Exercises – R Course

Fall 2013

Day 2: Basic models

1. Set the seed to today’s date (in MMDDYY format). Generate the following variables with for each variable and store them in a single data set:
   1. is a count variable from 1 to 30
   2. is a factor containing 3 levels corresponding to low, medium, and high

set.seed(052213)

y <- rnorm(30, 40, 4)

u <- runif(30, 0, 3)

z <- 1:30

x <- gl(3, 10, labels=c("Low", "Medium", "High"))

data1 <- data.frame(y, u, z, x)

1. Print the data from (1) and construct a plot of Y vs. U (label the axes and give an appropriate title to the graph).

data1

plot(u, y, main="N(40, 16) versus Uni(0,3)", ylab="Normal", xlab="Uniform")

1. Fit the appropriate linear model of Y on each of U, Z, X, and all first-order interactions between U, Z, and X.
   1. Print a summary of the linear model
   2. Extract the coefficients from the model
   3. Compute the residuals and the fitted values from the model, append both to the data set created in (1) and print the updated data set in the R window

lm.3 <- lm(y ~ u + z + x + u\*z + u\*x + z\*x, data= data1)

summary(lm.3)

coef(lm.3)

resid.lm3 <- residuals(lm.3)

fitted.lm3 <- fitted(lm.3)

data2 <- data.frame(data1, resid.lm3, fitted.lm3)

data2

1. Load the **frogs** data set from the **DAAG** package
   1. Print the first ten records of the data set
   2. Print a summary of each of the variables in the data set
   3. Construct a table of **pres.abs**. What kind of variable is this?

library(DAAG)

frogs[1:10,]

summary(frogs)

table(frogs$pres.abs)

1. Fit an appropriate generalized linear model to the **frogs** data set with **pres.abs** as the outcome variable and the logarithm of **distance**, logarithm of **NoOfPools**, **meanmin**, and **meanmax** as predictors
   1. Which predictors are significant at the 0.001 level of significance?
   2. ***BONUS***: Use **help()** to study the function **termplot()**. Use **termplot()** to display plots showing the contributions of the explanatory variables to the fitted values on the scale of the linear predictor. Try to display all the plots in a single graphics window.

glm.4 <- glm(pres.abs ~ log(distance) + log(NoOfPools) + meanmin + meanmax, family=binomial, data = frogs)

summary(glm.4)

par(mfrow=c(2,2))

termplot(glm.4)