

Relationship between transmission type and MPG

Summary

The Motor Trends data were analyzed to study the effect of transmission type on miles per gallon (MPG) in this document. First, we performed a t-test to see if there is any significant difference of MPG between the automatic and manual cars. The analysis result shows there is significant difference. Seoncd, we quantified the effect of transmission type on mpg using a linear regression model. It is shown that manual transmission can increase the mpg by 2.936 when other factors are held constant.

Question 1: Automatic or manual transmission is better for higher MPG?

We first transform the Motor Trends data by converting the transmission,cyl,vs,gear,carb columns to factor variables.

```
library(plyr)
cars <- mutate(mtcars,cyl=factor(cyl), vs=factor(vs),gear=factor(gear),carb=factor(carb),
               am=factor(ifelse(am,"manual","automatic"), levels=c("automatic","manual")))
```

Then we obtain the mpg data for automatic and manual transmission from the data set. It appears that automatic transmission has higher mpg, as shown in Figure 1 in Appendix. We then perform the student t-test and the results suggests that there is a significant difference since the pvalue is less than 0.05.

```
auto <- subset(cars,cars$am=="automatic",select=c(mpg))[,1]
man <- subset(cars,cars$am=="manual",select=c(mpg))[,1]
t.test(auto,man)
```

```
##
## Welch Two Sample t-test
##
## data: auto and man
## 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
```

Question 2: the effect of transmission type on mpg

We further transform the data by centering the non-categorical data such as mpg,disp,hp,drat columns.

```
cars <- mutate(cars,wt=(wt-mean(wt)), qsec=(qsec-mean(qsec)), disp=(disp - mean(disp)),
               hp=(hp - mean(hp)), drat=(drat-mean(drat)))
```

The first model considers all the variables in the data as independent variable. For this model, F-statistics p-value is less than 0.05 but the p-value for the coefficients indicate that the null hypothesis can not be rejected.

```
modelallfactor <- lm(mpg~.,data=cars)
summary(modelallfactor)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.98425     5.32412   3.378  0.00414 **
## cyl6         -2.64870     3.04089  -0.871  0.39747
## cyl8         -0.33616     7.15954  -0.047  0.96317
## disp          0.03555     0.03190   1.114  0.28267
## hp           -0.07051     0.03943  -1.788  0.09393 .
## drat          1.18283     2.48348   0.476  0.64074
## wt           -4.52978     2.53875  -1.784  0.09462 .
## qsec          0.36784     0.93540   0.393  0.69967
## vs1           1.93085     2.87126   0.672  0.51151
## ammanual      1.21212     3.21355   0.377  0.71132
## gear4         1.11435     3.79952   0.293  0.77332
## gear5         2.52840     3.73636   0.677  0.50890
## carb2        -0.97935     2.31797  -0.423  0.67865
## carb3         2.99964     4.29355   0.699  0.49547
## carb4         1.09142     4.44962   0.245  0.80956
## carb6         4.47757     6.38406   0.701  0.49381
## carb8         7.25041     8.36057   0.867  0.39948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF, p-value: 0.000124
```

Then we use the Akaike Information Criteria (AIC) to find a model that can be used to quantify the effect of transmission on mpg. Step function is used to compare the performance of different models.

The best model selected using step function consider “wt”, “qsec” and “automatic” as the independent variables. The individual p-values for the model parameters suggest to reject the hypothesis that the coefficients are null. The parameter for manual transmission suggests the mpg increase 2.936 if manual transmission is used instead of automatic transmission.

```
summary(bestmodel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.9387 -1.2560 -0.4013 1.1253 5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.96535    1.56753  13.375 3.65e-13 ***
## cyl6        -3.03134    1.40728   -2.154 0.04068 *
## cyl8        -2.16368    2.28425   -0.947 0.35225
## hp          -0.03211    0.01369   -2.345 0.02693 *
## wt          -2.49683    0.88559   -2.819 0.00908 **
## ammanual     1.80921    1.39630    1.296 0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The regression diagnostic plots for the best model is shown in Appendix.

Appendix

Summary for MPG Data

```
summary(cars[cars$am == "Automatic",])
```

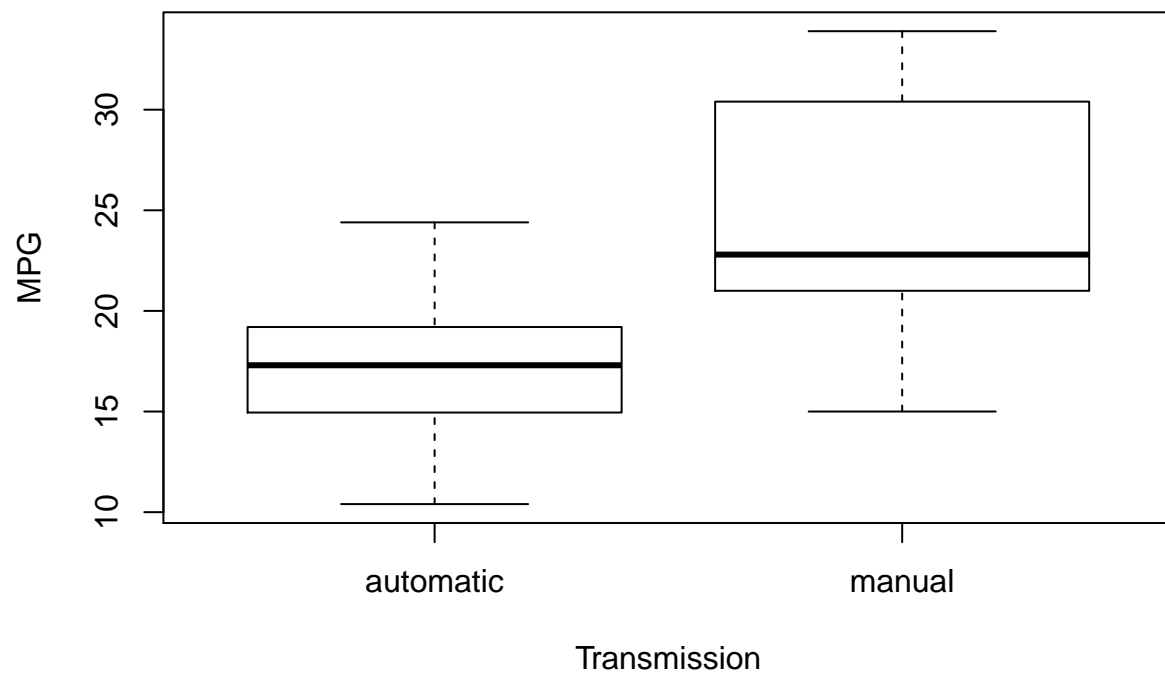
```
##      mpg      cyl      disp      hp      drat
## Min.   : NA    4:0   Min.   : NA   Min.   : NA   Min.   : NA
## 1st Qu.: NA    6:0   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
## Median : NA    8:0   Median : NA   Median : NA   Median : NA
## Mean   :NaN           Mean :NaN   Mean :NaN   Mean :NaN
## 3rd Qu.: NA      3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
## Max.   : NA      Max.   : NA   Max.   : NA   Max.   : NA
##      wt      qsec      vs      am      gear carb
## Min.   : NA   Min.   : NA   0:0   automatic:0   3:0   1:0
## 1st Qu.: NA   1st Qu.: NA   1:0   manual    :0   4:0   2:0
## Median : NA   Median : NA               5:0   3:0
## Mean   :NaN   Mean   :NaN               4:0
## 3rd Qu.: NA   3rd Qu.: NA               6:0
## Max.   : NA   Max.   : NA               8:0
```

```
summary(cars[cars$am == "Manual",])
```

```
##      mpg      cyl      disp      hp      drat
## Min.   : NA    4:0   Min.   : NA   Min.   : NA   Min.   : NA
## 1st Qu.: NA    6:0   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
## Median : NA    8:0   Median : NA   Median : NA   Median : NA
## Mean   :NaN           Mean :NaN   Mean :NaN   Mean :NaN
## 3rd Qu.: NA      3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
## Max.   : NA      Max.   : NA   Max.   : NA   Max.   : NA
##      wt      qsec      vs      am      gear carb
## Min.   : NA   Min.   : NA   0:0   automatic:0   3:0   1:0
## 1st Qu.: NA   1st Qu.: NA   1:0   manual    :0   4:0   2:0
```

```
## Median : NA      Median : NA      5:0    3:0
## Mean   :NaN      Mean   :NaN      4:0
## 3rd Qu.: NA      3rd Qu.: NA      6:0
## Max.   : NA      Max.   : NA      8:0
```

```
boxplot(cars$mpg~cars$am,xlab="Transmission",ylab="MPG")
```



```
plot(bestmodel)
```

