Analysis on relationship of transmission type and MPG on cars

Executive summary

This report is to investigate the relationship between a set of variables and miles per gallon(MPG). The following two questions will be addressed.

- Is an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

The data was collected from 32 automobiles(1973-1974 models). The key findings of this report are:

- Automatic transimission is not better than manual transmission
- The MPG difference for manual and automatic transmission is 1.8 mpg.

Data processing

Acquire the data set and cleaning it.

```
data(mtcars);mtcars$cyl<-factor(mtcars$cyl);mtcars$vs<-factor(mtcars$vs)
mtcars$am<-factor(mtcars$am);mtcars$gear<-factor(mtcars$gear)
mtcars$carb<-factor(mtcars$carb)
```

Exploratory analysis

Figure 1 is a boxplot of mpg vs am, it shows that automatic and manual transmission do seem to have different mpg values. Figure 2 is a pairwise graph to have a look at the relationships between each variables.

Model Building

The step function is used for automatically search of best regression. Start with a model that include all the predictors. The summary is shown in table 1, which demonstrates that not all the predictors are significant. Next, I am going to use step function to reduce the insignificant variables. The analysis is shown in table 2. According to the summary table, the most significant variables are cyl, hp, wt and am. The fitting yields an R-squared value of 84%, meaning that very high percentage of the variation can be explained by this model.

Residual and diagnosis

Figure 3 shows the residual plots for the best fitted model. From the plots, one can get the following information:

- The residuals vs fitted and scale-location plots have no patterns, indicate the model fits well.
- Normal Q-Q plot show the distribution of standardized residuals is very close to normal distribution
- The residuals vs leverage plot shows that there is no significant outliers with this fitting model

Statistical inference

A t test can be performed to examine the hyphothesis that the transmission type does not have effects on miles/gallon values for different cars. The result is shown in Table 3. The p-value of this test is 0.001374, which means that the probablility of accepting null hypothesis is low, therefore we should reject the null. In other words, the two types do influent the mpg.

Conclusion

From the above analysis, one can see that the transmission type has an effect on the measurement of miles/gallon. The manual transmission is better by 1.8 mpg than the automatic transmission. The fitting model used here has four predictors, which are cyl, hp, wt and am. The R squared value for the fitting model is 84%. In addition, the mean value of mpg for automatic transmission is 17.15 and that for manual transmission is 24.39. The difference between the means is about 7.2 mpg.

Appendix

Figure 1. Boxplot of mpg vs am

```
##
## Attaching package: 'ggplot2'
##
## The following object is masked _by_ '.GlobalEnv':
##
##     mpg
```

```
ggplot(mtcars,aes(x=factor(am,labels=c("Automatic","Manual")),y=mpg,fill=am))+
    geom_boxplot()+theme_bw()+xlab("Transmission type")+ylab("Miles/US gallon")+
    ggtitle("Transmission type vs Miles/(US) gallon of cars")+
    theme(legend.position="none")
```

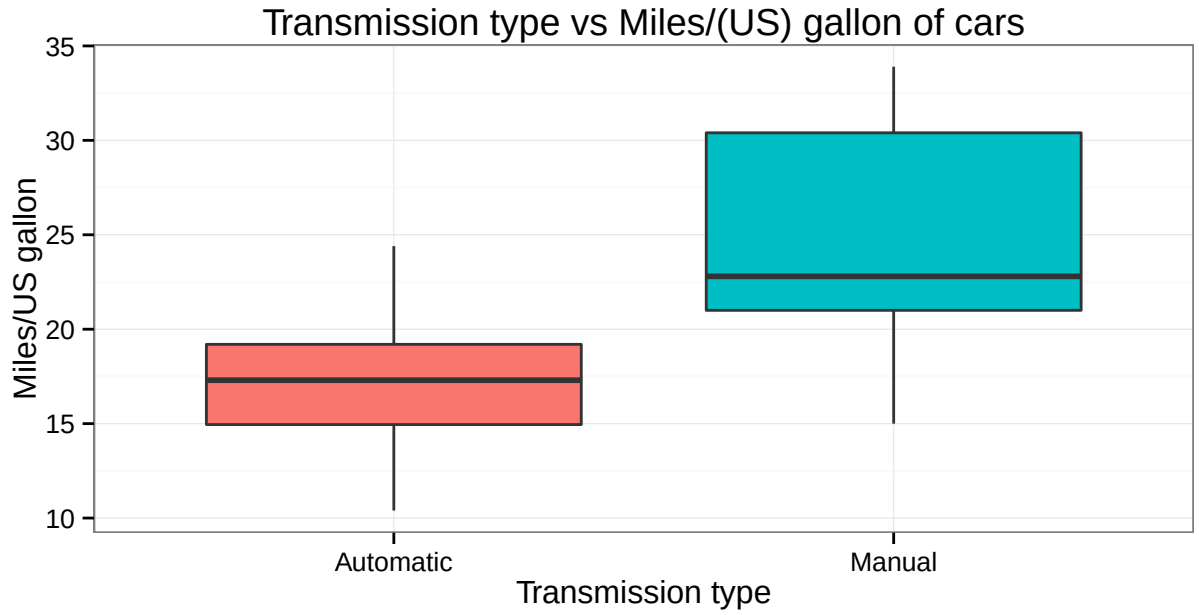

Figure 2. Pairwise graph of "mtcars" data

```
pairs(mtcars, panel=panel.smooth)
```

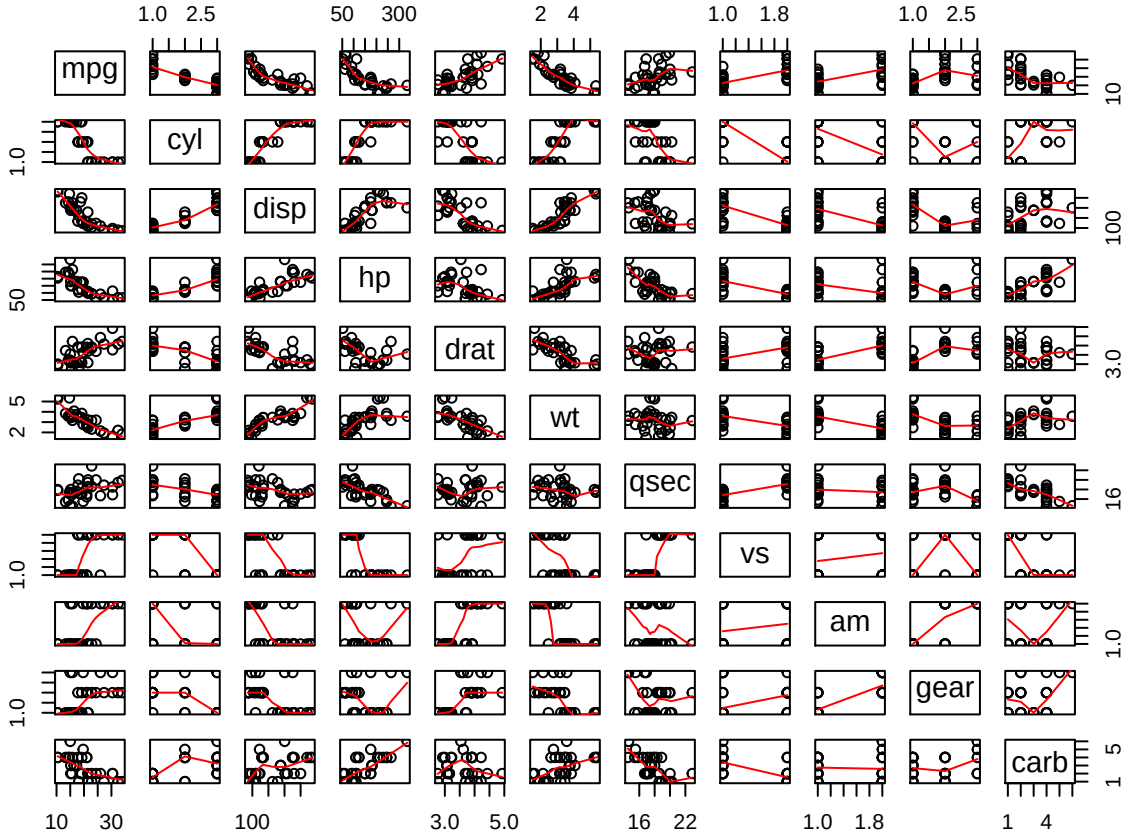

Table 1. Model fitted with all predictors

```
fitall <- lm(mpg~., mtcars)
summary(fitall)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.509 -1.358 -0.095  0.775  4.625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.8791    20.0658    1.19    0.253
## cyl6         -2.6487     3.0409   -0.87    0.397
## cyl8         -0.3362     7.1595   -0.05    0.963
## disp          0.0355     0.0319    1.11    0.283
## hp           -0.0705     0.0394   -1.79    0.094 .
## drat          1.1828     2.4835    0.48    0.641
```

```
## wt          -4.5298     2.5387   -1.78    0.095 .
## qsec         0.3678     0.9354    0.39    0.700
## vs1          1.9309     2.8713    0.67    0.512
## am1          1.2121     3.2135    0.38    0.711
## gear4        1.1144     3.7995    0.29    0.773
## gear5        2.5284     3.7364    0.68    0.509
## carb2       -0.9794     2.3180   -0.42    0.679
## carb3        2.9996     4.2935    0.70    0.495
## carb4        1.0914     4.4496    0.25    0.810
## carb6        4.4776     6.3841    0.70    0.494
## carb8        7.2504     8.3606    0.87    0.399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.83 on 15 degrees of freedom
## Multiple R-squared:  0.893,  Adjusted R-squared:  0.779
## F-statistic: 7.83 on 16 and 15 DF,  p-value: 0.000124
```

Table 2. Model with significant variables

```
fitbest <- step(fitall, direction="backward", trace=0)
summary(fitbest)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.939 -1.256 -0.401  1.125  5.051
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.7083     2.6049   12.94  7.7e-13 ***
## cyl6         -3.0313     1.4073   -2.15   0.0407 *
## cyl8         -2.1637     2.2843   -0.95   0.3523
## hp           -0.0321     0.0137   -2.35   0.0269 *
## wt           -2.4968     0.8856   -2.82   0.0091 **
## am1           1.8092     1.3963    1.30   0.2065
## ---
## 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:  0.866,  Adjusted R-squared:  0.84
## F-statistic: 33.6 on 5 and 26 DF,  p-value: 1.51e-10
```

Figure 3. The residual plots for the best fitted model.

```
par(mfrow=c(2,2))
plot(fitbest, col="cornflowerblue", pch=20, lwd=2)
```

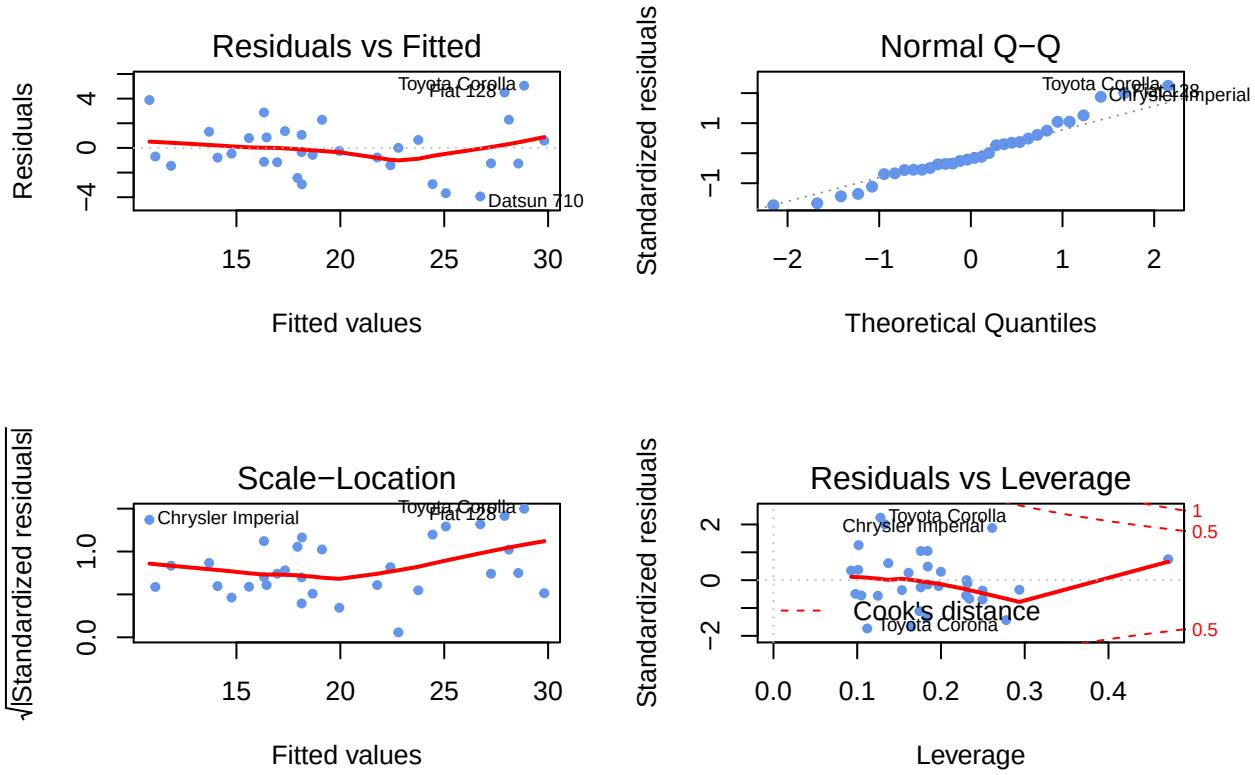

Table 3. P-value for t-test.

```
t.test(mpg~am, data=mtcars)$p.value
```

```
## [1] 0.001374
```