Tutorial 11: Diagnostic Techniques and Imputation of Missing Data

Jan Vogler (jan.vogler@duke.edu)

November 6, 2015

Today's Agenda

1. Diagnostic techniques
2. Functional form specifications
3. Imputation of missing data
4. The most useful R packages for applied work

1. Diagnostic techniques

2. Functional form specifications

3. Imputation of missing data

We often deal with missing data for some observations. Imputation allows us to make statistical inferences about the missing data values. Our guesses are based on the data that we have. For the imputation of missing data, we will use a package by Gary King called "Amelia".

install.packages("Amelia")

4. The most useful R packages for applied work

The following are some of the most useful packages for applied work. I recommend to get the related text books and download the documentations of these packages. These packages have many useful commands that can help you to deal with data management and data analysis.

1. car — Companion to Applied Regression Associated with Fox & Weisberg's book "Companion to Applied Regression" Most useful for regression diagnostics as demonstrated in this tutorial.
2. arm — Analysis of Regression and Multilevel Models Associated with Gelman & Hill's book on "Regression and Multilevel/Hierarchical Models" Most useful for simulations of regressions and plotting. (More in tutorial 13)
3. Zelig — by political science professor Gary King, see: http://zeligproject.org/ Has many different regression tools included
4. ggplot2 — introduced extensively in Chang's "R Graphics Cookbook" Allows to produce nicer graphics for visual presentation
5. stargazer — by Marek Hlavac Allows to easily generate LaTeX code of regression tables
6. reshape – by Hadley Wickham, see: http://had.co.nz/reshape/ Allows to reformat data