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

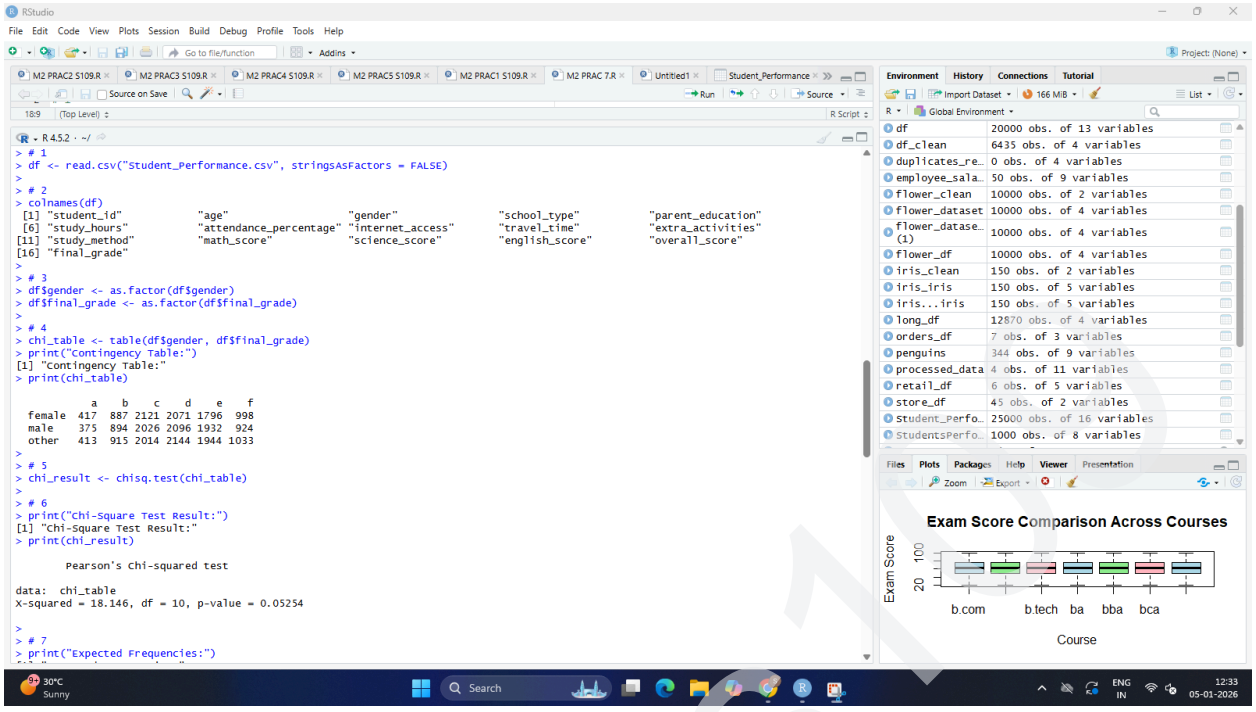
AIM- 7 Performing one-way ANOVA using aov() (R).

OUTPUT-

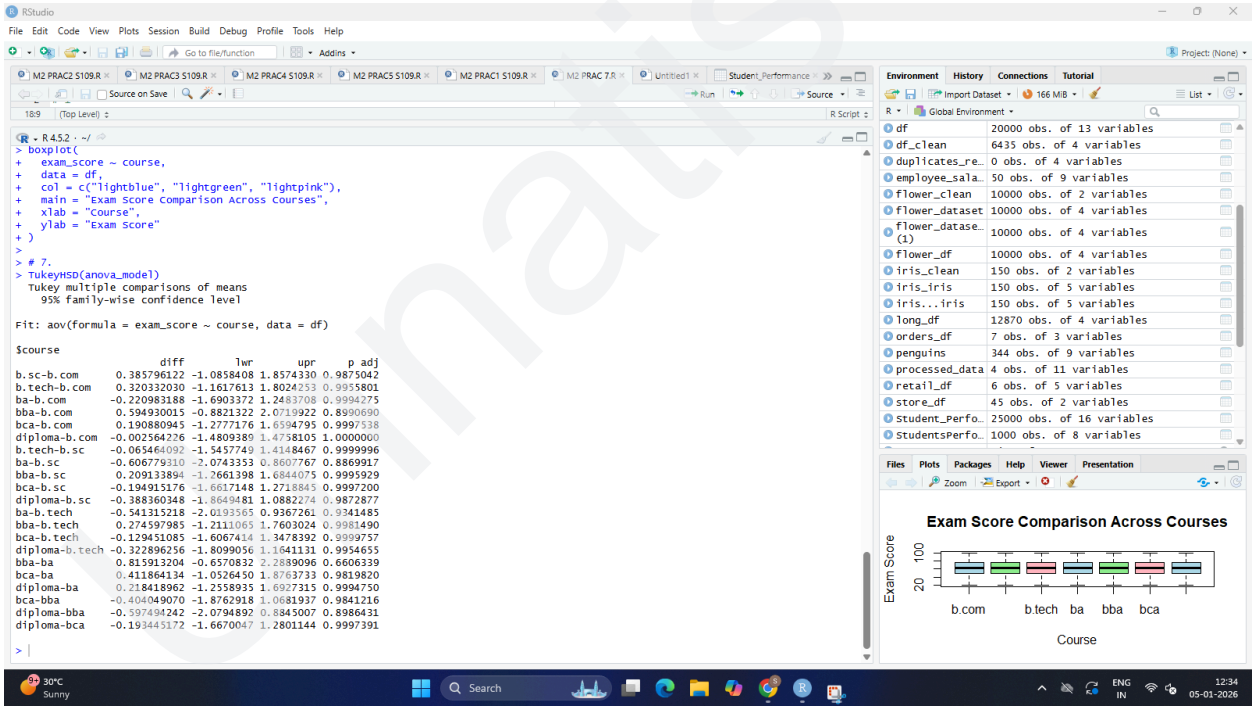


NAME - UNNATI RATHOD
ROLL NO - S109

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



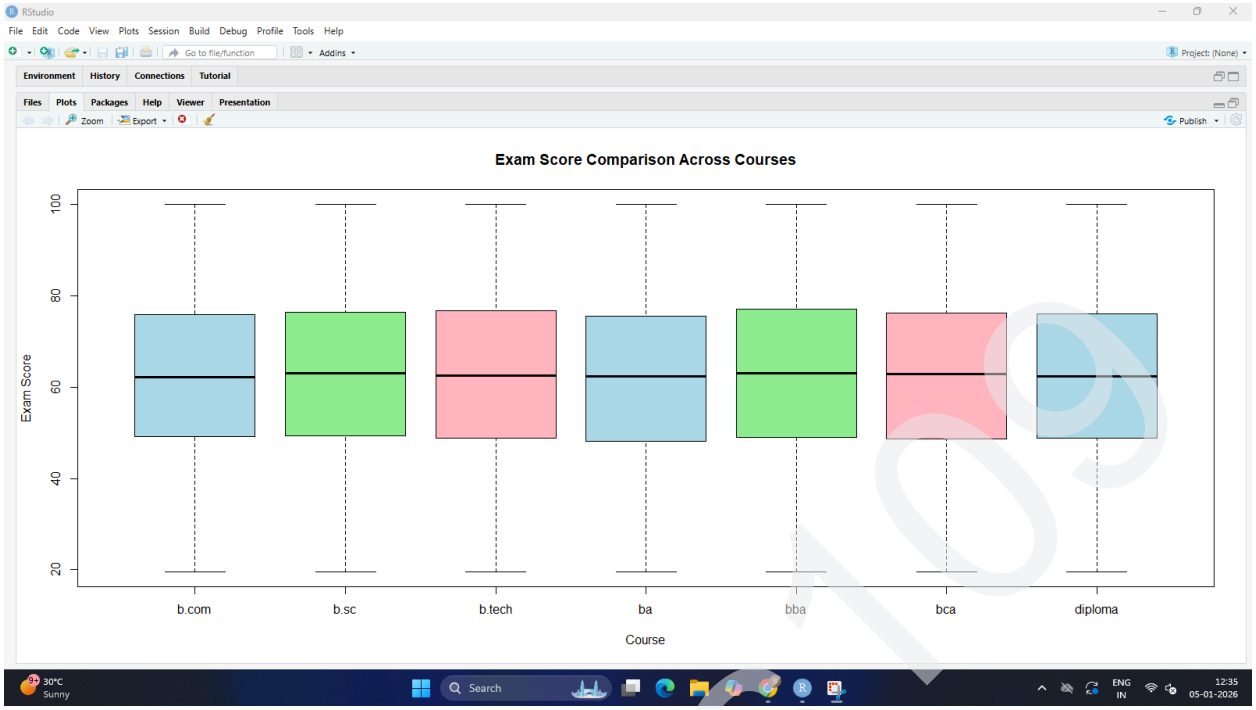
```
> # 1
> df <- read.csv("Student_Performance.csv", stringsAsFactors = FALSE)
>
> # 2
> colnames(df)
 [1] "student_id"          "age"                 "gender"              "school_type"         "parent_education"
 [6] "study_hours"         "attendance_percentage" "internet_access"     "travel_time"         "extra_activities"
[11] "study_method"        "math_score"          "science_score"       "english_score"       "overall_score"
[16] "final_grade"
>
> # 3
> df$gender <- as.factor(df$gender)
> df$final_grade <- as.factor(df$final_grade)
>
> # 4
> chi_table <- table(df$gender, df$final_grade)
> print("Contingency Table:")
[1] "Contingency Table:"
> print(chi_table)

            a    b    c    d    e    f
  female  417  887 2121 2071 1796  998
  male    375  894 2026 2096 1932  924
  other   413  915 2014 2144 1944 1033
>
> # 5
> chi_result <- chisq.test(chi_table)
>
> # 6
> print("Chi-Square Test Result:")
[1] "Chi-Square Test Result:"
> print(chi_result)

        Pearson's Chi-squared test

data:  chi_table
X-squared = 18.146, df = 10, p-value = 0.05254

>
> # 7
> print("Expected Frequencies:")
```



```
> boxplot(
+   exam_score ~ course,
+   data = df,
+   col = c("lightblue", "lightgreen", "lightpink"),
+   main = "Exam Score Comparison Across Courses",
+   xlab = "Course",
+   ylab = "Exam Score"
+ )
>
> # 7.
> 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 0.8990690
bca-b.com        0.190880945 -1.2777176 1.6594795 0.9997538
diploma-b.com   -0.002564226 -1.4809389 1.4758105 1.0000000
b.tech-b.sc     -0.065464092 -1.5457749 1.4148467 0.9999996
ba-b.sc         -0.606779310 -2.0743353 0.8607767 0.8869917
bba-b.sc         0.209133894 -1.2661398 1.6844075 0.9995929
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 -1.8099056 1.1641131 0.9954655
bba-ba           0.815913204 -0.6570832 2.2889096 0.6606339
bca-ba           0.411864134 -1.0526450 1.8763733 0.9819820
diploma-ba       0.218418962 -1.2558935 1.6927315 0.9994750
bca-bba         -0.404049070 -1.8762918 1.0681937 0.9841216
diploma-bba     -0.597494242 -2.0794892 0.8845007 0.8986431
diploma-bca     -0.193445172 -1.6670047 1.2801144 0.9997391

>
```

NAME - UNNATI RATHOD
ROLL NO - S109

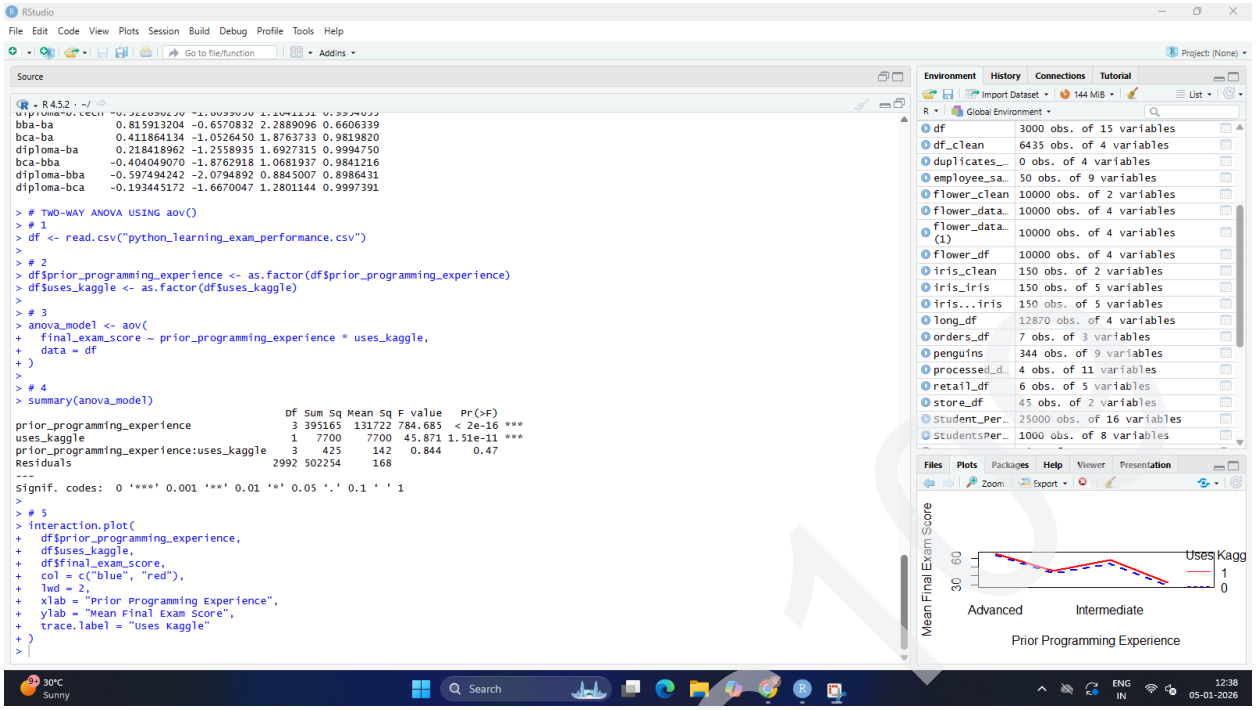
SHETH L.U.J AND SIR M.V COLLEGE
PRACTICAL NO - 7,8,9
SUBJECT -DATA ANALYSIS




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

NAME - UNNATI RATHOD
ROLL NO - S109

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



```
bba-ba           0.815913204 -0.6570832 2.2889096 0.6606339
bca-ba           0.411864134 -1.0526450 1.8763733 0.9819820
diploma-ba       0.218418962 -1.2558935 1.6927315 0.9994750
bca-bba         -0.404049070 -1.8762918 1.0681937 0.9841216
diploma-bba     -0.597494242 -2.0794892 0.8845007 0.8986431
diploma-bca     -0.193445172 -1.6670047 1.2801144 0.9997391

> # TWO-WAY ANOVA USING aov()
> # 1
> df <- read.csv("python_learning_exam_performance.csv")
>
> # 2
> df$prior_programming_experience <- as.factor(df$prior_programming_experience)
> df$uses_kaggle <- as.factor(df$uses_kaggle)
>
> # 3
> anova_model <- aov(
+   final_exam_score ~ prior_programming_experience * uses_kaggle,
+   data = df
+ )
>
> # 4
> summary(anova_model)
                                        Df Sum Sq Mean Sq F value   Pr(>F)
prior_programming_experience             3 395165  131722 784.685  < 2e-16 ***
uses_kaggle                              1   7700    7700  45.871 1.51e-11 ***
prior_programming_experience:uses_kaggle    3    425     142   0.844     0.47
Residuals                             2992 502254     168
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # 5
> interaction.plot(
+   df$prior_programming_experience,
+   df$uses_kaggle,
+   df$final_exam_score,
+   col = c("blue", "red"),
+   lwd = 2,
+   xlab = "Prior Programming Experience",
+   ylab = "Mean Final Exam Score",
+   trace.label = "Uses Kaggle"
+ )
>
```





NAME - UNNATI RATHOD
ROLL NO - S109

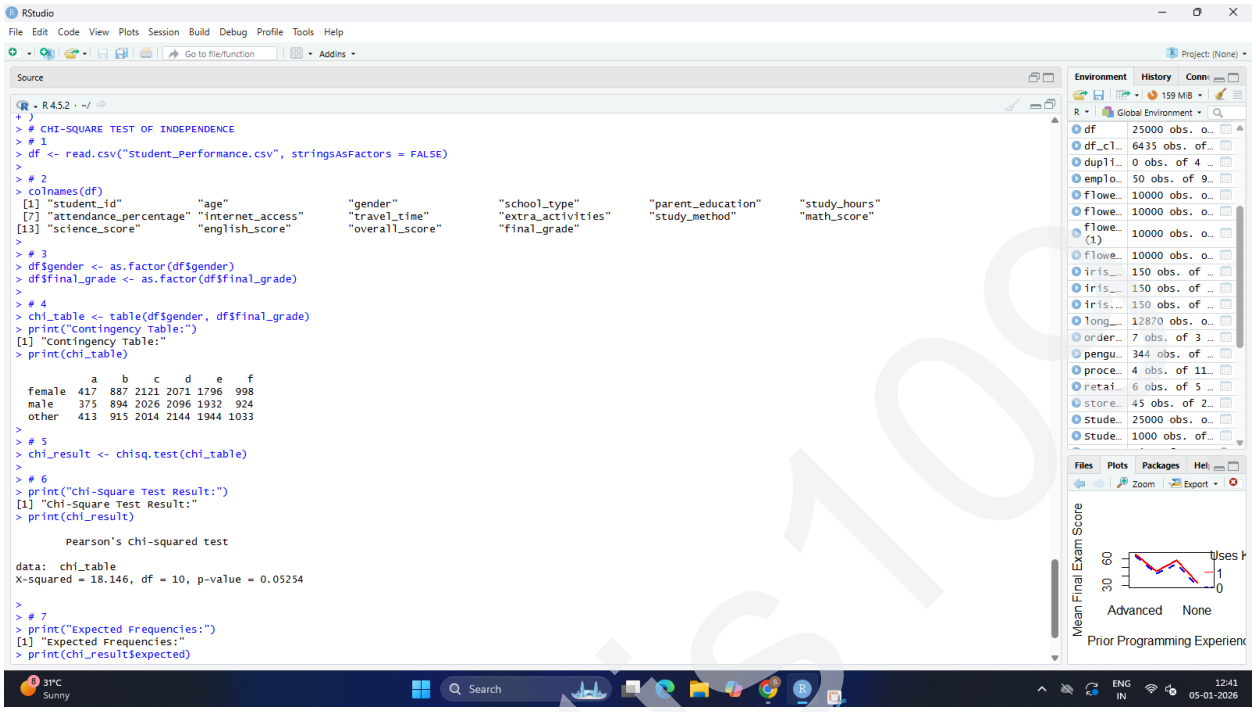
SHETH L.U.J AND SIR M.V COLLEGE

PRACTICAL NO - 7,8,9

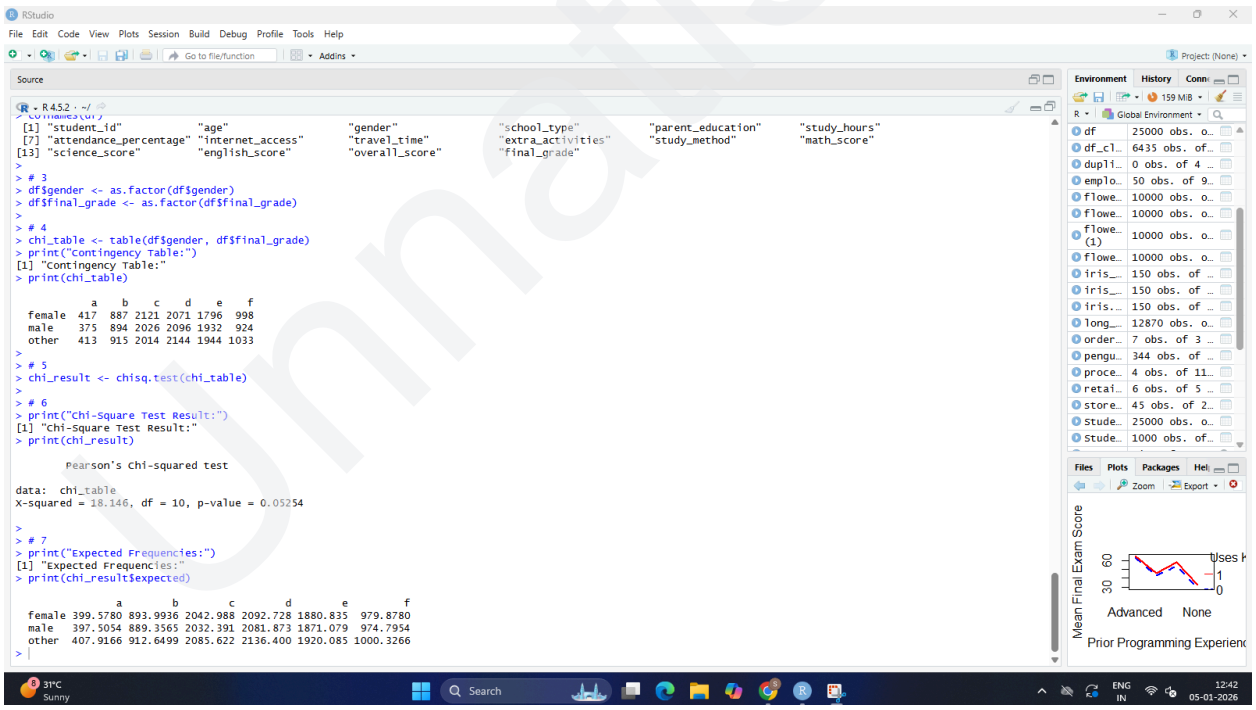
SUBJECT -DATA ANALYSIS

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

OUTPUT



```
> # CHI-SQUARE TEST OF INDEPENDENCE
> # 1
> df <- read.csv("Student_Performance.csv", stringsAsFactors = FALSE)
>
> # 2
> colnames(df)
 [1] "student_id"          "age"               "gender"            "school_type"       "parent_education"  "study_hours"
 [7] "attendance_percentage" "internet_access"  "travel_time"       "extra_activities"  "study_method"      "math_score"
[13] "science_score"       "english_score"     "overall_score"     "final_grade"
>
> # 3
> df$gender <- as.factor(df$gender)
> df$final_grade <- as.factor(df$final_grade)
>
> # 4
> chi_table <- table(df$gender, df$final_grade)
> print("Contingency Table:")
[1] "Contingency Table:"
> print(chi_table)

            a    b    c    d    e    f
  female  417  887 2121 2071 1796  998
  male    375  894 2026 2096 1932  924
  other   413  915 2014 2144 1944 1033
>
> # 5
> chi_result <- chisq.test(chi_table)
>
> # 6
> print("Chi-Square Test Result:")
[1] "Chi-Square Test Result:"
> print(chi_result)

        Pearson's Chi-squared test

data:  chi_table
X-squared = 18.146, df = 10, p-value = 0.05254

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



```
> # 3
> df$gender <- as.factor(df$gender)
> df$final_grade <- as.factor(df$final_grade)
>
> # 4
> chi_table <- table(df$gender, df$final_grade)
> print("Contingency Table:")
[1] "Contingency Table:"
> print(chi_table)

            a    b    c    d    e    f
  female  417  887 2121 2071 1796  998
  male    375  894 2026 2096 1932  924
  other   413  915 2014 2144 1944 1033
>
> # 5
> chi_result <- chisq.test(chi_table)
>
> # 6
> print("Chi-Square Test Result:")
[1] "Chi-Square Test Result:"
> print(chi_result)

        Pearson's Chi-squared test

data:  chi_table
X-squared = 18.146, df = 10, p-value = 0.05254

>
> # 7
> 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  979.8780
  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