

Final Project Part 1
Research Proposal and Data Introduction
Due: October 20, 2022 by 11:59PM ET

Goal of the Assessment:

The purpose of this assessment is two parts. First, to give you a head start with your final project by finding an area of interest to study and real-world data to work with. Next, to research a little into your area of interest to see what has been accomplished surrounding your question and to highlight the importance of your proposal.

The steps involved in completing this assignment encompass the general process of proposing a research question and will form the basis for a solid introduction section for your final project report (Part 3). Completing this assignment will also give you the chance to think about the appropriateness of linear regression as a tool for answering your proposed research question using your chosen data. Lastly, this assignment provides an opportunity to get some feedback on your writing and research question that can be used to improve your final report.

Instructions: **Assignment Project Exam Help**

1. Decide on one (or a few possible) areas of interest that you may want to explore. These areas of interest can be anything that matters or is of interest to you. Some examples could be (but are certainly not limited to) sports, medicine, public health, economics, video games, literature, etc. Pick something that you really care about.
2. Next, think about possible research questions you may want to study in these areas. What do you want to know about this area? You want to make sure that your question can be answered/studied using linear regression models. So, you'll want to frame your question to be something related to modelling a relationship or predicting a value based on this relationship. You'll also want to consider whether the variable of interest would allow the assumptions of linear regression to hold (see Module 3 content). See the workshop slides from September 23 for advice on framing your research question effectively.
3. After producing a research question, you will need to find some open-source data that you may use in your data analysis. You want to make sure that the data you find has both: 1) your response variable of interest (or has variables that could be used to create that variable), and 2) any other variable you may want to use as predictors. By looking for data online, you may realize you need to modify your research question slightly or pick another one if you can't quite find the data you're looking for. Alternatively, you can stick with your research question, but be sure to mention that you expect there to be many limitations to the dataset because it doesn't quite meet your needs. Step 4 can also help you decide what predictors might be needed for you to answer your question.

Examples of open data sources:

- <https://open.toronto.ca> for open-data from Toronto
- <https://data.ontario.ca> for open-data from Ontario
- <https://www150.statcan.gc.ca/n1/en/type/data?MM=1> for data collected by Statistics Canada
- <https://sports-statistics.com/sports-data/> for various sports-related datasets
- <https://data.oecd.org> for data on various country-level variables
- <https://mdl.library.utoronto.ca> for links to many other data portals through the University of Toronto library

4. Once you've found your dataset and have decided on your research question (or you can work on steps 2-4 simultaneously and use what you find in all of them to finalize your research question), you need to look at what others have studied in relation to your research question. Do a quick search on the University of Toronto library website or other databases that feature scholarly articles (see workshop slides from Sept. 23) to learn about anything related to your area of interest and research question. Look for academic papers (i.e., peer-reviewed work that has been published in reputable scholarly journals, not websites, blogs, or news articles, etc.) that studied the same research question or something related, that tells you more about what you may need to consider in your analysis. Also use the academic papers to justify why your research question is important.

- Focus on giving your reader a rough idea of how many academic papers have studied your research topic (or closely related concepts to your topic). This process of looking at the number of academic papers which describe a specific topic tells your audience how popular the area of research is and how much research has been done.
- Give examples from a few important papers about what was found or discovered to be important in relation to your question. This can be important variables, important results, surprising results, etc. The process of identifying and describing important papers tells your audience that you are aware of prior results and that you will be using these to plan your analysis.
- Think about how your research question fits into the general area of research about your topic. Is your research question different to research questions in other studies? If so, how? A novel research question consists of something that nobody has studied before, or studied in the way you are looking at, or in the population you hope to examine. The process of examining if your research question is novel tells your audience that you see the importance of what you are researching and can frame it against what has already been done.

Library resources:

- <https://guides.library.utoronto.ca/librarysearchtips/gettingstarted> for more details about searching for articles related to your question
- <https://guides.library.utoronto.ca/citing> for details about why and how to cite your references
- <https://guides.library.utoronto.ca/c.php?g=251103&p=1673071> for help getting the correct citation format

5. Lastly, perform a short exploratory data analysis of your chosen dataset. You will want to focus on identifying anything that you may need to consider moving forward. This includes identifying:
- skews,
 - statistical outliers,
 - variables with high spread or observations that don't make sense,
 - missing data

For section 5, you want to make sure you specifically mention the presence of any of the characteristics in 5a-d (or lack thereof) and what this means for the analysis you will eventually perform. For example, this may include describing how any of the characteristics in 5a-d might cause problems (or not) with the results of linear regression or generalizability. You will need to present numerical and/or graphical summaries describing the variables. Choose the options that highlight the features of the data that you want to point out but will also let your reader clearly understand the data that you will be working with.

Guidelines for Picking a Dataset

- Assignment Project Exam Help**
<https://powcoder.com>
Add WeChat powcoder
- Government data portals often contain many datasets about diverse topics – if one dataset doesn't have all the variables you might want to consider, feel free to combine different datasets together
 - Just make sure that each unit being measured is the same in both datasets (i.e. it's reasonable that both measurements are on the same unit)
 - There are many data repositories online – if you find a dataset there that is of interest to you, you **MUST** ensure that your question is different than what the dataset was originally used for.
 - YOU MAY NOT use any dataset that is part of any R package or library, or that is contained in a textbook. If you're not sure, please ask the instructor or one of the TAs.
 - You will need to make sure you have enough variables to be able to showcase the statistical methods that you will learn later in the course. This is because your final project replaces a final exam and so the teaching team needs to assess your knowledge on all topics covered in the course. Some topics the teaching team will require include model validation and model refinement so please ensure your dataset has at least 5 predictor variables.
 - You will also need to make sure you have enough observations to be able to validate your model, which will involve splitting your dataset into two roughly equal parts.
 - A good rule of thumb for a minimum number is to have about 10 observations per variable in each half of your dataset (e.g. 6 predictors x 10 observations/variable x 2 halves of the dataset = 120 observations in total)

Proposal Content Requirements:

Your proposal should be created to satisfy the following requirements:

- The proposal should be organized clearly (consider using headings or sections) and include the following information:
 - a. Your research question, why you chose it (i.e., why it's of interest to you), and why it may be of interest to others.
 - b. Summaries of academic papers related to your question or topic, highlighting similarities/differences to what you propose, and how you will incorporate this knowledge into your model/project.
 - c. Details and summaries on your chosen dataset including the variables collected, the number of observations and anything that stands out in the data that would need to be addressed/investigated further in your analysis.
 - d. A discussion about how and why a linear model fits your chosen data. This will allow you to answer your proposed research question, as well as whether you anticipate any problems that may arise in your analysis from EDA.
 - e. References for where you located the data, and your background research on your topic
- The proposal should be written/presented for an audience that has some statistics background but is not necessarily familiar with the area of your research question or linear regression models,
- The proposal should contain figures and/or tables with proper labels/titles as appropriate in your Data Description, Exploratory Data Analysis section,
- The proposal should have references listed in proper APA format, and
- The proposal itself should not contain R codes

Technical Requirements:

Your submission to Quercus should include the following:

1. A video that presents your proposed research area and question, the dataset you have chosen, and the exploration of your dataset.
 - The video should be no more than 5 minutes in length
 - You must display your U of T Student ID card (or other valid government-issued photo ID) at the beginning of your video The presenter's face **must** be visible throughout the video
 - The presentation should include an appropriate visual medium (e.g., slides) to display important information in an easily readable way.
 - The video should be hosted on a video-sharing service (e.g., MS Streams, MyMedia are supported by the university)
2. The proposed dataset you will use in your Final Project, as a csv or xlsx file, or if too large, as a link to cloud storage where the dataset is saved in csv or xlsx.
3. A copy of the slides/visual aids used in your presentation saved as a PDF document.
4. The R Markdown file containing the code used to produce your exploratory data analysis and tables/figures.

How to Upload:

- Link to Video Presentation – add as a comment to your submission
 - Instructions for uploading to MS Stream: <https://learn.microsoft.com/en-us/stream/portal-upload-video>
 - Instructions for uploading to MyMedia: https://ito-engineering.screenstepslive.com/s/ito_fase/a/1291600-how-do-i-upload-a-video-or-audio-file-to-mymedia
 - Both require you to log in with your UofT credentials.
- R Markdown File – as a file upload on Quercus
- Slides used in Presentation – as a file upload on Quercus
- Chosen Dataset – either as a file upload OR as a comment to your submission (best option if the file is large)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder