



Testing Benford's Law with Software Code Counts

Chuck Knight
Chris Kaldes

ICEAA Conference
June 2014



Agenda

- Introduction
- What is Benford's Law?
- Situation and Problem
- Testing the law with code counts
- Practical applications for cost estimating
- Questions



- 2 -

Introduction



Chuck Knight
CCE/A
chuknight@deloitte.com



- 5 years of cost estimating, earned value, and acquisition experience
- 2 years at Deloitte Consulting and 3 years at the Department of the Navy
- Currently providing cost estimation support in the Intel community

Chris Kaldes
CCE/A, PMP
ckaldes@deloitte.com



- 22 years of cost estimating experience
- 4 years at Deloitte Consulting
- CCE/A in 1997
- Currently providing cost estimation support at DoD and USPS



- 3 -

Basics of Benford's Law



Benford's Law looks at the leading digit of each data point in any real data set. The law simply states to then look at the frequency of each leading digit. In other words, how many data points start with one? How many start with two? And so on.

If you were to create a histogram of the frequency of each leading digit (one through nine), what pattern, or distribution, do you think you would see? (e.g. uniform, normal)

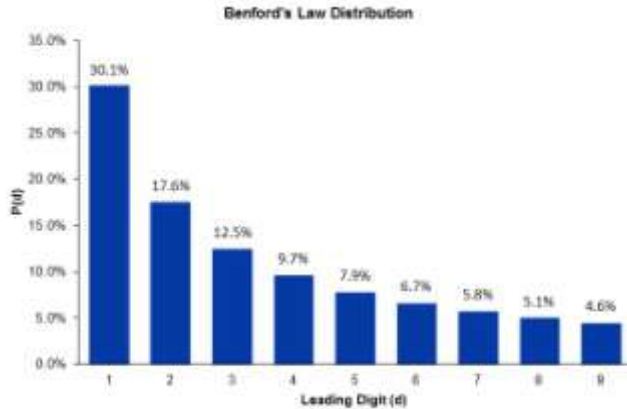


- 4 -

Basics of Benford's Law

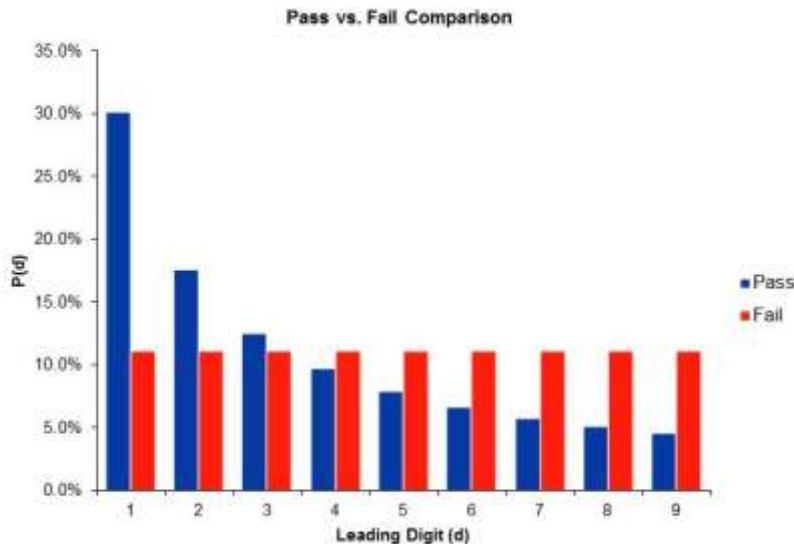


Given any real data set, Benford's Law states that the leading digit of each data point should exhibit a distinct frequency pattern where the number one occurs more often than the number two which occurs more often than the number three and so on.



- 5 -

Benford's Law: Pass vs. Fail Comparison



- 6 -

How can we use Benford's Law as cost estimators?



Situation: Tasked to create a life cycle cost estimate (LCCE) for a ground system which will primarily consist of software development. The primary basis of estimate are source lines of code (SLOC) count estimates provided in the Cost Analysis Requirements Description (CARD). Assume the CARD contains best available data.

Problem: How confident are you in the SLOC estimates provided in the CARD? If confidence is low, what basis do we have to challenge the estimates?

Hypothesis: Actual SLOC counts (real data set based on software that has already been developed) will follow Benford's Law.

Question: If hypothesis is true, should the same logic be applied to the SLOC estimates from the CARD that are used as primary basis of the LCCE?

Deloitte.

- 7 -

Ground Rules and Assumptions



General:

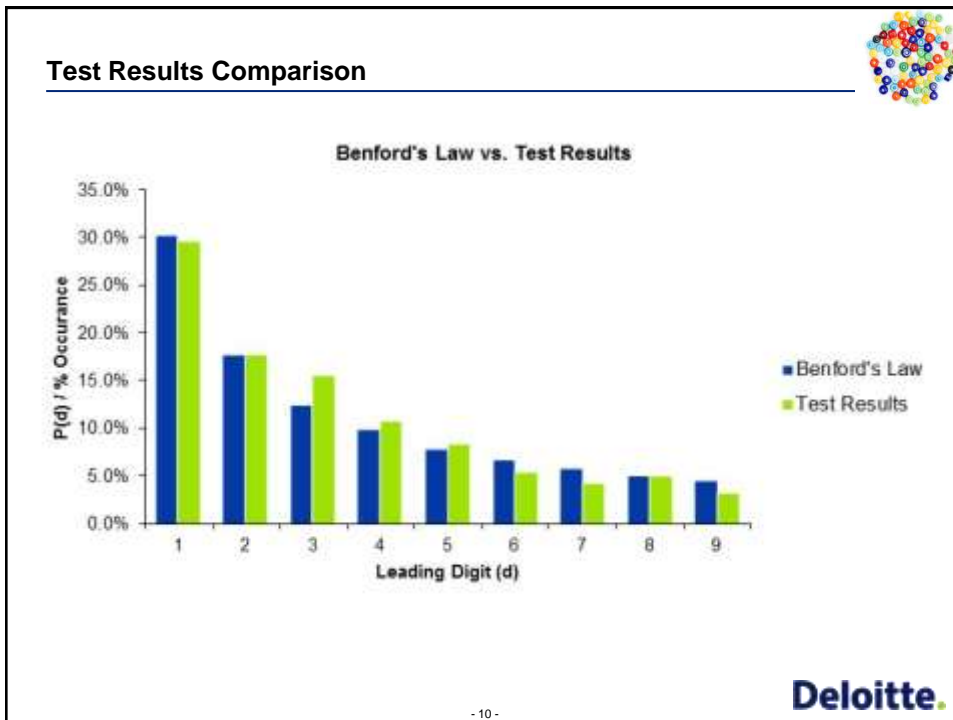
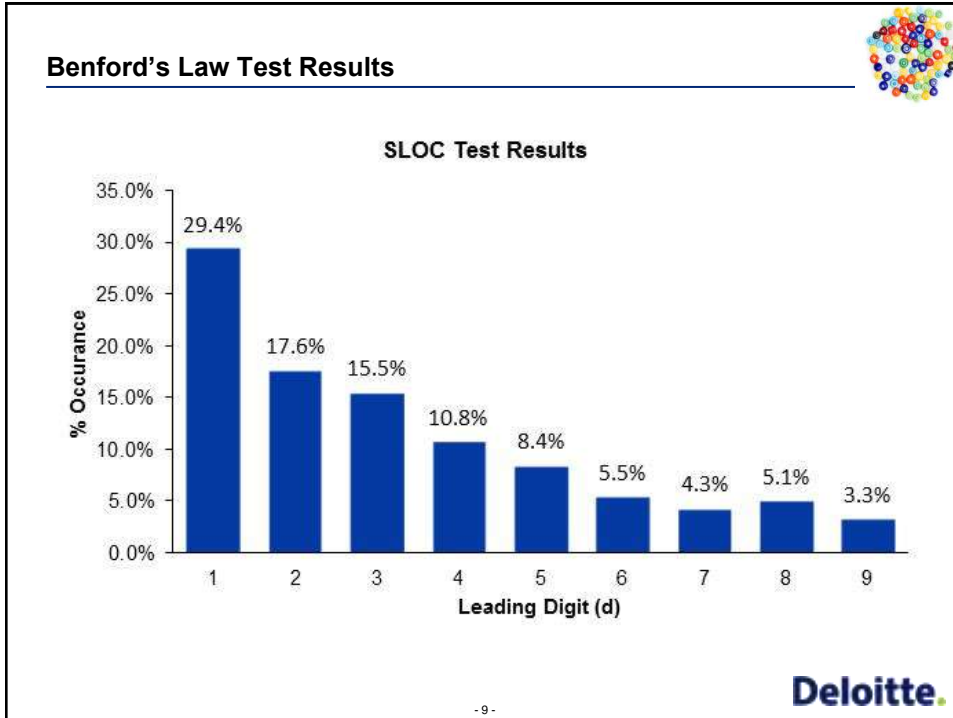
- Data set must consist of real data
- Each data point is independent of each other
- Zero or non-value data points are excluded

SLOC Test Specific:

- 510 data points (Actual SLOC counts from final delivered software)
- Counts are at lower levels (CSCI or CSC), not at a system or program level

Deloitte.

- 8 -



Conclusion



If we believe that the actual SLOC counts will follow the behavior consistent with Benford's Law, then we should stress that the SLOC estimates at the beginning of the LCCE development should also follow that behavior.

This test is meant to be a quick and easy cross check for cost estimators who may lack subject matter expertise in technical areas that they have been tasked to estimate.

If input data does not pass the Benford's Law test, it can provide the cost estimator with a starting point to go back to the engineers or SME's and explore the basis of estimate for those inputs.

At the very least, this test can help provoke additional thought around inputs which will help make an estimate more defensible.

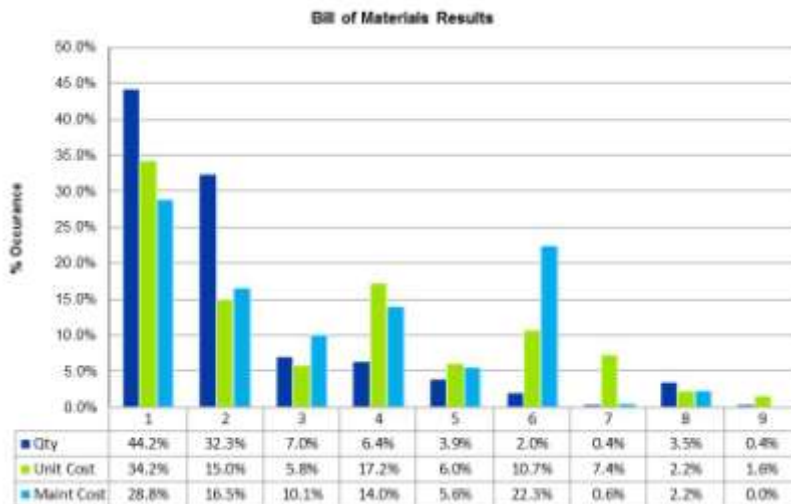
This test can help increase stakeholders' confidence in technical inputs



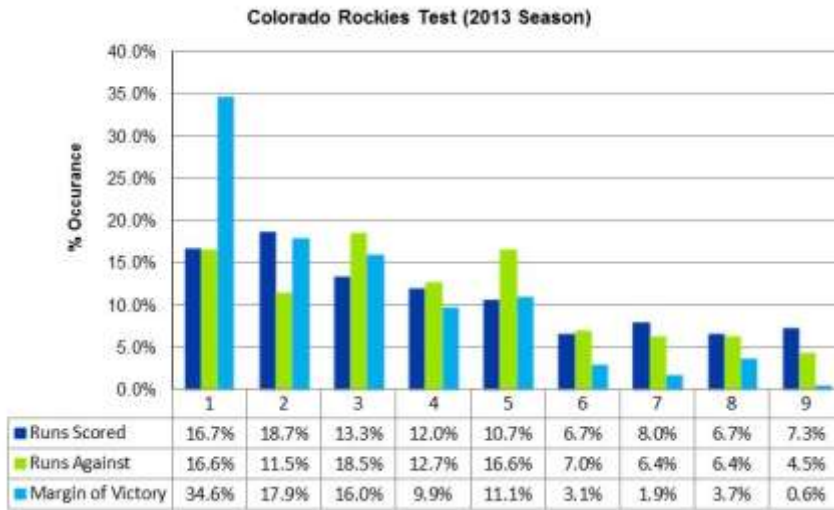
Discussion: Practical Applications for Cost Estimating



What other areas might we be able to use this test?



Other Benford's Law Examples



- 13 -



Wrap Up



Questions?

- 14 -

