Introduction to Statistics

Dr Tom Ilvento
Department of Food and Resource Economics
University of Delaware

Overview

- A brief introduction to the field of statistics
- Statistics as both a field of study and a set of tools used by many other disciplines
- The distinction between descriptive and inferential statistics

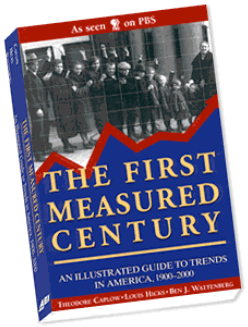
Why should you be required to take a statistics course?

- We are surrounded by data!
  - Voting behavior
  - Medical research
  - The state of our economy
- Making sense of this data is a challenge
- Most research depends upon the use of statistics to reach a conclusion

Two thoughts I want to share up front in this course

- Just because you see something in your data, doesn't mean it is so
- We often are data rich.... and model poor.

The First Measured Century



- http://www.pbs.org/fmc/index.htm

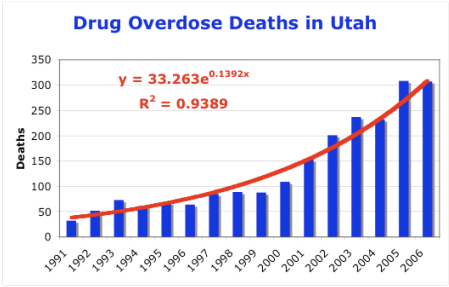
Americans work harder than they used to, so they have a lot less free time.

Yes or No?

- **NO!! We have more free time.**
- The change in the manufacturing work week is typical of the change, which reflects
  - Mechanization
  - The state of the economy
  - Labor and social movements
- We went from 10 hour days six days a week to the now familiar 40 hour week five days a week

Source: The First Measured Century

We are surrounding by data



What is Statistics?

- Statistics is the science of data
- It refers to every aspect of how we handle and use data
  - Collecting data
  - Classifying, summarizing, and organizing data
  - Analysis of data using summary measures and graphs
  - Interpretation of the results

Statistics

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

We will focus on two types of statistical applications

- Descriptive

Stem and Leaf Plot

6	5
7	5
8	0 0 0 5 5 5
9	0 0 0 5 5 5 5 5 5 5 8 8
10	0 0 0 0 5 5 5
11	0 0 0 0 5
12	0 0 0 5
13	0

stem unit 10's, leaf unit ones

- Inferential

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	98.67568	93.84142	103.5099	0.950
Std Dev	14.49915	11.79086	18.83382	

Descriptive Statistics

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

Parsimony - principle of economy: economy in the use of means to achieve something, especially the principle of endorsing the simplest explanation that covers a case

Look at the following data

- This is a small data set of the age of 100 randomly sampled drivers killed in traffic accidents
- Rather than looking at a set of numbers, we need an analysis strategy

16	17	18	20	24	28	34	44	53	73
16	17	18	20	24	28	35	44	56	74
16	17	18	20	24	29	36	45	57	76
16	17	18	21	24	29	37	45	58	76
16	17	18	21	25	30	37	45	65	77
16	17	18	21	26	30	38	45	69	81
16	18	18	22	27	31	38	51	69	86
16	18	18	23	27	31	38	51	70	87
17	18	18	24	27	32	40	51	73	87
17	18	19	24	28	34	42	52	73	88

Basic Descriptive Statistics

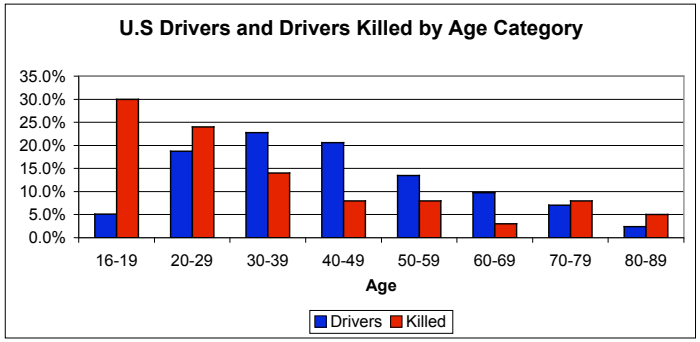
These sample statistics are from Excel

Age	
Mean	**35.9**
Standard Error	2.1
Median	**28.0**
Mode	**18.0**
Standard Deviation	**21.1**
Sample Variance	**447.1**
Kurtosis	0.0
Skewness	1.1
Range	72.0
Minimum	16.0
Maximum	88.0
Sum	3587.0
Count	**100.0**

Table Percentage by Age of a random sample of drivers killed in a car accident

Age Categories	Sample of Drivers Killed	Percent
16-19	30	30.0%
20-29	24	24.0%
30-39	14	14.0%
40-49	8	8.0%
50-59	8	8.0%
60-69	3	3.0%
70-79	8	8.0%
80-89	5	5.0%
TOTAL	100	

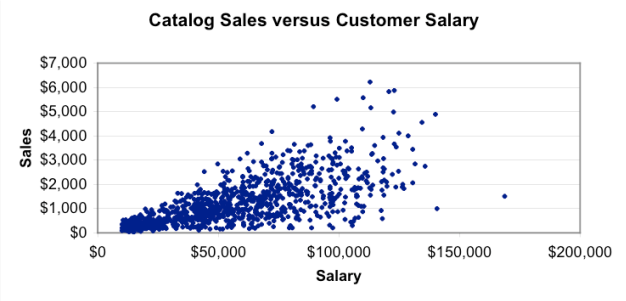
A Graph of the Table Data



Descriptive statistics can be powerful tools

- Descriptive Statistics can also involve relationships between variables or sets of variables
- And involve very sophisticated techniques – cross-tabulations and chi square analysis; ANOVA, regression, correlation, factor analysis, and logistic regression

A Scatterplot of Catalog Sales versus Customer Salary



Inferential Statistics

- Inferential statistics is concerned with making inferences from a sample to a population
- Whenever we deal with a sample, we deal with estimates and uncertainty
- Statistical inference helps us place our estimates in a probability framework that allows us to make conclusions from the data

Just because you see something in your sample data doesn't mean it is real!

Inferential Statistics

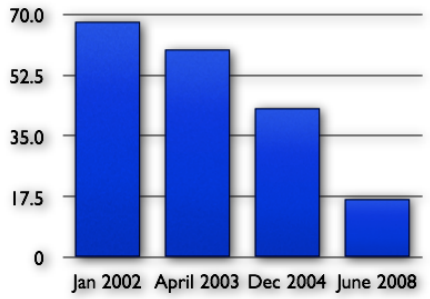
- Inferential statistics are a powerful tool for research
- It enables us to make statements about a large group (population) from a much smaller sample.
  - Survey
  - Designed Experiments
  - Observational studies

Based on principles of inferential statistics, we can survey 1,000 people and make statements about a population of 300 million

Is the country headed in the right direction?

Associated Press-Ipsos poll conducted by Ipsos Public Affairs. N=1,000 adults nationwide. **MoE ± 3.1.**

Generally speaking, would you say things in this country are heading in the right direction, or are they off on the wrong track?



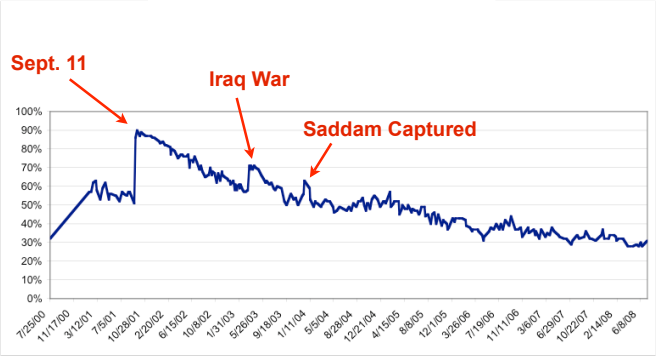
Let's look more closely at this survey example

- It was based on a telephone poll of 1,000 adult Americans taken for the Associated Press and conducted by Ipsos Public Affairs. June 12-16, 2008.
- *Generally speaking, would you say things in this country are heading in the right direction, or are they off on the wrong track?*
- **The Margin of Error (MoE) is ± 3.1%**

Here is my interpretation

- The survey is designed to represent adult Americans in June of 2008.
- **The estimate for the percent Right Direction is 17%**
- Because we are taking a sample, we have some error associated with our estimate.
- Since the sample was taken randomly, we have a method to estimate the error of our estimate
- In this case, we are reasonably sure **(95% sure) that the true percentage is within ± 3.1% points of our estimate**
- **Which means our interval inference is 13.9% to 20.1%**

What is this graph about?



Summary

- Statistics are an important part of research in many fields
- All statistics should be descriptive in nature to help us describe data
- Some statistics have inferential properties that help us make statements about a population from a sample