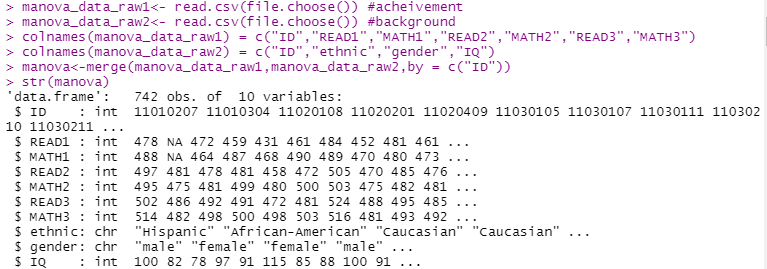
HW1

0859605陳冠景

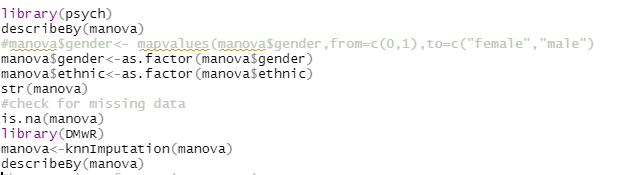
1. Merge the data



Merge two datasets by the “ID” column.

This data is about students’ reading and math scores and their personal information also included.

Since there are missing data , I use KNN function to fill them.



After that , the data size turn into 742.

There are 10 variables included ID ,IQ ,gender and their ethnic about their characteristics and three reading scores, three math scores.

Gender has 2 level ,female and male.

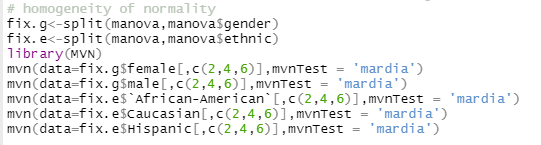
Ethnic has 3 level , Hispanic, African-American and Caucasian.

1. Controlling for students’ IQ scores , I choose to use Mancova statistics to find if any mean difference of three reading achievement measures among ethnic and gender groups. Ethnic and gender are IVs ,so it’s a 2 way mancova test.

Three reading achievement scores are dependent variables meaning it ‘s a multivariate analysis. IQ may be the covariant.

First of all, I need to check several assumptions before the analysis.

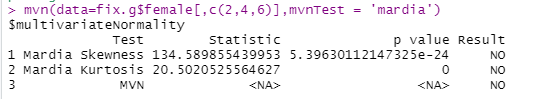
1. Multivariate normality.(Assumption1)



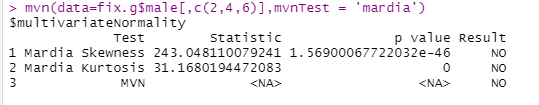
By using MVN function and Mardia test to check the multivariate normality of 5 independent variables.

Result：

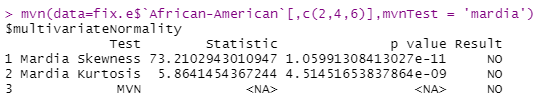
Gender-female



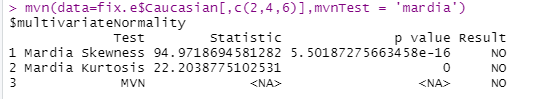
Gender-male



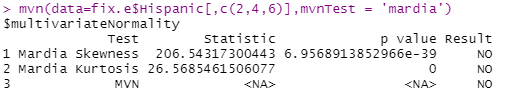
Ethnic - African-American



- Caucasian

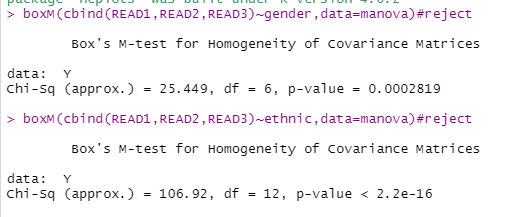


- Hispanic



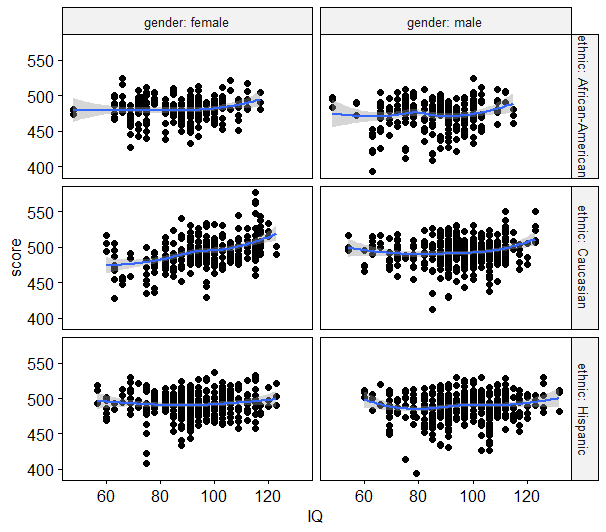
The 5 dependent variables all reject the normality .

1. Homogeneity of covariance matrices across groups. (Assumption2)



BoxM test is used to check the assumption about homogeneity of covariance matrix. The result of the test(=.0001) is significant(χ2gender(6)=25.45, *p* <.001, χ2ethnic(12)=16.94, *p* <.0001),means that gender and ethnic are both reject the the assumption of multivariate homogeneity of variance.

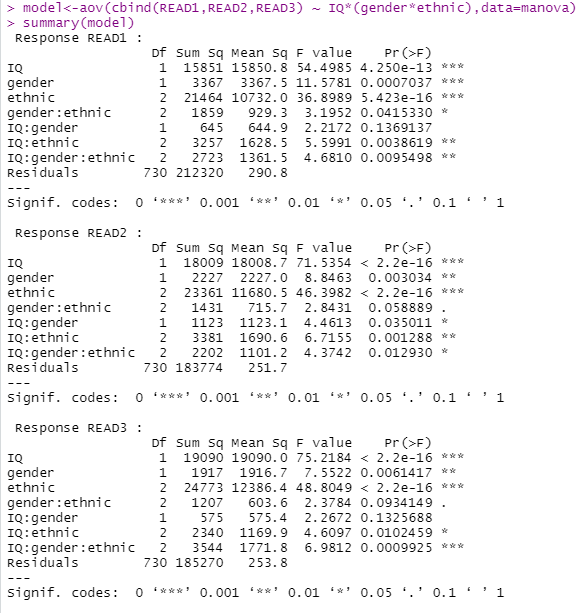
Since BOX M is sensitive to the violation of multivariate normality, I got the significant BOX M.

1. Linearity (Assumption3)

To check the linearity by create a scatter plot between the covariate(i,e.IQ) and the outcome variable (i,e.READ1.READ2.READ3 score) for each combination of the groups of the two grouping variables (gender, ethnic).

There was almost linear relationship between the covariate and the outcome variable for each group.

1. Homogeneity of regression slopes.( Assumption4)



To check is there any significant interaction between the covariate and the grouping variables before ancova analysis.

In READ1 ,between the covariate and grouping variables(ethnic ,gender and

Ethnic, *p*<.05 )were statistically significant.

(IQ:ethnic,*F*(2,730)=5.6,*p*=.004;IQ:gender:ethnic, ,*F*(2,730)=4.68,*p*=.009)

In READ2 ,between the covariate and grouping variables(gender ,gender and

ethnic *p*<.05, ethnic *p*<.01 )were statistically significant.

(IQ:gender,*F*(1,730)=4.46,*p*=.035;IQ:ethnic,*F*(2,730)=6.71,*p*=.001;IQ:gender:ethnic, *F*(2,730)=4.37,*p*=.013)

In READ3 ,between the covariate and grouping variables(ethnic , *p*<.05,gender

and ethnic, ethnic *p*<.001 )were statistically significant.

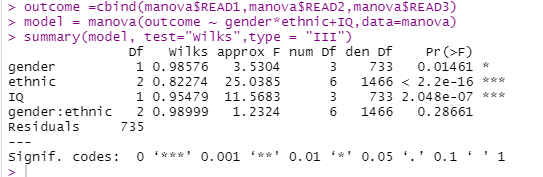
(IQ:ethnic,*F*(2,730)=4.6,*p*=.01;IQ:gender:ethnic, *F*(2,730)=6.98,*p*<.000)

There are interaction in READ1~3, as a result, the homogeneity of regression

slopes in three DVs are all rejected.

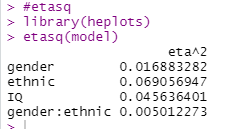
After check all the assumption of MANCOVA ,although both multivariate normality and homogeneity of regression seem to be reject , I still try to do the following analysis.

1. Do the Mancova



By using “Wilk’s test” to test the MANOVA model.

There were significant difference(wilks’ Λ(gender) =.99,*F* =3.53,*p*<.01; wilks’ Λ(gender) =.82,*F* =25.034,*p*<.001) on both gender and ethnic. Therefore, I continue to do post-hoc.



Use heplots function to calculate etaSquared

*Ƞ2* (gender)=.01; *Ƞ2* (ethnic)=.07

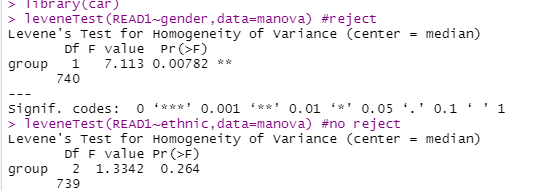
After Manova test, I use ancova as the post-hoc.

Before the post-hoc, it’s required to do the test of homogeneity of variance and homogeneity of regression.

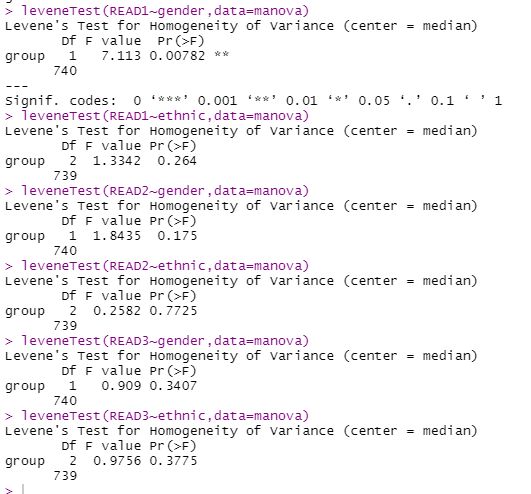
1. Homogeneity of variance

Ancova assumes that the variance of the residuals is equal for all groups.

Check this by Levene’s test.

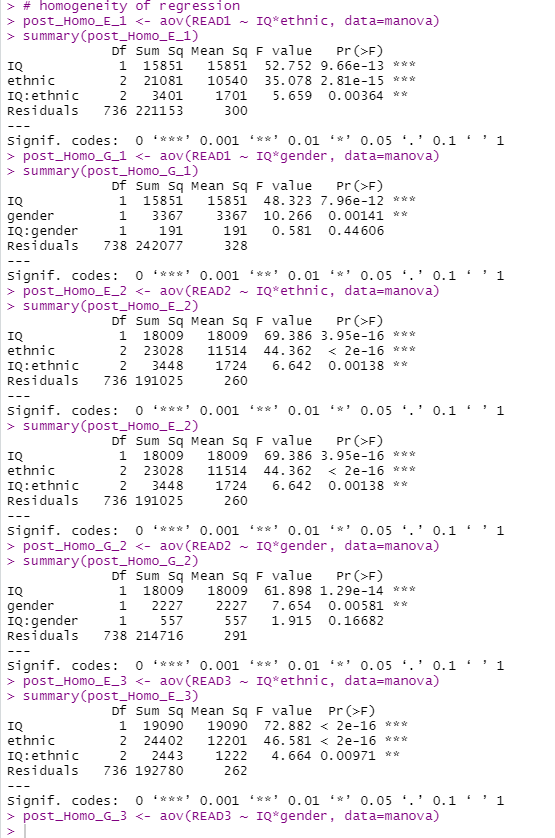


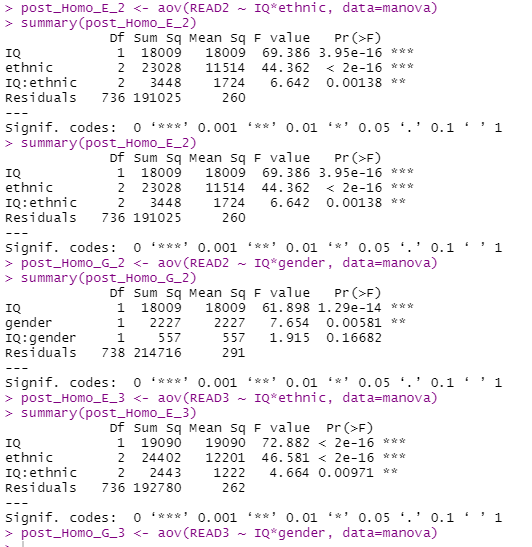
The Levene’s test was significant(Levene’s F = 7.11, *p* =.007),means that different genders have mean difference in READ1,but it was not significant mean difference on ethnic in READ1.



The Levene’s test was not significant in both gender and ethnic in READ2 and READ3.

1. Homogeneity of regression

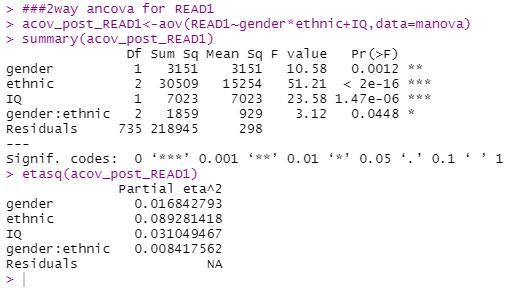




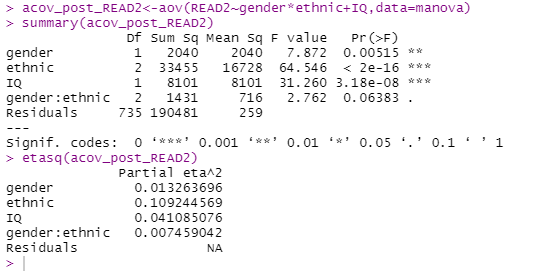
Reject the homogeneity of regression between IQ and ethnic in READ1; between IQ and ethnic , between IQ and gender in READ2; between IQ and ethnic in READ3.

Although I need to reject homogeneity of variance in READ1,and reject the homogeneity of regression in 3 reading, I still try to do post-hoc.

1. 2 way post-hoc ancova

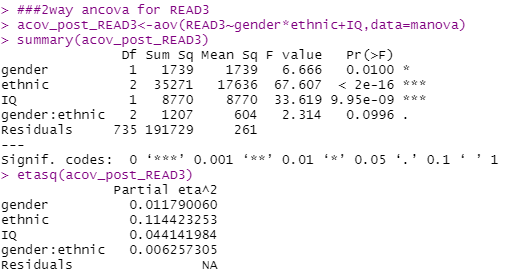


In READ1,there was significant difference in 2-way-ANCOVA,and both gender & ethnic have significant difference(gender:*F*(1,735)=10.58,*p*=.001, *Ƞ2*=.017;ethnic:*F*(2,735)=51.21,*p*<.001, *Ƞ2*=.089), the interactive effect was also significant.(*F*(2,735)=3.12,*p*=.045, *Ƞ2*=.008).



There was no significant interaction in 2-way-ANCOVA in READ2,and

both ethnic & gender have significant difference.(gender:*F*(1,735)=7.87,*p*=.005, *Ƞ2*=.013;ethnic:*F*(2,735)=64.55,*p*<.001, *Ƞ2*=.109).

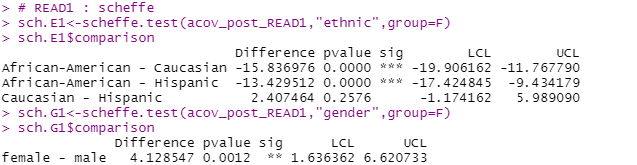


There was no significant interaction in 2-way-ANCOVA in READ3,and

both ethnic & gender have significant difference(*F*(1,735)=6.66,*p*=.01, *Ƞ2*=.012; *F*(2,735)=67.6,*p*<.001, *Ƞ2*=.114).

After 2-way-ANCOVA analysis, do scheffe analysis main effect test.

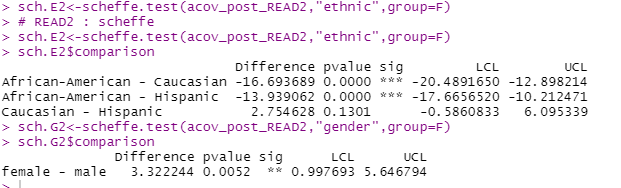
\*for READ1



READ1shows differences between African-American and Caucasian, African-American and Hispanic. The average score on the READ1 in Caucasian is higher than in African-American(ΔMA-C=-15.83,*p*<.001). The average score on the READ1 in Hispanic is higher than in African-American(ΔMA-H=-13.42,*p*<.001).

The average score on the READ1 of female is significantly higher than male(ΔMf-m=4.13,*p*=.001)

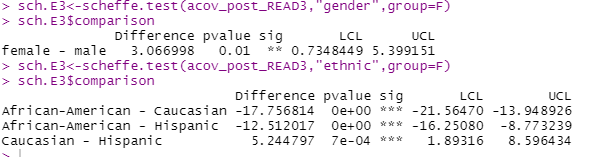
\*for READ2



The average score on the READ2 of female is significantly higher than male(ΔMf-m=3.32,*p*=.005)

READ2 shows differences between African-American and Caucasian, African-American and Hispanic. The average score on the READ2 in Caucasian is higher than in African-American(ΔMA-C=-13.69,*p*<.001). The average score on the READ1 in Hispanic is higher than in African-American(ΔMA-H=-13.93,*p*<.001).

\*for READ3



The average score on the READ3 of female is significantly higher than male(ΔMf-m=3.066,*p*=.01)

READ3 shows not only differences between African-American and Caucasian, African-American and Hispanic,but also between Caucasian and Hispanic .

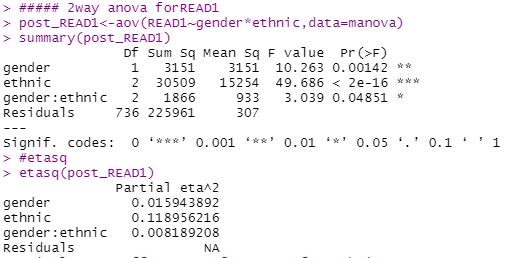
The average score on the READ3 in Caucasian is higher than in African-American(ΔMA-C=-13.69,*p*<.001). The average score on the READ1 in Hispanic is higher than in African-American(ΔMA-H=-13.93,*p*<.001).

The average score on the READ3 in Caucasian is higher than Hispanic(ΔMC-H=5.24,*p*<.001).

I try the 2 way ANOVA for the post-hoc as well.

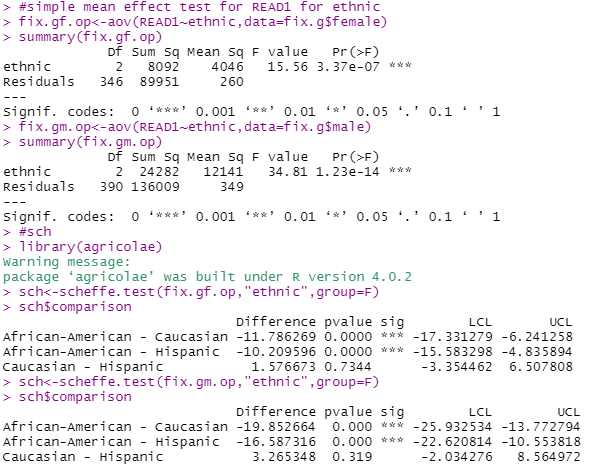
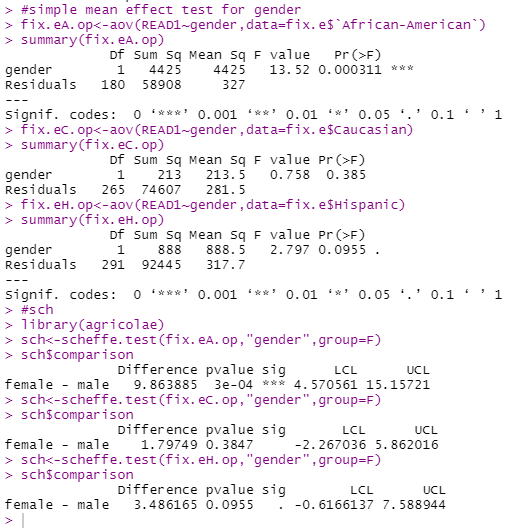
1. 2 way anova

\*forREAD1



For READ1, gender and ethnic have interaction, so simple main effect test is

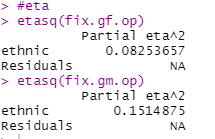
needed.

\*simple main effect for ethnic(fix gender)

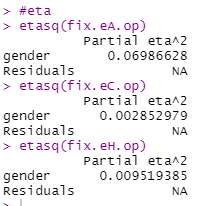
When gender is female ,READ1shows differences between African-American and Caucasian, African-American and Hispanic. The average score on the READ1 of female in Caucasian is higher than in African-American. The average score on the READ1 of female in Hispanic is higher than in African-American.

When gender is male ,READ1shows differences between African-American and Caucasian, African-American and Hispanic. The average score on the READ1 of female in Caucasian is higher than in African-American. The average score on the READ1 of female in Hispanic is higher than in African-American.



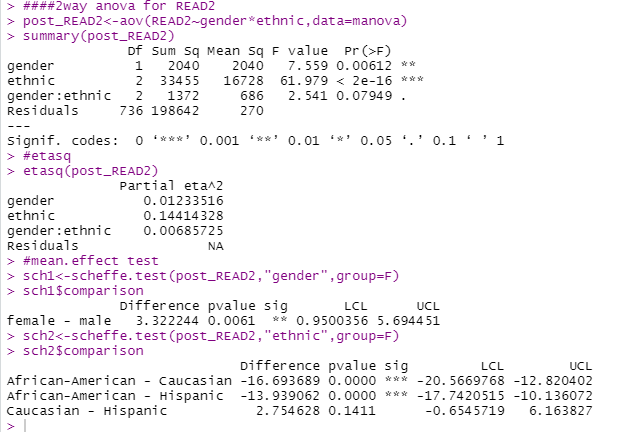
\*simple main effect for gender(fix ethnic)

When ethnic is African-American ,READ1shows significant differences between gender. The average score on the READ1 of female in African-American is higher than male in African-American. Other ethnics don’t have significant difference between gender when controlling ethnic .



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Simple main effect for gender & ethnic | | | | | | |
| Source | SS | df | MS | F | p | *Ƞ2* |
| Ethnic(fix gender) | | | | | | |
| Under female | 8092 | 2 | 4046 | 13.18 | <0.001 | .082 |
| Under male | 24282 | 2 | 14141 | 46.06 | <0.01 | .151 |
| Gender(fix ethnic) | | | | | | |
| African-American | 4425 | 1 | 28.52 | .09 | <0.001 | .069 |
| Caucasian | 213 | 1 | 11.31 | .03 | .38 | .002 |
| Hispanic | 888 | 1 | 888 | 2.89 | .09 | .009 |
| Total | 225961 | 736 | 307 |  |  |  |

\*READ2

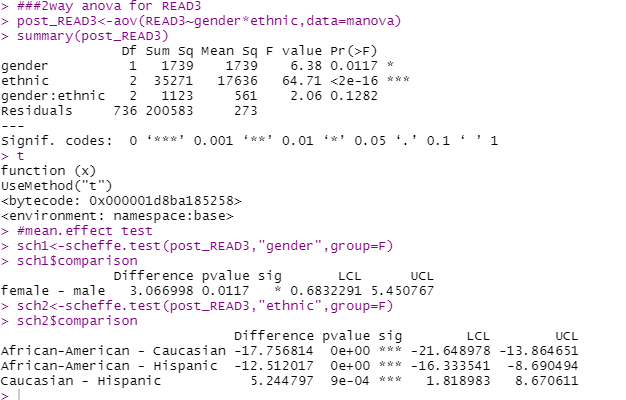


For READ2,gender and ethnic both have significant difference.

The average score on the READ2 of female is higher than male.

The average score on the READ2 in in Caucasian is higher than in African-American. The average score on the READ2 in Hispanic is higher than in African-American.The average score on the READ2 in Caucasian is significant higher than in Hispanic.

\*For READ3



For READ3 ,gender and ethnic both have significant difference.

The average score on the READ3 of female is higher than male. ( ΔMf-

m=3.06,*p*=.011)

The average score on the READ3 in in Caucasian is higher than in African-American (ΔMA-C=-17.75,*p*<.001). The average score on the READ3 in Hispanic is higher than in African-American(ΔMA-H=-12.51,*p*<.001).

The average score on the READ3 in Caucasian is higher than in Hispanic.(ΔMC-

H=5.244,*p*<.001)

1. Conclusion：

I’ve done 3 homogeneity test before MANCOVA. According to the analysis results, the homogeneity of the normality of 5 independent variables. They all reject the normality.

BoxM test (χ2gender(6)=25.45, p <.001, χ2ethnic(12)=16.94, p <.0001) also reject. The 2 main assumptions of MANOVA have been rejected. Otherwise, the third homogeneity of regression tests of three dependent variables are significant between covariate(IQ) and 2 grouping variables(gender, ethnic)(inREAD1,F(2,730)=4.68,p=.009;inREAD2,F(2,730)=4.37,p=.013;inREAD3, F(2,730)=6.98,p<.001).

As a result, I can't do the analysis since the assumptions all violate.

However, I try MANCOVA analysis, the result shows it is significant of the dependent variable on gender(wilks’ Λ(gender) =.99,F =3.53,p<.01,Ƞ2 (gender)=.01)and ethnic(wilks’ Λ(gender) =.82,F =25.034,p<.001,Ƞ2 (ethnic)=.07). It means the arithmetic mean of 3 reading score is different between different gender and different ethnic. The part of interactive effects isn't significantly different.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MANCOVA test table | | | | |
| Source | df | F | p | *Ƞ2* |
| gender | 1 | 3.53 | .014 | .017 |
| ethnic | 2 | 25.04 | <0.001 | .069 |
| IQ | 1 | 11.57 | <0.001 | .045 |
| Gender\*ethnic | 2 | 1.23 | .287 | .005 |
| residuals | 735 |  |  |  |
| Total | 736 |  |  |  |

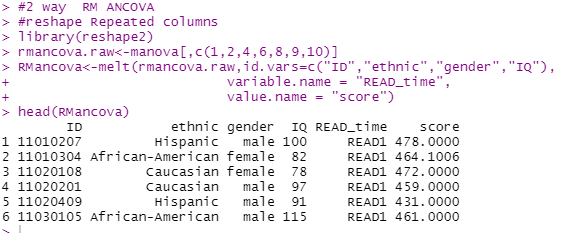
Therefore, I do post hoc for checking the difference between different genders and different ethnics. I use ANCOVA as a post hoc for MANCOVA.

In two way ANCOVA post-hoc, I find out that in 3 reading score, female got higher mean score than male(READ1: ΔMf-m=4.13,*p*=.001;READ2:ΔMf-m=3.32,*p*=.005;READ3:ΔMf-m=3.066,*p*=.01).In three reading score,both

shows significant differences between African-American and Caucasian, African-American and Hispanic. The average score in Caucasian is higher than in African-American(READ1:ΔMA-C=-15.83,*p*<.001;READ2:ΔMA-C=-13.69,*p*<.001;REA3: ΔMA-C=-13.69,*p*<.001). The average score in Hispanic is higher than in African-American(READ1: ΔMA-H=-13.42,*p*<.001;READ2:ΔMA-H=-13.93,*p*<.001;READ3: ΔMA-H=-13.93,*p*<.001).

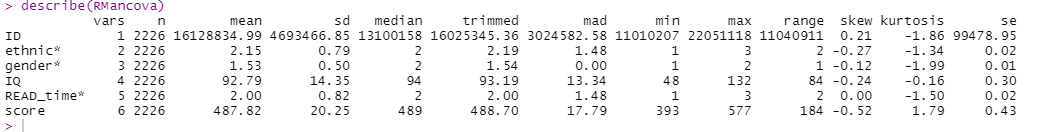
Only in READ3, The average score in Caucasian is significantly higher than Hispanic(ΔMC-H=5.24,p<.001).

1. Controlling for students’ IQ scores , to find if any mean difference of three reading repeated measures of reading achievement among ethnic and gender groups. Ethnic and gender are IVs ,so it’s a 2 way repeated measures ANCOVA.
2. Reshape the repeated measure as a new independent variable.



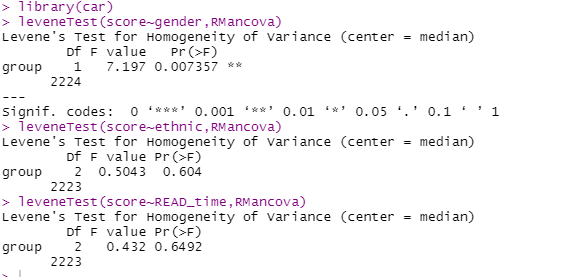
First of all, I need to check serveral assumptions before the analysis.

1. Linearity (see above)(assumption1)
2. Normality(assumption2)



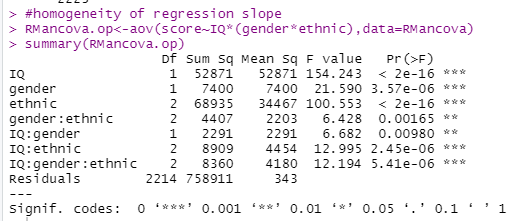
Total data size is 2226, check the skew and kurtosis, all variables have normal distribution.

1. Homogeneity of variance (assumption3)



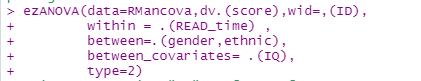
The Levene’s test was significant(Levene’s F = 7.2, *p* =.007),means that different genders have mean difference in reading score. However, it was not significant mean difference on ethnic and READ time, so I assume homogeneity of the residual variance.

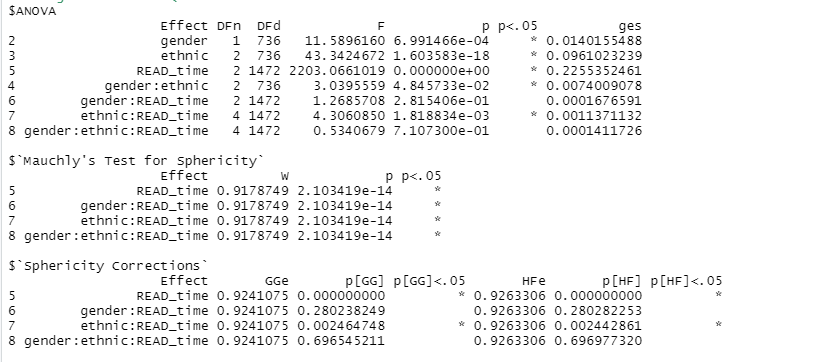
1. Homogeneity of regression slope



There are interaction between gender and IQ, between ethnic and IQ and between gender,ethnic and IQ, as a result, the homogeneity of regression slopes in reading score all reject.

1. Do RM ancova





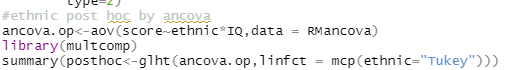
By ez function, gender and read time both have significant on score(reading).

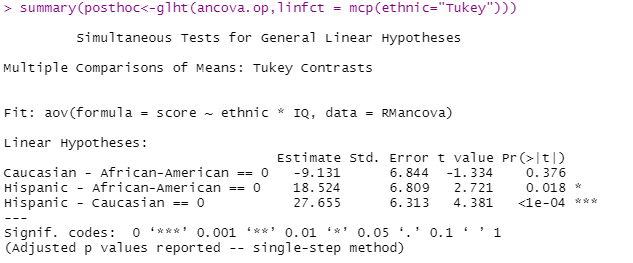
Ethnic was not only significant (*F* (ethnic)=43.34,*p*<.05) but also significantly interact with READ\_time(*F*(ethnic:READ\_time)=4.3,*p*<.05).

As a result, I use repeated measure anova analysis as the post-hoc on gender and READ\_time, but ancova analysis as the pos-hoc on ethnic.

1. Post-hoc

\*for ethnic

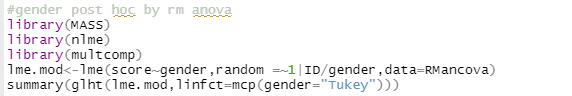


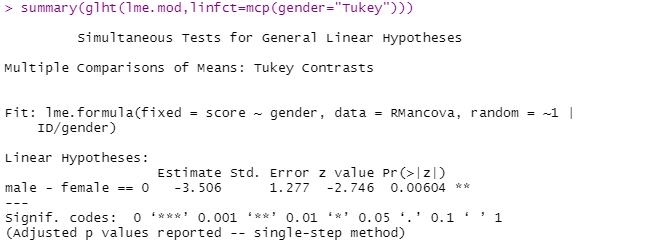


Result: Tukey comparisons shows that the average score on African-American

is higher than in Caucasian. The average score in Hispanic is higher than in African-American and Caucasian .

\*for gender

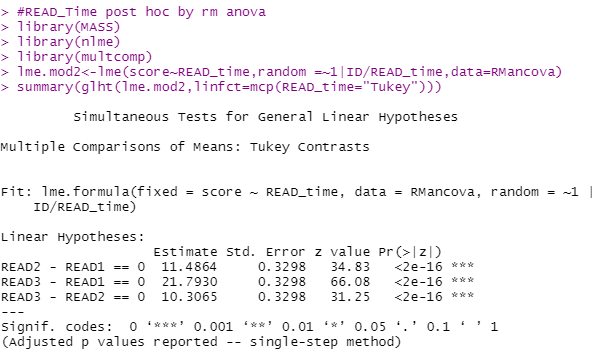




Result: Tukey comparisons shows that the average score of female

is higher than in male.

\*for READ\_time



Result: Tukey comparisons shows that the average score of three reading score is READ3>READ2>READ1.

1. Any difference between 2 analysis?

If I use MANOVA as analysis ,every reading score is a separate variable. W can find out difference of gender and ethnic in each reading score.

If I use repeated measure anova as analysis , the three reading score seem as a variable . We could only compare the mean difference of 3 reading scores in 3 grouping variables(gender, ethnic, read\_time)