

## DCCS 221 – Probability and Statistics

### Lab # 5

In this lab, we will be looking at a dataset that contains information about the maximum number of motor vehicles being operated during peak times by a transit agency and the total operating cost of the agency. These data were obtained from the Federal Transit Administration ([www.fta.dot.gov](http://www.fta.dot.gov)).

1. Open the data set ***veh.dat*** in Python. The vector  $x$  refers to the number of buses operated during peak service. The vector  $y$  refers to the total operating cost. The vector  $z$  is vector  $y$  divided by vector  $x$ , and represents the operating cost per peak period vehicle.
2. Create a histogram for vector  $x$  **`hist(x)`**
  - a. Use default settings (10 equal bins)
  - b. Use more/less bins **`hist(x,15)`**, **`hist(x,startpoint:by:endpoint)`**
  - c. Use unequal bins **`hist(x,n)`** where  $n$  is a vector, returns the distribution of  $x$  among `length(n)` bins with centers specified by  $n$ . For example, if  $n$  is a 5-element vector, `hist` distributes the elements of  $x$  into 5 bins centered on the  $n$ -axis at the elements in  $n$ , i.e.  $\{n = [500;1500;2500;4000;5000]\}$
  - d. **Challenge Question (Extra pts.)** - Comment on the *skewness*.
3. **Challenge Question (Extra pts.)** - Investigate the cumulative frequencies of vector  $z$ .
  - a. Create a frequency diagram and a cumulative frequency diagram (on a same canvas) (Hint: Use `np.histogram()` to get the count values and find PDF by dividing it with `sum(count)`. Then, use `np.cumsum()` to generate CDF.)
  - b. Using `np.percentile(z)`, find the quartiles, interquartile, and quantiles at 0.1 and 0.9 cumulative frequencies. And mark the point on the cumulative frequency diagram.
4. Look at vector  $z$  using a box plot.
  - a. Notice that the bottom and top edges of the box correspond to the 1<sup>st</sup> and 3<sup>rd</sup> quartile, respectively.
  - b. Identify any outliers and suggest reasons for their presence.
  - c. Also plot the box plot for vector  $x$ ,  $y$  on a same plane.
5. Compare vectors  $x$  and  $y$  (vehicles and total cost) by using a scatter plot.