1) Install and load the ‘car’ package. We’re going to use the ‘Leinhardt’ dataset included in ‘car’.

Check the dataset with ‘?Leinhardt’

Drop the ‘Europe’ region from the dataset and then compare ‘Americas’ to the remaining two other regions – code it accordingly. Infant-mortality rate per 1000 live births (infant) is the dependent variable.

a) Set up the model and interpret the results

b) Now add the variable ‘oil’ (Oil-exporting country). Code it so that America and non-oil exporting countries are the reference and add interaction among predictors. Interpret the results

c) Check the model assumptions and describe. You can look at the first two plots of plot(fittedmodel)

Bonus) Find a way to plot the means (infant on y axis) across oil (oil on x axis) for the different regions (different lines).

2) Logarithmic transformations: the dataset mcdonald.dat contains mortality rates and various environmental factors from 60 U.S. metropolitan areas (see McDonald and Schwing, 1973). For this exercise we shall model mortality rate (MORT) given nitric oxides (NOX), sulfur dioxide (SOx), and hydrocarbons (HC) as inputs. This model is an extreme oversimplification as it combines all sources of mortality and does not adjust for crucial factors such as age and smoking. We use it to illustrate log transforma-tions in regression.

a) Create a scatterplot of mortality rate (y-axis) versus level of nitric oxides (x). Do you think linear regression will fit these data well? Fit the regression and evaluate a residual plot from the regression.

b) Find an appropriate transformation that will result in data more appropriate for linear regression. Fit a regression to the transformed data and evaluate the new residual plot.

c) Interpret the slope coefficient from the model you chose in (b).

d) Now fit a model predicting mortality rate using levels of nitric oxides, sulfur dioxide, and hydrocarbons as inputs. Use appropriate transformations when helpful. Plot the fitted regression model through the scatter plot (eg. you can use abline(fittedmodel) for that line) and interpret the coefficients.

e) Run diagnostic plots: qqPlot (from the ‘car’ package) and plot fitted vs. residuals. Comment on the plots.

Bonus: You would like to estimate a model with two splines – i.e., with two regression lines that meet at a given knot point. The knot can be placed at x=0. Fit and plot the regression line overlaid on the scatter plot.