Regression Models Course Project

Daniel DeWaters

11/19/2019

Executive Summary

Here we run an analysis for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions

Cleaning

```
# Convert am and vs columns into factors
mtcars <-
  mtcars %>%
  mutate(vs = as.factor(vs)) %>%
  mutate(am = as.factor(am))

# Add meaningful labels to factor levels
levels(mtcars$vs) <- c("vshaped", "straight")
levels(mtcars$am) <- c("automatic", "manual")
```

Exploratory Analysis

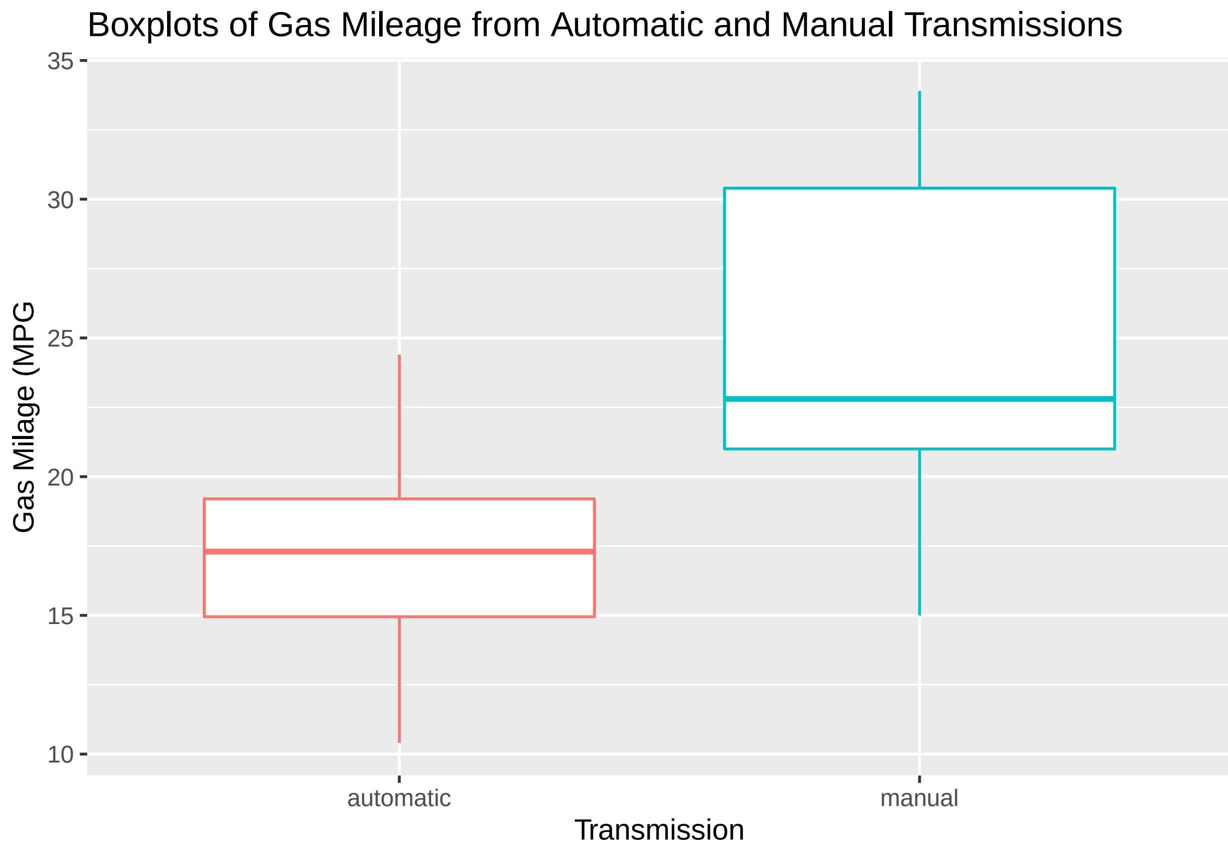
First we'll get a basic overview of all of the data:

```
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  : Factor w/ 2 levels "vshaped","straight": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "automatic","manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Then we'll get an overview of gas milage from automatic and manual transmissions:

```
# Make boxplot
mpgBP <- ggplot(mtcars, aes(x=as.character(am), y=mpg, color=am))
mpgBP + geom_boxplot(aes(group=am)) +
  ggtitle("Boxplots of Gas Mileage from Automatic and Manual Transmissions") +
  xlab("Transmission") + ylab("Gas Milage (MPG") +
  theme(legend.position="none")
```



Here we can see a pretty distinct difference in distribution of gas mileage between manual and automatic transmission. Manual transmission seems to be centered at higher values of mpg than automatic.

Models

Model 1

A linear model where "mpg" is the outcome and "am" is the predictor.

```
fit1 <- lm(mpg~am, mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## ammanual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
```

```
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

This model shows that a manual transmission will get 7.245mpg more than an automatic. The adjusted R-squared is 0.3385, so this model can explain 33.85% of the variance around mpg. This is not very high, so we should add more variables to the model so it can explain a larger portion of the variability

Model 2

"mpg" is the outcome and all of the other variables are the predictors

```
fit2 <- lm(mpg~., mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vsstraight   0.31776    2.10451   0.151   0.8814
## ammanual     2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

When including the other variables, the adjusted R-squared value increases to 0.8066, so it can explain 80.66% of mpg's variance. From the P-values shown here, it seems that the predictor variables don't have much significance since they are all above 0.05. In order to remove the least significant predictor variables, we can use the step function.

Model 3

```
fit3 <- step(fit2, direction = "backward", k=log(nrow(mtcars)), trace = F)
summary(fit3)
```

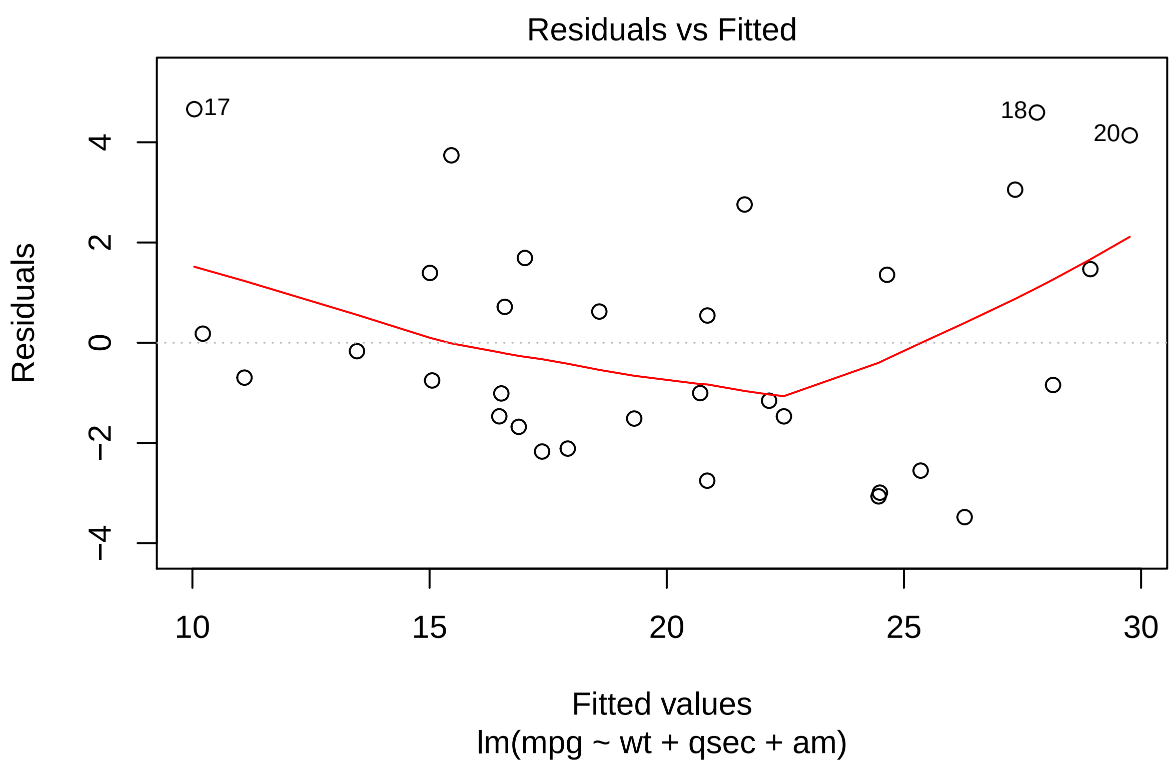
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

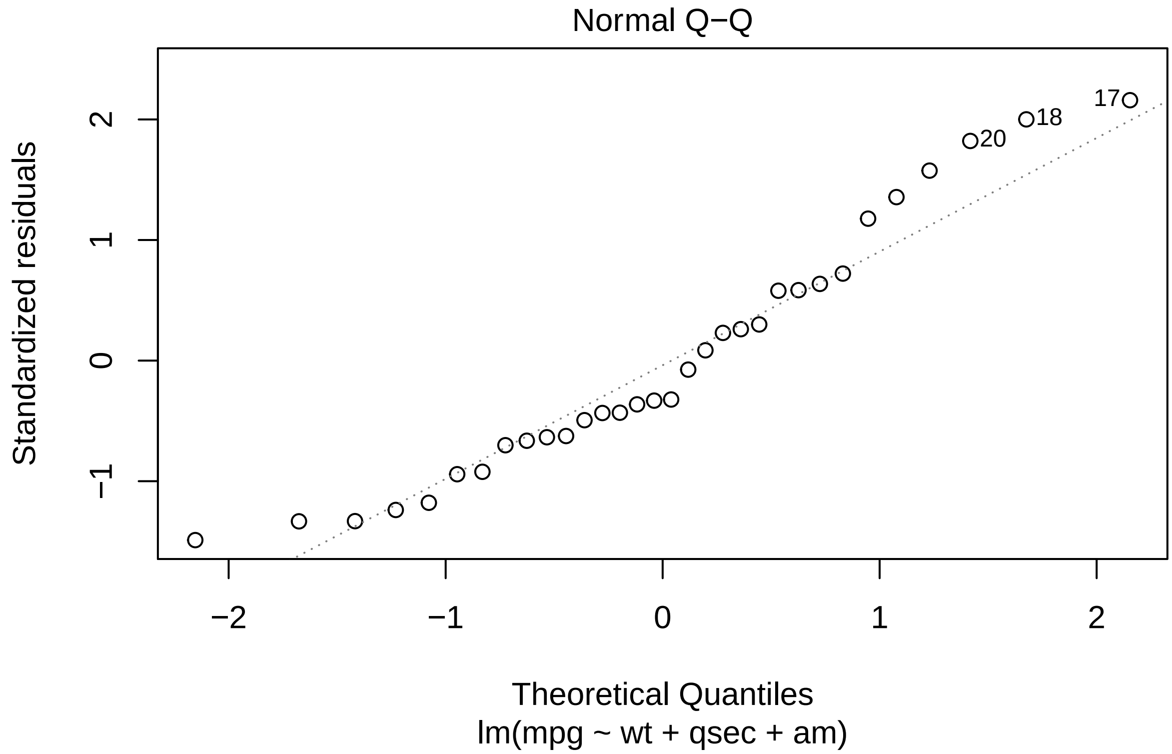
```
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ammanual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

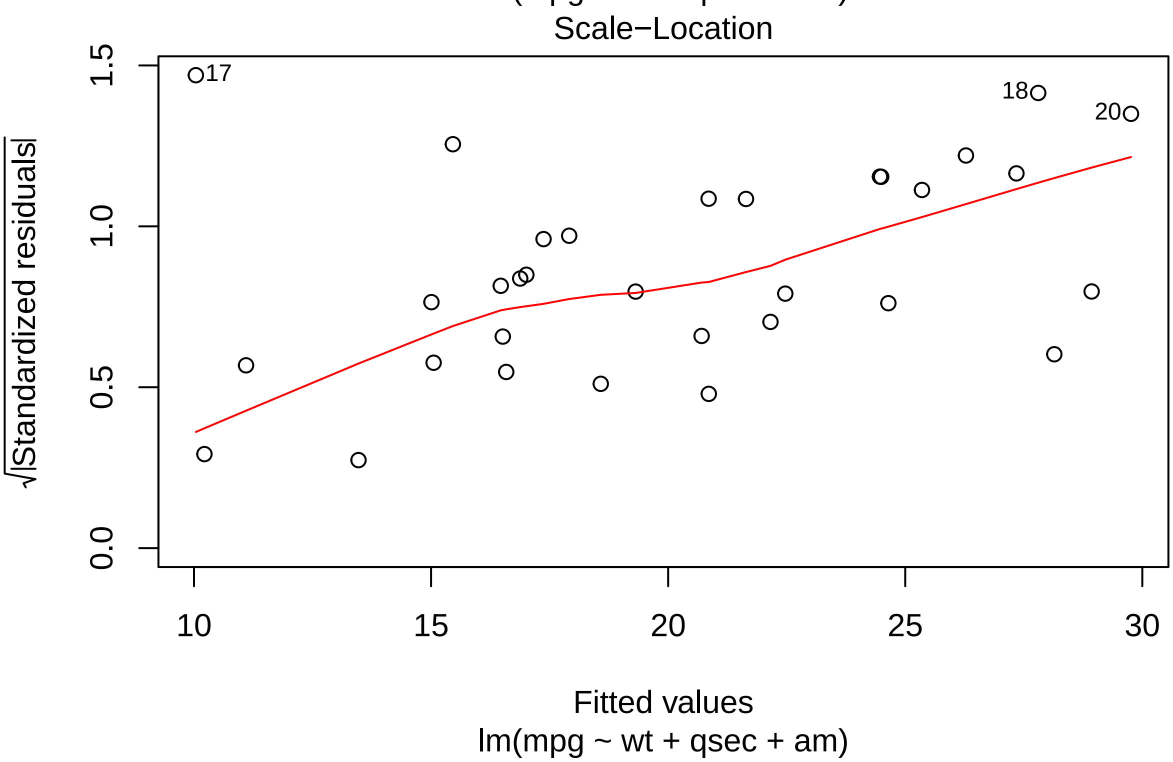
The step function selected "wt", "qsec", and "am" as the most significant variables. This increases our adjusted R-squared value again to 0.8336, so it can explain 83.36% of mpg's variance. This model also shows that a manual transmission will get more gas milage than an automatic, by 2.9358mpg.

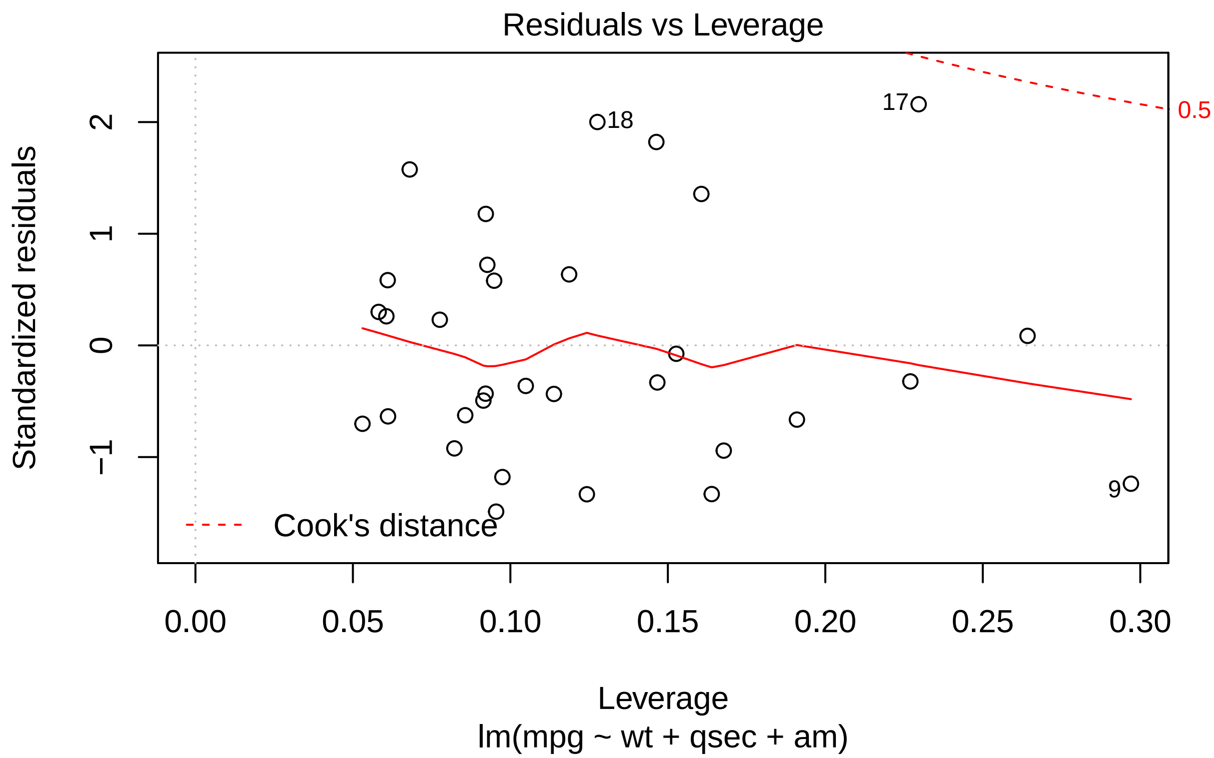
Residuals and Diagnostics

```r
plot(fit3)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ wt + qsec + am)
17
18
20
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
Fitted values
lm(mpg ~ wt + qsec + am)
17
18
20
0.0
0.5
1.0
1.5
10
15
20
25
30



Conclusion

Across all 3 of our models, manual transmissions were shown to have better gas milage. With "fit3" being the most reliable model since its adjusted R-squared value is the highest, we can conclude that manual transmissions will get 2.9358 more miles to the gallon on average if "wt" and "qsec" are held constant.

Model	Transmission w/ Higher MPG	Diff. in MPG	Adjusted R-squared Value
fit1	Manual	7.245	0.3385
fit2	Manual	2.52023	0.8066
fit3	Manual	2.9358	0.8336