Motor Trend Cars MPH Analysis

Rick Sanqui

Saturday, October 18, 2014

Executive Summary

The following analysis demonstrates that for the 1973-1974 car models, automatic transmission cars produced lower fuel efficiency than manual transmission cars by an average of 7.24 mpg (see Appendix A). By utilizing the *Motor Trend* US magazine data available in the *mtcars* R data set, we will also explore other variables, which in conjunction with transmission type, can help us predict *mpg*.

Analysis

To identify the most significant variables that can help us predict *mpg*, we will perform an initial linear regression on the *mtcars* 1974 data including all variables. The five most significant predictors on **mpg** will be identified from the **P** value in asscending order.

```
data(mtcars)
mtcars$am <- factor(mtcars$am, levels=c(0,1), labels=c('Automatic','Manual'))
fit <- lm(mpg ~ ., data=mtcars)
coef.p <- summary(fit)$coefficients
( all.coef <- as.data.frame( coef.p[ order(coef.p[,"Pr(>|t|)"] ) ,] )[1:5,] )
```

```
##             Estimate Std. Error    t value   Pr(>|t|)
## wt       -3.71530393 1.89441430 -1.9611887 0.06325215
## amManual  2.52022689 2.05665055  1.2254035 0.23398971
## qsec      0.82104075 0.73084480  1.1234133 0.27394127
## hp       -0.02148212 0.02176858 -0.9868407 0.33495531
## disp      0.01333524 0.01785750  0.7467585 0.46348865
```

In order of significance, five most significant predictors include

- wt – Weight (lb/1000)
- am – Transmission (0 = automatic, 1 = manual)
- qseq – 1/4 mile time
- hp – Gross horsepower
- disp – Displacement (cu.in.)

Creating linear models by adding the next significant predictor we will determine which model provides the most improvement while selecting the most parsimonious model.

```
fit.1 <- lm(mpg ~ wt , data=mtcars)
fit.2 <- lm(mpg ~ wt + am, data=mtcars)
fit.3 <- lm(mpg ~ wt + am + qsec, data=mtcars)
fit.4 <- lm(mpg ~ wt + am + qsec + hp , data=mtcars)
fit.5 <- lm(mpg ~ wt + am + qsec + hp + disp, data=mtcars)
anova( fit.1 , fit.2 , fit.3 , fit.4 , fit.5 )
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + am
## Model 3: mpg ~ wt + am + qsec
## Model 4: mpg ~ wt + am + qsec + hp
## Model 5: mpg ~ wt + am + qsec + hp + disp
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 278.32
## 2     29 278.32  1     0.002  0.0004 0.984604
## 3     28 169.29  1   109.034 18.4757 0.000214 ***
## 4     27 160.07  1     9.219  1.5622 0.222472
## 5     26 153.44  1     6.629  1.1232 0.298972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary( fit.3 )$coefficients
```

```
##              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
## amManual     2.935837  1.4109045  2.080819 4.671551e-02
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

```
(fit.3.adj.r.sqrd <- summary( fit.3 )$adj.r.squared)
```

```
## [1] 0.8335561
```

Model 3, which includes **wt**, **am**, and **qsec**, explains 83% of the data.

Is an automatic or manual transmission better for MPG?

To answer this question regarding the second most influential attribute in the *mtcars* data set, we examine a model with transmission as the only predictor.

```
summary( mpg.fit.am <- lm(mpg ~ am , data=mtcars) )$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

```
( mpg.fit.am.a.r.squared <- summary( mpg.fit.am )$adj.r.squared )
```

```
## [1] 0.3384589
```

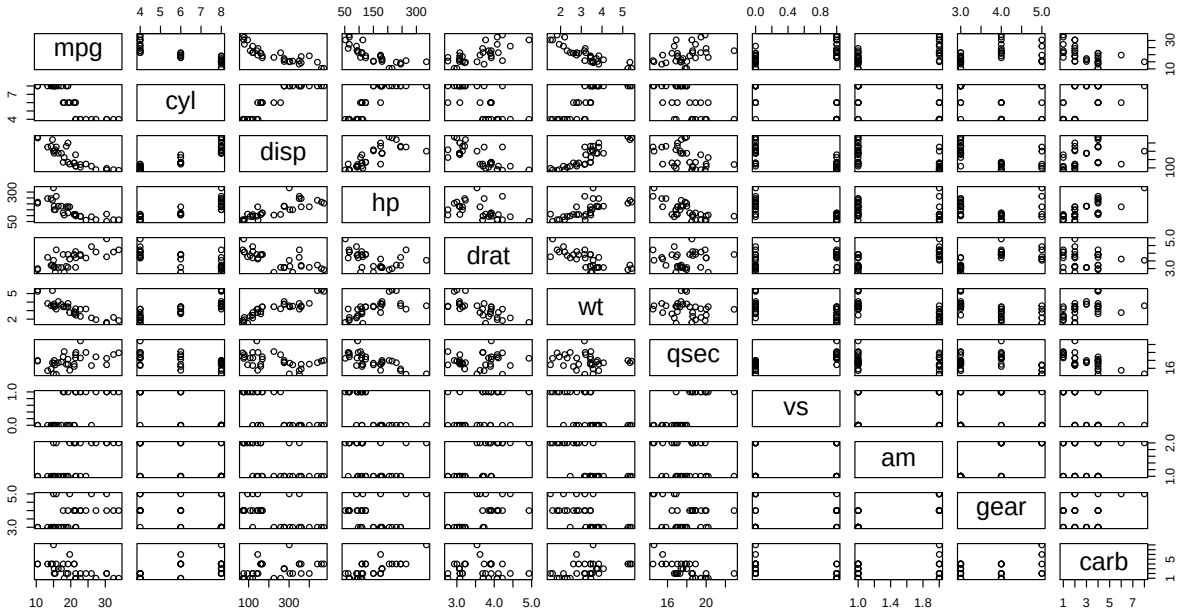
On average, manual transmissions provided a 17 mpg advantage over automatic tranmissions.
The transmission type P values indicates that transmission is a statistically significant predictor. However, the adjusted R^2 value of 0.338 tells us that transmission type, on its own, is not a very good predictor of mpg; only explaining 34% of the data.

```
intercept.confidence <- summary(mpg.fit.am)$coefficient[1, 1] + c(-1, 1) * qt(0.975,
    df = mpg.fit.am$df) * summary(mpg.fit.am)$coefficient[1, 2]
slope.confidence <- summary(mpg.fit.am)$coefficient[2, 1] + c(-1, 1) * qt(0.975,
    df = mpg.fit.am$df) * summary(mpg.fit.am)$coefficient[2, 2]
Tranny.Sigma <- summary(mpg.fit.am)$sigma
```

In this model using only transmission type as a predictor of **mpg**, we are 95% confident that

- Automatic transmissions mpg are between 14.85 and 19.44 mpg.
- Manual transmissions have a mpg advantage between 3.64 and 10.85 mpg over automatic transmissions.
- The Unbiased Standard Deviation, σ of the residual, is 4.9 mpg.

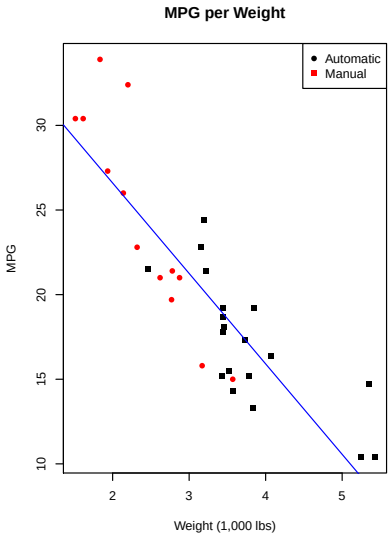
Appendix A – Correlations between attribute pairs

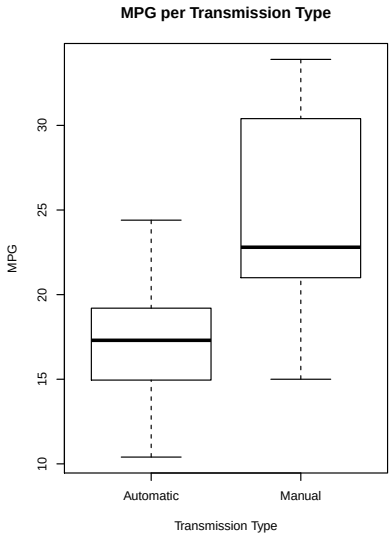


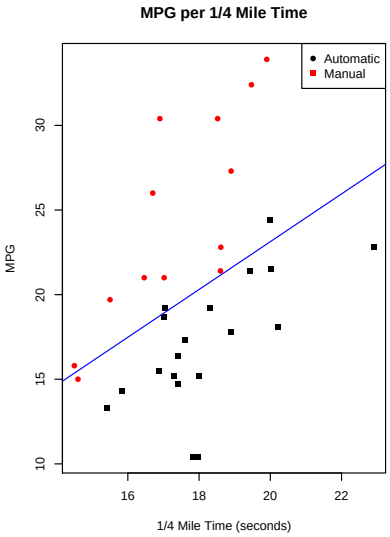
Appendix B – Correlations between attribute pairs

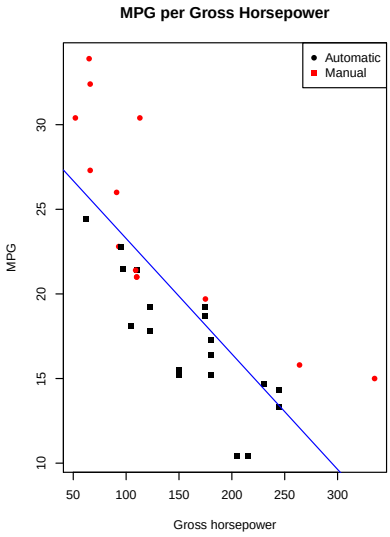
```
par(mfrow=c(2,3))
plot( mtcars$wt , mtcars$mpg , col=factor(mtcars$am, levels=c('Automatic','Manual'), labels=c(1,2)) ,
      pch=as.numeric( factor(mtcars$am, levels=c('Automatic','Manual')) ) + 14 , ylab="MPG",
      xlab="Weight (1,000 lbs)" , main="MPG per Weight")
legend("topright", c('Automatic','Manual'), col=c('black','red') , pch=c(16,15) )
abline(lm(mtcars$mpg ~ mtcars$wt), col="blue")

plot( mtcars$am , mtcars$mpg , main="MPG per Transmission Type" , xlab="Transmission Type", ylab="MPG")

plot( mtcars$qsec , mtcars$mpg , col=factor(mtcars$am, levels=c('Automatic','Manual'), labels=c(1,2)) ,
      pch=as.numeric( factor(mtcars$am, levels=c('Automatic','Manual')) ) + 14 , ylab="MPG",
```

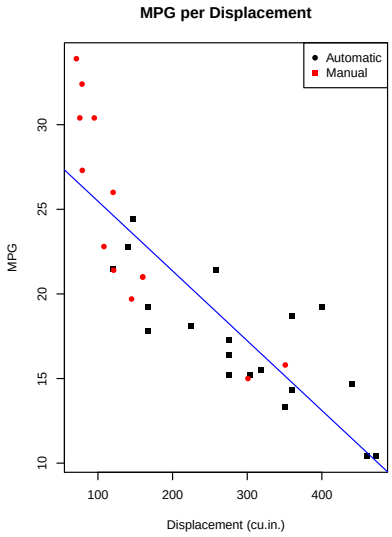
```
      xlab="1/4 Mile Time (seconds)" , main="MPG per 1/4 Mile Time")
legend("topright", c('Automatic','Manual'), col=c('black','red') , pch=c(16,15) )
abline(lm(mtcars$mpg ~ mtcars$qsec), col="blue")

plot( mtcars$hp , mtcars$mpg , col=factor(mtcars$am, levels=c('Automatic','Manual'), labels=c(1,2)) ,
      pch=as.numeric( factor(mtcars$am, levels=c('Automatic','Manual')) ) + 14 , ylab="MPG",
      xlab="Gross horsepower" , main="MPG per Gross Horsepower")
legend("topright", c('Automatic','Manual'), col=c('black','red') , pch=c(16,15) )
abline(lm(mtcars$mpg ~ mtcars$hp), col="blue")

plot( mtcars$disp , mtcars$mpg , col=factor(mtcars$am, levels=c('Automatic','Manual'), labels=c(1,2)) ,
      pch=as.numeric( factor(mtcars$am, levels=c('Automatic','Manual')) ) + 14 , ylab="MPG",
      xlab="Displacement (cu.in.)" , main="MPG per Displacement")
legend("topright", c('Automatic','Manual'), col=c('black','red') , pch=c(16,15) )
abline(lm(mtcars$mpg ~ mtcars$disp), col="blue")
```





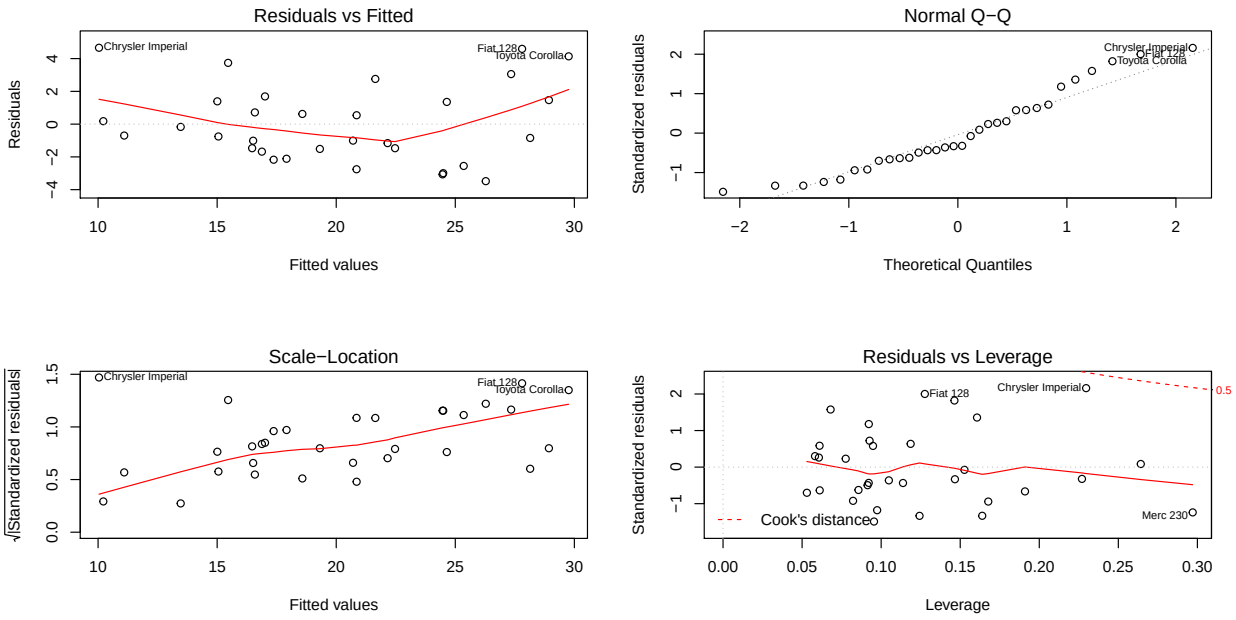






Appendix B – Residual Plots of 'mpg ~ wt + am + qsec'

```
par(mfrow=c(2,2)) ; plot( fit.3 )
```



Appendix C – Residual Plots of Transmission as predictor of MPG

```
par(mfrow=c(2,2)) ; plot( mpg.fit.am )
```