

Programming Exam 1

Instructions. This exam has one question. You have 50 minutes. The exam is *open course materials*, which includes the course textbook, the companion booksite, the course website, your course notes, and code you wrote for the course. Accessing other information or communicating with a non-staff member (such as via email, instant messenger, text message, Facebook, Piazza, phone, or Snapchat) is prohibited.

Submission. Submit your solution electronically, via the link on the *Class Meetings* page. Be sure to click the *Check All Submitted Files* button to verify your submission.

Grading. Your program will be graded for correctness, clarity (including comments), design, and efficiency. You will receive partial credit for a program that correctly implements some of the required functionality. You will receive a substantial penalty if your program does not compile or if you do not follow the prescribed input/output specifications.

Discussing this exam. Discussing or communicating the contents of this exam before solutions have been posted is a violation of the Honor Code.

This exam. You must turn in this exam. Print your name, NetID, and precept in the space below. Write and sign the Honor Code pledge before turning in this exam.

Name:

NetID:

Precept:

“I pledge my honor that I have not violated the Honor Code during this examination.”

Problem. Write a program `GStatistic.java` that prints the *degrees of freedom* and the *G-statistic* for a sequence of observed frequencies (x_1, x_2, \dots, x_n) and a sequence of expected frequencies (y_1, y_2, \dots, y_n) .

- The *degrees of freedom* is the one less than the number of non-zero entries among (y_1, y_2, \dots, y_n) .
- The *G-statistic* is given by the formula:

$$G = 2x_1 \ln\left(\frac{x_1}{y_1}\right) + 2x_2 \ln\left(\frac{x_2}{y_2}\right) + \dots + 2x_n \ln\left(\frac{x_n}{y_n}\right)$$

If $x_i = 0$, then you should treat the term $2x_i \ln(x_i/y_i)$ as 0. Here, $\ln()$ denotes the natural logarithm function; recall that `Math.log(z)` returns that natural logarithm of z .

Step-by-step calculation (for reference). Here is an example with $n = 6$, where the individual $2x_i \ln(x_i/y_i)$ terms are shown (to 4 digits of precision after the decimal point):

i	x_i	y_i	$2x_i \ln(x_i/y_i)$
1	125202	125000	404.3263
2	125144	125000	288.1658
3	124877	125000	-245.8789
4	125698	125000	1399.8904
5	124798	125000	-403.6734
6	374281	375000	-1436.6206

$$\begin{aligned} \text{degrees of freedom} &= 6 - 1 \\ &= 5 \end{aligned}$$

$$\begin{aligned} G\text{-statistic} &= 404.3263 + 288.1658 - 245.8789 + 1399.8904 - 403.6734 - 1436.6206 \\ &= 6.2096 \end{aligned}$$

Input specification. There are no command-line arguments. The input from standard input consists of an integer n followed by n lines, where each line contains two integers (the observed frequency, followed by the corresponding expected frequency), separated by whitespace. Here are two examples in the prescribed format: `loadedDie1.txt` corresponds to the example on the facing page; `loadedDie2.txt` includes some entries that are zero.

% more loadedDie1.txt		% more loadedDie2.txt	
6		6	
125202	125000	125314	125000
125144	125000	124746	125000
124877	125000	374396	250000
125698	125000	0	0
124798	125000	0	125000
374281	375000	375544	375000

Note: You may assume that all entries are nonnegative integers.

Output specification. The output to standard output is the degrees of freedom (an integer) followed by the G -statistic (a floating-point number, with 4 digits of precision after the decimal point).

```
% java-introcs GStatistic < loadedDie1.txt
degrees of freedom = 5
G-statistic        = 6.2096

% java-introcs GStatistic < loadedDie2.txt
degrees of freedom = 4
G-statistic        = 303612.0960
```

Do *not* print any other output on standard output.

Test files. For convenience, the files `loadedDie1.txt` and `loadedDie2.txt` are available at

```
http://introcs.cs.princeton.edu/java/loadedDie1.txt
http://introcs.cs.princeton.edu/java/loadedDie2.txt
```

Submission. Submit the single file `GStatistic.java` via Dropbox at

```
https://dropbox.cs.princeton.edu/COS126_S2016/Exam1
```

Context (optional). *You could use the G -statistic to determine whether a die is loaded according to a $(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{3}{8})$ -distribution by rolling it 1 million times and comparing the observed frequencies with the expected frequencies. A G -statistic far from zero (relative to the degrees of freedom) indicates that the observed frequencies are inconsistent with the distribution defined by the expected frequencies.*