

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

SUBJECT: DATA ANALYSIS WITH R

PRACTICAL NO. 7 TO 9 MOD 2

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

The screenshot shows the RStudio interface with the following code in the Source pane:

```
R > # Load dataset
> data <- read.csv("D:/S095 Aashka/student_exam_scores.csv")
> View(data)
>
> # Create attendance groups
> data$attendance_group <- cut(
+   data$attendance_percent,
+   breaks = c(0, 70, 85, 100),
+   labels = c("Low", "Medium", "High"))
>
> # Perform One-way ANOVA
> anova_result <- aov(exam_score ~ attendance_group, data = data)
>
> # Display ANOVA table
> summary(anova_result)
    Of Sum Sq Mean Sq F value Pr(>F)
attendance_group 2 493 246.37 5.591 0.00435 ***
Residuals 197 8681 44.06
---
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> # Load dataset
> data <- read.csv("D:/S095 Aashka/student_exam_scores.csv")
> View(data)
>
> # Create attendance groups
> data$attendance_group <- cut(
+   data$attendance_percent,
+   breaks = c(0, 70, 85, 100),
+   labels = c("Low", "Medium", "High"))
>
> # Perform One-way ANOVA
> anova_result <- aov(exam_score ~ attendance_group, data = data)
>
> # Display ANOVA table
> summary(anova_result)
    Of Sum Sq Mean Sq F value Pr(>F)
attendance_group 2 493 246.37 5.591 0.00435 ***
Residuals 197 8681 44.06
---
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

The Environment pane shows various datasets and objects. The User Library pane lists CRAN packages. The bottom status bar shows the date and time.

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

The screenshot shows the RStudio interface with the following code in the Source pane:

```
[4] "How.would.you.rate.your.study.load."
[5] "How.many.times.a.week.you.practice.extra.curricular.activities..."
[6] "How.would.you.rate.your.stress.levels."
>
> # Convert factors
> data$Gender <- as.factor(data$Gender)
Error in `\$<-` data.frame(`*tmp*`, Gender, value = integer(0)) :
replacement has 0 rows, data has 520

> # Load dataset
> data <- read.csv("D:/S095 Aashka/Student Stress Factors.csv")
>
> # Rename columns for simplicity (optional but recommended)
> colnames(data) <- c(
+   "Sleep_Quality",
+   "Headache_Frequency",
+   "Academic_Performance",
+   "Study_Load",
+   "Extracurricular_Activities",
+   "Stress_Level"
+ )
>
> # Convert variables to factors
> data$Sleep_Quality <- as.factor(data$Sleep_Quality)
> data$Stress_Level <- as.factor(data$Stress_Level)
>
> # Two-Way ANOVA
> anova_result <- aov(
+   Academic_Performance ~ Stress_Level * Sleep_Quality,
+   data = data
+ )
>
> # Display ANOVA table
> summary(anova_result)
    Df Sum Sq Mean Sq F value Pr(>F)
stress_Level        4   6.5   1.632   1.736  0.141
sleep.Quality       4  55.9   13.982  14.874 1.72e-11 ***
stress_Level:Sleep.quality 13  53.9   4.143   4.407 4.34e-07 ***
Residuals          498 468.1   0.940
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

The Environment pane shows various datasets and objects. The User Library pane lists CRAN packages. The bottom status bar shows the date and time.

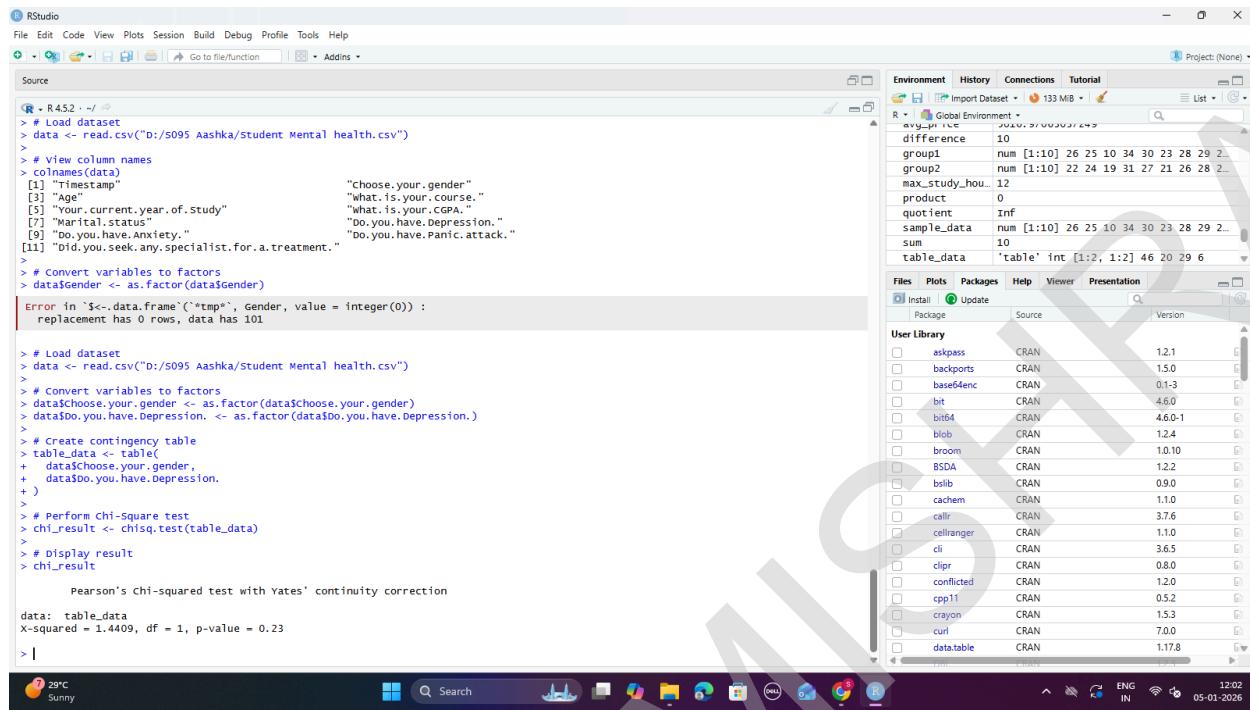
NAME: AASHKA MISHRA

ROLL NO: S095

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

SUBJECT: DATA ANALYSIS WITH R

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



The screenshot shows the RStudio interface with the following R code in the Source pane:

```

R - R 4.5.2 - ~
> # Load dataset
> data <- read.csv("D:/S095 Aashka/Student Mental health.csv")
>
> # view column names
> colnames(data)
[1] "timestamp"           "choose.your.gender"
[3] "Age"                 "what.is.your.course."
[5] "Your.current.year.of.study" "what.is.your.CGPA."
[7] "Marital.status"       "Do.you.have.Depression."
[9] "Do.you.have.Anxiety." "Do.you.have.Panic.attack."
[11] "Did.you.seek.any.specialist.for.a.treatment."
>
> # Convert variables to factors
> data$Gender <- as.factor(data$Gender)
>Error in `\$<-`(`data.frame`(*tmp*), gender, value = integer(0)) : 
replacement has 0 rows, data has 101

> # Load dataset
> data <- read.csv("D:/S095 Aashka/Student Mental health.csv")
>
> # convert variables to factors
> data$choose.your.gender <- as.factor(data$choose.your.gender)
> data$do.you.have.Depression. <- as.factor(data$do.you.have.Depression.)
>
> # Create contingency table
> table_data <- table(
+   data$choose.your.gender,
+   data$do.you.have.Depression.
+ )
>
> # Perform Chi-Square test
> chi_result <- chisq.test(table_data)
>
> # display result
> chi_result
Pearson's Chi-squared test with Yates' continuity correction

data: table_data
X-squared = 1.4409, df = 1, p-value = 0.23
> |

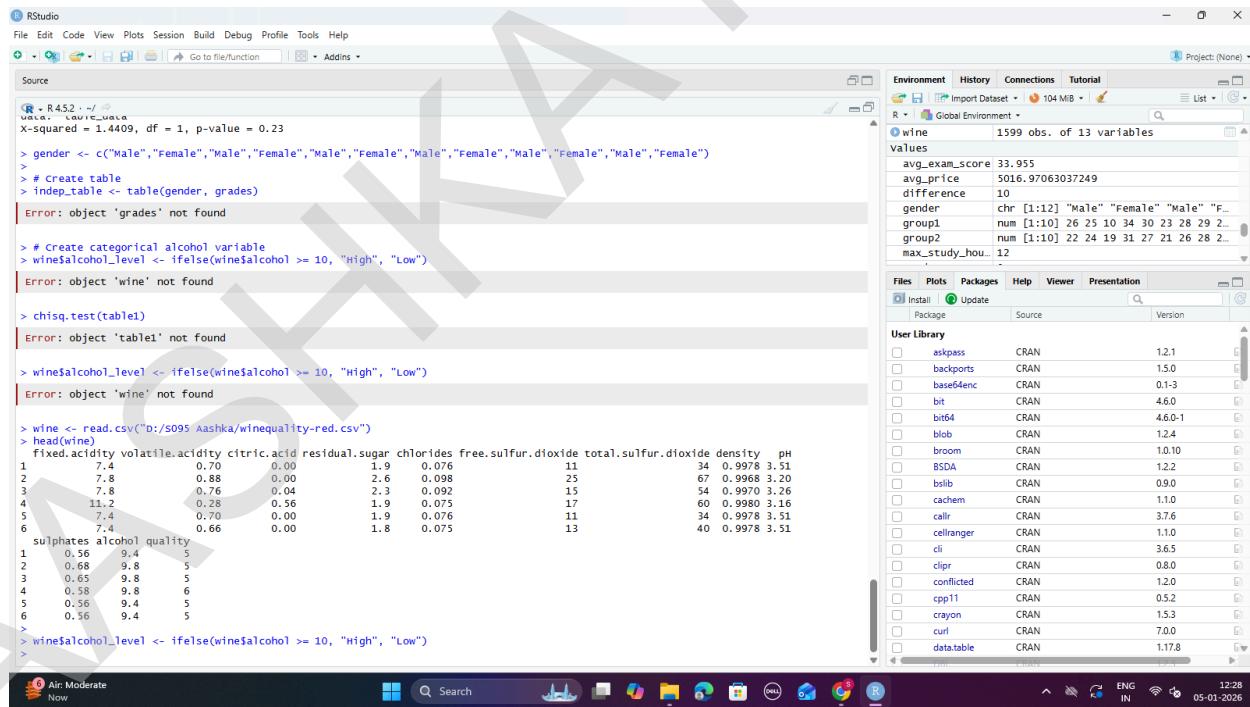
```

The Environment pane shows the following data frame:

	avg_exam_score	avg_price	difference
group1	[1:10]	26 25 10 34 30 23 28 29 2...	
group2	[1:10]	22 24 19 31 27 21 26 28 2...	
max_study_hou...	12		
product	0		
quotient	Inf		
sample_data	[1:10]	26 25 10 34 30 23 28 29 2...	
sum	10		
table_data	'table'	[1:2, 1:2] 46 20 29 6	

The User Library pane lists CRAN packages.

TYPE 1: INDEPENDENCE



The screenshot shows the RStudio interface with the following R code in the Source pane:

```

R - R 4.5.2 - ~
> wine <- read.csv("D:/S095 Aashka/winequality-red.csv")
> head(wine)
fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide density pH
1 7.4 0.70 0.09 1.9 0.076 11 34 0.998 3.51
2 7.8 0.88 0.09 2.6 0.088 25 67 0.9993 3.20
3 7.8 0.70 0.04 2.3 0.092 15 54 0.9970 3.16
4 11.2 0.28 0.56 1.9 0.075 17 60 0.9980 3.16
5 7.4 0.70 0.00 1.9 0.076 11 34 0.9978 3.51
6 7.4 0.66 0.00 1.8 0.075 13 40 0.9978 3.51
sulphates alcohol quality
1 0.56 9.4 5
2 0.68 9.8 5
3 0.65 9.8 5
4 0.58 9.8 6
5 0.56 9.4 5
6 0.56 9.4 5
> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
>

```

The Environment pane shows the following data frame:

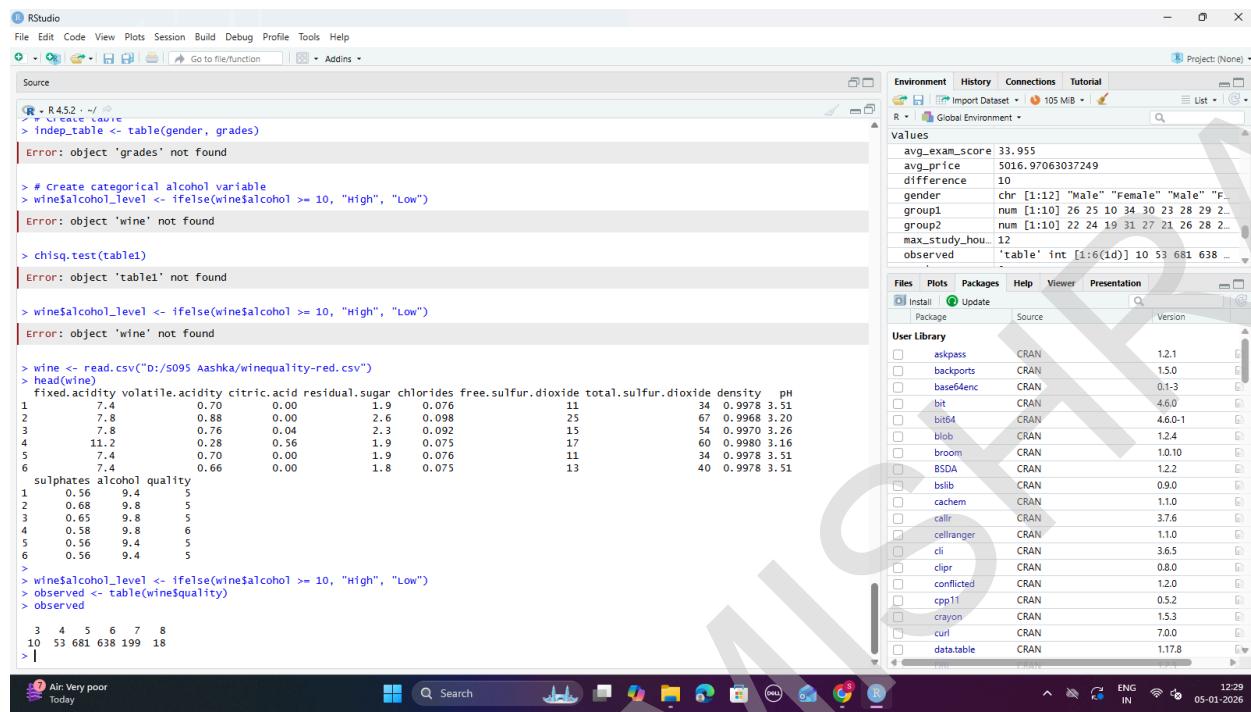
	avg_exam_score	avg_price	difference
gender	chr [1:12]	"Male" "Female" "Male" "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female" "Male"	
group1	num [1:10]	26 25 10 34 30 23 28 29 2...	
group2	num [1:10]	22 24 19 31 27 21 26 28 2...	
max_study_hou...	12		

The User Library pane lists CRAN packages.

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

SUBJECT: DATA ANALYSIS WITH R

TYPE 2: GOODNESS OF FIT



RStudio interface showing R code and output for Type 2 Goodness of Fit analysis.

```

R > R45.2 ->
> # Load wine quality dataset
> indp_table <- table(gender, grades)
>Error: object 'grades' not found

> # Create categorical alcohol variable
> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
>Error: object 'wine' not found

> chisq.test(table1)
>Error: object 'table1' not found

> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
>Error: object 'wine' not found

> wine <- read.csv("D:/S095 Aashka/winequality-red.csv")
> head(wine)
fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide density pH
1 7.0 0.70 0.00 1.9 0.076 11 34 0.9978 3.51
2 7.8 0.88 0.00 2.6 0.098 25 67 0.9968 3.20
3 7.8 0.76 0.04 2.3 0.092 15 54 0.9970 3.26
4 11.2 0.28 0.56 1.9 0.075 17 60 0.9980 3.16
5 7.4 0.70 0.00 1.9 0.076 11 34 0.9978 3.51
6 7.4 0.66 0.00 1.8 0.075 13 40 0.9978 3.51

sulphates alcohol quality
1 0.56 9.4 5
2 0.68 9.8 5
3 0.65 9.8 5
4 0.58 9.8 6
5 0.56 9.4 5
6 0.56 9.4 5
>
> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
> observed <- table(wine$quality)
> observed

3 4 5 6 7 8
10 53 681 638 199 18
>

```

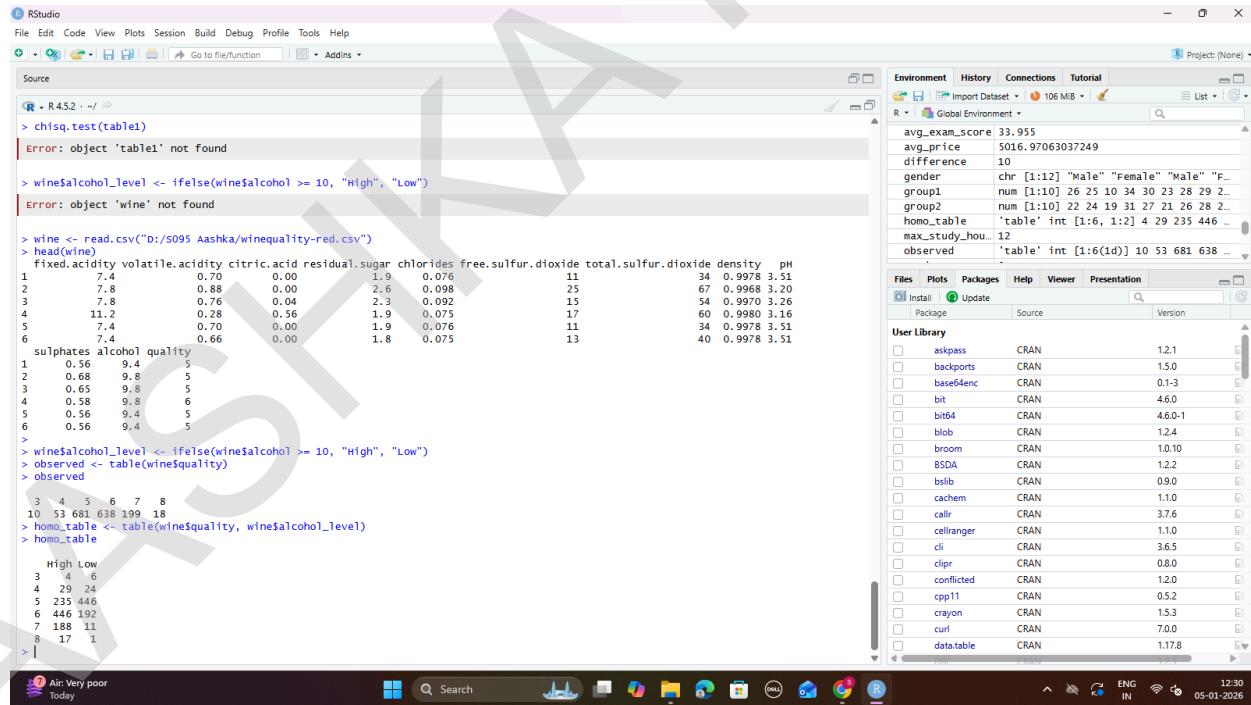
Environment pane shows variables:

- avg_exam_score: 33.955
- avg_price: 5016.97063037249
- difference: 10
- gender: chr [1:12] "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female"
- group1: num [1:10] 26 25 10 34 30 23 28 29 2...
- group2: num [1:10] 22 24 19 31 27 21 26 28 2...
- max_study_hours: 12
- observed: 'table' int [1:6(1d)] 10 53 681 638 ...

Session pane shows installed packages:

- askpass
- backports
- base64enc
- bit
- bit64
- blob
- broom
- BSDA
- bslib
- cachem
- callr
- cellranger
- cli
- clipr
- conflicted
- cpp11
- crayon
- curl
- data.table

TYPE 3: HOMOGENEITY



RStudio interface showing R code and output for Type 3 Homogeneity analysis.

```

R > R45.2 ->
> chisq.test(table1)
>Error: object 'table1' not found

> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
>Error: object 'wine' not found

> wine <- read.csv("D:/S095 Aashka/winequality-red.csv")
> head(wine)
fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide total.sulfur.dioxide density pH
1 7.0 0.70 0.00 1.9 0.076 11 34 0.9978 3.51
2 7.8 0.88 0.00 2.6 0.098 25 67 0.9968 3.20
3 7.8 0.76 0.04 2.3 0.092 15 54 0.9970 3.26
4 11.2 0.28 0.56 1.9 0.075 17 60 0.9980 3.16
5 7.4 0.70 0.00 1.9 0.076 11 34 0.9978 3.51
6 7.4 0.66 0.00 1.8 0.075 13 40 0.9978 3.51

sulphates alcohol quality
1 0.56 9.4 5
2 0.68 9.8 5
3 0.65 9.8 5
4 0.58 9.8 6
5 0.56 9.4 5
6 0.56 9.4 5
>
> wine$alcohol_level <- ifelse(wine$alcohol >= 10, "High", "Low")
> observed <- table(wine$quality)
> observed

3 4 5 6 7 8
10 53 681 638 199 18
> homo_table <- table(wine$quality, wine$alcohol_level)
> homo_table

  High Low
3 29 24
4 235 446
6 446 192
7 188 11
8 17 1
>

```

Environment pane shows variables:

- avg_exam_score: 33.955
- avg_price: 5016.97063037249
- difference: 10
- gender: chr [1:12] "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female" "Male" "Female"
- group1: num [1:10] 26 25 10 34 30 23 28 29 2...
- group2: num [1:10] 22 24 19 31 27 21 26 28 2...
- homo_table: 'table' int [1:6, 1:2] 4 29 235 446 ...
- max_study_hours: 12
- observed: 'table' int [1:6(1d)] 10 53 681 638 ...

Session pane shows installed packages:

- askpass
- backports
- base64enc
- bit
- bit64
- blob
- BSDA
- bslib
- cachem
- callr
- cellranger
- cli
- clipr
- conflicted
- cpp11
- crayon
- curl
- data.table