Regression in R - STA 521

Abbas Zaidi

1 Agenda

1. Producing LaTeX tables in R
2. Regression Diagnostics
3. Confidence and Prediction Intervals
4. Implementations in R
5. Longley Data in R
6. Readings

2 Lab Tasks

1. Your goal is to build a regression model for Gross National Product (GNP) based on two input variables: number of people employed and the total population using the longley data in R. Start by performing graphical exploratory data analysis: create univariate density estimates and scatterplots to understand the bivariate features of the data. Do you see anything interesting?
2. Build the regression model, and provide the coefficients of the model and details on their significance using xtable. Interpret your coefficients. Is the intercept meaningful? What can you do to make the intercept more meaningful?
3. Perform regression diagnostics via graphical methods: Assess normality of your residuals, constant variance, independence as well any potentially influential points using Cook's Distance with a threshold value of $\frac{4}{n}$. Be sure to detail what you see. Do you need to transform your data? If you were to transform your data, how would it impact the interpretation of your model?
4. Create a plot of Population against GNP that shows the fitted regression line holding Employment at its mean value. Add prediction and confidence intervals to your plot based on the same assumption in different colors. Where are the intervals narrowest? What do you expect will happened to the intervals as $n \to \infty$.

3 Directions

In general for Labs, at the top of any file you are asked to submit, please list the following:

1. First Name Last Name
2. Lab Date
3. Team Member(s)

With respect to any item for which you are asked to generate any output, please provide the actual R output as a part of your solution and any explanation needed as well. For any functions/ computations that you will write, please list the following as comments before the step in R:

1. Task number and descriptions.
2. Input(s) with descriptions.
3. Outputs(s) with descriptions.
4. Function/ output summary (along with intermediate step comments).

For Lab 6, please provide the following deliverable items:

1. Please provide your solutions using Markdown as a .pdf with the following naming convention: LastName_FirstName_Solutions_Lab6.pdf.
2. Provide your .Rmd file (this **MUST** compile) for the lab using the following naming convention: LastName_FirstName_Solutions_Lab6.Rmd