Regression Models Course Project - Transmission vs. MPG

Naman Gandhi

February 21, 2017

Executive Summary

Motor Trend, a magazine about the automobile industry is looking at a data set of a collection of cars, they are interested in exploring the relationship between transmission's influence on miles per gallon (MPG). We'll use the **mtcars** dataset from R and do the analysis with objective to understand:
1) **Is an automatic or manual transmission better for MPG**
2) **Quantify the MPG difference between automatic and manual transmissions**

Takeaways
1) Manual transmission is better for MPG by a factor of 1.8 compared to automatic transmission
2) Single term linear models cannot explain variance in the model, we need other variables to explain better
3) Means and medians for automatic and manual transmission cars are significantly different

Understanding the data - Exploratory Data Analysis

We'll look at structure and summary of data to understand how this data looks like

```
str(mtcars)
summary(mtcars)
```

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
## Converting the variables to factor for better modeling
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
## am - Transmission (0 = automatic, 1 = manual)
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual"))
```

Looking at the box-plots helps us understand that transmission indeed brings an effect on MPG and automatic transmission cars have lower MPG than the Manual ones. (Appendix : Figure 1)

Let's also look at **Pairwise correlations**, they will help us formulate model construct (Appendix : Figure 2)

Regression Analysis

Let's start with **Basic Linear Regression**

```
fit1 <- lm(mpg ~ am, data=mtcars)
coef(summary(fit1))
```

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

```
summary(fit1)$r.squared
```

```
## [1] 0.3597989
```

Insights -
i) The (Intercept) Estimate from the above coefficients show the mean MPG for automatic transmissions (our baseline) is 17.1 mpg while that of the manual tranmission as indicated by the "am" Estimate us 7.2 mpg higher than the baseline
ii) R^2 of 35% indicates a poor model fit

Let's throw in few more variables into the model, take reference from the pairwise correlation plot

Multivariate Linear Regression

```
fit2 <- lm(mpg~am + cyl + disp + hp + wt, data = mtcars)
coef(summary(fit2))
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 33.864276061 2.69541569 12.5636562 2.668321e-12
## amManual     1.806099494 1.42107933  1.2709350 2.154510e-01
## cyl6        -3.136066556 1.46909031 -2.1346996 4.277253e-02
## cyl8        -2.717781289 2.89814941 -0.9377644 3.573375e-01
## disp         0.004087893 0.01276729  0.3201848 7.514890e-01
## hp          -0.032480178 0.01398322 -2.3227963 2.862128e-02
## wt          -2.738694608 1.17597755 -2.3288664 2.824553e-02
```

```
summary(fit2)$r.squared
```

```
## [1] 0.8664276
```

```
f <- summary(fit2)$fstatistic # F-statistic
(p <- pf(f[1],f[2],f[3],lower.tail=F)) # p-value
```

```
##        value
## 8.861459e-10
```

```
## Let's do a quick analysis of variance
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + disp + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     25 150.41  5    570.49 18.965 8.637e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Insights and Conclusion -

i) This results in a p-value of 8.637e-08, and we can validate that fit2 model is significantly better than fit1 simple lm.
ii) Residuals for non-normality (Appendix) and can see they are all normally distributed and their is no evident heteroskedasticity. (Appendix : Figure 3)
iii) The model explains 86.64% of the variance and as a result, cyl, disp, hp, wt did affect the correlation between mpg and am. Thus, we can say the difference between automatic and manual transmissions is 1.81 MPG.

Appendix

Figure 1: Boxplot - Understanding Transmission as opposed to MPG

```
## Mean MPG of Manual and Automatic transmission cars is  24.39231 mpg and 17.14737 mpg respectively
```

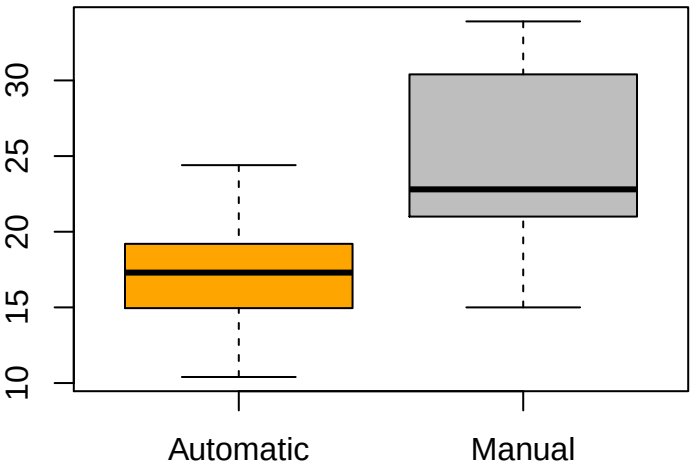

Figure 2: Pairwise correlations

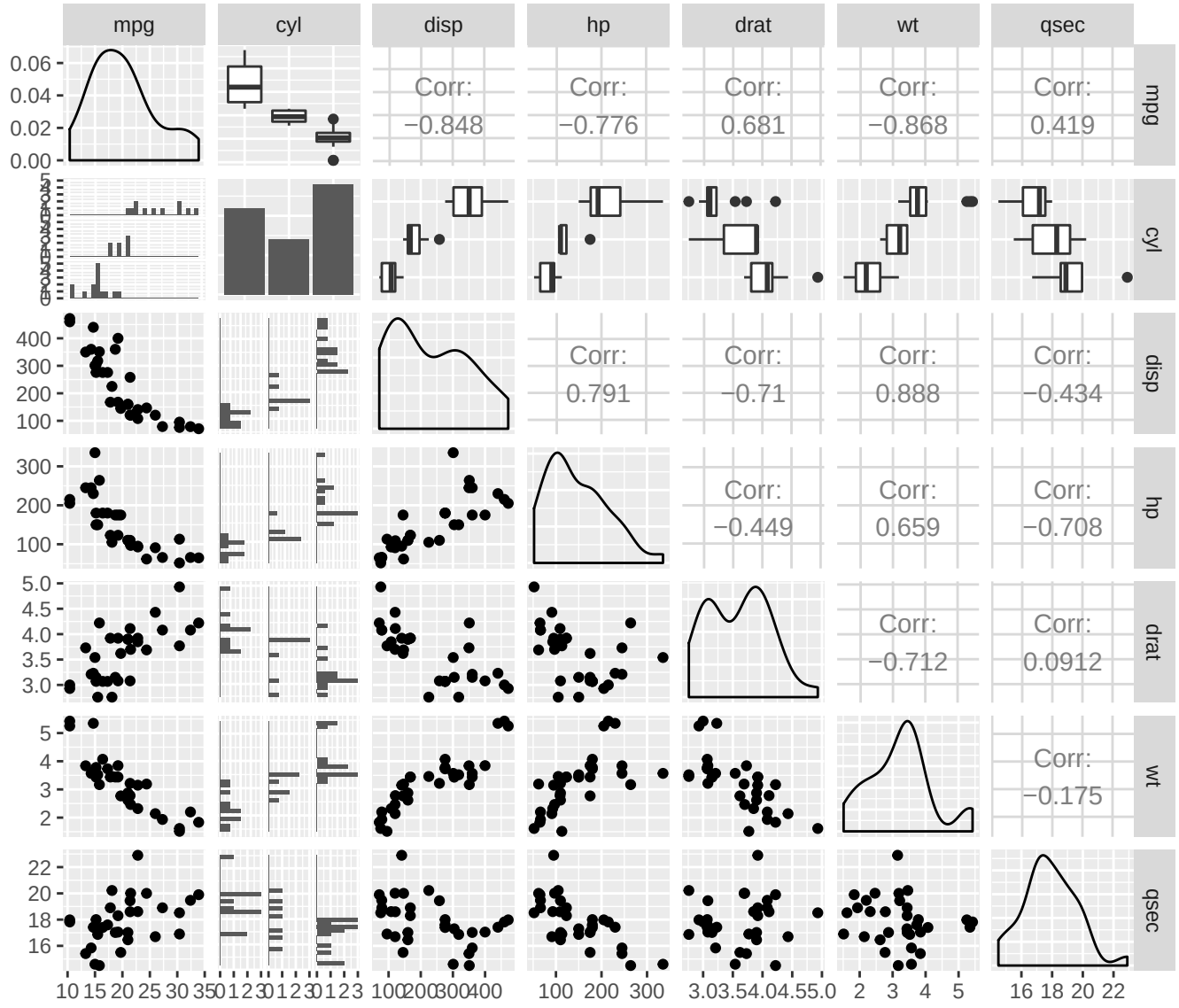

Figure 3: Model Residual Plots

```
## The Linear Regression is formulated as: mpg ~ am + cyl + disp + hp + wt
```