Some Rough Definitions:

Here are some definitions of technical terms. These are not to be memorized – rather use them to better understand the statistical issue related to the definition. Some of the definitions will be refined as these topics are revisited.

Experiment: A study in which the investigator assigns comparison features (sometimes called the "treatment") to the study units.

Observational study: A study in which the features to be compared are observed from the study units.

Random Sample: A sample selected from a population in such a way that all possible samples of the same size have the same chance of being selected.

Probability: The long run relative frequency.

Simulation: The use of probability models to generate data.

Probability Distribution: the set of probabilities from a particular probability model.

Data Distribution: The set of relative frequencies generated in a simulation, or from a sampling event, or from any other single source.

Average=Mean: arithmetic average

Standard Deviation=SD: There are two definitions, closely related. For this course the difference between them is irrelevant:

     Definition 1: SD = root mean square deviation from the mean

     Definition 2: SD = root mean square deviation from the mean $* (n/(n-1))^{1/2}$

     The first one is simpler but the second one is more common and is what we demonstrated in class.

Histogram: Frequency graph for quantitative data.

Bar Chart: Frequency graph for qualitative data.

p-values: Probability of observing a result as extreme as was observed, relative to the tested hypothesis

Hypothesis Test: Use of data to assess the credibility of a hypothesis

Random Walk: A series of values generated by independent and identical probability models. The default "random walk" is the symmetiric random walk which takes value +1 and -1 each with probability ½.

Time Series: A series of observations attached to a succession of times.

Reliability: The probability that an item will last (or work) a certain length of time.

Interaction: A relationship between two predictors that predict a third measurement, in which the effect of each of the two predictors on the third measurement cannot be stated separately.

Covariate: An uncontrolled variable that may have an effect on an outcome of interest, even though its effect is not itself of interest.

Normal Distribution: a particular probability distribution that is a good approximation to many data distributions. It is especially useful for describing the variability of averages.

Regression: An analytical technique used to predict one variable from one or more other variables. ("variable" here can be thought of as "measurement").

Residual Plot: Plot of the prediction errors from a regression.