mtcars_analysis

Executive summary

This report provides an analysis to answer if there is a difference between manual and automatic transmissions with respect to miles per (US) gallon. The methods of the analysis include exploratory analysis, regression model, nested models, anova tests, regression models with interaction between key regressor and covariants and diagnostics. The report found with a 95% confidence level that cars with manual transmission do ~3.30 to ~19.80 miles per (US) gallon better than the cars with automatic transmission on average, holding gross horsepower constant, but with weight interaction. This difference proved to be significant. The report found no significant difference in the average miles per (US) gallon between cars with automatic and manual transmission holding weight and gross horsepower constant.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.3

library(statsr)
library(car)

## Warning: package 'car' was built under R version 3.4.4
```

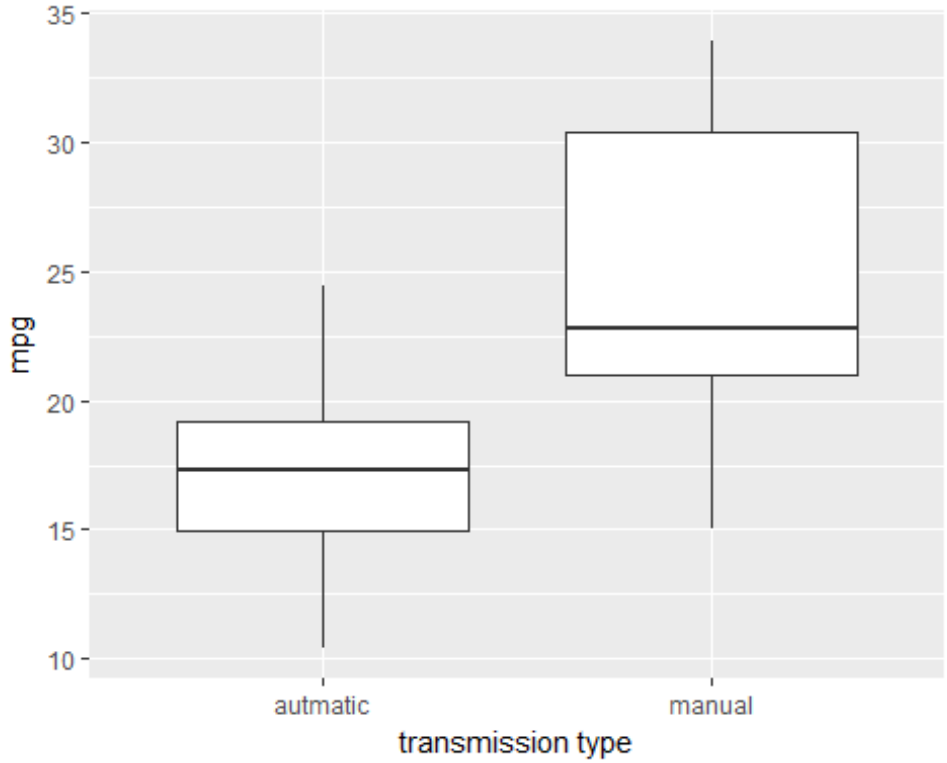
Exploratory Analysis

We can observe a difference the average miles per (US) gallon, between automatic and manual transmission. The manual transmission has higher median miles per (US) gallon.

```
mtcars<- mtcars %>% mutate(transmission = ifelse(mtcars$am==
1,'manual','autmatic'))

## Warning: package 'bindrcpp' was built under R version 3.4.3

mtcars<- transform(mtcars, transmission = factor(transmission))
ggplot(mtcars,aes(x =transmission, y =mpg)) + geom_boxplot() + labs(y ="mpg",
x = "transmission type")
```



Strategy

I will start with a simple regression model, the response variable will be mpg and the key regressor variable will be the transmission. I will then add other covariants and run an anova test in order to select the best model. After choosing the model I will interact the key regressor with other covariants and run an anova test between the adjusted and the interaction model to spot any significant difference.

Regression Model

In the first model, the data provide evidence of a significant difference in the average mpg between automatic and manual transmission, as the p-value is lower than 5% significant level. Cars with manual transmission do ~7.25 miles per (US) gallon better than the cars with automatic transmission on average.

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

##
## Call:
## lm(formula = mpg ~ transmission, 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 ***
## transmissionmanual    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
```

Nested models

In the fit2 model, the difference in the average mpg between automatic and manual transmission is no longer significant.

```
fit2<- lm(mpg~transmission + hp + wt, data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ transmission + hp + wt, 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 ***
## transmissionmanual  2.083710   1.376420   1.514 0.141268
## hp                 -0.037479   0.009605  -3.902 0.000546 ***
## wt                 -2.878575   0.904971  -3.181 0.003574 **
## ---
## 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
```

```
fit4<-update(fit2,mpg~transmission + hp + wt + as.factor(cyl) + disp +
as.factor(carb))
```

ANOVA

Anova results confirmed that the appearance of hp and wt variables in our original model proved to be necessary, as the p-value is below 5% significant level.

When running anova is necessary to run a Shapiro test in order to verify if the residuals our model are normally distributed. Because the p-value is higher than the 5% significant level we fail to reject the null hypothesis that our residuals are normally distributed.

```
anova(fit1,fit2,fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ transmission
## Model 2: mpg ~ transmission + hp + wt
## Model 3: mpg ~ transmission + hp + wt + as.factor(cyl) + disp +
as.factor(carb)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 180.29  2    540.61 39.257 1.189e-07 ***
## 3     20 137.71  8     42.58  0.773    0.6305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
shapiro.test(fit2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit2$residuals
## W = 0.9453, p-value = 0.1059
```

Interaction

In the new model, mdl2, we can observe a significant difference in average mpg between cars with an automatic and manual transmission.

Anova results confirmed that the interaction between the key regressor and the wt variable is necessary for our model. The residuals of the new model follow a normal distribution, the p-value is higher than 5% significant level.

```
mdl<- lm(mpg~ transmission * hp +wt, data = mtcars)
summary(mdl)
```

```
##
## Call:
## lm(formula = mpg ~ transmission * hp + wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.435 -1.510 -0.697  1.284  5.245
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           33.34196    2.79711  11.920 2.89e-12 ***
## transmissionmanual     3.55141    2.35742   1.506  0.14355
## hp                    -0.02918    0.01449  -2.014  0.05407 .
## wt                    -3.05617    0.94036  -3.250  0.00309 **
## transmissionmanual:hp -0.01129    0.01466  -0.770  0.44809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.556 on 27 degrees of freedom
## Multiple R-squared:  0.8433, Adjusted R-squared:  0.8201
## F-statistic: 36.33 on 4 and 27 DF,  p-value: 1.68e-10
```

```
mdl2<- lm(mpg~ transmission * wt + hp, data = mtcars)
summary(mdl2)
```

```
##
## Call:
## lm(formula = mpg ~ transmission * wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0639 -1.3315 -0.9347  1.2180  5.0822
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           30.947333   2.723411  11.363 8.55e-12 ***
## transmissionmanual    11.554813   4.023277   2.872  0.00784 **
## wt                    -2.515586   0.844497  -2.979  0.00605 **
## hp                    -0.026949   0.009796  -2.751  0.01048 *
## transmissionmanual:wt -3.577910   1.442796  -2.480  0.01968 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.332 on 27 degrees of freedom
## Multiple R-squared:  0.8696, Adjusted R-squared:  0.8503
## F-statistic: 45.01 on 4 and 27 DF,  p-value: 1.451e-11
```

```
anova(fit2,mdl2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ transmission + hp + wt
## Model 2: mpg ~ transmission * wt + hp
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     28 180.29
## 2     27 146.84  1    33.446 6.1496 0.01968 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
shapiro.test(mdl2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mdl2$residuals
## W = 0.91554, p-value = 0.01577
```

Wea are 95% confident that cars with manual transmission do ~3.30 to ~19.80 miles per (US) gallon better than the cars with automatic transmission on average, holding gross horsepower constant, but with weight interaction.

```
confint(mdl2)
```

```
##                               2.5 %       97.5 %
## (Intercept)            25.35935553 36.535310859
## transmissionmanual      3.29973130 19.809894626
## wt                     -4.24834926 -0.782821748
## hp                     -0.04704888 -0.006849817
## transmissionmanual:wt -6.53828181 -0.617537786
```

diagnostics

The residuals vs fitted plot do not present any specific pattern, the residuals follow approximately a normal distribution.

```
par(mfrow = c(2,2))
plot(mdl2)
```