*Project 7.4 Admissions

This project is a continuation of the problem begun in Section 7.6, and assumes that you have read that section.

Suppose you work in a college admissions office and would like to study how well admissions data (high school GPA and SAT scores) predict success in college.

Part 1: Get the data

We will work with a limited data source consisting of data for 105 students, available on the book website.[3] The file is named `sat.csv`.

Write a function

```
readData(fileName)
```

that returns four lists containing the data from `sat.csv`. The four lists will contain high school GPAs, math SAT scores, verbal (critical reading) SAT scores, and cumulative college GPAs. (This is an expanded version of the function in Exercise 7.6.3.)

Then a write another function

```
plotData(hsGPA, mathSAT, crSAT, collegeGPA)
```

that plots all of this data in one figure. We can do this with the `matplotlib` `subplot` function:

```
pyplot.figure(1)
pyplot.subplot(4, 1, 1) # arguments are (rows, columns, subplot #)

# plot HS GPA data here

pyplot.subplot(4, 1, 2)

# plot SAT math here

pyplot.subplot(4, 1, 3)

# plot SAT verbal here

pyplot.subplot(4, 1, 4)

# plot college GPA here
```

Question 7.4.1 *Can you glean any useful information from these plots?*

Part 2: Linear regression

As we discussed in Section 7.6, a linear regression is used to analyze how well an independent variable predicts a dependent variable. In this case, the independent variables are high school GPA and the two SAT scores. If you have not already, implement the linear regression function discussed in Section 7.6. (See Exercise 7.6.1.)

[3] Adapted from data available at `http://onlinestatbook.com/2/case_studies/sat.html`

Copyright Taylor and Francis, 2021

Then use the `plotRegression` function from Section 7.6 to individually plot each independent variable, plus combined (math plus verbal) SAT scores against college GPA, with a regression line. (You will need four separate plots.)

Question 7.4.2 *Judging from the plots, how well do you think each independent variable predicts college GPA? Is there one variable that is a better predictor than the others?*

Part 3: Measuring fit

As explained in Exercise 7.6.4, the coefficient of determination (or R^2 coefficient) is a mathematical measure of how well a regression line fits a set of data. Implement the function described in Exercise 7.6.4 and modify the `plotRegression` function so that it returns the R^2 value for the plot.

Question 7.4.3 *Based on the R^2 values, how well does each independent variable predict college GPA? Which is the best predictor?*

Part 4: Additional analyses

Choose two of the independent variables and perform a regression analysis.

Question 7.4.4 *Explain your findings.*