

Bias in Sampling

MATH 130, *Elements of Statistics I*

J Robert Buchanan

Department of Mathematics

Fall 2023

Bias

Definition

If the results of the sample are not representative of the population, then the sample has **bias**.

Remark: bias could mean to give preference to selecting some individuals over others; it could also mean that certain responses are more likely to occur in the sample than in the population.

Sources of Bias

There are three sources of bias:

1. Sampling Bias
2. Nonresponse Bias
3. Response Bias

Sampling Bias

Definition

Sampling bias means that the technique used to obtain the individuals to be in the sample tends to favor one part of the population over another.

Sampling Bias

Definition

Sampling bias means that the technique used to obtain the individuals to be in the sample tends to favor one part of the population over another.

Undercoverage is a type of sampling bias.

Definition

Undercoverage occurs when the proportion of one segment of the population is lower in a sample than it is in the population.

Example: 1936 Presidential Election

The magazine *Literary Digest* predicted that Alfred M. Landon would defeat Franklin D. Roosevelt in the 1936 presidential election. The *Literary Digest* conducted a poll based on a list of its subscribers, telephone directories, and automobile owners. On the basis of the results, the *Literary Digest* predicted that Landon would win with 57% of the popular vote. However, Roosevelt won the election with about 62% of the popular vote.

This election took place during the height of the Great Depression. In 1936, most subscribers to the magazine, households with telephones, and automobile owners were Republican, the party of Landon. There was undercoverage of Democrats.

Nonresponse Bias

Definition

Nonresponse bias exists when individuals selected to be in the sample who do not respond to the survey have different opinions from those who do.

Remark: Nonresponse can be improved through the use of callbacks or rewards/incentives.

Example: Nonresponse Bias

The federal government's Current Population Survey has a response rate of about 92%, but it varies depending on the age of the individual. For example, the response rate for 20- to 29-year-olds is 85%, and for individuals 70 and older, it is 99%.

Response rates in random digit dialing (RDD) telephone surveys are typically around 70%, e-mail survey response rates hover around 40%, and mail surveys can have response rates as high as 60%.

Response Bias

Definition

Response bias exists when the answers on a survey do not reflect the true feelings of the respondent.

Response Bias

Definition

Response bias exists when the answers on a survey do not reflect the true feelings of the respondent.

Types of response bias:

- ▶ Interviewer error
- ▶ Misrepresented answers (ask someone how many push-ups they can do in 1 minute versus having them show you how many push-ups they can do in 1 minute).
- ▶ Words used in survey question
- ▶ Order of the questions or words within the question

Words Used in Survey Question

Consider two ways of asking for opinions regarding estate taxes:

1. Do you oppose the reduction of estate taxes?
2. Do you favor or oppose the reduction of estate taxes?

The second question is considered **balanced**.

Words Used in Survey Question

Consider two ways of asking for opinions regarding estate taxes:

1. Do you oppose the reduction of estate taxes?
2. Do you favor or oppose the reduction of estate taxes?

The second question is considered **balanced**.

How would responses to the following two questions likely differ?

1. Do you think the United States should forbid public speeches against democracy?
2. Do you think the United States should allow public speeches against democracy?

Ordering of Questions or Words

Many surveys will rearrange the order of the questions within a questionnaire so that responses are not affected by prior questions. Consider the following two questions:

1. Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?
2. Do you think a Communist country such as Russia should let American newspaper reporters come in and send back to America the news as they see it?

Ordering of Questions or Words

Many surveys will rearrange the order of the questions within a questionnaire so that responses are not affected by prior questions. Consider the following two questions:

1. Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?
2. Do you think a Communist country such as Russia should let American newspaper reporters come in and send back to America the news as they see it?

Effect of question order:

- ▶ For surveys conducted in 1980 in which the questions appeared in the order (1, 2), 54.7% of respondents answered “yes” to (1) and 63.7% answered “yes” to (2).
- ▶ If the questions were ordered (2, 1), then 74.6% answered “yes” to (1) and 81.9% answered “yes” to (2).

Open Questions

Definition

An **open question** allows the respondent to express his or her response.

Example

What is the most important problem facing America's youth today?

Closed Questions

Definition

A **closed question** requires the respondent to choose from a list of predetermined responses.

Example

What is the most important problem facing America's youth today?

1. Drugs
2. Violence
3. Single-parent homes
4. Promiscuity
5. Peer pressure

Closed Questions

Definition

A **closed question** requires the respondent to choose from a list of predetermined responses.

Example

What is the most important problem facing America's youth today?

1. Drugs
2. Violence
3. Single-parent homes
4. Promiscuity
5. Peer pressure

Remark: In closed questions, the possible responses should be rearranged because respondents are likely to choose early choices in a list rather than later choices.

Sources of Error in Sampling

Definition

Non-sampling errors are errors that result from the survey process. They are due to the

- ▶ nonresponse of individuals selected to be in the survey,
- ▶ inaccurate responses,
- ▶ poorly worded questions, or
- ▶ bias in the selection of individuals to be given the survey.

Sources of Error in Sampling

Definition

Non-sampling errors are errors that result from the survey process. They are due to the

- ▶ nonresponse of individuals selected to be in the survey,
- ▶ inaccurate responses,
- ▶ poorly worded questions, or
- ▶ bias in the selection of individuals to be given the survey.

Definition

Sampling error is the error that results from using sampling to estimate the information regarding a population. This type of error occurs because a sample gives incomplete information about the population.