Analysis for Motor Trend magazine

Executive Summary

This paper explores the relationship between a set of variables and miles per gallon (mpg) using mtcars dataset. In particular, the focus is on answering two questions:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

Based on correlation analysis and linear regression model, transmission type is found to be a moderate predictor for MPG. Therefore no significant conclusions can be made regarding transmission type's correlation with MPG.

Other variables in mtcars dataset (like cylinders, weight) may have a stronger correlation with MPG but that requires a more detailed analysis that is outside the scope of this paper.

Data Exploration

Load Data

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

Relationship between MPG and Transmission type (single variable analysis)

The boxplot in Figure 1 in Appendix shows that MPG distribution is generally higher in cars with manual transmission as compared to MPG distribution in cars with automatic transmission. Figure 1 answers the first question about what transmission is better for MPG using a single variable relationship.

However, MPG may also be dependent on any of the other variables like number of cylinders, weight, displacement and horsepower. The correlation coefficients below show that number of cylinders (cyl), displacement (disp), horsepower (hp) and weight (wt) have a strong negative linear relationship with MPG. While rear axle ratio (drat) and transmission (am, vs) have strong to moderate positive linear relationship with MPG.

```
cor(mtcars, mtcars$mpg)[-1,]
```

```
##        cyl       disp         hp       drat         wt       qsec
## -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594  0.4186840
##         vs         am       gear       carb
##  0.6640389  0.5998324  0.4802848 -0.5509251
```

We will now fit a linear regression model with MPG as outcome and transmission as predictor.

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

The above coefficient summary show that cars with automatic transmissions achieve 17.15mpg fuel economy on average and that cars with manual transmission achieve 24.39mpg fuel economy on average.

Relationship between MPG and other variables (mutlivariate analysis)

Due to the 2-page requirement, we will limit the model to add weight (wt) as a predictor along with transmission. We chose weight because it shows the strongest relationship with the outcome (mpg). We will use anova() function to compare the two models.

```
fit2 <- lm(mpg ~ am + wt, data=mtcars)
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 46.115 1.867e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residuals

The residual plot in Figure 2 produces no identifying pattern.

Appendix

Figure 1: MPG values by Transmission type

```
library(ggplot2)
mtcars$trn <- factor(mtcars$am, labels=c("automatic", "manual"))
ggplot(mtcars, aes(x=trn, y=mpg, fill=trn)) +
  geom_boxplot() +
  geom_jitter(aes(colour=trn)) +
  xlab("Transmission") +
  ylab("Miles Per Gallon (MPG)") +
  ggtitle("Figure 1: MPG values by Transmission type") +
  scale_color_discrete(name="Transmission") +
  guides(fill=guide_legend(title="Transmission"))
```

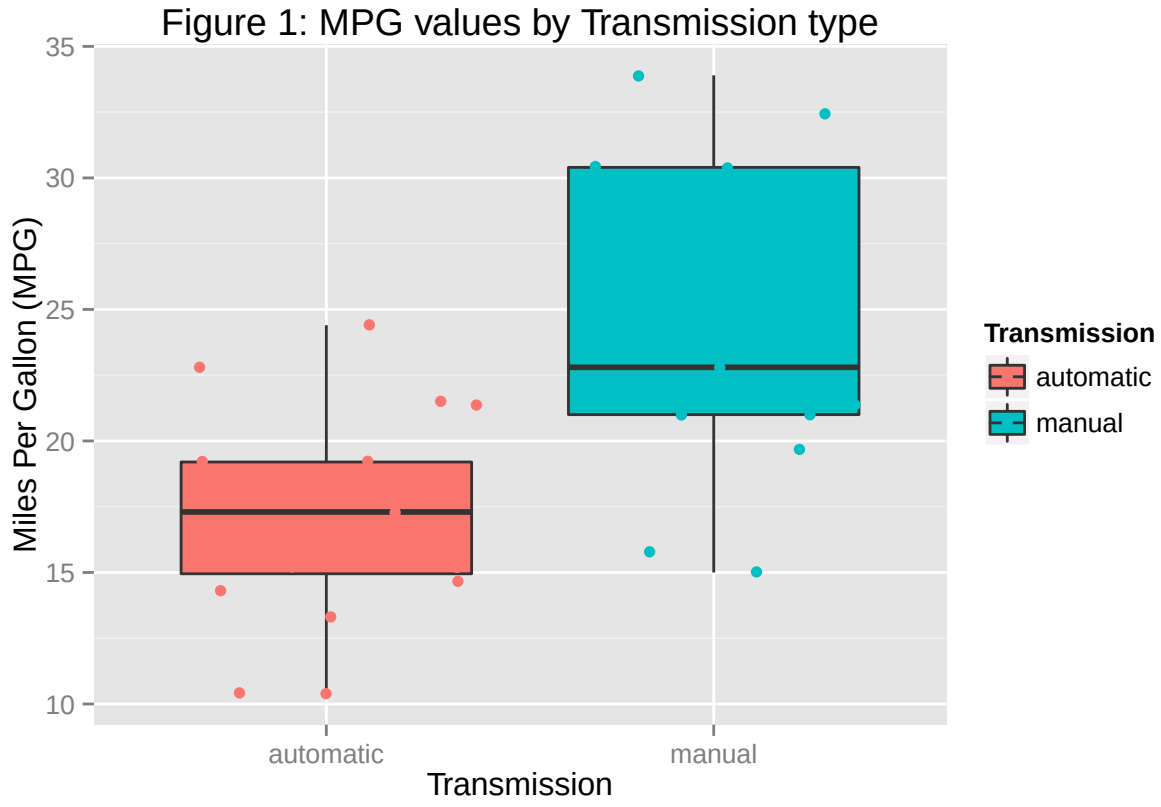
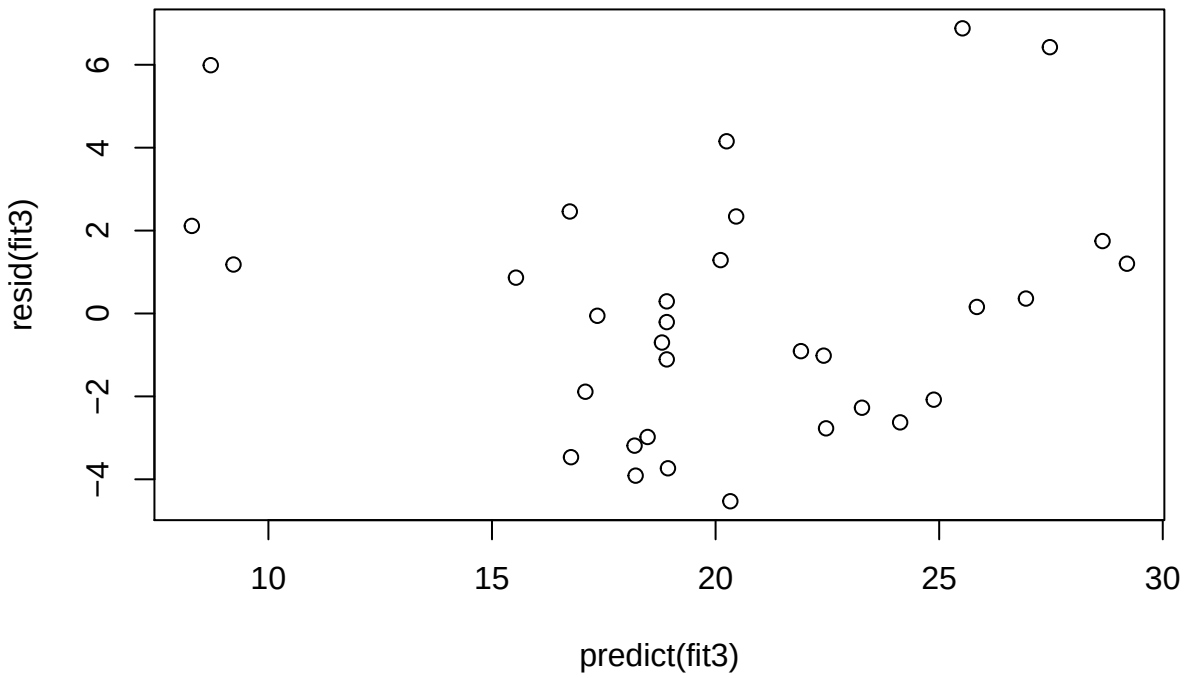


Figure 2: Residual plot

```
fit3 <- lm(mtcars$mpg ~ am + wt, data=mtcars)
plot(resid(fit3) ~ predict(fit3), title="Residual plot")
```

```
## Warning in plot.window(...): "title" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "title" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "title" is
## not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "title" is
## not a graphical parameter

## Warning in box(...): "title" is not a graphical parameter

## Warning in title(...): "title" is not a graphical parameter
```


resid(fit3)
6
4
2
0
−2
−4
10
15
20
25
30
predict(fit3)