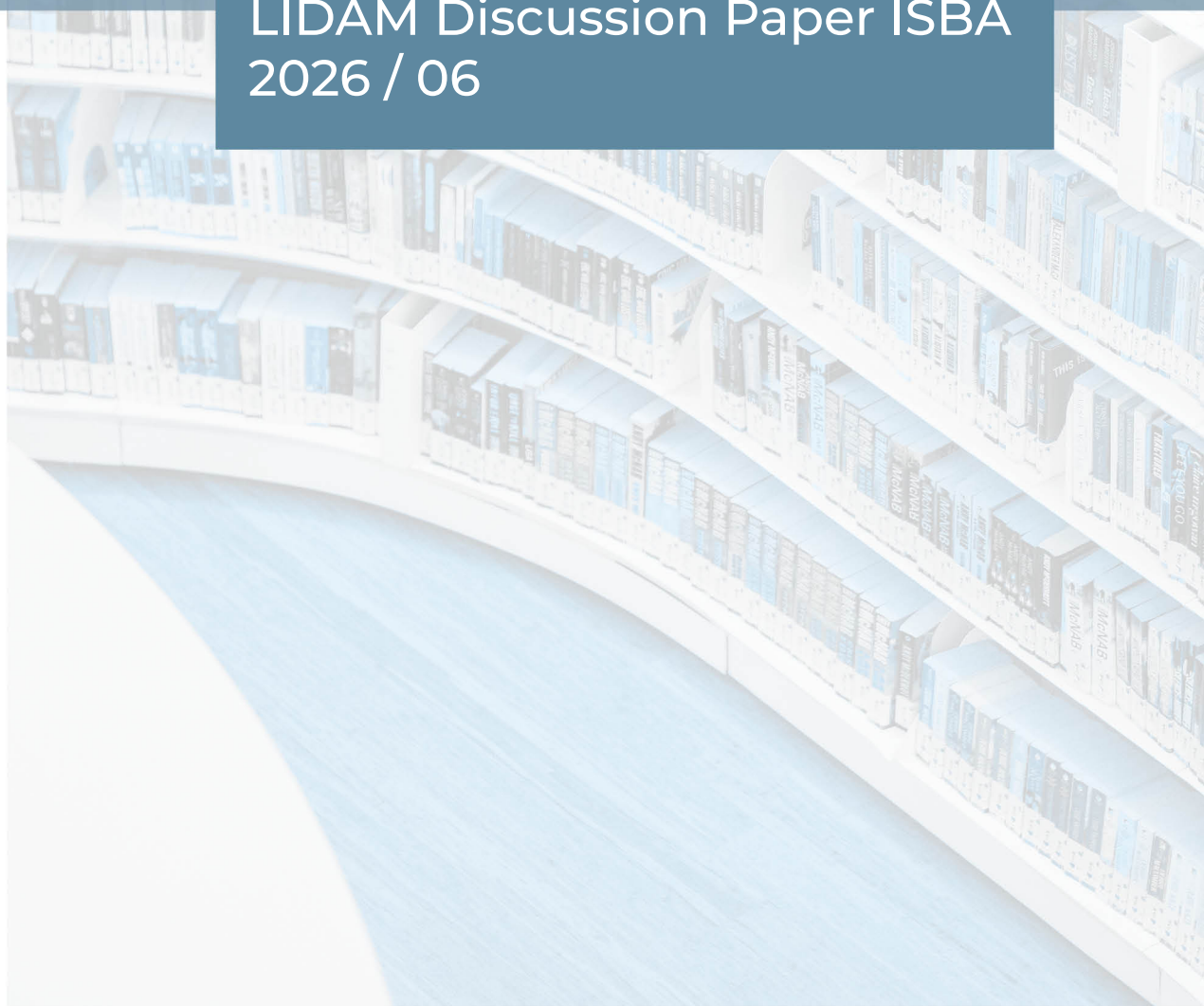
STATISTICS 101, 201, AND 202: THREE SHINY APPS FOR TEACHING PROBABILITY DISTRIBUTIONS, INFERENTIAL STATISTICS, AND SIMPLE LINEAR REGRESSION

Antoine Soetewey

LIDAM Discussion Paper ISBA
2026 / 06



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

The Journal of Open Source Education

Statistics 101, 201, and 202: Three Shiny Apps for Teaching Probability Distributions, Inferential Statistics, and Simple Linear Regression

Antoine Soetewey [1]

1 HEC Liège, ULiège, Rue Louvrex 14, 4000 Liège, Belgium

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Submitted: 01 January 1970
Published: unpublished

Summary

Statistics 101, 201, and 202 are three open-source interactive web applications built with R (R Core Team, 2025) and Shiny (Chang et al., 2025) to support the teaching of introductory statistics and probability. The apps are intended to help students carry out common statistical computations (computing probabilities from standard probability distributions, constructing confidence intervals, conducting hypothesis tests, and fitting simple linear regression models) without requiring prior knowledge of R or any other programming language. Each app provides numerical results, plots rendered with `ggplot2` (Wickham, 2016), and inline mathematical derivations typeset with MathJax (Cervone, 2012), so that computation and statistical reasoning appear side by side in a single interface.

The suite was developed during the author's doctoral research and teaching assistantship at UCLouvain and continues to be used as an invited lecturer at UCLouvain and UNamur in courses on probability, inferential statistics, and linear regression. The three apps are organised around a broad pedagogical progression: Statistics 101 introduces probability distributions and their properties; Statistics 201 addresses confidence intervals and hypothesis tests; and Statistics 202 covers the simple linear model. All three apps are freely accessible online and their source code is released under a CC-BY-4.0 license.

Statement of Need

Introductory statistics courses in economics, biostatistics, social sciences, and related disciplines commonly ask students to evaluate probabilities, construct confidence intervals, perform hypothesis tests, and fit linear regression models. Working through these computations by hand can be time-consuming and error-prone, while doing so in R presupposes a level of programming familiarity that beginners may not yet have. Printed or online statistical tables can help with probability computations, but typically cover only a limited number of distributions and offer no graphical feedback.

Several existing tools partially address this gap. Calculators bundled with commercial textbooks are often proprietary. General-purpose environments such as R, Python, or SPSS require students to write or adapt code. Web-based probability calculators (for instance those offered by Wolfram Alpha or individual course websites) tend to cover only a limited number of distributions, and many do not include hypothesis testing or regression modules, nor do they display the underlying mathematical expressions alongside their results. Shiny-based tools for teaching statistics have been explored in the literature (Potter et al., 2016; Soetewey et al., 2026), though these tend to focus on specific topics or workflows rather than covering the full introductory sequence in single interfaces.

The Statistics 101/201/202 suite is an attempt to complement these resources in a

JOSE
The Journal of Open Source Education

few ways. It brings together topics from a typical two-semester introductory statistics sequence (probability distributions, inference, and regression) within a single, consistent interface. Each numerical result is shown alongside a dynamically updated plot and the corresponding mathematical expression, with the aim of supporting statistical intuition alongside computation. The apps are fully open source, so other instructors are welcome to adopt, fork, or extend them. Their continued use across introductory and intermediate statistics courses at UCLouvain and UNamur suggests they may be useful beyond the context in which they were originally developed.

Description of the Apps

Statistics 101: Probability Distributions

Statistics 101 (github.com/AntoineSoetewey/statistics-101) is designed to help students compute probabilities from 18 probability distributions: Beta, Binomial, Cauchy, Chi-square, Exponential, Fisher, Gamma, Geometric (two parameterisations), Hypergeometric, Logistic, Log-Normal, Negative Binomial (two parameterisations), Normal, Poisson, Student's t, and Weibull. The app uses a single-page sidebar layout. Students select a distribution from a drop-down menu, and the sidebar dynamically renders the appropriate parameter inputs using conditional panels.

For each distribution, the student specifies the type of probability to compute:

- Lower tail: $P(X \leq x)$
- Upper tail: $P(X > x)$
- Interval: $P(a \leq X \leq b)$

The main panel displays three outputs:

1. Solution: the probability statement in full mathematical notation rendered via MathJax, for example $X \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ and $P(X \leq 1) = 0.8413$, computed to four decimal places using R's built-in distribution functions (`pbeta`, `pbinom`, `pnorm`, etc.).
2. Plot: a density or mass function plot with the relevant region shaded, rendered using `ggplot2` graphics and displayed with a loading spinner via `shinycssloaders` (Attali & Sali, 2024).
3. Details: the probability density function (PDF) or probability mass function (PMF) in display-style mathematical notation, followed by the expectation $E(X)$, standard deviation $SD(X)$, and variance $Var(X)$, all evaluated numerically at the current parameter values.

Statistics 201: Confidence Intervals and Hypothesis Tests

Statistics 201 (github.com/AntoineSoetewey/statistics-201) is intended to support the teaching of confidence intervals and hypothesis tests across seven inference settings, selectable from a drop-down menu:

- One mean: z-test (population variance σ^2 known) or one-sample t-test (σ^2 unknown)
- Two means, independent samples: z-test or two-sample t-test with either equal ($\sigma_1^2 = \sigma_2^2$) or unequal ($\sigma_1^2 \neq \sigma_2^2$) variances (Welch's correction)
- Two means, paired samples: paired t-test or z-test on the differences $D_i = X_{1i} - X_{2i}$
- One proportion: one-sample z-test
- Two proportions: two-sample z-test with an option to use a pooled standard error
- One variance: χ^2 test, using `EnvStats::varTest()` (Millard, 2013) for raw data and a custom helper function for summary statistics (n and s^2)
- Two variances: F-test, using `var.test()` from base R

JOSE
The Journal of Open Source Education

For all mean and variance settings, students may provide either raw data (comma-separated values pasted into a text field) or summary statistics (n, $\bar{x}$, s^2, etc.) directly. The significance level α is controlled via a slider (range 0.01 to 0.20, default 0.05). The null hypothesis value H_0 and the direction of the alternative hypothesis (two-sided, greater, or less) are set in the sidebar.

For each procedure, the main panel displays a structured four-step output:

1. Data: the entered observations (if raw data were provided) together with the relevant summary statistics (n, $\bar{x}$, s, etc.), typeset in MathJax.
2. Confidence interval: the interval formula with all quantities substituted numerically, for example $\bar{x} \pm \left(t_{\alpha/2,\,n-1} \times s/\sqrt{n}\right) = [\ell;\, u]$, in the appropriate one- or two-sided form.
3. Hypothesis test: a numbered four-step procedure consisting of (1) the statement of H_0 and H_1, (2) the test statistic formula evaluated numerically, (3) the critical value(s), and (4) the reject or do-not-reject conclusion.
4. Interpretation: a plain-language conclusion at the chosen significance level, including the p-value.

Below the text output, a `ggplot2` plot shows the relevant sampling distribution (t, $\mathcal{N}(0,1)$, χ^2, or F) with the rejection region shaded and the observed test statistic marked by a vertical line.

Statistics 202: Simple Linear Regression

Statistics 202 (github.com/AntoineSoetewey/statistics-202) is designed to help students fit a simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ to data entered directly in the browser. The app accepts comma-separated numeric values for the predictor x and the response y, checks that the two vectors are of equal length and that x contains more than one distinct value. The fitted model is computed via `lm()` in R.

The main panel presents the following sequence of outputs:

1. Data table: an interactive DT (Xie et al., 2025) table of the entered observations, with export buttons (copy, CSV, Excel, PDF, print).
2. Summary statistics: sample means $\bar{x}$, $\bar{y}$, and sample size n.
3. Step-by-step derivation: the OLS estimator $\hat{\beta}_1$ computed via the formula

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

   followed by $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, both rendered in MathJax and evaluated numerically on the student's own data (Fox, 2015).
4. R output: the full `summary(lm(...))` output, including coefficient estimates, standard errors, t-statistics, p-values, and adjusted R^2.
5. Interactive regression plot: a scatter plot with the fitted line and an optional confidence band, rendered with `ggplot2` and made interactive via plotly (Sievert, 2020).
6. Interpretation: a plain-language reading of $\hat{\beta}_0$ and $\hat{\beta}_1$, adapted to the user-supplied axis labels and conditioned on whether each coefficient is significantly different from zero at the 5% level.
7. Assumption diagnostics: a collapsible panel containing the six model-checking plots produced by `performance::check_model()` (Lüdecke et al., 2021), covering linearity, homoscedasticity, influential observations and normality of residuals.
8. Downloadable report: an HTML report of the analysis, generated from an R Markdown (Allaire et al., 2025; Xie et al., 2018, 2020) template, with an option to include or hide the underlying R code.

JOSE
The Journal of Open Source Education

Pedagogical Design

The three apps share a common design orientation. The interface is kept relatively minimal so that attention can remain on the statistical content rather than on navigating the software. No registration, installation, or command-line interaction is required; students open a URL and begin working straight away. Results are displayed at a level of mathematical abstraction close to what is used in lectures, with notation, formulas, and numerical output appearing together, so that the apps can serve as a complement to, rather than a replacement for, the conceptual material covered in class.

The step-by-step derivation in Statistics 202 was motivated by a difficulty that often arises in introductory regression courses: students may be able to write down a formula from a slide yet find it hard to connect it to the numerical output produced by R. Showing both the symbolic formula and its numerical evaluation side by side, on the student's own data, is one way of trying to make that connection more concrete. A related idea underpins the Details panel in Statistics 101, which displays the PDF or PMF alongside the computed probability, so that students can see how the shaded area in the plot corresponds to the analytical expression.

A comparable motivation applies to Statistics 201. Introductory statistics courses typically cover a large number of inference procedures, each with its own test statistic formula, degrees of freedom, and critical value. Students often find it difficult to locate the relevant formula quickly when working through an exercise, particularly when the course material is spread across a lengthy syllabus or a large set of slides. By displaying the appropriate formula for each procedure alongside its numerical evaluation, Statistics 201 is intended to serve as a single reference point where students can find, read, and apply the correct expression without having to search through their course materials.

The apps were originally developed in response to a practical difficulty observed during teaching-assistant sessions at UCLouvain, where students were spending a considerable share of class time on the mechanics of looking up or computing probabilities and test statistics, leaving relatively little time for interpretation and discussion. The hope was that shifting some of that computational work to a transparent, easy-to-use tool might free up more time for conceptual engagement.

Availability and Reproducibility

All three apps are hosted on shinyapps.io and their source code is publicly available on GitHub under a CC-BY-4.0 license:

- Statistics 101: github.com/AntoineSoetewey/statistics-101
- Statistics 201: github.com/AntoineSoetewey/statistics-201
- Statistics 202: github.com/AntoineSoetewey/statistics-202

Each app's GitHub repository includes a README file with instructions on how to use the app, the dependencies, instructions for local deployment and a link to the live version.

Acknowledgements

The author thanks Prof. Dominique Deprins and Prof. Eugen Pircalabelu for their comments on previous versions of the apps. The author is grateful to his PhD supervisors, Prof. Catherine Legrand and Prof. Michel Denuit, for their support and for allowing him to work on these apps during his doctoral studies at UCLouvain. Finally, the author thanks the students at UCLouvain, without whom these apps would not exist.

JOSE
The Journal of Open Source Education

References

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2025). *Rmarkdown: Dynamic documents for r*. https://github.com/rstudio/rmarkdown

Attali, D., & Sali, A. (2024). *Shinycssloaders: Add loading animations to a 'shiny' output while it's recalculating*. https://doi.org/10.32614/CRAN.package.shinycssloaders

Cervone, D. (2012). MathJax: A platform for mathematics on the web. *Notices of the AMS*, *59*(2), 312–316.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Aden-Buie, G., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2025). *Shiny: Web application framework for r*. https://doi.org/10.32614/CRAN.package.shiny

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage publications.

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. https://doi.org/10.21105/joss.03139

Millard, S. P. (2013). *EnvStats: An r package for environmental statistics*. Springer. ISBN: 978-1-4614-8455-4

Potter, G., Wong, J., Alcaraz, I., Chi, P., & others. (2016). Web application teaching tools for statistics using r and shiny. *Technology Innovations in Statistics Education*, *9*(1).

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://doi.org/10.32614/r.manuals

Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny*. Chapman; Hall/CRC. ISBN: 9781138331457

Soetewey, A., Heuchenne, C., Claes, A., & Descampe, A. (2026). AssociationExplorer: A user-friendly shiny application for exploring associations and visual patterns. *SoftwareX*, *33*, 102483. https://doi.org/10.1016/j.softx.2025.102483

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. ISBN: 9781138359338

Xie, Y., Cheng, J., Tan, X., & Aden-Buie, G. (2025). *DT: A wrapper of the JavaScript library 'DataTables'*. https://doi.org/10.32614/CRAN.package.DT

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC. ISBN: 9780367563837