IE 306.02 SIMULATION ASSIGNMENT 1 REPORT

GroupID: 20 - Emirhan Saraç - Berkay Kozan - Gökhan Tekel

6 May 2019

Contents

1 Question 1

We are asked to test inter-arrival times are distributed uniformly between 0 and 400 seconds and our test method will be the Kolmogorov-Smirnov test with a significance level of 0.05.

First we sort the both days in ascending order. Then we count number of samples, and maximum value. We compute (I/N) and R_i for both days. Then we compute $((I/N) - R_i)$ and $(R_i - ((I-1)/N))$, if values are less then 0, we put '-'. Highest of these both will be called "D". We will find D values for day 1 as $0,604795082$ and day 2 as $0,559835729$. We need to test this respectively $D_{0.05,N}$, N is the number of samples(488,487 respectively). We found testing $D_{0.05,N}$ values are $0,061478298$ and $0,061541385$ and since our D values from observations have higher values, we reject that inter-arrival times are distributed uniformly between 0 and 400 seconds.

2 Question 2

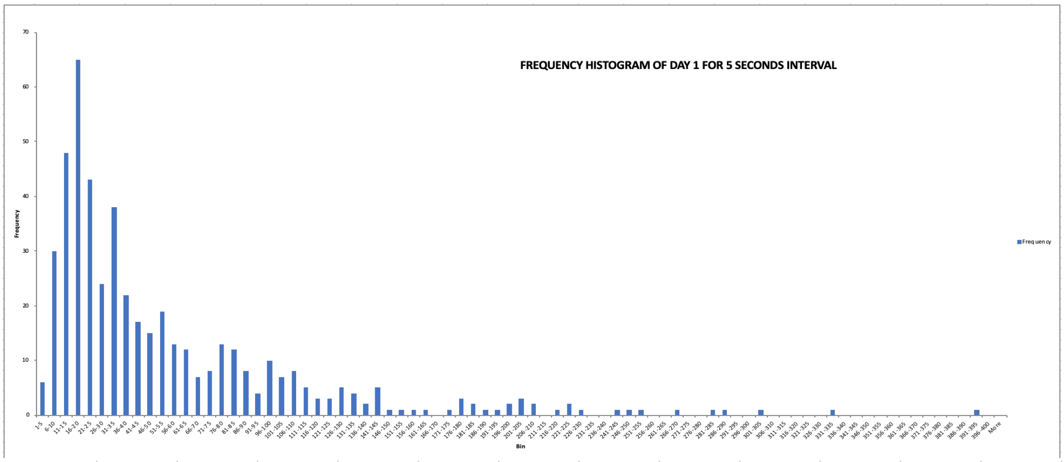
We are asked to find; sample mean, standard deviation and other descriptive statistics that you deem appropriate from both days.

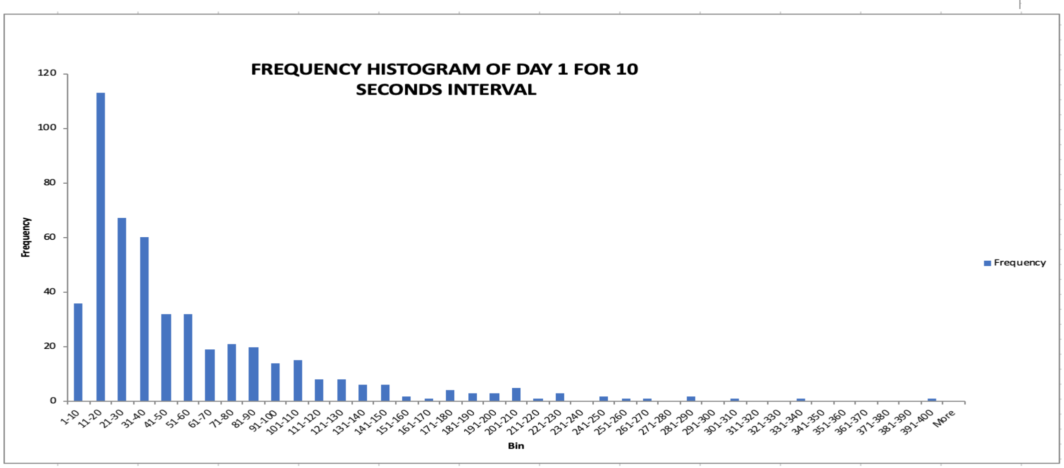
	Day 1	**Day 2**
Mean	55,16803279	64,2238193
STD	56,10850849	58,57513559
Variance	3141,713568	3424,00124
Max	391	332
Min	1	0
Mod	16	14
Median	34	45

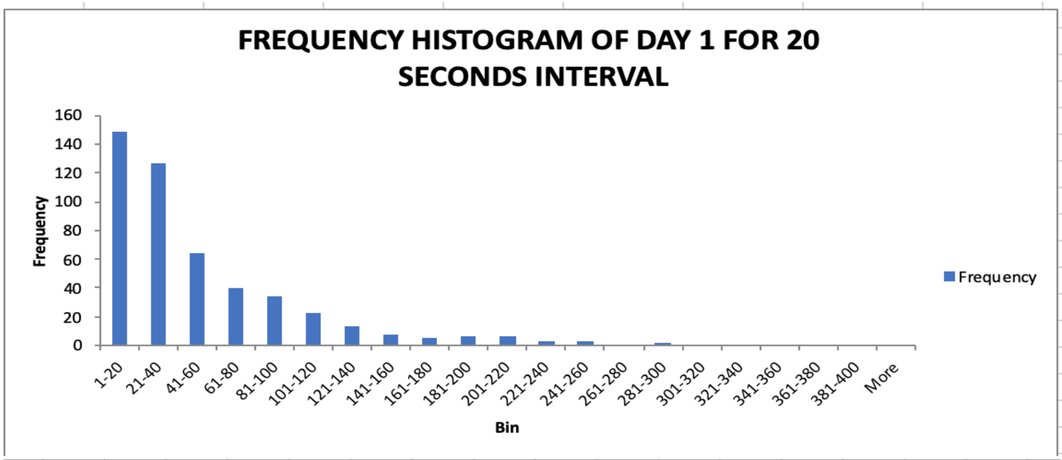
3 Question 3

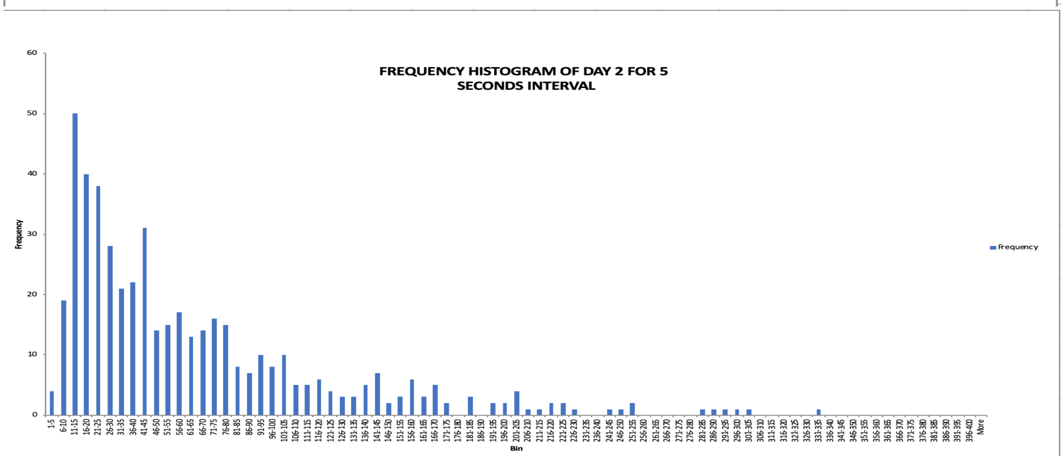
We are asked to draw frequency histograms of the data for 5, 10 and 20 second intervals. We use bins and then create these charts.

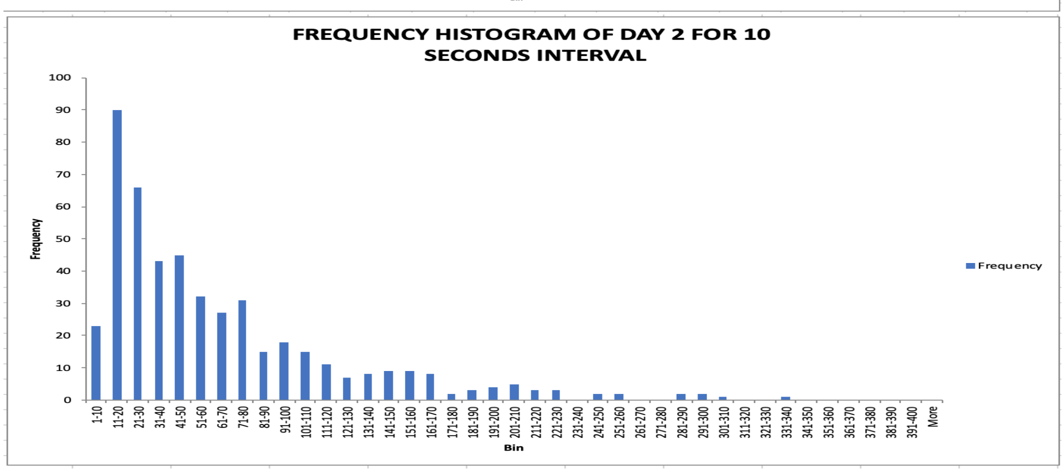
3.1 Figures

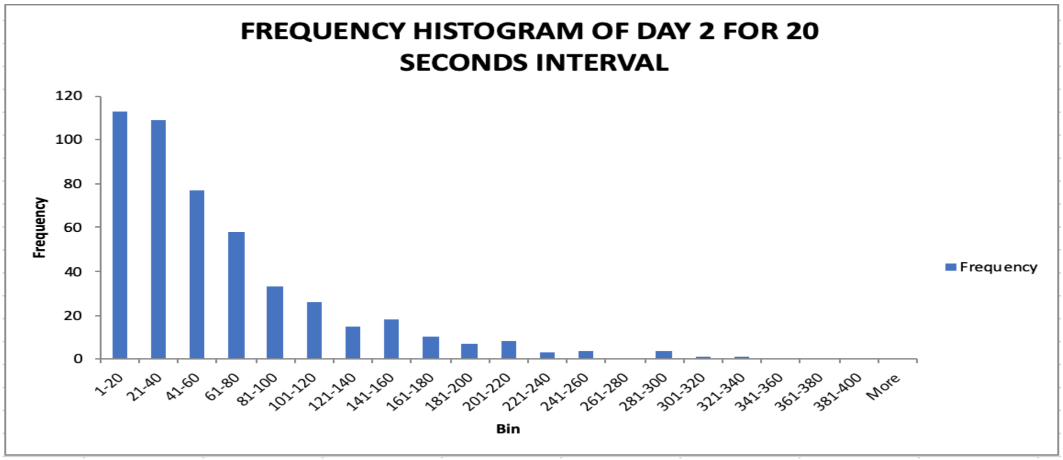





FREQUENCY HISTOGRAM OF DAY 1 FOR 20 SECONDS INTERVAL
Frequency
160
140
120
100
80
60
40
20
0
1-20
21-40
41-60
61-80
81-100
101-120
121-140
141-160
161-180
181-200
201-220
221-240
241-260
261-280
281-300
301-320
321-340
341-360
361-380
381-400
More
Bin
Frequency


FREQUENCY HISTOGRAM OF DAY 2 FOR 5 SECONDS INTERVAL
60
50
40
30
20
10
0
Frequency
Bin
Frequency


FREQUENCY HISTOGRAM OF DAY 2 FOR 10 SECONDS INTERVAL
Frequency
100
90
80
70
60
50
40
30
20
10
0
1-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
101-110
111-120
121-130
131-140
141-150
151-160
161-170
171-180
181-190
191-200
201-210
211-220
221-230
231-240
241-250
251-260
261-270
271-280
281-290
291-300
301-310
311-320
321-330
331-340
341-350
351-360
361-370
371-380
381-390
391-400
More
Bin
Frequency



3.2 Analysis

It looks like last "Day 2 For 20 second Interval" has a exponential distribution. Other figures also looks like exponential distribution except for the beginning part.

4 Question 4

We are asked to perform a chi-square test at a significance level of 0.05 with 10 second intervals to test whether the data comes from an exponential distribution where the mean is as found in Question 2. First, we got the bin-frequency table from the question 3, for 10 second intervals and each day. Then we created expected values of bin and frequency for both days. Expected value for an interval is $F(M, \lambda, True) - F(m-1, \lambda, True)$ where F is cumulative distribution function, M is the biggest element of interval and m is the smallest element of interval, and λ is $1/\mu$. This way, we can get the expected probability for each interval. Then we multiply this probabilities with total number of elements, and that gives us the expected number of minutes for each interval. Then we find

$$\tilde{\chi}^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

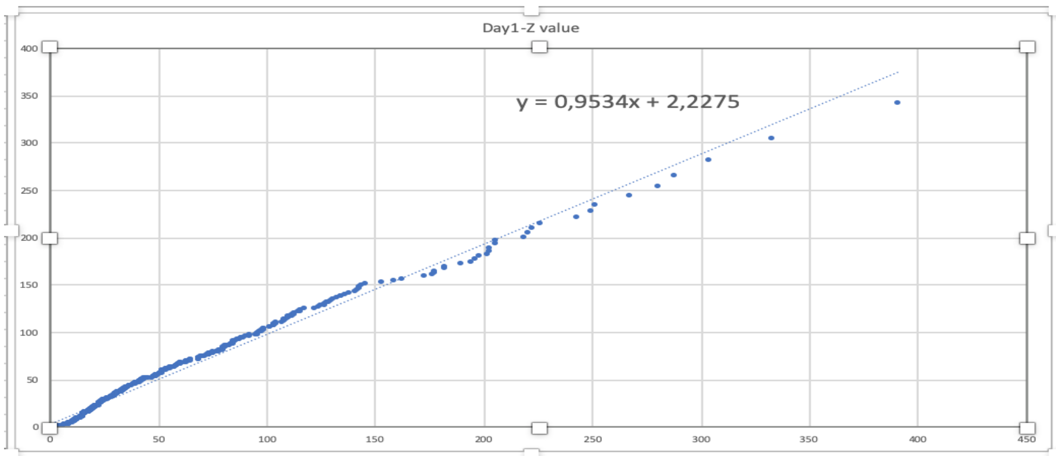
for each day, where n is number of intervals, in this case it is 40. We have 2 test statistics. Afterwards, we compare these test statistics with $\chi_{0,05,38}$. Degrees of freedom is 38 and significance level is 0.05. We infer if

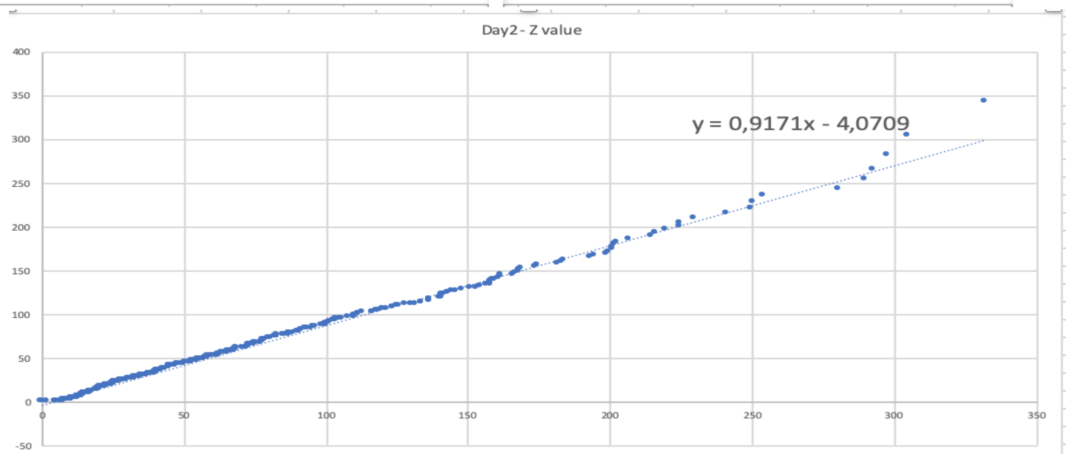
$$\tilde{\chi}^2 \geq \chi^2_{0,05,38}$$

reject that the data comes from an exponential distribution, else we do not reject. From the result, we see that the equation above holds for both days. Therefore for each day, we reject that the data comes from an exponential distribution.

5 Question 5

5.1 Figures





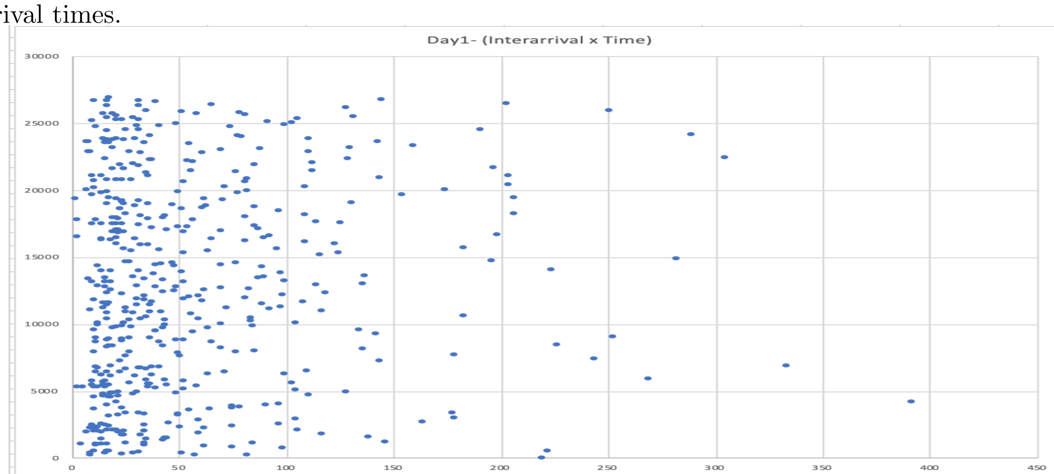
5.2 Analysis

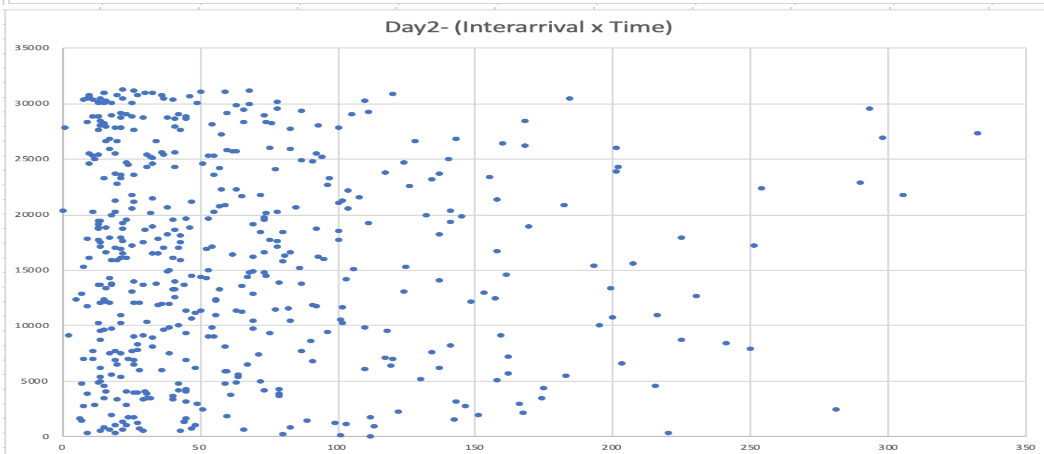
We know that if the general trend of the Q–Q plot is steeper than the line y = x, the distribution plotted on the vertical axis is more dispersed than the distribution plotted on the horizontal axis. Since our distribution is not steeper than exponential distribution but so close, we can say that our distribution can be called exponential distribution with less dispersed. So that QQ test "FAIL TO REJECT" case "Our distribution is exponentially distributed.

6 Question 6

6.1 Figures

In these figures, Y axis corresponds to time and X axis corresponds to inter-arrival times.



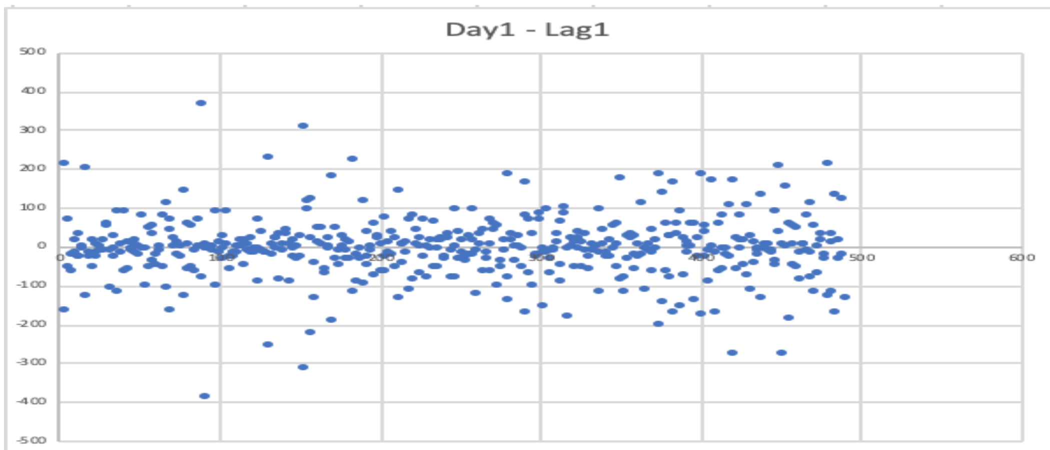


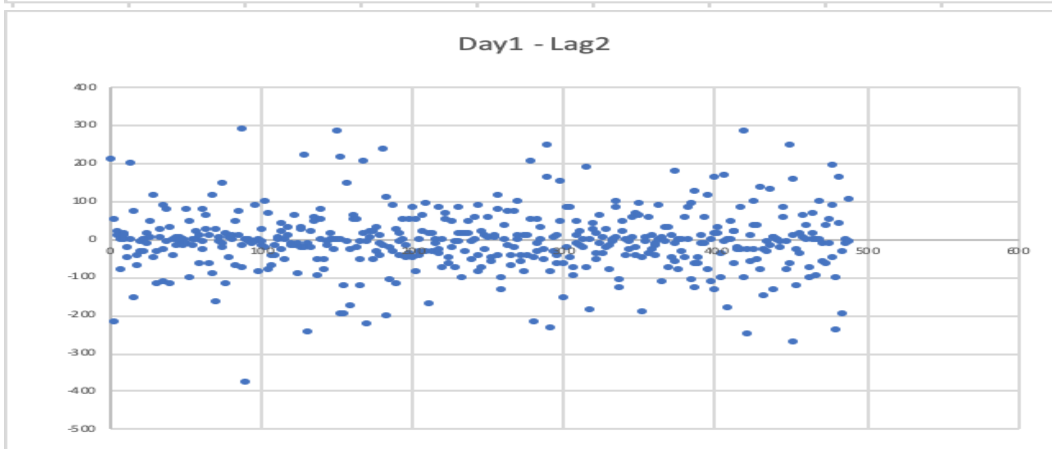
6.2 Analysis

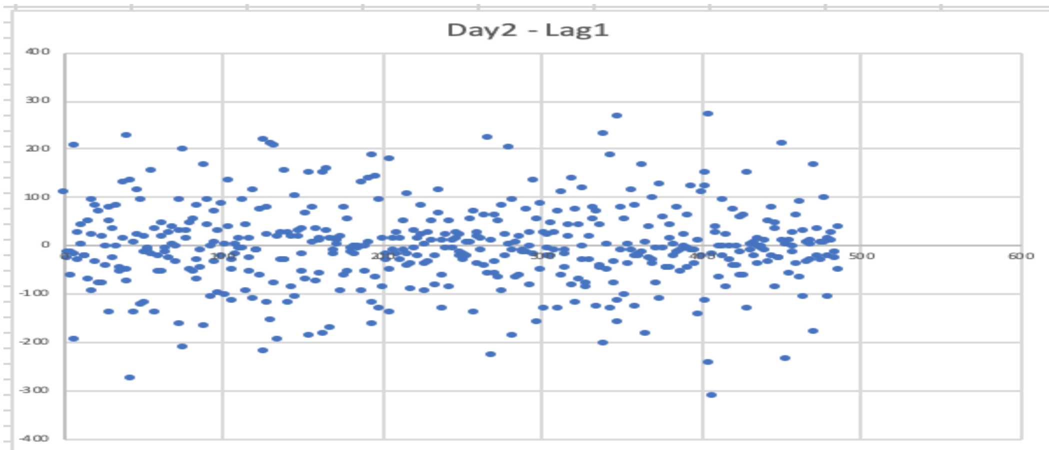
We can see that inter-arrival times are piled between 0-50 and as going further from 50 number of points are decreasing. It is kinda exponential distribution. Since the inter-arrival times are not changing by time(Y axis is time), we can say that inter-arrival data points are stationary.

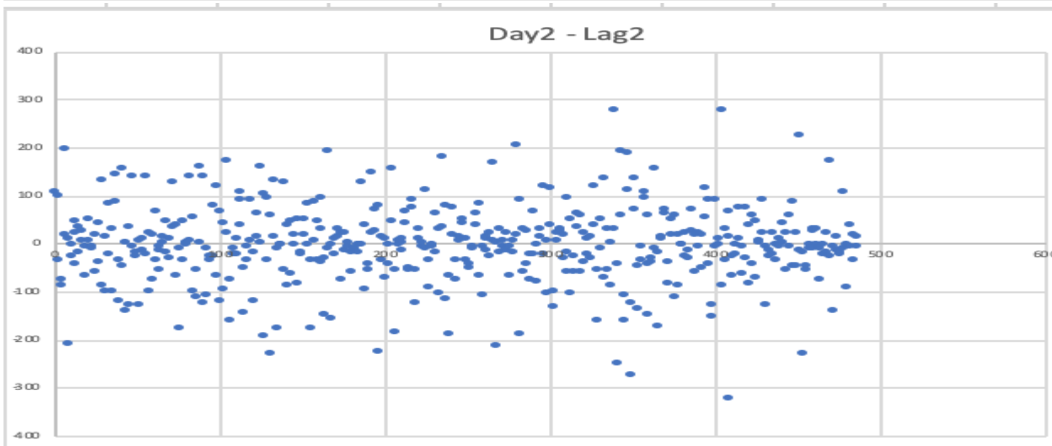
7 Question 6

7.1 Figures









7.2 Analysis

Day1, Lag1 Correlation value $-->0,027839541$
Day1, Lag2 Correlation value $-->-0,054755263$
Day2, Lag1 Correlation value $-->-0,020612497$
Day2, Lag2 Correlation value $-->0,081254542$

Auto-correlation results range from -1 to +1. At end points auto-correlations exist. Since our points are not close to end points, except day2-lag2, we can say that ther is no auto-correlation.