Name: _____*Caleb McWhorter — Solutions*_____

MAT 222

Spring 2019

Excel Lab 3: Ch. 9

Benford's Law (also known as the First-Digit Law or Law of Anomalous Numbers) is an observation that the frequency distribution of leading digits of 'real world' data follows a certain distribution. Specifically, it states that for 'real world' data, the leading digit of numbers follows a distribution given by $\log_{10}(1 + 1/n)$, where n is the leading digit. For example in 'real world' data, the digit '1' occurs as the first digit of a number in a data set with proportion $\log_{10}(1 + 1/1) = \log_{10}(2) \approx 0.301$.

This is often used in finance to detect fraudulent data; that is, if the leading digits of financial data do not follow Benford's Law, one might be suspicious that the data is fabricated, though it does not *prove* that it is fraudulent. In the file `leading_digits.xlsx`, using the calculations for Dataset 1 as a template, perform a χ^2 Goodness of Fit analysis to predict whether each of the datasets is most likely a 'real world' dataset or a faked dataset. [Do not forget to make the prediction for Dataset 1.] Record the X^2 test statistic for each dataset and your predictions below.

Dataset	X^2 Statistic	Real/Fake
1	0.4747	Real
2	≈ 0	Fake
3	≈ 0	Fake
4	0.8099	Real

Note: Dataset 1 was actually distances (in light years) from Earth to various stars taken from `http://www.atlasoftheuniverse.com/stars.html`*. Dataset 2 was average US city temperatures taken from* `https://www.currentresults.com/Weather/US/average-annual-temperatures-large-cities.php`*. The final two datasets were both faked, one completely randomly and the other, Dataset 4, was faked to seem real. This goes to show that while Benford's Law can help indicate whether a dataset is real or fake, it is not conclusive evidence of either. It is merely a tool.*