NYU TANDON SCHOOL OF ENGINEERING

Project 1

Business Analysis _ MG-GY 9753

Author:

Menghe Dou

Yichun Zhou

Yichi Zhang

Professor:

Dr. Gazi Md Daud Iqbal

February 18, 2020

We *Menghe Dou, Yichun Zhou, Yichi Zhang* did not give or receive any assistance on this project from other groups, and the report submitted is wholly by us.

Signature:_____

Section I

- ***Data collection***

Data Set 1(attached in Appendices I):

The quiz grades of one physics class, which includes more than 200 students.

Data Set 2(attached in Appendices II):

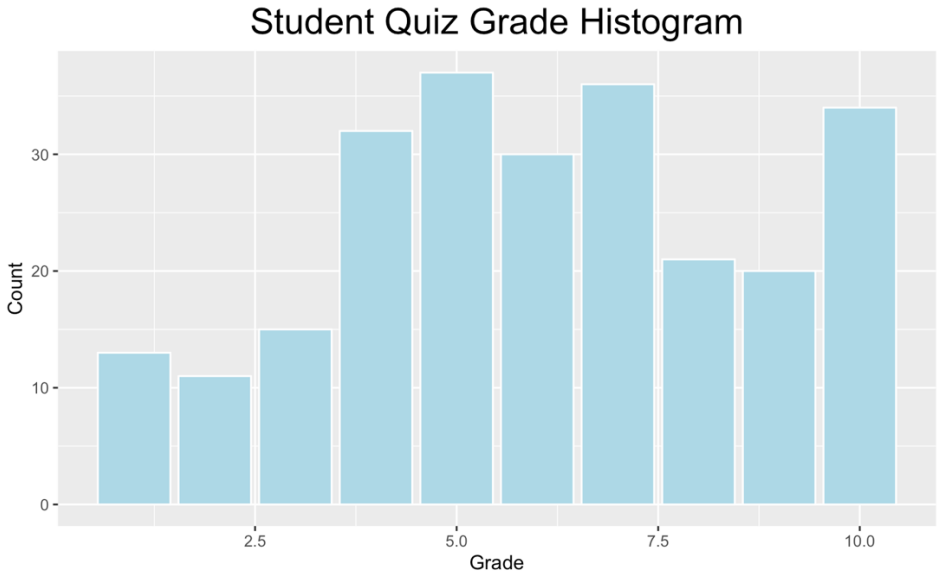
The interarrival time of customers in Five Guys at MetroTech during 13:00 to 18:00 on Feb. 16th 2020 (Sunday).

Section II

- ***Descriptive statistics analysis***

Data Set 1:

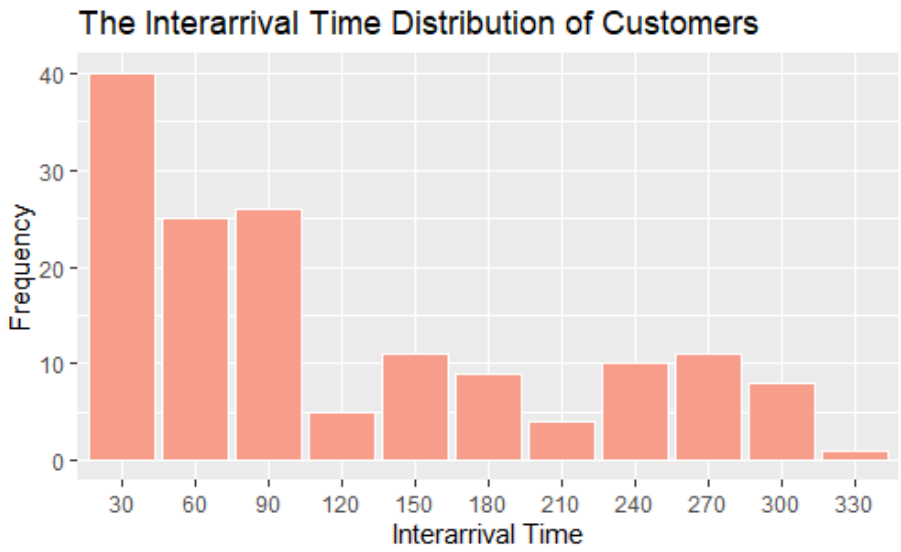
Min	0.000
1st Quartile	4.000
Median	6.000
Mean	6.042
3rd Quartile	8.000
Max	10.000



pic1. Histogram of data set1

Data Set 2:

Mean	103.1013
Standard Deviation	91.69912
1st Quartile	28.2675
Median	65.1200
3rd Quartile	165.4725



Pic2. Histogram of data set2

Section III

- ***Goodness-of-Fit Tests:***

  Data Set 1:

  H0: the Student Quiz Grade dataset follows the normal distribution

  H1: the Student Quiz Grade dataset does not follow the normal distribution

Lower	Upper	Bin	Frequency o_i	e_i
0	1	1	13	6.993059
1	2	2	11	8.666121
2	3	3	15	15.369981
3	4	4	32	23.656661
4	5	5	37	31.598726
5	6	6	30	36.628976
6	7	7	36	36.848571
7	8	8	21	32.170457
8	9	9	20	24.374352
9	10	10	34	32.693096

$$e_i = n * [\frac{1}{\sqrt{2\pi}\sigma}\left(e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} - e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2}\right)]$$

$n = 249$

$df = 9$

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = 15.59835$$

$\chi^2_{.05,9} = 16.91898$

$$\chi^2 < \chi^2_{.05,9}$$

Conclusion:
Chi_square_value < Test_statistics, ***Fail to reject H0.*** *Therefore, this dataset follows normal distribution.*

Data Set 2:
H0: the interarrival time follows exponential distribution
H1: the interarrival time does not follow exponential distribution

Lower	Upper	Bin	Frequency	e_i	Revised data	
0	30	1	40	37.869964	o_i	e_i
30	60	2	25	28.309069	40	37.869964
60	90	3	26	21.161980	25	28.309069
90	120	4	5	15.819290	26	21.161980
120	150	5	11	11.825451	5	15.819290
150	180	6	9	8.839922	11	11.825451
180	210	7	4	6.608138	9	8.839922
210	240	8	10	4.939805	14	11.547943
240	270	9	11	3.692670	11	3.692670
270	300	10	8	2.760395	9	6.366409
300	330	11	0	2.063488		
330	360	12	1	1.542527		

$e_i = n * (e^{-\lambda_2 x} - e^{-\lambda_1 x})$
$n = 150$
$df = 8$
$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = 25.14321$
$\chi^2_{.05,8} = 12.59159$

$$\chi^2 > \chi^2_{.05,8}$$

Conclusion:
Chi_square_value > Test_statistics, ***Reject H0.*** *And this dataset does not follow exponential distribution.*

Appendices I

Grade						
0	5	7	10	5	7	9
5	5	6	8	7	7	8
9	7	10	10	5	4	9
7	5	6	10	8	7	7
6	5	7	10	3	8	7
10	7	6	4	4	7	4
8	7	8	3	5	6	8
10	3	10	1	5	6	4
8	10	5	10	4	6	6
0	2	9.5	4	8	7	5
7	1	8	0	5	5	2
3	8	6	10	2	5	5
4	7	3	7	9	5	7
10	7	10	3	6	6	6
10	10	10	3	5	7	4
4	10	3	10	4	5	5
5	8	3	10	7	7	5
9	7	1	9	5	5	8
6	5	10	10	5	6	6
9	10	2	5	7	5	6
4	8	10	9	5	7	5
0	9	10	9	4	6	3
4	2	9	9	4	2	4
8	6	10	3	5	3	4
7	4	6	4	5	10	3
4	7	4	4	4	5	0
7	4	0	4	4	7	3
6	10	6	4	2	10	4
7	6	7	0	6	10	0
7	8	7	4	8	6	6
2	8	10	9	1	9	7
9	8	1	9	4	7	5
6	10	6	4	9	9	2
6	9	10	5	2	8	
7	8	9	3	4	10	
6	5	2	5	6	5	

Appendices II

Customer	Time						
1	60.2	39	31.4	77	86.19	115	180.98
2	46.8	40	19.9	78	138.68	116	54.9
3	10.02	41	226.05	79	28.47	117	288.55
4	45.99	42	237.41	80	42.88	118	77.14
5	56.32	43	181.24	81	2.22	119	23.09
6	31.7	44	218.44	82	181.95	120	60.47
7	51.02	45	2.39	83	257.2	121	14
8	18.8	46	2.04	84	211.4	122	4.56
9	64.58	47	17.28	85	63.27	123	235.71
10	49.29	48	191.14	86	109.31	124	2.8
11	2.6	49	155.73	87	137.17	125	2.36
12	81.92	50	64.9	88	71	126	3.57
13	2.14	51	150.53	89	253.03	127	280.4
14	19.06	52	31.78	90	31.65	128	259.57
15	72.84	53	62.99	91	23.36	129	54.25
16	30.89	54	2.65	92	60.9	130	6.2
17	139.5	55	15.91	93	242.56	131	2.97
18	28.2	56	22.31	94	56.62	132	283.29
19	2.8	57	23.31	95	2.94	133	351.17
20	278.92	58	21.53	96	284.48	134	87.25
21	62.88	59	36.52	97	55.53	135	38.57
22	86.37	60	215.96	98	268.92	136	285.23
23	3.46	61	105.98	99	230	137	47.05
24	268.7	62	26.6	100	214.31	138	155.85
25	57.6	63	137.98	101	3.33	139	265.25
26	125.41	64	49.39	102	59.15	140	152.36
27	135.92	65	274.78	103	102.86	141	64.54
28	252.49	66	257.66	104	139.92	142	54.15
29	230.99	67	64.56	105	169.24	143	107.48
30	65.34	68	72.59	106	69.55	144	70.66
31	36.64	69	28.78	107	18.85	145	43.96
32	54.58	70	129.81	108	20.96	146	2.93
33	268.12	71	178.7	109	82.35	147	1.18
34	78.69	72	10.35	110	129.3	148	26.2
35	138.32	73	66.83	111	171.51	149	17.4
36	166.79	74	110.72	112	23.31	150	161.52
37	146.09	75	77.25	113	237.03		
38	279.43	76	77.16	114	264.27		

Appendices III

Related codes & results in R

```
library(readxl)
setwd("/Users/joezhou/Desktop/2020 Spring/BA/project1") data <- read_excel ("clean_data.xlsx")
summary(data)
```

Descriptive Statistics Analysis

```
mean <- mean(data$Grade)
sd <- sd(data$Grade)
quantile <-quantile(data$Grade) mean
```

Including Plots

You can also embed plots, for example:

```
library(ggplot2)
ggplot(data, aes(x=Grade)) + geom_histogram(binwidth = 1,color="white", fill="lightblue") +

ylab('Count') +
ggtitle('Student Quiz Grade Histogram') + theme(plot.title = element_text(hjust = 0.5,size = 20))
```

Chi-square Tests

```
print('H0: the Student Quiz Grade dataset follows the normal distribution')

print('H1: the Student Quiz Grade dataset does not follow the normal distribution')
```

```
chi_matrix <- matrix(c(0,1,2,3,4,5,6,7,8,9,1,2,3,4,5,6,7,8,9,10),nrow=10,ncol=6) chi_matrix[1,3] <-
length(which(data >=0 & data <=1))

for (i in 2:10){
chi_matrix[i,3] <- length(which(data > i-1 & data <= i))

}
chi_matrix[,3]

hist.df <- as.data.frame(chi_matrix)
#hist.df[2]
count = hist.df[3]

chi_matrix[1,4] <- (1 - pnorm(chi_matrix[1,2], mean, sd,lower.tail = FALSE)) for (i in 2:9){

chi_matrix[i,4] <- (1 - pnorm(chi_matrix[i,2], mean, sd,lower.tail = FALSE) -( 1- pnorm( }

chi_matrix[10,4] <- pnorm(chi_matrix[9,2], mean, sd,lower.tail = FALSE) chi_matrix[,4]
```

```
for (i in 1:10){
chi_matrix[i,5] <- length(data$Grade) * (chi_matrix[i,4])

}
chi_matrix[,5]

for (i in 1:10){
chi_matrix[i,6] <- ((chi_matrix[i,3] -chi_matrix[i,5] )**2)/chi_matrix[i,5]

}
chi_matrix[,6]
Chi_square_value <- sum(chi_matrix[,6])
Chi_square_value

Test_statistics <- qchisq(.95, length(chi_matrix[,1])-1)
Test_statistics

print ('Chi_square_value < Test_statistics, Fail to reject H0. And this dataset follows normal distribution.')
```