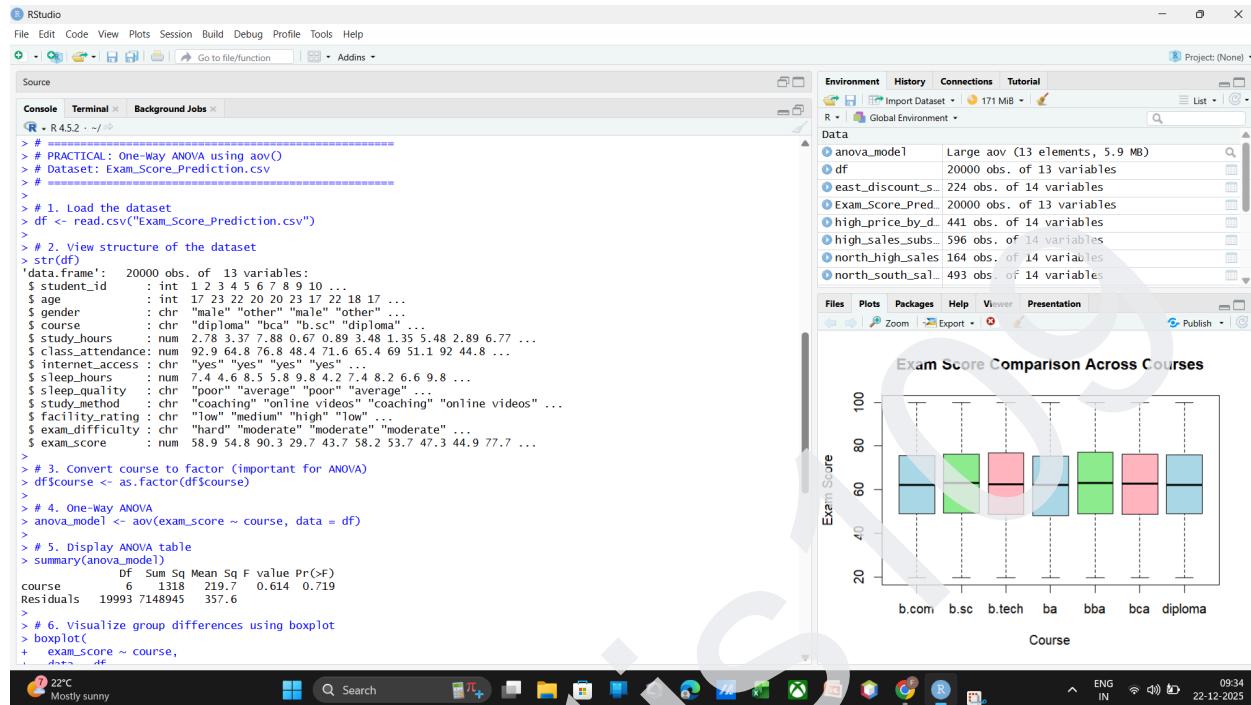


**SHETH L.U.J AND SIR M.V COLLEGE**  
**PRACTICAL NO - 7,8,9 MODULE 2**  
**SUBJECT - DATA ANALYSIS**

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

**OUTPUT -**



The screenshot shows the RStudio interface with the following details:

- Console:** Displays the R session history, including code for reading a CSV file, creating a data frame, and performing a one-way ANOVA.
- Data View:** Shows a list of objects in the global environment, including the generated data frame and the ANOVA model.
- Plots:** A boxplot titled "Exam Score Comparison Across Courses" is displayed. The y-axis is labeled "Exam Score" and ranges from 0 to 100. The x-axis is labeled "Course" and lists seven categories: b.com, b.sc, b.tech, ba, bba, bca, and diploma. Each category has a boxplot showing the distribution of exam scores.
- Environment:** Shows the current workspace structure.
- Session:** Includes information like the R version (R 4.5.2), memory usage (171 MB), and system details (Windows 10, ENG IN, 09:34, 22-12-2025).

```

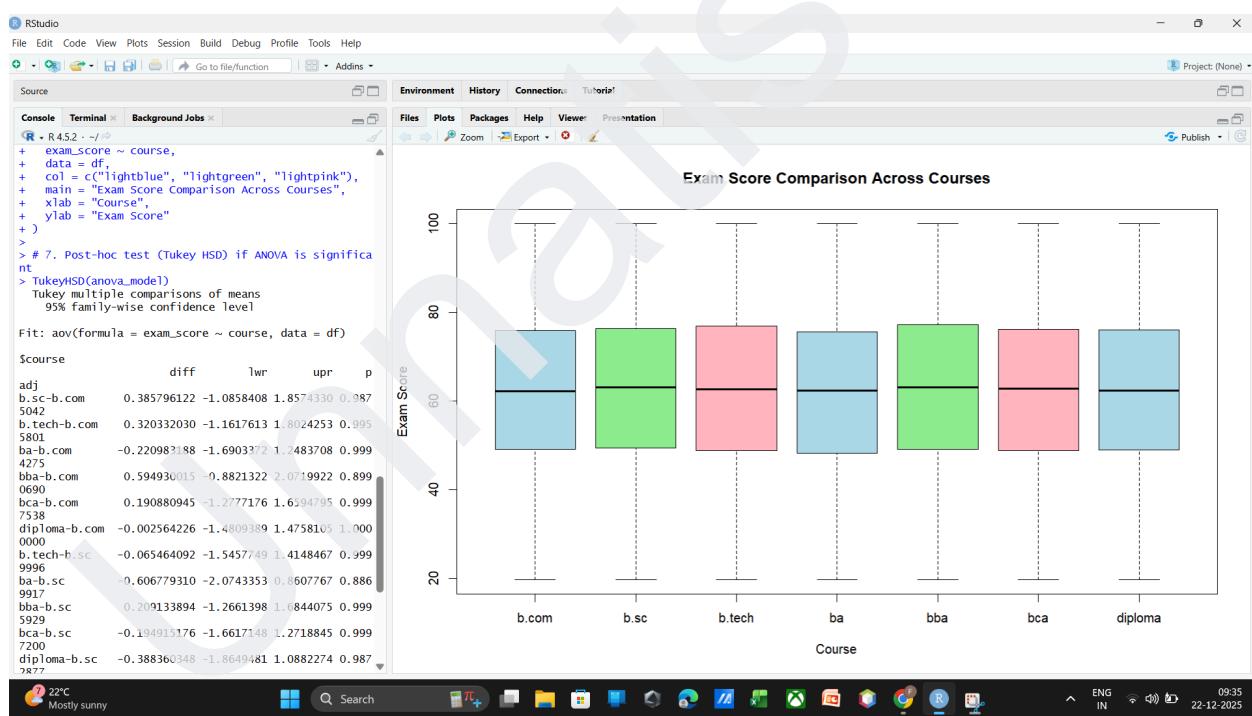
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
R 4.5.2 - /-
> # PRACTICAL: One-way ANOVA using aov()
> # Dataset: Exam_Score_Prediction.csv
> #
> #
> # 1. Load the dataset
> df <- read.csv("Exam_Score_Prediction.csv")
> #
> # 2. View structure of the dataset
> str(df)
'data.frame': 20000 obs. of 13 variables:
 $ student_id : int 1 2 3 4 5 6 7 8 9 10 ...
 $ age         : int 17 23 22 20 20 23 17 22 18 17 ...
 $ gender      : chr "male" "other" "male" "other" ...
 $ course      : chr "diploma" "bca" "b.sc" "diploma" ...
 $ study_hours : num 2.78 3.37 7.88 0.67 0.89 3.48 1.35 5.48 2.89 6.77 ...
 $ class_attendance: num 92.9 64.8 76.8 48.4 71.6 65.4 69 51.1 92 44.8 ...
 $ internet_access: chr "yes" "yes" "yes" "yes" ...
 $ sleep_hours : num 7.4 4.6 8.5 5.8 9.8 4.2 7.4 8.2 6.6 9.8 ...
 $ sleep_quality: chr "poor" "average" "poor" "average" ...
 $ study_method: chr "coaching" "online videos" "coaching" "online videos" ...
 $ facility_rating: chr "low" "medium" "high" "low" ...
 $ exam_difficulty: chr "hard" "moderate" "moderate" "moderate" ...
 $ exam_score   : num 58.9 54.8 90.3 29.7 43.7 58.2 53.7 47.3 44.9 77.7 ...
> #
> # 3. Convert course to factor (important for ANOVA)
> df$course <- as.factor(df$course)
> #
> # 4. One-way ANOVA
> anova_model <- aov(exam_score ~ course, data = df)
> #
> # 5. Display ANOVA table
> summary(anova_model)
    DF Sum Sq Mean Sq F value Pr(>F)
course       6   1318   219.7  0.614  0.719
Residuals 19993 7148945   357.6
> #
> # 6. Visualize group differences using boxplot
> boxplot(
+   exam_score ~ course,
+   data = df)

```

**NAME - UNNATI RATHOD**  
**ROLL NO - S109**

**SHETH L.U.J AND SIR M.V COLLEGE**  
**PRACTICAL NO - 7,8,9 MODULE 2**  
**SUBJECT - DATA ANALYSIS**

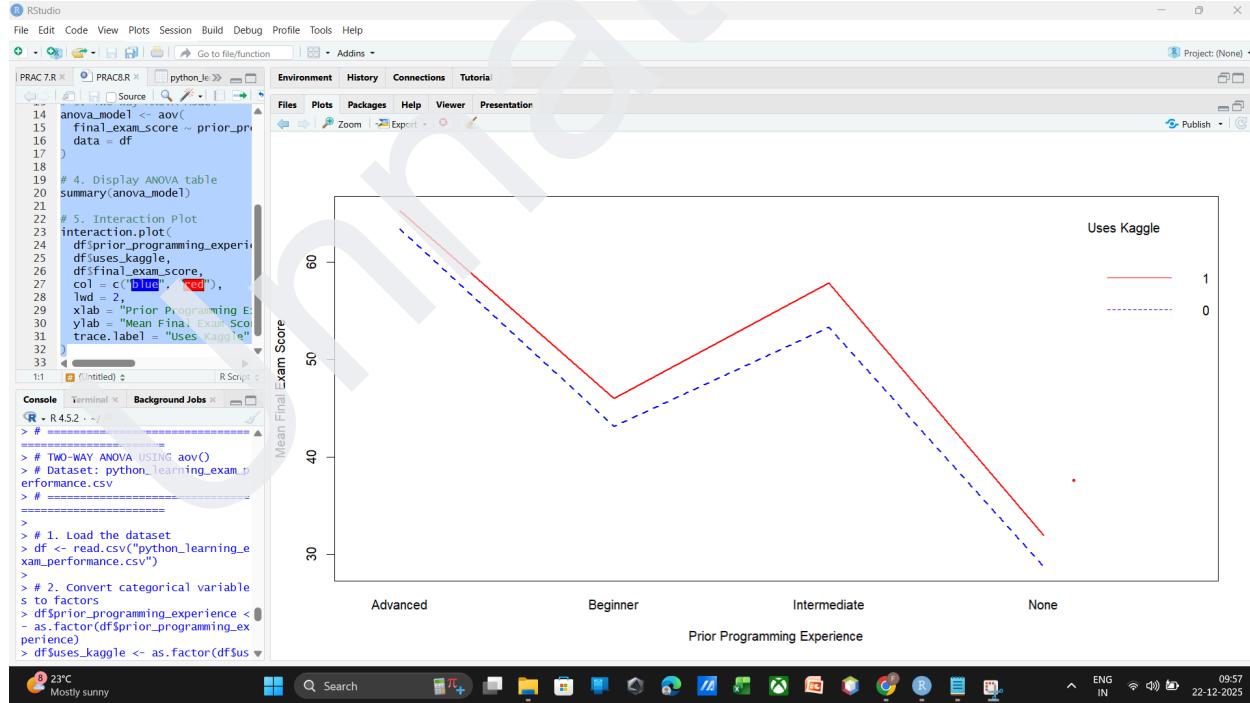
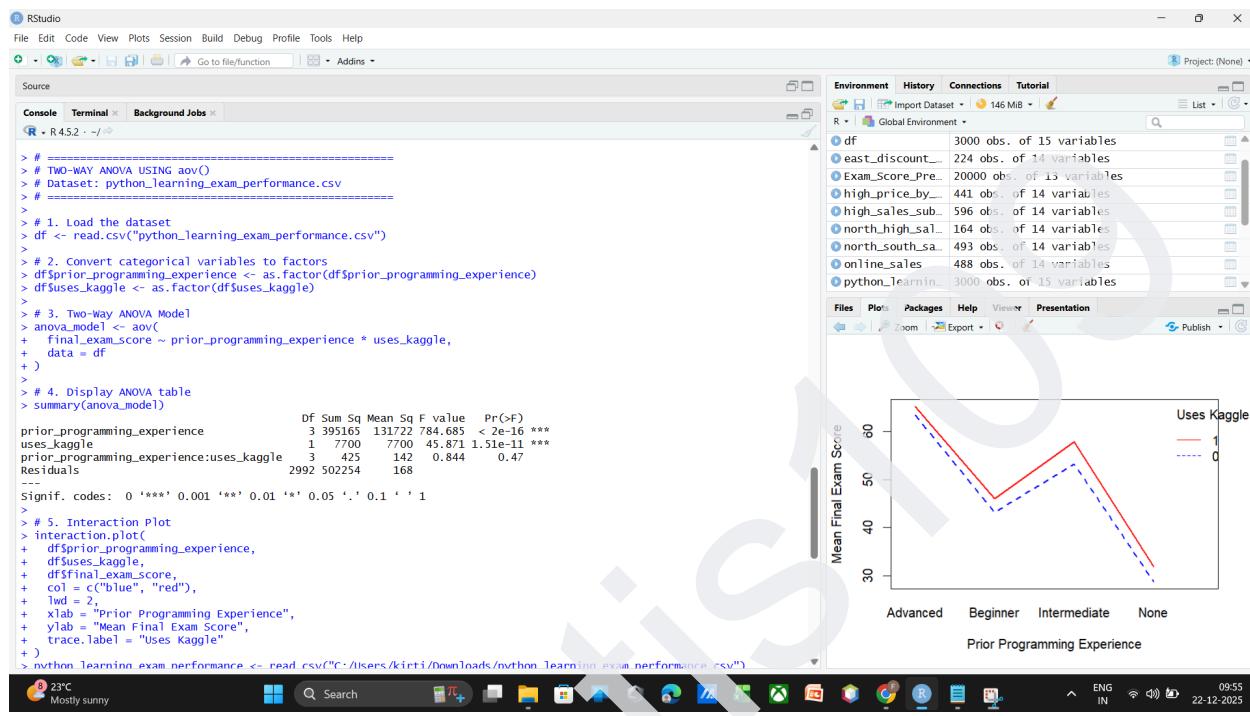
RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Console Terminal Background Jobs  
R 4.5.2 - / ~  
+ exam\_score ~ course,  
data = df,  
col = c("lightblue", "lightgreen", "lightpink"),  
main = "Exam Score Comparison Across Courses",  
xlab = "Course",  
ylab = "Exam Score"  
+ )  
>  
> # 7. Post-hoc test (Tukey HSD) if ANOVA is significant  
> TukeyHSD(anova\_model)  
Tukey multiple comparisons of means  
95% family-wise confidence level  
Fit: aov(formula = exam\_score ~ course, data = df)  
\$course  
diff lwr upr p adj  
b.sc-b.com 0.385796122 -1.0858408 1.8574330 0.9875042  
b.tech-b.com 0.320332030 -1.1617613 1.8024253 0.9955801  
ba-b.com -0.220983188 -1.6903372 1.2483708 0.9994275  
bba-b.com 0.594930015 -0.8821322 2.0719922 0.8996090  
bca-b.com 0.190880945 -1.2777176 1.6594795 0.9997538  
diploma-b.com -0.002564226 -1.4809384 1.4758105 1.0000000  
b.tech-b.sc -0.065464092 -1.5457749 1.4148467 0.9999996  
ba-b.sc -0.60679310 -2.0743353 0.8607767 0.889917  
bba-b.sc 0.201133894 -1.2661398 1.844075 0.9999790  
bca-b.sc -0.194915176 -1.6617148 1.2718845 0.9997200  
diploma-b.sc -0.388360348 -1.8649481 1.0882274 0.9872877  
ba-b.tech -0.541315218 -2.0193565 0.9367261 0.9341485  
bba-b.tech 0.274597985 -1.2111065 1.7603024 0.9981490  
bca-b.tech -0.129451085 -1.6067414 1.3478392 0.9999757  
diploma-b.tech -0.322896256 -1.8099053 1.1641131 0.9954655  
bba-ba 0.815913204 -0.6570832 2.2889096 0.6606339  
bca-ba 0.411864134 -1.0526453 1.8763733 0.9819820  
diploma-ba 0.218418962 -1.2558935 1.6927315 0.9994750  
bca-bba -0.404049070 -1.8762918 1.0681937 0.9842126  
diploma-bba -0.597494242 -2.0794892 0.8845007 0.8986431  
diploma-bca -0.193445172 -1.6670047 1.2801144 0.9997391  
> |



**SHETH L.U.J AND SIR M.V COLLEGE**  
**PRACTICAL NO - 7,8,9 MODULE 2**  
**SUBJECT - DATA ANALYSIS**

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

OUTPUT -

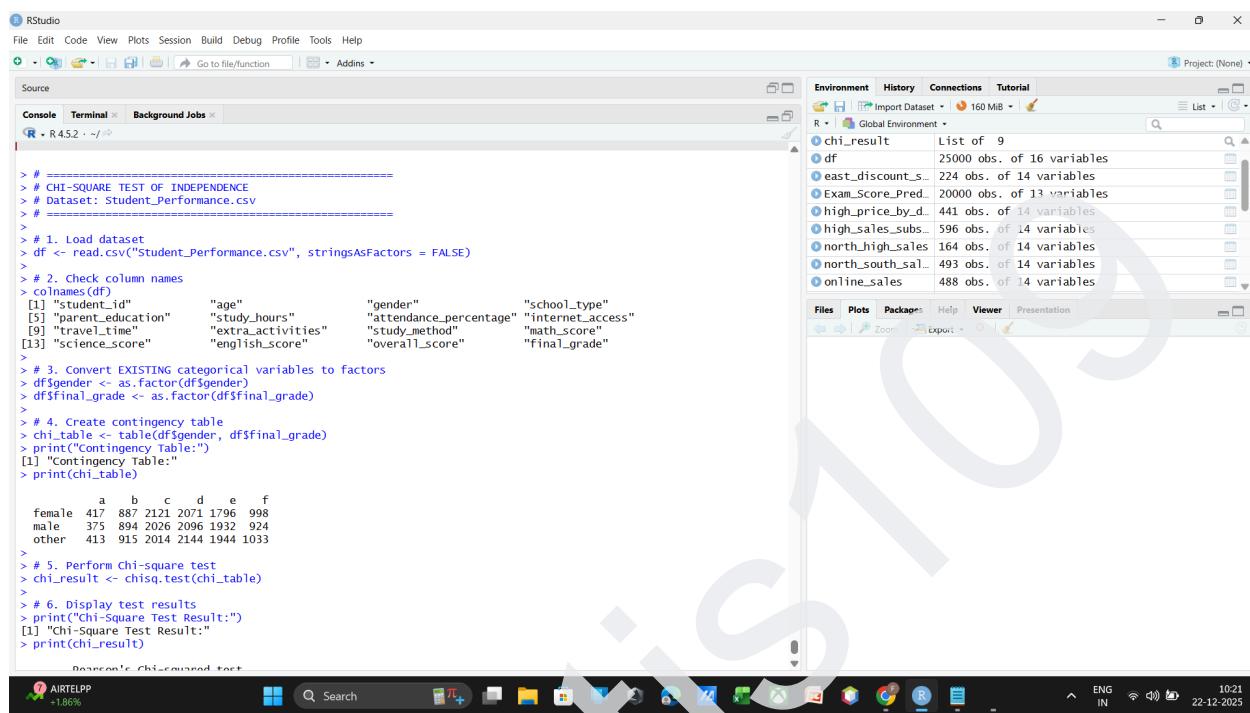


**SHETH L.U.J AND SIR M.V COLLEGE**  
**PRACTICAL NO - 7,8,9 MODULE 2**  
**SUBJECT - DATA ANALYSIS**

**AIM - 9**

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

**OUTPUT -**



```

> # =====
> # CHI-SQUARE TEST OF INDEPENDENCE
> # Dataset: Student_Performance.csv
> #
> #
> # 1. Load dataset
> df <- read.csv("Student_Performance.csv", stringsAsFactors = FALSE)
>
> # 2. Check column names
> colnames(df)
[1] "student_id"          "age"           "gender"         "school_type"
[5] "parent_education"    "study_hours"     "attendance_percentage" "internet_access"
[9] "travel_time"         "extra_activities" "study_method"      "math_score"
[13] "science_score"       "english_score"   "overall_score"    "final_grade"
>
> # 3. Convert EXISTING categorical variables to factors
> df$gender <- as.factor(df$gender)
> df$final_grade <- as.factor(df$final_grade)
>
> # 4. Create contingency table
> chi_table <- table(df$gender, df$final_grade)
> print("Contingency Table:")
[1] "Contingency Table:"
> print(chi_table)

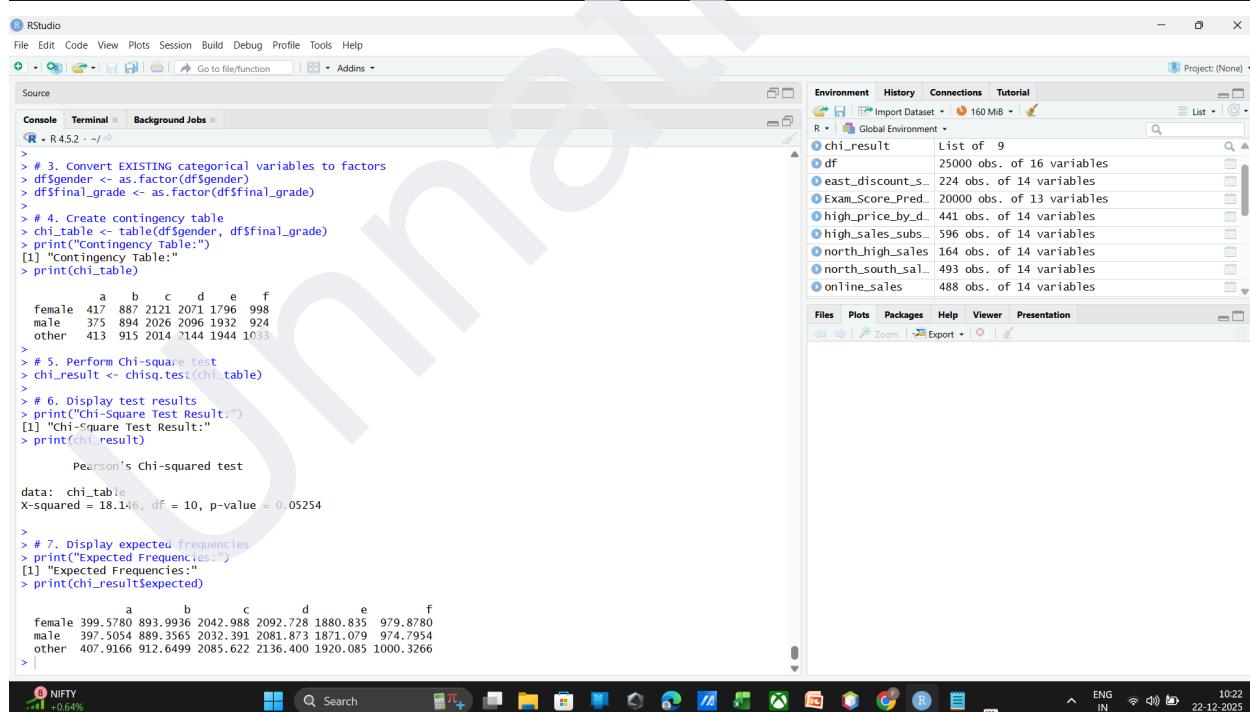
Pearson's Chi-squared test

data: chi_table
X-squared = 18.140, df = 10, p-value = 0.05254

> # 5. Perform Chi-square test
> chi_result <- chisq.test(chi_table)

> # 6. Display test results
> print("Chi-Square Test Result:")
[1] "Chi-Square Test Result:"
> print(chi_result)

```



```

Pearson's Chi-squared test

data: chi_table
X-squared = 18.140, df = 10, p-value = 0.05254

> # 7. Display expected frequencies
> print("Expected Frequencies:")
[1] "Expected Frequencies:"
> print(chi_result$expected)

      a      b      c      d      e      f 
female 399.5780 893.9936 2042.988 2092.728 1880.835 979.8798 
male   397.5054 889.3565 2032.391 2081.873 1871.079 974.7954 
other  407.9166 912.6499 2085.622 2136.400 1920.085 1000.3266 

```

**NAME - UNNATI RATHOD**  
**ROLL NO - S109**