Transmissions and MPG

Stacey Farr

2022-07-25

Executive Summary

The mtcars data set was explored and analyzed to determine whether a manual or automatic transmission was better for gas mileage. Several regression models were created and an ANOVA was run to determine the model that best explained miles per gallon. The analysis determined that cars with manual transmissions get better gas mileage than those with automatic transmissions overall,but weight has a significant effect and must be taken into account as well.

Loading and Transforming Data

The mtcars data set is included with R, so no file download is necessary, you can read the data directly from R. The tidyverse package is required for the analysis.

The documentation for mtcars tells us the variable for transmission type is 'am', with a value of 0 for automatic and 1 for manual. For reference, I added a variable (trans), using "automatic" and "manual" as values.

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb  trans
## Mazda RX4       21   6  160 110  3.9 2.620 16.46  0  1    4    4 manual
## Mazda RX4 Wag  21   6  160 110  3.9 2.875 17.02  0  1    4    4 manual
```

Exploratory Analysis

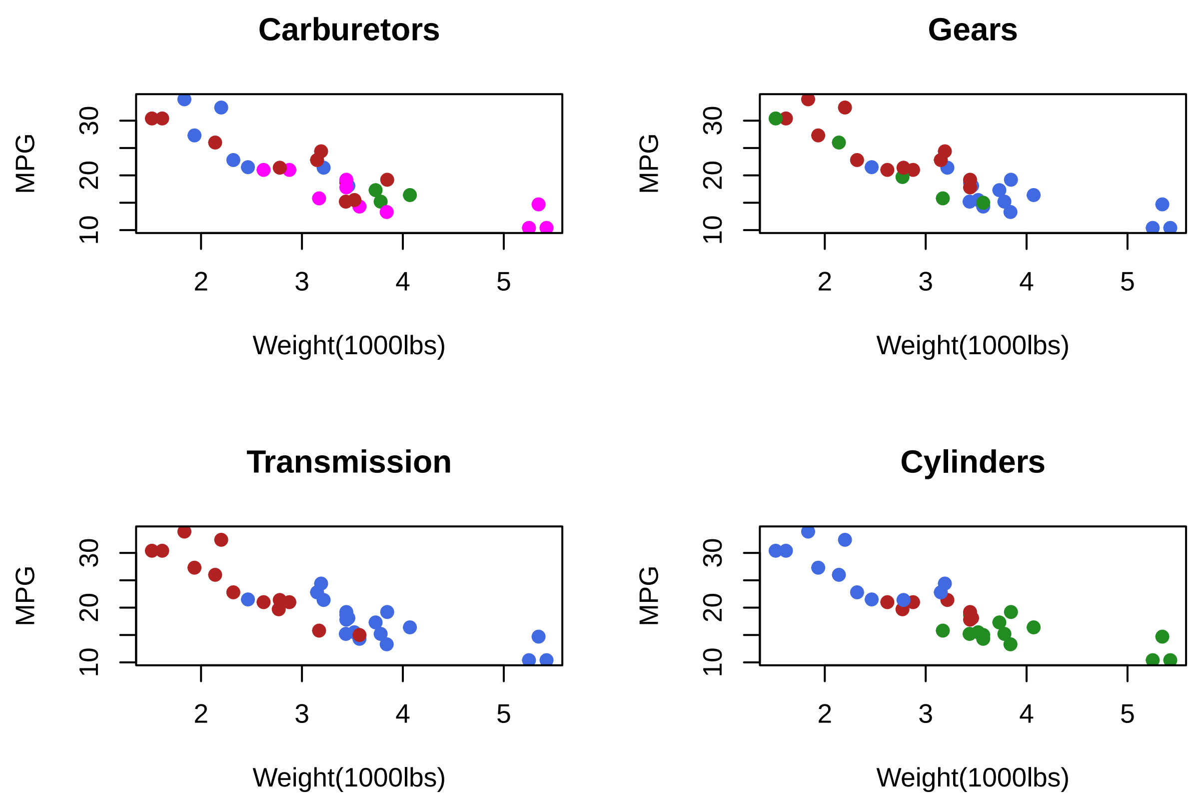
First, I would like to see how much of a difference (if any) there is between the miles per gallon for automatic and manual transmissions without taking other variables into account. I'll do this by finding the average mph for each transmission type.

```
## # A tibble: 2 x 2
##   trans       avg
##   <chr>     <dbl>
## 1 automatic  17.1
## 2 manual     24.4
```

Manual transmissions get an average of 7 miles per gallon more than automatic. Now I need to investigate other variables to see if they factor into that difference. I want to see how weight (wt), number of cylinders (cyl), gears (gear), and/or number of carburetors (carb) affect mpg for each transmission type. First, I want to check the correlations between these variables.

```
##              mpg        cyl         wt          am      gear        carb
## mpg   1.0000000 -0.8521620 -0.8676594  0.59983243  0.4802848 -0.55092507
## cyl  -0.8521620  1.0000000  0.7824958 -0.52260705 -0.4926866  0.52698829
## wt   -0.8676594  0.7824958  1.0000000 -0.69249526 -0.5832870  0.42760594
## am    0.5998324 -0.5226070 -0.6924953  1.00000000  0.7940588  0.05753435
## gear  0.4802848 -0.4926866 -0.5832870  0.79405876  1.0000000  0.27407284
## carb -0.5509251  0.5269883  0.4276059  0.05753435  0.2740728  1.00000000
```

As you can see, most of these variables are highly correlated with each other. Let's look at some plots to see if any patterns emerge. I will create a scatter plots of mpg and wt and use color to indicate other variables.



The Transmission and Cylinder plots show a clear color grouping. Carburetors and Gears, on the other hand, do not. I suspect the best fit model for mpg will have 'trans', 'wt', and possibly 'cyl' as regressors.

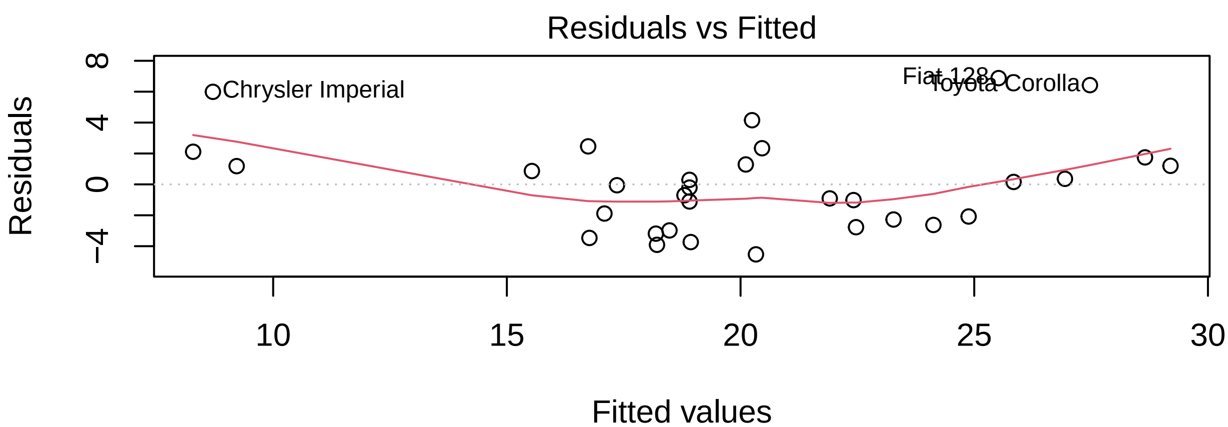
Regression Models

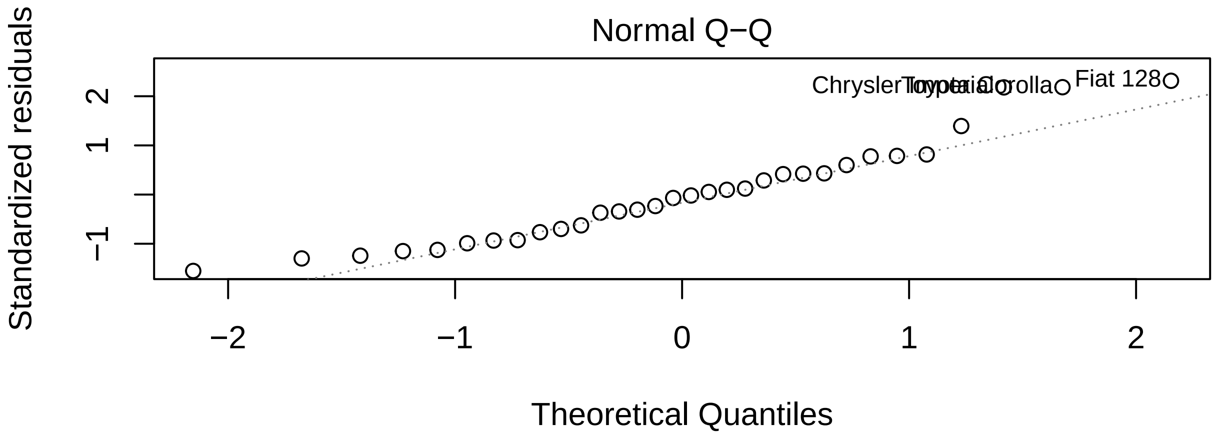
I will create nested models and use ANOVA to find the one that explains mpg the best. I will not be considering zero as the y-intercept for these models, so all formulas will include '-1'.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans - 1
## Model 2: mpg ~ trans + wt - 1
## Model 3: mpg ~ trans + wt + cyl - 1
## Model 4: mpg ~ trans + wt + cyl + gear - 1
## Model 5: mpg ~ trans + wt + cyl + gear + carb - 1
```

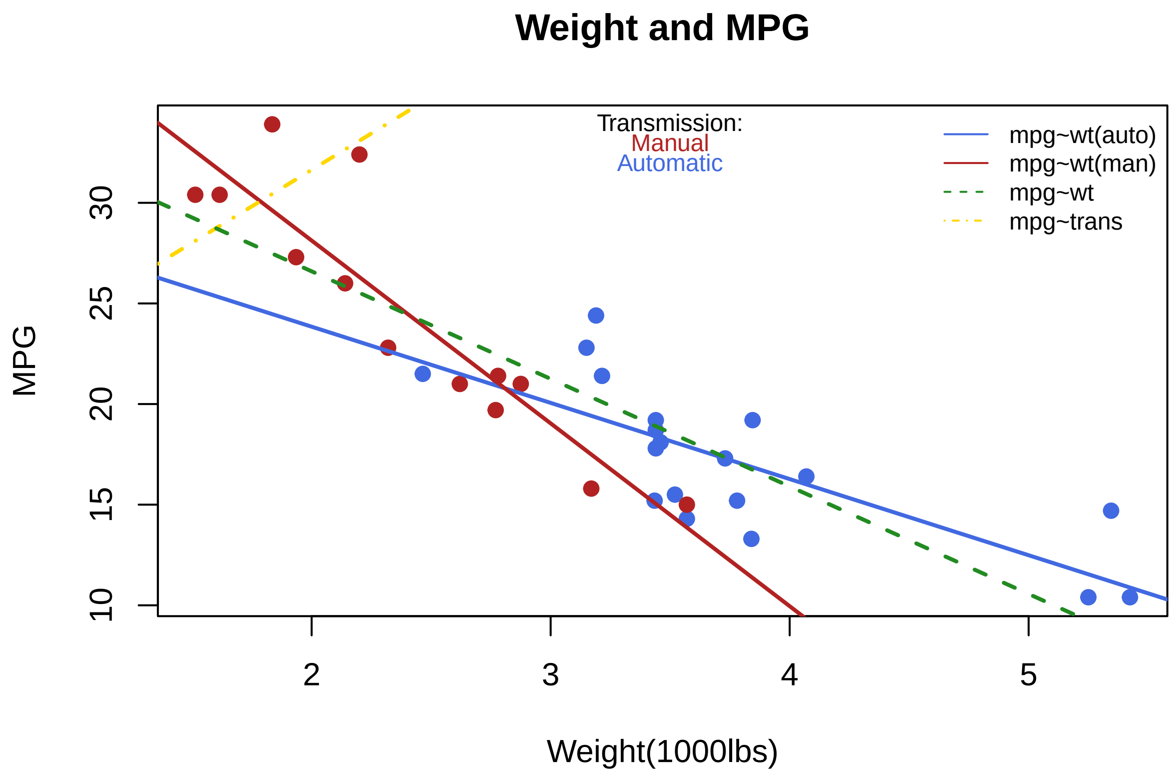
```
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                   
## 2     29 278.32  1    442.58 68.3830 9.453e-09 ***
## 3     28 191.05  1     87.27 13.4846  0.001093 ** 
## 4     27 183.93  1      7.12  1.0994  0.304040    
## 5     26 168.27  1     15.66  2.4195  0.131926    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA shows that model 2 is significantly better than the other models. To check this finding, I need to check the residuals' correlation to the fit (there should be none) and their normality.





The residuals show no clear pattern in the first plot, so there is no correlation to the fit. The QQ-Plot shows they are approximately normal. I can conclude that Model2 is the best fit. To get a better idea of how weight and transmission factor into miles per gallon, I have created a scatter plot of mpg and wt and overlayed fitted lines for different models. Note the slope of the gold line, which represents the model for mpg and trans without wt as a factor.



Conclusion

Transmission type has a major effect on miles per gallon, with manual averaging 7 miles more per gallon. However, weight also factors into mpg so it must be taken into account. There is a sharper drop in mpg for cars with manual transmissions as the cars get heavier. Specifically, there is a decrease of 9mpg for each 1000lbs increase in weight. Automatic transmissions, on the other hand, only have a 3.75mpg decrease for every 1000lb increase.

Appendix - Selected Code

```
par(mfrow=c(2,2))
colors <- c("royalblue","firebrick","forestgreen","magenta")
plot(cars$wt,cars$mpg, col=colors[factor(cars$carb)],pch=19, xlab="Weight(1000lbs)", ylab="MPG")
title(main="Carburetors")
plot(cars$wt,cars$mpg, col=colors[factor(cars$gear)],pch=19,xlab="Weight(1000lbs)", ylab="MPG")
title(main="Gears")
plot(cars$wt,cars$mpg, col=colors[factor(cars$trans)],pch=19, xlab="Weight(1000lbs)", ylab="MPG")
title(main="Transmission")
plot(cars$wt,cars$mpg, col=colors[factor(cars$cyl)],pch=19, xlab="Weight(1000lbs)", ylab="MPG")
title(main="Cylinders")
```

```
fit1 <- lm(mpg ~ trans - 1, cars)
fit2 <- update(fit1,mpg~trans+ wt - 1)
fit3 <- update(fit1,mpg~trans + wt + cyl - 1)
fit4 <- update(fit1,mpg~ trans + wt + cyl + gear - 1)
fit5 <- update(fit1,mpg~ trans + wt + cyl + gear + carb - 1)
anova(fit1,fit2,fit3,fit4,fit5)
```

```
par(mfrow=c(2,1), mar=c(4,4,1.5,1))
plot(fit2,which=1)
plot(fit2, which=2)
```

```
plot(cars$wt,cars$mpg, col=colors[factor(cars$trans)],pch=19,xlab="Weight(1000lbs)", ylab="MPG")
title(main="Weight and MPG")
text(3.5,34,"Transmission:",cex=.75)
text(3.5,33,"Manual",col="firebrick",cex=.75)
text(3.5,32,"Automatic",col="royalblue",cex=.75)
abline(lm(mpg[am==0]~wt[am==0],mtcars),col="royalblue",lwd=2)
abline(lm(mpg[am==1]~wt[am==1],mtcars),col="firebrick",lwd=2)
abline(lm(mpg~am,mtcars),col="gold",lty=4,lwd=2)
abline(lm(mpg~wt,cars),col="forestgreen",lty=2,lwd=2)
legend("topright",bty="n", legend = c("mpg~wt(auto)", "mpg~wt(man)", "mpg~wt","mpg~trans"),
       cex=.75,lty=c(1,1,2,4),lwd = 1, col = c("royalblue", "firebrick","forestgreen","gold"))
```