


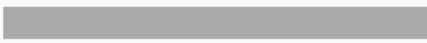

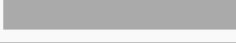





Lab Exercise #7

Assignment Overview

This lab exercise provides practice with lists, functions and namespaces in Python.

Benford's Law

Benford's Law (sometimes called the First-Digit Law) is an observation about the distribution of first digits in many data sets: the number 1 occurs as the leading digit about 30% of the time, while larger numbers occur in that position less frequently (for example, 9 occurs as the first digit less than 5% of the time). As noted on Wikipedia: This result has been found to apply to a wide variety of data sets, including electricity bills, street addresses, stock prices, population numbers, death rates, lengths of rivers, physical and mathematical constants, and processes described by power laws (which are very common in nature).

<i>d</i>	<i>P(d)</i>	Relative size of <i>P(d)</i>
1	30.1%	
2	17.6%	
3	12.5%	
4	9.7%	
5	7.9%	
6	6.7%	
7	5.8%	
8	5.1%	
9	4.6%	

Develop a program which will allow you to experiment with Benford's Law. The program will prompt the user for the name of an input file, read the contents of the file and count the leading digits, and then display the results. For simplicity, the program will assume that the data file contains integer numbers, with one value per line.

The program's output will be formatted as shown below (see below). **Hint:** use the header line to determine formatting values.

```
Enter a file name: city_all.txt
```

```
Digit Percent Benford
```

1:	32.7%	(30.1%)
2:	19.9%	(17.6%)
3:	13.3%	(12.5%)
4:	9.1%	(9.7%)
5:	5.8%	(7.9%)
6:	4.6%	(6.7%)
7:	5.2%	(5.8%)
8:	5.0%	(5.1%)
9:	4.5%	(4.6%)

Notes and Suggestions

1. The directory for this lab exercise contains three data files which you can use as you develop your program. Mimir will test all three input files.

<code>city_part.txt</code>	the first 10 lines of the following file (a small file for testing)
<code>city_all.txt</code>	the population of every city and town in Michigan (1500+ lines)
<code>warblers.txt</code>	decades of counts of warbler observations (540,000+ lines)

2. There are three phases to this program. What are they? Which should you do first?

3. **Requirement:** you need a counter for each digit so use a **list** of counters and use the digit read from the file as an index into your list of counters. **BEWARE:** be careful that you do not count values with a zero as the leading digit—that can throw off your percentage calculation!

4. **Requirement:** your program must have a function `open_file` that continuously prompts for a file name until it opens and returns the file pointer.

3. The string method “`strip`” is useful for discarding leading and trailing whitespace.

4. The string method “`isdigit`” is useful to ensure that your program only processes integer numbers.

5. The string method “`format`” is useful to align the output in columns.

6. Your program does not need to somehow compute the Benford values (e.g. 30.1%, 17.6%, ...). Instead, you may use the values from the chart (or sample output).