Homework 3

Problem 1: regression warm up

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. Note that while you fit the model in R, you might find it easier/lower overhead to just do the math for parts A and C "by hand," i.e. using a calculator or R as a calculator.

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: Modeling disease growth

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, with 95% bootstrapped confidence intervals for each.

2. An estimated growth rate and doubling time for Spain, with 95% bootstrapped confidence intervals for each.

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 tenth of a day (e.g. 4.7 days).

Problem 3: price elasticity of demand

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? Give a 95% bootstrapped confidence interval for this quantity. Briefly describe what you did to estimate the elasticity – no more than a few sentences, together with your estimate + interval.