Module A: The Foundations of Statistics

Slide Deck A4:

What *Not* to Do

The section in which we examine several sources of bias in statistical studies. The purpose is for you to know what to avoid doing and what to pay attention to when reading studies.

Today's Objectives

By the end of this slidedeck, you should

- learn the basic techniques for analyzing a report;
- be reflexive with this information; and
- understand the practical and ethical concerns that arise when conducting a study

Things to Consider about a Scientific Study

When thinking about a scientific study reported in the mainstream press, consider the following factors:

- The source of the reporting
- The variables involved (and their meanings)
- The conclusions drawn from the data

Consider the Source

As you listen to sound bites on the radio or television spouting off statistics from studies, you should ask yourself pertinent questions:

- Who paid for the study?
- Where were the data collected?
- When was the information collected?
- Who published the study?

- How knowledgeable is the reporter on science?

Note: The "source" could also refer to the population from which the data are collected.

Consider the Variables

Have you ever read a headline like "America's 100 Best Companies to Work For" or "Top 10 Things to Make You Happy"? You should read the fine print of the study to find out by what criteria best is defined or whose standard of happy they are talking about.

- Headlines like these can mean very different things to different people.
- In addition to considering what the variable of the study is measuring, we also need to think about those variables not directly being measured — ones which may have an effect on the results.

Note: A well-constructed study will make a point to identify and account for *reasonable* **confounding** variables.

Consider the Conclusions

In the end, it is the conclusions of the study that we should be most aware of:

- Do the data support the conclusion?
- Do the results present the whole picture or just a part?
- Could there be other reasonable conclusions drawn?
- Could there be other reasons for the same conclusion drawn?
- Does the study have any practical applications?

Why does this Matter?

Why does this matter to you as a scientist?

- Your research will be published
- Your research will be read by others
- Your research will be analyzed by others
- Your research will be used by others

Note: You need to ensure that your research is properly reported

Study Validity

There are two main types of study validity:

- **External validity** is the validity of applying the conclusions of a scientific study outside the context of that study. In other words, it is the extent to which the results of a study can be generalized to and across other situations, people, stimuli, and times.
- In contrast, **internal validity** is the validity of conclusions drawn within the context of a particular study.

Both are important. Internal validity concerns whether the drawn conclusions are appropriate, given the data. External validity concerns whether these findings can be extended to other conclusions.

Because general conclusions are almost always a goal in research, external validity is an important property of any study. Mathematical analysis of external validity concerns a determination of whether generalization across heterogeneous populations is feasible, and devising statistical and computational methods that produce valid generalizations.

Study Bias

In the context of a study's structure, **bias** refers to the how of unrepresentative the sample is of the population. There are several sources of bias in a study, broadly divided into

- researcher-based
- respondent-based

Note that **no study is perfect**. No sample will *perfectly* represent the population. Bias is only important when it is sufficiently severe to cause the researcher to draw incorrect conclusions.

Researcher-Caused Bias

Typical sources (and names) of bias include:

- **Selection bias** occurs when the selection of individuals for analysis (or groups or data) is done in such a way that proper randomization is not achieved.
- **Researcher bias** occurs when a researcher influences the results of a study, whether purposely or not.
- **Response bias** occurs when a researcher's behavior (or existence) causes a participant to alter his or her response or when a participant gives an inaccurate response.

Participant-Caused Bias

Beyond conscious biases, there are several that do not arise from planning errors. They arise from the actions of the participants. These include:

- **Participation bias** occurs when there is a problem with the participation — or lack thereof — of those chosen for the study.
  - As an example, **nonresponse bias** occurs when there is a lack of participation in a self-selected sample from certain segments of a population, when a person refuses to participate in a survey, or when a respondent omits questions when answering a survey.
  - Another example concerns respondents who **do not report the truth**.
- **Dropouts** are participants who begin a study but fail to complete it.
- **Nonadherents** are participants who remain in the study until the end but stray from the protocol (directions given).

Oopsie-Caused Bias

Finally, there are simple errors in the study. These tend to occur in the transcribing of the data into a form that the program can handle.

- **Processing errors** are errors that occur simply from the data being processed, such as typos when data are being entered.

More Information

The National Institutes of Health (NIH) is the primary agency of the United States government responsible for biomedical and public health research. As such, they focus on proper experimental procedures. Their extensive website contains *a lot of information* about sampling:

- `https://www.ncbi.nlm.nih.gov/books/NBK253184/table/background.t1/`

If you will be performing biomedical research, bookmarking this site is important.

Other research organs of the US government have their own websites aimed at ensuring good research is done, such as the National Institute of Standards and Technology (NIST):

- `https://www.nist.gov/itl`

Today's Objectives

Now that we have concluded this lecture, you should be able to

- Learn the basic techniques for analyzing a report
- Be reflexive with this information
- Understand the practical and ethical concerns that arise when conducting a study

Supplemental Readings

The following may be of interest to you in terms of today's topics:

- Hawkes Learning: Section 1.4
- Intro to Modern Statistics: Chapter 2
- R for Starters: Nothing