Heteroscedasticity

Srikar

2021-11-24

Introduction:

Heteroscedasticity means unequal scatter. In regression analysis, we talk about heteroscedasticity in the context of the residuals or error term. Specifically, heteroscedasticity is a systematic change in the spread of the residuals over the range of measured values. Heteroscedasticity is a problem because ordinary least squares (OLS) regression assumes that all residuals are drawn from a population that has a constant variance (homoscedasticity).

Aim: To plot the residuals and comment on the assumption of homoscedasticity. Also test for heteroscedasticity using Breusch Pagan test.

Data Description:

#The following table shows the annual consumption and disposable income for 30 households in India. The independent variable is income denoted by variable
X and the dependent variable is expenditure denoted by X. Since there is only one independent variable, we build a simple regression model.

```
library(readxl)

data=read_excel("C:/Users/Srikar/Desktop/SS/R/Sem 5/Linear
Regression/Practical 11/data.xlsx")
head(data)
```

```
## # A tibble: 6 x 3
##   `Sl No` Expenditure Income
##     <dbl>       <dbl>  <dbl>
## 1       1       10600  11000
## 2       2       11400  12000
## 3       3       12300  13000
## 4       4       13000  14000
## 5       5       13800  15000
## 6       6       13900  16000
```

Procedure:

```
#) Building the regression model

mod=lm(data$Expenditure~data$Income,data=data)
summary(mod)

##
## Call:
## lm(formula = data$Expenditure ~ data$Income, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1252.74  -448.90   -46.99   452.05  1548.22
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.234e+03  7.260e+02   4.454 0.000132 ***
## data$Income 7.010e-01  4.655e-02  15.060 1.17e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 700.6 on 27 degrees of freedom
## Multiple R-squared:  0.8936, Adjusted R-squared:  0.8897
## F-statistic: 226.8 on 1 and 27 DF,  p-value: 1.169e-14
```

We observe that Intercept is significant as its p-value is 0.05, the significance level. This means that if income was 0, the average expenditure will be intercept value. The regressor, income is also significant which means that income explains the variation of expenditure.

#2)Performing Breusch-Pagan Test.

```
library(lmtest)

## Warning: package 'lmtest' was built under R version 3.6.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
bptest(mod)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod
## BP = 7.9545, df = 1, p-value = 0.004797
```

#Since the p-value is less than 0.05, we accept the null hypothesis.We have sufficient evidence to say that heteroscedasticity is present in the regression model.

Conclusion:

The assumption of heteroscedasticity is