

W04-1 Regressions

This worksheet introduces linear regressions between variables. After completing this worksheet you should know how to compute a linear regression model between two variables (in general two columns of a data frame).

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](#)

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 “W04-1.R” and use it for the programming tasks of this worksheet.

After finding some relationships between the variables let's start with regressions!

😊 Please visualize once again the relationship between animal activity and vegetation coverage using the `plot()` function.

😊 Although it is not a perfectly clear relationship, let's fit a linear model using the `lm()` function anyway.

To save yourself some typing: Write the linear model in a variable and use that variable for further proceedings with the linear modell

😊 Please summarize the result to get all the important information about the linear model. You can use the `summary()` function. To interpret these results, you might want to check the distribution of the residuals for which you can use the `qqPlot()` function.

😊 After all the theoretical stuff, let's see what our linear model looks like using the `regLine()` function

⚠ R doesn't know the function `regLine()` yet. It's one of the functions of the package `car`. Remember that there are countless of packages and functions available for R ...but... most of them, you'll never use, as there are special functions for each special field. So if you won't use it, then you don't want those functions hanging around on your computer, right? That's why you need to install packages every once in a while. There are two steps to that:

1. You need to download the function from the internet and install it on your computer! It's as easy as: `install.packages("<function name without brackets>")` Great thing about that: you only need to do that one time for each function, because it's on your computer now.

2. If you want to use the function in a script, you have to load that package you just installed. Just write: `library(<function name without squared brackets>)` Unfortunately you have to do that every time you restart that script in R. That is why normally you will load all the packages you need right before you start reading the data.

☹ After analyzing the relationship between animal activity and vegetation coverage, please check how the animal activity is related to the diversity of agricultural vegetation (i.e. number of different species per plot) and natural vegetation.

☹ After all the technical stuff let's recap what we have done. Do our linear models make sense in an ecological manner?

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-04-1>

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

