STA 302 Fall 2024 Final Project Part 3
Detailed Report and Poster Overview
Due: November 29, 2024, by 8:00PM ET
Latest Acceptance: December 6, 2024, by 8:00PM ET

---

*Please note that if you intend to use **one of the three** NQA extensions, you **should not submit any documents prior to the posted deadline** as Quercus will not allow any changes or additions to the submission after the initial deadline. Instead, make sure you have all your documents prepared and ready to submit all at once.*

Goal of the Assessment:	Learning Outcomes being Assessed:
• Showcase your group's knowledge of the correct usage of methods and techniques from the course. • Write a formal report summarizing how data were analyzed to answer a proposed research question. • Experience the process of conducting a complete linear regression analysis on real data. • Think about the ethical responsibility of a statistician in the context of conducting a data analysis. • Demonstrate editing practices to produce concise and clear written work. • Summarize key message of analysis in presentation-worthy poster.	• Conduct a complete analysis starting from a research question to model that reasonably answers said research question. • Communicate the motivation, process, and results of an analysis in a written report. • Report the results of a full analysis of the assumptions and recommend a course of action. • Defend the decision to apply a transformation, the choice of transformation, and whether the violated assumption(s) is(are) adequately corrected. • Defend the choice of "best" model for a given dataset and research question. • Critique hypothesis test/confidence interval conclusions from an analysis as it pertains to violated assumptions. • Appropriately discuss the ethical considerations of opting in/out of using automated selection methods. • Interpret the results of an analysis involving linear models for technical and non-technical audiences.

Overview of assessment:

Using only methods and techniques presented in this course throughout the term, you and your group will answer your proposed research question from Part 1 (or modified based on the feedback received) by building an appropriate linear regression model on your chosen dataset (you do not need to start with the preliminary model from Part 1). You will want to ensure that your model was selected/built using evidence from the literature, dataset, and statistical methods. You will also discuss the limitations of your analysis (e.g. ethical, data, methodological, etc.) and what could be done differently in an ideal situation.

Your group will produce two deliverables:

1. A written report outlining the **technical aspects** of your analysis (i.e., what methods you used and how you made decisions to develop/select your model that answers your research question).
2. A poster that summarizes or provides an overview of the **key message** and results of your analysis and could be understood without reading the written report.

We are assessing you on whether you can correctly apply the course content and document statistical decisions with corresponding evidence while effectively communicating these results in a concise but clear manner. You've already had valuable practice in these areas through your coursework and by completing parts 1 and 2 of the project (see below for how to use the feedback from these in part 3). These skills are important for many academic and industry positions and will provide you with materials to use in a portfolio or list on a resume.

Why are we creating two deliverables?

Both deliverables allow you to practice skills needed in most workplaces but may not have much opportunity to experience in your courses. By completing both, you have two options for materials to showcase on your resume depending on what a job posting is looking for in a successful applicant.

- Report: a more standard option focusing on your ability to outline your process, techniques, and decisions in a concise yet easy to follow manner. The emphasis is on written communication and showcasing your statistical knowledge.
- Poster: a more creative option focussing on telling the overarching story/message of the analysis with more emphasis on plots and tables rather than text and on selection of important and relevant information to help readers see the message at a glance and without supplemental materials.

We are also looking to nominate a few groups to enter into the Canadian ISLP Poster Competition to showcase the amazing work that UofT undergraduate students are doing in their courses. Winning posters receive a cash prize and these groups will go on to represent Canada in the international poster competition of the International Statistical Institute (ISI). **Groups are asked to opt in to being considered for this competition when you submit your work and the teaching team will select the best ones to be nominated. Your grade is NOT impacted if you do not wish to be considered for this competition.**

There are many other opportunities if you wish to showcase the work you are doing in this project, with a few listed below:

- you can shorten your report and submit to the Undergraduate Statistics Class Project Competition (a group last year placed in the top 5 nationally)
- you can present your poster at the Canadian Statistics Student Conference coming this summer.

Goal of your statistical analysis:

Ultimately, your main goal is to answer your proposed research question from Part 1 using the most suitable linear regression model possible built on your chosen dataset. You can consider more predictors than in your preliminary model from Part 1 and/or you can reduce the number of predictors in your model based on statistical tests or other relevant information. You should ensure you have attempted to correct any violated assumptions and have assessed your assumptions at appropriate stages of your analysis. You should also perform full diagnostics on your chosen model, including checking for multicollinearity, problematic observations, and violated assumptions. Finally, you need to explicitly answer your proposed research question using the statistical tools learned in the course.

Components of Final Project Part 3:

Once you have completed your analysis, you will create your two deliverables. Below are instructions for what content from your analysis to include in each deliverable, as well as technical specifications/requirements for each:

1) REPORT REQUIREMENTS:

Remember that the report is your opportunity to showcase your technical mastery of the course material. Make sure the reader clearly understands what techniques you used, how they were used, and why they led you to your final model.

A. CONTENT OF REPORT:

- **Contributions:** each group member's name is listed and a description of their contribution to this project part is outlined (this does not count towards the word limit).
- **Introduction (250 words max.):** introduce the motivation/relevance of the project, state the research question of interest, connect your research question to the results of at least three peer-reviewed papers on the topic, and describe why linear regression can be used to answer the research question.

- **Methods (500 words max.):** outline the methods, tools, and techniques you use to conduct your analysis and arrive at your final model. You are encouraged to present a very simplified version of your flowchart from Part 2 as a figure (won't count towards figure limit) that shows only what broad steps you took and in what order but doesn't go into as much detail about how it is actually done or how you know the result of any step. The text of this section will describe what tools you use and how you would make conclusions from these tools. No data or results should be presented here.
- **Results (750 words max.):** present the important results of your analysis that led you to make the decisions to build and assess your models (following the methods you outlined earlier in the methods section), ultimately arriving at your final model. When referring to different models you may have fit, be sure that it is clear to your reader what variables are used in each model to make it easy to follow.
- **Conclusion and Limitations (350 words max.):** interpret your final model and describe why it answers the research question and why it is important, as well as discuss any limitations that still exist based on your results (e.g. violated assumptions, problematic points, data issues, etc.).
- **Ethics Discussion - only for L0101/L0201/L2001/L2002 students (200-250 words):** If you chose to use automated selection methods in your analysis, explain why you did not use manual selection methods. If you chose to use manual selection methods, explain why you did not use automated selection methods. Were the two methods ethically the same, and you chose one for purely practical reasons? (If so, make sure to explain why you think they are ethically the same.) Or did you think that one was more ethical than the other (e.g., because using the other would have been negligent or reckless?) In either case, use some of the material discussed in the second ethics module to defend your answer.

B. TECHNICAL REQUIREMENTS FOR THE REPORT

- Contains an informative title about your project and the contributions of each member of your group.
- Does not exceed maximum word limit by more than 15 words in each section
- No more than 5 figures or tables (total number combined) in the main text of your report. You may include up to 3 additional figures or tables (total number combined) in an appendix/supplemental at the end of your report.
    - "Figures" are a collection of individual plots or graphs that are similar or related to each other (e.g., a series of residual plots). Avoid making the individual plots too small.
    - Each plot should have proper axis labels and a title, and each figure should have a label (e.g. Figure 1) and a caption that tells the reader what is important to see in the figures. Captions do not count towards the word limit but should not be used to replace text in the main report.
- No R code or R output should be present in the report. These should be included in your Rmd file only. Create your own tables to display output like that of summary(model).
- Includes a reference list/bibliography in an appropriate citation style (e.g. APA) that at minimum contains the source information for your data.

2) POSTER REQUIREMENTS:

Remember that the poster is the overview of your whole analysis, from the question through what techniques you applied to the final answer to your question. Your focus should be on limiting the amount of text and selecting/creating figures/tables that contain the information needed to understand why this model answers your research question.

A. CONTENT OF POSTER

- **Motivation and Research Question:** Briefly state your research question and why this is an important problem or question to be studying (can be supported by literature where applicable).
- **Data Collection:** Where did you source your data (with appropriate citation), how was it originally collected (e.g. sample size, reliability of source, are measurements accurate, method of collection, etc.), how is data appropriate for answering research question (e.g. same population/ability to generalize to population of interest, contains meaningful information/variables, etc.).
- **Methods of Analysis:** Give the readers a rough idea of what methods you used to arrive at your final model, including whether you employed model selection techniques, model validation, diagnostics for violated

assumptions or other specific issues and remedies to these, and what tools/methods were used for each (a simple flow diagram could be useful here to replace excessive text).

- **Results and Conclusions:** Select figures and/or tables that highlight what prompted your final model to look as it does, display final model giving particular emphasis to anything that helps answer your research question, limit text to a discussion on how information in figures/tables led you to create final model and a concrete answer to your research question that is supported by the information presented.
- **Limitations:** a short discussion on at least one of: whether data were not entirely suitable and how, whether model is not entirely accurate or suitable for answering question (e.g. issues with diagnostics or assumptions), or whether model was not able to be validated (if appropriate). A suggestion is made for what could be done differently to improve the analysis overall.
- **References:** a mini-bibliography with citations for your data source and any other peer-reviewed work that was referenced in the poster.

B. TECHNICAL REQUIREMENTS FOR THE POSTER

- **Must not contain student names, group number, or university or class information.**
- Include a title that reflects your research question and/or analysis.
- The maximum size is **A1 (841 mm x 594 mm or 33.1 in x 23.4 in)** and the maximum file size is **10 MB**. For electronic posters, the font size should be such that if printed in A1, it would be readable from 2 metres (7 feet) away.
- The data source must be cited in the poster (under references).
- Minimal text should be used (instead consider using figures/tables and including captions that highlight what you want your reader to know).
- Feel free to use colour and be creative in the layout and design of your poster.

3) DEMONSTRATION OF EDITING REQUIREMENTS:

Two workshops are being held during lectures in this course in November around 1) creating your first draft of your report/poster (emphasis on organization and selecting relevant figures/tables), and 2) editing your written work (tips and techniques to modify written work for clarity and length). As part of your final project, your group will need to demonstrate how you have employed one of the editing techniques from the workshop to your **introduction section** of your report.

In 1 page or less, tell us:

- who was involved in the editing process.
- what technique from the workshop you chose to use to edit your work and why.
- what area did you focus on (clarity, conciseness, elegance)
- give us an example of your edited introduction by showing us
  - the original portion of text from your introduction you want to edit (up to 100 words)
  - the edited text portion of your introduction after applying your selected technique (up to 100 words). Ensure that the edited text is written in your own words.
- what went well and what was challenging with editing your introduction.

How to use feedback from Parts 1 & 2:

As part of the scaffolded nature of the final project, you and your group have already completed some of the components needed for this project:

Part 1 – Proposal: you motivated and proposed your research question, learned about your topic area through a literature summary, looked into the source of your dataset, summarized your dataset (numerically, graphically, in text), fit a preliminary model and assessed assumptions for violations.

- ↳ Use feedback to revise the research question.
- ↳ use to craft the introduction/motivation for report & poster,
- ↳ use to talk about data used, how it was collected, and its trustworthiness for poster,
- ↳ use to guide how to check and report violated assumptions for report.

Part 2 – Analysis flowchart: you outlined part of your analysis, making a direct connection between your research question and how it will be answered, while considering how assumptions impact the validity of the statistical tools you use.

- ↳ use to keep track of your analysis steps for drafting your report,
- ↳ use to ensure you correctly apply steps in your analysis,
- ↳ **use "action steps" to draft your methods section of report,**
- ↳ **use "decision steps" to decide what evidence to include in results section of report,**
- ↳ use whole chart to identify limitations to discuss and to answer your research question in both report & poster.

What to Submit to Quercus:

****Group members who do not participate in the creation of the report and poster will receive a grade of zero****

Only ONE member of the group should submit ALL required components. A complete submission to Quercus will include:

File uploads:

- ✓ The written report, outlining the full analysis with the above technical requirements, saved as a PDF.
- ✓ The poster, summarizing the key message and meeting the above technical requirements, saved as a PDF.
- ✓ The Rmd file containing all the code used to conduct the data analysis (this is not being graded but is helpful to understand your results, if needed).
- ✓ The one-page editing demonstration of your introduction section, saved as a PDF.

Submission comments:

- ✓ The dataset used in the data analysis as CSV files, uploaded to a cloud-based storage service (e.g. OneDrive), with the **shareable link** included as a **submission comment** on Quercus.
- ✓ One of the below statements indicating whether your group would like to be considered for nomination to the Canadian ISLP Poster Competition
  - *"Yes, our group would like to be considered for entry into the competition, and if selected, we give permission for this work to be displayed at the Statistical Society of Canada (SSC) and International Statistical Institute (ISI) conferences and special events, in publications and promotional material, and in electronic format on the Internet."*
  - *"No, our group does not wish to be considered for entry into the competition."*

Resources:

You may wish to consider the writing resources posted on the General Resources Quercus page and refer to the writing workshop materials presented in class (in November). Additionally, keep an eye on the course announcements for dedicated writing office hours with our English Language Learning TA, Dory. Or book time at the University of Toronto Writing Center.

For some **advice on writing a research report**, see our Tip Sheet for Writing a Research Report, designed by Dory Abelman, our English Language Learning TA. Feel free to peruse the cheat sheet on formatting a report if you want more general guidelines of what goes in each section of a report.

New to creating posters? The University of Toronto library has lots of information on How to Create a Poster, including a video tutorial for how to create it in Microsoft PowerPoint. You can find examples of posters submitted to the competition at the competition website, or examples of the different ways you can organize your poster.

If you want to export any of your plots from RStudio and aren't sure how, see this guide.

Criteria of Assessment (for Written Report)	Excellent (3 points)	Satisfactory (2 points)	Revision needed (1 point)	Incomplete (0 points)
Introduction: • The research question is stated with explicit mention of the response and predictors (or common predictor groups) in general (non-technical) terms, and it is explicitly mentioned why this question/topic is important/relevant. • The results of at least three peer-reviewed academic journal articles are used to describe what is already known about the topic and how this knowledge will be integrated into the study. • An appropriate justification is given on how a linear model can be used to answer this research question and whether the focus will be on interpretability/description or prediction.	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Methods: • How a final model for answering the research question is developed using the course material is explicitly and logically outlined. • All assumptions are explicitly assessed and options for mitigating the potential violations are explicitly mentioned. • All model diagnostics are explicitly stated and options for mitigating any issues are explicitly mentioned. • The order in which the model building, assessment of assumptions and model diagnostic steps would proceed is explicit. • Details are provided to describe what tools are used and how conclusions are drawn in each case above and are appropriate/correct.	All five criteria are met.	Only four criteria are met.	Only two to three criteria are met.	One or fewer criteria are met.
Results: • Process of determining a final model is organized logically, is easy to follow, is consistent with the methods section, and highlights important decisions made to arrive at the final model rather than documenting every small step in the analysis. • Appropriate evidence is provided to support conclusions and decisions to arrive at the final model, and this evidence is referenced explicitly. • Conclusions of model selection method(s)/tool(s) are correct and appropriate based on previous analysis results/steps. • Information is provided to confirm that model assumptions have been assessed and model diagnostics have been performed correctly and in appropriate places. • The final model that is used to answer the research question is presented in a professional and easily understandable way.	All five criteria are met.	Only four criteria are met.	Only two to three criteria are met.	One or fewer criteria are met.
Conclusion and Limitations: • An explicit answer to the research question is provided and is consistent with how a linear model was proposed to answer the question in the introduction. • One estimated coefficient is correctly interpreted formally using course terminology in the context of the research question. • A brief discussion is provided on whether the answer to the research question is surprising or consistent with the literature from the introduction. • Limitations of the analysis (or lack of) are identified (e.g. extreme observations, violated assumptions, limited data, etc.) and the impact of these are correctly identified.	All four criteria are met.	Only two to three criteria are met.	Only one criterion is met.	No criteria are met.

Use of Figures and Tables: • All figures/tables are correctly labelled (e.g., Figure 1) and have informative captions describing the main message of the figure/table. • All figures/tables have meaningful titles, axis labels (if applicable) and variable names. • All figures/tables presented in the report and the appendix are referred to in the text. • Individual plots that are combined into a single figure are appropriately sized and easy to read.	All four criteria are met.	Only three criteria are met.	Only one or two criteria are met.	No criteria are met.
Report Structure and Readability: • Each section of the report meets the maximum word count or exceeds it by only 15 words. • The report is well structured, with appropriate sections and appropriate usage of paragraphs for increased readability. • The report does not contain R code or output copied directly from R. • References are written in an appropriate and consistent citation format.	All four criteria are met.	Only three criteria are met.	Only one or two criteria are met.	No criteria are met.
Criteria for Assessment (for Poster)	**Excellent (3 points)**	**Satisfactory (2 points)**	**Revision needed (1 point)**	**Incomplete (0 points)**
Research Question and Data: • The research question is explicitly stated and a clear motivation for this topic is provided. • The source of the data is explicitly stated/provided and why it is appropriate to use to answer the question is correctly stated. • A discussion on how the data were originally collected and why it may be trustworthy or reliable is provided.	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.
Analysis and Conclusions: • The methods/techniques employed to answer the research question are outlined. • An answer to the research question is provided which is consistent with the analysis and evidence presented. • Sufficient evidence is presented that allows the reader to understand how the research question was answered and to evaluate whether the process was correct. • Limitations of the analysis are correctly noted with a suggestion for improvement for future studies.	All four criteria are met.	Only two to three criteria are met.	Only one criterion is met.	No criteria are met.
Poster appearance: • The poster does not contain a burdensome amount of text that could have been replaced with the use of figures or tables. • Figures and tables chosen are clear and convey an intentional message with appropriate labels and captions. • The poster is logical and understandable without the need for additional or supplemental materials (i.e. can be understood without reading the full report).	All three criteria are met.	Only two criteria are met.	Only one criterion is met.	No criteria are met.

Criteria for Ethics Discussion and Editing Demonstration	Excellent (2 points)	Satisfactory (1 points)	Revision needed (0 point)	
Ethics Discussion Question: • Answer meets the minimum and maximum word count. • Response makes a reasonably clear attempt to argue for its conclusion. • Answer refers to some concepts and ideas found in the second ethics module.	All three criteria are met.	Only two criteria are met.	Only one or fewer criteria are met.	
Editing Demonstration: • The editing technique employed was stated and a reason for this choice was provided. • Both an original and an edited writing sample are provided, and notable improvements were made reflecting the editing focus stated. • A short reflection is provided discussing what went well and what was challenging in editing the writing sample.	All three criteria are met.	Only two criteria are met.	Only one or fewer criteria are met.	
TOTAL:	**/31**			