Objectives:

- To understand the ***association*** between two variables and the notions of the ***direction*** and ***strength*** of the association.
- To learn to construct and to interpret ***scatterplots*** as graphical displays of relationships between two variables
- To discover the utility of including a 45° line on a scatterplot when working with paired data
- To become familiar with ***labeled scatterplots*** as devices for including information from a categorical variable into a scatterplot
- To use the computer/calculator to explore associations between variables of genuine data in a variety of applications.

Scatterplot The simplest means of displaying two measurement variables simultaneously. One variable occupies the horizontal axis the other the y. You place a dot at each intersection of the paired x-y observations.

Many times this type of plot is used to predict the value of one variable by knowing the value of the other. To do this you normally put the variable you are trying to predict on the y-axis (*dependent*) and the variable that is being used to predict is on the x-axis (*independent*)

Positive vs. negative association Two variables are said to be associated ***positively*** if as one variable gets larger the value of the other also gets larger. The association is said to be ***negative*** if the value of one gets larger while the other gets smaller.

Strength of association The ***strength*** of the association depends on how closely the observations follow the relationship observed. This strength of association reflects how accurately a prediction may be made from the *dependent* variable to the *independent* variable. (Note direction here – dependent to independent – *important*)

Tendency The association seen in a group of observations illustrates a ***tendency*** for the value of one variable to be paired with a value of the other variable. If there is a very strong relationship that does not mean that, the tendency is always correct. *e.g.* There is a tendency that if you score high on your SAT's you will be successful, this does not mean that if you score high on the SAT you will be successful, you are just more likely to be successful than someone who does not score high.

45° line Illustrates a tendency. If points tend to be above the line then the y value tends to be larger than the x-values, vice versa if more points are below the line. This is not a regression line or a line of best fit. For example if you were to plot the ages of couples with the man's age as the y axis and the woman's age on the x-axis, the points would tend to be above the line since in most couples the man is older.

Labeled scatterplot The scatterplot is for the relationship between two measurement variables, but often you want to compare categorical variables within a scatterplot. Labeling the points by different symbols depending on what symbol you use does this. *e.g.* You have data broken down into three variables *years of service*, *average salary*, and *gender*. You can compare how a mans salary over a length of time compares to a woman's by labeling the man's with x's and woman's with o's. This way you have in a sense two scatterplots on the same graph so they may be compared easily.