METRISTICS

SOEN 6611: SOFTWARE MEASUREMENT (GROUP C)

Anirudh Boddu	40225271
Ayushi Chaudhary	40224978
Riddhi Bhuva	40220969
Visnunathan Chidambaranathan	40230157
Yang Cao	26654029

October 2, 2023

Contents

List of Figures

List of Tables

Introduction

The purpose of descriptive statistics is to quantitatively describe a collection of data by measures of central tendency, measures of frequency, and measures of variability.

Let x be a random variable that can take values from a finite data set x1, x2, x3, ..., xn, with each value having the same probability.

- The minimum, m, is the smallest of the values in the given data set. (m need not be unique.)
- The maximum, M, is the largest of the values in the given data set. (M need not be unique.)
- The mode, o, is the value that appears most frequently in the given data set. (o need not be unique.)
- The median, d, is the middle number if n is odd, and is the arithmetic mean of the two middle numbers if n is even.
- The arithmetic mean, μ, is given by

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- The mean absolute deviation, MAD, is given by

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu_i|$$

- The standard deviation, σ, is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i - \mu^2}$$

Project URL:
https://github.com/itshisher/METRICSTICS
https://www.overleaf.com/read/gxkjzdqfbndm

Chapter 1

Problem 1

1.1 Goal of the METRICSTICS system

Increase the accuracy and reliability of statistical calculations performed by the calculator to achieve a margin of error of less than 1% and ensure a high level of reliability for all 7 used statistical functions by the end of November 2023.

Specific

The goal is specific about what needs improvement - the accuracy of statistical calculations. It specifies the desired outcome, which is achieving a margin of error of less than 1%.

Measurable

Accuracy can be measured by comparing the calculator's results with known correct results for various statistical functions. Achieving a margin of error of less than 1% provides a clear, quantifiable benchmark.

Attainable

Improving accuracy in statistical calculations is achievable through rigorous testing, validation, and refinement of the underlying algorithms and code in the calculator.

Realistic

Accuracy is a critical factor for a statistics calculator. Users rely on it for precise calculations in fields such as research, data analysis, and decision-making.

Timely

Achieve a margin of error of less than 1% within the next 2 months.

1.2 Questions with applied metrics

Question 1

How can we improve the accuracy of METRICSTICS in calculating minimum (m) and maximum (M) values for different data sets?

- Metric 1: Mean Absolute Error (MAE) between METRICSTICS' calculated minimum (m) and actual minimum.
- Metric 2: Mean Absolute Error (MAE) between METRICSTICS' calculated maximum (M) and actual maximum.

Question 2

How can we improve the correctness of METRICSTICS in determining mode (o) for data sets with large sample sizes?

- Metric 1: Mode calculation time for different data set sizes.
- Metric 2: Percentage of correctly identified modes for data sets with varying degrees of multimodality.

Question 3

What strategies can be implemented to effectively handle even and odd sample sizes when calculating the median (d)?

- Metric 1: Median calculation time for data sets with odd and even sample sizes
- Metric 2: Accuracy of METRICSTICS in calculating the median compared to established methods for different sample sizes.

Question 4

How can we improve the accuracy of METRICSTICS in calculating the arithmetic mean (μ) for data sets with outliers?

- Metric 1: Mean Absolute Error (MAE) between METRICSTICS' calculated mean (μ) and a robust mean calculation method (e.g., trimmed mean).
- Metric 2: Percentage of data sets where METRICSTICS' mean calculation is influenced by outliers.

Question 5

How can we reduce the mean absolute deviation (MAD) in METRICSTICS' descriptive statistics?

- Metric 1: MAD value for data sets processed by METRICSTICS compared to traditional MAD calculation.
- Metric 2: Percentage reduction in MAD achieved by implementing optimization techniques.

Question 6

How can we improve the accuracy of METRICSTICS in calculating the standard deviation (σ) for data sets with varying distributions? distributions?

- Metric 1: Mean Squared Error (MSE) between METRICSTICS' calculated standard deviation (σ) and actual standard deviation.
- Metric 2: Standard deviation error as a function of data set skewness and kurtosis.

Question 7

What strategies can be implemented to ensure the reliability and stability of METRICSTICS' calculations?

- Metric 1: Calculation stability over multiple runs with the same data.
- Metric 2: Frequency of software crashes or errors during calculations.

Question 8

How can we ensure METRICSTICS remains adaptable to different data types and formats?

- Metric 1: Number of supported data formats and data types.
- Metric 2: Frequency of updates and additions to accommodate new data structures.

Question 9

What is the team's capacity for maintaining and improving METRICSTICS over time?

- Metric 1: Team workload in terms of METRICSTICS-related tasks.
- Metric 2: Average time to address and implement user feedback and feature requests.

Question 10

How can we enhance the precision and numerical stability of METRICSTICS for very large or very small data values?

- Metric 1: Assessment of numerical precision using double precision arithmetic.
- Metric 2: Handling of extreme values in calculations.

Chapter 2

Problem 2

2.1 Use Case Model for METRICSTICS

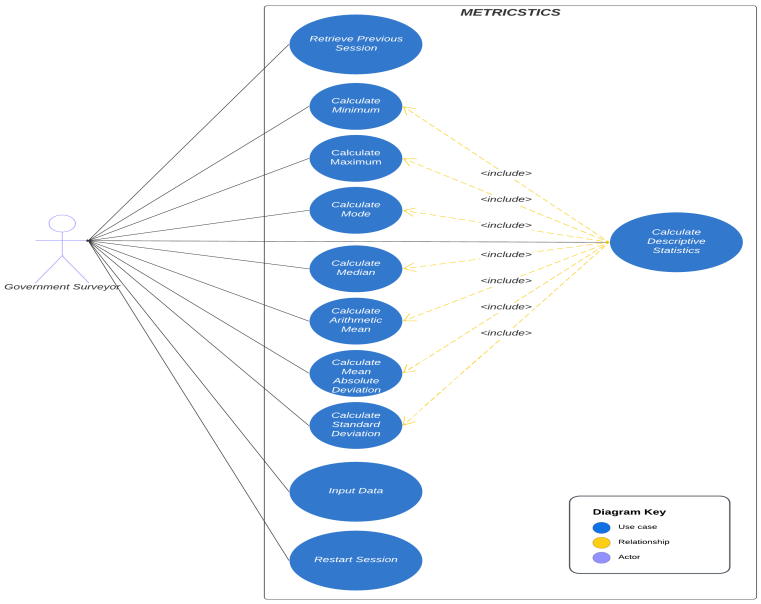

Figure 1: A user modeling graph

2.2 Actors

- **User:** Government Surveyor
  The primary user of the Metricstics system, representing a surveyor working for the government. This surveyor's responsibilities may include collecting and analyzing data from surveys, censuses, and various government-related studies.

2.3 Use cases

Use case 1

System: Metricstics	
Identifier:	UC-01 (Input Data)
Author(s):	Anirudh Boddu, anirudh.boddu@mail.concordia.ca
Version:	1.0
Priority:	High
Name:	Input Data
Pre-Condition(s):	1. The system is running and accessible. 2. The government surveyor has initiated a session.
Post-Condition(s):	1. Data has been successfully entered and saved by the system.
Trigger:	Government surveyor initiates data input.
Normal Flow:	1. Government surveyor selects "Input Data" option. 2. Surveyor chooses to upload a CSV file or manually enter data. 3. Surveyor provides data as per the selected method. 4. System validates and accepts the data. 5. Use case ends.
Exceptional Flow(s):	1. Surveyor cancels data input. 2. Data format is incorrect or incomplete. 3. System encounters an error.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Calculate Descriptive Statistics (All-in-One), Calculate Mean, Calculate Minimum, Calculate Maximum, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 1: Use Case 1: Input Data

Use case 2

System: Metricstics	
Identifier:	UC-02 (Calculate Descriptive Statistics (All-in-One))
Author(s):	Anirudh Boddu, anirudh.boddu@mail.concordia.ca
Version:	1.0
Priority:	High
Name:	Calculate Descriptive Statistics (All-in-One)
Pre-Condition(s):	1. Data input has been completed (Use Case 1). 2. The system has received valid data for analysis.
Post-Condition(s):	1. Descriptive statistics (minimum, maximum, mean, median, mode, MAD, and standard deviation) have been calculated and displayed to the user.
Trigger:	Government surveyor selects "Calculate Descriptive Statistics" option.
Normal Flow:	1. Government surveyor selects "Calculate Descriptive Statistics" option. 2. System processes the input data. 3. System calculates minimum, maximum, mean, median, mode, MAD, and standard deviation. 4. System displays the results to the user. 5. Use case ends.
Exceptional Flow(s):	1. Data input is missing or invalid. 2. System encounters an error during calculations.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Calculate Minimum, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 2: Use Case 2: Calculate Descriptive Statistics (All-in-One)

Use case 3

System: Metricstics	
Identifier:	UC-03 (Calculate Minimum)
Author(s):	Ayushi Chaudhary, ayushivihan@gmail.com
Version:	1.0
Priority:	High
Name:	Calculate Minimum
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The minimum value from the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Minimum" option.
Normal Flow:	1. Government surveyor selects the "Calculate Minimum" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the minimum value. 4. The minimum value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 3: Use Case 3: Calculate Minimum

Use case 4

System: Metricstics	
Identifier:	UC-04 (Calculate Maximum)
Author(s):	Ayushi Chaudhary, ayushivihan@gmail.com
Version:	1.0
Priority:	High
Name:	Calculate Maximum
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The maximum value from the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Maximum" option.
Normal Flow:	1. Government surveyor selects the "Calculate Maximum" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the maximum value. 4. The maximum value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Minimum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 4: Use Case 4: Calculate Maximum

Use case 5

System: Metricstics	
Identifier:	UC-05 (Calculate Mean)
Author(s):	Visnunathan Chidambaranathan, visnu8898@gmail.com
Version:	1.0
Priority:	High
Name:	Calculate Mean
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The arithmetic mean (average) of the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Mean" option.
Normal Flow:	1. Government surveyor selects the "Calculate Mean" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the arithmetic mean (average) 4. The mean value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 5: Use Case 5: Calculate Mean

Use case 6

System: Metricstics	
Identifier:	UC-06 (Calculate Median)
Author(s):	Visnunathan Chidambaranathan, visnu8898@gmail.com
Version:	1.0
Priority:	High
Name:	Calculate Median
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The median value from the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Median" option.
Normal Flow:	1. Government surveyor selects the "Calculate Median" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the median value. 4. The median value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 6: Use Case 6: Calculate Median

Use case 7

System: Metricstics	
Identifier:	UC-07 (Calculate Mode)
Author(s):	Yang Cao, caoyangtommy@hotmail.com
Version:	1.0
Priority:	High
Name:	Calculate Mode
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The mode (most frequent value) of the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Mode" option.
Normal Flow:	1. Government surveyor selects the "Calculate Mode" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the mode value. 4. The mode value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 7: Use Case 7: Calculate Mode

Use case 8

System: Metricstics	
Identifier:	UC-08 (Calculate Mean Absolute Deviation)
Author(s):	Yang Cao, caoyangtommy@hotmail.com
Version:	1.0
Priority:	High
Name:	Calculate Mean Absolute Deviation
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The mean absolute deviation (MAD) from the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Mean Absolute Deviation" option.
Normal Flow:	1. Government surveyor selects the "Calculate Mean Absolute Deviation" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the mean absolute deviation (MAD). 4. The MAD value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 8: Use Case 8: Calculate Mean Absolute Deviation

Use case 9

System: Metricstics	
Identifier:	UC-09 (Calculate Standard Deviation)
Author(s):	Riddhi Bhuva, riddhivinodbhuva@gmail.com
Version:	1.0
Priority:	High
Name:	Calculate Standard Deviation
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. A government surveyor has initiated a session and provided data.
Post-Condition(s):	1. The standard deviation from the dataset has been successfully calculated.
Trigger:	Government surveyor selects the "Calculate Standard Deviation" option.
Normal Flow:	1. Government surveyor selects the "Calculate Standard Deviation" option from the available choices. 2. METRICSTICS processes the data from the dataset. 3. METRICSTICS calculates the standard deviation. 4. The standard deviation value is displayed to the government surveyor. 5. Use case ends.
Exceptional Flow(s):	1. No Data Available: - If there is no data available in the dataset, METRICSTICS displays an error message indicating the unavailability of data for calculation. - Use case ends. 2. Error in Calculation: - If an unexpected error occurs during the calculation process (e.g., data format error), METRICSTICS displays an error message. - Government surveyor may retry the operation with corrected data. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	Input Data, Calculate Maximum, Calculate Mean, Calculate Median, Calculate Mode, Calculate Mean Absolute Deviation, Calculate Standard Deviation.

Table 9: Use Case 9: Calculate Standard Deviation

Use case 10

System: Metricstics	
Identifier:	UC-10 (Retrieve Previous Session)
Author(s):	Riddhi Bhuva, riddhivinodbhuva@gmail.com
Version:	1.0
Priority:	Medium
Name:	Retrieve Previous Session
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. 2. A government surveyor has previously initiated a session that was saved or stored.
Post-Condition(s):	1. The previous session data has been successfully retrieved, restoring the system to its state when the session was saved.
Trigger:	Government surveyor selects the "Retrieve Previous Session" option.
Normal Flow:	1. Government surveyor selects the "Retrieve Previous Session" option from the available choices. 2. METRICSTICS searches for and retrieves the previously saved session data. 3. The system restores the session, including all stored data and results. 4. The government surveyor can continue working with the restored session data. 5. Use case ends.
Exceptional Flow(s):	1. No Previous Session Found: - If there is no previously saved session data available, METRICSTICS displays a message indicating that there are no previous sessions to retrieve. - Use case ends. 2. Error in Retrieval: - If an unexpected error occurs during the retrieval process (e.g., data corruption), METRICSTICS displays an error message. - The government surveyor may choose to retry or contact support for assistance. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	None

Table 10: Use Case 10: Retrieve Previous Session

Use case 11

System: Metricstics	
Identifier:	UC-11 (Session Restart)
Author(s):	Ayushi Chaudhary, ayushivihan@gmail.com
Version:	1.0
Priority:	High
Name:	Session Restart
Pre-Condition(s):	1. The METRICSTICS system is running and accessible. 2. The government surveyor is currently in an active session.
Post-Condition(s):	1. The session has been successfully restarted, clearing all stored data and results. 2. METRICSTICS returns to its initial state.
Trigger:	Government surveyor initiates a session restart.
Normal Flow:	1. Government surveyor selects the "Restart Session" option. 2. METRICSTICS displays a confirmation prompt, asking the government surveyor to confirm their intent to restart the session. 3. Government surveyor confirms the restart by selecting "Yes." 4. METRICSTICS clears all stored data and results. 5. METRICSTICS returns to its initial state. 6. Use case ends.
Exceptional Flow(s):	1. Government surveyor cancels the session restart: - In step 3 of the normal flow, if the government surveyor selects "No" in response to the confirmation prompt, the session restart is canceled, and the system remains in its current state. - Use case ends. 2. Error in restarting the session: - If an unexpected error occurs during the session restart process (e.g., system error), METRICSTICS displays an error message and remains in its current state. - Government surveyor may retry the session restart. - Use case ends.
Related Actor(s):	Government Surveyor
Related Use Case(s):	None

Table 11: Use Case 11: Session Restart

Contribution table

Anirudh Boddu	40225271	1. Questions Formulation. 2. Collaborated with the team in defining the goal and framing questions. 3. Ensured goal is SMART (Specific, Measurable, Achievable, Relevant, Time-bound). 4. Defined Actor, Use Cases and Use Case Scenarios.
Ayushi Chaudhary	40224978	1. Co-ordinated team meetings. 2. Questions Formulation. 3. Defined Use Cases and Use Case Scenarios. 4. Latex documentation.
Riddhi Bhuva	40220969	1. Questions Formulation. 2. Ensure goal is SMART (Specific, Measurable, Achievable, Relevant, Time-bound). 3. Designed Use Case diagram. 4.Research and Analysis.
Visnunathan Chidambaranathan	40230157	1. Questions Formulation. 2. Metric Identification. 3. Ensure that metrics are suitable for measuring progress toward answering the questions. 4. Research and Analysis.
Yang Cao	26654029	1. Questions Formulation. 2. Metric Identification. 3. Ensure that metrics are suitable for measuring progress toward answering the questions. 4. Latex documentation.

Table 12: Contribution table

Bibliography

[1] Prof. Pankaj Kamthan, *INTRODUCTION TO FRAMEWORKS FOR SOFTWARE MEASUREMENT* Lecture Notes, SOEN 6611

[2] User Modelling Diagram https://www.lucidchart.com

[3] SMART Performance Metrics https://praxie.com/smart-performance-metrics-tools-templates/#:~:text=SMART%20is%20an%20acronym%20that,help%20it%20achieve%20that%20success

[4] Use-Case Diagrams https://www.ibm.com/docs/en/rational-soft-arch/9.6.1?topic=diagrams-use-case

[5] Victor R. Basili Gianluigi Caldiera1 H. Dieter Rombach2, *THE GOAL QUESTION METRIC APPROACH* Department of Computer Science, University Of Maryland https://www.ibm.com/docs/en/rational-soft-arch/9.6.1?topic=diagrams-use-case