Statistics, Day 2

Objectives: To find the mean and median of a data set

To make a histogram and box plot of a data set

Suppose we have a set of data, in this case the test scores of one student in Algebra 2. What might we conclude?

 55, 84, 99, 60, 73, 99, 76, 95, 91, 97

Sometimes it is convenient to have one number to describe a set of data. We call this number a measure of central tendency. The most commonly used measures of center are mean, median, and mode.

Mean – use when the data are spread out and you want an average value.

=

Median – arrange data numerically and take the middle value. If there are an even number of data points, average the two data in the middle.

 55, 60, 73, 76, 84, 91, 95, 97, 99, 99

It is best to use the median as a measure of center when the distribution has outliers or a long tail.

Mode – the number that appears most often. If no number is most often, the data has no mode. This measure of center is not widely used.

Which measure of center do most teachers use as grades? Does this seem fair?

Usually, a single number does not tell us all we want to know, and it is helpful to look at the data *graphically*.

Frequency Distributions and Histograms

A frequency distribution summarizes the number of data points that fall within a specific interval. Each interval has to have the same width. A histogram forms a picture of the distribution.

55, 60, 73, 76, 84, 91, 95, 97, 99, 99



|  |  |
| --- | --- |
| Interval | # of Data Pts |
| 50-60\* |  |
| 60-70\* |  |
| 70-80\* |  |
| 80-90\* |  |
| 90-100\* |  |

\*The upper endpoints are not

included in the interval

The way your histogram looks will depend greatly on the intervals you choose.

Box Plot (Box and Whisker Plot)

In a set of data, the quartiles divide the distribution into four equal parts. The first quartile, or Q1, is the median of the lower half of the data and the third quartile, or Q3, is the median of the upper half of the data. By definition, 25% of the data points are contained in the first quartile, 50% of the data is within the first two quartiles, and 75% of the data is within the first three quartiles.

To find Q1, Q2 (which is the same as the Median), and Q3, examine the data in numerical order:

55, 60, 73, 76, 84, 91, 95, 97, 99, 99

To make the box plot, place the quartile values over a number line, along with the minimum and maximum values. Draw a box from Q1 to Q3, a vertical line at the Median, and whiskers to the min and max values.



The minimum, Q1, Median, Q3 and maximum values are called the five number summary.