ANOVA & Testing Equality of Group Variances

SIT Wai Tang

```
rm(list=ls())
library(aod)
library(data.table)
library(leaps)
```

Description: 4 tutorial classes. All classes under same circumstance but taught by different tutors. The examination scores for random samples of students independently drawn from the four classes are given in the following table:

```
# input data
tutor1 = c(80,40,50,40,60,70)
tutor2 = c(80,60,80,90,80,60,60)
tutor3 = c(50,60,90,40,70,80,60,70)
tutor4 = c(70,80,90,80,90,70,100)

df = cbind(c(rep("Polly", length(tutor1)), rep("Pauline", length(tutor2)),
             rep("Paul", length(tutor3)), rep("Patsy", length(tutor4))),
           c(tutor1, tutor2, tutor3, tutor4))
colnames(df) = c("Tutor", "Score")
df = data.frame(df)
df$Score = as.numeric(as.character(df$Score))
df$Tutor = as.factor(as.character(df$Tutor))
df
```

```
##      Tutor Score
## 1    Polly    80
## 2    Polly    40
## 3    Polly    50
## 4    Polly    40
## 5    Polly    60
## 6    Polly    70
## 7  Pauline    80
## 8  Pauline    60
## 9  Pauline    80
## 10 Pauline    90
## 11 Pauline    80
## 12 Pauline    60
## 13 Pauline    60
## 14    Paul    50
## 15    Paul    60
## 16    Paul    90
## 17    Paul    40
## 18    Paul    70
## 19    Paul    80
## 20    Paul    60
## 21    Paul    70
## 22   Patsy    70
## 23   Patsy    80
## 24   Patsy    90
## 25   Patsy    80
## 26   Patsy    90
## 27   Patsy    70
## 28   Patsy   100
```

```
summary(df)
```

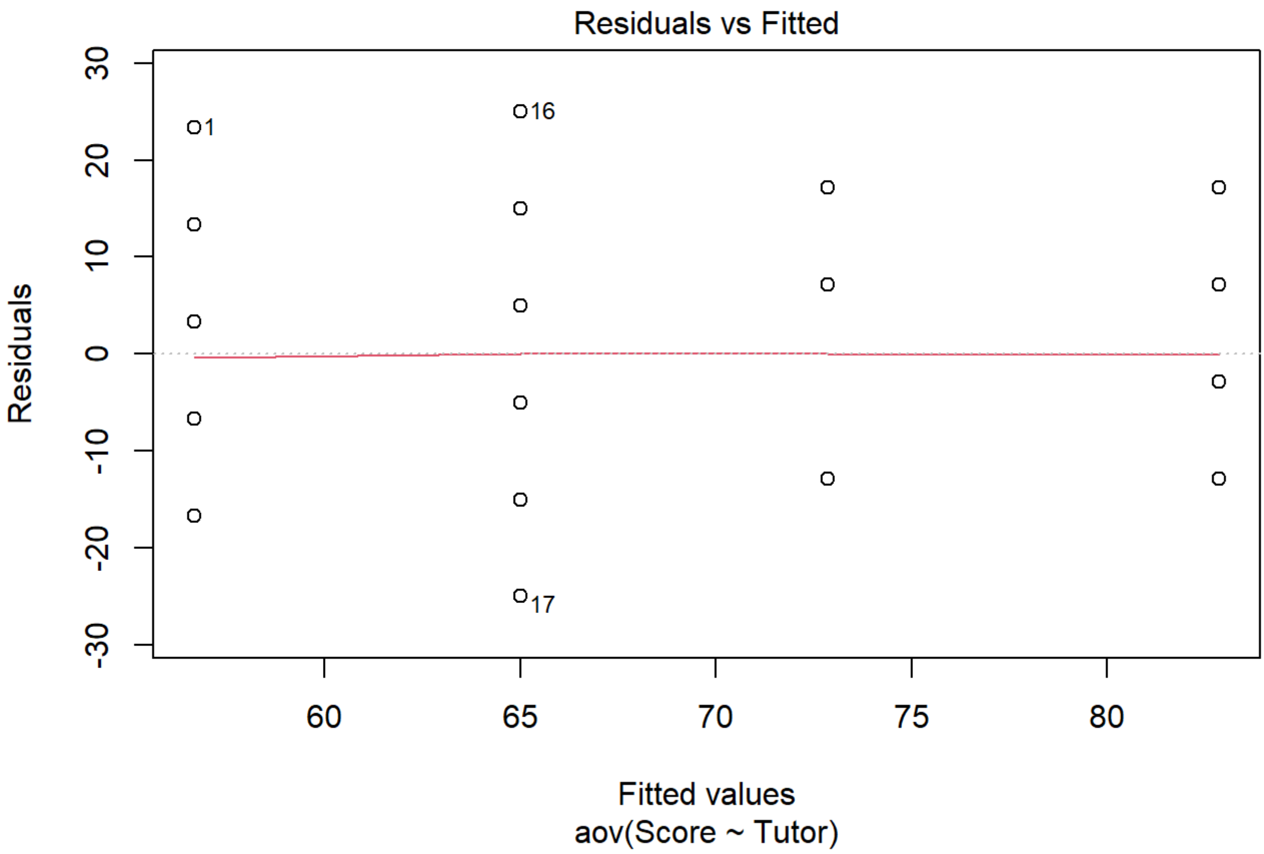
```
##      Tutor       Score
##  Patsy  :7   Min.   : 40.00
##  Paul   :8   1st Qu.: 60.00
##  Pauline:7   Median : 70.00
##  Polly  :6   Mean   : 69.64
##              3rd Qu.: 80.00
##              Max.   :100.00
```

Test whether the average scores of students from 4 classes are different

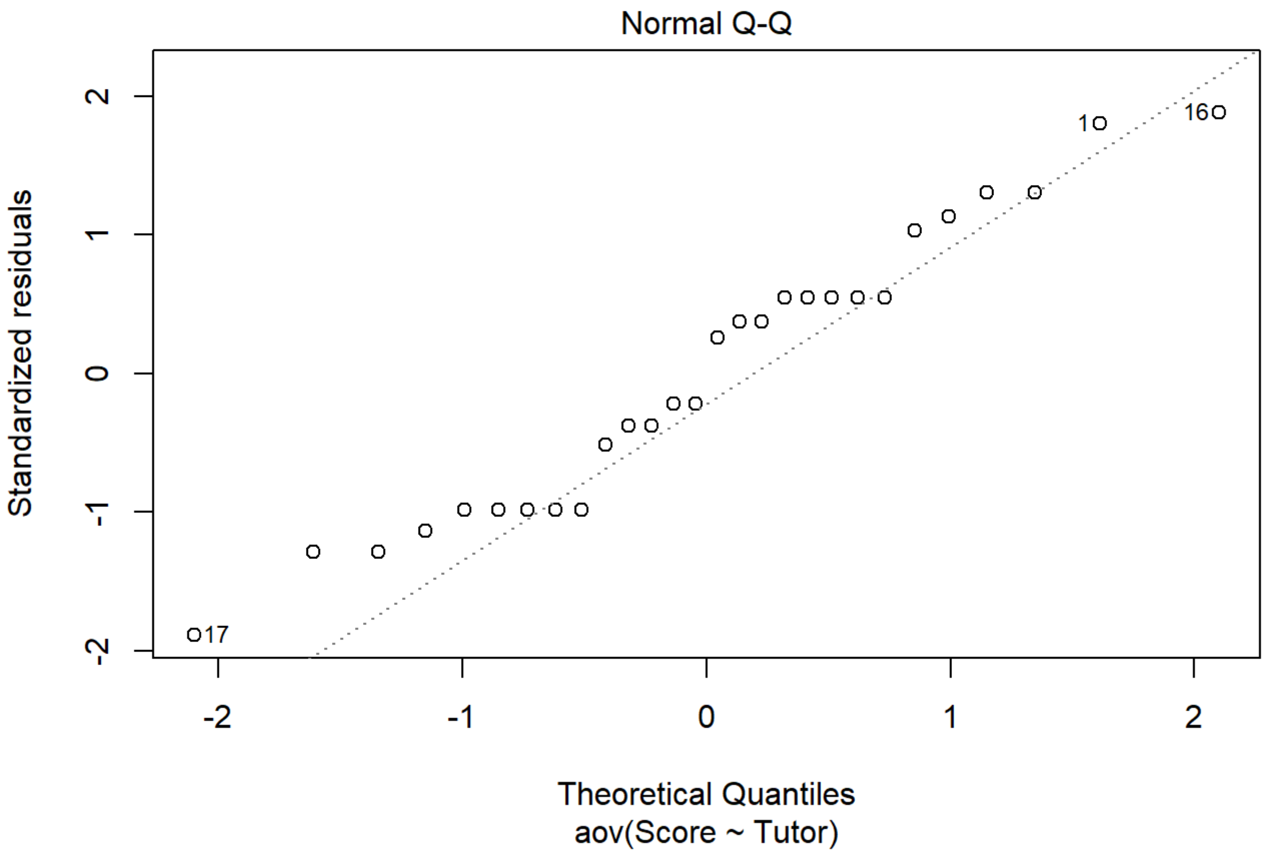
```
anova_one_way <- aov(Score~Tutor, data = df)
summary(anova_one_way)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Tutor        3   2477   825.8   4.113 0.0173 *
## Residuals   24   4819   200.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

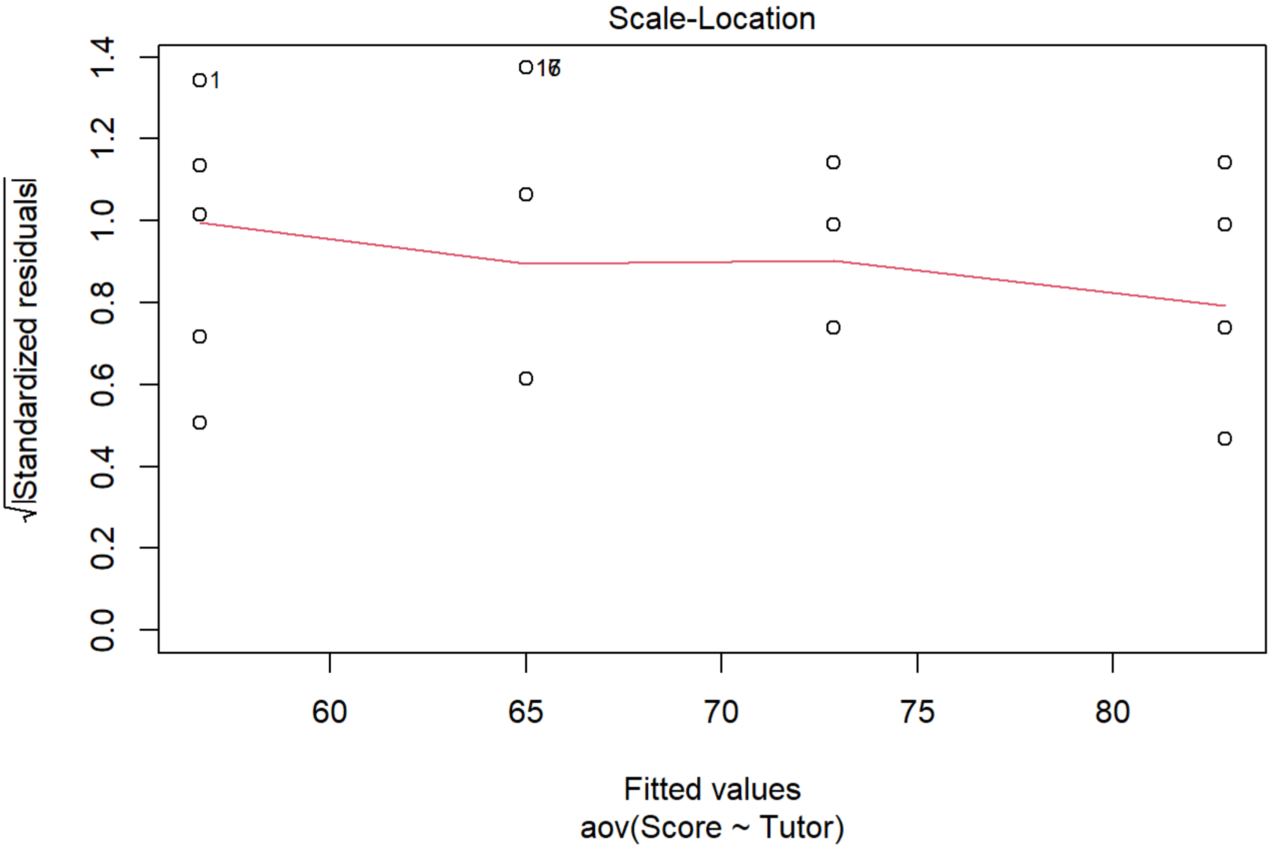
```
plot(anova_one_way)
```

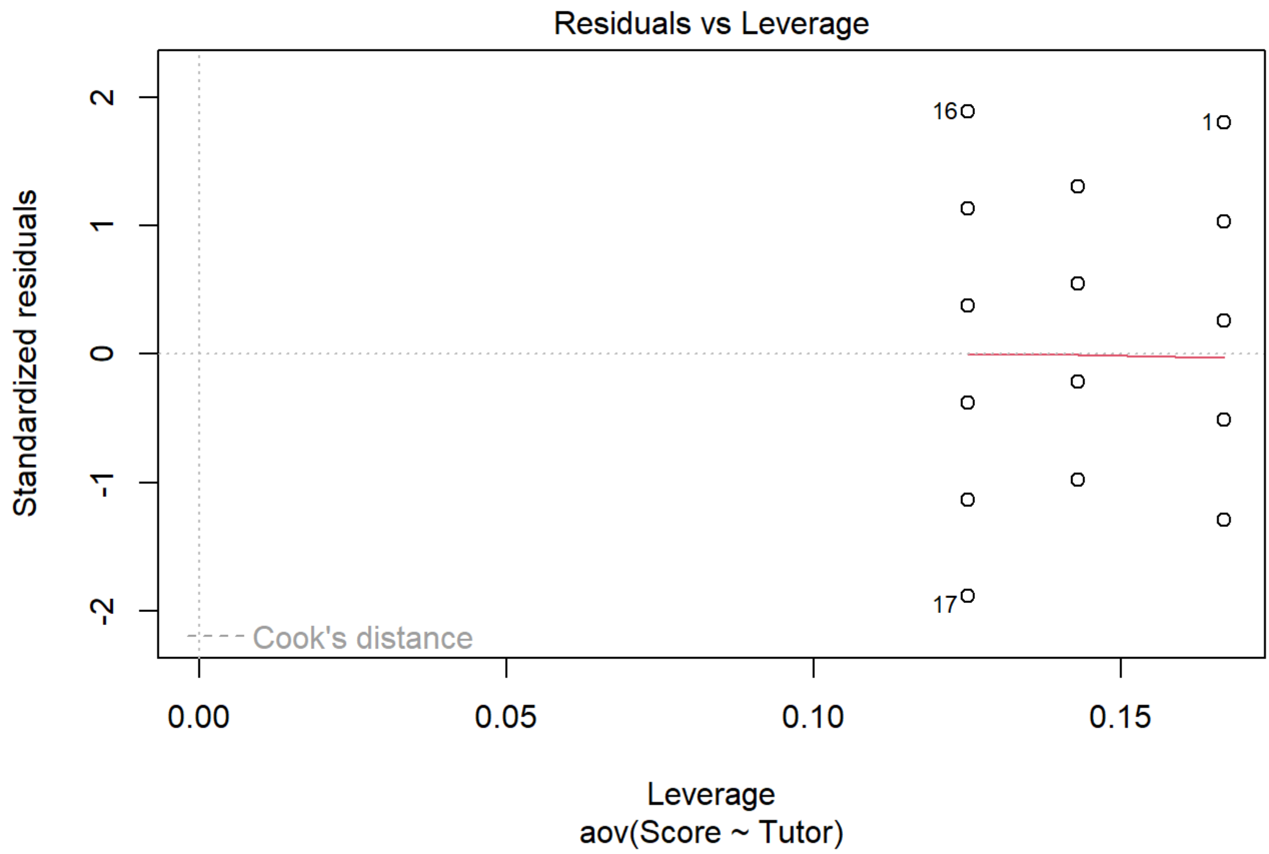


Residuals vs Fitted
Residuals
Fitted values
aov(Score ~ Tutor)
30
20
10
0
-10
-20
-30
60
65
70
75
80
1
16
17



Normal Q-Q
Standardized residuals
Theoretical Quantiles
aov(Score ~ Tutor)
2
1
0
-1
-2
-2
-1
0
1
2
1
16
17





one way

Anova

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

The p-value is 0.0173 < 0.05, therefore we reject the H_0 and conclude that the average scores of the students from the four classes are different.

the validity of assumptions:

1. Independence

As description mentioned, the samples are independently drawn.

2. Constant variance

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 == \sigma_4^2$

H_1 : at least two variances are different

Bartlett's Test

```
bartlett.test(Score~Tutor, df)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Score by Tutor
## Bartlett's K-squared = 1.1414, df = 3, p-value = 0.7671
```

p-value = 0.7671 is very large, so we do not reject the H_0 and can onclude that constant variance assumption is valid.

3. Normality

```
shapiro.test(df$Score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Score
## W = 0.94389, p-value = 0.1388
```

H_0:the data are normally distributed vs H_1:data are not normally distributed

p-value = 0.1388 is large enough (>0.05), therefore we do not reject the null and can conclude that the normality assumption is valid.

Model: $E(score) = \beta_0 + \beta_1 I(Tutor = Paul) + \beta_2 I(Tutor = Pauline) + \beta_3 I(Tutor = Polly)$

Estimate and standard error of β0, β1, β2, β3

```
coef(summary.lm(anova_one_way))
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)   82.85714   5.355820 15.470487 5.531147e-14
## TutorPaul    -17.85714   7.333758 -2.434924 2.270054e-02
## TutorPauline -10.00000   7.574273 -1.320259 1.992024e-01
## TutorPolly   -26.19048   7.883553 -3.322166 2.853174e-03
```

```
b = coef(summary.lm(anova_one_way))[1:4, 1]

se = coef(summary.lm(anova_one_way))[1:4, 2]

list("estimate:"=b,"standard error:"=se)
```

```
## $`estimate:`
##  (Intercept)    TutorPaul TutorPauline   TutorPolly
##     82.85714    -17.85714    -10.00000    -26.19048
##
## $`standard error:`
##  (Intercept)    TutorPaul TutorPauline   TutorPolly
##     5.355820     7.333758     7.574273     7.883553
```

Under some specified weighting

$0.5\beta_0 + \beta_1 - 2\beta_2 - \beta_3$

Estimate and standard error:

```
est = 0.5*b[1]+b[2]-2*b[3]-b[4]

se1 = sqrt(0.5**2*se[1]**2+se[2]**2+2**2*se[3]**2+se[4]**2)

list("estimate_1:"=est,"standard error_1:"=se1)
```

```
## $`estimate_1:`
## (Intercept)
##     69.7619
##
## $`standard error_1:`
## (Intercept)
##    18.77722
```

Multiple Comparisons for pairs that are significantly different.

Bonferroni method:

We have 4C2 = 6 comparisons

```
pairwise.t.test(df$Score, df$Tutor, p.adj = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$Score and df$Tutor
##
##         Patsy Paul  Pauline
## Paul    0.136 -     -
## Pauline 1.000 1.000 -
## Polly   0.017 1.000 0.306
##
## P value adjustment method: bonferroni
```

```
# significance threshold = 0.05

#pairwise.t.test(df$Score, df$Tutor, p.adj ="none")
# significance threshold = 0.05/4C2 = 0.05/6
# the Bonferroni method uses the Fisher’s LSD idea for examining c = k(k − 1)/2 comparisons, with α repl
aced by α/c
```

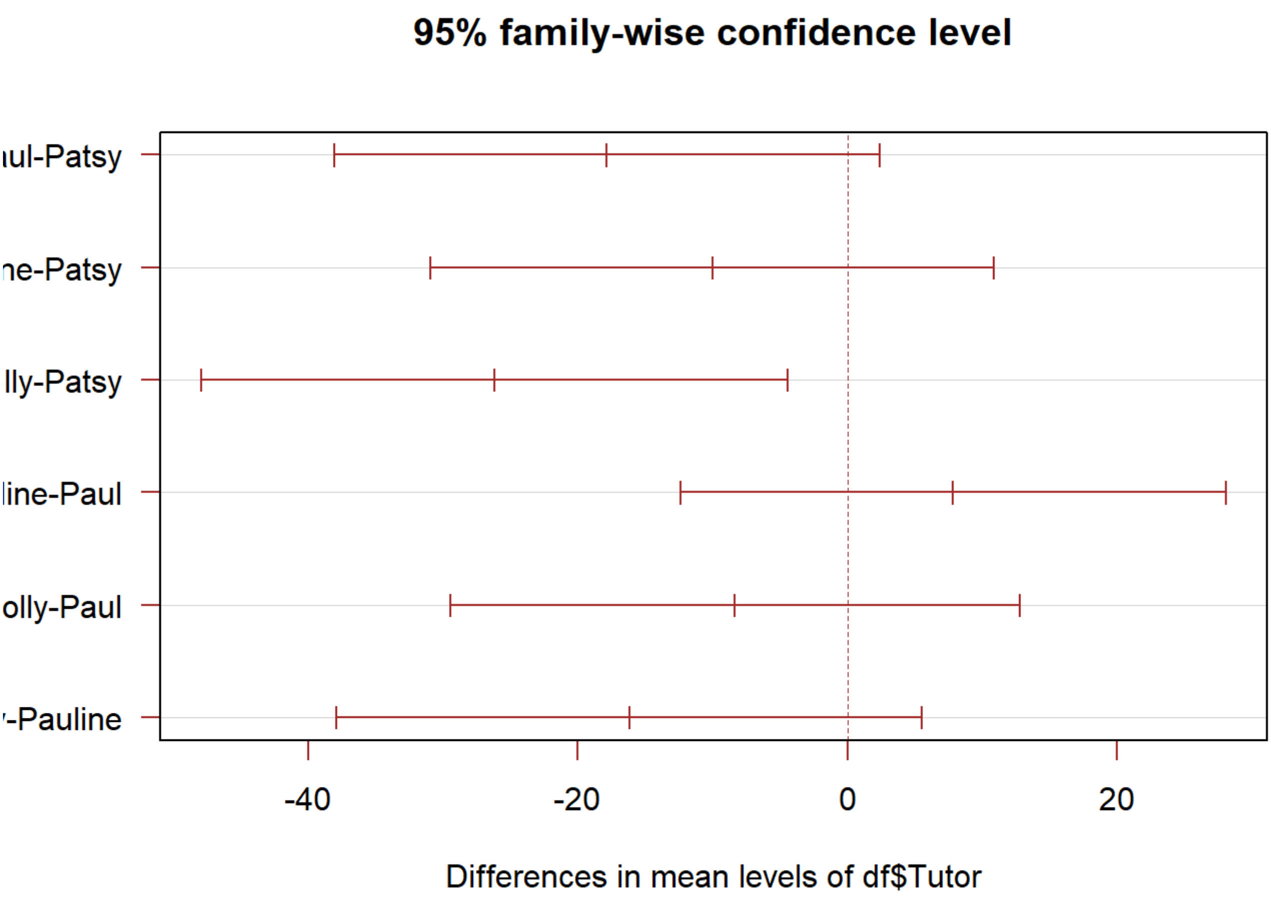
There is significant difference for Patsy and Polly. Other comparisons are not significantly different.

Tukey's method:

```
TUKEY <- TukeyHSD(aov(df$Score ~ df$Tutor), 'df$Tutor', conf.level=0.95)
TUKEY
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = df$Score ~ df$Tutor)
##
## $`df$Tutor`
##                    diff       lwr       upr     p adj
## Paul-Patsy    -17.857143 -38.08811  2.373828 0.0972265
## Pauline-Patsy -10.000000 -30.89446 10.894457 0.5594893
## Polly-Patsy   -26.190476 -47.93812 -4.442836 0.0141442
## Pauline-Paul    7.857143 -12.37383 28.088113 0.7098138
## Polly-Paul     -8.333333 -29.44432 12.777652 0.6994842
## Polly-Pauline -16.190476 -37.93812  5.557164 0.1970113
```

```
# Tuckey test representation :
plot(TUKEY , las=1 , col="brown")
```



The pair Polly-Pastsy (<0.05) is significantly different

```
#install.packages("DescTools")
#library(DescTools)
#DunnettTest(x=df$Score, g=df$Tutor)#,subset="4")
```

Conclusion: The pair Polly-Pastsy (<0.05) is significantly different.

Multiple compare the average score of Patsy's class with all other classes

```r
res = aov(df$Score ~ df$Tutor)
summary(res)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## df$Tutor     3   2477   825.8   4.113 0.0173 *
## Residuals   24   4819   200.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(agricolae)
least = LSD.test(df$Score, df$Tutor, DFerror=24, MSerror=200.8, p.adj="none")
least
```

```
## $statistics
##   MSerror Df     Mean       CV
##     200.8 24 69.64286 20.34723
##
## $parameters
##         test p.ajusted   name.t ntr alpha
##   Fisher-LSD      none df$Tutor   4  0.05
##
## $means
##         df$Score      std r      LCL      UCL Min Max  Q25 Q50  Q75
## Patsy   82.85714 11.12697 7 71.80310 93.91119  70 100 75.0  80 90.0
## Paul    65.00000 16.03567 8 54.65989 75.34011  40  90 57.5  65 72.5
## Pauline 72.85714 12.53566 7 61.80310 83.91119  60  90 60.0  80 80.0
## Polly   56.66667 16.32993 6 44.72693 68.60640  40  80 42.5  55 67.5
##
## $comparison
## NULL
##
## $groups
##         df$Score groups
## Patsy   82.85714      a
## Pauline 72.85714     ab
## Paul    65.00000      b
## Polly   56.66667      b
##
## attr(,"class")
## [1] "group"
```

```r
# group mean: average score for each tutor
tapply(df$Score, df$Tutor, mean)
```

```
##    Patsy     Paul  Pauline    Polly
## 82.85714 65.00000 72.85714 56.66667
```

Let $\mu_1, \mu_2, \mu_3, \mu_4$ denote the average score of Patsy's, Paul's, Pauline and Polly's classed, respectively.

Test 1: compare Polly and Patsy $H_0 : \mu_1 = \mu_4$

Since when the sample sizes are not equal, LSD.test function cannot return the value of LSD.

We can compute LSD by formula $t_{\alpha/2}(N-k)\sqrt{\hat{\sigma}^2(1/n_i + 1/n_j)}$ Where N is the total sample size (N=28), k is the number of groups (k=4)

```
(1-qt(0.05/2, 28-4))*sqrt(200.8*(1/6 + 1/7))
```

```
## [1] 24.15479
```

LSD = 24.15479

```
82.85714 - 56.66667
```

```
## [1] 26.19047
```

$|\bar{y_1} - \bar{y_4}|$ = |82.85714 - 56.66667| = 26.19 > LSD

Hence, we reject the null hypothesis. The average score of Polly's and Patsy's classes are significantly different.

Test 2: compare Pauline and Patsy $H_0 : \mu_1 = \mu_2$

```
(1-qt(0.05/2, 24))*sqrt(200.8*(1/7 + 1/7))
```

```
## [1] 23.20717
```

```
82.85714 - 72.85714
```

```
## [1] 10
```

LSD = 23.207

$|\bar{y_1} - \bar{y_2}|$ = 10 < LSD

Hence, we do not reject the null hypothesis. There is no significant difference between the average score of Polly's class and the average score of Patsy's class.

Test 3: compare Paul and Patsy $H_0 : \mu_1 = \mu_3$

```
(1-qt(0.05/2, 24))*sqrt(200.8*(1/8 + 1/7))
```

```
## [1] 22.47025
```

```
82.85714 - 65.00000
```

```
## [1] 17.85714
```

LSD = 22.47

$|\bar{y_3} - \bar{y_4}|$ = 17.86 < LSD

Hence, we do not reject the null hypothesis. There is no significant difference between the average score of Paul's class and the average score of Patsy's class.

Conclusion:

Average score of Pasty is is significantly different from Polly, but not significantly different from Paul and Pauline.