Motor Trend Analysis - MPG and Transmissions

Mari Coonley

January 19, 2017

Exectutive Summary

We have been asked to review the dataset from ***Motor Trend US*** magazine from 1974. This dataset is comprised of fuel consumption and ten other aspects of automobile design and performance for 32 automobiles, model years 1973 & 1974.

Questions of interest include:

Is an automatic or manual transmission better for MPG?

Quantify the MPG difference between automatic and manual transmissions.

When comparing only miles per gallon and the type of transmission using simple linear regression, I discovered the average MPG for a manual transmission is 24.39 miles per gallon of fuel. Whereas, the average MPG for an automatic transmission is 17.15. This suggests manual transmissions provide better fuel efficiency. Further, when using multivariate regression to take in to account the number of cylinders, gross horsepower and weight, after averaging each of those additional explanatory variables by centering, I concluded the manual transmission again, had a higher MPG, with 21.01, as compared with the automatic transmissionMPG measuring 19.46. Residual plots and diagnosis as well as t-testing further confirm this assumption.

Data Processing

I first created a secondary data set, *mtcars2*, including only the desired additional exploratory variables.

```
library(datasets)
data(mtcars)
mtcars2<-mtcars[c(-5,-7, -8, -10, -11)]
mtcars2$am<-as.factor(mtcars2$am)
library(plyr)
mtcars2$am<-revalue(mtcars2$am, c("0"="auto", "1"="man"))
str(mtcars2)
```

```
## 'data.frame':    32 obs. of  6 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ am  : Factor w/ 2 levels "auto","man": 2 2 2 1 1 1 1 1 1 1 ...
```

Next I created a tertiary dataset, *mtcars3*, centering the exploratory variables.

```
mtcars3<-NULL
mtcars3$mpg<-mtcars2$mpg; mtcars3$am<-mtcars2$am
mtcars3$disp<-mtcars2$disp-mean(mtcars2$disp)
mtcars3$cyl<-mtcars2$cyl-mean(mtcars2$cyl)
mtcars3$hp<-mtcars2$hp-mean(mtcars2$hp)
```

```
mtcars3$wt<-mtcars2$wt-mean(mtcars2$wt)
mtcars3<-data.frame(mtcars3$mpg, mtcars3$cyl, mtcars3$disp, mtcars3$hp, mtcars3$wt,mtcars3$am)
names(mtcars3)<- c("mpg", "cyl", "disp","hp", "wt", "am")
str(mtcars3)
```

```
## 'data.frame':    32 obs. of  6 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  -0.188 -0.188 -2.188 -0.188 1.812 ...
##  $ disp: num  -70.7 -70.7 -122.7 27.3 129.3 ...
##  $ hp  : num  -36.7 -36.7 -53.7 -36.7 28.3 ...
##  $ wt  : num  -0.59725 -0.34225 -0.89725 -0.00225 0.22275 ...
##  $ am  : Factor w/ 2 levels "auto","man": 2 2 2 1 1 1 1 1 1 1 ...
```

Exploratory Data Analysis

After comparing the correlations between all exploratory variables and MPG, as well as researching each topic individually, I selected cylinders, displacement, horsepower and weight as the exploratory variables having the most direct affect on MPG. (see Exp. 1 in Appendix) I next generated a boxplot depicting correlation between mpg and automatic and manual transmissions (see figure 1 in Appendix) as well as a second pairs plot with only the selected variables. (see figure 2 in Appendix).

Regression Models

Initially, I calculated a simple linear regression model comparing MPG to the factored transmissions variable.

```
fit1<-lm(mpg~am-1, mtcars2)
fit1$coefficients
```

```
##   amauto    amman
## 17.14737 24.39231
```

I next generated multivariate regression models using the original data, *mtcars2*, and the data centering the exploratory variables by averaging them, *mtcars3*.

```
fit2<-lm(mpg~am-1+cyl+disp+hp+wt, mtcars2)
fit3<-lm(mpg~am-1+cyl+disp+hp+wt, mtcars3)
compfits<-data.frame(fit2$coefficients, fit3$coefficients)
compfits
```

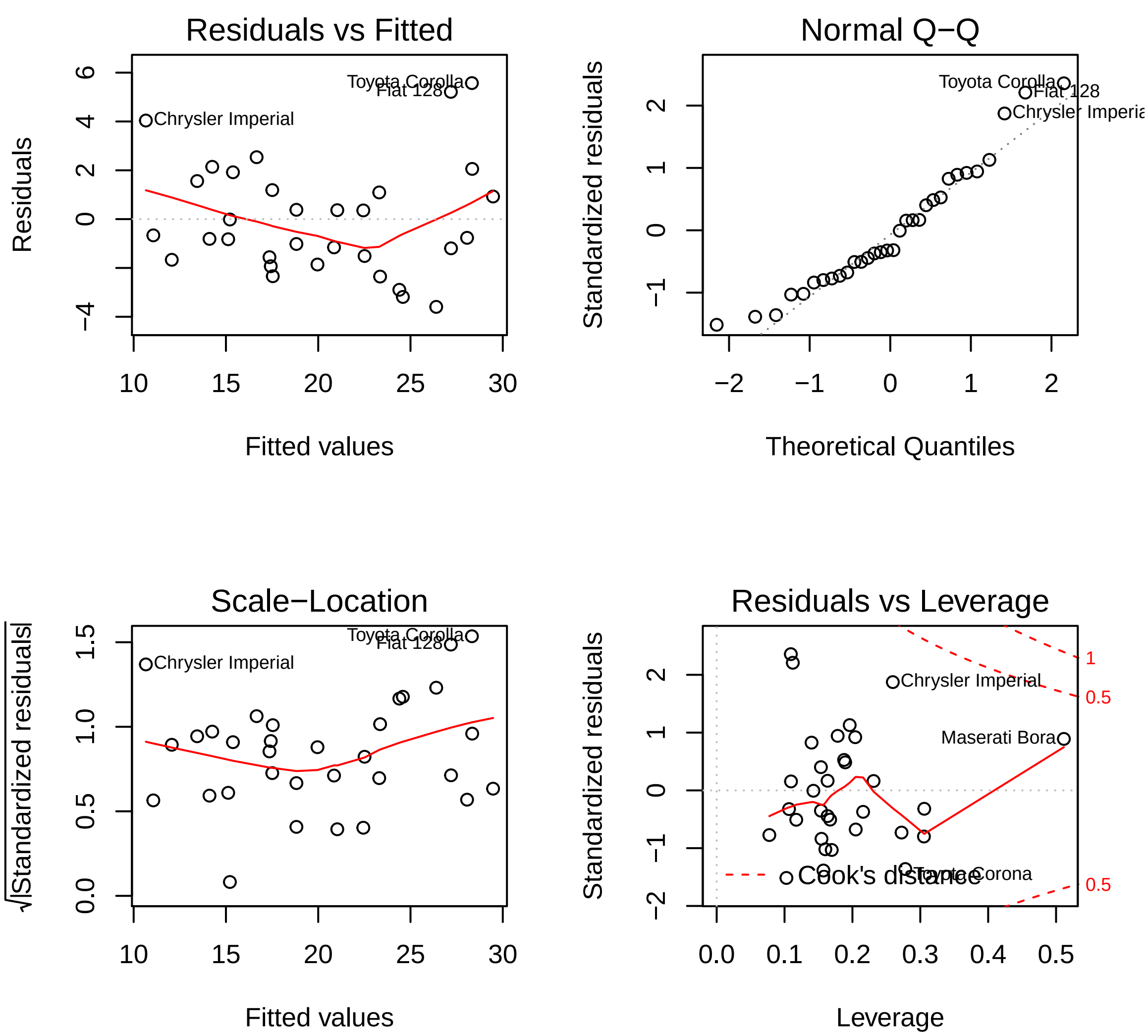
```
##        fit2.coefficients fit3.coefficients
## amauto       38.20279869       19.45830027
## amman        39.75929032       21.01479191
## cyl          -1.10637984       -1.10637984
## disp          0.01225708        0.01225708
## hp           -0.02796002       -0.02796002
## wt           -3.30262301       -3.30262301
```

Fit2 coefficients reflect the average MPG, per automatic and manual, when cylinders, displacement, horsepower and weight are at zero. On the other hand, *fit3* reflects MPG accounting for the averages of cylinders, horsepower, and weight. *fit3* is a more reasonable analysis than *fit2* simply because without cylinders, displacement and horsepower, which have weight, MPG would not be possible (i.e. when those exploratory variables equaling zero). Note the slopes of the exploratory variables don't change, only the intercepts for manual and automatic.

Residuals Plots and Diagnosis

Let's run the set of diagnostic plots for linear regression to assess leverage and influence.

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



Review each plot.

1. *Residual vs. Fitted* - no systematic pattern is evidence of fit.
2. *Normal Q-Q* - 3 outliers, still on the line, but outside cluster suggests high leverage but low influence, normally distributed.
3. *Scale-Location* - shows fairly equally spread residual both above and below the line, line is fairly horizontal, suggesting equal variance.

4. *Residual vs. Leverage* outliers in the upper and/or lower right hand corners of this plot around and outside dashed line, are considered influential as well as having leverage. There are none present, including the outliers visible in all the other plots. This suggest that while these outliers have leverage, they have little influence on the regression.

Inferences

Lastly, lets run a t-test to check significance.

```
t.test(mpg~am, mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by 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 0 mean in group 1
##        17.14737        24.39231
```

The relatively large t-value coupled with the small size of the p-value we can infer that the unstated null hypothesis that there is no significant difference in mpg of manual and automatic transmissions should be rejected. Further note, the calculated means of the two transmissions are the same as calculated using regression analysis.

Conclusion

In conclusion, based on the analysis data, the null hypothesis that there is no difference in the miles per gallon based on a vehicle having either a manual or automatic transmission can be rejected. There are some uncertainties to take into consideration with this analysis. First, the data set is fairly small, we have to consider how it would translate to a much larger data set. Also, even though the residual analysis shows a fairly good evidence of fit, would the inclusion of other exploratory variables, such as displacement or type of engine had any impact, or merely inflated the variances? While I am confident my analysis is correct, there is always room to question these and other concerns.

Appendix

Figure 1

```
disp<-cor(mtcars$mpg,mtcars$disp)
hp<-cor(mtcars$mpg,mtcars$hp)
wt<-cor(mtcars$mpg,mtcars$wt)
cyl<-cor(mtcars$mpg,mtcars$cyl)
data.frame(disp, hp, wt, cyl)
```

```
##         disp         hp         wt       cyl
## 1 -0.8475514 -0.7761684 -0.8676594 -0.852162
```

```r
rraxl<-cor(mtcars$mpg,mtcars$drat)
gear<-cor(mtcars$mpg,mtcars$gear)
carb<-cor(mtcars$mpg,mtcars$carb)
data.frame(rraxl, gear, carb)
```

```
##       rraxl      gear       carb
## 1 0.6811719 0.4802848 -0.5509251
```

Figure 2

```r
plot(mpg~am, mtcars2)
```

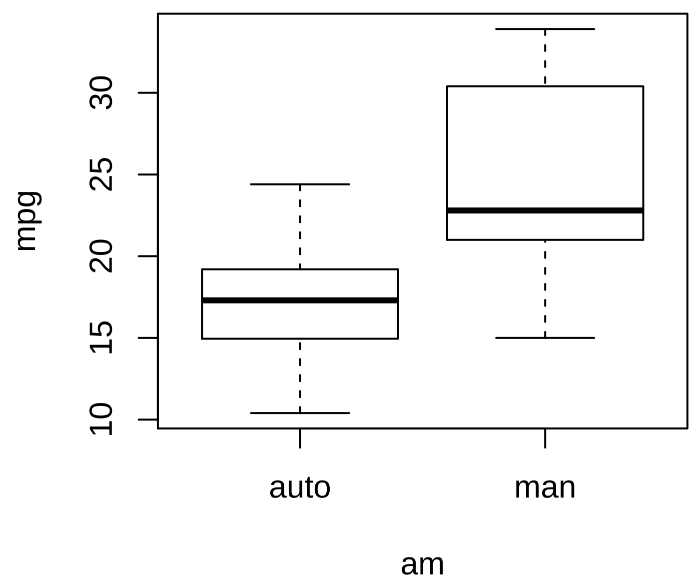
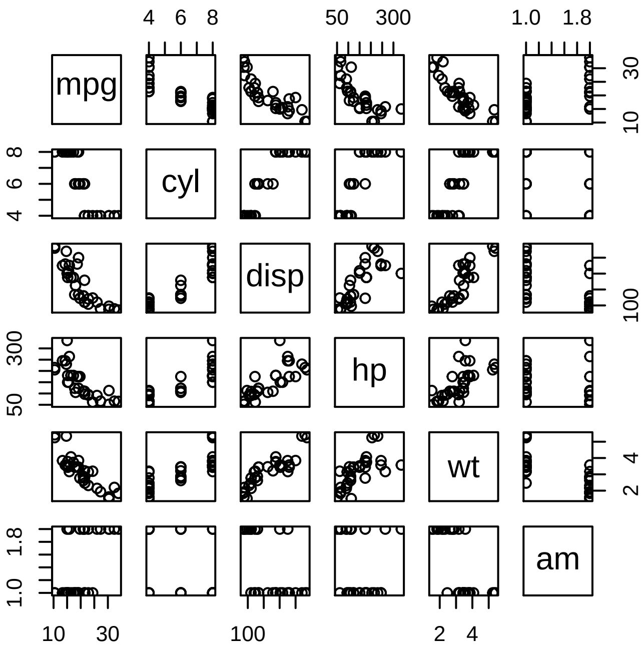


Figure 3

```r
pairs(mtcars2)
```


mpg
cyl
disp
hp
wt
am
4 6 8
50 300
1.0 1.8
30
10
8
6
4
100
300
50
4
2
1.8
1.0
10 30
100
2 4