REPORT

SOCIAL MEDIA ENGAGEMENT

*Submitted by*

**Nelson Joseph**

**29 MAY 2023**

Table of Contents

[Introduction 3](#_Toc136236281)

[Data Description 4](#_Toc136236282)

[Data Pre-processing 4](#_Toc136236283)

[Visualization of each columns subcategories vs data count 5](#_Toc136236284)

[Chi-Square Goodness-of-Fit Test for Data set 2 7](#_Toc136236285)

[Analysis on Dataset 9](#_Toc136236286)

[Engagement Rates 10](#_Toc136236287)

[Data set 1 10](#_Toc136236288)

[Appendix B 11](#_Toc136236289)

[Data set 2 11](#_Toc136236290)

[Appendix C 12](#_Toc136236291)

[Goodness of fit Calculations of Data set 1 12](#_Toc136236292)

[Goodness of fit Calculations of Data set 2 12](#_Toc136236293)

[References 13](#_Toc136236294)

# Introduction

The engagement rates of various post formats, including text, links, images, and videos, across multiple social media platforms were analysed in this paper.

The data was pre-processed to prepare it for analysis before being used in the analysis. To compare the engagement rates of various platforms, the data was then divided into subcategories. In particular, the engagement rates of well-known games like CSGO, DOTA 2, and Valorant on websites like YouTube, LinkedIn, Facebook, and Twitter were examined.

The analysis involves a careful assessment of the data, and the conclusions are explained in great depth in the pages that follow. The findings in this study seek to offer insightful data on the engagement rates of various post kinds.

# Data Description

The Goodness of Fit Test is used to determine whether the sample data chosen truly represents the data that is expected to be found in a population. The Chi-Square test is the goodness of fit test used in this project. The Chi-square goodness of fit test compares a randomly collected sample with a single categorical variable to a larger population. This test is most used to compare a random sample data to the population from which it was possibly drawn.

The test begins with the assumption of a **Null hypothesis (Ho)** and **Alternative hypothesis(H1).**

**Null Hypothesis (Ho):** The sample data collected is consistent with the population distribution.

**Alternative Hypothesis (H1):** The sample data collected is not consistent with the population distribution.

## Data Pre-processing

**χ2 =∑ (Oi – Ei)2/ Ei**

* **n** = Total number of observations or total frequency
* **χ2**= The **χ2 statistic** derived from the chi-square goodness of fit table.
* **Oi** = It is the observed value from the frequency table for each class interval.
* **Ei** = It is calculated by multiplying the class probabilities of each class by n.
* Each value of **(Oi - Ei)2/Ei** is referred to as a **" *χ2* Class Component."**
* Degree of Freedom **(v)** = number of groups - 1

There are several assumptions made by the Chi-Square test, including that the expected value of the observations should be more than 5, or else the groups should be combined appropriately. The Chi-Square test also presupposes that the level of significance for statistical analysis is set at 0.05

i.e., **α = 0.05**

Using the **significance level**(**α**) and **degree of freedom(v)**, the estimated **χ2 statistic** is compared with the **Tabulated value** that is derived from **Table A.5**.

* If **χ2 statistic** > **Tabulated value**

Conclusion: **Reject H0**

* **If χ2 statistic** < **Tabulated value**

Conclusion: **Fail to Reject H0**

## Visualization of each columns subcategories vs data count

The chi-square goodness of-Fit Test is performed using the descriptive statistical analysis results of Data set 1, the sample of 100 people's US shoe sizes. The Excel functions **AVERAGE (), MEDIAN (), STDEV.S (),** and **VAR.S ()** are used to calculate the sample mean, sample median, sample standard deviation, and sample variance, respectively.

|  |  |  |
| --- | --- | --- |
| **Statistics Value Units** | **Values** | **Units** |
| Sample Mean | 9.92 | US |
| Sample Median | 10.00 | US |
| Sample Variance | 3.57 | U*S* |
| Sample Standard Deviation | 1.89 | US |

Table 1.1 Descriptive Statistics Exploring the US shoe size of randomly selected 100 people within UTA Campus.

For the chi-square goodness of fit test, the sample mean and sample standard deviation are taken from **Table 1.1**.

Making a hypothesis regarding Sample Data Set 1 is the first stage of the test, which has a level of significance of 0.05.

**Null Hypothesis (H0):** The sample Data set 1 is sampled from Normal distribution with population mean equal to the sample mean and a population standard deviation equal to the sample standard deviation.

**Alternative Hypothesis (H1):** The sample Data set 1 is not sampled from Normal distribution with population mean equal to the sample mean and a population standard deviation equal to the sample standard deviation.

For performing the chi-square goodness of test on Data set 1, an expanded frequency table is constructed by adding the columns into 5 group intervals. Found class probability, Class expected value, and class chi-square components from the expanded frequency table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Observed Frequency(oi)** | **Class Probability** | **Expected value(ei)** | **χ2 Class Component** |
| X ≤ 8 | 23 | 0.154845 | 15.48 | 3.647703887 |
| 8 < X ≤ 10 | 41 | 0.362036 | 36.20 | 0.6354331503 |
| 10 < X ≤ 12 | 25 | 0.347568 | 34.76 | 2.738881448 |
| 12 < X ≤ 14 | 10 | 0.120115 | 12.01 | 0.3368633071 |
| X >14 | 1 | 0.015436 | 1.54 | 0.1914103344 |
| **Total** | 100 | 1.00 | 100.00 | 7.550292127 |

Table 1.2 Expanded Frequency table for performing Goodness of Fit Test on Data set 1

The observed Frequency is found by adding up the number of observations in the interval. The Class probability of each class intervals are calculated by:

The class probabilities are:

P [X ≤ 8] = 0.154845

P [8 < X ≤ 10] = P [X ≤ 10] - P [X ≤ 8] = 0.362036

P [10 < X ≤ 12] = P [X ≤ 12] - P [X ≤ 10] = 0.347568

P [12 < X ≤ 14] = P [X ≤ 14] - P [X ≤ 12] = 0.120115

P [X > 14] = 1 - P [X ≤ 14] = 0.015436

For the **Normal distribution**, the **sample mean (μ)** and the **sample standard deviation (σ)** are the two parameters taken to calculate the class probabilities. The excel formula **NORMDIST (x, μ, σ, 1)** is used to calculate **P [X ≤ x].**

Expected values are found by multiplying each class probabilities by total number of observations(n), here **n = 100**. But here the expected value in the class **X > 14** is less than 5, so the class is regrouped as below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Observed Frequency(oi)** | **Class Probability** | **Expected value(ei)** | **χ2 Class Component** |
| X ≤ 8 | 23 | 0.154845 | 15.48 | 3.647703887 |
| 8 < X ≤ 10 | 41 | 0.362036 | 36.20 | 0.6354331503 |
| 10 < X ≤ 12 | 25 | 0.347568 | 34.76 | 2.738881448 |
| X > 12 | 11 | 0.135551 | 13.56 | 0.4816236518 |
| **Total** | 100 | 1.00 | 100.00 | 7.503642138 |

Table 1.3 Expanded frequecny table with all the values with intervals combined.

P [X > 12] = 1 - P [X ≤ 12] = 0.135551

Chi-square class components are found by the chi-square formula **(Oi – Ei)2/ Ei** for each class and added up to get **χ2 test statistic, here the χ2 test statistic is** 7.503642138.

The Tabulated value for the Dataset 1 is found from **Table A.5,** Critical values of the Chi-Squared Distribution.

**Significance level(α)** = 0.05

**Degree of freedom(v) =** No of groups – 1 = 4-1 = 3

Therefore **χ2(α, v) = χ2(0.05, 3) = 7.815**

The **Tabulated value** is compared with **χ2 test statistic** to test our hypothesis.

Here the **χ2 test statistic < Tabulated value i.e.,** 7.503642138 < 7.815

**Conclusion: Fail to Reject Null Hypothesis(H0)**

Thus, a weak conclusion is obtained that Dataset 1 is sampled from Normal distribution with population mean equal to the sample mean and a population standard deviation equal to the sample standard deviation.

## Chi-Square Goodness-of-Fit Test for Data set 2

The chi-square goodness of-Fit Test is performed using the descriptive statistical analysis results of Data set 2, the samples of the time differences between two students entering the library through the main door. The Excel functions **AVERAGE (), MEDIAN (), STDEV.S (), QUARTILE (),** and **VAR.S ()** are used to calculate the sample mean, sample median, sample standard deviation, and sample variance, respectively.

|  |  |  |
| --- | --- | --- |
| **Statistics Value Units** | **Values** | **Units** |
| Sample Mean | **7.23** | S*econds* |
| Sample Median | **6.00** | S*econds* |
| Sample Variation | **17.82** | S*econds* |
| Sample Standard Deviation | **4.22** | S*econds* |

*Table 2.1 Descriptive Statistics Exploring the time interval between students entering the UTA Central Library*

**Table 2.1** displays the descriptive statistics values for the intervals between students entering the library. The 7.23 second obtained mean is used as **β** parameter for determining the class probabilities for each class.

The test is done with a level of significance of 0.05 and the first step is to make the hypothesis about the sample data set 2.

**Null Hypothesis (H0):** The sample data set 2 is sampled from an Exponential Distribution with a population mean equal to the sample mean.

**Alternative Hypothesis (H1):** The sample data set 2 is not sampled from an Exponential Distribution with a population mean equal to the sample mean.

For performing the chi-square goodness of test on Data set 2, an expanded frequency table is constructed by adding the columns into 6 group intervals. Found class probability, Class expected value, and class chi-square components from the expanded frequency table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Observed Frequency(oi)** | **Class Probability** | **Expected value(ei)** | **χ2 Class Component** |
| X ≤ 3 | 5 | 0.3396186 | 33.96 | 24.69798163 |
| 3 < X ≤ 6 | 56 | 0.2242778 | 22.43 | 50.25436362 |
| 6 < X ≤ 9 | 16 | 0.1481089 | 14.81 | 0.09546926 |
| 9 < X ≤ 12 | 16 | 0.0978084 | 9.78 | 3.95446870 |
| 12 < X ≤ 15 | **1** | 0.0645908 | 6.46 | 4.61390241 |
| X >15 | 6 | 0.1255955 | 12.56 | 3.42589529 |
| Total | 100 | 1 | 100.00 | 87.04208089 |

Table 2.2 Expanded frequency table for calculating Goodness of Fit Test on Data set 2

The observed Frequency is found by adding up the number of observations in the interval. The Class probability of each class intervals are calculated by:

The class probabilities are:

P [X ≤ 3] = 0.3396186

P [2 < X ≤ 6] = P [X ≤ 6] - P [X ≤ 3] = 0.2242778

P [6 < X ≤ 9] = P [X ≤ 9] - P [X ≤ 6] = 0.1481089

P [9 < X ≤ 12] = P [X ≤ 12] - P [X ≤ 9] = 0.0978084

P [12 < X ≤ 15] = P [X ≤ 15] - P [X ≤ 12] = 0.0645908

P [X > 15] = 1 - P [X ≤ 15] = 0.1255955

For the Exponential **distribution**, the **sample mean (μ)** is taken as β parameter to calculate the class probabilities. The excel formula GAMMADIST (x, **α**, β, 1) is used to calculate **P [X ≤ x]. α =1 as we are using the function for exponential distribution.**

Chi-square class components are found by the chi-square formula **(Oi – Ei)2/ Ei** for each class and added up to get **χ2 test statistic, here the χ2 test statistic is** 87.04208089.

The Tabulated value for the Dataset 2 is found from **Table A.5**, Critical values of the Chi-Squared Distribution.

**Significance level(α)** = 0.05

**Degree of freedom(v) =** No of groups - 1 = 6-1 = 5

Therefore **χ2(α, v) = χ2(0.05, 5) =** 11.070

The **Tabulated value** is compared with **χ2 test statistic** to test our hypothesis.

Here the **χ2 test statistic < Tabulated value i.e.,** 87.04208089 > 11.070

**Conclusion: Reject Null Hypothesis(H0)**

Thus, a make a strong conclusion is obtained that the sample Data set 2 is not sampled from an Exponential Distribution with a population mean equal to the sample mean.

# Analysis on Dataset

For data visualization and data presentation to be effective, descriptive statistics are essential. I conducted a statistical study on two datasets, starting with the Data set 1 that looked at 100 randomly selected students' US shoe sizes on November 15, 2022. The objective of this research was to check for any patterns in the 100 students whose shoe sizes were randomly selected. On November 15, 2022, between 3 and 4 PM, data for Data Set 2 was collected as the time difference between two students entering the library. Similar information may be acquired for other structures, allowing us to determine which ones draw students in. The assumptions that had previously been made based on the histogram, mean, median, and mode were put to the test using the chi-square goodness of Fit-Test. On Data Sets 1 and 2, the Goodness of Fit-Test was run to determine whether the samples were drawn from populations with normal or exponential distribution, respectively.

**Data set 1** was subjected to descriptive statistical analysis, and it was found that the average US shoe size is 9.92. Based on the mean shoe size of 9.92, mode 10, and median 10, the data set was approximated to a normal distribution earlier. The Chi-square goodness of Fit-Test on Data Set 1 also produced results that supported the initial hypothesis. The Data Set 1 is drawn from a population that is assumed to have a normal distribution, with sample mean equal to population mean and sample standard deviation equal to population standard deviation, as the result failed to reject the null hypothesis (H0). A tentative conclusion that the sample follows a Normal Distribution was drawn from the Goodness of Fit-Test, despite the sample data not being perfectly symmetrical with the mean centre.

For **Data Set 2**, descriptive statistics reveal that it takes a student, on average, 7.23 seconds to enter the library. The sample data appears to follow an exponential distribution in accordance with the previous analysis based on the histogram. The sample data had previously been divided into only four classes, which produced a frequency histogram that was exponentially declining. A non-exponential trend was observed in frequencies of each class as a result of the variations in class intervals count to 6 from 4. The findings of the Chi-square Goodness of Fit-Test benefited the more recent class interval observations. The null hypothesis (H0), that the sample data is drawn from a population that is expected to have an exponential distribution, was rejected as a result of the goodness of fit test. A strong conclusion that the sample Data Set 2 did not follow an exponential distribution was drawn from the goodness of fit test.

# Engagement Rates

## Data set 1

Below is a table of the raw data of data set 1 that was gathered with various values and used for the goodness of fit test and descriptive statistical analysis.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | **Race** | **Gender** | **US Shoe Size** |  | No | **Race** | **Gender** | **US Shoe Size** |
| 1 | Asian | Male | 10.5 |  | 52 | American | Male | 13 |
| 2 | Asian | Male | 12 |  | 53 | American | Female | 13 |
| 3 | Asian | Female | 8 |  | 54 | Asian | Male | 9.5 |
| 4 | American | Female | 9 |  | 55 | African American | Male | 10.5 |
| 5 | American | Male | 12 |  | 56 | Asian | Male | 9.5 |
| 6 | Asian | Female | 8 |  | 57 | American | Male | 10 |
| 7 | Asian | Male | 8.5 |  | 58 | Asian | Male | 11 |
| 8 | Asian | Female | 9 |  | 59 | American | Male | 11 |
| 9 | Asian | Female | 7.5 |  | 60 | American | Female | 11 |
| 10 | Asian | Male | 10 |  | 61 | American | Male | 12 |
| 11 | American | Female | 10 |  | 62 | American | Male | 13 |
| 12 | Asian | Female | 8 |  | 63 | Asian | Male | 10.5 |
| 13 | Asian | Male | 9 |  | 64 | Asian | Female | 7 |
| 14 | Asian | Female | 10.5 |  | 65 | Asian | Male | 10 |
| 15 | Asian | Male | 11 |  | 66 | Asian | Female | 7 |
| 16 | African American | Male | 11 |  | 67 | Asian | Female | 8 |
| 17 | American | Female | 12 |  | 68 | Asian | Female | 8 |
| 18 | Asian | Female | 6.5 |  | 69 | Asian | Female | 7 |
| 19 | Asian | Female | 7.5 |  | 70 | Asian | Female | 10 |
| 20 | Asian | Female | 7 |  | 71 | Asian | Female | 7.5 |
| 21 | Asian | Male | 14 |  | 72 | Asian | Male | 9.5 |
| 22 | Asian | Male | 13 |  | 73 | Asian | Male | 9.5 |
| 23 | American | Male | 12 |  | 74 | Asian | Female | 9 |
| 24 | American | Female | 10 |  | 75 | Asian | Male | 11 |
| 25 | Asian | Male | 10.5 |  | 76 | Asian | Male | 9 |
| 26 | American | Female | 11 |  | 77 | American | Female | 14 |
| 27 | American | Female | 10 |  | 78 | African American | Female | 12 |
| 28 | Asian | Female | 8 |  | 79 | Asian | Male | 9.5 |
| 29 | Asian | Male | 13 |  | 80 | Asian | Female | 9 |
| 30 | Asian | Female | 9 |  | 81 | Asian | Male | 10 |
| 31 | Asian | Male | 8 |  | 82 | African American | Female | 10 |
| 32 | Asian | Male | 11 |  | 83 | African American | Female | 9.5 |
| 33 | Asian | Male | 10 |  | 84 | American | Male | 10 |
| 34 | Asian | Male | 8 |  | 85 | Asian | Male | 10 |
| 35 | American | Male | 10 |  | 86 | Asian | Male | 10 |
| 36 | American | Female | 8 |  | 87 | Asian | Female | 10 |
| 37 | Asian | Male | 11.5 |  | 88 | Asian | Female | 7 |
| 38 | American | Male | 15 |  | 89 | Asian | Female | 7 |
| 39 | American | Female | 12 |  | 90 | Asian | Female | 9 |
| 40 | African American | Male | 10 |  | 91 | African American | Male | 14 |
| 41 | Asian | Female | 7 |  | 92 | Asian | Male | 10 |
| 42 | Asian | Female | 6 |  | 93 | Asian | Female | 10 |
| 43 | American | Male | 9 |  | 94 | Asian | Female | 9 |
| 44 | Asian | Male | 13 |  | 95 | Asian | Male | 9 |
| 45 | American | Female | 9 |  | 96 | Asian | Male | 11 |
| 46 | Asian | Male | 11 |  | 97 | Asian | Female | 9 |
| 47 | Asian | Female | 8 |  | 98 | American | Female | 10 |
| 48 | Asian | Male | 11 |  | 99 | Asian | Male | 8 |
| 49 | American | Male | 9 |  | 100 | African American | Male | 12 |
| 50 | American | Female | 9.5 |  | 99 | Asian | Male | 8 |
| 51 | Asian | Male | 13 |  | 100 | African American | Male | 12 |

# Appendix B

## Data set 2

Below is a table of the raw data of data set 2 that was gathered with various values and used for the goodness of fit test and descriptive statistical analysis.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Time** | **Time Difference in Seconds** | **Seconds** |  | **No** | **Time** | **Time Difference in Seconds** | **Seconds** |
| 1 | 15:45:29 | 0 | 0 |  | 52 | 15:52:31 | 00:00:05 | 5 |
| 2 | 15:45:46 | 00:00:17 | 17 |  | 53 | 15:52:35 | 00:00:04 | 4 |
| 3 | 15:46:08 | 00:00:22 | 22 |  | 54 | 15:52:39 | 00:00:04 | 4 |
| 4 | 15:46:20 | 00:00:12 | 12 |  | 55 | 15:52:50 | 00:00:11 | 11 |
| 5 | 15:46:25 | 00:00:05 | 5 |  | 56 | 15:53:03 | 00:00:13 | 13 |
| 6 | 15:46:31 | 00:00:06 | 6 |  | 57 | 15:53:09 | 00:00:06 | 6 |
| 7 | 15:46:52 | 00:00:21 | 21 |  | 58 | 15:53:13 | 00:00:04 | 4 |
| 8 | 15:47:10 | 00:00:18 | 18 |  | 59 | 15:53:19 | 00:00:06 | 6 |
| 9 | 15:47:33 | 00:00:23 | 23 |  | 60 | 15:53:25 | 00:00:06 | 6 |
| 10 | 15:47:42 | 00:00:09 | 9 |  | 61 | 15:53:29 | 00:00:04 | 4 |
| 11 | 15:47:48 | 00:00:06 | 6 |  | 62 | 15:53:34 | 00:00:05 | 5 |
| 12 | 15:47:54 | 00:00:06 | 6 |  | 63 | 15:53:40 | 00:00:06 | 6 |
| 13 | 15:48:00 | 00:00:06 | 6 |  | 64 | 15:53:48 | 00:00:08 | 8 |
| 14 | 15:48:05 | 00:00:05 | 5 |  | 65 | 15:53:51 | 00:00:03 | 3 |
| 15 | 15:48:10 | 00:00:05 | 5 |  | 66 | 15:53:58 | 00:00:07 | 7 |
| 16 | 15:48:16 | 00:00:06 | 6 |  | 67 | 15:54:03 | 00:00:05 | 5 |
| 17 | 15:48:28 | 00:00:12 | 12 |  | 68 | 15:54:09 | 00:00:06 | 6 |
| 18 | 15:48:35 | 00:00:07 | 7 |  | 69 | 15:54:17 | 00:00:08 | 8 |
| 19 | 15:48:47 | 00:00:12 | 12 |  | 70 | 15:54:21 | 00:00:04 | 4 |
| 20 | 15:48:56 | 00:00:09 | 9 |  | 71 | 15:54:25 | 00:00:04 | 4 |
| 21 | 15:49:00 | 00:00:04 | 4 |  | 72 | 15:54:37 | 00:00:12 | 12 |
| 22 | 15:49:07 | 00:00:07 | 7 |  | 73 | 15:54:43 | 00:00:06 | 6 |
| 23 | 15:49:12 | 00:00:05 | 5 |  | 74 | 15:54:47 | 00:00:04 | 4 |
| 24 | 15:49:30 | 00:00:18 | 18 |  | 75 | 15:54:59 | 00:00:12 | 12 |
| 25 | 15:49:34 | 00:00:04 | 4 |  | 76 | 15:55:02 | 00:00:03 | 3 |
| 26 | 15:49:43 | 00:00:09 | 9 |  | 77 | 15:55:12 | 00:00:10 | 10 |
| 27 | 15:49:47 | 00:00:04 | 4 |  | 78 | 15:55:16 | 00:00:04 | 4 |
| 28 | 15:49:51 | 00:00:04 | 4 |  | 79 | 15:55:20 | 00:00:04 | 4 |
| 29 | 15:49:55 | 00:00:04 | 4 |  | 80 | 15:55:30 | 00:00:10 | 10 |
| 30 | 15:50:03 | 00:00:08 | 8 |  | 81 | 15:55:34 | 00:00:04 | 4 |
| 31 | 15:50:09 | 00:00:06 | 6 |  | 82 | 15:55:38 | 00:00:04 | 4 |
| 32 | 15:50:20 | 00:00:11 | 11 |  | 83 | 15:55:42 | 00:00:04 | 4 |
| 33 | 15:50:32 | 00:00:12 | 12 |  | 84 | 15:55:48 | 00:00:06 | 6 |
| 34 | 15:50:37 | 00:00:05 | 5 |  | 85 | 15:55:54 | 00:00:06 | 6 |
| 35 | 15:50:43 | 00:00:06 | 6 |  | 86 | 15:55:58 | 00:00:04 | 4 |
| 36 | 15:50:52 | 00:00:09 | 9 |  | 87 | 15:56:03 | 00:00:05 | 5 |
| 37 | 15:50:59 | 00:00:07 | 7 |  | 88 | 15:56:07 | 00:00:04 | 4 |
| 38 | 15:51:10 | 00:00:11 | 11 |  | 89 | 15:56:12 | 00:00:05 | 5 |
| 39 | 15:51:20 | 00:00:10 | 10 |  | 90 | 15:56:20 | 00:00:08 | 8 |
| 40 | 15:51:25 | 00:00:05 | 5 |  | 91 | 15:56:22 | 00:00:02 | 2 |
| 41 | 15:51:32 | 00:00:07 | 7 |  | 92 | 15:56:30 | 00:00:08 | 8 |
| 42 | 15:51:37 | 00:00:05 | 5 |  | 93 | 15:56:37 | 00:00:07 | 7 |
| 43 | 15:51:43 | 00:00:06 | 6 |  | 94 | 15:56:39 | 00:00:02 | 2 |
| 44 | 15:51:50 | 00:00:07 | 7 |  | 95 | 15:56:51 | 00:00:12 | 12 |
| 45 | 15:52:00 | 00:00:10 | 10 |  | 96 | 15:57:03 | 00:00:12 | 12 |
| 46 | 15:52:06 | 00:00:06 | 6 |  | 97 | 15:57:09 | 00:00:06 | 6 |
| 47 | 15:52:10 | 00:00:04 | 4 |  | 98 | 15:57:13 | 00:00:04 | 4 |
| 48 | 15:52:14 | 00:00:04 | 4 |  | 99 | 15:57:18 | 00:00:05 | 5 |
| 49 | 15:52:18 | 00:00:04 | 4 |  | 100 | 15:57:21 | 00:00:03 | 3 |
| 50 | 15:52:22 | 00:00:04 | 4 |  | 101 | 15:57:32 | 00:00:11 | 11 |
| 51 | 15:52:26 | 00:00:04 | 4 |  |  |  |  |  |

# Appendix C

## Goodness of fit Calculations of Data set 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Observed Frequency(oi)** | **Class Probability Calculations** | **Expected Frequency (ei)** | **χ2 Class Component** |
| X ≤ 8 | Count the | NORMDIST (8,9.92,1.89,1) | Take the | **(oi-ei) ^2/ei** |
| 8 < X ≤ 10 | observations of | NORMDIST (10,9.92,1.89,1)-NORMDIST (8,9.92,1.89,1) | probability for |
| 10 < X ≤ 12 | each class from the | NORMDIST (12,9.92,1.89,1)-NORMDIST (10,9.92,1.89,1) | each class |
| 12 < X ≤ 14 | frequency table | NORMDIST (14,9.92,1.89,1)-NORMDIST (12,9.92,1.89,1) | and multiply it by n |
| X >14 |  | 1-NORMDIST (14,9.92,1.89,1) |  |
| **Total** | **n** | **1.00** | **n** | **χ2 statistic** |
|  |  |  |  |  |
| **Class** | **Observed Frequency(oi)** | **Class Probability Calculations** | **Expected Frequency (ei)** | **χ2 Class Component** |
| X ≤ 8 | Count the | NORMDIST (8,9.92,1.89,1) | Take the | **(oi-ei) ^2/ei** |
| 8 < X ≤ 10 | observations of | NORMDIST (10,9.92,1.89,1)-NORMDIST (8,9.92,1.89,1) | probability for |
| 10 < X ≤ 12 | each class from the | NORMDIST (12,9.92,1.89,1)-NORMDIST (10,9.92,1.89,1) | each class |
| X > 12 | frequency table | 1-NORMDIST (12,9.92,1.89,1) | and multiply it by n |
| **Total** | **n** | **1.00** | **n** | **χ2 statistic** |

The class probabilities are calculated using the excel functions.

P [X ≤ 8] = NORMDIST (8,9.92,1.89,1)

P [8 < X ≤ 10] = P [X ≤ 10] − P [X ≤ 8] = NORMDIST (10,9.92,1.89,1) - NORMDIST (8,9.92,1.89,1)

P [10 < X ≤ 12] = P [X ≤ 12] − P [X ≤ 10] = NORMDIST (12,9.92,1.89,1) - NORMDIST (10,9.92,1.89,1)

P [X > 12] = 1 - P [X ≤ 12] = 1-NORMDIST (12,9.92,1.89,1)

## Goodness of fit Calculations of Data set 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Observed Frequency(oi)** | **Class Probability Calculations** | **Expected Frequency (ei)** | **χ2 Class Component** |
| X ≤ 3 | Count the | GAMMADIST (3,1,7.23,1) | Take the | **(oi-ei) ^2/ei** |
| 3 < X ≤ 6 | observations of | GAMMADIST (6,1,7.23,1)-GAMMADIST (3,1,7.23,1) | probability for |
| 6 < X ≤ 9 | each class from the | GAMMADIST (9,1,7.23,1)-GAMMADIST (6,1,7.23,1) | each class |
| 9 < X ≤ 12 | frequency table | GAMMADIST (12,1,7.23,1)-GAMMADIST (9,1,7.23,1) | and multiply it by n |
| 12 < X ≤ 15 |  | GAMMADIST (15,1,7.23,1)-GAMMADIST (12,1,7.23,1) |  |
| X >15 |  | 1-GAMMADIST (15,1,7.23,1) |  |
| **Total** | **n** | **1.00** | **n** | **χ2 statistic** |

P [X ≤ 3] = GAMMADIST (3,1,7.23,1)

P [2 < X ≤ 6] = P [X ≤ 6] - P [X ≤ 3] = GAMMADIST (6,1,7.23,1)-GAMMADIST (3,1,7.23,1)

P [6 < X ≤ 9] = P [X ≤ 9] - P [X ≤ 6] = GAMMADIST (9,1,7.23,1)-GAMMADIST (6,1,7.23,1)

P [9 < X ≤ 12] = P [X ≤ 12] - P [X ≤ 9] = GAMMADIST (12,1,7.23,1)-GAMMADIST (9,1,7.23,1)

P [12 < X ≤ 15] = P [X ≤ 15] - P [X ≤ 12] = GAMMADIST (15,1,7.23,1)-GAMMADIST (12,1,7.23,1)

P [X > 15] = 1 - P [X ≤ 15] = 1-GAMMADIST (15,1,7.23,1)

# References

Benjamin Frimodig (2020, December 14). Chi Square (χ2) Test Statistic

https://www.simplypsychology.org/chi-square.html

Shaun Turney (2022, May 23). Chi-Square (Χ²) Tests | Types, Formula & Examples

<https://www.scribbr.com/statistics/chi-square-tests/>