

¹ biostats: Biostatistics and Clinical Data Analysis in R

² **Sebastian Quirarte-Justo**  ¹¶, **Angela Carolina Montaño-Ruiz**  ¹, and **José M. Torres-Arellano** 

⁴ **1** Laboratorios Sophia, S.A. de C.V., Jalisco, Mexico. ¶ Corresponding author

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

Software

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

Editor: ↗

Submitted: 26 January 2026

Published: unpublished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](#))

⁵ Summary



⁶ **biostats** is an R package ([R Core Team, 2024](#)) that provides a cohesive and structured set of tools for biostatistics and clinical data analysis. The package includes 14 specialized functions covering descriptive statistics, exploratory data analysis, sample size and power calculations, statistical analysis and inference, and data visualization. These functions aim to offer standardized, well-documented workflows that are frequently required in clinical studies, trial planning, and analysis. By consolidating these capabilities into a single framework, the package facilitates consistent, transparent, and reproducible analyses across studies.

¹⁴ This package serves both as an analytical toolkit for professional biostatisticians and clinical data analysts, and as an educational resource for researchers transitioning to R-based biostatistics, including professionals from other domains, clinical researchers, and medical practitioners involved in the development of clinical trials.

¹⁸ **biostats** is available on the Comprehensive R Archive Network (CRAN) and adheres to CRAN standards for documentation, testing, reproducibility, and long-term maintainability within the R ecosystem.

Descriptive Statistics and Exploratory Data Analysis

`clinical_data()`, `summary_table()`, `normality()`, `missing_values()`, `outliers()`

Sample Size and Power Calculation

`sample_size()`, `sample_size_range()`

Statistical Analysis and Inference

`omnibus()`, `effect_measures()`

Data Visualization

`plot_bar()`, `plot_line()`, `plot_hist()`, `plot_box()`, `plot_corr()`

Figure 1: Functions included in the `biostats` package.

21 Statement of need

22 Biostatistics is a fundamental component of clinical research, essential for validating trial designs,
23 methodologies, results, conclusions, as well as supporting submission to regulatory entities
24 (Ciolino et al., 2021; Dwivedi, 2022; Sagar et al., 2023). In practice, clinical data analysis
25 involves the execution of similar tasks across multiple studies and projects. Typical workflows
26 include the calculation of descriptive statistics and exploratory data analysis, assumption
27 validation, hypothesis testing, primary, secondary, and exploratory statistical analyses, effect
28 size estimation, sample size and power calculations, as well as data visualization.

29 Popular packages in this field include Hmisc (Harrell Jr, 2026) and tableone (Yoshida & Bartel,
30 2022) for descriptive statistics, pwr (Champely, 2020) for power and sample size calculations,
31 effectsize (Ben-Shachar et al., 2020) for effect size estimates, and ggplot2 (Wickham, 2016)
32 for data visualization, among others. While these packages are well-designed and widely
33 used, completing a clinical study workflow typically requires combining multiple packages with
34 different syntax conventions , output formats, and integration patterns. As a result, analysts
35 frequently develop custom code to connect results, automate recurring tasks, or standardize
36 outputs across studies. This fragmentation can lead to inconsistent implementations, duplicated
37 effort, and increased time spent on code development, validation, and quality control.

38 The **biostats** package addresses these challenges by providing a unified, clinically oriented
39 framework that consolidates commonly used biostatistical procedures into a single, coherent
40 toolkit. While users still retain full flexibility to write custom code tailored to study-specific
41 needs, **biostats** is designed to streamline repetitive and foundational tasks in biostatistics
42 and clinical data analysis through consistent syntax, harmonized outputs, and functions that
43 reflect standard clinical workflows. Its goal is to deliver a professional-grade toolset for
44 biostatisticians and clinical researchers while remaining accessible to data analysts from other
45 fields. In addition, **biostats** serves as an educational resource for users transitioning to R or
46 to biostatistics, offering a structured and reproducible approach aligned with contemporary
47 recommendations for transparent and rigorous statistical practice.

48 State of the field

49 Regarding the specific functions in this package, **biostats** differs from existing packages
50 such as ez (Lawrence, 2016), rstatix (Kassambara, 2023), ggblanket (Hodge, 2025), ggpqr
51 (Kassambara, 2025), extras (Hill & Thorley, 2025), SampleSize4ClinicalTrials (Qi & Zhu,
52 2021), TrialSize (Vicky Qian Wu ; Shein-Chung Chow ; Harry G.Zhang, 2024), TrialSimulator
53 (Zhang, 2025) and simtrial (Anderson et al., 2025) to name a few, due to its ease of use,
54 consistent syntax, clear and professional presentation of results without unnecessary complexity
55 in interpretation, and thorough, beginner-friendly documentation.

56 The functions *sample_size()*, *sample_size_range()*, *effect_measures()* and *normality()* propose
57 a composite approach to variable evaluation. In many existing packages these analyses are
58 implemented through separate functions dependent on specific statistical tests or methods. For
59 example, normality assessment via distinct tests (e.g. Shapiro–Wilk, Kolmogorov–Smirnov),
60 kurtosis measures, or independent graphical analyses; sample size calculations through functions
61 tailored to individual study designs; and effect measures evaluated separately for each type of
62 association. In contrast, the **biostats** package unifies these analysis within single functions,
63 providing a unified, consistent, and streamlined workflow.

64 The *omnibus()* function offers an integrated approach to determining whether parametric
65 linear models or non-parametric alternatives are appropriate. It evaluates data using minimally
66 specified parameters, returns the corresponding model's analysis, reports observed values per
67 each assumption, runs appropriate post-hoc tests, and presents the results in an clear and
68 easy-to-follow format.

- 69 The *missing_values()*, *outliers()*, and *summary_table()* functions present data and analysis in
- 70 a clean and organized format with professional visual outputs, as opposed to other alternatives
- 71 that only return raw values without formatting or graphical complements.
- 72 When compared to other available options, the *clinical_data()* function offers a simple but
- 73 realistic and clean dataframe of simulated clinical data, ideal for users who want sample data
- 74 without highly specialized parameters.
- 75 The ggplot2 wrapper functions included in this package are designed to require minimal
- 76 code and parameter specifications, while quickly producing professional publication-grade
- 77 visualizations and fully retaining the flexibility to further customize ggplot2 objects.

78 Software Design

- 79 The ***biostats*** package was designed to balance analytical rigor, usability, and reproducibility in
- 80 applied biostatistics and other analytical fields where these tools could also be useful. The
- 81 structure of the package follows a unified, workflow-oriented design, where each function
- 82 performs a complete analytical step and returns clear, structured outputs that can be
- 83 implemented as input for subsequent analysis with other functions. This approach prioritizes
- 84 transparency and auditability, enabling analyses to be inspected, reproduced, and reviewed in
- 85 a stepwise manner. To support chaining, reporting, and downstream reuse, parameters and
- 86 outputs are standardized across functions.
- 87 Visualization functions return native ggplot2 objects rather than static figures. This design
- 88 enables users to quickly produce professional, publication-grade visualizations with minimal
- 89 code, while retaining full flexibility to customize aesthetics and formatting to meet specific
- 90 reporting or journal requirements without modifying internal package logic.
- 91 Overall, this package aims to emphasize clarity, consistency, and reproducibility, supporting
- 92 both analytical workflows and educational use by researchers and professionals transitioning to
- 93 R-based biostatistics and clinical data analysis.

94 Research Impact Statement

- 95 The ***biostats*** package has been released on CRAN (current version: 1.1.1), ensuring standardized
- 96 installation, long-term availability, and seamless integration within the R ecosystem. It is also
- 97 publicly available and maintained on GitHub, where it is accompanied by reproducible examples,
- 98 detailed documentation, and an active issue tracker. Updates have been implemented based on
- 99 the authors' real-world use, as well as user feedback, supporting transparency, reproducibility,
- 100 and community-driven improvement.
- 101 Since its release, the software has demonstrated early but meaningful adoption within the
- 102 biostatistics and broader data analysis communities, reflected by package downloads, GitHub
- 103 stars, active engagement through comments, shares, and reactions across professional social
- 104 media platforms. In addition, the authors have received positive feedback and feature
- 105 suggestions from users across multiple disciplines, including data science, clinical research,
- 106 healthcare, and applied statistics, indicating relevance beyond a single application domain.
- 107 The package addresses a common challenge in applied research: the fragmentation of statistical
- 108 workflows across multiple scripts and tools. By providing a unified set of functions for core
- 109 biostatistical tasks, it promotes reproducible and transparent analysis, as well as providing
- 110 thorough documentation for educational purposes.

111 Key features

112 Descriptive Statistics and Exploratory Data Analysis

113 `clinical_data()` creates a simulated clinical trial dataset with subject demographics, multiple
 114 visits, treatment groups with different effects, numerical and categorical variables, as well as
 115 optional missing data and dropout rates.

116 `summary_table()` performs descriptive statistics with normality assessment (Shapiro–Wilk or
 117 Kolmogorov–Smirnov with Lilliefors' correction), selects appropriate tests such as Welch's
 118 t-test or Mann–Whitney U for numerical variables and chi-squared or Fisher's exact tests for
 119 categorical variables, and computes effect sizes including Cohen's d, Mann–Whitney U effect
 120 size (r), odds ratios, and Cramer's V.

121 `missing_values()` visualizes missing data patterns, `outliers()` identifies extreme values using
 122 Tukey's method with customizable thresholds, and `normality()` performs an assessment
 123 of distributions with Q–Q plots, histograms, and multiple diagnostic tests based on the
 124 recommendations mentioned by Mishra et al. (2019) and methods by Lilliefors (1967) and
 125 Dallal & Wilkinson (1986).

126 Sample Size and Power Calculation

127 `sample_size()` and `sample_size_range()` are specifically focused on sample size calculation for
 128 clinical trials based on the equations in Chow et al. (2017), supporting equality, equivalence,
 129 and non-inferiority/superiority hypothesis, with parallel or crossover designs, and evaluating
 130 outcomes specified in means or proportions.

131 Statistical Analysis and Inference

132 `omnibus()` performs multi-group hypothesis testing to evaluate overall differences among three
 133 or more groups, with the theory behind this function being influenced by the works of Blanca et
 134 al. (2017) and Field et al. (2012). This function automatically conducts assumption diagnostics
 135 and selects the appropriate statistical test based on data characteristics. It supports both
 136 independent and repeated-measures designs and applies one-way ANOVA, repeated-measures
 137 ANOVA, Kruskal–Wallis test, or Friedman test as appropriate. When significant effects are
 138 detected, `omnibus()` also performs post-hoc comparisons.

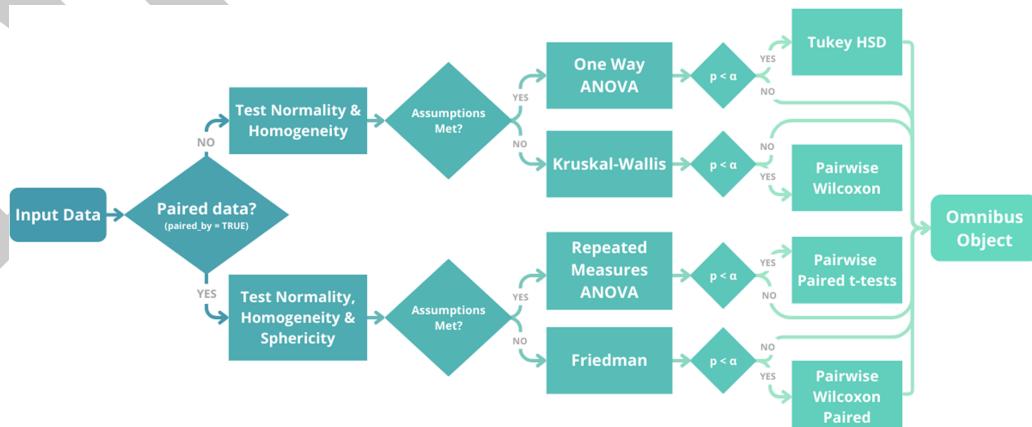


Figure 2: Omnibus function test selection algorithm for multi-group comparisons.

139 `effect_measures()` calculates effect measure indices commonly required in clinical research,
 140 including odds ratios, risk ratios, and number needed to treat or harm.

141 Data Visualization

142 The plotting functions `plot_bar()`, `plot_line()`, `plot_box()`, `plot_hist()`, and `plot_corr()`
 143 generate publication-ready visualizations tailored for clinical research. These functions display
 144 summary measures such as means, medians, standard errors, standard deviations, and 95
 145 percent confidence intervals, and they apply consistent formatting, grouping structures, and
 146 labeling to enhance interpretability. Each function returns a fully customizable ggplot2 object,
 147 allowing users to refine themes, annotations, scales, and other graphical elements.

148 Examples

```
# Simulate basic clinical data
clinical_df <- clinical_data()

head(clinical_df, 5)
#>   participant_id visit sex treatment age weight biomarker response
#> 1          001     1 Male Treatment 35  55.4    42.22 Complete
#> 2          001     2 Male Treatment 35  60.3    44.70    None
#> 3          001     3 Male Treatment 35  58.1    44.85 Partial
#> 4          002     1 Male Placebo  21  68.3    56.51    None
#> 5          002     2 Male Placebo  21  66.3    51.03    None
#> 6          002     3 Male Placebo  21  64.0    39.59    None

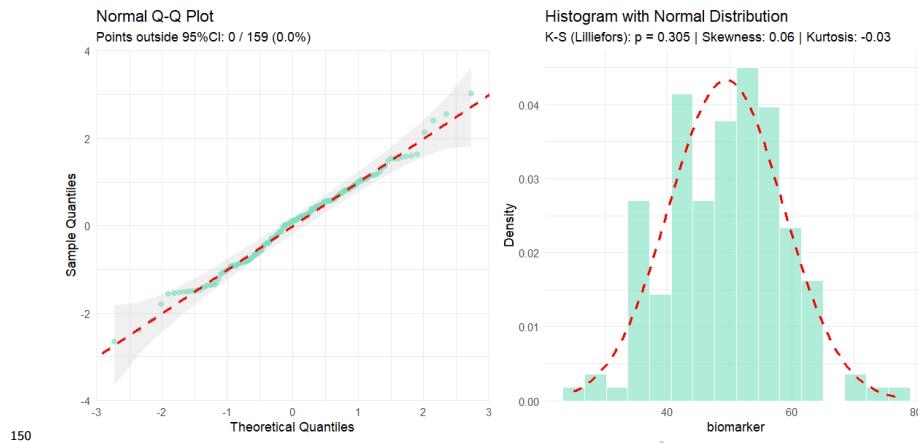
# Grouped summary by treatment group with all stats and effect size
summary_table(clinical_df, group_by = 'treatment', all = TRUE,
               effect_size = TRUE, exclude = c('participant_id', 'visit'))
```

variable	n	Placebo (Group A)	Treatment (Group B)	normality	test	p_value	effect_size	effect_param
sex	A:159, B:141	Male: 93 (58.5%); Female: 66 (41.5%)	Male: 72 (51.1%); Female: 69 (48.9%)	NA	Chi-squared	0.197	0.07	Cramer's V
age	A:159, B:141	Mean (SD): 43.68 (14.1); Median (IQR): 45.00 (35.0,53.0); Range: 18.00,70.00	Mean (SD): 42.09 (12.8); Median (IQR): 42.00 (33.0,53.0); Range: 18.00,66.00	A:<0.001, B:0.005	Mann-Whitney U	0.252	0.91	r
weight	A:159, B:141	Mean (SD): 72.56 (12.9); Median (IQR): 69.20 (62.5,83.6); Range: 50.60,100.30	Mean (SD): 71.79 (14.1); Median (IQR): 74.60 (59.8,81.9); Range: 45.00,102.20	A:<0.001, B:0.008	Mann-Whitney U	0.678	0.96	r
biomarker	A:159, B:141	Mean (SD): 49.44 (9.2); Median (IQR): 50.38 (42.2,55.3); Range: 24.92,77.26	Mean (SD): 46.71 (10.2); Median (IQR): 46.70 (40.4,52.9); Range: 16.64,71.46	A:0.546, B:0.832	Welch's t-test	0.016	0.28	Cohen's d
response	A:159, B:141	Complete: 31 (19.5%); Partial: 38 (23.9%); None: 90 (56.6%)	Complete: 59 (41.8%); Partial: 25 (17.7%); None: 57 (40.4%)	NA	Chi-squared	<0.001	0.24	Cramer's V

149

```
# Filter clinical data to Placebo arm
clinical_df_treat <- clinical_df[clinical_df$treatment == "Placebo", ]

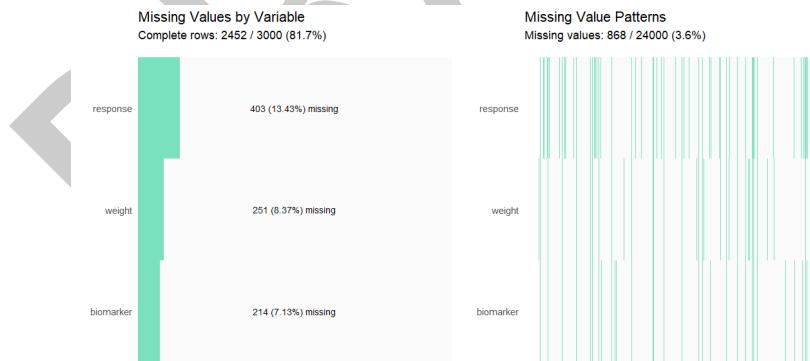
# Normally distributed variable
normality(data = clinical_df_treat, "biomarker")
#>
#> Normality Test for 'biomarker'
#>
#> n = 159
#> mean (SD) = 49.44 (9.2)
#> median (IQR) = 50.38 (13.1)
#>
#> Kolmogorov-Smirnov (Lilliefors): D = 0.054, p = 0.305
#> Shapiro-Wilk: W = 0.992, p = 0.546
#> Skewness: 0.06 (z = 0.30)
#> Kurtosis: -0.03 (z = -0.08)
#>
#> Data appears normally distributed.
```



```
# Simulate more complex clinical data
clinical_df_full <- clinical_data(n = 300, visits = 10, arms = c('A', 'B', 'C'),
dropout = 0.10, missing = 0.05)
```

Missing value analysis of only variables with missing values

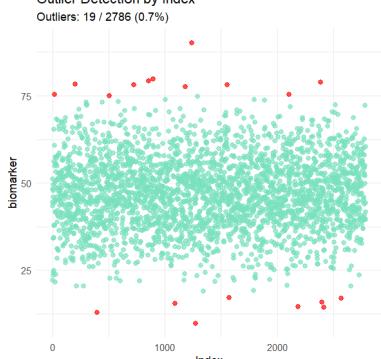
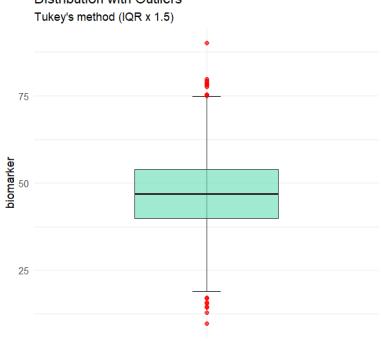
```
missing_values(clinical_df_full)
#>
#> Missing Value Analysis
#>
#> Complete rows: 2452 (81.7%)
#> Missing cells: 868 (3.6%)
#>
#> n_missing pct_missing
#> response      403   13.43
#> weight        251    8.37
#> biomarker     214    7.13
```



151

```
# Basic outlier detection
outliers(clinical_df_full, "biomarker")
#>
#> Outlier Analysis
#>
#> Variable: 'biomarker'
#> n: 2786
#> Missing: 214 (7.1%)
#> Method: Tukey's IQR x 1.5
#> Bounds: [18.971, 74.761]
#> Outliers detected: 19 (0.7%)
```

```
#>
#> Outlier indices: 27, 223, 440, 559, 795, 931, 973, 1175, 1277, 1346 (...)

Outlier Detection by Index
Outliers: 19 / 2786 (0.7%)

Distribution with Outliers
Tukey's method (IQR x 1.5)

152

# Two-sample parallel non-inferiority test for means with 10% expected dropout
sample_size(sample = 'two-sample', design = 'parallel', outcome = 'mean',
            type = 'non-inferiority', x1 = 5.0, x2 = 5.0,
            SD = 0.1, delta = -0.05, k = 1, dropout = 0.1)
#>
#> Sample Size Calculation
#>
#> Test type: non-inferiority
#> Design: parallel, two-sample
#> Outcome: mean
#> Alpha ( $\alpha$ ): 0.050
#> Beta ( $\beta$ ): 0.200
#> Power: 85.0%
#>
#> Parameters:
#> x1 (treatment): 5.000
#> x2 (control/reference): 5.000
#> Difference ( $x_1 - x_2$ ): 0.000
#> Standard Deviation ( $\sigma$ ): 0.100
#> Allocation Ratio ( $k$ ): 1.00
#> Delta ( $\delta$ ): -0.050
#> Dropout rate: 10.0%
#>
#> Required Sample Size
#> n1 = 55
#> n2 = 55
#> Total = 110
#>
#> Note: Sample size increased by 10.0% to account for potential dropouts.

# Compare numerical variable across treatments
omnibus(data = clinical_df_full, y = "biomarker", x = "treatment")
#>
#> Omnibus Test: One-way ANOVA
#>
#> Assumption Testing Results:
#>
#> Normality (Shapiro-Wilk Test):
#> A: W = 0.9985, p = 0.321
#> B: W = 0.9975, p = 0.237
```

```

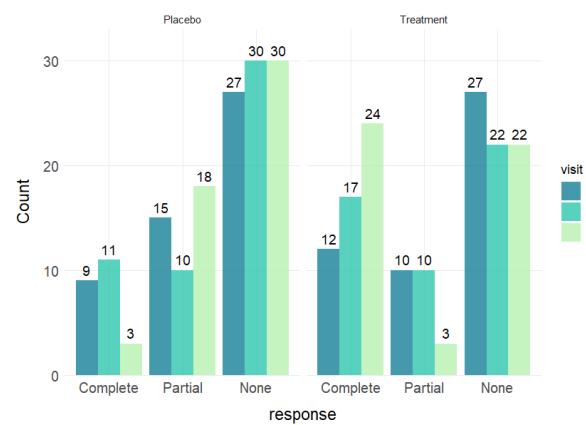
#> C: W = 0.9988, p = 0.733
#> Overall result: Normal distribution assumed.
#>
#> Homogeneity of Variance (Bartlett Test):
#> Chi-squared(2) = 1.3685, p = 0.504
#> Effect size (Cramer's V) = 0.0151
#> Result: Homogeneous variances.
#>
#> Test Results:
#> Formula: biomarker ~ treatment
#> alpha: 0.05
#> Result: significant (p = <0.001)
#>
#> Post-hoc Multiple Comparisons
#>
#> Tukey Honest Significant Differences (alpha: 0.050):
#> Comparison          Diff   Lower   Upper   p-adj
#> -----
#> B - A      -3.178  -4.296  -2.060  <0.001*
#> C - A      -5.542  -6.618  -4.466  <0.001*
#> C - B      -2.364  -3.468  -1.259  <0.001*
#>
#> The study groups show a moderately imbalanced distribution
#> of sample sizes ( $\Delta n = 0.214$ ).

effect_measures(exposed_event = 15,
                 exposed_no_event = 85,
                 unexposed_event = 5,
                 unexposed_no_event = 95)

#>
#> Odds/Risk Ratio Analysis
#>
#> Contingency Table:
#>           Event No Event   Sum
#> Exposed        15       85    100
#> Unexposed      5       95    100
#> Sum            20      185   200
#>
#> Odds Ratio: 3.353 (95% CI: 1.169 - 9.616)
#> Risk Ratio: 3.000 (95% CI: 1.133 - 7.941)
#>
#> Risk in exposed: 15.0%
#> Risk in unexposed: 5.0%
#> Absolute risk difference: 10.0%
#> Number needed to harm (NNH): 10.0
#>
#> Note: Correction not applied (no zero values).

# Grouped barplot of categorical variable by treatment with value labels
plot_bar(data = clinical_df, x = "response",
          group = "visit", facet = "treatment",
          title = "Response by visit and treatment", values = TRUE)

```



153

```
# Line plot with mean and standard error by treatment
plot_line(data = clinical_df_full, x = "visit", y = "biomarker",
           group = "treatment", stat = "mean", error = "se")
```

154

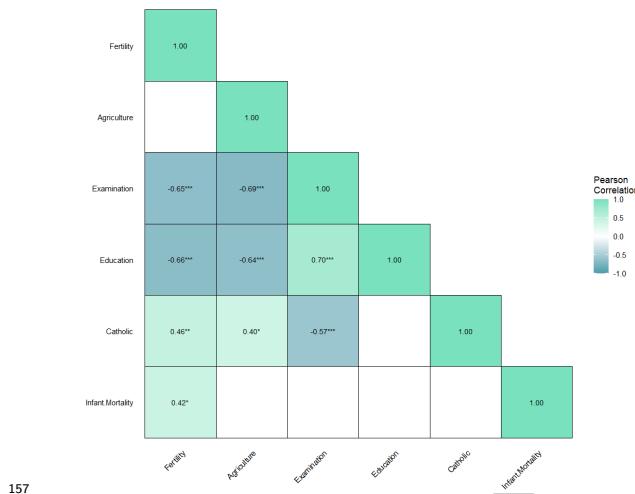
```
# Faceted histogram
plot_hist(clinical_df, x = "biomarker", facet = "treatment")
```

155

```
# Boxplot of biomarker by study visit and treatment
plot_box(clinical_df, x = "visit", y = "biomarker", group = "treatment")
```

156

```
# Lower triangle with significance indicators and filtering for R dataset 'swiss'
plot_corr(data = swiss, type = "lower", show_sig = TRUE, sig_only = TRUE)
```



158 License and Availability

159 The **biostats** package is distributed under an MIT license with source code available on [GitHub](#).

160 Acknowledgements

161 The authors wish to acknowledge the R open-source community for their ongoing maintenance of
 162 the packages upon which **biostats** depends, and for their continued commitment to transparency
 163 and reproducibility in scientific research. Gratitude is also extended to Laboratorios Sophia
 164 S.A. de C.V. for supporting the authors through salaries and employment, and for fostering an
 165 environment that promotes innovation, open-source development, and open science.

166 AI usage disclosure

167 Generative AI tools were used during the development of the **biostats** package to assist with
 168 code refinement, debugging, automated tests, and the configuration of continuous integration
 169 and continuous deployment (CI/CD) workflows through GitHub actions. These tools were also
 170 used to review and improve the final manuscript. All AI-generated suggestions were carefully
 171 reviewed, modified, and validated by the authors. The authors assume full responsibility and
 172 accountability for the reliability, integrity, and maintenance of the software provided.

173 References

- 174 Anderson, K., Zhao, Y., Blischak, J., & Zhang, Y. (2025). *Simtrial: Clinical trial simulation*.
<https://CRAN.R-project.org/package=simtrial>
- 175 Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). *effectsize*: Estimation of effect
 size indices and standardized parameters. *Journal of Open Source Software*, 5(56), 2815.
<https://doi.org/10.21105/joss.02815>
- 176 Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Effect of variance
 177 ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods*, 50(3),
 178 937–962. <https://doi.org/10.3758/s13428-017-0918-2>

- 182 Champely, S. (2020). *Pwr: Basic functions for power analysis*. [https://CRAN.R-project.org/
package=pwr](https://CRAN.R-project.org/package=pwr)
- 183
- 184 Chow, S.-C., Shao, J., Wang, H., & Lokhnygina, Y. (2017). *Sample size calculations in clinical
research* (3rd ed.). Chapman; Hall/CRC. <https://doi.org/10.1201/9781315183084>
- 185
- 186 Ciolino, J. D., Spino, C., Ambrosius, W. T., Khalatbari, S., Cayetano, S. M., Lapidus, J. A.,
187 Nietert, P. J., Oster, R. A., Perkins, S. M., Pollock, B. H., Pomann, G. M., Price, L. L., Rice,
188 Tosteson, T. D., Lindsell, C. J., & Spratt, H. (2021). Guidance for biostatisticians on
189 their essential contributions to clinical and translational research protocol review. *Journal
of Clinical and Translational Science*, 5(1), e161. <https://doi.org/10.1017/cts.2021.814>
- 190
- 191 Dallal, G. E., & Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors'
192 test for normality. *The American Statistician*, 40(4), 294–296. [https://doi.org/10.1080/
00031305.1986.10475419](https://doi.org/10.1080/
00031305.1986.10475419)
- 193
- 194 Dwivedi, A. K. (2022). How to write statistical analysis section in medical research. *Journal
of Investigative Medicine*, 70(8), 1759–1770. <https://doi.org/10.1136/jim-2022-002479>
- 195
- 196 Field, A. P., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.
- 197 Harrell Jr, F. E. (2026). *Hmisc: Harrell miscellaneous*. [https://CRAN.R-project.org/package=Hmisc](https://CRAN.R-project.org/package=
Hmisc)
- 198
- 199 Hill, N., & Thorley, J. (2025). *Extras: Helper functions for bayesian analyses*. [https://CRAN.R-project.org/package=extras](https://CRAN.R-
project.org/package=extras)
- 200
- 201 Hodge, D. (2025). *Ggblanket: Simplify 'ggplot2' visualisation*. [https://CRAN.R-project.org/
package=ggblanket](https://CRAN.R-project.org/
package=ggblanket)
- 202
- 203 Kassambara, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests*. [https://CRAN.R-project.org/package=rstatix](https://
CRAN.R-project.org/package=rstatix)
- 204
- 205 Kassambara, A. (2025). *Ggpubr: 'ggplot2' based publication ready plots*. [https://CRAN.R-project.org/package=ggpubr](https://CRAN.R-
project.org/package=ggpubr)
- 206
- 207 Lawrence, M. A. (2016). *Ez: Easy analysis and visualization of factorial experiments*. [https://CRAN.R-project.org/package=ez](https://
CRAN.R-project.org/package=ez)
- 208
- 209 Lilliefors, H. W. (1967). On the Kolmogorov–Smirnov test for normality with mean and
210 variance unknown. *Journal of the American Statistical Association*, 62(318), 399–402.
211 <https://doi.org/10.1080/01621459.1967.10482916>
- 212
- 213 Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive
214 statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1),
67–72. https://doi.org/10.4103/aca.ACA_157_18
- 215
- 216 Qi, H., & Zhu, F. (2021). *SampleSize4ClinicalTrials: Sample size calculation for the comparison
of means or proportions in phase III clinical trials*. [https://CRAN.R-project.org/package=SampleSize4ClinicalTrials](https://CRAN.R-project.org/package=
SampleSize4ClinicalTrials)
- 217
- 218 R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation
219 for Statistical Computing. <https://www.R-project.org/>
- 220
- 221 Sagar, K. P., P., R., & B., S. (2023). Biostatistical methodologies in clinical trials: An overview
222 of recent developments and pitfalls. *International Journal of Clinical Biostatistics and
Biometrics*, 9, 051. <https://doi.org/10.23937/2469-5831/1510051>
- 223
- 224 Vicky Qian Wu ; Shein-Chung Chow ; Harry G.Zhang, E. Z.; (2024). *TrialSize: R functions
for chapter 3,4,6,7,9,10,11,12,14,15 of sample size calculation in clinical research*. [https://CRAN.R-project.org/package=TrialSize](https://
CRAN.R-project.org/package=TrialSize)
- 225
- 226 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
227 ISBN: 978-3-319-24277-4
- 228

- ²²⁸ Yoshida, K., & Bartel, A. (2022). *Tableone*: Create 'table 1' to describe baseline characteristics
²²⁹ with or without propensity score weights. <https://CRAN.R-project.org/package=tableone>
- ²³⁰ Zhang, H. (2025). *TrialSimulator*: Clinical trial simulator. <https://CRAN.R-project.org/package=TrialSimulator>

DRAFT