DSA211 Statistical Learning with R

Homework 5

Use R functions and data file (Business.csv) to solve the following problems:

1. A financial analyst engaged in business valuation obtained financial data on 72 drug companies to predict the price-to-book value. The Business.csv contains the following variables:

   PB Ratio: Price-to-book value ratio
   ROE: Return on equity (in %)
   Growth: Growth rate
   Local: Whether a local or international company (Yes or No)
   Size: Size of company (Small, Medium or Large)

   (a) Find the multiple regression equation with all the four main independent variables.

   (b) Determine whether there is a significant linear relationship between the price-to-book value ratio and the four independent variables at the 0.05 level of significance.

   (c) Determine whether each independent variable makes a significant contribution to the regression model at the 0.05 level of significance.

   (d) Find the multiple regression equation with only two main independent variables: ROE and Local.

   (e) Find the multiple regression equation with only two main independent variables: ROE and Local, with their interaction effect.

   (f) When comparing three estimated models in parts (a), (d) and (e), explain the reason why model in part (e) is the best.

   (g) Perform a residual analysis on the model in part (e) and determine whether the regression assumptions are valid. (Note that there is no need to plot the interaction effect versus residuals as Local variable is a qualitative variable.)

   (h) Predict the price-to-book value ratio if a local company has the ROE of 24.1% under the model in part (e).

   (i) Construct a 95% prediction interval estimate for the the price-to-book value ratio of an international company if it has the ROE of 24.1% under the model in part (e).

   (j) Based on the model in part (e), write down the simple linear regression equation for predicting all local companies' price-to-book value ratio.

-END-