11th MALAYSIA STATISTICS CONFERENCE 2024
Data and Artificial Intelligence: Empowering the Future
Sasana Kijang, Bank Negara
19th September 2024

Healthcare Insurance Fraud Detection using Benford's Law

Wei Han Yap[1]; Kee Huong Lai[1,*]

[1] School of Accounting and Finance, Taylor's Business School, Taylor's University, No. 1, Jalan Taylor's, 47500 Subang Jaya, Selangor Darul Ehsan, Malaysia.

weihan.yap03@sd.taylors.edu.my, keehuong.lai@taylors.edu.my

Abstract:

Fraudulent insurance claims pose a significant threat to the financial sustainability of insurance companies. While machine learning models have been employed to detect fraud, these methods are typically sophisticated and computationally expensive. This work examines the use of Benford's law in the context of healthcare insurance claims. Benford's law was adopted in this study due to its straightforward and easy-to-apply statistical technique. We utilized a synthetic dataset obtained from Kaggle and conducted three statistical tests, namely, the *z*-test, the chi-square test, and the mean absolute difference, on the dataset to measure the dataset's conformity to Benford's law. Specifically, we compared the observed and expected frequency of occurrence for each of the nine leading digits. The results were presented both graphically and in tabular form. The study found that the inpatient reimbursement amount, outpatient reimbursement amount, and outpatient gross claim conform to Benford's law, whereas the inpatient gross claim does not.

Keywords: Benford's law; insurance fraud detection; unsupervised learning

1. Introduction:

It is estimated that one fifth of the insurance claims submitted in the USA are fraudulent, resulting in an alarming annual loss of $308.6 billion due to insurance fraud (Kilroy, 2024). To cover the increased risk and losses due to fraud, actuaries adopt more conservative pricing strategies. This approach transfers the financial burden onto policyholders in the form of higher premiums (Chen et al., 2020).

While it is impossible to eliminate fraud entirely, the reduction of fraudulent cases is achievable through effective detection methods. The increased implementation of automated fraud detection technologies, such as machine learning models (Nabrawi & Alanazi, 2023), has led to a significant decrease in fraudulent activities.

Although machine learning models have achieved considerable detection accuracy, they are often overly sophisticated and computationally expensive due to the large amount of input data required. In this regard, Benford's law serves as a viable alternative for fraud detection. Benford's law is a statistical tool that is straightforward and easy to apply. Additionally, it is relatively less computationally expensive compared to machine learning models. It has been found widespread successful implementations in various fields, including forensic accounting (Druică et al., 2018) and electoral fraud detection (Gueron & Pellegrini, 2022).

Mathematically speaking, Benford's law states that the probability of occurrence of the first, or the leading digit, *d*, of numerical values follows the following probability function:

$$P(d) = \log_{10}\left(1+\frac{1}{d}\right), \text{ where } d = \{1, 2, ..., 9\}. \quad (1)$$

For instance, the digit one will appear approximately 30.1% of the time, whereas the digit nine will appear only about 4.58% of the time. It is important to note that Benford's Law applies exclusively to naturally occurring numbers, such as insurance claim amounts, stock prices, and death rates. The law is not applicable to manipulated or pre-assigned numbers, such as phone numbers, identification numbers, and aggregate claim amounts after the policy limit is applied.

2. Methodology:

The synthetic dataset utilized in this study was obtained from the Kaggle website (Gupta, 2019). It comprises 63,968 observations, detailing the annual reimbursement amounts for Medicare, a health insurance program implemented in the USA. Specifically, this research focused on data from four primary columns: IPAnnualReimbursementAmt, IPGrossClaim, OPAnnualReimbursementAmt, and OPGrossClaim, where IP, OP, and Amt stand for inpatient, outpatient, and amount, respectively.

Both Microsoft Excel and the R programming language were used to perform the experiments. During the data preprocessing stage, the first digits of the respective inpatient and outpatient reimbursement amounts and gross claims were initially extracted. Subsequently, the total number of observations for each of the nine leading digits was tallied.

The distribution of leading digits for each of the four categories was analyzed to see how well it conformed to Benford's Law. Combo charts were generated for visualization purposes, including error bars to identify any significant deviations between the observed and expected values.

Three performance metrics were used to score the conformity of the dataset to Benford's law; they are the *z*-test, the chi-squared test, and the mean absolute deviation (MAD).

The *z*-test is a statistical hypothesis test that quantifies how many standard deviations the observed values are from the expected values. We used $\alpha = 0.05$ and the corresponding critical value of $z_{\alpha/2} = 1.96$. The formula of the *z*-statistics is given as follows:

$$Z = \frac{|AP - EP| - \frac{1}{2N}}{\sqrt{\frac{EP(1-EP)}{N}}}, \quad (2)$$

where the abbreviation AP, EP, and N stand for the actual proportion, the expected proportion, and the number of records, respectively.

The chi-squared goodness-of-fit test compares the observed values to the expected values, assessing any significant deviations between the two. If the observed values are close to the expected values, the chi-squared test statistic will be small, indicating a good fit. Conversely, if the observed values deviate significantly from the expected values, the chi-squared test statistic will be large, leading to the rejection of the null hypothesis. The formula of the chi-squared test statistics is given as follows:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(O_i - E_i\right)^2}{E_i}, \tag{3}$$

where O_i and E_i are the observed and expected frequencies of the i^{th} category, respectively, and the degree of freedom is given by the formula $df = n-1,$ where n is the number of categories.

The mean absolute deviation (MAD) is a measure of conformity to Benford' law (Nigrini, 2012). The absolute value function $|\cdot|$ is incorporated in the formula to prevent negative deviations from offsetting positive deviations. The formula for MAD is given as follows:

$$MAD = \frac{\sum_{i=1}^{k} \left|AP - EP\right|}{k}, \tag{4}$$

where the abbreviation AP and EP stand for the actual proportion and the expected proportion, respectively, and k is the number of leading digits. A greater value of MAD implies that the observed values differ significantly from the expected values. While no critical scores or values are proposed for the MAD test, a range of scores was suggested by Nigrini (2012) to categorize the results. These categories are as follows: close conformity (0.000 to 0.006), acceptable conformity (0.006 to 0.012), marginally acceptable conformity (0.012 to 0.015), and non-conformity (above 0.015).

3. Result:

Figure 1 shows the visualization of the distribution of the data from the four categories, plotted against the expected distribution predicted by Benford's law. From the visual representation, it is observed that among the four categories, the inpatient reimbursement amount (InReimburse) category conforms closely to Benford's Law for all digits except digit 2, where there is a noticeable deviation observed from the height of the bar and the error bar plotted. For the outpatient reimbursement amount (OutReimburse) and outpatient gross claim (OutGrossClaim) categories, all nine digits conform closely to Benford's Law, with no significant deviations observed. Lastly, for the inpatient gross claim (InGrossClaim) category, it is observed from Figure 1 that its distributions for the digits 1, 2, and 3 significantly differ from the distribution predicted by Benford's law.

Table 1 presents the performance metrics for the inpatient reimbursement amount and inpatient gross claim categories. The first column lists the leading digits, followed by the expected proportion predicted by Benford's Law in the second column.

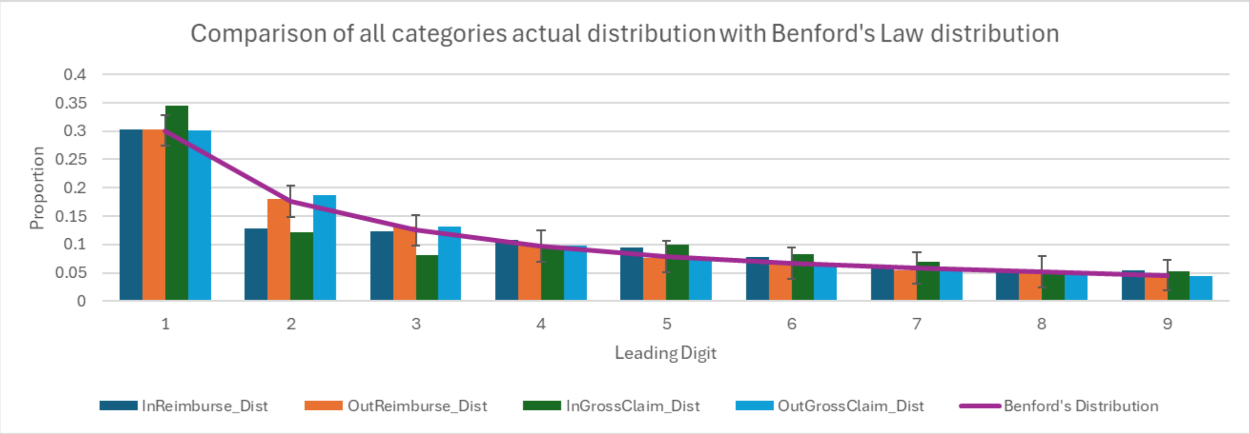

Figure 1. Comparison of all categories with Benford's law distribution

Table 1. Performance metrics of the inpatient categories

Digit	Benford's law	Inpatient Reimbursement	*z*-stat	Inpatient Gross Claim	*z*-stat
1	0.3010	0.3024	0.4212	0.3455	13.7297*
2	0.1761	0.1280	17.7637*	0.1214	20.3182*
3	0.1249	0.1232	0.7501	0.0803	19.0966*
4	0.0969	0.1073	4.9091*	0.0980	0.5256
5	0.0792	0.0948	8.1215*	0.0990	10.3752*
6	0.0669	0.0771	5.7159*	0.0823	8.6848*
7	0.0580	0.0577	0.1366	0.0686	6.4413*
8	0.0512	0.0559	3.0317*	0.0517	0.3400
9	0.0458	0.0536	5.2667*	0.0531	4.9286*
	χ^2 *p*-value	1.494e-83	χ^2 *p*-value	6.662e-216	* $Z > 1.96$
	MAD	0.0111	MAD	0.0221	
	Interpretation	Acceptable conformity	Interpretation	Non-conformity	

Table 2. Performance metrics of the outpatient categories

Digit	Benford's law	Outpatient Reimbursement	*z*-stat	Outpatient Gross Claim	*z*-stat
1	0.3010	0.3029	0.9958	0.3011	0.0496
2	0.1761	0.1796	2.3193*	0.1870	7.1578*
3	0.1249	0.1317	5.1056*	0.1322	5.5429*
4	0.0969	0.0978	0.7309	0.0984	1.2497
5	0.0792	0.0754	3.4714*	0.0749	3.9737*
6	0.0669	0.0653	1.6151	0.0619	5.0904*
7	0.0580	0.0542	4.0552*	0.0533	5.0172*
8	0.0512	0.0487	2.7934*	0.0475	4.1746*
9	0.0458	0.0444	1.6193	0.0437	2.4504*
	χ^2 *p*-value	1.442e-11	χ^2 *p*-value	1.287e-29	* $Z > 1.96$
	MAD	0.00288	MAD	0.00439	
	Interpretation	Close conformity	Interpretation	Close conformity	

The third column shows the actual proportion from the observed values of the inpatient reimbursement amount category, with the corresponding *z*-stat values in the fourth column. The fifth and sixth columns are defined similarly for the inpatient gross claim category. Any *z*-stat value greater than 1.96 is indicated with an asterisk. Additionally, the *p*-value of the chi-squared statistics and the mean absolute deviation (MAD), along with its interpretation, are provided in the final three rows of the table. Table 2 reports the performance metrics for the outpatient reimbursement amount and outpatient gross claim categories. The table is interpreted in a similar manner as Table 1.

4. Discussion and Conclusion:

It is noteworthy to highlight the difference between the inpatient gross claim and inpatient reimbursement amount categories. While the inpatient gross claim represents the raw loss data submitted by policyholders, the inpatient reimbursement amount refers to the revised amount after the deductible has been applied to the gross claim. Intuitively speaking, the inpatient gross claim (raw data) would conform to Benford's law, whereas the inpatient reimbursement amount (data that has been manipulated) would not. However, Figure 1 showed otherwise, which contradicts Benford's law.

The *z*-test is applied to each of the nine leading digits in the four categories. Among the four categories, the number of cases that deviate significantly from the predicted values of Benford's Law ranges between 5 and 7. While this statistical test provides granular insight into each of the nine individual leading digits, it does not reveal the overall conformity. With regard to the *p*-values of the chi-squared test statistics, all four categories showed values less than 0.05. Due to the large sample size, the *p*-values obtained are very small, while the test statistics have large values. This phenomenon is known as the excess power problem encountered by the chi-squared test (Kossovsky, 2021). It has also been argued that the chi-squared test may not be appropriate for testing conformity to Benford's Law if the sample size is too large (Kossovsky, 2021). Lastly, the interpretations of the four MAD values obtained indicated that the outpatient reimbursement amount and outpatient gross claim categories exhibited close conformity to Benford's Law. In contrast, the inpatient reimbursement amount category demonstrated acceptable conformity, whereas the inpatient gross claim category showed non-conformity.

For large sample sizes, it is inappropriate to rely solely on the chi-squared test to determine a dataset's conformity to Benford's Law. Therefore, other statistical tests, such as the mean absolute deviation (MAD), should be used to complement the chi-squared test to provide additional insights.

It is also interesting to compare the results obtained by the chi-squared test and the MAD test. For smaller sample sizes, even if the chi-squared test concludes that a given dataset conforms to Benford's Law, it does not necessarily imply that the MAD test will produce the same conclusion. This is because the chi-squared test tolerates large deviations from the expected distribution when the sample size is small. Conversely, if the chi-squared test concludes that a given dataset does not conform to Benford's Law, it is highly likely that the MAD test will arrive at the same conclusion, as both tests indicate a statistically significant deviation from the expected distribution.
This study has revealed the contradictory conclusions drawn from the graphical representation, the *z*-score test, and the chi-squared test. Specifically, the chi-squared test suffers from the excess power problem, which tolerates small deviations as the number of observations increases.

The main limitation identified in this study is the lack of accessibility to real-world insurance datasets. Due to regulatory policies and the need to protect policyholders' privacy, insurance companies are prohibited from sharing their data with third parties. Since the study utilized a synthetic data set, the analysis could not accurately quantify the accuracy of Benford's law. Furthermore, deviations from Benford's Law do not necessarily imply that the datasets are fraudulent. It is important to note that Benford's Law serves as a preliminary statistical tool to quickly assess the authenticity of a given dataset. However, more advanced tests and algorithms should be employed for further investigation.

Simulated datasets that are realistic and representative of actual insurance datasets (Campo & Antonio, 2023) could be adopted in future studies. Additionally, complementing the results from Benford's Law with more advanced machine learning models presents another valuable avenue for exploration. While this study has incorporated datasets with deductibles, it would be insightful to examine insurance datasets that include co-payments. The correlation between the claim amounts submitted by healthcare providers and fraud detection should also be investigated. This could help identify and mitigate unnecessary treatments performed for higher profitability. Research in this area also aligns with the government's goal to reduce false claims, which would otherwise lead to increased premiums for policyholders.

References:

1. Kilroy, A. (2024). Insurance Fraud Statistics 2024. *Forbes Advisor*. Retrieved July 1, 2024, from https://www.forbes.com/advisor/insurance/fraud-statistics/
2. Chen, Z. X., Hohmann, L., Banjara, B., Zhao, Y., Diggs, K., & Westrick, S. C. (2020). Recommendations to protect patients and health care practices from medicare and medicaid fraud. *Journal of the American Pharmacists Association*, *60*(6), e60-e65.
3. Nabrawi, E., & Alanazi, A. (2023). Fraud detection in healthcare insurance claims using machine learning. *Risks*, *11*(9), 160.
4. Gueron, E., & Pellegrini, J. (2022). Application of Benford–Newcomb law with base change to electoral fraud detection. *Physica A: Statistical Mechanics and its Applications*, *607*, 128208.
5. Druică, E., Oancea, B., & Vâlsan, C. (2018). Benford's law and the limits of digit analysis. *International Journal of Accounting Information Systems*, *31*, 75-82.
6. Gupta, R. A. (2019). Healthcare Provider Fraud Detection Analysis. Retrieved May 1, 2024, from https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis/code
7. Zdraveski, D., & Janeska, M. (2021). Application of Benford's law for detecting manipulation in the financial statements in Macedonian companies. *Annals of Constantin Brancusi University of Targu-Jiu, Economy Series, 6*, 4-13.
8. Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley & Sons.
9. Kossovsky, A. E. (2021). On the mistaken use of the chi-square test in Benford's law. *Stats*, *4*(2), 419-453.
10. Campo, B. & Antonio, K. (2023). Insurance fraud network data simulation machine: Generating synthetic fraud network data sets to develop and to evaluate insurance fraud detection strategies. In *Insurance Data Science Conference, Location: London, United Kingdom*.