

**SETH L.U.J. COLLEGE OF ARTS & SIR M.V. COLLEGE OF SCIENCE & COMMERCE**  
**PRACTICAL 07 TO 09**

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

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 # Set working directory
2 setwd("c:/Users/IT-30/Downloads/DATASETS")
3
4 # Load dataset
5 iris_data <- read.csv("Iris.csv")
6
7 # Check structure
8 str(iris_data)
9
10 # One-way ANOVA
11 anova_one-way <- aov(Sepal.Length ~ Species, data = iris_data)
12
13 # ANOVA summary
14 summary(anova_one-way)
15
```

The console shows the output of the code:

```
R - R 4.5.2 - C:/Users/IT-30/Downloads/DATASETS/
> source("c:/Users/IT-30/Desktop/UJALA/S127 RM2 7.R")
'data.frame': 150 obs. of 6 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Sepal.LengthCm: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.WidthCm : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.LengthCm: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.WidthCm : 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" ...
>
```

The Environment pane shows the following objects:

Object	Class	Size
Global Environment		
anova_one-way	List of 13	
College_Marks_Dat...	1000 obs. of 8 variables	
cross_tab_df	4 obs. of 3 variables	
df	6607 obs. of 21 variables	
df_clean	1000 obs. of 10 variables	
Iris	150 obs. of 6 variables	
iris_data	150 obs. of 6 variables	

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

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 # Load dataset
2 college_data <- read.csv("College_Marks_Dataset.csv")
3
4 # Check structure
5 str(college_data)
6
7 # Convert categorical variables to factors
8 college_data$class <- as.factor(college_data$class)
9 college_data$grade <- as.factor(college_data$grade)
10
11 # Two-way ANOVA
12 anova_two-way <- aov(College_Marks ~ Class * Grade, data = college_data)
13
14 # Summary
15 summary(anova_two-way)
16
```

The console shows the output of the code:

```
R - R 4.5.2 - C:/Users/IT-30/Downloads/DATASETS/
> source("c:/Users/IT-30/Desktop/UJALA/S127 M2 8.R")
'data.frame': 1000 obs. of 8 variables:
 $ Student_ID   : chr  "s1000" "s1001" "s1002" "s1003" ...
 $ Name        : chr  "Student_0" "Student_1" "Student_2" "Student_3" ...
 $ Class       : chr  "Commerce" "Commerce" "Science" "Science" ...
 $ SSC_Marks   : int   535 494 542 441 427 520 504 509 499 411 ...
 $ HSC_Marks   : int   452 535 460 483 544 539 573 481 474 450 ...
 $ College_Marks : int   692 551 634 686 569 519 646 504 668 636 ...
 $ Attendance_Percentage: num  84.7 82 92.1 79.3 92 ...
 $ Grade       : chr   "C" "D" "B" "D" ...
>
```

The Environment pane shows the following objects:

Object	Class	Size
Global Environment		
anova_two-way	List of 13	
college_data	1000 obs. of 8 variables	
College_Marks_Dat...	1000 obs. of 8 variables	
cross_tab_df	4 obs. of 3 variables	
df	6607 obs. of 21 variables	
df_clean	1000 obs. of 10 variables	
Iris	150 obs. of 6 variables	
iris_data	150 obs. of 6 variables	

**SETH L.U.J. COLLEGE OF ARTS & SIR M.V. COLLEGE OF SCIENCE & COMMERCE**  
**PRACTICAL 07 TO 09**

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

The screenshot shows the RStudio interface with a script editor on the left containing the following code:

```
1 # Set working directory
2 setwd("C:/Users/IT-30/Downloads/DATASETS")
3
4 # Load dataset
5 emp_data <- read.csv("employee_salary_dataset.csv")
6
7 # check structure (optional)
8 str(emp_data)
9
10 # Create contingency table between Department and Gender
11 emp_table <- table(emp_data$Department, emp_data$Gender)
12
13 # Display the table to check counts
14 print(emp_table)
15
16 # Perform Fisher's Exact Test (safe for small samples)
17 fisher_result <- fisher.test(emp_table)
```

The console on the right shows the output of the script:

```
$ Education_Level : chr "Master" "Bachelor" "High School" "Phd" ...
$ Age             : int  53 25 51 44 36 50 57 34 53 28 ...
$ Gender          : chr  "Female" "Female" "Female" "Male" ...
$ City            : chr  "Delhi"  "Bangalore" "Hyderabad" "Delhi" ...
$ Monthly_salary  : int  111416 95271 69064 95091 132450 65818 70525 44830 42429 31893
...
```

Below the console output, the contingency table is displayed:

	Female	Male
Finance	6	4
HR	3	4
IT	5	5
Marketing	7	6
Operations	6	4

Below the table, the text "Fisher's Exact Test for Count Data" is displayed.

The screenshot shows the RStudio interface with a script editor on the left containing the following code:

```
1 # Set working directory
2 setwd("C:/Users/IT-30/Desktop/UJALA/S127 M2 9.R")
3
4 # Load dataset
5 emp_data <- read.csv("employee_salary_dataset.csv")
6
7 # check structure (optional)
8 str(emp_data)
9
10 # Create contingency table between Department and Gender
11 emp_table <- table(emp_data$Department, emp_data$Gender)
12
13 # Display the table to check counts
14 print(emp_table)
15
16 # Perform Chi-square test
17 chisq_result <- chisq.test(emp_table)
```

The console on the right shows the output of the script:

```
$ EmployeeID      : int  1 2 3 4 5 6 7 8 9 10 ...
$ Name            : chr  "Employee_1" "Employee_2" "Employee_3" "Employee_4" ...
$ Department       : chr  "Marketing" "Operations" "IT" "Operations" ...
$ Experience_Years : int  15 7 12 8 15 3 14 17 4 18 ...
$ Education_Level  : chr  "Master" "Bachelor" "High School" "Phd" ...
$ Age             : int  53 25 51 44 36 50 57 34 53 28 ...
$ Gender          : chr  "Female" "Female" "Female" "Male" ...
$ City            : chr  "Delhi"  "Bangalore" "Hyderabad" "Delhi" ...
$ Monthly_salary   : int  111416 95271 69064 95091 132450 65818 70525 44830 42429 31893
...
```

Below the console output, the contingency table is displayed:

	Female	Male
Finance	6	4
HR	3	4
IT	5	5
Marketing	7	6
Operations	6	4

Below the table, the text "Fisher's Exact Test for Count Data" is displayed.

Below the console output, the results of the Chi-square test are displayed:

```
data: emp_table
p-value = 0.9711
alternative hypothesis: two.sided
```