




AssumpSure: a user-friendly R Shiny package for automated validation of statistical assumptions and appropriate test selection

Ahmed Bargheet ¹✉

¹ Host-Microbe Interactions Research Group, Department of Medical Biology, UiT The Arctic University of Norway, Tromsø, Norway ✉ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Chris Vernon](#)  

Reviewers:

- [@ZekeMarshall](#)
- [@adithirgis](#)

Submitted: 04 August 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

AssumpSure is an open-source R Shiny package that enforces mandatory assumption checks to improve statistical rigor and guide appropriate test selection. It supports a wide array of statistical tests. The app automatically evaluates key assumptions such as normality (Shapiro–Wilk) and homogeneity of variance (Levene’s test), using color-coded indicators (green for passed, red for violated) with clear interpretive summaries. When assumptions are not met, it recommends alternatives like Kruskal–Wallis for non-normal ANOVA data. *AssumpSure* also prevents invalid variable types, provides preprocessing tools (log, Box–Cox, centered log-ratio transformations), and generates downloadable diagnostic plots and publication-ready outputs.

Statement of Need

Rigorous statistical analysis depends on checking underlying assumptions and choosing appropriate tests. Yet in practice, these steps are often skipped, misunderstood, or incorrectly applied. Reviews of social science literature report that fewer than 25% of studies explicitly mention assumption checks, and even fewer report them correctly ([Shatz, 2024](#)). This contributes to widespread misuse of statistical tests and undermines reproducibility. One common issue is the misinterpretation of assumption tests. For example, many users mistakenly interpret a significant Shapiro–Wilk test result ($p < 0.05$) as evidence for normality rather than deviation from it ([Ghasemi & Zahediasl, 2012](#); [Gosselin, 2024](#); [Midway & White, 2025](#)). Moreover, in small samples, where the Shapiro–Wilk test lacks power, borderline results are easily overlooked or misjudged.

AssumpSure directly addresses these problems by making assumption checking a required part of the analysis workflow. Through an intuitive web interface, users are guided step-by-step: assumptions are tested automatically, results are color-coded (green for met, red for violated, orange for borderline), and clear recommendations are given. For example, if ANOVA assumptions are not met, Kruskal–Wallis is suggested. By complementing color-coded squares with tooltips and clear messages, *AssumpSure* helps reduce misinterpretation of common assumption tests and guides the user on how to proceed.

Comparison to other Packages

Tools like *Jamovi* ([Şahin & Aybek, 2019](#)) and *JASP* ([Love et al., 2019](#)) offer user-friendly interfaces but lack integrated workflows that enforce mandatory assumption diagnostics, appropriate test selection, and actionable recommendations, leaving users to manually interpret

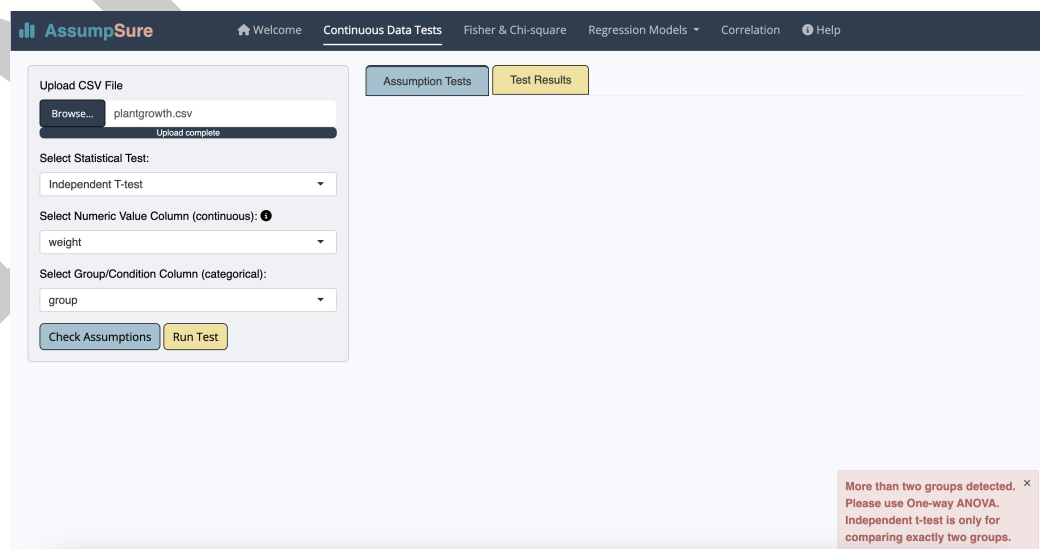
39 complex outputs. This increases errors and undermines reproducibility. Additionally, non-R
40 applications such as assumption-sheriff (Python) (Root, 2024) also aim to simplify statistical
41 assumption checking and test selection. However, these alternatives lack fully integrated,
42 interactive diagnostic workflows or do not provide enforced guidance on test appropriateness
43 for each step.

44 Software Description

45 *AssumpSure*'s modular R Shiny architecture organizes statistical analyses into continuous
46 variable tests, categorical tests, correlations, and regression models. Users upload CSV data,
47 and the interface automatically detects variable types, preventing invalid selections with clear
48 error messages (e.g., suggesting one-way ANOVA if a t-test is chosen for a categorical variable
49 with three levels). Mandatory assumption diagnostics (normality via Shapiro-Wilk, variance
50 homogeneity via Levene's test) are visualized through diagnostic plots (QQ plots, histograms,
51 scatter plots), with tooltips to guide the user how to interpret the plots.

52 For correlations, a tooltip explains the centered log-ratio (CLR) transformation's importance,
53 conducted by compositions R package (Boogaart et al., 2013), in reducing spurious correlations
54 in compositional data (e.g., microbiome profiles). Prevalence filtering is available for correlations
55 to enhance reliability. Regression models leverage the performance package (Lüdecke et al.,
56 2021), providing transformations like log or Box-Cox that are offered to improve residual
57 normality and reduce heteroscedasticity, with guidance on their role in ensuring model validity.

58 The package supports independent/paired t-tests, Welch test, one-way ANOVA, Mann-Whitney
59 U, Wilcoxon signed-rank, Kruskal-Wallis, chi-square, Fisher's exact, linear, logistic, multinomial,
60 Poisson, negative binomial, zero-inflated negative binomial, and mixed-effects models, plus
61 Pearson, Spearman, Kendall, and biweight midcorrelation methods. Post-hoc comparisons
62 (e.g., Tukey HSD, Dunn's test) and multiple testing corrections (e.g., Benjamini-Hochberg)
63 are included where applicable. Outputs, including p-values, confidence intervals, effect sizes,
64 and publication-ready plots, are exportable in PNG, PDF, or CSV formats for transparent
65 reporting as suggested by researchers (Shatz, 2024). *AssumpSure* integrates R libraries like
66 *rstatix* (Kassambara, 2019), *performance* (Lüdecke et al., 2021), *lmerTest* (Kuznetsova et al.,
67 2015), *MASS* (Ripley et al., 2013), and *bestNormalize* (Peterson & Peterson, 2020) for robust
68 functionality, with all used packages listed in the DESCRIPTION file.



69
70 Figure 1: The workflow from data upload to test selection, guiding users to switch to the
71 appropriate statistical test.

72

Illustrative Example

73 A researcher uploads a CSV dataset, imported using the readr package (Wickham et al.,
74 2024) with bacterial alpha diversity (Shannon index) across three countries. This example
75 uses the infants.csv dataset included with AssumpSure. AssumpSure detects variable types,
76 restricting invalid selections (e.g., excluding count variables for ANOVA). Selecting one-way
77 ANOVA triggers normality (Shapiro-Wilk) and homogeneity (Levene's) checks using the rstatix
78 package (Kassambara, 2019), visualized via QQ plots and histograms with ggplot2 (Wickham,
79 2011) with tooltips implemented using the shiny package (Chang et al., 2015) (which provides
80 the icon() function and HTML integration for Bootstrap tooltips). A red indicator signals
81 non-normality and heterogeneous variances, recommending Kruskal-Wallis (Figure 2).

Assumption Test Results

Levene's Test (Equal Variance)

df1	df2	statistic	p
2	247	32.29271	0

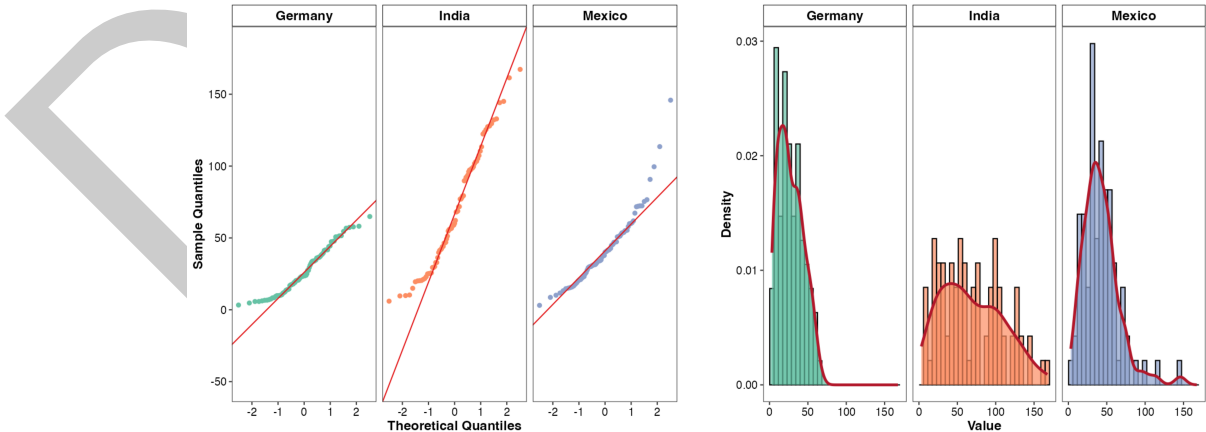
Variances are significantly different (Heterogeneous).

Shapiro-Wilk Test (Normality)

group	variable	statistic	p
Germany	value	0.9515254	0.0031726
India	value	0.9589311	0.0095929
Mexico	value	0.9067978	0.0000173

One or more groups deviate significantly from normality.

Diagnostic Plots



Assumption Check Summary

✗ Assumptions Not Met
Consider a non-parametric alternative: Kruskal-Wallis test.

82
83 Figure 2: The assumption checking report.

84 Switching to this test, the user obtains significant results ($p < 0.05$) and effect size using
85 Kruskal-Wallis eta-squared (Figure 3), also based on rstatix (Kassambara, 2019).

Assumption Tests		Test Results			
.y.	n	statistic	df	p	method
value	250	59.77617	2	0	Kruskal-Wallis

Statistically Significant Difference				
Effect size (Kruskal-Wallis eta-squared ($\eta^2[H]$))				
.y.	n	effsize	method	magnitude
value	250	0.2339116	eta2[H]	large

86
87 Figure 3: Kruskal-Wallis test and effect sizes estimation using Kruskal-Wallis eta-squared.

88 The user can run Dunn's with p-value correction, such as Benjamini-Hochberg, based on
89 rstatix R package (Kassambara, 2019), and download a publication-ready boxplot (Figure 4).
90 This workflow, requiring no coding, ensures valid test selection and transparent reporting for
91 researchers with minimal statistical expertise.

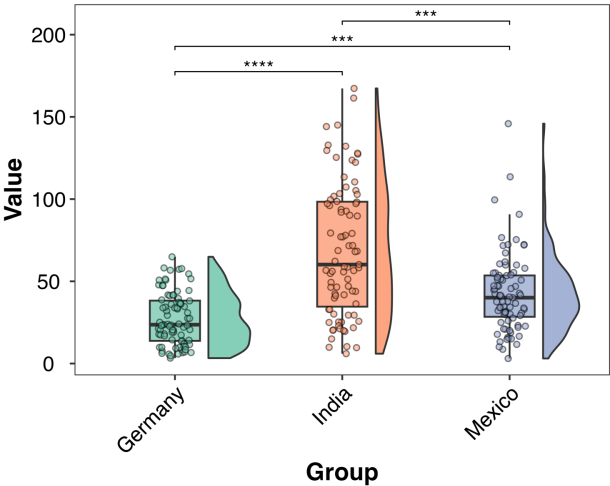
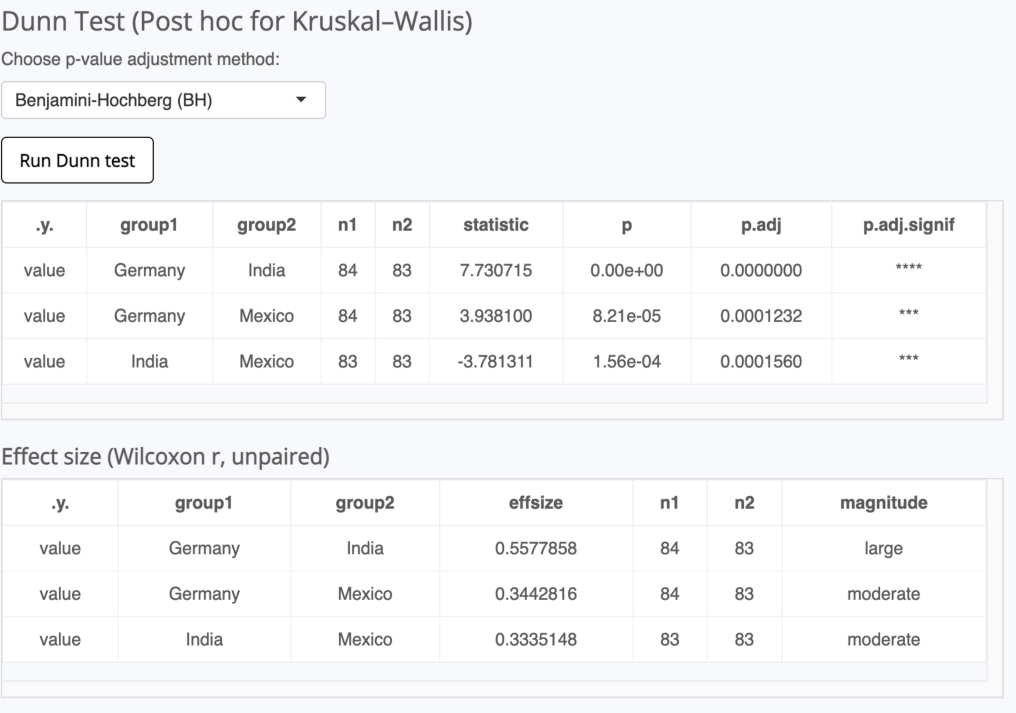


Figure 4: Dunn test and Wilcoxon r unpaired test for effect size, and publication-ready boxplot.

Example Usage and Reproducibility

AssumpSure provides example datasets for all modules. infants.csv is downloadable within the app (Help tab), and the remaining datasets are available in the GitHub repository (inst/extdata/) for manual testing.

- Continuous data tests: t-test.csv, plantgrowth.csv, and infants.csv (for independent/paired t-tests, one-way ANOVA, and other continuous analyses).
- Fisher and Chi-square tests: chi_fisher.csv and infants.csv for categorical comparisons.
- Regression models: infants.csv for linear, logistic, and other regression analyses.
- Correlation analyses: correlation.csv and bacteria_for_correlation.csv for standard and compositional correlation methods.

Limitations

AssumpSure accepts UTF-8, comma-delimited CSV files with a header row (parsed by *readr* package (Wickham et al., 2024)). Unusual delimiters or encodings are not supported. There is no strict size limit, but large files may slow performance. *AssumpSure* visualizes assumption checks using diagnostic plots (e.g., from the *performance* package (Lüdtke et al., 2021)) and provides general guidance via tooltips, but does not generate automated, context-specific recommendations. Users should interpret plots with the provided guidance.

Licensing and Availability

AssumpSure is licensed under the GPL-3 license, with all source code stored at GitHub (<https://github.com/Ahmedbargheet/AssumpSure>). In the spirit of honest and open science, we encourage requests, tips for fixes, feature updates, as well as general questions and concerns via direct interaction with the developer.

Acknowledgments

AssumpSure acknowledges the open-source R community for maintaining the packages that support the application's statistical and graphical functionality.

References

- Boogaart, K. G. van den, Tolosana, R., Bren, M., & Boogaart, M. K. G. van den. (2013). Package "compositions." *Compositional Data Analysis Ver, 1*, 40–41. <https://doi.org/10.32614/CRAN.package.compositions>
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). Package "shiny." See <http://CiteSeerX.Ist.Psu.Edu/Viewdoc/Download>. <https://doi.org/10.32614/CRAN.package.shiny>
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486. <https://doi.org/10.5812/ijem.3505>
- Gosselin, R.-D. (2024). Testing for normality: A user's (cautionary) guide. *Laboratory Animals*, 58(5), 433–437. <https://doi.org/10.1177/00236772241276808>
- Kassambara, A. (2019). Rstatix: Pipe-friendly framework for basic statistical tests. *CRAN: Contributed Packages*. <https://doi.org/10.32614/CRAN.package.rstatix>
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & others. (2015). Package "lmerTest." *R Package Version*, 2(0), 734. <https://doi.org/10.32614/CRAN.package.lmerTest>
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., & others. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, 88, 1–17. <https://doi.org/10.18637/jss.v088.i02>
- Lüdtke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60). <https://doi.org/10.21105/joss.03139>
- Midway, S., & White, J. W. (2025). Testing for normality in regression models: Mistakes abound (but may not matter). *Royal Society Open Science*, 12(4), 241904. <https://doi.org/10.1098/rsos.241904>

- 145 Peterson, R. A., & Peterson, M. R. A. (2020). Package “bestNormalize.” *Normalizing Trans-*
146 *formation Functions. R Package Version, 1.* [https://doi.org/10.32614/CRAN.package.](https://doi.org/10.32614/CRAN.package.bestNormalize)
147 [bestNormalize](https://doi.org/10.32614/CRAN.package.bestNormalize)
- 148 Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B.
149 (2013). Package “mass.” *Cran r*, 538(113-120), 822. [https://doi.org/10.32614/CRAN.](https://doi.org/10.32614/CRAN.package.MASS)
150 [package.MASS](https://doi.org/10.32614/CRAN.package.MASS)
- 151 Root, B. (2024). Assumption-sheriff: Automated assumption checking for statistical tests. In
152 *Python Package Index. PyPI.* <https://pypi.org/project/assumption-sheriff/>
- 153 Şahin, M., & Aybek, E. (2019). Jamovi: An easy to use statistical software for the social
154 scientists. *International Journal of Assessment Tools in Education*, 6(4), 670–692. <https://doi.org/10.21449/ijate.661803>
155
- 156 Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualiza-
157 tion and effect size in statistical diagnostics. *Behavior Research Methods*, 56(2), 826–845.
158 <https://doi.org/10.3758/s13428-023-02072-x>
- 159 Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2),
160 180–185. <https://doi.org/10.1002/wics.147>
- 161 Wickham, H., Hester, J., Francois, R., Bryan, J., Bearrows, S., Jylänki, J., & Jørgensen, M.
162 (2024). Package “readr.” *Read Rectangular Text Data. Available Online: Hhttps://Cran.*
163 *R-Project. Org/Web/Packages/Readr/Readr. Pdf (Accessed on 23 August 2023).* <https://doi.org/10.32614/CRAN.package.readr>
164