Regression Models - Motor Trend Project

Sarah Elsayed

Executive summary

In this project we will use Motor Trend "mtcars" data set. Explore the relationship between a set of variables and miles per gallon (MPG) (outcome). This project will answer the following questions "Is an automatic or manual transmission better for MPG?" "Quantify the MPG difference between automatic and manual transmissions?"

```
library(ggplot2)
data(mtcars)
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual")) #give meaningful name to the transmissi
# Exploratory Analysis
```

Firt to understand the relation between the miles per gallon (MPG) and automatic or manual transmission we will do some exploratory analysis. This can be done by scatter plot or box plot as in Appendix plot 1,2. These plots shows definite impact on MPG by the transmission type, where the manual transmission (type 1) have a greater miles per gallon than automatic transmission (type 0).

Linear Regression Model

We have visually seen that manual transmission is better for MPG, now we will quantify this difference by calculating the mean MPG values for cars with both transmission types using regression model.

```
fit<-lm(mpg~ am,data=mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, 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)   17.147      1.125  15.247 1.13e-15 ***
## amManual       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
```

It is noticed the the Manual transmission has on average 7 MPG more than Automatic. Also the R squared value explanes that this model can cover around 33% of the variance. The P-Value is < 0.05 which means that the transmission type is highly participating, and we will not reject the reject the hypothesis

Multivariable Regression Model

As the linear regression model with MPG as a response and am as a predictor covers only 33% of the variance, we will include other predictors to enhance the result.

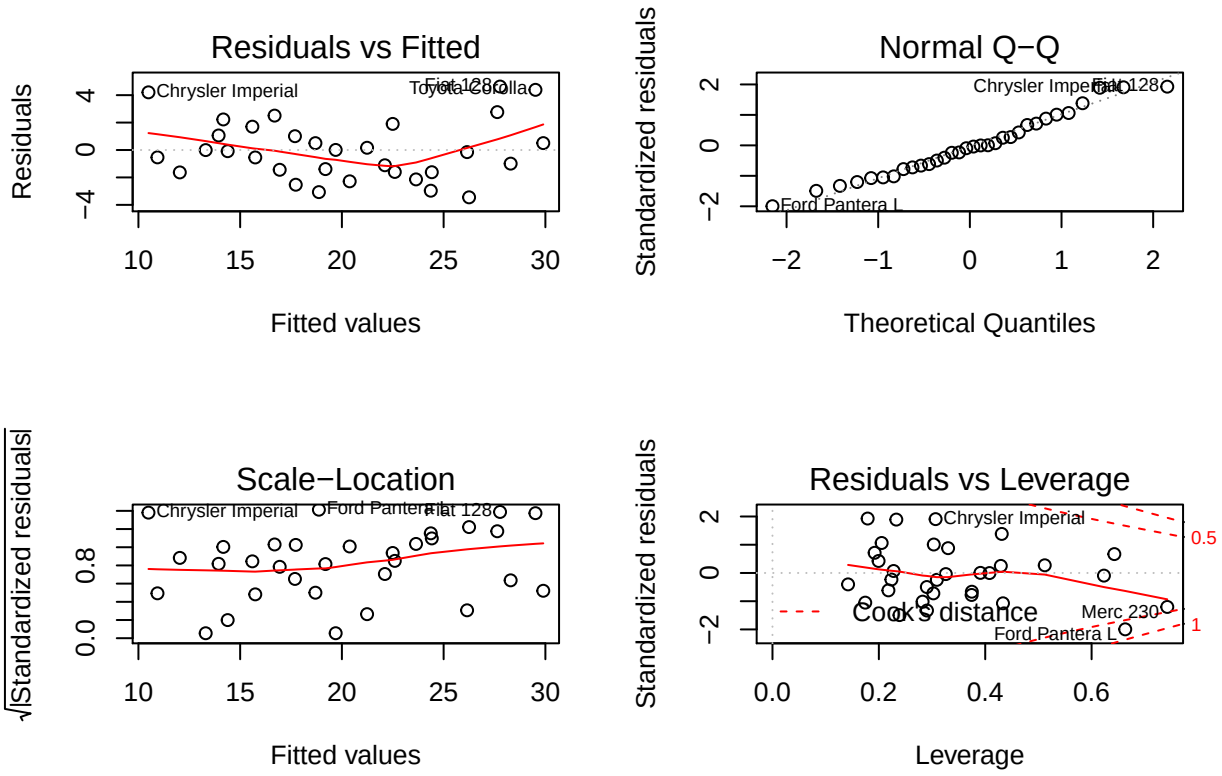
```
fit2<-lm(mpg~ .,data=mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## amManual     2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

now the R squared value explanes that the Multivariable model can cover around 80% of the variance, and still we can reject the null hypothesis.

Residual plot

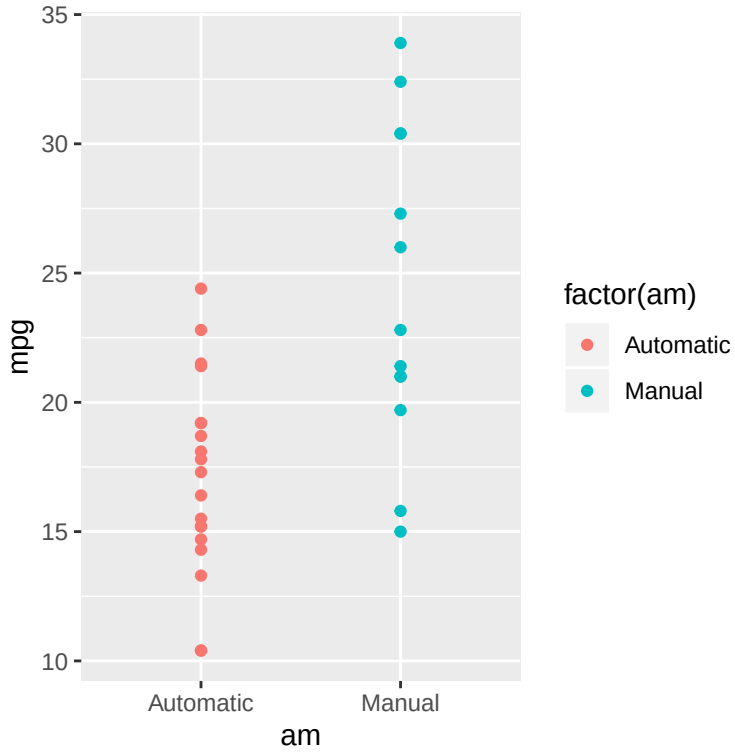
```
par(mfrow = c(2,2))
plot(fit2)
```



Appendix

Plot 1,2

```
ggplot(aes(x=am, y = mpg), data = mtcars) +
       geom_point(aes(color = factor(am)))
```




```
boxplot(mpg ~ am, data = mtcars,ylab = "Miles Per Gallon", xlab = "Transmission Type")
```