

Introduction: Statistics, Data and Statistical Thinking

FREC 408
Dr. Tom Ilvento
213 Townsend Hall
ilvento@udel.edu
<http://www.udel.edu/FREC/ilvento>

The First Measured Century



<http://www.pbs.org/fmc/index.htm>

Statistics

- **Statistics** (Def 1.1 p24) is the science of data
- It refers to
 - Collecting data
 - Classifying, summarizing, and organizing data
 - Analysis of data
 - Interpretation of data

Statistics

- Statistics is both a field of study
- ...and a set of tools used by many disciplines
 - Social Sciences
 - Biological Sciences
 - Physical Sciences

We will focus on two types of statistical applications

- Descriptive
- Inferential

Descriptive Statistics

- Descriptive statistics uses summary measures, graphs, and measures of association to show relationships in data.
- The focus is on describing the data
- With an emphasis on **parsimony**

Descriptive Statistics

- Rather than looking at a set of numbers,
- 0, 0, 2, 2, 3, 3, 3, 4, 5, 2, 1, 3, 2, 2, 1, 1, 3, 1, 1, 2, 5, 7, 8, 10, 12

Descriptive Statistics

- we want to find summary measures which describe the data adequately and succinctly
- Be they a
 - Percentage
 - Average
 - Range from highest to lowest
 - mode

Descriptive Statistics

- Descriptive Statistics also involve relationships between variables or sets of variables
- And they can involve very sophisticated techniques – regression, principle components, factor analysis, Logistic Regression, Probit Analysis

Inferential Statistics

- Inferential statistics takes it a step further
- Now we use some of the same techniques to make estimates, decisions, predictions, or generalizations about a population from a smaller subset or sample

Inferential Statistics

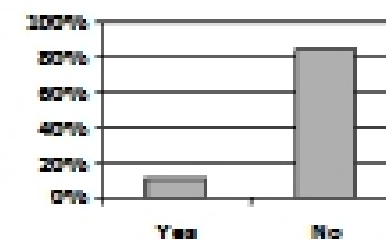
- Inferential statistics are a powerful tool for research
- It enables us to make statements about a large group from a much smaller sample.
 - We can survey 1,000 people and make statements about 280 million people

Did the public care if George W. Bush used cocaine in his 20s?

A Time/CNN Poll found:

If Bush did use cocaine in his 20s, should that disqualify him from being President?

Yes 11%
No 89%



Let's look closer at this survey example

- It was based on a telephone poll of 942 adult Americans taken for Time/CNN on August 19th by Yankelovich Partners, Inc.
- The sampling error is $\pm 3.3\%$
- What does this mean?

Here's my interpretation

- The survey is designed to represent adult Americans in August of 1999
- Because we are taking a sample, we have some error associated with our estimate.

Here's my interpretation

- Since the sample was taken randomly, we have a method to estimate the error of our estimate
- In this case, we are reasonably sure that the true percentage is within $\pm 3.3\%$ points of our estimate
- Which means our interval is 7.7% to 14.3%

We need some terms

- A **Population** (or **U** or **P₂₈**) is the total number of units involved in the research question. The units are the members (or elements) of the population.
- Populations could be:
 - People
 - Animals
 - Plants
 - Courses
 - Objects

A POPULATION IS DEFINED BY

- Purpose of the study
- The units and elements involved
- Geographic coverage
- Time frame

Population Example

- If I was interested in understanding current household consumption of chicken in the Mid-Atlantic states, I might define the population as:
- All households in in the Mid-Atlantic states (DE, MD, PA, NJ, NY) in the Fall of 2002