Definitions for Chapter 1– Introduction to Statistics
Math 1040-1

Section 1.1

- **Data** are information coming from observations, counts, measurements, or responses.
- **Statistics** is the science (and art) of designing studies, as well as analyzing and interpreting the data that those studies produce.
- A **population** is the set of all subjects in which we are interested.
- A **sample** is a subset of a population. This is the group from which we have data.
- A **parameter** is a numerical description of a population characteristic.
- A **statistic** is a numerical description of a sample characteristic.
- **Descriptive statistics** refers to methods for summarizing the data collected.
- **Inferential statistics** involves using data from a sample to make conclusions about a population.

Section 1.2

- **Qualitative data** consist of attributes, labels, and other non-numerical entries.
- **Quantitative data** consist of numerical measurements or counts.
- Data at the **nominal level of measurement** are qualitative. No mathematical computations can be carried out.
- Data at the **ordinal level of measurement** are quantitative or qualitative. They can be arranged in order (ranked), but differences between entries are not meaningful.
- Data at the **interval level of measurement** are quantitative and can be ordered. Meaningful differences between data entries can be calculated. The zero entry represents a position on a scale, but the entry is not inherently zero.
- Data at the **ratio level of measurement** satisfy the requirements for data at the interval level, except that the zero entry is an inherent zero.

Section 1.3

- A **variable** is any characteristic that is recorded for subjects in a study.
- In an **observational study**, a researcher measures and observes the variables of interest without changing existing conditions.
- In an **experiment**, a researcher assigns a **treatment** and observes the **response**. Sometimes, a **control group** may be used to compare the effectiveness of a treatment.
- A **simulation** uses a mathematical, physical, or computer model to replicate the conditions of a process or situation.
- A **survey** is used to investigate characteristics of a population. It is frequently used when the subjects are people, and questions are asked of them.

- A **confounding variable** occurs when an experimenter cannot tell the difference between the effects of different factors on a variable.
- The **placebo effect** occurs when a subject reacts favorably to a placebo when no medicated treatment has been given.
- **Blinding** is a technique used to make the subjects "blind" to which treatment they are being given.
- A **double-blind** experiment is one in which neither the experimenter nor the subjects know which treatment is being given.
- **Randomization** is a process of randomly assigning subjects to treatment groups:
  - A **completely randomized design** assigns subjects to different treatment groups through random assignment.
  - A **randomized block design** is sometimes used to make sure that subjects with certain characteristics are assigned to each treatment.
- A **matched pairs design** pairs up subjects according to similarities. One subject in the pair receives one treatment, while the other receives a different treatment.
- **Sample size** is the number of participants in the experiment.
- **Replication** is the ability to reproduce the experiment (and results) under similar conditions.
- A **sampling error** is the difference between the results of the sample and those of the population. Even with the best sampling techniques, this is possible.
- A **biased sample** is one that is not representative of the entire population. We want to avoid bias.
- A **random sample** is one in which every member of the population has an equal chance of being chosen.
- A **simple random sample** (SRS) is a sample in which every possible sample of the same size has the same chance of being collected.
- A **stratified sample** first splits the population into segments (called "strata"), then a predetermined number of subjects is chosen from each of the strata.
- **Cluster sampling** can be used when the population naturally falls into subgroups with similar characteristics. First, determine the clusters, then select all the members of one or more of the clusters.
- **Systematic sampling** selects sample members by choosing the first member randomly, then selecting subsequent members at regular intervals after the starting number.
- A **convenience sample** consists only of available members of the population, but this often leads to biased studies.
- A **volunteer sample** is a kind of convenience sample in which only volunteers participate.