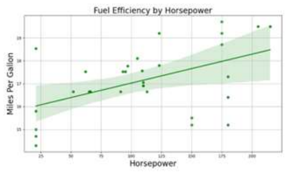
Correlation Reversal Manipulation Revealed By Benford's Law and Random Forest

Supervisor: Assoz. Prof. PD Dr. Jakob Müllner
Vienna University of Economics, Austria

Co-supervisor: Assoz.Prof. Dr. Nysret Musliu
Technical University of Vienna, Austria

Poj Netsiri

INTRODUCTION

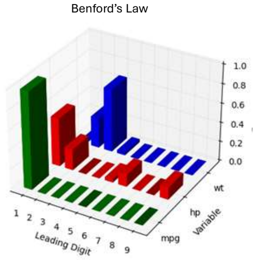


Manipulated dataset showing a misleading positive correlation (r = +0.5, p < 0.001) between horsepower and fuel efficiency. This counterintuitive trend contradicts known automotive principles and serves as a potential indicator of data manipulation.

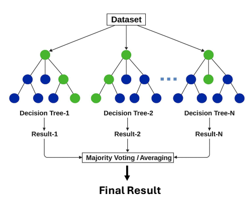
Unethical data manipulation—whether through falsification, p-hacking, or selective reporting—poses a serious threat to research integrity, often leading to misleading conclusions. This study investigates a manipulated dataset of 32 datapoints, a sample size commonly used in scientific and engineering research, reporting a statistically significant but counterintuitive positive Pearson correlation (r = +0.5, p < 0.001) between car horsepower and fuel efficiency. Benford's Law [1] revealed significant anomalies in digit distribution (χ^2 > 30.58, p < 0.0001), and a Random Forest classifier [3] also flagged irregularities. Both methods were then used to reconstruct the original data, revealing a corrected negative correlation (r = −0.57, p < 0.001) consistent with automotive principles, but also consistent with the direction of the known original data prior to manipulation, even if not matching its exact values. While Benford's Law offers robust anomaly detection without requiring training, the Random Forest algorithm excels at regression when trained on clean data. This study suggests that a hybrid approach can effectively detect manipulation and identify correlation direction reversal—even in small datasets—by combining statistical and machine learning techniques.

METHODS

Benford's Law



Random Forest



This study combines Benford's Law [1] and a Random Forest algorithm [3] to detect data manipulation and identify correlation direction reversal. For classification, Benford's Law evaluates deviations in leading digit distributions using the χ^2 test, serving as a blind detector of anomalies [5]. In parallel, a Random Forest classifier is trained on labeled original and manipulated datasets to learn patterns indicative of tampering [4]. For prediction, Benford-guided stochastic simulation generates synthetic values by sampling expected leading digits, applying scaling, and adding noise to reflect natural variability. Separately, a Random Forest regressor, trained on clean data, estimates plausible values in manipulated datasets. Together, these methods uncover whether statistical relationships—such as correlation direction—have been intentionally reversed, providing insight into the presence and nature of manipulation.

RESULTS

Detection of Data Manipulation

Dataset	χ^2	p-value	Classification
Manipulated	74.300	6.81×10^{-13}	Anomaly
Unmanipulated	23.383	0.00291	Acceptable

Chi-squared test results showing a significant deviation from Benford's Law in the manipulated dataset (χ^2 = 74.300, p < 0.0001), while the unmanipulated dataset remains within acceptable limits.

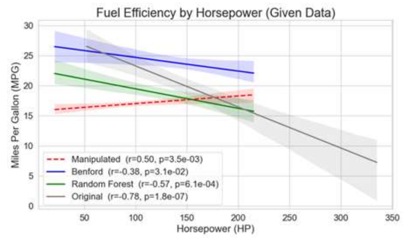
Dataset	Prediction	Classification
Manipulated	71.9 %	Anomaly
Unmanipulated	6.2%	Acceptable

Random Forest classifier predictions showing a high probability (71.9%) of manipulation in the manipulated dataset and a low probability (6.2%) in the original dataset, supporting effective anomaly detection.

Benford's Law [1] and a Random Forest classifier [3] were used to distinguish between manipulated and unmanipulated data. The manipulated dataset showed a significant deviation from the expected Benford digit distribution (χ^2 = 74.3, p < 0.0001), while the original data conformed acceptably (χ^2 = 23.4, p = 0.003). The Random Forest classifier, trained on labeled original and manipulated records, achieved 92% accuracy, correctly identifying 12 out of 13 cases with high precision and recall. These findings demonstrate both methods effectively detect data manipulation, with Benford's Law flagging digit-level anomalies and Random Forest classifying structural inconsistencies in the dataset.

RESULTS

Identification of Direction Reversal



Regression trends for the manipulated data compared with original and reconstructed versions. The manipulated dataset (red dashed line) shows a strong false positive correlation (r = +0.50), while the original data (gray line) displays a strong negative correlation (r = −0.78). Benford-based reconstruction (blue) reduces the anomaly (r = -0.38), but the Random Forest regressor (green) more closely restores the true trend (r = -0.57). This demonstrates the Random Forest's effectiveness in recovering the correct correlation direction.

To estimate the original, unmanipulated data and assess whether manipulation caused a reversal in correlation direction, both Benford's Law [1] and a Random Forest regressor [3] were used. A stochastic simulation guided by Benford's Law was applied to the manipulated dataset to reconstruct plausible values. The simulated data showed improved conformity to Benford's digit distribution and revealed a reversal in correlation from a misleading positive (r = +0.50) to a more realistic negative value (r = –0.38), consistent with known automotive behavior. Similarly, a Random Forest regressor trained on the original data was used to predict corrected values from the manipulated dataset. The reconstructed output not only reversed the correlation direction to r = –0.57 but also demonstrated good predictive performance (R^2 = 0.66). Together, these results highlight how both statistical and machine learning methods can detect manipulation and reveal hidden direction reversal in small datasets.

DISCUSSION

The above figure compares regression trends among original, manipulated, and reconstructed datasets. In both cases, the original data (gray line) shows a strong negative correlation between horsepower and miles per gallon, reflecting expected vehicle behavior. In contrast, the manipulated datasets display misleading positive correlations—moderate in the given data (r = +0.50)—indicating significant distortion. The Benford-reconstructed data (blue line), generated through stochastic simulation based on digit frequencies, partially recovers the correct trend but produces weaker correlations, particularly in the given dataset (r = -0.38). The Random Forest-reconstructed data (green line), however, closely mirrors the original relationship, recovering a consistent and meaningful negative correlation (r = −0.57). These results support a hybrid approach: Benford's Law [1], which requires no prior training, serves as an effective blind detector for identifying manipulation through digit anomalies. In contrast, the Random Forest regressor [3] requires access to the original data for training but excels in accurately reconstructing the underlying structure of the original dataset. Together, these complementary methods form a robust framework for detecting and correcting data manipulation.

Limitations: This study is based on a small dataset (n = 32), which may limit generalizability. Additionally, Random Forest models can overfit on small datasets if not carefully tuned.

Future Work: Extending this methodology to larger, real-world datasets in domains such as epidemiology, finance, or public policy could further evaluate its robustness and scalability.

These complementary findings reinforce the importance of using both statistical and machine learning approaches to ensure data integrity in research contexts

REFERENCES

1. Benford, F. (1938). The Law of Anomalous Numbers. Proceedings of the American Philosophical Society, 78(4), 551–572.
2. Pearson, K. (1895). Note on Regression and Inheritance in the Case of Two Parents. Proceedings of the Royal Society of London, 58, 240–242.
3. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
4. Liaw, A., & Wiener, M. (2002). Classification and Regression by Random Forest. R News,2(3), 18–22.
5. Zhou, W., & Tam, K. H. (2018). Detecting Data Manipulation with Benford's Law: A Review of Methodologies. Forensic Science International, 289, 45–58.