- **Math 2050, Probability and Statistics**

A First Course in Statistics by McClave and Sincich

CHAPTER 1: STATISTICS, DATA, AND STATISTICAL THINKING

- ***statistics*** is the science of data which involves collecting, classifying, summarizing, organizing, analyzing, and interpreting numerical information

- ***descriptive statistics*** utilizes numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present that information in a convenient form

- ***inferential statistics*** utilizes sample data to make estimates, decisions, predictions, or other generalization about a larger set of data

- ***experimental unit is an object*** (person, thing, transaction, event, etc.) upon which we collect data

- ***population*** – a set of units (usually, people, objects, transactions, or events)

- ***variable*** – a characteristic or property of an individual population unit; any particular characteristic may “vary" among the units in a population

- ***census*** – measuring a variable for every unit of a population

- ***sample*** – a subset of the units of the population

- a ***statistical inference*** is an estimate or prediction or some other generalization about a population based on information contained in a sample

- a ***measure of reliability*** is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference

- four elements of descriptive statistical problems
  1) the population or sample of interest
  2) one or more variables that are to be investigated
  3) tables, graphs, or numerical summary tools
  4) identification of patterns in the data

- five elements of inferential statistical problems
    1) the population of interest
    2) one or more variables that are to be investigated
    3) the sample of population units
    4) the inference about the population based on information contained in the sample
    5) a measure of reliability for the inference

- ***quantitative data*** are measurements that are recorded on a naturally occurring numerical scale

- ***qualitative data*** are measurements that cannot be measured on a natural numerical scale; they can only be classified into one of a group of categories

- types of data by measurement scale (from lower to higher)

nominal data – qualitative data that can be classified into categories
ordinal data – qualitative data can be ordered or ranked
interval data – numerical data with meaningful differences in measurements
ratio data – numerical data; division provides meaningful ratios and zero exists

qualitative – includes nominal and ordinal data types
quantitative – includes interval and ratio data types

- ***data collection***: a) data from a published source
  b) data from a designed experiment
  c) data from a survey
  d) data collected observationally

- a ***representative sample*** exhibits characteristics typical of those possessed by the target population

- a ***random sample*** is a sample collected in such a way that every subset of fixed size in the population has the same chance of being included in the sample

- ***statistical thinking*** involves applying rational thought to assess data and the inferences made from them critically

- ***selection bias*** - when a subset of the population has no chance of being sampled
- ***nonresponse bias*** - when subset of the sample does not respond to survey
- ***measurement error*** - inaccuracies in the values of the data recorded;

may be due to ambiguous or leading questions or interviewer's effect