

Final Exam Questions (66 Points Total)

Contents

Introduction	2
Exam Questions	3
Set Up Your Project Folder (2 pts)	3
Set Up Your R Markdown File (10 pts)	3
Prepare Document for Analyses (5 pts)	3
Descriptive Statistics	3
Explore Data (6 pts; 7 after extra credit)	3
Exploratory Visualization (Extra Credit - 2 pts)	4
Statistical Analyses (21 pts)	4
Visualizing Results (11 pts; 14 pts after extra credit)	4
Take-Home Summary (8 pts)	5
Extra Credit: README File (4 pts)	5

Introduction

Basic Goal: Choose a dataset. Use R and R markdown to go from nothing to an attractive statistical report.

What you'll do: Create your own R project and R Markdown files from scratch, read in your data, choose a hypothesis and some variables to work with, clean and explore the data, report results and visualizations, and provide a project summary, all in a professional-looking PDF document knitted from your R Markdown file. The details for what should be in the knitted PDF are explained under Exam Questions, starting on the next page.

In the end, you **will turn in an R project folder** via email that contains the following contents:

1. The R project itself
2. The dataset
3. The R knitted PDF report containing answers to all of the Exam questions (the questions below will tell you what to put in the report).
4. The R Markdown file that generated the report.

Points Breakdown:

- Exam R Markdown and PDF: 63 points (up to 69 points with extra credit)
- Project File: automatic 3 points for including everything in a project folder
- Extra Credit (4 points): include an R Markdown-knitted README file (see the end under Extra Credit)
- **Total: 66 points (up to 76 points with extra credit)**

Exam Questions

Set Up Your Project Folder (2 pts)

- (1pt) Create a new R project inside of a project folder. Name the folder, and the R Studio project, `YOUR_NAME_Final`.
- (1pt) Retrieve the dataset (you can pick the one you want to work on), and move it into the project folder.
 - The datasets, along with their descriptions, are on Blackboard.

Set Up Your R Markdown File (10 pts)

- (1pt) Create a new R Markdown file. Name it `Statistical_Report`.
- Make the following changes in the YAML:
 - (1pt) Give the document an informative title.
 - (1pt) Add a table of contents.
 - (1pt) Set the “depth” of the table of contents to 2.
 - (1pt) Add a setting to change the author, and make yourself the author.
 - (1pt) Add a section to show the date, and put today’s date.
- (1pt) Use R Markdown syntax to create a new level-1 section heading called “Project Description”.

Knit this document. After you do that, type the text “`\newpage`” somewhere between the YAML and your section header. You don’t have to tell me what happened; just note the difference for yourself. Leave it there, and continue on...

Here is what you will put in your document under the “Project Description” heading:

- (1pt) what hypothesis do you have about this data (pick your own)?
- (1pt) what variables will you use/explore to test this hypothesis (pick your own)?
- (1pt) what statistical analysis will you use to test the hypothesis (pick your own)?

Prepare Document for Analyses (5 pts)

Here, you’ll be doing what you would normally do on the homeworks. Make sure to do everything necessary to: * read the data. * use the packages required by your cleaning/analyses/visualization code (you can pick which packages you want to use and load). * inspect the data to make sure the variables are without errors and in the correct format for their level of measurement (for example, continuous data should be in “numeric” or “integer” format). If there are errors (there will be), find the errors and fix them. * Re-code or relabel data if you have to (for example, you might have to turn a scale from “strongly disagree to strong agree” to a scale that ranges from 1-5)

Importantly, while this code will be necessary for you to run your analysis and create your visualizations, **you will hide the code and output for this section**. In other words: the code in the chunks at the beginning of the document, and their output, should not be visible on the knitted PDF.

Descriptive Statistics

Explore Data (6 pts; 7 after extra credit)

- (2pt) First, create a second level-1 section heading called “Exploring the Data”. Under this heading, create a level-2 heading called “Descriptives”.
 - (4pt) Here, you will provide descriptive statistics for the variables you chose. If stuck about what kind of statistics to provide, think about what we’ve done on previous homeworks and challenges. What are some ways to get a quick-and-dirty summaries of the trends among your variables? Consider also the levels of measurement of variables you have (continuous vs categorical, etc).

- * Print the output. EXTRA CREDIT (1 pts): Use xtable. Make sure that the code does not show (only the output), and make sure that, if you do use xtable, to provide a caption for your table and to change the chunk settings so that your table will be rendered. *Hide your code.*

Exploratory Visualization (Extra Credit - 2 pts)

- Second, create another level-2 heading called “Visualization”.
 - Here, you will choose a visualization or two that give a quick-and-dirty sense of what the trends in the variables are. Again, be sure to consider the level of measurement of the variable when choosing a visualization. Depending on the level of measurement, you may want to look at frequency counts or something similar. . . or group means, or visualization of linear relationships. It will depend on what variables you choose and which visualizations you personally think are most useful for exploring them.
 - Show the visualization (i.e. output of your code), but *hide your code.*

Statistical Analyses (21 pts)

- (1pt) Create a level-1 Section header called “Results”.
- (1pt) Underneath that, create a level-2 section header called “Hypothesis”.
 - (4pts) Here, provide a few sentences where you describe the expected relationships between the variables. Make sure to articulate the important aspects of the hypothesis: the direction of the relationship, the magnitude or how strong you think the relationship will be, which variables are the independent variable (predictor) and which the dependent variable (outcome), and why you expect the relationship you do. Address these in whatever capacity or order you personally thinks gets the point across best. Make sure the description is appropriate for the level of measurement of the variables. You can be technical here.
- (1pt) Create another level-2 section heading called “Analyses”.
- (1pt) Print the results (but *hide the code*) of the analyses using xtable:
 - (1pt) Add a caption
 - (1pt) Change the alignment of the columns
 - (1pt) Put the caption on the top
- (10pt) After the table, add a brief summary of what you found (4-6 sentences). Explain the extent to which the test supported or failed to support your hypothesis. Consider p-values, confidence intervals, the size of the estimate, and effect sizes, each where applicable. Basically, use what you have learned in this class to provide a well-informed written summary of the results. While you can use technical language here, be sure to also briefly explain the concept behind each element of the results. When summarizing the confidence intervals, for example, provide a brief definition of what confidence intervals do.

Visualizing Results (11 pts; 14 pts after extra credit)

- (6pt) Think of a visualization that best demonstrates the results of your study, and create one. Show the visualization in this same section, but *hide your code.*
- (1pt) Add your own title.
- (4pt) Explain how to interpret the visualization to a layperson. Consider questions a person seeing this plot for the first time might ask. What is the y-axis, or vertical axis, and what does it represent? Things like that.
- EXTRA CREDIT:
 - Make sure the visualization looks professional by adding (half point each):
 - * Axis titles
 - * Change the font
 - * Change the theme
 - * Customize the theme somehow (like moving the legend if there is one, or realigning the title, etc.)
 - * Add colors or fills

- * Customize the colors or fills

Take-Home Summary (8 pts)

- (1pt) Create a level-1 section header called “Project Summary”. In this section:
 - (2pt) Remind the reader of the hypothesis (in simple language) and the variables you used to test it.
 - (2pt) Explain in simple language why you chose the test you did.
 - (3pt) Explain in simple language whether the test supported or failed to support, and to what degree, the hypothesis.

Extra Credit: README File (4 pts)

Your mission, if you choose to accept it: create an additional R Markdown file, called `README`.

Imagine that I am a collaborator opening up R Studio for the first time. You just emailed me your project folder, and now I want to add my own section of analyses at the end of the report you made. Your `README` file should explain how to do that. For example, how would I:

- View the code that generated the results in your `Statistical_Report` PDF?
- Create my own section heading at the end of your report?
- Run and record my own analyses?
- Ensure that results of my new code are displayed in the knitted document, along with my code?

Make sure to use YAML settings to add authorship, title, and date. Use R Markdown syntax to organize the document. You will include the R Markdown file and the knitted PDF file.