

The First Digit — Benford's Law, What It Catches and What It Cannot

AICE Library · benfords-law

12 chapters · single-language edition

Every number, name and year in this volume is traceable to a cited source.

01. The Worn Pages — How a Library Book Gave It Away

The Astronomer and the Logarithm Tables

In 1881, the American astronomer and mathematician Simon Newcomb made a peculiar discovery while working with logarithm tables in a library.

He observed that the pages at the beginning of the book were significantly more worn and tattered than the pages located toward the end.

This physical deterioration revealed a mathematical truth: the ten digits do not occur with equal frequency as the leading digits of numbers.

Newcomb noted that the digit 1 appears as the first significant digit far more often than the digit 9 in natural numbers.

He published this observation in the American Journal of Mathematics, suggesting that the mantissae of logarithms are equally probable in their distribution.

This foundational insight remained largely unremarked and unexplained for 57 years, existing in the literature as an evident but unexamined phenomenon.

The Physicist and the Law of Anomalous Numbers

In 1938, the physicist Frank Benford, working at the GE Research Laboratory in Schenectady NY, provided the empirical proof that solidified this observation.

Benford compiled 20,229 first-digit observations across 20 disparate datasets, including river areas, population counts, physical constants, and even street addresses.

His measurements demonstrated that the frequency of digits follows a logarithmic law, starting at 30.6% for the digit 1 and declining to 4.7% for the digit 9.

These observed frequencies closely matched the predicted logarithmic values, which range from 30.1% for the digit 1 down to 4.6% for the digit 9.

This distribution is deeply surprising to everyone who encounters it, as human intuition falsely suggests that each digit should appear with equal frequency.

Benford's comprehensive study transformed an obscure observation into a recognized law, detailing the specific frequencies for all digits from 1 through 9.

The Question of Theoretical Derivation

Benford's empirical compilation of these 20,229 observations is entirely uncontested and remains a standard in the field of statistical analysis.

However, his own original theoretical derivation of the law, which relied on a geometric-series argument, was later superseded by more rigorous work.

Specifically, the theoretical foundations he proposed were replaced by the mathematical proofs developed by Pinkham in 1961 and Hill in 1995.

Example: Expecting digits to appear with equal frequency is like assuming every shelf in a library holds an identical number of books regardless of its location.

Verdict. The leading digits of numbers in natural datasets are not evenly distributed, with the digit 1 appearing significantly more frequently than the digit 9.

Overtured if: A comprehensive replication of the 20,229 observations fails to show the logarithmic decline from 30.1% to 4.6% across diverse datasets.

Sources. 1881 · Simon Newcomb, "Note on the Frequency of Use of the Different Digits in Natural Numbers" — [PDF](#) · **1938** · Frank Benford, "The Law of Anomalous Numbers" — [PDF](#)

02. The Mantissa Circle — Why the Law Exists At All

The Anatomy of a Number

Any positive number can be expressed as a product of an integer power of ten and a fractional part known as the mantissa.

The mantissa is the value that remains once you strip away the decimal shifts, essentially mapping every number onto a circular scale between zero and one.

R.M. Fewster, in her 2009 work, explains that the first digit of a number is determined entirely by where this mantissa falls on the logarithmic cycle.

If a data set is spread uniformly across this circular mantissa space, the distribution of leading digits will inevitably conform to the specific logarithmic frequencies of Benford's Law.

Theodore P. Hill reinforced this principle in 1995 by demonstrating that base-invariance is the unique mathematical signature of this specific logarithmic distribution across different number systems.

This mechanism explains why world population data, which spans nine orders of magnitude as measured in 2009, tends to conform so naturally to these predicted probabilities.

Scale, Growth, and Convergence

The distribution holds its power because it is the only one that remains invariant when a set of numbers is multiplied by a constant factor.

Roger S. Pinkham established in 1961 that if a set follows Benford's Law, multiplying every value by a nonzero constant preserves that same characteristic distribution of digits.

Theodore P. Hill later provided a statistical derivation in 1995, showing how samples drawn from diverse distributions converge toward this law when they are pooled together.

While this provides a rigorous foundation, Fewster noted in 2009 that this convergence offers little diagnostic insight into why specific real-world datasets land precisely where they do.

Measurements of US Powerball jackpots from 2009, which span only 1.5 orders of magnitude, demonstrate that a narrow range often leads to a measurable deviation from the law.

In that case, the data showed 36 percent of values starting with one, far exceeding the expected 30 percent, while only 2.6 percent started with nine.

The Heuristic Dispute

A popular contemporary claim asserts that data must span several orders of magnitude to follow the law, but Theodore P. Hill identifies this as a mega-error.

Hill proves that a distribution spanning only a single order of magnitude can be perfectly Benford, while wide ranges do not guarantee conformity if the shape is wrong.

This represents an in-house dispute among specialists, contrasting the popular-science shorthand of the Election Integrity Partnership against Hill's formal mathematical rebuttal published in 2020.

Example: A clock face represents the mantissa, where the position of the hand dictates the leading digit regardless of how many times the minute hand has circled.

Verdict. The law describes the shape of the underlying distribution, not the raw span of the data across multiple orders of magnitude.

Overtured if: A formal mathematical proof demonstrates that a log-uniform distribution on a single order of magnitude fails to produce the Benford digit frequencies.

Sources. 2009 · R.M. Fewster, "A Simple Explanation of Benford's Law" — [PDF](#) · **1995** · Theodore P. Hill, "Base-Invariance Implies Benford's Law" — [Cal Poly DigitalCommons](#)

03. Units and Bases — Settling the Argument

The claim that is too strong

A common statement, repeated by textbooks and by chatbots, is that river lengths obey Benford's law in any unit and any base.

The objection is immediate: converting kilometres to feet multiplies every number by 3280.84, so the digits must move.

The objection is correct, and the popular statement is false as usually phrased.

What Roger Pinkham proved in 1961 runs the other way: if a digit law is to be independent of the unit, it must be Benford's, and no other law has that property.

That is a conditional about laws, not a promise about datasets, and the difference is the whole confusion.

Benford's law is unit-independent; a pile of numbers is not, unless it already happens to be Benford.

What the experiment shows

Running the test on real data settles it, and the book ships the script.

The first 300 Fibonacci numbers span 62.35 orders of magnitude and conform closely to Benford under every multiplier tested and in base 8 and base 16, with a mean absolute deviation of 0.0023 to 0.0040.

That is scale invariance: the unit changes and the answer does not.

The 132 longest rivers span only 0.84 orders of magnitude, because such a list has a floor built into it, and they fail Benford in every unit tested.

Their digits also move: 65.2 per cent lead with 1 in kilometres, 37.1 per cent with 3 in feet, 31.1 per cent with 5 in furlongs.

Mean adult height across 195 countries spans 0.06 orders of magnitude and leads with 1 for 100 per cent of countries in metres and with 5 for 99.0 per cent in feet.

Why units rotate the circle

Keep only the fractional part of a number's base-ten logarithm, and every number becomes a point on a circle of circumference one.

Multiplying everything by a constant adds that constant's logarithm to every point, rotating the circle and changing nothing else.

Benford's law is exactly the statement that those points are spread uniformly, and a uniform distribution is the only one a rotation leaves unchanged.

So the question is never which unit, but how uniformly the data is spread on the log scale: Fibonacci sits 0.0106 from uniform, the rivers 0.3764, the heights 0.7357, and the digit tables follow those three numbers exactly.

Example: Rotating a plain ring leaves it identical; rotating a ring set with one diamond moves the diamond into view — only the featureless ring survives the turn.

Verdict. Benford's law is unit-invariant; datasets are not, and a dataset is Benford in every unit only if it was already Benford in one.

Overtaken if: a dataset far from uniform on the log-mantissa circle keeps the same leading-digit frequencies under a change of unit, breaking Pinkham's uniqueness result.

Sources. 1961 · Roger S. Pinkham, "On the Distribution of First Significant Digits," *Annals of Mathematical Statistics* 32(4) — [JSTOR](#) · **1995** · Theodore P. Hill, "Base-Invariance Implies Benford's Law," *Proceedings of the AMS*

123(3) — [AMS](#) · **2026** · AICE, "The units experiment" — [source/
experiment.py](#)

04. When the Law Does Not Apply

The limits of scale invariance

Benford's law is frequently misunderstood as a universal guarantee that applies to any numerical dataset provided the numbers span a sufficient range of magnitudes.

Theodore P. Hill (2020) argues in his arXiv paper that this common shorthand regarding orders of magnitude is actually a widespread error in statistical analysis.

To test this, the 2026 AICE units experiment examined 132 of the world's longest rivers, which span only 0.84 orders of magnitude in their measurements.

The resulting mantissa KS distance of 0.3764 in this 2026 study demonstrates that narrow-range data often fails to conform to the expected Benford distribution.

Conformity is not an inherent property of data scale but rather a conditional outcome that requires specific underlying mathematical distributions to function as predicted.

The assumption that a wide span of numbers automatically ensures Benford-like behavior is a heuristic that frequently misleads those who rely on it for fraud detection.

The problem of assigned numbers

Assigned numbers such as phone numbers, ZIP codes, or specific identification numbers are inherently structured by human design rather than natural processes or growth.

Because these identifiers are typically formatted to satisfy bureaucratic requirements, they almost never exhibit the logarithmic distribution characteristic of Benford's law.

The 2026 AICE units experiment further illustrated this by measuring the mean height of individuals across 195 countries, finding a span of only 0.06 orders of magnitude.

This specific dataset produced a mantissa KS distance of 0.7357, showing how imposed minimums or narrow physical constraints lead to total nonconformity in the distribution.

When the leading digits shift drastically based on the chosen unit—such as 100% leading-1s in metres versus 99.0% leading-5s in feet—the failure of the law is absolute.

Human-assigned data remains fundamentally incompatible with the law because the digits are chosen for convenience or classification rather than emerging from a multiplicative process.

The dispute over the rule

The popular rule requiring several orders of magnitude for compliance is contested by Theodore P. Hill (2020) who labels it a mega-error.

This in-house dispute targets the shorthand currently employed by organizations such as Wolfram Research, the Election Integrity Partnership, and Reuters Fact Check.

The reliance on this flawed heuristic risks misapplying the law in contexts where it lacks the necessary mathematical foundation to provide valid evidence.

Example: Expecting a list of ZIP codes to obey Benford's law is like expecting a list of names to obey the laws of planetary motion.

Verdict. The law does not apply to assigned, bounded, or narrow-range data.

Overtured if: A rigorous mathematical proof demonstrates that narrow-range datasets with high KS distances still reliably predict digit frequencies in a meaningful way.

Sources. **2020** · Theodore P. Hill, "A Widespread Error in the Use of Benford's Law to Detect Election and Other Fraud" — [arXiv:2011.13015](#) · **1995** · Theodore P. Hill, "A Statistical Derivation of the Significant-Digit Law," Statistical Science 10(4) — [Project Euclid](#) · **2026** · AICE, "The units experiment" — [source/experiment.py](#)

05. Hill's Theorem — Why Mixtures Converge

The Mathematical Foundation of Mixtures

Theodore P. Hill provided a rigorous statistical explanation for why disparate datasets often conform to the logarithmic distribution of significant digits.

His derivation, published in the 1995 volume of *Statistical Science*, Vol. 10, No. 4, pp. 354–363, demonstrates that Benford's law is a natural consequence of data aggregation.

When probability distributions are picked at random and a random sample is drawn from each, the combined significant digits converge to Benford's law regardless of shape.

This theorem explains why Benford's own 20 disparate datasets, which lacked a singular origin, tended toward a consistent and predictable logarithmic pattern of occurrence.

The work of Theodore P. Hill (1995) suggests that the law acts as a mathematical attractor for mixtures of independent processes that might otherwise appear unrelated.

By treating the underlying distributions as random variables, the theorem shows that the law is not a coincidence but a byproduct of mixing many distributions.

Implications for Real-World Bookkeeping

Real-world bookkeeping often involves the aggregation of multiple, independent sources of financial data, which mirrors the mechanism described by Theodore P. Hill in 1995.

Because financial statements are constructed from various accounts, the resulting mixture of distributions naturally tends toward the logarithmic distribution of digits over time.

However, the explanatory reach of this mathematical phenomenon is qualified by R.M. Fewster (2009), who argues that the theorem is rigorous without being diagnostic.

Fewster (2009) notes that while the derivation explains why Benford's law appears in nature, it gives us little insight into why it applies or when.

Consequently, analysts must recognize that satisfying the requirements for a mixture does not confirm the absence of human intervention or potential manipulation in the books.

The statistical derivation provides a baseline for expected behavior but does not inherently validate the specific integrity of every individual entry within a ledger.

The Limits of the Theorem

The theorem is not contested as a valid mathematical result, yet R.M. Fewster (2009) disputes its practical utility for identifying anomalies in specific datasets.

Fewster (2009) maintains that the theorem remains a descriptive tool rather than a diagnostic one, limiting its use as a definitive detector for fraud.

The mathematical derivation is a statement about convergence in the limit, which may not always manifest in datasets with small sample sizes or specific constraints.

Example: A chorus of many different singers hitting random notes will eventually produce a harmonious average, even if no single singer is perfect.

Verdict. The theorem explains why Benford's law emerges from complexity but cannot inherently distinguish between natural growth and carefully concealed human manipulation.

Overtaken if: A rigorous replication fails to show convergence when mixing random samples from a broad, diverse set of well-defined, non-Benford probability distributions.

Sources. **1995** · Theodore P. Hill, "A Statistical Derivation of the Significant-Digit Law" — [Project Euclid](#) · **2009** · R.M. Fewster, "A Simple Explanation of Benford's Law" — [PDF](#)

06. The Auditor's Tool — Nigrini and Forensic Accounting

From Curiosity to Instrument

Mark J. Nigrini transformed Benford's law from a mathematical curiosity into a rigorous forensic-accounting screening tool in his 1999 article published in the Journal of Accountancy.

This transition required establishing that digit-frequency analysis could serve as a systematic method for detecting irregularities within large, complex financial datasets used by professional accountants.

By 1999, Nigrini demonstrated that auditors could identify potential anomalies by comparing the observed distribution of leading digits against the expected logarithmic patterns defined by the law.

The adoption of these tests by the accounting profession shifted the focus from anecdotal suspicion to a measurable, data-driven approach for identifying problematic entries.

This evolution turned abstract probability into a practical, repeatable instrument that allows forensic professionals to prioritize their investigative efforts on specific segments of financial information.

The Admissibility of Deviation

The legal standing of this methodology was solidified in the 2015 case of *United States v. Channon*, heard in the U.S. District Court for the District of New Mexico.

During these proceedings, the court reviewed a consolidated motion to exclude the testimony of forensic accountant Janet McHard, which challenged the validity of her digit-frequency analysis.

The court denied this motion on December 2, 2015, ruling that the analysis developed by Nigrini was generally accepted within the relevant community and had survived rigorous peer review.

Under the standards of Daubert and Rule 702, the court recognized that these tests are scientifically sound enough to serve as expert testimony when challenging spreadsheet evidence.

However, the court emphasized that a deviation from the expected distribution is simply an indicator that a problem exists within the data, rather than a definitive verdict of criminal fraud.

This distinction is vital because the instrument serves to flag items for a closer, manual inspection rather than functioning as an automated judge of human intent or guilt.

The Base Rate and Forensic Reality

There is no serious dispute found regarding the admissibility of these methods in professional forensic practice or within the U.S. federal court system.

The methodology remains a screening tool for identifying potential anomalies that warrant further investigation by a qualified professional.

Because a deviation is not proof of fraud, the base rate problem reminds us that an flagged entry is merely a signal to investigate further.

Example: A metal detector identifies the presence of dense material, but it cannot distinguish between a hidden gold coin and a common iron nail.

Verdict. A deviation provides objective grounds for a closer look at the underlying data but does not constitute a verdict of guilt.

Overtaken if: A court ruling determines that the digit-frequency analysis fails the Daubert standard or is no longer considered generally accepted by the forensic accounting community.

Sources. 1999 · Mark J. Nigrini, "I've Got Your Number" — [Journal of Accountancy](#) · **2015** · U.S. District Court, D.N.M., Memorandum Opinion and Order, United States v. Channon — [PDF](#)

07. National Accounts — Transparency Between States

The Stability of Numbers

Governments report specific macroeconomic data to Eurostat under the Stability and Growth Pact to maintain transparency and fiscal discipline across the diverse European Union member states.

Researchers Bernhard Rauch, Max Götttsche, Gernot Brähler, and Stefan Engel investigated whether Benford's law could serve as a diagnostic tool for detecting potential manipulation of these figures.

They examined deficit-criteria data for EU member states between 1999 and 2009 to see if the reported digits followed the expected mathematical distribution of natural numbers.

The study identified that Greek data showed the most significant deviation from Benford's law among all euro-area states during the period leading up to the 2009 revelation.

This specific finding suggests that statistical anomalies in reported accounts can serve as an early signal for auditors to investigate the underlying integrity of governmental economic claims.

The Limits of Statistical Inference

In 2009, Jesus Gonzalez-Garcia and Gonzalo C. Pastor of the IMF conducted a broader assessment by testing various macroeconomic data series against Benford's law distributions.

They compared these statistical results against independent IMF data-quality ROSC ratings to determine if nonconformity reliably pointed toward poor quality or intentional tampering with the records.

The IMF researchers discovered that a lack of conformity with Benford's law did not reliably track with the actual, independently-rated quality of the macroeconomic data series.

Instead, they observed that these deviations often reflect ordinary structural shifts within a complex economy rather than evidence of systematic deception or poor data management practices.

Their work demonstrates that the law is not a universal audit tool, as macroeconomic series frequently possess internal characteristics that naturally diverge from the expected digit distribution.

A Contested Diagnostic

The findings of Gonzalez-Garcia and Pastor directly qualify and contest the headline results presented by Rauch et al. regarding the efficacy of this statistical method.

The IMF authors explicitly state that nonconformity with Benford's law should not be interpreted as a reliable indication of poor quality in macroeconomic data reporting.

The simple act of publishing these statistics acts as a deterrent because governments know their figures are subject to public scrutiny and potential mathematical verification.

Example: A deviation from the law is like a fingerprint found at a scene; it warrants further investigation but cannot prove the crime by itself.

Verdict. Benford's law provides a checkable claim for auditors but serves as a poor substitute for a comprehensive and independent review of fiscal data.

Overtaken if: A rigorous study demonstrates that a specific, consistent deviation in deficit-criteria data perfectly predicts confirmed cases of government accounting fraud across all monitored states.

Sources. 2011 · Bernhard Rauch, Max Götttsche, Gernot Brähler, Stefan Engel, "Fact and Fiction in EU-Governmental Economic Data" — [University of Regensburg record](#) · **2009** · Jesus Gonzalez-Garcia, Gonzalo C. Pastor, "Benford's Law and Macroeconomic Data Quality" — [IMF Working Paper 09/10](#)

08. The Election Error — Where the Law Was Abused

The Seduction of Simple Statistics

Benford's Law describes how the leading digits of datasets spanning several orders of magnitude often follow a specific, predictable logarithmic distribution.

In 2020, this mathematical curiosity was repurposed as a tool to allege widespread electoral fraud within the United States precinct-level vote counts.

Walter R. Mebane, Jr., a prominent expert in election forensics, investigated these claims by applying his own rigorous statistical models to specific jurisdictions.

He examined the 2020 Milwaukee and Allegheny County precinct-level vote data to see if the alleged anomalies held any real mathematical weight.

His analysis yielded an estimated probability of "no fraud" equal to .9797 for Milwaukee and .9839 for Allegheny County, refuting the fraud narrative.

The application of the law had become a slogan used to manufacture a false sense of certainty for an audience unfamiliar with its limits.

The Failure of the First-Digit Test

The primary error in these election-related claims lies in the naive application of first-digit tests to data that does not meet the necessary criteria.

Theodore P. Hill explained in 2020 that there is a widespread error in the use of Benford's Law when the data lacks sufficient orders of magnitude.

Mebane clarified that the first digit in precinct-level vote counts is largely determined by the number of voters registered or present in each distinct precinct.

This structural feature of election data has nothing whatsoever to do with any kind of election fraud, despite what the popular, simplified narratives suggested.

Because the precinct sizes are restricted by geography and population density, they fail to span the wide numerical range required for the law to apply.

Mebane's own 2020 research found essentially no evidence that election frauds occurred in the specific jurisdictions he analyzed using his forensic models.

A Dispute of Methodology

The debate surrounding these findings is not a clash between mainstream science and fringe critics, but rather a rejection by the field's own methodologists.

Walter R. Mebane, Jr. explicitly contested the validity of the naive first-digit applications that were being circulated to characterize the 2020 election results.

Theodore P. Hill similarly contested the statistical soundness of applying these tests to election data that lacks the required orders of magnitude for validity.

Example: Using a ruler meant for inches to measure the weight of an object and claiming the resulting error proves the object is missing mass.

Verdict. The application of first-digit Benford tests to precinct-level vote counts is scientifically invalid and fails to provide evidence of electoral manipulation.

Overtured if: A study demonstrates that precinct-level vote counts naturally span multiple orders of magnitude across all jurisdictions, satisfying the fundamental requirement for Benford's Law.

Sources. 2020 · Walter R. Mebane, Jr., "Inappropriate Applications of Benford's Law Regularities to Some Data from the 2020 Presidential Election in the United States" — [PDF](#) · **2020** · Theodore P. Hill, "A Widespread Error in the Use of Benford's Law to Detect Election and Other Fraud" — [arXiv:2011.13015](#)

09. Science and Clinical Data — Integrity Screening

The Application of Distribution Tests

Digit-distribution tests serve as a diagnostic tool for researchers investigating the integrity of published scientific literature.

Andreas Diekmann noted in his 2007 study, "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data," that these methods provide a unique lens for identifying fabrication.

His work demonstrated that while faked first digits might mimic the expected Benford's Law decline, the second, third, and fourth digits often reveal significant deviations.

This research highlights that a screening process must look beyond the most prominent digit to gain a clearer understanding of the underlying numerical structure.

By examining these deeper distribution patterns, scientists can better evaluate whether the reported data aligns with the statistical expectations of natural phenomena.

The utility of this instrument lies in its ability to highlight anomalies that warrant closer inspection by the broader scientific community.

Screening Clinical Trial Results

John B. Carlisle provided a rigorous demonstration of these principles in his 2012 paper, "The Analysis of 168 Randomised Controlled Trials to Test Data Integrity."

He examined 168 randomised controlled trials by a single author, covering the period between 1991 and 2011, to determine if the reported statistics were consistent with chance.

The results were striking: 28 of the 33 tested variables, representing 85% of the data, showed distributions that were statistically inconsistent with the expected patterns.

This multi-method statistical screen, which incorporated digit-based tests, served as the primary trigger for a deeper scrutiny of the author's work.

The subsequent investigation led to a significant mass-retraction case, proving the effectiveness of these screening tools in maintaining the standards of clinical research.

The methodology serves as a warning that aggregated data should behave in predictable ways, and departures from these norms can signal a need for institutional review.

The Limits of Statistical Suspicion

There is no serious dispute found regarding the findings presented in the papers authored by either Andreas Diekmann or John B. Carlisle.

Both researchers explicitly frame their statistical results as a trigger for further scrutiny rather than as a definitive verdict of fraud.

The danger remains that a statistic strong enough to convince a reader of an anomaly is also strong enough to unfairly convict the innocent.

Example: A smoke detector indicates the presence of fire but does not identify the specific source of the heat or the intent of the person who lit the match.

Verdict. A positive result on a Benford's Law screen represents a call for a manual audit, not a final judgment of academic misconduct.

Overtaken if: A replication of the clinical trial data shows that the reported distributions were actually consistent with the underlying trial design and variables.

Sources. 2007 · Andreas Diekmann, "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data" — [Journal of Applied Statistics 34\(3\)](#) · **2012** · John B. Carlisle, "The Analysis of 168 Randomised Controlled Trials to Test Data Integrity" — [Anaesthesia, Wiley](#)

10. Second Digits and Beyond — The Fuller Law

The Expansion of Statistical Resolution

The examination of numerical distributions is not confined to the leading digit alone, as the underlying logic of the law extends to every subsequent decimal position.

Walter R. Mebane, Jr. demonstrated in 2006 that the second-digit test, known as the 2BL test, provides an additional layer of scrutiny for analyzing data sets.

By applying this method to precinct vote counts, researchers can identify patterns that might otherwise remain obscured when observing only the primary digit of a number.

However, this increased granularity does not come without a cost, as it introduces a higher frequency of false positives that require careful and nuanced interpretation.

In his 2006 study, Mebane noted that the 2BL test frequently flagged jurisdictions that were actually experiencing normal political phenomena rather than genuine instances of electoral fraud.

This illustrates the fundamental tension between achieving higher statistical resolution and maintaining the practical utility of an analytical tool in real-world environments.

Comparing the Power of Multi-Digit Tests

Dieter W. Joenssen conducted a comprehensive Monte Carlo comparison in 2013 to evaluate how effectively various tests perform across a range of different sample sizes.

His research measured performance across sample sizes from 25 to 10,000, utilizing 1,000,000 replications per cell to ensure the robustness of the statistical conclusions drawn.

The testing of the joint first-two-digit distribution gained between 1 and 3 percentage points of power with the Kolmogorov-Smirnov and Freedman-Watson tests against alternative distributions.

Despite these gains, the method is not universally superior, as Joenssen found that for the chi-squared test, one-digit testing remained favored at larger sample sizes.

The complexity of selecting the correct test-and-digit-count combination means that analysts must be wary of assuming that more digits will always yield a clearer result.

Ultimately, the trade-off between statistical power and the risk of misidentifying innocent data remains a central challenge for those applying these sophisticated numerical tests.

The Consensus on Complexity

There is no serious dispute among experts that incorporating more digits into an analysis adds information that can be valuable for detecting specific anomalies.

The literature currently disagrees on which specific test-and-digit-count combination performs best across the diverse variety of real-world scenarios encountered by investigators.

This lack of uniformity suggests that the selection of a testing protocol must be tailored to the specific nature and size of the dataset being examined.

Example: Relying on the second digit to confirm a finding is like using a magnifying glass to check a signature that already appeared blurry to the naked eye.

Verdict. Adding digits increases the sensitivity of the test but necessitates a higher threshold for evidence to avoid convicting innocent parties of fraudulent behavior.

Overtaken if: A replication of the 2013 Joenssen study demonstrates that the chi-squared test consistently outperforms multi-digit tests regardless of the sample size provided.

Sources. 2006 · Walter R. Mebane, Jr., "Election Forensics: The Second-digit Benford's Law Test and Recent American Presidential Elections" — [PDF](#) · **2013** · Dieter W. Joenssen, "Two Digit Testing for Benford's Law" — [PDF](#)

11. What It Cannot Catch

The Honest Boundary of the Instrument

Benford's law serves as a statistical signal rather than an absolute truth that can serve as the sole foundation for an accusation of fraud.

William Goodman noted in 2016 that many published claims of non-conformance in real records are merely reports of deviations within the range of ordinary variation.

These instances function as false alarms rather than evidence of misconduct because the underlying data still falls within expected statistical bounds for natural distributions.

When researchers rely on the law to detect anomalies, they often mistake natural statistical noise for the deliberate manipulation of numerical records by a human agent.

The instrument is sensitive to the structure of datasets, yet it lacks the capacity to distinguish between a creative accountant and a simple statistical outlier.

Understanding this limit is essential for any practitioner who wishes to use the distribution of leading digits to flag potential discrepancies in financial or scientific records.

The Problem of Small Samples and Estimation

Andrea Cerasa and Miguel A. Fernández conducted a rigorous Monte Carlo simulation study in 2022 to examine how these tests function at very small sample sizes.

Their research demonstrated that existing conformity tests behave unreliably when the number of observations is limited, often leading to misleading conclusions about the data integrity.

In response to these findings, Cerasa and Fernández proposed a combined three-test procedure to restore statistical power and improve the reliability of the analysis process.

Legitimate rounding and estimation artefacts can easily skew the distribution, creating patterns that appear suspicious even when the numbers were generated through entirely honest, standard processes.

Fraud that is carefully constructed to preserve the natural digit distribution remains invisible to the test, allowing a sophisticated actor to bypass the check entirely.

A test that is strong enough to convince an observer is often strong enough to convict the innocent, especially when the sample size is dangerously small.

The Status of the Findings

There is no serious dispute found in the academic literature regarding the cautionary nature of these recent findings by Goodman, Cerasa, and Fernández.

Both papers act as corrective measures within the Benford literature, emphasizing that these tests are meant to clarify the evidence rather than reject the law itself.

The consensus remains that the law is a valid regularity, provided the analyst recognizes the boundary where the mathematical model ceases to be a reliable guide.

Example: A Benford test is like a metal detector on a beach: it alerts you to something beneath the sand, but it cannot tell a gold coin from a discarded pull-tab.

Verdict. Benford conformity is a red flag, not proof of fraud, and should never be treated as a self-executing verdict of criminal intent.

Overtured if: A future meta-analysis proves that standard conformity tests maintain high sensitivity and low false-positive rates when applied to datasets containing fewer than fifty items.

Sources. 2016 · William Goodman, "The Promises and Pitfalls of Benford's Law" — [Significance, Wiley](#) · **2022** · Andrea Cerasa, Miguel A. Fernández, "Testing for Benford's Law in Very Small Samples" — [PLOS One](#)

12. The Procedure — Running the Test Yourself

The Preliminary Assessment

A researcher must begin by checking the range and the mantissa spread of the data before applying any mathematical expectations regarding digit frequency.

R.M. Fewster in 2009 established that these properties provide the necessary foundation for any analysis, as seen in the method described in her paper.

In 2026, the AICE institution applied this logic to three distinct datasets, specifically measuring Fibonacci numbers, 132 rivers, and 195 countries' mean heights.

These measurements confirm that the span of the data determines whether the test is even applicable to the set in question before proceeding further.

Failing to verify the mantissa spread invites error, as the distribution is not a universal law but a property dependent on specific data characteristics.

By evaluating these metrics first, one avoids the common mistake of assuming that all numerical data must conform to the expected logarithmic distribution.

The Necessity of Unit Variation

The procedure demands that the test be run in more than one unit to ensure that the findings are not merely artifacts of measurement.

As demonstrated by AICE in 2026, testing the three datasets across six length units revealed how sensitive the leading-digit output is to scale.

The Fibonacci numbers showed close conformity with a KS 0.0106 distance, remaining stable across every unit and base tested during the experiment.

Conversely, the 132 rivers with a KS 0.3764 and 195 countries' heights with a KS 0.7357 both displayed clear nonconformity in every unit used.

For those failing sets, the leading-digit answer shifted whenever the unit changed, providing concrete proof that a single-unit test proves nothing on its own.

An investigation that relies on a single set of units ignores the possibility that the apparent pattern is an illusion created by the scale.

The Interpretation of Results

Conformity is not an absolute proof of truth or accuracy, but merely an observation that the data follows a specific mathematical regularity.

No serious dispute exists regarding these findings, as the distinction between conforming and nonconforming datasets is supported by the rigorous 2009 method.

One must treat the result as a question to be investigated further, never as a final answer or a definitive judgment on the data.

Example: A ruler that changes length depending on the temperature is a poor tool for measuring the structural integrity of a bridge.

Verdict. A test result is only as robust as the variety of units used to confirm its stability across different scales of measurement.

Overtaken if: A dataset currently marked as nonconforming, such as the 195 countries' mean heights, were to show consistent conformity across all tested units.

Sources. 2009 · R.M. Fewster, "A Simple Explanation of Benford's Law" — [PDF](#) · **1999** · Mark J. Nigrini, "I've Got Your Number" — [Journal of Accountancy](#)