STT 810

Homework 5

Due Monday, November 14 at 11:59:59pm

1. In the car_multi.csv file, there is a field called acceleration. Calculate
   a. The mean value
   b. The standard deviation
   c. The 5^{th} and 95^{th} quantile
2. For this acceleration field, assuming it is normally distributed, conduct a hypothesis test at the 95% confidence level to determine whether you can say the mean value is greater than 15.2. State the null hypothesis. Conduct the hypothesis test by
   a. Constructing an appropriate confidence interval
   b. Calculating a p-value, both by
      i. Analytically
      ii. Running an appropriate simulation
3. A Poisson process generates the following data: {3, 5, 6, 7, 10, 4, 5, 6, 4, 3}, covering 2 second intervals. Run a Monte Carlo simulation to test the hypothesis that the rate parameter is greater than 0.1.
4. For the 4 values you calculated in #1, construct 95% confidence intervals using
   a. Regular bootstrapping
   b. Bayesian bootstrapping
5. Import the data in the dataset called "longtail.csv." Calculate the standard deviation and the 0.99 quantile. Then create 95% confidence intervals for these two quantities, using both regular and Bayesian bootstrapping. Describe what you see with these results.
6. (a) Find the eigenvectors and eigenvalues for the matrix

$$A = \begin{bmatrix} 1 & 7 & 3 \\ 7 & 4 & 5 \\ 3 & 5 & 0 \end{bmatrix}$$

   (b) Express the vectors x = <1, 3, 1> and y = <-1, 4, 9> in terms of the eigenvectors basis for the above matrix.

   (c) Find the inner product of x and y in the original coordinates. Then find the inner product of x and y in terms of the eigenvector basis. Do you get the same value?

7. For this problem we will use the nndb dataset available in the sample data. There are 45 columns, 38 of which are numerical.
   a. Calculate the covariance matrix for the numerical data.
   b. The eigenvalues of the covariance matrix are called the **principal component values**. How many of the 38 principal components are within 0.1% of the largest component?
   c. Transform the data into the eigenvector basis, also called the **principal component basis**. Calculate the covariance matrix in the new basis. What structure can you see from this matrix?