Homework 1

Statistical Modeling and Learning

due: **Friday, January 16, 2026 at 10:00a**

Instructions: Please submit solutions on Populi as a knitted markdown pdf. For handwritten work, please scan your solution pages and include in the final pdf. Homeworks can be completed individually or in groups of up to three people. Names of all group members must be included on the writeups.

Problem 1: ISL Exercises:

Chapter 2 Exercises – Problems 1, 2, 4a, 4b, 8

Chapter 3 Exercises – Problems 6, 7, 11

Problem 2: This problem will get you comfortable with loading data and performing basic data exploration in your coding environment. Download the file `drone_strikes_venezuela.csv` from the course website and load it into your workspace.

a. Provide summary statistics for all numeric variables and identify any missing values.

b. Compare drug boat strikes versus military strikes in terms of casualties, altitude, strike type, and damage level. Include visualizations.

c. Create a scatter plot of altitude versus casualties. Does this relationship differ between drug boat and military strikes?

d. Do strikes in clear weather differ from non-clear weather in terms of casualties, damage level, or strike type? Include summaries and visualizations.

e. Which region has the highest average casualties? Create a visualization and discuss whether other variables might explain regional differences.

Problem 3: Download the file `stock_returns.csv` from the course website. The dataset contains monthly returns (2004-2025) for the S&P 500 ETF (SPY), Google (GOOGL), Amazon (AMZN), and Coca-Cola (KO).

Brief background: The Capital Asset Pricing Model (CAPM) suggests that a stock's return can be decomposed into a component that moves with the overall market (captured by the stock's "beta") and an idiosyncratic component. The market return is often proxied by a broad market index like SPY.

a. Regress GOOGL returns on KO returns (without controlling for the market). Report and interpret the coefficient on KO.

b. Now regress GOOGL returns on both KO returns and SPY returns. How does the coefficient on KO change? Explain why this happens.

c. Estimate the beta for GOOGL, AMZN, and KO by regressing each stock's returns on SPY returns. Which stock is most sensitive to market movements?

d. Create a scatter plot of GOOGL returns versus KO returns, and overlay the fitted lines from both regression models (with and without SPY). What do you observe?