Assignment #2: Linear Regression (4 pts)

Group Submission

Due: Wednesday February 27 2:00 pm.

Perform a linear regression analysis on data Assignment2.csv.

1: $Units = \beta_1 Hours + \beta_2 Lines + \beta_3 Workers + \beta_4 Region + \beta_0 + \epsilon$

2: $Units = \beta_1 Hours + \beta_2 Lines + \beta_3 Workers + \beta_4 Region + \beta_5 Region * Workers + \beta_0 + \epsilon$

In this assignment, we predict *Units* using the other variables.

a. Fit the above two models using least squares to all data. Compute AIC and adjusted R^2 for two models. Which one is a better model?
b. Write out each model in equation form, being careful to handle the qualitative variables properly.
c. Use the sample() function to split the original data into one training set with 70% of the original observations and one testing set with the rest of observations. Compute the prediction MSE associated with each model. Which model is a better one in terms of the prediction MSE?
d. Compute the ten-fold cross-validation error (MSE) associated with each model. Which model is a better one?
e. Select the "better" model as the final model. Which predictors appear to have a statistically significant relationship to the response (Units)? How does each predictor affect the response?
f. Is there evidence of outliers in the model selected from (e)? Please justify your answer.

Deliverables

1. Group submission. Each group submits one set of report and code. Please include a cover page with all team members' names.
2. Two files: R code file and the report are submitted to Blackboard.
3. The report should contain the answers to each question. No R code or raw outputs except plots. Please pay attention to the presentation of your tables and figures (if any).