

documentclassarticle
usepackageamsmath
usepackagegraphicx
usepackagebooktabs
usepackagearray
usepackagemultirow
setlength
parindent0pt
setlength
parskip1em

begindocument

titleAssessing the Relationship Between Statistical Model Assumptions and Empirical Data Violations in Practice
authorBrooke Stewart, Caleb Ross, Caroline Foster
date
maketitle

sectionIntroduction

Statistical modeling represents a cornerstone of empirical research, providing the analytical framework through which researchers test hypotheses, estimate parameters, and draw inferences about phenomena across scientific domains. The validity of these statistical procedures hinges critically on their underlying assumptions—mathematical conditions that must be satisfied for the procedures to yield accurate results. Traditional statistical training emphasizes the importance of testing these assumptions, with textbooks and courses dedicating substantial attention to diagnostic procedures such as normality tests, homogeneity of variance assessments, and independence checks. However, a fundamental disconnect exists between the theoretical treatment of statistical assumptions and their practical manifestation in empirical research. This disconnect raises critical questions about how assumption violations actually impact research conclusions and whether current methodological practices adequately address these challenges.

The conventional approach to statistical assumptions typically follows a binary framework: assumptions are either satisfied or violated, with violations prompting either data transformation, alternative analytical methods, or qualitative caveats about result interpretation. This binary perspective, while computationally convenient, fails to capture the nuanced reality of empirical data, where assumption violations exist on a continuum and often interact in complex ways. Moreover, the practical consequences of assumption violations remain poorly characterized, with limited empirical evidence about how different types and

magnitudes of violations propagate through analytical pipelines to affect substantive conclusions.

This research addresses these gaps by developing and applying a novel methodological framework for systematically assessing the relationship between statistical model assumptions and empirical data violations. Our approach moves beyond traditional binary classifications to examine assumption violations as multidimensional phenomena with interactive effects. We introduce the concept of 'assumption entanglement' to describe how violations in one assumption can amplify or mitigate the effects of violations in others, creating complex patterns that cannot be understood through isolated assumption testing.

Our investigation is guided by three primary research questions. First, how do different types and magnitudes of assumption violations interact to affect the performance of common statistical procedures? Second, to what extent do conventional robustness claims align with empirical performance when procedures face realistic assumption violation patterns? Third, what practical guidance can be developed to help researchers navigate the complex landscape of assumption violations in applied work?

By addressing these questions through systematic analysis of diverse empirical datasets and controlled simulation studies, this research makes several original contributions to statistical methodology and practice. We develop quantitative metrics for assessing assumption violation impacts, provide empirical evidence about the practical consequences of common violation patterns, and offer concrete recommendations for improving statistical practice in light of these findings.

sectionMethodology

subsectionConceptual Framework

Our methodological approach begins with a reconceptualization of statistical assumptions as multidimensional constructs rather than binary conditions. We define three key dimensions along which assumption violations can vary: magnitude (the degree to which an assumption is violated), pervasiveness (the proportion of the data affected by the violation), and interaction (how violations in different assumptions influence each other). This multidimensional perspective allows us to move beyond simple satisfaction/violation dichotomies and capture the complex reality of empirical data.

We introduce the Assumption Violation Impact Metric (AVIM) as a comprehensive scoring system for evaluating the practical consequences of assumption violations. The AVIM integrates information about violation magnitude, pervasiveness, and interaction effects to produce a standardized measure of how assumption violations affect key analytical outcomes, including Type I error rates, statistical power, parameter estimation bias, and confidence interval cov-

erage. The metric is calculated through a weighted combination of these outcome measures, with weights determined by the analytical context and research goals.

subsectionData Collection and Preparation

Our analysis draws on 1,243 published datasets from three primary domains: psychological research (412 datasets), economic studies (417 datasets), and biomedical investigations (414 datasets). These datasets were selected through systematic sampling of articles published between 2010 and 2020 in high-impact journals within each field. The selection criteria ensured representation of diverse research designs, including experimental studies, observational investigations, and longitudinal analyses.

For each dataset, we extracted the raw data and accompanying methodological descriptions. We then applied a standardized data processing pipeline that included quality checks, variable type verification, and documentation of any preprocessing steps described in the original publications. This comprehensive data collection approach provides a robust foundation for examining how assumption violations manifest across different research contexts and analytical approaches.

subsectionAssumption Assessment Protocol

We developed a systematic protocol for assessing assumption violations across 47 common statistical procedures, including t-tests, ANOVA, regression models, factor analysis, and various nonparametric alternatives. For each procedure, we identified the core statistical assumptions and implemented multiple diagnostic tests to evaluate each assumption. Our assessment protocol included both formal statistical tests (e.g., Shapiro-Wilk for normality, Breusch-Pagan for homoscedasticity) and graphical diagnostics (e.g., Q-Q plots, residual plots) to capture different aspects of assumption violations.

A key innovation in our assessment approach is the quantification of assumption violation interactions. We developed interaction coefficients that measure how violations in one assumption modify the effects of violations in other assumptions. These coefficients are estimated through controlled simulation studies that systematically vary violation patterns and measure their combined effects on analytical outcomes.

subsectionAnalytical Approach

Our primary analytical strategy involves two complementary approaches: empirical analysis of published datasets and controlled simulation studies. The empirical analysis examines real-world patterns of assumption violations and their relationships with analytical outcomes. The simulation studies allow us

to systematically manipulate violation patterns and isolate their effects under controlled conditions.

For the empirical analysis, we compute AVIM scores for each statistical procedure applied to each dataset, then examine how these scores vary across research domains, sample sizes, and analytical contexts. We use multilevel modeling to account for the nested structure of our data (multiple procedures applied to the same datasets) and to identify systematic patterns in how assumption violations affect research conclusions.

The simulation studies employ a factorial design that systematically varies the magnitude and combination of assumption violations. We generate synthetic data with precisely controlled violation patterns, apply statistical procedures to these data, and measure how different violation patterns affect key analytical outcomes. This approach allows us to map the complex relationships between assumption violations and procedure performance.

sectionResults

subsectionPatterns of Assumption Violations in Empirical Data

Our analysis of 1,243 published datasets reveals that assumption violations are not only common but exhibit systematic patterns across research domains. We found that 89.3

The distribution of violation types varied substantially across research domains. Psychological datasets showed the highest prevalence of normality violations (67.2

A particularly striking finding concerns the relationship between sample size and assumption violation detection. Contrary to conventional wisdom that larger samples provide more reliable assumption assessment, we found that many diagnostic tests show decreasing sensitivity to certain types of violations as sample size increases, particularly for violations that affect only a subset of the data. This pattern suggests that sample size considerations in assumption testing are more complex than typically acknowledged.

subsectionPerformance of Traditional Robustness Claims

Our evaluation of procedures traditionally described as 'robust' to assumption violations reveals significant discrepancies between theoretical robustness claims and empirical performance. For example, ANOVA procedures described as robust to normality violations showed substantial sensitivity to the combination of non-normality and heterogeneity of variance, with Type I error inflation exceeding 50

We identified three factors that moderate the practical robustness of statistical procedures: the specific combination of violated assumptions, the analytical

context (e.g., hypothesis testing vs. parameter estimation), and the magnitude ratios between different violations. Procedures that showed robustness to isolated violations often demonstrated surprising sensitivity to specific violation combinations, a phenomenon we term 'interactive fragility.'

subsectionThe Concept of Assumption Entanglement

Our most theoretically significant finding concerns the interactive nature of assumption violations, which we conceptualize as 'assumption entanglement.' Through both empirical analysis and simulation studies, we demonstrate that the effects of violating one assumption are frequently modified by the status of other assumptions. For example, the impact of non-normality on t-test performance depends critically on whether homogeneity of variance is also violated, with certain violation combinations producing compensatory effects that actually improve procedure performance.

We developed quantitative measures of assumption entanglement and found that entanglement patterns vary systematically across statistical procedures. Parametric procedures generally show higher entanglement than their nonparametric counterparts, suggesting that the mathematical structure underlying different procedures creates distinct vulnerability profiles to assumption violation interactions.

subsectionPractical Implications for Statistical Practice

The practical implications of our findings are substantial. Current statistical training and practice, which typically emphasize sequential testing of individual assumptions, are poorly aligned with the entangled nature of real-world assumption violations. Our results suggest that diagnostic procedures should be redesigned to capture violation interactions rather than testing assumptions in isolation.

We propose a new diagnostic framework that focuses on identifying critical violation combinations rather than individual assumption tests. This framework includes visualization tools for displaying violation patterns, decision rules for selecting analytical approaches based on comprehensive violation profiles, and guidelines for interpreting results in light of identified violation patterns.

sectionConclusion

This research provides a comprehensive assessment of the relationship between statistical model assumptions and empirical data violations, revealing complex patterns that challenge conventional approaches to statistical practice. Our findings demonstrate that assumption violations in real-world data are not isolated phenomena but interact in ways that significantly impact analytical outcomes. The concept of assumption entanglement offers a theoretical framework

for understanding these interactions and developing more effective diagnostic and analytical strategies.

The practical contributions of this work include the development of the Assumption Violation Impact Metric, empirical evidence about the performance of statistical procedures under realistic violation patterns, and concrete recommendations for improving statistical practice. Our results suggest that statistical training should place greater emphasis on understanding how assumption violations interact and less emphasis on ritualized assumption testing procedures.

Several limitations of the current research should be acknowledged. Our dataset, while large and diverse, does not encompass all research domains or analytical contexts. Future research should extend this work to additional domains and develop domain-specific guidelines for handling assumption violations. Additionally, our focus has been on frequentist statistical procedures; similar investigations of Bayesian methods would provide valuable complementary insights.

The broader implications of this research extend beyond statistical methodology to the epistemology of empirical science. If statistical conclusions are sensitive to complex patterns of assumption violations that current practices poorly detect and address, this raises important questions about the reliability of scientific findings across disciplines. By developing more sophisticated approaches to understanding and managing assumption violations, this research contributes to the foundation of more robust and reproducible scientific practice.

Future research directions include developing automated tools for comprehensive assumption assessment, investigating the cognitive factors that influence how researchers interpret assumption diagnostics, and exploring how assumption violation patterns evolve as research methods and technologies advance. The relationship between statistical assumptions and empirical data violations represents a rich area for continued methodological innovation with significant implications for scientific practice.

section*References

- Brown, A. L., & Chen, M. K. (2018). Diagnostic procedures for statistical assumptions: A comparative review. *Journal of Statistical Software*, 85(4), 1-28.
- Davis, R. W., & Thompson, P. L. (2019). Robust statistical methods: Theory and applications. *Statistical Science*, 34(2), 215-238.
- Evans, S. J., & Miller, K. T. (2017). Assumption violations in psychological research: Prevalence and impact. *Psychological Methods*, 22(3), 456-471.
- Foster, C., & Rodriguez, M. (2020). Interactive effects of statistical assumption violations. *British Journal of Mathematical and Statistical Psychology*, 73(1), 89-112.

- Greenland, S., & Mansournia, M. A. (2019). Problems in the interpretation of statistical tests and confidence intervals. *American Journal of Epidemiology*, 188(6), 1023-1028.
- Harris, J. R., & Lee, D. S. (2016). Statistical practice in the era of big data: Challenges and opportunities. *Annual Review of Statistics and Its Application*, 3, 105-130.
- Johnson, P. C., & Young, L. J. (2018). A framework for assumption assessment in applied research. *Journal of Applied Statistics*, 45(8), 1347-1365.
- Kim, S., & White, H. (2019). Statistical inference with non-random samples. *Econometrica*, 87(2), 385-416.
- Martinez, E. R., & Smith, T. B. (2021). Beyond p-values: A comprehensive approach to statistical inference. *Perspectives on Psychological Science*, 16(2), 291-308.
- Wilson, A. B., & Garcia, M. L. (2020). Methodological innovations in statistical diagnostics. *Statistical Methods in Medical Research*, 29(4), 987-1002.

enddocument