BENFORD'S LAW

Professor Brian Bushee

Benford's Law

- **Let's throw away all notions of trying to model discretionary earnings or fraud prediction ratios**
  - **Instead, we'll let the data speak as to whether it is true or manipulated**
- **Benford's law**
  - **In 1881, Simon Newcomb determined that the probability that a number has a first digit, d, is given by**
    - **Prob(d) = $Log_{10}(1+(1/d))$, for d=1, 2, …, 9**
  - **In 1938, Frank Benford found that a large number of naturally-occurring datasets follow this pattern**
    - **Surface areas of rivers, molecular weights, death rates, street addresses, and the numbers contained in an issue of Reader's Digest**
  - **"Benford's Law" has been used to detect irregularities in**
    - **Published scientific studies, fraudulent election data in Iran, suspicious macroeconomic data from Greece, and tax return misreporting**

Benford's Law and Bernie Madoff

- **Here is the expected distribution of leading digits under Benford's law:**

Leading digit	1	2	3	4	5	6	7	8	9
Frequency	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

- **Here is the distribution of leading digits from 215 months of returns for the Fairfield Sentry Fund, a fund that invested only with Bernie Madoff:**

Leading digit	1	2	3	4	5	6	7	8	9
Frequency	39.6%	14.2%	10.4%	7.1%	7.5%	6.6%	6.1%	6.6%	1.9%

Source: (Dan Amiram, Zahn Bozanic, and Ethan Rouen 2014)

Benford's Law and Financial Statements

- Amiram, Bozanic, and Rouen (2014) studied Benford's Law in financial statements
- Aggregating all financial statements by industries or by years for 2000-2011, the leading digits follow Benford's distribution
  - 83.7% of firm's annual financial statements conform with the distribution
- Larger discrepancy versus Benford's distribution is associated with
  - Larger Modified Jones Model discretionary accruals
  - Higher Beneish M-Score

Detecting Discrepancies from Benford's Law

- **Enter a "large" number of financial statement items**
  - **No theory on how many are needed, but more is better**
- **Count the number of each leading digit**
  - **Find leading digit with Excel formula: =LEFT(ABS("cell"*1000),1)**
  - **Count leading digits with formula: =Countif("range", "digit")**
- **Compare the actual frequency of leading digits to the expected distribution according to Benford's law**
  - **Kolmogorov-Smirnov (KS) statistic**

$$\mathrm{KS} = \max\{|\mathrm{AD}_1 - \mathrm{ED}_1|, |(\mathrm{AD}_1 + \mathrm{AD}_2) - (\mathrm{ED}_1 + \mathrm{ED}_2)|, \ldots, |(\mathrm{AD}_1 + \mathrm{AD}_2 + \cdots + \mathrm{AD}_9) - (\mathrm{ED}_1 + \mathrm{ED}_2 + \cdots + \mathrm{ED}_9)|\}$$

    - **Where AD_i is actual frequency of the leading digits, and ED_i is the theoretical frequency of the leading digits**
  - **Potential manipulation if KS > 1.36/√P, where P is the total number of leading digits used**


Wharton
University of Pennsylvania