BERD R Short Course
Session 5 Homework Problems

There are two problems in this homework assignment. The first problem is about linear regression and the second problem is about logistic regression.

Problem 1

The data file name is RMR.csv. Data were collected on 44 healthy women. There were 5 variables in the data set: weight, RMR (resting metabolic rate), athlete (1 for athletes and 0 otherwise), age, and height. The objective is to conduct multiple linear regression with RMR as the dependent variable and the other 4 variables as the covariates.

1. Read in the data and create a data frame called RMR.
2. Generate pairwise scatter plots among 5 variables to check for collinearity among covariates and linearity between dependent variable and the 4 covariates.
3. Fit a full model with rmr as the dependent variable, and weight, athlete, age, height, athlete\*age, as the covariates. Save the regression results to Full.
4. Perform a backward model selection on FULL to obtain a more parsimonious model, say, Final.
5. Plot studentized residuals vs. fitted values to examine linearity and constant variance assumptions
6. Plot QQ plot to examine the normality assumption.

Problem 2

The data file name is sepsis.csv. Data were collected on 106 sepsis patients. There were 5 variables in the data set:
death: patients either died or suffered from organ failures
shock: patients were in a state of shock
malnutr: patients were mal-nutritioned
alcohol: patients were alcoholic
bowelinf: patients had bowel infarction

The objective is to conduct logistic regression with death as the dependent variable and the other 4 variables as the covariates

1. Read in the data and create a data frame called Sepsis
2. Fit a full model with shock, malnutr, alcohol, bowelinf, and malnutr\*alcohol interaction. Save the regression results to Full
3. Perform a backward model selection on FULL to obtain a more parsimonious model, say, Final.
4. Produce the Hosmer-Lemeshow goodness of fit test