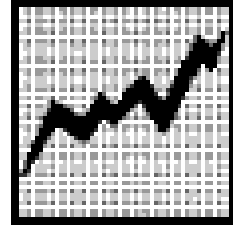# LAB EXERCISE

STATISTICS

Background:

1. Your instructor will provide you with a text file, (*Numbers.txt*), containing a large amount of integers. The integers range in value from 0 to 100. The text file has been created with each value separated by at least one space.

2. The number of integers in the file is unknown. You must read the text file until the "EOF" marker is encountered.

3. Your program must find the average, standard deviation, and mode of the list of numbers. The mode is defined as the value(s) present with the highest frequency. Calculating the standard deviation consists of the following steps:

   a. Find the average of the list of numbers.
   b. Determine the difference of each number from the average, and square each difference. Sum all the differences.
   c. Divide this sum by (the number of values - 1).
   d. Take the square root of the above division problem from step c.

   Example, given this list of numbers: 7 4 5 9 10

   a. The average $= 7$

   b. Sum of square of differences:
   $$(7 - 7)^2 \quad + \quad (4 - 7)^2 \quad + \quad (5 - 7)^2 \quad + \quad (9 - 7)^2 \quad + \quad (10 - 7)^2$$
   $$0 \quad + \quad 9 \quad + \quad 4 \quad + \quad 4 \quad + \quad 9 \quad = 26$$

   c. $\dfrac{26}{(5\text{-}1)} = 6.50$

   d. $\sqrt{6.50} = 2.55$

4. Miscellaneous info about standard deviation (not part of lab): For a normal distribution, 68.3% of the data will lie within one standard deviation of the average, while 95.4% will lie within two standard deviations – remember that bell curve thingy?

Assignment:

1. Your program should print out the average, standard deviation, and mode(s) of the data in *Numbers.txt*. <u>Format the real numbers to print with 2 decimal places</u>. Clearly label all output.

2. Your program must utilize proper modular design and parameter passing.

(See back for **Bonus**)

Bonus: Print a Histogram of the frequency counts. It is easiest to describe what to do by example:

```
                        *
     *                  *                                           *
*    *                  *                       *                   *
*    *      *           *   *                   *                   *
00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
<this is where a single blank line goes>
 *
 *                 *
 *           *     *
26 27 28 29 30 31 …50
<this is where a single blank line goes>
 *
 *   *
51…75
<this is where a single blank line goes>
      *
      *
 *    *
76..100
```

Since the range of numbers is 0..100, you'll need to break up the lines as shown
above (0…25, 26…50, 51…75, 76…100). There should only be one blank line between any
row of numbers and the highest star in the row under the numbers.