Statistical Thinking for the 21st Century

Chapter 17

Doing reproducible research

Most people think that science is a reliable way to answer questions about the world. When our physician prescribes a treatment we trust that it has been shown to be effective through research, and we have similar faith that the airplanes that we fly in aren't going to fall from the sky. However, since 2005 there has been an increasing concern that science may not always work as well as we have long thought that it does.

In this chapter we will discuss these concerns about reproducibility of scientific research and outline the steps that one can take to make sure that our statistical results are as reproducible as possible.

17.6 Doing reproducible data analysis

An important aspect of reproducibility is to be able to reproduce someone's analyses on their own data. This requires that researchers share both their data

"Code" refers to the formulas and functions you use to analyze your data.

and their analysis code, so that other researchers can both try to reproduce the result as well as potentially test different analysis methods on the same data. There is an increasing move in psychology towards open sharing of code and data; for example, the journal *Psychological Science* now provides "badges" to papers that share research materials, data, and code, as well as for pre-registration.

The ability to reproduce analyses is one reason that we strongly advocate for the use of formula-based analyses rather than using a "point-and-click" software package. It's also a reason that we advocate the use of free and open-source software as opposed to commercial software packages, which require others to buy the software in order to reproduce any analyses.

This statement is why you learned to "code" in this course (using formulas and functions) rather than learning to use a "point-and-click" software package such as SPSS.

17.7 Conclusion: Doing better science

It is every scientist's responsibility to improve their research practices in order to increase the reproducibility of their research. It is essential to remember that the goal of research is not to find a significant result; rather, it is to ask and answer questions about nature in the most truthful way possible. Most of our hypotheses will be wrong, and we should be comfortable with that, so that when we find one that's right, we will be even more confident in its truth.

- Describe the concept of pre-registration and how it can help protect against questionable research practices

17.9 Suggested Readings

- Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions, by Richard Harris
- Improving your statistical inferences - an online course on how to do better statistical analysis, including many of the points raised in this chapter.