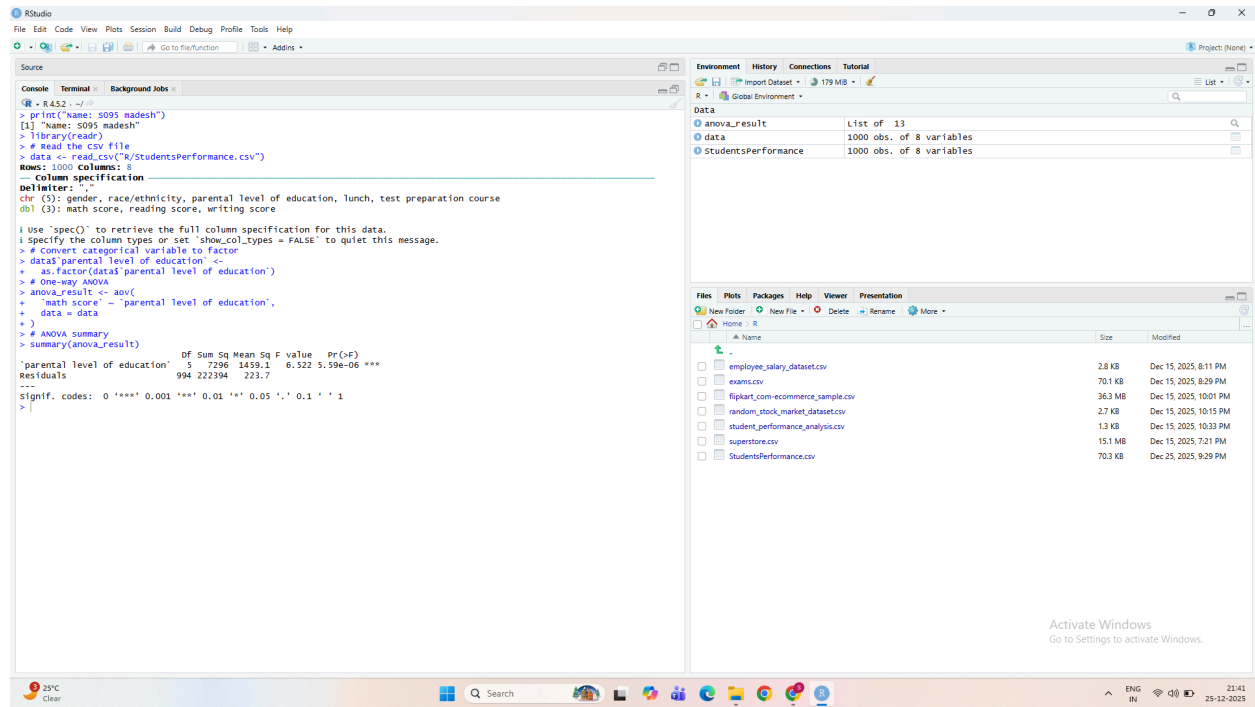


# SHETH L.U.J. AND SIR M.V. COLLEGE

## PRACTICAL NO 7

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



```
R > R452 -> /
> print("Name: S095 madesh")
[1] "Name: S095 madesh"
> library(readr)
> # read the csv file
> data <- read_csv("R/StudentsPerformance.csv")
Rows: 1000 Columns: 8
# Column specification
Delimiter: ","
chr (5): gender, race/ethnicity, parental level of education, lunch, test preparation course
dbl (3): math score, reading score, writing score

I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # Convert categorical variable to factor
> data$'parental level of education' <-
+ as.factor(data$'parental level of education')
> # One-way ANOVA
> anova_result <- aov(
+   'math score' ~ 'parental level of education',
+   data = data
+ )
> # ANOVA summary
> summary(anova_result)
              Df Sum Sq Mean Sq F value    Pr(>F)
'parental level of education'  5   7296   1459.1    6.522 5.59e-06 ***
Residuals                    994 222394    223.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

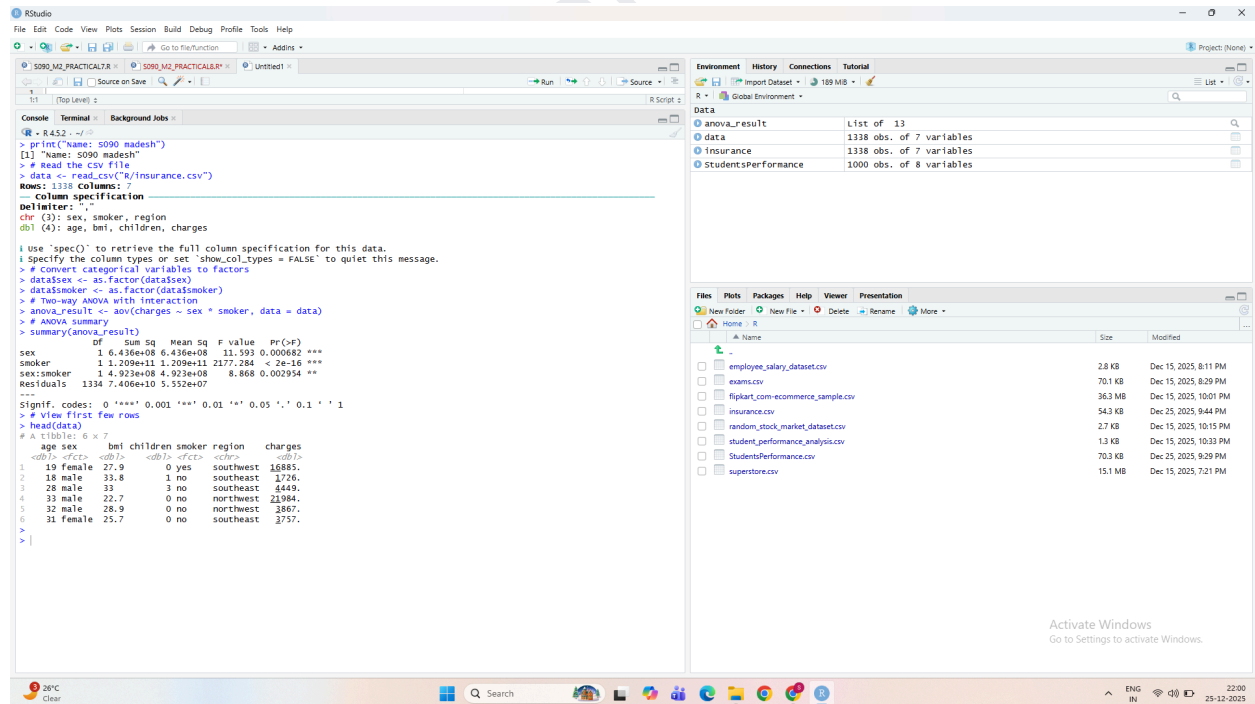
Environment: Global Environment

Object	Class	Size
anova_result	List of 13	
data	1000 obs. of 8 variables	
StudentsPerformance	1000 obs. of 8 variables	

Files: employee\_salary\_dataset.csv (2.8 KB), exams.csv (70.1 KB), flipkart\_commerce\_sample.csv (36.3 MB), random\_stock\_market\_dataset.csv (2.7 KB), student\_performance\_analysis.csv (1.3 KB), superstore.csv (15.1 MB), StudentsPerformance.csv (70.3 KB)

## PRACTICAL NO 8

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



```
R > R452 -> /
> print("Name: S090 madesh")
[1] "Name: S090 madesh"
> library(readr)
> # read the csv file
> data <- read_csv("R/insurance.csv")
Rows: 1338 Columns: 7
# Column specification
Delimiter: ","
chr (3): sex, smoker, region
dbl (4): age, bmi, children, charges

I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # Convert categorical variables to factors
> data$sex <- as.factor(data$sex)
> data$smoker <- as.factor(data$smoker)
> # Two-way ANOVA with interaction
> anova_result <- aov(charges ~ sex * smoker, data = data)
> # ANOVA summary
> summary(anova_result)
              Df Sum Sq Mean Sq F value    Pr(>F)
sex              1  6.436e+08  6.436e+08    11.593 0.000682 ***
smoker           1  1.209e+11  1.209e+11 2177.284 < 2e-16 ***
sex:smoker       1  4.932e+08  4.932e+08    8.868 0.002954 **
Residuals      1334  7.406e+10  5.552e+07
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # view first few rows
> head(data)
# A tibble: 6 x 7
  age sex    bmi children smoker region charges
<dbl> <fct> <dbl> <dbl> <fct> <fct> <dbl>
1   19 female  27.9     0 yes  southwest 16885.
2   18 male   33.8     1 no   southeast 1726.
3   28 male   33     1 no   southeast 4449.
4   33 male  22.7     0 no   northwest 21984.
5   32 male  28.9     0 no   northwest 3867.
6   31 female 25.7     0 no   southeast 2157.

> |
```

Environment: Global Environment

Object	Class	Size
anova_result	List of 13	
data	1338 obs. of 7 variables	
insurance	1338 obs. of 7 variables	
StudentsPerformance	1000 obs. of 8 variables	

Files: employee\_salary\_dataset.csv (2.8 KB), exams.csv (70.1 KB), flipkart\_commerce\_sample.csv (36.3 MB), insurance.csv (54.3 KB), random\_stock\_market\_dataset.csv (2.7 KB), student\_performance\_analysis.csv (1.3 KB), StudentsPerformance.csv (70.3 KB), superstore.csv (15.1 MB)

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

## PRACTICAL NO 9

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

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for loading a dataset, creating a categorical variable, and performing a chi-square test.
- Console:** Shows the execution of the code, including the output of the chi-square test.
- Environment:** Lists the objects in the R environment, including 'anova\_result', 'categoricaldataset', 'chisq\_result', 'data', 'insurance', and 'studentsperformance'.
- Files:** Shows a list of files in the project directory, including 'CategoricalDataset.csv', 'employee\_salary\_dataset.csv', 'exams.csv', 'flipkart\_commerce\_sample.csv', 'insurance.csv', 'random\_stock\_market\_dataset.csv', 'student\_performance\_analysis.csv', 'StudentsPerformance.csv', and 'superstore.csv'.

**R Code in Source Editor:**

```
# R452 - w/...
> print("Name: S095 MADESH")
[1] "Name: S095 MADESH"
> data <- read_csv("R/CategoricalDataset.csv")
Rows: 7494 Columns: 240
# Column specification
Delimiter: ","
dbl (240): Team 1_Afghanistan, Team 1_Australia, Team 1_Bangladesh, Team 1_Bermuda, Team 1_Canada, Team 1_East Africa...

i use 'spec()' to retrieve the full column specification for this data.
i specify the column types or set 'show_col_types = FALSE' to quiet this message.
> table_data <- table(data$Team 1_India, data$Team 2_Pakistan)
> chisq_result <- chisq.test(table_data)
> chisq_result

Pearson's chi-squared test with Yates' continuity correction

data: table_data
X-squared = 6.2942, df = 1, p-value = 0.01211

> head(data)
# A tibble: 6 x 240
#   Team 1_Afghanistan Team 1_Australia Team 1_Bangladesh Team 1_Bermuda Team 1_Canada Team 1_East Africa
#   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 0 0 0 0 0 0
2 0 0 0 0 0 0
3 0 0 0 0 0 0
4 0 0 0 0 0 0
5 0 0 0 0 0 0
6 0 0 0 0 0 0
# i 234 more variables: Team 1_England <dbl>, Team 1_Hong Kong <dbl>, Team 1_India <dbl>, Team 1_Ireland <dbl>,
# Team 1_Kenya <dbl>, Team 1_Namibia <dbl>, Team 1_Netherlands <dbl>, Team 1_New Zealand <dbl>,
# Team 1_P.N.G. <dbl>, Team 1_Pakistan <dbl>, Team 1_Scotland <dbl>, Team 1_South Africa <dbl>,
# Team 1_Sri Lanka <dbl>, Team 1_U.A.E. <dbl>, Team 1_U.S.A. <dbl>, Team 1_West Indies <dbl>,
# Team 1_Zimbabwe <dbl>, Team 2_Afghanistan <dbl>, Team 2_Australia <dbl>, Team 2_Bangladesh <dbl>,
# Team 2_Bermuda <dbl>, Team 2_Canada <dbl>, Team 2_East Africa <dbl>, Team 2_England <dbl>,
# Team 2_Hong Kong <dbl>, Team 2_India <dbl>, Team 2_Ireland <dbl>, Team 2_Kenya <dbl>, ...
# i use 'colnames()' to see all variable names
> |
```

**Console Output:**

```
data: table_data
X-squared = 6.2942, df = 1, p-value = 0.01211

# A tibble: 6 x 240
#   Team 1_Afghanistan Team 1_Australia Team 1_Bangladesh Team 1_Bermuda Team 1_Canada Team 1_East Africa
#   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 0 0 0 0 0 0
2 0 0 0 0 0 0
3 0 0 0 0 0 0
4 0 0 0 0 0 0
5 0 0 0 0 0 0
6 0 0 0 0 0 0
# i 234 more variables: Team 1_England <dbl>, Team 1_Hong Kong <dbl>, Team 1_India <dbl>, Team 1_Ireland <dbl>,
# Team 1_Kenya <dbl>, Team 1_Namibia <dbl>, Team 1_Netherlands <dbl>, Team 1_New Zealand <dbl>,
# Team 1_P.N.G. <dbl>, Team 1_Pakistan <dbl>, Team 1_Scotland <dbl>, Team 1_South Africa <dbl>,
# Team 1_Sri Lanka <dbl>, Team 1_U.A.E. <dbl>, Team 1_U.S.A. <dbl>, Team 1_West Indies <dbl>,
# Team 1_Zimbabwe <dbl>, Team 2_Afghanistan <dbl>, Team 2_Australia <dbl>, Team 2_Bangladesh <dbl>,
# Team 2_Bermuda <dbl>, Team 2_Canada <dbl>, Team 2_East Africa <dbl>, Team 2_England <dbl>,
# Team 2_Hong Kong <dbl>, Team 2_India <dbl>, Team 2_Ireland <dbl>, Team 2_Kenya <dbl>, ...
# i use 'colnames()' to see all variable names
> |
```

**Environment:**

Object	Class	Attributes
anova_result	List of 13	
categoricaldataset	7494 obs. of 240 variables	
chisq_result	List of 9	
data	7494 obs. of 240 variables	
insurance	1338 obs. of 7 variables	
studentsperformance	1000 obs. of 8 variables	

**Files:**

Name	Size	Modified
CategoricalDataset.csv	3.4 MB	Dec 25, 2025, 10:22 PM
employee_salary_dataset.csv	2.8 KB	Dec 15, 2025, 8:11 PM
exams.csv	70.1 KB	Dec 15, 2025, 8:29 PM
flipkart_commerce_sample.csv	36.3 MB	Dec 15, 2025, 10:01 PM
insurance.csv	54.3 KB	Dec 25, 2025, 9:44 PM
random_stock_market_dataset.csv	2.7 KB	Dec 15, 2025, 10:15 PM
student_performance_analysis.csv	1.3 KB	Dec 15, 2025, 10:33 PM
StudentsPerformance.csv	70.3 KB	Dec 25, 2025, 9:29 PM
superstore.csv	15.1 MB	Dec 15, 2025, 7:21 PM