MATH 245

Statistics
- Is the science of learning from data.
- Is a science that deals with the collection, analysis, interpretation, and presentation of data.
- Is a bunch of methods used for the collection, analysis, interpretation, and presentation of data.

Two kinds of Statistics:
1. *Descriptive Statistics* - Methods for organizing and summarizing information.
2. *Inferential Statistics* - Methods for drawing and measuring the reliability of conclusions about a *population* based on information obtained from a *sample* of the population.

- **Population** - A set of units (people, objects, transactions, or events) that we are interested in studying.
  1. Concrete populations
  2. Conceptual or hypothetical populations

- **Sample -** A part of the population from which information is collected.

Some examples of statistical problems:

1. The president of the Student Council of UWL wants to determine the proportion of the student population who are in favor of making the UWL campus a non-smoking zone.
2. A politician wants to know his chance of winning in the coming election.
3. For more than a century, normal body temperature for humans has been accepted to be 98.6^oF. Is it really? Researchers want to estimate the average temperature of healthy adults in the United States.
4. A city engineer wants to estimate the average weekly water consumption for single-family dwellings units in the city.
5. Leaders of women's rights groups want to check if there is a significant difference in the salaries of male professors and female professors.
6. A union leader wants to know if people in this state are earning less or more than those living in other states?
7. An environmentalist group wants to determine the number of deer living in a certain region or the number of fish in a lake.

   *Capture-Recapture Method***:**
   Step 1. Capture a sample of the population, mark them, and release back to the population.

   Step 2. After a certain period of time, capture another sample from the same population.

   Step 3. Compute the proportion of marked individuals, and use it to estimate the population size.

8. A pharmaceutical company wants to check if their new drug is really better than the existing one.
9. Credit card companies want to know the likelihood that a person with certain known characteristics will be able to pay a loan.
10. Medical doctors want to determine the characteristics of a patient that affect his/her likelihood of surviving a surgery.

Methods of Acquiring Information:

I.   **Public Source**: The data set of interest has already been collected and is available to the public.
  1. *Statistical Abstract of the United States* – A comprehensive summary of statistics on the social, political, and economic organization of the United States (yearly).
  2. *Survey of Current Business* – Data on the economy of the United States (monthly).
  3. *The Wall Street Journal* – Financial data.
  4. *The Sporting News* – Sports information.
  5. *The Internet*

II. **Census** – Information is obtained from the whole population.

III. **Sampling** – Information is obtained from a small group (*sample*) of objects/individuals taken from the *population*. The sample should be a *representative sample*, that is, it should reflect as closely as possible the relevant characteristics of the population under consideration.

   Simple Random Sampling – is a sampling procedure for which each possible sample of a given size is equally likely to be the one obtained. A sample obtained in this way is called a *Simple Random Sample* (SRS).

   - *Observational Study* – researchers simply observe characteristics and take measurements, as in a sample survey. Can only establish **association**.
   - *Designed Experiment* – researchers impose treatments and controls and then observe characteristics and take measurements. Can establish **causal link**.

Other Common Sampling Designs:

1. **Systematic Random Sampling**.
   Step 1. Divide the population size by the sample size and round the result down to the nearest whole number, m.
   Step 2. Use a random-number generator (table, computer or any similar device) to obtain a number k, between 1 and m.
   Step 3. Select for the sample those numbers of the population that are numbered k, k+m, k+2m, ...

2. **Cluster Sampling**.
   Step 1. Divide the population into groups (clusters).
   Step 2. Obtain a simple random sample of the clusters.
   Step 3. Use all the members of the clusters in Step 2 as the sample.

3. **Stratified Random Sampling with Proportional Allocation**.
   Step 1. Divide the population into subpopulations (strata)
   Step 2. From each stratum, obtain a simple random sample of size proportional to the size of the stratum; that is, the sample size for a stratum equals the total sample size times the stratum size divided by the population size.
   Step 3. Use all the members obtained in Step 2 as the sample.

Homework: Read Section 1.1.