# Regression Models: Course Project

## Effect of Transmission Type on Car Fuel Consumption

### Executive Summary

Using the mtcars data set, the purpose of this analysis, is to study the effect of the transmission type on fuel consumption. The study is aiming to answer following two questions:

• Is an automatic or manual transmission better for MPG?

• Quantify the MPG difference between automatic and manual transmissions

To answers these questions, we use exploratory data analysis and regression models.

### Data Processing

In this step, we load and read the data and prepare it for analysis. Looking at the data headers, the field "am" is going to be the predictor variable that guides the study. This variable can be converted to a factor class with better descriptive labels: "Automatic" and "Manual".

```
data(mtcars)
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

### Exploratory Data Analysis

To ensure that our regression model will be accurate, we will analyze and plot the "mpg" dependent variable to check its distribution.

```
par(mfrow = c(1, 2))
xMpg <- mtcars$mpg
h<-hist(xMpg, breaks=10, col="blue", xlab="Miles Per Gallon (mpg)", main="Distribution of Miles per $
xfit<-seq(min(xMpg),max(xMpg),length=40)
yfit<-dnorm(xfit,mean=mean(xMpg),sd=sd(xMpg))
yfit <- yfit*diff(h$mids[1:2])*length(xMpg)
lines(xfit, yfit, col="blue", lwd=2)
den <- density(mtcars$mpg)
plot(den, xlab = "Miles per Gallon (mpg)", main ="Miles per Gallon (mpg) Density")
```

See Fig. 1 in Appendix

The plots show that the distribution is acceptably clean or normal and there are no skewing outliers.

Now we analyze and compare the transmissions: Automatic vs Manual

```
boxplot(mpg~am, data = mtcars,
        col = c("blue", "green"),
        xlab = "Transmission Type",
        ylab = "Miles per Gallon (mpg)",
        main = "Miles per Gallon (mpg) per Transmission Type")
```

See Fig. 2 in Appendix

Manual transmissions show better utilization of fuel than automatic transmissions.

## Hypothesis

In this section, we will throw a hypothesis but first we have to get the mean of each:

```
aggregate(mpg~am, data = mtcars, mean)
```

```
         am      mpg
Automatic 17.14737
   Manual 24.39231
```

The difference is 7.245 MPGs in favor of manual transmissions.

Does this difference stand to be statistically significant?

To test this, the alpha-value is set to 0.5, and a t-test is run to test the hypothesis:

```
autoData <- mtcars[mtcars$am == "Automatic",]
manualData <- mtcars[mtcars$am == "Manual",]
t.test(autoData$mpg, manualData$mpg)
```

```
        Welch Two Sample t-test

data:  autoData$mpg and manualData$mpg
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 of x mean of y
 17.14737  24.39231
```

The null hypothesis is rejected due to the p-value of 0.001374.

## The Model

First, we create a correlation matrix for the mtcars dataset and look at the row for mpg to decide on the predictors to be used in the model.

```
data(mtcars)
sort(cor(mtcars)[1,])
```

```
        wt        cyl       disp         hp       carb       qsec       gear         am         vs
-0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251  0.4186840  0.4802848  0.5998324  0.6640389
      drat        mpg
 0.6811719  1.0000000
```

It is determined that wt, cyl, disp, and hp are highly correlated with mpg, therefore they can be candidates for the model. It is also determined that cyl and disp are are highly correlated with each other so they were excluded from the model.

## Regression Analysis

We first fit a simple linear regression for mpg on am.

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

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

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 ***
am             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
```

Then, we do a multivariate linear regression fit for mpg on am, wt, and hp.

```
bestfit <- lm(mpg~am + wt + hp, data = mtcars)
anova(fit, bestfit)
```

```
Analysis of Variance Table

Model 1: mpg ~ am
Model 2: mpg ~ am + wt + hp
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1     30 720.90                                  
2     28 180.29  2    540.61 41.979 3.745e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is rejected due to the p-value of 3.745e-09.

Now we show the plots and summary

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

See figure 3 in Appendix

```
summary(bestfit)
```

```
Call:
lm(formula = mpg ~ am + wt + hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4221 -1.7924 -0.3788  1.2249  5.5317

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
am           2.083710   1.376420   1.514 0.141268
wt          -2.878575   0.904971  -3.181 0.003574 **
hp          -0.037479   0.009605  -3.902 0.000546 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.538 on 28 degrees of freedom
Multiple R-squared:  0.8399,    Adjusted R-squared:  0.8227
F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

## Conclusion

We conclude that cars fitted with manual transmissions have better fuel efficiency than cars with automatic transmissions.

## Appendix

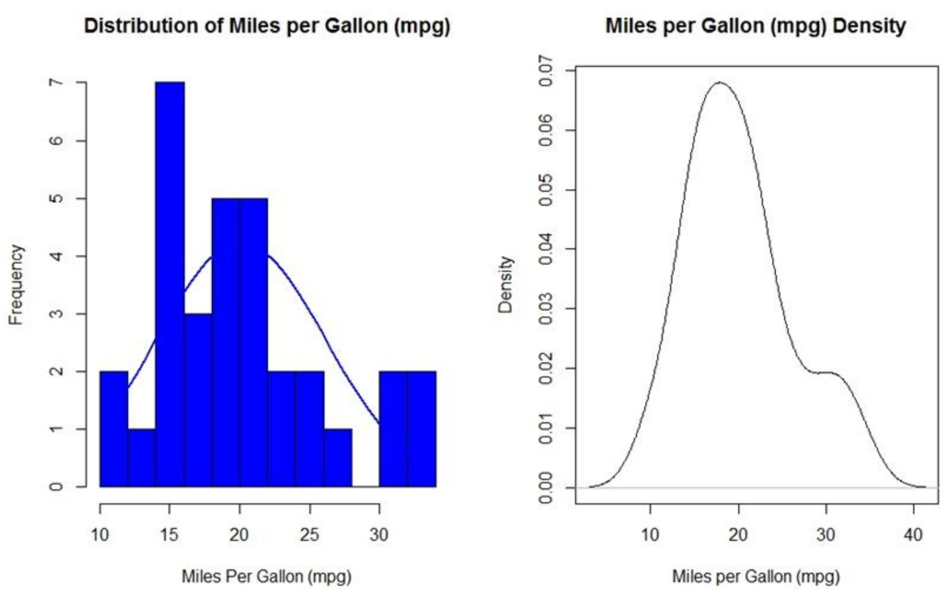


Figure 1

----------------------------------------------------------------------------------------------------

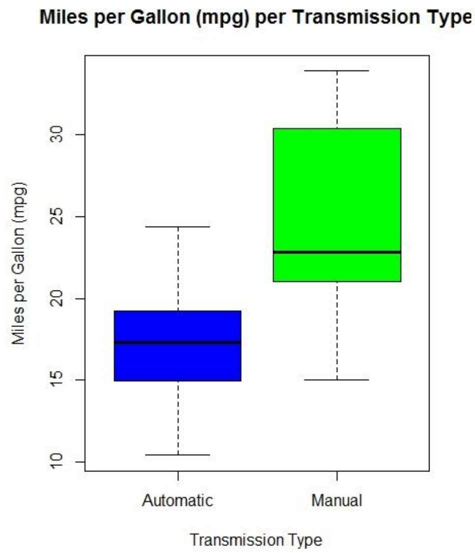


Figure 2

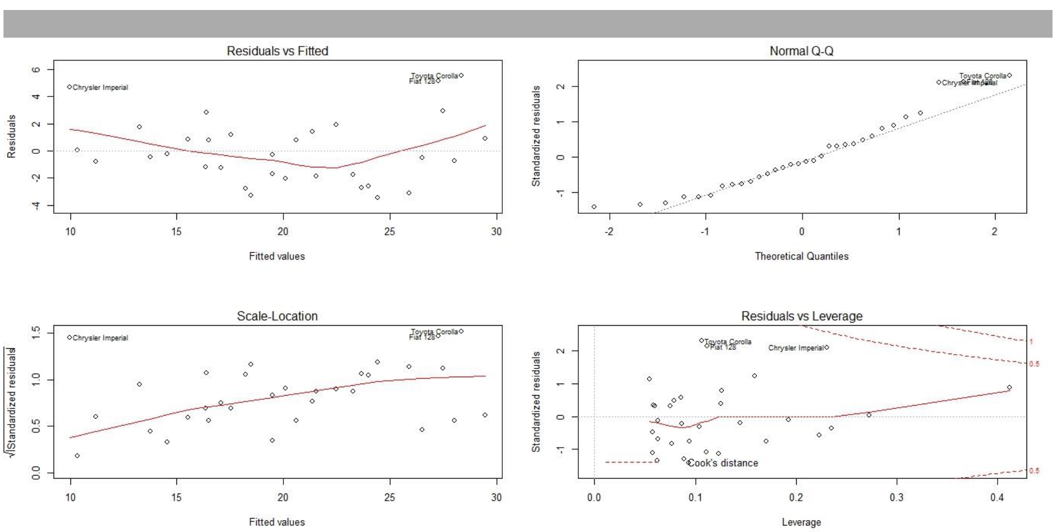


Figure 3