Homework 3

Problem 1

Download the data in `creatinine.csv`. Each row is a patient in a doctor's office. The variables are:

- age: patient's age in years.
- creatclear: patient's creatine clearance rate in mL/minute, a measure of kidney health (higher is better).

Use this data, together with your knowledge of linear regression, to answer three questions:

A) What creatinine clearance rate should we expect for a 55-year-old? Explain briefly (one or two sentences + equations) how you determined this.

B) How does creatinine clearance rate change with age? (This should be a single number whose units are ml/minute per year.) Explain briefly (one or two sentences) how you determined this.

C) Whose creatinine clearance rate is healthier (higher) for their age: a 40-year-old with a rate of 135, or a 60-year-old with a rate of 112? Explain briefly (a few sentences + equations) how you determined this.

Problem 2

A very important model that finance professionals use to understand asset prices is called the Capital Asset Pricing Model (CAPM). The basic assumption of the model is that the rate of return on an individual asset (like a stock or a mutual fund) is linearly related to the rate of return on the overall stock market. That is, each asset's rate of return is assumed to follow a linear regression model:

$$Y_t^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} X_t + e_t^{(k)} \, ,$$

where $Y_t^{(k)}$ is the rate of return of an individual stock (k) in some given time period t; X_t is the rate of return of the entire stock market in that same time period; and $e_t^{(k)}$ is the residual for stock k in that time period. The superscript (k)'s here are simply denoting the different stocks (Apple, Target, etc), while the subscript t's are denoting the different time periods. Note that the market rate of return (X_t) is a predictor common to all stocks. (The rate of return can be interpreted similarly to an interest rate. For example, if a stock was worth \$100 yesterday and \$102 today, then it gained 2%, for an implied daily rate of return of 0.02.)

The β_1 (slope) term in this regression model is super important to finance professionals; they just call it "beta", and they refer to the β_0 (intercept) term as "alpha." Please watch this short YouTube video to understand how beta is used to think about different stocks.

Once you've watched the video, please turn to the data in `marketmodel.csv`, which contains information on the daily returns for the S&P 500 stock index, denoted SPY, along with the returns for 6 individual stocks: Apple (AAPL), Google (GOOG), Merck (MRK), Johnson and Johnson (JNJ), Wal-Mart (WMT), and Target (TGT). (We can think of the return of the S&P 500 as a proxy for the whole market.) The data start from the beginning of 2019. The entries are interpretable as percentage returns, expressed on a 0-to-1 decimal scale—for example, if the S&P 500 gained 1.5% in value on a given day, the corresponding entry in the data frame would be 0.015.

Regress the returns for each of the 6 stocks individually on the return of S&P 500 (which is like X_t, the market return, in the equation above). Make a clean, professional looking table (e.g. in Excel, or creating your own `tibble` in R that can be displayed via RMarkdown) that shows the ticker symbol, intercept, slope, and R^2 for each of the 6 regressions.

In your write-up, you should include:

- a short introduction, in your own words, on what the "beta" of a stock is measuring and how it is calculated. (Watch the video and summarize it in your own words, making sure to connect it to the regression model we've written down above—this is a bridge you will have to make yourself, using what you know about regression models.) A reasonable aim for your summary is a couple of hundred words here, but this is approximate; nobody on our end is breaking out the word counter.

- the table itself, along with an informative caption below the table, no more than 2-3 sentences in length, to give readers the information necessary to interpret the table.

- a conclusion that answers two questions: in light of your analysis, which of these six stocks has the *lowest* systematic risk? And which has the *highest* systematic risk? (Again, watch the video to understand how this is measured using the regression model.)

Problem 3

The file `covid.csv` contains data on daily reported COVID-19 deaths for Italy and Spain—two of the hardest-hit European countries—during the first pandemic wave in February and March of 2020. The columns in this data frame are:

- *date*: the calendar date
- *country*: Italy or Spain
- *deaths*: the number of reported COVID-19 deaths in that country on that day
- *days_since_first_death*: the number of days elapsed since the first death in that country

Your task is to fit two exponential growth models, one for Italy and one for Spain, using `days_since_first_death` as the time variable. Use the results of your model to characterize the growth rate and doubling time of the daily death total in each country.

Please include the following in your write-up:

1. An estimated growth rate and doubling time for Italy.
2. An estimated growth rate and doubling time for Spain.
3. A line graph showing reported daily deaths over time (using `days_since_first_death`, rather than calendar date, as the relevant time variable) in each country. Your line graph should have two lines, one for each country, distinguished by their color.

Please round the growth rate to three decimal places (e.g. 0.022) and the doubling time to the nearest integer day.

Problem 4

The data in `milk.csv` comes from something called a "stated preference" study, which is intended to measure people's sensitivity to the price of a good or service. The basic framework is that participants are given a fixed budget and presented with a menu of goods, including milk, at varying prices. The key here is that the prices of milk (and other goods) are varied across different participants.

For example, one group of participants might see a gallon of milk priced at \$2, while another group sees it priced at \$4. Each participant has to decide how much milk to buy, along with other goods, within their given budget constraint. By observing how the quantity of milk purchased varies with its price across different groups, the economist can determine how sensitive consumers are to changes in the price of milk.

If participants with the higher milk price buy significantly less milk than those with the lower price, this indicates a higher elasticity, showing that demand for milk decreases as the price increases. On the other hand, if the quantity of milk purchased does not vary much between the different price levels, it suggests that the demand for milk is relatively inelastic with respect to its price.

This approach, by incorporating a fixed budget and a broad choice set, attempts to mimic real-world purchasing decisions more closely and can provide a relatively (though not perfectly) realistic estimation of how consumers would react to price changes in an actual market scenario. It acknowledges that consumers' choices are influenced by their overall budget and the relative prices of all goods they consume, not just the price of a single item.

In `milk.csv`, there are two columns of data arising from this experiment:

- price, representing the price of milk on the menu
- sales, representing the number of participants willing to purchase milk at that price.

The economists' power-law model is $Q = KP^{\beta}$, where P is price, Q is quantity demanded by consumers at that price, where β is the price elasticity of demand.

In light of the data, what is the estimated price elasticity of demand for milk? Briefly describe what you did – no more than a few sentences, together with your estimate.