mtcars data analysis - Regression Models Course Project

MSibomana

October 23, 2015

Executive summary

Looking at the provided mtcars dataset, I can not draw a strong conlusion whether automatic transmission cars give better MPG outcome or the opposite because the samples are not matched for the typical car classifications (weight and/or horsepower). The manual cars in the sample are smaller cars, which give higher MPG than larger cars in general. Looking at a subset of midsize cars with similar weight, automatic cars tend to have a better MPG outcome, but the subset is too small and larger sample is needed to draw a final conclusion.

Exploratory analysis

```
data(mtcars)
am_count <- c(sum(mtcars$am==0), sum(mtcars$am==1))
am_mpg <- tapply(mtcars$mpg,factor(mtcars$am),mean);
am_wt <- tapply(mtcars$wt,factor(mtcars$am),mean)
```

The dataset contains 19 automatic and 13 manual cars. The average Miles Per Gallon (MPG) is respectively 17.1 and 24.4 for automatic and manual transmission cars.

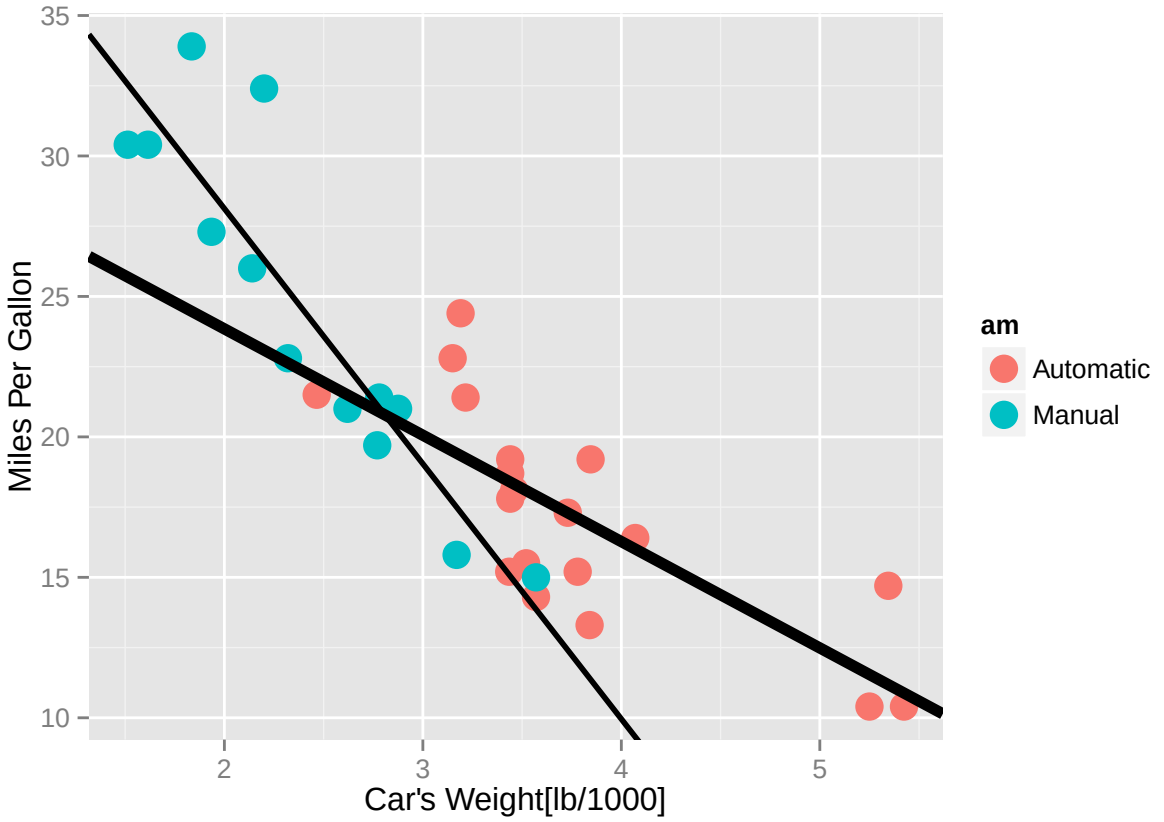
```
library(ggplot2)
mtcars$am <- factor(mtcars$am,levels=c(0,1),labels=c("Automatic","Manual"))
fit <- lm(mpg~wt*am,data=mtcars)
g <- ggplot(mtcars,aes(y=mpg,x=wt,colour=am))+geom_point(size=5)
g <- g+labs(x="Car's Weight[lb/1000]", y="Miles Per Gallon")
```

The assumption that smaller cars (lower weight) give a better MPG outcome than larger cars (higher weight) can be observed in the dataset for both automatic and manual transmission cars (see figure 1).

Looking at the small subset of automatic and manual in the mid-size range (weight close to 3), there is a trend that automatic cars have better MPG outcome.

Detailed analysis

```
g1 <- g+geom_abline(intercept=coef(fit)[1],slope=coef(fit)[2],size=2)
g1+geom_abline(intercept=coef(fit)[1]+coef(fit)[3],slope=coef(fit)[2]+coef(fit)[4],size=1)
```




```
library(xtable)
xt <- xtable(summary(fit))
print(xt)
```

% latex table generated in R 3.2.2 by xtable 1.7-4 package % Sun Oct 25 17:51:51 2015

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.4161	3.0201	10.40	0.0000
wt	-3.7859	0.7856	-4.82	0.0000
amManual	14.8784	4.2640	3.49	0.0016
wt:amManual	-5.2984	1.4447	-3.67	0.0010

```
xCross <- (-coef(fit)[3]/coef(fit)[4])*1000
```

As shown on the fit coefficients table above, there is a positive effect `14.9` on the intercept for the manual cars compared to automatic (reference) but but a negative effect `-5.3` on the slope. The fit model shows a better MPG for manual cars for smallsize cars group before the lines crossing wt=`2808`lb and the opposite midsize and larger cars with weight higher than `2808`lb. However, the number of manual cars with higher weight and significant leverage is only 2.

Looking at the residual plots in the appendix, the model looks OK because there is no visible pattern in the Residual-vs-Fitted and no high amplitude residual with high leverage on the Residual-vs-Leverage plot. The cars with high Residuals are 2 higher efficient small cars (Fiat128, Toyota Corolla) and Merc240D which has a Diesel engine.

Appendix: residual plots

```
par(mfrow=c(2,2))
plot(lm(mpg~wt*am,data=mtcars))
```