Math 67 Unit 6
Statistics

Dear Parents

In this unit students will begin the study of statistics beginning with numerical data sets, different methods of organizing the data, and analyzing the data sets.

Concepts Students will Use and Understand

- Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers
- Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- Understand that numerical data can be displayed in plots on a number line, including dot plots, histograms, and box plots.
- Summarize numerical data sets in relation to their context, such as by:
- Reporting the number of observations.
- Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
- Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation (Math 67 students)), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

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.
- **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.
- **Mode**- The number that occurs the most often in a list. There can be more than one mode, or no mode.
- **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.
- **Stem and Leaf Plot**- A graphical method used to represent ordered numerical data. Once the data are ordered, the stem and leaves are determined. Typically the stem is all but the last digit of each data point and the leaf is that last digit.

Try http://intermath.coe.uga.edu/dictnary/homepg.asp or http://www.amathsdictionaryforkids.com/ for further examples.

Formula

Mean

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Symbols

Q_1 = quartile 1

Q_2 = median

Q_3 = quartile 3

Example 1

The number of wins that a high school lacrosse team had each of the last five seasons is shown in the table below.

Season	1	2	3	4	5	6
Number of Wins	27	18	24	25	12	x

What number of wins does the team need in season 6 to have a mean of 21 wins for all six seasons?

a) 18   b) 20   c) 22   d) 24

Example 2

What is the median and interquartile range of the set of data displayed in the box and whisker plot?
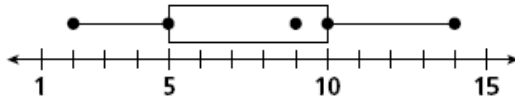
Example 3

Scores from a Unit 1 Math Quiz:
70, 85, 72, 68, 75, 85, 100, 94, 85, 96, 70, 65, 97, 72, 88

1) Organize the data in a stem and leaf plot.
2) What conclusion(s) can you draw from the data?

Key

Example 1

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$21 = \frac{27 + 18 + 24 + 25 + 12 + x}{6}$$

$$21 = \frac{106 + x}{6}$$

$$6 \times 21 = \frac{106 + x}{6} \times 6$$
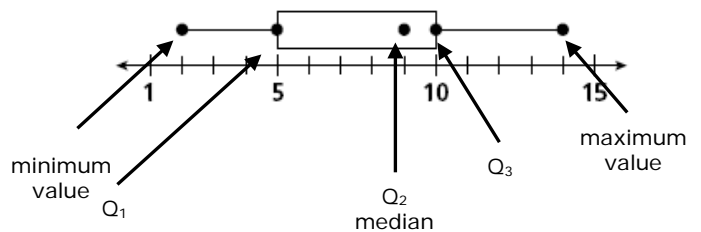
$$126 = 106 + x$$

$$20 = x$$

106	x
126	

Example 3

Stem	Leaf
6	5, 8
7	0, 0, 2, 2, 5
8	5, 5, 5, 8
9	4, 6, 7
10	0

Sample Response:
85 is the median score. Students are still struggling with a few concepts on the quiz.

Example 2

The median is value is 9. The interquartile range is 5.