Analysis of Relationship Between Weight, Transmission and Fuel Economy

Graham King

February 2015

Overview

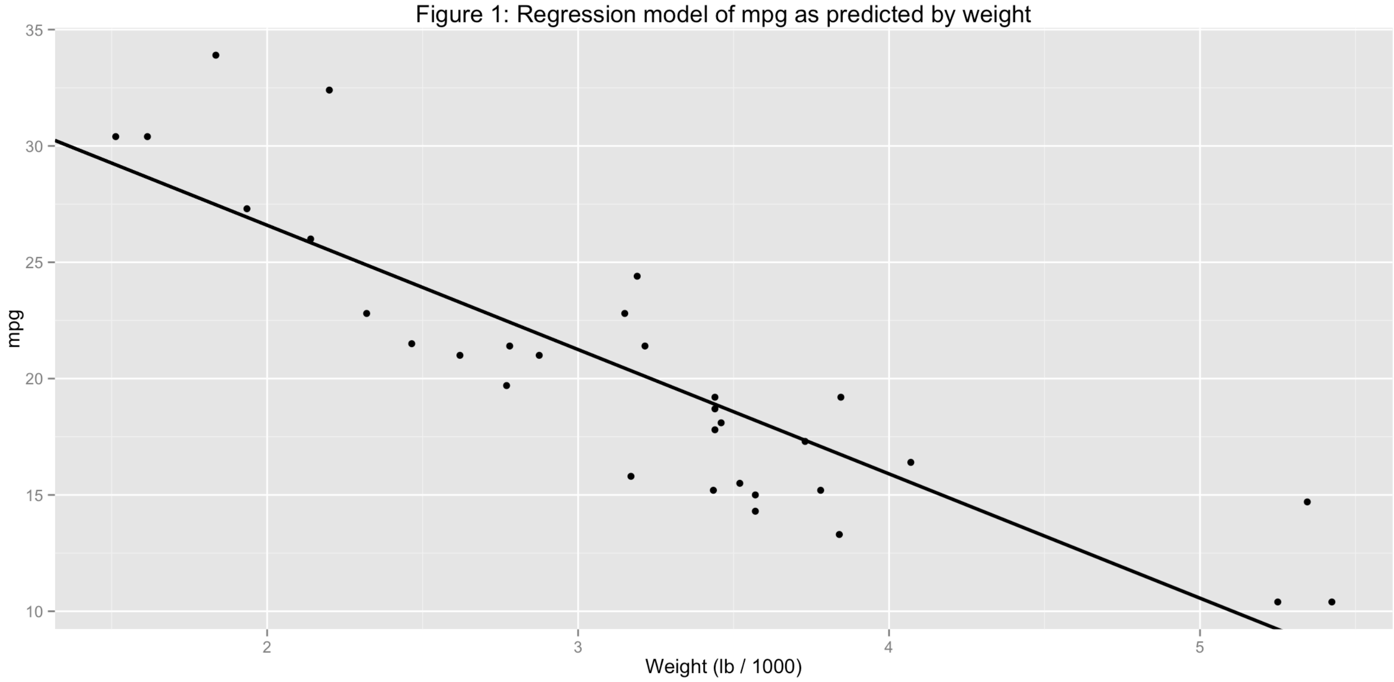
In this report, I will use the `mtcars` dataset to look at the effect that the weight of a car has on its fuel economy, and what impact the transmission has on this.

The data shows that lighter manual transmission cars have better fuel economy, but this advantage is transferred to automatic cars as the cars get heavier.

Content

An initial look at the complete dataset (Figure 1) shows that there is a clear linear relationship between car weight and fuel economy.

Figure 1: Regression model of mpg as predicted by weight



There is a certain degree of distance between some values and the regression line. The sum of these squared vertical distances is:

```
[1] 278.3219
```

and the residual variance is calculated as:

```
[1] 9.277398
```

When looking at Automatic and Manual transmission data separately, there is a distinct difference in the regression lines:

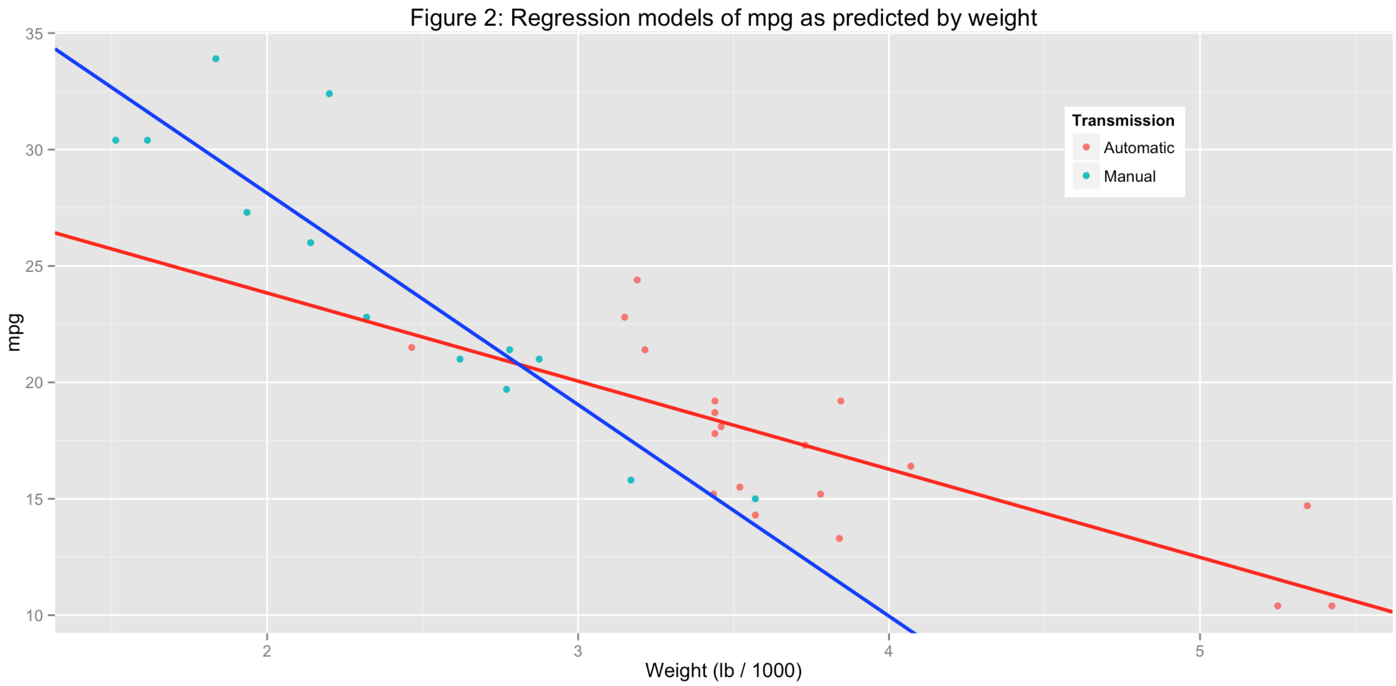


Figure 2 indicates that fuel economy (mpg) reduces faster for manual cars as the weight increases. However at lower weights, the fuel economy is superior than for automatic cars.

We can demonstrate that this model is a better fit by looking at the sum of all squares of the distance of residuals from their corresponding regression line:

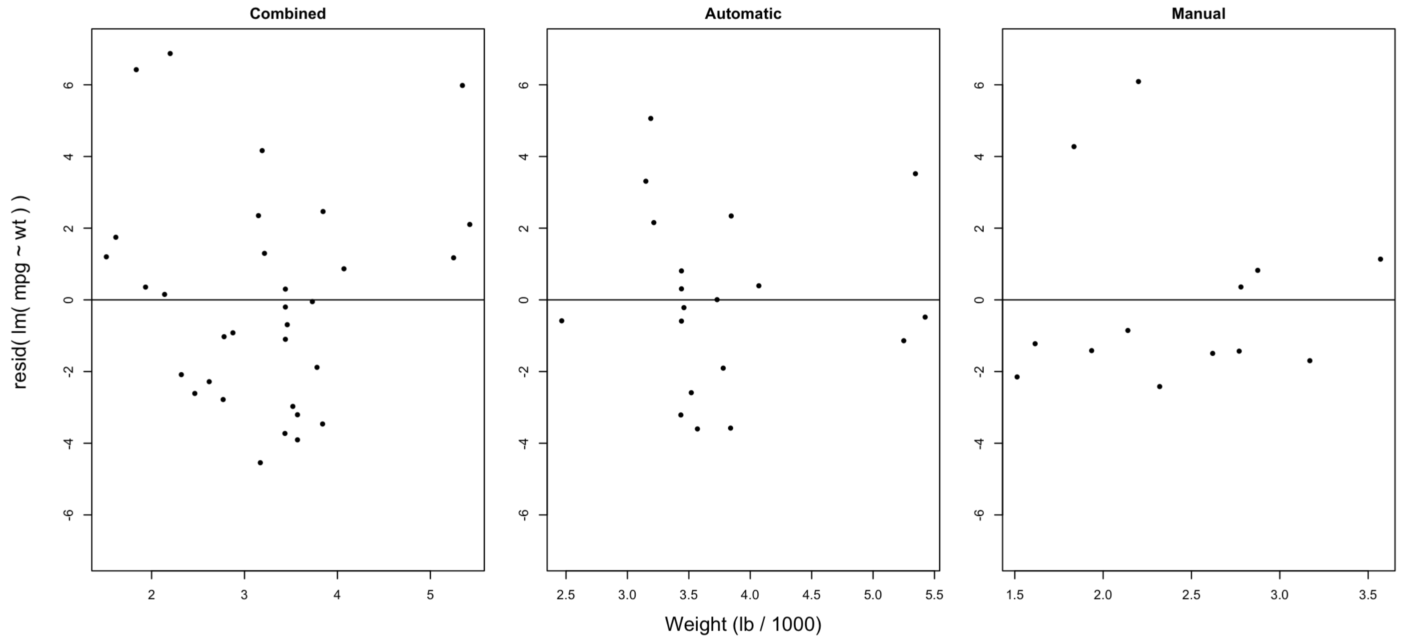
```
[1] 188.0077
```

So the total of all distances from the regression lines is less than in the combined dataset. And the corresponding residual variances are also both lower than the same statistic in the combined database:

```
     Automatic   Manual
[1,]    6.3922 7.212752
```

The difference in residual variances can be better visualised in a comparison of the residual plots, with the dots spread farther away from zero (the regression model) in the combined dataset:

Figure 3: Residual Plots



The regression model coefficients for intercept and slope are as follows:

```
          (Intercept)        wt
Automatic    31.41606 -3.785908
Manual       46.29448 -9.084268
```

Specifically, for each 1,000lb increase in weight, the manual cars have a fuel economy reduction of 9.1 mpg, compared with a reduction of 3.8 mpg for automatic cars.

We can use this model to predict fuel economy for hypothetical car weights:

```
              0lbs 1,000lbs  4,000lbs
Automatic 31.41606 27.63015 16.272425
Manual    46.29448 37.21021  9.957406
```

Under this model, a hypothetical manual transmission car weighing 1,000lbs gets 9.6 more miles to the gallon than an automatic car of the same weight. Conversely, a hypothetical manual car weighing 4,000lbs gets 6.3 fewer miles to the gallon than an automatic car of the same weight.

However, this model cannot be extended to extremes: despite what the model says, a theoretical car weighing 0lbs will never get that kind of fuel economy, regardless of whether it’s a manual or automatic!

Appendix

```r
library(datasets)
data(mtcars)

library(ggplot2)
```

Start by converting am to a factor variable:

```
mtcars$am <- as.factor(mtcars$am)
```

Plot full dataset with regression linear model:

```
fit1 <- lm(mpg ~ wt, data = mtcars)
g1 <- ggplot(data = mtcars, aes(x=wt, y=mpg)) +
    geom_point() +
    geom_abline(intercept=coef(fit1)[1], slope=coef(fit1)[2], colour="black", size=1) +
    ggtitle("Figure 1: Regression model of mpg as predicted by weight") +
    labs(x = "Weight (lb / 1000)")
print(g1)
```

Calculate sum of squared errors:

```
sume <- sum(resid(fit1)^2)
print(sume)
```

Calculate residual variance:

```
n <- nrow(mtcars)
resid.var <- 1/(n-2) * sum(resid(fit1)^2)
print(resid.var)
```

Create subsets of data for automatic and manual transmition and plot again:

```
a <- subset(mtcars, am==0, select=c(wt, mpg))
m <- subset(mtcars, am==1, select=c(wt, mpg))
fita <- lm(mpg ~ wt, data = a)
fitm <- lm(mpg ~ wt, data = m)
g2 <- ggplot(data = mtcars, aes(x=wt, y=mpg, colour=am)) +
    geom_point() +
    geom_abline(intercept=coef(fita)[1], slope=coef(fita)[2], colour="red", size=1) +
    geom_abline(intercept=coef(fitm)[1], slope=coef(fitm)[2], colour="blue", size=1) +
    ggtitle("Figure 2: Regression models of mpg as predicted by weight") +
    labs(x = "Weight (lb / 1000)") +
    scale_colour_discrete(name="Transmission", labels=c("Automatic", "Manual")) +
    theme(legend.position=c(0.8,0.8))
print(g2)
```

Re-calculate sum of all squared errors, from respective models:

```
sume.am <- sum(resid(fita)^2) + sum(resid(fitm)^2)
print(sume.am)
```

Calculate residual variance for each separate linear model:

```
na <- nrow(a)
nm <- nrow(m)
resid.var.am <- cbind(
    1/(na-2) * sum(resid(fita)^2),
    1/(nm-2) * sum(resid(fitm)^2))
colnames(resid.var.am) <- c("Automatic", "Manual")
print(resid.var.am)
```

Residual plots:

```
ylim <- c(-7, 7);
par(mfrow=c(1,3), pch=20, oma=c(4,4,4,0), mar=c(1, 2, 2, 2))
plot(mtcars$wt, resid(fit1), ylim=ylim, ylab="", xlab=""); abline(h=0)
title(main="Combined")
plot(a$wt, resid(fita), ylim=ylim, ylab="", xlab=""); abline(h=0)
title(main="Automatic")
plot(m$wt, resid(fitm), ylim=ylim, ylab="", xlab=""); abline(h=0)
title(main="Manual")
mtext("resid( lm( mpg ~ wt ) )", side = 2, outer = TRUE, line = 2)
mtext("Figure 3: Residual Plots", side = 3, outer = TRUE, line = 2)
mtext("Weight (lb / 1000)", side = 1, outer = TRUE, line = 2)
```

Display coefficients:

```
coefTable <- rbind(fita$coefficients, fitm$coefficients)
rownames(coefTable) <- c("Automatic", "Manual")
print(coefTable)
```

Calculate predictions:

```
predictions <- rbind(
    predict(fita, newdata=data.frame(wt=c(0, 1,4))),
    predict(fitm, newdata=data.frame(wt=c(0, 1,4))))
rownames(predictions) <- c("Automatic", "Manual")
colnames(predictions) <- c("0lbs", "1,000lbs", "4,000lbs")
print(predictions)
```