HW5

Jiankun (Bob) Dong CM3226

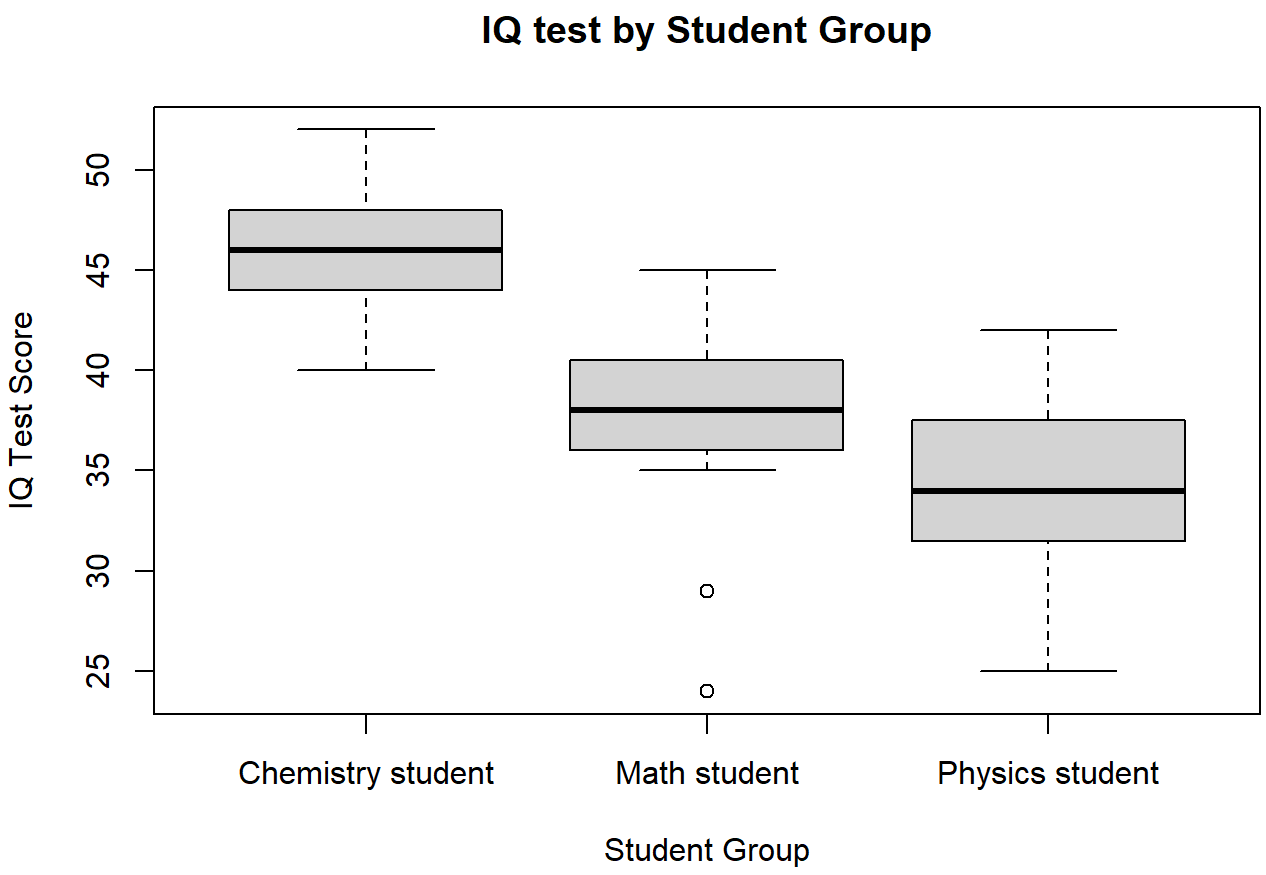
2023-11-18

```
dataSet <- read.csv("./A05.csv")
dataSet$group <- as.factor(dataSet$group)
attach(dataSet)
summary(dataSet)
```

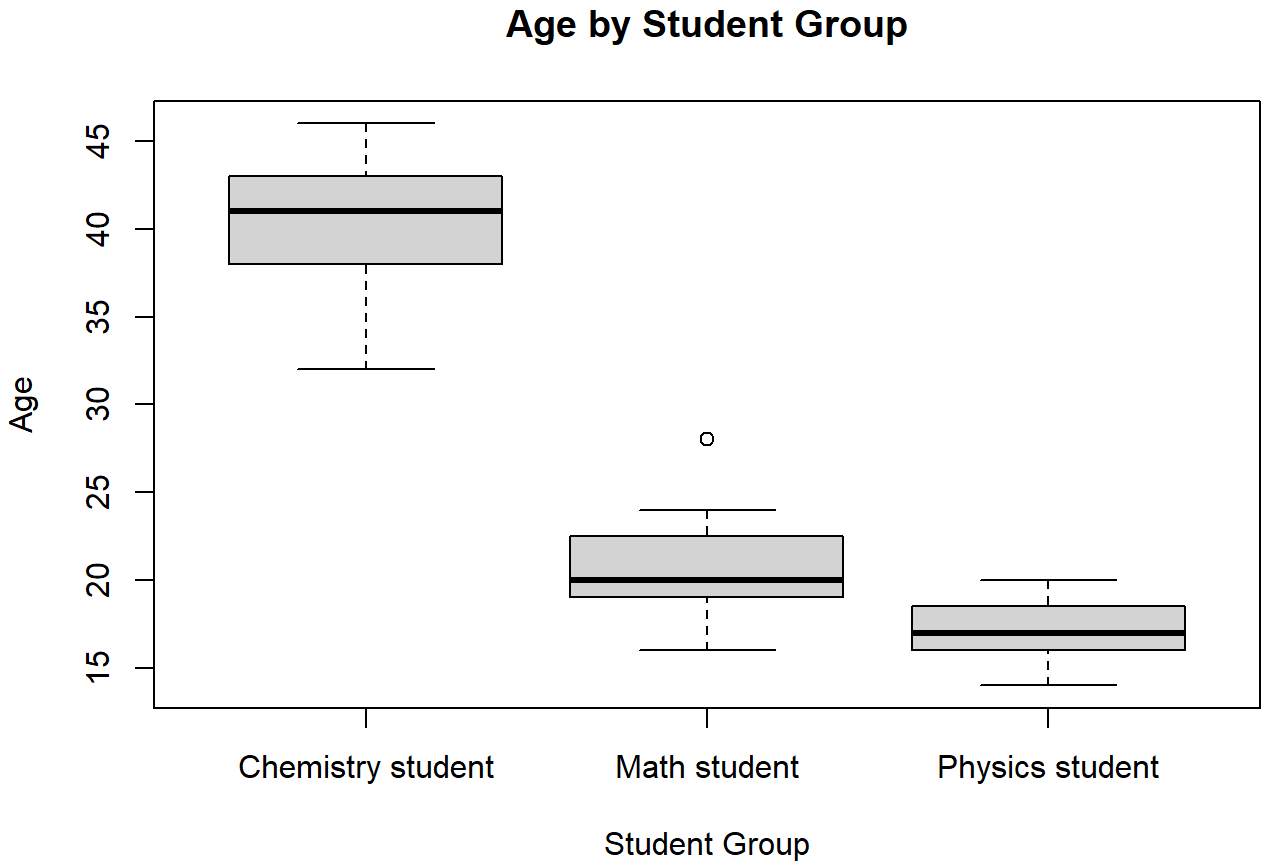
```
##                group          iq              age
##  Chemistry student:15   Min.   :24.00   Min.   :14.00
##  Math student     :15   1st Qu.:34.00   1st Qu.:18.00
##  Physics student  :15   Median :39.00   Median :20.00
##                         Mean   :39.33   Mean   :25.98
##                         3rd Qu.:45.00   3rd Qu.:38.00
##                         Max.   :52.00   Max.   :46.00
```

1. Based on the summary of the data set, there are 15 students in each group. Here are the graph summary of the IQ vs student group and age vs student group.

```
boxplot(iq~group,data = dataSet, main="IQ test by Student Group", xlab="Student Group",
        ylab="IQ Test Score")
```




```
boxplot(age~group,data = dataSet, main="Age by Student Group", xlab="Student Group",
        ylab="Age")
```



2)

Step1:
$H_0 : _phy = _math = _chem $
$H_1 : \beta_i \neq \beta_j for some i and j$
$\alpha = 0.05$
Step2:
$F = \frac{MSB}{MSW}$ with k-1 = 2 and n-k = 42 degrees of freedom
Step3:
Decision rule: reject H_0 if F>3.2199423
Step4:
Compute F:

```
summary(aov(iq~group,data = dataSet))
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## group        2 1171.7   585.9   26.57 3.5e-08 ***
## Residuals   42  926.3    22.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step5:
Because 26.57>3.2199423, we have significant evidence that at $\alpha = 0.05$ that there is a difference in IQ test score among Chemistry, Math and Physics students.

```
m <- aov(iq~group, data = dataSet)
TukeyHSD(m)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = iq ~ group, data = dataSet)
##
## $group
##                                         diff        lwr        upr     p adj
## Math student-Chemistry student     -8.666667 -12.832756 -4.5005778 0.0000262
## Physics student-Chemistry student -12.133333 -16.299422 -7.9672445 0.0000000
## Physics student-Math student       -3.466667  -7.632756  0.6994222 0.1194835
```

At $\alpha = 0.05$ level, we do not have enough evidence that the mean IQ score between Physics and Math students differs, but we have significant evidence that the mean IQ score between Physics - Chemistry and Math - Chemistry student differs. 3)

```
dataSet$g0 <- ifelse(group == 'Chemistry student', 1, 0)
dataSet$g1 <- ifelse(group=='Math student', 1, 0)
dataSet$g2 <- ifelse(group=='Physics student', 1, 0)
m1 <- lm(iq ~ g1+g2,data=dataSet)
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: iq
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## g1         1   67.60   67.60  3.0652   0.08729 .
## g2         1 1104.13 1104.13 50.0651 1.133e-08 ***
## Residuals 42  926.27   22.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = iq ~ g1 + g2, data = dataSet)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6000  -2.1333  -0.1333   2.7333   7.8667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   46.267      1.213  38.157  < 2e-16 ***
## g1            -8.667      1.715  -5.054 8.93e-06 ***
## g2           -12.133      1.715  -7.076 1.13e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.696 on 42 degrees of freedom
## Multiple R-squared:  0.5585, Adjusted R-squared:  0.5375
## F-statistic: 26.57 on 2 and 42 DF,  p-value: 3.496e-08
```

With chemistry student as reference, math student group has p value of .087 while physics student has p value of 1.133*10^-8. This means that math student is not a group with significantly different mean.This agrees with the previous result that only one of physics and math group is significantly different. For the β it means that the expected IQ test score would be 8.667 lower if it's a student from the math students group, and 12.133 lower if it's from the physics students group.

4)

```
m2<-lm(iq~group+age,data=dataSet)
Anova(m2,type=3)
```

```
## Anova Table (Type III tests)
##
## Response: iq
##             Sum Sq Df F value   Pr(>F)
## (Intercept) 152.74  1  7.8294 0.007797 **
## group        21.89  2  0.5610 0.574969
## age         126.42  1  6.4804 0.014763 *
## Residuals   799.84 41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANCOVA test analysis differs from the previous analysis in that it takes the effect of age as the covariance out of the ANOVA analysis. It shows that if the different we see between group of students is or is not due to age difference. Yes we obtained different result: the difference in result we see in the group of students is due to age difference.

```
(m2.emm <- emmeans(m2, 'group'))
```

```
##  group             emmean   SE df lower.CL upper.CL
##  Chemistry student   38.6 3.24 41     32.0     45.1
##  Math student        40.5 1.60 41     37.2     43.7
##  Physics student     39.0 2.22 41     34.5     43.5
##
## Confidence level used: 0.95
```

When adjusted for the covariance age, the mean value of IQ test for chemistry students is 38.6, for math students is 40.5 and for physics students is 39.0.