

Regression-Course Project

Scot Shields

Wednesday, March 18, 2015

Executive Summary:

In this report we look at a data set of a collection of cars from the mtcars data. We are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). We are particularly interested in the following two questions:

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

PreProcessing:

```
cache = TRUE
data(mtcars)
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'))
```

Exploratory Data Analysis:

```
head(mtcars)
str(mtcars)
summary(mtcars)
```

The data contains 32 observations of 11 variables. The minimum mpg is 10.4 mpg and the maximum is 33.90 mpg, with a mean of 20.09 mpg. From the pairs scatterplot in the appendix, it appears that cyl, disp, hp, drat, and wt seem to be strongly correlated with mpg. The boxplot of mpg vs transmission type indicates manual transmissions have over all higher mpg. This may indicate mpg and am are correlated as well.

Models:

```
fit1<-lm(mpg~.,mtcars)
fit2<-lm(mpg~am,mtcars)
fit3<-lm(mpg~am+cyl+disp,mtcars)
fit4<-lm(mpg~am+cyl+disp+hp,mtcars)
fit5<-lm(mpg~am+cyl+disp+hp+drat,mtcars)
fit6<-lm(mpg~am+cyl+disp+hp+drat+wt,mtcars)
fit7<-lm(mpg~am+cyl+disp+hp+drat+wt,mtcars)
```

Model Comparison:

The anova results in the appendix indicate that fit3 gives the best model with independent variable am and confounders cyl and disp. From the Summary(fit3) in the appendix we see the adjusted R-squared value for the model is .765 indicating that more than 76% of the variance in the data is explained by the model.

Residuals & Diagnostics:

Looking at the results in the appendix we see that the Hornet 4 Drive is the only car shown in the residual plots and the diagnostics.

Inference:

The small p-value from the t.test in the appendix causes us to reject the null hypothesis that there is no difference between manual and automatic transmissions.

Conclusions

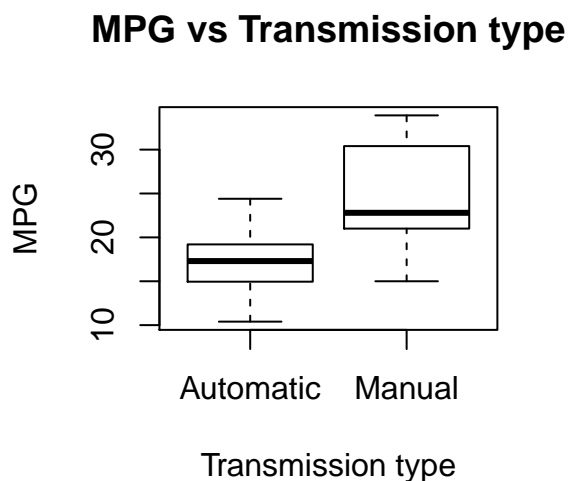
‘Manual’ transmissions get better mpg than ‘Automatic’ transmissions.

Our fit3 model indicates the following:

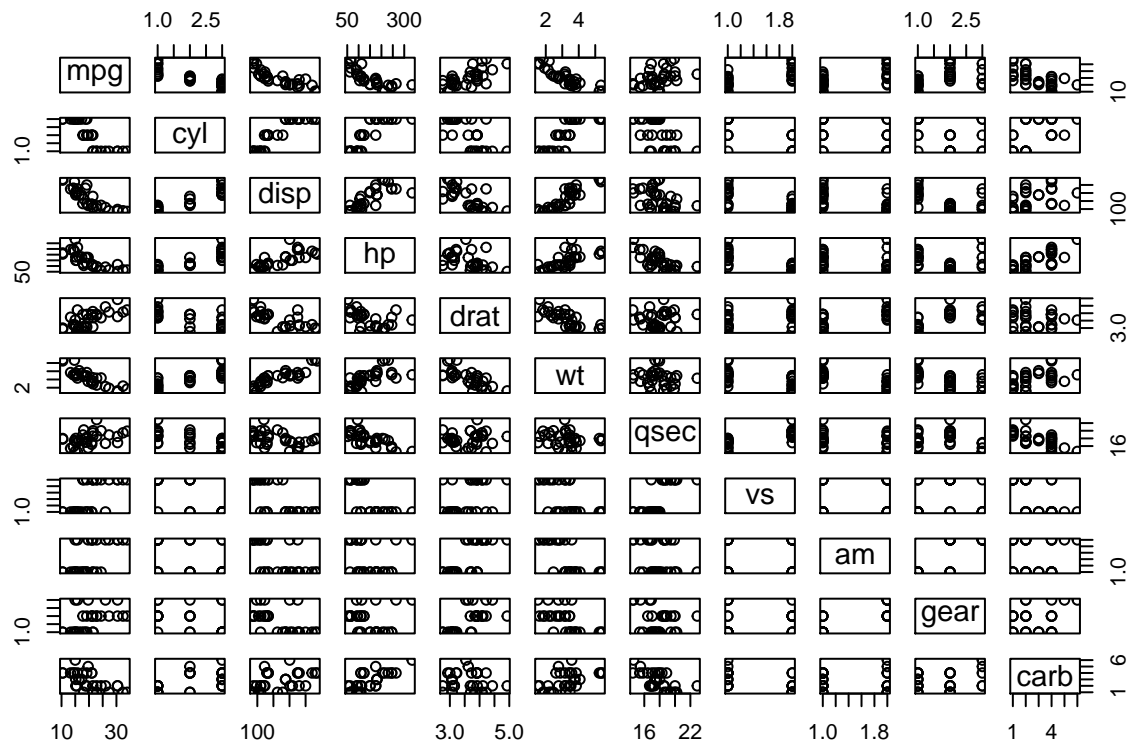
- mpg will increase by 1.64 in cars with ‘Manual’ transmission compared to cars with ‘Automatic’ transmission (adjusted by cyl and disp).
- mpg will decrease by a factor of 4.7 and 5 if number of cylinders increases from 4 to 6 and 8, respectively (adjusted by disp, and am).
- disp will decrease by a factor of .02 (adjusted by cyl, and am).

Appendix:

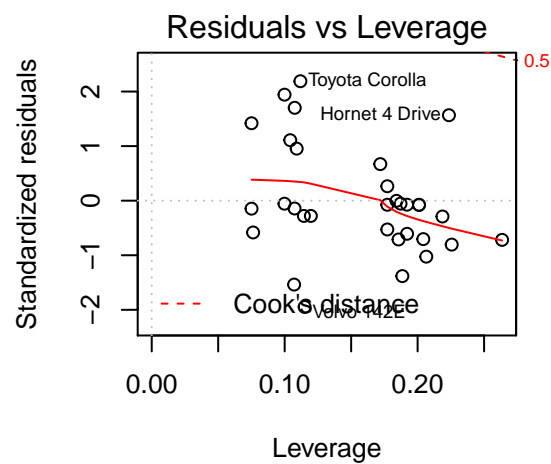
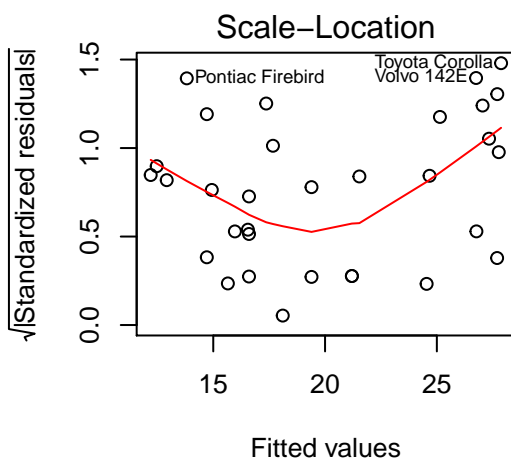
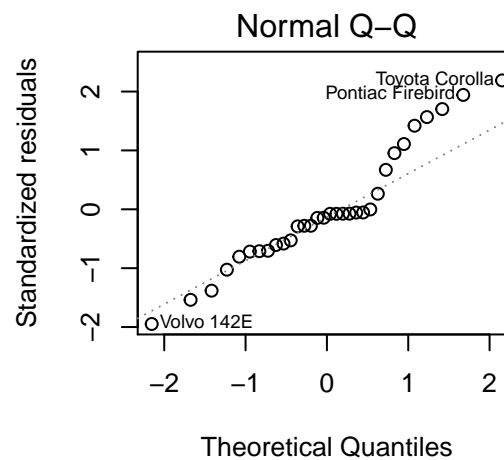
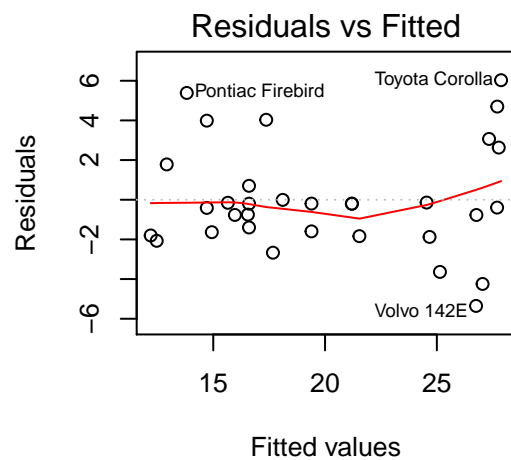
```
boxplot(mpg ~ am, data=mtcars, main="MPG vs Transmission type", xlab="Transmission type", ylab="MPG")
```



```
pairs(mtcars)
```



```
par(mfrow=c(2,2))
plot(fit3, which=1)
plot(fit3, which=2)
plot(fit3, which=3)
plot(fit3, which=5)
```



```
anova(fit1,fit2,fit3,fit4,fit5,fit6,fit7)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 2: mpg ~ am
## Model 3: mpg ~ am + cyl + disp
## Model 4: mpg ~ am + cyl + disp + hp
## Model 5: mpg ~ am + cyl + disp + hp + drat
## Model 6: mpg ~ am + cyl + disp + hp + drat + wt
## Model 7: mpg ~ am + cyl + disp + hp + drat + wt
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      15 120.40
## 2      30 720.90 -15   -600.49  4.9874 0.001759 **
## 3      27 230.46   3    490.44 20.3665 1.512e-05 ***
## 4      26 183.04   1     47.42  5.9078 0.028091 *
```

```
## 5      25 182.38   1      0.66 0.0820 0.778551
## 6      24 150.10   1     32.28 4.0216 0.063310 .
## 7      24 150.10   0      0.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3542 -1.6772 -0.2978  0.9712  6.0274
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.83036    1.96997  14.127 5.44e-14 ***
## amManual     1.63580    1.31741   1.242 0.22503
## cyl6        -4.67997    1.63631  -2.860 0.00807 **
## cyl8        -5.05008    2.86683  -1.762 0.08947 .
## disp        -0.02241    0.01122  -1.997 0.05601 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.922 on 27 degrees of freedom
## Multiple R-squared:  0.7953, Adjusted R-squared:  0.765
## F-statistic: 26.23 on 4 and 27 DF,  p-value: 5.869e-09
```

```
tail(sort(hatvalues(fit3)),3)
```

```
##      Hornet 4 Drive Lincoln Continental  Cadillac Fleetwood
##      0.2235715          0.2256513          0.2635558
```

```
tail(sort(dfbetas(fit3)[,5]),3)
```

```
## Chrysler Imperial  Pontiac Firebird  Hornet 4 Drive
##      0.2267650          0.3413307          0.4006992
```

```
t.test(mpg ~ am, data = 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 Automatic    mean in group Manual
##      17.14737          24.39231
```