Lecture 1 - Key

Selection Bias

A sampling scheme displays **bias** *if the sample systematically favors certain parts of the population over other parts. If certain parts of the population are overrepresented in the sample, while other parts are underrepresented, then sampling bias exists.*

Selection bias occurs whenever the researchers select their sample in a way that makes it unrepresentative of the population. Lohr states *"Selection bias occurs when some part of the target population is not in the sampled population, or, more generally, when some population units are sampled at a different rate than intended by the investigator."*

Selection of whichever individuals are easiest to reach is called *convenience sampling. The resulting sample (not surprisingly) is called a sample of convenience. This will often lead to selection bias.*

Give an example of a sampling scheme for the Bridger Bowl question that is a convenience survey that will lead to selection bias.

How might political polling using landlines be subject to selection bias?

Participation bias often occurs in *voluntary response surveys (or self-selected surveys) in which people decide for themselves whether or not to be included in the survey. That is, the participants are not selected by the researchers conducting the study.*

Voluntary (self-selected) responses tend to *over-represent certain portions of the population, and are commonly form people with very strong opinions (often negative opinions) about an issue or they desire a change from the status quo. The opinions of the respondents are unlikely to accurately represent the opinions of the majority which is less emotionally attached to the issue.*

Selection bias can also occur when a *judgement sample is taken. A judgement sample occurs when the researcher uses his/her so-called 'expert' judgement to select what he/she judges to be a representative sample.*

- For example, a wildlife biologist decides to collect information about a bird species from only those areas that she considers to be typical or representative of the habitat type the bird would be found in.

A sample suffers from *undercoverage if the sample fails to represent the target population because it is not possible for specific member of the population to be included in the sample.*

- You want to estimate the size of the elk population in a study area in Yellowstone. Certain areas are too remote to access and so are not in the sample by choice (not by random design). From the data that is collected, you provide an estimate for the entire study area (which includes the inaccessible parts).

A sample suffers from *overcoverage} when sampling units not in the target population are included in the sample.*

- A local newspaper wants to estimate the proportion of Gallatin County voters who support a referendum to introduce a 'tourist tax' which imposes a sales tax on hotels and goods commonly purchased by tourists. The voters are required to vote online through the newspapers website. There is no way to guarantee the online participants are eligible voters in Gallatin Counts.

Nonresponse and Measurement Errors

Nonresponse errors occur when the sampling unit *does not yield a response of interest*

- A questionnaire respondent refuses to provide certain information (e.g., regarding age, income, gender, ect. . .) or refuses to participate at all (e.g., phone surveys).
- The measurement device breaks down. It gets too dark to take measurements. You run out of money, You fail to find the sampling unit.

The problem of nonresponse is *not solved by starting with an excess number of cases to allow for a certain level of nonresponse. A sample is no longer a probability sample if it is affected by nonresponse. There is no perfect substitute for response.*

However, if certain cases are strongly correlated with a tendency for nonresponse, *it would not be surprising for nonresponse to seriously bias results and inferences about the population.*

Suppose the question on a survey is 'Have you ever driven while intoxicated?' I would expect that those who have actually driven while intoxicated are less likey to respond than thos who have not.

Note that bias due to nonresponse is not a form of selection bias. *That is, the bias does not arise from the procedure used for collecting the sample. It occurs after sampling units are selected.*

Measurement errors or observational errors *occur when the information from a sampling unit is faulty (does not yield the true response).* - In a study area, the presence of a bird's nest is missed and 'absence' is recorded. This also happens when studying elusive populations. - The measuring device is uncalibrated or its user is improperly trained. - People lie when asked about sensitive issues (e.g., drug use, criminal history, income, voting preference, etc..)

Response bias occurs when the responses taken from the sampling units tend to differ from the truth in one particular direction. *That is, there is a tendency to have either more overstated values or more understated values in the sample in comparison to the true values.* - A field ecologist records an estimate of the distance between a snowmobile and a group of elk. This ecologist tends to underestimate the true distance, so the recorded distances are understated values of the true distances.

Response bias occurs in surveys when the response given by the respondent *may have been influenced by improperly or poorly worded survey questions of by the behavior of the interviewer.*