ARIZONA Auditor General
Making a Positive Difference

FRAUD PREVENTION ALERT
Data Analytics: Part 3
July 2019

Why are we issuing this alert?

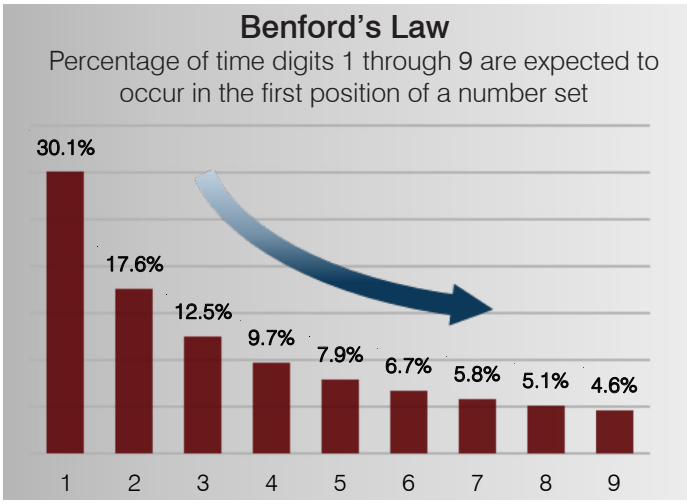
As presented in our previous 2 Fraud Prevention Alerts, using risk-based assessments and data analytic techniques to identify trends, patterns, anomalies, and abnormal relationships in data can be an effective anti-fraud control.[1] Data analytic techniques can be separated into 3 categories for practical applications: basic, statistical, and advanced. This alert, the third in a 4-part series, focuses on the practical applications for statistical techniques and describes how using these techniques on an organizations' data and investigating associated findings can help public officials identify fraudulent activity before it becomes significant.

What are statistical data analytic techniques?

Statistical data analytic techniques identify data that deviates from the expected values. Deviations could indicate an error, irregularity, or fraud, and should be further investigated. Although there are many statistical data analytical techniques, they all generally involve comparing actual data with a set of expected values for that data. This alert focuses on 2 methods: the first, known as Benford's Law, identifies anomalies in large data sets, and the second, called regression analysis, determines the relationship between 2 or more variables.

Benford's Law

Benford's Law maintains that within large data sets of real-life numbers such as invoice amounts, numbers are statistically more likely to begin with small digits like 1 or 2 rather than large digits like 8 or 9. Benford's distribution gives a probability of obtaining digits 1 through 9 in the leading position of a number set. For example, as shown in the chart, data that follows Benford's Law should have a leading digit of 1 about 30.1 percent of the time, a leading digit of 2 about 17.6 percent of the time, and each subsequent numeral, 3 through 9, will be the leading digit with decreasing frequency.[2] Two common tests that utilize Benford's Law are described below.



First digit test—The first digit test using Benford's Law focuses on the first digit in number amounts. Specifically, a transaction of $147.65 has a first digit of 1, $29.74 has a first digit of 2, and so on. The first digit test is a high-level test of reasonableness and will identify only obvious anomalies. To illustrate, if a Benford analysis of 5,000 randomly selected vendor invoices showed that 25 percent of transactions had a first digit of 4 (more than the expected distribution curve of 9.7 percent), that group of transactions is anomalous and has a higher risk of fraud. That said, it is not always efficient to use the first digit test to select targets for sampling, as the sample size will usually be too large, and the results should be further reviewed as described in the first 2 digits test below.

---

[1] See Office of the Auditor General, *Fraud Prevention Alert—Data Analytics Part 1*, November 2017, Report 17-406 and *Fraud Prevention Alert—Data Analytics Part 2*, October 2018, 18-406.

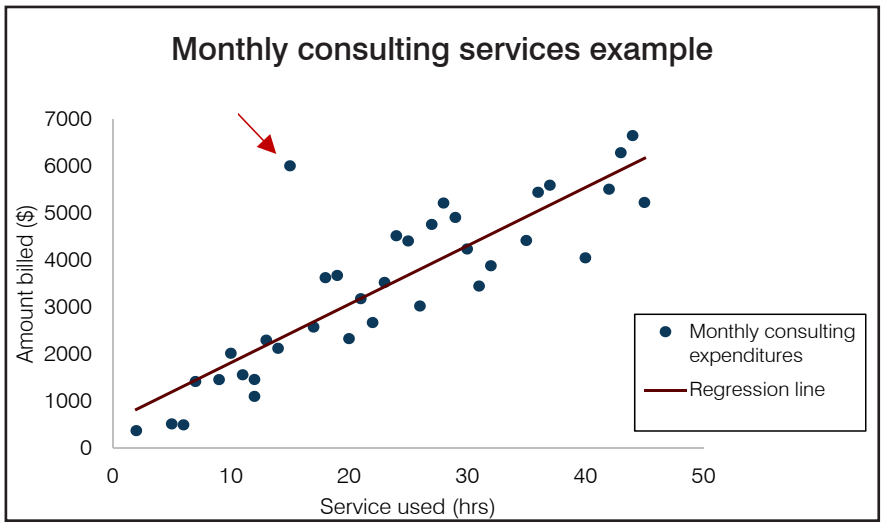
[2] Collin, J.C. (2017). *Using Excel and Benford's Law to detect fraud*. Retrieved 5/6/2019 from https://www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.

First 2 digits test—The first 2 digits test is a more focused application of Benford's Law that may be used to select a smaller data set. For example, if a first digit test on 5,000 vendor invoices showed more invoices beginning with 4 than expected, the first 2 digits test could find anomalies on that subset of invoices. As with the first digit test, an expected probability distribution for the first 2 digits of transactions can be used to compare actual data with their expected values. Anomalous transactions at this filtered level produce a more efficient audit sample to investigate.

Regression analysis

Regression analysis is a statistical method that examines the relationship between 2 or more variables. Analysts try to understand or predict one of the variables—the outcome variable, while expecting the other variables—predictor variables—to impact the outcome variable.

For example, regression analysis can evaluate the reasonableness of an organization's monthly consulting service costs (outcome variable). As shown in the graphic, monthly consulting service billed hours (predictor variable) are plotted on a graph along with the associated billing amount establishing a line of best fit, also known as a regression line. The regression line allows an organization to see any months in the period under examination where the relationship significantly deviated from expectations. As the graphic illustrates, points above the regression line represent cases where billed amounts were higher than expected, given the number of hours billed. Moreover, one clear outlying point shows where approximately 15 hours were billed and expenses were $6,000, equating to about $400 per hour. The analysis shows that the expected per hour rate should be about $144. Therefore, further investigation is required to determine the monthly cost's propriety.



Recommendations

To help protect public monies, public officials should establish proactive data monitoring and analysis programs designed to deter and detect fraud. Specifically, public officials should:

- Review all areas of operation and determine areas that may be susceptible to fraud, using a risk-based approach to prioritize the most vulnerable areas.
- Develop and regularly perform data analyses to detect fraud in the identified areas.
- Investigate findings in a timely manner using trained, knowledgeable employees to interpret results.
- Communicate the results to management, who should take appropriate action.