Life Technology Solutions - Coding Exercise

**Statistics Generator**

# Goal

Write a command line tool that calculates a set of specified statistics on some given data.

# Definitions

Scenario – A possible course of action or events. Specific scenarios might be: Interest Rate Up, Interest Rate Down.

Variable – A specific named value within a scenario. Specific variables are: Average Policy Loan Yield, Cash Premium.

Period – A specific point in time, at a certain interval. Periods can be Monthly, Quarterly, Yearly, or anything in-between. The columns should be read as the value (of the variable for the scenario) in the 000th period, etc.

Value – A numeric value for the variable within a scenario for a specific period.

# Input

There are two files consumed as input to the command line tool. The TotalTemp file contains the data that will be calculated; the Configuration file contains the calculations that will need to be performed.

## TotalTemp file

The TotalTemp file is output that is created by a different tool. It represents values of variables within scenarios across periods. The TotalTemp file is in the following format:

* Tab-delimited text file.
* .TXT file extension.
* The first line of the file is the header.
* Each subsequent line contains the values for the scenario/variable combination for the periods.
* The scenario is identified by the ScenId column.
* The variable is identified by the VarName column.
* The periods are identified by the Value### columns.

The number of scenarios, and the variables within the scenarios, differ from file to file. Also, although a single file will contain the same number of periods for each row, the number of periods differ from one file to the next.

## Configuration file

The Configuration file will instruct your tool about what calculations to perform on which variables and values. The Configuration file is in the following format:

* Tab delimited text file.
* .TXT file extension.
* No header.
* Each line contains the variable name, statistic calculation, and period choice, as follows:

Statistic Calculation

* + MinValue
  + MaxValue
  + Average

Period Choice

* + FirstValue (the value in the first Period - i.e. Value000)
  + LastValue (the value in the last Period - i.e. ValueNNN)
  + MinValue (the minimum value in any Period for the scenario/variable combination)
  + MaxValue (the maximum value in any Period for the scenario/variable combination)

# Output

## Results file

The implemented tool will produce a results file within the following specifications:

* Tab delimited text file.
* .TXT file extension.
* No header.
* Each line contains the variable name, statistics calculation, and calculated value.

# Constraints

Imagine that reading the TotalTemp file is expensive (in terms of time, storage access, or computational cost). For this reason, we are imposing restrictions that you will have to work around to accomplish your goal, including:

* You must only read the TotalTemp file once.
* You must not store the entire contents of the file into memory. (The easiest way, and an acceptable solution, is to process the TotalTemp file one line at a time, for example, iterating over System.IO.File.ReadLines(string).)

# Example

Let’s walk through a simple example to help clarify how to process the data. Below is a simplified extract of the data. The column headers displayed below are different (for clarity of this definition) than the actual TotalTemp file. There are also a limited number of scenarios and variables, again for clarity for this example.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ScenId** | **VarName** | **Value000** | **Value001** | **Value002** | **Value003** | **Value004** | **Value005** |
| 1 | AvePolLoanYield | 0.00 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 |
| 1 | CashPrem | 0 | 165215335.38 | 130922548.81 | 107196660.00 | 92462698.42 | 84655947.13 |
| 1 | ResvAssumed | -27923645.44 | -28437248.89 | -29893491.30 | -31761676.09 | -34092668.16 | -36815307.05 |
| 2 | AvePolLoanYield | 0.00 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 |
| 2 | CashPrem | 0 | 0 | 130922548.81 | 107196444.36 | 92462698.42 | 84655914.86 |
| 2 | ResvAssumed | -27923645.44 | -28437248.89 | -29893531.02 | -31762115.98 | -34094542.44 | -36821010.24 |
| 3 | AvePolLoanYield | 0.00 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 |
| 3 | CashPrem | 0 | 0 | 0 | 107196660.00 | 92462698.42 | 84655947.13 |
| 3 | ResvAssumed | -27923645.44 | -28437248.89 | -29893482.02 | -31761477.70 | -34091316.73 | -36811494.23 |

We load into memory the Configuration file, which contains a single line: ‘CashPrem,Average,MaxValue’. This means that, for each CashPrem row in the TotalTemp file, we want to determine the maximum value in the periods for that row. From the resulting values, we want to calculate the average.

We begin by looping over the TotalTemp file one row at a time (as per our constraints). The first row is for AvePolLoadYield, but since we do not have any calculations to do for this variable, we can move on to the next one.

The second row is for CashPrem, which is something we are interested in. Now we need to determine which Period’s value to use. The Configuration file specified MaxValue, so we need to remember 165215335.38 (from Value001).

We continue to loop through the TotalTemp file until we reach the end. When we finish, we have the following information:

|  |  |  |
| --- | --- | --- |
| **ScenId** | **VarName** | **{MaxValue}** |
| 1 | CashPrem | 165215335.38 |
| 2 | CashPrem | 130922548.81 |
| 3 | CashPrem | 107196660.00 |

Now we want to calculate the Average of the values:

|  |  |
| --- | --- |
| **VarName** | **{Average}** |
| CashPrem | 134444848.1 |

So the output for our example would be: ‘CashPerm,Average,134444848.1’.