Homework 11

Xiangyu Ren

1. Consider an experiment to determine whether or not having an automatic transmission affects gas mileage in cars. It is also known that the horsepower of a car affects gas mileage, so the horsepower of each car used is also recorded. The following R code will read in the data to R

```
library(lsmeans)
library(car)
library(multcompView)
library(lme4)
library(lmerTest)
library(jtools)
library(interactions)
```

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

```
##                    am  mpg  hp
## Mazda RX4           1 21.0 110
## Mazda RX4 Wag       1 21.0 110
## Datsun 710          1 22.8  93
## Hornet 4 Drive      0 21.4 110
## Hornet Sportabout   0 18.7 175
## Valiant             0 18.1 105
```

```
## type "cars" into R to see the full data
cars$am = as.factor(cars$am)
cars
```

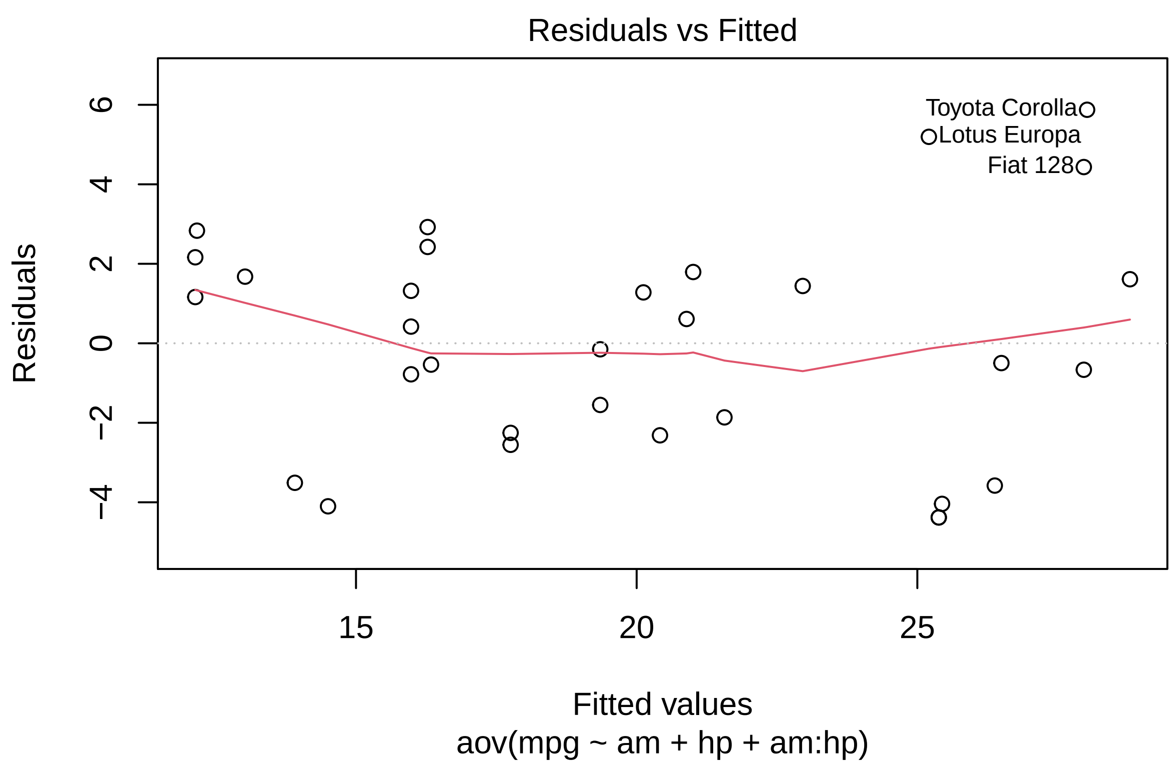
```
##                     am  mpg  hp
## Mazda RX4            1 21.0 110
## Mazda RX4 Wag        1 21.0 110
## Datsun 710           1 22.8  93
## Hornet 4 Drive       0 21.4 110
## Hornet Sportabout    0 18.7 175
## Valiant              0 18.1 105
## Duster 360           0 14.3 245
## Merc 240D            0 24.4  62
## Merc 230             0 22.8  95
## Merc 280             0 19.2 123
## Merc 280C            0 17.8 123
## Merc 450SE           0 16.4 180
## Merc 450SL           0 17.3 180
## Merc 450SLC          0 15.2 180
## Cadillac Fleetwood   0 10.4 205
```

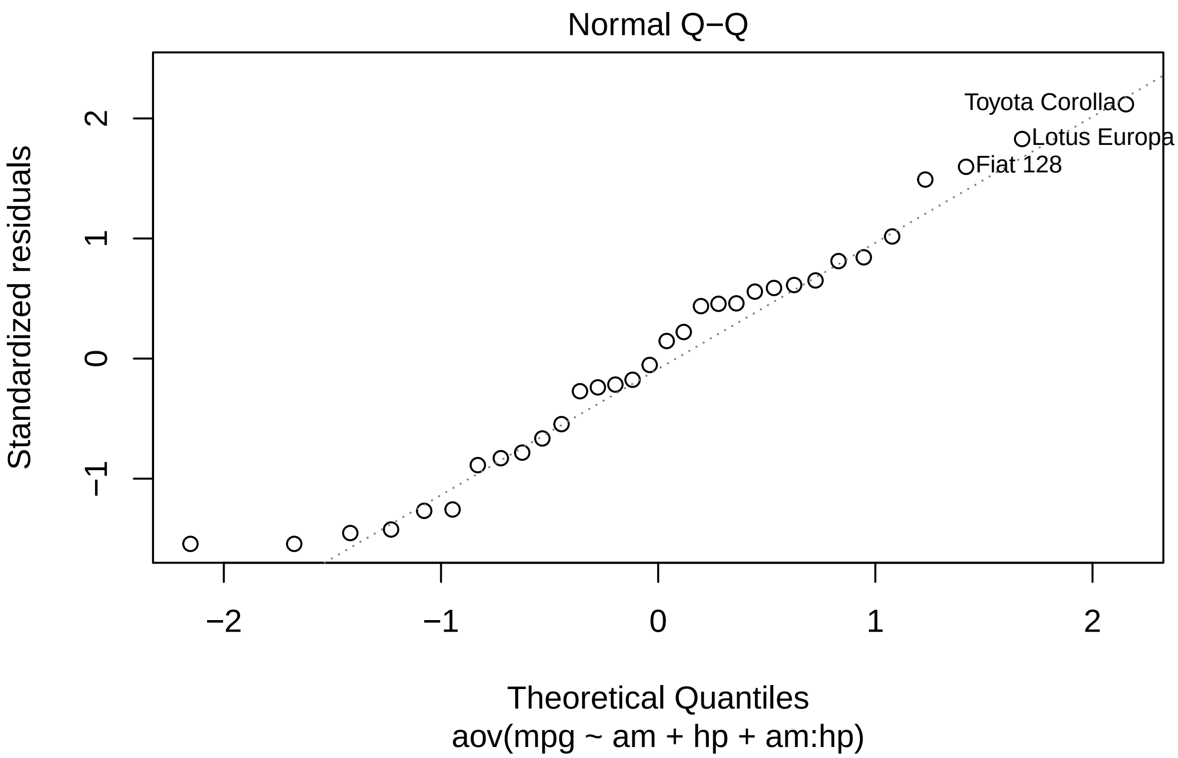
```
## Lincoln Continental  0 10.4 215
## Chrysler Imperial    0 14.7 230
## Fiat 128             1 32.4  66
## Honda Civic          1 30.4  52
## Toyota Corolla       1 33.9  65
## Toyota Corona        0 21.5  97
## Dodge Challenger     0 15.5 150
## AMC Javelin          0 15.2 150
## Camaro Z28           0 13.3 245
## Pontiac Firebird     0 19.2 175
## Fiat X1-9            1 27.3  66
## Porsche 914-2        1 26.0  91
## Lotus Europa         1 30.4 113
## Ford Pantera L       1 15.8 264
## Ferrari Dino         1 19.7 175
## Maserati Bora        1 15.0 335
## Volvo 142E           1 21.4 109
```

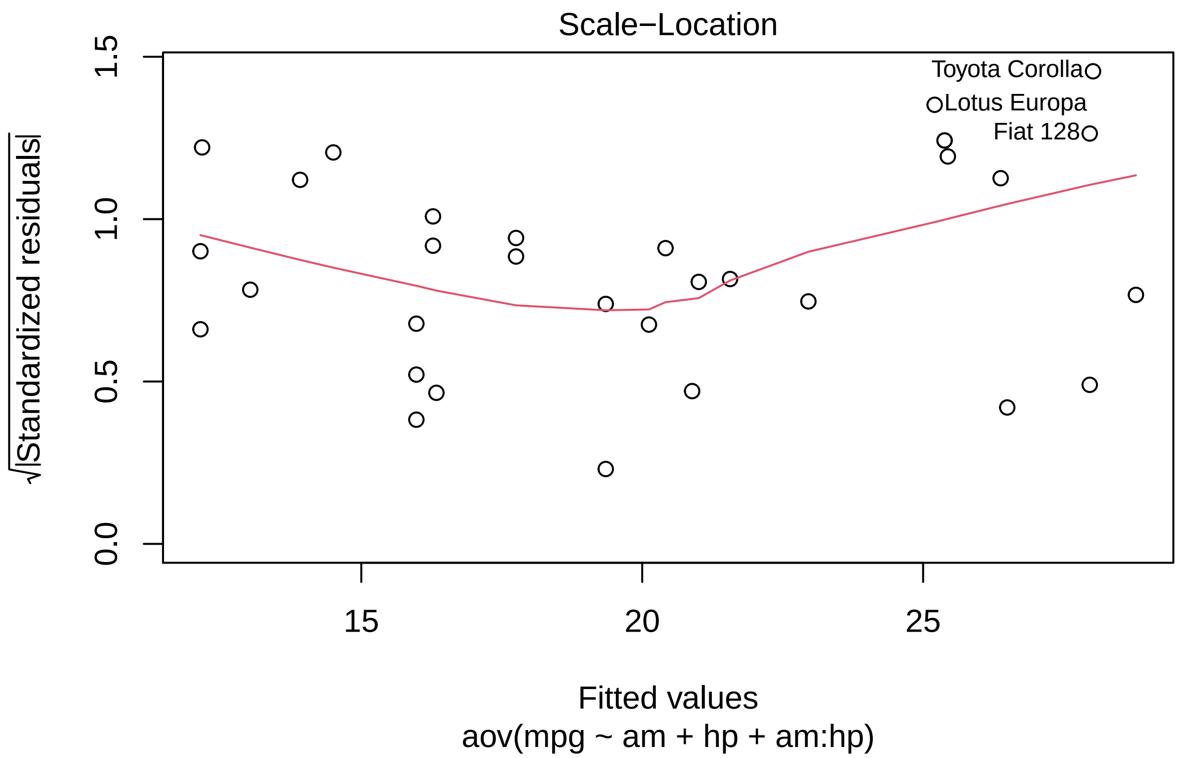
Fit an appropriate model to this data, and interpret the results.

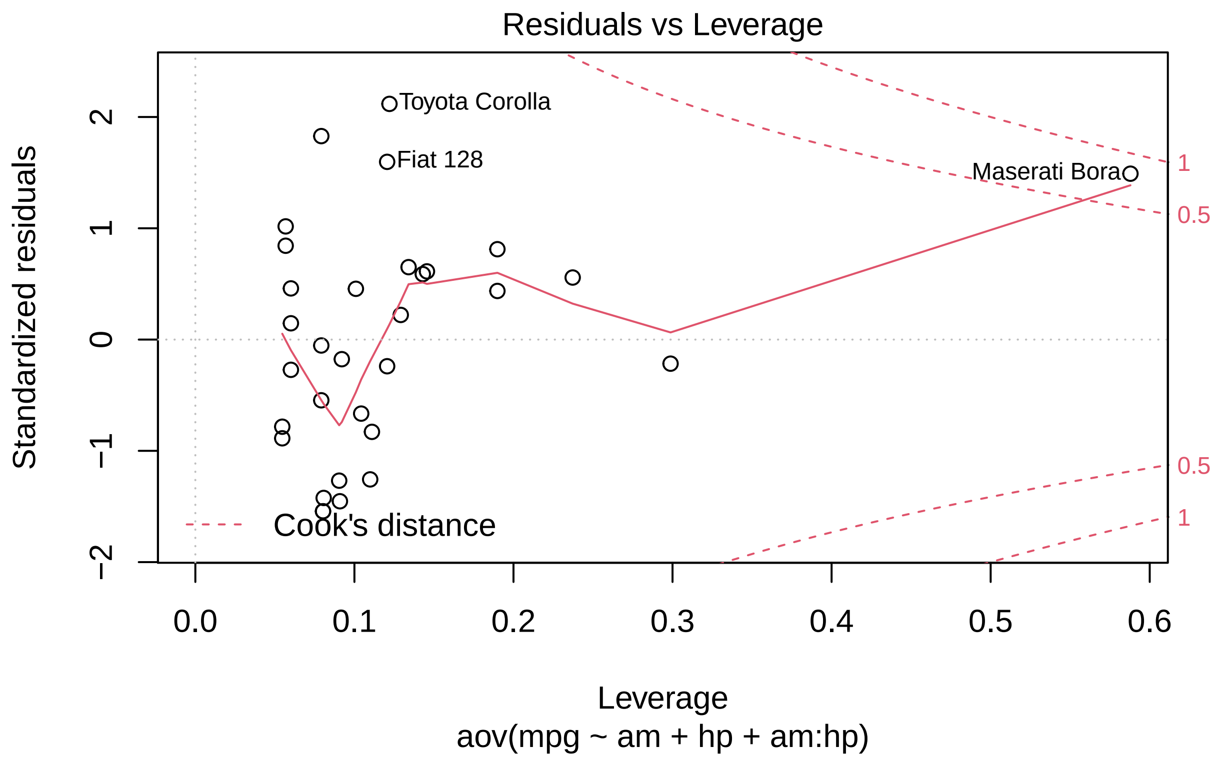
$$Y_{it} = \mu + \alpha_i + \beta * x_j + (\alpha\beta)_i * x_{it} + \epsilon_{ijt}, \quad \epsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

```
fit.ancova = aov(mpg ~ am + hp + am : hp, data = cars)
plot(fit.ancova)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
aov(mpg ~ am + hp + am:hp)
Toyota Corolla
Lotus Europa
Fiat 128
−1
0
1
2
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
Toyota Corolla
Lotus Europa
Fiat 128
Fitted values
aov(mpg ~ am + hp + am:hp)




```r
anova(fit.ancova)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## am         1 405.15  405.15 46.2210 2.197e-07 ***
## hp         1 475.46  475.46 54.2419 5.088e-08 ***
## am:hp      1   0.01    0.01  0.0006    0.9806
## Residuals 28 245.43    8.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypothesis test with $\alpha = 0.05$

'am:hp' line tests with

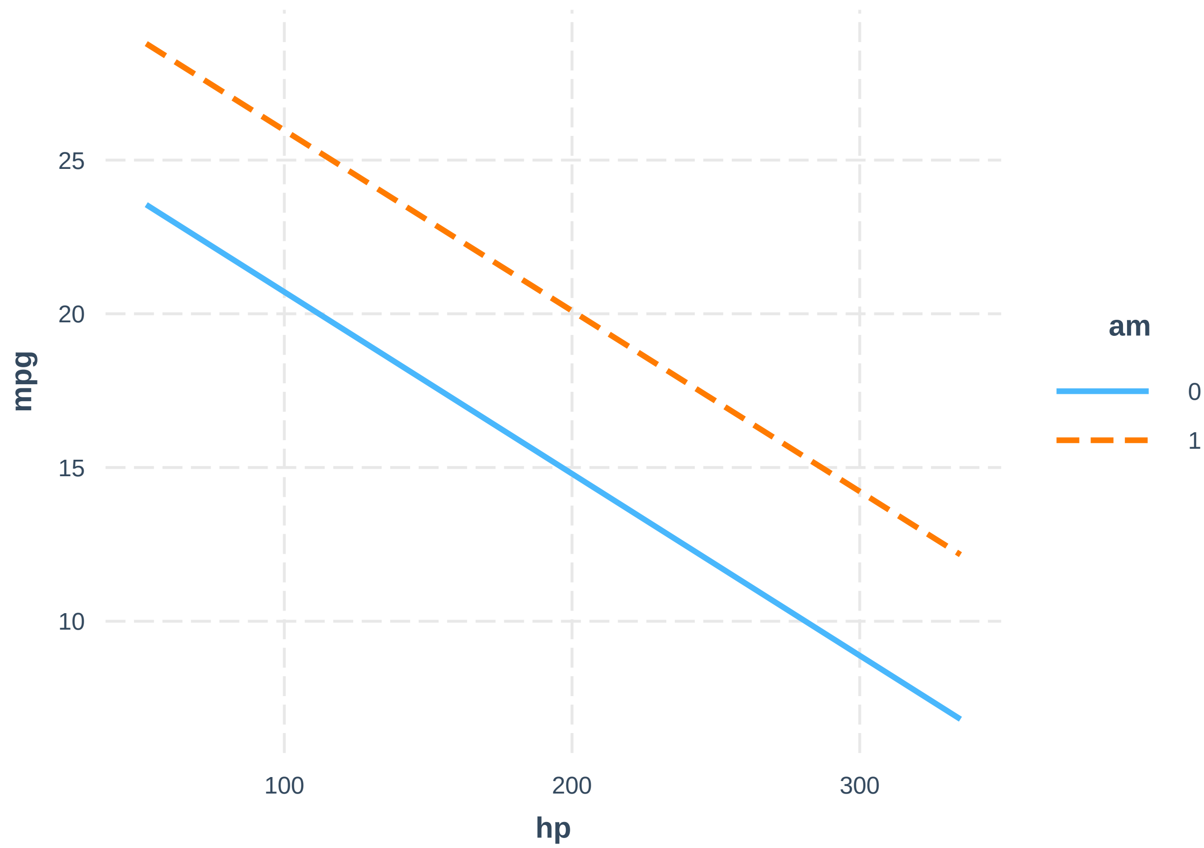
$$H_0: \ (\alpha\beta)_i = 0 \text{ for all } i \ \ vs. \ \ H_a: \ \text{ at least one of the intercept terms } (\alpha\beta)_i \neq 0$$

'hp' line tests with

$$H_0: \ \ \beta = 0 \ \ vs. \ \ H_a: \ \ \beta \neq 0$$

By the table, we can see that the p-value of am:hp is greater than α, so we fail to reject the null, and conclude that the interaction between am and hp is not significant; the p-values of either am or hp are less than α, so these two variables are significant.

```r
interact_plot(fit.ancova, pred ="hp", modx="am")
```




```r
lsmeans(fit.ancova, ~ hp)
```

```
## NOTE: Results may be misleading due to involvement in interactions

##    hp lsmean    SE df lower.CL upper.CL
##   147   20.6 0.549 28     19.5     21.7
##
## Results are averaged over the levels of: am
## Confidence level used: 0.95
```

$\beta = 20.6$ If cars have same hp, then cars with $am = 1$ have higher mpg than cars with $am = 0$. If cars have same am, then mpg decreases as hp decreases.

2. The "high school and beyond" survey looked at differences in socioeconomic status, race, gender, and types of schools and students' SAT scores. The data can be read into R using the following code:

The goal of this study is to see if socio-economic status (ses) and school type (schtyp) affect students math scores, after accounting for how well they do on their writing score. Student's writing score should be treated as a continuous covariate. Conduct a full analysis of this data. Fit an appropriate model, and interpret the results.

```r
hsb2= read.table("hsb2.csv")## from https://stats.idre.ucla.edu/stat/data/hsb2.csv
head(hsb2)
```

```
##   ses schtyp write math
## 1   1      1    52   41
## 2   2      1    59   53
## 3   3      1    33   54
## 4   3      1    44   47
## 5   2      1    52   57
## 6   2      1    52   51
```

```r
## type "hsb2" into R to see the full data
hsb2$ses = as.factor(hsb2$ses); hsb2$schtyp = as.factor(hsb2$schtyp)
hsb2
```

```
##     ses schtyp write math
## 1     1      1    52   41
## 2     2      1    59   53
## 3     3      1    33   54
## 4     3      1    44   47
## 5     2      1    52   57
## 6     2      1    52   51
## 7     2      1    59   42
## 8     2      1    46   45
## 9     2      1    57   54
## 10    2      1    55   52
## 11    2      1    46   51
## 12    2      1    65   51
## 13    3      1    60   71
## 14    3      1    63   57
## 15    1      1    57   50
## 16    1      1    49   43
## 17    3      1    52   51
## 18    2      2    57   60
## 19    3      1    65   62
## 20    2      1    39   57
## 21    2      1    49   35
## 22    2      1    63   75
## 23    2      1    40   45
## 24    3      1    52   57
## 25    2      1    44   45
## 26    2      1    37   46
## 27    3      1    65   66
## 28    2      2    57   57
## 29    3      2    38   49
## 30    1      1    44   49
## 31    2      1    31   57
## 32    3      1    52   64
## 33    3      2    67   63
## 34    2      1    41   57
## 35    3      2    59   50
## 36    3      1    65   58
## 37    2      2    54   75
```

```
## 38      3           1       62     68
## 39      1           1       31     44
## 40      2           1       31     40
## 41      2           2       47     41
## 42      2           2       59     62
## 43      2           1       54     57
## 44      1           1       41     43
## 45      3           1       65     48
## 46      1           1       59     63
## 47      3           1       40     39
## 48      2           1       59     70
## 49      2           2       59     63
## 50      2           1       54     59
## 51      2           1       61     61
## 52      3           1       33     38
## 53      2           1       44     61
## 54      2           2       59     49
## 55      2           1       62     73
## 56      3           1       39     44
## 57      1           1       37     42
## 58      2           1       39     39
## 59      2           2       57     55
## 60      2           1       49     52
## 61      2           2       46     45
## 62      3           1       62     61
## 63      1           1       44     39
## 64      2           1       33     41
## 65      3           2       42     50
## 66      2           1       41     40
## 67      2           1       54     60
## 68      1           1       39     47
## 69      1           1       43     59
## 70      2           1       33     49
## 71      2           1       44     46
## 72      3           1       54     58
## 73      2           1       67     71
## 74      2           1       59     58
## 75      2           1       45     46
## 76      1           1       40     43
## 77      3           1       61     54
## 78      3           1       59     56
## 79      2           1       36     46
## 80      3           1       41     54
## 81      3           1       59     57
## 82      1           1       49     54
## 83      2           2       59     71
## 84      1           1       65     48
## 85      2           1       41     40
## 86      3           1       62     64
## 87      3           1       41     51
## 88      3           1       49     39
## 89      2           1       31     40
## 90      3           1       49     61
## 91      2           1       62     66
```

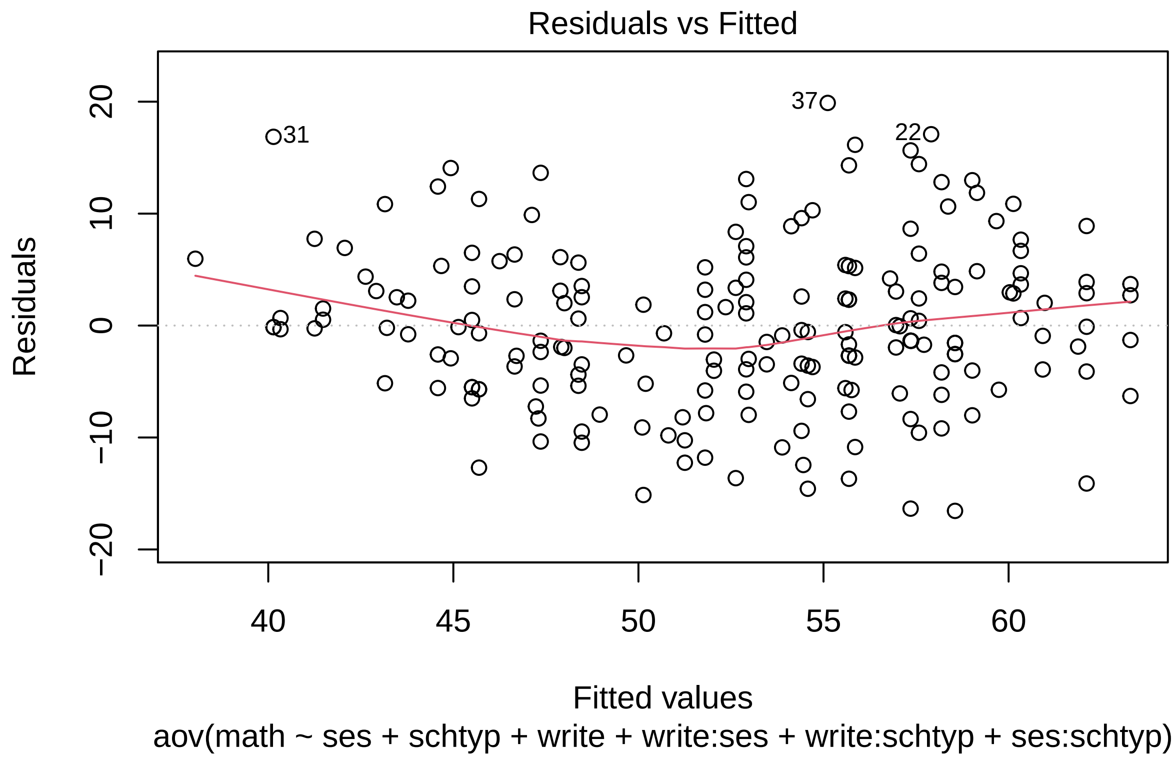
```
## 92     1        1      49    49
## 93     3        1      62    65
## 94     1        1      44    52
## 95     1        1      44    46
## 96     1        1      62    61
## 97     2        1      65    72
## 98     3        1      65    71
## 99     1        1      44    40
## 100    3        2      63    69
## 101    3        1      60    64
## 102    3        1      59    56
## 103    1        1      46    49
## 104    3        1      52    54
## 105    2        1      59    53
## 106    2        1      54    66
## 107    3        1      62    67
## 108    1        1      35    40
## 109    1        2      54    46
## 110    3        2      65    69
## 111    2        1      52    40
## 112    1        1      50    41
## 113    3        1      59    57
## 114    1        1      65    58
## 115    3        2      61    57
## 116    2        1      44    37
## 117    3        1      54    55
## 118    3        1      67    62
## 119    1        1      57    64
## 120    1        1      47    40
## 121    1        2      54    50
## 122    2        1      52    46
## 123    2        1      52    53
## 124    2        1      46    52
## 125    1        1      62    45
## 126    3        1      57    56
## 127    2        1      41    45
## 128    2        1      53    54
## 129    3        1      49    56
## 130    1        1      35    41
## 131    2        1      59    54
## 132    1        1      65    72
## 133    2        1      62    56
## 134    2        1      54    47
## 135    1        1      59    49
## 136    3        1      63    60
## 137    2        2      59    54
## 138    2        1      52    55
## 139    2        1      41    33
## 140    2        2      49    49
## 141    1        1      46    43
## 142    3        1      54    50
## 143    2        1      42    52
## 144    2        1      57    48
## 145    2        1      59    58
```

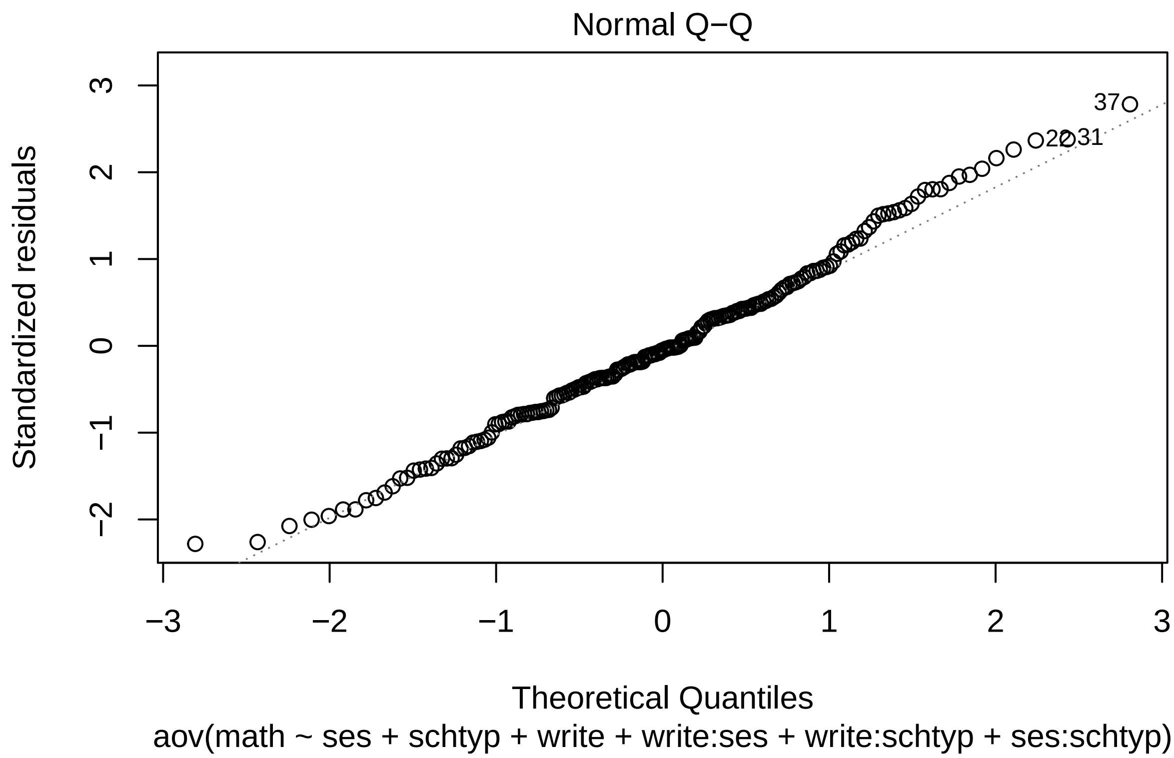
```
## 146   3        2     52    43
## 147   2        1     62    41
## 148   2        2     52    43
## 149   1        1     41    46
## 150   1        1     55    44
## 151   1        1     37    43
## 152   3        1     54    61
## 153   2        1     57    40
## 154   2        1     54    49
## 155   2        1     62    56
## 156   2        1     59    61
## 157   2        1     55    50
## 158   2        1     57    51
## 159   2        1     39    42
## 160   3        1     67    67
## 161   1        1     62    53
## 162   2        1     50    50
## 163   3        2     61    51
## 164   1        1     62    72
## 165   2        1     59    48
## 166   1        1     44    40
## 167   2        1     59    53
## 168   1        1     54    39
## 169   2        2     62    63
## 170   1        1     60    51
## 171   1        1     57    45
## 172   2        1     46    39
## 173   3        1     36    42
## 174   3        1     59    62
## 175   1        1     49    44
## 176   1        1     60    65
## 177   2        1     67    63
## 178   2        1     54    54
## 179   3        1     52    45
## 180   1        1     65    60
## 181   2        1     62    49
## 182   2        2     49    48
## 183   3        1     67    57
## 184   2        1     65    55
## 185   3        1     67    66
## 186   1        1     65    64
## 187   2        1     54    55
## 188   2        1     44    42
## 189   3        2     62    56
## 190   1        1     46    53
## 191   1        1     54    41
## 192   3        2     57    42
## 193   2        2     52    53
## 194   3        1     59    42
## 195   2        2     65    60
## 196   2        2     59    52
## 197   2        1     46    38
## 198   2        2     41    57
## 199   2        1     62    58
```

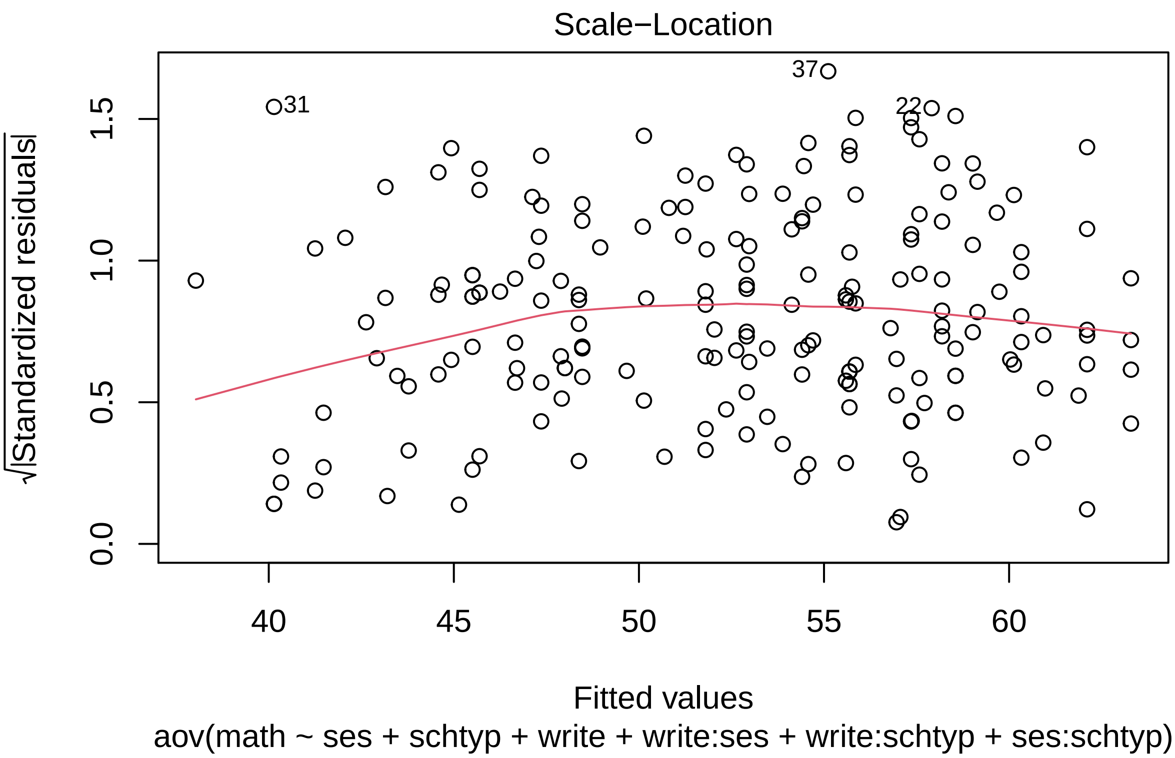
```
## 200    3       1     65    65
```

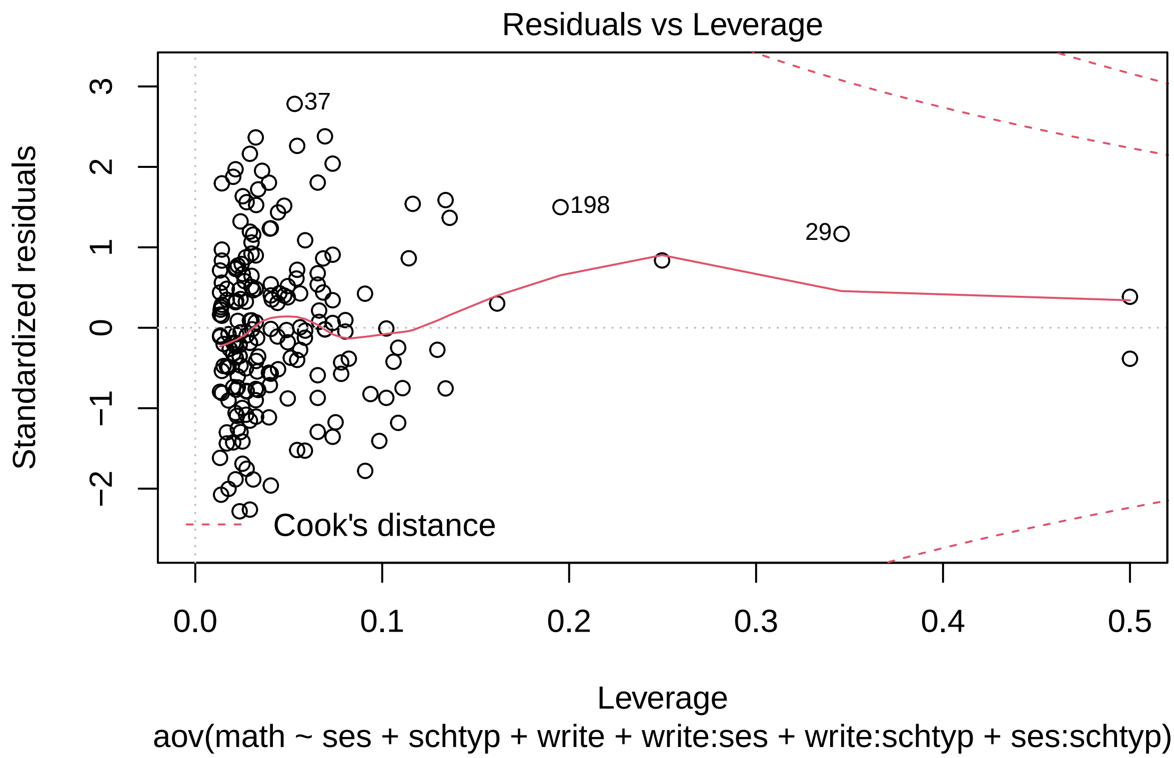
$$Y_{ijkt} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + \beta * x_j + (\alpha\beta)_i * x_{it} + (\gamma\beta)_k * x_{kt} + \epsilon_{ijkt}, \quad \epsilon_{ijt} \overset{iid}{\sim} N(0, \sigma^2)$$

```r
fit.ancova = aov(math ~ ses + schtyp + write + write : ses + write : schtyp + ses : schtyp, data = hsb2)
plot(fit.ancova)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
37
22
31
−3
−2
−1
0
1
2
3
aov(math ~ ses + schtyp + write + write:ses + write:schtyp + ses:schtyp)


Scale−Location
37
31
22
1.5
1.0
0.5
0.0
√|Standardized residuals|
40
45
50
55
60
Fitted values
aov(math ~ ses + schtyp + write + write:ses + write:schtyp + ses:schtyp)



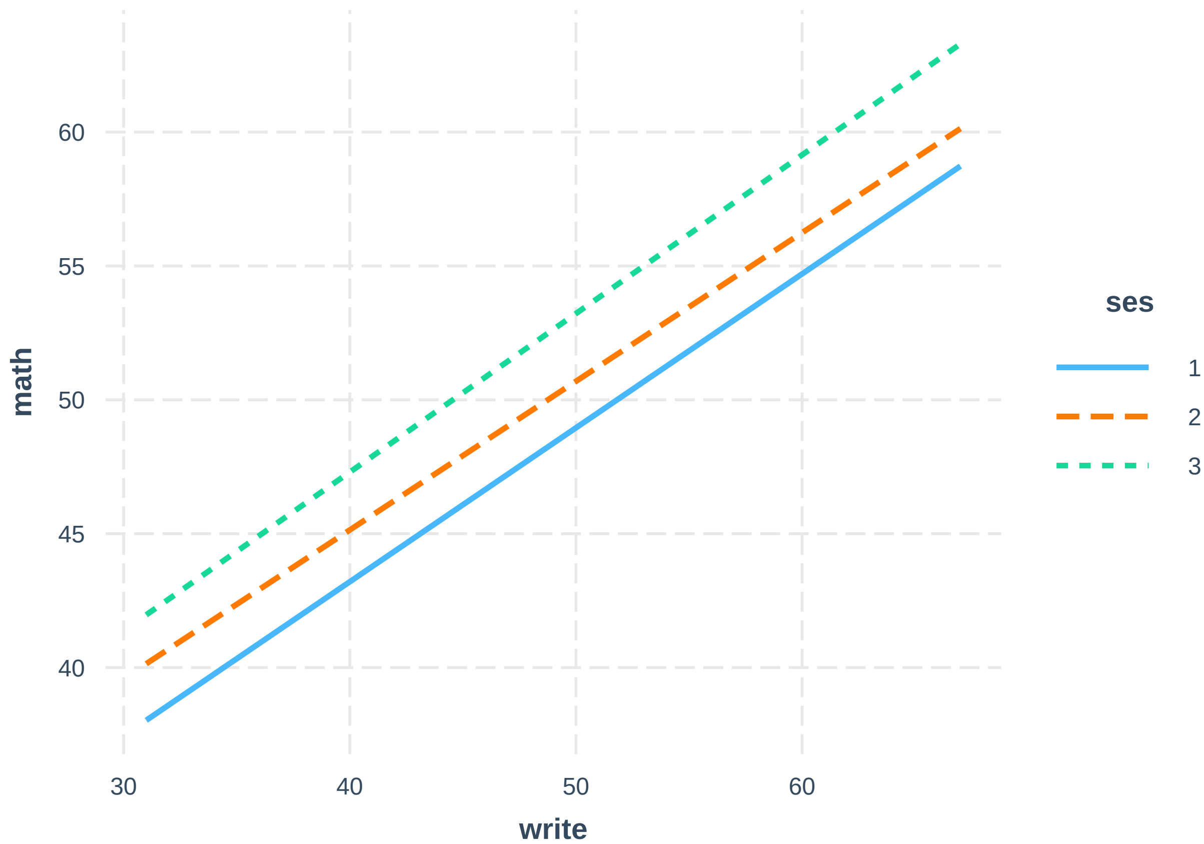
By the graphs, it appears to be both non-constant variance and non-normality in the residuals under the untransformed data, doesn't satisfy the assumption of ANCOVA.

```r
anova(fit.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 ***
## ses:write       2     1.8     0.9   0.0167    0.9834    
## schtyp:write    1     0.4     0.4   0.0079    0.9294    
## 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
```

By the table, p-values of interactions of ses and write, and schtyp and write are larger than α, so fail to reject the null and conclude that they are not significant, p-values for the rest:ses, schtyp, write, interaction between ses adn schtyp are smaller than α, thus, they are significant.

```r
interact_plot(fit.ancova, pred = "write",modx = "ses")
```




```r
lsmeans(fit.ancova, ~ write)
```

```
## NOTE: Results may be misleading due to involvement in interactions

##  write lsmean   SE  df lower.CL upper.CL
##   52.8   51.8 1.05 190     49.7     53.9
##
## Results are averaged over the levels of: ses, schtyp
## Confidence level used: 0.95
```

$\beta = 3.94$ If students have same write, then math of students with $ses = 3 > ses = 2 > ses = 1$. If students have same ses, then math increases as write increases.