

Imperial College of Science, Technology and Medicine	University of London
Computer Science (CS) / Software Engineering (SE)	BEng and MEng Examinations Part I
Department of Computing	Integrated Laboratory Course
Laboratory work is a continuously assessed part of the examinations and is a required part of the degree assessment. Laboratory work must be handed in for marking by the due date. Late submissions may not be marked.	

Exercise: 8	Working: Individual
Title: Simple Statistical Analysis	
Issue date: 1st Dec 2003	Due date: 8th Dec 2003
System: Linux	Language: Kenya/Java

## Aims

- To develop a program for performing statistical analysis on a set of input data.
- To design and write an imperative program comprising several methods.
- To gain practice in handling arrays with associated loops.

## The Problem

- Each student in a class is given a mark in the range **0..10** for a piece of work. Each mark is given as a whole number. The number of students in a class can vary, but can never be less than one. The marks for a piece of work are to be analysed statistically and the **mean**, **median** and **standard deviation** output on the screen. The data is also presented graphically in the form of a bar chart.

## Definitions Used

- The *mean* of a set of values is the sum of the values divided by the number of values.
- The *median* (or central value) is the central value in a data set. It is calculated differently depending on whether the number of items in the set is odd or even. Assuming the items are arranged in ascending order, the median of an odd numbered set is the  $\frac{1}{2}(n+1)^{th}$  item; the median of an even numbered set is the mean of the  $\frac{1}{2}n^{th}$  and the  $\frac{1}{2}(n+1)^{th}$  item.

- *Standard deviation* is a measure of variation from the mean in a data set. It is calculated as follows:

$$S.D = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

where  $n$  is the number of items,

$\bar{x}$  is the mean

and  $x_i$  are the individual data items.

- Write a Kenya/Java program **Statistics.k/Statistics.java** to perform the following tasks in order:
  1. **Input** a set of marks from the keyboard (“standard input”). Marks are separated by spaces, and/or newline(s). Input is terminated by an end of file. This is (detectable using the boolean method **isEOF()** in Kenya) and is generated by clicking on the **EOF** button in the I/O window.
  2. Find the **mean**, **median** and **standard deviation** of the data.
  3. **Output** the data on the screen (“standard output”), in the form of the mean, median and standard deviation in the above order, one to a line, and then print a vertical bar chart. The chart should be displayed using “|”, “\*” and “ ” characters,
- Your program can include I/O such as prompts to the user and explanations of the program output.

## Submit by Monday 8th December 2003

### What To Do

- You can compare the output of your program with a lab version of the program by typing **LabStatistics** at your Unix prompt. You can then type the marks (in the proper format). To indicate end-of-file when typing your input type *control-d* after typing a newline.
- Design and code **Statistics.k/Statistics.java**. To do this you need to consider what data structures are required. You can design the types as you choose; we recommend the use of an 11-element array containing the numbers of students associated with each possible mark from 0..10. This will allow you to input and process any number of students’ marks.

We also recommend you write a separate method for each of the phases of your program, *e.g.* a method **getMarks** to read in a set of marks and the methods **findMean**, **findStdDeviation**, **findMedian** to process the data and a method **printBarChart** to display the marks.

Your main program can simply call these methods in order.

- Kenya provides two **integer** operators **/** and **%** which may be useful for the calculation of the median value. For integer division you can use **/** ( **7 / 3** gives 2), and to find the remainder after a division you can use **%**. ( **16 % 5** gives 1 ) and ( **3 % 3** gives 0.) Kenya also provides the pervasive method **sqr**t for finding the square root of a number.

For example `sqrt( 100 )` gives 10. The parameter passed to **sqrt** can be either an *int* or a *double* however the return type is always a *double*.

- After printing out the mean, median and standard deviation your program should display the data in bar chart form: one vertical bar for each person with that mark, tipped by a star at the top.

As an example given the input :

```
7 9 8 7 5 9 8 4 8 6 0 9 10 6 10 9 7 8
```

your program should output the following bar chart:

```

          * *
        * | |
      * | | | *
    *   * * | | | |
  | * * * | | | | |
0 1 2 3 4 5 6 7 8 9 10
```

- Note that each mark in the bar chart is separated by a space character.

## Testing

- You should **test your code thoroughly** before submitting.
- Test your program on all the different types of mark sets, *e.g.* a single mark, odd numbers of marks, even numbers of marks, all students having same mark, *etc. etc.*

Testing is extremely important and you should try as hard as possible to detect and fix any bugs in your program before you submit it. Your program should work as specified for any legal input. A program which works “most of the time” cannot arguably be said to work at all.

- You should get in the habit of testing your program as you write it. This will make bugs much easier to find and fix. If you acquire the habits of good style, commenting and incremental testing when you write small programs you will save yourself a great deal of time and effort later when you write large programs.

## Submission

- Submit **Statistics.k** and **Statistics.java** by typing **submit 8**.

## Assessment

Finding the mean	1.0
Finding the standard deviation	1.0
Finding the median	1.5
I/O (bar chart etc.)	1.5
Design, style, readability	5.0
Total	10.0