Second Assignment

by PsychoStatisticia

For the second assignment, you would need to use two datasets: *GermanCredit* and *Hitters*. You don't need to download these 2 datasets by yourself because they are embedded in two R packages: *caret* and *ISLR*, respectively.

You can first install the packages before loading them if you don't have them:

```
install.packages("caret")
install.packages("ISLR")
library("caret")
library("ISLR")
```

Now, you can load the 2 datasets by executing the following code:

```
data(GermanCredit)
data(Hitters)
```

You should be able to see the datasets in your R studio now! You may be interested in the meaning of each variable/column name of the datasets. By code below, you can simply get access to the information page of the datasets:

```
?GermanCredit
?Hitters
```

Please read the information pages carefully and choose one dataset to perform the regression analysis on. This choice can be made completely subjectively (i.e which dataset you prefer to dealing with), so there is no good or bad choice.

Once you have chosen the dataset you prefer, you can follow the following steps to start your regression analysis:

1. Please choose 2 variables (1 outcome variable and 1 continuous predictor). The relationship between those 2 variables will be the primary interest throughout the analysis. Also remember to clearly state your research question!

2. Please run a complete simple linear regression to investigate the relationship between the outcome and the predictor you chose. Ask yourself which components are essential to define a complete simple linear regression.

3. After the simple linear regression analysis, please choose 4-5 additional predictors (at least 1 categorical predictor) that may influence the primary relationship you have investigated. Please briefly motivate your choices.

4. Fit different possible models on the set of predictors you chose to reach the decision of "final" model. You may use different ways to reach that "final" model by making use of the knowledge you acquired in lectures.

5. Interpret your final model properly. Also, don't forget to test model assumptions.

6. (Challenging) Try to perform model diagnostics we mentioned in last lecture. For example, outliers detection, etc.

7. Based on your interpretation, try to make a conclusion against your research question. Please carefully pay attention to the model assumption violations and discuss the restriction of your analysis explicitly!