Sequoia

Curiosity Corner

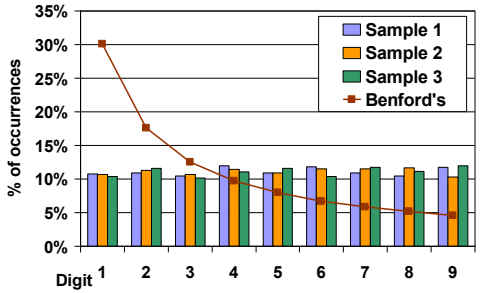
Logarithms, rivers and the company accounts

How robust is your sales data? A little-known mathematical theory could help you find out.

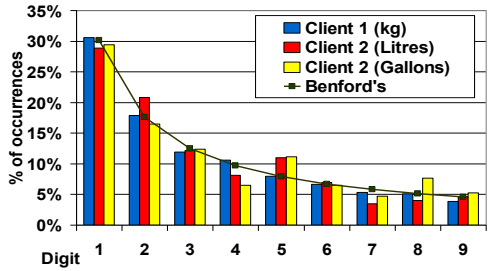
Take a set of figures, like monthly production costs, or areas of the world's rivers, or the specific heat of chemical compounds. If you look only at the first, or leading, digit of each number in the set, you might reasonably expect that one-ninth would be 1s, one-ninth would be 2s and so on. Yet this is often not the case. In fact, it's more likely that nearly a third of all the leading digits will be 1s, while less than 5% will be 9s. This is because many data sets conform to Benford's Law, which states that the probability of each digit occurring is not random but is governed by the logarithmic formula:

$$P(d) = \log_{10}(1+1/d)$$

where P is the probability and d is the digit. Although academics are still debating why this should be, they seem to agree that the law is most useful when considering data formed from a combination of other data – for example, manufacturing costs, which are composed of various fixed and variable elements; or financial accounts. In fact, Benford's Law is increasingly used by auditors to analyse financial data for fraud, especially since the Enron scandal brought about the collapse of Arthur Andersen. Just to show that it works, the graphs below show Benford's predictions for the distribution of leading digits compared against three sets of randomly-generated numbers:



and against three sets of real production data from our clients:



The Client data shows a good fit to the Benford's Law prediction, while the random numbers don't. So, if you're worried about the accuracy of your production figures, or concerned about your financial accounts, there's a simple way to start the validation process – check the leading digits against Benford's Law to see if they fit.