Assignment 3: Fundamental Concepts from Statistics

Please complete this assignment in Jupyter Notebook with Python 3.

File retention.txt (attached) contains the data that we looked at in class (there are 170 data points in the file). The first line contains the names of the eight variables (described below). Please, look at it carefully using a statistical package of your choice. Write down any interesting observations. Are the data normal and are the interactions linear? As a minimum, you should:

(a) generate descriptive statistics and plot histograms for the following three columns: apret, tstsc, and salar.

(b) perform linear regression of apret on tstsc and salar separately and then of apret on both tstsc and salar.

If you find it useful, the meaning of the columns is as follows:

spend - average spending per student (in dollars)

apret - average retention rate (i.e., percentage of students making it through the studies)

top10 - percentage of incoming freshmen who were among the top 10% students in their high schools

rejr - school's rejection rate (percentage of applicants denied admission)

tstsc - average test scores of incoming freshmen

pacc - percent of admitted applicants who accept university's offer

strat - student-teacher ratio

salar - average faculty salary (in dollars)

Each row is for one of the 170 colleges for which the data was measured.