Assignment 1

Your name

Date

You will be asked to perform some basic data manipulation and analysis tasks. You will likely want to use some functions from the tidyverse and janitor packages that we’ve focused on in the course. The data you’ll be working with are available at this URL:

https://clanfear.github.io/ioc\_iqa/\_data/street\_crime.csv

You’ll want to load them into R (perhaps call the object street\_crime) as we’ve done before in class. You’ll probably want to take a quick look at them to get an idea for what is there. These data describe 500 adults and juveniles convicted of particular types of crime and receiving different lengths of sentences. It contains the following measures:

* sex: Recorded sex of the individual
* age: Age of individual at time of sentencing
* crime\_type: Type of crime they were sentenced for
* sentence: Length of sentence they were given in months (can be 0 if suspended sentence)

You may write answers as normal text in this Word doc below each question. Alternatively, if you’d prefer to make a second document with only your answers (or an RMarkdown / Quarto doc for advanced students), go ahead. You will upload this to the assignment on Moodle. For questions that ask for some output from R, like a table or plot, you can insert them into this document however works best for you. Copy-pasting output from the console is fine, so long as it is passably formatted (using a fixed-width/monospace font like Courier or Monaco helps). Cropped screenshots are fine for console output or plots too. You can also use the Export button on the Plots pane to save plots if you’d prefer (screenshots are easy though). Written answers should be brief and to the point!

Also include any code used to solve these problems in a separate R script file (a .R file, like what we use in class time). Divide this code into sections using comments. You can get started with [this template here](https://clanfear.github.io/ioc_iqa/_assignments/assignment-1.R) if you would like some guidance. Avoid writing a lot of code if possible—most problems can be solved with just a few lines of code using what we’ve learned in lecture—but any amount of code that solves the problem you have is good enough!

# Fundamentals

## Central Tendency

Calculate one **measure of central tendency** for each variable in the dataset. Report that value and explain why you chose that measure of central tendency in no more than a sentence per measure.

*Answer here.*

## Tabulation

Create a **frequency table** or **table of counts** of crime\_type and include it here. What are the most and least common types of crime?

*Answer here.*

## Cross-Tabulation

Create a **cross-tabulation** (**crosstab**) of crime\_type and sex and include it here. In about one sentence, explain whether you think there is some association between these variables. No need for a statistical test—just describe what you see.

*Answer here.*

## Linear Model

Create a **linear model** predicting sentence using crime\_type, sex, and age. Display the estimated coefficients here, preferably with the standard errors / t-values / p-values (you’ll use those later).

*Answer here.*

# Moderate Problems

## Group Means

Calculate the **mean** of sentence for *each* crime\_type. Show these here and comment on what you see in about one sentence total.

*Answer here.*

## Histogram

Create a **histogram** of sentence using either the base R function or ggplot2. Display the plot here. Comment on what you see in about one sentence.

*Answer here.*

## Cross-tab Chi-Square

Run a **chi-square test** on the cross-tab you created in 1.3. What does this result indicate about the presence of an association? Does it agree with your impression from 1.3? No need to write more than 1-2 sentences.

*Answer here.*

## Linear Model Interpretation

In 1.4, you created a linear model predicting sentence using crime\_type, sex, and age. Provide a short **interpretation** of each coefficient (estimate) in the model including its statistical significance. Keep in mind the type of each variable (e.g., categorical) while you do this so that you correctly interpret them.

*Answer here.*

# Harder Problems

## Regression Plot

Create a **scatterplot** of age predicting sentence. Include a regression line.

*Answer here.*

## Prediction

Using your model from 1.4 and 2.4 of sentence predicted by crime\_type, age, and sex, what average sentence would we expect for an age 20 male convicted of an assault?

*Answer here.*

## DAG to Estimator



Given the DAG above, run a **linear model** that identifies the effect of sex on sentence. Present the output here and interpret the coefficient on sex including its statistical significance.

*Answer here.*

# How’s it goin?

In anything from one word to a paragraph, let me know how you’re feeling about the course so far. Overwhelmed? Doing alright? No problem at all? Are there topics that leave you confused? Let me know if there are things that might make the course work better for you.