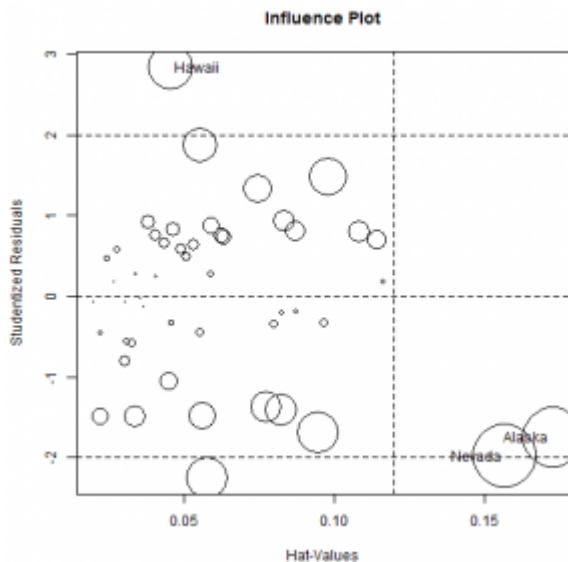


Multiple Regression (Part 2)

– Diagnostics



Multiple Regression is one of the most widely used methods in statistical modelling. However, despite its many benefits, it is oftentimes used without checking the underlying assumptions. This can lead to results which can be misleading or even completely wrong. Therefore, applying diagnostics to detect any strong violations of the assumptions is important. In the exercises

below we cover some material on multiple regression diagnostics in R.

Answers to the exercises are available [here](#).

If you obtain a different (correct) answer than those listed on the solutions page, please feel free to post your answer as a comment on that page.

Multiple Regression (Part 1) can be found [here](#).

We will be using the dataset `state.x77`, which is part of the state datasets available in R. (Additional information about the dataset can be obtained by running `help(state.x77)`.)

Exercise 1

- Load the state datasets.
- Convert the `state.x77` dataset to a dataframe.
- Rename the `Life Exp` variable to `Life.Exp`, and `HS Grad` to `HS.Grad`. (This avoids problems with referring to these variables when specifying a model.)
- Produce the correlation matrix.

e. Create a scatterplot matrix for the variables Life.Exp, HS.Grad, Murder, and Frost.

Exercise 2

- a. Fit the model with Life.Exp as dependent variable, and HS.Grad and Murder as predictors.
- b. Obtain the residuals.
- c. Obtain the fitted values.

Exercise 3

- a. Create a residual plot (residuals vs. fitted values).
- b. Create the same residual plot using the plot command on the lm object from Exercise 2.



Learn more about multiple linear regression in the online courses [Linear regression in R for Data Scientists](#), [Statistics with R – advanced level](#), and [Linear Regression and Modeling](#).

Exercise 4

Create plots of the residuals vs. each of the predictor variables.

Exercise 5

- a. Create a Normality plot.
- b. Create the same plot using the plot command on the lm object from Exercise 2.

Exercise 6

- a. Obtain the studentized residuals.
- b. Does there appear to be any outliers?

Exercise 7

- a. Obtain the leverage value for each observation and plot them.
- b. Obtain the conventional threshold for leverage values. Are any observations influential?

Exercise 8

- a. Obtain DFFITS values.

- b. Obtain the conventional threshold. Are any observations influential?
- c. Obtain DFBETAS values.
- d. Obtain the conventional threshold. Are any observations influential?

Exercise 9

- a. Obtain Cook's distance values and plot them.
- b. Obtain the same plot using the plot command on the lm object from Exercise 2.
- c. Obtain the threshold value. Are any observations influential?

Exercise 10

Create the Influence Plot using a function from the car package.