Autocorrelation

Srikar

2021-11-11

Introduction:

Autocorrelation refers to the degree of correlation between the values of the same variables across different observations in the data
For example, one might expect the air temperature on the 1st day of the month to be more similar to the temperature on the 2nd day compared to the 31st day. If the temperature values that occurred closer together in time are, in fact, more similar than the temperature values that occurred farther apart in time,
the data would be autocorrelated.Autocorrelation might mostly refers to either autocorrelation in errors, or also more generally to time series models where variables are related to their past realizations.
When it comes to autocorrelation in error term then it means that different errors are correlated with each other (usually across time but spatial autocorrelation can sometiems exists as well).

The differene between multicolinearity and autocorrelation are that
Autocorrelation refers to a correlation between the values of an independent variable, while multicollinearity refers to a correlation between two or more independent variables.

Aim:

i) To fit a regression model of data
ii) To check the assumption of no autocorrelation

Data Description:

The dataset contains the information about the purity of oxygen produced by a fractionation process is thought to be related to the percentages of hydrocarbons in the main condenser of the processing unit. The dataset contains 20 observations sampled from a population. Here the two variables are Purity (given in %) and Hydrocarbon (given in %). The independent varibale (X) is hydrocarbon which is thought to be effecting the production of pure oxygen.

```
library(readxl)
data <- read_excel("C:/Users/Srikar/Desktop/SS/R/Sem 5/Linear
Regression/Practical 10/1940834_Practical 10.xlsx")
head(data)
```

```
## # A tibble: 6 x 2
##   Purity Hydrocarbon
##    <dbl>       <dbl>
## 1   86.9        1.02
## 2   89.8        1.11
## 3   90.3        1.43
## 4   86.3        1.11
## 5   92.6        1.01
## 6   87.3        0.95
```

Hypothesis Statement

Null Hypothesis- Errors are not autocorrelated
Alternate Hypothesis- Errors are autocorrelated

Procedure:

1)Constructing the regression Model

```
mod=lm(data$Purity~data$Hydrocarbon)
summary(mod)
```

```
##
## Call:
## lm(formula = data$Purity ~ data$Hydrocarbon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6724 -3.2113 -0.0626  2.5783  7.3037
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        77.863      4.199  18.544 3.54e-13 ***
## data$Hydrocarbon   11.801      3.485   3.386  0.00329 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.597 on 18 degrees of freedom
## Multiple R-squared:  0.3891, Adjusted R-squared:  0.3552
## F-statistic: 11.47 on 1 and 18 DF,  p-value: 0.003291
```

We obtain the model :

Y = 77.863+ 11.801*X where

Y is the purity of oxygen obtained and X is the amount of hydrocarbons present.

We observe that the intecept p-value is below the significance value (0.05) and hence we can say that the intecept is significant in the prediction.This means that if the values of the regressors were all zero, the intercept would predict the average purity of oxygen if there were no hydrocarbon present at all

The p-values of the regressor Hydrocarbons shows significance as its below 0.05. That means these variables describe the linear relationship with the independent variable. Also the overall p-value is also lesser than significane level (0.05) and hence we can say that the model is significant.

The R-squared value is 0.3891 which means that only 38.91% of the variation of the Y variable (purity of oxygen) is explained by X variable (Hydrocarbon)

2)Checking the assumptions of autocorrelation

```
require(stats)

#Using the Durbin-Watson test to check for autocorrelation for errors

library(lmtest)

dwtest(data$Purity~data$Hydrocarbon)
```

```
##
##  Durbin-Watson test
##
## data:  data$Purity ~ data$Hydrocarbon
## DW = 1.9108, p-value = 0.3665
## alternative hypothesis: true autocorrelation is greater than 0
```

#Since the p-value is greater than 0.05 (significance level), we accept the null hypothesis and say that there is no autocorrelation between the errors or the residuals.

Conclsuion:

1) We obtain the regression model : # Y = 77.863+ 11.801*X where

Y is the purity of oxygen obtained and X is the amount of hydrocarbons present. The model is not ideal fit as the independant variable hardly
#explains the depedant variable.

2) The assumption of no autocorrealtion of errors has been fulfilled. This means that we can proceed further with the experiment