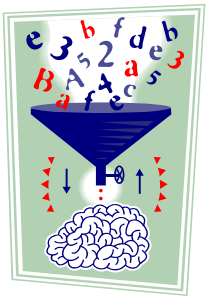
Numerical Data Analysis (pp. 1 of 5)



Before analyzing collected data, it must be determined if the data are **qualitative** (categorical) or **quantitative** (numerical). If numerical or quantitative data have been collected, descriptive statistics involves analysis of data numerically and graphically.

1. On the **Student Survey**, which questions were qualitative (categorical) and which were quantitative (numeric)?

Numerical Analysis

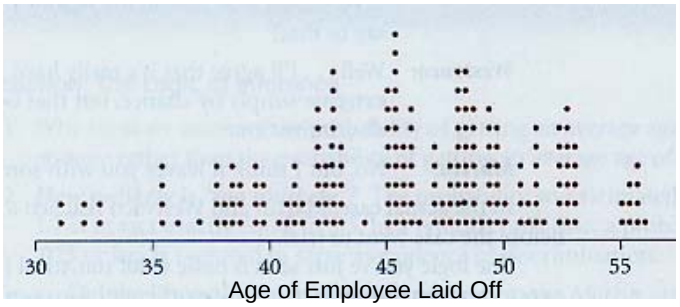
The first step in numerical analysis is to organize the data in a tabular format. Data can be put into sorted lists or condensed using frequency tables. Then the **range** of the data (maximum value – minimum value) and the **measures of central tendency** can be determined.

a. Mode – most often repeated item (modal, bimodal, etc.)
b. Median – center or middle of a list of items after they have been placed in order
c. Mean – calculation found by totaling items and dividing by the number of items in the set ($\bar{x}$, read as *x*-bar)

2. Use the graphing calculator to analyze the data in questions #9-24 on the original **Student Survey**. Organize the data in an appropriate table, determine the range, and determine the measures of central tendency (mode, median, and mean). Record results on the **Chart for Analysis of Numerical Data**. Which measure of central tendency may not occur in a set of data? Which central tendency is the most affected by outliers (non-resistant measure)?

Graphical Analysis

Four main types of graphs are used to represent numerical univariate data: line plot, stem and leaf, box and whisker plot, and a histogram. Graphs require a title and axes labeled and scaled.

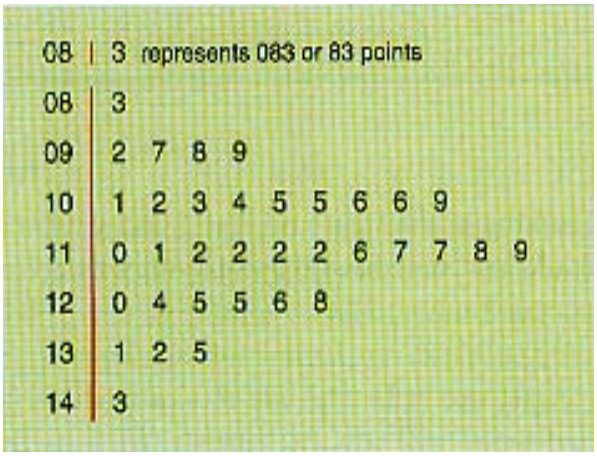


Line Plots (Dot Plots)

- Use with small data sets.
- Draw a horizontal axis and scale appropriately.
- Locate each value in the data set along the scale with a dot, stacked vertically if more than one value in a position.

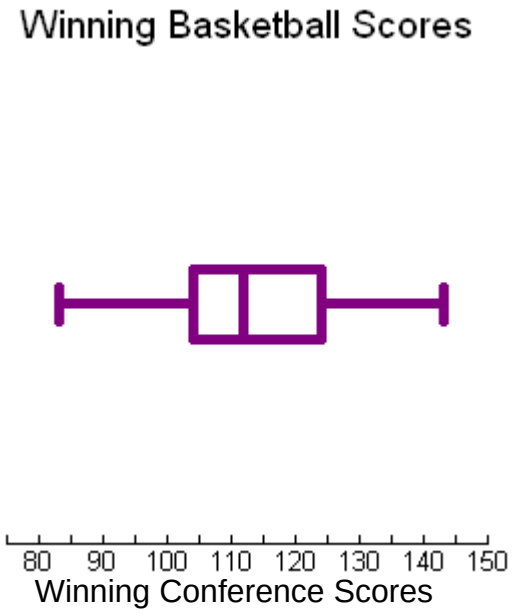
Numerical Data Analysis (pp. 2 of 5)

Points Scored in Last 36 Games



Stem and Leaf Plots

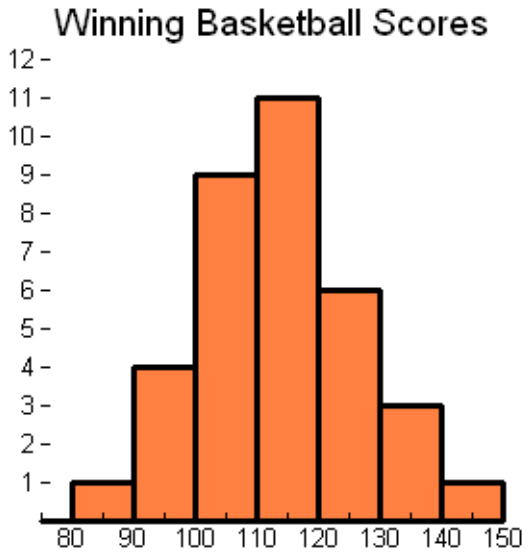
- Use with small to moderate data sets.
- Select one or more leading digits for the stem. The last digit will be the leaf. List stems in a vertical column. Record the leaf for every observation beside the corresponding stem value. Give a key for the units for stems and leaves in the display.



Box and Whiskers Plot

- Use with all sizes of data sets.
- Draw a horizontal measurement scale. Construct a rectangular box whose left or lower edge is at the Quartile 1 value and the right or upper edge is at the Quartile 3 value. Draw a vertical line segment inside the rectangular box at the median value. Extend horizontal lines from the box on the left and right to the minimum and maximum values. The Quartile 1 value is the median of the lower half of the data. The Quartile 3 value is the median of the upper half of the data.

Numerical Data Analysis (pp. 3 of 5)



Histogram
- Use with all sizes of data sets.
- Mark boundaries of group intervals on the horizontal axis. Use frequency on the vertical axis. Draw a rectangle for each class so that the height of each rectangle is the number of observations in that group. The graphing calculator includes the smaller boundary number in the group but excludes the larger boundary number.

3. Construct a line plot and stem and leaf plot for the numerical data collections from the **Student Survey**. Use the graphing calculator to sketch a box and whisker plot and histogram. Label the measures of central tendency on each graph. Record results on the **Chart for Analysis of Numerical Data**.

Analysis of Variability

A final numerical analysis involves the **variability** or spread of the data. The range is the spread of the entire data set. The spread around the mean or median is also useful in analyzing data. The **standard deviation** measures the spread of data around the mean. The **five-number summary** measures the spread of the data around the median.

Numerical Data Analysis (pp. 4 of 5)

Standard deviation

- Definition – average amount either above or below the mean by which the data deviate from the mean
- Calculation – subtract each data value from the mean, square the result, add up all the resulting values, divide by one less than the number of values.

$$s = \sqrt{\frac{\sum \left(x - \bar{x}\right)^2}{n - 1}}$$

4. Explain what each variable in the formula represents and how the formula orders the calculations.

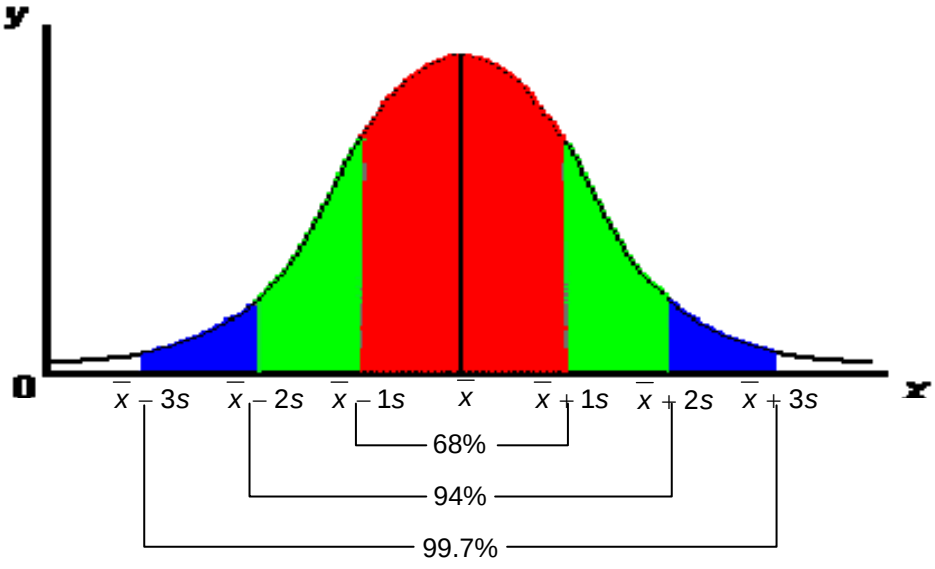
5. Compare the mean and standard deviation of the following sets of numbers. {0,1,2,3,4 5,6,7,8,9} {3,4,4,5,5,5,5,6,6,7}

6. Give examples of data samples where standard deviations might be expected to be large.

7. Give examples of data samples where standard deviations might be expected to be very small.

8. Where is standard deviation found on the graphing calculator?

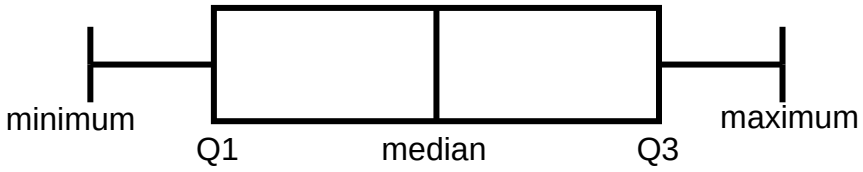
Normally distributed data form a bell shaped curve. The larger the number of samples taken, the more normal data becomes. For normal data, approximately 68% of the data fall within one standard deviation on either side of the mean, 95% of the data fall within two standard deviations of the mean, and 99.7% fall within three standard deviations of the mean.



Numerical Data Analysis (pp. 5 of 5)

Five-number summary

The five-number summary consists of the divisions of the box and whisker plot.



Twenty-five percent of the data values lie in each section denoted by the five-number summary.

9. Use the graphing calculator to determine the standard deviation and five-number summary. Record results on The **Chart for Analysis of Numerical Data**. Super-impose the bell curve and 1, 2, and 3 standard deviation positions on the histogram.

Independent Practice

The following is a list of the number of home runs Babe Ruth had each year for the fifteen years he was with the New York Yankees from 1920 to 1934.

54 59 35 41 46 25 47 60 54 46 49 46 41 34 22

Create an appropriate list for the data.

1. Find the range of the data.
2. Determine the measures of central tendency of the data.
3. If one year Babe Ruth had not hit a home run, which measure of central tendency would be most affected? Explain your reasoning.
4. Graph the data using a line plot, stem and leaf, box and whisker plot, and a histogram. Be sure to title and label and scale axes.
5. Determine the standard deviation and five-number summary of the data.
6. Find the locations of 1, 2, and 3 standard deviations, and super-impose the bell curve on the histogram.