Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 3, No. 2, pp. 61-66, 2025

Application of Benford's Law to Detect Fraud in Customers' Ending Balances Using First Digit Test, Second Digit Test, and First Two Digits Test

Muhammad Ilham Firdaus[1], Agung Prabowo[2*], Najmah Istikaanah[3]

[1]*Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman*
[2]*Soedirman Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Indonesia.*
[3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman*

**Corresponding author email: agung.prabowo@unsoed.ac.id*

Abstract

Bank fraud involves several actions such as manipulating duplication, forgery, changing accounting records and so on. This study aims to detect the potential for fraud in banking reports on customer final balances. The types of tests used to detect potential fraud in this study are the First Digit Test, Second Digit Test and First Two Digit Test Benford's Law. Benford's law states that the proportion of occurrences of numbers in certain numbers is not the same. In addition to the three Benford's Law tests, further statistical tests were carried out to determine the magnitude of the deviation between the actual proportion of occurrences and the expected proportion of Benford's Law using the Mean Absolute Deviation (MAD), Chi-square test, and Z test. This study uses secondary data on the final balance of customer deposits as of July 2023 as much as 20,105 data. The results showed that there were indications of fraud in the form of rounding and duplication of data on the customer's final balance. MAD results show that the proportion of occurrence of actual numbers is quite consistent with the proportion of occurrence of Benford's Law expectations. Based on the Z test, the balance that has the potential for fraud is the value with the first digit '5', the second digit '3' and the first two digits '23'. These numbers can be found in balances with a nominal value of Rp5000 and Rp5636 in the first digit '5', and Rp23038 in the second digit '3' and the first two digits '23'.

Keywords: Customer balance, Duplication, Digit Test, Logarithm, Mean Absolute Deviation

1. Introduction

A bank is a financial institution whose main activity is to collect funds from the public and redistribute them back to the public, as well as provide several other services (Bossone, 2001). Banks have an obligation to safeguard customer/depositor funds from the risk of fraud. Bank fraud is described as an intentional act, deception, or manipulation committed by certain individuals within the bank, resulting in losses to the bank itself, customers, or other parties (Sherpa, 2021). In general, according to Murphy & Dacin (2011), fraud is a deliberate deviation by certain parties in order to gain personal benefit. A survey conducted by the Association of Certified Fraud Examiners (ACFE) showed the percentage of fraud detection by internal auditors was 23.4% and by external auditors was 9.6%. According to Kanu et al. (2023), fraud occurring in the banking sector includes credit disbursement (43%), manipulative actions (19.6%), and forgery (18.6%). Manipulations include attempts to alter, hide, or create data information in an illegal manner that can harm various parties.

Benford's Law, first proposed by Frank Benford in 1938, is a supporting tool for analyzing the risk of potential fraud. Benford's Law estimates the probability of the appearance of a number in a series of data (Bauer & Gross, 2011). Three types of Benford's Law analysis commonly used include the first-digit test, second-digit test, and first-two-digit test (Nigrini & Mittermaier, 1997). The first-digit test is used for data with extreme anomalies. The second-digit test provides an overview of data involving number rounding. The first-two-digit test is used to detect data duplication.

Alimirruchi & Chariri (2023) conducted a study to identify potential data manipulation in several account items of Jiwasraya insurance. The use of Benford's Law proved that there was significant fraud in the 2014 and 2015 analysis on the account items of Future Policy Benefits Liabilities, Estimated Claim Liabilities, and Unit Link Mutual Funds. Wachira (2022) conducted a study to detect fraud indications in operational transaction data at ABC Hotel. The

detection focused on the market department, consisting of room service sales transactions, accounts receivable, accounts payable, and expenses. The study showed a discrepancy between data patterns and Benford's Law expectations. The first-digit test indicated that sales transactions, accounts receivable, and accounts payable data differed from Benford's Law expectations (Alali & Romero, 2013). The second-digit test showed that the overall transaction data differed from Benford's Law expectations. The first-two-digit test showed discrepancies of 26% in room service sales data, 33% in accounts receivable data, 23% in accounts payable data, and 21% in expense data.

Based on the aforementioned explanation and the referenced articles, this study applies Benford's Law to detect the potential occurrence of fraud in customer end balance data. This study uses the first-digit test, second-digit test, and first-two-digit test, as well as the Z-test, Chi-square test, and mean absolute deviation (MAD).

2. Materials and Methods

2.1. Research Data

The data used in this study is secondary data, namely customer balance data as of July 31, 2023, totaling 20,105 customers. The research object consists of the entire final balance data.

2.2. Research Steps

This study involves several procedures to analyze the potential for fraud in customer final balances:

a. The research data used is the entire customer balance, totaling 20,105 entries;
b. Analyzing the balance data requirements based on predetermined criteria;
c. Classifying the first digit, second digit, and first two digits using the pivot table feature in Microsoft Excel;
d. Calculating all Benford's Law probabilities for the first digit, second digit, and first two digits using the probability mass function for each digit;
e. Comparing and interpreting the probability values of the first digit, second digit, and first two digits with the actual values in a graph;
f. Calculating the Z, Chi-square, and MAD values for each digit;
g. Conducting an analysis of probabilities that indicate fraud;
h. Drawing conclusions from the calculations and analysis obtained.

3. Results and Discussion

3.1. Data Description

The research was conducted to show that Benford's Law can be used to detect indications of fraud in customer final balances. The data used in this research is final customer balance data that has previously gone through the calculation process of debit transactions, credit transactions, and interest, with a total of 20,105 data entries. The following is a data description using descriptive statistics.

Table 1: Data Description of Customer Balances Using Descriptive Statistics

Description	Value (IDR)
Mean	5,308,135
Median	285,905
Mode	23,038
Minimum	671
Maximum	1,039,352,510
Total	20,105
Skewness	16.115
Kurtosis	495.843

The probability test data for the occurrence of digits in Benford's Law has several rules as follows:

- The data is a whole unit and represents a similar phenomenon.
  This study uses a complete sample of the final customer balance data from one bank. The data used as the object of this research reflects the same phenomenon and is not mixed with the final balance data from other banks, thus meeting the first requirement.

- The data does not have a maximum and minimum limit at certain digits.
  The data taken is the actual data recorded in the reports. This research does not take maximum and minimum limits on intervals or specific ranges of numbers, thus meeting the second requirement.
- The data has a large size (at least 1,000 data points).
  The sample used in this study consists of 20,105 data points, which are the final customer balance data from the bank, thus meeting the third requirement.
- The data used consists of natural numbers (not a symbol/code of any kind).
  The data in this study is the result of the accumulation of the initial balance, which has undergone credit and debit transactions as well as interest deductions. Therefore, the final customer balance data, which is the object of this research, forms a natural number resulting from calculations, thus meeting the fourth requirement.

3.2. Benford's Law Data Analysis

Benford's Law is a law of first digits that predicts the frequency of occurrence in a series of numerical data. According to Nigrini & Miller (2009), the application of Benford's Law can be performed through three basic tests: first-digit test, second-digit test, and first-two-digits test.

3.2.1First-Digit Test

The first-digit test is an examination with a high-level nature, focusing only on clear irregularities. The formula for the first-digit test is as follows:

$$P(D_1 = d_1) = log\left(1 + \frac{1}{d_1}\right) \quad (1)$$

The results of the first-digit test can be seen in the following table:

Table 2: Result of First-Digit Test

First Digit	Actual Data Count	Benford Probability	Actual Probability	Z Value
1	5675	0.30103	0.28227	5.79
2	4083	0.17609	0.20308	10.04
3	2254	0.12494	0.11211	5.49
4	1940	0.09691	0.09649	0.19
5	2046	0.07918	0.10177	11.85
6	1310	0.06695	0.06516	1.00
7	1080	0.05799	0.05372	2.58
8	923	0.05115	0.04591	3.36
9	794	0.04576	0.03949	4.23
MAD	***Chi-S***			
0.01102	298.065			

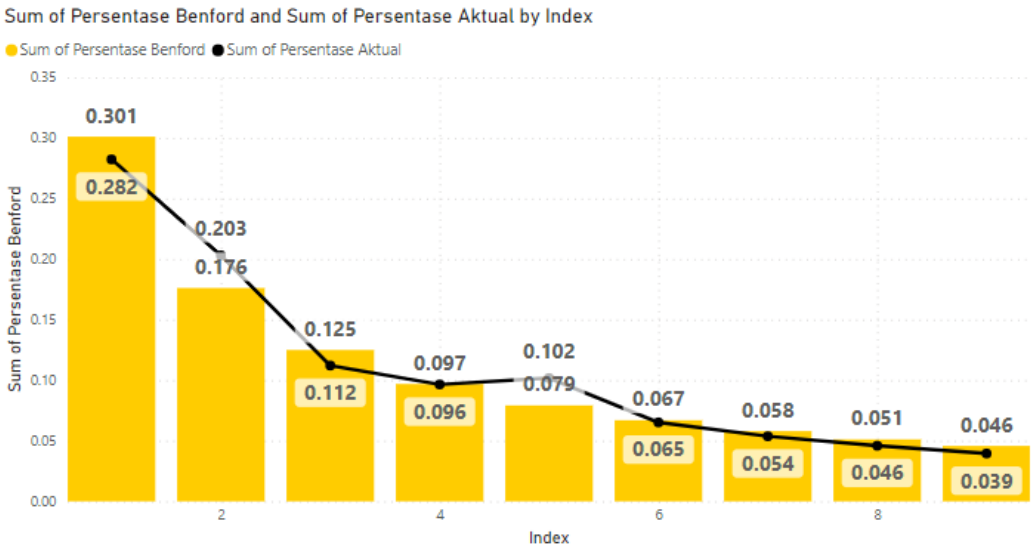

Figure 1: Comparison Diagram of First-Digit Test Proportions

Table 2 and Figure 1 show that overall, the customer final balances follow the proportions expected by Benford's Law. The average difference (MAD) between the actual data proportion and the expected Benford proportion is 0.01102, indicating that the balance data is consistent with Benford's Law. The Chi-square value obtained is 298.065, while the table Chi-square value is 15.507. This indicates that there is a significant spread of data in the final customer balances.

When viewed through the Z test, seven of the nine digits in the first digit have a Z value greater than 1.96. Some balances indicated by the Z test with the first digits '5' and '2' include IDR 5000, IDR 5636, and IDR 23038, showing many duplications that suggest balance manipulation.

3.2.2 Second-Digit Test

The second-digit test is a testing method using the second digit of a series of data in its calculation. The second-digit test shares the same high-level overview characteristic as the first-digit test, so this test is effective when there are obvious irregularities. The formula for the second-digit test is as follows:

$$P(D_2 = d_2) = \sum_{d_1=1}^{9} log\left(1 + \left(\frac{1}{10d_1 + d_2}\right)\right), d_2 = 0,1,2,\dots,9 \quad (2)$$

The results of the second-digit test can be seen in the following table:

Table 3: Result of Second-Digit Test

Second Digit	Actual Data Count	Expected Data Count	Z Value
0	3033	2406	13.60889
1	2382	2290	2.03670
2	1861	2188	7.39031
3	2590	2098	11.34975
4	1725	2017	6.83706
5	2004	1944	1.42589
6	1855	1877	0.52610
7	1710	1816	2.60734
8	1447	1761	7.81172
9	1498	1709	5.32139
MAD	***Chi-S***		
0.0126	463.911		

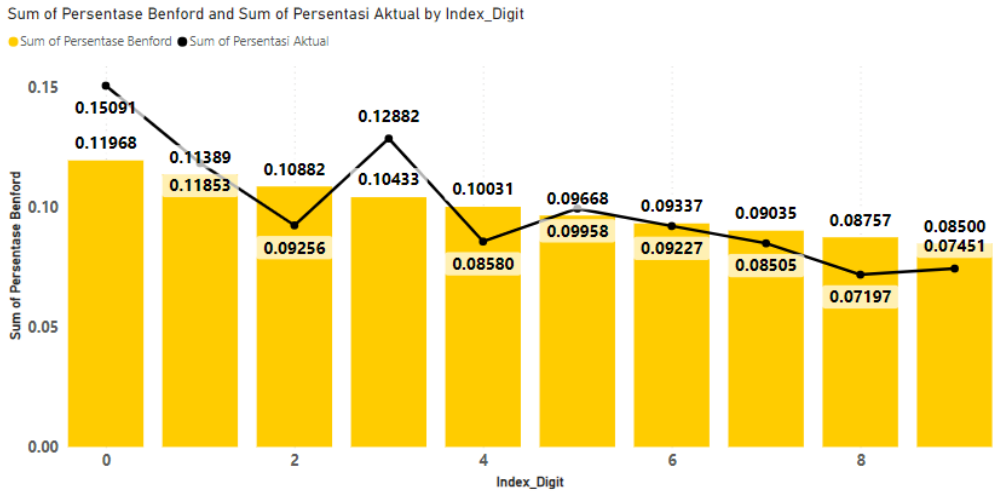

Figure 2: Comparison Diagram of Second-Digit Test Proportions

Table 3 and Figure 2 show that visually, the customer final balances do not match the expectations of Benford's Law. This is supported by the MAD value of 0.0126, indicating that the pattern of occurrence of the balance digits does not match what is expected by Benford's Law. The Chi-square value obtained is 463.911, while the table Chi-square value is 16.919. This suggests there is a significant spread of data in the final customer balances.

Based on the results of the Z test, almost all digits have significant proportional differences. The type of fraud indicated is the rounding of numbers performed by the financial statement preparers. For example, a balance of IDR 5105 may have been changed by rounding it down to IDR 5000, leading to more occurrences of '0' in the second digit.

3.2.3 First-Two Digit Test

The first two-digits test is a testing method involving the first two digits of a series of data and is a continuation of the previous two tests. This test concludes that if the actual probabilities differ significantly from the expected Benford probabilities, there are many duplications. The formula for the first-two-digits test is as follows:

$$P(D_1 = d_1, D_2 = d_2) = log\left(1 + \frac{1}{10d_1 + d_2}\right) \quad (3)$$

The results of the first-two digit test can be seen in the following table:

Table 4: Result of First-Two Digit Test

First Two Digit	Actual Data Count	Benford Probability	Z Value
10	888	832	1.95790
11	756	760	0.11979
12	563	699	5.21280
13	656	647	0.33674
⋮	⋮	⋮	⋮
94	80	92	1.24059
95	77	91	1.46015
96	69	90	2.21083
97	77	90	1.27662
98	68	89	2.14440
99	76	88	1.20403
MAD	***Chi-S***		
0.00202	2074.057		

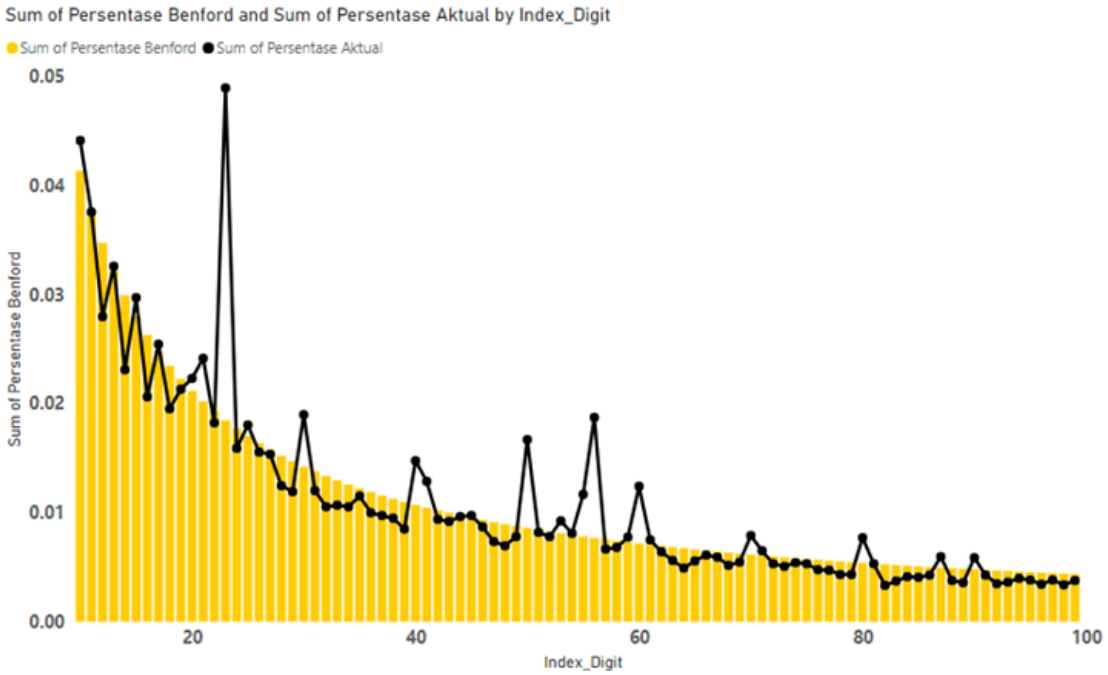

Figure 1: Comparison Diagram of First-Two Digit Test Proportions

Table 4 and Figure 3 visually demonstrate that the distribution of the first two digits of the customers' ending balances aligns fairly well with the expectations of Benford's Law. This is supported by the Mean Absolute Deviation (MAD) value of 0.00202, indicating that the observed frequency pattern of the digit sequences closely matches the expected distribution. However, the obtained Chi-square value is 2,074.057, which significantly exceeds the critical

Chi-square value of 112.022, suggesting a statistically significant deviation in the distribution of the first two digits of the ending balances.

Based on the Z-test results, a portion of the data shows a significant difference in proportions compared to what is expected under Benford's Law, indicating potential fraud. Specifically, 48 digit combinations in the first two digits show no significant difference, while 42 combinations exhibit significant discrepancies. One possible indication of fraud is the high frequency of duplicated balances, which creates anomalies in the data. This is further supported by the presence of identical balances associated with different customer identities. The most frequently duplicated balance is IDR 23,038, which appears 368 times, reinforcing suspicions of irregularities in the dataset.

4. Conclusion

The application of Benford's Law aids in identifying patterns in sequences of numbers that may have been deliberately manipulated, serving as an early detection tool to assist auditors in selecting audit samples and further investigating the ending balance figures in financial reports as potential indicators of fraud. In detecting fraud indications in customers' ending balances, three types of tests were applied: the First Digit Test, the Second Digit Test, and the First Two Digits Test. The First Digit Test showed that the distribution of digits generally followed the expected Benford pattern, as indicated by a MAD value of 0.01102, although the Chi-square and Z-tests revealed anomalies, particularly for the digits '5' and '2'. The Second Digit Test, on the other hand, indicated a deviation from Benford's expected distribution, supported by a MAD value of 0.01265 and further confirmed through Chi-square and Z-tests, with digits '0' and '3' being the most notable outliers. The First Two Digits Test suggested a fairly good alignment with Benford's expectations, as shown by a MAD value of 0.00202, but some deviations still emerged, particularly for the digit combination '23' according to the Z-test. For future research, it is recommended to include additional Benford tests such as the third digit test, the first three digits test, and the last two digits test to enhance fraud detection accuracy and effectiveness in analyzing customers' ending balances.

References

Alali, F. A., & Romero, S. (2013). Benford's Law: Analyzing a decade of financial data. Journal of Emerging Technologies in Accounting, 10(1), 1-39.

Alimirruchi, W., & Chariri, A. (2023). Revealing Red Flags of Insurance Fraud: A Case Study Research of PT Jiwasraya Indonesia. Theory and Practice of Forensic Science and Criminalistics, 30(1), 23-49.

Bauer, J., & Gross, J. (2011). Difficulties detecting fraud? The use of benford's law on regression tables. Jahrbücher für Nationalökonomie und Statistik, 231(5-6), 733-748.

Bossone, B. (2001). Do banks have a future?: A study on banking and finance as we move into the third millennium. Journal of banking & finance, 25(12), 2239-2276.

Kanu, C., Nnam, M. U., Ugwu, J. N., Achilike, N., Adama, L., Uwajumogu, N., & Obidike, P. (2023). Frauds and forgeries in banking industry in Africa: a content analyses of Nigeria Deposit Insurance Corporation annual crime report. Security journal, 36(4), 671-692.

Murphy, P. R., & Dacin, M. T. (2011). Psychological pathways to fraud: Understanding and preventing fraud in organizations. Journal of business ethics, 101, 601-618.

Nigrini, M. J., & Miller, S. J. (2009). Data diagnostics using second-order tests of benford's law. Auditing: A Journal of Practice & Theory, 28(2), 305-324.

Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's Law as an Aid in Analytical Procedures. Auditing: A journal of practice & theory, 16(2).

Sherpa, P. (2021). Legal and Regulatory issues of Cyber Fraud in Transnational Banking with special reference to European Union (Doctoral dissertation).

Wachira, D. (2022). Influence Of Internal Control Systems On The Financial Performance Of Classified Hotels In The Nairobi County (Doctoral dissertation, KCA University).