1. Use data 'Cars93' in R package 'MASS'. This data is for 93 types of automobiles sold in the United States in 1993. Among the various variables, MPG.highway is used as a response variable, and a regression model with EngineSize, Weight, Price, Width, Length, Horsepower, and Wheelbase as explanatory variables is constructed.

library (MASS)

Data1 <- data (Cars93)

1-1) Describe the correlation coefficient matrix and explain the relationship between the variables.

1-2) Create a regression model using all explanatory variables and find out which variables are significant at significance level α = 0.05.

1-3) Create a regression model using Forward Selection, Backward Elimination, and Stepwise Selection. In the selection process, list which explanatory variables are added or removed in order, and the AIC value and the regression coefficient of the variables when finally selected.

ex) Forward: variable 1 -> variable 1 + 3 -> ... ..

   Backward: 1 + 2 + 3 + 4 + 5 + 6 variables -> 1 + 3 + 4 + 5 + 6 variables -> ... ..

2. The data 'KBO1' is a survey of 162 players who played in more than one match in the first round of 2011. The variables are ranked in order of Y (annual salary, in thousands of won), X1 (homerun), X2 (homerun), X3 (other score), X4 (number of points), X5 (number of walks) (Career experience), club name, and player name. Y is used as a response variable, and X1 to X7 are used as explanatory variables.

Data2 <- load ("KBO1.rda")

2-1) Convert the annual salary to the base 10 log, then show the scatter plot. What explanatory variables are closely related to salary?

2-2) Create a regression model using Ridge and Lasso (training data 70%, test data 30% separation). Express a regression coefficient that is estimated according to another λ value. For optimal λ values, show the regression coefficient and explain what the difference between Ridge and Lasso is.

\* Since the response variable is continuous data, the parameter [family = "binomial"] is not used in the glmnet () function.

2-3) Create a regression model using Forward, Backward, Stepwise, Ridge, and Lasso and Elastic Net (training data 70%, test data 30% separation). Evaluate the regression model using RMSE, MAE, and MAPE. Slide Slide See page 51 to rank the performance and variability of the regression model from the first to the sixth.