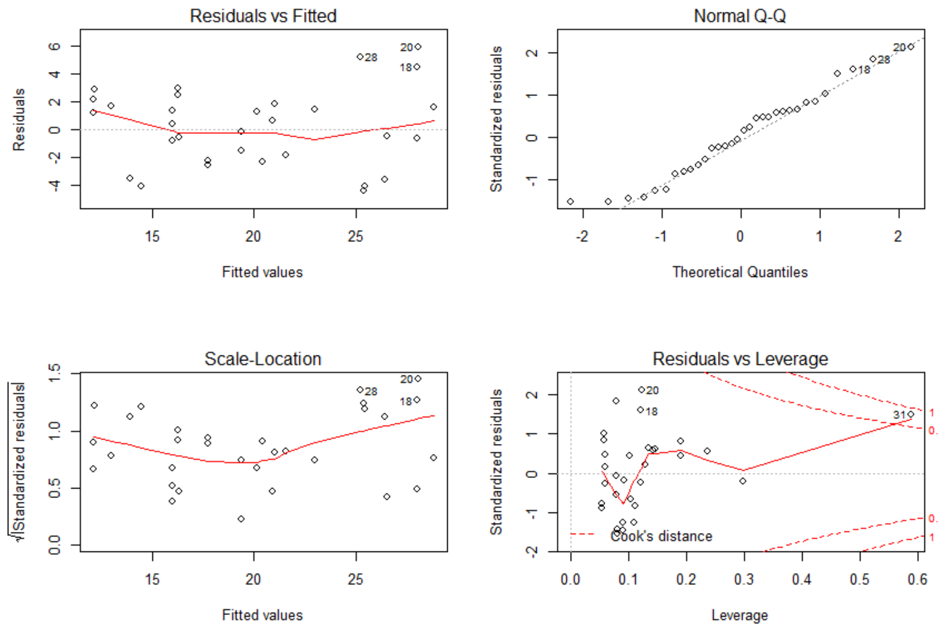
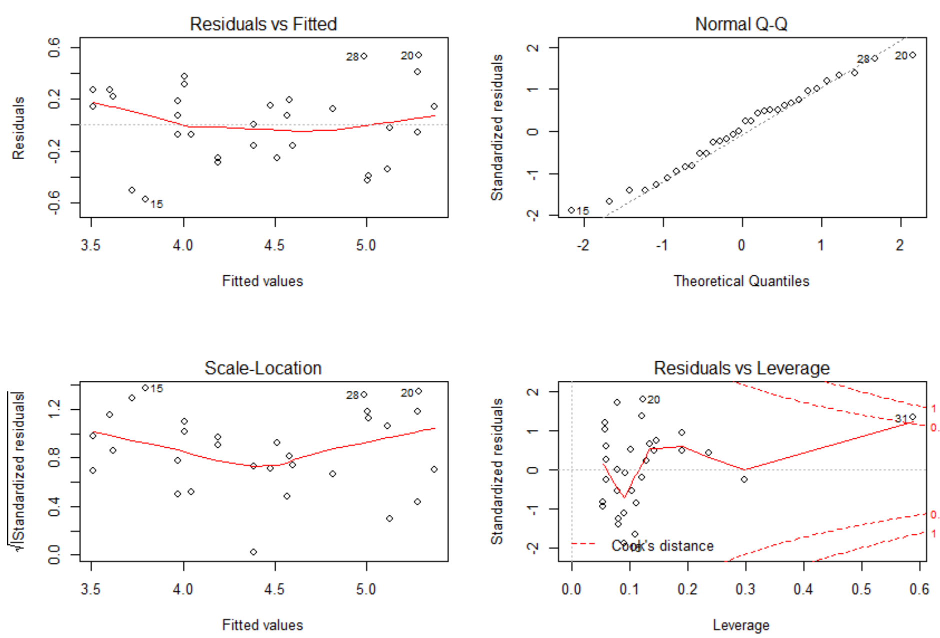
1.



From the graph, we can see that the Q-Q plot shows the normality, but the hat the residuals do not show a constant variance Thus, assumptions are not all satisfied. Then, we want to use transformation to make the model better



From the graph, we can see that the residuals show a constant variance and the Q-Q plot shows the normality. Thus, assumptions are now all satisfied. We can use this model for further study.

```
> anova(cars.ancova)
Analysis of Variance Table

Response: sqrt(mpg)
          Df Sum Sq Mean Sq F value    Pr(>F)    
am         1 4.7674  4.7674 47.6577 1.671e-07 ***
hp         1 5.9555  5.9555 59.5349 2.096e-08 ***
am:hp      1 0.0264  0.0264  0.2641    0.6113    
Residuals 28 2.8009  0.1000                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

<u>For Interaction term</u>:

H_0: $(\alpha\beta)_i = 0$

Test Statistic = $F_{1,28} = 0.02641$

p-value = 0.6113 > α = 0.05

Since the p-value of Horsepower is larger than α = 0.05, we fail to reject null hypothesis and conclude that the combination of Auto Transmission and Horsepower has no significant impact on mean gas mileage.

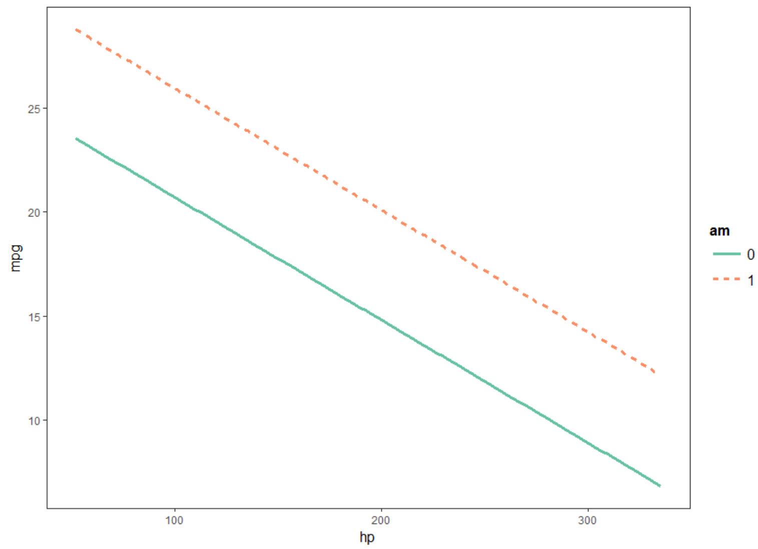
<u>For Automatic Transmission</u>:

H_0: automatic transmission has no significant impact on mean gas mileage

Test Statistic = $F_{1,28} = 47.6577$

p-value = $1.671 \times 10^{-7} < \alpha = 0.05$

Since the p-value of Automatic Transmission is smaller than α = 0.05, we reject null hypothesis and conclude that Automatic Transmission has significant impact on mean gas mileage.



By observing the graph above, we can see that two lines (with/without automatic transmission) have negative slope, which indicates that as horsepower increases, mpg will decrease. Also, the graph shows that cars with automatic transmission (represented by line 1) has higher mpg than cars without automatic transmission (represented by line 0). No interaction effects are shown.

```
> cld(cars.am,alpha=0.05)
 am  lsmean         SE df lower.CL upper.CL .group
 0  4.21256 0.07494891 28 4.059034 4.366085  1
 1  4.77747 0.09032865 28 4.592440 4.962499   2

Results are given on the sqrt (not the response) scale.
Confidence level used: 0.95
significance level used: alpha = 0.05
```

By observing the chart above, we can see that cars that have automatic transmission have more impacts on mean gas mileage than cars that do not have automatic transmission have.

For Horsepower:

H_0: $\beta = 0$

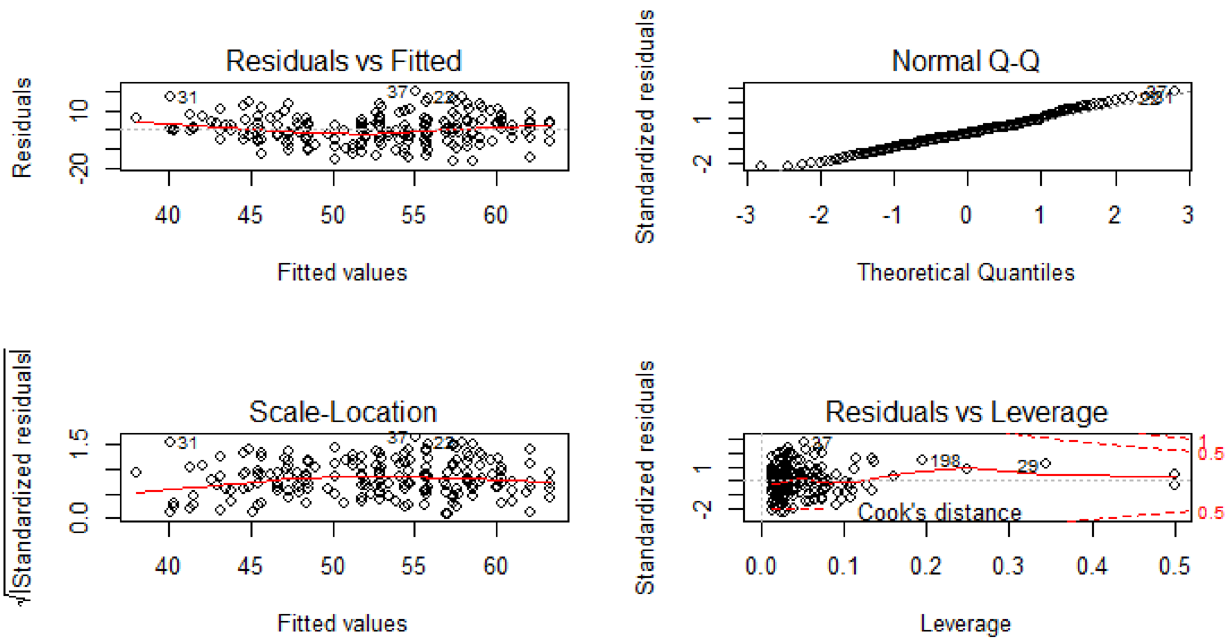
Test Statistic = $F_{1,28} = 59.5349$

p-value = $2.096 \times 10^{-8} < \alpha = 0.05$

Since the p-value of Horsepower is smaller than $\alpha = 0.05$, we reject null hypothesis and conclude that Horsepower has significant impact on mean gas mileage.

In conclusion, the automatic transmission and horsepower of a car can affect its gas mileage, but the combination of these two factors does not have any impact on the gas mileage.

2.



From the graph, we can see that the residuals show a constant variance and the Q-Q plot shows the normality. Thus, assumptions are satisfied. We can use this model for further study and no transformation is needed at this time.

```
> anova(hsb2.ancova)
Analysis of Variance Table

Response: math
              Df  Sum Sq Mean Sq  F value    Pr(>F)    
ses            2  1307.1   653.5  12.1203 1.112e-05 ***
schtyp         1    73.6    73.6   1.3646    0.2442    
write          1  5663.3  5663.3 105.0280 < 2.2e-16 ***
schtyp:write   1     0.8     0.8   0.0158    0.9002    
ses:write      2     1.4     0.7   0.0128    0.9873    
ses:schtyp     2   174.5    87.2   1.6180    0.2010    
Residuals    190 10245.1    53.9                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

<u>For Interaction term between School Type and Write</u>:

H_0: combination of School Type and Write has no significant impact on mean math grade ($(\alpha\gamma) = 0$)

Test Statistic = $F_{1,190} = 0.0158$

p-value = 0.9002

Since the p-value of combination of School Type and Write is larger than $\alpha = 0.05$, we fail to reject null hypothesis and conclude that the combination of School Type and Write has no significant impact on mean math grade.

<u>For Interaction term between Socio-economic Status and Write</u>:

H_0: combination of Socio-economic Status and Write has no significant impact on mean math grade

Test Statistic = $F_{2,190} = 0.0128$

p-value = 0.9873

Since the p-value of combination of Socio-economic Status and Write is larger than $\alpha = 0.05$, we fail to reject null hypothesis and conclude that the combination of Socio-economic Status and Write has no significant impact on mean math grade.

<u>For Interaction term between Socio-economic Status and School Type</u>:

H_0: combination of Socio-economic Status and School Type has no significant impact on mean math grade

Test Statistic = $F_{2,190} = 1.6180$

p-value = 0.2010

Since the p-value of combination of Socio-economic Status and School Type is larger than $\alpha = 0.05$, we fail to reject null hypothesis and conclude that the combination of Socio-economic Status and School Type has no significant impact on mean math grade.

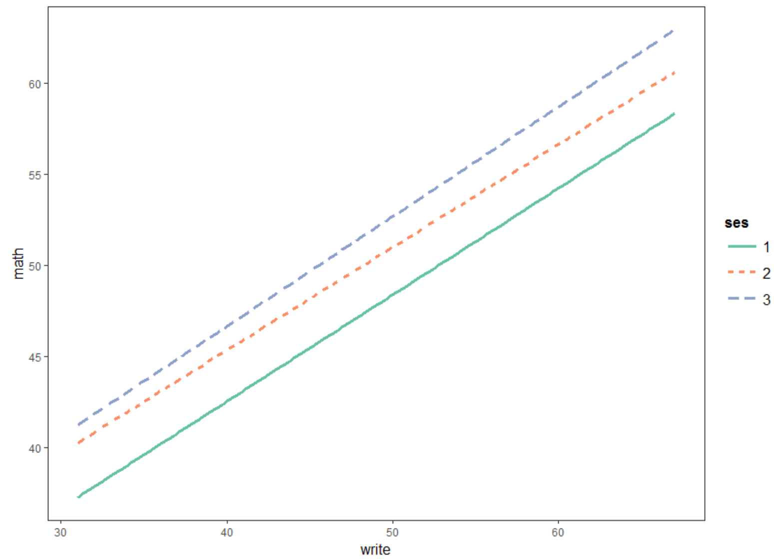
<u>For Socio-economic Status</u>:

H_0: Socio-economic Status has no significant impact on mean math grade

Test Statistic = $F_{2,193} = 12.1203$

p-value = 1.112×10^{-5}

Since the p-value of Socio-economic Status is smaller than $\alpha = 0.05$, we reject null hypothesis and conclude that Socio-economic Status has significant impacts on mean math grade.



By observing the graph above, we can see that three lines (with socio-economic status 1, 2, 3) have positive slope, which indicates that as writing score increases, math score will increase as well. Also, at high writing score, the graph shows that students with socioeconomic status 3 (represented by line 3) has higher math score than students with socioeconomic status 2 (represented by line 2); students with socioeconomic status 2 (represented by line 2) has higher math score than students with socioeconomic status 1 (represented by line 1). In addition, at low writing score, students with socioeconomic status 3 and 2 tend to have similar math score while students with socioeconomic status 1 tend to have lower math score. No interaction effects between write and socio-economic status are shown.

```
> cld(hsb2.ses,alpha=0.05)
 ses   lsmean        SE  df lower.CL upper.CL .group
   1 48.88601 2.6568070 190 43.64538 54.12663  1
   3 53.28281 1.3179080 190 50.68320 55.88243  1
   2 53.29589 0.9634366 190 51.39548 55.19629  1

Results are averaged over the levels of: schtyp
Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05
```

By observing the chart above, we can see that <u>all Socio-economic Status (1, 2, 3) have the same impacts on the mean math grade.</u>

<u>For School Type</u>:

H_0: school type has no significant impact on mean math grade

Test Statistic = $F_{1,193}$ = 1.3646

p-value = 0.2442

Since the p-value of School Type is larger than $\alpha = 0.05$, we fail to reject null hypothesis and conclude that the School Type has no significant impact on mean math grade.

<u>**For Write**</u>:

H_0: Write has no significant impact on mean math grade ($\gamma = 0$)

Test Statistic = $F_{1,193} = 105.0280$

p-value $< 2.2 \times 10^{-16}$

Since the p-value of Write is smaller than $\alpha = 0.05$, we reject null hypothesis and conclude that Write has significant impacts on mean math grade.

In conclusion, the socio-economic status and writing sores can affect the SAT math grade of a student, but the combination between socio-economic status and writing sores does not have any impacts on the SAT math grade.

R code:

```
#install.packages("lmerTest")
#install.packages("lsmeans")
#install.packages("car")
#install.packages("multcompView")
#install.packages("lme4")
#install.packages("jtools")

library(lsmeans)
library(car)
library(multcompView)
library(lme4)
library(lmerTest)
library(jtools)
options(contrasts = c("contr.sum", "contr.poly"))

#1---------------------------------------------------
cars <- mtcars[,c("am","mpg","hp")]
head(cars)

am<-as.factor(cars$am)
mpg<-cars$mpg
hp<-cars$hp
df<-data.frame(am=am,mpg=mpg,hp=hp)

cars.ancova=aov(mpg~am+hp+hp:am,data=df)
par(mfrow=c(2,2))
plot(cars.ancova)

cars.ancova=aov(sqrt(mpg)~am+hp+hp:am,data=df)
par(mfrow=c(2,2))
plot(cars.ancova)

anova(cars.ancova)

interact_plot(cars.ancova,pred="hp",modx="am")

cars.am=lsmeans(cars.ancova,~as.factor(am))
cld(cars.am,alpha=0.05)

#2---------------------------------------------------
hsb2= read.table("hsb2.csv")
head(hsb2)

ses<-as.factor(hsb2$ses)
schtyp<-as.factor(hsb2$schtyp)
write<-hsb2$write
```

```
math<-hsb2$math
df2=data.frame(ses=ses,schtyp=schtyp,write=write,math=math)

hsb2.ancova=aov(math~ses+schtyp+write+schtyp:write+ses:write+ses:schtyp,data=df2)
par(mfrow=c(2,2))
plot(hsb2.ancova)

shapiro.test(hsb2.ancova$residuals)

anova(hsb2.ancova)

interact_plot(hsb2.ancova,pred="write",modx="ses")

hsb2.ses=lsmeans(hsb2.ancova,~as.factor(ses))
cld(hsb2.ses,alpha=0.05)
```