Sample surveys in real world

1 Quick recap of sampling errors from previous two chapters:

- Non-Response Bias: When an individual chosen for a sample cannot be contacted or decides to not participate in the study or research. This type of bias occurs after the sample has been selected and can create potential bias in the data collected.
- Response Bias: This type of bias can occur when a person does not understand a question or feels influenced to respond to a question in a certain way. Response bias can also occur as a result of the wording of questions that are of a sensitive nature.
- Voluntary Response Bias: The sample is not random or representative of the population. The people who volunteer for a study or survey may be more inclined to respond to questions or report certain behaviors. Examples: online reviews and post-call surveys.
- Convenience sampling: Selection of whichever individuals are easiest to reach is called convenience sampling. The sample is composed of individuals who are easily accessible to the researcher and are willing to be a part of the study.

2 Non-sampling errors:

Non-sampling errors are errors not related to the act of selecting a sample from the population. They can be present even in a census. The most common non-sampling errors are:

- Undercoverage: When some groups of the population are left out of the sampling process and the individuals in these groups do not have an equal chance of being selected for the sample. Example: a sample survey of households in a country may miss people who are homeless, prison inmates, or students living in dorms.
- Multiple inclusions: Multiple inclusions occur if some population members appear multiple timed in the sampling list so they have higher chances of being selected. Example: if you are collecting a sample from a list of telephone numbers, some members of the population can have multiple telephone lines.
- Processing errors: Mistakes in entering responses into a computer or other mechanical tasks failure. Fortunately with the development of technology processing errors are less common than in the past.

3 Selection bias

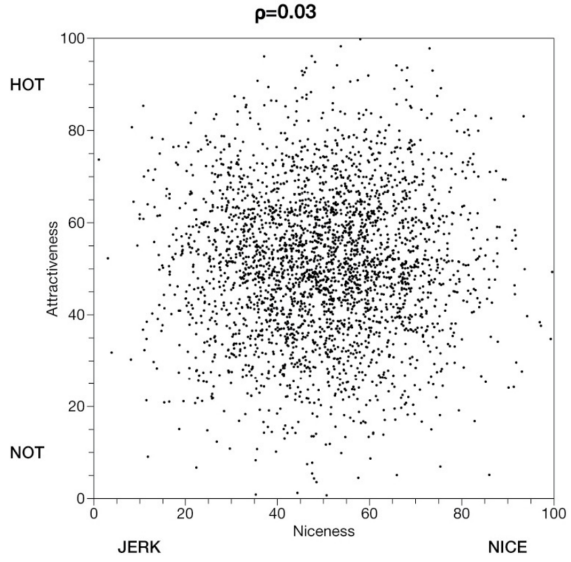
Selection bias occurs when the sample used in a study or analysis is not representative of the overall population due to the method of selection. Selection bias can arise when certain groups are over represented or under represented in the sample, often because of the criteria used to select participants. Examples: Examples: In recruitment, a company might favor candidates from prestigious universities, potentially overlooking other qualified applicants. Similarly, surveys on sensitive topics, such as drug use or sexual behavior, may suffer from selection bias if those willing to participate differ significantly from those who opt out.

Selection bias can lead to false correlations and in real-world applications, this can result in ineffective policies, treatments, or business strategies based on misleading data. Recognizing and accounting for selection bias is important in any form of analysis to ensure that conclusions are valid and representative of the true population.

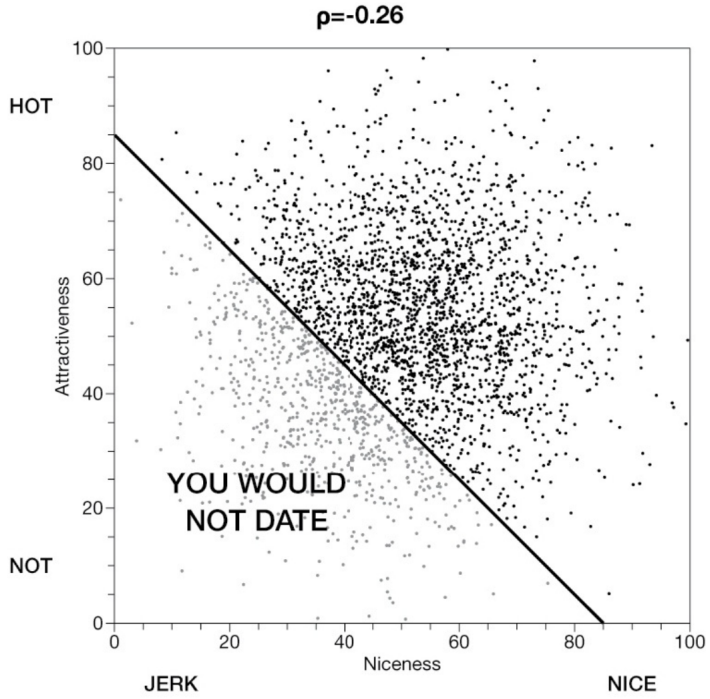
4 (Adapted from the book "Calling bullshit.") Fun example for Selection bias

There is a common observation in the dating world explained by Berkson's Paradox (a specail case of selection bias). Often, people complain that when they go on a date with someone who is very attractive, the person turns out to be a jerk, and conversely, when they date someone who is nice, the person tends to be less attractive. A common explanation for this pattern is that attractive people can "afford" to be less kind because they are in high demand as partners.

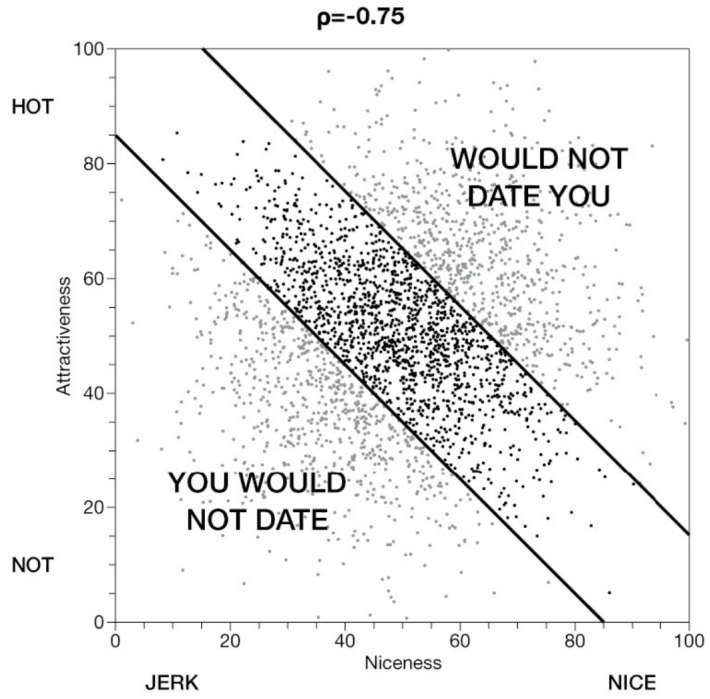
Imagine placing potential partners on a two-dimensional plot, with how nice they are on the horizontal axis and how attractive they are on the vertical axis. That might give you a plot of individuals that looks something like the next chart.



In this image, attractiveness and niceness are basically uncorrelated. So far, we have no reason to think that nice guys are unattractive and attractive guys are not nice. But now let's look at what happens once you consider whom you'd actually be willing to date. Certainly not anyone who is both unattractive and a jerk. Maybe you'd compromise in the looks department if a fellow was a really great person, or vice versa. So in the space of guys, the people you might date would be in the region above the diagonal.



Among the guys you would date, there is now a modest negative correlation between attractiveness and niceness. A really hot guy is less likely to be nice than the average guy that you would date, and a really nice guy is less likely to be hot than the average guy you would date. By selecting for both niceness and attractiveness, you have created a negative correlation between niceness and attractiveness among the people whom you would be willing to date. A similar process happens in the opposite direction. Not only do you pick whom you are willing to date, people pick whether they are willing to date you. Apologies for putting this bluntly, but unless you are David Beckham, there will be some people who will have better options. So this rules out another set of possible people that you might date.



Who is left? Your dating pool is now restricted to a narrow diagonal band across the space

of potential partners. This is selection bias, or Berkson's Paradox, at work: by selecting for both niceness and attractiveness, we create an apparent negative correlation between niceness and attractiveness among the people we consider as dating prospects. This perceived trade-off is not a true reflection of the general population but rather a product of the way we select potential partners.

5 Ideal sampling method

Simple Random Sample (SRS): The 'purest' form of sample, and the simplest, is the 'simple random sample'. A simple random sample takes a small, random portion of the entire population to represent the entire data set, where each member has an equal probability of being chosen. A simple random sample is the ideal way to sample as it has no potential for biases. There are also sampling methods that are unbiased but not truly random such as stratified random sampling, cluster sampling, systematic sampling, quota sampling which you might see in sample survey reports.

6 Wording of the question:

Words make a big difference. A final influence on the results of a sample survey is the exact wording of questions. It is surprisingly difficult to word the questions so that they are completely clear. Examples:

- You are writing an opinion poll question about a proposed amendment to the Constitution. You can ask if people are in favor of "changing the Constitution" or "adding to the Constitution" by approving the amendment. Changing suggests there is something wrong with the Constitution. Adding suggests there is something missing. More people will be in favor of adding to the constitution than changing it.
- In light of sky-rocketing gasoline prices,we should consider opening up a very small amount of Alaskan wilderness for oil exploration as a way of reducing our dependence on foreign oil. Do you agree or disagree? The statement is slanted towards a desired response of 'yes'.

7 Questions to ask before you believe a Poll:

As a final recap we have learnt from last three chapters, here are some questions to ask before you pay much attention to poll results.

- Who carried out the survey? Even a political party should hire a professional sample survey firm whose reputation demands that they follow good survey practices.
- What was the population? That is, whose opinions were being sought?
- How was the sample selected? Look for mention of random sampling.
- How large was the sample? Even better, find out both the sample size and the sampling variability.
- What was the response rate? That is, what percentage of the original subjects actually provided information?

- How were the subjects contacted? By telephone? Mail? Face-to-face interview?
- What were the exact questions asked?