Regression Models Course Project

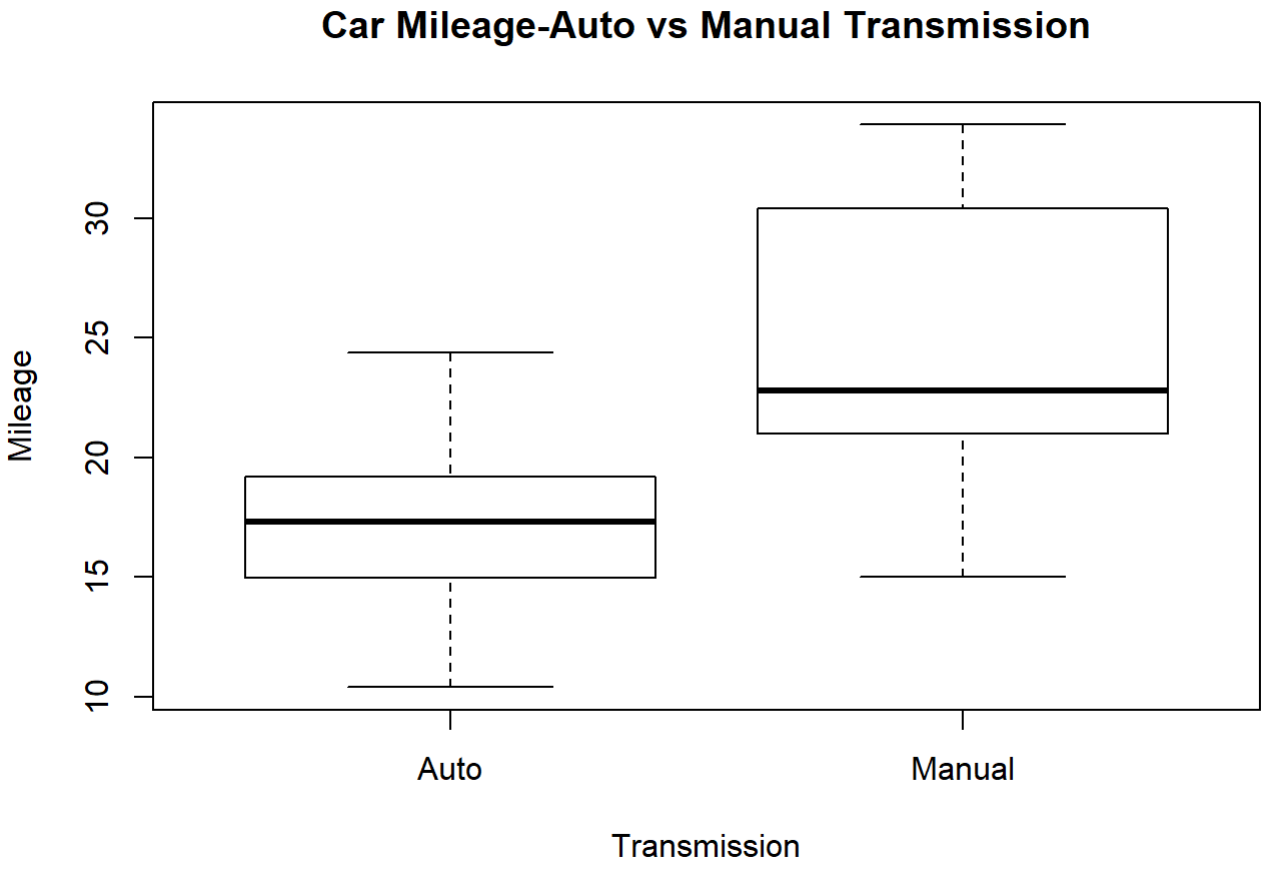
EvaLiang

2020/02/12

Brief summary: According to subsequent analysis, we found that the mpg is largely determined by the interplay between weight, acceleration and transmission. This could give us some instincts towards whether an automatic or manual transmission better for MPG.

1. Is an automatic or manual transmission better for MPG

```
plot(mtcars1$Transmission, mtcars1$mpg, main = "Car Mileage-Auto vs Manual Transmission", x
lab ="Transmission", ylab = "Mileage")
```



A boxplot

can be seen above, a car with Manual transmission has better mileage than a Auto transmission.

2. Quantify the MPG difference between automatic and manual transmissions We fit the effect of only the transmission on miles per gallon(mpg) to quantify the MPG difference between automatic and manual transmissions. For a two sided hypothesis with p value = 0.05, the null hypothesis is defined that the transmission has no impact on miles per gallon(mpg).

```
fitlm1 <- lm(mpg~am, mtcars)
summary(fitlm1)
```

```
##
## 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
```

From the summary of above, we see that both transmission types are positively related to mileage. Also the manual transmission has a edge over auto, i.e. a higher gas mileage of 7.2449393 compared to auto transmission.

The p value of both transmission types < 0.05, we can reject the null hypothesis and state that both transmission types have impact on the mileage.

To construct a model with good fit, we use stepwise selection.

```
stepmodel = step(lm(formula = mpg ~ wt + qsec + am , data = mtcars),trace=0,steps=10000)
summary(stepmodel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

In this case, we include three variables, i.e. wt, qsec and am, and the interaction between wt and am. which take up 85% of total variance.
Next, we can examine mpg ~ wt + qsec correlation with am.

```
model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)
```

```
##
## Call:
```

```
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.9692     5.7756   2.419  0.02259 *
## factor(am)0:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)1:wt    -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)0:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)1:qsec   1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

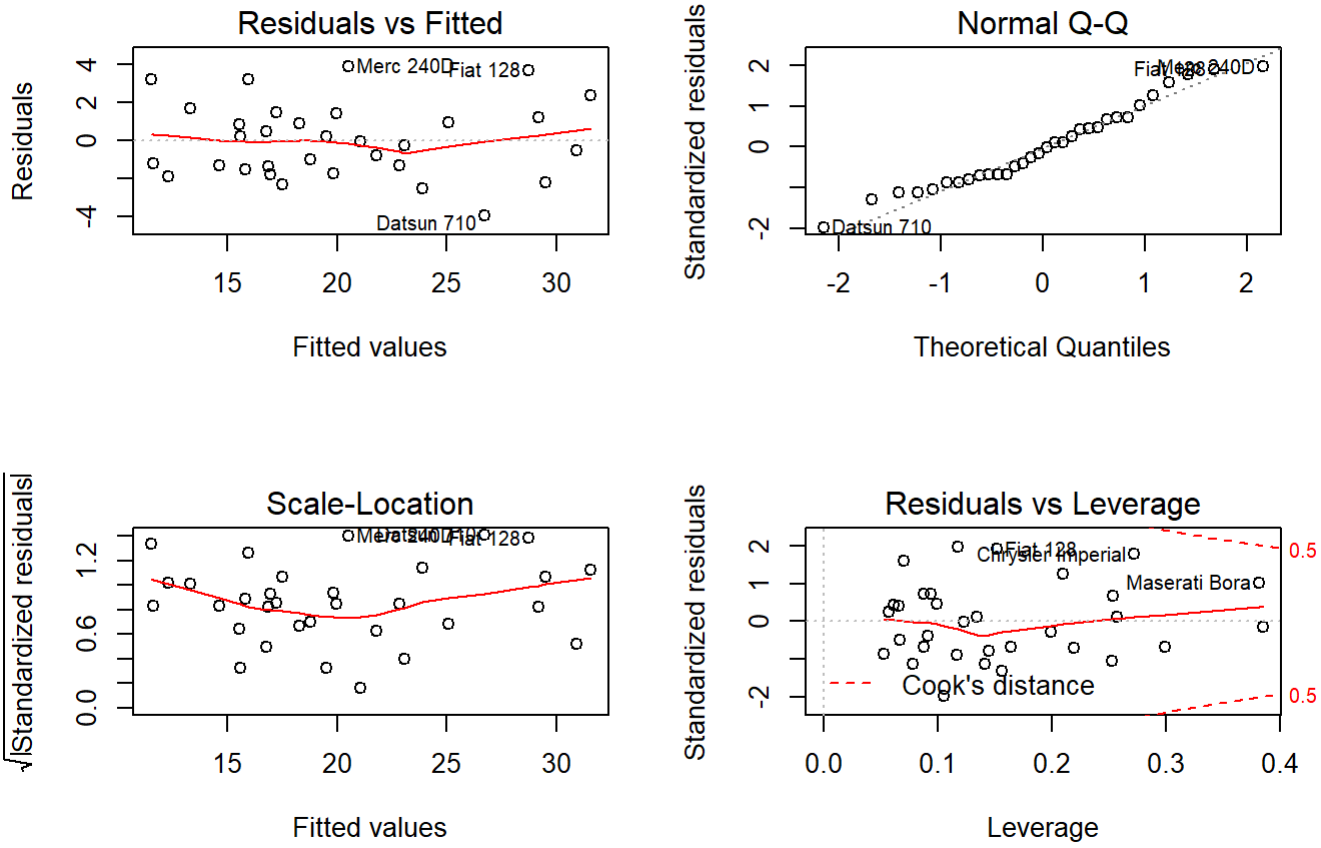
3. Conclusion In total, this model has a 89.5% total variance with an adjusted variance of 0.879. For coefficients interpretation: A car with Manual transmission has higher gas mileage of 7.2449393 compared to auto transmission.

The mpg is mainly determined by the interplay between weight, acceleration and transmission. When the weight increased by 1000 lbs, the mpg decreased by -3.1759 for automatic transmission cars, and -6.0992 for manual transmission cars so with increasing car weight we should choose manual transmission cars. When the acceleration speed dropped, and 1/4 mile time increased (by 1 sec), the mpg factor increased by 0.8338 miles for automatic transmission cars, and 1.446 miles for manual transmission cars so with lower acceleration speed, but same weight, manual transmission cars are better for mpg.

4. Appendix Appendix 1. Residual check and diagnostics plot

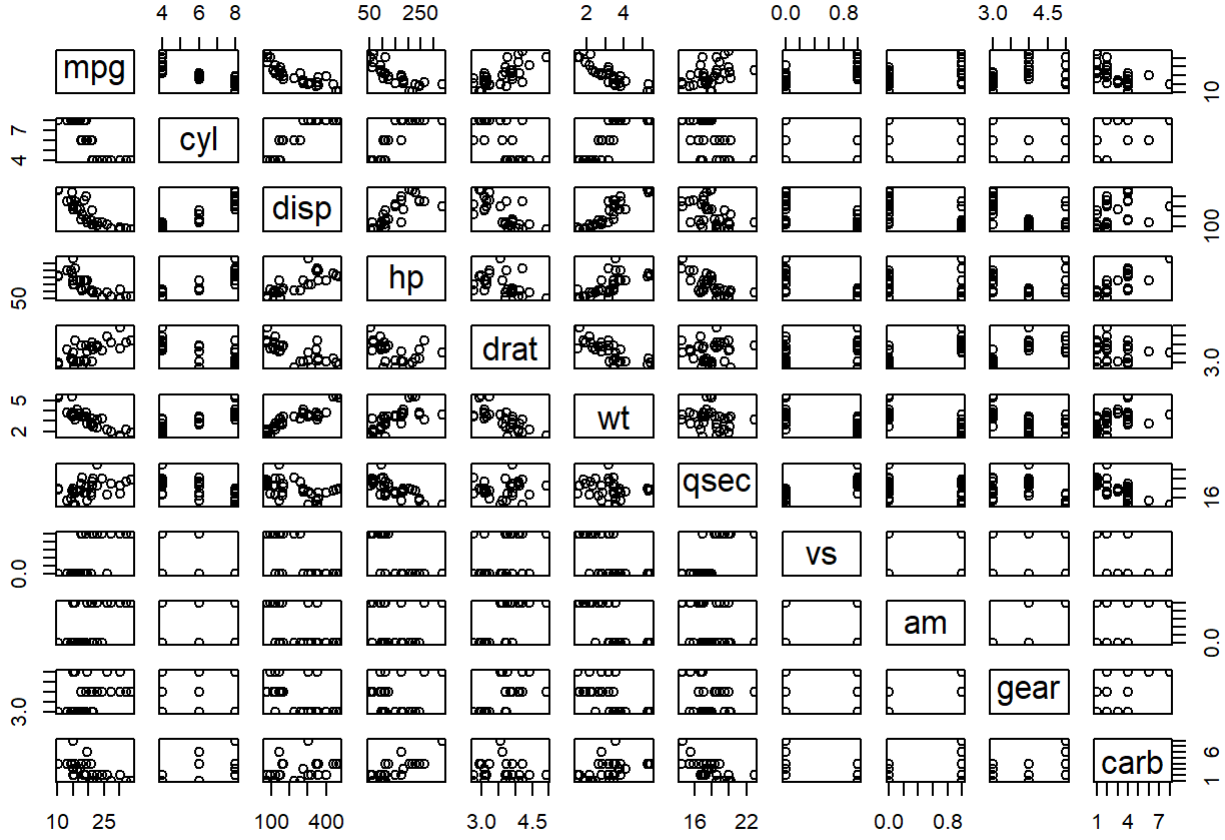
```
par(mfrow=c(2,2))
plot(model)
```

Appendix



2. Scatter plot

```
pairs(mpg ~ ., data = mtcars)
```



Appendix

3. Density and histogram

```
par(mfrow=c(2,1))
hist(mtcars$mpg, breaks=10, xlab="MPG", main="MPG histogram")
plot(density(mtcars$mpg), main="kernel density", xlab="MPG")
```