1. About Data

Data (the plural of Datum), in this class, is information on people, places, companies, almost anything. This is the foundation on which economists, businesses, and others base their decisions or thinking back to last time – analysis and generalization (forecasting, predicting, drawing conclusions). GIGO.

2. Some Vocabulary

- *OBSERVATIONS or CASES* – the individuals, animal, company, place, etc which comprise the data. Some might call it the "unit of analysis" the things you plan to study.

- *VARIABLE*: a characteristic of a person, animal, place, thing that can be expressed symbolically with a number or a character (alphabetical).

- *VALUE:* A value is the actual number or descriptor associated with the variable.

3. Types of Data

- *Cross Sectional – provides information on many people, companies, places, whatever, at a single point in time.*

- *Time Series – information on a single entity over different time periods*

- *Panel or Longitudinal Data – many people, companies are measured at different points in time. It differs from a time series in that the time series only follows a single "thing" over time. It differs from a cross-section in that there are different time periods involved.*

- *Simulated Data – data that is generated by software or by hand to simulate/imitate a real process. Video poker is an example. Allows one to play many times and keeps the cost low. Things can be learned from the information that is generated.*

4. Types of Variables

Quantitative - have a "natural" ordering which may be discrete like the spots on a single die or the number of packages shipped by UPS. They might be binary (1,0) (yes/no) (on/off)(dead/alive). They may also be continuous like time, temperature, unemployment rate. Another word for "quantitative variable" is a "numeric variable".

Qualitative - do not have a "natural" ordering examples are major, occupation, name brand. Additionally, there is not necessarily any agreement on how to measure the "distance" between the values. These types of variables may also be called "categorical".

Variables that can be ordered (1st, 2nd, 3rd) (Love, Like, Neutral, Dislike, Hate) may have problems with "distance". These types of variables are not really quantitative (but could be treated as such) nor are they really qualitative (because they do have some order)

5. Additional Words

Random – patternless, without definite aim, not exactly predictable, we might call it "luck", non-deterministic.

Bias, Biased – bent, tendency (think "unfair"), unrepresentative

Population (the "all" or everything being studied) We do not work with the population directly, were more interested in outcomes (samples) that come from it. If we have knowledge about the population we will use it to make statements about the smaller samples that come from it.

Sample (a part of the all) (see Figure 1-1 in your text, page 7) A smaller more manageable piece of the population. Ultimately, one hopes to generalize information derived from the sample to make a statement about the larger population.

Looking at some of the examples handed out recently, can you identify the ones that involve biased samples?