Practical no 10

Aim: Use of R Markdown and RStudio Cloud (Store mini project in RStudio Cloud)

Theory:

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability

found inside a data set into two parts: systematic factors and random factors. The systematic factors have a

statistical influence on the given data set, while the random factors do not.

Analysts use the ANOVA test to determine the influence that independent variables have on the dependent

variable in a regression study.

The t- and z-test methods developed in the 20th century were used for statistical analysis until 1918, when

Ronald Fisher created the analysis of variance method. ANOVA is also called the Fisher analysis of variance,

and it is the extension of the t- and z-tests. The term became well-known in 1925, after appearing in Fisher's

book, "Statistical Methods for Research Workers". It was employed in experimental psychology and later

expanded to subjects that were more complex.

- Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests.
- A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.
- If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

One-Way ANOVA Versus Two-Way ANOVA

There are two main types of ANOVA: one-way (or unidirectional) and two-way. One-way or two-way refers to the number of independent variables in your analysis of variance test. A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as salary and skill set.

The following practical is performed by ANOVA method i.e., Analysis of Variance on a dataset in the R Studio.

Code:

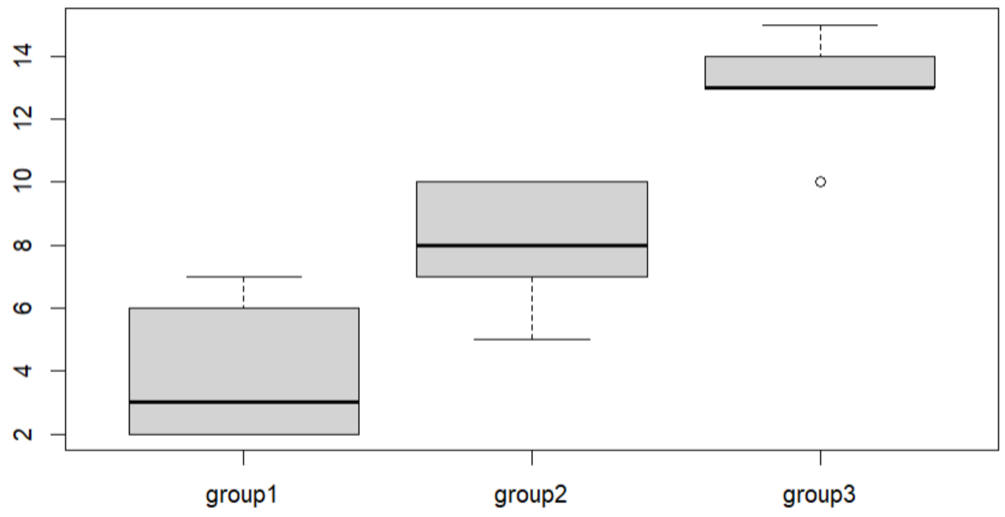
```
group1=c(2,3,7,2,6)
group2=c(10,8,7,5,10)
group3=c(10,13,14,13,15)
cg=data.frame(cbind(group1,group2,group3))
cg
boxplot(cg)
stacked_g=stack(cg)
stacked_g
av=aov(values~ind,data=stacked_g)
summary(av)
g1=c(29,30,31,31,29)
g2=c(28,29,27,30,29)
g3=c(25,28,29,27,29)
cg1=data.frame(cbind(g1,g2,g3))
cg1
stacked_g=stack(cg)
stacked_g
av=aov(values~ind,data=stacked_g)
av1=aov(values~ind,data=stacked_g)
summary(av1)

```

Output:

```
> group1=c(2,3,7,2,6)
> group2=c(10,8,7,5,10)
> group3=c(10,13,14,13,15)
> cg=data.frame(cbind(group1,group2,group3))
> cg
  group1 group2 group3
1      2     10     10
2      3      8     13
3      7      7     14
4      2      5     13
5      6     10     15
> boxplot(cg)
> stacked_g=stack(cg)
> stacked_g
   values    ind
1       2 group1
2       3 group1
3       7 group1
4       2 group1
5       6 group1
6      10 group2
7       8 group2
8       7 group2
9       5 group2
10     10 group2
11     10 group3
12     13 group3
13     14 group3
14     13 group3
15     15 group3
> av=aov(values~ind,data=stacked_g)
> summary(av)
            Df Sum Sq Mean Sq F value   Pr(>F)    
ind          2  203.3   101.7   22.59 8.54e-05 ***
Residuals   12   54.0     4.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> g1=c(29,30,31,31,29)
> g2=c(28,29,27,30,29)
> g3=c(25,28,29,27,29)
> cg1=data.frame(cbind(g1,g2,g3))
> cg1
  g1 g2 g3
1 29 28 25
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
ind          2  203.3   101.7   22.59 8.54e-05 ***
Residuals   12   54.0     4.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> g1=c(29,30,31,31,29)
> g2=c(28,29,27,30,29)
> g3=c(25,28,29,27,29)
> cg1=data.frame(cbind(g1,g2,g3))
> cg1
  g1 g2 g3
1 29 28 25
2 30 29 28
3 31 27 29
4 31 30 27
5 29 29 29
> stacked_g=stack(cg)
> stacked_g
   values    ind
1       2 group1
2       3 group1
3       7 group1
4       2 group1
5       6 group1
6      10 group2
7       8 group2
8       7 group2
9       5 group2
10     10 group2
11     10 group3
12     13 group3
13     14 group3
14     13 group3
15     15 group3
> av=aov(values~ind,data=stacked_g)
> av1=aov(values~ind,data=stacked_g)
> summary(av1)
            Df Sum Sq Mean Sq F value   Pr(>F)    
ind          2  203.3   101.7   22.59 8.54e-05 ***
Residuals   12   54.0     4.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```



Conclusion: Hence we have successfully performed the mini project in R Studio.