Regression Models Project

Tonya MacDonald

2/21/2021

Overview

This assignment is to analyze a dataset to see what factors have the largest impact on mile per gallon, in paticular to see if the following questions can be answered:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

Data Prep

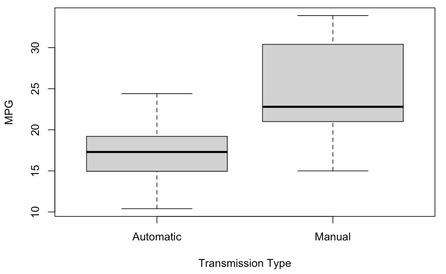
```
# load packaes and dataset
library(ggplot2)
data(mtcars)
```

```
# convert to factors
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$transmissiontype <- factor(mtcars$am, labels=c("Automatic","Manual"))
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
##                   transmissiontype
## Mazda RX4                   Manual
## Mazda RX4 Wag               Manual
## Datsun 710                  Manual
## Hornet 4 Drive           Automatic
## Hornet Sportabout        Automatic
## Valiant                  Automatic
```

Data Analysis

```
# create a boxplot of tranmission type vs mpg
boxplot(mpg ~ transmissiontype, data = mtcars, ylab = "MPG", xlab = "Transmission Type")
```



This boxplot shows a clear indication that manual transmissions have higher mpg than automatic.

We will review the data to see if there are other factors that impact mpg.

```
# create a linear regression of mpg vs transmission type
linreg <- lm(mpg ~ factor(am), data=mtcars)
summary(linreg)
```

```
##
## Call:
## lm(formula = mpg ~ factor(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 ***
## factor(am)1    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
```

Here we can see that the p-value is very low (p-value: 0.000285), so transmission type has a big effect on mpg.

However, the R-squared value is 0.3598, so this means that about 36% of the mpg variance is from transmission type.

```
# see what other factors impact the mpg variance
mpgvar <- aov(mpg ~ ., data = mtcars)
summary(mpgvar)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## cyl          1  817.7   817.7 102.591 2.3e-08 ***
## disp         1   37.6    37.6   4.717 0.04525 *
## hp           1    9.4     9.4   1.176 0.29430
## drat         1   16.5    16.5   2.066 0.16988
## wt           1   77.5    77.5   9.720 0.00663 **
## qsec         1    3.9     3.9   0.495 0.49161
## vs           1    0.1     0.1   0.016 0.90006
## am           1   14.5    14.5   1.816 0.19657
## gear         2    2.3     1.2   0.145 0.86578
## carb         5   19.0     3.8   0.477 0.78789
## Residuals   16  127.5     8.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

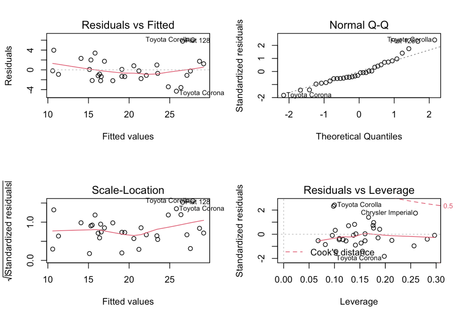
Here we see that number of cylinders (cyl), weight (wt), and displacement (disp) have low p-values and we should take a closer look.

```
# create a linear regression of mpg vs number of cylinders, weight, displacement, and transmission type
multilinreg <- lm(mpg ~ cyl + disp + wt + am, data = mtcars)
summary(multilinreg)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.318 -1.362 -0.479  1.354  6.059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.898313   3.601540  11.356 8.68e-12 ***
## cyl         -1.784173   0.618192  -2.886  0.00758 **
## disp         0.007404   0.012081   0.613  0.54509
## wt          -3.583425   1.186504  -3.020  0.00547 **
## am           0.129066   1.321512   0.098  0.92292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8079
## F-statistic: 33.59 on 4 and 27 DF,  p-value: 4.038e-10
```

Including the additional variables accounts for about 83% of the variance now (R-squared = 0.8327)

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



Summary

Looking at only transmission type, the average mpg The average mpg by transmission type is about 17 mpg for automatic transmissions, and 24 mpg for manual.

Transmission type was a large factor for mpg, however not the only one. Looking at the comparison of number of cylinders, weight, displacement, and transmission type you can see that they all have an impact on mpg.

```
aggregate(mtcars$mpg,by=list(mtcars$transmissiontype),FUN=mean)
```

```
##      Group.1        x
## 1 Automatic 17.14737
## 2    Manual 24.39231
```