Introduction to Statistics I

Textbook: Elementary Statistics (4th Edition, by Navidi and Monk), and Workshop Statistics (4th Edition, by Rossman and Chance).

Previous Lecture

- Explanatory/Response/Confounding/Lurking Vars
- Observational Studies & Cause/Effect
- Anecdotal Evidence
- Experimental design: Random assignment
- Placebo/Blindness
- Control/Treatment groups

§1.4: Bias in Studies

Example: Elvis Presley died on August 16th, 1977.
Pretend you are in 1989, when there was a conspiracy theory that Elvis faked his death.

RQ: What % of American adults believe Elvis Presley faked his death?

Obv-unit?/var/var type?

Sampling

What's best way to gather this data?

Instead, listeners of 100+ radio stations called a 900 number ($2.50 per call) to voice their opinion of whether Elvis was really dead.

Will this **sampling method** answer the original RQ?

Example: Call-in survey found 56% of callers thought Elvis was still alive. Stat/Param?
Does this reflect the views of the population?

56% was calculated from people who called in (sample), and so is a statistic.

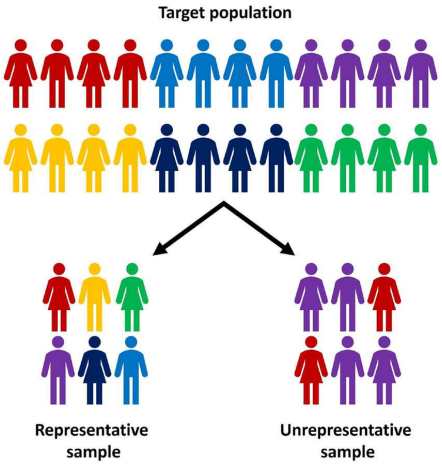
True % of Americans who believed Elvis was alive in 1989 is the paramater.

Parameter is unknown!

Ideally, the statistic gives us knowledge about the parameter.

Drawing Conclusions

With our sample, can we draw a conclusion?



Sample Bias: For our statistic to be a good estimate of our parameter, our sample must be **representative** of population.

If sampling is done in a way that makes the sample *not* representative (it systematically over- or under-represents certain segments of population), then the sampling method is **biased**.

What leads to biased samples?

- Sampling from only a (nonrepresentative) part of the population

- Convenience Samples (lazy sampling)

- Voluntary Samples (self-selection)

- Allowing non-response

These result in members of the population having unequal chances of being in the sample.

Ideally you have a list of *all* obv-units in population!
This list (if it exists) is called the **sampling frame**.

Sampling frame: boys/girls

Having a **representative sample** is a necessary ingredient in drawing good conclusions.

Other sources of bias:

Social Acceptability Bias: "Do you tend to eat healthily?" versus

"What challenges to eating healthfully have you faced this year?"

Or, "What sources of whole grains do you consume on a daily basis?"

Leading Question Bias:

"Given his awful position on policy A, will you vote for candidate X?" versus just "Will you vote for candidate X?"

What did we learn?

- Sampling from a population
- Sampling/question bias