CS301/360 PROJECT #1
DESCRIPTIVE STATISTICS

CS301/360, AY 03-2

1. Project Objectives

1.1. **Ada 95 programming constructs.**

1. Demonstrate proficiency with sequence, selection, and iteration methods of program control.
2. Apply Ada95 exception-handling constructs.
3. Organize data using Ada95 arrays.
4. Define and utilize Ada95 subprograms – procedures and functions.

1.2. **Software development approach.** Apply a logical, structured approach to the development of this program.

2. Project Points

Project # 1 is worth 100 points toward your final course grade.

3. Deadline

3.1. **Interim submission (5 points).** You must submit a hardcopy of your stub code, i.e., your main subprogram and auxilliary subprogram shells, to the section marcher before the start of lesson 18.

3.2. **Final submission (95 points).** Hard and softcopy must be turned in to your instructor and submitted electronically not later than 1500 on 27 March.

4. Deliverables

4.1. **Hard copy (project folder).** A brown project folder is required. You may bring it to class, or drop it in the box outside your instructor's office (TH1115 = LTC Nash, TH1108 = Dr. Blair).

4.2. **Soft copy.** Create a directory named `proj1` and ensure all source files for your project are stored there. Electronically copy the contents of your `proj1` directory to the CS301 turn-in directory by doing the following:

1. Change your working directory to the one that contains the `proj1` directory. In other words, you should be able to execute an `ls -F` command, and see in the resulting listing an entry `proj1/` (the trailing slash indicates that the entry is a directory).
2. Execute the command `sub cs301 proj1`.

5. Background

Mathematics is the language of science, and the language of nature. We use mathematical analyses to try to understand the things that we sense in our environment. As a prelude to understanding, we can employ mathematical means simply to *describe* the phenomena that we observe, and then use those descriptions to draw conclusions.

Descriptive statistics are values which one calculates to quanitify (usually in a somewhat coarse fashion) important aspects of a set of observations. We are most often concerned with three things: the data's central tendency (the value a given observation tends to be close to, more often than not); its variability (the extent to which observations stray from the usual value); and correlation (the degree of relatedness between pairs of simultaneous observations).

Suppose that we are given a set of observations $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$, where x values an independent variable and the y values are dependent (that is, the y values are the what we are trying to measure). Then we have the *sample mean*, $\bar{y}$, defined by Equation 1:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

Likewise, the *sample variance*, s^2, is defined by Equation 2:

$$s^2 = \frac{\left(\sum_{i=1}^{n} y_i^2\right) - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n-1} \tag{2}$$

Finally, *Pearson's correlation coefficient* (r) can be defined by the following terms:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{4}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{5}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \tag{6}$$

6. Requirements

Write a main subprogram named `proj1` which allows a user to input a data set of observations, each observation consisting of two floating point values. We may therefore think of each observations as being plottable on an x-y coordinate system. For each data set, you must calculate the mean, variance, and coefficient of variation, and display these quantities with 5 decimal digits of precision (five digits following the decimal point). Specifically, your program must:

1. Display a menu that allows a user to repeatedly select one of these alternatives, until the Exit alternative is chosen:
   (a) Accept input of a data set.
   (b) Display the most recently input data set.
   (c) Display only the mean.
   (d) Display only the standard deviation.
   (e) Display only Pearson's coefficient of correlation.
   (f) Display only the number of observations.
   (g) Display all descriptive statistics.
   (h) Display a ten-bin histogram.
   (i) Exit the program.
2. Inputs:
   (a) X-Y coordinate pairs. All points must be within the range $[-100.0, 100.0]$ in both the x and y coordinates. There must be at least two pairs, and no more than 25.
   (b) The user's indication of whether or not they wish to enter another data point (either the single character `y` or `n`).
   (c) The user's menu choice.

   If an input value is entered improperly, your program should output an appropriate message and re-prompt for the value.

   The ten bins of the histogram divide the range input values into equal sub-intervals. The first bin of the histogram counts the values that are greater than or equal to the smallest input, the second bin counts the number of values greater than or equal to the smallest plus the width of the interval, and so on.
3. Write an algorithm for the main program; you need not submit any algorithms that you may write for subprograms; only the algorithm for the main program will be graded. Express the algorithm using the pseudocode primitives set forth in Brookshear and those we spoke about during class.
4. Implement the design using Ada95. As an interim submission, create the stub code for the main program and all subprograms The stub code should contain the main subprogram's and supplementary subprograms' structures, including known declarations and parameters, less their body code. For example, the interim submission's programs only require a `NULL;` statement between their `BEGIN` and `END` statements.

5. Prepare a test plan as part of your design and complete the test plan as you implement your solution.

7. Sample program output

```
This program calculates descriptive statistics for data sets that
you input.

Make your choice from the menu below:

1.  Accept input of a new data set.
2.  Display the most recently input data set.
3.  Display only the mean.
4.  Display only the standard deviation.
5.  Display only Pearson's coefficient of correlation.
6.  Display only the number of observations.
7.  Display all descriptive statistics.
8.  Display a ten-bin histogram.
9.  Exit the program.

Enter your choice: 1

Enter the next X value: 1.0
Enter the next Y value: 2.7

Enter another observation? y

Enter the next X value: 5.0
Enter the next Y value: 2.8

Enter another observation? y

Enter the next X value: 2.3
Enter the next Y value: 4.1

Enter another observation? n

Make your choice from the menu below:

1.  Accept input of a new data set.
2.  Display the most recently input data set.
3.  Display only the mean.
4.  Display only the standard deviation.
5.  Display only Pearson's coefficient of correlation.
6.  Display only the number of observations.
7.  Display all descriptive statistics.
8.  Display a ten-bin histogram.
9.  Exit the program.
```

```
Enter your choice: 8

2.70: **
2.84:
2.98:
3.12:
3.26:
3.40:
3.54:
3.68:
3.82:
3.96: *

1.  Accept input of a new data set.
2.  Display the most recently input data set.
3.  Display only the mean.
4.  Display only the standard deviation.
5.  Display only Pearson's coefficient of correlation.
6.  Display only the number of observations.
7.  Display all descriptive statistics.
8.  Display a ten-bin histogram.
9.  Exit the program.

Enter your choice: 8
```