

Q1-

Sample Statistics

MEAN	232,884097035019000
STANDARD ERROR	10,983732750992600
MEDIAN	129,332246424698000
MODE	#N/A
STANDARD DEVIATION	299,193331773552000
VARIANCE	89516,649777758800000
Kurtosis	12,043704957215900
Skewness	2,982381442909400
RANGE	2292,327775173440000
MINIMUM	0,280512375578876
MAXIMUM	2292,608287549020000
SUM	172799,999999984000000
COUNT	742

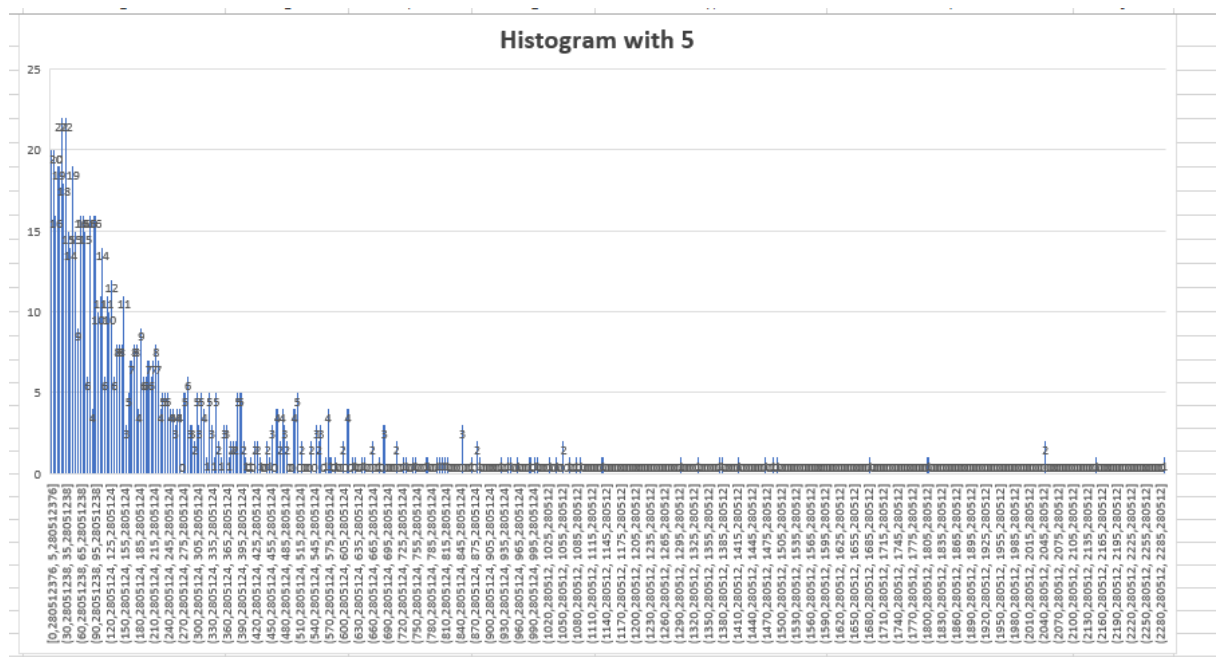
Q2-

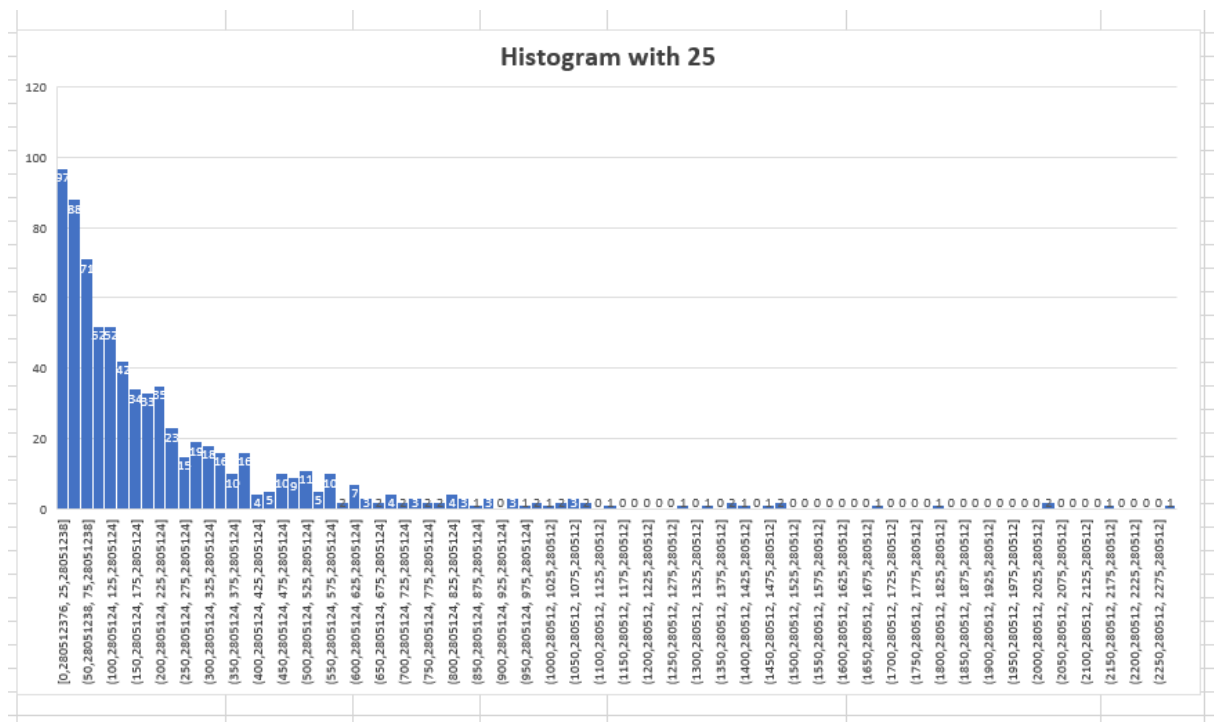
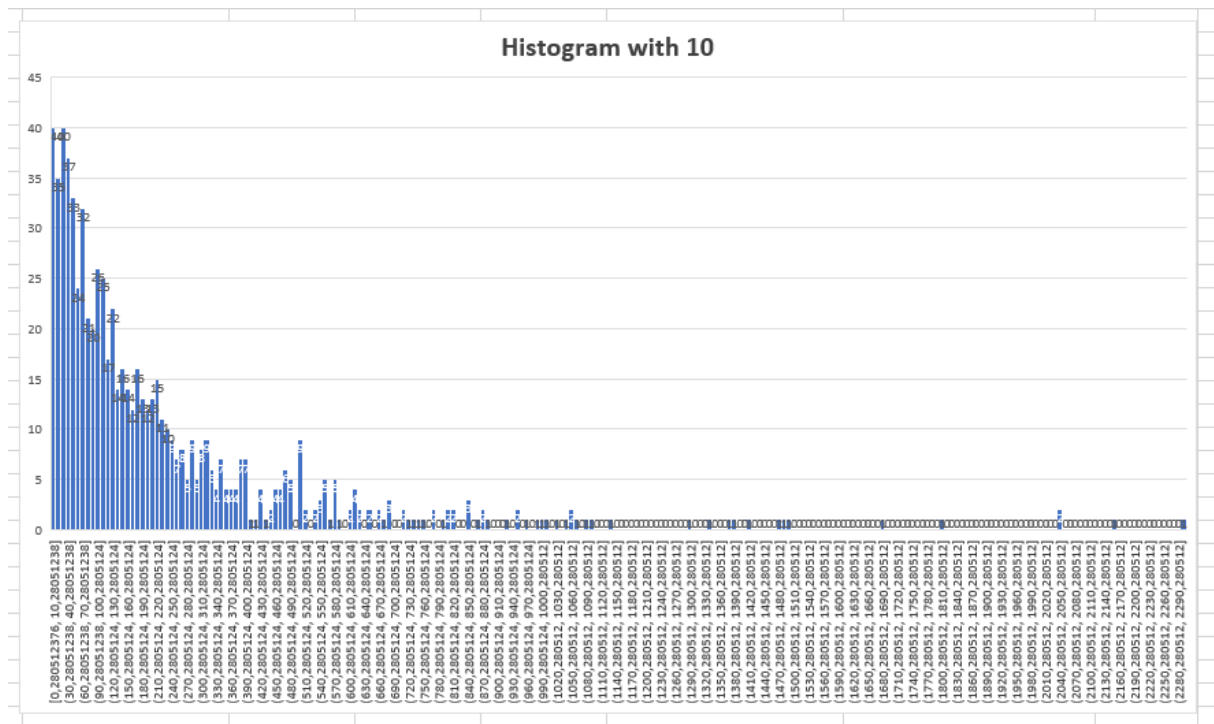
Kolmogorov-Smirnov test yields $D=0,423249547201150$

For %5 confidence level and 740 freedom (-2 for parameter estimation of mean and standard deviation), D critical = $D_{0.05, 740} = 0,049994594302379$

Since $D \geq D_{critical}$ and D is way bigger than D critical, we Reject H_0 . Interarrival times most probably **not** normally distributed with mean 200 seconds and standard deviation 50 seconds.

Q3-





As the interval size increases, the number of observations we see in each bin will increase obviously.

As the interval size increases, the graph becomes smoother, nicer.

By looking at the plot, it seems like distribution is exponential but we can't be sure unless we do some statistics test such as chi-square or k-s test.

Q4-

Chi-square value we found is 2118,79271184054

Chi-square critical value is Chi-square 0.05, 228=264,2242178. Freedom is 230-1-1=228 since there is a parameter estimation related to lambda for exponential distribution.

Since $\text{Chi-square} \geq \text{Chi-square Critical}$ and Chi-square we found is way bigger than the critical value, mostly probably the distribution is not exponential. So, we reject H_0 .

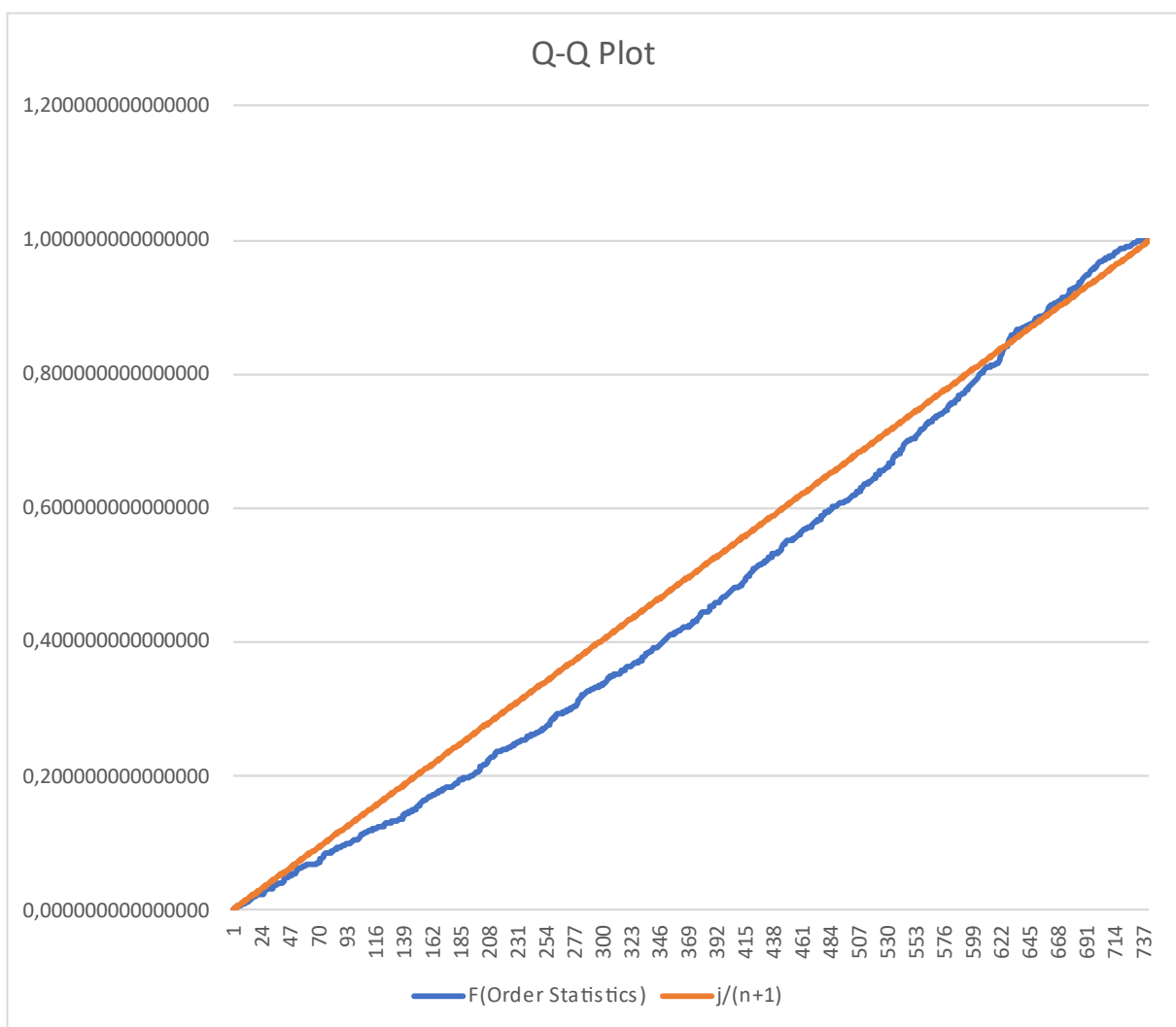
However, on the other hand, there are 3 outliers in our dataset which are the followings and causing us to reject H_0 . In my opinion, we should ignore these outliers and shouldn't reject H_0 .

595,778347881153000

325,690359442805000

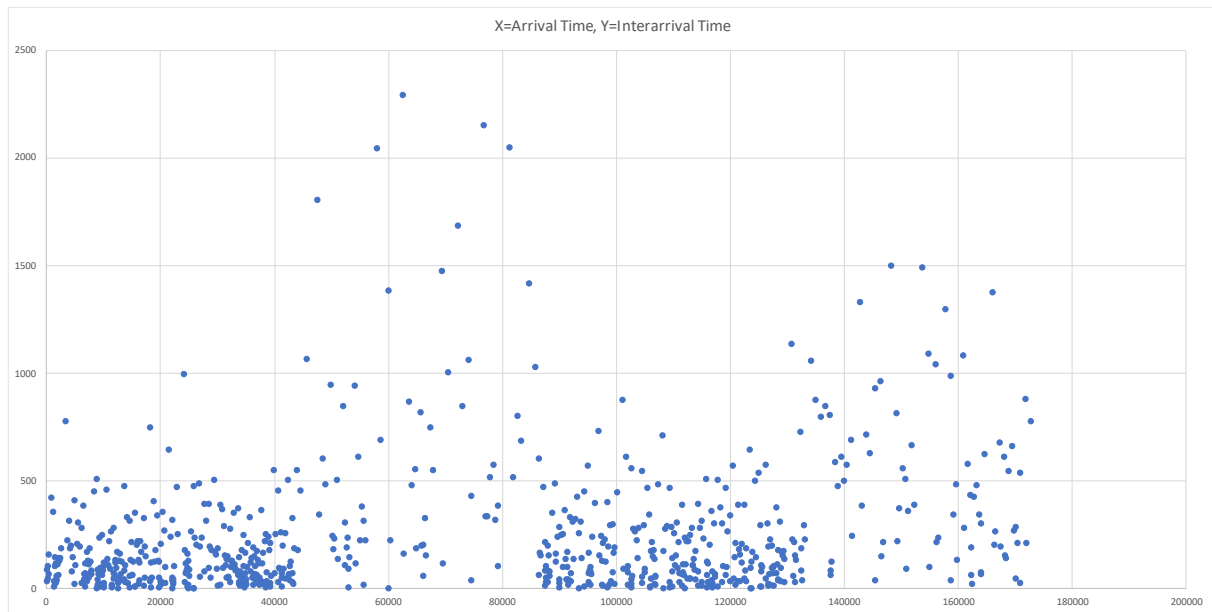
813,313064940807000

Q5-



By looking at the plot, we can say that dataset resembles exponential distribution. This relation is not strong (since they don't collide) but let's say it is %80 percent similar.

Q6-



Yes, there is an obvious pattern as it can be seen from the figure.

0-40000 and 80000-120000 behaves similarly.

40000-80000 and 120000-160000 behaves similarly.

So, the pattern is that 0-80000 is a periodic graph somehow.

Since there is a pattern, data is close to being stationary (such as 0-40000 and 80000-120000 will have similar mean).

Q7-

We calculated lag1, lag2, lag3 and for each of them, we found that that there is biased. So, we reject H_0 .

Lag1's Z-value is 3,856452221

Lag2's Z-value is 9,662343776

Lag3's Z-value is 25,70590307

For confidence interval %5 (Z critical=1.96 and -1.96), all the three fails the test. Therefore, we reject H_0 . There is relationship between a variable's instant value at that time and previous past value.

Q8-

Since all tests failed, we are not able to apply theoretical distribution for generating 10 days of interarrival time. So, using empirical continuous distribution makes sense. Therefore, we generated 10 days of interarrival time by the help of empirical continuous distribution that can be found in excel Q8 file.

