IE306 Assignment 2

Spring 2020

Muhsin Etki - 2015400093

Mehmet Berk Kemaloğlu - 2015400081

Ali Batır - 2015400261

**Question 1 :**

We tried to test whether it is safe to assume that inter-arrival times are distributed uniformly between 0 and 400 seconds by using the Kolmogorov-Smirnov test with a significance level of 0.05.

First of all , we sort day1 and day2 in ascending order. Then we count number of samples, and maximum value. We compute (i/N) and Ri for both days. Then we compute ((i/N) − Ri) and (Ri − ((i− 1)/N)) . Maximum of the calculated values is D. We found D for Day 1 as 0,647495446 and Day 2 as 0,597320128 . We found testing value as 0,061564307 and since our D values from observations have higher values, we reject that inter-arrival times are distributed uniformly between 0 and 400 seconds.

**Question 2 :**

We found sample mean, standard deviation and other descriptive statistics.

|  |  |  |
| --- | --- | --- |
|  | Day 1 | Day 2 |
| **Sample Mean:** | 45 | 54 |
| **Standart Deviation:** | 52,812353 | 55,0840901 |
| **Lambda:** | 0,018934964 | 0,01815406 |
| **Variance:** | 2783,429169 | 3028,03924 |
| **Mod:** | 10,22222222 | 20 |
| **Median:** | 26 | 35 |
| **Max:** | 434 | 339 |
| **Min:** | 1 | 0 |
| **Range:** | 433 | 339 |
| **Sum:** | 21829 | 26240 |
| **Count:** | 488 | 488 |

**Question 3 :**

We drew frequency histograms of the data for 5, 10 and 20 second intervals

* *5 second interval*
* *10 second interval*
* *20 second interval*

*Comment on the shape of the histograms :*

All of them looks like exponential distribution ,especially Day1\_20 and Day2\_20 .

**Question 4 :**

We performed 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. In this question the intervals are constant however expected numbers are not constant . We need to calculate expected numbers corresponding to their intervals . Expected value for an interval is F(M,λ,True)−F(m−1,λ,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/μ. 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

nesne, saat içeren bir resim

Açıklama otomatik olarak oluşturuldu

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 . Degrees of freedom is 38 and significance level is 0.05. We infer if

reject that the data comes from an exponential distribution, else we do not reject.

For Day1;

According to the chi-square test we reject the null hypothesis that the data comes from an exponential distribution.

For Day2;

According to the chi-square test we fail to reject the null hypothesis that the data comes from an exponential distribution.

We used 38 as degree of freedom. When we use degrees of freedom formula (n-k-1, n is number of intervals, k is the number of values we obtained from sample which is 1 since we obtained mean from sample), it gives us 38.

**Question 5 :**

We drew the QQ-plot to test whether the data comes from an exponential distribution.

Q-Q plot show us that the data is nearly distributed exponentially.

**Question 6 :**

We plot the inter-arrival times with respect to observation times.

Since all the values are stacked between 0-50 then gradually decreased so we can say that this is exponential distribution.

**Question 7 :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Day 1** | **Day 2** | **Lag** |
| Correlation | 1 | 1 | 0 |
| Correlation | 0,002393 | -0,01665 | 1 |
| Correlation | -0,01529 | -0,00848 | 2 |

When computing autocorrelation, the resulting output can range from 1 to -1, in line with the traditional correlation statistic.

Since all the results are near to zero , we can say that there is no auto-correlation.