

Q. 1 The company is accessing the difference in time to complete the task between two groups of employees . State the hypothesis and do the variance testing (ANOVA) for the given dataset (20)

Group-I: Experience (0–1 year)

Group-II: Experience (1–2 year)

Time taken by Group1: 85,95,100,80,90,97,104,95,88,92,94,99

Time taken by Group2: 83,85,96,92,100,104,94,95,88,90,93,94

Marks Distribution (5 marks each)

- Loading the data and visualization
- Model building
- Model testing
- Inference / Prediction

## Solution

```
Group1=c(83,85,96,92,100,104,94,95,88,90,93,94)
```

```
Group2=c(85,95,100,80,90,97,104,95,88,92,94,99)
```

```
Comb_group<-data.frame(cbind(Group1,Group2))
```

```
Stack_group<-stack(Comb_group)
```

```
Stack_group
```

```
Anova_Result<-aov(values~ind,data=Stack_group)
```

```
Anova_Result
```

```
summary(Anova_Result)
```

Q. 2 Load the iris data from datasets library of R . Do the cluster analysis of species by using Hierarchical clustering technique . (20)

Marks Distribution ( 5 Marks each)

- Loading the data and visualization

- Model building
- Model testing
- Inference / Prediction

## Steps

- 1) Import the data
- 2) Create a scatter plot.
- 3) Normalize the data.
- 4) Calculate Euclidean Distance.
- 5) Create a dendrogram.

## Solution

```
df<-iris
head(df)
plot(Petal.Length~Petal.Width,df)          #Scatter Plot
plot(Sepal.Length~Sepal.Width,df)          #Scatter Plot
df <- na.omit(df)
df <- scale(df)                             #Remove Noise
d <- dist(df1, method = "euclidean")        #calculate distance
hc1 <- hclust(d, method = "complete" )      # plotting dendrogram using complete linkage
plot(hc1)
hc1 <- hclust(d, method = "average" )       # plotting dendrogram
plot(hc1)
```