**Module 2 Project 2**

****

**By**

**Sunil Raj Thota**

**Date: 03/15/2021**

**Title: Project 2 Report**

**ALY 6050 – Introduction to Enterprise Analytics**

**Prof. Tamir Hegazy**

**Introduction**

In this assignment, I am using R and R Studio to generate different random numbers in and execute Chi-Square Goodness of Fit test. It helps us to identify and verify how well a statistical model fits a set of observations and whether these values have its place to the probability distribution or not? This test gives an idea between the Expected values and Observed values of each problem. The random numbers are generated by using runif() function in R, be in unform distribution or non-uniform distribution or normal distribution.

Also, I have to define how the local emergency facilities like hospitals that are able to withstand with natural disasters. I have to deal with the 5 hospitals on the problems given. I have also calculated the time taken for the victims to the hospitals. The total transport is calculated by the assumptions that one emergency vehicle comes to the campus and other vehicle comes to hospital. It is nothing but the sum of each victims to the hospital.

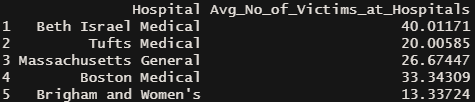
To perform the analysis, we do not have any data to utilize, so we use the triangular probability function which has 3 values like minimum, maximum, and mode. The distribution is in triangular plot. I also performed law of large numbers on the Beth Israel Medical Hospital and calculated 95% CI to the total time taken to travel to Beth Israel Medical Hospital.

I will also tend to determine the type of the distribution for each problem and showcase the plots, results, and findings. In this report, I have implemented 2 Problems and related questions.

**Analysis**

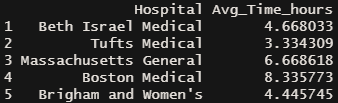
**Problem 1(i):**

In this problem, I have to work on generating 5000 random numbers in R by using the runif() function. Here, we took minimum, maximum, and mode values as given 20, 300, and 80. Then I combined the names of the hospitals, time taken in hours and mins for the transportation. Also, randomly generated 5000 simulations to allocate to the disaster. After this, I had averaged the victims that can come to the hospital.



**Problem 1(ii):**

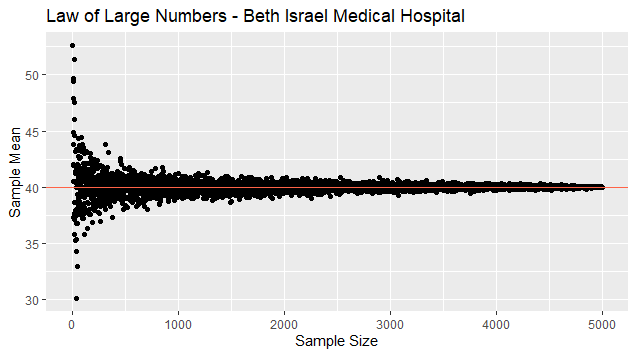
In this problem, I need to find the Avg. total times to the hospitals by each victim in hours. From the table below, we can see that the Beth Israel Medical hospital is having an Avg. Time in hours as 4.668033. Th highest is Boston Medical Hospital with 8.335773 and the lowest is with the Tufts Medical as 3.334309. By this data, its not possible to make an inference. So, there is no sufficient evidence to support this claim.



**Problem 1(iii):**

In this problem, we need to create a chart that showcases the Law of Large Numbers in action for the Beth Israel Medical Hospital. Law of large numbers is the number of trials becomes larger, the observed averages approach to the theoretical average. I had used the ggplot in R to plot the graph.

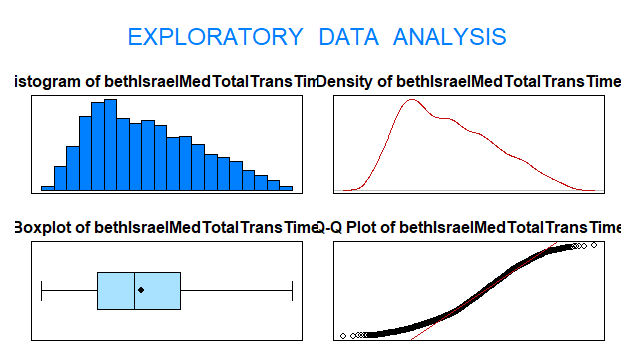
After analysis, I got a mean of 40.01171 which is evident on the plot. Hence, the Law of Large numbers has been proved as True.



From the above plot, we can depict that the increase in number of the trials shows an interest in approaching to the theoretical avg.

**Problem 1(iv):**

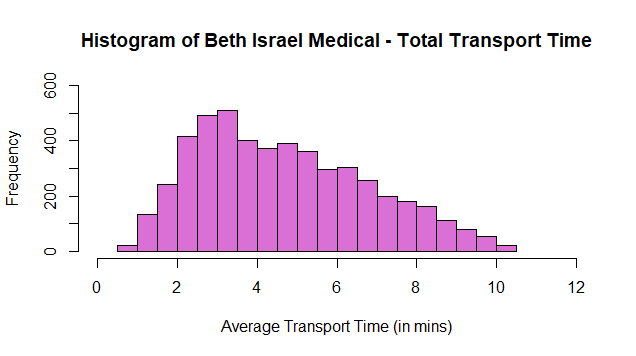
In this problem , I have found out the 95% CI for the time taken to the Beth Israel Medical Hospital as 4.726886 upper limit and 4.609180 as lower limit. I also see that this belongs to the Exponential Probability Distribution. After, this, lets conduct a chi-squared goodness of fit test to support our claim and we do not reject the Ho which means the data belongs to this distribution.



From the above Exploratory Data Analysis plot, for the Beth Israel Medical hospital, we can calculate the CI. I also used the MASS package to determine the best fit probability distribution using fitdistr(). After analyzing I got to see the Gamma Probability distribution is the best fit and checked the Chi-square Goodness of fit test.

**Problem 1(v):**

In this problem, let’s perform the Exploratory Data Analysis using the Histogram and summarize the results. From the plot, we can depict that the greater number of simulates happened in between 3.6 to 5.2. This also looks like an exponential probability distribution plot.



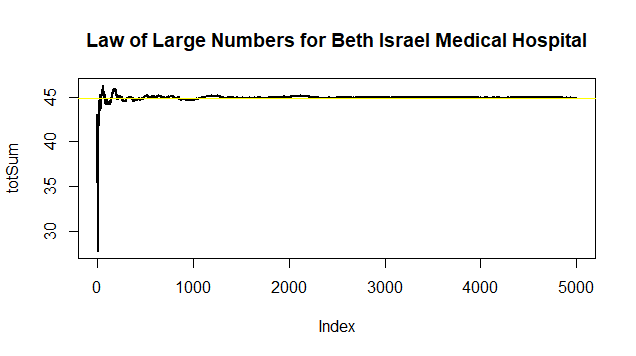
**Problem 2(i):**

In this problem, I have to work on generating 5000 random numbers in R by using the runif() function. Here, we took mean of 150, and standard deviation of 50 for a normally distributed probability function. After this, I had averaged the victims that can come to the hospital. After calculating the average for the Beth Israel Medical Hospital, I got to found as 44.90. The lowest mean is Brigham and Women’s Medical Hospital with 14.967. The highest is with the Beth Hospital.

**Problem 2(ii):**

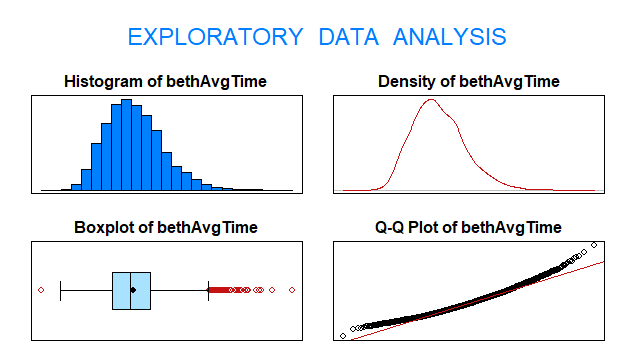
In this problem, I have calculated the average time of victims to the hospitals in mins. We can see that the value of Tufts Medical Hospital mean is around 3.755 and the highest is Boston Medical with 9.444. The Beth Israel Medical Hospital is having a mean of 5.251. Hence, there is no supporting evidence to claim this.

**Problem 2(iii):**



From the above chart, we can depict that the increase in the trials shows a tendency to the avg theoretical values.

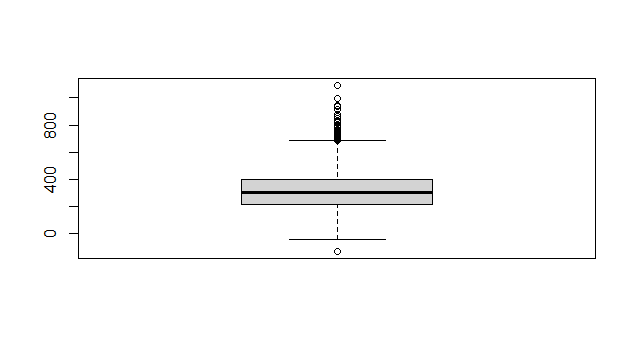
**Problem 2(iv):**



In this problem , I have found out the 95% CI for the time taken to the Beth Israel Medical Hospital as 319.0555 upper limit and 311.1409 as lower limit. I also see that this belongs to the Normal Probability Distribution.

**Problem 2(v):**

In this problem, the mean is 12.72 and the total time in mins for all the victims is in Exponential Probability Distribution manner.



From the Problem 1 and Problem 2, I can see that there’s a significant difference in the probability distribution functions for the best fit. Also, we can see the changes in the frequency distributions plotted using Histograms.

**Conclusion**

From this project I have gained a lot of knowledge and got familiar with the histograms, relative frequencies, Chi-Square Goodness Fit Test and determined the relationships and correlations between them. Also, using the random number generation helped to predict the results with no data. We utilized the triangular probability distribution to work on the random numbers. I previously did not work on these distributions and tests, but from this Assignment I had a great time in learning them. In coming days and assignments, I will make sure to use this knowledge and apply on the problems.

**References**

[1] Chi-squared Distribution was retrieved from http://www.r-tutor.com/elementary-statistics/probability-distributions/chi-squared-distribution

[2] Datacamp, Triangular: The Triangular Distribution was retrieved form https://www.rdocumentation.org/packages/EnvStats/versions/2.4.0/topics/Triangular