

STA2023 Application: Sample Data, Statistics, and the Empirical Rule

The completed application should be submitted to the Assignments link in Falcon Online.

Purpose: The purpose of this assignment is to organize a random sample of data values and create statistics, tables, and a graph based on the data. Then the sample data will be compared to the Empirical Rule. The information will then be analyzed in a written summary.

Part 1: Random Data, Statistics, and the Empirical Rule

Methods: Use Excel (or similar software) to create the tables and graph. Then copy the items and paste them into a Word document. The tables should be formatted vertically, have borders, and be given the labels and titles stated in the assignment. The proper symbols should be used. Do not submit this assignment as an Excel file. The completed assignment should be a Word (or .pdf) document.

1. The data values and relevant information are posted in the course website. Use the data set (P, Q, R, S, or T) assigned to you by your instructor to complete this application.

For the purpose of this application, treat the data set as if it represented a certain random variable and was a valid random sample gathered by a researcher from a normally distributed population. The sample data was actually found with an online Gaussian random number generator that creates normally distributed data values. The random number generator simulates the results of a researcher finding those values through observation or experimentation.

2. Use technology (Excel, graphing calculator, etc.) to sort the sample data values from low to high. Use Excel or similar software to put the data into a table with about 5 or 6 columns. Label this "Table 1: Sorted Set of Sample Data."

3. Using 5 to 10 class intervals, organize the sample data as a frequency distribution in a table. The intervals of the frequency distribution should be rounded to the tenths so that they match the data. Label this "Table 2: Frequency Distribution."

4. Use Excel (or similar software) to construct a frequency histogram to illustrate the data. Give the axes the proper titles. Label this "Graph 1: Histogram."

5. Use Table 2, the frequency distribution, to find the midpoints of each class interval. Create a new frequency distribution with the midpoints in the left column and the frequencies in the right column. Label this "Table 3: Frequency Distribution with Midpoints."

6. Use technology to find the mean, median, standard deviation, and variance of the sample data organized in Table 3 (from step 5 above). Put these values into a table with the proper symbol in the left column and the value of the statistic in the right column. Also, from the original data set, put the values of the range and sample size in the table. The median and range do not generally have symbols so the terms "Median" and "Range" can be used in the left

column. Identify the modal class (the one with the highest frequency). Put the terms "Modal Class" in the left column and the class interval in the right column. The statistics should be rounded properly (one more decimal place than the data). Label this "Table 4: Summary Statistics"

7. Use the sample mean and standard deviation to find the values related to the Empirical Rule.

The Empirical Rule: For a set of data whose distribution is approximately normal,

- about 68% of the data are within one standard deviation of the mean.
- about 95% of the data are within two standard deviations of the mean.
- about 99.7% of the data are within three standard deviations of the mean.

Use the value of n and the percents listed above to find how many data values should be within each category. Then use the sample mean and standard deviation to find the lower and upper cut-off values in each category. Then use the sorted list of data to determine how many values are actually in each category. Put the values into a table as shown in the example and label it "Table 5: The Empirical Rule."

Part 2: Written Introduction and Summary

1. Write an introduction to this application. Discuss the random variable and the source of the sample data. Refer to the textbook or class notes to describe the basic components of the application and the statistical concepts that are applied. The introduction should be at least 50 words and be written with proper grammar and spelling.

2. Write a summary of the application, considering the following topics. The summary should be at least 150 words and be written with proper grammar and spelling. Refer to the tables and graph (by label and number) throughout the summary. Use the proper statistical terms and symbols in the summary. Write about the concepts in the application rather than the steps followed. Do not number the parts or steps in the application except for the proper numbering of the tables and graph.

- Discuss the difference between population and sample values
- Discuss how the frequency distribution was created, specifically referring to the intervals and class width.
- Describe the features of the histogram, including the axes and the shape of the distribution, especially whether it meets the criteria of an approximately normal distribution
- Discuss measures of center and variation as related to the sample data; given the values used in the random number generator, compare the sample values to those of the population
- Discuss the Empirical Rule and compare the actual number of values within each category of the Empirical Rule to the number that should be there
- Discuss how the Empirical Rule is related to the criteria for significantly low or high values.
- Discuss characteristics of the sample data such as significantly low/high values and skewness, providing support and examples. Use both the histogram and measures of center to identify whether there is skewness.

Part 3: Format Requirements

- Include a title page with your name, STA2023 Application 1, the word count for the introduction, the word count for the summary, and which data set is used.
- The introduction and summary should be written in paragraphs that are typed and double-spaced, with 1-inch margins and a readable font type and font size. The preferred font size could be 12 Times New Roman.
- The introduction by itself should be at least 50 words.
- The summary should be at least 150 words.
- The introduction and summary should be college-level writing with proper grammar and spelling. Do not use first or second person (I, you, etc.).
- Do not write about the parts of the assignment or the use of a calculator or excel to complete steps in the application. Write about the concepts used in the application.
- Throughout the introduction, tables, graph, and summary, proper statistical symbols and terms should be used.
- The tables should be formatted vertically, have 2 columns (unless specified otherwise), have borders, and be given the labels and titles stated in the assignment.
- The title page, introduction, tables, graphs, and summary should be a single document (.doc, .docx, or .pdf) which is submitted to the Assignments link in Falcon Online by the due date.
- Assemble the application in the following order to create a single document.
 - Title page
 - Introduction
 - Tables and Graph
 - Summary

Part 4: Grading: The application is worth up to 30 points, based on the following criteria.

- The **statistical content** (Part 1) is worth up to 10 points, based on completion of all the listed requirements.
- The **introduction and written summary** (Part 2) are worth up to 10 points, based on completion of all the listed requirements.
- The **format** (Part 3) of the complete document submitted to the Assignments link is worth up to 10 points, based on completion of all the listed requirements.

Rubric/allocation of 10 points:

- 0 points for completing none of the requirements
- 2 points for attempting the requirements
- 4 points for completing some of the requirements
- 8 points for completing most of the requirements
- 10 points for completing all of the requirements