The Nature of Statistics

“Statistics” First appeared in the English language in 1787.

Statistical Thinking Will One Day Be As Necessary For Efficient Citizenship as The Ability To Read & Write.

~ H. G. Wells

Lies, Damn Lies, & Statistics

~ Benjamin Disraeli –or– (Mark Twain)
   (1804-1881)   (1906)

http://www.york.ac.uk/depts/mathsmaths/histstat/lies.htm
What is Statistics?

DECEMBER 3, 2012

What is Statistics?

When many people hear the word “statistics,” they think of either sports-related numbers or the college class they took and barely passed. While statistics can be thought about in these terms, there is more to the relationship between you and statistics than you probably imagine.

So, what is statistics? Several informal definitions are offered in the book A Career in Statistics: Beyond the Numbers by Gerald Hahn and Necip Doganaksoy:

- The science of learning from (or making sense out of) data
- The theory and methods of extracting information from observational data for solving real-world problems
- The science of uncertainty
- The quintessential interdisciplinary science
- The art of telling a story with [numerical] data
Two Kinds of Statistics

Descriptive

Consists of methods for organizing and summarizing information in a clear & effective way.

Examples:

* Comparison of motorcar sales of GM v/s Ford
* Air pollution levels in different cities
* Voting results of the 2000 Presidential elections

Inferential

Consists of methods of drawing conclusions about a population based on information obtained from a sample of the population.

Examples:

* Political polling
* Experiment to determine the number of fish in a body of water
* Experiment to estimate the life of a GE light bulb
* Evaluation of Polio vaccine --- Does it work?

What is Population?

The collection of all individuals, items, or data under consideration in a statistical study.

What is a Sample?

That part of the population from which information is collected.

What is a Parameter?

A numerical measurement describing some characteristic of a population. Generally denoted by a Greek letter.

What is a Statistic?

A numerical measurement describing some characteristic of a sample. Generally denoted by Roman letters.
Development Of Inferential Statistics

Pierre de Fermat (1601-1665)
  Co-inventor of analytic geometry
  Number theory
  Contributed to probability theory

Blaise Pascal (1623-1662)
  Theory of conic sections
  Invented the calculating machine
  Contributed to probability theory

Jakob Bernoulli (1634-1705)
  Leibnitz calculus

Pierre-Simon, Marquis de La Place (1749-1827)
  Physics
  Celestial mechanics
  Pure mathematics

Lambert Adolphe Jules Quetlet (1796-1874)
  Mathematical sociology

Francis Galton (1822-1911)
  Fingerprints

Karl Pearson (1857-1936)
  Founder of 20th century statistics

William Sealy Gosset (1876-1937)
  Small sample theory -- t-ratio

Leonard Aylmer Fisher (1890-1962)
  Theory of estimation & experimental design
  often called the father of modern statistics
2013 International Year of Statistics

http://www.statistics2013.org/

http://www.youtube.com/watch?v=nTBZuQR7dRc&feature=player_embedded
Views of Students and Faculty on “What constitutes a Good Statistics Department?”

A country’s evaluation of Statistics as a department and a field of study:
<< page 3 >>

Examples of Areas of Statistical Research

- Applied Statistics
- Medical Statistics
- Financial Statistics
- Bayesian and Computational Statistics
- Applied Probability and Operations Research
- Extreme Value


What makes a Statistician? Why Statistics from a Statistician?

Getting A Job as a Statistician

http://www.youtube.com/watch?v=D4FQsYTBLoI&feature=player_embedded
http://www.amstat.org/careers/index.cfm

The Statistical Revolution

http://www.youtube.com/watch?v=oGGYIw_pIj8

Examples of Questions We Will Answer in STA308
Simple Random Sampling

Definition: A simple-random-sampling procedure is a sampling procedure for which each possible sample is equally likely to be the one selected. A sample obtained by the simple-random-sampling procedure is called a simple random sample.

Using The Random-Number Tables
Other Sampling Procedures

1. Systematic Random Sampling
   a. Divide the population size by the sample size and **round down** to the nearest whole number, m
   b. Use a random-number table (or a similar device) to obtain a number, k, between 1 and m.
   c. Select for the sample those members of the population that are numbered, k, k+m, k+2m, …

2. Cluster Sampling
   a. Divide the population into groups (clusters)
   b. Obtain a simple random sample of the clusters
   c. Use ALL the members of the clusters as the sample

3. Stratified Sampling
   a. Divide the population into “strata”
   b. Sample within each “strata”

4. MultiStage Sampling
   a. Combination of sampling methods
Experimental Design

Observational Studies V/S Designed Experiments

- In an observational study researchers simply observe characteristics and take measurements.
- In a designed experiment researchers impose treatments and controls and then observe characteristics and take measurements.

Principles of Experimental Design

1. Control
2. Randomization
3. Replication

Terminology of Experimental Design

1. Experimental Units / Subjects
2. Response Variable: The characteristic of the experimental outcome that is measured or observed.
3. Factor: A variable whose effect on the response variable is of interest in the experiment
4. Levels: The possible values of a factor.
5. Treatment: Each experimental condition.
Statistical Designs

1. Completely Randomized Design – all experimental units are assigned randomly among all the treatments
2. Randomized Block Design – experimental units are assigned randomly among all the treatments separately within each block.