

STA 712: HW 1

Due: Friday, September 8, 12:00pm (noon) on Canvas

Instructions: Submit your work as a single PDF. Your document should be created using LaTeX; see the course website for a homework template file and instructions on getting started with LaTeX and Overleaf.

1 Practice reading a research paper

If you've every felt stressed around finals, you might have seen "animal therapy" opportunities organized by your school, in which students decompress by interacting with cute and cuddly animals like puppies and bunnies. Does interacting with animals actually help participants' wellbeing? In "The Importance of Client–Canine Contact in Canine-Assisted Interventions: A Randomized Controlled Trial" <https://www.tandfonline.com/doi/abs/10.1080/08927936.2021.1944558> (Binfet *et al.* 2022, *Anthrozoos*), the authors investigate whether interacting with dogs leads to increases in participant wellbeing, and decreases in participant illbeing. This activity will help you read the paper by Binfet *et al.*, and think about important study design and data collection questions.

Reading a research paper, particularly in a field in which you are not an expert, can be challenging. The trick is to skim the paper for the most relevant information, and skip over technical details that are not essential to understanding the key take-aways. The questions below will guide you to the most important sections in the paper by Binfet *et al.*

Questions

The Abstract and Introduction

A good place to start is often with the Abstract and Introduction, which allow you get an overview of the paper, and usually don't contain too many technical details. The Abstract

is more succinct than the Introduction, but it also provides less motivation. When the Introduction is long, you may want to skim for key details.

Read the Abstract, and skim the Introduction (I recommend focusing on paragraphs 1, 2, 6, and 7). Then answer the following questions.

1. In the abstract, the researchers explain that participants volunteered for the study. However, the title of the paper describes the study as a randomized controlled trial. Explain why these two statements do not conflict – which part of the study design is randomized?
2. What is the specific research question the researchers want to study, and what are their three hypotheses about this research question?

Study Participants

Read the *Student Participants* and *Procedure* subsections of the Methods, and then answer the following questions.

3. How many students participated in the study?

When reporting on human subjects research, it is very important that study participants should not be *identifiable* in the data – that is, we should not be able to determine the identities of individual participants from the information reported in the paper. This is key to protecting participants' privacy.

4. Which details about participants have been omitted in the paper, which make participants identifiable if they were reported?

Potential research subjects must meet certain criteria, defined by the researchers, to participate in a study. *Inclusion* criteria define requirements for inclusion in the study (e.g., a target age range or social group), while *exclusion* criteria are reasons a subject would be asked not to participate (e.g., certain medical conditions).

5. What are the inclusion/exclusion criteria for this study?

Data Collection and Analysis

Now that we know who was studied, we want to know what data was collected about each participant, and how it was analyzed. Read the *Procedure*, *Analytic Plan*, *Hypothesis 1*, and *Hypothesis 2 and Hypothesis 3* subsections of the Methods, and skim the subsection headings for measures of wellbeing and measures of illbeing. Then answer the following questions.

6. What are the three treatment groups, and how were participants assigned to each treatment?
7. Summarize the different measures used to capture wellbeing and illbeing (you don't need to read about these measures in detail yet).
8. The researchers describe *pre-registering* their research plan. Pre-registration means that the researchers formally commit to a specific plan of analysis before collecting data. Why might pre-registration be important?
9. Summarize the methods used to test the three hypotheses. Why are paired-sample *t*-tests appropriate for Hypothesis 1, but not for Hypotheses 2 and 3?

Results

Finally, let's see what the researchers concluded from their statistical models. Details are provided in the *Hypothesis 1* and *Hypothesis 2 and Hypothesis 3* subsections of the Results. Read these subsections, then answer the following questions.

10. What should the researchers conclude about their three hypotheses?
11. The researchers randomly assigned participants to the three treatment groups. The benefit of random assignment is that we no longer need to worry about confounding variables, because no explanatory variable can be systematically associated with the treatment. So why do the researchers collect demographic information about their participants, and compare the demographics for the three groups in Table 1?

2 Reproducibility

When we do research, we want our work to be *reproducible*: other researchers should be able to reproduce our results, given the same data. Unfortunately, published research is not always reproducible. In this part of the assignment, you will read about why reproducibility

is important, and some good practices for making your own work more reproducible.

To begin, read the following articles at the links provided:

- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226-1227.
- Broman, K., Cetinkaya-Rundel, M., Nussbaum, A., Paciorek, C., Peng, R., Turek, D., & Wickham, H. (2017). Recommendations to funding agencies for supporting reproducible research. American Statistical Association.

Then answer the following questions.

12. Briefly summarize why reproducibility is important in computational research.
13. What steps can a researcher take to make their work more reproducible?
14. Revisit the two papers you have read so far this semester (the dengue paper and the canine contact paper). Assess the reproducibility of these two papers. What factors are limiting their reproducibility, and how could reproducibility be improved?
15. The articles here focus mostly on reproducibility in academic research, particularly published academic research. Suppose you are performing statistical analysis for a private company instead. Why might reproducibility (perhaps by another statistician in your company) still be important?

3 Reading a statistics research paper

In the first section of this homework assignment, you read an applied paper that used statistics to analyze the data. What about research papers in statistics itself?

Research papers in statistics are written a bit differently than applied papers in the sciences and social sciences. There is more variability in how the papers are organized, and the content depends somewhat on the journal. Often, however, you can expect a statistics research paper to include some or all of the following information (the order and organization may differ):

- Background and motivation
- A description of methods

- Theoretical results
- Simulations
- An application to real data

In the final part of this assignment, you will read a short research paper on randomized quantile residuals (the original paper, from the authors of your textbook!) and answer some questions.

The paper:

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236-244.

The questions:

16. What is the motivation for the authors' work?
17. Which previous papers (cited in the article) have developed similar definitions of residuals?
18. How do the authors demonstrate the value of their proposed residuals?