Math 78
Unit 2: Inferences

Dear Parents,
In this unit your child will learn how to quantify data, and how to organize and display it in different ways. Your child will also learn about **populations** and **samples**. When information is being gathered about a group, such as all registered voters, that entire group is called a population. Because it can be difficult or impossible to research all members of a population, we often choose a part of a population, called a sample, to study. The way a sample is chosen affects how well it represents the entire population. A **random sample**, in which each member of the population has an equal chance of being chosen, is more representative than a **biased sample**. Students will also make use of the box and whiskers plot which displays the upper and lower extreme, the upper and lower quartile, and the median to enable them to make inferences.

Inferences

By the end of this unit, students will:
- Understand how statistics is used to gain information about populations.
- Use data from a random sample to draw inferences about populations.
- Assess the degree of visual overlap of 2 numerical data distributions with similar variability.
- Use measures of center & measures of variability to draw informal comparative inferences about 2 populations.

Vocabulary

- **Box and Whisker Plot**: A diagram that summarizes data using the median, the upper and lowers quartiles, and the extreme values (minimum and maximum). Box and whisker plots are also known as box plots. It is constructed from the five-number summary of the data: Minimum, Q1 (lower quartile), Q2 (median), Q3 (upper quartile), Maximum.
- **Frequency**: the number of times an item, number, or event occurs in a set of data
- **Grouped Frequency Table**: The organization of raw data in table form with classes and frequencies.
- **Histogram**: a way of displaying numeric data using horizontal or vertical bars so that the height or length of the bars indicates frequency
- **Inter-Quartile Range (IQR)**: The difference between the first and third quartiles. (Note that the first quartile and third quartiles are sometimes called upper and lower quartiles.)
- **Maximum value**: The largest value in a set of data.
- **Mean Absolute Deviation:** the average distance of each data value from the mean. The MAD is a gauge of "on average" how different the data values are form the mean value.

$$MAD = \frac{total\ distance\ from\ the\ mean\ for\ all\ values}{number\ of\ data\ values}$$

- **Mean**: The "average" or "fair share" value for the data. The mean is also the balance point of the corresponding data distribution.

$$arithmetic\ mean = \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n}$$

- **Measures of Center**: The mean and the median are both ways to measure the center for a set of data.
- **Measures of Spread**: The range and the mean absolute deviation are both common ways to measure the spread for a set of data.
- **Median**: The value for which half the numbers are larger and half are smaller. If there are two middle numbers, the median is the arithmetic mean of the two middle numbers. Note: The median is a good choice to represent the center of a distribution when the distribution is skewed or outliers are present.
- **Minimum value**: The smallest value in a set of data.

- **Mutually Exclusive**: two events are mutually exclusive if they cannot occur at the same time (i.e., they have not outcomes in common).
- **Outlier**: A value that is very far away from most of the values in a data set.
- **Range**: A measure of spread for a set of data. To find the range, subtract the smallest value from the largest value in a set of data.
- **Sample:** A part of the population that we actually examine in order to gather information.
- **Simple Random Sampling**: Consists of individuals from the population chosen in such a way that every set of individuals has an equal chance to be a part of the sample actually selected. Poor sampling methods, that are not random and do not represent the population well, can lead to misleading conclusions.

Textbook Connection

McGraw Hill Georgia Math Grade 7: Chapter 7
McGraw Hill Textbook Online: connected.mcgraw-hill.com

Web Resources

Vocabulary Online: List One   List Two   Online Quiz Measures of Center
Instructional Videos:
- Mean, Median, Mode, Range
- Choosing the best method to describe a set of data
- The effect of outliers

Online Sites:
http://moodle.oakland.k12.mi.us/os/mod/book/tool/print/index.php?id=25497 (measures of center)
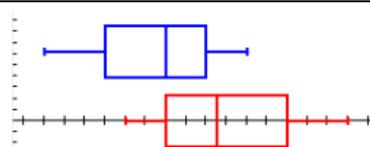https://www.youtube.com/watch?v=TdzrQqqFXW4 (measures of variation)

Practice Problem

In a golf tournament, the top 6 men's & women's scores are given. Compare the spread of the data.
Men's: 65, 68, 70, 72, 73, 75
Women's: 69, 71, 73, 74, 77, 80

A n s w e r

Men's=blue   Women's=red

The range of the women's scores is greater than the range of men's scores by 1. Neither graphs appear symmetrical. The men's skews left while the women's skews right and the men's median is lower than the women's. 50% of the men's lowest scores fall in the range of the lowest 25% of women's scores. The interquartile range for the men's is 5 with a median of 71, while the women's interquartile range is 6 with a median of 73.5. Overall, the men had better scores than the women.