**Module 2: Week 2 Introduction to Enterprise Analytics**

**Author’s Name:** Shivani Sharma

**Course Number:** ALY6050

**Course Title:** Intro to Enterprise Analytics

**Academic Term:** Fall 2019 CPS Analytics

**Instructor’s Name:** Dr Alex Huang

**Assignment Completion Date:** 01-22-2020

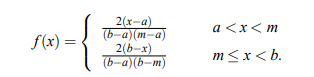


**Introduction**

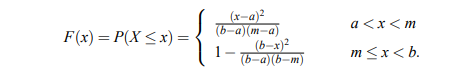
This assignment is based on the concept of simulation which helps in decision making in conditions like emergency and crisis conditions of accidents, draughts etc. We have been provided with the problem statement of a emergency facilities in like how effectively the local hospitals can handle the situation in case of any natural calamity like storm or fire .The victims are university students and the target location is provided as nearest 5 hospitals.

We have to determine the approximate number of victims in each hospital during that disaster situation and time required to drop off the victims to hospital. We have no data set provided so we need to use the concept of triangular probability distribution.

This distribution is used as an approximate model when no data values are provided. Triangular distribution is used to indicate the random variable having triangular distribution with 3 random parameters. And it is denoted by following function:



And the cumulative distribution function for X is as follows:



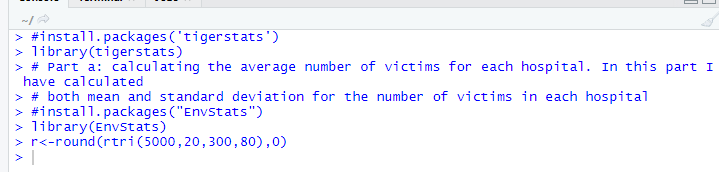
We have been provided with 2 problem statements and both are having different data which I would like to explain in analysis section.

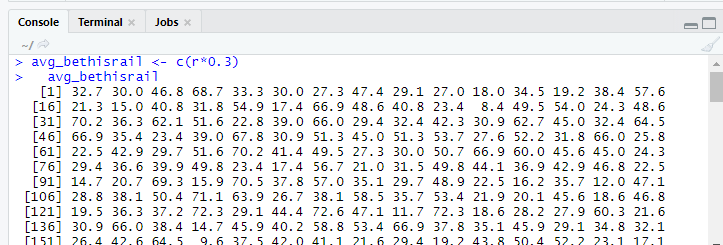
We have been provided with the assumption that each hospital is having 2 vehicles and at a time if one leaves the campus then another one leaves the hospital so total transport time will be the addition of transporting each victim to the hospital.

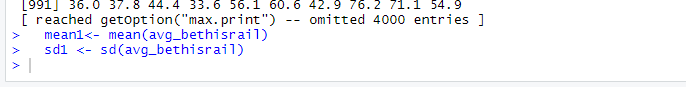
**Analysis**

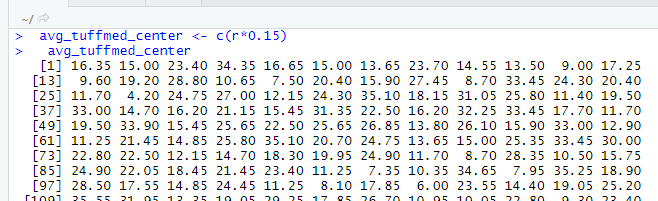
**Problem 1:** In this problem we have been provided with the minimum, maximum and mode values as 20,300 and 80 respectively. We have been also provided with the allocation of disaster victims for all the hospitals, using which we can calculate the average number of victims. We have been provided with the assumption that the transport time of each hospital is exponentially distributed with a average that depends on hospital.

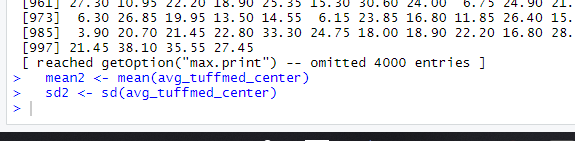
a.) Now first of all we have to calculate the expected average number of victims in each hospital. Here is the logic for the same with output.

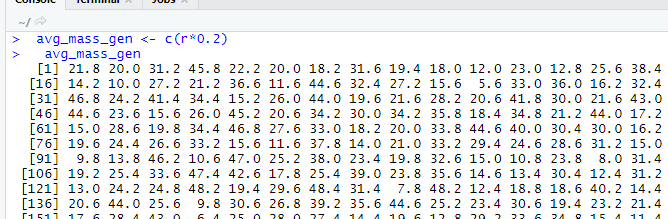


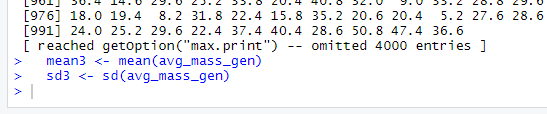


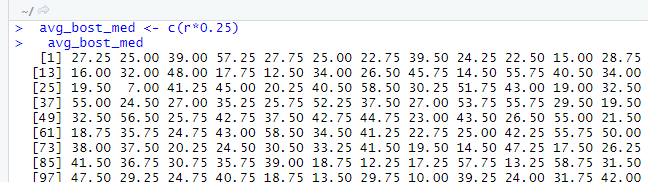


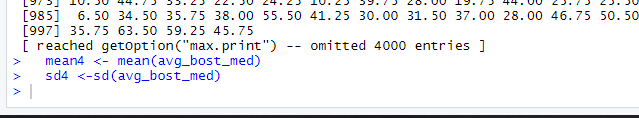


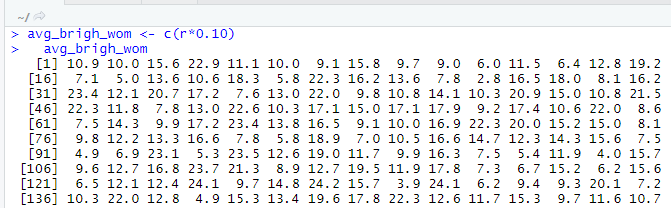


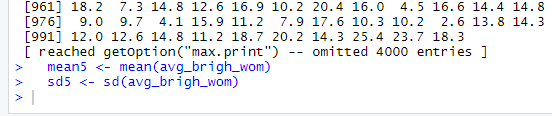


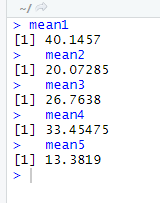






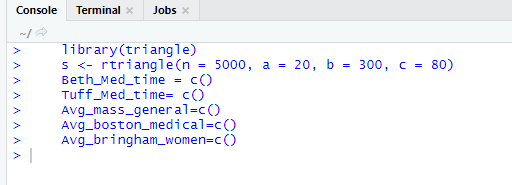


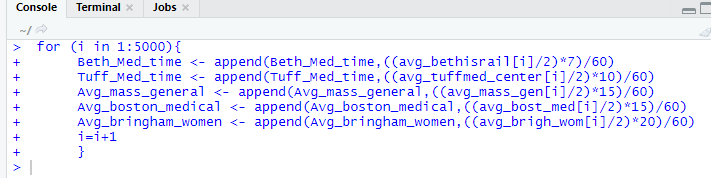


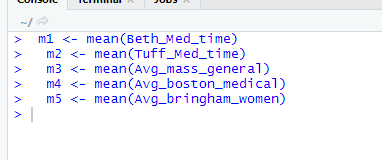


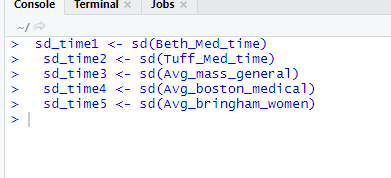
From the above values we can observe the average number of expected victims for all the hospitals.

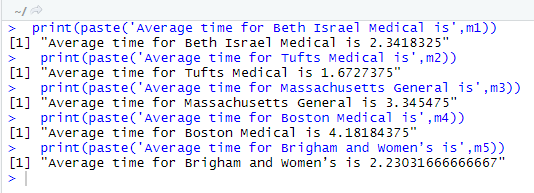
b.)Moving further we have to find out the average total time (in hours) required to transport all the victims for each hospital.





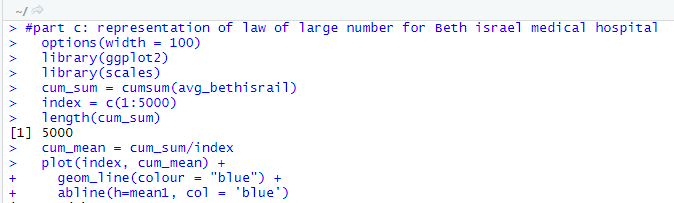




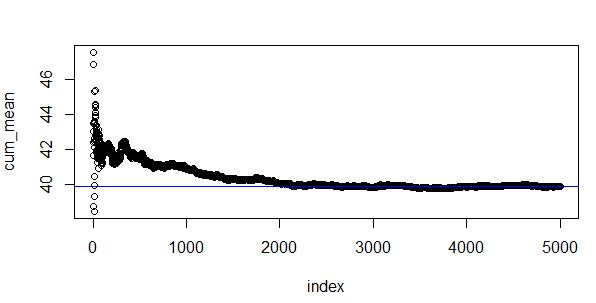


From the above results we can see that the average time for transporting victims from campus to hospitals respectively.

c.)From the average number of victims we have to plot a chart displaying “Law of Large numbers” for “Beth Israel Medical School” and I have used following logic for the same:

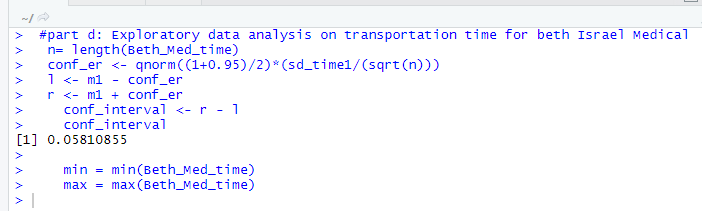


Here is the shape of curve:



From the above curve we can observe that the law of large numbers curve is showing a behavior that as the number of victims are increasing the relationship is becoming constant Initially it was at a peak value then it became constant.

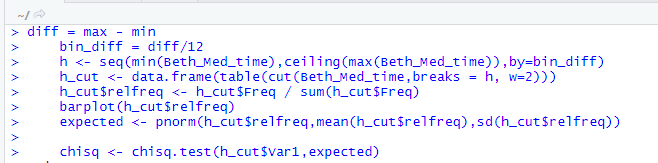
d.)Here we need to do a exploratory data analysis of the total transport time and the confidence interval is 95% for the total transport time and here is my logic for the same:

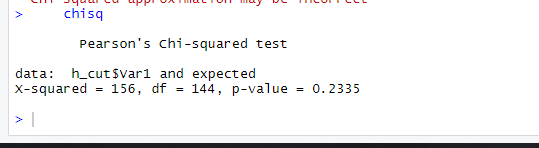


The confidence interval is coming as 0.058 which is coming as correct as per the significance level.

In order to do hypothesis test i.e the theoritical values and expected values are same I have used chi- squared test . Here H0 : is the hypothesis that the observed and expected values are same.

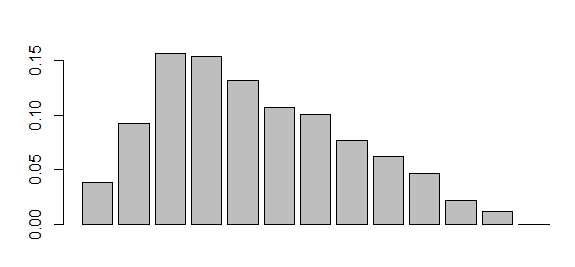
Ha: is the alternative hypothesis that the observed and expected values are not same.





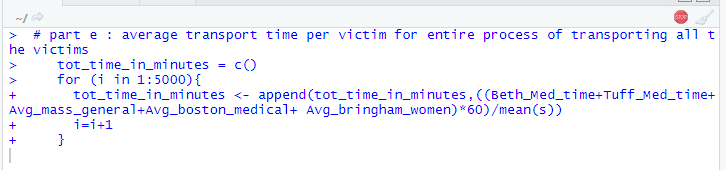
From the above results we can conclude that the null hypothesis cannot be rejected as p value is more than the significance level. i.e the observed values and expected values are same.

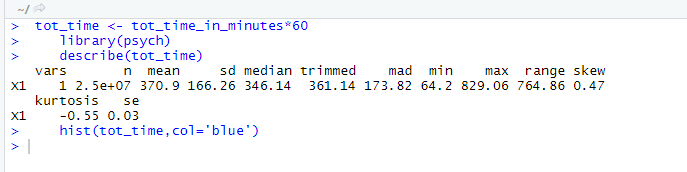
The distribution is also coming as follows:

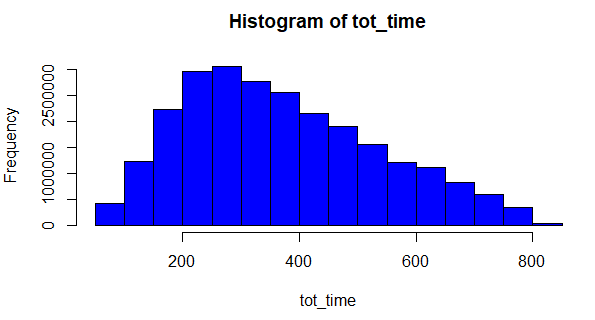


Which is showing a slight skewness in the shape towards left.

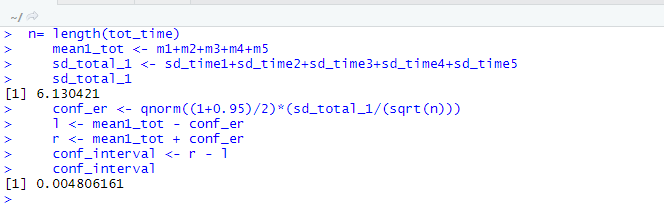
e.) Finally we need to calculate the average transport time per victim for the complete process of transportation of the victims. In order to do that I am using the transport time calculated in the part b of this question. And for performing the exploratory analysis for the transport time I am using the “psych” library and finding the analysis using the describe function.

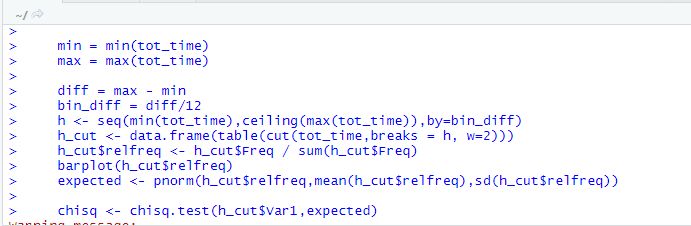


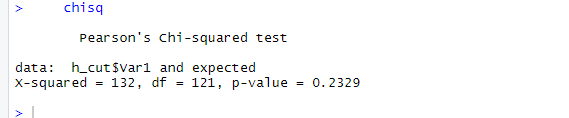




From the above results we can observe that there is slight skewness in the results towards left.

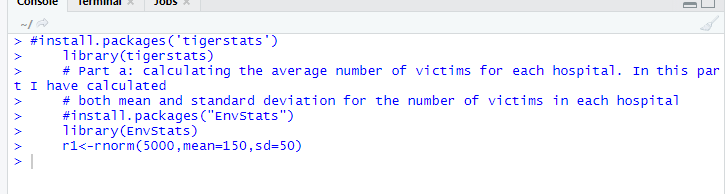


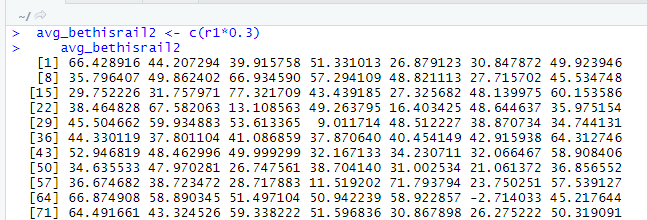


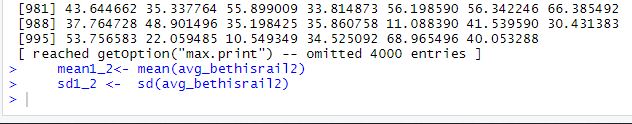


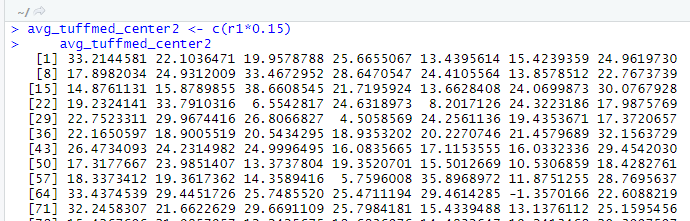
From the above results of the exploratory data analysis we can observe that the p-value is coming as 0.2329 which is more than the 0.05. So, we are accepting the null hypothesis i.e the actual and expected values are same. Also, there is slight skewness in the shape of the histogram towards left.

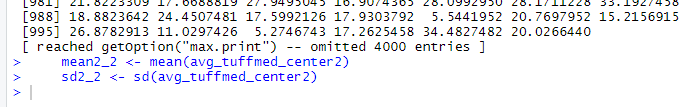
**Problem 2:** In this problem we have been provided with the mean value as 150 and standard deviation as 50. Also, we have been provided that the transport time is normally distributed with the different mean . Also, we have to perform the transport time calculations in minutes. Standard deviations are also provided to us. Rest of the logic and code implementation is same as problem statement 1.

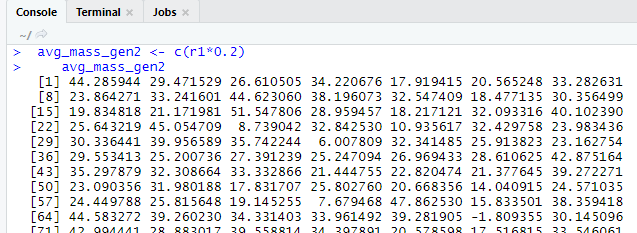


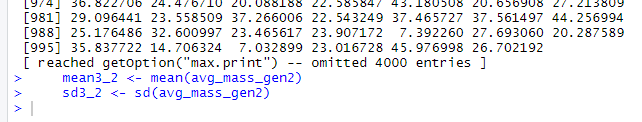


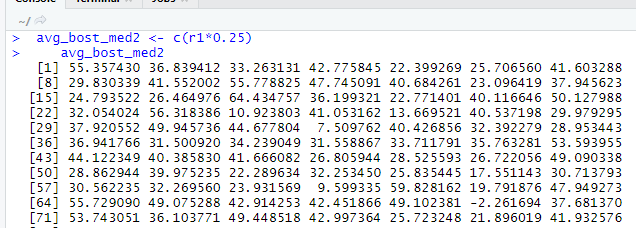


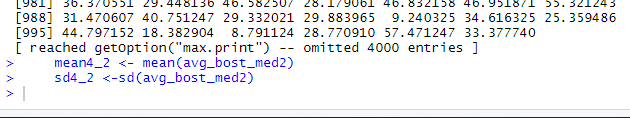


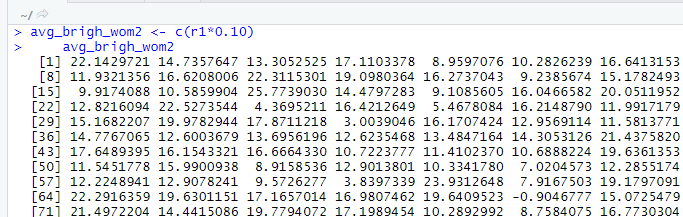


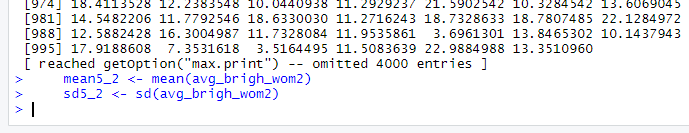


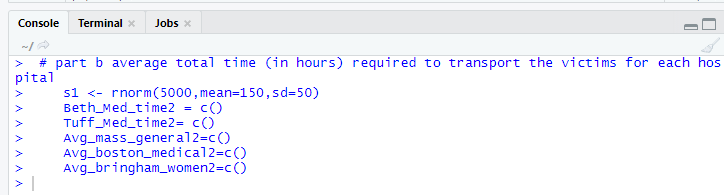


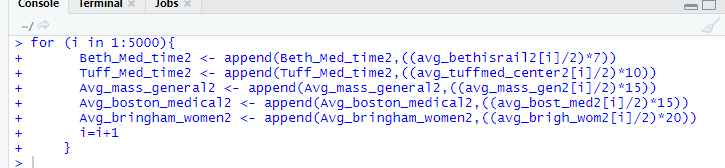


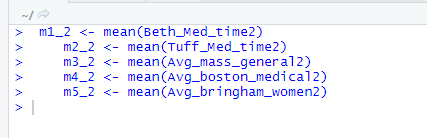


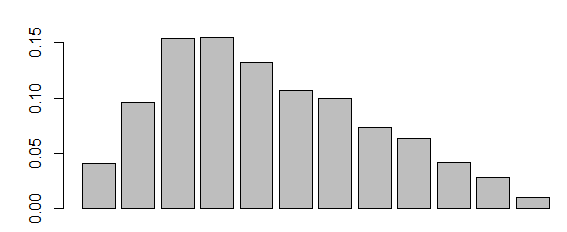




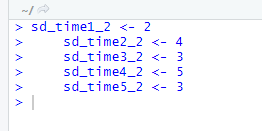


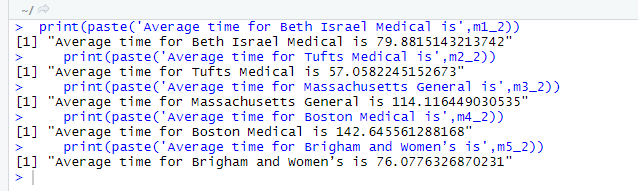


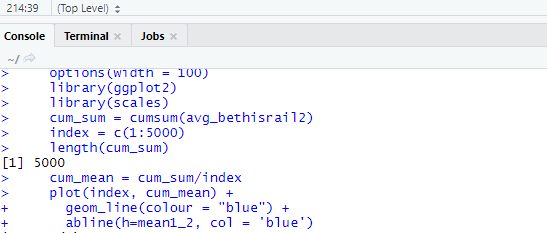


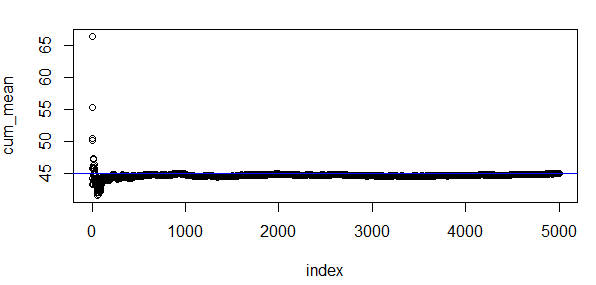


From the shape of the above histogram we can observe that there also exist left skewness in the behavioral pattern of the sample values.

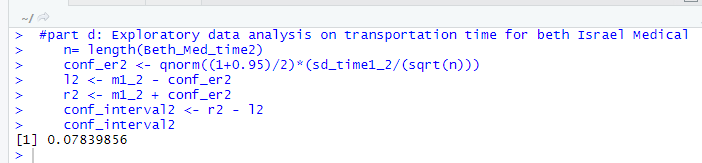




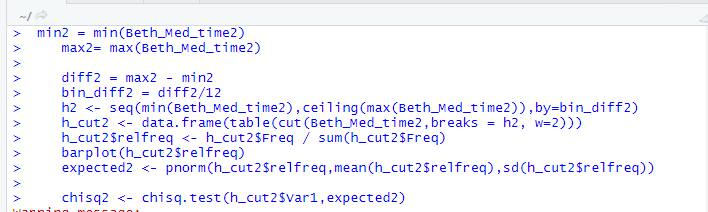


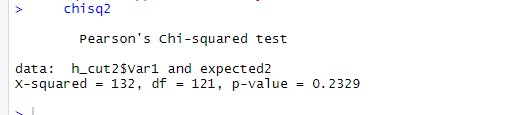


From the above plot we can conclude that initially the law of large number was showing less values then it became constant as the number of victims increased.

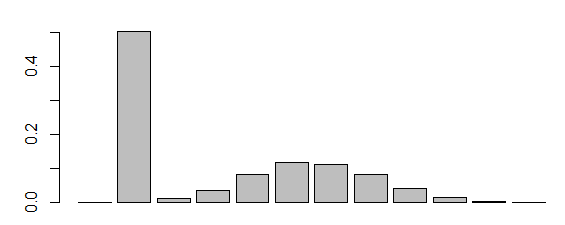


The confidence interval value is coming as 0.078 which is near to 0.05 i.e significance value.

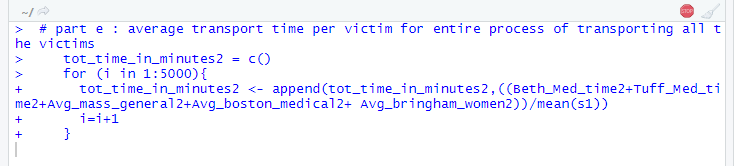


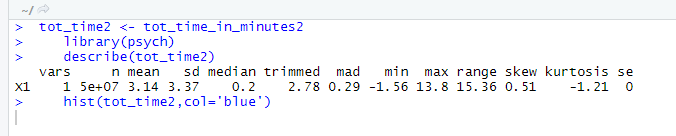


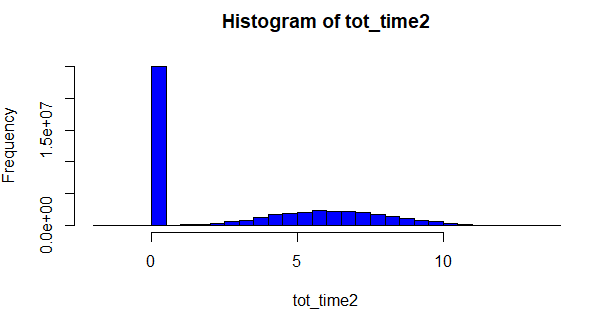
The p value of the chi-squared test is coming as 0.23 which is more than the significance level value i.e 0.05.



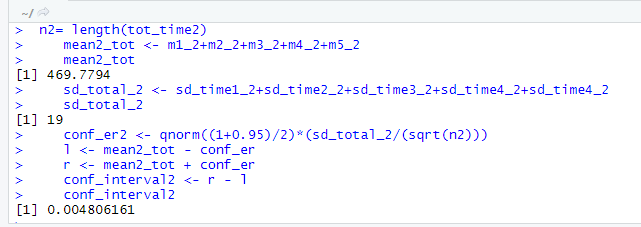
The behavioral pattern for the histogram after chi squared test is showing a slight left skewness.



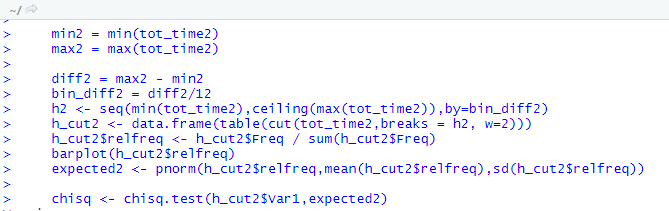


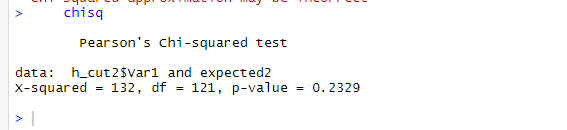


We can observe that there is only single value which is showing the maximum peak value at 0 and the rest of the pattern is showing approximately uniform distribution.

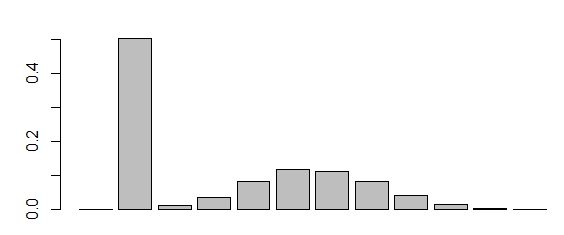


The confidence interval is showing a value as 0.005 which is same as significance value for total time.





The p-value is coming as 0.2329 which is same as obtained in problem 1 so here also we are accepting the null hypothesis.



The graph is showing a slight skewness towards left.

**Conclusion**

In the end I would like to conclude that I have used triangular distribution for creating the data set values in problem 1 and I have used “rnorm” in problem 2 and reason behind this is we have been provided that there exists a normal distribution in problem 2. Also, we can conclude that there is slight difference in the behavior of mean values, standard deviation and total timing attributes however the p-value is coming as same for both the problems i.e 0.23. From the simulation results obtained for both the problems we can conclude that this information can be used for calculating the time and effort required for estimating the values of resources required like manpower for driving the vehicles, how many drivers are required for driving the cabs, how to decide the route for driving cab. In order to get the additional useful information from the simulation we can change the maximum data range value which can be collected from any historical event. Also, we can apply the different simulation techniques if required.

# References

1. “Stats.” *Function | R Documentation*, [www.rdocumentation.org/packages/stats/versions/3.6.2/topics/sd](http://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/sd).
2. “Axis Scale in R Software : Minimum, Maximum and Log Scale.” *STHDA*, [www.sthda.com/english/wiki/axis-scale-in-r-software-minimum-maximum-and-log-scale](http://www.sthda.com/english/wiki/axis-scale-in-r-software-minimum-maximum-and-log-scale).
3. “Sign In.” *RPubs*, rpubs.com/shyambv/sampling\_distributions.
4. Irizarry, Rafael A. “Introduction to Data Science.” *Chapter 2 R Basics*, 18 Jan. 2020, rafalab.github.io/dsbook/r-basics.html.
5. Ryu, Choonghyun. *Exploratory Data Analysis*, 9 Jan. 2020, cran.r-project.org/web/packages/dlookr/vignettes/EDA.html.
6. University Health and Counselling Services. Northeastern University. Emergencies. Retrieved from: https://www.northeastern.edu/uhcs/medical-services/emergencies/ 2.

7. Taylor III, B.W. (2011) Introduction to Management Sciences. Prentice Hall. Upper Saddle River, NJ