CE888 Lab 2 Analysis

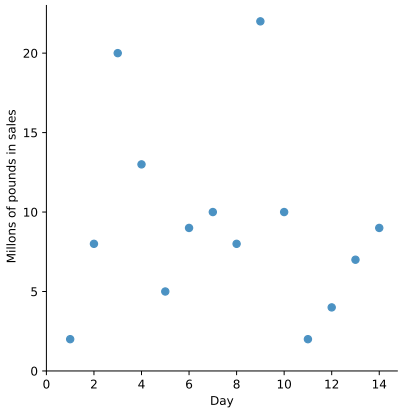
Ogulcan Ozer

January 2019

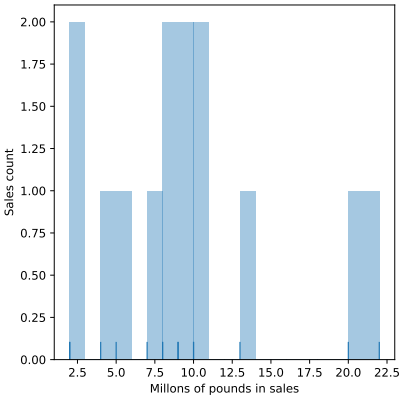
1 Work

First, vehicles.py was implemented. The script reads the vehicles.csv file, outputs the scatter plot of two features, and outputs the histogram of each feature using the seaborn library. Then, the bootstrap function was defined in bootstrap.py to return the mean, lower and upper values of a given sample. Same bootstrap function was also used to get the mean, lower and upper values of the features in the vehicles.csv to see if they are similar, by examining if their confidence intervals overlap. Finally, power.py was implemented, which calculates the power for a given size and alpha value. The two samples are first passed to the perm_test function to get a p-value by applying permutation test. Then the returned p-value is compared to the alpha, which is used to calculate the power after a number of repetitions.

1.1 salaries.py results



(a) Scatter plot of days vs. millions of pounds

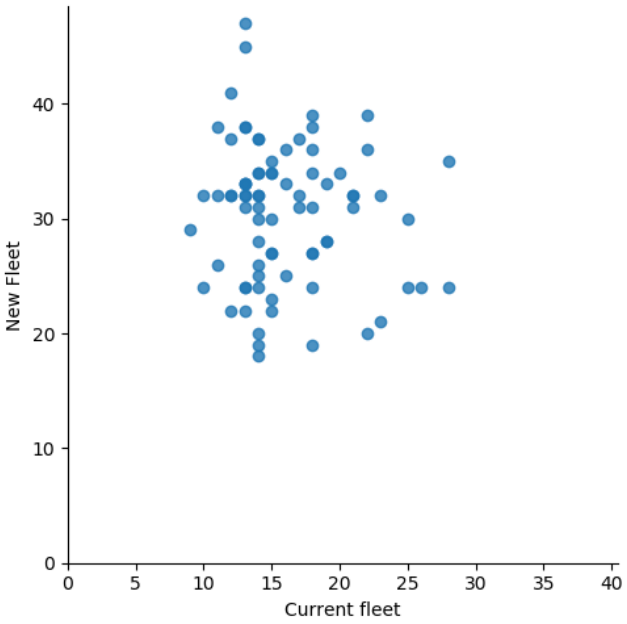


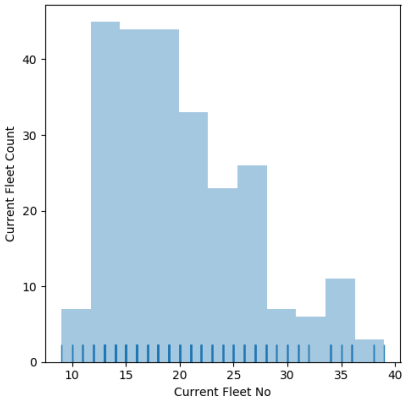
(b) Histogram of millions of pounds

Figure 1: Output figures of the salaries.py

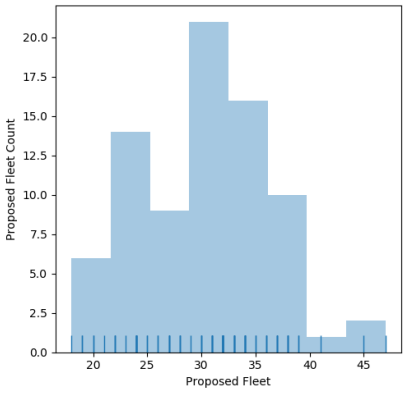
1.2 vehicles.py results

(a) Scatter plot of the Proposed Fleet and Current Fleet features.





(b) Histogram of the Current Fleet.



(c) Histogram of the Proposed Fleet.

Figure 2: Output figures of the vehicles.py.

1.3 bootstrap.py results

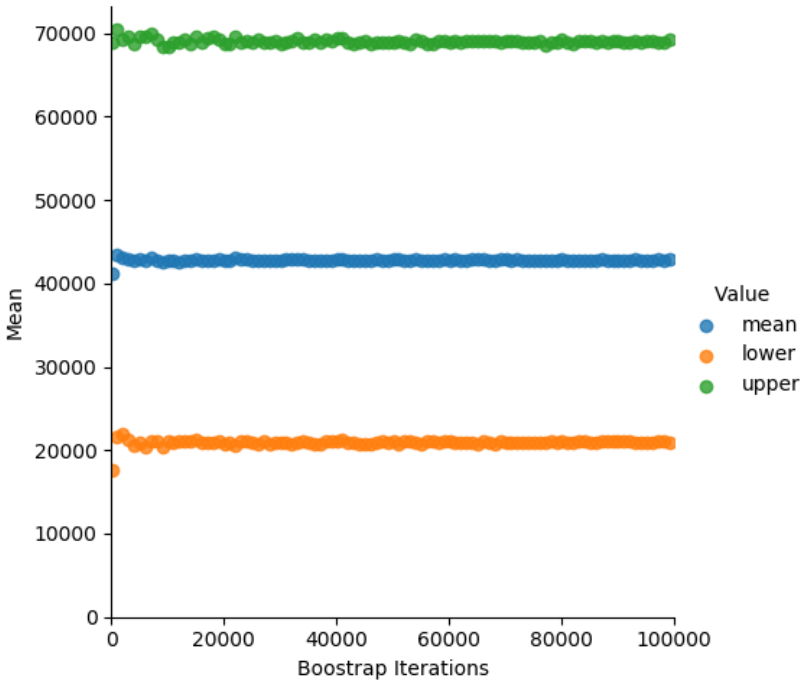


Figure 3: Bootstrap Exercise 1 Confidence Plot.

	Upper	Mean	Lower
Difference	10.879	10.337	9.790

Table 1: Upper, Mean, Lower differences of the New Fleet and Current Fleet