Regression Project

January 17, 2015

Executive summary

In this report data gathered by Moto Trends is explored and analyzed in order to determine the relationship between a set of variables and miles per gallon of automobiles. In particlular, the following two issues will be addressed:

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

Take and process data

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

Exploratory analysis

A boxplot(Figure 1) shows the difference between automatic and manual in terms of MPG, manual transmission produces more MPG than automatic transimission. A pairwise graph (Firgure 2) shows what other variables may be of interest.

Model Selection

Basic model using tramsmission as predictors

The baisc model only uses tramsimission as predictors. So a significant relationship does exist with a p-value < 0.001 when other factors dosen't be considered. Vehicles with a manual transmission got on average 7.245 more miles per gallon.

```
basicModel <- lm(mpg ~ am, data = mtcars)
```

A Two Sample t-test was conducted between the different transmission types. The p-value is 0.001374 and difference in means is not equal to 0, the mean in group manual is 24.39231 and the mean in group automatic is 17.14737, it shows that manual transmission has significantly more MPG than automatic.

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

Stepwise-selected model

The step function is used to produces list of the best predictors.In this case, the stepwise model process shows the most significant predictors are wt,qsec,and am. The result is shown in Figure3. The Figure 4 shows mpg comparision with respect to wt and qsec.

Compare basic model with stepwise-selected model

Compared with basic model which only use tramsmission type as predictor, p-value of stepwise-selected model is 1.688435e-08, so, the predictors(wt, qsec)are added into model are significant to improving the model's accuracy.

```
compare <- anova(basicModel, stepwiseModel)
compare$Pr
```

```
## [1]           NA 1.550495e-09
```

Residuals analysis

Figure4 also shows the residuals for the StepWise-selected model. The Residuals vs Fitted plot indicates that this regression model is well fit because no pattern between the residuals and fitted values. The points line up as expected meaning in the Normal Q-Q plot, indicating that the distribtion is normal and the predictions are accurate. In the Scale-Location plot and the Residuals vs Leverage plots, the points are not too far from the center shows no point had too much leverage.

Conclusion

The transimission type has a significant effect on fuel efficiency, manual transmission produces more miles per gallon than automatic transmission does. Moreover, weight,and qsec also have influential effect on fuel efficiency.

Appendix

Figure1. Boxplox for MPG vs Transmission

```
boxplot(mpg~am, data = mtcars, main = "MPG vs Transmission",
        xlab = "Transmission Type", ylab = "MPG",
        names = c("Automatic", "Manual"),
        col = c("lightblue", "pink"))
```

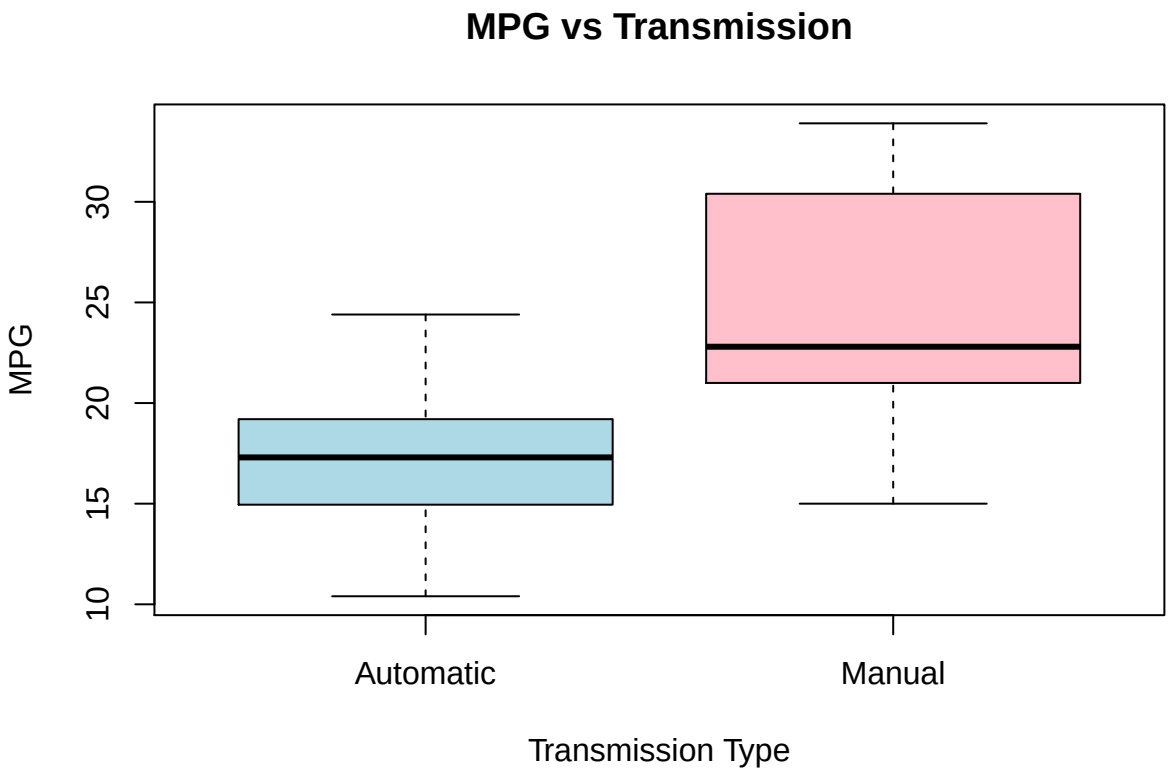


Figure2. Pairwise graph of mtcars

```
plot <- pairs(mtcars, panel = panel.smooth, main ="Pairwise plot of mtcars data")
```

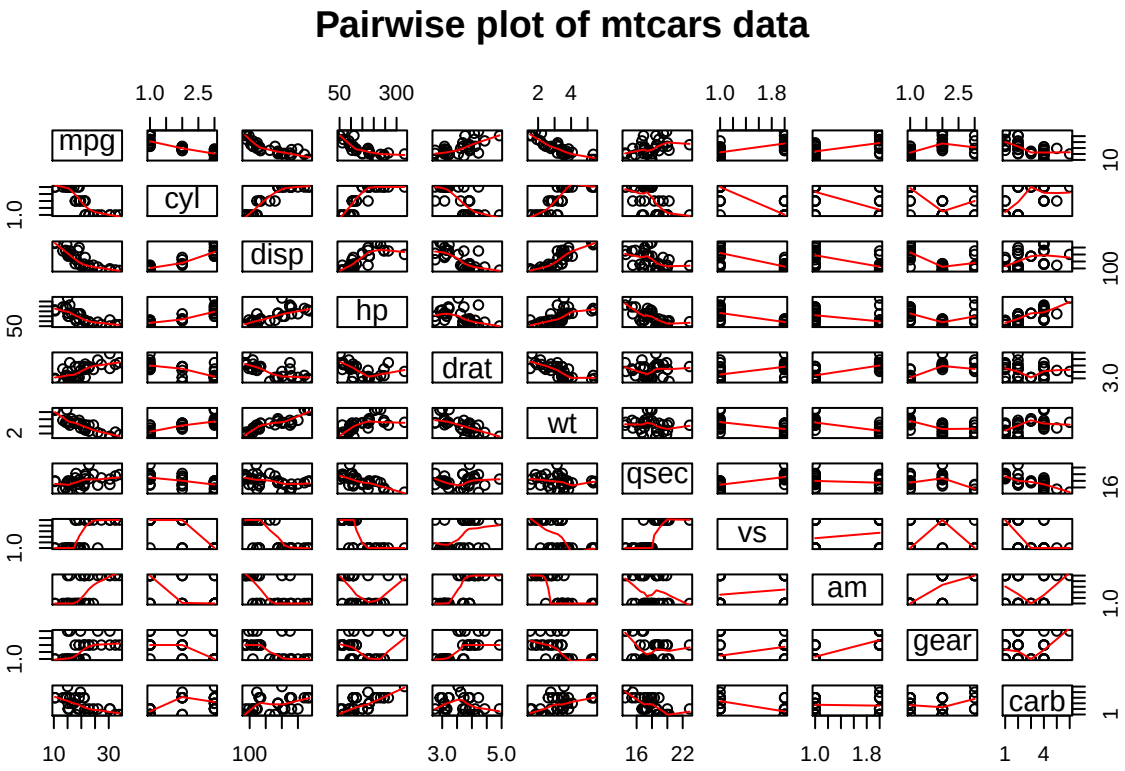


Figure3. StepWise-selected model

```
stepwiseModel <- step(lm(mpg ~ ., data = mtcars), trace = 0)
summary(stepwiseModel)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## ammanual     1.80921138 1.39630450  1.295714 2.064597e-01
```

```
stepwiseModel$anova
```

```
##     Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1         NA         NA        15   120.4027 76.40339
## 2 - carb  5 13.5988573        20   134.0015 69.82769
## 3 - gear  2  5.0215145        22   139.0230 67.00492
## 4 - drat  1  0.9672159        23   139.9903 65.22678
## 5 - disp  1  1.2473996        24   141.2377 63.51066
## 6 - qsec  1  2.4420033        25   143.6797 62.05921
## 7   - vs  1  7.3459298        26   151.0256 61.65483
```

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

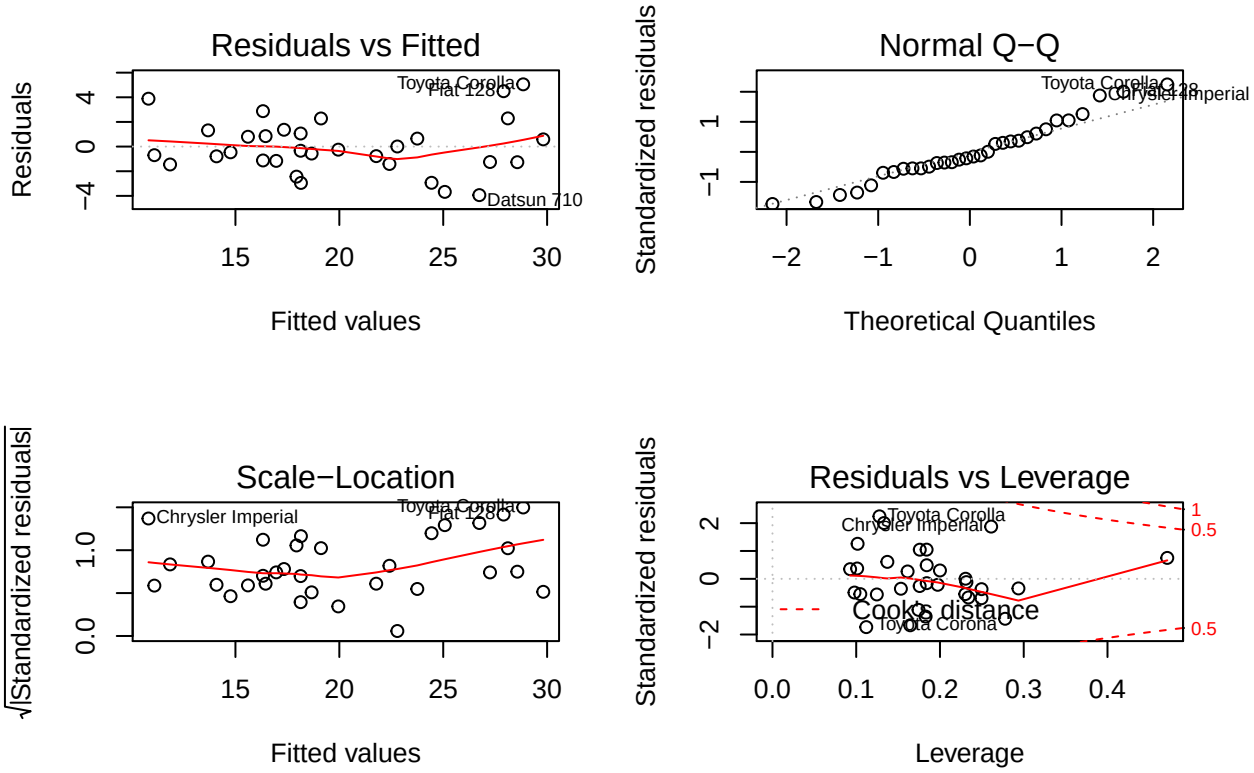
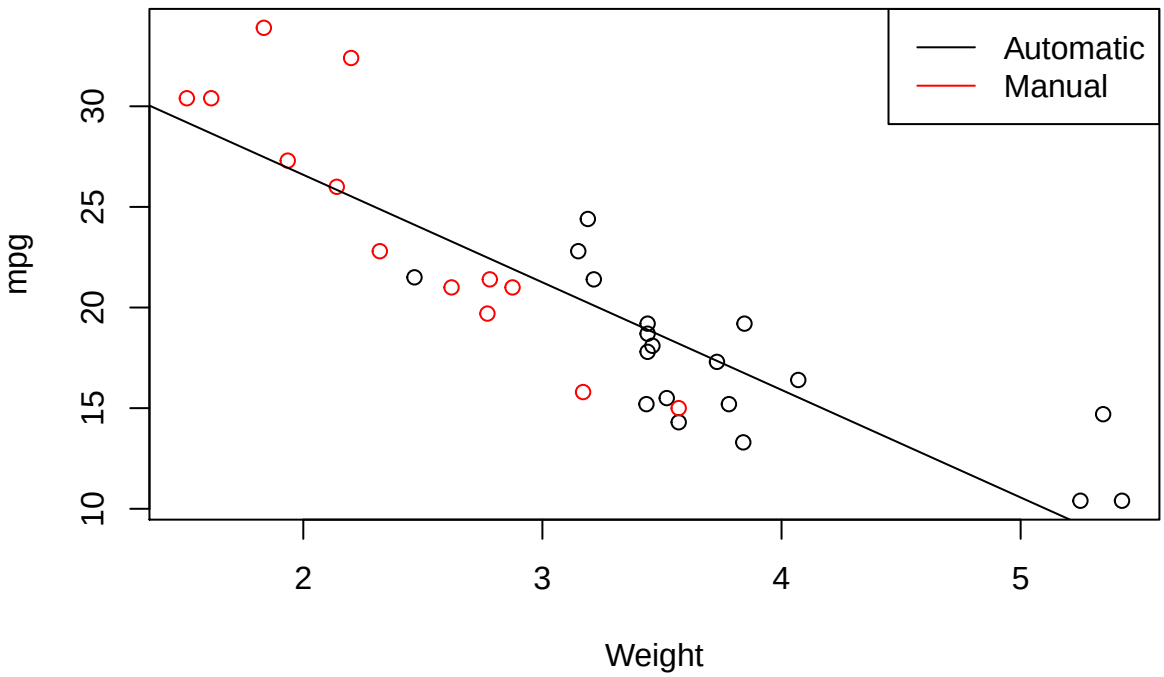


Figure4: MPG vs Weight, MPG vs qsec(1/4 mile time).

```
plot(mpg ~ wt, mtcars, col = am, xlab = 'Weight', ylab = "mpg")
abline(lm(mpg ~ wt, mtcars))
legend('topright', lty = c(1,1), col = 1:2, legend = c('Automatic', 'Manual'))
```




```
plot(mpg ~ qsec, mtcars, col = am, xlab = "qsec(1/4 mile time)", ylab = "mpg")
abline(lm(mpg ~ qsec, mtcars))
legend('topright', lty = c(1,1), col = 1:2, legend = c('Automatic', 'Manual'))
```