Benjamin Arancibia

CUNY IS 698

Data Collection and Analysis

**Using Benford’s Law to Detect Fraud in Foreign Financial Assistance Transactions**

**Project Summary**

The main focus of this research is to apply Benford’s law to foreign financial aid transactions and answer the following questions:

* What foreign financial aid transactions have a high likelihood of corruption during project implementation?
* What type of organizations, multilateral or bilateral, have more foreign financial aid transactions flagged as possibly fraudulent?

**Methods**

Transaction level data from the International Aid Transparency Initiative (IATI) was used for the analysis foreign aid transactions and this data can be easily accessible via IATI’s API or IATI’s datastore, which allows users to query the desired data. When the data is queried and downloaded there are several different qualitative and quantitative fields within the dataset. The full IATI dataset has 74 variables and 471,395 transactions, but many columns are missing values. The percentages of missing values for columns range from 0.07% (transaction value) to 100% (Transaction Recipient Region). Since there are a large amount of variables with missing data, the raw dataset was filtered to only contain columns that are relevant to the investigation. The relevant data columns with description can be seen in Table 1.

|  |  |
| --- | --- |
| Variable | Description |
| Transaction Type | Type of transaction |
| Default Currency | Currency value |
| Transaction Value | Numeric Amount |
| Transaction Value Date | Date of transaction |
| Transaction Provider Organization | Organization providing funding |
| Transaction Receiver Organization | Organization receiving funding |
| Reporting Organization | Organization reporting funding |
| Title | Title of Project that transaction is part of |
| Description | Description of project that transaction is part of |
| Start Planned Date | Planned start date |
| End Planned Date | Planned end date |
| Start Actual Date | Actual start date |
| End Actual Date | Actual end date |
| Recipient Country | Recipient country of transaction |
| Sector Vocabulary | Sector of transaction |

Table 1: Filtered Data Columns

After filtering the data set the transaction amounts were transformed to one currency. There are 18 different currencies, including no currency identified, within the dataset and these were all transformed to United States Dollars (USD). The transformation was done by filtering the transactions by years and then transforming the transaction amount based on the currency exchange rate for that year into USD. Those transactions without identified currencies were assumed to be in USD. After performing this the data was then removed of all transactions that have no values, which there are 339. The summary statistics of USD transaction amounts are seen below in Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean (SD) | Median | Min | Max |
| USD Transaction Amount | 8,811,817 USD (1108042891 USD) | 30,000 USD | - 2,147,483,648 USD | 166,298,000,000 USD |

Table 2: Summary Statistics

One aspect of the dataset that must be discussed are when multiple countries have been entered into the recipient country variable. As an example, a transaction might have Kenya, Ethiopia, and Uganda as the listed recipient countries. There are two ways to deal with these sort of cases, divide into three equal transaction values or leave it as it is. It was decided to leave these transactions alone because it is not a safe assumption that the transaction values are split evenly.

Once the data has been transformed it can be analyzed using Benford’s Law via the Benford Analysis package available on CRAN (Cinelli, 2015). This package provides a robust tool for using Benford’s Law to investigate a specific dataset. Its main purposes are to find out where the dataset deviates from Benford’s Law and to identify suspicious data that need further verification. The outputs of this tool are a Chi Square test and the Mantissa Arc Test. The Mantissa Arc Test is a commonly used digit forensic tool looks to see if mantissas are uniformly distributed and if they are uniformly distributed then the result is a perfect circle with a radius of 1 and a center of gravity of (0,0). The following is the output of the Benford Law Analysis Package.

Benford object:

Data: benfords$usd.conversion

Number of observations used = 456111

Number of obs. for second order = 292817

First digits analysed = 2

Mantissa:

Statistic Value

Mean 0.4968

Var 0.0848

Ex.Kurtosis -1.1999

Skewness -0.0087

The 5 largest deviations:

digits absolute.diff

1 10 4461.34

2 50 2738.37

3 20 1950.33

4 30 1872.78

5 15 1774.79

Stats:

Pearson's Chi-squared test

data: benfords$usd.conversion

X-squared = 7442.799, df = 89, p-value < 2.2e-16

Mantissa Arc Test

data: benfords$usd.conversion

L2 = 1e-04, df = 2, p-value < 2.2e-16

Mean Absolute Deviation: 0.001038434

Distortion Factor: -0.6463139

What the output is showing is that the results appear to be significant based on the resulting p-values, which are highlighted. Plotting the results also give a good overview of what is occurring in the dataset, which can be seen in Figure 1 below.

Figure 1: Benford’s Law Analysis

Based on this initial results, the digits by decreasing order of discrepancy are in Table 3.

|  |  |  |
| --- | --- | --- |
| Rank | Digits | Absolute Difference |
| 1 | 10 | 4461.3410 |
| 2 | 50 | 2738.3671 |
| 3 | 20 | 1950.3276 |
| 4 | 30 | 1872.7791 |
| 5 | 15 | 1774.7908 |
| 6 | 17 | 1201.3096 |
| 7 | 40 | 1073.7270 |
| 8 | 13 | 1037.7881 |
| 9 | 11 | 940.7783 |
| 10 | 25 | 804.9066 |

Table 3: Digits by Decreasing Order of Discrepancies

These results show that the first two digits of 10 in that combination appear the most often together and occur more often than they should. Table 4 shows the largest number of duplicate values in the dataset.

|  |  |  |
| --- | --- | --- |
| Rank | Numbers | Duplicate Counts |
| 1 | 100000 | 1517 |
| 2 | 500000 | 959 |
| 3 | 50000 | 884 |
| 4 | 1000000 | 840 |
| 5 | 1 | 752 |
| 6 | 200000 | 739 |
| 7 | 300000 | 646 |
| 8 | 10000 | 598 |
| 9 | 250000 | 519 |
| 10 | 150000 | 503 |

Table 4: Duplicates by Decreasing Order

These results show that the value 100000 appears the most often in the dataset, exactly 1,517 times. This makes sense in comparison to the leading digits that were flagged for discrepancies. The package also is able to get data that is suspicious based on the digits groupings by employing the following command:

suspects <- getSuspects(benford.data, benfords, how.many=2)

suspects

The function results in creating a dataframe with 30,002 observations that are suspicious. When that dataframe is investigated the most frequent suspicious countries can be seen in Table 5.

|  |  |  |
| --- | --- | --- |
| Rank | Country | Counts |
| 1 | Blank | 8182 |
| 2 | Tanzania | 643 |
| 3 | India | 607 |
| 4 | Kenya | 590 |
| 5 | Mozambique | 575 |
| 6 | Uganda | 538 |
| 7 | Ethiopia | 463 |
| 8 | Sudan | 432 |
| 9 | Afghanistan | 427 |
| 10 | Indonesia | 423 |

Table 5: Suspicious Countries by Decreasing Order

By far the most suspicious transactions in the dataset that is flagged by Benford’s Law analysis are transactions with no recipient country. The following analysis is performed for organizations and the results can be seen in Table 6.

|  |  |  |
| --- | --- | --- |
| Rank | Organization | Counts |
| 1 | Blank | 20161 |
| 2 | Department for International Development | 1059 |
| 3 | Bill and Melinda Gates Foundation | 1032 |
| 4 | Ministry of Foreign Affairs (DGIS) | 838 |
| 5 | International Development Association | 500 |
| 6 | International Bank for Reconstruction and Development | 441 |
| 7 | Oxfam Novib | 282 |
| 8 | GlobalGiving | 273 |
| 9 | UNICEF (FOR GR Allocations Only) | 265 |
| 10 | United National Development Programme | 246 |

Table 6: Suspicious Organizations by Decreasing Order

The complete code and data is available upon request.

# References

Cinelli, C. (2015, 11 17). *Package 'benford.analysis'.* Retrieved from CRAN: https://cran.r-project.org/web/packages/benford.analysis/benford.analysis.pdf