CourseReport: Automatic or Manual Transmission for MPG?

Xudong Zhang

August 15, 2014

Executive Summary

In this report, we try to answer the following two question: (1) Is an automatic or manual transmission better for MPG (2) Quantify the MPG difference between automatic and manual transmissions.

With some analysis, we came to a conclusion that manual transmission is better than automatic for MPG. More specifically, changing from automatic to manual will bring about an increase of 2.94 in mpg.

Exploratory Analysis

First we took a look at the data format and distribution of mtcars.

```
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 ...
```

The plotting of dataset is shown in Figure 1. We noticed that four attributes should be regarded as factor: cyl, vs, am and gear(card attribute have too many kinds of values and we do not treat it as factor variable). Then we perform T test on the model which takes mpg as outcome and factorized am as predictor. The two-tailed Student's T-Test for the groups shows that automatic and manual cars are drawn from populations with different means (95% CI), i.e. 24.39 for manual and 17.15 for automatic transmission.

```
t.test(mpg ~ factor(am), data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by factor(am)
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
## mean in group 0 mean in group 1
##           17.15           24.39
```

The plotting of this regression model is shown in Figure 2, which shows that considering only am as regressor, cars with manual transmission have higher average mpg that those with automatic transmission.

Regression Analysis

We first use all attributes as regressors to fit the linear model. The following result shows that all of the regressior is not significant(p-value > 0.05).

```
fit <- lm(mpg ~ factor(cyl) + disp + hp + drat + wt + qsec + factor(vs) + factor(am)
          + factor(gear) + carb, data = mtcars)
summary(fit)$coef
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.09262   17.13627  0.8807  0.38946
## factor(cyl)6  -1.19940    2.38736 -0.5024  0.62116
## factor(cyl)8   3.05492    4.82987  0.6325  0.53460
## disp           0.01257    0.01774  0.7085  0.48727
## hp            -0.05712    0.03175 -1.7992  0.08789
## drat           0.73577    1.98461  0.3707  0.71494
## wt            -3.54512    1.90895 -1.8571  0.07887
## qsec           0.76801    0.75222  1.0210  0.32008
## factor(vs)1    2.48849    2.54015  0.9797  0.33956
## factor(am)1    3.34736    2.28948  1.4621  0.16007
## factor(gear)4 -0.99922    2.94658 -0.3391  0.73824
## factor(gear)5  1.06455    3.02730  0.3516  0.72897
## carb           0.78703    1.03599  0.7597  0.45677
```

Therefore, we need to pick up several influential attributes by AIC in a stepwise algorithm:

```
step_res <- step(fit,trace=0,steps=1000, direction="both")
step_res$call
```

```
## lm(formula = mpg ~ wt + qsec + factor(am), data = mtcars)
```

Three attributes are chosen by the stepwise algorithm: wt, qsec and am. Next we fit the regression model using the above three attributes:

```
fit2 <- lm(formula = mpg ~ wt + qsec + factor(am), data = mtcars)
coef <- summary(fit2)$coef
coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## factor(am)1    2.936     1.4109   2.081 4.672e-02
```

```
intv <- coef[4, 1] + c(-1, 1) * qt(0.975, df = fit$df) * coef[4, 2]
intv
```

```
## [1] -0.01722  5.88889
```

The result shown that

- All three regressors is significant.
- For wt, one unit increase in wt will lead to about 4 unit decrement in mpg.
- For qsec, one unit increase in qsec will lead to about 1.23 unit increment in mpg.
- For wt, switching from automatic to manual transmission will lead to about 2.94 unit increment in mpg (note that for am, 0 means automatic and 1 means manual).

We validate the above regression model by using the plot summary of residuals of the model, shown in Figure 3. The "Residuals vs Fitted" and "Scale-Location" charts show that there is no trend to the residuals. The Q-Q plot shows that the the errors are approximately normally distributed.

Conclusion

From the above the analysis, we can tell that manual transmission is better than automatic for mpg with 2.94, with 95% confidence, change from -0.017 to 5.89.

Appendix

Figure 1 The plotting of dataset mtcars

```
plot(mtcars, panel = panel.smooth)
```

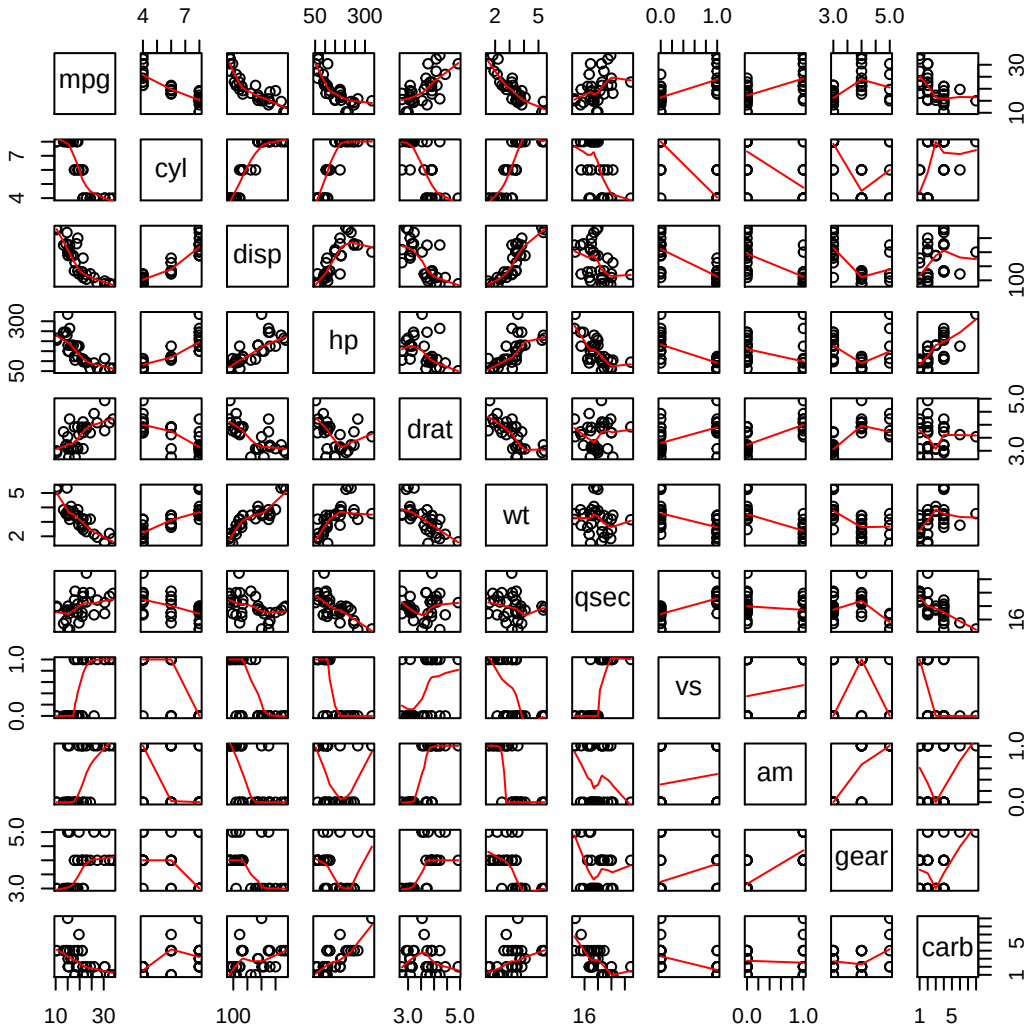


Figure 2 Box plot for the regression model mpg~factor(am)

```
plot(mpg ~ factor(am), data = mtcars)
```

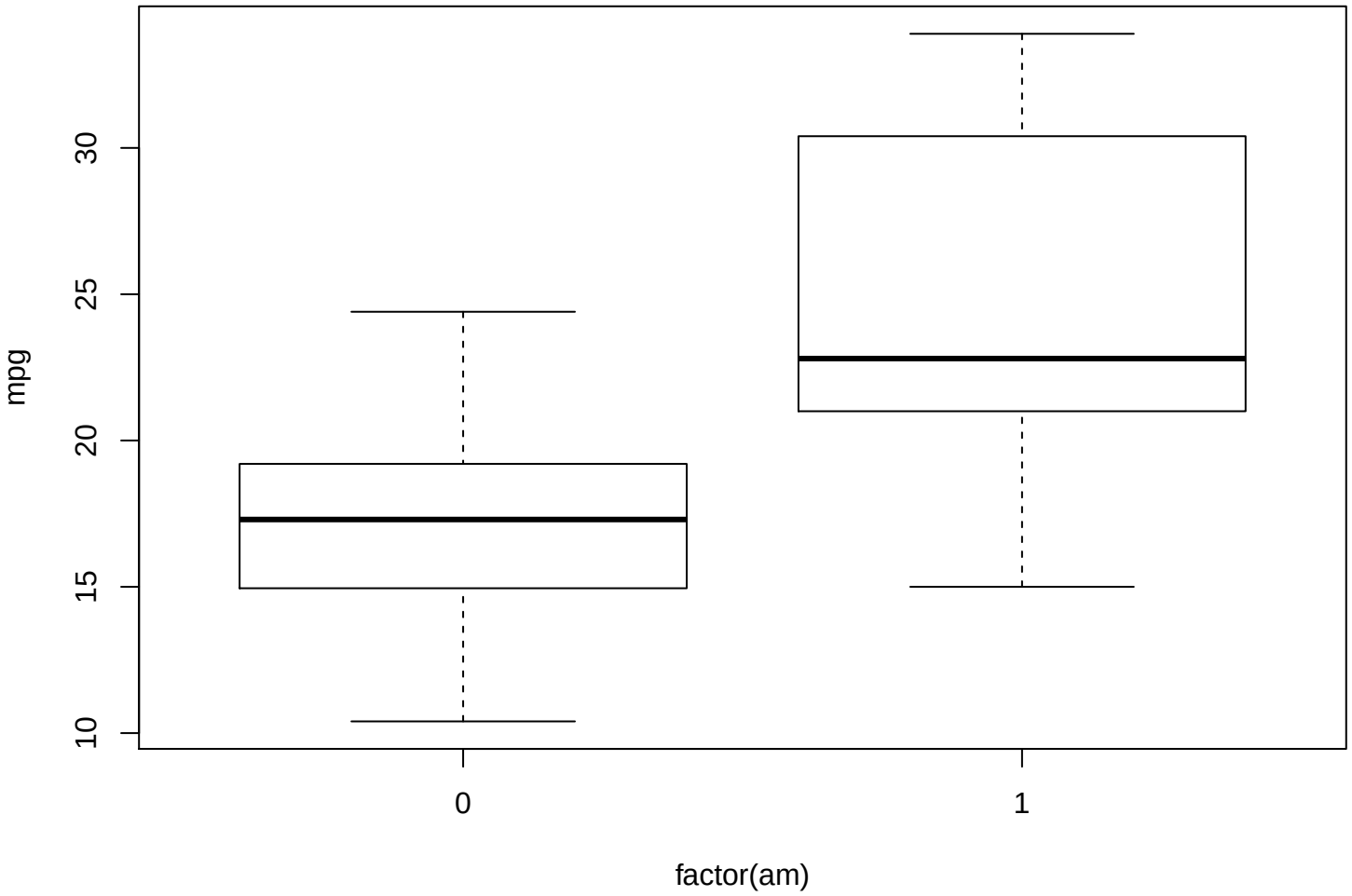
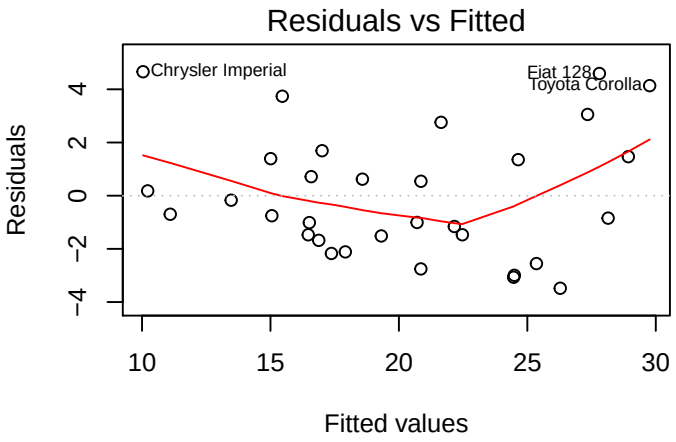
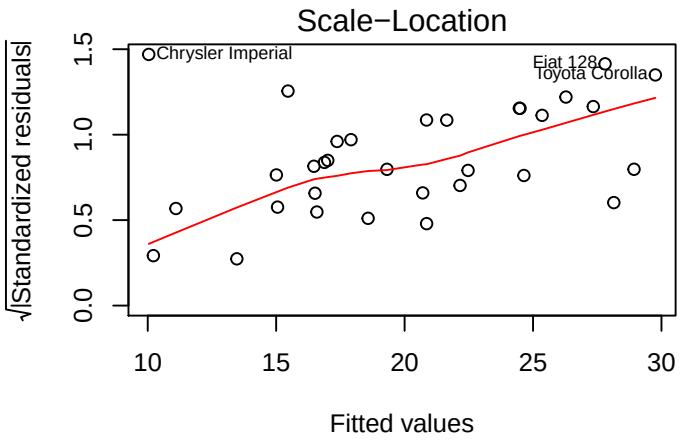
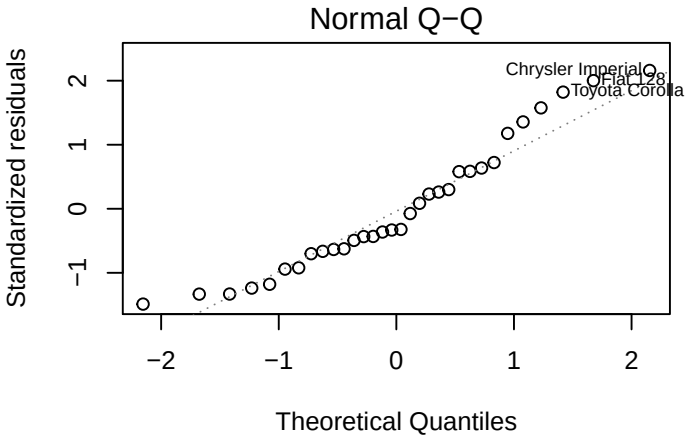


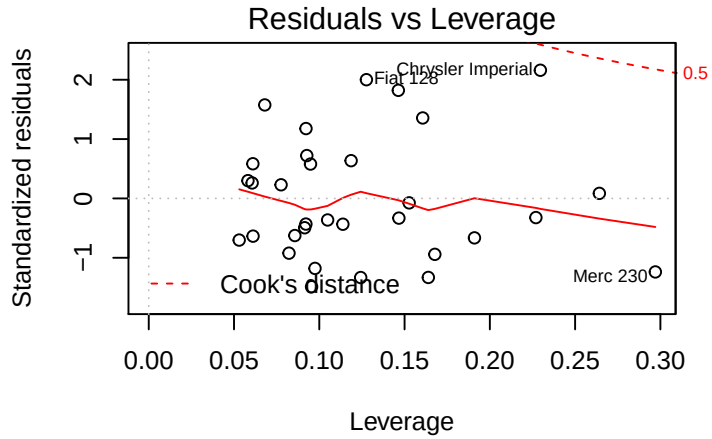
Figure 3 The plotting the regression with qec, wt, am as predictors and mpg as outcome

```
layout(matrix(c(1,2,3,4),2,2))
plot(fit2)
```


Residuals vs Fitted
Chrysler Imperial
Fiat 128
Toyota Corolla
Residuals
Fitted values


Scale−Location
Chrysler Imperial
Fiat 128
Toyota Corolla
√|Standardized residuals|
Fitted values


Normal Q−Q
Chrysler Imperial
Fiat 128
Toyota Corolla
Standardized residuals
Theoretical Quantiles


Residuals vs Leverage
Chrysler Imperial
Fiat 128
Merc 230
Cook's distance
0.5
Standardized residuals
Leverage