A Shiny App for Learning Logistic Regression

Marius Keute

March 15, 2022

A Shiny App for Learning Logistic Regression

Marius Keute

Introduction

One of my jobs is statistical consulting for medical researchers. Time and again, clients come to me with regression tables that some friend or colleague helped them generate with IBM SPSS or other statistical software. Typically, however, my clients have no idea how to proceed with the information contained in the output. They know that they have to look for "something significant", but do not seem to have much understanding of what a regression model does and how the output table can be read. This confusion tends to be even worse when it comes to logistic regression (LR). Medical data is often categorical (sick vs. not sick; ventilated vs. not ventilated; Hodgkin vs. Non-Hodgkin lymphoma etc.), so LR models are quite prevalent. For a person with little statistical training (which, in my experience, most medical researchers are, with notable exceptions), the idea of interpreting a regression weight as a log-odds-ratio, which reflects a probability prediction, is not very intuitive - after all, there was nothing probabilistic about the original data, just samples belonging to categories. To make the concepts behind LR more tangible, I created an educational Shiny App that explains the core ideas of LR. I drew inspiration from an existing educational app on linear regression (Rivera Baena, no year), but I was not able to find a similar, existing Shiny App for LR.

App Design

The app comprises two R scripts, `ui.R` and `server.R`. Within `ui.R`, the `shinyUI` function is called, containing all elements of the graphical user interface (GUI), whereas in `server.R`, a call to `shinyServer` contains the backend and input/output functionality of the app. This is the standard design for a Shiny App (Wickham, 2020). The app consists of several tab pages, each corresponding to one didactic step (see below). UI-to-server and server-to-UI communication is controlled through the `input` and `output` data structures, as usual in Shiny apps (Wickham, 2020).

GUI

The App uses the built-in `spacelab` theme without modifications. It is split into five pages that can be navigated using `tabPanels`. Graphics are generated using `ggplot2`. User Interaction with figures and data is realized by several `sliderInput` instances. The app's GUI combines explanatory texts and interactive/animated graphics.

Backend

The core of the app's backend is the data-generating function `get_toydata`. It creates a simple medical dataset, with a clinical global impression (CGI) score, t-cell count and systolic blood pressure as continuous variables, and sex (m/f) and smoker (yes/no) as categorical variables. As an additional categorical variable, CGI is thresholded (if CGI < 50: sick; if CGI >= 50: not sick). The three continuous variables are drawn randomly from a multivariate normal distribution. Users can manipulate relationships between variables in the dataset via `sliderInput` instances. The backend function makes sure that n impossible covariance structures (singular matrices) are imposed.

Most text elements have been refactored into Markdown files, rather than being hardcoded in the source code. This simplifies modifications of the texts, e.g., translations to other languages, and improves code readability.

Didactic Structure

On the first page, there is a short introductory text and the user can browse the dataset.

On the second page, the user can manipulate relationships between the variables in the dataset, and the user is invited on the subsequent pages to return to this page and play around with these relationships to understand the effects on the regression model.

The third page contains a recap of linear regression, since I have the impression that basic concepts of regression models are easier to explain with linear models instead of jumping to logistic models right away. An animated figure gives the user an idea of least-squares fitting.

The fourth page introduces logistic regression, with a very similar structure and example model as for linear regression.

While the third and fourth pages only covered bivariate models, the last slide covers the generalization to models with more than one explanatory variable.

Possible Improvements

A number of things could be improved about this app that I did not have the time to do within the scope of this semester project:

- The app uses a standard design theme, and the visual appearance could clearly be more pleasing
- The same is true for the figures
- The explanatory texts might still be a bit too technical for people without prior statistical training.
- Some more (animated/interactive) figures could help, e.g. to demonstrate why a linear model would output poor predictions for binary data
- The page on logistic regression should also get an interactive element, so that users can experiment with the regression weights
- There could be the option for users to read in their own datasets (e.g. as a csv file)

Conclusion

I have described a simple Shiny App that can be used as a tutorial for logistic regression. Since it was the first Shiny App that I have built, it was primarily a learning experience for me, and many of the programming concepts (e.g., programming the reactive functions) feel very unfamiliar, given that I usually write R scripts for data analysis that just run linearly without any user interaction. While many things about this app are still rudimentary and need further development, I am surprised how simple the development was, once I had understood some principles of Shiny programming. Hadley Wickham's e-book (Wickham, 2020) was a great resource to get started, and after a short onboarding period, I really enjoyed building this app and I will continue to improve it in the future, when I find the time.

The most up-to-date version is available at

https://github.com/mkeute/learn-logistic-regression

References

Rivera Baena, O.D. (no year). Didactic modeling process: Linear regression https://shiny.rstudio.com/gallery/didacting-modeling.html [10 March 2022]

Wickham, H. (2020). Mastering Shiny. O'Reilly Media. https://mastering-shiny.org/ [10 March 2022]