Automatic transmission versus manual tranmission based on MPG

Susana Arias Laso

Executive summary

The relationship between some of the variables in the mtcars dataset and the outcome, miles per gallon, of a set of regression models is explored by developing a regression analysis of the relevant statistical factors. More specifically, the questions whether automatic or manual transmission is better for mpg is addressed below.

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
require("cowplot")
```

```
## Loading required package: cowplot
```

```
## Warning: package 'cowplot' was built under R version 3.3.2
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

The "am" variable in the mtcars dataset is of binary type, indicating manual tranmission with "1" and automatic transmission with "0". The box plot can be seen in the appendix (Figure 1) indicating that manual transmission is more effective in terms of miles/gallon, the box plots indicate the media of mpg for the three number of cylinders in the dataset. With exception of 8-cylinder cars, which are independent of the type of transmission as the blue boxes indicate, one can say that manual cars are better in terms of miles/gallon.

Fit a linear model with transmission as the only variable, without taking into consideration the number of cylinders or the weight of the cars.

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

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

This model assumes one line through the data. As the slope of the linear model indicates, there is a 7.2449393 increase in miles/gallon going from automatic transmission to manual.

```
confInterval <- confint(fit)
confInterval
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## am           3.64151 10.84837
```

The confidence interval indicates that with a 95% confidence, there is an increase from 3.6415096 to 10.848369 miles/gallon going from automatic transmission to manual.

Multivariable regression analysis

An additional variable, the weight of the cars, in included in the analysis. In this way it is possible to model how other factors may interfere in the outcome of this study, which is miles/gallon. See figure 2 in the appendix.

```
fit2 <- lm(mpg ~ am + wt, data = mtcars)
coef2 <- summary(fit2)$coefficients
coef2
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

Adding a second variable to the model, the weight of the cars, namely, wt one finds that the coefficient in front of the binary variable, am, is the change in the intercept between automatic and manual transmission. This model generates two parallel lines with different interceps and the same slope (-5.3528114), which indicates the change in mpg per unit increase (1000 lbs) in the car weight. The coefficient that multiplies the binary variable, -0.0236152 represents the change in the intercept between automatic and manual transmission. However, having added the weight of the cars as a second variable one can summarize that the transmission type is not really relevant for fuel efficiency in comparison to the weight, as one can read from the t-probability 0.9879146.

Residual plots

The data sets corresponding to the two values of the binary variable are created, data1 and data0, to construct the regression lines with their residuals. The regression lines and residual plots are shown in Figure 3 in the Appendix.

```
y1 <- mtcars$mpg[mtcars$am == 1]
x1 <- mtcars$wt[mtcars$am == 1]
data1 <- data.frame(x1, y1)
fit1 <- lm(y1 ~ mtcars$am[mtcars$am == 1] + x1)
coef1 <- summary(fit1)$coefficients
```

```
y0 <- mtcars$mpg[mtcars$am == 0]
x0 <- mtcars$wt[mtcars$am == 0]
data0 <- data.frame(x0, y0)
fit0 <- lm(y0 ~ mtcars$am[mtcars$am == 0] + x0)
coef0 <- summary(fit0)$coefficients
```

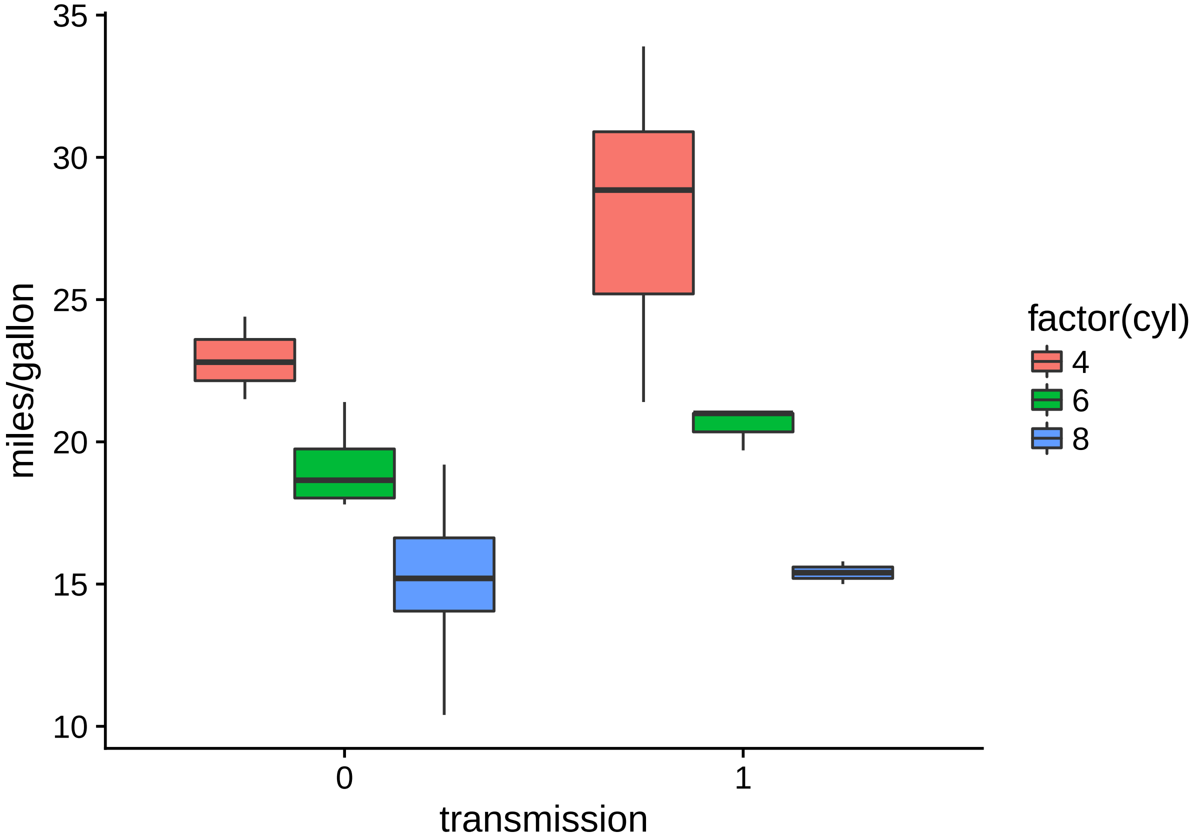
Apendix

This appendix contains supplementary figures and their respective R codes.

Box plot of the mpg(outcome) as a function of the type of transmission of the car. The number of cylinders was included as a second variable in the plot an indicates that, except for 8-cylinder cars, it is not relevant to consider this variable for a regression analysis.

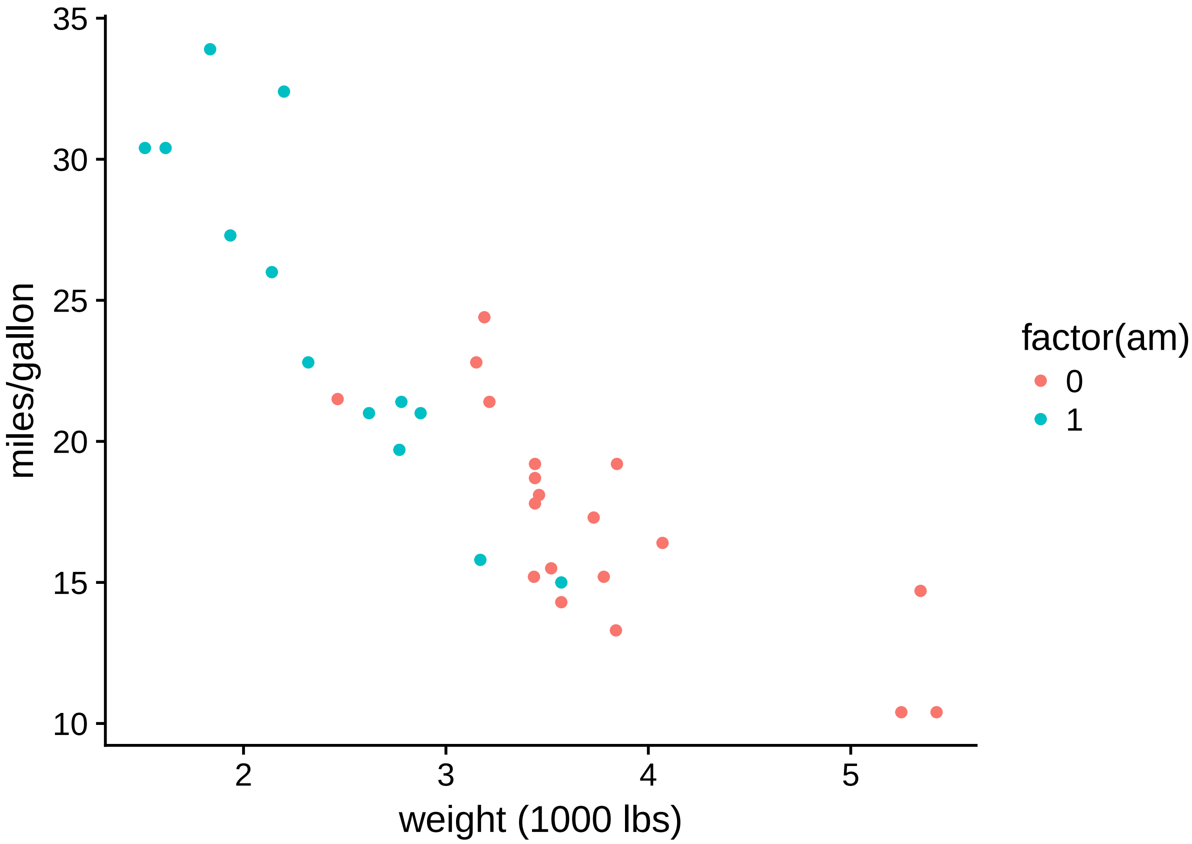
```
g <- ggplot(data = mtcars, aes(y = mpg, x = factor(am), fill = factor(cyl)))
g <- g + geom_boxplot()
```

```
g <- g + xlab("transmission") + ylab("miles/gallon")
g
```



This figure shows a point plot that indicates the direct influence the variable weight has on the mpg outcome. Therefore, this is a variable worth considering when analyzing whether automatic or manual transmission is better for mpg, given that manual cars weight less than automatic ones.

```
g <- ggplot(data = mtcars, aes(y = mpg, x = wt, color = factor(am)))
g <- g + geom_point() + xlab("weight (1000 lbs)") + ylab("miles/gallon")
g
```

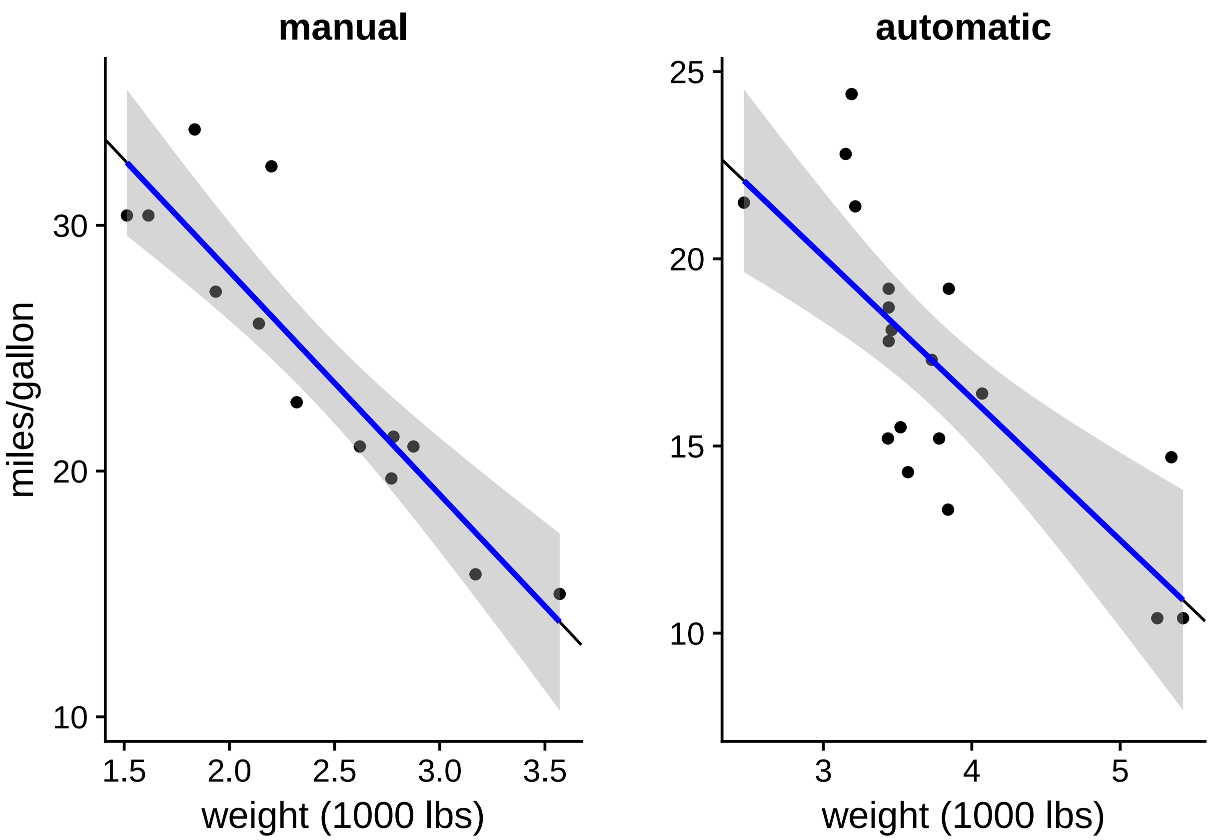


The effect of the weight on the mpg variable is shown in these figures for the two possible values of the binary variable, the transmission type.

```
plot_manual <- ggplot(data1, aes(x = x1, y = y1)) + geom_point() +
    geom_abline(intercept = coef1[1, 1], slope = coef1[2, 1]) +
    stat_smooth(method = "lm", col = "blue") +
    xlab("weight (1000 lbs)") + ylab("miles/gallon") +
    ggtitle("manual")

plot_auto <- ggplot(data0, aes(x = x0, y = y0)) + geom_point() +
    geom_abline(intercept = coef0[1, 1], slope = coef0[2, 1]) +
    stat_smooth(method = "lm", col = "blue") +
    xlab("weight (1000 lbs)") + ylab("") +
    ggtitle("automatic")

plot_grid(plot_manual, plot_auto)
```


manual
automatic
miles/gallon
10
20
30
1.5
2.0
2.5
3.0
3.5
weight (1000 lbs)
10
15
20
25
3
4
5
weight (1000 lbs)