

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 7

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

#CODE:

Practical 7: One-Way ANOVA

```
print("Siya Poipkar S105")
```

```
mall_data <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Mall_Customers.csv")
```

```
str(mall_data)
```

```
colnames(mall_data) <- c(
```

```
  "CustomerID", "Gender", "Age", "AnnualIncome", "SpendingScore"
```

```
)
```

```
mall_data$AnnualIncome <- as.numeric(mall_data$AnnualIncome)
```

```
summary(mall_data$AnnualIncome)
```

```
mall_data$Income_Group <- cut(
```

```
  mall_data$AnnualIncome,
```

```
  breaks = c(15, 40, 70, 137),
```

```
  labels = c("Low", "Medium", "High"),
```

```
  include.lowest = TRUE
```

```
)
```

```
mall_data$Income_Group <- droplevels(mall_data$Income_Group)
```

```
table(mall_data$Income_Group)
```

```
anova_one <- aov(SpendingScore ~ Income_Group, data = mall_data)
```

```
summary(anova_one)
```

#OUTPUT:

The screenshot shows the RStudio interface with the following content:

Console:

```
R - R4.5.2 - /...
> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
> mall_data <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Mall_Customers.csv")
> str(mall_data)
'data.frame':   200 obs. of  5 variables:
 $ CustomerID   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Gender       : chr  "Male" "Male" "Female" "Female" ...
 $ Age         : int  19 21 20 23 31 22 35 23 64 30 ...
 $ Annual.Income.k.. : int  15 15 16 16 17 17 18 18 19 19 ...
 $ Spending.Score..1.100.: int  39 81 6 77 40 76 6 94 3 72 ...
> colnames(mall_data) <- c(
+   "CustomerID", "Gender", "Age", "AnnualIncome", "SpendingScore"
+ )
> mall_data$AnnualIncome <- as.numeric(mall_data$AnnualIncome)
> summary(mall_data$AnnualIncome)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 15.00  41.50   61.50  60.56   78.00  137.00
> mall_data$Income_Group <- cut(
+   mall_data$AnnualIncome,
+   breaks = c(15, 40, 70, 137),
+   labels = c("Low", "Medium", "High"),
+   include.lowest = TRUE
+ )
> mall_data$Income_Group <- droplevels(mall_data$Income_Group)
> table(mall_data$Income_Group)
   Low Medium  High
    50     76    74
> anova_one <- aov(SpendingScore ~ Income_Group, data = mall_data)
> summary(anova_one)

            Df Sum Sq Mean Sq F value Pr(>F)
Income_Group  2     44    21.9   0.032  0.968
Residuals    197 132660    673.4
```

Environment:

Object	Class	Attributes
AI_Impact_o...	List of 13	
anova_one	List of 13	
anova_result	List of 13	
anova_two	List of 13	
Cardiovascu...	1000 obs. of 14 variables	
category_df	4 obs. of 2 variables	
chisq_result	List of 9	
Cleaned_BMW...	99 obs. of 11 variables	
Cleaned_Car...	199 obs. of 10 variables	
College_Mar...	1000 obs. of 8 variables	
data	9 obs. of 2 variables	
data2	8 obs. of 3 variables	
data7	200 obs. of 5 variables	
df	99 obs. of 12 variables	
mall_data	200 obs. of 6 variables	
placement	num [1:3, 1:2] 20 15 25 10 15...	
sales_data	1000 obs. of 14 variables	
t_test_one	List of 10	
t_test_pair...	List of 10	
t_test_two	List of 10	
category.co...	'table' int [1:4(1d)] 268 246 ...	
cross_tab	'table' int [1:20, 1:3] 0 0 15...	

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 7

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

#CODE:

Practical 8: Two-Way ANOVA

print("Siya Poipkar S105")

ai_data <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/AI_Impact_on_Jobs_2030.csv")

str(ai_data)

colnames(ai_data) <- gsub(" ", "_", colnames(ai_data))

ai_data\$Average_Salary <- as.numeric(ai_data\$Average_Salary)

ai_data\$Education_Level <- as.factor(ai_data\$Education_Level)

ai_data\$Risk_Category <- as.factor(ai_data\$Risk_Category)

ai_data\$Education_Level <- droplevels(ai_data\$Education_Level)

ai_data\$Risk_Category <- droplevels(ai_data\$Risk_Category)

table(ai_data\$Education_Level)

table(ai_data\$Risk_Category)

anova_two <- aov(Average_Salary ~ Education_Level * Risk_Category, data = ai_data)

summary(anova_two)

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 7

#OUTPUT:

```
R - R4.5.2 - ~/R
> # Practical 8: Two-Way ANOVA
> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
> ai_data <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/AI_Impact_on_Jobs_2030.csv")
> str(ai_data)
'data.frame':   3000 obs. of  18 variables:
 $ Job_Title       : chr  "Security Guard" "Research Scientist" "Construction Worker" "Software Engineer" ...
 $ Average_Salary  : int  45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
 $ Years_Experience: int  28 20 2 13 22 11 23 12 12 2 ...
 $ Education_Level : chr  "Master's" "PhD" "High School" "PhD" ...
 $ AI_Exposure_Index: num  0.18 0.62 0.86 0.39 0.52 0.29 0.67 0.2 0.3 0.01 ...
 $ Tech_Growth_Factor: num  1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
 $ Automation_Probability_2030: num  0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
 $ Risk_Category   : chr  "High" "Low" "High" "Medium" ...
 $ Skill1_1        : num  0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
 $ Skill1_2        : num  0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
 $ Skill1_3        : num  0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ Skill1_4        : num  0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ Skill1_5        : num  0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ Skill1_6        : num  0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
 $ Skill1_7        : num  0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
 $ Skill1_8        : num  0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
 $ Skill1_9        : num  0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
 $ Skill1_10       : num  0 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
> colnames(ai_data) <- gsub(" ", "", colnames(ai_data))
> ai_data$Average_Salary <- as.numeric(ai_data$Average_Salary)
> ai_data$Education_Level <- as.factor(ai_data$Education_Level)
> ai_data$Risk_Category <- as.factor(ai_data$Risk_Category)
> ai_data$Education_Level <- droplevels(ai_data$Education_Level)
> ai_data$Risk_Category <- droplevels(ai_data$Risk_Category)
> table(ai_data$Education_Level)
```

```
R - R4.5.2 - ~/R
> colnames(ai_data) <- gsub(" ", "", colnames(ai_data))
> ai_data$Average_Salary <- as.numeric(ai_data$Average_Salary)
> ai_data$Education_Level <- as.factor(ai_data$Education_Level)
> ai_data$Risk_Category <- as.factor(ai_data$Risk_Category)
> ai_data$Education_Level <- droplevels(ai_data$Education_Level)
> ai_data$Risk_Category <- droplevels(ai_data$Risk_Category)
> table(ai_data$Education_Level)

Bachelor's High School  Master's  PhD
765                    784        716

> table(ai_data$Risk_Category)

High  Low Medium
740   739  1521

> anova_two <- aov(Average_Salary ~ Education_Level * Risk_Category, data = ai_data)
> summary(anova_two)

              Df Sum Sq Mean Sq F value Pr(>F)
Education_Level      3  3.855e+09  1.285e+09  1.076  0.3581
Risk_Category        2  4.985e+09  2.493e+09  2.086  0.1243
Education_Level:Risk_Category  6  1.304e+10  2.173e+09  1.819  0.0915
Residuals          2988  3.570e+12  1.195e+09

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
```

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 7

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

#CODE:

Practical 9: Chi-Square Test

`print("Siya Poipkar S105")`

`heart_data <-`

`read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cardiovascular_Disease_Dataset.csv")`

`str(heart_data)`

`colnames(heart_data) <- gsub(" ", "_", colnames(heart_data))`

`heart_data$gender <- as.factor(heart_data$gender)`

`heart_data$target <- as.factor(heart_data$target)`

`heart_data$gender <- droplevels(heart_data$gender)`

`heart_data$target <- droplevels(heart_data$target)`

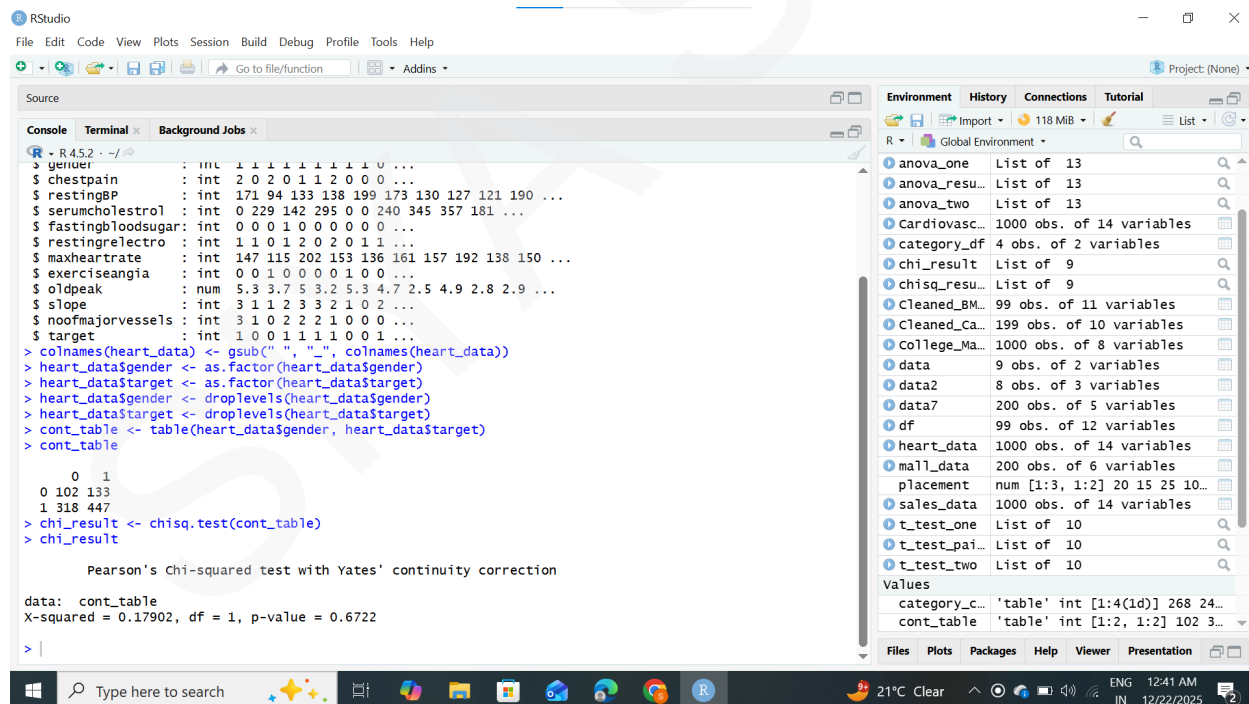
`cont_table <- table(heart_data$gender, heart_data$target)`

`cont_table`

`chi_result <- chisq.test(cont_table)`

`chi_result`

#OUTPUT:



SIYA POIPKAR S105