

W05-1 Predictions

This worksheet puts linear regressions on the next level by actually using them to predict values of the dependent variable based on the independent one. After completing this worksheet you should know how to predict the values of the dependent variable using linear regression models.

Things you need for this worksheet

- R — the interpreter can be installed on any operation system. For Linux, you should use the r-cran packages supplied for your Linux distribution. If you use Ubuntu, [this](#) is one of many starting points. If you use windows, you could install R from the official [CRAN](#) web page.
- R Studio — we recommend to use R Studio for (interactive) programming with R. You can download R Studio from the official [web page](#).
- your script and data from [W02-1: Reading CSV files](#)

What's the plan?

What we want to do in this worksheet is to take the vegetation coverage of each plot (because we have that for all 161 plots) and predict the animal activity (because we have that only for 50 of those plots). The model will be built by using those 50 plots with information about animal activity and the corresponding vegetation coverage. That model will be used to predict the animal activity for the rest of the plots.

Learning log assignments

🚨 First things first: the following analysis is build on top of your script from [W02-1](#). Please copy your script "W02-1.R", rename the copy to "W05-1.R" and use it for the programming tasks of this worksheet.

😬 Let's have a look at the distribution of the coverage. You can use the `hist()` function for that. Does this plot suit your expectations of normally distributed data? If not, you could try to plot a histogram of the square root. The function `sqrt()` will help you with that. Please also check the distribution for the animal activity in the same way.

😬 Now it's time to redo our regression. You could peek at [W04-1](#) and do the regression in the same way, only this time please use the square root.

Don't forget: if you are working in a new script (not in the one from [W04-1](#)) you have to load the package `car` again.

😬 Let's use our linear model now to predict the animal activity by the square root of coverage. Please use the `predict()` function and save the result in a variable. Remember that you still have the square root of the value so please convert it back to the original ones. An exponent in R can be calculated by using `^x`.

😊 Great, now our values are back to animal activity. Let's see what we achieved, please visualize the relation between coverage and our newly calculated variable.

From: <http://bisfogo.environmentalinformatics-marburg.de/> - **BIS-Fogo**

Permanent link: <http://bisfogo.environmentalinformatics-marburg.de/doku.php?id=en:learning:schools:s01:worksheets:ba-ws-05-1>

Last update: **2015/09/22 15:22**

