

Regression Modelling

Regression Modelling Assignment

Is an automatic or manual transmission better for MPG?

Synopsis

In this report we aim to show whether manual or automatic transmission is better for mile per gallon consumed (MPG). We also intend to quantify the MPG difference between automatic and manual transmissions. A linear regression model with MPG as the outcome is fitted with transmission, weight and horse power used as response variables. We have predicted that the weight of the car was a significant confounder in our analysis, and the choice of manual or automatic depends on it. We conclude that the cars with a automatic transmission are better for mpg than cars with an automatic transmission.

Data

This assignment makes use of the dataset Motor Trend Car Road Tests (mtcars) data. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74) models.

Loading and Summary of the Data

The data has 19 automatic cars with an average MPG of 17.15 compared to the mean MPG of manual cars of 24.39. Table 1.0 shows the summary of the MPG for each transmission with the standard deviation.

Plot 1 in the appendix section shows that motor vehicles with high number of cylinders have a lower mean MPG consumption for both transmissions. Plot shows weight as a confounder of the data since, manual cars have lower weight but a higher MPG compared to automatic cars with high weight but lower MPG consumption.

Regression Model

Model Overview

Only analyzing the transmission factor, we could conclude that an automatic car is better than a manual one as model summary in the appendix shows. However, there could be more factors that influenced in the car efficiency and act as a confounder variable in the results obtained. To get a preview which variables could have more influence in the outcome of our regression model, we then build a model with all the variables included and selected the ones who had a significant participation, i.e. a significant p-value or big residual standard deviation.

Model Selection

Considering the model with all the variables as predictor variables, we obtained a high R^2 statistic of 0.89, meaning that this model explained 89.31% of the data variance. We could have selected this model but, it has many linear dependent variables, so this increases the standard error, hence variance, of the coefficients, named variance inflation.

We used a stepwise selection of the model and used anova to check whether the best fit model has a significance difference from the others. The variables we selected from the model with all the variables are

- **hp** with a p-value of 0.0939
- **wt** with a p-value of 0.0946
- **am** the outcome of interest
- **wt:am** and **hp:am**

To validate the model we did a plot of the standardized residuals and the its normal qq plot (plot3 in the appendix). This shows tha our residuals are normally distributed.

Discussion

```
#include the key variables
model2 <- lm(mpg ~ am + wt + hp + wt:am + hp:am ,data= dataCars)
sModel2 <-summary(model2)
sModel2

##
## Call:
## lm(formula = mpg ~ am + wt + hp + wt:am + hp:am, data = dataCars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.987 -1.447 -0.535  1.261  5.599
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   30.7039     2.6752   11.48 1.1e-11 ***
## amManual      13.7400     4.2234    3.25  0.0032 **
## wt           -1.8559     0.9451   -1.96  0.0603 .
## hp            -0.0409     0.0136   -3.00  0.0058 **
## amManual:wt   -5.7689     2.0720   -2.78  0.0099 **
## amManual:hp    0.0278     0.0192    1.45  0.1598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.29 on 26 degrees of freedom
## Multiple R-squared:  0.879, Adjusted R-squared:  0.856
## F-statistic: 37.9 on 5 and 26 DF, p-value: 3.9e-11
```

From the model above, we can observe that the main variables are transmission , the interactions between horse power and transmission and weight and transmission. The model has an r-squared of 0.88 and a p-value of 4.913e-10

A manual motor vehicle has a high MPG 30.7. Both weight and horse power have a negative effect on the MPG efficiency. With other effects held constant, a unit increase in weight will reduce the MPG by -1.86 and the horse power by -0.04.

```
#include the key variables
model3 <- lm(mpg ~ wt:factor(am) + hp:factor(am) ,data= dataCars)
sModel3 <-summary(model3)
sModel3
```

```
##
## Call:
## lm(formula = mpg ~ wt:factor(am) + hp:factor(am), data = dataCars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.679 -2.066 -0.305  0.941  5.878
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      36.2167      2.4096   15.03 1.2e-14 ***
## wt:factor(am)Automatic -3.3682      0.9579   -3.52  0.0016 **
## wt:factor(am)Manual    -3.2964      1.4861   -2.22  0.0351 *
## factor(am)Automatic:hp -0.0385      0.0158   -2.43  0.0221 *
## factor(am)Manual:hp    -0.0330      0.0141   -2.35  0.0265 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.66 on 27 degrees of freedom
## Multiple R-squared:  0.83,    Adjusted R-squared:  0.805
## F-statistic:  33 on 4 and 27 DF,  p-value: 4.91e-10
```

Checking on the weight, if we keep the horse power constant, for each unit increase, which is equal to 1000 lbs , the number of miles per gallon will decrease -3.3 for manual cars and -3.37 for automatic cars. Regarding the horse power, for each additional horse power car , it will make-0.03 more miles for manual cars and -0.04 for automatic cars.

Finally we did an anova to compare the models and shows model1 is better than fit but our last model (model2) has no significant difference. But due to variance inflation we settled on model2.

```
#compare the models
#using anova
anova(fit, model1, model2)
```

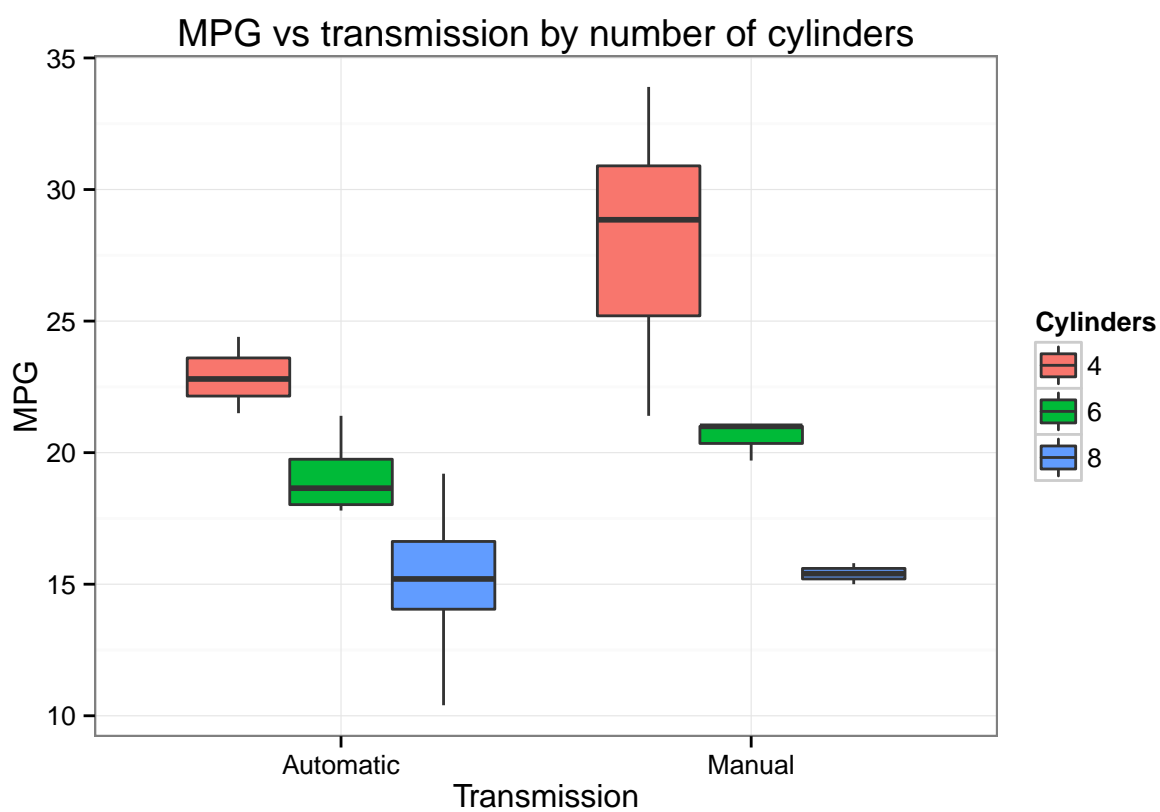
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 3: mpg ~ am + wt + hp + wt:am + hp:am
##   Res.Df RSS    Df Sum of Sq    F Pr(>F)
## 1      30 721
## 2      15 120  15      600 4.99 0.0018 **
## 3      26 136 -11      -15 0.18 0.9970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix

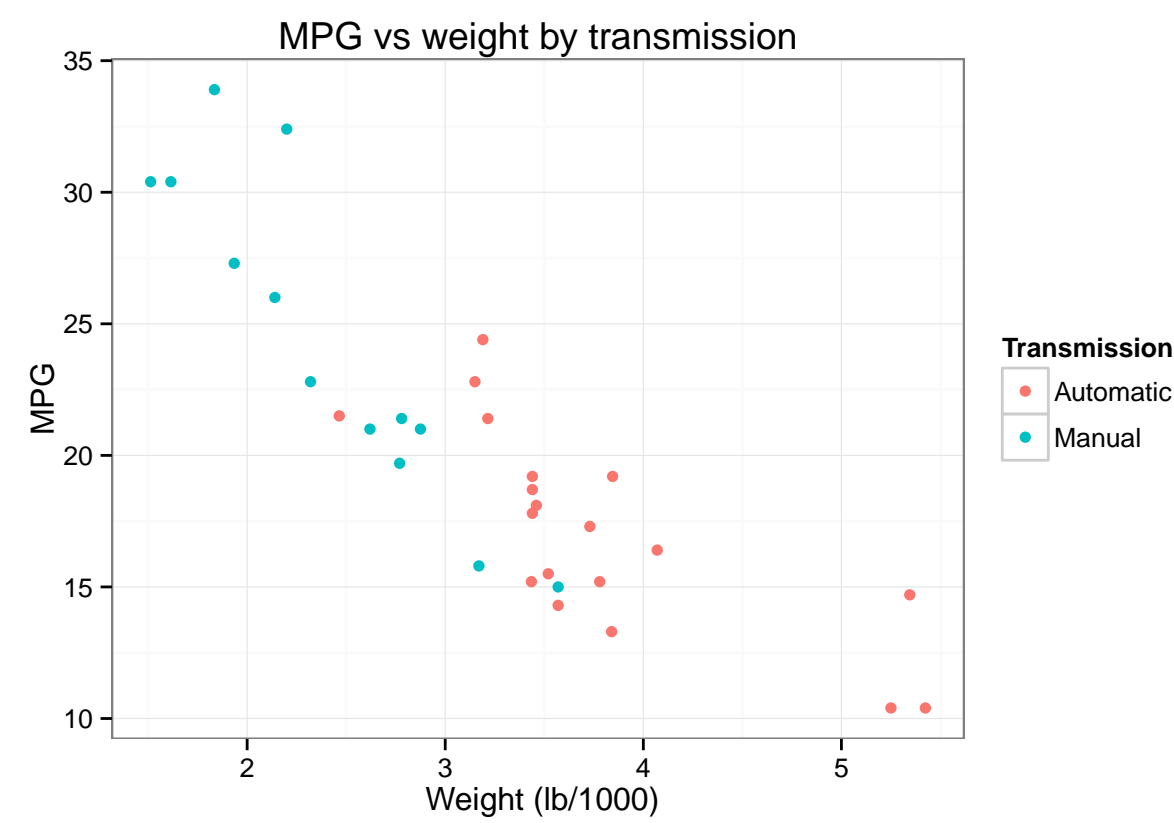
Table 1

am	count	mean	sd
Automatic	19	17.15	3.834
Manual	13	24.39	6.167

Plot1



Plot2



Plot3

