Regression Models Project

Alex Baur

January 21, 2016

In this project the mtcars dataset from the datasets library will be analyzed for a correlation between transmission type (automatic or manual) and the effect that each type has on miles per gallon (mpg). We are asked whether an automatic or manual transmission is better for mpg and to quantify that difference, including supporting statistics. ## Loading the data

```
setwd("C:/Users/alexb/Desktop/Coursera/RegModels")
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.2.3
```

```
library(datasets)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
cars <- mtcars
```

Cleaning the Data

The am transmission column needs to be converted to a factor for analysis

```
cars$am <- as.factor(cars$am)
levels(cars$am)[1] <- "automatic"
levels(cars$am)[2] <- "manual"
```

For later exploration, the manual and automatic data needs to be isolated in seperate objects.

```
Manual <- subset(cars, am=="manual")
Auto <- Auto <- subset(cars, am=="automatic")
```

Effect of Transmission Type on Miles Per Gallon

To get an idea of the spread of the data, the manual and automatic data were plotted in a box plot as **Plot 1** in the appendix below. The manual transmission cars are generally more fuel efficient, with a significantly higher median than the automatic transmission cars. There is a significant amount of overlap however with both boxes' bounds extending past the other's medians. Now we need to get an idea of exactly how much greater the averages of the manuals are compared to the automatics.

```
median(Manual$mpg) - median(Auto$mpg)
```

```
## [1] 5.5
```

```r
mean(Manual$mpg) - mean(Auto$mpg)
```

```
## [1] 7.244939
```

The miles per gallon for manual transmission cars are indeed on average greater than those of automatics (by 5.5 comparing medians and 7.24 when comparing means). ## Regression Analysis of Transmission Type Effect on MPG First we examine whether a linear model would be a good comparison for the transmission types, shown as **Plot 2** in the appendix.

There is a clear linear relationship when looking at the Normal Q-Q graph, indicating that the points are distributed approximately normally and can be examine dusing a linear model. The linear model is created comparing the mpg to the transmission type in the cars dataset and see a summary of the fit.

```r
fit <- lm(mpg~am, data = cars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = cars)
##
## 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
```

The coefficient for the ammanual estimate matches our calculation for the difference in the means between the transmission types calculated above, 7.24 mpg greater for manuals over automatics. Likewise, the intercept of 17.147 matches the mean value of the automatic tranmission mpg. Next we take a look at the confidence intervals and how much error is introduced in our calculations.

```r
t.test(mpg~am, data= cars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group automatic    mean in group manual
##                17.14737                24.39231
```

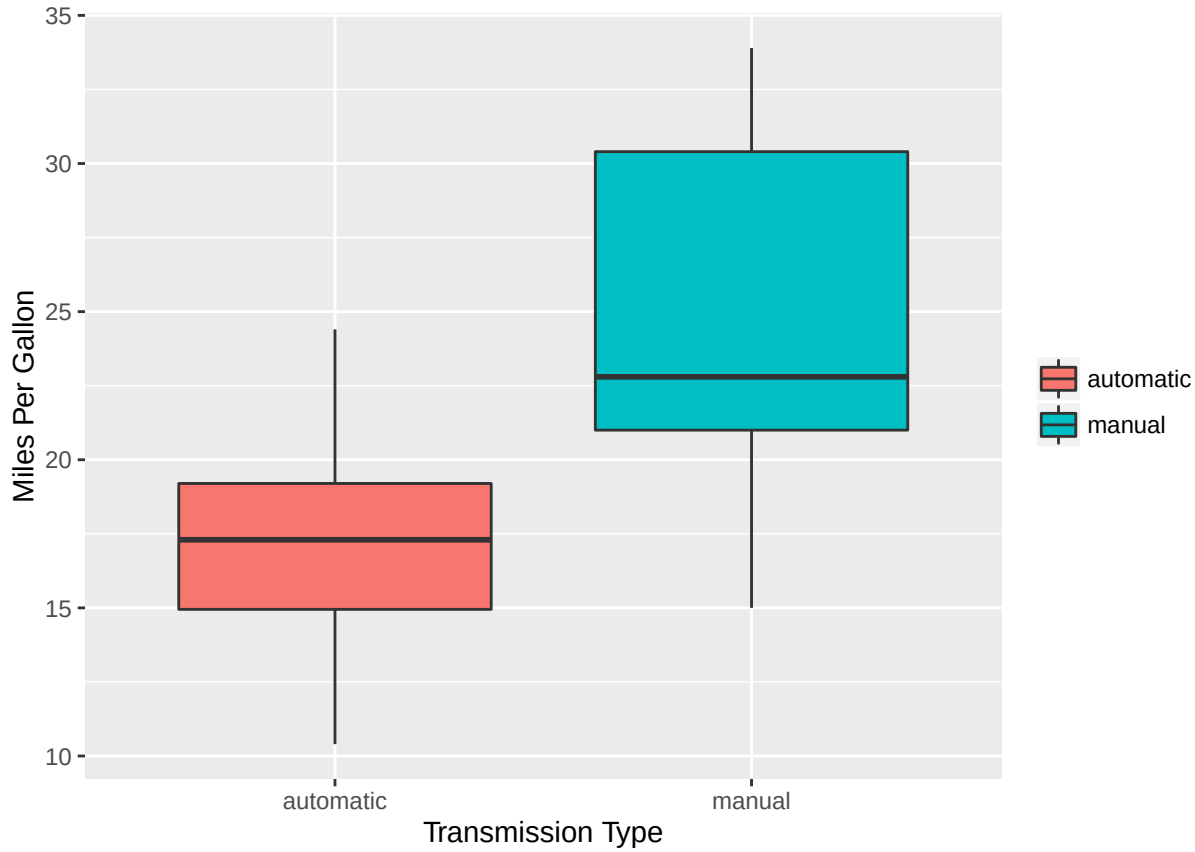
From Student's T-test the 95% confidence interval is between -11.28 and -3.21 which does not contain 0, therefore we can reasonably assume that the null hypothesis (there is no difference between automatic and manual transmissions) is false and that there is a difference between the tranmission types' fuel economy. Furthermore, with a p-value of 0.0014, our sampling error would only affect .14% of our observations which is negligible.

Finally, our residuals are plotted and examined in **Plot 2** in the appendix. The residuals for the manual transmission cars vary more than those for the automatics, and the median for the manuals is quite a bit further from 0 than the automatics, although both are close which indicates a relatively close fit. There are some outliers however, with some approaching +/- 10 which could be approved upon by examining more relationships with am and mpg. ## Summary Overall, we are confident that the **manual transmission cars** are: 1. Better for MPG 2. On average 7.245 MPG more efficient than **automatic transmission cars** Both results are supported by our regression analysis, t-test, and residuals. As could be seen from the summary of our fit model, the p-value was adequately low but the R^2 values were not conclusive, leading to a realtively low correlation when other features are ignored.

Appendix

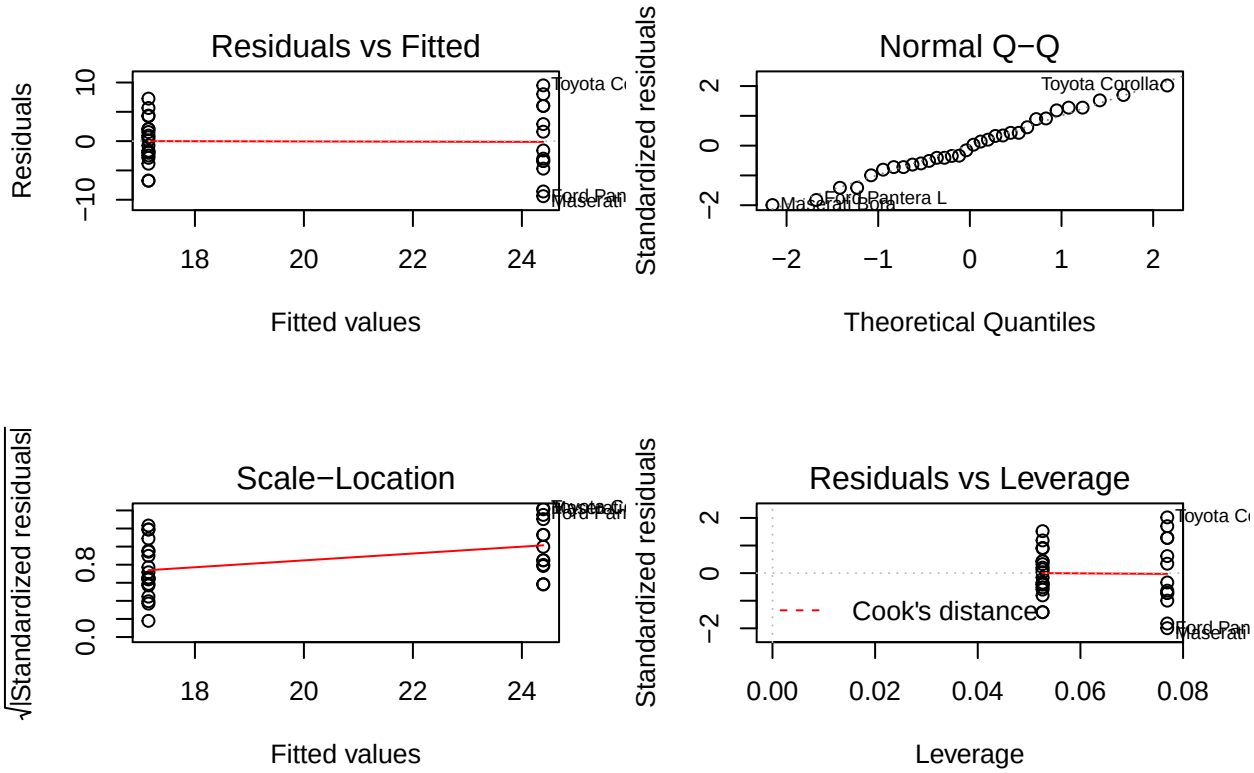
Plot 1

```
g <- ggplot(data = cars, aes(x=am, y=mpg, fill = am))+geom_boxplot() +ylab("Miles Per Gallon")+xlab("Tra
g
```



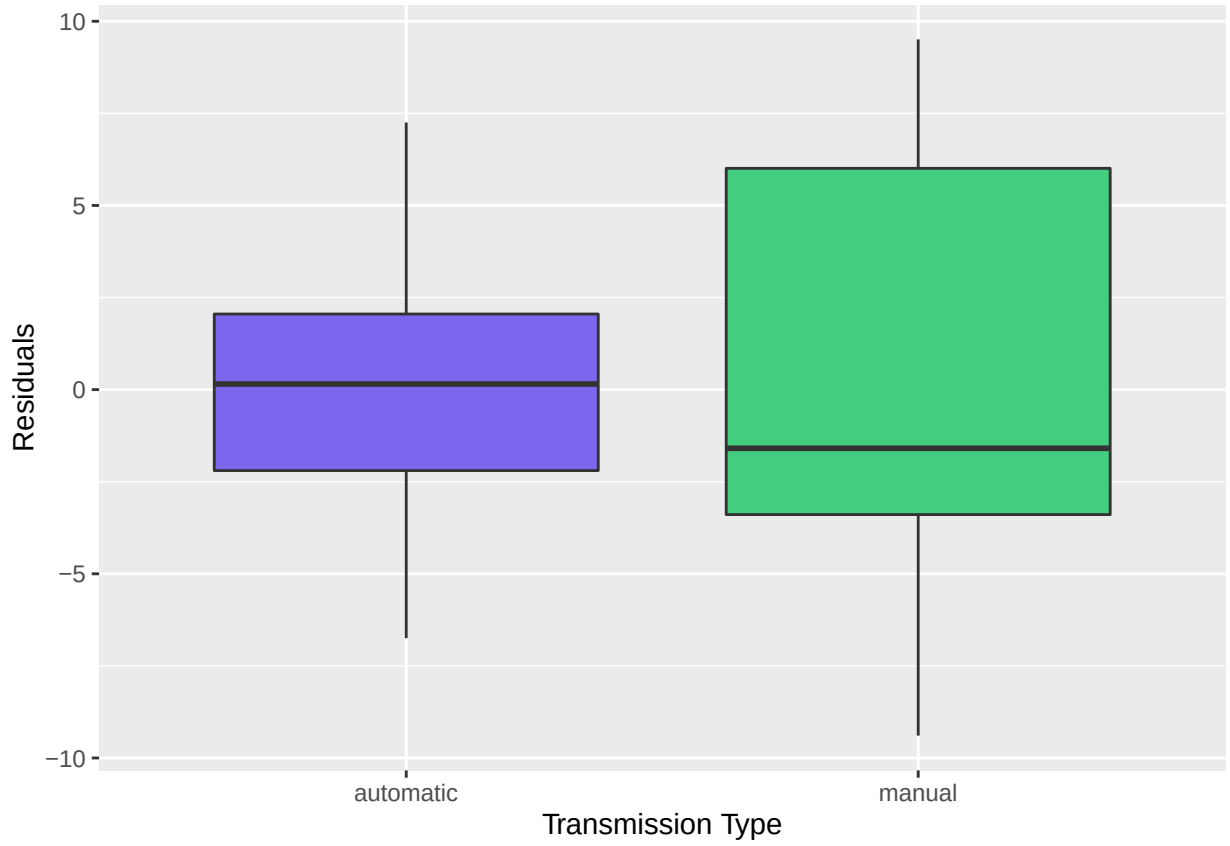
Plot 2

```
par(mfrow=c(2,2))
plot(fit)
```



Plot 3

```
ggplot(data = cars, aes(x=am, y=resid(fit)))+geom_boxplot(fill=c("slateblue2", "seagreen3")) +xlab("Tra
```


10
5
0
−5
−10
Residuals
automatic
manual
Transmission Type