Motor Trend Analysis

Damien Edwards

February 20, 2015

Is an automatic or manual transmission better for gas mileage(MPG)?

Executive Summary This report for Motor Trend will use data collected to determine the relationship between a set of variables and miles per gallon(MPG)outcome. We will need to determine is an automatic or manual transmission better for MPG. Also, we will need to quantify the MPG difference between automatic and manual transmissions.

Data Preparation The mtcars data set is used for the the data analysis

```
library(datasets)
head(mtcars)
```

Model

Do a multivariable reqgression on the mtcars dataset

```
summary(lm(mpg ~ ., data = mtcars))$coefficients
```

The backward-elimination strategy is used to eliminate the unrelated variables one-at-a-time. Remove cyl from the data set because of the high p-value and refit the model. (See Appendix)

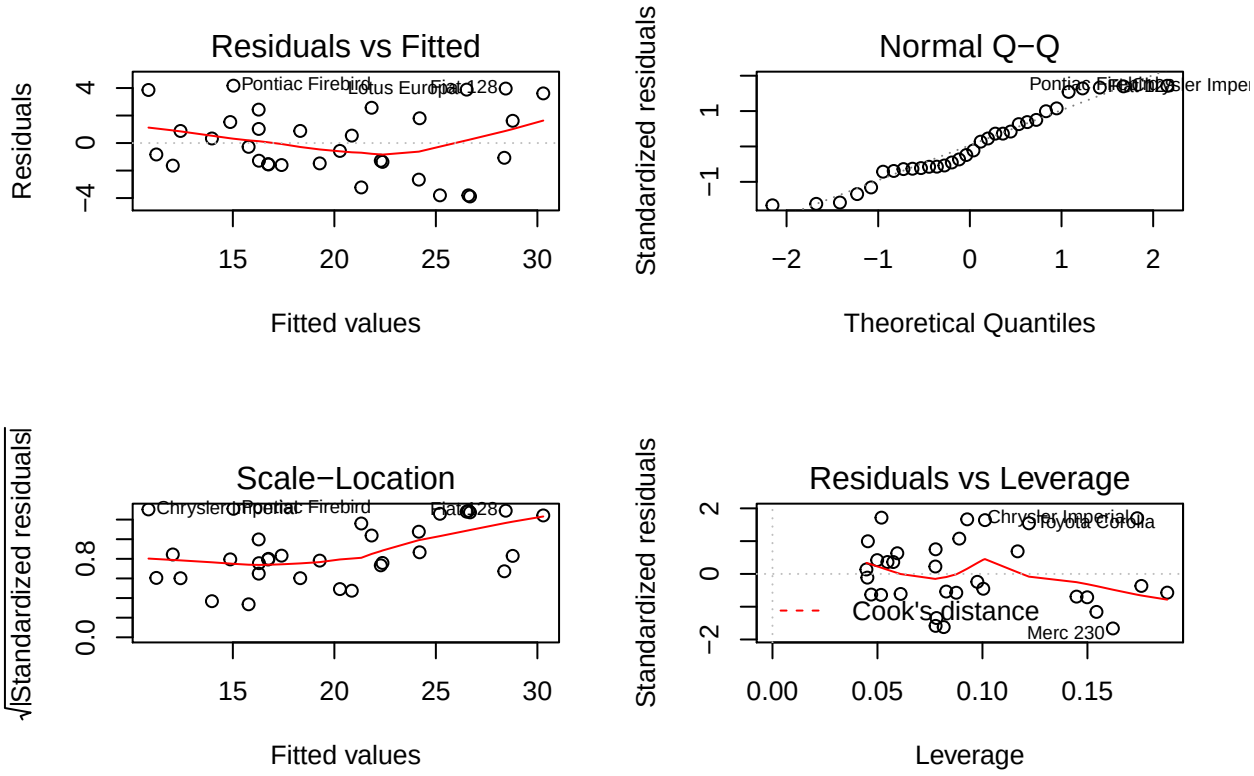
The variables used are wt,qsec,am in the final model. Those variables are used to predict the miles per gallon of a vehichle.

```
d1 <- mtcars[, c("mpg", "wt", "qsec", "am")]
fit <- lm(mpg ~ . - 1, data = d1)
summary(fit)$coefficients
```

```
##       Estimate Std. Error   t value     Pr(>|t|)
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## am    4.299519  1.0241147  4.198279 2.329423e-04
```

Exploratory Analysis

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



Statistical Inference

Using a t-test.(p-value = 0.001374) The null Hypothesis that the transmission doesn't have an effect on the MPG is thrown out if the p=value is greater than .05,

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

```
Coef <- summary(fit)$coefficients
Coef_intv <- Coef["am", 1] + c(-1, 1) * qt(0.975, df = fit$df) * Coef["am",
    2]
Coef_intv
```

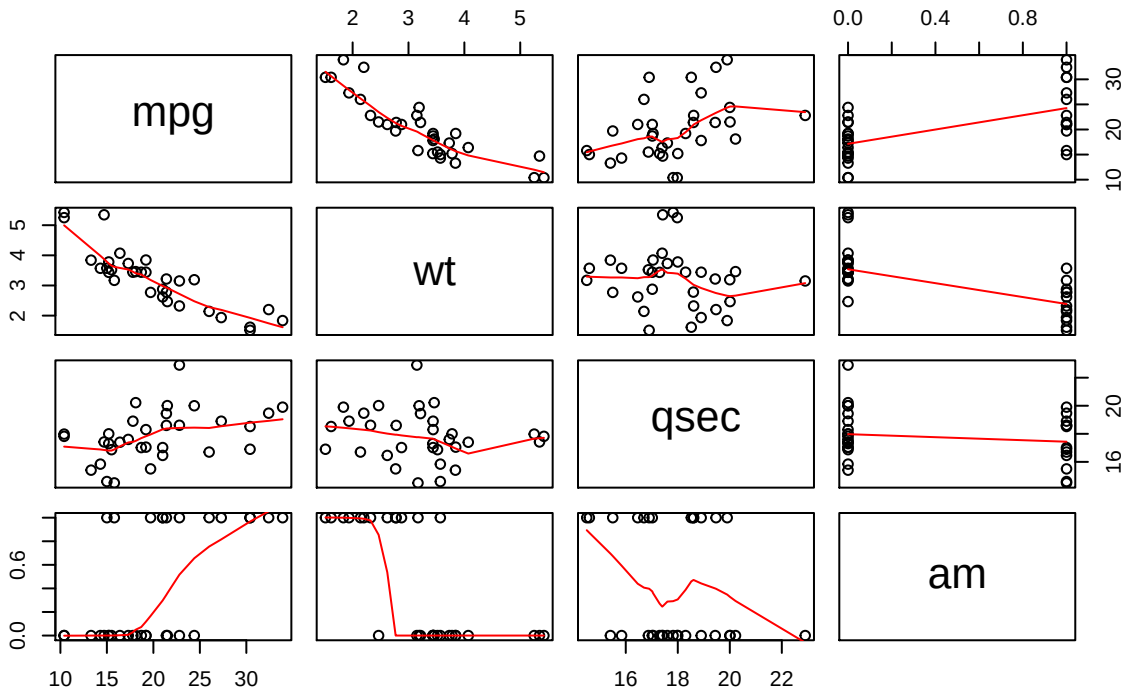
```
## [1] 2.204969 6.394069
```

Conclusion

The transmision of a car has a significicant effect on miles per gallon.The change from automatic to manual transmissions results in 2.2 to 6.39 increase in mpg for cars with 95% confidence. The manaul transmission is better for mpg.

Appendix

```
pairs(d1, panel = panel.smooth, main = "mtcars dataset")
```




```
d1 <- mtcars
d1 <- d1[, names(d1) != "cyl"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 10.96007405 13.53030251  0.8100391 0.42659327
## disp         0.01282839  0.01682215  0.7625891 0.45380797
## hp          -0.02190885  0.02091131 -1.0477031 0.30615002
## drat         0.83519652  1.53625251  0.5436584 0.59214373
## wt          -3.69250814  1.83953550 -2.0073046 0.05715727
## qsec         0.84244138  0.68678068  1.2266527 0.23291993
## vs           0.38974986  1.94800204  0.2000767 0.84325850
## am           2.57742789  1.94034563  1.3283344 0.19768373
## gear         0.71155439  1.36561933  0.5210489 0.60753821
## carb        -0.21958316  0.78855537 -0.2784626 0.78325783
```

```
d1 <- d1[, names(d1) != "vs"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept)  9.76827789 11.89230469  0.8213949 0.41985460
## disp         0.01214441  0.01612373  0.7532010 0.45897019
## hp          -0.02095020  0.01992567 -1.0514175 0.30398892
## drat         0.87509822  1.49112525  0.5868710 0.56300717
## wt          -3.71151106  1.79833544 -2.0638592 0.05049085
## qsec         0.91082822  0.58311935  1.5619928 0.13194532
## am           2.52390094  1.88128007  1.3415870 0.19282690
## gear         0.75984464  1.31577205  0.5774896 0.56921947
## carb        -0.24796312  0.75933250 -0.3265541 0.74695821
```

```r
d1 <- d1[, names(d1) != "carb"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate  Std. Error    t value    Pr(>|t|)
## (Intercept)  9.19762837 11.54220381  0.7968693 0.433339841
## disp         0.01551976  0.01214235  1.2781513 0.213420001
## hp          -0.02470716  0.01596302 -1.5477746 0.134763097
## drat         0.81022794  1.45006779  0.5587518 0.581507634
## wt          -4.13065054  1.23592980 -3.3421401 0.002717119
## qsec         1.00978651  0.48883274  2.0657097 0.049814778
## am           2.58979984  1.83528342  1.4111171 0.171042438
## gear         0.60644020  1.20596266  0.5028681 0.619640616
```

```r
d1 <- d1[, names(d1) != "gear"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate  Std. Error    t value    Pr(>|t|)
## (Intercept) 10.71061639 10.97539399  0.9758753 0.338475309
## disp         0.01310313  0.01098299  1.1930387 0.244054196
## hp          -0.02179818  0.01465399 -1.4875257 0.149381426
## drat         1.02065283  1.36747598  0.7463772 0.462401185
## wt          -4.04454214  1.20558182 -3.3548467 0.002536163
## qsec         0.99072948  0.48002393  2.0639168 0.049550895
## am           2.98468801  1.63382423  1.8268110 0.079692318
```

```r
d1 <- d1[, names(d1) != "drat"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 14.36190396 9.74079485  1.474408 0.152378367
## disp         0.01123765 0.01060333  1.059823 0.298972150
## hp          -0.02117055 0.01450469 -1.459565 0.156387279
## wt          -4.08433206 1.19409972 -3.420428 0.002075008
## qsec         1.00689683 0.47543287  2.117853 0.043907652
## am           3.47045340 1.48578009  2.335779 0.027487809
```

```r
d1 <- d1[, names(d1) != "disp"]
summary(lm(mpg ~ ., data = d1))$coefficients
```

```
##                Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 17.44019110  9.3188688  1.871492 0.072149342
## hp          -0.01764654  0.0141506 -1.247052 0.223087932
## wt          -3.23809682  0.8898986 -3.638726 0.001141407
## qsec         0.81060254  0.4388703  1.847021 0.075731202
## am           2.92550394  1.3971471  2.093913 0.045790788
```

```r
d1 <- d1[, names(d1) != "hp"]
summary(lm(mpg ~ ., data = d1))$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
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am           2.935837  1.4109045  2.080819 4.671551e-02
```

The p-values are smaller for the remaining variables so we can stop removing variables..