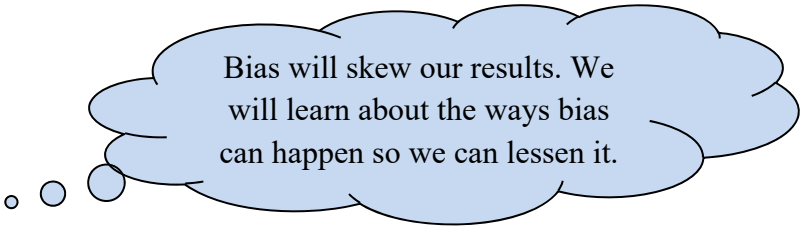

General Education Statistics
Class Notes
Bias in Sampling (Section 1.5)

A study can be biased toward one view or another. Sometimes the bias is intentional but often not.

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

There are three sources of bias. They are

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

1. 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.

Definition: Undercoverage results in sampling bias. It occurs when the proportion of one segment of the population is lower in a sample than it is in the population. This can happen when you leave out major groups in your sample, like women or transsexuals or vegetarians.

Phone surveys can suffer from this because not everyone has a phone (homeless people come to mind) and many will not pick up the phone if they do not recognize who is calling. If these people are different than those who do get sampled, then we have undercoverage.

Have you ever gotten a survey in your email and ignored it? If so, you are included in the next type of bias.

2. 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.

All surveys suffer from nonresponse. However, nonresponse can be improved through the use of callbacks or rewards/incentives. Samples often attach a small reward for participation to ease this bias.

3. Definition: Response bias exists when the answers on a survey do *not* reflect the true feelings of the respondent. This can happen in many different ways.

Types of Response Bias

1. Interviewer error
2. Misrepresented answers
3. Wording of questions
4. Order of questions or words

Let's look at some examples of these biases.

expl 1: (Wording of questions) In the early 1990's, Gallup asked Americans whether they supported the US bombing Serbian forces in Bosnia. In this survey, 35% of respondents supported the idea. The very same day, ABC News asked whether Americans would support the US, along with its allies in Europe, bombing Serbia forces in Bosnia. In this survey, 65% supported the idea. Explain the difference in the wording of the question. What does this suggest?

expl 2: (Misrepresented answers) Ask a group of people how many push-ups they can do and then ask them to actually do them. How accurate do you think the survey would be?

expl 3: (Order of questions or words) Consider a survey given in 1980 that contained **both** of the questions below.

a.) Do you think the US should let Communist reporters from other countries come in here and send back to their papers the news as they see it?

b.) 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?

If you were taking this survey, how would you respond? What if your survey had the two questions in reverse order? Would that change your opinion?

In fact, this survey was given in two different forms, one with question *a* first and one with question *b* first.

When question *a* was asked first, 54.7% of respondents answered "yes" to question *a* and 63.7% then answered "yes" to question *b*.

But when they asked question *b* first, 81.9% answered "yes" to *b* and 74.6% then answered "yes" to *a*.

Why do you think that is?

To summarize:	**When question *a* was asked first…**	**When question *b* was asked first…**
Percent who said "yes" to question *a**	55%	75%
Percent who said "yes" to question *b**	64%	82%

*I rounded these to whole number percents for ease of discussion.

Reputable surveys will often word their questions so they can rotate the options. This helps to eliminate **response bias**. For instance, consider the question "Do you favor or oppose the reduction of estate taxes?" They will ask the same question in half the surveys but phrase it as "oppose or favor".

Definition: Data-entry 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 may need to be entered into a computer, which could result in input errors. Or, a respondent may make a data entry error. For example, 39 may be entered as 93. It is imperative that data be checked for accuracy.

Pollsters conducted an exit poll of voters leaving the 2004 presidential election. These exit polls predicted that John Kerry would win over George W. Bush. When asked what went wrong, the pollsters cited interviewer error. Many of the interviewers were young and so interviewed young people. These young voters tended to vote for Kerry. Also, they were prone to data entry errors. Also, it was concluded, that more women were selected for the exit poll and they tended to vote for Kerry as well.

As we will see, most surveys have errors. We try to minimize them. Even censuses have errors.

Definition: Nonsampling errors are errors that result from sampling bias, nonresponse bias, response bias, or data-entry error. Such errors could also be present in a complete census of the population.

This is opposed to **Sampling error** which is an error that 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. These errors are *not* mistakes per se.

Worksheet: Sampling questions:
This worksheet focuses on questions that you should ask of any survey you read about. Who carried out the survey? How was the sample selected? And more… Answering these questions helps you determine how much bias is a factor. The worksheet uses a study that explored how HIV/AIDS is viewed in the African American population.