

Sample Surveys in the Real World

So far, we've discussed what a sample survey is and how important it is to avoid possible sources of bias by having a good sampling method with random sampling at its core. How does this look in a realistic setting?

First, let's recognize that getting a list of the entire population so we can assign each population element its own unique code for random sampling is often not feasible or even possible. For example, is it possible to get a list that includes *all* working mothers on it? In practice, we take our sample from a **sampling frame**.

Definition

A **sampling frame** is the list of population elements from we actually draw our sample. The ideal setting is a sampling frame containing every population element, but often it does not.

After attempting to recruit the chosen members of your sample (by sampling from the sampling frame), there are still problems. Consider the following example:

Example 1

A link in a Pew Research online report, <http://poq.oxfordjournals.org/content/70/5/759>, talks about some problems in obtaining responses to surveys.

"Where the 1997 Rigorous study completed 1,201 interviews after making 31,385 calls, the 2003 Rigorous study completed 1,089 interviews after 72,485 calls. ".....

Of the "2,626 respondents who began interviews in both the Standard and Rigorous samples, 282 (or 11 percent) broke off the interview at some point prior to the last question. Of these, our interviewers were subsequently able to persuade 115 to complete the survey, while 167 were never completed."

It is expensive and takes strategy to conduct surveys. There are many issues to resolve in the process of collecting responses and we want the most accurate and truthful representation of our population possible. Polls may claim a random sample (which avoids bias) with a small margin of error (small variability in our estimate), but is that enough? What are some pitfalls to avoid?

Errors that happen in Sampling

1. Sampling Errors (Caused by the act of taking a sample and make the results from a sample different from those of a census)

- a. Random Sampling Error
 - Deviation between the statistic and parameter
 - Caused by chance in selecting a random sample
 - ONLY error accounted for in the margin of error in a confidence statement
- b. Bad Sampling Methods
 - Convenience and voluntary response samples
 - An incomplete sampling frame can cause **undercoverage** where certain groups of the population are left out

2. Non-sampling Errors (Error not related to the act of selecting a sample from the population and can even be present in a census)

- a. Processing errors
 - Mistakes in data entry or arithmetic
- b. Poorly worded questions
 - Poor wording can introduce
- c. Response error
 - Response from an individual in the survey that is inaccurate from
 - Lying
 - Bad memory
 - Interviewer/Respondent bias (interviewer sways responses or respondent say what they think the interviewer wants to hear)
- d. Non-Response
 - Failing to obtain a response from an individual selected for the survey

Example 2

Identify the error.

The subject lies about past drug use.

- A. Sampling Error: Bad Sampling Method
- B. Non Sampling Error: Response Error
- C. Non Sampling Error: Non Response Error
- D. Non Sampling Error: Processing Error

A typing error is made in recording the data.

- A. Sampling Error: Bad Sampling Method
- B. Non Sampling Error: Response Error
- C. Non Sampling Error: Non Response Error
- D. Non Sampling Error: Processing Error

Data are gathered by asking people to go to a website and answer questions online.

- A. Sampling Error: Bad Sampling Method
- B. Non Sampling Error: Response Error
- C. Non Sampling Error: Non Response Error
- D. Non Sampling Error: Processing Error

The subject cannot be contacted after five calls.

- A. Sampling Error: Bad Sampling Method
- B. Non Sampling Error: Response Error
- C. Non Sampling Error: Non Response Error
- D. Non Sampling Error: Processing Error

Interviewers choose people on the street to interview.

- A. Sampling Error: Bad Sampling Method
- B. Non Sampling Error: Response Error
- C. Non Sampling Error: Non Response Error
- D. Non Sampling Error: Processing Error

Example 3

Poorly worded questions. Look at the difference with a few changed words!

Is our government providing too much money for welfare programs?

– 44% said “yes”

Is our government providing too much money for assistance to the poor?

– 13% said yes