Benford's Law

Introduction

Benford's Law (which was first stated by Simon Newcomb in 1881) states that if you randomly select a number from a table of physical constants or statistical data, the probability that the first digit will be a "1" is about 0.301, rather than 0.1 as we might expect if all digits were equally likely. In general, the "law" says that the probability of the first digit being a *d* is

$$P(d) = \log(1 + \tfrac{1}{d})$$

This implies that a number in a table of physical constants is more likely to begin with a smaller digit than a larger digit. It was published by Newcomb in a paper entitled "Note on the Frequency of Use of the Different Digits in Natural Numbers", which appeared in The American Journal of Mathematics (1881) 4, 39-40. It was re-discovered by Benford in 1938, and he published an article called "The Law of Anomalous Numbers" in Proc. Amer. Phil. Soc 78, pp 551-72.

1. Use this formula to find the probability of the first digit being a 1, 2, …,9.

2. Using a table of physical or statistical data that contains at least 100 data values (an atlas would be useful here), determine the number of leading digits that are a 1, a 2, etc. Compare your results to the results expected if Benford's Law applies.

3. Visit the website to learn how Benford's Law is used to detect corporate and scientific fraud.

Downloading a TI-8x program file

Download a TI-8x program to a Windows computer as follows:
1. Right-click on the program name.
2. Click on "Save Target As …"
3. Change to the correct directory.
4. In the File edit box, type the filename, *with the extension* (e.g. ben.83p)
5. Change "Save as Type:" to "All Files"
6. Click on [Save]. The file is now saved
7. Use the appropriate GraphLink program to upload the program to your graphics calculator.

The Programs

To use the programs BEN.83P, BENFORD.83P, BEN.8XP and BENFORD.8XP:
1. Enter your data in list L1.
2. Run the program.

The programs BEN.83P (for the TI-83) and BEN.8XP (for the TI-83 Plus) calculate the relative frequency of each leading digit. List L1 contains the original data, list L2 contains the digits 1 to 9 and list L3 contains the relative frequency of each of these digits.

You can draw a scatterplot of the data using Xlist = L2 and Ylist = L3. If you set Y1 = log(1 + 1/x) you can see how well the dataset fits the model proposed by Newcomb and Benford.

The programs BENFORD.83P (for the TI-83) and BENFORD.8XP (for the TI-83 Plus) do the following:
- ❖ calculate the relative frequency of each leading digit.
- ❖ draw the scatterplot of the relative frequency of each leading digit, along with the function Y1 = log(1 + 1/x)
- ❖ draw the residual plot using the function Y1 = log(1 + 1/x) as the line of best fit.