Motor Trend Data Analysis

Philip Bulsink

2017-02-18

How does Transmission Type Effect Milage?

Executive Summary

This report will outline the effect of transmission type (automatic vs. manual) on gas milage. The influence of other variables will be investigated as well as confounding factors. It is shown that

Introduction

The data to be studied is available in `R` in the `mtcars` frame. This data contains information on 32 automobiles from 1973-1974, listing a total of 10 variables on each car. These variables include the number of cylinders, the engine size (displacement), the horsepower, the weight, the number of gears, and the transmission type, amongst others. This is easily loaded into `R` with one command.

```
data("mtcars")
```

Data Preparation

A few of the variables we will be investigating in this report can be considered 'factors', that is, of a non-numeric or a counted type. This includes transmission type, and may include number of gears, number of carburetors, number of cylinders, or if the engine is straight or V shaped.

```
mtcars$am<-factor(mtcars$am, labels = c("Automatic", "Manual"))
mtcars$vs<-factor(mtcars$vs, labels = c("V", "Straight"))
mtcars$cyl<-as.factor(mtcars$cyl)
mtcars$carb<-as.factor(mtcars$carb)
mtcars$gear<-as.factor(mtcars$gear)
```

Data Exploration

A first glance at the data suggests that there is an effect on milage (mpg) by transimission type (am), whereby manual transmissions have better (higher) gas milage. See a plot of this data in Appendix 1.

We can quantify this by regressing the data over this factor:

```
fit1<-lm(mpg ~ am, data=mtcars)
summary(fit1)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     17.1        1.1    15.2  1.1e-15
## amManual         7.2        1.8     4.1  2.9e-04
```

Thus, we can see that, taking no other factors into play, the average milage of cars with a manual transmission is approximately 7.24 better than that of an automatic transmission, which has an average milage of 17.15. The p value for this estimate is very low, at 2.85×10^{-4}, allowing us to reject the null hypothesis that there is no difference.

Detailed Data Analysis

What if the enhancement of milage in manual transmissions is due to some other factor? For example, in Appendix 2, we show the milage as a function of transmission type, but broken up by cylinder. For 8 cylinder cars, it looks like manual may not be any better than automatic, contrary to our above discovery. We'll test for other factors like this as well, before validating which matter.

```
fit2 <- lm(mpg~am + cyl, data=mtcars)
```

If we include cylinders in our model, we then find that cars with a manual transmission is approximately 2.56 better than that of an automatic transmission, which has an average milage of 24.8. The benefit of a manual transmission is much reduced when considering cylinder count.

```
fit3 <- lm(mpg~am + cyl + wt, data=mtcars)
fit4 <- lm(mpg~am + cyl + wt + gear, data=mtcars)
fit5 <- lm(mpg~am + cyl + wt + gear + hp, data=mtcars)
fit6 <- lm(mpg~am + cyl + wt + gear + hp + carb, data=mtcars)
fit7 <- lm(mpg~am + cyl + wt + gear + hp + carb + disp, data=mtcars)
fit8 <- lm(mpg~am + cyl + wt + gear + hp + carb + disp + drat, data=mtcars)
fit9 <- lm(mpg~am + cyl + wt + gear + hp + carb + disp + drat + vs, data=mtcars)
fit10 <- lm(mpg~am + cyl + wt + gear + hp + carb + disp + drat + vs + qsec, data=mtcars)

anova(fit1, fit2, fit3, fit4, fit5, fit6, fit7, fit8, fit9, fit10)[1:6]
```

```
##    Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1      30 721
## 2      28 264  2       456 28.43 7.9e-06 ***
## 3      27 183  1        82 10.16  0.0061 **
## 4      25 175  2         8  0.51  0.6106
## 5      24 150  1        25  3.13  0.0972 .
## 6      19 140  5        10  0.25  0.9329
## 7      18 128  1        12  1.46  0.2456
## 8      17 126  1         2  0.26  0.6165
## 9      16 122  1         4  0.52  0.4839
## 10     15 120  1         1  0.15  0.6997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see which of these factors are confounding using

```
#code
```

Final Model

Knowing that we can fit the best model with by including , we do so:

```
#code
fitfinal<-lm(mpg~am+cyl, data=mtcars)
summary(fitfinal)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     24.8        1.3    18.8  2.2e-17
## amManual         2.6        1.3     2.0  5.8e-02
## cyl6            -6.2        1.5    -4.0  4.1e-04
## cyl8           -10.1        1.5    -6.9  1.5e-07
```

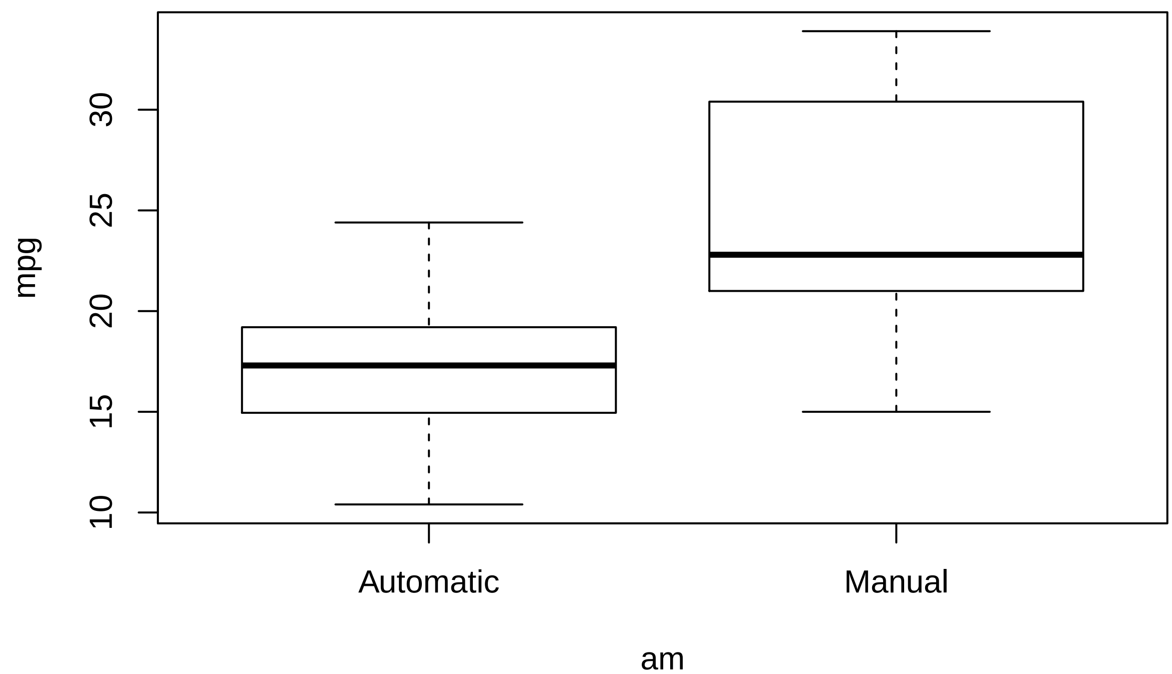
Thus, we can see that, with appropriate factors taken into consideration, the average manual transmission has a mpg of 2.56 higher than that of an automatic transmission at -6.16. This is in addition to other factors, but has a p value of 2.85×10^{-4}, showing its significance.

Analysis of the residuals shows that there is no obvious trend, and the QQ plot appears approprate.

Appendices

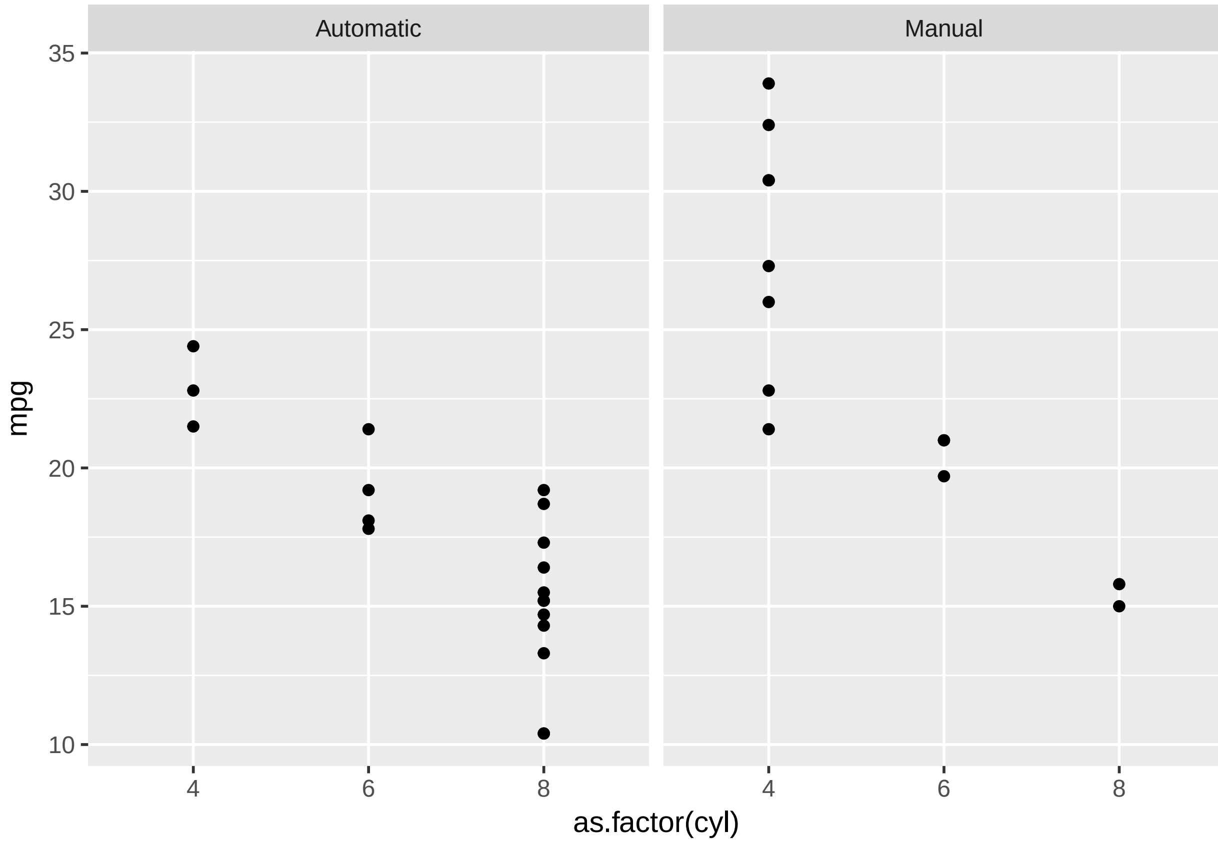
Appendix 1: Plot of milage vs. transmission type:

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



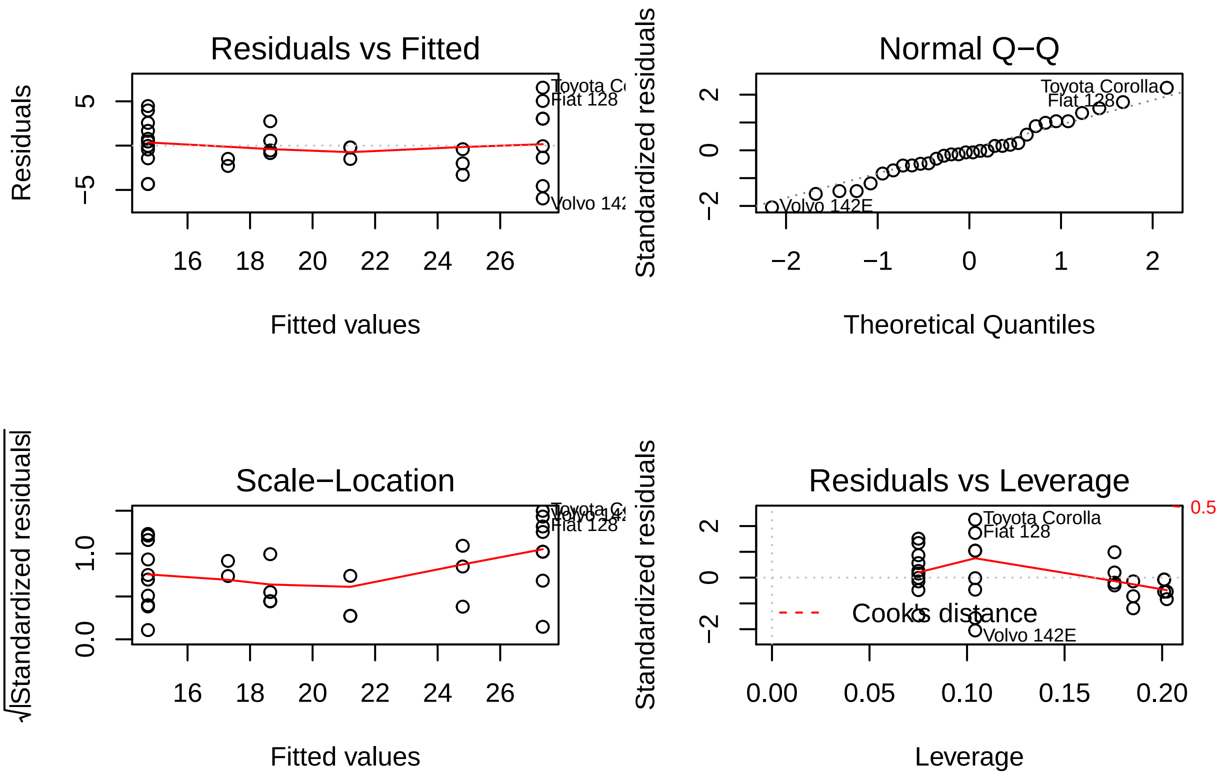
Appendix 2: Plot of milage vs. cylinder count and transmission type:

```r
ggplot(mtcars, aes(x=as.factor(cyl), y=mpg)) + geom_point() + facet_grid(.~am)
```



Appendix 3: Residuals, QQ, Leverage

```r
par(mfrow=c(2,2))
plot(fitfinal)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Co
Fiat 128
Volvo 142
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Volvo 142E
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5