Peer Assessment 1

Summary

The aim of this analysis is to state which transmission, automatic or manual, is better for MPG, and to quantify the MPG difference between automatic and manual transmissions.

Although straightforward analysis by simple regression indicated the manual vs automatic MPG difference of at least 3.6 (lower CI limit), when controlling for wt and qsec in a multivariable regression model, the difference diminished considerably and may be even as small as 0.05. When the multivariable model with interactions is applied, the CI for the manual vs. automatic MPG difference becomes very wide, and the sign of the difference can no longer be stated (the difference is statistically nonsignificant). In such a case both questions remain unanswerable. Collecting more data may help to narrow the obtained CI and the sign of the difference may become visible.

Data loading and exploration

```
data(mtcars)
```

MPG seems negatively correlated with cyl, disp, hp, wt and carb, and positively with drat, qsec, vs and am (App. 1). MPG seems normally distributed (App. 2).

Analysis

MPG in automatic vs. manual - simple comparison disregarding other variables.

This may be achieved by t-test (not shown) or simple linear regression.

```
mtcars$am<-as.factor(mtcars$am)
levels(mtcars$am)<-c("Automatic", "Manual")
m1 <- lm(mpg~am, data=mtcars)
m2 <- lm(mpg~am-1, data=mtcars)
confint(m1, parm="amManual")
```

```
##          2.5 % 97.5 %
## amManual 3.642  10.85
```

```
confint(m2, parm=c("amManual", "amAutomatic"))
```

```
##             2.5 % 97.5 %
## amManual    21.62  27.17
## amAutomatic 14.85  19.44
```

The 95% confidence interval (CI) for mean MPG is [14.8, 19.4] and [21.6, 27.2] for automatic and manual cars, respectatively. The CI for the manual-automatic difference is [3.64, 10.8], so automatic cars seem to have significantly lower MPG, by at least 3.6.

MPG in automatic vs. manual - multivariate linear regression

Stepwise algorithm is used to find the best model for MPG.

```
bestmodel1 <- step(lm(mpg~., data=mtcars), trace=0)
bestmodel1$call
```

```
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
```

The model explains 85% of MPG variance by variables wt, qsec and am. The model seems good (see App. 3).

```
bestmodel2<-lm(mpg~wt+qsec+am-1, data=mtcars)
confint(bestmodel1, parm="amManual")
```

```
##             2.5 % 97.5 %
## amManual 0.04573  5.826
```

```
confint(bestmodel2, parm=c("amManual", "amAutomatic"))
```

```
##                2.5 % 97.5 %
## amManual      0.1457  24.96
## amAutomatic -4.6383  23.87
```

When controlling for the wt and qsec variables, the 95% CI for the mean MPG is [-4.64, 23.87] and [0.15, 24.96] for automatic and manual cars, respecatively. The CI for the manual-automatic MPG difference is [0.05, 5.83]. The difference is statistically significant (p-val=0.047), however, it is pretty small.

MPG in automatic vs. manual - multivariate linear regression with interaction

```
twoslopes <- lm(mpg~wt*am+qsec*am, data=mtcars)
twoslopes2 <- lm(mpg~wt:am+qsec:am, data=mtcars)
confint(twoslopes, parm="amManual")
```

```
##            2.5 % 97.5 %
## amManual -17.11  34.96
```

```
confint(twoslopes2)
```

```
##                    2.5 % 97.5 %
## (Intercept)        2.119 25.820
## wt:amAutomatic    -4.481 -1.870
## wt:amManual       -8.086 -4.112
## amAutomatic:qsec  0.300  1.368
## amManual:qsec     0.894  1.999
```

The model with interactions explains almost 90% of MPG variance and is good (App. 4). The 95% CI for the manual vs. automatic MPG difference is [-17.1, 35.0]. It is not statistically significant (CI includes 0), so we cannot state the sign of the difference in MPG and the two questions become unanswerable. Besides, when the weight increases 1000 lbs, the MPG decreases by [1.87, 4.48] for automatic cars, and by [4.11, 8.09] for manual cars. In case of the 1 sec increase of the 1/4 mile time (qsec), the MPG increases by [0.3, 1.37] for automatic cars and by [0.89, 2.00] for manual cars (numbers in brackets are 95% CI limits).

Appendix

App. 1

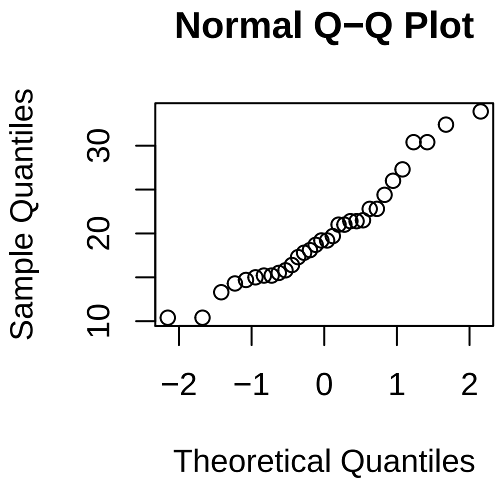
Analysis of correlations between mtcars variables. Additional analysis was achieved by cor() function (not shown due to lack of space):

```
pairs(mtcars, pch=16)
```

App. 2

Check for the normality of mpg distribution. The distribution seems normal, which was further confirmed by Shapiro-Wilk test (not shown due to lack of space).

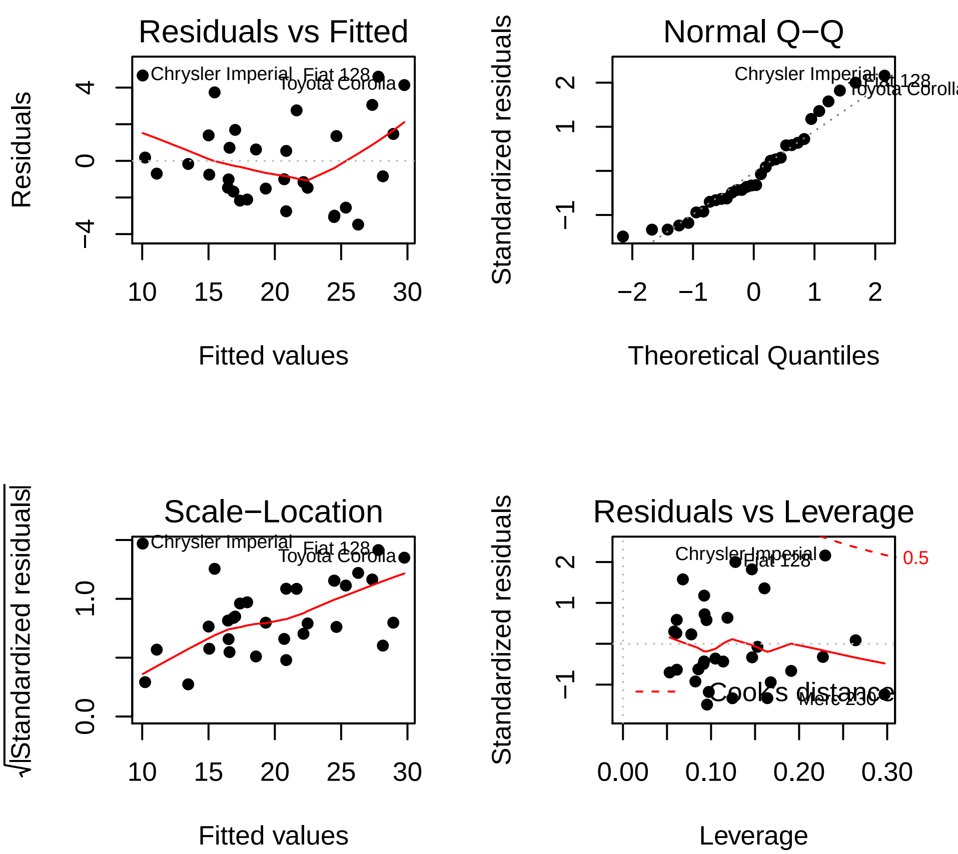
```r
qqnorm(mtcars$mpg)
```



App. 3

Analysis of residuals shows they are normally distributed (Q-Q plot) and homoskedastic (residuals vs fitted).

```r
par(mfrow = c(2,2))
plot(bestmodel1, pch=16)
```



App. 4

Analysis of residuals shows they are normally distributed (Q-Q plot) and homoskedastic (residuals vs fitted).

```r
par(mfrow = c(2,2))
plot(twoslopes, pch=16)
```