Regression Models Course Project

Vehicle Mileage as a function of Automatic or Manual Transmission

Executive Summary

Here we'll look at the mtcars dataset and see the effect on vehicle mileage (mpg) of vehicle transmission (am). There are only two vehicle transmission types: automatic or manual. In the data, automatic transmission is coded as 0 and manual transmission is 1.

We'll answer to basic questions: 1) Is an automatic or manual transmission better for MPG? 2) What is the MPG difference between automatic and manual transmissions?

Data Analysis & Discussion

```
library(UsingR)
data(mtcars)
```

First, let's look at average mileage for manual and automatic transmission seperately.

```
mean(mtcars[mtcars$am==0, 'mpg'])
```

```
## [1] 17.14737
```

```
mean(mtcars[mtcars$am==1, 'mpg'])
```

```
## [1] 24.39231
```

From the above analysis and from the first plot in Appendix 1, we can see that vehicles with manual transmission clearly have better mileage than vehicles with automatic transmission. Not only is the mean mileage much higher for manual transmission vehicles, even the spreads do not overlap, meaning the worst manual transmission vehicles have better mileage than the best automatic transmission vehicles.

Now let's create a linear model of this data to examine it further.

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

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

From the above summary of the linear model, we can see that our p-values show a significant difference between vehicles with manual versus automatic transmission. The Estimates shown in the Coeffecients correspond to the mean mileage (mpg) for manual transmission (factor(am)0) and automatic transmission (factor(am)1).

But what if other variables are confounding the manual versus automatic data? What if weight or the number of cylinders are very different for manual versus automatic transmission vehicles? Let's add these other variables in and see what happens.

```
mpgvsam2 <- lm(mpg ~ factor(am) + wt, data=mtcars)
mpgvsam3 <- lm(mpg ~ factor(am) + wt + factor(cyl), data=mtcars)
mpgvsam4 <- lm(mpg ~ factor(am) + wt + factor(cyl) + wt*factor(cyl), data=mtcars)
anova(mpgvsam, mpgvsam2, mpgvsam3, mpgvsam4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + factor(cyl)
## Model 4: mpg ~ factor(am) + wt + factor(cyl) + wt * factor(cyl)
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 71.9823 7.916e-09 ***
## 3     27 182.97  2     95.35  7.7541  0.002399 **
## 4     25 153.71  2     29.26  2.3793  0.113263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the analysis of variance here, we see that the p values are quite significant. This means that vehicle weight and the number of cylinders are strong confounding factors in the relationship between vehicle mileage (mpg) and transmission type (am).

However, including an interaction term for weight and the number of cylinders yields a p-value that is insignificant. This suggests it's best to leave out the interaction term.

Let's look at the linear model that relates mpg to transmission, weight, and the number of cylinders, without any interaction terms. Let's use this as our definitive model.

```
summary(mpgvsam3)$coef
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)   33.7535920  2.8134831 11.9970836 2.495549e-12
## factor(am)1    0.1501031  1.3002231  0.1154441 9.089474e-01
## wt            -3.1495978  0.9080495 -3.4685309 1.770987e-03
## factor(cyl)6 -4.2573185  1.4112394 -3.0167231 5.514697e-03
## factor(cyl)8 -6.0791189  1.6837131 -3.6105432 1.227964e-03
```

This model leads us to believe that automatic transmission vehicles have slightly better mileage for a given weight and number of cylinders than manual transmission vehicles. However, the p-value for this is not that small, meaning we can't make this statement with the highest confidence.

Let's also look at the hat values of our linear model to see if there are any outliers biasing the data. The hat values are shown in Appendix 2. Looking at them, none of our points stand out as being outliers. Therefore we choose not to omit any vehicle from our linear models.

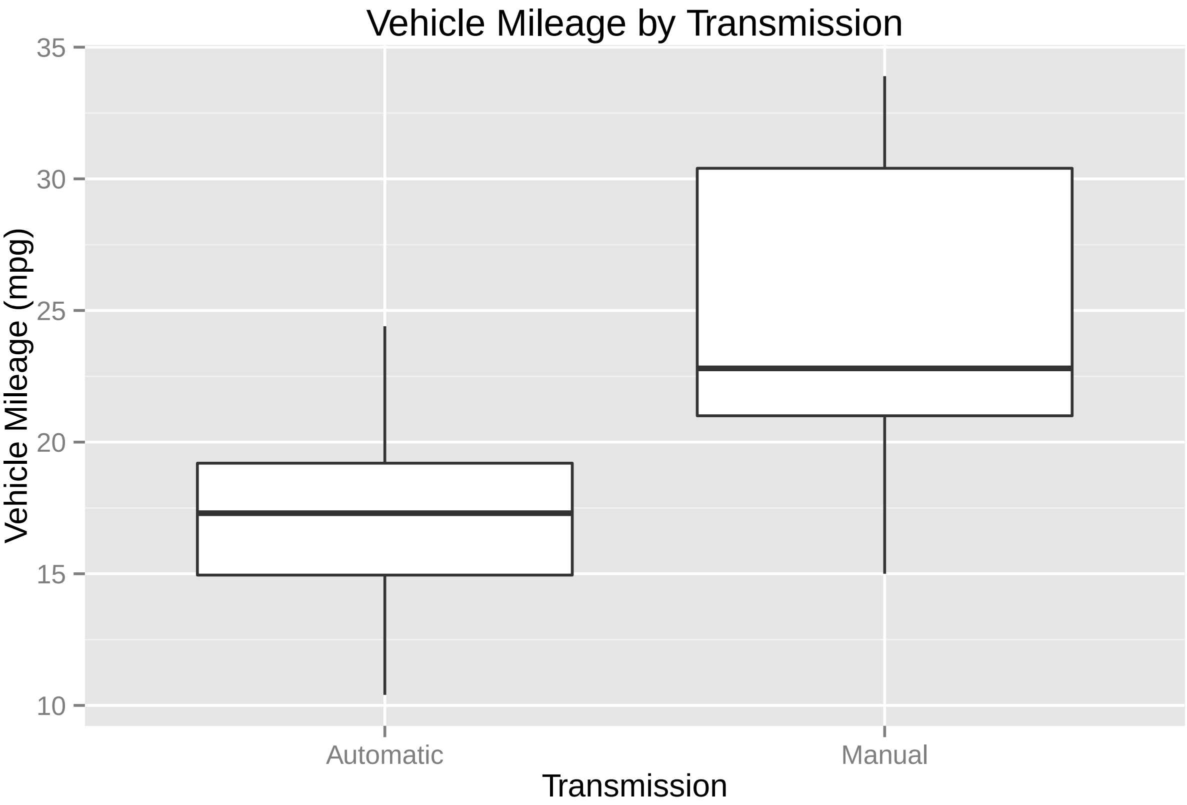
Conclusions

Let's wrap up by answering our two original questions: 1) Manual transmissions seem better if one just looks at raw data. However, a more thorough analysis including confounding variables, such as weight and number of cylinders, points to the fact that automatic transmissions are better. 2) Including weight and the number of cylinders in the linear model, the mileage of automatic vehicles is actually

Appendices

Appendix 1 Boxplot of mileage (mpg) versus transmission (am), without any other variables included.

```
library(ggplot2)
library(scales)
mpgamplot <- qplot(factor(mtcars$am), mtcars$mpg, geom="boxplot") +
  labs(title="Vehicle Mileage by Transmission", x="Transmission", y="Vehicle Mileage (mpg)")+
  scale_x_discrete(breaks=c("0", "1"), labels=c("Automatic", "Manual"))
print(mpgamplot)
```



Appendix 2 Hat values of our selected model, mpgvsam3.

```
hatvalues(mpgvsam3)
```

```
##          Mazda RX4       Mazda RX4 Wag          Datsun 710
##         0.20149516          0.20568847          0.11134826
##     Hornet 4 Drive   Hornet Sportabout             Valiant
##         0.18203504          0.12944546          0.17562241
##         Duster 360           Merc 240D            Merc 230
##         0.11035260          0.19990502          0.19671403
##           Merc 280           Merc 280C          Merc 450SE
##         0.17559835          0.17559835          0.07524667
##         Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##          0.09249950          0.08819801          0.23360825
## Lincoln Continental   Chrysler Imperial            Fiat 128
##          0.28562638          0.26109577          0.10600585
##         Honda Civic      Toyota Corolla       Toyota Corona
##          0.13014404          0.11129580          0.20249595
##    Dodge Challenger         AMC Javelin          Camaro Z28
##          0.11720930          0.13026193          0.08383928
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##          0.08351560          0.10662207          0.10464875
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##          0.14287912          0.20603900          0.20204540
##       Maserati Bora          Volvo 142E
##          0.20862936          0.16429081
```