informative title

Clara le Draoulec, Maeva Kameni

March 22, 2022

Introduction:

quick background presentation + scientific question define all the variable:

CITATION: The first variable considered is the degree of unionization of the manufacturing labor force. Second, a dummy variable is used for the classifications right-to-work states and union- shop states. The third variable is a structural factor indicating the importance of relatively non-unionized industries in the economy of a given state. The fourth variable is the size of the agricultural labor force relative to the to- tal labor force.

Explanation of the variables in the text ## Pre-processing: During this step, we will open the data file and make sur that every elements contained is properly set and that the data variable is ready to be used for further analysis.

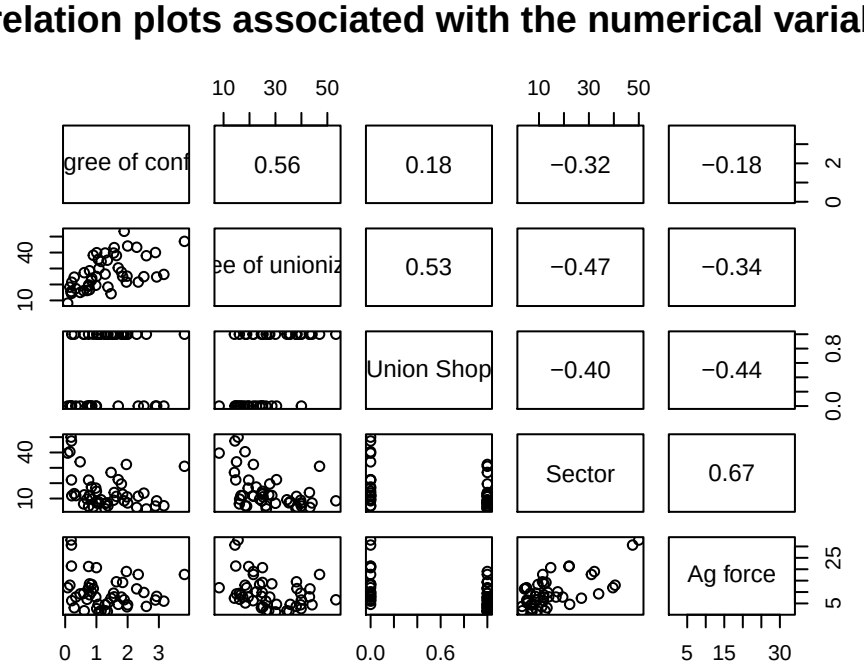
Exploratory Data Analysis:

Since there are some outliers, we will try to fit the model with or without them to access if these are influential points. (slide 68 Lecture 2)

For the dummy variable, a barplot is preferred to assess the distribution

Now we will analysis the distribution of the differents variable factors according to our groups namely Right-to-work states and Union-shop States

Let's see how the variables are linked to each other before beginning the model fitting. It could help us make sure that the model fits properly afterwards.



Statisitics and model fitting:

First step:

- The first step in the analysis is a direct comparison of the observed levels of industrial conflict in the right-to-work stated and union-shop states for the period 1957-1962
- the measure of industrial conflict is the proportion of time lost due to work stoppages (man hours relative to total estimated working time)

(RESULT PAPER:

AVERAGE PER YEAR: right-to-work : 0.19% // non-right to work:0.24% –> significative at 5%

AVERAGE TOTAL MAN HOURS LOST: right to work: 267,151 hours per year // union-shop : 712,603 hours per year –> significative –> causative factors: conceivably be favtor other than the existaence or non existence of a right to work law)

–> test the statements (better test t-test independant ? Paired ?) :

- Agriculture > industry in right to work whereas industry > agriculture.

" Where the relatively non-unionized agricultural sector is of major importance in the economy of a state, the theoretical level of industrial conflict should be less than where this is not the case".

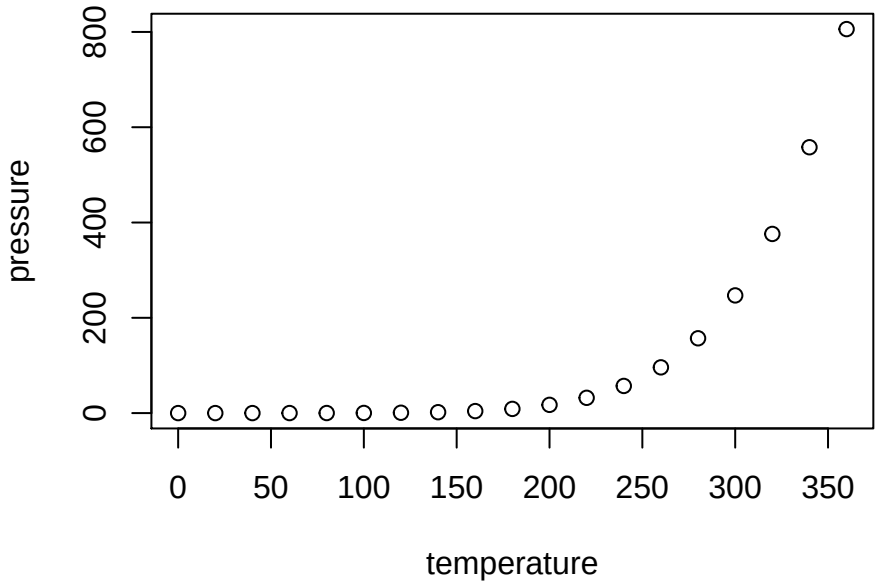
Second step:

- Try to explain the outcome variable using the variable factors (variance explained R^2 maybe adj-R^2) –> multiple linear regression (2 digits and ^ on estimates) –>state the assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

Model Assessment:

- State how the model fitted (Ordinary Least square) degree_of_conflict ~ remaining variable factors
- check the assumptions –> qqplot of the residuals and residuals vs. fitted



#rmarkdown::render('R2-XX.Rmd', 'pdf_document')