

Project Report Instructions

BSTA 513/613

Nicky Wakim

Directions

! Project template

You *may* use [this project template](#) to get started on the report. It is your responsibility to meet the formatting guidelines below!!

DO NOT USE SITE PAGE (“Project Report Instructions”, current page) as your template!!

Purpose

Project reports serve as a great way to communicate the knowledge learned in our class and connect it to context within research. It is important that we can take a step back from the numbers and analysis to see what questions linear regression can help us answer.

It is really important for you to look back through your labs and connect the work!!

Formatting guide

- The report will be written in Quarto. Turn in both the `qmd` and `html` files
 - **No code** should appear in the `html` document
 - * This means all R code chunks should have `## echo: false`
 - * This also means warnings and messages should be turned off
- The report should be 10 - 14 paragraphs long
 - **In 512, many lab reports were a little too long!**
 - **Remember, I know mostly what you did in the labs! This report is meant to synthesize that work into a coherent message/story!**

* **This often means details of our analysis are lost.**

- Tables and figures should NOT have variable names as they appear in the data frame
 - Variable names should be understood by a reader
 - Variable names should be written in full words
 - Include a title or caption for all figures
 - Figure and tables appear on same page or close to same page where they are first referenced
 - Tables and figures are an appropriate size in the html - Nicky is able to read all words in figures and tables
- Writing, spelling, and grammar should be admissable
 - This means I can generally follow your thought/what you are trying to communicate
 - Some spelling and grammar mistakes are allowed
 - * I will not take off points if there are a few sprinkled in
 - * If *every or close to every sentence* has mistakes, then I will take off
- Sectioning of the report
 - Main sections that are required: Introduction, Statistical Methods, Results, Discussion, Conclusion, **Reflection**, and References
 - * You may have an appendix to include additional figures!
 - Other sections that might help group specific methods or results
- Title information at the top of the `html`
 - This includes the title itself, your name, and the date

i The project report is a separate file from the labs

You can save tables and figures from labs or separate files, then load them in the report

- Save R objects in analyses file:
 - Suppose you named the Table 1 as `table1`
 - `save(table1, file = "table1.Rdata")`
- Load R objects in report file: `load(file = "table1.Rdata")`

Examples of reports

The following are examples of reports from BSTA 513 with the feedback that I gave them.

Please note that 513 uses a different type of outcome than our class. These examples are meant to help guide you with the formatting and some appropriate content.

Also note that these were converted to PDFs so I could write in feedback. Some of the tables and figure sizes were distorted. They need to be legible in the `html`.

- [Report 1 with my feedback](#)
- [Report 2 with my feedback](#)

The above reports have code showing in their `html`. Remember that I am asking you to **hide** all code, warnings, and messages.

Grading

The project report is out of 36 points. Note that the Statistical Methods and Results sections are graded on an 8-point scale, while all other components are graded on a 4-point scale.

Rubric

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|------------|--|---|--|--|--|
| Formatting | Lab submitted on Sakai (or by email if late) with <code>.html</code> file. Report is written in complete sentences with very few grammatical or spelling errors. With little editing, the report can be distributed. | Lab submitted on Sakai (or by email if late) with <code>.html</code> file. Report is written in complete sentences with some (around 2 per section) grammatical or spelling errors. With some editing, the report can be distributed. | Lab submitted on Sakai (or by email if late) with <code>.html</code> file. Report is written in complete sentences, but have many grammatical or spelling errors. With major editing, the report can be distributed. | Lab submitted on Sakai (or by email if late) with <code>.html</code> file. Report is written in complete sentences, but are very hard to follow due to grammar mistakes. | Lab not submitted on Sakai (or by email if late) with <code>.html</code> file. Report is not written with complete sentences. With major editing, the report can be distributed. |

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|------------------|---|--|---|--|---|
| Figures and work | All requested output is displayed, including 2 required figures and tables, and at least one additional figure. Figures and tables look professional, are easily interpreted by the reader, and easily convey the intended message. | All requested output is displayed, including 2 required figures and tables, and at least one additional figure. For the most part, figures and tables look professional, are easily interpreted by the reader, and easily convey the intended message. A few mistakes in the figures are made. | All requested output is displayed, including 2 required figures and tables, and at least one additional figure. Figures and tables look semi-professional, are not so easily interpreted by the reader, and convey the intended message but after some work by the reader. Some mistakes in the figures are made. | All requested output is displayed, including 2 required figures and tables, and at least one additional figure. Figures and tables do not look professional, are not easily interpreted by the reader, and/or do not convey the intended message. Many mistakes in the figures are made. | Requested output is not displayed, Missing one or more figures. |

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|--------------|--|--|--|---|---------------------------|
| Introduction | Provides a good background for the research question, includes motivation for the question, and references previous research that justifies this analysis. | Provides a decent background for the research question and includes motivation for the question. Previous research is mentioned, but feels disconnected to the current analysis. | Provides a decent background for the research question and includes motivation for the question. Previous research is mentioned, but feels disconnected to the current analysis. | Does not provide a background that connects to the research question. Motivation and previous research are not mentioned. | No introduction included. |

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|--------------------|---|---|---|--|----------------------|
| Methods (8 points) | Describes statistical methods concisely and highlights pertinent information to the reader (listed Sections below). Demonstrates proper analyses were performed. | Describes statistical methods and highlights pertinent information to the reader (listed Sections below). Details were omitted or added that were not needed to explain the overarching methods. Demonstrates proper analyses were performed. | Describes statistical methods and highlights pertinent information to the reader (listed Sections below). Details were omitted or added that were not needed to explain the overarching methods. Some incorrect analyses included in the description. | Describes statistical methods, but lacks clarity. Demonstrates a lack of understanding about the overall process of regression analysis. Incorrect analyses included in the description. | No methods included. |
| Results (8 points) | Correctly interprets coefficients for the explanatory variable and identifies any other interesting trends. Highlights pertinent results to the reader (listed Sections below). | Correctly interprets coefficients, but does not correctly incorporate the interaction (if in the model). Highlights pertinent results to the reader (listed Sections below). | Incorrectly interprets coefficients. Highlights pertinent results to the reader (listed Sections below). | Incorrectly interprets coefficients. Omits pertinent results to the reader (listed Sections below). | No results included. |

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|---------------------------|--|---|--|---|--|
| Discussion | Thoroughly and concisely discusses limitations and considerations of the results, and their consequences. | Discusses limitations and considerations of the results and their consequences, but misses some big considerations. | Discusses limitations and considerations of the results, but does not discuss the consequences. | Discusses limitations and considerations of the results, but misses many considerations and does not discuss consequences. | No discussion included. |
| Conclusion and References | For the conclusion, main research question is answered and statistical caveats described to non-technical person. References are mostly cited consistently within the report, and in the Reference section. This includes the data source! | For the conclusion, main research question is answered and statistical caveats described to non-technical person. References are sometimes cited consistently within the report, and in the Reference section. This includes the data source! | For the conclusion, main research question is somewhat answered (but focus is not on the research question) and statistical caveats described to non-technical person. References are sometimes cited consistently within the report, and in the Reference section. This includes the data source! | For the conclusion, main research question is somewhat answered (but not the focus at all) and statistical caveats are not described. References are not cited consistently within the report, and in the Reference section. This includes the data source! | For the conclusion, main research question is not answered. Or references are not included at all. |

| | 4 points | 3 points | 2 points | 1 point | 0 points |
|------------|--|----------------------------|----------------------------|----------------------------|------------------------|
| Reflection | Discusses all the labs. Reflection demonstrates an understanding of each lab's purpose and overarching connection from the labs to overall project. | Discusses 3 out of 4 labs. | Discusses 2 out of 4 labs. | Discusses 1 out of 4 labs. | No reflection included |

More clarifications

- In formatting, an example of a report with little editing needed is one that has zero to some grammar or spelling mistakes, no code chunks showing, and no output warnings nor messages showing.
- Professional figures mean
 - I can read the words and numbers in the html
 - * Variable names are converted from the data frame version to readable text
 - * For example: `iam_001` does not show up on axes, instead something like: Response to "Currently, I am..."
 - Colors are only used if conveying information
 - Intended message of the figure is easily understood
 - * If you are trying to show a trend of mean IAT vs. an ordered categorical variable, then the variable is ordered on the x-axis
- For the references
 - I will not be overly critical about the formatting
 - By consistency, I mean that you if you are citing things like (Last Name, Year) it doesn't suddenly change to number citations.

- If you would like to use Quarto’s citation tool, you can! I actually pair it with Zotero and it works beautifully! (But I would not embark on this if you haven’t used Zotero before)

Sections

Title

- **Purpose:** Create an identifiable name for your research project that includes the main research question’s variables and gives some context to the analysis or results

Introduction

- **Length: 1-2 paragraphs**
- **Purpose:** Introduce the project motivation, data, and research question. It also includes any background information relevant for understanding the analysis and relevant previous work.
- This section is non-technical.
 - By reading just the introduction, someone without a technical background should have an idea of what they study was about, and why it is important
- **You may start with the introduction written in Lab 1, but you should edit it and make sure it flows into your report well!**
- Should contain some references
- Should include a sentence that states your research question (but NOT using a question). For example: “This study investigates the association between food insecurity and age.”

Statistical Methods

- **Length: 3-5 paragraphs**
- **Purpose:** Describe the analyses that were conducted and methods used to select variables and check diagnostics
- **Important to keep in mind:** methods typically describe your approach and process, *not the results of that process*
 - For example: I might say “We investigated the linearity of each continuous covariate visually. If continuous variables were not linear, then we divided the variable into categories using existing guidelines from <insert reference here> or creating quartiles.”

- * In the methods section, I would NOT say: “We investigated the linearity of each continuous covariate visually. We found that age was not linearly related to IAT scores. Thus, we categorized age into the following groups: _____, _____, _____, _____, and _____.”
 - The last two sentences about age would be more appropriate in the Results section
- **Some important methods to discuss** (You may divide these into your sections, not necessarily with these names)
 - General approach to the dataset
 - * 3-5 sentences
 - * Did you need to do any quality control?
 - * **Missing data: complete case analysis or imputation based on Lab 3 choices**
 - 1 sentence
 - Can be included in the Exploratory data analysis section
 - Variable transformations
 - * This includes a description of analyses for Table 1 and what statistics were used to summarize the variables
 - More on creation of Table 1, not discussing the results of Table 1
 - * Includes (not required)
 - Categorizing a continuous variable (even if performed in model selection)
 - Did you use family size/household size/number of children as categorical or continuous?
 - Did you make any transformations to income?
 - Using scoring for an ordered categorical variable (that is not your explanatory variable)
 - * 1 sentence per **changed** variable
 - Making a categorical variable a factor in our R code is NOT a change!
 - Model building: we performed purposeful selection OR LASSO regression
 - * 3 sentences max
 - * **This is where we discuss our process from Lab 4**
 - * For purposeful selection, includes:
 - Describe purposeful selection: combining existing literature, clinical significance, and analysis
 - How did you build the model? Describe the process in ONE SENTENCE
 - Did you consider confounders and effect modifiers?
 - * For LASSO:
 - Mention that you used LASSO regression with important facts like the penalty used, percent testing set, included interactions in the selection process