CONCORDIA UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

SOEN 6611: SOFTWARE MEASUREMENT: SECTION AA
SUMMER 2020

PROJECT DESCRIPTION

1. INTRODUCTION

The purpose of descriptive statistics is to quantitatively describe a collection of data.

Let x be a **random variable** that can take values from a finite data set x_1, x_2, x_3, ..., x_n, 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

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n} | x_i - \mu | .$$

The **standard deviation**, σ, is given by

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

Let there be a system, called DESCRIPTIVE-STATISTICS, for finding m, M, o, d, μ, MAD, and σ. The system should take as input a random number of data values and output its descriptive statistics.

The purpose of the project is to create a **set of interrelated artifacts** for conducting certain measurements related to DESCRIPTIVE-STATISTICS. In the rest of the document, 'DESCRIPTIVE-STATISTICS' stands for the name of both the project and the product, unless otherwise stated. The work on DESCRIPTIVE-STATISTICS has been divided into a **collection of related parts**.

2. DELIVERABLE 1 (D1)

PART 1 [20 MARKS]

Using the **Goal-Question-Metric (GQM)** approach (or one of its extensions), present one goal specific to DESCRIPTIVE-STATISTICS and articulate 2N questions related to that goal, where N is the team size. Discuss whether any metrics help answer those questions.

Note. The goals must aim to be SMART.

PART 2 [30 MARKS]

Using the given description, construct a **use case model** for DESCRIPTIVE-STATISTICS.

Note. There can be several use cases, including saving data in memory, restarting a session, and so on. A **statistical calculator**[1] could be used as a motivation to 'elicit' necessary use cases.

PART 3 [30 MARKS]

(a) Using the **use case points (UCP)** approach (or one of its extensions), provide an estimate of the effort towards the project.

(b) Using Basic COCOMO 81, provide an estimate of the effort towards the project.

(c) Comment on the difference in estimates using the UCP approach and COCOMO 81, and the actual effort towards the project.

Note. It should be understood that an estimation should be dependable, but, at the same time, an estimation is not an exactimation.

PART 4 [60 MARKS]

(a) Using the **Java** programming language, implement DESCRIPTIVE-STATISTICS from **scratch**. This means, apart from the functions related to input and output and basic arithmetic, an implementation of DESCRIPTIVE-STATISTICS should **not** make use of any reuse mechanism (such as built-in functions, libraries, or APIs) provided natively by the programming language or otherwise. This may lead to **recursive implementation** of certain **primary** functions, the result of which would be **secondary** functions.

[1] URL: https://www.socscistatistics.com/descriptive/ .

(b) Using input data that consists of (at least) 1000 values, randomly distributed between 0 and 1000, test DESCRIPTIVE-STATISTICS.

Note. It is better to aim for **optimal but feasible** algorithms. The implementation of a GUI is encouraged, but not required.

PART 5 [20 MARKS]

(a) Calculate the **cyclomatic number** (also known as cyclomatic complexity) of DESCRIPTIVE-STATISTICS.

(b) Comment on the qualitative conclusions that can be drawn with respect to the quantitative thresholds of the metric.

Note. It is better to aim for a small cyclomatic number.

PART 6 [30 MARKS]

(a) Calculate the object-oriented metrics, WMC, CF, and LCOM*, for each of the classes of DESCRIPTIVE-STATISTICS. For WMC, assume that the weights are not normalized. Show your calculations in detail, manually or automatically (using a tool), as applicable.

(b) Comment on the qualitative conclusions that can be drawn with respect to the quantitative thresholds of the respective metrics.

Note. It is better to aim for values that are within the respective thresholds allowed.

PART 7 [20 MARKS]

Calculate the Physical SLOC and Logical SLOC for DESCRIPTIVE-STATISTICS. State your counting scheme, and show your calculations, manually or using a tool, as applicable.

Note. The source code for any purpose is part of the count. The test code, if any, should not be part of the count.

PART 8 [30 MARKS]

(a) Using **Scatter Plot**, carry out an analysis of the correlations between the data for Logical SLOC and WMC obtained from DESCRIPTIVE-STATISTICS.

(b) Using a **correlation coefficient**, carry out an analysis of the correlations between the data for Logical SLOC and WMC obtained from DESCRIPTIVE-STATISTICS.

Note. It is important that the conclusions are sensible.

CONSTRAINTS [20 MARKS]

QUALITY OF PROGRAMMING

The submission must include **source code, test data, and documentation**, named and organized in appropriate directories that are easily locatable.

DESCRIPTIVE-STATISTICS should aim for **generality**.

The source code of DESCRIPTIVE-STATISTICS should be **modular**, and aim for **readability** and **maintainability**.

The source code of DESCRIPTIVE-STATISTICS should be **independent** of any particular Java processor or editor or, more generally, any IDE.

The source code of DESCRIPTIVE-STATISTICS should aim to conform to an established **Java programming style**.

QUALITY OF DOCUMENTATION

The documentation for DESCRIPTIVE-STATISTICS must be expressed in **LATEX**.

The documentation of DESCRIPTIVE-STATISTICS should be **independent** of any particular LATEX processor or editor or, more generally, any IDE.

The documentation for DESCRIPTIVE-STATISTICS should, according to established norms of **technical writing**, be **organized**, and aim to be **clear and concise**.

The documentation for DESCRIPTIVE-STATISTICS should aim to be as **self-contained** as possible.

The documentation for DESCRIPTIVE-STATISTICS must **state explicitly any assumptions** made in the project.

The documentation for DESCRIPTIVE-STATISTICS must include brief **descriptions of any algorithm(s)** used and the **rationale** for selecting those algorithm(s).

For clarity and to minimize the potential for ambiguity, the documentation for DESCRIPTIVE-STATISTICS must have a **glossary**.

A team must strive for the highest standard of **academic ethics**. The documentation for DESCRIPTIVE-STATISTICS must have **citations and references at appropriate places** corresponding to any non-original work in the project (that is, any work external to the team). A comprehensive collection of resources on citing and referencing is available[2]. For example, **ACM, APA, and IEEE** provide standard formats for citing and referencing. It is important not to make claims that cannot be substantiated, and not to copy others' work verbatim regardless of whether it is cited. A copied work does not earn any credit.

[2] URL: http://library.concordia.ca/help/howto/citations.html .

3. DELIVERABLE 2 (D2) [40 MARKS]

This involves, in a poster, slides, or video presentation, defending major decisions made in D1, explaining briefly why those decisions were critical, and providing a retrospective (such as, by presenting a perspective on the lessons learned by doing the project).

NOTES

The **team in the audience** must **prepare (in real-time or otherwise) two questions and pose those questions** to the team that is presenting.