Course project, Regression Models

explore the mtcars dataset

First here are the brief explanations for all variables.

```
?mtcars
mpg     Miles/(US) gallon
cyl     Number of cylinders
disp    Displacement (cu.in.)
hp    Gross horsepower
drat    Rear axle ratio
wt    Weight (lb/1000)
qsec    1/4 mile time
vs    V/S
am    Transmission (0 = automatic, 1 = manual)
gear    Number of forward gears
carb    Number of carburetors
```

```
data(mtcars)
# convert categorical variables into factor
mtcars$am = factor(mtcars$am)
mtcars$cyl = factor(mtcars$cyl)
mtcars$vs = factor(mtcars$vs)
# summary(mtcars)
# variable selection
summary(lm(mpg ~ . , data = mtcars))$coefficient
```

```
##                Estimate  Std. Error     t value   Pr(>|t|)
## (Intercept) 17.81984325 16.30602324  1.09283809 0.28745417
## cyl6        -1.66030673  2.26229662 -0.73390320 0.47152449
## cyl8         1.63743980  4.31573451  0.37941162 0.70838075
## disp         0.01391241  0.01740176  0.79948296 0.43340363
## hp          -0.04612835  0.02712018 -1.70088685 0.10446190
## drat         0.02635025  1.67648954  0.01571752 0.98761549
## wt          -3.80624757  1.84664309 -2.06117121 0.05252853
## qsec         0.64695710  0.72195025  0.89612421 0.38084614
## vs1          1.74738689  2.27267212  0.76886889 0.45095593
## am1          2.61726546  2.00474936  1.30553251 0.20653091
## gear         0.76402917  1.45668015  0.52450029 0.60569589
## carb         0.50935118  0.94244181  0.54045902 0.59484874
```

```
fit <- lm(mpg ~ . , data = mtcars)
# calculate the variance inflation factor (VIF)
library(car)
sqrt(vif(fit))
```

```
##          GVIF       Df GVIF^(1/(2*Df))
## cyl  6.028697 1.414214        1.566953
## disp 4.650961 1.000000        2.156609
```

```
## hp   4.009813 1.000000        2.002452
## drat 1.933020 1.000000        1.390331
## wt   3.896435 1.000000        1.973939
## qsec 2.782022 1.000000        1.667940
## vs   2.470153 1.000000        1.571672
## am   2.157224 1.000000        1.468749
## gear 2.317650 1.000000        1.522383
## carb 3.282641 1.000000        1.811806
```

It seems that the weight is the best predictor for mpg. Therefore, I will include only weight and am as the covariates in my following analysis.

model fitting and selection

```
# to select the best model, I'll fit 3 models and compare them
fit1 = lm(mpg ~ am, data = mtcars)
summary(fit1)$coefficient
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
fit2 = lm(mpg ~ wt + am, data = mtcars)
summary(fit2)$coefficient
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## am1         -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

```
fit3 = lm(mpg ~ wt * am, data = mtcars)
summary(fit3)$coefficient
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 31.416055  3.0201093 10.402291 4.001043e-11
## wt          -3.785908  0.7856478 -4.818836 4.551182e-05
## am1         14.878423  4.2640422  3.489277 1.621034e-03
## wt:am1      -5.298360  1.4446993 -3.667449 1.017148e-03
```

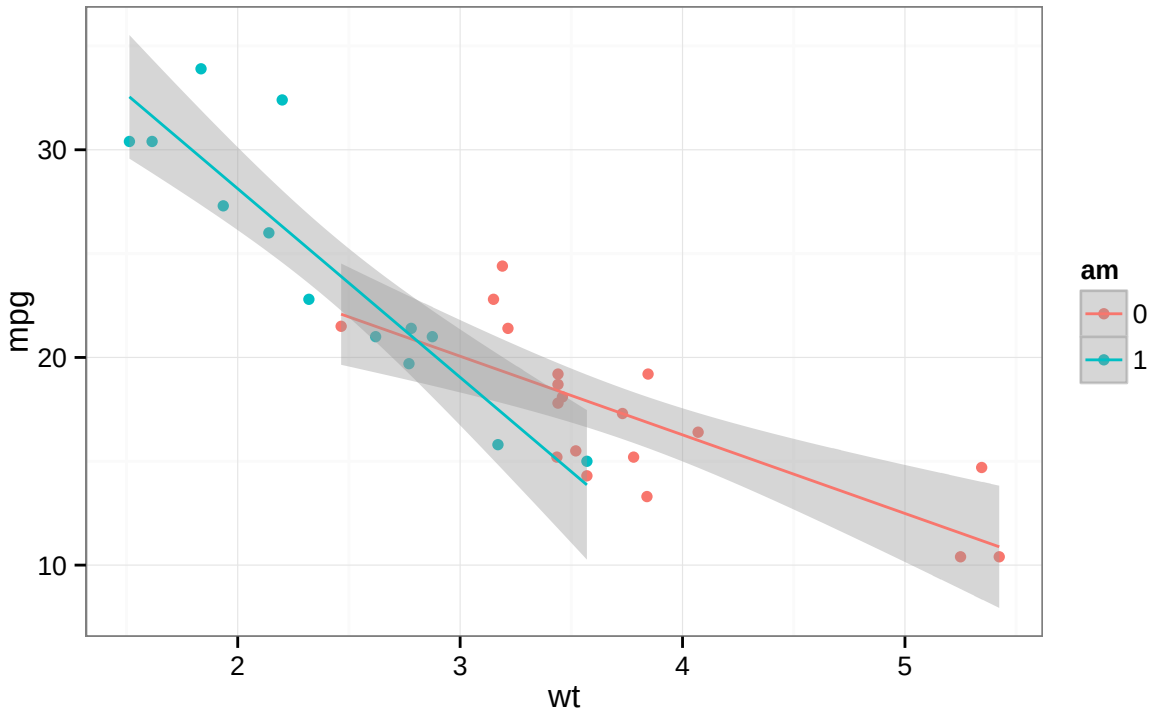
```
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + am
## Model 3: mpg ~ wt * am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
```

```
## 2     29 278.32  1    442.58 65.913 7.717e-09 ***
## 3     28 188.01  1     90.31 13.450  0.001017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From model fit1, I found that manual transmission has positive effect on mpg comparing with automatic transmission. The coefficient 7.2 means the manual transmission has 7.2 mpg increase than automatic transmission on average.
From model fit3, I also found that the interaction between weight and transmission is significant. The intercept 31.4 is the average mpg of cars from the reference group am 0 (automatic transmission) at weight 0. -3.8 is the change of mpg for each 1000 pounds from the reference group am 0. 14.9 is the increase of mpg comparing am1 cars with am0 cars at weight 0.

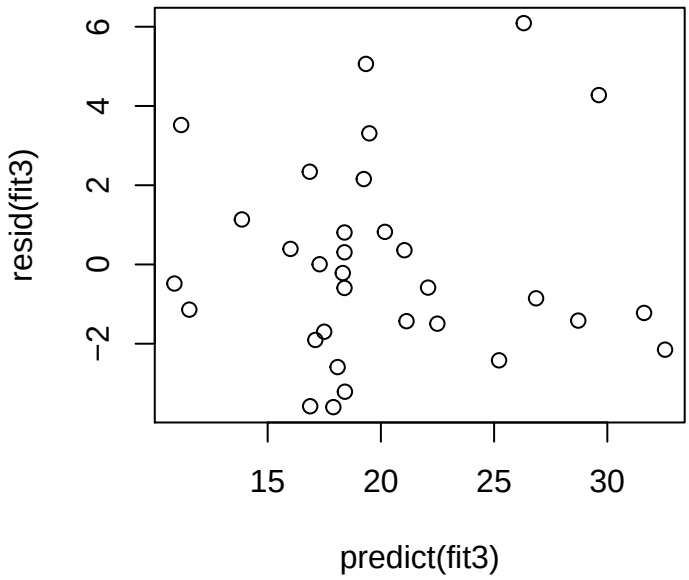
```
# to show that there is interaction, I plotted the regression within each am group
library(ggplot2)
theme_set(theme_bw())
qplot(wt, mpg, data = mtcars, geom="point", colour = am) + geom_smooth(method="lm")
```



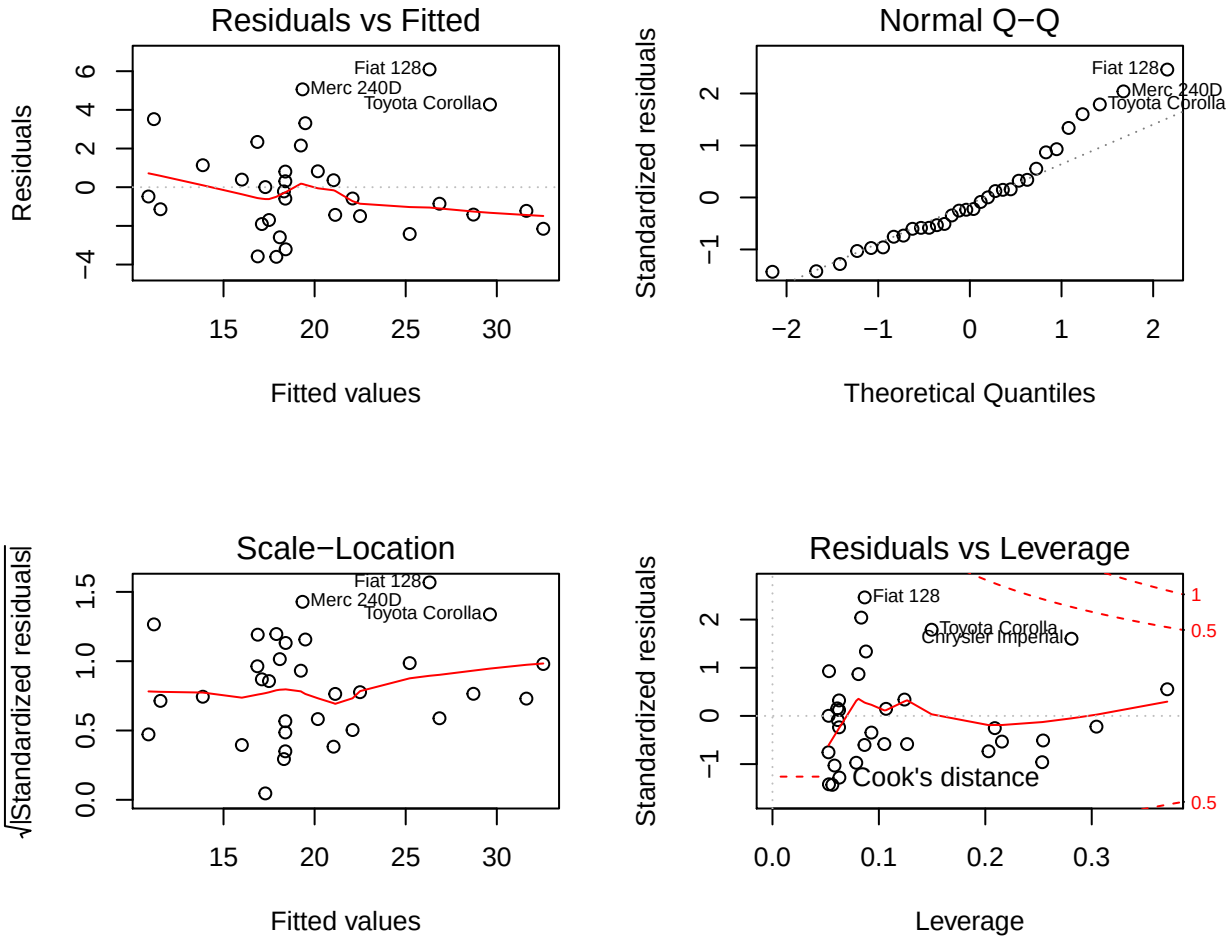
I can also observe the intersection of regressin line, suggesting the interaction between weight and transmission. The effect of weight on the mpg is dependant on the transmission type of cars. That is why the slopes of two regression lines are different.

residues and diagnostics

```
plot(predict(fit3), resid(fit3))
```




```
par(mfrow = c(2, 2))
plot(fit3)
```



```r
# round(dfbetas(fit3)[, 2], 3)
round(hatvalues(fit3), 3)
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##               0.086               0.124               0.079
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##               0.081               0.063               0.061
##          Duster 360           Merc 240D            Merc 230
##               0.056               0.083               0.088
##            Merc 280           Merc 280C          Merc 450SE
##               0.063               0.063               0.061
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##               0.053               0.053               0.254
## Lincoln Continental   Chrysler Imperial            Fiat 128
##               0.304               0.281               0.087
##         Honda Civic      Toyota Corolla       Toyota Corona
##               0.216               0.150               0.209
##    Dodge Challenger         AMC Javelin          Camaro Z28
##               0.058               0.063               0.053
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##               0.053               0.127               0.093
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##               0.253               0.203               0.105
##       Maserati Bora          Volvo 142E
##               0.371               0.107
```

There is no abnormal pattern found in the residues plot.

executive summary

1. There is significant evidence suggesting that an manual transmission is better for MPG than an automatic transmission.
2. The MPG difference between automatic and manual transmissions is 7.2 mpg disregarding the weight, while the difference is 14.9 mpg regarding the weight.