Rochester Institute of Technology

Boston Duck Boat Tour Simulation

Design of Experiments

Nick Morris

2015

Problem Statement

Customers are expected to have a long wait following the purchase of their ticket. An expected long wait takes away from the intended positive experience. Six factors are involved in this study to determine their effect and how to minimize wait time.

Objective

This experiment will model a simulation of the process from the costumers' perspective, from the ticket purchase to the finish of the tour. The goal is to screen through various terms and then minimize the costumer waiting time by determining the optimal level for each significant design factor.

Design Factor Adjustments

Table 1 below represents the changes made to the two levels of each factor studied in this experiment. The numbers of boats for each of the three locations in the model are set at 3 and 8. Previously the model represented the whole system instead of each of the three locations that a customer can start the tour. The distribution process at the high level was changed to Exponential to represent a standard distribution process for service modes. The mean process times are set at 40 and 100 minutes. The low level of 40 minutes was chosen to represent days when the water isn't accessible due to weather conditions, such as ice, and the tour time is reduced. The high level of 100 minutes was chosen to represent the typical 90 minute tour and 10 minutes of load time. Down times are set at 10 and 60 minutes to represent preventative maintenance procedures (10 minutes) and issues that require technical trouble shooting (60 minutes). The occurrences of these down times are based on the number of tours completed for each location. A discrete distribution was used such that down times occur 20% of the time after 5 tours, 50% of the time after 10 tours, and 30% after 15 tours. The boat capacity low level was changed to 26 to represent slower days where boats will begin the tour without utilizing their full capacity of 36. The arrival rates are set at 10 and 15 people every 30 minutes for each location. Arrivals have a poisson distribution to represent a standard distribution process for arrival modes.

Table 1

Design Factors	Range of Levels		Units
	Low Level	*High Level*	
Number of Boats	3	8	*[boats]*
Distribution Process	Normal	Exponential	*[type]*
Mean Process Time	40	100	*[minutes]*
Down Time	10	60	*[minutes]*
Boat Capacity	26	36	*[people]*
Arrival Rate	10	15	*[People/30 min]*

Design Chosen

The design chosen for this experiment was a 2^6 full factorial. This was chosen because time and resources weren't constraints by building a simulation model to run real time in much quicker simulation time. The use of a 2^6 full factorial design with replication allows the analysis of terms up through order 6 without an alias structure present because it is an orthogonal design. This makes for a screening process that filters out significant terms from 63 candidates to result in a model with higher R^2_{ADJ} and R^2_{PRED} values.

Table 2 below represents the Power and Sample Size results from Minitab. There was a pilot study of 10 replications completed to estimate factor effects and the response standard deviation. These values were inputted into Minitab with a target power of 0.8 to calculate the required replications for each factor. There were 38 replications for a total of 2432 total runs completed in the simulation software Simio 7 for this study. The value of 38 was chosen based on a weighted average such that the weight of each factor's required replications was based on their effect value. The product sum of the effect weight column and reps column in Table 2 resulted in 38 replications.

Table 2

Factor	Center Points	Effect	Reps	Total Runs	Target Power	Actual Power	Effect Weight	Reps* 38
Number of Boats	0	0.8807	65	4160	0.8	0.80	0.0533	
Distribution Process	0	2.9442	6	384	0.8	0.81	0.1783	
Mean Process Time	0	0.8270	74	4736	0.8	0.80	0.0501	
Down Time	0	0.1141	3861	247104	0.8	0.80	0.0069	
Boat Capacity	0	1.6057	20	1280	0.8	0.81	0.0973	
Arrival Rate	0	10.1371	2	128	0.8	1.00	0.6140	

Blocking is not included in this experiment due to the homogenous nature of a simulation model. Also, the response of interest is the overall average waiting time for the whole system, as opposed to blocking by location. The experiment won't take any longer by blocking because it isn't present.

Table 3 below represents the results of the first 33 runs and the last run of the experiment in coded units and standard order. The average waiting time column is the grand mean of the average waiting time for each of the 3 locations in every row. The unit of the average waiting time is minutes.

Table 3

Number of Boats x1	Distribution Process x2	Mean Process Time x3	Down Time x4	Boat Capacity x5	Arrival Rate x6	Average Waiting Time
-1	-1	-1	-1	-1	-1	34.74358974
1	-1	-1	-1	-1	-1	39.87179487
-1	1	-1	-1	-1	-1	40.44871795
1	1	-1	-1	-1	-1	43.14102564
-1	-1	1	-1	-1	-1	34.74358974
1	-1	1	-1	-1	-1	39.87179487
-1	1	1	-1	-1	-1	41.67871245
1	1	1	-1	-1	-1	43.14102564
-1	-1	-1	1	-1	-1	34.74358974
1	-1	-1	1	-1	-1	39.87179487
-1	1	-1	1	-1	-1	40.44871795
1	1	-1	1	-1	-1	43.14102564
-1	-1	1	1	-1	-1	34.74358974
1	-1	1	1	-1	-1	39.87179487
-1	1	1	1	-1	-1	41.67871245
1	1	1	1	-1	-1	43.14102564
-1	-1	-1	-1	1	-1	56.75925926
1	-1	-1	-1	1	-1	56.75925926
-1	1	-1	-1	1	-1	58.42592593
1	1	-1	-1	1	-1	58.42592593
-1	-1	1	-1	1	-1	56.75925926
1	-1	1	-1	1	-1	56.75925926
-1	1	1	-1	1	-1	58.42592593
1	1	1	-1	1	-1	58.42592593
-1	-1	-1	1	1	-1	56.75925926
1	-1	-1	1	1	-1	56.75925926
-1	1	-1	1	1	-1	58.42592593
1	1	-1	1	1	-1	58.42592593
-1	-1	1	1	1	-1	56.75925926
1	-1	1	1	1	-1	56.75925926
-1	1	1	1	1	-1	58.42592593
1	1	1	1	1	-1	58.42592593
-1	-1	-1	-1	-1	1	29.03846154
…1	…1	…1	…1	…1	…1	34.01388889

Running the Experiment

Figure 1 below is a snapshot of the model mid-run where all design factors are set at their high level. There are 3 parallel processes representing the three locations, Prudential Center, Museum of Science, and New England Aquarium. Each location has two sources, one for the one-time creation of duck boats and the second is for the arrival of customers. The duck boats and customers converge at the starting location where customers can board duck boats at 30 minute increments throughout the day from 8am – 4:30pm. This 30 minute increment represents the time slots online that a ticket can be bought. This is controlled in the model by three resources, represented by the three people standing by each starting location, which are only available every 30 minutes to allow customers to board the duck boat and begin the tour. The duck boat will leave the starting location and travel the time length of the tour, then drop customers off at the next node where they will exit the system at the far end of Figure 1. After dropping customers off, the duck boat will make a stop at a node for a maintenance check before making a round trip back to their starting location. Once at the starting location the duck-boat is available for another tour.

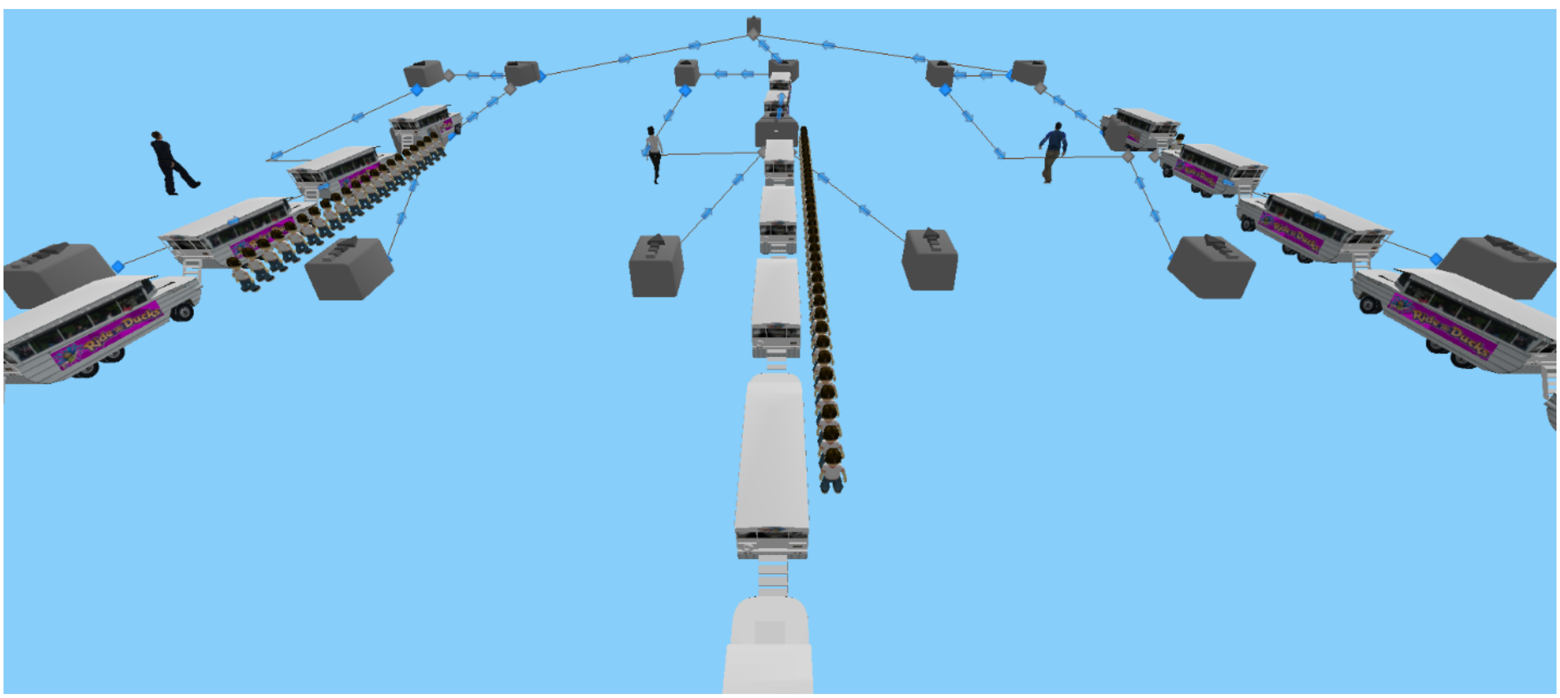

Figure 1

Figure 2 below represents the average waiting time (TIQ - Time in Queue) for each location (P - Prudential Center, M - Museum of Science, A - New England Aquarium) during one run (Standard Order Run #64) where all factors were set at their high level. This chart was used to determine the system's Warm-Up Period. The data during this period is blocked from the results of this experiment so the results aren't skewed by the initial time necessary to load the system with customers. The black dotted vertical line is the chosen warm-up period of 3 hours. This was chosen by finding the nearest hour to the starting time (8 am) where the proceeding trend line shows a stable slope. The total run length is still 8.5 hours in real time beginning at 11am instead of 8am.

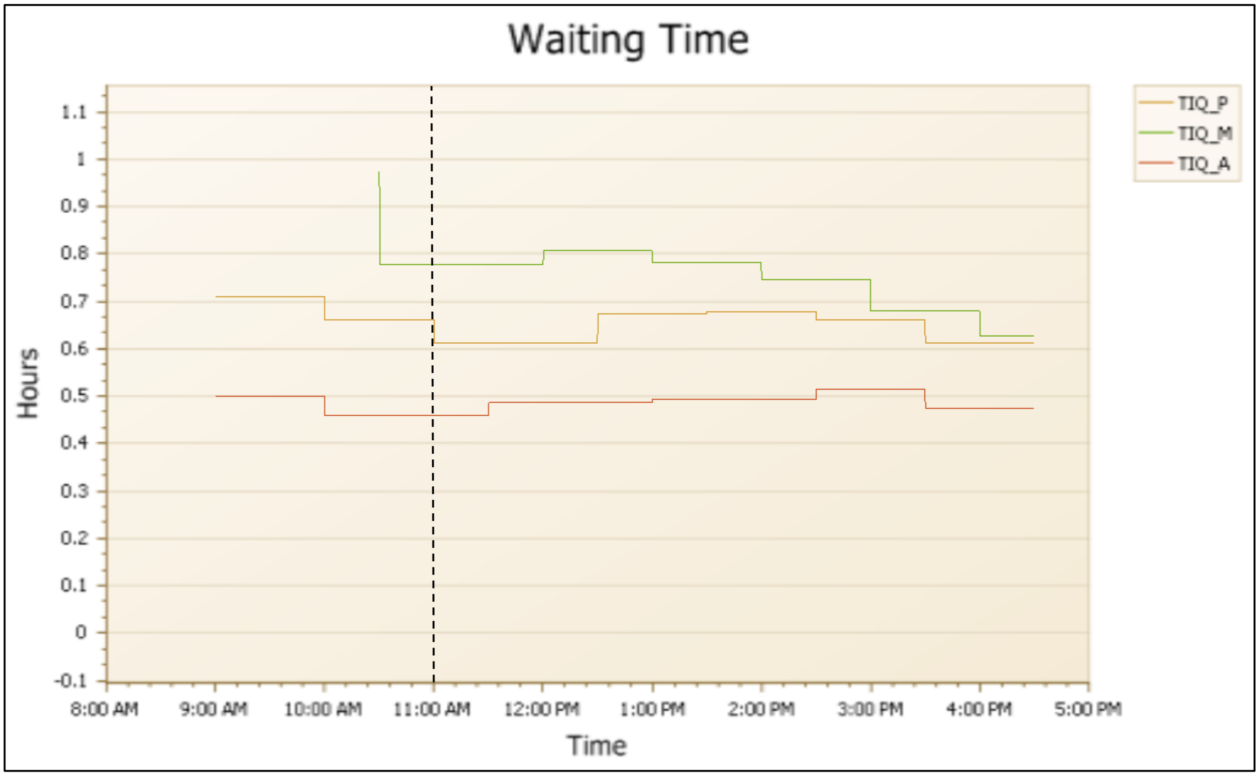

Figure 2

The modeling time took hours to complete over the course of weeks. The entire experiment of 2432 total runs took no more than 2 minutes to obtain results due to computing speed. The problems faced during the modeling process were the selection of levels, incorporating randomness to avoid identical response values across runs, and controlling the 30 minute tour departure increments. The issues that arose from the selection of levels were response results that had minimal difference in value between levels and some run scenarios wouldn't even record response values. This was handled by iteratively trouble shooting trials until a set of viable levels were determined, and then the levels were chosen based on engineering logic of which ones best represented the known information of the real system. The issue of identical responses was due to the arrival rate and down time factors being deterministic inputs. This was handled by using random discrete distributions, as previously mentioned, to model the natural variance of a real-world system. The problem of modeling 30 minute departures was due to lack of modeling experience in Simio 7 which was handled by trouble-shooting. I would try out various ideas and possibilities until I could visually identify this desired function.

Statistical Analysis

The hypothesis test of the effect on the response for each term analyzed in the model is shown below. The factors are rewritten in terms of x_i such that {Number of Boats, Distribution Process, Mean Process Time, Down Time, Boat Capacity, Arrival Rate} = {x1, x2, x3, x4, x5, x6} as displayed in Table 3.

Hypothesis Test for Each Model Term:

H_o: $\beta_i = 0 \quad \forall \quad i$ € Model {x1, x2, x3, x4, x5, x6 up through order 6}

H_a: $\beta_i \neq 0 \quad \forall \quad i$ € Model {x1, x2, x3, x4, x5, x6 up through order 6}

Level of Significance $\alpha = 0.05$

The ANOVA results from Minitab are shown below. The yellow highlighted terms will be eliminated from the model per the screening process due to P-Values > 0.15. The term x4 and all higher order interaction terms involving x4 will be eliminated from the model. This shows that the model still maintains its hierarchal form such that Down Time (x4) is insignificant up through order 6. The Model P-Value of > 0.000 < level of significance 0.05 indicates that at least one term explains Average Waiting Time.

The Model Summary below shows a moderately strong relationship between Average Waiting Time and the 63 model terms due to a R^2 Value of 82.26% where a perfect value is 100%. Therefore 82.26% of the total variance in Average Waiting Time can be explained by the linear relationship between Average Waiting Time and the corresponding model terms. The regression equation can be found at the bottom of the statistical output. Notice that the R^2 value agrees with the Model P-Value such that at least one factor explains Average Waiting Time. The R^2_{ADJ} value of 81.79% shows that despite the noise added to the regression equation by the insignificant terms, highlighted in yellow, the equation maintains accuracy due to < 1% difference from R^2. The R^2_{PRED} value of 81.29% shows that the regression equation captures 81.29% of the total variance when removing a known response value and predicting it with the remaining response values.

The coded coefficients below shows that Arrival Rate (x6) has the largest magnitude of impact on Average Waiting Time due to the highest absolute value of -14.3219 [minutes] in the Effects column. The remaining magnitudes of impact that stand out are Boat Capacity (x5) and x5*x6 at 11.6187 and -2.9587 [minutes] respectively. These top three terms based on effect are highlighted in green. Notice that the magnitude of the Effect values and corresponding P-Values are inversely related.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	63	219268	3480	174.26	0.000
Linear	6	208283	34714	1738.11	0.000
x1	1	783	783	39.23	0.000
x2	1	344	344	17.22	0.000
x3	1	369	369	18.48	0.000
x4	1	0	0	0.00	0.965
x5	1	82076	82076	4109.50	0.000
x6	1	124711	124711	6244.22	0.000
2-Way Interactions	15	8167	544	27.26	0.000
x1*x2	1	513	513	25.69	0.000
x1*x3	1	355	355	17.79	0.000
x1*x4	1	0	0	0.00	0.965
x1*x5	1	434	434	21.72	0.000
x1*x6	1	322	322	16.12	0.000
x2*x3	1	373	373	18.69	0.000
x2*x4	1	0	0	0.00	0.944
x2*x5	1	318	318	15.93	0.000
x2*x6	1	124	124	6.22	0.013
x3*x4	1	0	0	0.00	0.997
x3*x5	1	196	196	9.81	0.002
x3*x6	1	208	208	10.44	0.001
x4*x5	1	0	0	0.00	0.990
x4*x6	1	0	0	0.00	0.979
x5*x6	1	5322	5322	266.49	0.000
3-Way Interactions	20	2068	103	5.18	0.000
x1*x2*x3	1	351	351	17.58	0.000
x1*x2*x4	1	0	0	0.00	0.944
x1*x2*x5	1	270	270	13.50	0.000
x1*x2*x6	1	289	289	14.50	0.000
x1*x3*x4	1	0	0	0.00	0.997
x1*x3*x5	1	185	185	9.24	0.002
x1*x3*x6	1	203	203	10.15	0.001
x1*x4*x5	1	0	0	0.00	0.990
x1*x4*x6	1	0	0	0.00	0.979
x1*x5*x6	1	185	185	9.25	0.002
x2*x3*x4	1	0	0	0.00	0.961
x2*x3*x5	1	198	198	9.93	0.002
x2*x3*x6	1	202	202	10.12	0.001
x2*x4*x5	1	0	0	0.00	0.996
x2*x4*x6	1	0	0	0.00	0.983
x2*x5*x6	1	90	90	4.51	0.034
x3*x4*x5	1	0	0	0.00	0.971
x3*x4*x6	1	0	0	0.00	0.987
x3*x5*x6	1	95	95	4.75	0.029
x4*x5*x6	1	0	0	0.00	0.976

Source	DF	Adj SS	Adj MS	F-Value	P-Value
4-Way Interactions	15	654	44	2.18	0.005
x1*x2*x3*x4	1	0	0	0.00	0.961
x1*x2*x3*x5	1	182	182	9.12	0.003
x1*x2*x3*x6	1	209	209	10.46	0.001
x1*x2*x4*x5	1	0	0	0.00	0.996
x1*x2*x4*x6	1	0	0	0.00	0.983
x1*x2*x5*x6	1	83	83	4.15	0.042
x1*x3*x4*x5	1	0	0	0.00	0.971
x1*x3*x4*x6	1	0	0	0.00	0.987
x1*x3*x5*x6	1	90	90	4.51	0.034
x1*x4*x5*x6	1	0	0	0.00	0.976
x2*x3*x4*x5	1	0	0	0.00	0.996
x2*x3*x4*x6	1	0	0	0.00	0.962
x2*x3*x5*x6	1	90	90	4.51	0.034
x2*x4*x5*x6	1	0	0	0.01	0.932
x3*x4*x5*x6	1	0	0	0.00	0.981
5-Way Interactions	6	95	16	0.79	0.575
x1*x2*x3*x4*x5	1	0	0	0.00	0.996
x1*x2*x3*x4*x6	1	0	0	0.00	0.962
x1*x2*x3*x5*x6	1	95	95	4.74	0.030
x1*x2*x4*x5*x6	1	0	0	0.01	0.932
x1*x3*x4*x5*x6	1	0	0	0.00	0.981
x2*x3*x4*x5*x6	1	0	0	0.01	0.918
6-Way Interactions	1	0	0	0.01	0.918
x1*x2*x3*x4*x5*x6	1	0	0	0.01	0.918
Error	2368	47294	20		
Total	2431	266562			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.46903	82.26%	81.79%	81.29%

Regression Equation

```
Average Waiting Time =
38.4049 - 0.5676 x1 + 0.3760 x2 + 0.3896 x3 + 0.0039 x4 + 5.8093 x5
        - 7.1610 x6 - 0.4593 x1*x2 - 0.3822 x1*x3 + 0.0039 x1*x4
        + 0.4224 x1*x5 - 0.3639 x1*x6 + 0.3918 x2*x3 + 0.0063 x2*x4
        - 0.3617 x2*x5 + 0.2261 x2*x6 + 0.0003 x3*x4 - 0.2838 x3*x5
        + 0.2928 x3*x6 + 0.0011 x4*x5 + 0.0024 x4*x6 - 1.4794 x5*x6
        - 0.3800 x1*x2*x3 + 0.0063 x1*x2*x4 + 0.3330 x1*x2*x5
        - 0.3450 x1*x2*x6 + 0.0003 x1*x3*x4 + 0.2755 x1*x3*x5
        - 0.2886 x1*x3*x6 + 0.0011 x1*x4*x5 + 0.0024 x1*x4*x6
        + 0.2756 x1*x5*x6 - 0.0044 x2*x3*x4 - 0.2856 x2*x3*x5
        + 0.2883 x2*x3*x6 - 0.0005 x2*x4*x5 - 0.0019 x2*x4*x6
        - 0.1925 x2*x5*x6 + 0.0033 x3*x4*x5 + 0.0015 x3*x4*x6
        - 0.1975 x3*x5*x6 + 0.0027 x4*x5*x6 - 0.0044 x1*x2*x3*x4
        + 0.2737 x1*x2*x3*x5 - 0.2931 x1*x2*x3*x6 - 0.0005 x1*x2*x4*x5
        - 0.0019 x1*x2*x4*x6 + 0.1845 x1*x2*x5*x6 + 0.0033 x1*x3*x4*x5
        + 0.0015 x1*x3*x4*x6 + 0.1924 x1*x3*x5*x6 + 0.0027 x1*x4*x5*x6
        - 0.0005 x2*x3*x4*x5 + 0.0044 x2*x3*x4*x6 - 0.1926 x2*x3*x5*x6
        + 0.0078 x2*x4*x5*x6 + 0.0022 x3*x4*x5*x6 - 0.0005 x1*x2*x3*x4*x5
        + 0.0044 x1*x2*x3*x4*x6 + 0.1973 x1*x2*x3*x5*x6
        + 0.0078 x1*x2*x4*x5*x6 + 0.0022 x1*x3*x4*x5*x6
        - 0.0093 x2*x3*x4*x5*x6 - 0.0093 x1*x2*x3*x4*x5*x6
```

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		38.4049	0.0906	423.79	0.000	
x1	-1.1351	-0.5676	0.0906	-6.26	0.000	1.00
x2	0.7520	0.3760	0.0906	4.15	0.000	1.00
x3	0.7791	0.3896	0.0906	4.30	0.000	1.00
x4	0.0079	0.0039	0.0906	0.04	0.965	1.00
x5	11.6187	5.8093	0.0906	64.11	0.000	1.00
x6	-14.3219	-7.1610	0.0906	-79.02	0.000	1.00
x1*x2	-0.9186	-0.4593	0.0906	-5.07	0.000	1.00
x1*x3	-0.7644	-0.3822	0.0906	-4.22	0.000	1.00
x1*x4	0.0079	0.0039	0.0906	0.04	0.965	1.00
x1*x5	0.8447	0.4224	0.0906	4.66	0.000	1.00
x1*x6	-0.7277	-0.3639	0.0906	-4.02	0.000	1.00
x2*x3	0.7836	0.3918	0.0906	4.32	0.000	1.00
x2*x4	0.0127	0.0063	0.0906	0.07	0.944	1.00
x2*x5	-0.7233	-0.3617	0.0906	-3.99	0.000	1.00
x2*x6	0.4521	0.2261	0.0906	2.49	0.013	1.00
x3*x4	0.0007	0.0003	0.0906	0.00	0.997	1.00
x3*x5	-0.5676	-0.2838	0.0906	-3.13	0.002	1.00
x3*x6	0.5855	0.2928	0.0906	3.23	0.001	1.00
x4*x5	0.0022	0.0011	0.0906	0.01	0.990	1.00
x4*x6	0.0048	0.0024	0.0906	0.03	0.979	1.00
x5*x6	-2.9587	-1.4794	0.0906	-16.32	0.000	1.00
x1*x2*x3	-0.7599	-0.3800	0.0906	-4.19	0.000	1.00
x1*x2*x4	0.0127	0.0063	0.0906	0.07	0.944	1.00
x1*x2*x5	0.6659	0.3330	0.0906	3.67	0.000	1.00
x1*x2*x6	-0.6900	-0.3450	0.0906	-3.81	0.000	1.00
x1*x3*x4	0.0007	0.0003	0.0906	0.00	0.997	1.00
x1*x3*x5	0.5510	0.2755	0.0906	3.04	0.002	1.00
x1*x3*x6	-0.5773	-0.2886	0.0906	-3.19	0.001	1.00
x1*x4*x5	0.0022	0.0011	0.0906	0.01	0.990	1.00
x1*x4*x6	0.0048	0.0024	0.0906	0.03	0.979	1.00
x1*x5*x6	0.5512	0.2756	0.0906	3.04	0.002	1.00
x2*x3*x4	-0.0089	-0.0044	0.0906	-0.05	0.961	1.00
x2*x3*x5	-0.5711	-0.2856	0.0906	-3.15	0.002	1.00
x2*x3*x6	0.5767	0.2883	0.0906	3.18	0.001	1.00
x2*x4*x5	-0.0010	-0.0005	0.0906	-0.01	0.996	1.00
x2*x4*x6	-0.0038	-0.0019	0.0906	-0.02	0.983	1.00
x2*x5*x6	-0.3850	-0.1925	0.0906	-2.12	0.034	1.00
x3*x4*x5	0.0066	0.0033	0.0906	0.04	0.971	1.00
x3*x4*x6	0.0030	0.0015	0.0906	0.02	0.987	1.00
x3*x5*x6	-0.3950	-0.1975	0.0906	-2.18	0.029	1.00
x4*x5*x6	0.0054	0.0027	0.0906	0.03	0.976	1.00
x1*x2*x3*x4	-0.0089	-0.0044	0.0906	-0.05	0.961	1.00
x1*x2*x3*x5	0.5474	0.2737	0.0906	3.02	0.003	1.00
x1*x2*x3*x6	-0.5862	-0.2931	0.0906	-3.23	0.001	1.00
x1*x2*x4*x5	-0.0010	-0.0005	0.0906	-0.01	0.996	1.00
x1*x2*x4*x6	-0.0038	-0.0019	0.0906	-0.02	0.983	1.00
x1*x2*x5*x6	0.3691	0.1845	0.0906	2.04	0.042	1.00
x1*x3*x4*x5	0.0066	0.0033	0.0906	0.04	0.971	1.00
x1*x3*x4*x6	0.0030	0.0015	0.0906	0.02	0.987	1.00
x1*x3*x5*x6	0.3848	0.1924	0.0906	2.12	0.034	1.00
x1*x4*x5*x6	0.0054	0.0027	0.0906	0.03	0.976	1.00
x2*x3*x4*x5	-0.0010	-0.0005	0.0906	-0.01	0.996	1.00
x2*x3*x4*x6	0.0087	0.0044	0.0906	0.05	0.962	1.00
x2*x3*x5*x6	-0.3851	-0.1926	0.0906	-2.12	0.034	1.00
x2*x4*x5*x6	0.0155	0.0078	0.0906	0.09	0.932	1.00
x3*x4*x5*x6	0.0043	0.0022	0.0906	0.02	0.981	1.00
x1*x2*x3*x4*x5	-0.0010	-0.0005	0.0906	-0.01	0.996	1.00
x1*x2*x3*x4*x6	0.0087	0.0044	0.0906	0.05	0.962	1.00
x1*x2*x3*x5*x6	0.3946	0.1973	0.0906	2.18	0.030	1.00
x1*x2*x4*x5*x6	0.0155	0.0078	0.0906	0.09	0.932	1.00
x1*x3*x4*x5*x6	0.0043	0.0022	0.0906	0.02	0.981	1.00
x2*x3*x4*x5*x6	-0.0186	-0.0093	0.0906	-0.10	0.918	1.00
x1*x2*x3*x4*x5*x6	-0.0186	-0.0093	0.0906	-0.10	0.918	1.00

Figure 3 below shows which of the 63 model terms are significant (red box) and are not significant (blue circle) based on a Normal Plot of their effects. There are many significant terms which cannot be displayed due to the clustering. All of the not significant terms are the ones that are highlighted in yellow above in the ANOVA results. Also the three significant terms which stand out from the rest of the data are labeled and are the ones highlighted in green above in the coded coefficients results.

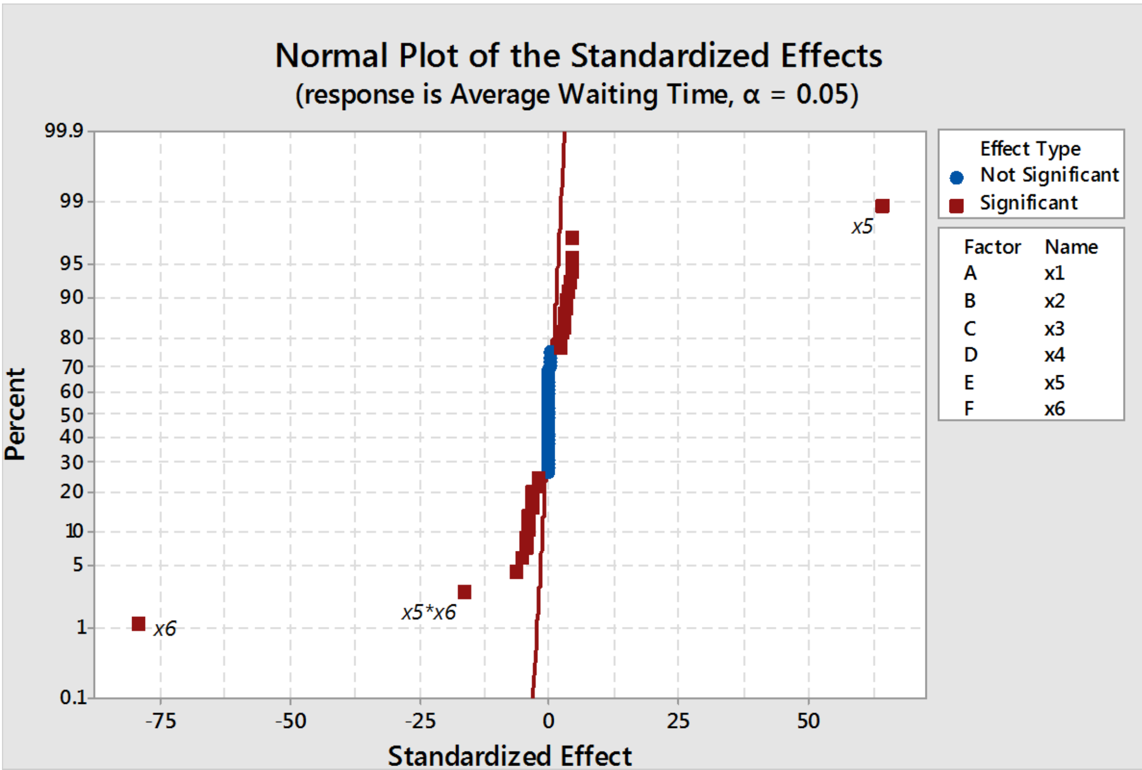

Figure 3

The ANOVA output below represents the re-fitted model with all insignificant terms removed, resulting in 31 significant terms with P-Values < level of significance 0.05. The Lack-of-Fit P-Value of < 1.000 > level of significance 0.05 indicates that there are no available terms missing from the model that are significant to Average Waiting Time.

The Model Summary below shows slightly different results from the previous Model Summary for each Coefficient of Determination. The R^2 value remains the same out to 4 decimal places, the R^2_{ADJ} value increases by 0.24%, the absolute difference between R^2 and R^2_{ADJ} decreases by 0.24%, and the R^2_{PRED} value increases by 0.49%. The R^2 value can be assumed to decrease when the Regression model is refitted because the total number of factors in the model has decreased. The R^2_{ADJ} value and R^2_{PRED} value increase because the Regression Model has been refitted with only significant terms; therefore the Coefficient of Determination isn't penalized the same as in the previous model for including insignificant factors at a 0.15 level of significance.

The regression equation below represents the best fit model for Average Waiting Time due to screening with ANOVA and seen by the improvement in the Coefficients of Determination.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	31	219267	7073	358.92	0.000
Linear	5	208283	41657	2113.85	0.000
x1	1	783	783	39.75	0.000
x2	1	344	344	17.45	0.000
x3	1	369	369	18.73	0.000
x5	1	82076	82076	4164.92	0.000
x6	1	124711	124711	6328.42	0.000
2-Way Interactions	10	8167	817	41.44	0.000
x1*x2	1	513	513	26.04	0.000
x1*x3	1	355	355	18.03	0.000
x1*x5	1	434	434	22.02	0.000
x1*x6	1	322	322	16.34	0.000
x2*x3	1	373	373	18.95	0.000
x2*x5	1	318	318	16.14	0.000
x2*x6	1	124	124	6.31	0.012
x3*x5	1	196	196	9.94	0.002
x3*x6	1	208	208	10.58	0.001
x5*x6	1	5322	5322	270.08	0.000
3-Way Interactions	10	2068	207	10.49	0.000
x1*x2*x3	1	351	351	17.82	0.000
x1*x2*x5	1	270	270	13.68	0.000
x1*x2*x6	1	289	289	14.69	0.000
x1*x3*x5	1	185	185	9.37	0.002
x1*x3*x6	1	203	203	10.28	0.001
x1*x5*x6	1	185	185	9.37	0.002
x2*x3*x5	1	198	198	10.06	0.002
x2*x3*x6	1	202	202	10.26	0.001
x2*x5*x6	1	90	90	4.57	0.033
x3*x5*x6	1	95	95	4.81	0.028
4-Way Interactions	5	654	131	6.64	0.000
x1*x2*x3*x5	1	182	182	9.25	0.002
x1*x2*x3*x6	1	209	209	10.60	0.001
x1*x2*x5*x6	1	83	83	4.20	0.040
x1*x3*x5*x6	1	90	90	4.57	0.033
x2*x3*x5*x6	1	90	90	4.58	0.033
5-Way Interactions	1	95	95	4.81	0.028
x1*x2*x3*x5*x6	1	95	95	4.81	0.028
Error	2400	47296	20		
Lack-of-Fit	32	1	0	0.00	1.000
Pure Error	2368	47294	20		
Total	2431	266562			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.43920	82.26%	82.03%	81.78%

Regression Equation

```
Average Waiting Time =
            38.4049 - 0.5676 x1 + 0.3760 x2 + 0.3896 x3 + 5.8093 x5 - 7.1610 x6
                    - 0.4593 x1*x2 - 0.3822 x1*x3 + 0.4224 x1*x5 - 0.3639 x1*x6
                    + 0.3918 x2*x3 - 0.3617 x2*x5 + 0.2261 x2*x6 - 0.2838 x3*x5
                    + 0.2928 x3*x6 - 1.4794 x5*x6 - 0.3800 x1*x2*x3 + 0.3330 x1*x2*x5
                    - 0.3450 x1*x2*x6 + 0.2755 x1*x3*x5 - 0.2886 x1*x3*x6
                    + 0.2756 x1*x5*x6 - 0.2856 x2*x3*x5 + 0.2883 x2*x3*x6
                    - 0.1925 x2*x5*x6 - 0.1975 x3*x5*x6 + 0.2737 x1*x2*x3*x5
                    - 0.2931 x1*x2*x3*x6 + 0.1845 x1*x2*x5*x6 + 0.1924 x1*x3*x5*x6
                    - 0.1926 x2*x3*x5*x6 + 0.1973 x1*x2*x3*x5*x6
```

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		38.4049	0.0900	426.64	0.000	
x1	-1.1351	-0.5676	0.0900	-6.31	0.000	1.00
x2	0.7520	0.3760	0.0900	4.18	0.000	1.00
x3	0.7791	0.3896	0.0900	4.33	0.000	1.00
x5	11.6187	5.8093	0.0900	64.54	0.000	1.00
x6	-14.3219	-7.1610	0.0900	-79.55	0.000	1.00
x1*x2	-0.9186	-0.4593	0.0900	-5.10	0.000	1.00
x1*x3	-0.7644	-0.3822	0.0900	-4.25	0.000	1.00
x1*x5	0.8447	0.4224	0.0900	4.69	0.000	1.00
x1*x6	-0.7277	-0.3639	0.0900	-4.04	0.000	1.00
x2*x3	0.7836	0.3918	0.0900	4.35	0.000	1.00
x2*x5	-0.7233	-0.3617	0.0900	-4.02	0.000	1.00
x2*x6	0.4521	0.2261	0.0900	2.51	0.012	1.00
x3*x5	-0.5676	-0.2838	0.0900	-3.15	0.002	1.00
x3*x6	0.5855	0.2928	0.0900	3.25	0.001	1.00
x5*x6	-2.9587	-1.4794	0.0900	-16.43	0.000	1.00
x1*x2*x3	-0.7599	-0.3800	0.0900	-4.22	0.000	1.00
x1*x2*x5	0.6659	0.3330	0.0900	3.70	0.000	1.00
x1*x2*x6	-0.6900	-0.3450	0.0900	-3.83	0.000	1.00
x1*x3*x5	0.5510	0.2755	0.0900	3.06	0.002	1.00
x1*x3*x6	-0.5773	-0.2886	0.0900	-3.21	0.001	1.00
x1*x5*x6	0.5512	0.2756	0.0900	3.06	0.002	1.00
x2*x3*x5	-0.5711	-0.2856	0.0900	-3.17	0.002	1.00
x2*x3*x6	0.5767	0.2883	0.0900	3.20	0.001	1.00
x2*x5*x6	-0.3850	-0.1925	0.0900	-2.14	0.033	1.00
x3*x5*x6	-0.3950	-0.1975	0.0900	-2.19	0.028	1.00
x1*x2*x3*x5	0.5474	0.2737	0.0900	3.04	0.002	1.00
x1*x2*x3*x6	-0.5862	-0.2931	0.0900	-3.26	0.001	1.00
x1*x2*x5*x6	0.3691	0.1845	0.0900	2.05	0.040	1.00
x1*x3*x5*x6	0.3848	0.1924	0.0900	2.14	0.033	1.00
x2*x3*x5*x6	-0.3851	-0.1926	0.0900	-2.14	0.033	1.00
x1*x2*x3*x5*x6	0.3946	0.1973	0.0900	2.19	0.028	1.00

The residuals don't follow a normally independent distribution because not all assumptions are met according to the Residual Plots in Figure 4. The assumption of normality is violated due to the residuals trailing far off the normal line at the upper percentile end, the P-Value of < 0.010 < Level of Significance 0.05, and the RJ-Value of 0.922 < RJ-Critical Value 0.9835 in the Normal Probability Plot. Therefore we reject the normality assumption. The assumption of independence is met due to the visual identification of sequential variation in residual values within the Versus Order chart. Therefore we fail to reject the independence assumption. The mean of zero assumption is met due to a P-Value of < 1.000 > Level of Significance 0.05 and T-Value of > 0.00 < T-Critical Value 1.96 which is visually displayed in the Histogram by the Confidence Interval of residuals capturing a residual value of zero about the zero bin. Therefore we fail to reject the mean of zero assumption. The consistent variance assumption is violated due to a Levene's Test P-Value of > 0.000 < Level of Significance 0.05 which is visually displayed in the Test for Equal Variance graph by all confidence intervals not intersecting each other for all levels of each factor. Therefore we reject the consistent variance assumption.

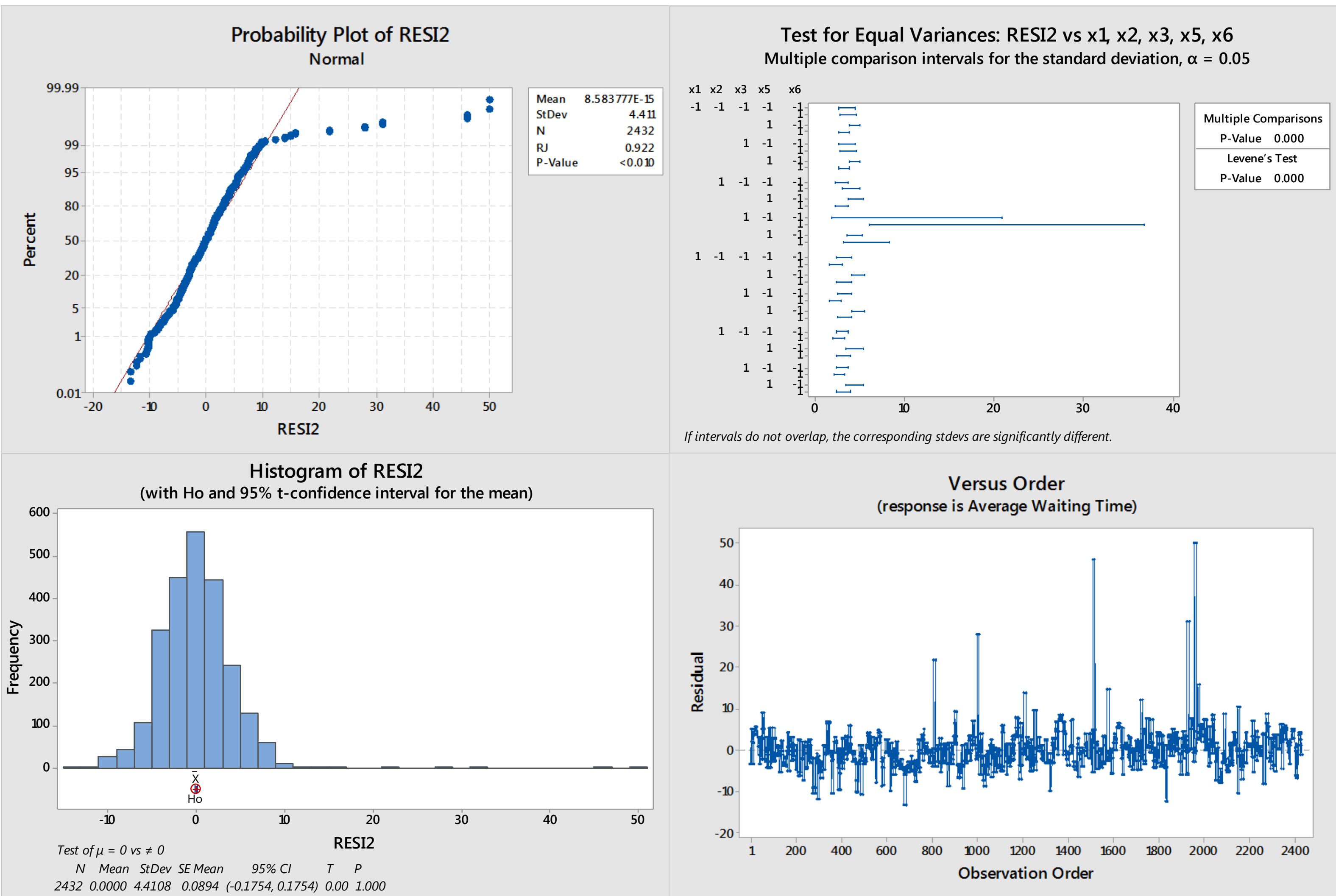

Figure 4

The Optimization output below from Minitab minimizes the value for Average Waiting Time based on the regression equation above in the re-fitted model. The smallest possible value for Average Waiting Time based on the data gathered is 16.0256 [minutes] which is set as the target. The solution below results in a minimal Average Waiting Time of 25.0798 [minutes] with a 95% Confidence Interval of (24.081, 26.078) [minutes]. This solution comes from the design factors {x1, x2, x3, x5, x6} set at {1, 1, -1, -1, 1}. These coded values represent {Number of Boats, Distribution Process, Mean Process Time, Boat Capacity, Arrival Rate} = {8, Exponential, 40, 26, 15}.

```
Parameters
Response               Goal     Lower   Target    Upper  Weight  Importance
Average Waiting Time   Minimum          16.0256  85.6713       1           1

Solution
                     Average
                     Waiting
                        Time    Composite
x1  x2  x3  x5  x6       Fit  Desirability
1   1   -1  -1  1    25.0798      0.869996

Multiple Response Prediction
Response                  Fit  SE Fit       95% CI            95% PI
Average Waiting Time   25.080   0.509  (24.081, 26.078)  (16.318, 33.842)
```

The Optimization Plot below was run with Mean Process Time (x3) held at a coded value of 1 (100 [minutes]) to maintain the typical tour time length while minimizing Average Waiting Time. The solution below results in a minimal Average Waiting Time of 25.1580 [minutes] with a 95% Confidence Interval of (24.159, 26.157) [minutes]. This solution comes from the design factors {x1, x2, x3, x5, x6} set at {1, 1, 1, -1, 1}. These coded values represent {Number of Boats, Distribution Process, Mean Process Time, Boat Capacity, Arrival Rate} = {8, Exponential, 100, 26, 15}.

```
Parameters
Response               Goal     Lower   Target    Upper  Weight  Importance
Average Waiting Time   Minimum          16.0256  85.6713       1           1

Solution
                     Average
                     Waiting
                        Time    Composite
x1  x2  x3  x5  x6       Fit  Desirability
1   1   1   -1  1    25.1580      0.868874

Multiple Response Prediction
Response                  Fit  SE Fit       95% CI            95% PI
Average Waiting Time   25.158   0.509  (24.159, 26.157)  (16.396, 33.920)
```

Statistical Validation

ANOVA was used to analyze the data because it is applicable to a full factorial design where it can analyze higher order factor interactions and be used for screening factors. This is a standard analysis incorporated into the DOE analysis of a factorial design in Minitab. A total of 32 terms were removed from the original model by screening with the ANOVA table. The Model Summary allowed the model as a whole to be quantified. This was used to compare the original model, which included all possible terms for a 2^6 full factorial design with replication, with the re-fitted model based on screening factors from the original ANOVA results. The refitted model shows superiority based on a slight improvement in R^2_{ADJ}, $| R^2 - R^2_{ADJ} |$, and R^2_{PRED} values. The residuals of the re-fitted model were assessed with an analytical approach to determine if the residuals are Normally Independently Distributed. The Ryan-Joiner Normal Probability Plot results show that the normality assumption is violated. The Versus Order chart shows that the independence assumption is met. The Test for Equal Variances with a Levene's Test shows that the equal variance assumption is violated. The Histogram with a 1-sample T-Test shows that the zero assumption is met. The residuals of the re-fitted model are determined to not be Normally Independently Distributed due to not meeting all four assumptions.

Discussion

The results of this study show that Number of Boats, Distribution Process, Mean Process Time, Boat Capacity, and Arrival Rate are the significant design factors whereas Down Time isn't and was left out of the final model. The final model summary shows a moderately strong relationship between Average Waiting Time and the 31 significant terms with all Coefficients of Determination taking on values less than 1% different from each other and each greater than 80%. The effects of these model terms, especially Boat Capacity, Arrival Rate and Boat Capacity*Arrival Rate, on the response are key in determining the minimal Average Waiting Time. A response optimizer was run twice with the second run restricting the Mean Process Time at the typical 100 [minutes], which resulted in a value less than 6 [seconds] longer than the first unrestricted run.

Conclusion

The results of the screening process through the Analysis of Variance determined the significant design factors and provided a regression equation with sufficient strength to predict Average Waiting Time. The results of the Response Optimization constrained by expected tour time length is 95% confident that a customer will on average wait 24.159 to 26.157 minutes. Based on the final model, there should be 8 duck boats allocated for each of the three starting locations. Process controls should exist to control the tour time length to follow an exponential distribution with a mean of 100 minutes. Tours should begin once the duck boats are loaded with 26 people. The Arrival Rate has the largest effect and is difficult to control, but should be a poisson distribution at 15 people per 30 minutes. The results of this study met the objective by screening through 63 model terms to obtain significant terms and a regression equation with high values for the Coefficients of Determination. The regression equation was then used to determine the minimal Average Waiting Time to meet the second half of the objective.

Final Remarks

The first outcome of this study showed that the difference between the model summary of each model was negligible. This may be due to the large degrees of freedom present which prevented the effect of noise from insignificant terms. Regardless, a model with clearly insignificant terms is questionable and should be re-fitted per a screening process. The lack-of-fit supports the re-fitted model with a high P-Value to assure that there are no available terms unpresented that have a significant effect on Average Waiting Time. The outcome of the analysis of the residuals from the re-fitted model proved to not be Normally Independently Distributed. But there were approxiamtely10 responses, out of 2432 responses, that were outliers as seen by the Normal Probability Plot and Histogram that surely were the reason for the violation on normality. The outcome of the response optimizer resulted in a 95% CI with a small width. The chosen values for each design factor were set at one of their two levels which indicates the potential for the true optimal to exist outside the bounds of the design area.

This study has the potential for improvement by incorporating center points for calculating squared power terms. This would be useful for increasing the values of the Coefficients of Determination, especially with plenty of degrees of freedom left over to use in the model. Due to the results of the response optimization choosing a corner point, a response surface design would be worth considering especially with optimality as an objective. A central composite design with an axial length > 1 would allow for movement among the greater design region to improve the solution. The final thought for improvement would be to conduct a time study on the actual arrival process and tour length, weather manually, through automation, or from a data base. This would allow for a move accurate mean, variance, and distribution identification of Arrival Rate and Mean Process Time.

Works Cited

"Our 2015 Season Has Begun!" *Boston Duck Tours Splash Page*. Web. 21 Mar. 2015. <http://www.bostonducktours.com/>.

"Ride The Ducks of Seattle | It's a Party on Wheels and in the Water!" *Ride The Ducks of Seattle*. Web. 21 Mar. 2015. <http://www.rideteducksofseattle.com/>.

Appendix

Assumptions

1. There will be 3 parallel processes to represent each starting location.
2. Resources are divided equally amongst the locations.
3. Resources are not allowed to be shared between locations.
4. Costumers purchase a ticket on the same day they receive service.
5. Costumers purchase a ticket for the next available tour.
6. Costumers are not allowed to switch starting locations.
7. No more than 600 customers for each location can buy a ticket per day, based on boat capacity and available online time slots found online.
8. Arrival Rate follows a Poisson Distribution
9. Down Time follows a Discrete Distribution
10. Mean Process Time follows a Normal Distribution or Exponential Distribution

Table 4

808 Appendix Tables

Table A.11 Critical Values for the Ryan–Joiner Test of Normality

		α		
		.10	**.05**	**.01**
n	**5**	.9033	.8804	.8320
	10	.9347	.9180	.8804
	15	.9506	.9383	.9110
	20	.9600	.9503	.9290
	25	.9662	.9582	.9408
	30	.9707	.9639	.9490
	40	.9767	.9715	.9597
	50	.9807	.9764	.9664
	60	.9835	.9799	.9710
	75	.9865	.9835	.9757