Regression Models Course Project

Lisa Hlmsch

5 11 2017

Executive Summary

Exploration of the mtcars dataset has revealed that there is an obvious difference between the gas mileage for automatic and manual transmission. The data suggestes that weight and acceleration are confounding variables in the relationship between transmission and gas milage. Including these confounding variables, a manual transmission car has better fuel efficiency of 2.94 miles/gallon than automatic transmission cars.

Introduction

In this project I am going to explore the mtcars data set, which comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). The following question will be addressed:

Is an automatic or manual transmission better for MPG (Miles/(US) gallon)?

Setup

```
library(ggplot2)
library(datasets)

set.seed(123)
data(mtcars)
```

Dataset exploration

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Converting transmission variable (am) from numeric to factor variable:

```
mtcars$am <- factor(mtcars$am, labels = c("automatic", "manual"))
str(mtcars$am)
```

```
##  Factor w/ 2 levels "automatic","manual": 2 2 2 1 1 1 1 1 1 1 ...
```

A histogram of milage counts per transmission type (see appendix) lets us assume that average miles/gallon is higher for manual transmission than for automatic. The boxplot (see appendix) also shows an obvious difference between the gas mileage for automatic and manual transmission.

```
aggregate(mpg~am, data=mtcars, mean)
```

```
##          am      mpg
## 1 automatic 17.14737
## 2    manual 24.39231
```

This assumption is true, as this table shows that average miles/gallon of automatic transmission is 17, which is roughly 7 mpg lower than that of manual transmission cars.

Model Selection

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

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## ammanual     7.244939   1.764422  4.106127 2.850207e-04
```

7.24 ist the estimated average miles/gallon comparing manual to automatic transmission This relationship is highly statistical significant. (p: 2.850207e-04 < 0.05, so we reject H0 that the coefficient of am is 0.)

```
summary(fit1)$r.squared
```

```
## [1] 0.3597989
```

R-squared is 0.36 which indicated that this model only explains 36% of the variance, attributed to transmission type alone.

For model improvement we include other variables, using the step function and build a multivariate linear regression. The step function only selects the most correlated variables of the dataset

```
fit2 <- step(lm(mpg ~ ., data = mtcars), trace=0) # Multivariate Linear Regression
summary(fit2)$coef
```

```
##              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
## ammanual     2.935837  1.4109045  2.080819 4.671551e-02
```

The model includes car weight, acceleration and transmission. A manual transmission car has better fuel efficiency of 2.94 miles/gallon than automatic transmission cars.

```
summary(fit2)$r.squared
```

```
## [1] 0.8496636
```

85% of mpg variance is explained by the model, which is an improvement to the first model.

Comparison of both models using ANOVA

```
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value indicates that we should reject the null hypothesis that the means from both models are the same. We can thus conclude that weight and acceleration are confounding variables in the relationship between transmission and gas milage.

The Residual vs Fitted plot (see appendix) hows that the residuals are homoscedastic and normally distributed, with the exception of some outliers.

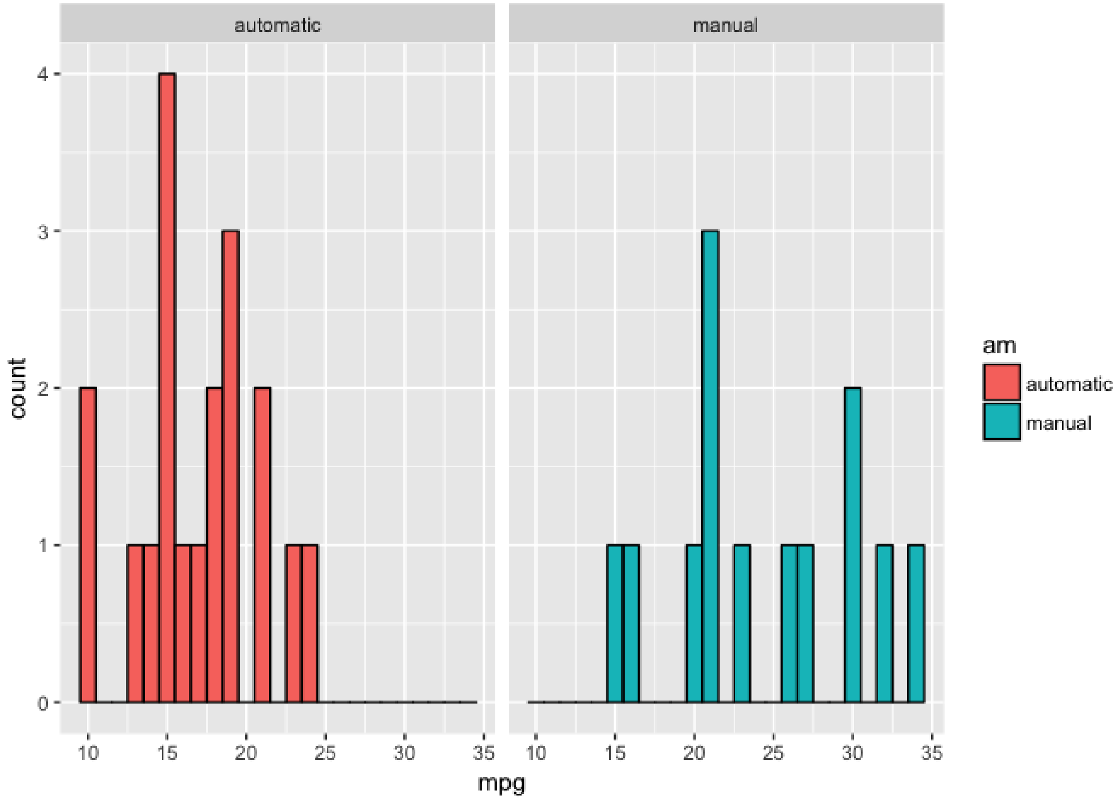
Conclusion

Data supports the claim that automatic transmission is better for MPG (Miles/(US) gallon) than manual

Appendix

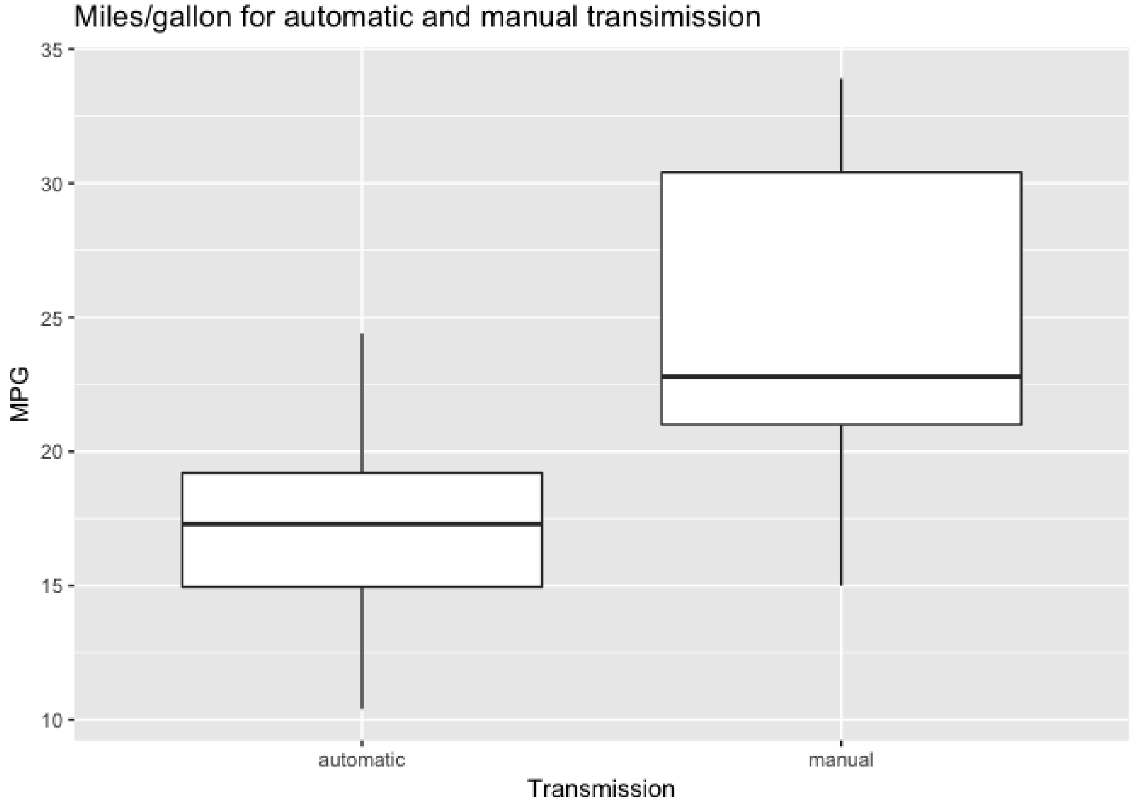
Histogram of transmission counts:

```
ggplot(mtcars, aes(x = mpg, fill = am)) +
  geom_histogram(colour = "black", binwidth = 1) +
  facet_grid(.~am)
```



Boxplot of MPGs per transmission type

```
ggplot(mtcars, aes(x = am, y = mpg)) +
  geom_boxplot() +
  labs(x= "Transmission", y = "MPG", title = "Miles/gallon for automatic and manual transimission")
```



Residuals plot:

Residual represents variation left unexplained by our model. Whenn plotting the residuals we are able to examine any heteroskedacity between the fitted and residual values or any abnormality.

```
par(mfrow = c(2,2))
plot(fit2, col = 4)
```