G51PRG Exercise Four-B: Statistics Processing

Steven R. Bagley

Introduction

This exercise involves writing a program that reads in a list of integers between `0` and `9`, and counting the number of occurences of each number. The program should stop reading in when the end of file is reached. Once the data has been read the program should calculate the mean, median and mode values of the data as well as printing out a histogram of the frequencies.

Number frequency

For this exercise, you are to develop a C program that can open a text file (using `fopen()` and with the name of the file passed in on the command line) and to read in a series of numbers from `0` to `9`. If a value out of the range `0` to `9` is entered, the program should print an error message (on the standard error output) and continue to read in values.

You'll need an array of 10 `int`s to store the count of each digit. Initialise all its elements to zero, and then increment each element appropriately as digits are read in

The maximum length of any bar in the histogram should be 70 units. If this limit is exceeded, the program should print an error message saying the length was too 'long' to the *standard error* output stream, and exit the program.

Output

Once you've successfully processed the entire file, you should output the count of each digit as a horizontal histogram, like this:

```
0 ( 6)  :======
1 ( 4)  :====
2 ( 8)  :========
3 (20)  :====================
4 ( 6)  :======
5 ( 1)  :=
6 ( 2)  :==
7 (30)  :==============================
8 ( 0)  :
9 ( 4)  :====
```

The histogram should be followed by a line telling you the mean, median (the middle value if sorted) and mode (most common) values of the data. You can either build these up as you read in the data, or calculate them from the data stored in your array, or a combination of both approaches as appropriate.

A selection of text files for you to test your program on can be found at:

`http://g51prg.cs.nott.ac.uk/Distribution/Coursework/cswk4b.zip`