Motor Trend Data Analysis

John Letteboer

11/25/2018

Executive Summary

In this project, we analyze the `mtcars` (Motor Trend Car Road Tests) data set and explore the relationship between a set of variables and miles per gallon (MPG). 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). The main focus of the project is to answer the following questions:

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

Data prep

Load the data and do some checks.

```
data(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

```
head(mtcars,3)
```

```
##                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
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 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 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Transform some variables into factors.

```
mtcars_org <- mtcars
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am,labels=c("Automatic","Manual"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Analysis

The plots created in this section can be found in the **Appendix: Figures** section at the end of the document.

```
x <- mtcars_org[1]
y <- mtcars_org[2:11]
cor(x, y)
```

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

There are stronger correlations between variables: "cyl", "disp", "hp", and "wt".

Statistical Inference

To see a significant difference in the mean of "mpg" for automatic and manual transmission let's do a T-Test and show the estimate.

```
t <- t.test(mpg ~ am, data=mtcars)
t$estimate
```

```
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

```
t$p.value
```

```
## [1] 0.001373638
```

The p-value is 0.0013736, so we reject the null hypothesis. The mean of manual transmission is about 7 MPG more than that of an automatic transmission.

Regression Analysis

```
initModel <- lm(mpg ~ ., data=mtcars)
summary(initModel)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
```

```
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## amManual     1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

The residual standard error is 2.8331687 on 15 degrees of freedom. The Adjusted R-Squared value is 0.7790215, which is interpreted to mean this model can explain about 77.9% of the variance of the MPG. All of the coefficients are above the 0.05 significant level, so we can conclude none of the coefficients are significant.

Backward elimination can be used to used to find the most statistically significant predictor variables. The `step()` function starts with all the predictors in the model and then drops the one with the largest p-value (least significant). Then the model is re-fitted and this is repeated until only variables that are statistically significant remain.

```r
bestModel <- step(initModel, k=log(nrow(mtcars)), trace=FALSE)
summary(bestModel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The best model shows that "wt", "qsec", and "amManual" are statistically significant variables. The Adjusted R-Squared value is 0.8336, which is interpreted to mean this model can explain about 83.36% of the variance of the MPG.

Since the focus of this project is on the relationship between transmission and MPG, the next model will fit transmission as a predictor of MPG.

```r
mpg_amModel <- lm(mpg ~ am, data=mtcars)
summary(mpg_amModel)
```

```
##
## 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
```

This model shows that a car with an automatic transmission has an average of 17.147 mpg, and manual transmission increases mpg by 7.245. However, this model has an Adjusted R-Squared of 0.3385, which means the model can only explain about 33.85% of the variance of the MPG. Other variables should be added in to get a higher Adjusted R-Squared value.

Lastly, let's compare the model with only "am" as the predictor variable with the best model obtained through backwards elimination.

```r
anova <- anova(mpg_amModel, bestModel)
```

This results in a p-value of 1.5504951 x 10-9, which is highly significant. Therefore, we reject the null hypothesis that variables "wt", "qsec", and "am" do not contribute to the prediction of mpg.

The best fitted model is `lm(formula = mpg ~ wt + qsec + am, data = mtcars)`

```r
confint(bestModel)
```

```
##                   2.5 %    97.5 %
## (Intercept) -4.63829946 23.873860
## wt          -5.37333423 -2.459673
## qsec         0.63457320  1.817199
## amManual     0.04573031  5.825944
```

```r
summary(bestModel)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## amManual     2.935837  1.4109045  2.080819 4.671551e-02
```

This model shows that when "wt" and "qsec" remain constant, manual transmission cars get an average of 2.94 more MPG than those with automatic transmission.

Residual Plots & Diagnostics

Refer to Figure 4 in the Appendix: Figures section for the residual plots. Interpretation of the residual plots:

- The Residuals vs Fitted plot - the points are randomly scattered on the plot confirming the independence condition.
- The Normal Q-Q plot - the majority of points fall on the line indicating normal distribution of the residuals.
- The Scale-Location plot - there is a random band around the line with no clear pattern indicating constant variance.
- The Residual vs Leverage plot - there are some outliers (leverage points) that may indicate values of increased leverage of outliers.

Next, regression diagnostics can be used to further investigate our model.

The hatvalues() function is used to find values far from the average (usually 2-3x the average) because these may have substantial influence on the regression parameters.

```
hv_mean<- mean(hatvalues(bestModel))
hv <- hatvalues(bestModel)
tail(sort(hv),3)
```

```
##   Chrysler Imperial Lincoln Continental            Merc 230
##           0.2296338           0.2642151           0.2970422
```

The average hatvalue is 0.125, and from the above step we get the cars that are about 2 times the average or greater. These are the same cars that we see mentioned on the residual plots.

Conclusion

Cars with manual transmission get better miles per gallon compared to those with automatic transmission.

- The t-test shows that manual transmission gets an average of 7 MPG more than cars with automatic transmission.
- Several linear regression models were fitted to evaluate different aspects that could impact MPG. The best fitted model lm(formula = mpg ~ wt + qsec + am, data = mtcars) showed that when "wt" (Weigth (lb/1000)) and "qsec" (1/4 mile time) remain constant, manual transmission cars get an average of 2.94 more MPG than those with automatic transmission.

Appendix: Figures

Figure 1 - Boxplot of MPG vs. Transmission

```
boxplot(mtcars$mpg ~ mtcars$am, xlab="Transmission", ylab="MPG", col=c("blue","green"))
```

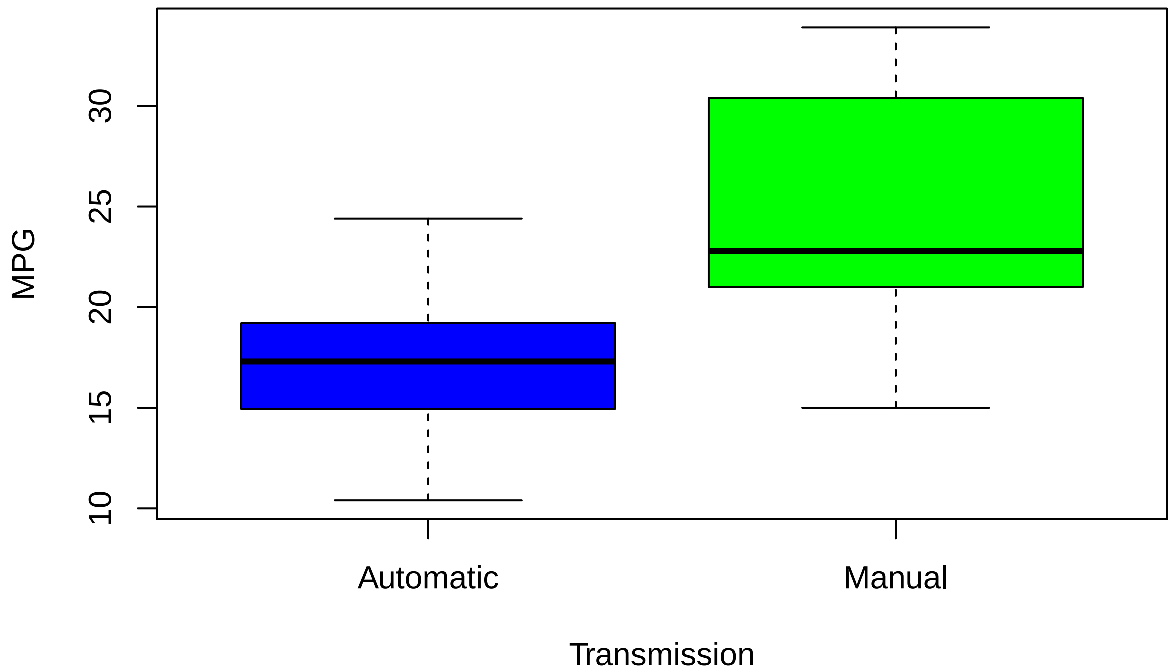


Figure 2 - Pairs Histograms and correlations from psych

```
library(psych)
pairs.panels(mtcars, method = "pearson", hist.col = "#00AFBB", ellipses = FALSE)
```

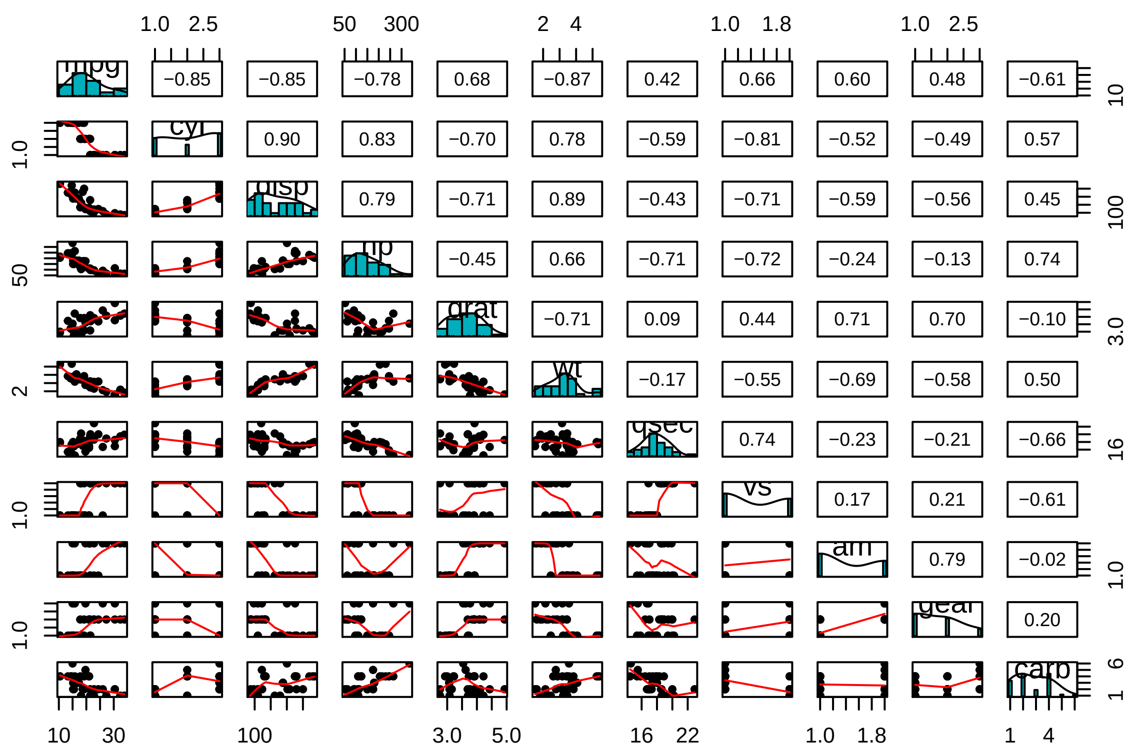


Figure 3 - Correlation Plot

```
require(corrplot)
corrplot.mixed(cor(mtcars_org), number.cex = .7, tl.col='black')
```

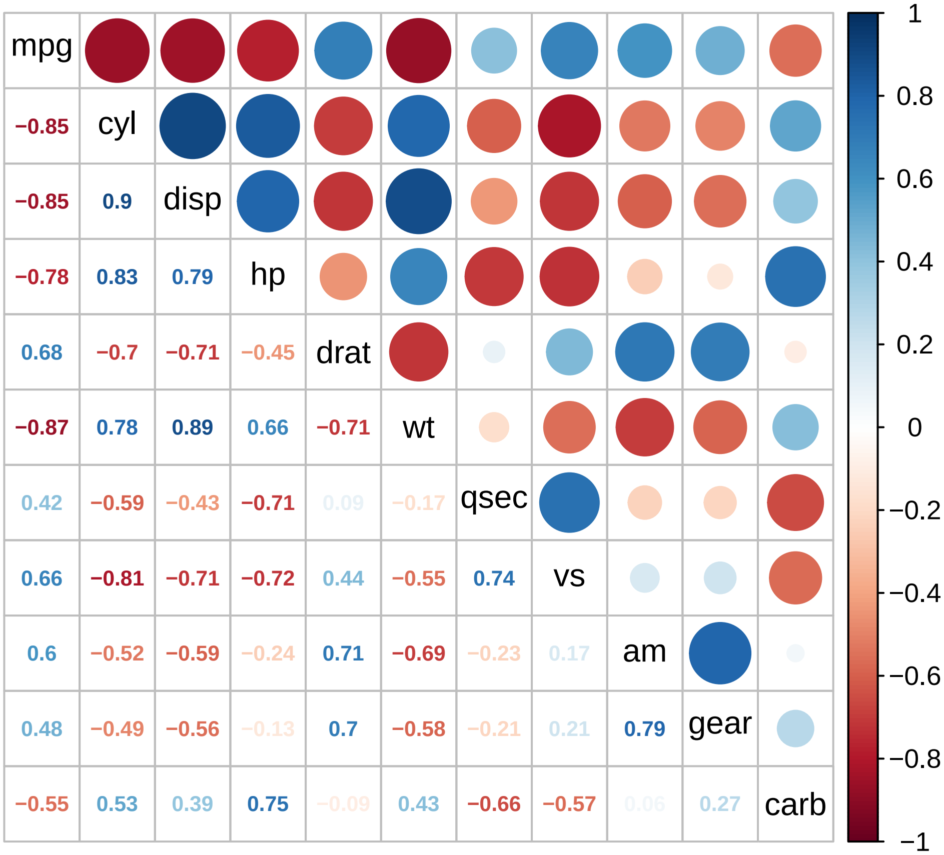
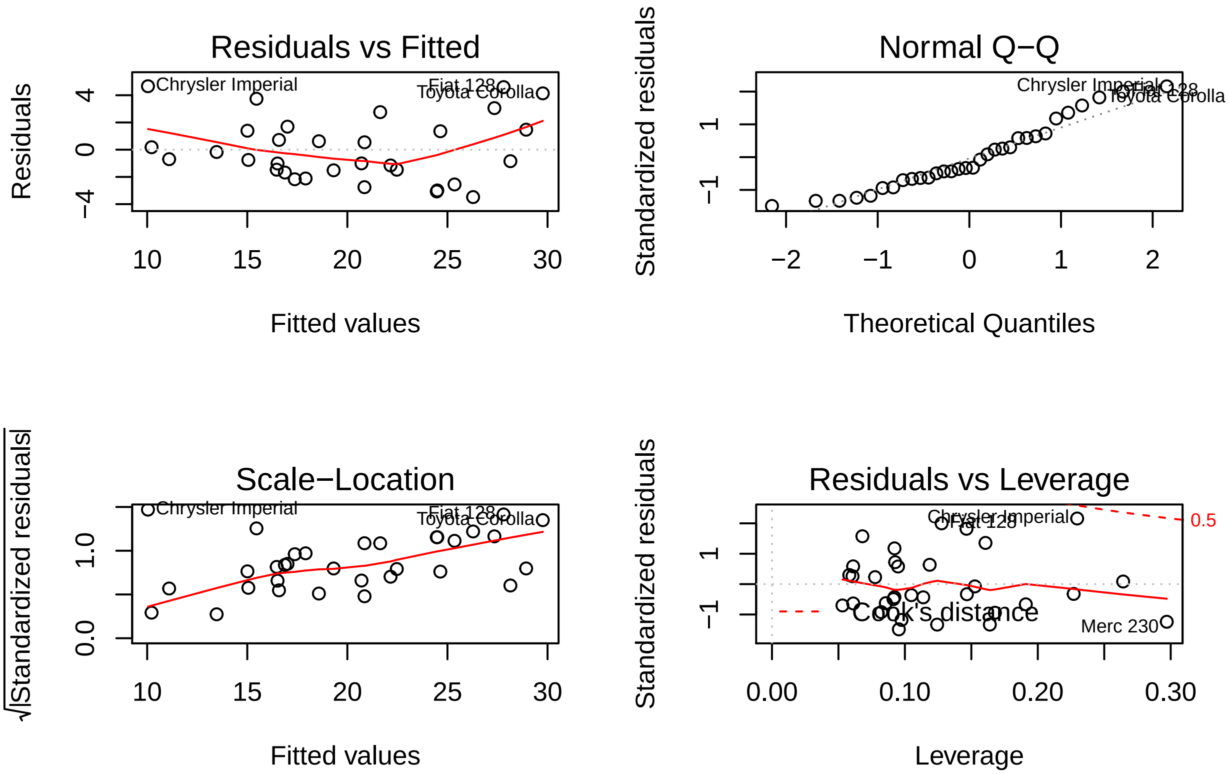


Figure 4 - Residual Plots

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


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5
Merc 230