Home Work Data Analysis 1

Md Shamsuzzaman

09.29.2017

1 Executive Summary

This data set (two-way) is a hypothetical sample of 27 participants who have got three types of treatments (mental, physical and medicinal) towards stress ruction. The participants were categorised into three groups as young, mid and old based on their age. Findings from the histogram, boxplot, q-q plot, and shapiro test, we can see that the data set is normally distributed having no extreme value or outliers. I have my statistical hypothesis to support the analysis of the data set, the null hypothesis iss $H_0 : \alpha_{mental} = \alpha_{physical} = \alpha_{medical}$ and H_a : at least one of them is different. The Anova test reveals the outcome through the P value (less than 0.05) which rejects the null hypothesis and indicate that there are differences in effects levels. It also indicates that there is no relation between the factors ages and treatment. The interaction plot conclude that the effectiveness of the treatments was more on young people rather than mid and old. On the other hand, mental treatment was more effective than physical and medical. At the same time the Pairwise tests reveal that there is no significant pairwise differences between the treatment groups and on the other hand the comparisons between age groups are significant. I have gone through the data set and found replicated values which are not valid for Friedman test. That is why the replicated data have been analysed through their mean value. Finally, from the P values of the Friedman test, I can see that these are almost close to effect level 0.05. But according to the statistical methods, I am accepting the P values (0.04979) as less than 0.05. So, these values indicate that there are no difference in age effect with treatment as same as for methods. On the other hand, the MS.test supports the statement from Friedman test outcomes. Through the statistical analysis, it has been concluded that the effectiveness of the treatments are more on young people rather than mid and old. On the other hand, mental treatment is more effective than physical and medical.

2 Analysis

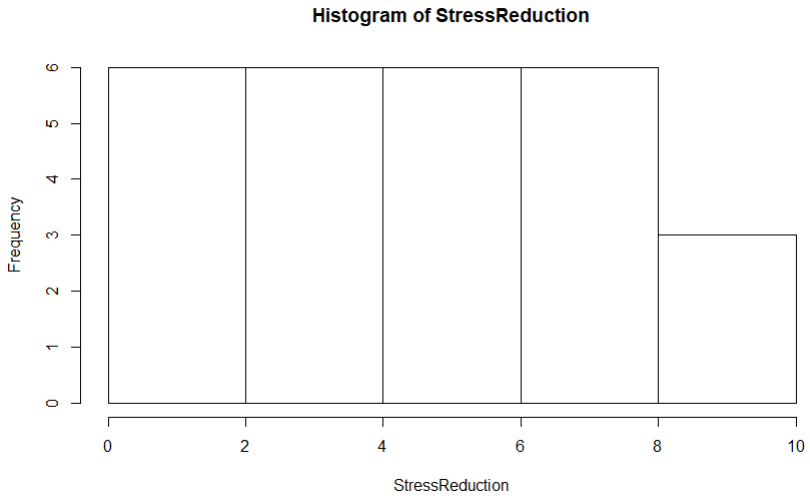

Fig 1. Histogram of the stress reduction data

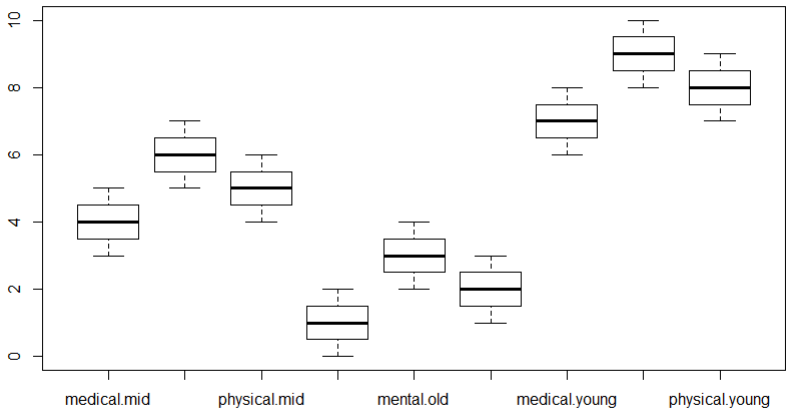

Fig 2. Boxplot of the Two Way data set

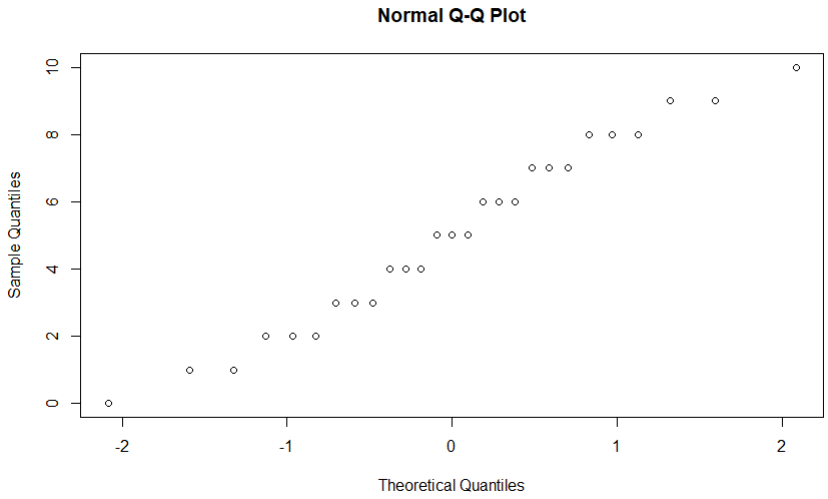

Fig 3. Q-Q plot of stress reduction data set

```
Shapiro-Wilk normality test
data:  StressReduction
W = 0.96814, p-value = 0.5535
```

From the histogram, boxplot, q-q plot, and shapiro test, I can see that the data set is normally distributed having no extreme value or outliers. For the analysis of the data set, I have the null hypothesis $H_0 : \alpha_{mental} = \alpha_{physical} = \alpha_{medical}$ and H_a : at least one of them is different.

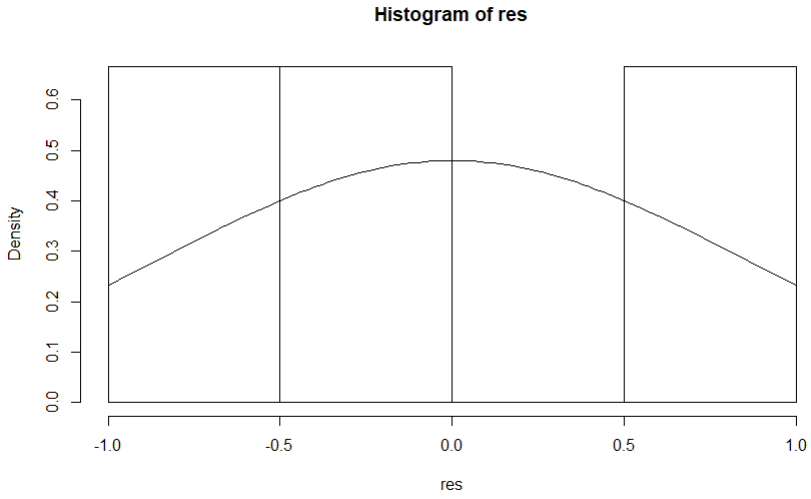

Fig 4. Histogram of the residuals

2.1 Solution of question 1

```
> anova (lm(StressReduction~Treatment+Age))
Analysis of Variance Table

Response: StressReduction
          Df Sum Sq Mean Sq F value    Pr(>F)
Treatment  2     18   9.000      11 0.0004883 ***
Age        2    162  81.000      99     1e-11 ***
Residuals 22     18   0.818
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

P value less than 0.05, of the anova test indicates, I reject null hypothesis. So, I can say that the effects of the levels are different.

2.2 Solution of question 2

```
> anova (lm(StressReduction~Treatment*Age))
Analysis of Variance Table

Response: StressReduction
              Df Sum Sq Mean Sq F value   Pr(>F)
Treatment      2     18       9       9 0.001953 **
Age            2    162      81      81    1e-09 ***
Treatment:Age  4      0       0       0 1.000000
Residuals     18     18       1
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

P value, less than 0.05, of the anova test indicates, there is no interaction between these two factors (age and treatment).

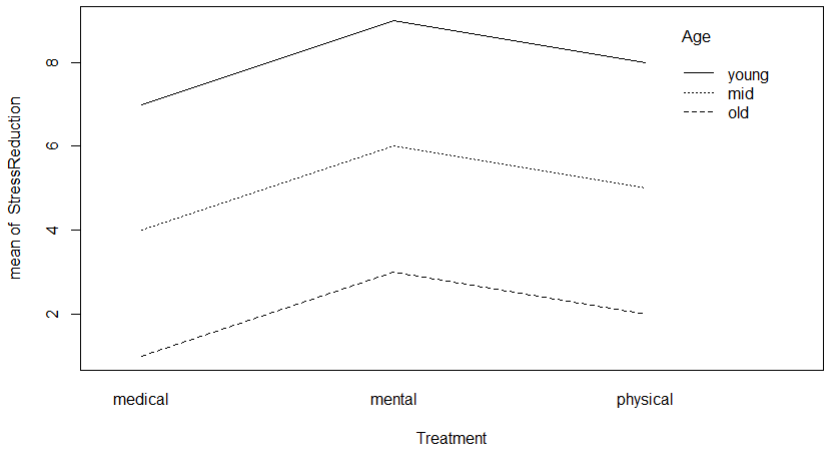

Fig 4. Interaction plot

From the interaction plot (fig. 4), I can conclude that the effectiveness of the treatments is more on young people rather than mid and old. On the other hand, mental treatment is more effective than physical and medical. I can express the effectiveness following below.

Treatment effectiveness on age: young >mid> old

Treatment effectiveness : mental > physical >medical

2.3 Solution of question 3

Pairwise tests are being carried out due to the rejection of null hypothesis.

```
> pairwise.t.test(StressReduction,Treatment, p.adj = 'none')

Pairwise comparisons using t tests with pooled SD

data:  StressReduction and Treatment

         medical mental
mental   0.13    -
physical 0.45    0.45

P value adjustment method: none

> pairwise.t.test(StressReduction,Age, p.adj = 'none')
```

```
Pairwise comparisons using t tests with pooled SD

data:  StressReduction and Age

      mid     old
old   2.5e-05 -
young 2.5e-05 2.3e-10

P value adjustment method: none
```

Pairwise tests reveal that there is no significant pairwise differences between the treatment groups and on the other hand the comparisons between age groups are significant.

2.4 Solution of question 4

Friedman test is applicable for non-replicated block data analysis, but our data contain replication and that is why we have taken the mean value of the replicated values to carry this data.

```
> friedman.test(StressReduction2~Age2|Treatment2,data=two_way1)

Friedman rank sum test

data:  StressReduction2 and Age2 and Treatment2
Friedman chi-squared = 6, df = 2, p-value = 0.04979

> friedman.test(StressReduction2~Treatment2|Age2,data=two_way1)

Friedman rank sum test

data:  StressReduction2 and Treatment2 and Age2
Friedman chi-squared = 6, df = 2, p-value = 0.04979

> MS.test(age, treat, reps=3)
  df MS.test.stat P(Chi.sq>MS)
1  2     11.26667  0.003576633
```

From the P values of the Friedman test, I can see that these are almost close to effect level 0.05. But according to the statistical methods, I am accepting the P values as less than 0.05. So, these values indicate that there are no difference in age effect with treatment as same as for methods. The outcomes of the MS.test confirms the statement from Friedman test outcomes.

3 Conclusion

Here, I can conclude that the effectiveness of the treatments was more on young people rather than mid and old. On the other hand, mental treatment is more effective than physical and medical. Here through this anova analysis, I have learnt to conduct the test for multiple factors and response of a data set.

4 Appendix

4.1 R Code

```
attach(two_way)
str(two_way)
summary(two_way)
hist(StressReduction)
boxplot(StressReduction~Treatment+Age)
shapiro.test(StressReduction)
qqnorm(StressReduction)
anova (lm(StressReduction~Treatment+Age))
anova (lm(StressReduction~Treatment*Age))
interaction.plot(Treatment, Age, StressReduction)
pairwise.t.test(StressReduction,Treatment, p.adj = 'none')
pairwise.t.test(StressReduction,Age, p.adj = 'none')
attach(two_way1)
friedman.test(StressReduction2~Age2|Treatment2,data=two_way1)
friedman.test(StressReduction2~Treatment2|Age2,data=two_way1)
young<-c(10,9,8,9,8,7,8,7,6)
mid<-c(7,6,5,6,5,4,5,4,3)
old<-c(4,3,2,3,2,1,2,1,0)
age<-cbind(young,mid,old)
library(asbio)
treat<-c(rep(1,3),rep(2,3),rep(3,3))
MS.test(age, treatt, reps=3)
```

4.2 Log File

```
> library(readr)
> two_way <- read_csv("/Volumes/SHOHAG/US Semester/Data Analysis by R/HW 3/two-way.csv")
Parsed with column specification:
cols(
  Treatment = col_character(),
```

```
  Age = col_character(),
  StressReduction = col_integer()
)
> View(two_way)
> library(readr)
> two_way1 <- read_csv("C:/Users/RASEL/Desktop/Fall 2017/Data analysis I/Homework/Hw04/two-
Parsed with column specification:
cols(
  Treatment2 = col_character(),
  Age2 = col_character(),
  StressReduction2 = col_integer()
)
> View(two_way1)
> attach(two_way)
> str(two_way)
Classes ?tbl_df?, ?tbl? and 'data.frame': 27 obs. of  3 variables:
 $ Treatment      : chr  "mental" "mental" "mental" "mental" ...
 $ Age            : chr  "young" "young" "young" "mid" ...
 $ StressReduction: int  10 9 8 7 6 5 4 3 2 9 ...
 - attr(*, "spec")=List of 2
  ..$ cols   :List of 3
  .. ..$ Treatment      : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ Age            : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ StressReduction: list()
  .. .. ..- attr(*, "class")= chr  "collector_integer" "collector"
  ..$ default: list()
  .. ..- attr(*, "class")= chr  "collector_guess" "collector"
  ..- attr(*, "class")= chr "col_spec"
> summary(two_way)
  Treatment             Age            StressReduction
 Length:27          Length:27          Min.   : 0
 Class :character   Class :character   1st Qu.: 3
 Mode  :character   Mode  :character   Median : 5
                                       Mean   : 5
                                       3rd Qu.: 7
                                       Max.   :10
> hist(StressReduction)
> boxplot(StressReduction~Treatment+Age)
> shapiro.test(StressReduction)
```

```
Shapiro-Wilk normality test

data:  StressReduction
W = 0.96814, p-value = 0.5535

> qqnorm(StressReduction)
> anova (lm(StressReduction~Treatment+Age))
Analysis of Variance Table

Response: StressReduction
          Df Sum Sq Mean Sq F value    Pr(>F)
Treatment  2     18   9.000      11 0.0004883 ***
Age        2    162  81.000      99     1e-11 ***
Residuals 22     18   0.818
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
> anova (lm(StressReduction~Treatment*Age))
Analysis of Variance Table

Response: StressReduction
              Df Sum Sq Mean Sq F value   Pr(>F)
Treatment      2     18       9       9 0.001953 **
Age            2    162      81      81    1e-09 ***
Treatment:Age  4      0       0       0 1.000000
Residuals     18     18       1
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
> interaction.plot(Treatment, Age, StressReduction)  ################Interaction plot
> pairwise.t.test(StressReduction,Treatment, p.adj = 'none')

Pairwise comparisons using t tests with pooled SD

data:  StressReduction and Treatment

         medical mental
mental   0.13    -
physical 0.45    0.45

P value adjustment method: none
> pairwise.t.test(StressReduction,Age, p.adj = 'none')
```

```
Pairwise comparisons using t tests with pooled SD

data:  StressReduction and Age

      mid     old
old   2.5e-05 -
young 2.5e-05 2.3e-10

P value adjustment method: none
> attach(two_way1)
> friedman.test(StressReduction2~Age2|Treatment2,data=two_way1)

Friedman rank sum test

data:  StressReduction2 and Age2 and Treatment2
Friedman chi-squared = 6, df = 2, p-value = 0.04979

> friedman.test(StressReduction2~Treatment2|Age2,data=two_way1)

Friedman rank sum test

data:  StressReduction2 and Treatment2 and Age2
Friedman chi-squared = 6, df = 2, p-value = 0.04979
> library(asbio)
> young<-c(10,9,8,9,8,7,8,7,6)
> mid<-c(7,6,5,6,5,4,5,4,3)
> old<-c(4,3,2,3,2,1,2,1,0)
> age<-cbind(young,mid,old)
> treat<-c(rep(1,3),rep(2,3),rep(3,3))
> MS.test(age, treat, reps=3)
  df MS.test.stat P(Chi.sq>MS)
1  2     11.26667  0.003576633
```