Outbreak, Surveillance, Investigation & Response (OSIR) Journal
Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, https://he02.tci-thaijo.org/index.php/OSIR

The Grammar of Science: How "Robust" Are Your Study Results?

Jaranit Kaewkungwal*

Mahidol University, Thailand

*Corresponding author email: jaranitk@biophics.org

Received: 14 May 2024; Revised: 5 Jun 2025; Accepted: 9 Jun 2025
https://doi.org/10.59096/osir.v18i2.275906

All statistical estimates have some degree of uncertainty due to sampling variability. The process of statistical modelling and interpretation typically requires implicit assumptions about random sampling and data distribution.[1–3] But as we know in real life that data were quite often deviated from these model assumptions.

The standard error (SE) quantifies the uncertainty around a sample estimate.[4] When the underlying assumptions are violated, the calculated SE may be incorrect. Instead of using classic SE, some researchers may decide to use "robust" SE which are robust to violations of certain assumptions.[2–4] You may run into clinical and epidemiological papers that used robust SEs. Robust options can be applied in various statistical context including: estimating descriptive statistics (e.g., mean, proportion), hypothesis testing (e.g., t-test, ANOVA), regression model fitting (i.e., linear, logistic, Poisson, Cox), and repeated measures or clustered data analysis (e.g., generalized estimating equation, multilevel mixed model).

This paper examines the concept of robust standard errors—what they are, how they are calculated, and the reasons for using or avoiding them.

Definition of "Robust"

The term "robust" in statistics refers to the resilience of an estimator or statistical model under conditions that deviate from ideal assumptions.[5] A robust model maintains its accuracy and reliability even when assumptions are only partially met while the results can still yield meaningful insights despite such imperfections.[6] In essence, a robust statistic resists provides trustworthy results even in less-than-ideal analytical conditions.[5]

Certain statistical methods are considered robust under specific conditions. For example, t-test and ANOVA assume normally distributed data; however, they still perform reliably when this assumption is violated—so long as each group includes a sufficiently large sample size. This robustness is supported by the central limit theorem, which assumes the statistics remain unbiased in a wide variety of probability distributions.[2,5] Similarly, nonparametric methods are robust by nature, offering resistance to both distributional deviations and the presence of outliers.[2,3] In regression analysis, issues like outliers and heteroscedasticity (non-constant variance of residuals) can undermine model validity. (Heteroscedasticity can be detected through formal tests such as the Breusch-Pagan test or simple residual plots as illustrated in Figure 1.[7,8] Robust regression techniques, however, are designed to accommodate such violations, offering more dependable results when standard assumptions fail.

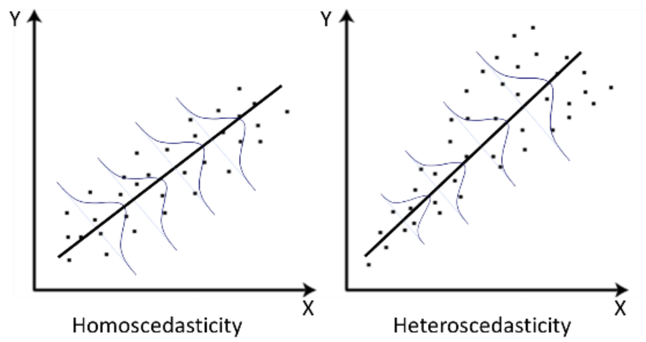


Figure 1. Homoscedasticity vs. heteroscedasticity in regression residuals

From a data-analytic perspective, robust statistics represent an extension of traditional parametric methods. These techniques acknowledge that statistical models are, at best, approximations of reality.[6] Rather than requiring the stochastic (random) component of a model to be precisely specified, robust procedures aim to capture the main structure of the data while flagging anomalous points or substructures for further investigation. In case of the dataset containing outliers, the goal of the analysis is not to eliminate outliers, but to model the majority of the data effectively.[1,5]

Review of Standard Error

SE is a fundamental concept of inferential statistics, measuring how accurately a sample statistic represents the corresponding population parameter. To review the concept of SE, let's go back to some basic statistics—SE of mean (SEM). Mean ($\bar{X}$) is a measure of central tendency that represents the average value within a dataset. The standard deviation (SD) quantifies the spread or variability of data points around the mean. While SD describes the dispersion of individual data points within a sample, SE measures the precision of the sample mean—or other statistics—relative to the true population value.[9,10] Suppose the average value in population is known (μ) and we collect a sample data drawn from that population and calculate the sample mean ($\bar{X}$) and its standard deviation (SD). In theory, if we repeat sampling the data from the same population, we will obtain several sample means and SDs, so-called sampling distribution. The average of sample means ($\bar{\bar{X}}$) is approximately the population mean (μ) and the distribution of sample means is the SE. (Figure 2) In practice, SEM can be estimated by dividing the sample's standard deviation (s) by the square root of the sample size (n). This value reflects how much the sample mean is expected to vary from the true population mean.

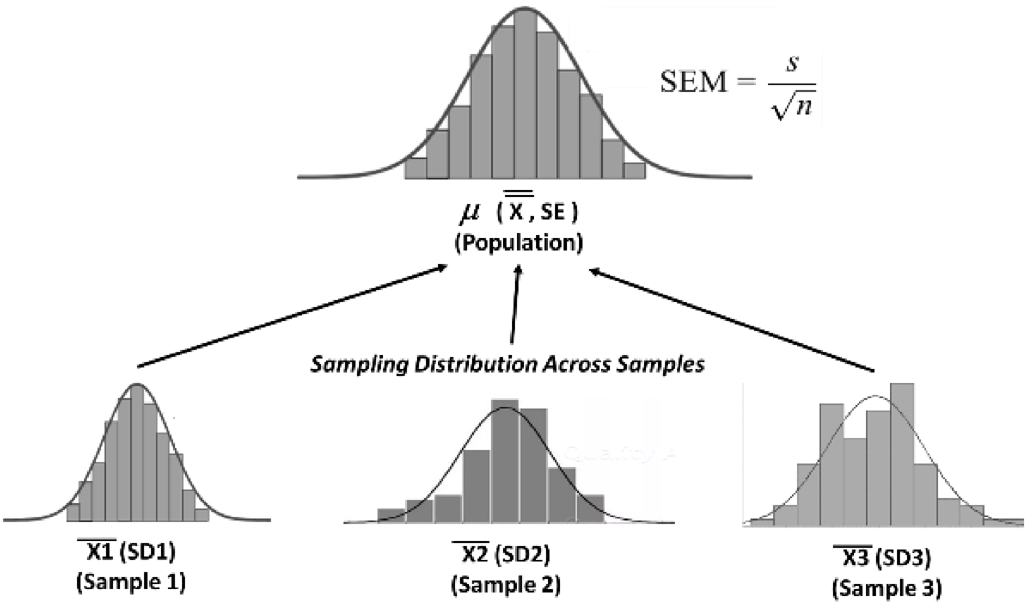


Figure 2. Sampling distribution and SEM

SE can be calculated not just for means, but for a wide variety of statistics and models. Here are some common examples:[9–12]

Simple SE for the Proportion

Calculated from the sample proportion (p) and the sample size (n). It is used to estimate the variability of a sample proportion from the true population proportion.

$$SE_p = \sqrt{\frac{p(1-p)}{n}}$$

SE of the Difference between Means

Used in comparing two means in independent t-test is calculated from the SE of the two samples being compared.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

SE of the Regression Coefficient

Derived from variance-covariance matrix of the estimated coefficient in a regression model (β). It is used to test the significance of the estimate of the regression coefficient.

$$SE(\hat{\beta}) = \sqrt{Var(\hat{\beta})}$$

SE of the Regression Estimate

Calculated from the deviation of the observed values and predicted values divided by the sample size (n) and the numbers of predictors in the regression model (k). It Indicates how well the regression model fits the data.

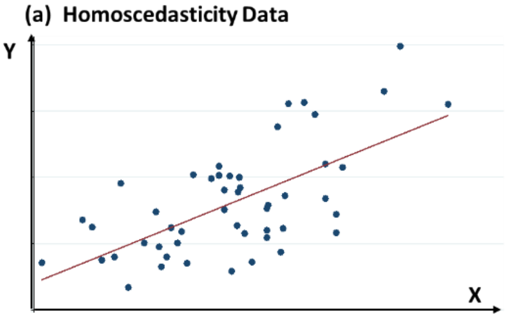
$$SE = \sqrt{\frac{\sum(y_i - \hat{y}_i{}^2}{n-k-1}}$$

In general, SE quantifies the variability of an estimated statistic around the true, but often unknown, population parameter.[13] Since we rarely know the actual population value, we rely on sample statistics to make educated guesses. These estimates are usually reported with confidence intervals (CIs) that incorporate the SE. A common approach is to construct a 95% CI using the formula: Estimate ±1.96 × SE. A narrow CI suggests high precision and greater confidence in the estimate, while a wide CI may indicate insufficient data or poor sampling methods.[1,10,13] In addition to describing precision, SE plays an essential role in hypothesis testing and statistical modeling, as outlined above.

Robust Standard Errors

You can see that SE calculations rely on the assumption that the sample is both random and representative. When a sample is biased, collected improperly, or too small, the SE might be underestimate or not accurately capture the true level of uncertainty. Thus, it will distort confidence intervals and lead to incorrect conclusions in hypothesis testing. SE calculations also assume that the underlying data follows a specific distribution. If the data is skewed or contains outliers, the SE may not reliably reflect the actual variability in the estimate.[1,13] On the other hand, robust SEs can still yield meaningful insights even when data don't perfectly meet ideal assumptions. They tend to hold up under varied distributions and can accommodate atypical values, making them a practical choice for analyzing real-world data.[2]

As an example, in a linear regression analysis of homoscedasticity and heteroscedasticity datasets running by Stata 14 with and without the robust SE option (Figure 3). While both methods produce identical coefficient estimates, the standard errors differ. This variation affects the width of the 95% confidence intervals and the *p*-values. In heteroscedasticity data, applying robust SEs leads to wider confidence intervals and higher *p*-values that are not statistically significant (0.061). Without the robust option, the *p*-values may appear close to marginal significance (0.048), potentially giving a misleading impression of the results. However, it's important to recognize that using the robust option doesn't always produce different outcomes or ensure the "correct" conclusion. When discrepancies do appear, we should further explore the data and assess model fit indicators to better understand the source of the variation.

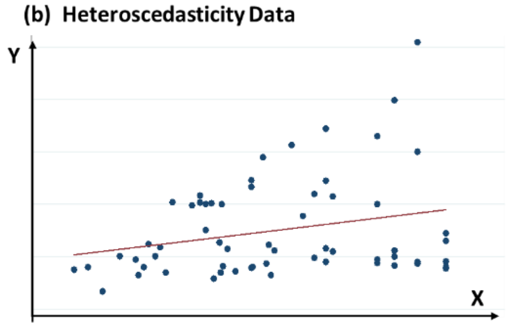


Regression Model with SE

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	65.10005	11.28348	5.77	0.000	42.40061	87.79949
_cons	-2.830365	29.89827	-0.09	0.925	-62.97792	57.31719

Regression Model with Robust SE

Y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	65.10005	11.8156	5.51	0.000	41.33013	88.86997
_cons	-2.830365	27.66772	-0.10	0.919	-58.49063	52.8299



Regression Model with SE

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	26.3917	13.06357	2.02	0.048	.2515674	52.53184
_cons	69.71653	41.86503	1.67	0.101	-14.05521	153.4883

Regression Model with Robust SE

Y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	26.3917	13.84254	1.91	0.061	-1.307165	54.09057
_cons	69.71653	36.49338	1.91	0.061	-3.306562	142.7396

Figure 3. Linear regression models with and without robust options

Types of Robust SE in Statistical Modelling

Robust SEs are often referred to as heteroscedasticity-consistent standard errors when used to address violations of the homoscedasticity assumption.[14] Beyond addressing heteroscedasticity, robust SEs can also help correct for certain forms of model misspecification in regression analysis. Various techniques have been developed to adjust standard errors, as suggested in the literature. Below are some commonly used methods:

Robust SE Options

The methods account for varying variances of the residuals in the model instead of assuming homoscedasticity. Two common methods are: (1)

Huber-White Sandwich Estimator (i.e., also known as sandwich estimator, adjusted for the covariance matrix and some forms of model misspecification) and (2) Newey-West Standard Errors (i.e., robust SE adjusted for both heteroscedasticity and autocorrelation in time-series data).[14–16]

Clustered SE

This method is appropriate when observations are collected within clusters (e.g., students within schools, patients within hospitals, or people within regions) or from the same subjects over time periods. Such data are, in theory, correlated; thus, the adjustment allows for intra-cluster correlation while assuming independence between clusters.[17]

Weighted Least Squares (WLS)

This technique adjusts by weighting each observation inversely to its error variance, thereby reducing the bias of heteroscedasticity.[18]

Bootstrapped SE

This is a resampling technique that is often used when the data do not meet the assumptions of traditional statistical methods (e.g., normality, homoscedasticity). By repeatedly resampling the data with replacement, the SE is derived from the variability in the estimates across the "bootstrapped samples".[19]

Finite Sample Adjustments

This method is commonly used with small sample sizes, so-called "small-sample correction" technique. It adjusts the covariance matrix of regression models.[20]

Generalized Least Squares (GLS) and Feasible GLS (FGLS)

The methods adjust for both heteroscedastic or autocorrelated among residuals in regression model. GLS accounts for the known residual structure while FGLS is used when the exact covariance structure is unknown.[21]

Delta Method

The method is approximate the standard error of a nonlinear transformation of estimated coefficients. It's useful in cases involving ratios, exponentials, or other nonlinear functions.[22]

Sampling-Weighted SE

This method is common in survey analysis. It is a design-based SE, accounted for sampling weights, levels of sampling strata, and clustering of observations.[23–24]

So—When to Use Robust SE?

There are various methods available for calculating robust SEs, each designed to handle different issues such as non-constant variance, outliers, autocorrelation, and other model irregularities. In the linear regression example discussed earlier, we saw how extreme data values combined with heteroscedasticity can influence statistical significance and ultimately affect study conclusions. This often raises the question: which model is truly the "best fit" or even "correct"—with or without the robust SE option? So–when should we use robust SEs?

There are diagnostic tools available for evaluating regression models, though a detailed discussion of these methods is beyond the scope of this paper. For interested readers, please refer to Zellner's papers and other references.[8,21,25,26] One key indicator that robust SE may be appropriate is the presence of large residuals or high-leverage points. We can say that if your model is approximately correct, conventional SEs are generally sufficient, and using robust SE is unlikely to add much value. However, if the model is seriously in error, robust adjustments may improve the estimation of variance, but the parameters being estimated are still controversial and require caution in interpretation. Even with robust SEs, the model might be overfitting or underfitting the data, especially when assumptions are clearly violated.[27,28]

Robust SE should not be used as a superficial safeguard against reviewer criticism or assumed to correct all problems.[29] Simply choosing to report only classical or only robust SEs can be misleading. We should take a closer look when the classical SE and robust SE differ substantially. A larger discrepancy between the two types of SEs means that you are potentially have a misspecification model.[29]

A Word of Caution

Robust statistics are not a replacement for classical methods.[6] Misunderstanding this can lead to misuse. While robust SEs are valuable for addressing certain issues in traditional models, their use should be justified, not automatic. You should carefully consider whether it is necessary or not.

Choosing the right approach to SEs depends not only on your data and sampling method but also on your broader research goals—how you intend to generalize the results across units and over time, and how you will use the model estimates. You should pick a method to adjust or not adjust your SE when you carefully consider your model's purpose and how the study would be replicated.[10,30] As emphasized in the literature, robust SEs do not substitute for careful model specification.[14,29] Use all available diagnostic tools, ensure your model fits the data well, test its predictions, and refine it based on those insights. When a model is well-specified, both classical and robust SEs are expected to converge.[29]

Acknowledgements

An AI tool, ChatGPT (OpenAI, 2025), was used to search for statistical formulas and references related to the topic and generate language suggestions during the preparation of this manuscript, and all outputs were reviewed for accuracy and appropriateness.[31] The author reviewed, edited, and took responsibility for the final content.

Suggested Citation

Kaewkungwal J. The grammar of science: how "robust" are you study results? OSIR. 2025 Jun;18(2):127–32. doi:10.59096/osir.v18i2.275906.

References

1. Analytical Methods Committee. AMC technical brief [Internet]. London: Royal Society of Chemistry; 2001 Apr [cited 2025 May 1]. p. 28–9. <https://cscjp.co.jp/fera/document/amc_technical_brief_robust_statistics_a_method_of_coping_with_outliers.pdf >
2. Frost J. What are robust statistics? [Internet]. [place unknown]: Statistics by Jim; [cited 2025 May 1]. <https://statisticsbyjim.com/basics/robust-statistics/>
3. Huber PJ, Ronchetti EM. Robust statistics. New York: Wiley; 1982.
4. Mansournia MA, Nazemipour M, Naimi AI, Collins GS, Campbell MJ. Reflections on modern methods: demystifying robust standard errors for epidemiologists. Int J Epidemiol. 2021;50(1):346–51. doi:10.1093/ije/dyaa260.
5. Taylor C. Robustness in statistics [Internet]. New York: ThoughtCo; 2023 Apr 5 [cited 2025 May 5]. 41 p. <https://www.thoughtco.com/what-is-robustness-in-statistics-3126323>
6. Ronchetti E. Introduction to robust statistics [Internet]. Geneva: University of Geneva; [cited 2025 May 1]. <https://orfe.princeton.edu/sites/g/files/toruqf1286/files/documents/ronchetti-lecture1.pdf>
7. Wooldridge JM. Introductory econometrics: a modern approach. 5th ed. Mason (OH); South-Western, Cengage Learning. 881 p.
8. Stock JH, Watson MW. Introduction to econometrics. 4th ed. London: Pearson Education Limited; c2020.
9. Moore DS, McCabe GP, Craig BA. Introduction to the practice of statistics. New York: W. H. Freeman; 2016. 725 p.
10. Altman DG, Bland JM. Standard deviations and standard errors. BMJ. 2005;331(7521):903. doi:10.1136/bmj.331.7521.903.
11. Agresti A. Statistical methods for the social sciences. London: Pearson Education Limited; 2018.
12. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. New York: McGraw-Hill Irwin; 2005.
13. Kenton W. Standard error (SE) definition: standard deviation in statistics explained [Internet]. New York: Investopedia; [updated 2025 Apr 19; cited 2025 May 1]. <https://www.investopedia.com/terms/s/standard-error.asp>
14. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica. 1980;48 (4):817–38. doi:10.2307/1912934.
15. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability; 1965 June 21–July18 and 1965 Dec 27–1966 Jan 7; Berkeley, USA. Oakland (CA): University of California Press; 1967. p. 221–34.
16. Newey WK, West KD. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica. 1987;55(3):703–8. doi:10.2307/1913610.
17. Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. Journal of Human Resources. 2015;50(2):317–72. doi:10.3368/jhr.50.2.317.
18. Greene WH. Econometric Analysis. 8th ed. Essex: Pearson Education; 2018.
19. Efron B, Tibshirani R J. An introduction to the bootstrap. Boca Raton: CRC Press; 1993.
20. Arellano M. Panel data econometrics. Oxford: Oxford University Press; 2003.
21. Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge: MIT Press; 2010.

22. Gourieroux C, Monfort A. Statistics and econometrics. Cambridge: Cambridge University Press; 1995.

23. Binder DA. On the variances of asymptotically normal estimators from complex surveys. International Statistical Review. 1983;51(3): 279–92. doi:10.2307/1402588.

24. Lohr SL. Sampling: design and analysis. 2nd ed. New York: Chapman and Hall; 2019. 610 p. doi:10.1201/9780429296284.

25. Zeileis A, Hothorn T. Diagnostic checking in regression relationships. R News [Internet]. 2002 [cited 2025 May 5];2:7–10. <https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf>

26. Zeileis A. Object-oriented computation of sandwich estimators. Journal of Statistical Software. 2006;16(9):1–16. doi:10.18637/jss.v016.i09.

27. Ford C. Understanding Robust Standard Errors [Internet]. Charlottesville: University of Virginia Library; 2020 Sep 27 [cited 2025 May 1]. <https://library.virginia.edu/data/articles/understanding-robust-standard-errors>

28. Freedman DA. On the so-called "Huber sandwich estimator" and "robust standard errors". The American Statistician. 2006;60(4): 299–302. doi:10.1198/000313006X152207.

29. King G, Roberts ME. How robust standard errors expose methodological problems they do not fix, and what to do about it. Political Analysis. 2015;23(2):159–79. doi:10.1093/pan/mpu015.

30. Gelman A. What is a standard error? [Internet]. New York: Statistical Modeling, Causal Inference, and Social Science, Columbia University; 2023 Aug 5 [cited 2025 May 1]. <https://statmodeling.stat.columbia.edu/2023/08/05/what-is-a-standard-error/>

31. OpenAI. ChatGPT [Internet]. 2025 [cited 2025 May 5]. <https://chat.openai.com>