

Statistical Traps and Trickery

Week 4: Calling Bullshit

Thoughts on Last Week's Reading?

- Robert Mathews. Storks deliver babies ($p=0.008$). *Teaching Statistics* 22:36-38, 2000.
- Case study: Traffic improvements
- Example of logical fallacies?

Thoughts on Sharia Law in the Courts

(referencing last week's discussion)

-
- “[Ban on Sharia Law](#)” and “[Sharia](#)” from Wikipedia (good starting points)
 - “[Sharia](#)” from Teaching Tolerance, a project of the Southern Poverty Law Center

 - Please consider going to [Factcheck.org](#), [Snopes.com](#), and [Politifact.com](#) and searching for “sharia.”

Three Ways to Spot a Bad Statistic

Base Rate Fallacy: mistakenly judge the likelihood of a situation by not taking into account all relevant data

Thought Experiment:

Suppose that the rate of disease D is three times higher among homosexuals than among heterosexuals, that is, the percentage of homosexuals who have D is three times the percentage of heterosexuals who have it. Suppose, further, that Pat is diagnosed with the disease, and this is all that you know about Pat. In particular, you don't know anything else about Pat's sexual orientation; in fact, you don't even know whether Pat is male or female. What is the likelihood that Pat is homosexual?

Prosecutor's Fallacy

A fallacy of statistical reasoning that assumes the prior probability of a random match is equal to the probability that the defendant is guilty.

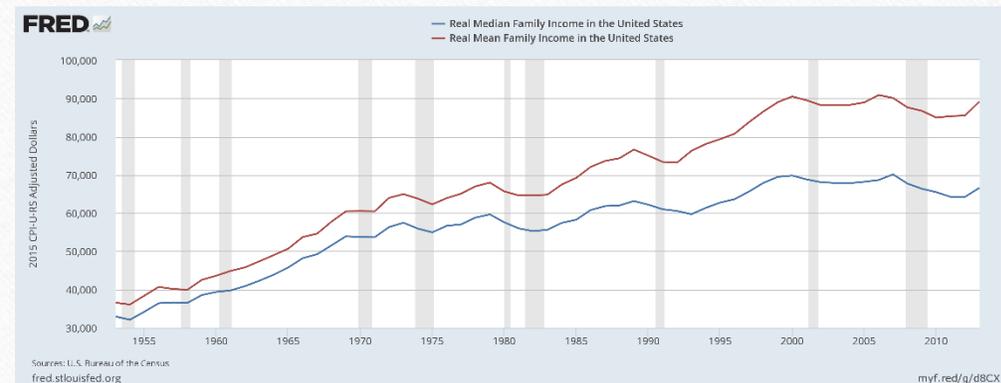
Example:

- Sally Clark was a British solicitor, and in 1996 his first son died, only weeks after his birth. In 1998 the same pattern occurred, when his second son died only weeks after his birth. Clark was tried for the murder of her two sons, and the prosecutor used the following argument: “According to an expert pediatrician, the chance of a newborn dying in a similar circumstance is 1 in 8500. So the chance of two newborns of the same mother dying in this same manner is $1/8500 * 1/8500$, which is equal to 1 in 72,250,000. In other words, this event is extremely unlikely and therefore Sally Clark must be guilty of murder.” The argument seemed plausible, and Clark was convicted and went to jail. She served over one year before people started questioning the statistics involved in her trial. In 2001 her process was revised and she was released.

What is “Average” in a statistic?

The “mean” is the “average” you’re used to, where you add up all the numbers and then divide by the number of numbers. The “median” is the “middle” value in the list of numbers. You want to use the median to describe the middle of a set of data that has an outlier.

Example on right: U.S. Bureau of the Census



Assigned Reading

- [Simpson's paradox](#): an interactive data visualization from VUDlab at UC Berkeley.
- Case studies: [Musicians and mortality](#) and [Track Records](#)
- *Optional* (academic article): Koehler, Jonathan J. [One in Millions, Billions and Trillions: Lessons from the People V. Collins \(1968\) for People V. Simpson \(1995\)](#), Journal of Legal Education (V. 47).

Next Week: Visualizations

Data can be powerful tools. Images and their origins.