Mustafa Abdullayev

Emergency Facilities Readiness Project

AY6050 – Intro to Enterprise Analytics

Term: Winter 2020

Instructor: Roy Wada

03/03/2020

# Introduction

This is Microsoft Word Report accompanying R Script. In my Script, my main aim was to perform simulation analysis. I was asked to use simulation analysis in order measure and observe readiness and competency of a local emergency facilities. As a data source, I utilized the Random values (Triangular Distribution and Normal Distribution) on R and performed statistical analysis using R. I generated the different random values and used Chi-Square Goodness of Fit Test to verify whether the generated values belong to a particular probability distribution. Also, I utilized powerful R built-in functions and graphs, such as bar charts, to dive deeper to observe hidden patterns and visually communicate my findings to the audience. Since I also provided R script with all the codes and comments, I removed some of the codes and comments from my report (such as package loading). Also, I did not incorporate all of my R code to report since it was too lengthy. However, all necessary outputs (Test results and graphs) are added. It is due to keep my report brief, succinct and to the point.

# Problem 1 – Simulation with triangular distribution

In first part of my analysis, I assumed that number of victims is best approximated by a triangular distribution. I assumed that minimum, peak and maximum values 20, 85 and 350 ,respectively. Moreover, transport time for hospitals has exponential distribution with different rate for each hospital.

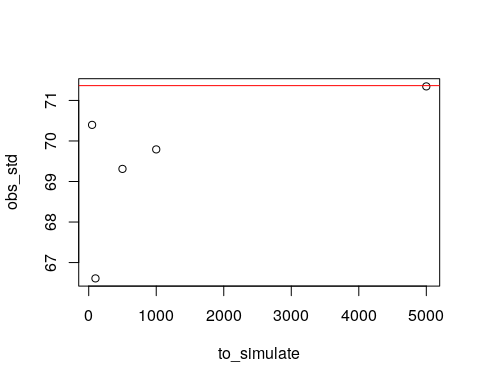
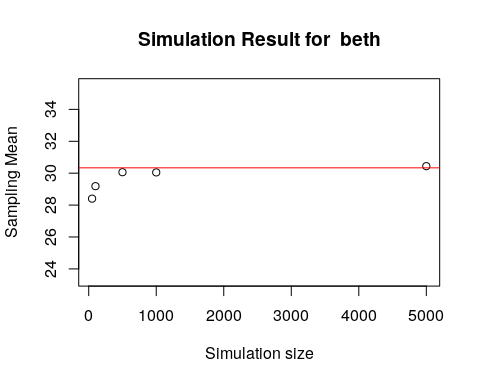
## Part A

In the first part, I used simulation analysis to find the expected number of victims and the expected variance in each hospital. I utilized various number of simulations (50, 100, 500, 1000, 5000) in order to see the effect of simulation number on sampling means and variance. In order to observe patterns easily, I utilized graphs with sampling means and variances for each number of visualizations. Also, I added theoretical expected mean and variance values on each graph (red line). From graphs below (Figure .1 – Figure 1.5) we can see that, as the number of simulations increase , both of sampling mean and the sampling variance are getting closer to theoretical mean and variance. (Exact values for sampling mean and variance are in the output part). One interesting aspect is that, we can see that, there is no dramatic increase in the accuracy of sampling mean after simulation value = 500. So, we can use elbow method to conclude that, 500 is the optimal number for simulations (since increased number of simulation’s cost is higher. )

#################### Problem 1a ####################

# Check R script for exact function  
  
find\_sim(1)

Figure .1 – Simulated values (mean and variance) vs Theoretical values for Beth Israel Hospital

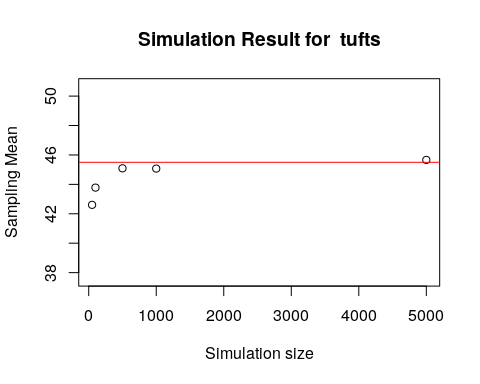


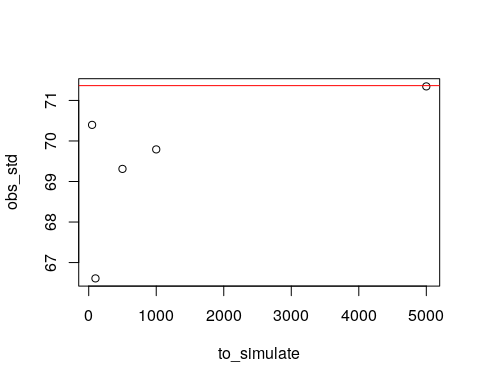
## [1] "Expected vs Observed for means"  
## [1] 30.33333  
## [1] 29.62776  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

######################################################

find\_sim(2)

Figure .2 – Simulated values (mean and variance) vs Theoretical values for Tufts Medical

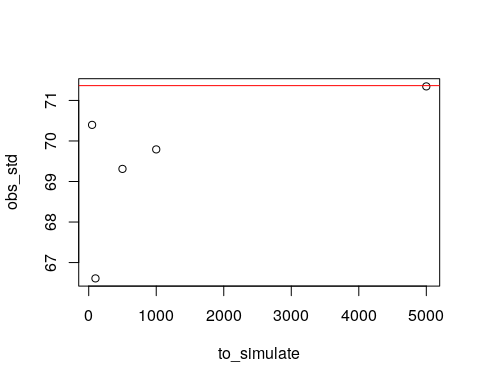
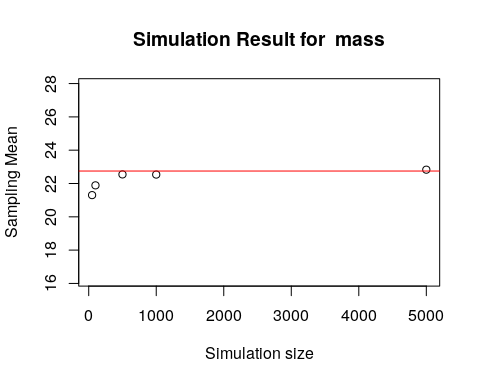




## [1] "Expected vs Observed foe means"  
## [1] 45.5  
## [1] 44.44164  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

#####################################################

find\_sim(3)

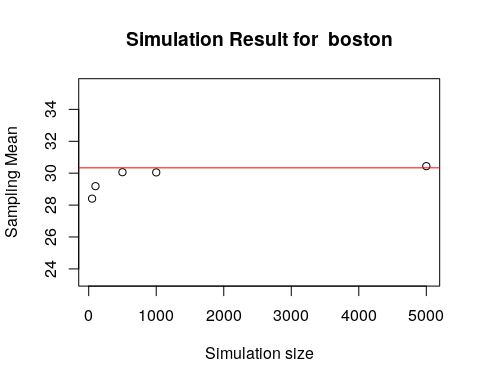
Figure .3 – Simulated values (mean and variance) vs Theoretical values for Massachusetts General

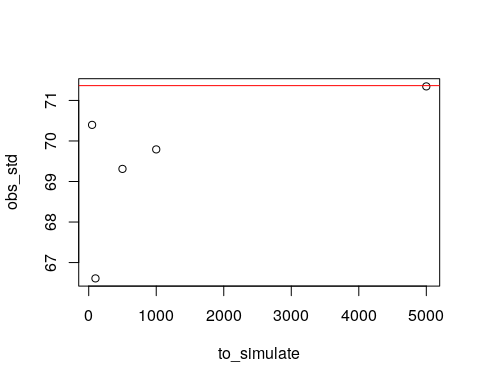
## [1] "Expected vs Observed foe means"  
## [1] 22.75  
## [1] 22.22082  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

##################################################

find\_sim(4)

Figure .4 – Simulated values (mean and variance) vs Theoretical values for Boston Medical



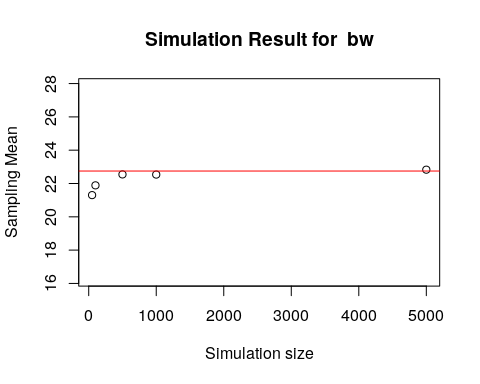


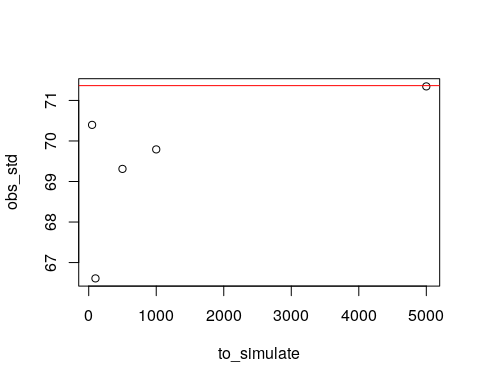
## [1] "Expected vs Observed foe means"  
## [1] 30.33333  
## [1] 29.62776  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

################################################

find\_sim(5)

Figure .5 – Simulated values (mean and variance) vs Theoretical values for Brigham and Women’s





## [1] "Expected vs Observed foe means"  
## [1] 22.75  
## [1] 22.22082  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

## Part B

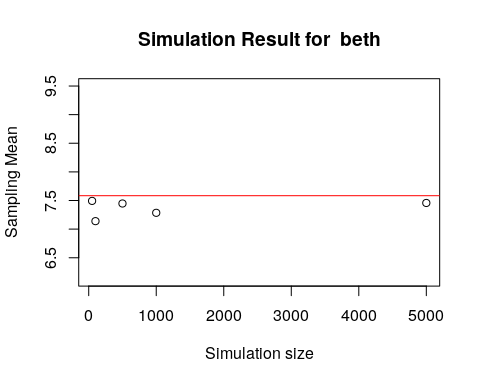
In this part of my analysis, I tried to predict average total time (in hours) that needed to carry all the victims to hospitals. Theoretical transportation time for each victim for Beth Israel, Tufts, Mass General, Boston Medical and Brigham Hospitals were exponential random variables with the rate of 15, 8, 25, 10, 12 minutes respectively. Again, I utilized various number of simulations (50, 100, 500, 1000, 5000) in order to see the effect of simulation number on sampling means. In order to observe patterns easily, I utilized graphs with sampling means for each number of visualizations. Also, I added theoretical expected mean and variance values for average time needed to transport all victims on each graph (red line). From graphs below (Figure 2.1 – Figure 2.5) we can see that, as the number of simulations increase , sampling mean for the average total time is getting closer to theoretical mean . (Exact values for sampling mean and variance are in the output part). Moreover, checking exact values for sampling variance demonstrate that, they get extremely close to theoretical value after 5000 simulations.

############################## Problem 1b ######################

# Check R script for function details

(find\_sim\_time(1))

Figure2.1 – Simulated average total transportation time vs Theoretical time for Beth Israel Hospital

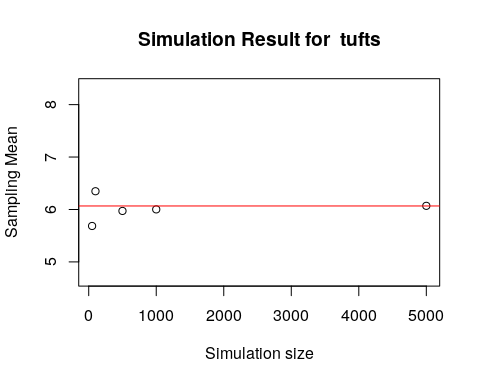


## [1] "Expected vs sampling for means"  
## [1] 7.583333  
## [1] 7.364397  
## [1] "Expected vs observed for variance"  
## [1] 15  
## [1] 14.47315

#####################################################

find\_sim\_time(2)

Figure2.2 – Simulated average total transportation time vs Theoretical time for Tufts Medical

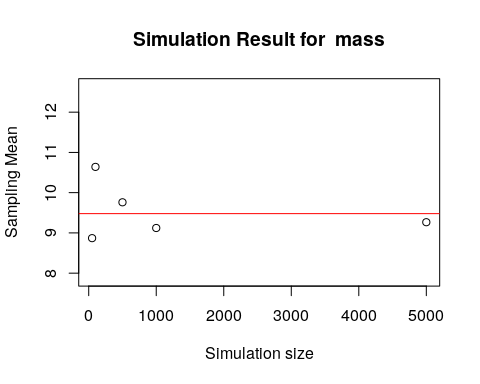


## [1] "Expected vs sampling for means"  
## [1] 6.066667  
## [1] 6.015172  
## [1] "Expected vs observed for variance"  
## [1] 8  
## [1] 7.899238

#########################################

find\_sim\_time(3)

Figure2.3 – Simulated average total transportation time vs Theoretical time for Massachusetts General

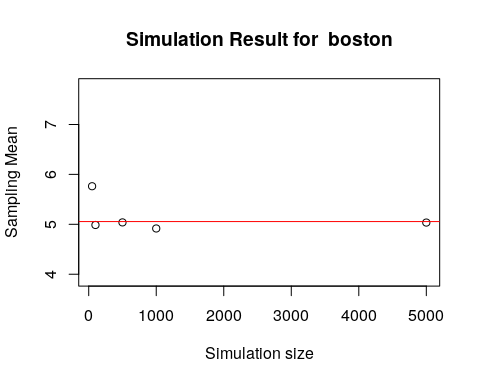


## [1] "Expected vs sampling for means"  
## [1] 9.479167  
## [1] 9.531747  
## [1] "Expected vs observed for variance"  
## [1] 25  
## [1] 23.71595

#############################################

find\_sim\_time(4)

Figure2.4– Simulated average total transportation time vs Theoretical time for Boston Medical

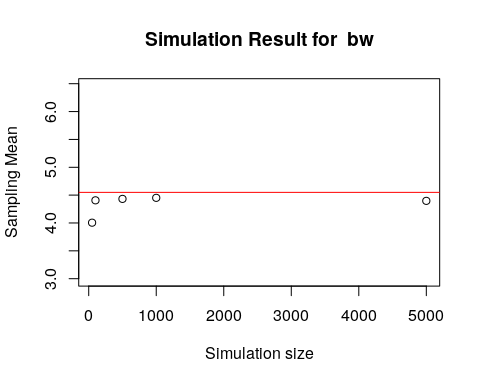


## [1] "Expected vs sampling for means"  
## [1] 5.055556  
## [1] 5.14776  
## [1] "Expected vs observed for variance"  
## [1] 10  
## [1] 9.650249

#######################################################

find\_sim\_time(5)

Figure2.5 – Simulated average total transportation time vs Theoretical time for Brigham and Women’s



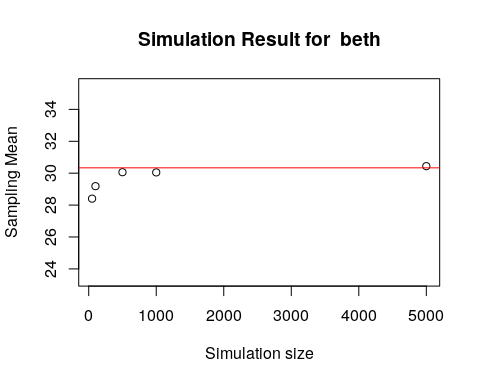
## [1] "Expected vs sampling for means"  
## [1] 4.55  
## [1] 4.338742  
## [1] "Expected vs observed for variance"  
## [1] 12  
## [1] 11.60258

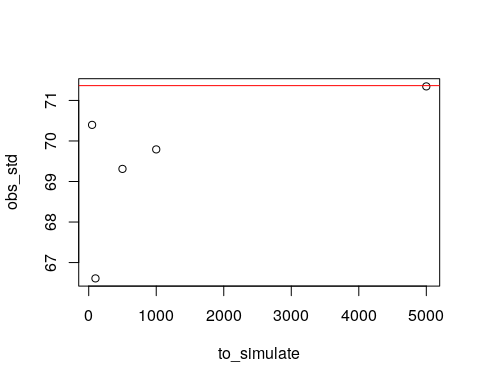
## Part C

In part c of my analysis, I was concerned with the Law of Large Numbers. It states that, as the number of trials becomes larger, the observed averages approach to the theoretical average. Again, I utilized various number of simulations (50, 100, 500, 1000, 5000) in order to see the effect of simulation number on sampling means and variance of the average number of victims for Beth Israel Hospital. In order to observe patterns easily, I utilized graphs with sampling means for each number of visualizations. Also, I added theoretical expected mean and variance values for average number of victims on graph (red line). From graph below (Figure 3.1) we can see that, as the number of trials increase , sampling mean and variance for the average number of victims for Beth Israel Hospital is getting closer to theoretical mean and variance . (Exact values for sampling mean and variance are in the output part). Also, as the number of trials gets larger, variance of sampling means getting smaller. So, I can say that, as the simulation number increase, our observed value and getting closer to theoretical value. Also, decreasing variance of sampling means shows that we are getting more confident about our observed value as the simulation size increases.

#################### problem 1c ####################  
  
find\_sim(1)

Figure3.1 –Observing Law of Large Numbers





## [1] "Expected vs Observed for means"  
## [1] 30.33333  
## [1] 29.62776  
## [1] "Expected vs observed for variance"  
## [1] 71.36565  
## [1] 70.39739 66.60887 69.31099 69.79211 71.34690

## Part D

In this part, I tried to do exploratory analysis of the total transportation time for the Beth Israel Hospital. First, I tried to construct a 95% Confidence Interval for total transport time. I used sampling means and variance that I obtained from different number of simulations and compared them. I started with 50 simulation and obtained interval of [5.925 , 8.7476]. Fast forward to 5000 simulations, we can see that, our interval got smaller with boundaries of [6.154255 , 8.850296]. Reason for that can be explained by the Law of Large Numbers. Since number of trails are getting bigger, we obtain better prediction for average total transportation time. Also, our variance for sampling means getting closer to zero. So, these two factors are resulted in better (smaller) confidence interval for the actual (theoretical) value. In the output part, one can observe exact confidence intervals for all number of simulations.

Secondly, from below graphs (Figure 4.1 – Figure 4.5) , We can see that as the number of simulations increase distribution of average total transportation time values is getting like triangular distribution. It is obvious in Figure 4.5. In order to dive deeper, I constructed Probability Plot for these values (Figure 4.6). Indeed, we can see that there is an almost perfect with along the line. This further enhance my opinion that distribution of these values is triangular. In order to formally prove that, I utilized Chi-Square Goodness of Fit Test with significance of 0.05. My hypotheses are as follow :

H0 : there is no significant difference between the observed and the expected value.

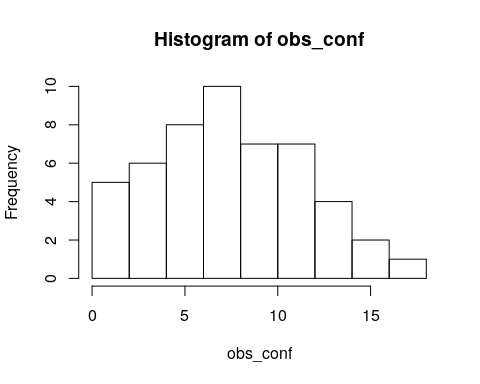
H1 : there is a significant difference between the observed and the expected value.

As a result of this test, I obtained p values to be 0.354178 (> 0.05) which means I am unable to reject my Null hypothesis. So, I do not have enough evidence in order to assume that distribution of average total transportation time is not triangular. Thus, I assume average total transportation time values for Beth Israel Hospital have triangular distribution.

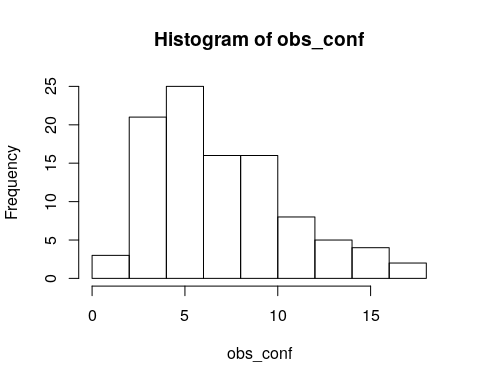
#################### problem 1d ####################  
  
for (val in to\_simulate){  
 print("Sampling means and CI - actual expected values is equal to 7.55 hours")  
 print(my\_sim\_confidence(val))  
}

## [1] "Sampling means and CI - actual expected values is equal to 7.55 hours"

Figure4.1 –Average transportation values distribution after 50 simulations

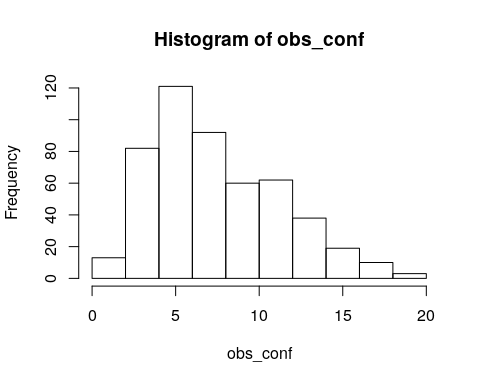


## [1] 7.336492  
## [1] 5.925403 8.747580  
  
## [1] "Sampling means and CI - actual expected values is equal to 7.55 hours"

*Figure4.2 –Average transportation values distribution after 100 simulations*

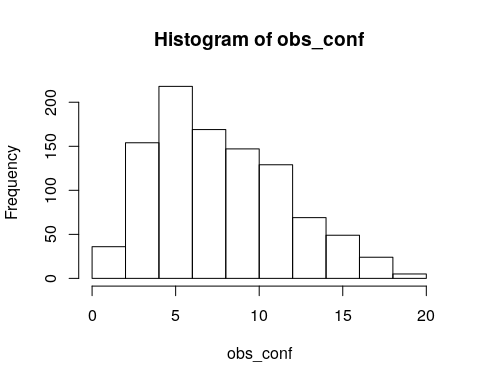
## [1] 6.891702  
## [1] 5.566976 8.216428  
  
## [1] "Sampling means and CI - actual expected values is equal to 7.55 hours"

*Figure4.3 –Average transportation values distribution after 500 simulations*

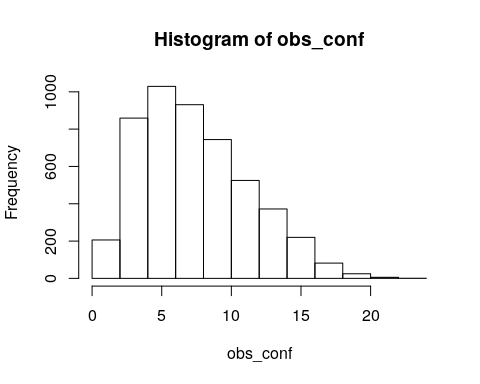


## [1] 7.494166  
## [1] 6.130865 8.857466  
  
## [499] 2.6894386 4.7469529  
## [1] "Sampling means and CI - actual expected values is equal to 7.55 hours"

*Figure4.4 –Average transportation values distribution after 1000 simulations*



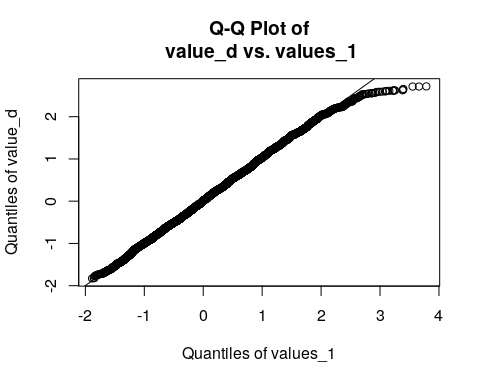
## [1] 7.637866  
## [1] 6.274700 9.001032  
##   
## [1] "Sampling means and CI - actual expected values is equal to 7.55 hours"

*Figure4.5 –Average transportation values distribution after 5000 simulations*

## [1] 7.424208  
## [1] 6.046329 8.802086

## [1] 7.502275  
## [1] 6.154255 8.850296

value\_d <- vector()  
r <- runif(length(values\_1))  
A <- lower + sqrt((upper-lower)\*(peak-lower)\*r)  
B <- upper - sqrt((upper-lower)\*(upper-peak)\*(1-r))  
C <- (peak - lower) / (upper - lower)  
  
x <- ifelse(r <= C, A, B)  
value\_d <- append(value\_d,x)  
  
value\_d <- (value\_d - mean(value\_d))/sd(value\_d)  
  
qqPlot(values\_1,value\_d,add.line = TRUE)

*Figure4.6 –Probability plot for average transportation values vs Triangular values*

my\_test\_1 <- chisq.test(values\_1,value\_d)

my\_test\_1$p.value

## [1] 0.354178

## Part E

In the last part of problem 1, I did similar analysis with Part D. But this time I did for average transportation time for all victims, not just for Beth Israel Hospital. With same procedure, I obtained 95% Confidence Interval for average time per victim for all hospitals to be [12.70630 ,13.24261] for 50 simulations. Fast forward to 5000 simulations, it becomes [12.56836, 12.99186]. Considering theoretical value is 12.95 minutes, we can see that we obtained way better estimation with 5000 simulations. Again, that is because of Law of Large Numbers.

This time, contrary to part d, we can see that transportation time per victim values has distribution of Normal (Figure 5.1 – Figure 5.5). It is obvious in the Figure 5.5. To dive deeper I constructed Probability plot for these values’ vs Normal Values (Figure 5.6). Indeed, there is an almost prefect with between them. In order to formally prove that, I utilized Chi-Square Goodness of Fit Test with significance of 0.05. My hypotheses are as follow :

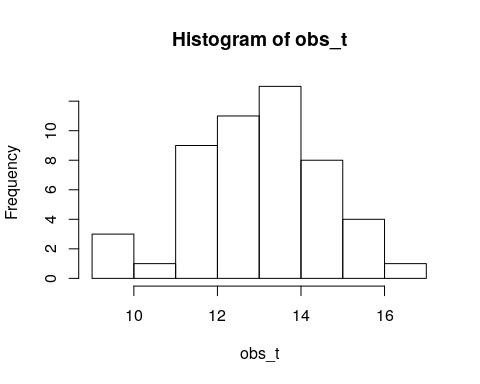
H0 : there is no significant difference between the observed and the expected value.

H1 : there is a significant difference between the observed and the expected value.

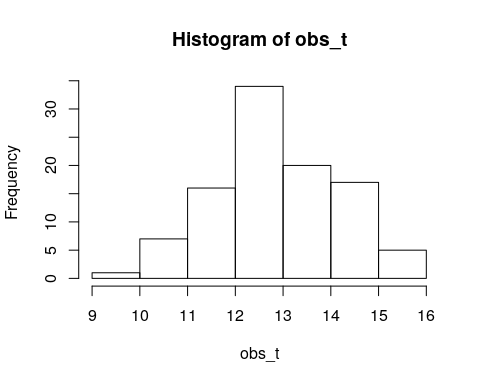
As a result of this test, I obtained p values to be 0.23985 (> 0.05) which means I am unable to reject my Null hypothesis. So, I do not have enough evidence in order to assume that distribution of transportation time per victim for all hospitals is not Normal. Thus, I assume transportation time per victim for all hospitals have Normal distribution.

#################### problem 1e ####################  
  
for (val in to\_simulate){  
 print("Expected time is 12.95 minutes")  
 print(my\_sim\_total(val))  
}

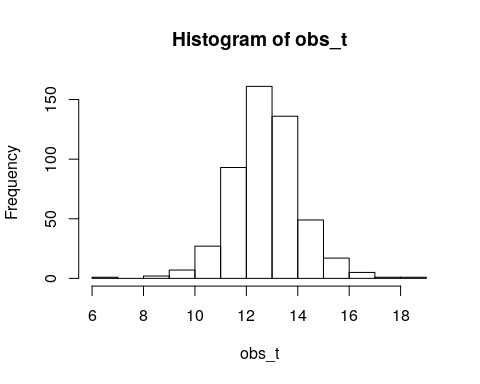
## [1] "Expected time is 12.95 minutes"  
## [1] 12.97446

*Figure5.1 –Transportation time values distribution after 50 simulations*

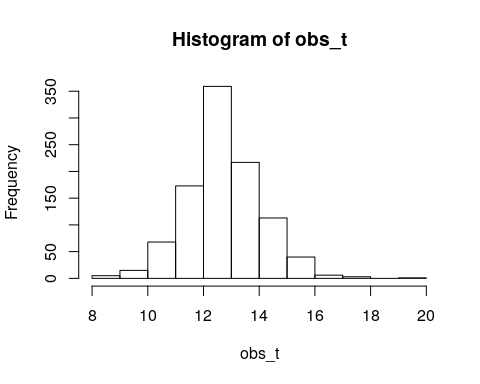
## [1] 12.70630 13.24261  
  
## [1] "Expected time is 12.95 minutes"  
## [1] 12.86827

*Figure5.2 –Transportation time values distribution after 100 simulations*

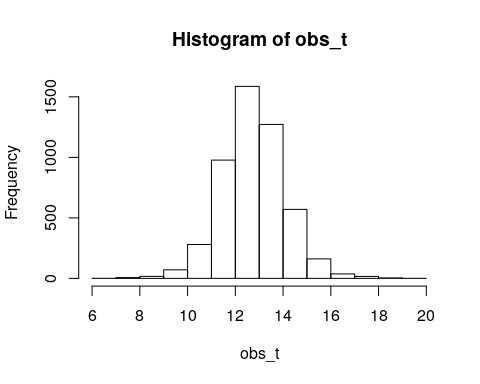
## [1] 12.66626 13.07028  
  
## [99] 11.761780 14.146583  
## [1] "Expected time is 12.95 minutes"  
## [1] 12.7696

*Figure5.3 –Transportation time values distribution after 500 simulations*

## [1] 12.55697 12.98224  
  
## [1] "Expected time is 12.95 minutes"  
## [1] 12.73617

*Figure5.4 –Transportation time values distribution after 1000 simulations*

## [1] 12.52592 12.94641  
## [1] "Expected time is 12.95 minutes"  
## [1] 12.76496

*Figure5.5 –Transportation time values distribution after 5000 simulations*

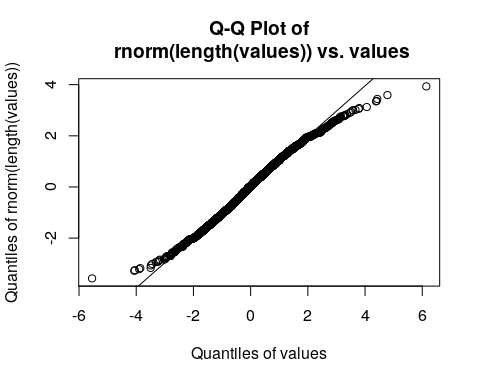
## [1] 12.55603 12.97388  
  
## [4999] 14.150887 13.603807

values <- my\_sim\_total(5000)

## [1] 12.78011

## [1] 12.56836 12.99186

values <- (values-mean(values))/sd(values)  
  
  
qqPlot(values,rnorm(length(values)),add.line = TRUE)

*Figure 5.6 –Probability plot for transportation time values per victim vs Normal values*

my\_test\_2 <- chisq.test(values,rnorm(length(values)))

my\_test\_2$p.value

## [1] 0.23985

# Problem 2 – Simulation with normal distribution

For the second problem, I did the exact same analysis with problem 1. But this time my assumptions were different. First of I assumed that, number of victims were normally distributed with mean 175 and standard deviation of 63. Also, transportation time for different hospitals were normally distributed. (Exact values in my code script or in assignment outline paper). Finally , I only utilized 5000 simulations for this part.

## Part A

With same procedure with Problem 1-part a, I obtained sampling mean and deviation values for average number of victims in each hospital after 5000 simulations. Here , hospitals are coded as (1,2,3,4,5) for Beth Hospital, Tufts Medical, Mass General, Boston Medical and Brigham and Women’s, respectively. (Exact values in the code part). As we can see, after 5000 simulations, our predictions are almost the same with the theoretical values

set.seed(123)  
  
#################### Problem 2a ####################  
  
avg = 175  
std = 63  
  
  
  
find\_sim\_2(1)

## [1] "Mean comparison"  
## [1] 35.12332  
## [1] 35  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

find\_sim\_2(2)

## [1] "Mean comparison"  
## [1] 52.68498  
## [1] 52.5  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

find\_sim\_2(3)

## [1] "Mean comparison"  
## [1] 26.34249  
## [1] 26.25  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

find\_sim\_2(4)

## [1] "Mean comparison"  
## [1] 35.12332  
## [1] 35  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

find\_sim\_2(5)

## [1] "Mean comparison"  
## [1] 26.34249  
## [1] 26.25  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

## Part B

With same procedure with Problem 1-part b, I obtained sampling mean and deviation values for average time for transportation of all victims for each hospital (in hours) after 5000 simulations. Here , hospitals are coded as (1,2,3,4,5) for Beth Hospital, Tufts Medical, Mass General, Boston Medical and Brigham and Women’s, respectively. (Exact values in the code part). As we can see, after 5000 simulations, our predictions are almost the same with the theoretical values

############################## Problem 2b ######################  
my\_sim\_time\_2(5000,1)

## [1] "Expected vs Observed means"  
## [1] 8.75  
## [1] 8.63782  
## [1] "Expected vs observed standard deviations"  
## [1] 105  
## [1] 104.8883

my\_sim\_time\_2(5000,2)

## [1] "Expected vs Observed means"  
## [1] 7  
## [1] 6.901302  
## [1] "Expected vs observed standard deviations"  
## [1] 105  
## [1] 104.2864

my\_sim\_time\_2(5000,3)

## [1] "Expected vs Observed means"  
## [1] 10.9375  
## [1] 10.77468  
## [1] "Expected vs observed standard deviations"  
## [1] 157.5  
## [1] 156.8402

my\_sim\_time\_2(5000,4)

## [1] "Expected vs Observed means"  
## [1] 5.833333  
## [1] 5.749415  
## [1] "Expected vs observed standard deviations"  
## [1] 122.5  
## [1] 122.206

my\_sim\_time\_2(5000,5)

## [1] "Expected vs Observed means"  
## [1] 5.25  
## [1] 5.107534  
## [1] "Expected vs observed standard deviations"  
## [1] 65.625  
## [1] 64.73311

## Part C

With same procedure with Problem 1-part c, I observed the Law of Large Numbers for Beth Israel Hospital after 5000 simulations. As we can see, after 5000 simulations, our predictions are almost the same with the theoretical values. That shows that, indeed, as number of simulations getting bigger, we obtain more and more accurate estimates. Also, standard deviation am0ng sampling means are getting smaller enabling us to be more accurately construct confidence intervals.

#################### problem 2c ####################  
  
find\_sim\_2(1)

## [1] "Mean comparison"  
## [1] 35.12332  
## [1] 35  
## [1] "Standard deviations"  
## [1] 63.77851  
## [1] 63

## Part D

Again, for this part, I tried to construct 95% confidence interval for total transport time for Beth Israel Hospital. After 5000 simulations, I found this interval to be [8.040944, 9.199970] in hours. Knowing that, theoretical value for total transport time is 8.75 hours, It looks we found perfect Confidence Interval with 95% Confidence. Moreover, from Figure 6.1, we can see that these values are distributed like Normal Values. In order to dive deeper, I plotted a probability plot for total transport time for Beth Israel Hospital vs Random values (Figure 6.2). Indeed, we can see that there is an almost perfect fit. In order to formally prove that, I utilized Chi-Square Goodness of Fit Test with significance of 0.05. My hypotheses are as follow :

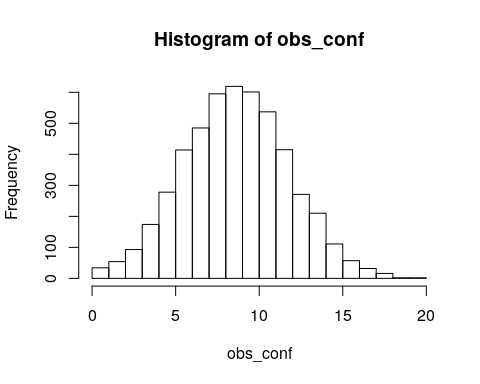
H0 : there is no significant difference between the observed and the expected value.

H1 : there is a significant difference between the observed and the expected value.

As a result of this test, I obtained p values to be 0.785463 (> 0.05) which means I am unable to reject my Null hypothesis. So, I do not have enough evidence in order to assume that distribution of total transportation time for Beth Israel Hospital is not Normal. Thus, I assume total transportation time values for Beth Israel Hospital, in hours, have Normal distribution.

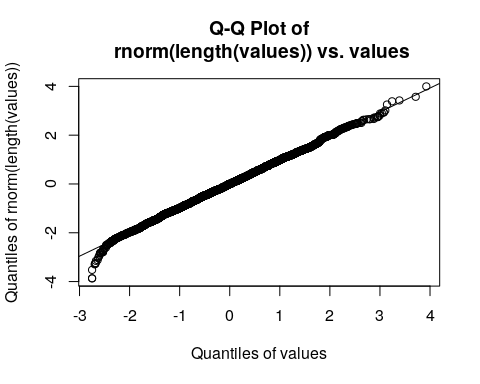
#################### problem 2d ####################  
  
my\_sim\_confidence\_2(5000)

Figure 6.1 –Total transportation time values for Beth Israel Hospital



## [1] 8.620457  
## [1] 8.040944 9.199970

qqPlot(values,rnorm(length(values)),add.line = TRUE)

*Figure 6.2 –Probability plot for total transportation time values for Beth Israel Hospital*

my\_test\_3 <- chisq.test(values,rnorm(length(values)))

my\_test\_3$p.value

## [1] 0.785463

## Part E

Finally, for Part e of problem 2, I did similar thing as part d. This time, I tried to find average transportation time per victim for all hospitals in minutes. After 5000 simulations, I found that value to be 12.7414. Also, I constructed 95% Confidence Interval for average time value which turns out to be [12.69995 , 12.78288]. Considering that actual theoretical value is 12.75 minutes, I think after 5000 simulations, I constructed near perfect 95% confidence interval and found near perfect sampling mean. Moreover, from Figure 7.1, we can see that average transportation time per victim values are distributed like Normal values. In order to analyze further, I constructed Probability lot for average transportation time vs Normal Values (Figure 7.2). Indeed ,we can see that there is an almost perfect fit. In order to formally prove that, I utilized Chi-Square Goodness of Fit Test with significance of 0.05. My hypotheses are as follow :

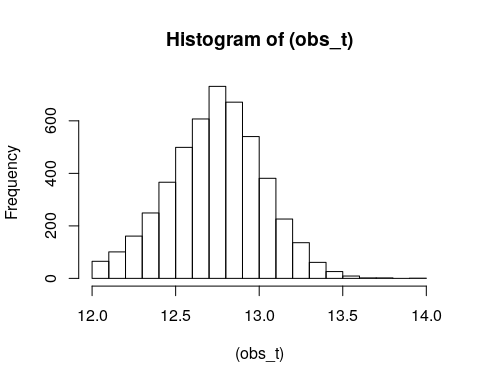
H0 : there is no significant difference between the observed and the expected value.

H1 : there is a significant difference between the observed and the expected value.

As a result of this test, I obtained p values to be 0.586321 (> 0.05) which means I am unable to reject my Null hypothesis. So, I do not have enough evidence in order to assume that distribution of transportation time per victim for all hospitals is not Normal. Thus, I assume transportation time per victim for all hospitals have Normal distribution.

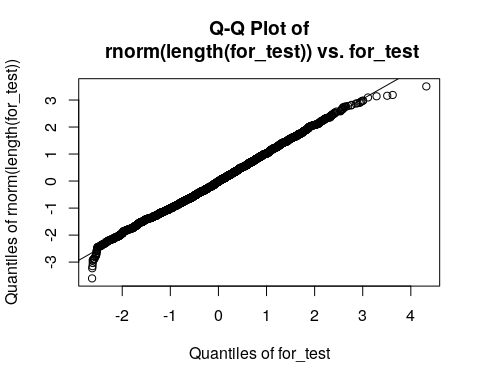
#################### problem 2e ####################  
  
for\_test <- my\_sim\_total\_2(5000)

## [1] 12.74142

*Figure 7.1 –Transportation time values distribution after 5000 simulations*

## [1] 12.69995 12.78288

for\_test<- (for\_test-mean(for\_test))/sd(for\_test)  
  
qqPlot(for\_test,rnorm(length(for\_test)),add.line = TRUE)

*Figure 7.2 –Probability plot for transportation time values per victim vs Normal values*

my\_test\_4 <- chisq.test(x=for\_test,y=rnorm(length(for\_test)))

my\_test\_4$p.value

## [1] 0.586321

# Conclusion

To conclude, I utilized the Random values on R and performed simulation analysis using R. In the first problem, I utilized triangularly distributed values. On the other hand, In the problem 2 I utilized Normally distributed values. Considering results, there was no huge quantitative and qualitative differences between these two simulation analyses. This can be attributed to Law of Large Numbers. Since my simulation analysis consisted up to 5000 simulations, I was able to generate correct r-experimental results in both analyses. (Since I had large number of trials, experimental results got very close to theoretical results for both distributions.). The only noticeable difference was in average number of victims that each hospital expected. Since I used differently distributed values with different means, this difference was understandable. Finally, this analysis is very helpful for the real world. We can use results from this simulation analysis in order to effectively distribute emergency tools like ambulance or medical stuff among different hospitals. Also, universities or other local agencies can utilize simulation analysis like this in order to prepare for the worst-case scenarios.

# References

Emergencies - University Health and Counseling Services. (n.d.). Retrieved from <https://www.northeastern.edu/uhcs/medical-services/emergencies/>