

What Is/Are Statistics?

Ray Block, Jr.
PS 585
Research Methods

Today's Blueprint

Last Class

- Research Ethics
 - Definitions
 - Documents, Regulations, and Guidelines
 - Roles and Responsibilities

Today's Class

- Statistics
 - What it "is"
 - What they "are"
 - How to lie with them
 - How to tell the truth with them

What "Is" Statistics?

What "Is" Statistics?

- Statistics: Word origins
 - *Statisticus* (New Latin): "state craft" of "state affairs"
 - *Statistik* (German): the science of politics
 - *Statista* (Italian): person skilled in statecraft
 - *Stato* (Old Italian): state; of the state
 - *Status* (Old Italian): position, form of government

What "Is" Statistics?

- Statistics = State-istics:
 - The collection of information vital to the state
 - The study of political, economic, and population facts and figures

What "Is" Statistics?

- The world is an uncertain place
 - We don't know everything
 - If we are lucky, we at least know how much we don't know
 - We can never be fully certain of anything

What "Is" Statistics?

- Total uncertainty is bad
- But some certainty (even if just a little) is better than no certainty at all

What "Is" Statistics?

- Conducting research in an uncertain world means dealing with uncertainty
- Statistics allow us to deal with this uncertainty

What "Is" Statistics?

- Dealing with the world's uncertainty requires skills in the following 3 areas:
 - *Data analysis*: Gathering, displaying, & summarizing
 - *Probability*: Laws of chance
 - *Statistical Inference*: Drawing conclusions based on properties 1 and 2

What "Are" Statistics?

What “Are” Statistics?

- The goal is to describe what is going on in a population without actually observing the population
- To do so, we base all we know on samples
- Samples don’t tell us everything, but they are at least better than total uncertainty

What “Are” Statistics?

- Statistics are numbers describing sample characteristics
- What purpose do they serve?
 - Statistics summarize:
 - The distributions of values on variables
 - The relationships between variables

How to Lie with Statistics

How to Lie with Statistics

- The Case of the Disappearing Baseline
- The Pictogram Trap
- Never mind the data – look at the peaks!
- Where’s the data?
- Design dominates the data
- The effect of 3-D shading

How to Lie with Statistics

- In Class Exercise:
 - We will preview an example of each
 - We will then discuss each example in turn
 - Then we will look at some examples of “honest” graphs

[Examples 1 thru 6 Here]

How to Lie with Statistics

- Disappearing Baseline:
 - Day Mines Inc., *1974 Annual report* (Reported in Tufte, E. 1983, p. 54.)

How to Lie with Statistics

- Problem:
 - There is no vertical scale
- Suggestion:
 - Should use clear, detailed, and thorough labeling to avoid graphical distortion and ambiguity.

How to Lie with Statistics

- Pictogram Trap:
 - Drinking up - Australian wine exports, *The Age*, 27 April 1998.

How to Lie with Statistics

- Problem:
 - Pictures distort story
- Suggestion:
 - Pictures should be proportional to the numerical quantities represented

How to Lie with Statistics

- Look at the peaks
 - Bank of Melbourne Peak rates advertising brochure

How to Lie with Statistics

- Problem:

- Pictures distort story
- Suggestion:
 - Fix proportions

How to Lie with Statistics

- Where's the data?
 - Language spoken at home, from the Brunswick Sentinel, August 1, 1994.

How to Lie with Statistics

- Problem:
 - There is so much detail that you can hardly see where the data bars end
- Suggestion:
 - Avoid chart junk

How to Lie with Statistics

- Design Dominates Data
 - Athens Olympic contract (Organizing Committee for Olympic Games, Athens 2004)

How to Lie with Statistics

- Problem:
 - Observations are obscured by columns
- Suggestion:
 - Avoid chart junk (excess grids, lines, etc.)

How to Lie with Statistics

- 3-D Shading
 - New York State Budget data, 1966-1978

How to Lie with Statistics

- Problem:
 - 3-D scaling over-emphasizes the difference between the lowest and highest bars
- Suggestion:
 - Use 2-D when appropriate

Telling the Truth with Statistics

Telling the Truth with Statistics

Two Easy Rules to Live By:

- Never quote data out of context
- Scales should correctly represent data

Telling the Truth with Statistics

- Quoting Stats Out of Context:
 - Connecticut Traffic Deaths before (1955) and After (1956)
 - Stricter Law Enforcement by the Police Against Exceeding the Speed Limit

Telling the Truth with Statistics

- Problem:
 - Exaggerates the [alleged] negative causal relationship between law enforcement and traffic deaths
- Suggestion:
 - Need to know the context

Telling the Truth with Statistics

- Choice of scales:
 - Australian stock exchange

- Fluctuations in a particular stock prices (BHP)

Telling the Truth with Statistics

- Problem:
 - Shows changes in prices in one day (by hour)
- Suggestion:
 - Modify scales to better show trends

To Recap:

- Statistics is a mathematical tool used to conduct research under less than certain conditions
- Statistics are numbers used to talk about samples and make inferences about populations
- People can use statistics (both the approach and the numbers) for honest or dishonest reasons

References

- FYI:
 - Huff, Darrell and Irving Geis. 1954. How to Lie with Statistics. New York, NY: W.W. Norton & Company, Inc.
 - Paine, Albert B. (Ed). 1924. Mark Twain's Autobiography, Volume 1. New York and London: Harper Brothers.
 - Manheim, Jarol B., Richard Rich, and Lars Willant. 2001. Empirical Political Analysis: Research Methods in Political Science, 5th Edition. New York, NY: Longman Publishers.
 - Walsh, Anthony and Jane C. Ollenburger. 2001. Essential Statistics for the Social and Behavioral Sciences: A Conceptual Approach. Upper Saddle River, NJ: Prentice Hall, Inc.
 - Wonnacott, Thomas H. and Ronald J. Wonnacott. 1990. Introductory Statistics, 5th Edition. New York, NY: John Wiley & Sons.