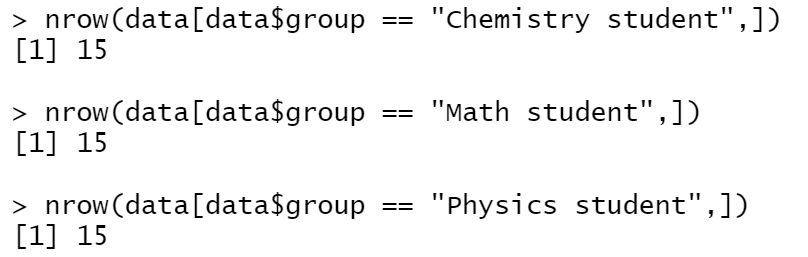
Yiduo Feng

CS 555

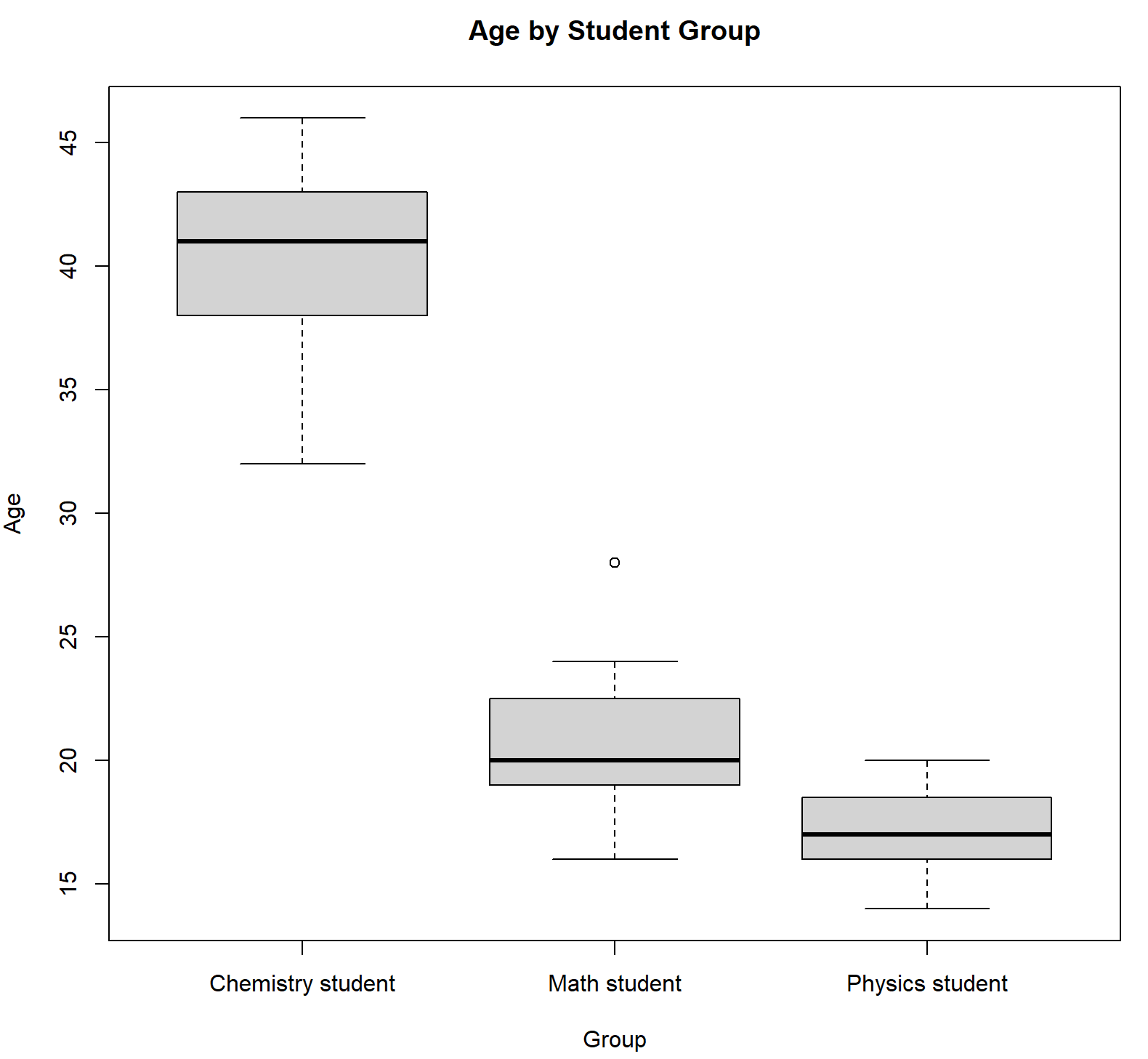
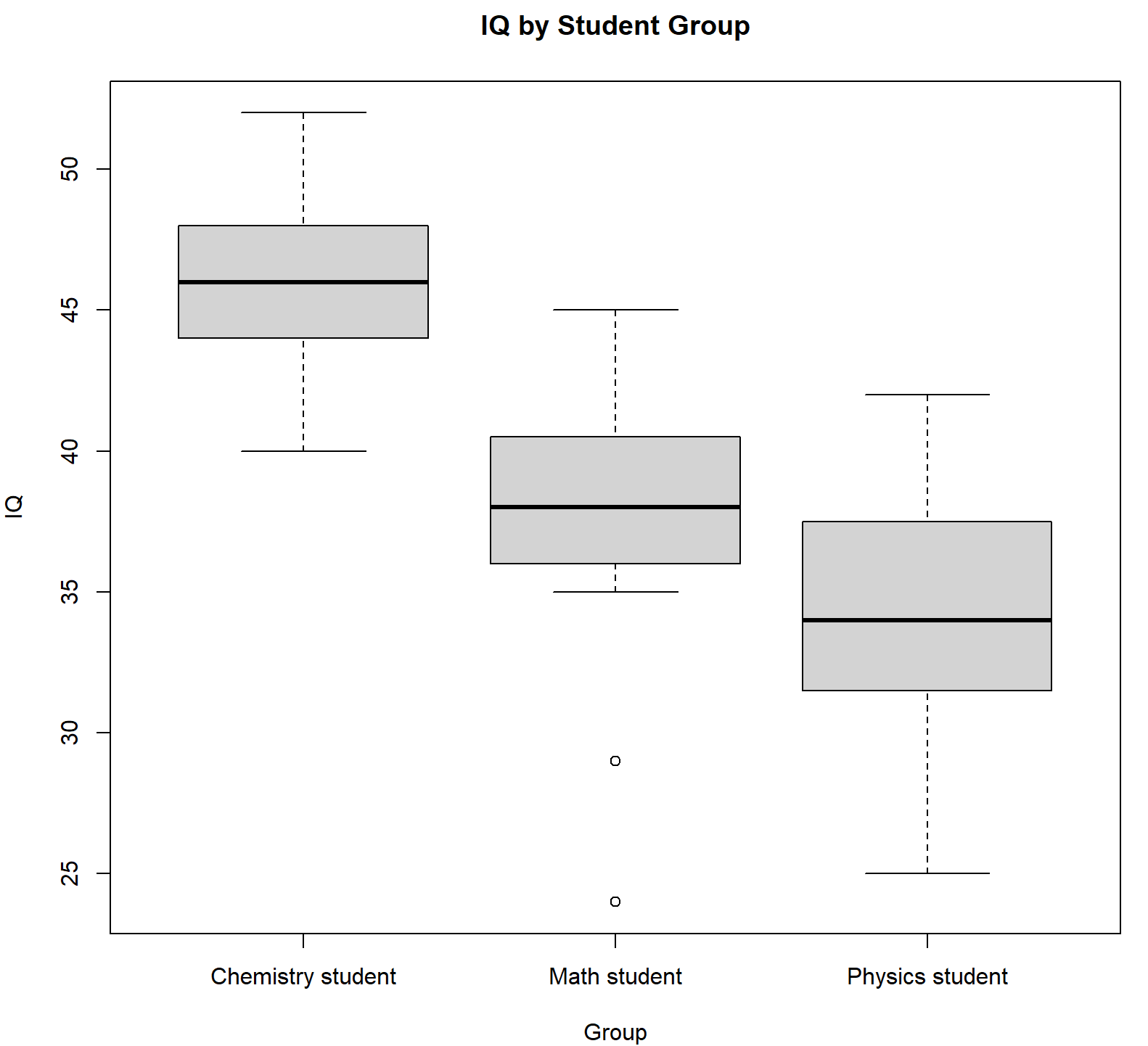
Homework 4

08/10/2022

**(1) How many students are in each group? Summarize the data relating to both test score and age by the student group (separately). Use appropriate numerical and/or graphical summaries. (3 points)**

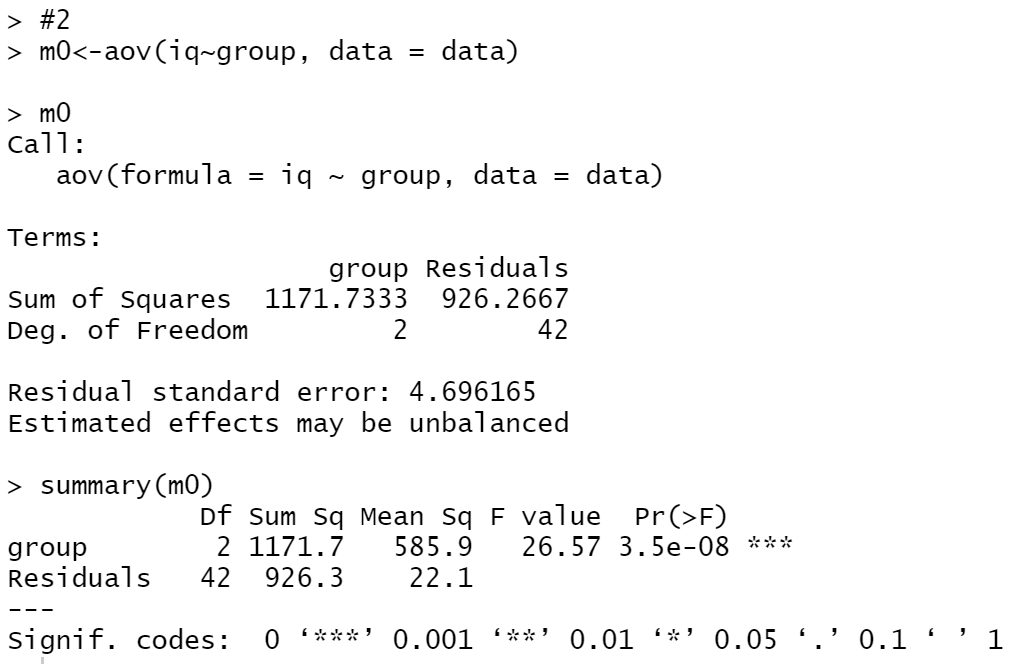


There are 15 students in each group.



The mean of both IQ and age of chemistry student are largest, and that of physics students are lowest. Only group of Math students has outliers in IQ and age.

**(2) Do** **the test scores vary by student group? Perform a one way ANOVA using the aov or Anova function in R to assess. Use a significance level of α=0.05. Summarize the results using the 5-step procedure. If the results of the overall model are significant, perform the appropriate pairwise comparisons using Tukey’s procedure to adjust for multiple comparisons and summarize these results. (7 points )**



1. Set up the hypotheses and select the alpha level

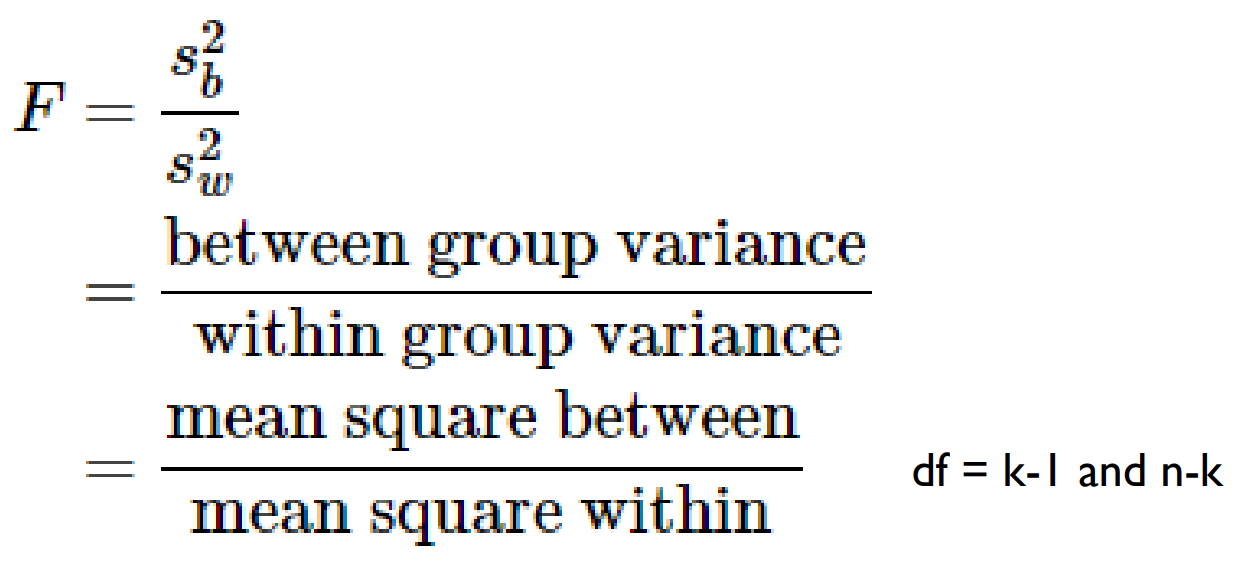
H0 : 𝛽i= 𝛽0 (the test scores don’t vary by student group)

H1 : 𝛽i≠ 𝛽0 (the test scores vary by student group)

𝛼 = 0.05

1. Select the appropriate test-statistic

ANOVA Test

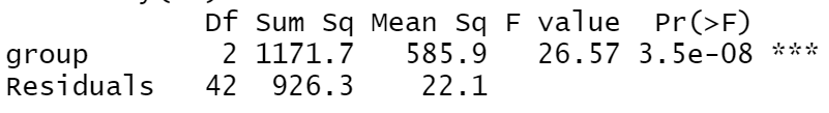


1. State the decision rule

Decision Rule: Reject H0 if 𝑝 ≤ 𝛼. Otherwise, do not reject H

1. Compute the test statistic

According to the code above,



P < 0.05

1. Conclusion

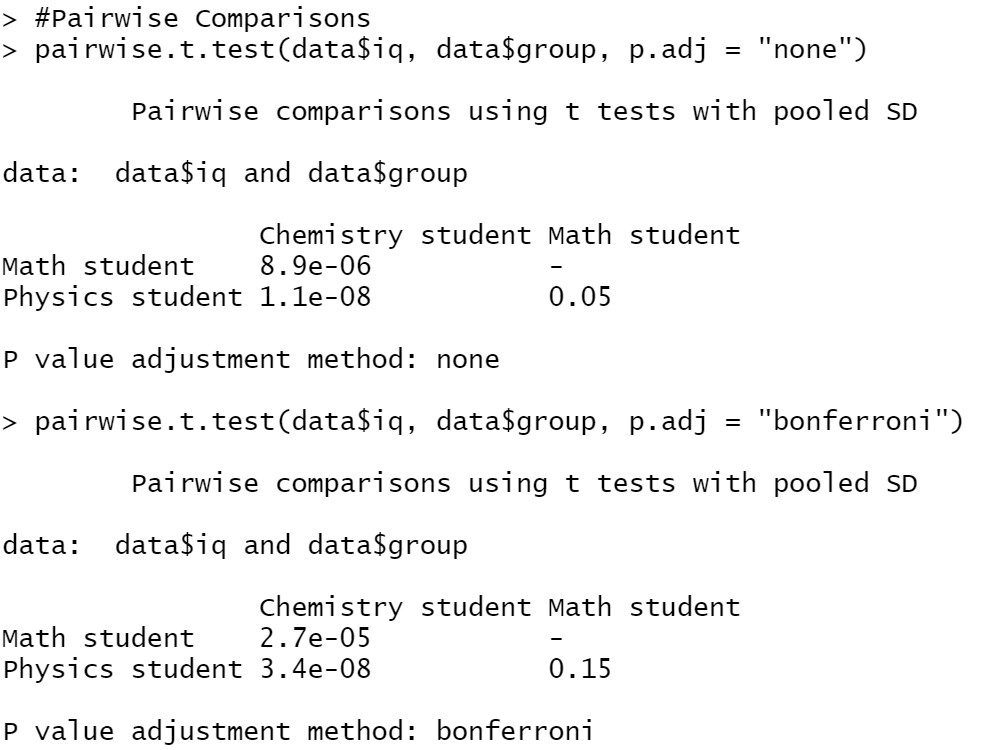
Reject H0

since 𝑝 ≤ 𝛼. We have significant evidence at the 𝛼 = 0.05 level

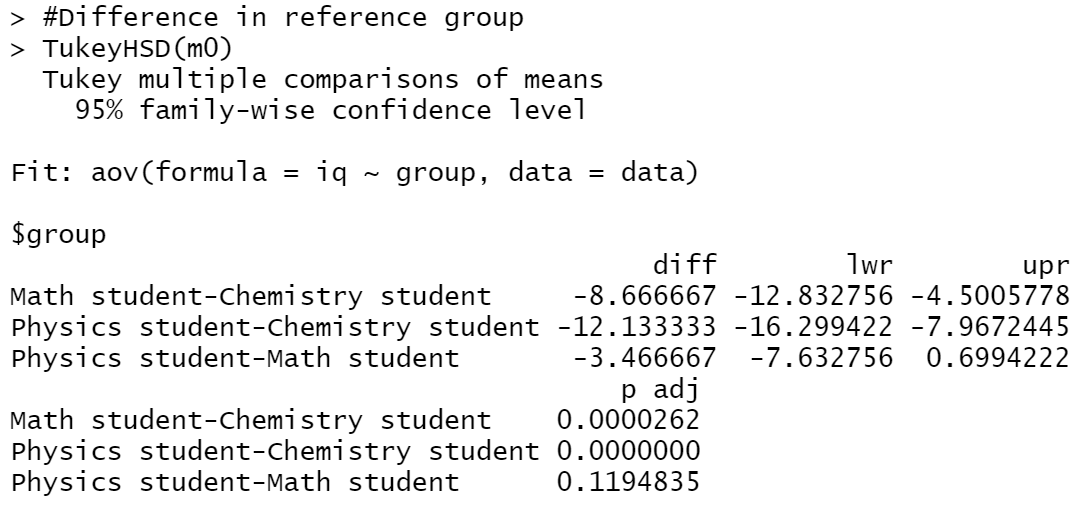
that the test scores vary by student group.

That is, there is evidence that the test scores vary by student group.

Pairwise Comparisons

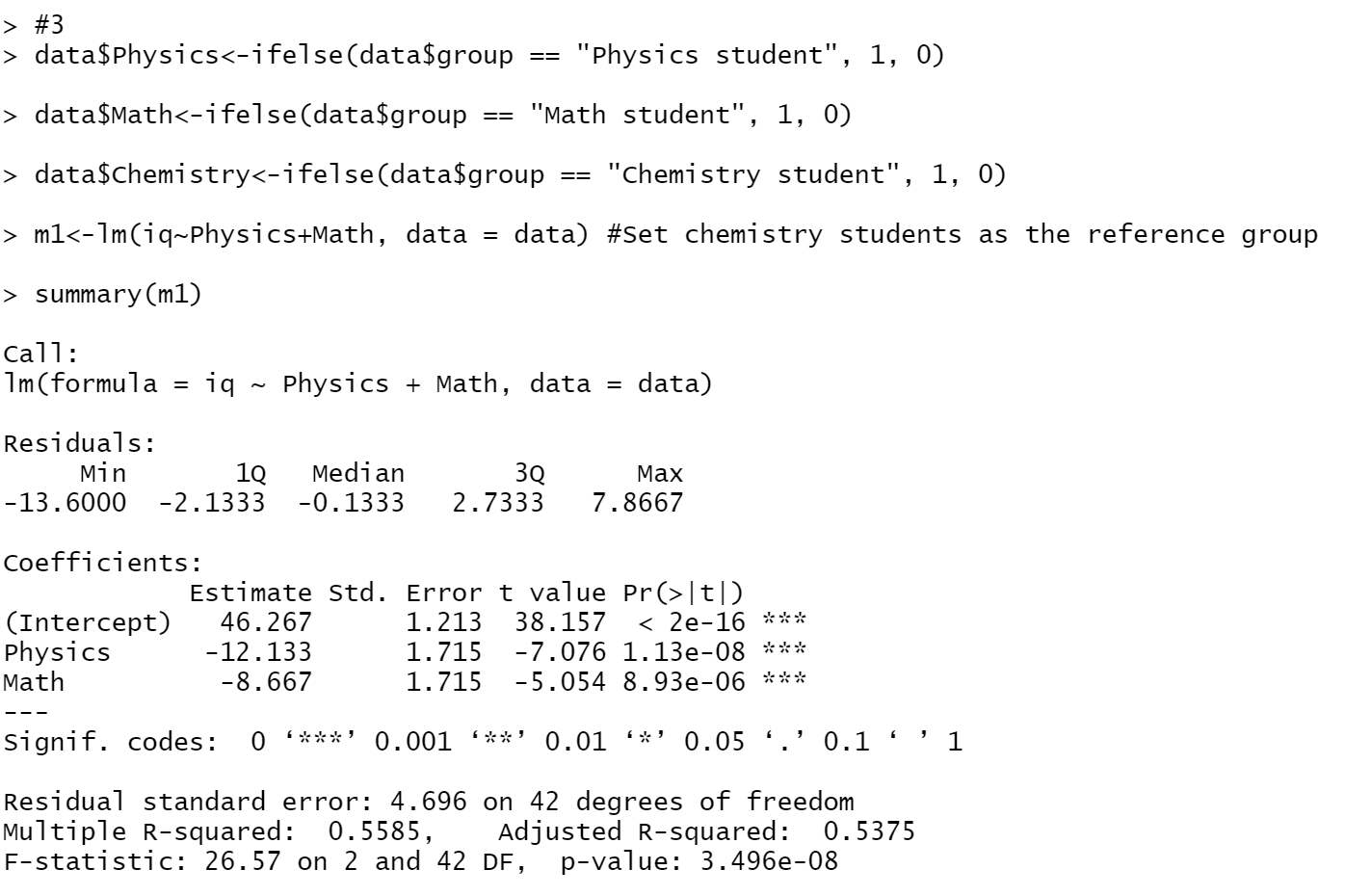


Difference in reference group

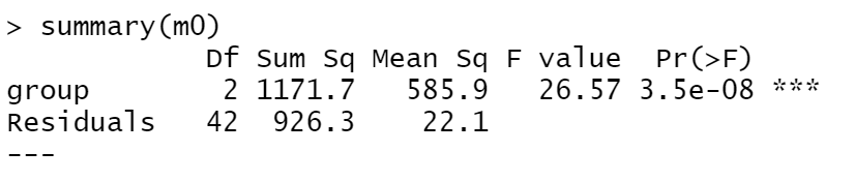


**(3) Create an appropriate number of dummy variables for student group and re-run the one-way ANOVA using the lm function with the newly created dummy variables. Set chemistry students as the reference group. Confirm the results are the same (specifically point out** **test statistics, p-values, etc. that show the results are equivalent). What is the interpretation of the beta estimates from the regression model? (4 points)**

Set chemistry students as the reference group

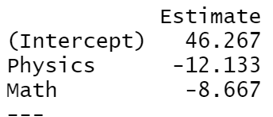


According to the results in question 2



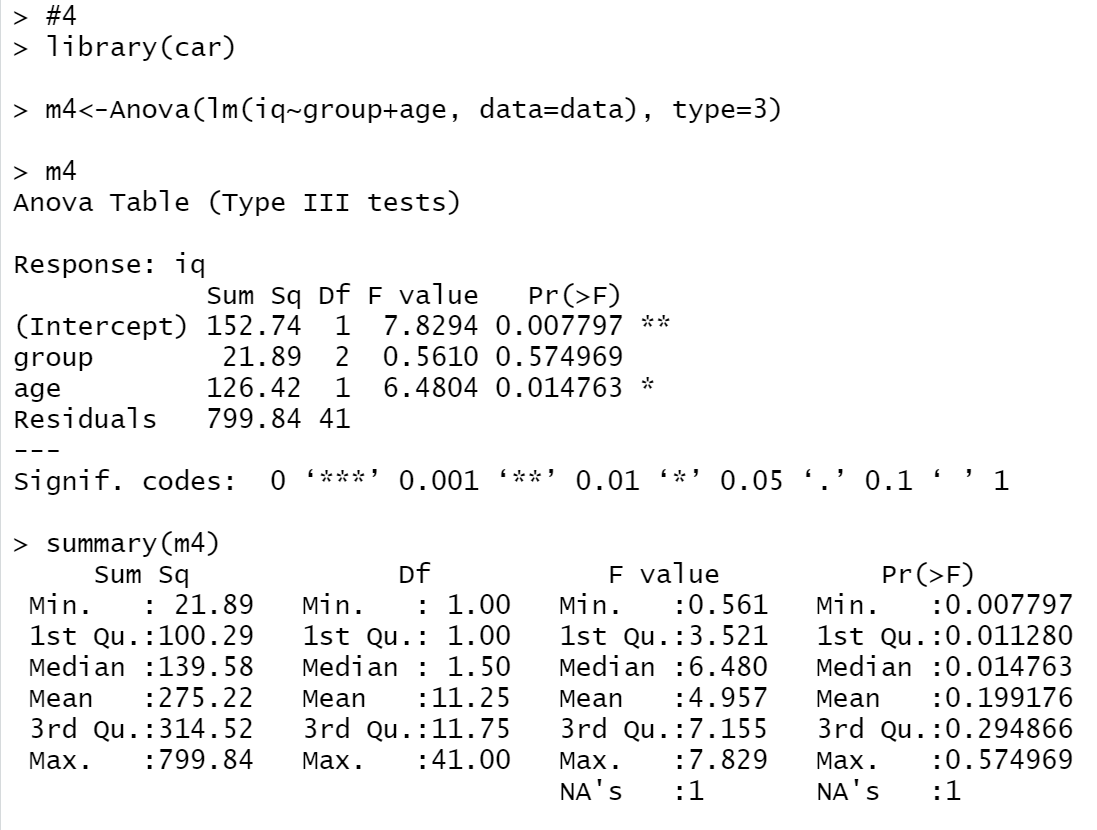
The test statistics, p-values are the same.

From beta estimate,



It shows that iq of Physics student is smaller than chemistry students at 12.133, and iq of Math student is smaller than chemistry students at 8.667.

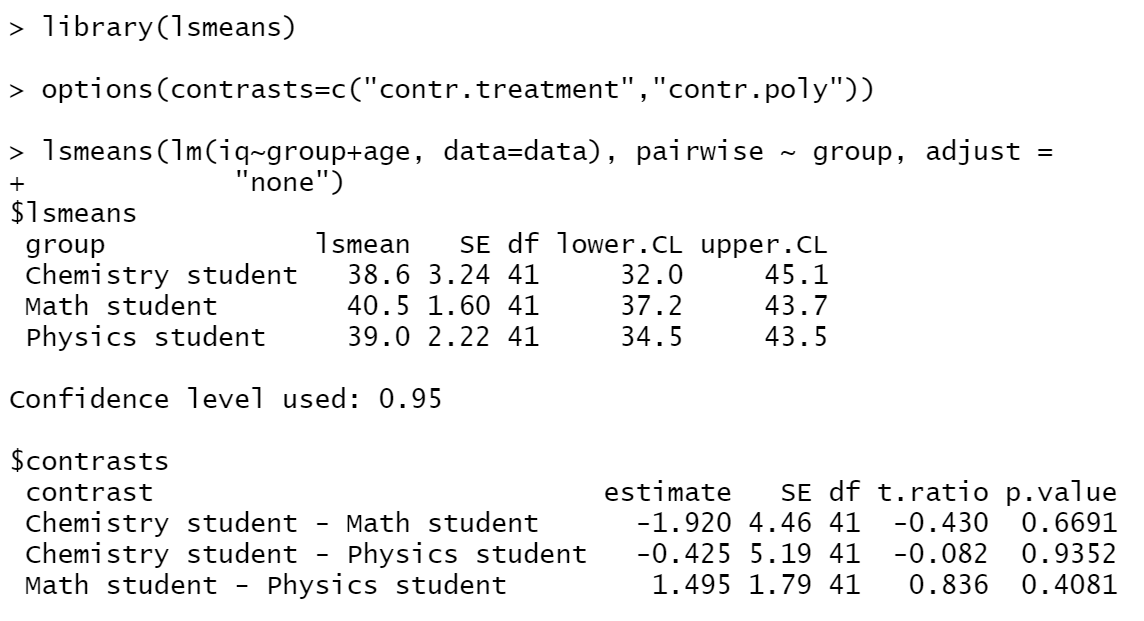
**(4) Re-do the one-way ANOVA adjusting for age (ANCOVA). Focus on the output relating to the comparisons of test score by student type. Explain how this analysis differs from the analysis in step 2 above (not the results but how does this analysis differ in terms of the questions it answers as opposed to the one above). Did you obtain different results? Summarize briefly (no need to go through the 5 –step procedure here). Lastly, present the least square means and interpret these. (6 points)**



By comparing with the analysis in question 2, we used Anova and lim function instead of avo function directly. Summary function gives us more information about f value, p value etc such as min and max. When we analyze these, we use both age column and iq column with group. We also use lsmeans package instead of dummy variables to find estimate.

The mean of p value is larger than 0.05, there are different results

The least square means are shown below.



The least square means of Math group is largest, so that means the uncertainty is highest.

**Code:**

data <- read.csv(file = 'C:/Users/Yidow/Desktop/IQ.csv', fileEncoding="UTF-8-BOM")

nrow(data[data$group == "Chemistry student",])

nrow(data[data$group == "Math student",])

nrow(data[data$group == "Physics student",])

boxplot(iq~group, data = data, main = "IQ by Student Group",

xlab = "Group",

ylab = "IQ")

boxplot(age~group, data = data, main = "Age by Student Group",

xlab = "Group",

ylab = "Age")

#2

m0<-aov(iq~group, data = data)

m0

summary(m0)

#Pairwise Comparisons

pairwise.t.test(data$iq, data$group, p.adj = "none")

pairwise.t.test(data$iq, data$group, p.adj = "bonferroni")

#Difference in reference group

TukeyHSD(m0)

#3

data$Physics<-ifelse(data$group == "Physics student", 1, 0)

data$Math<-ifelse(data$group == "Math student", 1, 0)

data$Chemistry<-ifelse(data$group == "Chemistry student", 1, 0)

m1<-lm(iq~Physics+Math, data = data) #Set chemistry students as the reference group

summary(m1)

m2<-lm(iq~Chemistry+Math, data = data) #Set Physics students as the reference group

summary(m2)

m3<-lm(iq~Physics+Chemistry, data = data) #Set Math students as the reference group

summary(m3)

#4

library(car)

m4<-Anova(lm(iq~group+age, data=data), type=3)

m4

summary(m4)

library(lsmeans)

options(contrasts=c("contr.treatment","contr.poly"))

lsmeans(lm(iq~group+age, data=data), pairwise ~ group, adjust =

"none")