IE 306 Assignment 2

Ömer Faruk Özdemir
Serkan Özel
İsmet Dağlı

There are 7 part of this assignment. They sorted according to the assignment order. Also, there are some figures in order to explain more clearly.

Part - 1

Step 1: Day 1 and Day 2 data is merged and they are sorted in an increasing manner.

Step 2: Then, we used this formula to maximum. Each of the values are calculated in the excel. Negative values are discarded. They will be evantually unimportant because the counter part of the Max will be higher(positive).

Result is: the max Value is **1.8775.**

$$D = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)}, R_{(i)} - \frac{i-1}{N} \right\}$$

R(i) = the Ith from sorted data
N = Number of Data

D(0,05 , 976) = 0.043471722
Significance level = 0.05
N = 976

Evaluation : 1.8775 is higher 0.043471722. **So, Reject H0, reject uniform claim**

Part – 2

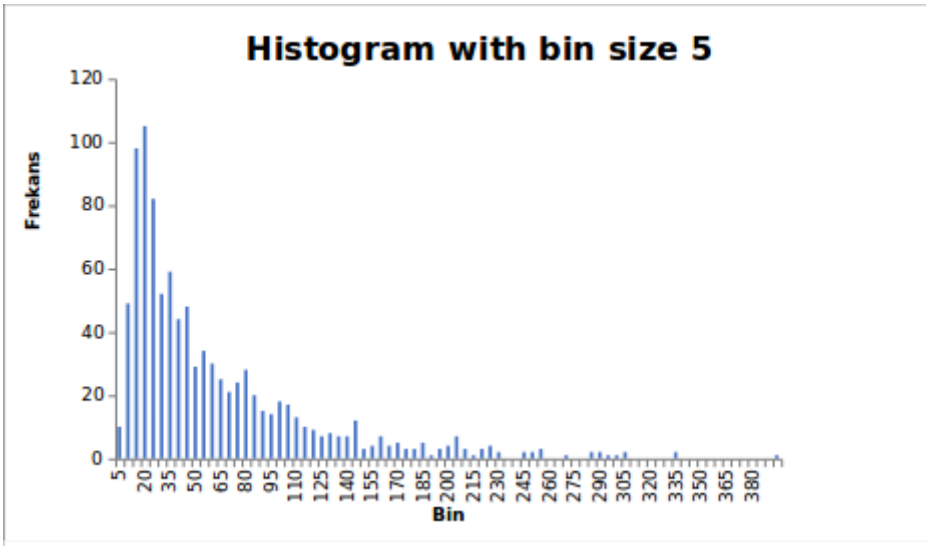
We used some descriptive function and calculated some values. We calculated the values. Here are the table of the values.

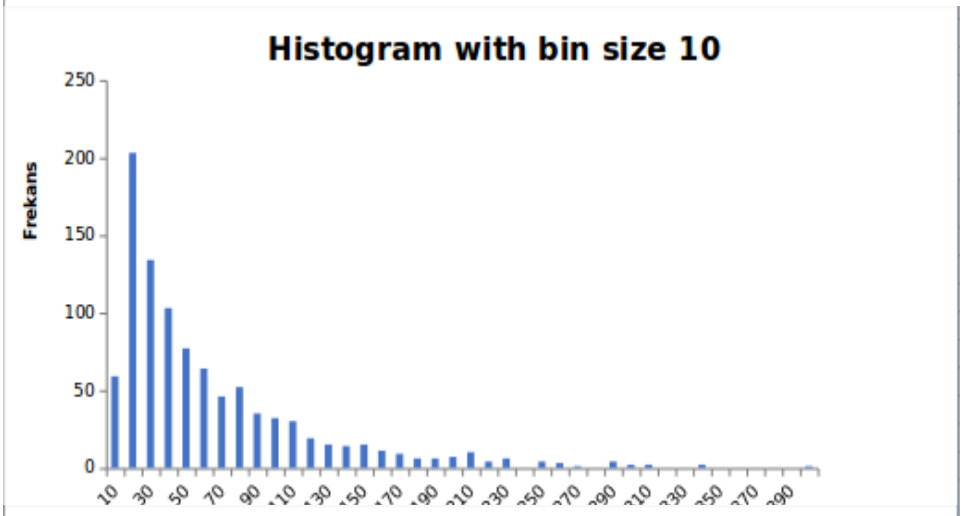
	Day 1	Day 2	Merged Data
Mean	55.168	64.224	59.653
Standart Deviation	56.109	58.575	57.486
Median	34	45	39
Mode	16	14	16
Standart Error	2.540	2.654	1.840
Sample Variance	3148.165	3431.047	3304.670
Kurtosis	6.364	3.121	4.494
Skewness	2.251	1.708	1.954
Range	390	332	391

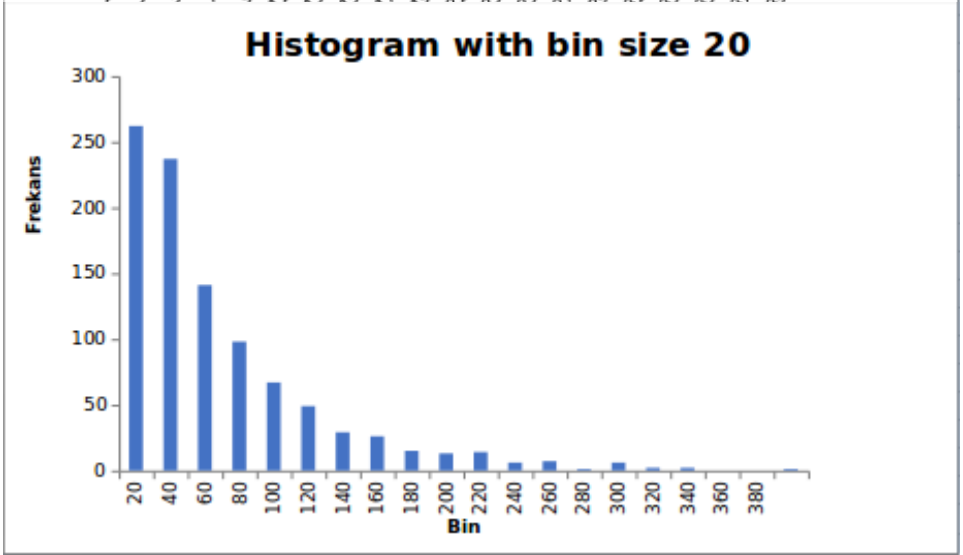
Minimum	1	0	0
Maximum	391	332	391
Sum	26922	31277	58221
Count	488	487	976

Part – 3

The frequency of the data is drawn as a histogram. There are 3 separate histograms for 5-10-20 seconds. Here they are :







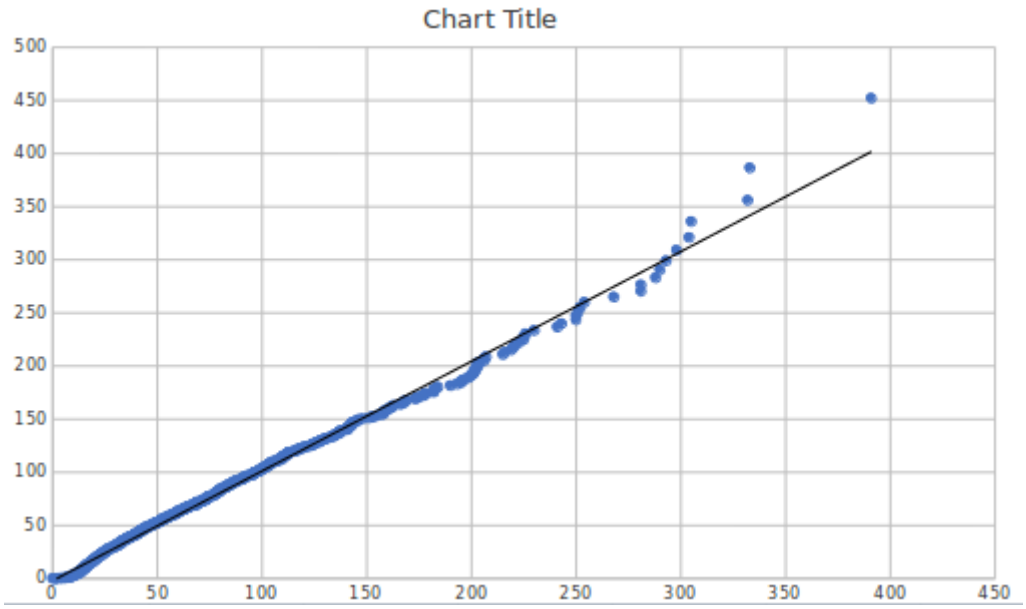
Part -4

We created bin with 20 intervals from 0 to 400. Then, we have observed values between the intervals. We calculated the expected value if the distribution is exponential. We have mean = 59.653 and Lambda = 0.092. Our function is F(x) = 1-e^-(BIN*x). BIN refers to probability of bin. Then, We used the formula of (Oi-Ei)^2/Ei for each interval.
Degree of Freedom = 975
X^2(0,05,975)=1048

So, we ended up test this result: Fail to reject exponential distribution with chi squared test

Part -5

We calculated the (j-0.5)/n and invCumExp((j-0.5)/n). Then we plotted the graph as :

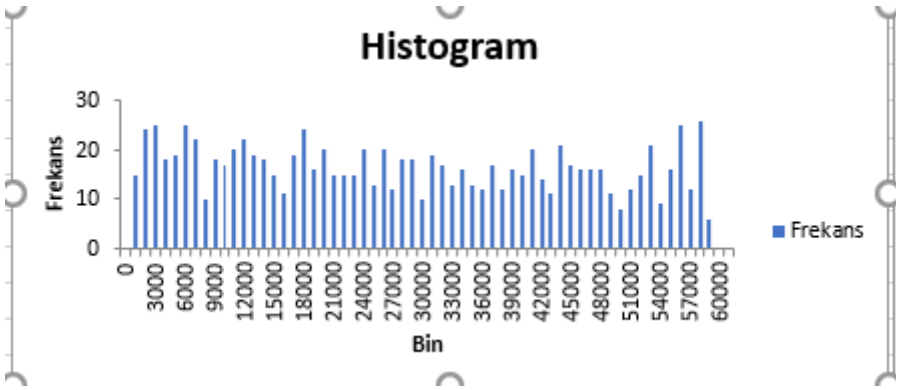


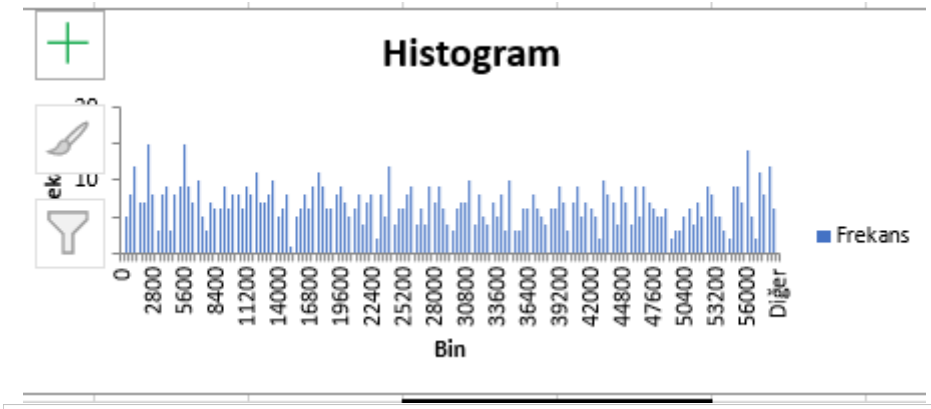
We concluded that Y axis shows inverse exponential cdf values given (j-0.5)/n, x axis shows data points. Since this plot is close to a line we can say that this data comes from exponential distribution.

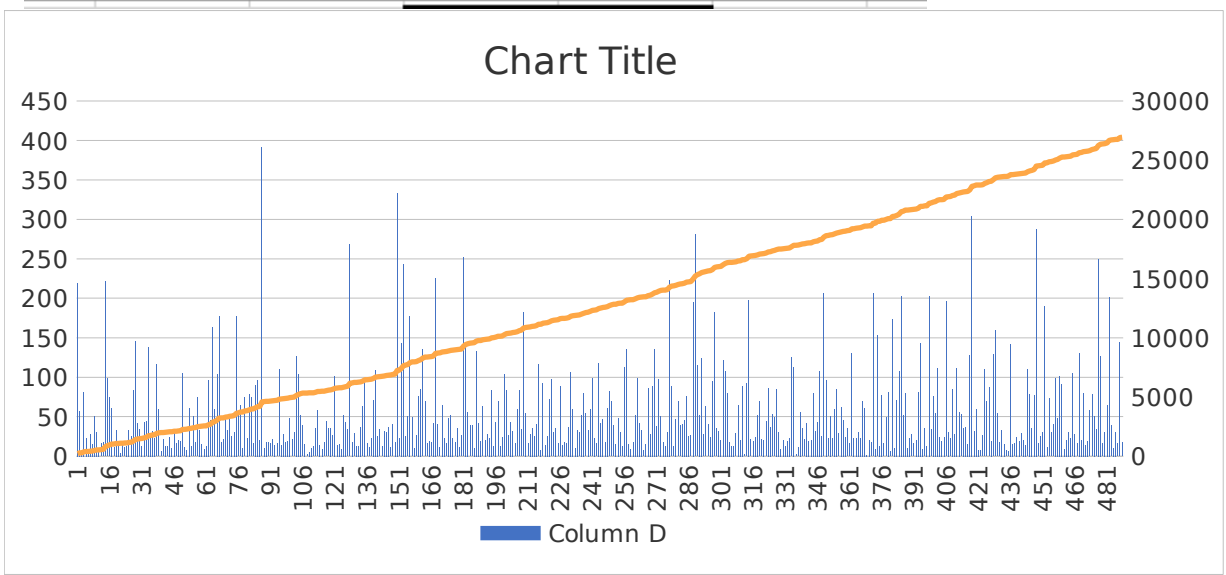
Part – 6

We sum up the data an increasing manner. Then, we created two different histogram with bin 0-1000 and bin 0-400. There is no obvious pattern. We can't observe any pattern with two different bin. Here are the histogram showing its values:

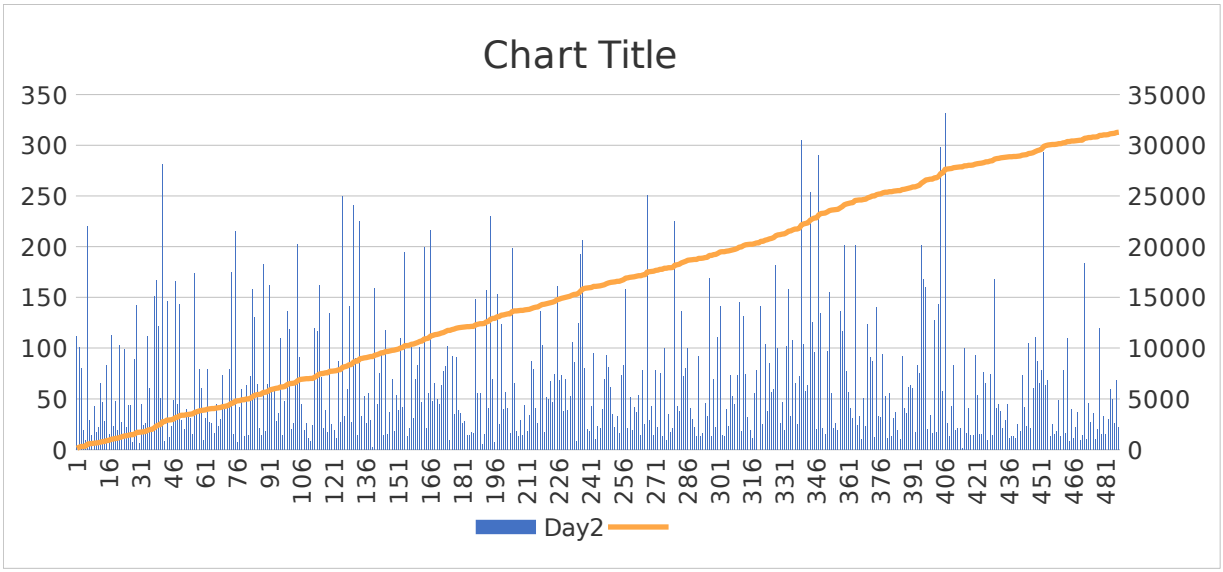
There is **no obvious pattern.**







Here are the day1 and day2 chart values. If we analyze the histograms and charts, we can easily say that **the data is stationary**



Part – 7

We merged the data and created the lag1 and lag2 data by deleting first and first+second data, respectively. Then, we use the mean and variance that we calculated at part 2 . After that, we evaluated the part one by one.
For lag1: The result of (first data – mean) times the result of (second data – mean). All data will be evaluated this way. Then, we found the coveriance by summing of the all results and dividing by N(number of data). To find the correlation, coveriance is divided by variance.
For lag2: The result of (first data – mean) times the result of (third data – mean). All data will be evaluated this way. The rest is repeated by the same manner. Our result are listed above:

Mean: 59.653
Variance: 3304.670

For Lag1
Coveriance : 28.672
Correlation: 0.0087

For Lag2
Coveriance : 76.381
Correlation: 0.0231

The correlation results are so close to 0. Therefore, we don't have any clue about the data is autocorrelated. So, we can't say its autocorrelated.