

SHETH L.U.J. AND SIR M.V. COLLEGE
PRACTICAL NO.(7 TO 9) MODULE 2
Data Analysis with SAS / SPSS /R

AIM:

7) Performing one-way ANOVA using aov() (R).

INPUT:-

```
library(dplyr)

df <- read.csv("IRIS (1).csv")

str(df)

df$species <- as.factor(df$species)

anova_model <- aov(sepal_length ~ species, data = df)

summary(anova_model)
```

OUTPUT:-

```
> library(dplyr)
>
> df <- read.csv("IRIS (1).csv")
>
> str(df)
'data.frame': 150 obs. of 5 variables:
 $ sepal_length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ sepal_width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ petal_length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ petal_width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ species      : chr  "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...
>
> df$species <- as.factor(df$species)
>
> anova_model <- aov(sepal_length ~ species, data = df)
>
> summary(anova_model)
              Df Sum Sq Mean Sq F value Pr(>F)
species        2   63.21   31.606   119.3 <2e-16 ***
Residuals    147   38.96    0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

8) Performing two-way ANOVA using aov() (R).

INPUT:-

```
library(dplyr)

df <- winequality.red

str(df)

df$quality <- as.factor(df$quality)
```

SHETH L.U.J. AND SIR M.V. COLLEGE
PRACTICAL NO.(7 TO 9) MODULE 2
Data Analysis with SAS / SPSS / R

```
df$alcohol_group <- as.factor(ifelse(df$alcohol > median(df$alcohol), "High",  
"Low"))
```

```
anova_model <- aov(fixed.acidity ~ quality * alcohol_group, data = df)
```

```
summary(anova_model)
```

OUTPUT:-

```
> library(dplyr)
>
> df <- winequality.red
>
> str(df)
'data.frame': 1599 obs. of 12 variables:
 $ fixed.acidity : num 7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
 $ volatile.acidity : num 0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...
 $ citric.acid : num 0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
 $ residual.sugar : num 1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
 $ chlorides : num 0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071 ...
 $ free.sulfur.dioxide : num 11 25 15 17 11 13 15 15 9 17 ...
 $ total.sulfur.dioxide : num 34 67 54 60 34 40 59 21 18 102 ...
 $ density : num 0.998 0.997 0.997 0.998 0.998 ...
 $ pH : num 3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...
 $ sulphates : num 0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...
 $ alcohol : num 9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...
 $ quality : int 5 5 5 6 5 5 5 7 5 ...
>
> df$quality <- as.factor(df$quality)
> df$alcohol_group <- as.factor(ifelse(df$alcohol > median(df$alcohol), "High", "Low"))
>
> anova_model <- aov(fixed.acidity ~ quality * alcohol_group, data = df)
>
> summary(anova_model)
              Df Sum Sq Mean Sq F value    Pr(>F)
quality          5      94   18.737    6.348 7.61e-06 ***
alcohol_group     1      29   29.293    9.925  0.00166 **
quality:alcohol_group  5      37    7.423    2.515  0.02813 *
Residuals      1587   4684    2.952
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

9) Conducting Chi-square tests using chisq.test() (R)

INPUT:-

```
library(dplyr)
```

```
df <- read.csv("IRIS (1).csv")
```

```
str(df)
```

SHETH L.U.J. AND SIR M.V. COLLEGE
PRACTICAL NO.(7 TO 9) MODULE 2
Data Analysis with SAS / SPSS /R

```
df$sepal_length_group <- cut(  
  df$sepal_length,  
  breaks = c(4, 5, 6, 8),  
  labels = c("Small", "Medium", "Large")  
)
```

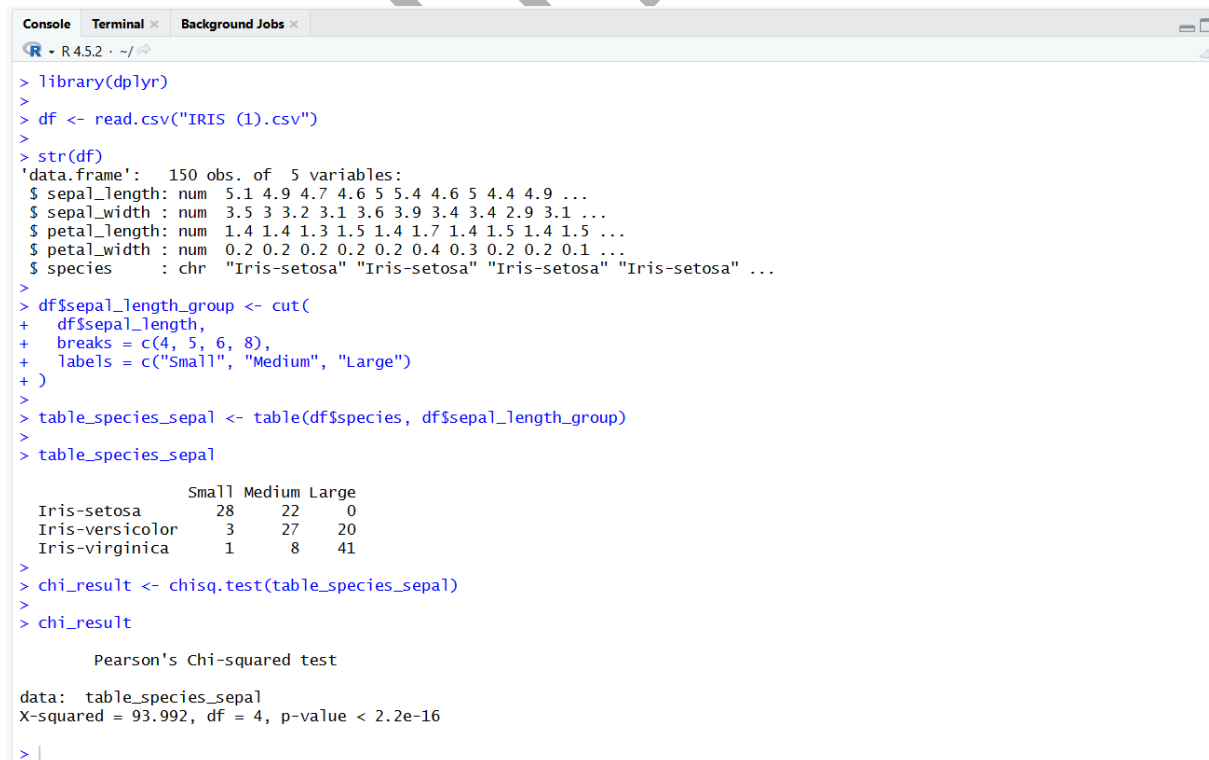
```
table_species_sepal <- table(df$species, df$sepal_length_group)
```

```
table_species_sepal
```

```
chi_result <- chisq.test(table_species_sepal)
```

```
chi_result
```

OUTPUT:-



```
Console Terminal Background Jobs  
R • R 4.5.2 • ~/R  
> library(dplyr)  
>  
> df <- read.csv("IRIS (1).csv")  
>  
> str(df)  
'data.frame': 150 obs. of 5 variables:  
 $ sepal_length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
 $ sepal_width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
 $ petal_length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
 $ petal_width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
 $ species : chr "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...  
>  
> df$sepal_length_group <- cut(  
+ df$sepal_length,  
+ breaks = c(4, 5, 6, 8),  
+ labels = c("Small", "Medium", "Large")  
+ )  
>  
> table_species_sepal <- table(df$species, df$sepal_length_group)  
>  
> table_species_sepal  
      Small Medium Large  
Iris-setosa      28      22      0  
Iris-versicolor   3      27     20  
Iris-virginica    1       8     41  
>  
> chi_result <- chisq.test(table_species_sepal)  
>  
> chi_result  
Pearson's Chi-squared test  
  
data: table_species_sepal  
X-squared = 93.992, df = 4, p-value < 2.2e-16  
> |
```