Note from instructors: This prompt, particularly sections D and E which all students completed, is designed to address all 3 QCR ILOs.

[redacted] Individual Project

AUTHOR NAME HERE

2025-11-17

Pledge

This portion of the project is different from the group project. You will include a statement of the pledge:

> *Pledge I pledge that on this examination I have neither given nor received assistance, and that I have seen no dishonest work.*
>
> Signed: **TYPE YOUR NAME HERE**

Inclusion of the above Pledge statement in the document that you submit on Moodle will serve as your electronic signature. Include either the pledge statement, or "I have intentionally not signed the pledge."

By including the Pledge, you certify that you conformed to the following guidelines for the [redacted] individual project:

- You are to work completely independently on this project
- You should save work only to your personal Student Folder. No work for this part of the project should be saved in the Projects or other folders.
- You may use any course materials on Moodle or the R server: our textbook, class notes, in-class activities, R script files, and assignments and answer keys from this semester. No other books or materials, paper or electronic, are allowed unless approved.
- You are not allowed to use google searches, or AI/LLM software like ChatGPT or Gemini.
- You may not discuss any aspect of this exam with any other person (except your Professor). This means that you *may not* seek out help or guidance from other students, graders, or tutors. All questions should be directed to your instructor.
- Exams are due by by your section's Final Exam Time (See Moodle for specific days/times). Make sure to leave plenty of time to knit to PDF and submit in Moodle. Your instructor CANNOT accept any work submitted after the scheduled final exam time.

By completing this individual project, you will be able to:

- Generate questions that can be answered with provided data.
- Perform EDA for

  - a quantitative response variable and a binary explanatory variable
  - a quantitative response variable and a categorical explanatory variable (with more than two groups)
  - two (binary) categorical variables.
  - two categorical variables, where at least one variable has more than two groups

- Perform inference for two means, several means, two proportions, and two categorical variables.
- Write accurate inferential conclusions and interpretations for each of these procedures.

Useful resources for this project:

- Book chapters 16, 17, 18, 20, 21, and 22.
- In class activities from Unit 2: Exploratory Data Analysis, Unit 4: Proportions, Unit 5: Means, and Unit 6: Chi-Square
- Project_data_sets.Rmd (in the Class>Project_Info Folder)

Introduction

As in the Group Project we will be using one of four datasets provided for you. Note that these datasets have been updated to include more categorical variables. Be sure to read the updated Project Data Descriptions for more details.

- Data Set 1: Marathon Runners
- Data Set 2: Canine Assisted Interventions
- Data Set 3: Song characteristics from Spotify
- Data Set 4: Champions League soccer statistics

Notice that the fifth dataset from the previous project (`Strength and Conditioning`) won't be available for this project.

The datasets are available under the folder [redacted]`_F25/Class/Project_Info/Project_data`.

Descriptions of the four datasets are available under the folder [redacted]`_F25/Class/Project_Info` in the `Project_data_sets` file.

Overview

- You will be exploring at least two quantitative variables and at least two categorical variables.
- You are encouraged to select variables that can contribute to a broader narrative related to the dataset you chose.

- At the end of the assignment you will write a paragraph summarizing the findings from your analyses so having a thread that connects these analyses will make this stronger (and easier).

- You can build on questions your group identified in the first project submission, but the questions and analyses presented here should be entirely your own.

- **You need only complete two of sections A,B, and C, But each project must complete sections D and E.**

Section: Set-up

0. Clearly identify the dataset you will be working with. You may continue with the same dataset or choose a different one to explore. Load in that dataset and provide a brief overview of origin of the data. You should answer the "what, where, who, how, why" of the data collection in this description.

Section A: Two-means

1. Select a quantitative response variable and binary categorical explanatory variable. Identify a question that can be answered using these two variables. Clearly state this question for a general audience.
2. Explore and describe the relationship *between* the two variables with appropriate summary statistics. Provide one plot and one sentence about the relationship (supported by summary stats).
3. Perform the appropriate hypothesis test and report the

- test statistic
- p-value
- 95% confidence interval

4. Comment on the role of assumptions/conditions for the test.
5. State your statistical conclusion in context.
6. Provide an interpretation of the confidence interval in context.

Section B: Several Means

1. Select a quantitative response variable and a categorical explanatory variable with more than two groups. Identify a question that can be answered using these two variables. Clearly state this question for a general audience.
2. Explore and describe the relationship *between* the two variables with appropriate summary statistics. Provide one plot and one sentence about the relationship (supported by summary stats).
3. Perform the appropriate hypothesis test and report the

- test statistic
- p-value

4. Comment on the role of assumptions/conditions for the test.
5. State your statistical conclusion in context.
6. Perform any additional tests, report your test statistics and p-values, and summarize your statistical conclusions in context. Interpret at least one appropriate 95% confidence intervals in context.

Section C: Two Proportions

1. Identify a question that can be answered with two binary categorical variables in the dataset. Clearly state this question for a general audience, identify the explanatory and response variable, and explain the two variables in context of the data collection.
2. Explore and describe the relationship *between* the two variables with appropriate summary statistics. Provide one plot and one sentence about the relationship (supported by summary stats).
3. Perform the appropriate hypothesis test and report the

- test statistic
- p-value
- 95% confidence interval

4. Comment on the role of assumptions/conditions for the test.
5. State your statistical conclusion in context.
6. Provide an interpretation of the confidence interval in context.

Section D: Categorical Variables

1. Identify a question that can be answered with two categorical variables in the dataset. At least one of these variables will have more than two groups. Clearly state this question for a general audience, and identify the explanatory and response variable.
2. Explore and describe the relationship *between* the two variables with appropriate summary statistics. Provide one plot and one sentence about the relationship (supported by summary stats).
3. Perform the appropriate hypothesis test and report the

- test statistic
- p-value

4. Comment on the role of assumptions/conditions for the test.
5. State your statistical conclusion in context.

Section E: Conclusion and Figure Caption

1. Write a paragraph summarizing your findings from your analyses to a general audience.

- In addition to summarizing conclusions, you should consider
  - how broadly the results apply (generalizability);
  - whether or not there may be confounding variables or if you can make causal conclusions;
  - and if there are any shortcomings or limitations to the study or conclusions.
- When presenting findings to a general audience it is common to report confidence intervals and statements of significance (or not) but you should refrain from using words like "null", "alternative", "p-value", "test-statistic", etc.
- A strong conclusion should reflect connections between the variables and questions you have chosen to address.

2. Select one Figure from your analyses above that appropriately summarizes an important conclusion from your analyses. For this figure provide a full Figure Caption.

Recall: Figure Captions

A good figure caption has **four** parts:

1. A figure number, which is referred to in the main text (R should automatically assign these).
2. Declarative sentence that summarizes the main findings or conclusions of the analysis.
3. A brief description of the data and data collection methods necessary to understand the figure. This includes the way the data were measured and the units used.
4. A brief description of statistical tests used and other relevant information such as p-values, test statistics, or sample sizes.

Knit this file. Submit your pdf to Moodle.