7.3	8.0	7.5	7.8	8.0
Brand B:	6.1	6.0	6.5	7.2	7.5	7.5	7.5	7.5

Analyze and compare the data for both sets.

A: 6 6.5 7 7.3 7.5 7.8 8 8

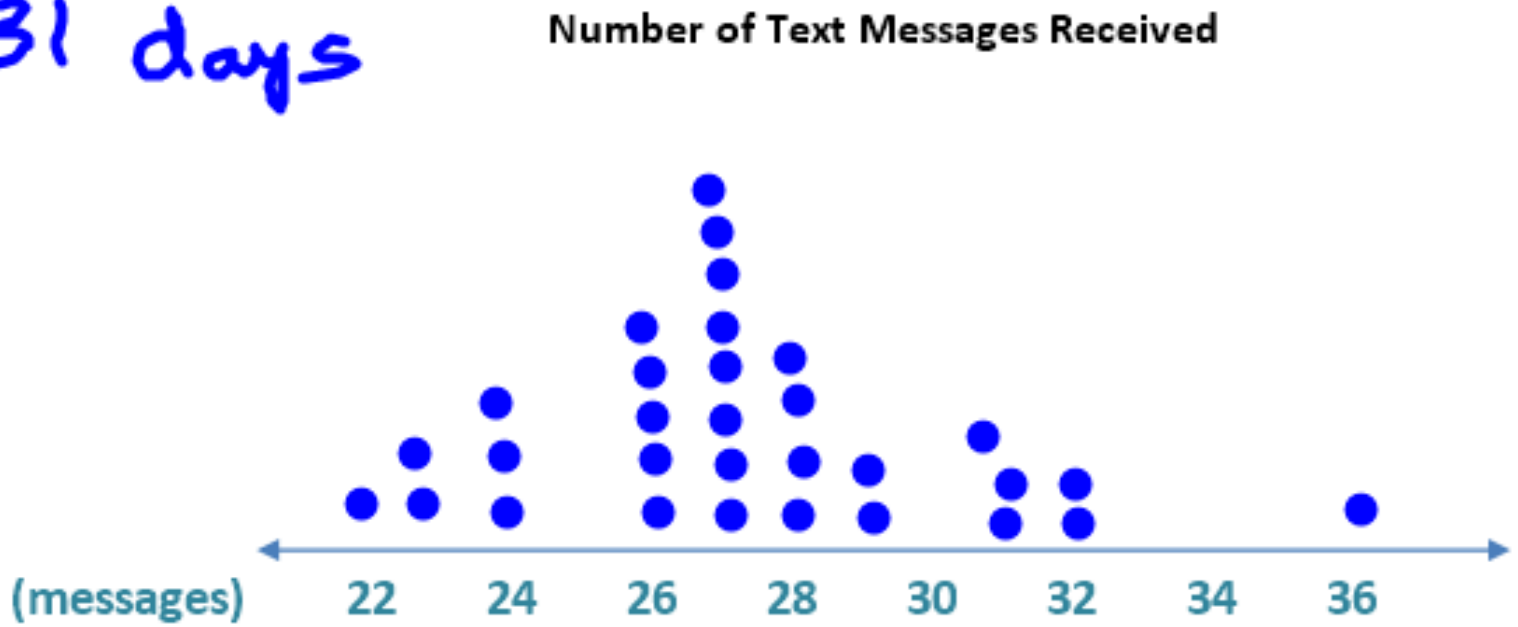
B: 6.0 6.1 6.5 7.2 7.5 7.5 7.5 7.5

	A	B
Mean	7.263	6.975
Median	7.4	7.35
Mode	8	7.5
Range	2	1.5

Example 2:

Alaina recorded the number of text messages she receives each day for a month on a line plot:

31 days



a) On how many days did she receive 32 texts?

2

b) What is the mode of Alaina's data set?

27

c) What is the range of the data set?

$36 - 22 = 14$

d) Determine the median of the data.

27

e) How many text messages did she receive in total over this month?

849

f) Calculate the mean of the data.

27.387

a) $\bar{x} = 15.6$

median = 16.5

mode: 18

Example 3:

c) mean

Sarah's teacher gave her class 10 quizzes, each worth 20 marks. Sarah's marks on the tests are shown below:

16	15	18	16	17
4	18	19	18	15

a) Determine the **measures of central tendency** for this data.

b) Which mark might be considered to be an **outlier**? Discard the outlier, and then determine the measures of central tendency once more.

c) Which measure of central tendency is most affected by an outlier?

b) outlier 4, $\bar{x} = 16.8$

median 17

mode 18

ASSIGNMENT:

Handout #1: #1 – 8

Handout #2: #1 – 10