

L05: Prediction

“This is the key to a new order. This code disk means freedom.”

Tron, Tron

Things we cover in this session

- Transforming variables used in linear regression models
- Predicting variable values using linear models

Things you need for this session

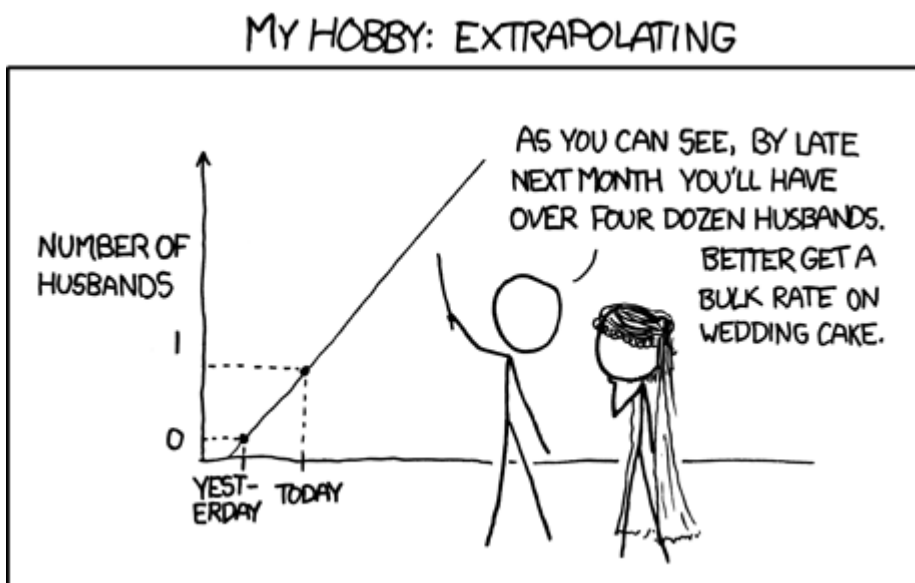
- [W05-1 Predictions](#)

Things to take home from this session

At the end of this session you should be able to

- visually check if the residuals of a linear model are normally distributed
- transform input variables of regression models if necessary
- predict values based on linear models

Prediction



Predicting data values based on a linear regression model is quite straight forward. Given that the independent variable x_i is available for position i (e.g. a research plot location), one can estimate (i.e. predict) the value of the dependent variable y_i using

$$y_i = a x_i + b$$

with a and b as the slope and intercept of the linear regression model.

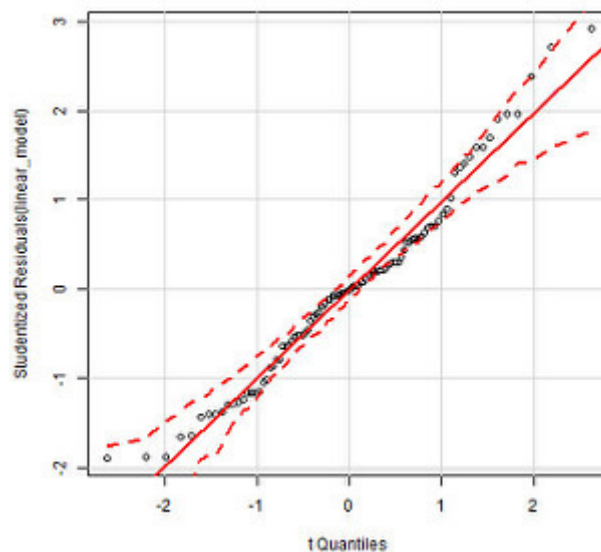
Before we have a look at that, let's revisit the coefficient of determination.

Reliability of r squared

R^2 can be interpreted as the percentage degree of how much of the variability of the dependent data can be explained by a linear regression model. But this interpretation requires that the residuals of the model are normally distributed (and also homoscedastic but this is a different story).

Hence, before using a linear model for predictions, one has to make sure that the R^2 is actually meaningful. This is especially true when no further validation analysis of the goodness of the prediction should be performed.

A simple visual test is to look at a Q-Q plot where the residuals of the individual data pairs should fall on or close to the 1:1 line.



If such a behavior is not given, it is quite likely that the input variables are not normally distributed. One possible solution might be the transformation of one or both input variables e.g. by computing its square root or more complicated stuff like the inverse sinus of the square root of the normalized variable. The latter would be a good choice for percentage coverage values (i.e. normalization transforms values between 0 and 100 to between 0 and 1 in this case).

Prediction

Computing a prediction based on a linear model is straight forward. Simply apply the linear model equation to the set of dependent variables for which the prediction of the independent variable is required.

If the input variables for the linear regression model have been transformed before, the resulting values from the prediction have to be transformed back by inverting the transformation function (i.e. square if the original transformation was square root etc.).

We came back on validating predictions in the next session.

Time for practice

W05-1 Predictions

🤖 If you need more examples, have a look at [C05-1 Predicting observations with linear models](#)

Note on data used for illustrating analysis The analysis used for illustration on this site are based on data from a field survey of areas in the Fogo natural park in 2007 by K. Mauer. For more information, please refer to [this report](#).

From:

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

Permanent link:

<http://bisfogo.environmentalinformatics-marburg.de/doku.php?id=en:learning:schools:s01:lecture-notes:ba-ln-05>

Last update: **2017/10/30 09:21**

