Regression Model

kuriboh

9/27/2020

This chunk loads the libraries used in this report

```
library(ggplot2)
library(datasets)
```

Executive summary

This report analyzes the mtcars database. The main purpose is to see if there is any direct relationship between mpg and transmission.

First analysis

- Skewing checking
  Exploring the dataset, I want to check if the data is skewed on the am variable. If these is any on the either class, the result might be not correct.

```
data(mtcars)
table(mtcars$am)
```

```
##
## 0 1
## 19 13
```

There are 19 autos and 13 manuals, there is no skews.

- Exploring correlations
  As the following step, I checked the correlation between mpg and the other variables.

```
cor(mtcars$mpg, mtcars)
```

```
##        mpg       cyl       disp        hp      drat        wt      qsec
## [1,]     1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##              vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

We all know that the heavier cars consume more fuel than the light ones. And according to the above result, the wt (weight) has the "nearest" correlation with the mpg, and followed by cyl (cylinders), this really makes sense. For the convenience, I will name this observation (1)
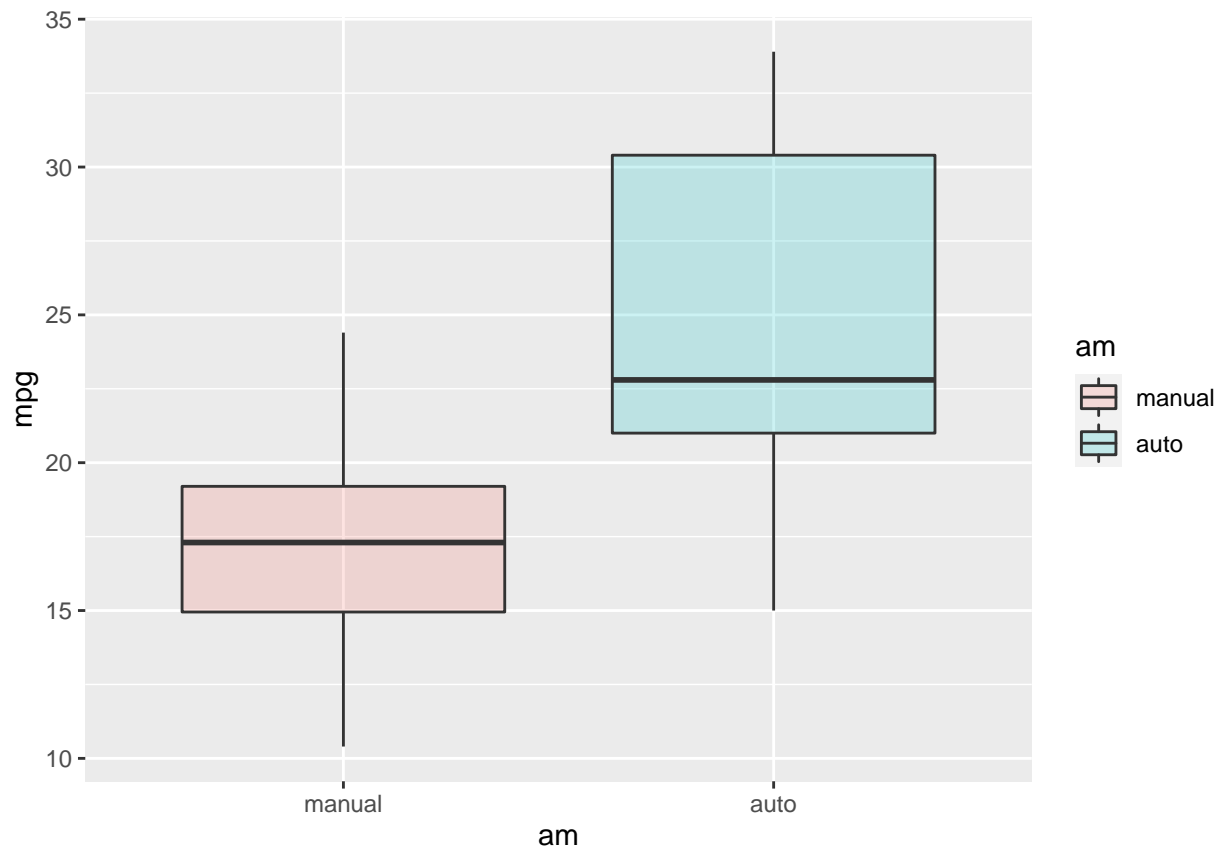
- For better visualization, I will name the tranmission types

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("manual", "auto")
```

Trying models

At first, I will plot the mpg by the type of transmission.

```
ggplot(mtcars, aes(x=am, y=mpg, fill=am)) + geom_boxplot(alpha=0.2)
```


We can easily see that the manual cars have better mpg than automatic cars. Now comes to the models to verify our analysis.

- First model
  To verify, I will use a simple linear model.

```
model_1 <- lm(mtcars$mpg ~ mtcars$am)
summary(model_1)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$am)
##
## 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 ***
## mtcars$amauto    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
```

This model shows that manual cars has 7.245mpg better than automatic ones.

** Second model
Looks back at the observation (1), I want to check if the cylinders affect mpg.

```r
model_2 <- lm(mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt + mtcars$cyl)
summary(model_2)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt +
##     mtcars$cyl)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            39.2414     3.7219  10.543 2.96e-11 ***
## as.numeric(mtcars$am)   0.1765     1.3045   0.135  0.89334
## mtcars$wt              -3.1251     0.9109  -3.431  0.00189 **
## mtcars$cyl             -1.5102     0.4223  -3.576  0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

The weight affects the mpg in 3.12mpg per weight unit, the number of cylinders also has something to say but the transmission has a very low impact.

- Third model
  Let's look at observation (1) again, I will check if there is any correlation between weights and transmission types.
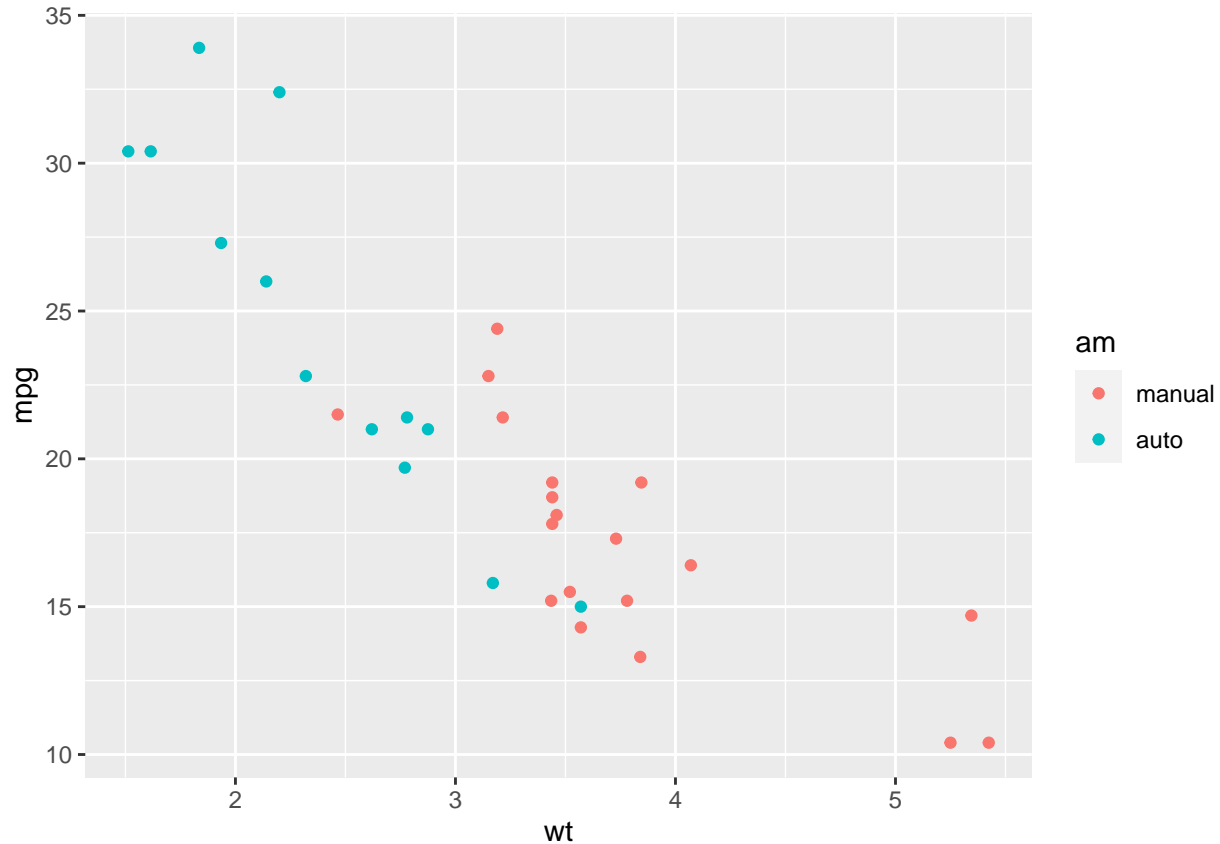
```r
cor(as.numeric(mtcars$am), mtcars$wt)
```

```
## [1] -0.6924953
```

We can see there is a strong correlation between weight and transmission types. The manual cars are lighter than the automatic ones.

Now lets see on a plot.

```r
ggplot(mtcars, aes(x=wt, y=mpg, color=am)) + geom_point()
```


From this, I want to try to use another model, similar to the previous ones, but with both weight and transmission types as predictor for mpg.

```r
model_3 <- lm(mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt)
summary(model_3)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           37.34517    4.36918   8.547 2.04e-09 ***
## as.numeric(mtcars$am) -0.02362    1.54565  -0.015    0.988
## mtcars$wt             -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

Weight influences -5.35 over mpg but transmission is neutral (0.02).

The last step, comparing the three models:

```r
anova(model_1, model_2, model_3)
```

```
## Analysis of Variance Table
##
## Model 1: mtcars$mpg ~ mtcars$am
## Model 2: mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt + mtcars$cyl
## Model 3: mtcars$mpg ~ as.numeric(mtcars$am) + mtcars$wt
##   Res.Df    RSS Df Sum of Sq      F     Pr(>F)
## 1     30 720.90
## 2     28 191.05  2    529.85 38.828 8.428e-09 ***
## 3     29 278.32 -1    -87.27 12.791  0.001292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe that the third model is the best among the threes.

Residual diagnosing

Model 3: mpg ~ wt

```r
par(mfrow=c(2, 2))
plot(model_3)
```
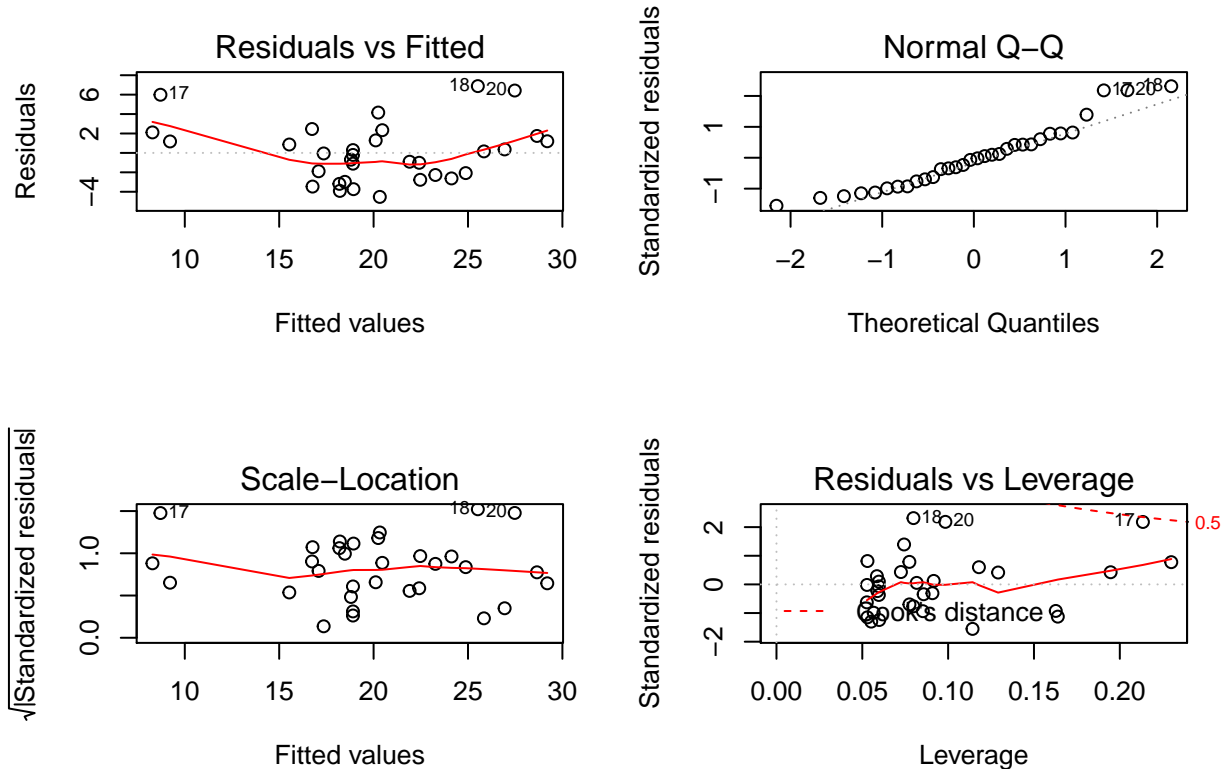
We can see that there's a good fit for the residuals, even considering some outliers, I am confident that our model is accurate.

Conclusion

- We cannot conclude that the cars' type of transmission affects the fuel consumption.

- The lighter cars have better mpg than heavier ones.

- Automatic transmission cars tend to be heavier than manual cars, and that's the reason why they show lower mpg (worse fuel consumption).

- I am wondering if there is any manual transmission Lamborghini car for me to buy?