NetSciX2026: International School and Conference on Network Science
February 17th-20th, 2026 - Auckland, New Zealand

Testing Most Influential Sets

influence, data attribution, model sensitivity, fairness, interpretable

Extended Abstract

Small subsets of data with disproportionate influence on model outcomes can dramatically impact conclusions, with just a few data points sometimes overturning key findings. While recent work has developed methods to identify these most influential sets [1, 2], no formal theory exists to determine when their influence reflects genuine problems rather than natural sampling variation. We address this gap by developing a principled framework for assessing the statistical significance of most influential sets.

Problem and Motivation. Machine learning models can be highly sensitive to small subsets of data. In many applications, just a handful of samples can overturn key conclusions: two countries nullify the estimated effect of geography on development, a single outlier flips the sign of a treatment effect, or a small group of individuals drives disparate outcomes in algorithmic decision-making. Current practice relies on domain expertise and ad-hoc sensitivity checks, while approximate methods such as influence functions systematically underestimate the impacts of sets and extreme cases [3]. What remains missing is a principled method to distinguish natural sampling variation from genuinely excessive influence.

Approach. We develop a statistical framework for assessing the significance of most influential sets by focusing on linear regression — a tractable, interpretable, and widely-used setting. We derive exact asymptotic distributions of maximal influence using extreme value theory. Our key theoretical contributions show that two distinct regimes emerge depending on the size of the influential set:

1. Constant-size sets: When the size k remains fixed as sample size N grows, maximal influence converges to a heavy-tailed Fréchet distribution

2. Relative-size sets: When k grows proportionally with N, maximal influence converges to a well-behaved Gumbel distribution

For exact influence computation, we derive a closed-form solution showing that the influence of set $\mathscr{S}$ on parameter θ is: $\Delta(\mathscr{S}) = \frac{\sum_{i\in\mathscr{S}} x_i r_i}{\sum_{n\notin\mathscr{S}} x_n^2}$, where x are feature values and r are residuals. This formula reveals the additive structure of individual contributions in the numerator and multiplicative adjustment from remaining data in the denominator, enabling efficient computation without explicitly forming leverage scores.

Results. We validate our theoretical predictions through controlled simulations across different distributional assumptions and demonstrate practical utility through real-world applications. In economics, we resolve the controversial 'Blessing of Bad Geography' [4] finding by showing that the Seychelles exerts statistically excessive influence ($p < 0.001$). In biology, we identify two data points in sparrow morphology data that excessively influence head-tarsus correlations (both $p < 0.001$). Across machine learning benchmarks (Law School, Adult Income, Boston Housing, Communities & Crime), we can distinguish between natural variation and problematic influence patterns.

Significance and Impact. This work provides the first rigorous statistical framework for assessing when most influential sets represent genuine problems rather than natural sampling variation. By establishing that maximal influence follows predictable extreme value distributions, we enable practitioners to move beyond ad-hoc rules and domain-specific judgment. Our results enable principled hypothesis tests for excessive influence, replacing current ad-hoc diagnostics with rigorous statistical procedures.

Conclusion. By transforming the assessment of most influential sets from art to science, our method offers clear guidance to practitioners—when small sets overturn results of interest, our tests reveal whether this influence is statistically excessive, enabling more robust and transparent decision-making in settings where reliability matters.

References

[1] Tamara Broderick, Ryan Giordano, and Rachael Meager. *An automatic finite-sample robustness metric: Can dropping a little data make a big difference?* 2021.

[2] Y. Hu et al. "Most Influential Subset Selection: Challenges, Promises, and Beyond". In: *Conference on Neural Information Processing Systems (NeurIPS 2024)*. Vol. 38. 2024.

[3] Pang Wei Koh et al. *On the Accuracy of Influence Functions for Measuring Group Effects*. arXiv:1905.13289 [cs]. 2019. URL: `http://arxiv.org/abs/1905.13289`.

[4] Nathan Nunn. "The historical roots of economic development". In: *Science* 367.6485 (2020). Publisher: American Association for the Advancement of Science, eaaz9986.

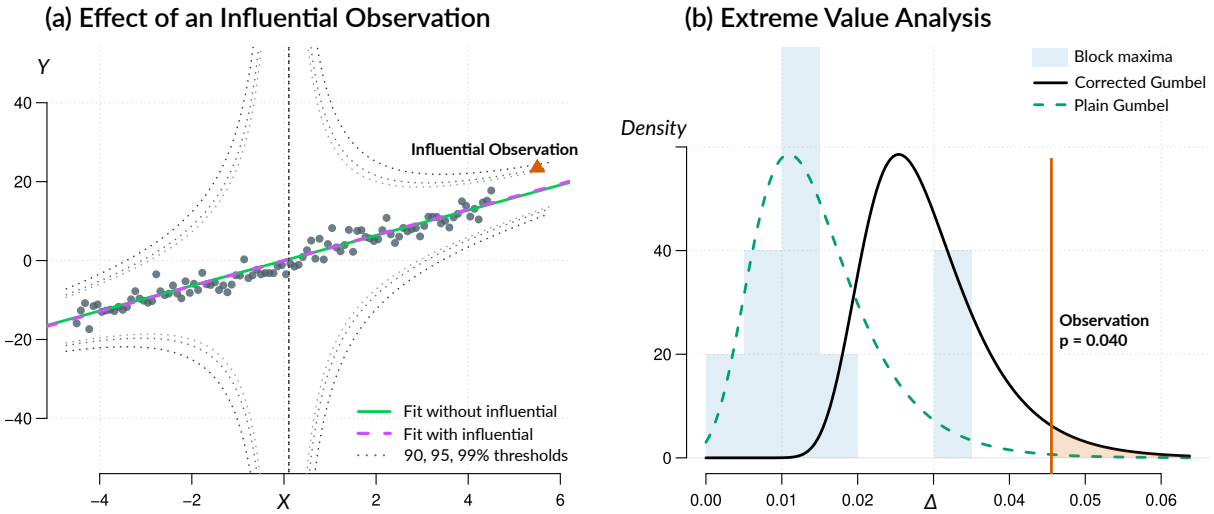

Figure 1: Illustration of our methodology on a simple linear regression with a moderately influential observation. Panel A depicts observations, estimated regression lines with and without the influential point, and conditional significance regions at the 10, 5, and 1% levels (dotted lines). Panel B shows the extreme value analysis: a histogram of block maxima, fitted Gumbel distribution with (solid) and without (dashed) bias correction, and the resulting p-value for the observation of interest.