**MSPP PPOL 501–02 & -06: Fall 2014**
**Course Notes #2: Populations & Samples**
**Professor Carolyn Hill**

---

## I. POPULATIONS AND SAMPLES

- Population: all elements of a group; the group about which we want to generalize; all cases of interest in this group.

- Sample: subset of the population of interest; a smaller group of the population of interest to be studied.

  -- an important type of sample is the simple random sample. Here, each element of the population has the same probability of being selected for the sample.

- Examples of populations and simple random samples from these populations…

- Other types of sampling are sometimes used, e.g.

  - systematic sampling (randomly select the first case, then select every $k^{th}$ case following that one);

  - stratified sampling (first divide the population along some characterisic(s) of interest, then randomly draw sample members from within each group)

  - cluster sampling (nested model – randomly select groups/areas from the population; then randomly select units from within that group; and units from within that group)

  -- All of these types of sampling are part of the class of EPSEM samples (equal probability of selection method)

  -- Sampling techniques are an area of study in and of themselves. We'll touch on these again in Quant 2, and GPPI offers a class in survey design, which covers sampling.

- Why draw a sample instead of use the entire population? Because it is seldom possible (due to expense or other kinds of difficulty) to observe all the units or elements in a population.

- If a sample is not representative of the population from which it is drawn, then we can't generalize the results we find to that particular population as a whole, let alone to other populations of interest.

- We distinguish between *descriptive statistics,* which are useful for, well, describing available data. This basic descriptive task may be just what is needed in certain settings.

- Often, though, we want to understand or make statements about ideas, constructs, or groups, etc. beyond the specific ones that are in the dataset at hand.

  -- Admittedly, this may at first seem like an odd enterprise: how can we make statements about ideas, constructs, groups, etc. that we don't actually observe, i.e., whose data we don't have in hand?

  In other words, if we only have information about a sample of the entire group we are interested in, what conclusions can we make about the entire group based on what we observe in the sample?

  -- Fortunately, great minds before us have pondered this question and have produced the kinds of tools that we'll be using. In particular, we'll be using *inferential statistics.*

- In a few class meetings, we'll start going through a battery of basic statistical tests based on inferential statistics. We'll now loop back to make sure you have a foundation in levels of measurement for the data, basic descriptive statistics based on these levels of measurement. To help you understand what lies behind statistical inference, we'll talk about probability, probability distributions, random variables, sampling distributions, and the Central Limit Theorem.