**Introduction to Enterprise Analytics**

# ALY6050 Module 2 Assignment

# Faculty: Prof. Richard He

# 2020 Winter CPS Quarter

# Image result for neu cps

Shraddha Gopalakrishnan [NUID: 001376444]

**Introduction**

Simulation technique involves a model that mimics or copies the operation of an already existing system or process. Decision making is another effective process of making choices by identifying a solution out of two or more courses of action. [2] The aim of this assignment is to acquaint us with the simulation process integrated with decision making. We must make use of simulation method to evaluate certain measures in the real world and analyze one solution by learning to use decision making. Also, simulation modeling is very frequently used to support decision making in different business areas. The purpose of this project is to understand such scenarios and provide solutions to a wide range of issues at strategic, operational and tactical level.

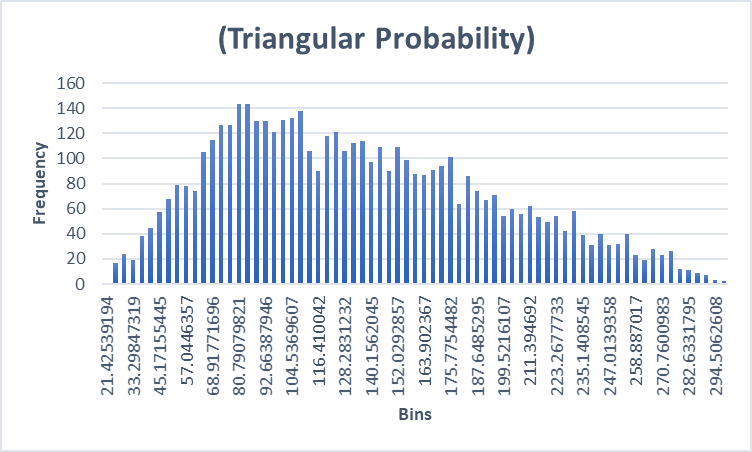
**Analysis**

The gist of the analysis in this assignment is dealing with how effectively the local emergency facilities i.e., local hospitals can handle natural disasters. Consider 5 hospitals Beth Israel Deaconess Medical Center, Tufts Medical Center, Massachusetts General Hospital, Boston Medical Center, Brigham and Women’s Hospital. Our deliverables are to find how many victims each hospital might receive, and time taken to transport victims from the disaster location to hospitals. Let us now move to the actual analysis part of the project.

**Part 1 (Q1, Q3)**

The initial work in part 1 is to generate 5000 random numbers using **RAND ()** function of Excel. Triangular probability is also known as a “a lack of knowledge distribution”, and since in our case there is lack of historical data for disasters, we make use of the triangular distribution while conducting the simulation. For triangular distribution, we need a lower value **A**, an upper value **B** and a peak i.e., mode **C** between those two values which are respectively given as 20,300 and 80.

Using the following formulae, we figure out our X2.

1. A= a+sqrt ((b-a) \*(c-a) \*r2)
2. B= b-sqrt ((b-a) \*(b-c) \*(1-r2))
3. C= (c-a)/(b-a)
4. X2 = ifelse (r2<C, A,B)

Where r2 is the standard uniform random value and X2 column denotes the random variables for further use.

The whole X2 column is selected to show the Distribution of the Simulation. The frequency histogram is triangular in nature indicating triangular probability.

With the histogram we even obtain the bins and frequency values for X2.

**a.]** For the average value of victims at each hospital, we multiply the percentage of every hospital given in the table with the frequency obtained for X2 and using **AVERAGE()** formula estimate the following numbers-



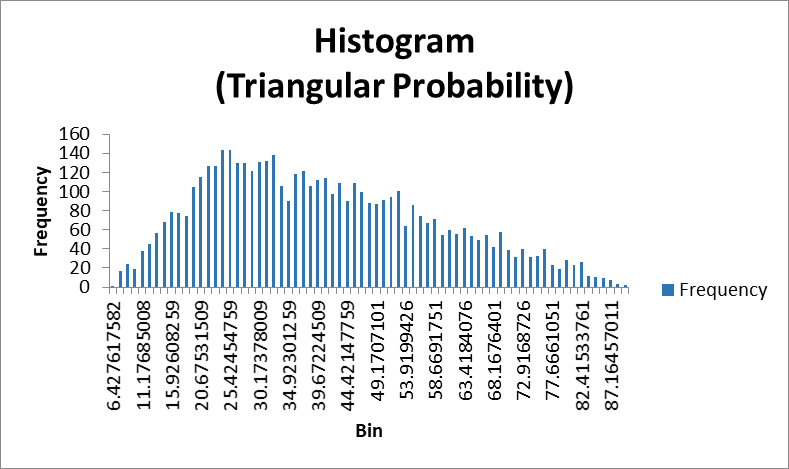


We consider the approximate number of victims as 21,10,14,17 and 7 for every hospital.

**b.]** To find the average total time required to transport all victims in hours we use the **CONVERT()** function in Excel converting time given in minutes to hours.

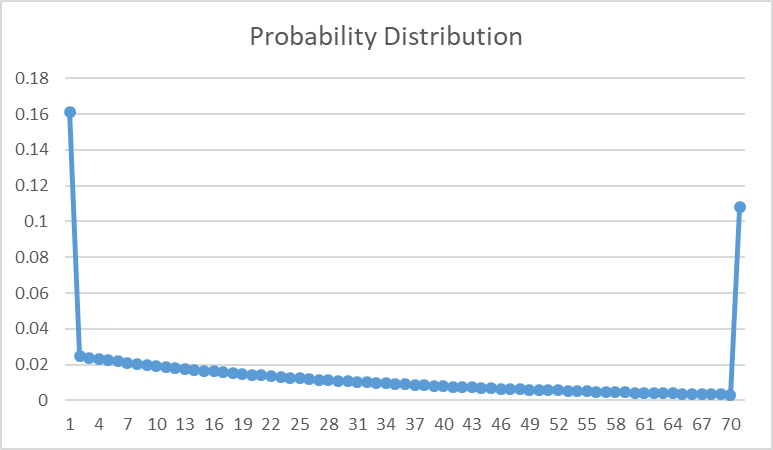


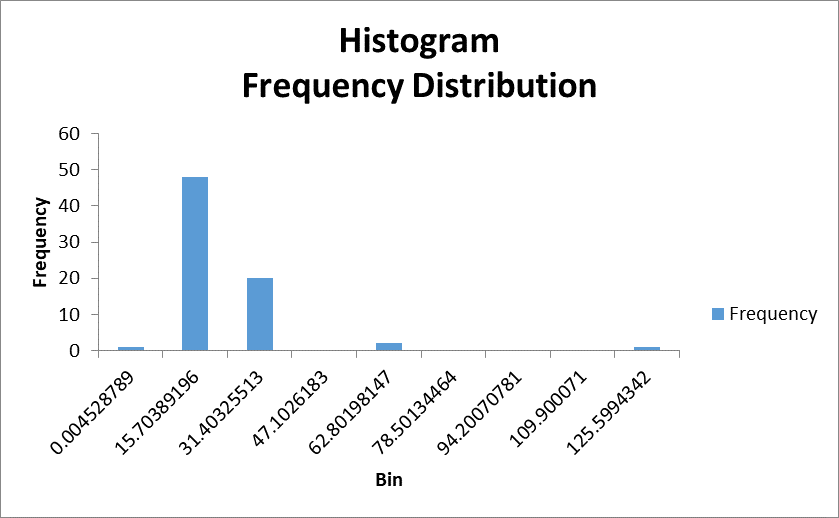


**c.]** For Beth Israel Medical, X2 is multiplied with 30% victims as mentioned. Next according to the number of bins, average values are evaluated for observed and expected frequencies of Beth Israel Medical hospital. Here we see that as the number of trials becomes larger i.e. 100 to 2000, the observed averages (38.70393, 38.98692…, 40.24686) lean towards the theoretical average (40). Thus, we successfully implemented the law of large numbers. A chart for the same theory is shown above.

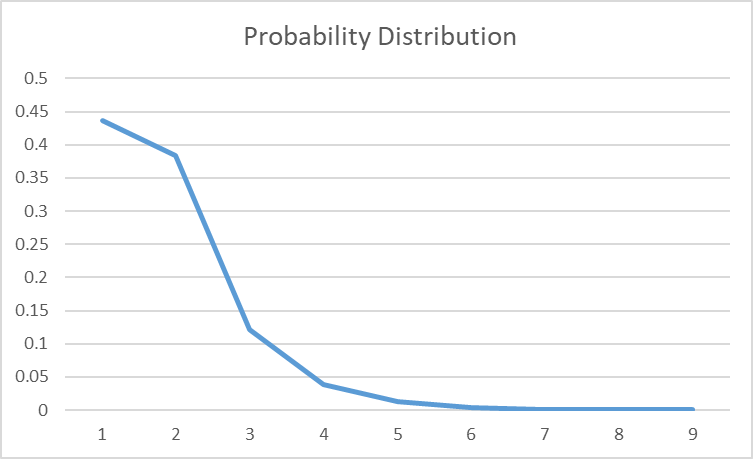
**d.]** Using column for Beth hospital, derive descriptive statistics to use it in the further analysis. Find the bins and frequency and calculate the left and right tails using formula **Bin1-(Bin1-Bind2)/2** and **Bin1+(Bin1-Bind2)/2**. Relative frequency is just frequency divided by the summation of observed frequencies 5000. Exponential distribution best fits as exponential distributions are generally good with time elapsed events. **EXPON.DIST()** gives us the exponential probability column which is multiplied by the count of simulation 5000 and this gives us the expected frequency. Chi-square formula is very basic -

(**Observed Freq – Expected Freq)^2/(Expected Freq)**.

Test statistic is summation of all Ch-square values (3557.094807). Level of significance is 0.05 because 95% confidence interval is mentioned for the total transport time. Degree of freedom is count of simulations -1-1 which is (71-1-1=69). P-value is estimated using **1-CHISQ.DIST(t,df,1)** which appears as 0.00E+00 very small. Hence, we reject the hypothesis. Our data is not exponentially distributed (Graph shown besides).

**e.]** For data analysis of average transport time, we made use of **CONVERT()** function inside which the number parameter is **Total time / Total number of victims**. We start out with the descriptive statistics following the measurements of bins, frequencies, left-right tails, relative frequency, exponential probability, expected frequency and finally the Chi-square values.

Frequency Histogram graph shows up



Test statistic- 138.1526853

Level of significance – 0.05

Df – 8-1-1 = 6

P-value turns out to be very small and we must reject the null hypothesis.

Our data of average transport time is therefore not exponential as can be clearly seen from the graph.

**Part 2 (Q2, Q3)**

Part 2 follows the same steps as Part 1 the only exception is assuming our data is normally distributed with a mean of 150 victims and standard deviation of 50. Also, table for reference involves not only the average transport time in minutes but also the standard deviation in minutes for each hospital. For implementing the same, we must use **NORM.INV(Random, mean, sd)** as the basic given data should be put to use. In the excel sheet, the histogram is also shown denoting a normal curve.

**a.]** Using X, bins and frequency is found. The frequency value of the tuple is multiplied by the probability percentage of every hospital (data given in part 1). The output derived is-

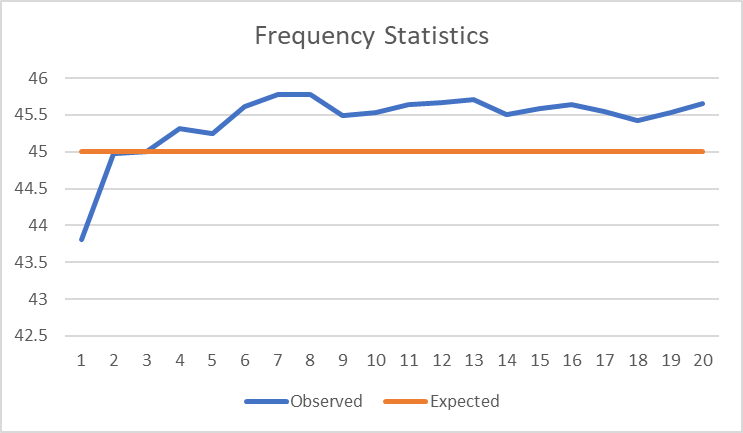




**b.]** This is the most important arithmetic of Part 2. Take the Total victim’s column and multiply it by **NORM.DIST(frequency, mean, sd,1)** where mean and standard deviation is different for all 5 hospitals. ([7,2],[10,4],[15,3],[15,5],[20,3]). The outcome follows-

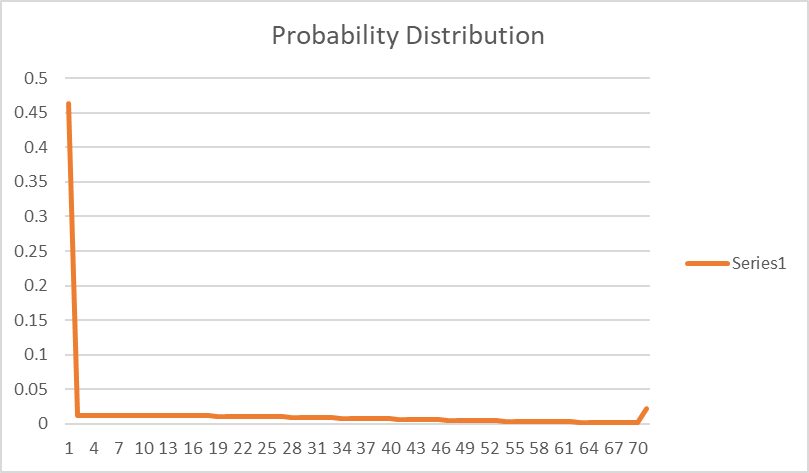




**c.]** Only the Beth hospital data is considered for the law of large numbers. As you can see from the graph, “the observed averages move towards the theoretical average with increasing trials”. Observed data is 43.80283, 44.97161……., 45.64859. The same math is indicated in the graph.

In the Excel sheet, I have shown a histogram of frequencies of Beth Israel Medical hospital.

**d.]** Exploratory Data Analysis for total transport time involves normal probability calculations with the equation **NORM.DIST(x, mean=0,sd,1)**. Mean is zero as the definition of standard normal distribution itself is 0. Finally, we get the following research-



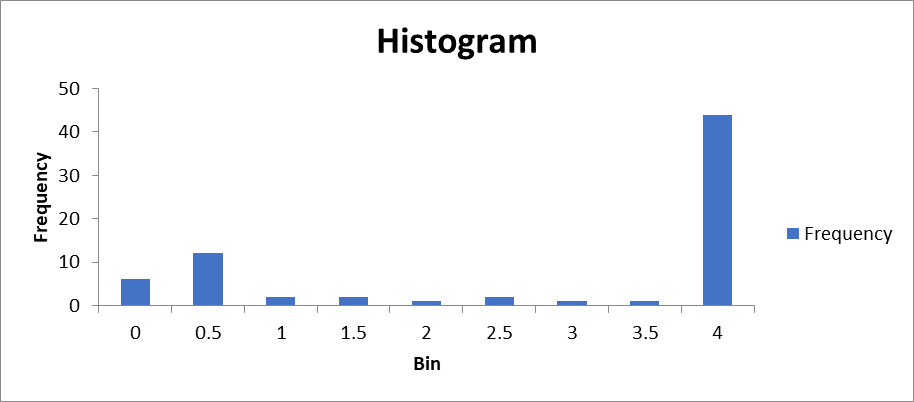
Test statistic- 13266.08886

Level of significance – 0.05

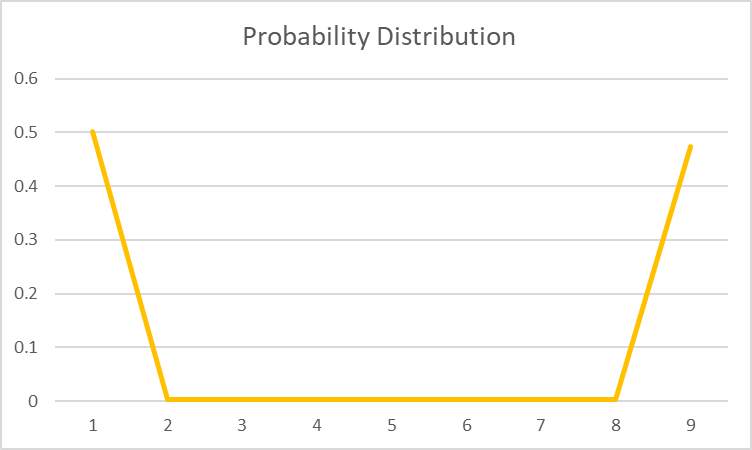
Df – 70-1-1=68

P-value turns out to be very small and we must reject the null hypothesis.

Our data of average transport time is therefore not normal which is evident from the plot.

**e.]** The average transport time per victim is measured as **Total time / Total number of victims**. Using this column, descriptive statistics is found. The Data Analysis Tool Pak gives us 9 bins and frequencies and with normal probability distribution we figure out the chi-square values.

A frequency distribution is automatically generated.



Test statistic- 549.1823923

Level of significance – 0.05

Df – 9-1-1=7

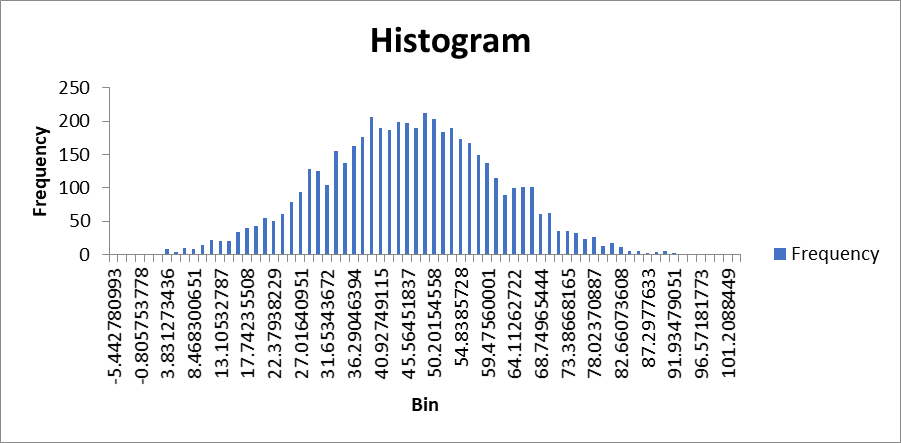
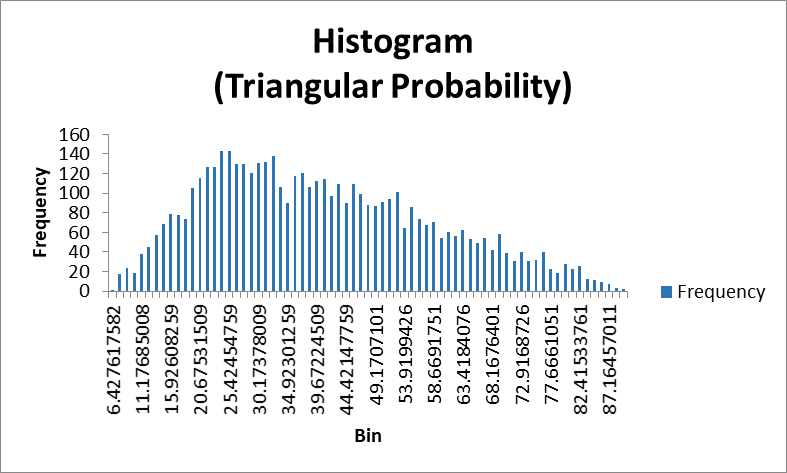
P-value turns out to be very small, so we reject the null hypothesis.

Our data distribution of average transport time is not normal and is visible from the diagram.

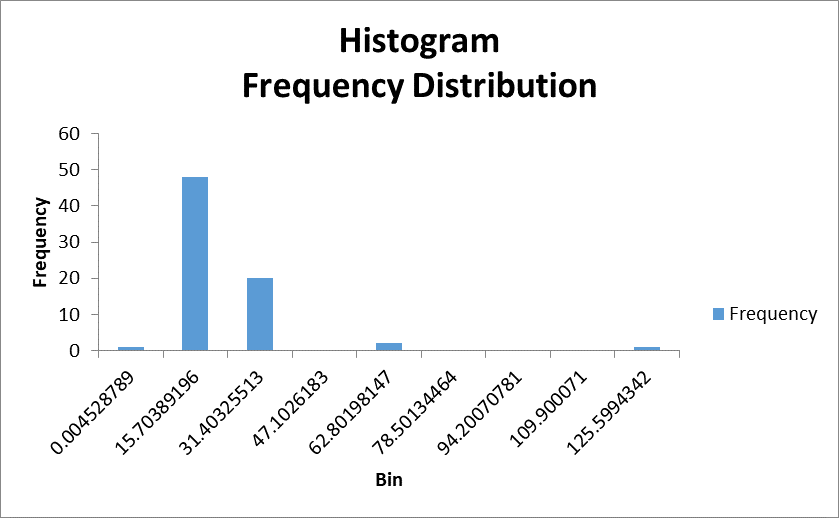
**Q4)** My opinion regarding a few major quantitative and qualitative differences between the outputs of simulations:

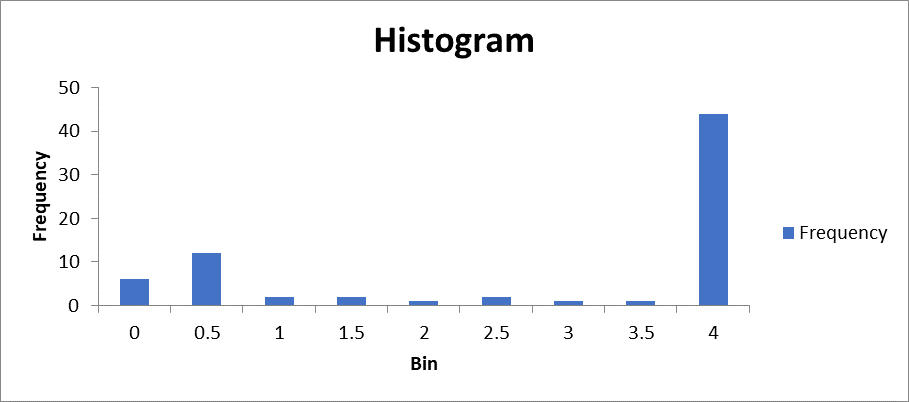
The qualitative difference in simulation 1 (S1) and simulation 2 (S2) are the graphs and visuals. In the exploratory data analysis of Beth Israel Medical Hospital, the frequency histogram for S1 is of Triangular Probability and S2 is Normal Probability. [1]

**S1 S2**



For the data analysis of average transport time, the histogram differs as there is a difference in the evaluation of total time.

 **S1 S2**

****

The quantitative differences are basically the ones dealing with numbers and statistics. [1] There is difference in the average total time in hours required to transport all the victims of each hospital. S1- 2, 1, 3, 4, 2

S2- 1, 1, 1, 1, 0 (69, 69, 69, 68, 0 converted to hours)

From this, I see that hospitals in S2 takes comparatively lesser time to transport victims from the disaster area. Hence, according to me the local emergency facility with simulation 2 will be quicker and much efficient.

Other difference is in the descriptive statistics of the average transport time i.e. Mean , Standard deviation, Kurtosis, Skewness and Count of simulations.

**Statistics S1 S2**

|  |  |  |
| --- | --- | --- |
| Mean | 13.665 | 2.702 |
| SD | 17.261 | 1.755 |
| Kurtosis | 25.106 | -1.367 |
| Skewness | 4.287 | -0.74 |
| Count | 72 | 71 |

**Q5, Q6**

Explanation of how the information obtained from our simulation can be used for planning purposes and how each simulation can be changed to provide additional useful information.

Both the simulations differ in their empirical distributions as the mean and standard deviation is involved in the simulation 2, unlike simulation 1. But common stuff is both can be used in business strategies. Our focus in this project was on the average wait time of victim’s treatment. Hence the simulations help us in finding the average wait time, how sooner the hospitals become vacant for the remaining patients to be treated. If the wait time seems to be very high, then 5 hospitals will not be enough to serve the purpose of the disaster accidents. Therefore, our conclusion from this is setting up more hospitals to cater the disaster victims.

**Conclusion**

The project helped us in practicing random number simulations with sampled data. We learnt to perform exploratory data analysis with different probability distribution formulae. Visual representations like frequency histograms and charts for exponential and normal probability distributions were encountered. With this assignment we were able to infer statistical conclusions as acknowledged in the analysis. The readiness and competency of local emergency medical facilities were studied. When simulation is intended as an aid to decision making by individuals or organizations, special problems arise in connection with designing and using the model. This project defines the use of simulation, explores factors related to its application, and suggests implications for decision making of business strategies. [2]

**References**

[1] Hocaoğlu, & Fatih, M. (2018, January 1). Qualitative Reasoning for Quantitative Simulation. <https://new.hindawi.com/journals/mse/2018/7842402/>

[2] Schultz, R. L. (2007, January 3). The use of simulation for decision making. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830190507>