AE 10: Multiple Linear Regression

Credit Cards

Driver: _____, Reporter: _____, Gopher: _____

> **!** Important
>
> - Open RStudio and create a subfolder in your AE folder called "AE-10".
> - Go to the Canvas and locate your `AE-10` assignment to get started.
> - Upload the `ae-10.qmd` file into the folder you just created. The `.qmd` and PDF responses are due in Canvas. You can check the due date on the Canvas assignment.

Packages + data

```
library(tidyverse)
library(ggformula)
library(broom)
library(knitr)
library(ISLR2)
library(GGally)
library(yardstick)
```

The data for this AE is from the `Credit` data set in the **ISLR2** R package. It is a simulated data set of 400 credit card customers. We will focus on the following variables:

Predictors

- `Income`: Annual income (in 1000's of US dollars)
- `Rating`: Credit Rating

Response

- `Limit`: Credit limit

Analysis goal

The goals of this analysis are to fit a linear regression model that has the following predictors:

- `Income`
- `Rating`
- An interaction term between the two

Exercise 0

What is a credit rating and what is a credit limit as it applies to a credit card? The primary credit rating in the US is called a FICO score. Based on the data, do you think that `Rating` corresponds to the borrower's FICO score?

Exercise 1

Use the function `ggpairs` from the `GGally` package (already loaded) to create a matrix of plots and correlations for our three variables of interest. Note that you will have to use `select` to select the four columns you are interested in. Which variable do you think will be the best predictor of `Limit`?

Exercise 1B (Optional)

If you have extra time, examine the variables individually and comment on anything that you think is relevant.

Exercise 2

Fit a linear model with just `Income` as the predictor and get the p-value associated with it's coefficient. Is it statistically significant?

Exercise 3

Fit a linear model with just `Rating` as the predictor and get the p-value associated with it's coefficient. Is it statistically significant?

Exercise 4

Fit a model with both `Income` and `Rating` as predictors. Find a spot on the white board to write down an equation representing the fitted model. How do the coefficients and p-values of `Income` and `Rating` compare to those in the two models above? Discuss what you see and the possible reasons you see them.

Exercise 5

Interpret all coefficients in the model.

Exercise 6

What is the predicted credit limit for an *single* borrower with a credit rating of 700 and an annual income of \$59,000? Include a 90% confidence interval. Hint: make a new `tibble` and use the `predict` function. How would you interpret this interval in context?

Exercise 7

Add an interaction term between `Rating` and `Income`. Interpret all coefficients in context. What do you notice about the p-values now?

Exercise 8

What is the predicted credit limit for an single borrower with a credit rating of 700 and an annual income of \$59,000? Include a 90% confidence interval. How does it compare to your answer to Exercise 6.

Exercise 9 (Optional)

Note that this data set only considers borrowers who have actually been granted loans. How does this impact the generalizability of our analysis?

To submit the AE

> **!** Important
>
> - Render the document to produce the PDF with all of your work from today's class.
> - Upload your QMD and PDF files to the Canvas assignment.