**Section 1.5 Bias in Sampling**

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

There are **three** sources of bias in sampling:

1. Sampling bias
2. Nonresponse bias
3. Response bias

**Sampling bias**

means that the technique used

to obtain the samples individuals

tends to favor one part of the population over another.

Any convenient sampling has sampling bias

because the individuals are not chosen through a random sample.

**Sampling bias** also results due to

under coverage, which occurs when

the proportion of one segment of the population

is lower in a sample then it is in the population.

**Undercoverage** can result if the frame used to obtain the sample

is incomplete or not representative

of the population.

Some frames, such as the list of all registered voters,

may seem easy to obtain.

But even this frame may be incomplete since people

who recently registered to vote may not

be on the publish list of registered voters.

Sampling bias can lead to incorrect predictions.

For example, 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.**

Therefore, the choice of the frame

used to conduct the survey led to an incorrect prediction

due to sampling bias.

**Essentially, there was an undercoverage of Democrats.**

It is difficult to gain access to a complete list

of the individuals in a population.

For example, public opinion polls

often use random telephone surveys,

which implies that the frame is all households with telephones.

This method of sampling excludes households

without telephones, as well as homeless people.

If these people differ in some way

from people with telephones or homes,

the results of the sample may not be valid.

The federal government prohibits the random calling

of cellular telephones.

However, a study done by Pew Research Center

suggests that excluding cellphone only homes

from telephone surveys does not significantly

affect the results.

1. What is sampling bias?

**means that the technique used**

**to obtain the samples individuals**

**tends to favor one part of the population over another.**

1. Does a convenience sample have sampling bias?

**Probably, Literary Digest example……Landon over Roosevelt**

1. What is under coverage?

**Undercoverage can result if the frame used to obtain the sample**

**is incomplete or not representative**

**of the population.**

**Nonresponse Bias**

Non-response bias exists when individuals selected

to be in the sample who do not respond to the survey

have different opinions from those who do.

**Non-response can occur because individuals selected**

**for the sample do not wish to respond**

**or the interviewer was unable to contact them.**

All surveys will suffer from non-response.

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's 99%.

Response rates in random digit dialing telephone surveys

are typically around 70%, email survey response rates

hover around 40%, and mail surveys

can have response rates as high as 60%.

Non-response bias can be controlled using callbacks.

For example, if a mail questionnaire was not returned,

a callback might mean phoning the individual

to conduct the survey.

If an individual is not at home, a callback

might mean returning to the home at other times in the day.

Another method to improve non-response

is using rewards such as cash payments

for completing a questionnaire or incentives such as a cover

letter that states that the responses to the questionnaire

will determine future policy.

For example, I received $1 with a survey regarding

my satisfaction with a recent purchase.

The $1 payment was made to make me feel guilty enough

to fill out the questionnaire.

As another example, a city may send out

questionnaires to households and state in a cover letter

that the responses to the questionnaire

will be used to decide pending issues within the city.

Let's consider the Literary Digest poll again.

The Literary Digest mailed out more than 10

million questionnaires and 2.3 million people responded.

The rather low response rate contributed

to the incorrect prediction.

After all, Roosevelt was the incumbent President,

and only those who are unhappy with his administration

were likely to respond.

By the way, in the same election,

the 35-year-old George Gallup predicted

that Roosevelt would win the election

in a survey involving 50,000.

When does nonresponse bias exist?

**3. Non-response can occur because individuals selected**

**for the sample do not wish to respond**

**or the interviewer was unable to contact them.**

1. List two causes of nonresponse bias.

Younger age groups have lower response rates.

People are not available during the time the survey was attempted.

1. List one tool that can be used to control nonresponse bias?

Cash Incentive.

RESPONSE BIAS

Response bias exists when the answers on a survey

do not reflect their true feelings of the respondent.

Response bias can occur in a number of ways.

A trained interviewer is essential to obtain

accurate information from a survey.

A skilled interviewer can elicit responses from individuals

and make the interviewee feel comfortable enough

to give truthful responses.

For example, a good interviewer can obtain truthful answers

to questions as sensitive as, have you ever cheated

on your taxes?

Do not be quick to trust surveys conducted

by poorly trained interviewers.

Do not trust survey results if the sponsor has

a vested interest in the results of the survey.

Would you trust the survey conducted

by a car dealer that reports 90% of customers

say they would buy another car from the dealer?

Some survey questions result in responses

that misrepresent facts or are flat out lies.

For example, a survey of recent college graduates

may find that self-reported salaries are inflated.

Also, people may overestimate their abilities.

For example, ask people how many push

ups they can do in one minute.

And then ask them to do the push ups.

How accurate were there?

The way a question is worded can lead

to response bias in a survey, so questions must always

be asked in balance form.

For example, the yes/no question,

INCORRECT

**do you oppose the reduction of the state taxes**

should be written,

CORRECT

**do you favor or oppose**

**the reduction of the state taxes?**

The second question is balanced.

Do you see the difference?

Consider the following report based

on studies from Schuman and Presser

who asked the following two questions—

1. do you think the United States should forbid

public speeches against democracy?

1. Do you think the United States should allow public speeches

against democracy?

**For those respondents presented with question A,**

**21.4% gave yes responses.**

**While for those given question B, 47.8% gave no responses.**

**The conclusion you may arrive that is that most people are**

**not necessarily willing to forbid something but more**

**people are willing not to allow something.**

These results illustrate how wordy a question

can alter a survey's outcome.

Another consideration in wording a question is not to be vague.

The question,

**How much do you study?**

**This question is too vague!**

**Does the researcher mean how much?**

**Do I study for all my classes or just for statistics?**

**Does the researcher mean per day or per week?**

**The question should be written, how many hours**

**do you study statistics each week?**

**Many surveys will rearrange the order of the questions**

**within a questionnaire so the responses are not**

**affected by prior questions.**

**Consider an example from Schuman and Presser**

**in which of the following two questions were asked,**

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?**

1. **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?**

***For surveys conducted in 1980 in which the questions appeared***

***A then B, 54.7% of respondents answered yes to A***

***and 63.7% answered yes to B.***

***If the questions were ordered***

***B then A, then 74.6% percent answered yes to A***

***and 81.9% answered yes to B.***

**When Americans are first asked**

**if US reporters should be allowed to report**

**Communist news, they are more likely to agree that Communists**

**should be allowed to report American news.**

**Questions should be rearranged as much as**

**possible to help reduce effects of this type.**

Pollsters will also rearrange words within a question.

For example, the Gallup organization routinely

asked the following question of 1,017 adults age 18 or older.

**The words "approve" and "disapprove"**

**are rotated to remove the effect that**

**may occur by writing the word "approve" first**

**in the question.**

One of the first considerations in designing a question

is determining whether the question

should be open or closed.

An open question allows the responded

to choose his or a response.

**For example, what is the most important problem**

**facing America's youth today?**

**A closed question requires the respondent**

**to choose from a list of predetermined of responses.**

**For example, what is the most important problem**

**facing America's youth today?**

And five answers are given.

**CLOSED QUESTIONS**

In closed questions, the possible responses

should be rearranged, because respondents

are likely to choose early choices in a list

rather later choices.

An open question should be phrased

so that the responses are similar.

You don't want a wide variety of responses.

This allows for easy analysis of the responses.

Closed questions limit the number of responded choices,

and therefore, the results are much easier to analyze.

The limited choices, however, do not always

include a respondent's desired choice.

In that case, the respondent will

have to choose a secondary answer or skip the question.

Survey designers recommend conducting pre-test surveys

with open questions and then using the most popular answers

as the choices on closed question surveys.

Another issue to consider the closed question design

is the number of possible responses.

The option "no opinion" should be

omitted, because this option does not

allow for meaningful analysis.

The goal is to limit the number of choices in a closed question

without forcing respondents to choose an option they do not

prefer, which would make the survey have response bias.

**DATA ERROR**

Although not technically a result of response bias,

data entry error will lead to results

that are not representative of the population.

Once data are collected, the results typically

must be entered into a computer which

could result in input errors.

For example, 39 maybe entered as 93.

It is imperative that data be checked for accuracy.

In this text, we present some suggestions

for checking for data error.

1. Under what conditions does response bias exist?

Response bias exists when the answers on a survey

do not reflect their true feelings of the respondent.

Response bias can occur in a number of ways.

A trained interviewer is essential to obtain

accurate information from a survey.

A skilled interviewer can elicit responses from individuals

and make the interviewee feel comfortable enough

to give truthful responses.

For example, a good interviewer can obtain truthful answers

to questions as sensitive as, have you ever cheated

on your taxes?

Do not be quick to trust surveys conducted

by poorly trained interviewers.

Do not trust survey results if the sponsor has

a vested interest in the results of the survey.

**Note:** Response bias can occur through interviewer error, misrepresented answers, wording of questions, ordering of questions or words, type of question, or data-entry error.

**Note:** An open questionallows the respondent to choose his or her response (free response).

**Note:** A closed question requires the respondent to choose from a list of predetermined responses (multiple choice).

Objective 1, Page 7

**Note: Can a Census Have Bias?**

A question on a census form could be misunderstood, thereby leading to response bias in the results. It is often difficult to contact each individual in a population. For example, the U.S. Census Bureau is challenged to count each homeless person in the country, so the census data published by the U.S. government likely suffers from nonresponse bias.

Objective 1, Page 8

Define the following terms.

1. Nonsampling Error:

Sampling error:

**Can a Census Have Bias?**

Yes!

The discussion so far has focused on bias in samples, but bias can also occur when conducting a census.

How?

A question on a census form could be misunderstood, thereby leading to response bias in the results.

We also mentioned that it is often difficult to contact each individual in a population. For example, the U.S. Census Bureau faces challenges in counting each homeless person in the country, so the census data published by the U.S. government likely suffers from nonresponse bias.

**Sampling Error versus Nonsampling Error**

Nonresponse bias, response bias, and data-entry errors are types of *nonsampling error*.

However, when a sample is used to learn information about a population, *sampling error* is also likely to occur.

**DEFINITION**

**Nonsampling errors** result from undercoverage, nonresponse bias, response bias, or data-entry error. Such errors could also be present in a census.

**Sampling error** results from using a sample to estimate information about a population. This type of error occurs because a sample gives incomplete information about a population.

**In Other Words**

[](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj1_8_eaa86c3e-b4f5-584d-c05f-c0f87a607082)

What do we mean when we say sampling results in "incomplete information"? We mean that the individuals in the sample cannot reveal all the information about the population. Suppose we want to determine the average age of students enrolled in an introductory statistics course. To do this, we find a simple random sample of four students and ask them to write their age on a sheet of paper and turn it in. The average age of these four students is 23.3 years. Assume that no students lied about their age or misunderstood the question and the sampling was done appropriately. If the actual average age of all 30 students in the class (the population) is 22.9 years, then the sampling error is 23.3−22.9=0.4 year. Now suppose the same survey is conducted again, but this time one student lies about his age. The results of that survey will also have nonsampling error.