Collaborative Course Project Information

1. Data Analysis Component Task

This collaborative project gives you the opportunity to carry out a statistical investigation using real-world data. You will work in small groups (up to 6), choose a dataset from one of the open sources listed below, and will explore statistical questions of interest.

Step 1: Choose a Dataset

Select a dataset from one of the following sources:

- Odesi/ Statistics Canada Open Data. (e.g., Canadian Community Health Survey (CCHS), General Social Survey (GSS))
- Census at School (youth survey data)
- City of Mississauga Open Data (e.g., traffic, environment, recreation, housing)
- City of Toronto Open Data (e.g., fire incidents, transit, neighbourhood well-being)
- Kaggle (platform for data science datasets)
- TidyTuesday (weekly social data project)

Each dataset has its own subjects, variables, and collection methods. Before choosing, spend some time browsing variables that interest you.

To learn about the Canadian Census, read chapter 10 in Guide to the Census of Population, 2021

Step 2: Formulate and Propose Statistical Research Questions

With your group, propose between 3 to 5 statistical research questions.

Your research questions should be clear, measurable, and statistically investigable with the dataset you chose.

Examples:

- *Is there a relationship between neighbourhood income level and reported fire incidents in Toronto?*
- *How does life satisfaction vary by gender and age among Canadian youth (CCHS)?*

Step 3: Conduct Statistical Analysis in R

You will analyze your dataset using R and R Markdown.

For each research question, include:

- **Summary statistics** (e.g., conditional percentages, means, proportions).
- **Graphical displays** (bar charts, histograms, boxplots, etc.).
- **Statistical tests of association** (e.g., chi-square tests, t-tests, ANOVA, logistic regression depending on the variable types).
- **Interpretations in context** (explain your findings clearly in plain language).

Step 4: Prepare Your Statistical Report

Your report will be written in R Markdown and submitted as a PDF. It should include the following sections:

1. **Cover Page**
   - Project title (based on your research topic)
   - Group member names
2. **Abstract**
   - One paragraph summary of the purpose, dataset, main findings, and interpretation.
3. **Introduction**
   - Brief description of the dataset (subjects, when/where/how collected, purpose).
   - Your research questions.
   - Mini codebook of variables selected (include details on recoding values, missing values, and variable treatment).
4. **Statistical Results**
   - One subsection per research question.
   - Include summary statistics, tables, graphs, and statistical tests.
   - Write a clear narrative ("tell the story of your data").
5. **Conclusion and Future Direction**
   - Summarize main findings.
   - Discuss limitations of your dataset.
   - Suggest possible extensions or improvements.
6. **Appendix**
   - Include your R code here. Do **not** put R code in the main body of the report.

Step 5: Submission Guidelines

- Submit one PDF report per group (produced from R Markdown).
- Save as: Data Analysis Report.pdf
- Assign one member (group manager) to upload the file on Quercus (using this assignment page).
- Ensure all members review the final report before submission.

Tips

- Keep your questions clear enough to be answerable with your dataset.
- Compare conditional vs. marginal percentages when analyzing categorical variables.
- Use data visualizations to strengthen your arguments.
- Write for clarity: imagine explaining your findings to someone without a statistics background.

Recommended Books

- *Communicating with Data: The Art of Writing for Data Science.* Deborah Nolan and Sarah Stoudt.
- *An Introduction to Categorical Data Analysis, 3rd Edition, by* Alan Agresti.
- *Statistical Methods for the Social Sciences, 5th Edition, by* Alan Agresti.

Asal' Notes on:

- **Simple Linear Regression:** to be posted
- **Logistics Regression:** to be posted
- **ANOVA:** to be posted

---

2. Infographic Component Task

An infographic is a visual representation of information, data, and knowledge designed to effectively communicate complex concepts clearly and creatively. Make an infographic based on your analysis of song durations data. Students in the past have used CanvaLinks to an external site. to design an infographic. But you can use other tools that you wish. Use a tool that is free to use and is a recommended tool (as you learned about them in the infographic module).

The head of graphics design at Statistics Canada, Miriam Kilby, visited one of our statistics classes and provided us with great ideas to think about when making an infographic. These ideas are described below:

- Ensure to communicate important statistical findings in a clear way.
- Do not use any statistical jargons (technical terms).
- Choose font sizes and colors wisely (think about how colors and fonts impact your infographic).
- Statistics Canada infographics division ensures to use no more than 150 words on an infographic. Let's adopt this approach.

For professional examples of infographics, refer to **Statistics Canada Infographics.**

Submission Guidelines

- Submit one file/link per group.
- Save as: Infographics.pdf
- Assign one member (group manager) to upload the file/link on Quercus (using this assignment page).
- Ensure all members review the final work before submission.

Recap of Required Submissions

Using this assignment page, as a group (submitted by one member in your group, for example the manager) upload two files:

1. PDF file of your data analysis.

2. Infographic file or the URL link to viewing it.