Final project

Lewis Kouassi

Wednesday, Dec. 14, 2022

Introduction

I was working with the USA-arrests data and this data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. The reason why I chose this day is I want to know the relationship between Murder , Assault , Urban Pop and Rate , by doing so I was trying to run some multiple regression to answer the question. But before that I run a couple graphs , summary statistics , and plots to see how the data look like.

Approach

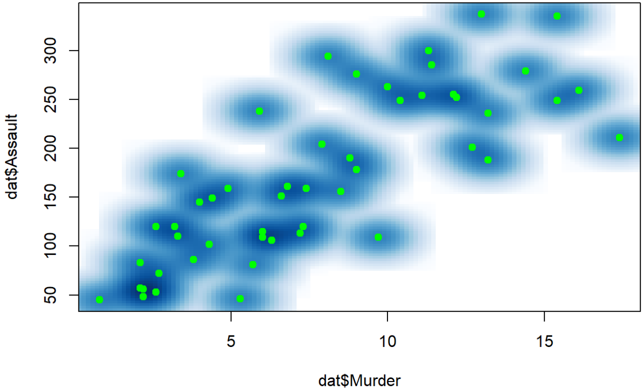
First of all I did some data exploration by using some commands in the software , then I use some data visualization , then finally data summary . For the final project I use R software especially R studio which is an integrated development environment for R, a programming language for statistical computing and graphics. The reason why I chose R , and R markdown is because RStudio is designed to make it easy to write scripts, RStudio makes it convenient to view and interact with the objects stored in your environment. and finally in R markdown it makes easy to knit to HTML , PDF documents and Words .RStudio is an integrated development environment for R, a programming language for statistical computing and graphics.

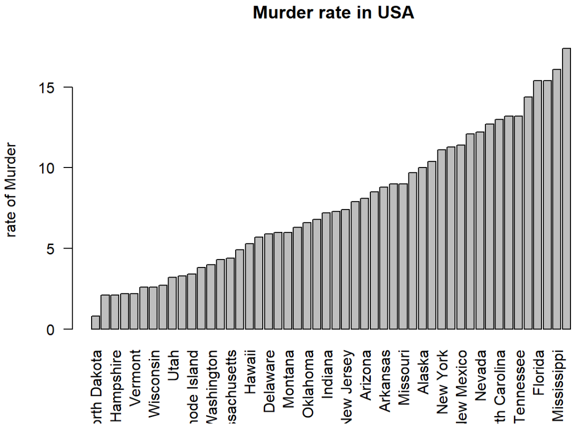
Maths and Graphs

The multiple regression equation

$$Murder = \beta_0 + \beta_1 Assault + \beta_2 UrbanPop + \beta_3 Rape.$$

$$Murder = -0.612 + 0.064 Assault - 0.067 UrbanPop + 0.347 Rape.$$




Murder rate in USA
rate of Murder
0
5
10
15
orth Dakota
Hampshire
Vermont
Wisconsin
Utah
node Island
Washington
ssachusetts
Hawaii
Delaware
Montana
Oklahoma
Indiana
New Jersey
Arizona
Arkansas
Missouri
Alaska
New York
lew Mexico
Nevada
th Carolina
Tennessee
Florida
Mississippi

Code in Rmardown

```
View(USArrests)
head(USArrests)
str(USArrests)
mean(USArrests$Murder)
smoothScatter(dat$Murder , dat$Assault , pch =19  ,
col = "green")
state.names = row.names(dat)
barplot(dat$Murder, names.arg = state.names , las = 2 ,
ylab = " rate of Murder " ,main = " Murder rate in USA")
fit<-lm(Murder˜Assault + UrbanPop + Rape + Assault*Rape ,
data = dat)
fit
summary(fit)
```

Results and Conclusion

Residual standard error: 2.461 on 45 degrees of freedom Multiple R-squared: 0.7069, Adjusted R-squared: 0.6809 F-statistic: 27.13 on 4 and 45 DF, p-value: 1.719e-11 When Assault,Urban Pop and Rape = 0 , Murder = -0.612 . R-squared: 0.7069 means the proportion of the variance in the Murder can be explained by Assault , Urban Pop and Rape , in other words 0.7069 of the variance of the model is explained by the predictor variable in the model.The Adjusted R-squared: 0.6809 Adjusted R-squared tells us how well a set of predictor variables is able to explain the variation in the response variable, adjusted for the number of predictors in a model. And P =1.719e-11 ¡ 0.05 tell us there is no significant different between variables.

Other perspective

In this course I have learned a lot especially in terms of software (Latex , R and sage) but also learned how to work with groups of people at school . I learned to feel comfortable with software and also how to write some simple line of code in different languages. As far as concerned my final project what I would done differently is in terms of my multiple regression analysis I would done different interaction between predictor variables and see how those variables interact between each other , calculate each F-test , and t-test and draw conclusion about it.