Week 3 Lab: Lab 1 Followup and Lab 2

Tom Hanna

University of Houston

tlhanna@uh.edu

February 1, 2021

Overview

Lab 1 Stats Question 1

Question

...the criteria to decide to run one or another test [of influence] to analyze the data. I guess that experience analyzing different models will give me a better understanding of what do I have to choose...

Answer

Now: Ease of interpretation. Later: peculiarities with the model or the data. Overall: Yes, experience matters which is part of the object.

Lab 1 Stats Question 2

Question

Why do we take the natural log instead of the log in the linear regression

- Mathematical convention: natural log is the default. Natural it uses the base e. e has unique qualities that make it as valuable as π or Einstein's c.
- Regression: Log transform X and Y, we can express things as percentage changes: A 1% percent change in X_1 results in a β_1 change y."
- Useful reference: http://www.cazaar.com/ta/econ113/interpreting-beta

Dummy variable significance

Yes

The dummy variable had a significant effect with a p-value lower than .05. I probably should have specified statistical significance and a specific p-value, but .05 is our default in social and life sciences.

Codebooks

Before you start any computations, you should always look at the **Codebook** or description for the dataset. In this case:

CaliforniaTestScores.pdf

It's all about regression

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{(n-1)SD(x)SD(y)}$$

Figure: Correlation coefficient calculation

$$b = \frac{\Sigma xy - n\bar{x}\bar{y}}{(n-1)SD(x)^2}$$

Figure: Regression calculation

Homework stats questions

Questions

Questions?

Lab 1 R Question 1

Question

1. Here is the following error that I attempted to resolve. ¡/p¿ ¡br¿ ¡p¿> plot(data$Density, data$Deaths.per.100k, pch=19)Error in data$Density : object of type 'closure' is not subsettable

Answer

I checked the object type of data$Density using the typeof() command.
typeof(data$Density)
I got type "double," which is correct, and the code worked fine.
So...

R Question 1: Cont....

- View(data) #capital V on View
- If you get an error or something looks wrong
- Change the dataframe name to something like "data2"
- You'll need to change it everywhere in the script that you see a reference to the original name ("data")

R Question 1: Cont...

1. Restart R Studio and DO NOT SAVE on exit
2. Start a new R session once R Studio is restarted
3. Add the following line
   rm(list=ls())
   just ahead of
   data ¡- read.csv("accidents50.csv")
4. Rerun script from the beginning including the rm(list...) command.
5. Try restarting everything completely with a new copy of the script and data file.

The End