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/maths/histstat/lies.htm

What is Statistics?

DECEMBER 3, 2012

in 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:
http://statmath.wu.ac.at/research/talks/resources/YildizEvren2009.pdf
<< 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

http://www.statistics2013.org/files/2012/12/STAT2013Poster.pdf

What makes a Statistician? Why Statistics from a Statistician?

http://www.nkd-group.com/sta308/readings/statistics and statistician.pdf

Getting A Job as a Statistician

http://www.nkd-group.com/sta308/notes/gettingajobinstatistics.pdf

http://www.youtube.com/watch?v=D4FQsYTbLoI&feature=player_embedded
http://www.amstat.org/careers/index.cfm

The Statistical Revolution

http://www.nkd-group.com/sta308/notes/nextgenerationstatisticians.pdf

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

Examples of Questions We Will Answer in STA308

http://www.statistics2013.org/2012/11/26/winatschoolquiz/

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

Line number	Column number									
	00–09		*10–19*		*20–29*		*30–39*		*40–49*	
00	15544	80712	97742	21500	97081	42451	50623	56071	28882	28739
01	01011	21285	04729	39986	73150	31548	30168	76189	56996	19210
02	47435	53308	40718	29050	74858	64517	93573	51058	68501	42723
03	91312	75137	86274	59834	69844	19853	06917	17413	44474	86530
04	12775	08768	80791	16298	22934	09630	98862	39746	64623	32768
05	31466	43761	94872	92230	52367	13205	38634	55882	77518	36252
06	09300	43847	40881	51243	97810	18903	53914	31688	06220	40422
07	73582	13810	57784	72454	68997	72229	30340	08844	53924	89630
08	11092	81392	58189	22697	41063	09451	09789	00637	06450	85990
09	93322	98567	00116	35605	66790	52965	62877	21740	56476	49296
10	80134	12484	67089	08674	70753	90959	45842	59844	45214	36505
11	97888	31797	95037	84400	76041	96668	75920	68482	56855	97417
12	92612	27082	59459	69380	98654	20407	88151	56263	27126	63797
13	72744	45586	43279	44218	83638	05422	00995	70217	78925	39097
14	96256	70653	45285	26293	78305	80252	03625	40159	68760	84716
15	07851	47452	66742	83331	54701	06573	98169	37499	67756	68301
16	25594	41552	96475	56151	02089	33748	65289	89956	89559	33687
17	65358	15155	59374	80940	03411	94656	69440	47156	77115	99463
18	09402	31008	53424	21928	02198	61201	02457	87214	59750	51330
19	97424	90765	01634	37328	41243	33564	17884	94747	93650	77668

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

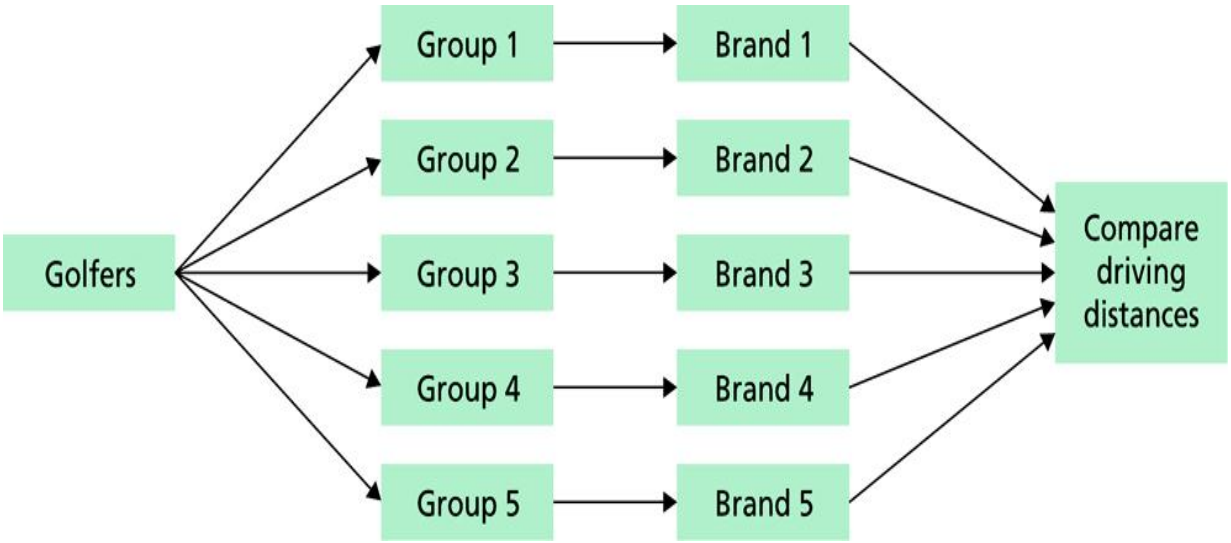
1. Control
2. Randomization
3. Replication

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.