Populations and Samples

• **population**
  the entire group of individuals about which we are interested, typically very large and inaccessible

• **sample**
  subset of the population which is actually measured; if sample = population, this is a census, but typically samples are small fractions of the populations from which they are drawn

• **parameter**
  a numerical measurement (like a mean value $\mu$) associated with a population, often impossible to determine with 100% accuracy because of the inaccesibility of the entire population; we typically use Greek letters to denote parameters

• **statistic**
  a numerical measurement (like a mean value $\bar{x}$) associated with a sample data set, often used to estimate a parameter; we typically use Roman letters to denote statistics
Sample selection

• **representative**
  a sample is said to be representative of its population if its statistics accurately estimate the corresponding population parameters

• **bias**
  the major difficulty in sampling: the tendency for a sample to systematically differ from the population with respect to some relevant characteristic, making it unrepresentative in some way
Sources of bias

- **voluntary response**
  when a sample is populated by data volunteered by subjects; tends create samples that include responses from people with extreme and strongly-held opinions

- **undercoverage**
  excludes certain segments of the population

- **convenience sample**
  sample chosen primarily for ease of access; exclude harder-to-reach segments of the population

- **nonresponse**
  tendency for substantial fractions of a population to ignore appeals for inclusion in the sample or refuse to cooperate

- **response bias**
  tendency of respondents to give inaccurate answers to questions, or of investigators to solicit certain responses and suppress others; often caused by...

- **wording effects**
  confusing terminology or leading questions can produce wildly different responses to attempts to measure the same characteristic from subjects
Methods for combating bias

• **randomization**
  the major tool against bias, as it provides each individual an equal chance at selection into the sample

• **matching**
  an attempt to force a sample to resemble certain specified attributes of the population; can improve the representability of a sample, but is no substitute for randomization

• **simple random sample (SRS)**
  selects individuals from the population entirely at random for inclusion in the sample

• **sampling frame**
  list of individuals from the population from which the sample is drawn; individuals not on the list cannot be selected

• **systematic sample**
  identifies individuals for the sample according to a predetermined selection method
• multistage sample
divides the population into subgroups of similar individuals, called strata, then chooses an SRS from each stratum and combines these to form the complete sample

• cluster sample
divides the population into subgroups of heterogeneous individuals, called clusters, selects certain clusters at random, then collects data from all the individuals in the chosen clusters to form the sample