Regression Cars

Neel

10/6/2020

Summary

The following analysis is a regression analysis of the mtcars dataset from R's datasets. We begin with loading and cleaning the data by changing respectable variables into factor variables. From this point we take a t-test of the mpg vs am variable to see that manual cars have a significant difference in mpg than automatic cars. From this point we perform a multivariate regression using a stepwise method in which we find cyl, hp, wt, and am contributing the most to mpg for cars. From this we can quantify that manual cars 1.80921 more mpg than automatic cars as seen by the coefficient from the multivariate regression.

Is automatic or manual transmission better for MPG?

Data preprocessing and Exploratory data analysis

Load and import the data and necessary plotting package.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
library(datasets)
data(mtcars)
```

Change some of the variables into factor variables if they are coded for specific types of car categories.

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am, labels = c('automatic', 'manual'))
summary(mtcars)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
```

```
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930
##        wt             qsec       vs             am     gear   carb
##  Min.   :1.513   Min.   :14.50   0:18   automatic:19   3:15   1: 7
##  1st Qu.:2.581   1st Qu.:16.89   1:14   manual   :13   4:12   2:10
##  Median :3.325   Median :17.71                         5: 5   3: 3
##  Mean   :3.217   Mean   :17.85                                4:10
##  3rd Qu.:3.610   3rd Qu.:18.90                                6: 1
##  Max.   :5.424   Max.   :22.90                                8: 1
```

With the clean dataframe, use a t-test to see if there is a significant difference between automatic and manual cars at the 95% confidence level. Our null hypothesis is that there is no significant difference in mpg between automatic vs manual cars. Our alternative hypothesis is that there is a significant difference in mpg between automatic vs manual cars.

```
t.test(mtcars$mpg ~ mtcars$am, alternative = c("two.sided"), conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group automatic    mean in group manual
##                17.14737                24.39231
```

From the above statistics we see that at the 95% confidence level, we reject the null hypothesis in favor of the alternative hypothesis as $0.001374 < 0.05$ such that there is a difference in mpg between automatic and manual cars. As see see that manual vs automatic cars mean mpg is 24.39231 vs 17.14737 we conclude that manual cars are better for mpg than automatic cars.

Quantify the MPG difference between automatic and manual transmissions

Regression analysis

We perform a single variable regression with mpg vs automatic or manual transmission.

```
first_fit <- lm(mpg ~ am, mtcars)
summary(first_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
```

The above summary shows that automatic cars get 17.147 mpg while we know that manual cars get 7.245 more mpg(24.292). We also see that the R^2 value is only 0.3385 which suggests that the type of transmission only accounts for 33.85% of the variance. Since this does not account for most of the variance we will perform a multivariate regression to find the most useful variables that account for the maximum amount of variance for mileage.

We perform a multivariate step regression between mpg and all other variables.

```
best_fit <- step(lm(mpg ~ ., mtcars), trace = 0)
summary(best_fit)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## ammanual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The above summary shows that the best fit for mpg includes cars with 6 and 9 cylinders, horsepower, weight, and manual transmission contributing the most to mpg with these variables explaining 84.01% of the variation as seen by the R^2 value of 0.8401. We see that 6 and 8 cylinder cars take 3.03134 and 2.16368 miles away from the mileage of a car, horsepower takes away 0.03211 mpg and weight takes away 2.49683 mpg. The manual transmission on the other hand shows a 1.80921 mpg increase over that of an automatic transmission through the coefficient seen.

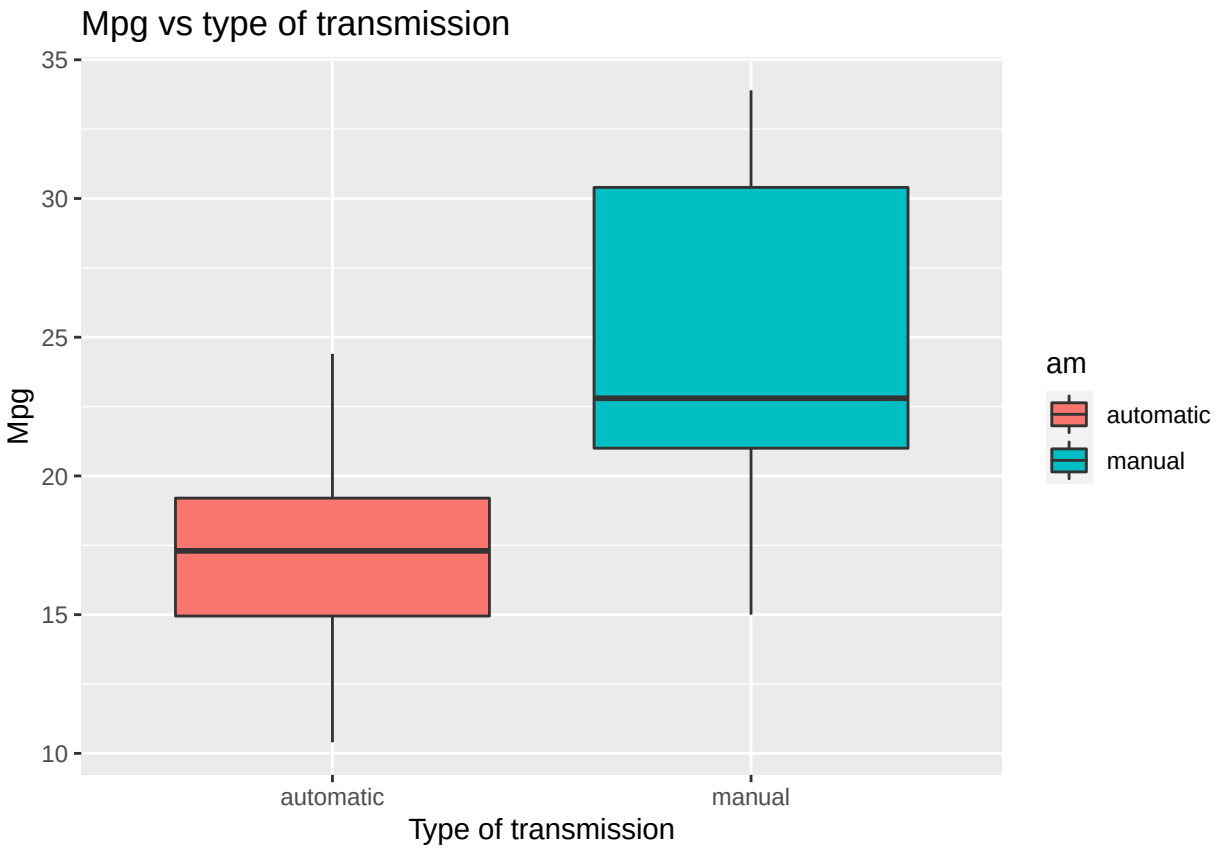
Conclusion

From the exploratory data analysis we see that manual transmission cars are significantly better than automatic cars. From the multivariate regression we can quantify that manual transmission cars have on average 1.80921 more mpg than an automatic car, but we also see that there are other factors such as cylinders, horsepower, and weight that can influence the mpg a car is able to get.

Appendix

Appendix 1: Boxplot of mpg vs automatic or manual transmission cars from ggplot2

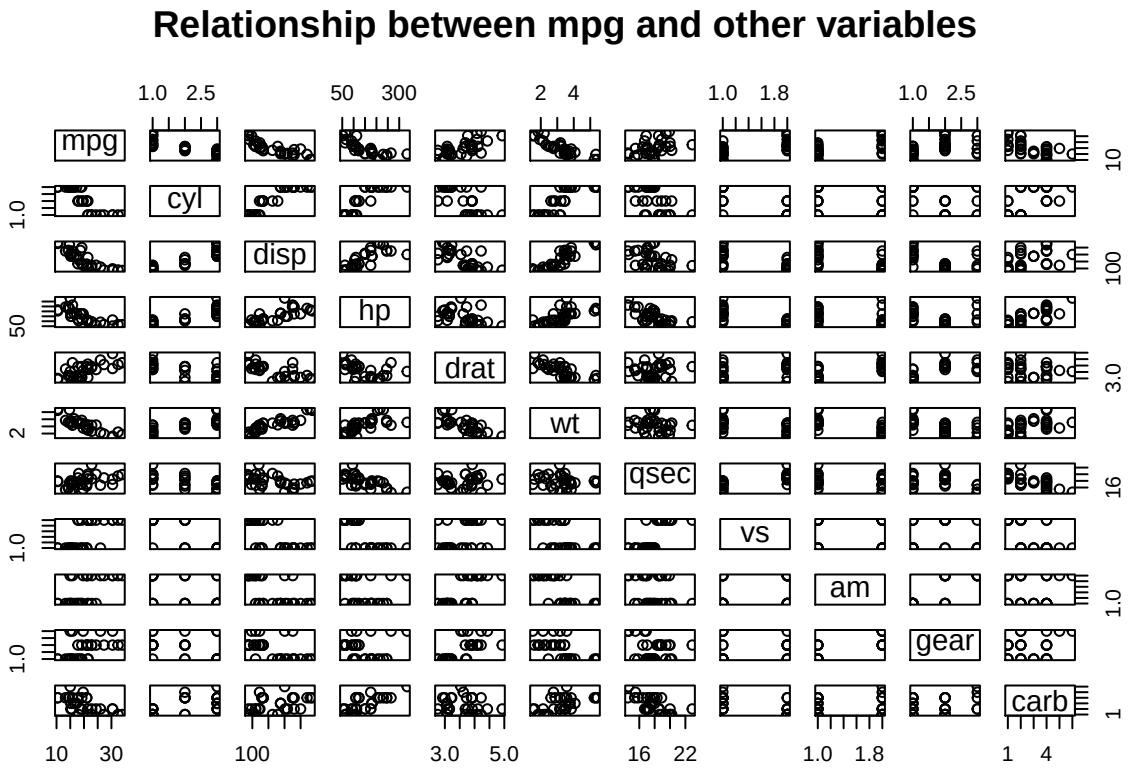
```
g <- ggplot(mtcars, aes(x = am, y = mpg, fill = am))
g <- g + geom_boxplot() + xlab('Type of transmission') + ylab('Mpg') + ggtitle('Mpg vs type of transmis
g
```



The above boxplot shows a visual of the significat difference we saw from the t-test suggesting manual transmission cars are better for mpg than automatic cars.

Appendix 2: Pair plot of mpg and all other variables

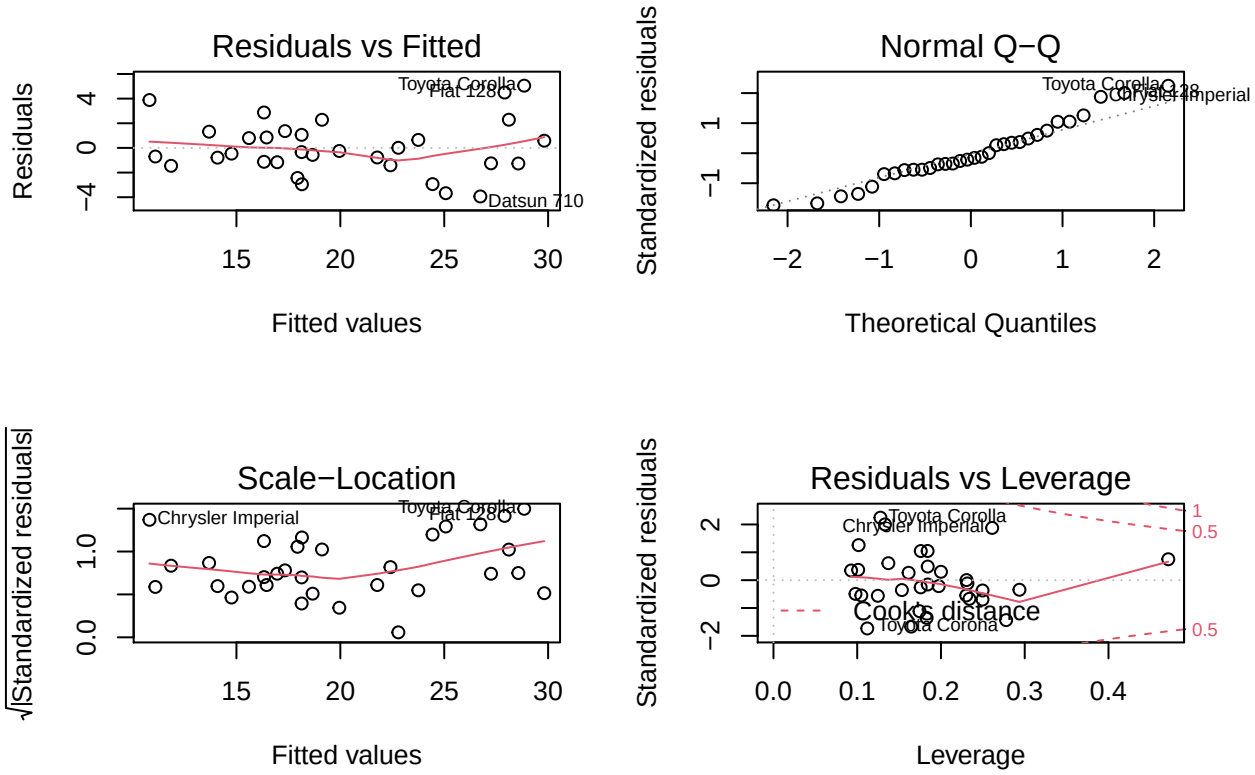
```
pairs(mpg ~ ., data = mtcars, main = "Relationship between mpg and other variables")
```



The above matrix of pair plots show that a lot of variables arent necessary linear such that for better overall analysis we may need to perform additional transformations to the data to ensure linearity and thus help account for more variance that we didn't manage to get from the above multivariate linear regression.

Appendix 3: Residual plots from multivariate regression

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



Further evidence that additional transformations may have been needed for the multivariate regression for better linearity. Regardless, the data looks nearly normal and thus is acceptable to continue with the analysis done above.