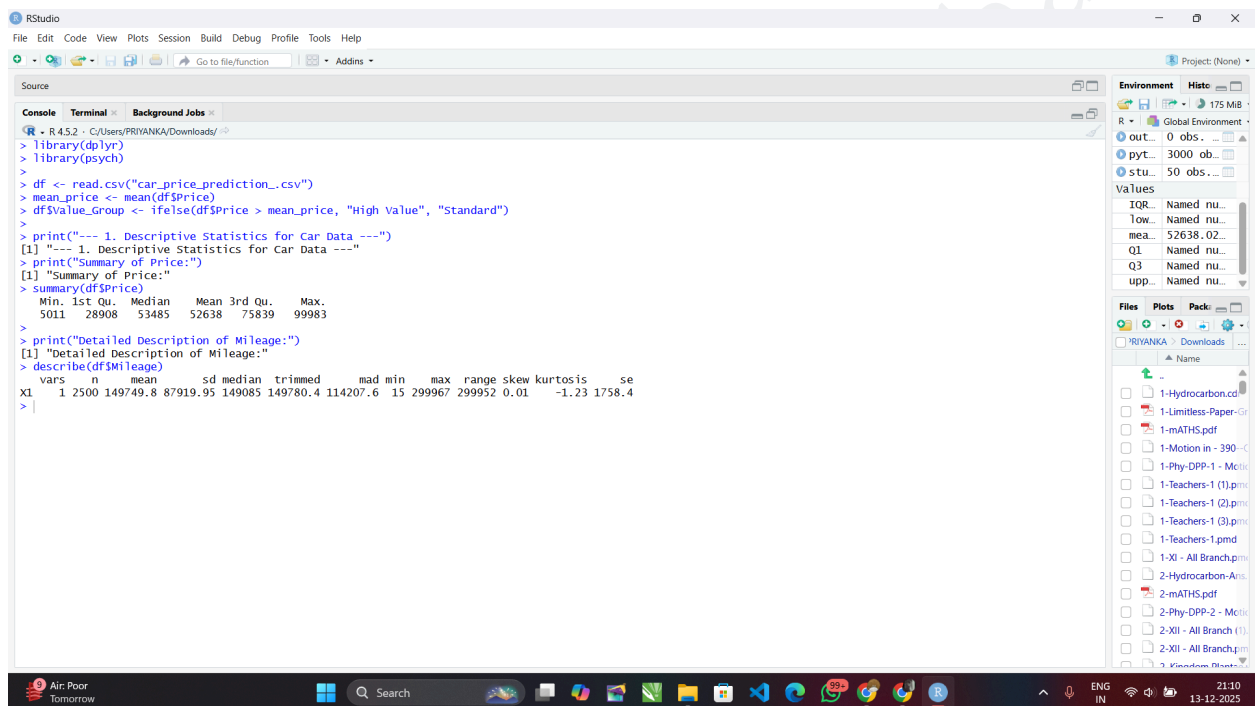


**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Module 2 Practical 1-6**

**Aim:** Generating descriptive statistics using summary() or describe() (R).

**OUTPUT:**



The screenshot shows the RStudio interface with the following content:

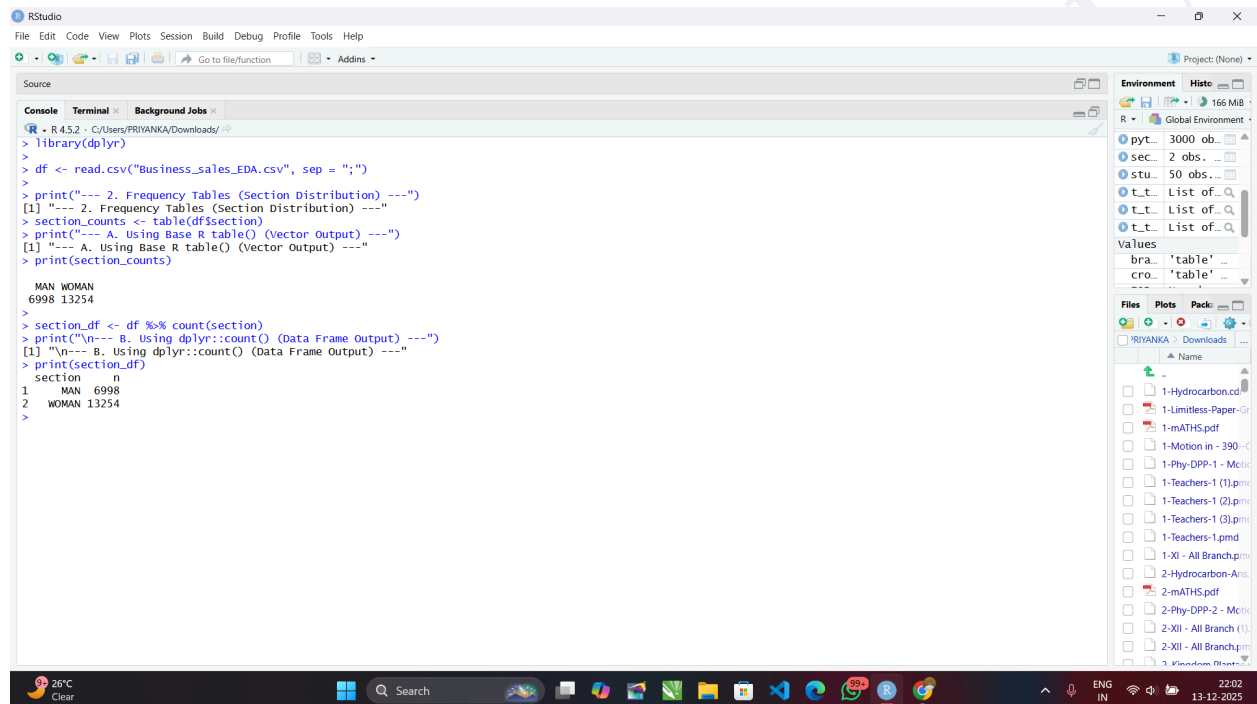
```
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
> library(dplyr)
> library(psych)
>
> df <- read.csv("car_price_prediction_.csv")
> mean_price <- mean(df$Price)
> df$Value_Group <- ifelse(df$Price > mean_price, "High Value", "Standard")
>
> print("--- 1. Descriptive Statistics for Car Data ---")
[1] "--- 1. Descriptive Statistics for Car Data ---"
> print("Summary of Price:")
[1] "Summary of Price:"
> summary(df$Price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5011   28908   53485   52638   75839   99983
>
> print("Detailed Description of Mileage:")
[1] "Detailed Description of Mileage:"
> describe(df$Mileage)
   vars      n      mean       sd median trimmed   mad min    max range skew kurtosis   se
xl     1 2500 149749.8 87919.95 149085 149780.4 114207.6  15 299967 299952 0.01  -1.23 1758.4
>
```

The Environment pane on the right shows the Global Environment with variables: out... (0 obs...), pyt... (3000 obs...), and stu... (50 obs...). The Files pane shows a list of files in the Downloads folder.

**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Aim:** Generating frequency tables using table() or count() (R).

**OUTPUT:**



The screenshot shows the RStudio interface with the following content:

```
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
> library(dplyr)
> 
> df <- read.csv("Business_sales_EDA.csv", sep = ";")
> 
> print("---- 2. Frequency Tables (Section Distribution) ----")
[1] "---- 2. Frequency Tables (Section Distribution) ----"
> section_counts <- table(df$section)
> print("---- A. Using Base R table() (Vector Output) ----")
[1] "---- A. Using Base R table() (Vector Output) ----"
> print(section_counts)

MAN WOMAN
6998 13254
> 
> section_df <- df %>% count(section)
> print("\n--- B. Using dplyr::count() (Data Frame Output) ---")
[1] "\n--- B. Using dplyr::count() (Data Frame Output) ---"
> print(section_df)
  section     n
1     MAN 6998
2    WOMAN 13254
>
```

The Environment pane on the right shows the following objects:

- pyt... 3000 obs...
- sec... 2 obs...
- stu... 50 obs...
- t.t... List of...
- t.t... List of...
- t.t... List of...
- Values
- bra... 'table'...
- cro... 'table'...

The Files pane on the right shows the following files:

- 1-Hydrocarbon.co...
- 1-Limitless-Paper-G...
- 1-mATHS.pdf
- 1-Motion in - 390-C...
- 1-Phy-DPP-1 - Moti...
- 1-Teachers-1 (1).pm...
- 1-Teachers-1 (2).pm...
- 1-Teachers-1 (3).pm...
- 1-Teachers-1.pmd
- 1-XI - All Branch.p...
- 2-Hydrocarbon-Ans...
- 2-mATHS.pdf
- 2-Phy-DPP-2 - Moti...
- 2-XII - All Branch (1)...
- 2-XII - All Branch.p...
- 3-Kinodrom Dis...

**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Aim:** Creating cross-tabulations and two-way tables using table() (R).

**OUTPUT:**

The screenshot shows the RStudio interface. The console displays the following R code and its output:

```
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
> df <- read.csv("Age-standardized suicide rates.csv")
> 
> print("--- 3. Cross-Tabulation (Country vs. Sex) ---")
[1] "--- 3. Cross-Tabulation (Country vs. Sex) ---"
> cross_tab <- table(df$country, df$sex)
> 
> print(cross_tab)
```

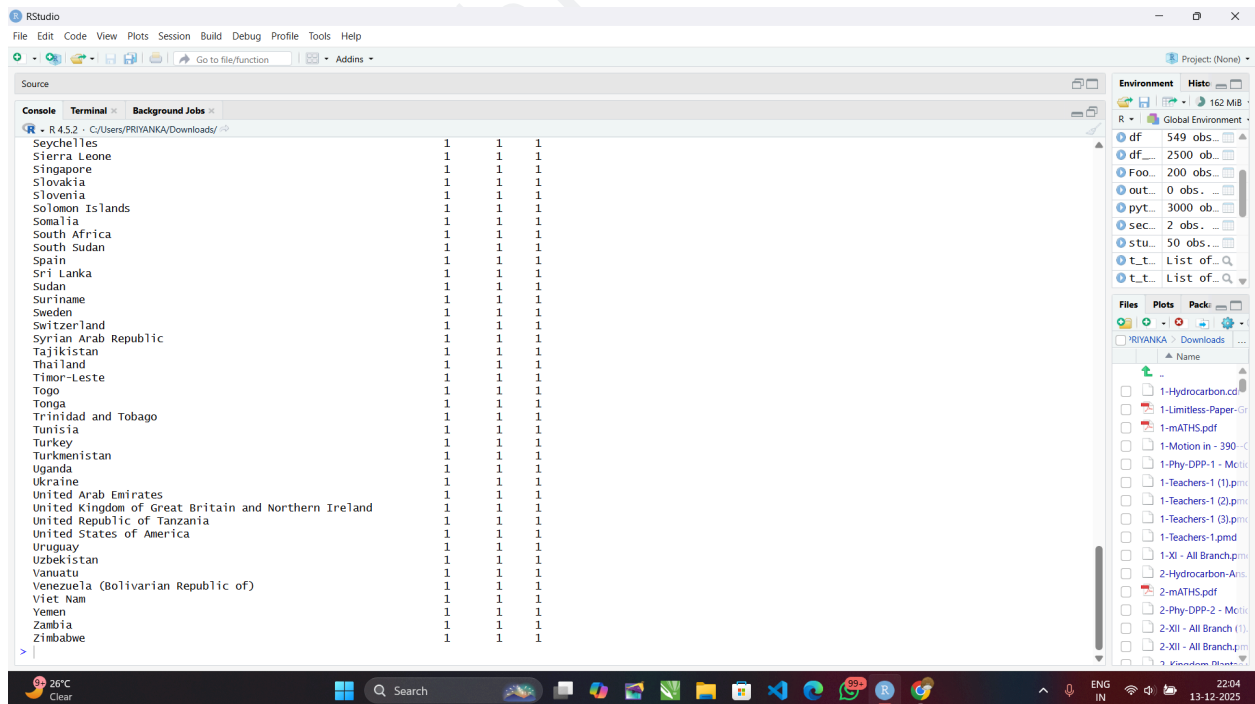
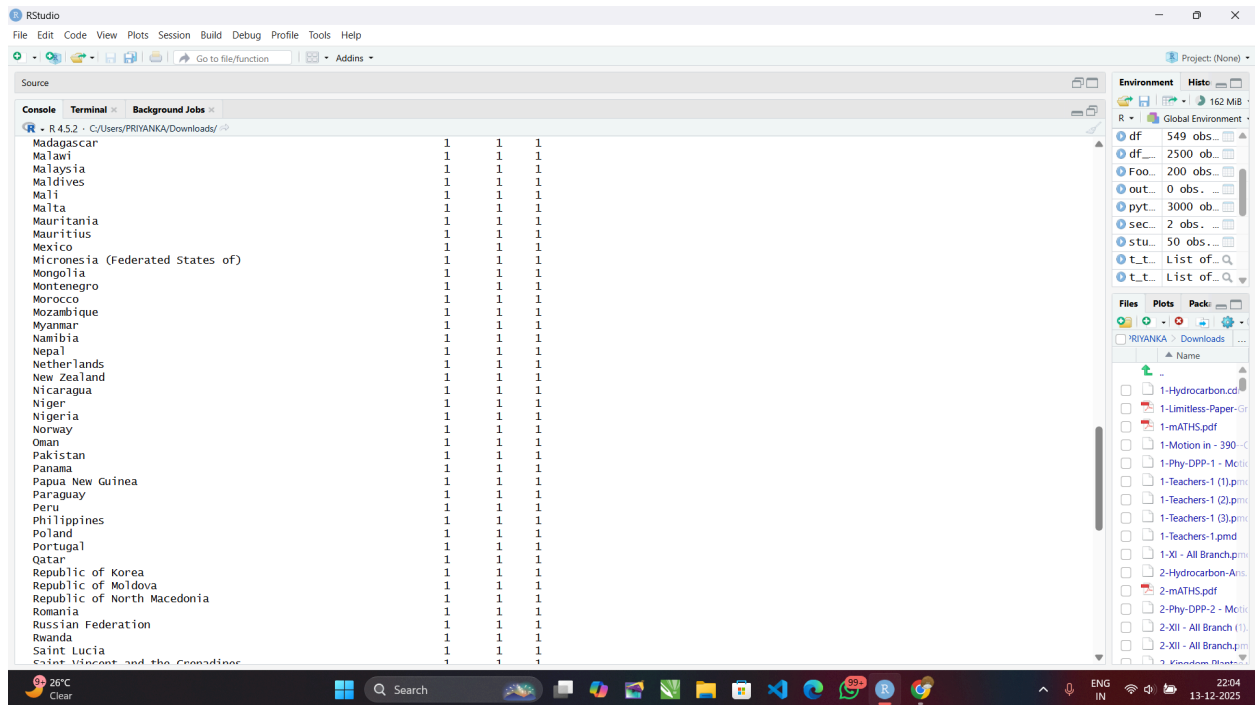
	Both sexes	Female	Male
Afghanistan	1	1	1
Albania	1	1	1
Algeria	1	1	1
Angola	1	1	1
Antigua and Barbuda	1	1	1
Argentina	1	1	1
Armenia	1	1	1
Australia	1	1	1
Austria	1	1	1
Azerbaijan	1	1	1
Bahamas	1	1	1
Bahrain	1	1	1
Bangladesh	1	1	1
Barbados	1	1	1
Belarus	1	1	1
Belgium	1	1	1
Belize	1	1	1
Benin	1	1	1
Bhutan	1	1	1
Bolivia (Plurinational State of)	1	1	1
Bosnia and Herzegovina	1	1	1
Botswana	1	1	1
Brazil	1	1	1
Brunei Darussalam	1	1	1
Bulgaria	1	1	1
Burkina Faso	1	1	1
Burundi	1	1	1
Cabo Verde	1	1	1
Cambodia	1	1	1
Cameroon	1	1	1
Canada	1	1	1
Central African Republic	1	1	1

The screenshot shows the continuation of the RStudio output. The console displays the following R code and its output:

```
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
Croatia
Cuba
Cyprus
Czechia
Democratic People's Republic of Korea
Democratic Republic of the Congo
Denmark
Djibouti
Dominican Republic
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Estonia
Eswatini
Ethiopia
Fiji
Finland
France
Gabon
Gambia
Georgia
Germany
Ghana
Greece
Grenada
Guatemala
Guinea
Guinea-Bissau
Guyana
Haiti
Honduras
Hungary
Iceland
India
Indonesia
Iran (Islamic Republic of)
Iraq
Ireland
Israel
```

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

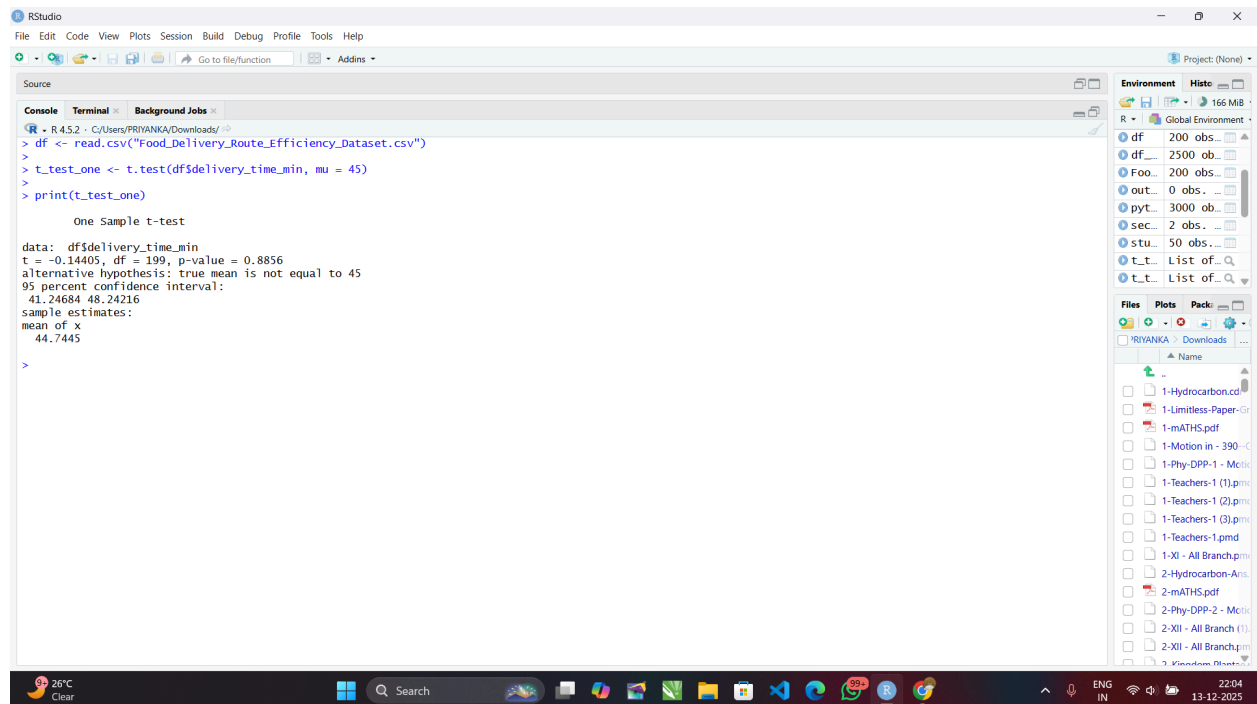
## SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R



**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Aim:** Performing one-sample t-tests using t.test() (R).

**OUTPUT:**



The screenshot shows the RStudio interface with the following content:

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R 4.5.2 - C:/Users/PRIVANKA/Downloads/
> df <- read.csv("Food_Delivery_Route_Efficiency_Dataset.csv")
>
> t_test_one <- t.test(df$delivery_time_min, mu = 45)
>
> print(t_test_one)

One Sample t-test

data: df$delivery_time_min
t = -0.14405, df = 199, p-value = 0.8856
alternative hypothesis: true mean is not equal to 45
95 percent confidence interval:
 41.24684 48.24216
sample estimates:
mean of x
 44.7445
>
```

The Environment pane on the right shows the following objects:

- df: 200 obs. ...
- df...: 2500 obs. ...
- Food...: 200 obs. ...
- out...: 0 obs. ...
- pyt...: 3000 obs. ...
- sec...: 2 obs. ...
- stu...: 50 obs. ...
- t.t...: List of...
- t.t...: List of...

The Files pane shows the following files:

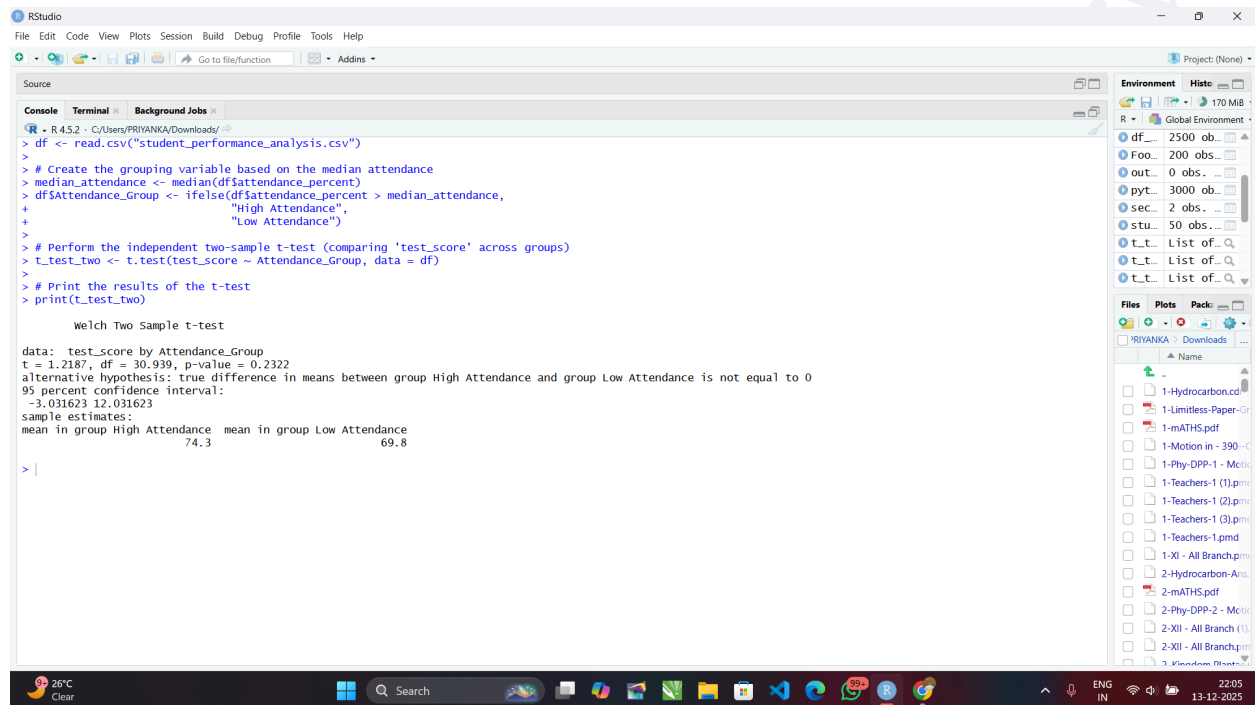
- 1-Hydrocarbon.co
- 1-Limitless-Paper-Gr
- 1-mATHS.pdf
- 1-Motion in - 390-C
- 1-Phy-DPP-1 - Moti
- 1-Teachers-1 (1).pm
- 1-Teachers-1 (2).pm
- 1-Teachers-1 (3).pm
- 1-Teachers-1.pmd
- 1-XI - All Branch.p
- 2-Hydrocarbon-Anis
- 2-mATHS.pdf
- 2-Phy-DPP-2 - Moti
- 2-XII - All Branch (1)
- 2-XII - All Branch.p
- 3-Kinematics\_Diagra

The Windows taskbar at the bottom shows the date and time as 13-12-2025, 22:04.

**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Aim:** Performing independent two-sample t-tests using `t.test()` with grouping (R).

**OUTPUT:**



The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains the R script for reading data, creating a grouping variable, and performing a t-test.
- Console:** Shows the output of the t-test, including the Welch Two Sample t-test results.
- Environment:** Lists the objects created in the R session, including the data frame and the t-test results.
- Files:** Shows the file explorer with various documents and folders.

```
> df <- read.csv("student_performance_analysis.csv")
>
> # Create the grouping variable based on the median attendance
> median_attendance <- median(df$attendance_percent)
> df$Attendance_Group <- ifelse(df$attendance_percent > median_attendance,
+                               "High Attendance",
+                               "Low Attendance")
>
> # Perform the independent two-sample t-test (comparing 'test_score' across groups)
> t_test_two <- t.test(test_score ~ Attendance_Group, data = df)
>
> # Print the results of the t-test
> print(t_test_two)
```

Welch Two Sample t-test

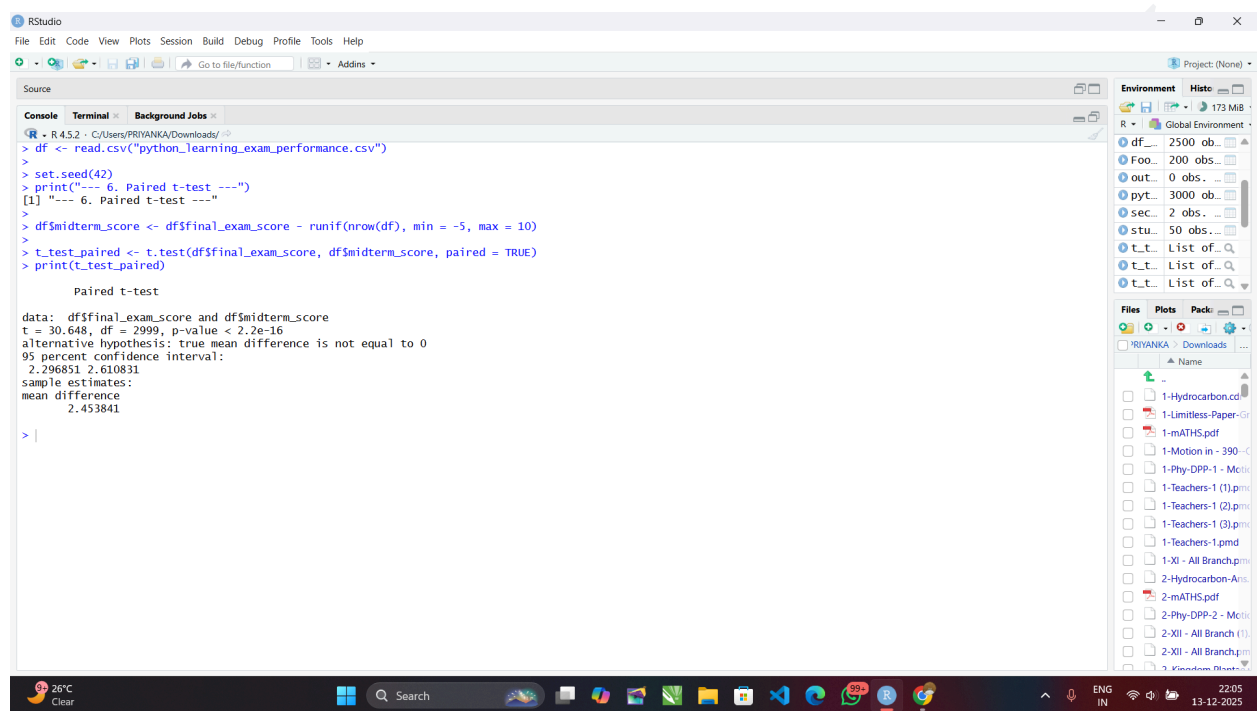
data: test\_score by Attendance\_Group  
t = 1.2187, df = 30.939, p-value = 0.2322  
alternative hypothesis: true difference in means between group High Attendance and group Low Attendance is not equal to 0  
95 percent confidence interval:  
-3.031623 12.031623  
sample estimates:  
mean in group High Attendance mean in group Low Attendance  
74.3 69.8

```
> |
```

**SHETH L.U.J AND SIR M.V. COLLEGE**  
**SUBJECT NAME: DATA ANALYSIS WITH SAS/SPSS/R**

**Aim:** Performing paired t-tests using `t.test(paired = TRUE)` (R).

**OUTPUT:**



The screenshot shows the RStudio interface with the following content:

```
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
> df <- read.csv("python_learning_exam_performance.csv")
>
> set.seed(42)
> print("---- 6. Paired t-test ----")
[1] "---- 6. Paired t-test ----"
>
> df$midterm_score <- df$final_exam_score - runif(nrow(df), min = -5, max = 10)
>
> t_test_paired <- t.test(df$final_exam_score, df$midterm_score, paired = TRUE)
> print(t_test_paired)

Paired t-test

data: df$final_exam_score and df$midterm_score
t = 30.648, df = 2999, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 2.296851 2.610831
sample estimates:
mean difference
      2.453841

> |
```

The Environment pane on the right shows the following objects:

- df: 2500 obs...
- Foo: 200 obs...
- out: 0 obs...
- pyt: 3000 obs...
- sec: 2 obs...
- stu: 50 obs...
- t.t: List of...
- t.t: List of...
- t.t: List of...

The Files pane shows the following files:

- 1-Hydrocarbon-Gr...
- 1-Limitless-Paper-Gr...
- 1-mATHS.pdf
- 1-Motion in - 390-C...
- 1-Phy-DPP-1 - Mote...
- 1-Teachers-1 (1).pm...
- 1-Teachers-1 (2).pm...
- 1-Teachers-1 (3).pm...
- 1-Teachers-1.pmd
- 1-XI - All Branch.p...
- 2-Hydrocarbon-Aris...
- 2-mATHS.pdf
- 2-Phy-DPP-2 - Mote...
- 2-XI - All Branch (1)...
- 2-XI - All Branch (2)...
- 2-XI - All Branch (3)...

The Windows taskbar at the bottom shows the date and time as 13-12-2025, 22:05.