The Journal of Open Source Software

easystats: A Unified Framework for Statistical Analysis in R

Daniel Lüdecke[1], **Mattan S. Ben-Shachar**[2], **Indrajeet Patil**[3], **Brenton M. Wiernik**[4], **Etienne Bacher**[5], **Rémi Thériault**[6], and **Dominique Makowski**[7, 8]

1 University Medical Center Hamburg-Eppendorf, Germany **2** Independent Researcher, Ramat Gan, Israel **3** Carl Zeiss AG, Germany **4** Independent Researcher, Tampa, FL, USA **5** Luxembourg Institute of Socio-Economic Research (LISER), Luxembourg **6** Department of Psychology, New York University, New York, NY, USA **7** School of Psychology, University of Sussex, Brighton, UK **8** Sussex Centre for Consciousness Science, University of Sussex, Brighton, UK

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

The **easystats** project is a collection of R packages that provides a unified and intuitive framework for data wrangling and statistical analysis. The ecosystem is built around a "tidyverse-like" philosophy of consistency, user-friendliness, and interoperability between packages. It aims to simplify the process of preparing, conducting, interpreting, and reporting statistical analyses by offering tools for a wide range of common tasks. These tasks include, for instance, data wrangling (`{datawizard}`, Patil et al., 2022), model assessment (`{performance}`, Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021), understanding and describing model parameters (`{parameters}`, Lüdecke, Ben-Shachar, Patil, & Makowski, 2020), including Bayesian models (`{bayestestR}`, Makowski, Ben-Shachar, & Lüdecke, 2019), computation of effect sizes (`{effectsize}`, Ben-Shachar, Lüdecke, & Makowski, 2020), calculating and visualizing marginal effects (`{modelbased}`, Makowski et al., 2025), and generating publication-ready figures (`{see}`, Lüdecke, Patil, et al., 2021) or reports of statistical models (`{report}`, Makowski et al., 2023). The **easystats** ecosystem is designed to be accessible to both novice and experienced R users, promoting reproducible research and a more seamless workflow from data exploration to result communication.

Statement of Need

The R programming language is a powerful and dominant tool for statistical computing in academic research. However, its power is distributed across a vast and fragmented ecosystem of packages. For a researcher to conduct a comprehensive statistical analysis—from fitting a model to checking its assumptions, interpreting its parameters, and visualizing the results—they often need to learn and combine numerous packages, each with its own unique syntax, design principles, and output structures. This fragmentation presents a significant barrier to entry for new users and creates an inefficient and often confusing workflow, even for experienced analysts.

The **easystats** ecosystem addresses this challenge by providing a unified, consistent, and intuitive framework for data wrangling, data exploration, statistical summaries and modeling, visualization, and reporting in R. Its primary goal is to make robust and transparent statistical analysis more accessible to a broad range of users, from students to seasoned researchers. **easystats** achieves this through a suite of lightweight and interoperable packages, each designed to handle a specific stage of the statistical workflow. These packages share a consistent syntax and work seamlessly together, hence dramatically lowering the cognitive load required to conduct such analytical workflows.

Other software toolkits often serve different purposes. The `{tidyverse}` (Wickham et al., 2019), for example, provides a world-class environment for data manipulation and general-purpose plotting but does not focus on the intricacies of statistical model interpretation and reporting. Specialist packages like `{lme4}` (Bates, Mächler, Bolker, & Walker, 2015) for mixed-effects models or `{marginaleffects}` (Arel-Bundock, Greifer, & Heiss, 2024) for marginal effects and predictions are essential tools, but **easystats** serves as a complementary "meta-layer" that provides a single, easy-to-learn interface for interacting with the outputs from these and many other modeling packages. This allows researchers to focus on their scientific questions rather than on the technical idiosyncrasies of different software implementations. Hence, **easystats** fills a critical need in the R ecosystem for a user-friendly and cohesive suite of tools dedicated to the statistical modeling pipeline.

The modularity of the **easystats** packages enables developers to select and use only the necessary components. For example, `{insight}`, a dependency-free package for retrieving model information, is utilized by 45 other CRAN packages, and `{parameters}` is used by 22. In contrast, the `{easystats}` meta-package provides users with a cohesive experience, granting access to the entire ecosystem and its consistent design principles without needing to know the specific package of each function.

A Harmonized and Integrated Workflow

A key design principle of the **easystats** ecosystem is the harmonization and integration of different packages into a simple, sequential workflow. The typical workflow for a statistical analysis using `{easystats}` starts with importing data and bringing the data into shape for the next step - fitting a model - and then sequentially using different functions to obtain a comprehensive understanding of the model.

Let's demonstrate this with an example, where the user starts by preparing some data and then fits a simple linear model:

```r
# we don't load each package individually,
# but rather the entire ecosystem
library(easystats)
data(coffee_data, package = "modelbased")

# dichotomize outcome variable
coffee_data$alertness <- categorize(coffee_data$alertness, lowest = 0)

# rename variable
coffee_data <- data_rename(coffee_data, select = c(treatment = "coffee"))

# fit model
model <- glm(
  alertness ~ treatment,
  data = coffee_data,
  family = binomial()
)
```

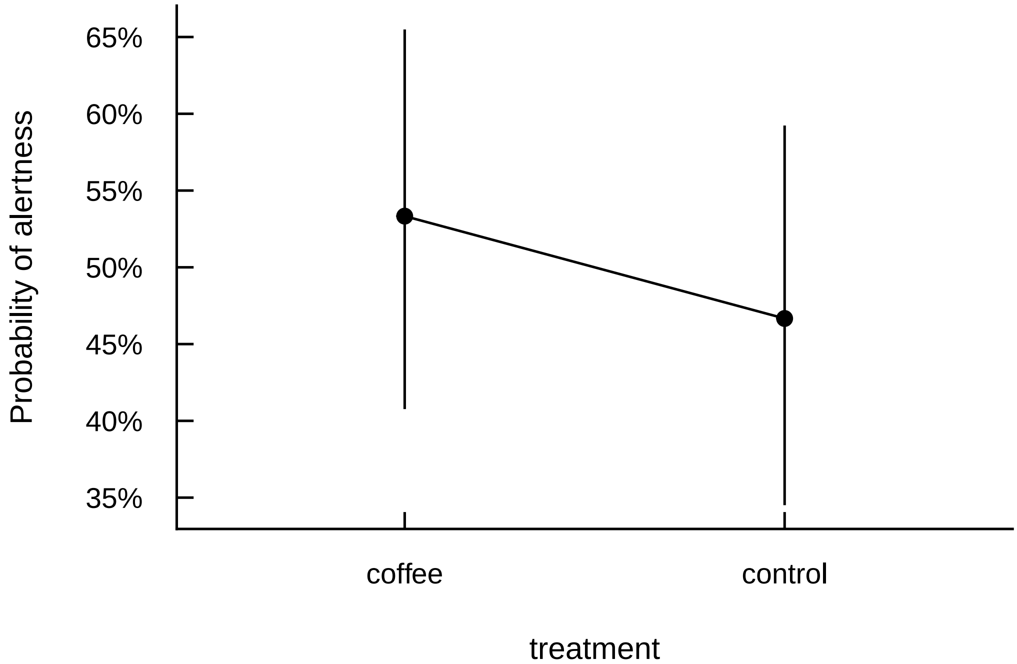

Figure 1: Predicted probability of alertness by treatment group.

The `model` object can then be passed to functions from different **easystats** packages. For instance, the user can get a summary of the model parameters using the `{parameters}` package:

```
model_parameters(model)
#> Parameter            | Log-Odds |   SE |         95% CI |     z |     p
#> ----------------------------------------------------------------------
#> (Intercept)          |     0.13 | 0.26 | [-0.37, 0.65] |  0.52 | 0.606
#> treatment [control] |    -0.27 | 0.37 | [-0.99, 0.45] | -0.73 | 0.466
```

Then, the performance of the model can be assessed with the `{performance}` package:

```
model_performance(
  model,
  metrics = c("AIC", "BIC", "R2", "PCP")
)
#> # Indices of model performance
#>
#> AIC    |   BIC | Tjur's R2 |   PCP
#> ---------------------------------
#> 169.8 | 175.4 |     0.004 | 0.502
```

The results can be visualized using the `{see}` package, for example by plotting the model's predictions with `{modelbased}`:

```
predictions <- estimate_means(model, "treatment")
plot(predictions) + theme_modern(show.ticks = TRUE) # add nice theme
```

Finally, a full report of the analysis can be generated with the `{report}` package:

```r
report(model)
#> We fitted a logistic model (estimated using ML) to predict alertness with
#> treatment (formula: alertness ~ treatment). The model's explanatory power is
#> very weak (Tjur's R2 = 4.44e-03). The model's intercept, corresponding to
#> treatment = coffee, is at 0.13 (95% CI [-0.37, 0.65], p = 0.606). Within this
#> model:
#>
#>   - The effect of treatment [control] is statistically non-significant and
#> negative (beta = -0.27, 95% CI [-0.99, 0.45], p = 0.466; Std. beta = -0.27, 95%
#> CI [-0.99, 0.45])
#>
#> Standardized parameters were obtained by fitting the model on a standardized
#> version of the dataset. 95% Confidence Intervals (CIs) and p-values were
#> computed using a Wald z-distribution approximation.
```

This seamless integration between packages allows users to move from model fitting to interpretation and reporting in a fluid and intuitive manner, without having to learn different syntaxes or data structures.

Licensing and Availability

`{easystats}` is licensed under the MIT-License, with all source code stored at GitHub (https://github.com/easystats/easystats), and with a corresponding issue tracker for bug reporting and feature enhancements. 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 contributors and developers.

Acknowledgments

`{easystats}` is part of the collaborative **easystats** ecosystem. Thus, we thank all members of easystats, contributors, and users alike.

References

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to Interpret Statistical Models Using **marginaleffects** for *r* and *python*. *Journal of Statistical Software*, *111*(9). doi:10.18637/jss.v111.i09

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

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. doi:10.21105/joss.02815

Lüdecke, D., Ben-Shachar, M. S., Patil, I., & Makowski, D. (2020). parameters: Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, *5*(53), 2445. doi:10.21105/joss.02445

Lüdecke, 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), 3139. doi:10.21105/joss.03139

Lüdecke, D., Patil, I., Ben-Shachar, M., Wiernik, B., Waggoner, P., & Makowski, D. (2021). See: An R Package for Visualizing Statistical Models. *Journal of Open Source Software*, *6*(64), 3393. doi:10.21105/joss.03393

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. doi:10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Patil, I., Thériault, R., & Lüdecke, D. (2025). modelbased: An R package to make the most out of your statistical models through marginal means, marginal effects, and model predictions. *Journal of Open Source Software*, *10*(109), 7969. doi:10.21105/joss.07969

Makowski, D., Lüdecke, D., Patil, I., Thériault, R., Ben-Shachar, M. S., & Wiernik, B. M. (2023). Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption. *CRAN*. doi:10.32614/CRAN.package.report

Patil, I., Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Bacher, E., & Lüdecke, D. (2022). datawizard: An R package for easy data preparation and statistical transformations. *Journal of Open Source Software*, *7*(78), 4684. doi:10.21105/joss.04684

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686