

Section Handout #5—Arrays

Since you just took the midterm this week, it makes sense to use the opportunity to introduce two problems relating to statistical processing using arrays. The first uses arrays in their most conventional application of storing a list of data values; the second also makes use of an array for tabulation, as described in Section 11.4.

For each of these problems, you should assume that you have access to a data file named `Midterm.txt`, which contains the scores on a hypothetical midterm exam, as shown on the right. You also have access to a method

```
public double[] readScoresArray(String filename)
```

that reads the contents of a file consisting of numeric scores and returns an array whose elements are initialized to those values. Thus, after calling

```
double[] scores = readScoresArray("Midterm.txt");
```

the array `scores` will be initialized to a 50-element array that looks something like this:

scores								
58	73	93	82	93.5	62	80	...	90
0	1	2	3	4	5	6		49

(The implementation of `readScoresArray` depends on familiarity with data files, which we will cover later in the course.)

1. Calculating exam statistics

Write a program that reads in the array of scores and then calculates two statistics for the data:

1. The *mean*, which is more conventionally known as the average.
2. The *standard deviation*, which measures the difference from the mean.

To calculate the standard deviation of a distribution x_1, x_2, \dots, x_n , you need to perform the following steps:

- a) Calculate the mean of the distribution (which you have already done).
- b) Go through the individual data items in the distribution and calculate the square of the difference between each data value and the mean. Add all these values to a running total.
- c) Take the total from step b and divide it by the number of data items.
- d) Take the square root of that value to get the standard deviation.

Midterm.txt

```
58
73
93
82
93.5
62
80
53
52
92
75
65
95
23
100
75
38
80
77
92
60
98
95
62
87
97
73.5
78
72
55
58
42
31
78
70
78
74
70
60
72.5
75
84
87
62
17
92
78
74
65
90
```

In mathematical form, the standard deviation (σ) is given by the formula

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

where μ signifies the mean. The Greek letter sigma (Σ) represents a summation of the quantity that follows, which in this case is the square of the difference between the mean and each individual data point.

2. Creating a histogram

Write a program that takes the score array read from the file `Midterm.txt` and then displays a histogram of those numbers, divided into the ranges 0–9, 10–19, 20–29, and so forth, up to the range containing only the value 100. If, for example, `Midterm.txt` contains the data shown on the preceding page, your program should be able to generate a histogram that matches the following sample run:

