CSCI 5015 Assignment 1
Due Feb 8, 11:59 PM.
50 points

Objective

Use Python to compute basic statistics for data in a file.

Background

You've been given a data file, data.txt, which contains data from experiments for a new environmental sensor. The sensor is still being developed and is being tweaked to find the settings for optimal performance. The project's engineers are running experiments to find settings that allow the sensor to work reliably and are recording their data in the file you've been given.

The data file has five columns of data. Each column represents the data collected during an experiment with the sensor. Column 1 represents experiment 1, column 2 represents experiment 2, and so on. All the values in the file are positive numbers greater than or equal to 0. Each number is a float and all values have 5 digits of precision after the decimal point.

The engineers want you to calculate the following statistics for each column / experiment:

- The maximum value, max, found in each column
- The minimum value, min, found in each column
- The range of values, $range$, found in each column, where
  $range = max - min$
- The sum, sum, of all the values in each column, where
  $sum = \sum_{i=1}^{N}x_i$
- The average value, $\bar{x}$, of each column, where
  $\bar{x} = \frac{\sum_{i=1}^{N}x_i}{N}$
- The standard deviation, s, of each column, where
  $s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$
- The variance, s^2, of each column, where the variance is square of the standard deviation, s.

Coding instructions

Create a Python file named **assignment1.py** and place the data file **data.txt** in the same folder. Add a comment at the top of your file with your name, for example

```
# Written by John Nicholson
```

Your program should open the data file for reading and compute the seven required values for each column/experiment. Since there are 5 columns, you will be computing 35 values in total.

Output the results to a second file named **results.txt**. Ensure that each value is labeled so that the person reading the file knows which statistic - max, min, $range$, sum, $\bar{x}$ (average), s (standard deviation), and s^2 (variance) - and which experiment - 1, 2, 3, 4, and 5 - is associated with each value. Add a statement at the top of the file with your name. Output the data in a format similar to the following (obviously replacing XXXX for real numbers):

```
Prepared by John Nicholson

Experiment 1
   max:      XXXX
   min:      XXXX
   range:    XXXX
   sum:      XXXX
   average:  XXXX
   stddev:   XXXX
   variance: XXXX

Experiment 2
   max:      XXXX
   min:      XXXX
   range:    XXXX
   ...
```

The purpose of this assignment is to ensure you can code basic Python, so you should not not use any external library. Therefore, there is one restriction on your program: you may not have any `import` statements in your code, except for the one required to use `sqrt()`, the square root function. Here is some sample code demonstrating the function:

```
from math import sqrt
```

```
x = 4
y = sqrt(x)

print("The square root of {} is {}".format(x, y))
```

Otherwise, you can complete this assignment using the topics we discussed in the first two lectures. You can use other features of Python if you wish, as long as they don't require an import statement.

Tips and help

The data file is big. You might want to create a similar, yet smaller file, to test you program on. This way you can place values in the small file that you know the answers for. Write your program and test it on that file. Once you are sure your program works on the small file, change it to work on the big file.

You can always come see me in my office or send me email. If you send me email, you should send your code as an attachment. Don't copy/paste your code into the message because that will make it harder for me to debug your code.

Submission

When your program is correct, log into D2L and locate the Dropbox for assignment 1. Upload the two files

- assignment1.py
- results.txt

into the D2L dropbox.

Contact me if you have any problems.