<u>A Statistical Analysis of the Number of Ingredients in Baking Recipes and the Interval Arrival Times of UTA Shuttles Arriving at the UC Stop</u>

IE3301 Fall 2024

Prepared by Preston Loera and Mary-Rose Tracy

“I, Preston Loera, did not give or receive any assistance on this project, and the report submitted is wholly my own.”

“I, Mary-Rose Tracy, did not give or receive any assistance on this project, and the report submitted is wholly my own.”

Table of Contents:

List of Figures:

List of Tables:

Section 1: Data

1.1 Introduction

The purpose of this project is to conduct a statistical analysis of two real-world data sets: the number of ingredients in baking recipes and the interval arrival times of UTA shuttles at the UC stop. These data sets are chosen to explore patterns and distributions in everyday contexts, one concerning food preparation and the other concerning transportation. The first data set, focusing on recipes, allowing us to study variability in the number of ingredients based on the complexity of the dish, while the second data set provides insights into the timing and efficiency of shuttle service.This study will utilize statistical tools to describe and interpret the data, and ultimately analyze whether these data sets follow expected patterns.

1.2 Methodology

To collect the data for this study, two distinct processes were followed. For the first data set, we gathered information on the number of ingredients used in a variety of baking recipes from 101cookbooks [3]. This involved sampling at least 100 unique recipes ensuring that each recipe was diverse in terms of type and complexity. For the second data set, we recorded the exact arrival times of UTA shuttles at the UC stop, using a stopwatch to capture the time to the nearest second. The intervals between consecutive shuttle arrivals were calculated by subtracting the arrival times of each successive shuttle, providing a continuous data set of 99 inter-arrival times.

1.3 Execution

Data collection for the number of ingredients in baking recipes involved manually counting recipes from *101cookbooks* [3]. Each recipe's ingredients were counted and logged in an excel spreadsheet, ensuring that the data was varied and not limited to a single category of baked goods. For more complex recipes such as cakes, these recipes can have multiple parts to them such as a filling or icing. During counting if the separate parts shared common ingredients such as butter, the count would include both instances of butter to the count. This decision was made to be more accurate to real life usage of ingredients. Additionally during collection there were ingredients labeled as "optional" for some recipes logged. These optional ingredients were not included in the final count. For the UTA shuttle arrival times, we stationed ourselves at the UC stop on September 27th at 10:00 AM and recorded the exact time each shuttle arrived and came to a full stop over the course of seven hours, taking care to note any factors that might affect the arrival intervals, such as delays in shuttle arrival times or multiple shuttles arriving simultaneously. Data collection took place over multiple hours instead of multiple days to ensure the reliability and accuracy of the observed time intervals.

1.4 Calculations

To prepare the data for analysis, we performed several key calculations for both data sets. The number of ingredients in each recipe was recorded as discrete values for set one, and for the UTA shuttle data, the inter-arrival times were computed by subtracting the previous shuttle's arrival time from the current one. These intervals formed the basis of our second data set. In both cases, the descriptive statistics were calculated to summarize the data, including measures such as the mean, standard deviation, and quartiles.

1.4.2. Cookbook ingredient values

The calculation for the number of ingredients was analyzed by a python program [1] to be able to get the descriptive statistics from the 100 data values. From the program the descriptive statistics, histogram, and boxplot were generated and outputted.

1.4.1. Shuttle Turnover Time

The calculation of turnover time for UTA shuttles focused on the intervals between arrivals. By converting the recorded times into seconds and finding the difference between successive arrivals, we were able to determine how frequently the shuttles arrived at the stop. This metric is crucial for understanding the service efficiency and identifying patterns in shuttle scheduling. The difference in vehicle arrivals, calculated as the time difference between successive shuttles, provided us with an important continuous data set. Each time difference was logged and analyzed by a python program [2] to understand variability in the shuttle service. These inter-arrival times allowed us to explore whether the timing followed an exponential distribution, as is common in arrival time studies for transportation services.

Section 2: Descriptive Statistics

2.1 Data Set 1: Number of Ingredients in Baking Recipes

2.1.1 Descriptive Statistics

The first data set consists of the number of ingredients in various baking recipes, which was analyzed to provide a comprehensive statistical summary. The mean number of ingredients across the sampled recipes is calculated to be 9 with a standard deviation of 3.6735, indicating the average deviation from the mean. The quartiles (Q1, Q2, and Q3) were also calculated, showing that the median number of ingredients is 10. These statistics provide insight into the overall complexity of the baking recipes sampled, with the interquartile range (IQR) highlighting the spread of the central 50% of the data.

```
==============Discrete Data Set==============
Num Data in list (Ignore this) 100
Range of number of ingredients (Ignore this) 17
Mean:  9
Median:  10.0
Mode:  12
Standard Deviation:  3.6735472631979973
25th percentile of data:  6.0
50th percentile of data:  10.0
75th percentile of data:  12.0
```

Figure 1: Descriptive Statistics for Data Set 1

Figure 1 presents the descriptive statistics for Data Set 1, summarizing the core metrics such as the mean, median, and quartiles.

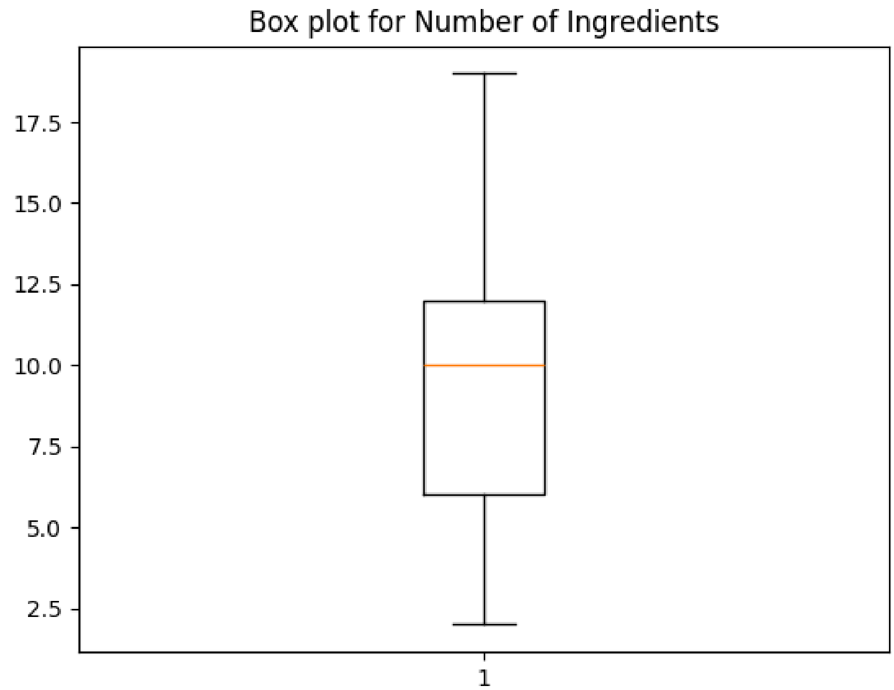

Figure 2: Box and Whisker Plot for Data Set 1

In addition, Figure 2 shows the box-and-whisker plot for Data Set 1, which visualizes the distribution of the number of ingredients. The box plot indicates that the data is slightly skewed, with a few outliers representing recipes that use an unusually high or low number of ingredients. The whiskers extend to the minimum and maximum values, providing a clear picture of the spread of the data.

2.1.2 Data Set 1 Frequency Distribution

Bins	Frequency	Relative Freq.	Cumulative Freq.
[2.00, 3.89)	4	0.040	0.040
[3.89, 5.78)	13	0.130	0.170
[5.78, 7.67)	17	0.170	0.340
[7.67, 9.56)	15	0.150	0.490
[9.56, 11.44)	20	0.200	0.690
[11.44, 13.33)	18	0.180	0.870
[13.33, 15.22)	7	0.070	0.940
[15.22, 17.11)	4	0.040	0.980
[17.11, 19.00)	2	0.020	1.000

```
PS E:\Python_Code\IE_PROJECT>
```

Figure 3: Frequency Distribution table for Set 1

The frequency distribution for the number of ingredients was calculated to further explore the pattern of ingredient use across the sampled recipes. Figure 3 displays the frequency histogram for Data Set 1, showing that the majority of recipes use 7 ingredients, with 17 of the recorded recipes containing exactly 7 ingredients. The distribution is slightly left-skewed, as indicated by the relationship between the mean, median, and mode (mean < median < mode), which suggests a tendency toward fewer ingredients in some recipes.

This left-skewed distribution reflects the variability in the complexity of baking recipes. While most recipes fall around the 7-ingredient mark, there are some simpler recipes with fewer ingredients and more complex ones with a larger number. It is possible that with a larger sample size, the data might trend towards a more normal distribution, as baking recipes tend to have a central tendency around a typical range of ingredients.

In addition to the frequency histogram, the relative frequencies and cumulative frequencies were calculated to understand the proportions of recipes falling within specific ingredient ranges. These provide further insight into the general complexity of the recipes and how ingredient counts are distributed across the data set.

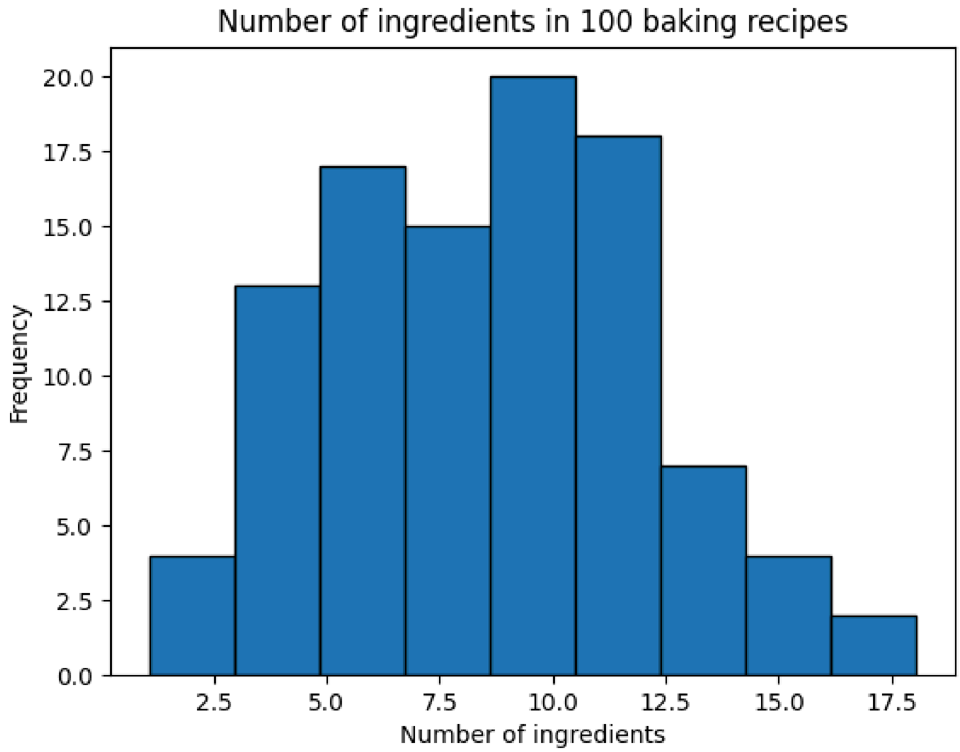

Figure 4: Frequency Histogram of Data Set 1

2.2 Data Set 2: UTA Shuttle Inter-Arrival Times

2.2.1: Descriptive Statistics

```
==============Continuous Data Set==============
Num Data in list (Ignore this) 100
Mean:  03:41
Median:  02:58
Mode:  0
Standard Deviation:  03:04
25th percentile of data:  01:07
50th percentile of data:  02:58
75th percentile of data:  05:13
     Bins       Frequency   Relative Freq.   Cumulative Freq.
[ 00:00, 01:41)     33         0.330            0.330
[ 01:41, 03:23)     23         0.230            0.560
[ 03:23, 05:04)     18         0.180            0.740
[ 05:04, 06:46)     11         0.110            0.850
[ 06:46, 08:27)      7         0.070            0.920
[ 08:27, 10:09)      4         0.040            0.960
[ 10:09, 11:50)      2         0.020            0.980
[ 11:50, 13:32)      1         0.010            0.990
[ 13:32, 15:14)      1         0.010            1.000
PS E:\Python_Code\IE_PROJECT>
```

Figure 5: Descriptive Statistics for Data Set 2

Figure 4 provides a summary of the descriptive statistics for Data Set 2, offering a numerical snapshot of the shuttle arrival behavior. The second data set, consisting of UTA shuttle inter-arrival times, was analyzed to assess the variability in shuttle scheduling. The mean inter-arrival time between shuttles is 03:41, with a standard deviation of 03:04, indicating some inconsistency in the shuttle intervals. The quartiles reveal that the median inter-arrival time is 02:58, and the interquartile range (IQR) demonstrates that the central 50% of the data falls between 1:07 and 5:13. This spread shows a wide range of inter-arrival times, suggesting that shuttle arrivals are not uniformly spaced.

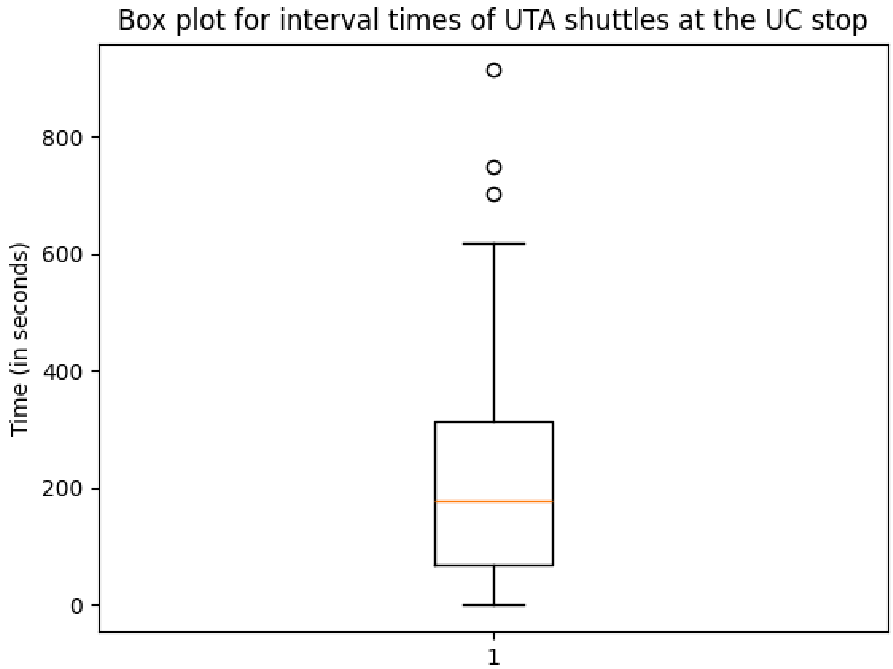

Figure 6: Box and Whisker Plot for Data Set 2

The box-and-whisker plot in Figure 5 illustrates a right-skewed distribution, as indicated by several outliers representing unusually long intervals between shuttle arrivals. Most data points fall below 230 seconds, showing that the majority of shuttle intervals are less than the mean of 03:41. The yellow line on the box plot represents the mean, while the box itself indicates the spread of the central 50% of the data, or the interquartile range. This plot emphasizes that while shuttles generally arrive with moderate regularity, there are occasional long gaps that push the distribution to the right.

2.2.2 Data Set 2 Frequency Distribution

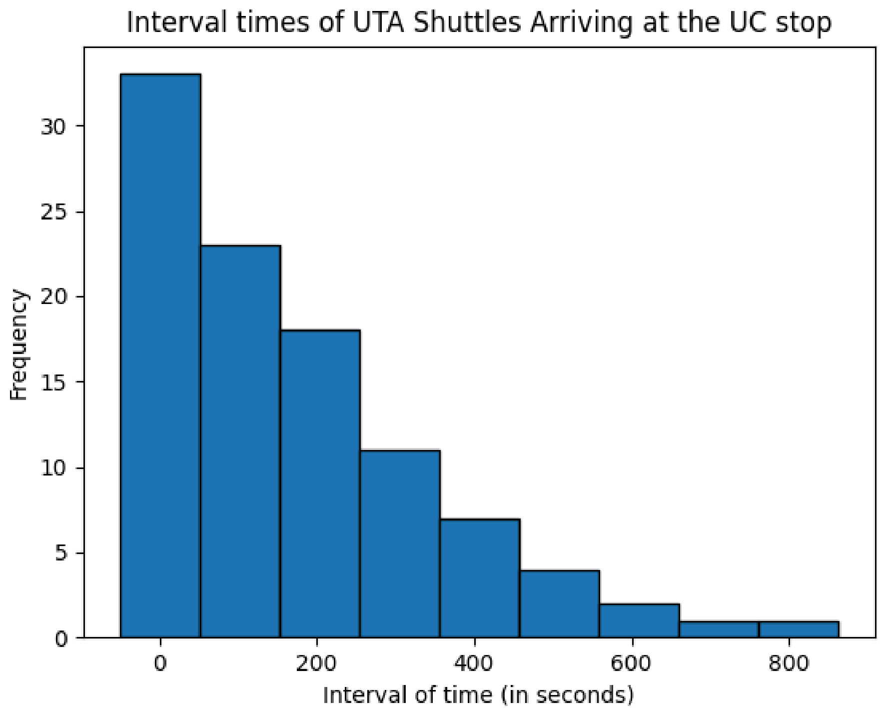

Figure 7: Frequency Histogram for Data Set 2

Figure 6 displays the frequency histogram for the inter-arrival times of UTA shuttles. The histogram shows that most shuttles arrive within an average time of 03:41, with wait times varying based on the standard deviation of 03:04. This means that while most passengers experience a wait time around the mean, some may experience shorter or longer waits due to this variability.

The inter-arrival times follow an exponential distribution, as expected for arrival-time data. This distribution suggests high variability in shuttle scheduling, with a few long gaps between shuttles contributing to longer wait times. External factors, such as traffic congestion, accidents, or other disruptions like ambulance activity, could influence these longer-than-average wait times. As a result, while most shuttles arrive within a reasonable timeframe, occasional delays push the overall distribution towards longer intervals, creating a right-skew in the data.

Section 3: Goodness-of-Fit Tests

3.1 Data Set 1: Chi-Square Goodness-of-Fit Test for Normal Distribution

To determine whether the number of ingredients in baking recipes follows a Normal Distribution, a Chi-Square Goodness-of-Fit Test was conducted. The null hypothesis (H0) posited that the data follows a Normal Distribution with a mean of 9 and a standard deviation of 3.6735, based on the descriptive statistics calculated in Part 1. The alternative hypothesis (H1) proposed that the data does not follow a Normal Distribution.

The data was divided into seven classes as determined by the histogram, ensuring that each class contained enough observations for reliable chi-square analysis. For each class, observed frequencies were recorded, and expected frequencies were calculated using the cumulative distribution function (NORM.DIST) in Excel. The chi-square components, calculated as (Oi−Ei)2/Ei were summed to produce the total chi-square statistic.

The results of the test indicated a chi-square statistic of 5.71, with 7 degrees of freedom. At a significance level of 0.05, the critical value was 11.07. Since the chi-square statistic was less than the critical value, we failed to reject the null hypothesis, concluding that the number of ingredients in baking recipes likely follows a Normal Distribution. This conclusion aligns with the observed central tendency and variability in the data, as seen in the descriptive statistics. Detailed calculations and Excel formulas can be found in **Appendix VI**.

3.2 Data Set 2: Chi-Square Goodness-of-Fit Test for Exponential Distribution

The second data set, comprising inter-arrival times of UTA shuttles, was tested to determine whether it follows an Exponential Distribution. The null hypothesis (H0) proposed that the data follows an Exponential Distribution with a mean equal to the sample mean (μ=221.8). The alternative hypothesis (H1) suggested that the data does not follow an Exponential Distribution.

For the analysis, the data was divided into six classes based on the histogram. Observed frequencies were recorded, while expected frequencies were calculated using the cumulative distribution function (GAMMA.DIST) in Excel. As with the first test, chi-square components were calculated for each class and summed to determine the total chi-square statistic.

The test produced a chi-square statistic of 2.09, with 6 degrees of freedom. At a significance level of 0.05, the critical value was 9.49. As the chi-square statistic was less than the critical value, we failed to reject the null hypothesis, concluding that the inter-arrival times of UTA shuttles are consistent with an Exponential Distribution. This finding is consistent with expectations for inter-arrival time data, which often exhibit memoryless properties and variability. Full calculations and formulas are included in **Appendix VII**.

Section 4: Conclusion

This study analyzed two real-world data sets: the number of ingredients in baking recipes and the inter-arrival times of UTA shuttles. Through descriptive statistics and Chi-Square Goodness-of-Fit Tests, we evaluated whether these data sets followed hypothesized distributions.

For the first data set, the results supported the hypothesis that the number of ingredients follows a Normal Distribution. The bell-shaped distribution aligns with the central tendency and variability observed in the descriptive statistics, suggesting that most recipes fall within a typical range of ingredient counts. This finding provides a basis for further analyses, such as estimating ingredient usage trends or average costs for baking.

For the second data set, the inter-arrival times of UTA shuttles were found to follow an Exponential Distribution, consistent with the variability expected in transportation scheduling. While most intervals fell within a moderate range, occasional delays extended the distribution's tail, reflecting real-world factors like traffic and operational disruptions. These findings can be used to model and optimize shuttle schedules, improving service reliability.

Overall, this analysis demonstrates the utility of statistical methods in interpreting real-world data, while also highlighting areas for further research. Expanding the sample size or incorporating additional variables, such as time of day or external conditions, could refine these conclusions and support more precise applications.

Appendix Ⅰ:

```
import matplotlib.pyplot as plt
import statistics
import numpy as np

print("=============Discrete Data Set=============")
numberOfIngredients = [3, 9, 18, 11, 12, 5, 19, 6, 11, 8, 13, 13, 3, 6, 16,
7, 12, 9, 14,                14, 11, 7, 6, 11, 12, 6, 8, 5, 12, 14, 8, 10, 8,
15, 11, 5, 9, 5, 9, 12, 16, 6, 16, 10, 6, 16, 13, 15, 5, 5, 10, 10, 5, 11,
11, 12, 6, 6, 13, 10, 2, 13, 8, 8, 6, 12, 8, 5, 4, 9, 15, 3, 12, 12, 7, 12,
11, 10, 6, 11, 14, 12, 11, 10, 7, 10, 4, 10, 5, 10, 12, 7, 9, 5, 6, 12, 8, 5,
8, 6]

print("Num Data in list (Ignore this)",len(numberOfIngredients))
print("Range of number of ingredients (Ignore this)",max(numberOfIngredients)
- min(numberOfIngredients))

mean = round(statistics.mean(numberOfIngredients))
median = statistics.median(numberOfIngredients)
mode = statistics.mode(numberOfIngredients)
std =  statistics.stdev(numberOfIngredients)
print("Mean: ",mean)
print("Median: ",median)
print("Mode: ",mode)
print("Standard Deviation: ", std)

#Quartiles
print("25th percentile of data: ", np.percentile(numberOfIngredients, 25))
print("50th percentile of data: ", np.percentile(numberOfIngredients, 50))
print("75th percentile of data: ", np.percentile(numberOfIngredients, 75))

plt.title("Box plot for Number of Ingredients")
plt.boxplot(numberOfIngredients)
plt.show()

#Histogram and Frequency Data
frequency, bins, patches = plt.hist(numberOfIngredients, bins = 9, ec =
'black', align = 'left')

#Continued to the next page.
#(Not in original program comment is here for sake of the reader)
```

```
#Convert to a list we are able to iterate over and grab data (returned as
np.array)
b = bins.tolist()
f = frequency.tolist()

print('     Bins      Frequency   Relative Freq.   Cumulative Freq.')
prev = 0
for x in range(len(b) - 1):
    relativeFreq = f[x]/len(numberOfIngredients)
    cumulativeFreq = prev + relativeFreq
    print("[ %2.2f, %2.2f)     %2d           %.3f             %.3f" % (b[x],
b[x + 1], f[x], relativeFreq, cumulativeFreq))
    prev = cumulativeFreq

plt.title("Number of ingredients in 100 baking recipes")
plt.xlabel("Number of ingredients")
plt.ylabel("Frequency")
plt.show()
```

Figure 8: Python Program used to get Descriptive Statistics for Set 1

Description:

The following program takes the number of recorded ingredients from 100 unique recipes and gets the descriptive statistics using libraries such as numpy and statistics. Matplotlib is used to be able to provide the images of the box plot, frequency table, and histogram. The following program is provided in a GitHub Gist in the link below.

Link to Gist of Discrete_Data_Code.py

Appendix Ⅱ:

```
import matplotlib.pyplot as plt
import statistics
import numpy as np
import scipy.stats

print("=============Continuous Data Set=============")
intervalTimesForUTAShuttle = [
    0, 95, 107, 284, 171, 110, 76, 392, 360, 63, 245, 338, 197, 195,
    619, 21, 70, 267, 37, 541, 44, 383, 188, 278, 48, 277, 76, 463,
    288, 63, 133, 595, 137, 56, 124, 452, 152, 0, 119, 152, 549, 43,
    78, 46, 207, 580, 14, 42, 703, 60, 326, 109, 246, 87, 33, 473,
    154, 245, 133, 313, 348, 31, 914, 18, 53, 346, 284, 285, 433,
    261, 149, 430, 33, 100, 500, 67, 49, 172, 244, 200, 315, 68,
    264, 302, 184, 196, 356, 17, 225, 51, 272, 459, 256, 62, 141,
    62, 748, 168, 122, 338
]

def convert(time):
    minutes = time // 60
    seconds = time % 60
    return '%02d:%02d' % (minutes, seconds)

print("Num Data in list (Ignore this)", len(intervalTimesForUTAShuttle))
mean = statistics.mean(intervalTimesForUTAShuttle)
median = statistics.median(intervalTimesForUTAShuttle)
mode = statistics.mode(intervalTimesForUTAShuttle)
std =  statistics.stdev(intervalTimesForUTAShuttle)
print("Mean: ",convert(mean))
print("Median: ",convert(median))
print("Mode: ",mode)
print("Standard Deviation: ", convert(std))

#Quartiles
print("25th percentile of data: ",
convert(np.percentile(intervalTimesForUTAShuttle, 25)))
print("50th percentile of data: ",
convert(np.percentile(intervalTimesForUTAShuttle, 50)))
print("75th percentile of data: ",
convert(np.percentile(intervalTimesForUTAShuttle, 75)))

#Continued to the next page.
#(Not in original program comment is here for sake of the reader)
```

```
plt.title("Box plot for interval times of UTA shuttles at the UC stop")
plt.boxplot(intervalTimesForUTAShuttle)
plt.ylabel("Time (in seconds)")
plt.show()

#Histogram and Frequency Data
frequency, bins, patches = plt.hist(intervalTimesForUTAShuttle, bins = 9, ec
= 'black', align = 'left')

#Convert to a list we are able to iterate over and grab data (returned as
np.array)
b = bins.tolist()
f = frequency.tolist()

print('     Bins      Frequency   Relative Freq.   Cumulative Freq.')

prev = 0
for x in range(len(b) - 1):
    relativeFreq = f[x]/len(intervalTimesForUTAShuttle)
    cumulativeFreq = prev + relativeFreq
    print("[ %s, %s)     %2d           %.3f             %.3f" %
(convert(b[x]), convert(b[x + 1]), f[x], relativeFreq, cumulativeFreq))
    prev = cumulativeFreq

plt.title("Interval times of UTA Shuttles Arriving at the UC stop")
plt.xlabel("Interval of time (in seconds)")
plt.ylabel("Frequency")
plt.plot(scipy.stats.norm.pdf(statistics.mean(intervalTimesForUTAShuttle),
statistics.stdev(intervalTimesForUTAShuttle)))
plt.show()
```

Figure 9: Python Program used to get Descriptive Statistics for Set 2

Description:

The following the python program was designed to analyze the continuous dataset recorded to generate the descriptive statistics, box plot, and histogram. The program takes the 99 recorded interval times in seconds and uses a function called “convert” to convert the time into a “MM:SS” format when printing to the terminal. The program utilizes libraries such as matplotlib, numpy, statistics, and scipy allowing for functions defined in these libraries to get mean, median, mode, etc. The following program can be found in the GitHub Gist Hyperlink below, to allow the reader to download and compile the program for themselves.

Link to Gist of Continuous_Data_Code.py

Appendix Ⅲ:

[101cookbooks about me]

[101cookbooks desserts]

Description:

The provided links are where the recipes for data set 1 were collected from. *101cookbooks* is a website created by Heidi Swanson and is a website used to be able to provide healthy and simple recipes on a website. Additionally the website also has recipes taken or adapted from cookbooks that Heidi Swanson personally owns. The website serves as a way to document culinary experiments and provide others with recipes for a variety of dishes without the need to buy cookbooks.

Appendix Ⅳ:

Number of Ingredients for Dessert Recipes	Number of Ingredients
Frozen Yogurt	3
Berry Pie	9
Yellow Cake with chocolate frosting	18
Strawberry Scones	11
Oatmeal Peanut Butter Cookies	12
Grapefruit Sorbet	5
Cheesecake Bars	19
Chocolate Cookies	6
Homemade Coconut Cream Pie	11
Grapefruit Curd with Ginger	8
Itsy Bitsy Chocolate Chip Cookies	13
Cranberry Cake	13
Sparkling Cranberries	3
Chocolate Energy Bites	6
Pumpkin Pie	16
Sicilian Pistachio Cookies	7
Brown Sugar Sandwich Cookies	12

Coconut Chocolate Pudding	9
Chocolate Bundt Cake	14
Oatmeal Muffins	14
Shaker Apple Pie	11
Shaker Lemon Pie	7
All-Butter Flaky Pie Crust	6
Vegan Berry Swirl Ice Cream	11
Mesquite Chocolate Chip Cookies	12
Candied Walnuts	6
Smoked Chocolate Mousse	8
Classic Shortbread Cookies	5
Chickpea Chocolate Chip Cookies	12
Gingerbread Cookies	14
Whole Vanilla Bean Cookies	8
Aran's Double Chocolate Buckwheat crinkle cookies	10
Swedish Rye Cookies	8
Middle Eastern Millionaires Shortbread	15

Rosemary Olive Oil Cake	11
Flourless Chocolate Cake	5
Walnut Nutmeg Butter Cake	9
Frosty Lime Sherbet	5
Turkish Coffee Chocolate Brownies	9
Coconut Rum Cake	12
Heidi's Coffee Cake Recipe	16
Tapioca Pudding	6
Black Sticky Gingerbread Cake	16
Summer Berry Crisp	10
Fresh Mint Chip Frozen Yogurt	6
Chocolate Fudge & Tahini Cake	16
Chocolate Dipped Biscotti	13
Chocolate Almond Swirl Cake	15
An Incredible No Bake Chocolate Cake	5
Super Swiss Meringues	5
Rosewater Shortbread Cookies	10
One Bowl Banana Bread	10

Rhubarb & Rosewater Syrup	5
Saffron Vanilla Snickerdoodle Cookies	11
Classic Berry Swirl Ice Cream	11
Strawberry Rhubarb Crumble	12
Roasted Strawberries	6
Glissade Chocolate Pudding	6
Buttermilk Berry Muffins	13
Healthful Double Chocolate Cookies	10
Two-ingredient Candied Citrus Lollipops	2
Triple Ginger Cookies	13
Fluffy Vanilla Nougat	8
4 o'clock No-bake Energy Bites	8
Simple Red Fruit Salad	6
Anzac Cookies	12
Quinoa Hemp Snack Balls	8
Broiled Saffron Dates	5
Black Berry Saffron Honey	4

Thinnest Oatmeal Cookies	9
Lemony Olive Oil Banana Bread	15
Lillet Buttermilk Shakes	3
Brown Butter Spice Cake	12
Whole Wheat Oatmeal Chocolate Chip Cookies	12
Quinoa Cloud Cookies	7
Chocolate Cherry Brownies	12
Bittersweet Chocolate Tart	11
Apple and Carrot Shortbread Recipe	10
Peanut Butter Krispy Treats	6
Old-Fashioned Blueberry Cake Recipe	11
Marathon Cookies	14
Breton Buckwheat Cake with Fleur de Sel	12
Yogurt Tartlets	11
Basic Chocolate Cake	10
Fantasy-ish Fudge	7

Unfussy Apple Cake	10
Caramel Apples	4
Nikki's Healthy Cookie	10
Coco Choco Clusters	5
Cherry Cobbler	10
Banana Chip Cookies	12
Macaroon Cherry Tart	7
Amazing Black Bean Brownie	9
Espresso Caramels	5
Hearst Castle Shortbread Cookie	6
Plum and Peach Crisp	12
Mexican Wedding Cookies	8
Reese's Cup Remix with Brazil Nut Butter	5
The Madame's Souffle	8
Peppermint Semifreddo	6

Table 1: Raw Data for Data Set 1

Appendix Ⅴ:

Bus #:	Time of Event:	Difference of times
Bus 1	10:25	0:00:00
Bus 2	10:26	0:01:35
Bus 3	10:28	0:01:47
Bus 4	10:33:06	0:04:44
Bus 5	10:35:57	0:02:51
Bus 6	37:47.0	0:01:50
Bus 7	39:03.0	0:01:16
Bus 8	10:45:35	0:06:32
Bus 9	10:51:35	0:06:00
Bus 10	10:52:38	0:01:03
Bus 11	10:56:43	0:04:05
Bus 12	11:02:21	0:05:38
Bus 13	11:05:38	0:03:17
Bus 14	11:08:53	0:03:15
Bus 15	11:19:12	0:10:19
Bus 16	11:19:33	0:00:21
Bus 17	11:20:43	0:01:10
Bus 18	11:25:10	0:04:27
Bus 19	11:25:47	0:00:37
Bus 20	11:34:48	0:09:01
Bus 21	11:35:32	0:00:44
Bus 22	11:41:55	0:06:23
Bus 23	11:45:03	0:03:08
Bus 24	11:49:41	0:04:38
Bus 25	11:50:29	0:00:48
Bus 26	11:55:06	0:04:37
Bus 27	11:56:22	0:01:16
Bus 28	12:04:05	0:07:43
Bus 29	12:08:53	0:04:48

Bus 30	12:09:56	0:01:03
Bus 31	12:12:09	0:02:13
Bus 32	12:22:04	0:09:55
Bus 33	12:24:21	0:02:17
Bus 34	12:25:17	0:00:56
Bus 35	12:27:21	0:02:04
Bus 36	12:34:53	0:07:32
Bus 37	12:37:25	0:02:32
Bus 38	12:37:25	0:00:00
Bus 39	12:39:24	0:01:59
Bus 40	12:41:56	0:02:32
Bus 41	12:51:05	0:09:09
Bus 42	12:51:48	0:00:43
Bus 43	12:53:06	0:01:18
Bus 44	12:53:52	0:00:46
Bus 45	12:57:19	0:03:27
Bus 46	13:06:59	0:09:40
Bus 47	13:07:13	0:00:14
Bus 48	13:07:55	0:00:42
Bus 49	13:19:38	0:11:43
Bus 50	13:20:38	0:01:00
Bus 51	13:26:04	0:05:26
Bus 52	13:27:53	0:01:49
Bus 53	13:31:59	0:04:06
Bus 54	13:33:26	0:01:27
Bus 55	13:33:59	0:00:33
Bus 56	13:41:52	0:07:53
Bus 57	13:44:26	0:02:34
Bus 58	13:48:31	0:04:05
Bus 59	13:50:44	0:02:13
Bus 60	13:55:57	0:05:13
Bus 61	14:01:45	0:05:48

Bus 62	14:02:16	0:00:31
Bus 63	14:17:30	0:15:14
Bus 64	14:17:48	0:00:18
Bus 65	14:18:41	0:00:53
Bus 66	14:24	0:05:46
Bus 67	14:29:11	0:04:44
Bus 68	14:33:56	0:04:45
Bus 69	14:41:09	0:07:13
Bus 70	14:45:30	0:04:21
Bus 71	14:47:59	0:02:29
Bus 72	14:55:09	0:07:10
Bus 73	14:55:42	0:00:33
Bus 74	14:57:22	0:01:40
Bus 75	15:05:42	0:08:20
Bus 76	15:06:49	0:01:07
Bus 77	15:07:38	0:00:49
Bus 78	15:10:30	0:02:52
Bus 79	15:14:34	0:04:04
Bus 80	15:17:54	0:03:20
Bus 81	15:23:09	0:05:15
Bus 82	15:24:17	0:01:08
Bus 83	15:28:41	0:04:24
Bus 84	15:33:43	0:05:02
Bus 85	15:36:47	0:03:04
Bus 86	15:40:03	0:03:16
Bus 87	15:45:59	0:05:56
Bus 88	15:46:16	0:00:17
Bus 89	15:50:01	0:03:45
Bus 90	15:50:52	0:00:51
Bus 91	15:55:24	0:04:32
Bus 92	16:03:03	0:07:39
Bus 93	16:07:19	0:04:16

Bus 94	16:08:21	0:01:02
Bus 95	16:10:42	0:02:21
Bus 96	16:11:44	0:01:02
Bus 97	16:24:12	0:12:28
Bus 98	16:27:00	0:02:48
Bus 99	16:29:02	0:02:02
Bus 100	16:34:40	0:05:38

Table 2: Raw Data for Data Set 2

<u>Appendix Ⅵ:</u>

Normal

Class	Frequency	Class Probability	Expected Value	Chi-square
x ≤ 3.89	4	0.082108761	8.210876123	2.159511051
3.89 < x ≤ 5.78	13	0.108260332	10.82603315	0.436552502
5.78 < x ≤ 7.67	17	0.168289033	16.82890326	0.001739513
7.67 < x ≤ 9.56	15	0.201922398	20.19223978	1.335134398
9.56 < x ≤ 11.44	20	0.186140925	18.61409252	0.103187386
11.44 < x ≤ 13.33	18	0.134019173	13.40191731	1.577562668
x > 13.33	13	0.119259379	11.92593786	0.096731133
Total	100	1	100	5.710418651

Class Probability Calculations	
x ≤ 3.89	NORMDIST(3.89, 9, 3.67354726319799, 1)
3.89 < x ≤ 5.78	NORMDIST(5.78, 9, 3.67354726319799, 1) - NORMDIST(3.89, 9, 3.67354726319799, 1)
5.78 < x ≤ 7.67	NORMDIST(7.67, 9, 3.67354726319799, 1) - NORMDIST(5.78, 9, 3.67354726319799, 1)
7.67 < x ≤ 9.56	NORMDIST(9.56, 9, 3.67354726319799, 1) - NORMDIST(7.67, 9, 3.67354726319799, 1)
9.56 < x ≤ 11.44	NORMDIST(11.44, 9, 3.67354726319799, 1) - NORMDIST(9.56, 9, 3.67354726319799, 1)
11.44 < x ≤ 13.33	NORMDIST(13.33, 9, 3.67354726319799, 1) - NORMDIST(11.44, 9, 3.67354726319799, 1)
x > 13.33	1 - NORMDIST(13.33, 9, 3.67354726319799, 1)

Hypothesis:

H0	Follows a Normal distribution
H1	Does not follow a Normal distribution

N = 100

Mean = 9

Std = 3.6735472631979973

Degrees of freedom = 7

Appendix Ⅶ:

Exponential

Class	Frequency	Class Probability	Expected Value	Chi-square
x ≤ 101.556	33	0.367370257	36.73702575	0.380144042
101.556 < x ≤ 203.111	23	0.232409351	23.24093514	0.002497737
203.111 < x ≤ 304.667	18	0.147029068	14.70290681	0.739365597
304.667 < x ≤ 406.223	11	0.093014962	9.301496152	0.310156052
406.223 < x ≤ 507.777	7	0.058844031	5.884403115	0.211500875
x > 507.777	8	0.10133233	10.13323303	0.449085021
Total	100	1	100	2.092749324

Class probability Calculations

x ≤ 101.556	GAMMADIST(101.555555555555, 1, 221.8,1)
101.556 < x ≤ 203.111	GAMMADIST(203.111111111111, 1, 221.8,1) - GAMMADIST(101.555555555555, 1, 221.8,1)
203.111 < x ≤ 304.667	GAMMADIST(304.666666666666, 1, 221.8,1) - GAMMADIST(203.111111111111, 1, 221.8,1)
304.667 < x ≤ 406.223	GAMMADIST(406.222222222222, 1, 221.8,1) - GAMMADIST(304.666666666666, 1, 221.8,1)
406.223 < x ≤ 507.777	GAMMADIST(507.777777777777, 1, 221.8,1) - GAMMADIST(406.222222222222, 1, 221.8,1)

x > 507.777	1 - GAMMADIST(507.777777777777, 1, 221.8,1)

Hypothesis:

H0	Follows a exponential distribution
H1	Does not follow a exponential distribution

N = 100

Mean = 221.8 (in seconds)

Std = 184.1370262724052 (in seconds)

Degrees of freedom = 6