Mileage vs Transmission type : Regression Models(Coursera)

Debjyoti Paul

Sunday, December 21, 2014

Executive Summary

In this project, the dependency of mileage on Transmission (automatic or manual) in an automobile is analyzed using **mtcars** data https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html. By fitting a regression model, inference is drawn on the influence on mileage due to transmission mode.

Analysis

mtcars data frame is loaded and the data in **am** variable is stored as a logical **autm** variable.

```
data( mtcars )
mtcars$autm <- mtcars$am == 0
attach( mtcars )
```

Some exploratory data analysis on the distribution of the automatic variable is depicted in (Appendix).

Initially the data is analzed with the Ordinary Least Squares (OLS) model.

```
ls <- lm( mpg ~ autm, data=mtcars )
summary( ls )
```

```
##
## Call:
## lm(formula = mpg ~ autm, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.392      1.360  17.941  < 2e-16 ***
## autmTRUE      -7.245      1.764  -4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

The estimation clearly indicates a strong negative influence on mileage due to automatic transmission.

The model is estimated using **Weighted Least Squares** (WLS), using the weight (**wt**) varible to facilitate better fitting.

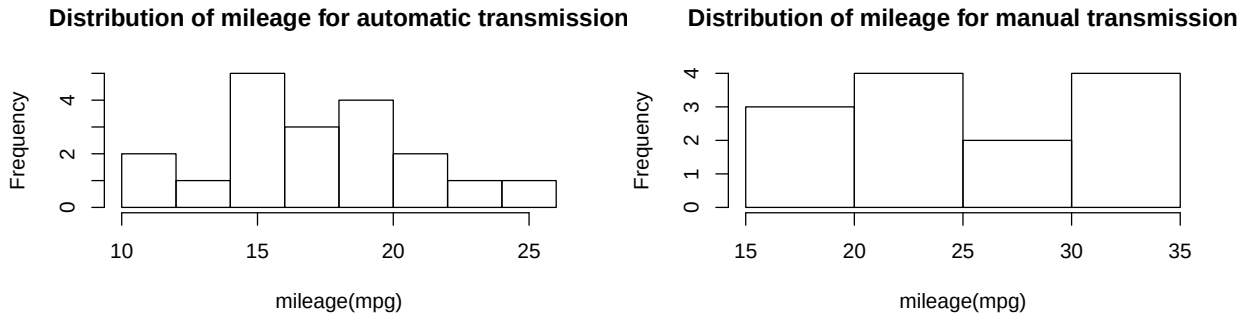
```r
wls = lm (mpg ~ autm, data=mtcars,weights=(1/mtcars$wt^2))

summary(wls)
```

```
##
## Call:
## lm(formula = mpg ~ autm, data = mtcars, weights = (1/mtcars$wt^2))
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4984 -1.4391 -0.4219  0.4841  3.8202
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   26.890      1.090  24.678  < 2e-16 ***
## autmTRUE      -8.748      1.829  -4.782 4.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.79 on 30 degrees of freedom
## Multiple R-squared:  0.4326, Adjusted R-squared:  0.4136
## F-statistic: 22.87 on 1 and 30 DF,  p-value: 4.308e-05
```

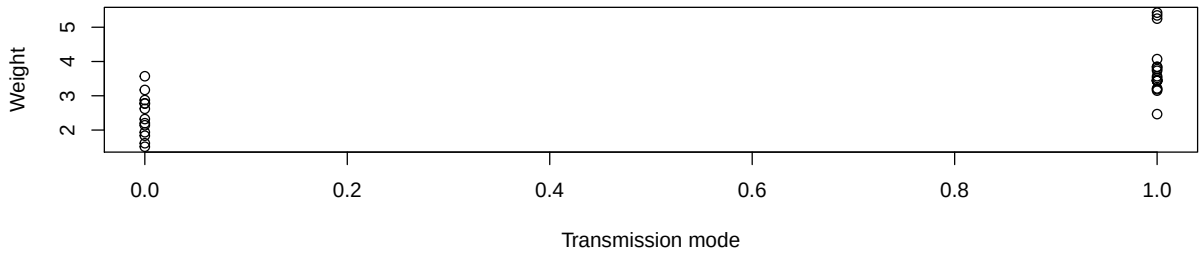
The summary depicts a stronger negative influence and the coeffecients are more significant than the that with Ordinary Least Square. The residuals for both estimations is plotted in (Appendix).

Appendix: Figures

```r
par( mfrow=c(1,2) )
hist( mpg[which(autm == TRUE)] , main="Distribution of mileage for automatic transmission",xlab="mileag
hist( mpg[which(autm == FALSE)],main="Distribution of mileage for manual transmission",xlab="mileage(mp
```



There is a strong correlation between mode of transmission and the weight of the automible. This dependence can be exploited as wieghts to yeild better results.



The residuals of the Ordinary Least Square and Weighted Least Sqaure estimations.

```
par( mfrow=c(1,2) )
hist( ls$residuals , main="Distribution of residuals of ordinary least square",xlab="residuals" )
hist(wls$residuals , main="Distribution of residuals of weighted least square",xlab="residuals")
```