Check for updates

Received: 14 October 2024 | Revised: 18 August 2025 | Accepted: 18 August 2025

DOI: 10.1111/test.70012

ORIGINAL ARTICLE

WILEY

This is not normal! (Re-) Evaluating the lower *n* guidelines for regression analysis

David Randahl

Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden

Correspondence
David Randahl, Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden.
Email: david.randahl@pcr.uu.se

Funding information
Riksbankens Jubileumsfond, Grant/Award Number: M21-0002; European Research Council, Grant/Award Number: H2020-ERC-2015-AdG 694640

Abstract

The rule of thumb that a sample size of $n \geq 30$ is sufficient for valid regression analysis is often cited but rarely scrutinized. This paper evaluates the minimum number of observations required by examining how distributional characteristics like skewness and kurtosis affect the convergence of estimated *t*-statistics to the *t*-distribution in linear models. Through simulations involving over 22 billion regressions, sample sizes from 4 to 10,000 are tested across various combinations of distributions. Results show that it is sufficient that either the dependent or independent variable follow a symmetric distribution for the *t*-statistics to converge at much smaller sample sizes than $n = 30$, unless the other variable is extremely skewed. However, when both variables are highly skewed, much larger sample sizes are required. These findings suggest the $n \geq 30$ rule is overly conservative in some cases and insufficient in others, offering revised guidelines for determining minimum sample sizes.

KEYWORDS

central limit theorem, linear regression, normality, number of observations, simulation

1 | INTRODUCTION

When I was a student and attended my first quantitative methodology course at university, I was told that for regression modeling to be appropriate we would either have to assume that the distribution of the error term (ϵ) was normally distributed, *or* that the sample size was large enough so that the central limit theorem would ensure that the resulting statistics of interest would follow an approximate normal distribution. When the professor was questioned about the required sample size they replied that a common rule of thumb was $n = 30$. At the time, I did not think much more about that but simply accepted the $n \geq 30$ rule of thumb.

A decade on, however, as I was preparing to teach the same material to a new cohort of students in political science, I decided to demonstrate the Central Limit Theorem in action by progressively increasing the sample size (n) of dice throws and calculating the mean. During this exercise, I realized that the distribution of the mean of dice throws converged to the normal distribution at much lower n than 30. I also found that other, more skewed, distributions took much longer to converge. Around the same time, a group of master's students I was supervising began pestering me for a reference on the $n \geq 30$ rule of thumb to use in their thesis. This was, of course, not the first time I had been asked for this, but these students were unusually persistent as they were collecting their own data using time consuming in-depth interviews and wanted to know the absolute minimum number of interviews they needed to conduct.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2025 The Author(s). *Teaching Statistics* published by John Wiley & Sons Ltd on behalf of Teaching Statistics Trust.

Certain that there would be plenty of systematic tests of this rule of thumb, I decided to look into where the rule of thumb came from, and what alternative rules of thumb existed in order to provide my students with a reference. To my surprise, while both references to and criticisms of the $n \geq 30$ and other rules of thumb were rife, few systematic investigations of these rules of thumb for regression analysis or similar methods relying on the central limit theorem or the normality assumption for the inferences (e.g., *t*-tests of means, differences in means, ANOVA, etc.) seemed to exist. In addition, with one notable exception, the investigations that did exist did not focus on the distributional assumptions of the regression analysis but rather on other (important) characteristics such as power. In this article I aim to critically examine the lower bound for the number of observations in regression analysis using simulation, and to provide guidance on the lower bound of n for linear regression analysis. The results show that we should likely revise both what we teach about the lower bound of observations in regression and under what circumstances we need to report distributional characteristics such as kurtosis and skewness when running regression analysis.

2 | THE DISTRIBUTIONAL ASSUMPTION OF LINEAR REGRESSION

The classical linear regression model assumes that the dependent variable Y is a linear function of the independent variable(s) X and an error term ϵ: $Y = \beta_0 + \beta_1 X + \epsilon$. The parameters are commonly estimated using the ordinary least squares (OLS) method, which minimizes the sum of the squared residuals. Under the condition that the error term, ϵ, is normally distributed, the resulting estimated coefficients, $\widehat{\beta}$, follow a normal distribution, in repeated sampling. When estimating the standard errors of the estimated coefficients and standardizing them, the resulting quantity, $\left(\widehat{\beta}_j - \beta_j\right)/\mathrm{SE}\left(\widehat{\beta}_j\right)$, follows a t distribution with $n-k$ degrees of freedom in repeated sampling, where k is the number of parameters in the model (Hogg et al., 2013; Wooldridge, 2012).

Inferences about the population parameters can then be made using the t distribution and its associated critical values or p-values. Crucially, the inferences about the parameters are generally made by testing the **null hypothesis** that the population parameter is equal to zero versus the **alternative hypothesis** that the population parameter is different from zero (other hypothesis tests are also possible, but the null hypothesis that the population parameter is equal to zero is the most common in regression analysis). These tests are conducted under the assumption that the **null hypothesis** is true, and thus that the estimated t-statistic follow a t-distribution with a mean of zero. A result is considered statistically significant at a specific significance level, α, (typically 0.05) if the absolute value of the estimated t-statistic exceeds the critical value from the t-distribution (or, equivalently, if the corresponding p-value falls below the chosen significance level), indicating that such an extreme result would be unlikely if the null hypothesis is true,. If the t-statistic is smaller than the critical value (or the p-value is higher than the threshold α), we fail to reject the null hypothesis.

In this setting, the distributional assumption of the error term ϵ is essential for the validity of inference in the classical linear regression model. While the OLS estimator remains unbiased and consistent under weaker conditions (such as zero-mean and homoskedastic errors), the normal distribution of ϵ is what justifies the use of the t-distribution when conducting hypothesis tests on the regression coefficients. Without this assumption, there is no guarantee that the estimated t-statistic will follow a t-distribution in repeated samples (see for instance Hesterberg, 2008).

If the t-statistics deviate from the expected t-distribution, the resulting critical values and p-values may not be valid. This can have serious consequences for interpretation. For example, a p-value of 0.03 might appear to provide strong evidence against the null hypothesis at the conventional significance level of 0.05, but if the estimated t-statistic does not follow a t-distribution, this p-value does not accurately reflect the true probability of observing such a result under the null. In such cases, we risk drawing incorrect conclusions: rejecting a true null hypothesis (inflated type I error rate) or failing to detect real effects (loss of power).

However, the normality assumption of linear regression can be, and typically is, relaxed when the sample size is sufficiently large. This is because of the central limit theorem which ensures that when the sample size is large $\widehat{\beta}$ will follow an approximate normal distribution and the estimated t-statistics will similarly follow an approximate t-distribution in repeated sampling, even if the error term ϵ is not normally distributed (Wooldridge, 2012).

3 | GUIDELINES FOR *N* IN LINEAR REGRESSION

For inference to be valid in linear regression it is thus necessary for either the error term ε to be normally distributed, or the sample size to be large enough for the central limit theorem to apply. However, what constitutes

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

large enough is less clear. A number of different rules of thumb have been proposed in the literature, based on different criteria such as a minimum overall number of observations, or a minimum number of observations per parameter.

Among the most commonly cited rules of thumb is the $n \geq 30$ rule, which suggests that a sample size of at least 30 observations is required for the central limit theorem to apply and thus for the inferences from parametric tests, such *t*-tests of means or difference in means, linear regression, and similar, to be valid. While this rule of thumb is widely used and cited in a variety of textbooks and papers on statistics and quantitative research methods across several different disciplines (see for instance Hogg et al., 2013; Ross, 2017; Kiernan, 2014; Carsey & Harden, 2013; Hesterberg, 2008; Cohen, 1992; Wooldridge, 2012; Thompson, 2023; Kwak & Kim, 2017), its justification seems only loosely related to the specific distributional assumptions of regression analysis. While some textbooks note that this rule of thumb may be too conservative or too permissive in certain cases (e.g., Hogg et al., 2013; Ross, 2017; Wooldridge, 2012), few provide any explicit motivation for this rule of thumb.

Others, such as Kwak and Kim (2017), suggest that the rule of thumb is based on the fact that the *t* distribution is nearly indistinguishable from the normal distribution when $\nu \geq 30$ degrees of freedom. A keen reader will, however, note that the convergence characteristics of the *t*-distribution to the normal distribution do not say anything about whether or not the statistics of interest themselves have converged to the *t* distribution. Yet another possible origin for the rule of thumb is discussed in a highly interesting thread on cross-validated which cites the fact that a *t*-table with 30 degrees of freedom neatly fit on a single page as a potential origin for the rule of thumb.

Beyond the $n \geq 30$ rule of thumb and other similar rules of thumb based on fixed thresholds, such as the $n \geq 50$ or $n \geq 200$ rules of thumb, there are also rules of thumb for the minimum sample size based on the number of parameters in the model. These rules of thumb, for example, the $n \geq 20 \cdot k$ or $n \geq 50 + k$ rules, are based on the idea that the sample size should be large enough to ensure that the estimated coefficients are approximately normally distributed. These rules of thumb are often cited to ensure that effects of certain sizes are captured, that is, to ensure that the regression model has sufficient *power* (for a longer discussion on these rules of thumb, see [Green, 1991]). Similarly, minimum desired sample sizes can also be calculated based on the desired power of the model, that is, the probability of correctly rejecting the null hypothesis when it is false, and the estimated, expected, or minimal relevant effect size (see for instance Gelman & Carlin, 2014; Gelman et al., 2021). However, once again, the *power* of the model does not tell us anything about the convergence characteristics of the statistics of interest.

The $n \geq 30$ rule is not without its critics and it has long been noted that distributional characteristics such as the *skew* affect the utility of this rule of thumb. Already back in 2008, Hesterberg, argued that the "$n \geq 30$ and not too skewed" rule of thumb should be retired in favor of bootstrapping techniques which are not reliant on these distributional characteristics. Yet, Hesterberg shied away from actually subjecting the $n \geq 30$ rule of thumb to a rigorous test.

Motivated by the lack of systematic evidence for the $n \geq 30$ rule of thumb for parametric statistics in general, Thompson (2023) embarked on a simulation study to investigate the utility of this rule of thumb when estimating the parameters of a set of common distributions. In this simulation study, Thompson utilizes a set of distributional tests and a convolutional neural network to ascertain for what n the distribution of the parameters becomes approximately normal. The results show that the skew of the distribution is important for the n that should be considered the lower bound.

This article builds on a similar type of simulation design, but focuses exclusively on regression analysis and how the distributional characteristics of the dependent and independent variables interact with n to affect what a reasonable lower bound of n is for regression analysis.

4 | SIMULATION SETUP

To evaluate the validity of distributional rules of thumb and to offer guidance on the lower bound of n in regression analysis, I conduct a large-scale simulation study. The design of the study is as follows:

1. Generate the variables under the null hypothesis, given a specific sample size and specific distributions for y and x.
2. Estimate a linear regression model with y as the dependent variable and x as the independent variable using ordinary least squares (OLS).
3. Extract the estimated *t*-statistics from the regression model.
4. Repeat steps 1–3 1000 times to produce a distribution of estimated *t*-statistics.
5. Test whether the estimated *t*-statistics are *t*-distributed with $\nu = n - 2$ degrees of freedom, where n is the sample size and 2 is the number of parameters in the model, using the Anderson–Darling test $\alpha = 0.05$ significance level.

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6. Calculate the proportion of estimated t-statistics which falsely reject the null hypothesis of the t-test at the $\alpha = 0.05$ level.
7. Repeat steps 1–6 500 times and calculate the proportion of times the null hypothesis of the Anderson–Darling test is rejected.
8. Repeat steps 1–7 for a range of sample sizes n and combinations of distributions for y and x.

The simulation design includes several important features. First, all variables are generated such that y is independent of x, and its theoretical mean is subtracted to center the distribution. This ensures that $\boldsymbol{beta} = 0$, that is, that the null hypothesis of the regression model is true.

Second, since the data are generated under the null, the resulting t-statistics should follow a t-distribution with $n-2$ degrees of freedom. Step 5 formally tests this using the Anderson–Darling test. As a robustness check, I also apply the Kolmogorov–Smirnov and Cramér–von Mises tests. Results from these complementary tests are reported in Appendix S1.

Third, I estimate the empirical type I error rate for each coefficient across simulations, allowing an assessment of whether the t-statistics are inflated (leading to false positives) or deflated (indicating loss of power). From an inferential standpoint, inflated t-statistics are more concerning, as they can lead to incorrect conclusions, whereas deflated values primarily reduce statistical power.

4.1 | Sample sizes and distributions

To evaluate how sample size and distributional characteristics, particularly skewness and kurtosis, affect the convergence of t-statistics to the t-distribution, I simulate data using a broad range of combinations of sample sizes and distributions for both the independent and dependent variables. These include leptokurtic (kurtosis greater than the normal), mesokurtic (kurtosis similar to the normal), and platykurtic (kurtosis less than the normal) symmetric distributions as well as a range of variously skewed distributions. The tested distributions with their corresponding skewness and kurtoses can be found in Table 1 in Appendix S1.

The simulation covers 54 unique sample sizes, ranging from $n = 4$ to $n = 10{,}000$, with denser sampling at smaller values of n (see Table 2 in Appendix S1). In total, 6534 combinations of n, x, and y distributions are tested, each using 500 replicates of 1000 regression models—for a total of approximately 3.3 billion simulations.

As a further extension of the simulation, I also introduce a second independent variable z, generated independently from both x and y, using a smaller set of unique distributions. In this multivariate setup, I test an additional 37,800 combinations, bringing the overall total to approximately 22.2 billion simulated regression models across both bivariate and multivariate settings.

5 | RESULTS

The results for a select few combinations of x and y distributions are shown in Figure 1 below. Results for all combinations of distributions are available in Appendix S1, as are the results using the alternative distributional tests. These results highlight some interesting patterns. First, it is not only when the residuals of the regression model are normal that regression modeling seems to perform well with a low number of observations n. Rather, as long as both x and y are *symmetric* the distribution of the t-statistics converges to the t-distribution already at $n = 4$, except in the most extreme platykurtic case (Beta 0.1,0.1) where convergence is reached at $n = 8$. Similarly, if *either* x or y are symmetric and the other is skewed the distribution of t-statistics also converge to the t-distribution at $n = 4$ unless the skewness of the other variable is severe (e.g., log-normal with $\sigma \geq 1$). This highlights that, contrary to theory, both the distribution of x and y matter for the convergence of the t-statistics to the t-distribution.

When both y and x are skewed, the rate of convergence seems to depend on both the skew of x and of y. In the most extreme cases, where both x and y follow log-normal distributions with parameter $\sigma \geq 1$, the t-statistics do not converge to the t-distribution even at $n = 10{,}000$. Except when y or x are normal, the distribution of t-statistics do not converge to the t-distribution at $n = 10{,}000$ when the other variable follows a log-normal distribution with $\sigma = 2$.

5.1 | Effects of non-convergence to the t-distribution

To assess the effects of the t-statistics not converging to the t-distribution I computed the type-I error rate across simulations. The results for a select few combinations of x and y distributions are shown in Figure 2 below. Results for all combinations of distributions are available in Appendix S1. As discussed earlier, the main concern about the t-statistics not converging to the t-distribution is that the p-values obtained from the regression analysis are no longer valid. This is especially problematic if the

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

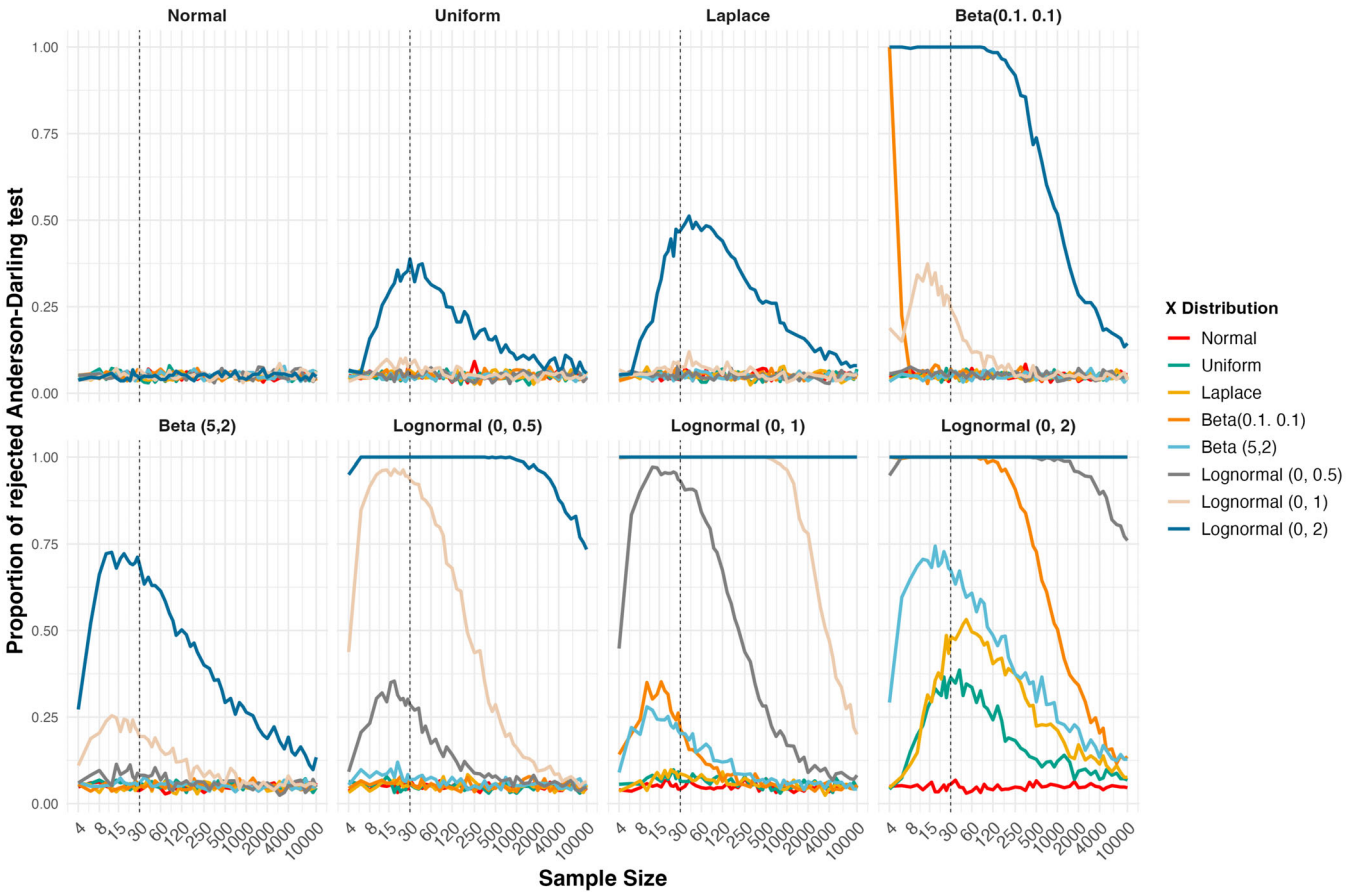

FIGURE 1 Anderson–Darling rejection rate at $\alpha = 0.05$ for different combinations of x and y distributions and n. Facets represent different distributions of y. [Colour figure can be viewed at wileyonlinelibrary.com]

risk of type-I errors is *higher* than α since we then risk rejecting null hypotheses when our evidence does not support this rejection.

The results from the simulation show that the consequences of the t-statistics not converging to the t-distribution depends on the skewness and kurtosis of the variables involved, and particularly the kurtosis. For the cases where the type-I error rate is not at the appropriate level and either of the variables are platykurtic, the type-I error rate is *lower* than α, except when $n < 8$. If neither of the variables are platykurtic but are *symmetric*, the type-I error rate is *higher* than α. The most extreme cases, where both x and y are log-normal with $\sigma \geq 1$, exhibit a strange pattern where the type-I error rate is higher than alpha for low n, but lower than α for higher n. The results also show that for some combinations of x and y distributions where the Anderson–Darling test has an appropriate rejection rate, the type-I error rate has not converged to α (e.g., when x or y are normal and the other distribution is log-normal with $\sigma = 1$).

In general, however, the results also show that the effect of the t-statistics not converging to the t-distribution is not as severe as one might expect. For example, even in the worst case where both x and y are log-normal with $\sigma = 2$, the type-I error rate is only approximately 0.06 when $n = 30$, that is, only 0.01 higher than α. While the effect is larger on the platykurtic distributions, the type-I error rate for these are *lower* than α, which is less problematic than if the type-I error rates are higher than α since this simply constitutes a more conservative test.

5.2 | The multivariate case

To assess the effect of having a second independent variable in the regression model I also ran simulations where z was added to the regression model. z was generated under different distributions to assess whether the skewness and kurtosis of z affected the convergence of the t-statistics associated with $\widehat{\beta}_x$ to the t-distribution. The results of this evaluation, presented in Appendix S1, shows that the distribution of z did not have any substantial effect on the convergence of the t-statistics associated with $\widehat{\beta}_x$ regardless of the distribution of z. This suggests that the distribution of additional independent variables,

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

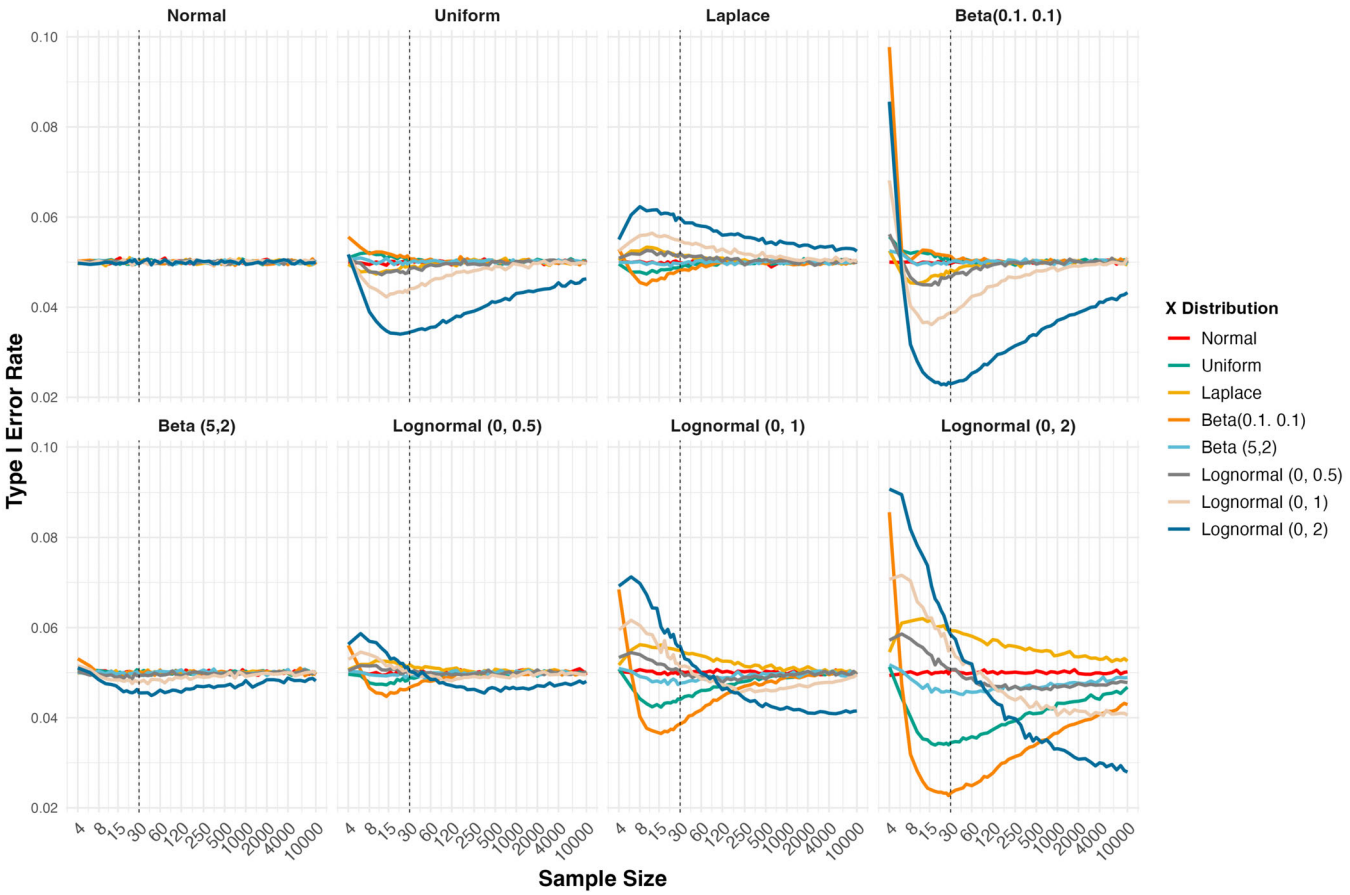

FIGURE 2 Type I errors for β_1 at $\alpha = 0.05$ for different combinations of x and y distributions and n. Facets represent different distributions of y. [Colour figure can be viewed at wileyonlinelibrary.com]

under the null hypothesis, does not affect the convergence of the t-statistics associated with $\widehat{\beta}_x$ to the t-distribution and thus that the results from the simulation in the bivariate case can be generalized to the multivariate case.

6 | DISCUSSION

The results of this article suggest that it is time to update what we teach our students regarding the lower limit of the number of observations. Specifically, the assumption that the residuals follow a normal distribution *or* that the number of observations is sufficiently large can actually be relaxed to that the distribution of either of the variables involved is **symmetric** *or* that the number of observations is sufficiently large, unless either of the variables exhibits extreme skewnewss. In cases where neither of the variables is symmetric, the results show that the number of observations needed largely depend on the skewness of the variables involved. Furthermore, the results also show that the impact of t-statistics not converging to the t-distribution is not as severe as one might expect. The type-I error rate typically deviates only slightly, either above or below, from the nominal level α in most cases. Overall, this suggests that unless either or both variables exhibit a high degree of skewness (e.g., a log-normal distribution), the $n \geq 30$ rule might be overly conservative and suggests that reliable inference is often possible with smaller sample sizes.

What are the implications of these findings for those who use and teach regression analysis? The findings imply that it may be more defensible than previously thought to run regression analysis with small samples, provided that the variables involved are not strongly skewed. This means that in situations where researchers or students may have preferred to run less powerful non-parametric tests, such as the Mann–Whitney U test, due to few observations and non-normality, it may instead be advisable to run a regression analysis if the variables are symmetric or at least not too skewed. At the other end of the spectrum, the results also suggest that the $n \geq 30$ rule of thumb may be too permissive in some cases, such as when the variables are highly skewed.

The results also suggest that the kurtosis of the variables involved affects how violations of the normality

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assumption impact inference. In particular, if either of the distributions of the variables is platykurtic (i.e., has a lower kurtosis than the standard normal distribution), the type-I error rate is lower than α even when the t-statistics do not converge to the t-distribution. This is less problematic than if the type-I error rate were higher than α, since it simply constitutes a more conservative test.

Taken together, these results suggest that if either of the distributions of the variables is symmetric, sample sizes of 20 or even 10 observations may be sufficient to trust the inferences from regression analysis. Further, if false rejection of the null hypothesis is the primary concern (the most common case in standard null-hypothesis testing), we can also trust the inferences from regression analysis with small sample sizes when either of the variables is platykurtic, as the type-I error rate, when biased, is lower than α. Apart from the cases where either the dependent or independent variables are symmetric or platykurtic, the results suggest that the $n \geq 30$ rule of thumb holds for most combinations of distributions, except when the distributions of either or both of the variables are highly skewed.

From a teaching perspective, these results also hold important lessons. When teaching quantitative research methods and regression to undergraduate students and others at the beginning of their methodological careers, we often resort to simple rules of thumb, such as requiring $n \geq 30$, to reassure students that classical inference methods are reliable. While the importance of such rules has declined in recent years, as data have become increasingly ubiquitous across almost all applied fields and very small-n studies exceedingly rare, they still retain practical value as accessible heuristics, particularly for students who are attempting to collect their own data.

However, the findings suggest that it is important to frame these guidelines as starting points rather than rigid thresholds. In particular, students should be encouraged to move beyond rote application of rules of thumb and instead learn to evaluate the distributional characteristics of their data—especially skewness and kurtosis—as they develop more advanced quantitative skills. The results of this paper could serve as a starting point for broader discussions about the conditions under which classical methods yield valid inferences, how violations of regression assumptions affect those inferences, and what it practically means when we say that inferences are incorrect—namely, that the actual type-I error rate deviates from the nominal α level. By encouraging students to think diagnostically about the assumptions underlying their analysis, rather than relying solely on fixed sample size rules, we can help them develop a more robust and adaptable understanding of statistical inference that better prepares them for real-world data challenges.

A key takeaway from this study is that reporting skewness and kurtosis *can* be helpful when running regression analysis, but this is not a blanket recommendation. These distributional summaries are particularly informative when the sample size is low or when there is reason to suspect that the variables may be far from symmetric. In typical instructional settings or applied work with moderate-to-large samples, such diagnostics may not be necessary. But in small-n contexts, they provide a rough yet useful gauge for whether t-distribution-based inference can be trusted.

Importantly, however, the fact that regression may yield valid inferences in smaller samples under certain conditions does not imply that it is always advisable to use regression under these circumstances, or that small sample sizes are necessarily sufficient to achieve the desired research goals. Consideration of statistical power remains essential and should always be weighed carefully when conducting regression analysis with small samples. The goal of this article is not to encourage underpowered analysis, but to show that the inferential assumptions we often treat as brittle are, in fact, more forgiving than traditionally taught, especially when variables are symmetric or platykurtic.

Future research should aim to better characterize the relationship between skewness and convergence to the t-distribution in regression, potentially developing more nuanced guidelines for the lower sample size limits. Further simulation studies could also explore how these distributional characteristics affect the power of regression analysis when the null hypothesis is in fact not true, and whether the same conclusions hold in cases where there are multiple, correlated independent variables with different distributions.

ACKNOWLEDGMENTS

The research was funded by the European Research Council, project H2020-ERC-2015-AdG 694640 (ViEWS) and Riksbankens Jubileumsfond, grant M21-0002 (Societies at Risk). The computations were conducted on the Dardel PDC using resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS).

ORCID

David Randahl https://orcid.org/0000-0003-1069-6067

REFERENCES

T. M. Carsey and J. J. Harden, *Monte Carlo simulation and resampling methods for social science*, SAGE Publications, Thousand Oaks, CA, 2013.

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

J. Cohen, *Things I have learned (so far)*, Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, Amer. Psychol. Assoc, 1992.

A. Gelman and J. Carlin, *Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors*, Perspect. Psychol. Sci. **9** (2014), no. 6, 641–651.

A. Gelman, J. Hill, and A. Vehtari, *Regression and other stories*, Cambridge University Press, Cambridge MA, 2021.

S. B. Green, *How many subjects does it take to do a regression analysis*, Multivar. Behav. Res. **26** (1991), no. 3, 499–510.

T. Hesterberg, *It's time to retire the n 30 rule*, Proc. Joint Statistical Meetings, 2008.

R. V. Hogg, E. A. Tanis, and D. Zimmerman, *Probability and statistical inference*, 9th ed., Pearson, Upper Saddle River, NJ, 2013.

D. Kiernan, *Natural resources biometrics*, Open SUNY Textbooks, Milne Library, New York, 2014.

S. G. Kwak and J. H. Kim, *Central limit theorem: The cornerstone of modern statistics*, Korean J. Anesthesiol. **70** (2017), no. 2, 144–156.

S. M. Ross, *Introduction to statistics*, 4th ed., Academic Press, Elsevier, 2017.

J. B. Thompson, *Approximating infinity: An investigation of the $n = 30$ rule of thumb*, Master's thesis, California State University, Long Beach, 2023.

J. Wooldridge, *Introductory econometrics: A modern approach*, 5th ed., CENGAGE Learning Custom Publishing, Mason, OH, 2012.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: D. Randahl, *This is not normal! (Re-) Evaluating the lower n guidelines for regression analysis*, Teach. Stat. **48** (2026), 130–137, DOI 10.1111/test.70012

14679639, 2026, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/test.70012 by Uppsala University Library, Wiley Online Library on [29/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License