

Math 401: Statistical Modeling

Lab Report Guidelines

With each lab, each *individual* will hand in a short report (2–3 pages) that explains the results of your statistical analysis. The report should contain:

1. **Title**

Include the title of the project, your name, your partner's name, the course number, and the date.

2. **Introduction**

Explain the purpose of the project (or the question(s) it is trying to answer). Provide some background on the problem. Why is this problem important/interesting? If the lab came with a specific question to answer, e.g. “is the difference significant?” then you should clearly state the answer as early as possible in the introduction, preferably right after stating the question you were investigating. Write your introduction to a general audience, i.e. one without a background in advanced statistics. Put yourself in the shoes of a statistical consultant who has been asked to make a report to a CEO or board of directors. Keep it to 1-2 short paragraphs.

3. **Results and Conclusions**

Explain your results. If the lab contains Questions, this is where they should be answered. Include graphs and numerical outputs to support your conclusions, and *explain how to interpret the graphs and numbers*. Do not include too many figures: carefully select the graphs and statistics that best describe the patterns in the data. All figures and tables should be labeled and should be referenced in the text. If your analysis required any assumptions, state these clearly when drawing conclusions.

If you used R or StatKey to carry out computations, include screenshots (or use R Markdown to include the numerical outputs directly) and explain how to interpret them. It might be helpful to circle the parts of the numerical output you want your reader to focus on, as is done in the slides for the first week of our course. In addition, please place any R code you wrote/used into the shared drive as part of your submission.

4. **Methods**

Discuss where the data came from, and any sources of bias that might exist in the data. If you suspect bias, discuss how this affects your results. For each statistical technique or model you use, describe briefly why you used it, how it works, and how you assessed the conditions required before using it. Discuss the extent to which the data can be assumed to be a random sample, a statistical

experiment, or an observational study.

If you needed to massage or wrangle the data, explain what you did to it. For example, if you discarded any outliers or data points with missing values, justify those decisions here. Can you explain how/why those outliers might have gotten into the data? For example, if a datum is 100 times more than all those around it, perhaps a missing decimal point is to blame. If one member of a team is making orders of magnitude more money than the others, perhaps she is a supervisor and should not be part of the comparison group.

If you wrote code to parse or clean the data, discuss that code here, and include the code in your submission in the shared drive. If you ran the data through other programs (e.g. Excel), document the steps you took here. This is important both to demonstrate to your reader that you have not allowed your own biases to affect the analysis, and also to aid future researchers in replicating your analysis.

Finally, talk about what you learned from the project, how it affected your overall view of statistical modeling, and how well or poorly it appears to mesh with the content covered in class, the readings, and the videos.

5. Group Notes

In a sentence or two, indicate how you and your partner worked together on the project. Did you split up the work or did you always work together? Was your partner an equal participant or did one person do more work than the other? This information will be kept confidential, so please be frank.

Writing your individual report

Although you will complete each project with a partner, you will hand in your own *individual project report*. It is natural (and encouraged) that you will talk to your partner about all aspects of the project while you are working on it together, including the analysis and code. However, when it comes time to write your project report, you are on your own. For example, since you and your partner analyzed the same data set, I would expect you to end up with the same p-value. However, it would not be appropriate for both partners to have an identical (or even too similar) sentences in their reports describing the results, the interpretations, the Questions, or the project as a whole in the introduction. In summary, work closely together on everything up to the report. When it comes time to write the report, do this individually and separately (no longer sitting next to or communicating with your partner). You are encouraged to turn in the same graphs and numerical output as your partner. I would expect that once the data has been wrangled, and the graphs and numerical outputs have been produced, that writing the report should only take an hour or two. Please try to keep it to 2-3 pages.