

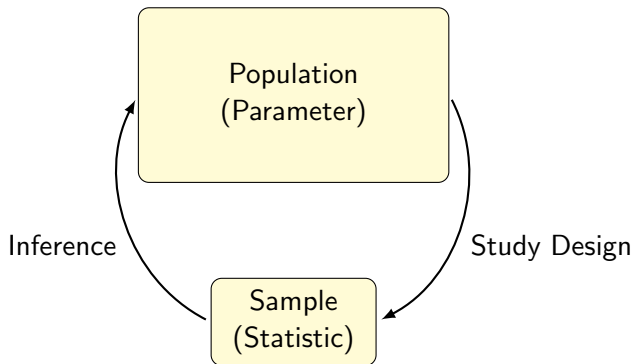
Study Design and Bias

February 02, 2021

In the previous class, we considered topics in the domain of descriptive statistics:

- Measures of centrality
- Measures of dispersion
- Measures of association
- Histograms, box plots, and scatter plots

Statistical Process

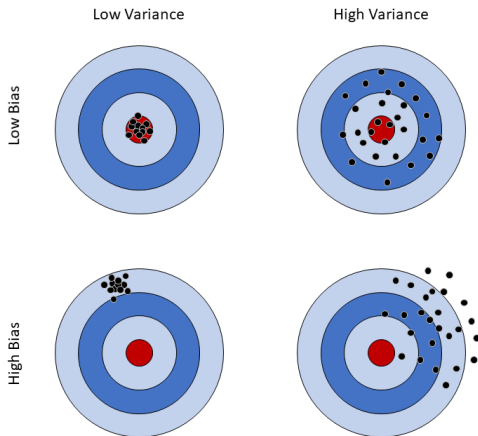


Variance and Bias

Suppose a population has a true mean value of μ which we hope to estimate by taking a sample and determining the sample mean \bar{X} . Here, we are referring to the variance and bias of the statistic rather than the population.

- The *variance* is a measure of dispersion. If variance is high, the range of values \bar{X} may fall in tends to be large, though we are just as likely to overestimate the value as we are to underestimate it. At any rate, the expected value is $E(\bar{X}) = \mu$
- The *bias*, on the other hand, is a systematic departure from the true value. If our statistic is biased, regardless of sample size, we have $E(\bar{X}) = \mu + \text{Bias}$

Variance and Bias



Sample Bias

- The 1936 US Presidential election between incumbent FDR and Republican Alfred Landon, six years into the Great Depression.
- The *Literary Digest* magazine, which had correctly predicted each election since 1916, mailed 10 million questionnaires to addresses gathered from telephone books and club membership lists, with 2.4 million responses returned
- The *Digest* predicted a landslide victory for Landon: 57% to 43%
- In the actual election, Roosevelt won 62% to 38%

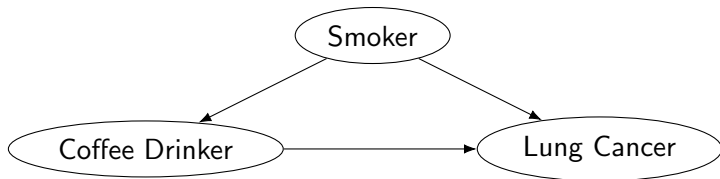
Extrapolation Bias

For practical, ethical, and economic reasons clinical trials of new treatments usually only involve adults, with only about 25% of drugs subject to pediatric studies.

Propofol is a sedative that has consistently been shown to be safe in adults. In 1992, the British government recommended against using it on patients under 16 after a number of children receiving propofol died in the ICU. A 2001 controlled trial found that 9.5% of children died after receiving propofol, compared with 3.8% on a different sedative

Confounding

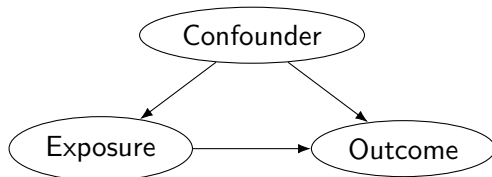
- Many epidemiologic studies have shown that coffee drinkers have an increased risk of lung cancer
- However, what researchers also noticed is that smokers are more likely to drink coffee



- Once researchers controlled for smoking status, they no longer found a change in lung cancer risk due to drinking coffee ([article](#))

Confounding

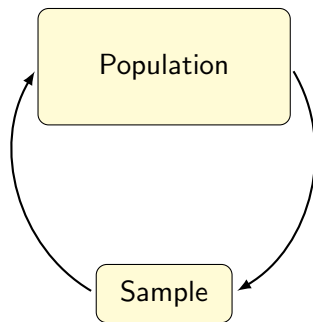
- A confounder (lurking variable) is a third variable, which is related to both exposure and outcome
- Because of this, confounders distort the relationship between exposure and outcome



Ideal Study Design

Ideally, for a given population we can:

- Obtain a list of everyone in the population
- Select from this list completely at random
- Each person is equally difficult/costly to sample
- Everybody selected responds to the survey



Randomized & Controlled

The "gold standard" of study design is a double-blind randomized and controlled clinical trial

- Subjects are assigned to treatment or control groups *at random*
- Randomized groups should be as similar as possible
- Neither investigator nor subjects know which group they are in with use of placebo
- Intent to Treat (ITT) analyzes results based on treatment assigned rather than treatment received

Intent to Treat

The Coronary Drug Project Research Group published an article in the *New England Journal of Medicine* (1980) describing a randomized controlled double-blind experiment involving the drug clofibrate, which reduces the level of cholesterol in the blood

	Clofibrate	
	Number	Deaths
Adherers	708	15%
Nonadherers	357	25%
Total	1,103	20%

Subjects who took more than 80% of their prescribed medicine were called “adherers”

Clofibrate and Placebo Results

	Clofibrate		Placebo	
	Number	Deaths	Number	Deaths
Adherers	708	15%	1,813	15%
Nonadherers	357	25%	882	28%
Total	1,103	20%	2,789	21%

- Taking into account the placebo results as well, clofibrate no longer looks effective
- One possibility is that adherers are more concerned with their health, and take better care of themselves in general
- Take-home message: comparing subjects *as they were randomized* is the only completely valid way of carrying out a controlled experiment; all other comparisons are subject to confounding and bias

Summary

- Samples should be representative of population in question
- Bias occurs when a (little s) statistic systematically differs from the population parameter
 - i. Sample and selection bias
 - ii. Non-response bias
 - iii. Extrapolation bias
 - iv. Confounding
- Well designed studies are conducted in such a way as to minimize the potential for bias