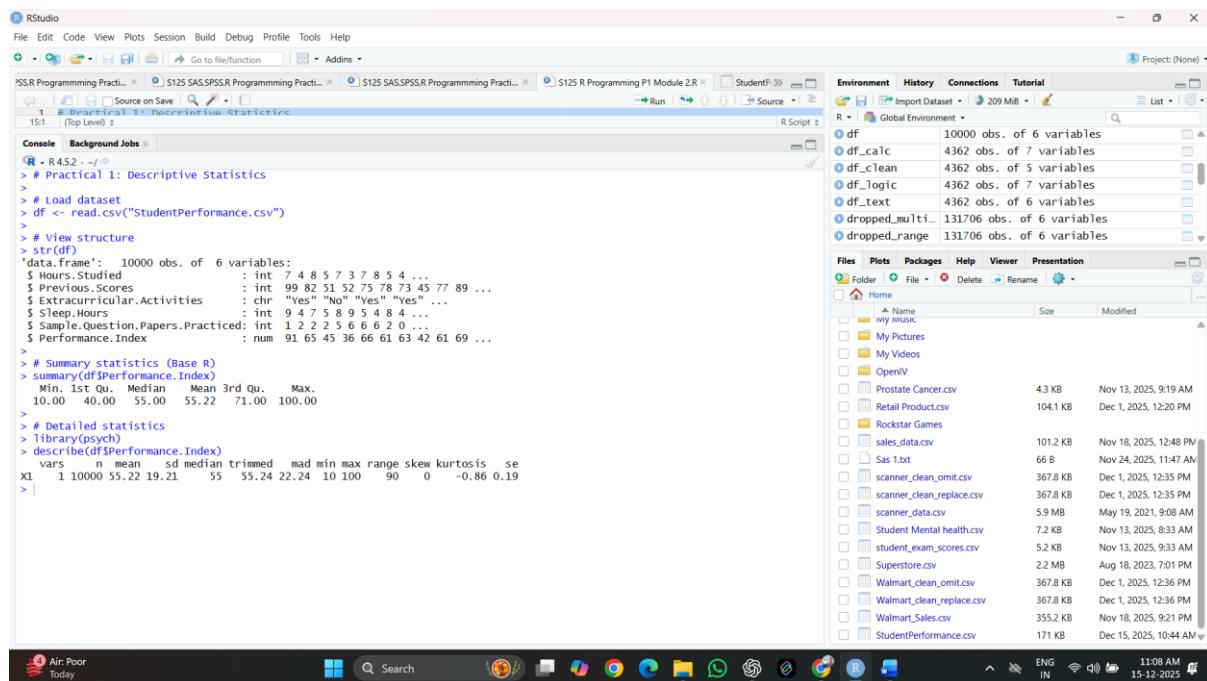


**Sheth L.U.J & Sir M.V College**  
**SAS?SPSS?R Programming**  
**Practical No:1 to 6 Module 2**

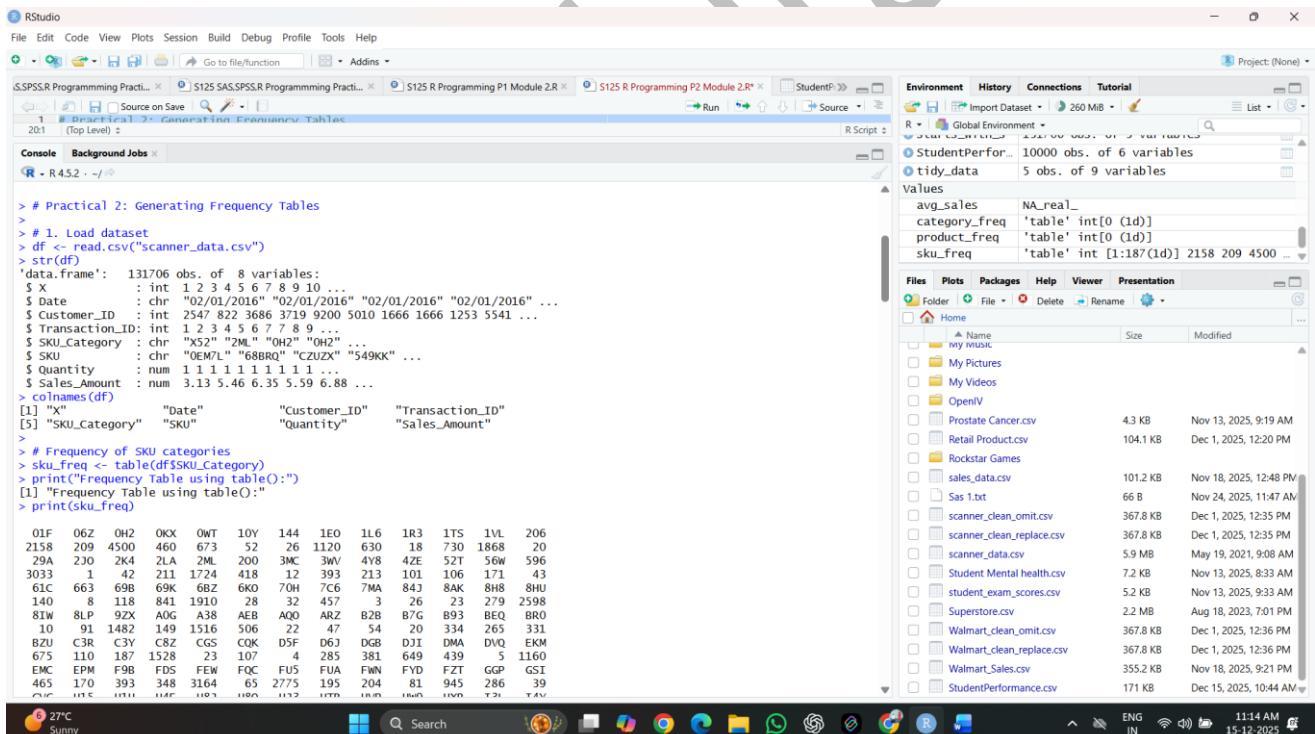
**Aim 1: Generating descriptive statistics using summary() or describe() (R).**



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Go to file/function Addins
S125 SAS,SPSS,R Programming Pract... S125 SAS,SPSS,R Programming Pract... S125 SAS,SPSS,R Programming Pract... S125 R Programming P1 Module 2.R StudentP...
15c1 # Practical 1: Descriptive Statistics
R Script t
Console Background Jobs
R - R 4.5.2 - ~
> # Practical 1: Descriptive statistics
>
> # Load dataset
> df <- read.csv("StudentPerformance.csv")
>
> # View structure
> str(df)
'data.frame': 10000 obs. of 6 variables:
 $ Hours.Studied : int 7 4 8 5 7 3 7 8 5 4 ...
 $ Previous.Scores : int 99 82 51 52 75 78 73 45 77 89 ...
 $ Extracurricular.Activities : chr "Yes" "No" "Yes" "Yes" ...
 $ Sleep.Hours : int 9 4 7 5 8 9 5 4 8 4 ...
 $ Sample.Question.Papers.Practiced: int 1 2 2 2 5 6 6 2 0 ...
 $ Performance.Index : num 91 65 45 36 66 61 63 42 61 69 ...
>
> # Summary statistics (Base R)
> summary(df$Performance.Index)
Min. 1st Qu. Median Mean 3rd Qu. Max.
10.00 40.00 55.00 55.22 71.00 100.00
>
> # Detailed statistics
> library(psych)
> describe(df$Performance.Index)
vars n mean sd median trimmed mad min max range skew kurtosis se
X1 1 10000 55.22 19.21 55 55.24 22.24 10 100 90 0 -0.86 0.19
> |
```

The screenshot shows the RStudio interface with multiple tabs open. The main pane displays R code for loading a dataset and calculating descriptive statistics. The Environment pane on the right shows various objects defined in the session, including the loaded dataset and summary statistics.

**Aim 2: Generating frequency tables using table() or count() (R).**



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Go to file/function Addins
S125 SAS,SPSS,R Programming Pract... S125 SAS,SPSS,R Programming Pract... S125 R Programming P1 Module 2.R S125 R Programming P2 Module 2.R StudentP...
20:1 # Practical 2: Generating Frequency Tables
R Script t
Console Background Jobs
R - R 4.5.2 - ~
> # Practical 2: Generating Frequency Tables
>
> # 1. Load dataset
> df <- read.csv("scanner_clean_data.csv")
> str(df)
'data.frame': 131706 obs. of 8 variables:
 $ X : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Date : chr "02/01/2016" "02/01/2016" "02/01/2016" "02/01/2016" ...
 $ Customer_ID : int 2547 822 3678 3719 9200 5010 1666 1666 1253 5541 ...
 $ Transaction_ID : int 1 2 3 4 5 6 7 7 8 9 ...
 $ SKU_Category : chr "X52" "ZM" "OH2" "OH2" ...
 $ SKU : chr "0EM\1" "68BRQ" "CZUZ" "549KK" ...
 $ Quantity : num 1 1 1 1 1 1 1 1 1 ...
 $ Sales_Amount : num 3.13 5.46 6.35 5.59 6.88 ...
> colnames(df)
[1] "X"          "Date"        "Customer_ID" "Transaction_ID"
[5] "SKU_Category" "SKU"         "Quantity"    "Sales_Amount"
>
> # Frequency of SKU categories
> sku_freq <- table(df$SKU.Category)
> print("Frequency Table using table():")
[1] "Frequency Table using table():"
> print(sku_freq)
OIE      06Z     OH2     0KX     0WY     10Y     144     1E0     116     1R3     1TS     1VL     206
2158    209    4500    460    673     52     26    1120    630     18    730    1868    20
29A     210    284     24    2M     200    3MC    3WV    4Y8     4Z     527    569    596
303     1     42    211    174     418     12    392    213     101    106    171     43
61C     663    698    59K    68Z    6K0     7H     7CG    7644    843    8AK    81U
140     8     118    841    1910    28     32    457     3     26    23    279    2598
8IW     8LP    92X    A0G    A3B    AEB    A0Q    ARZ    B2B    B7G    B93    BEQ    BRO
10     91    1482    149    1316    506     22    47     54     20    334    265    331
BZU     C3R    C3Y    C8Z    CGS    COK    D5F    D6J    D8L    D9A    DVQ    EKM
675     110    187    1528    23    107     4    285    381    649    439     5    1160
EMC     EPM    F9B    FDS    FEW    FQC    FUS    FUA    FWN    FYD    FZT    GGP    GSI
465     170    393    348    3164    65    2775    195    204     81    945    286     39
OIE     OIE    OIE    OIE    OIE    OIE    OIE    OIE    OIE    OIE    OIE    OIE    OIE
ENG IN 11:08 AM 15-12-2025
```

The screenshot shows the RStudio interface with multiple tabs open. The main pane displays R code for generating frequency tables using the `table()` function. The Environment pane on the right shows the resulting frequency table object and its components.

# Sheth L.U.J & Sir M.V College

## SAS?SPSS?R Programming

### Practical No:1 to 6 Module 2

RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/folder Addins

Project: (None)

ISPPSS.R Programming Practice... S125 SAS,SPSS.R Programming Practice... S125 R Programming P1 Module 2.R\* S125 R Programming P2 Module 2.R\* StudentPerformance.R

1 Practical > Generating Frequency Tables

201 (Top Level) t R Script

Console Background Jobs

R 4.5.2 - ✓

	ZLZ	2853	3Z	485b	>4b	3b/	ZLU	B6	-48	/45	2858	1811	14
QLB	QDN	QOR	QV7	R68	R6E	RF5	RML	RU6	S30	SFC	SIM	S35	
566	1	24	521	169	5099	14	497	393	14	1381	328	1899	
SKG	T80	TEU	TIF	TVL	TW8	TZB	U3N	USF	UZI	V9H	VDO		
97	383	170	108	547	1913	1241	2	4570	22	12	575	9	
VQD	W7	W41	WHL	WWM	X52	XAE	XG4	XVK	XX1	XZC	YJM	Z23	
483	50	100	94	9	1852	284	1995	196	90	30	1406	606	
Z40	Z99	Z3G	ZX5	ZYU									
708	48	28	139	131									

> # Load dplyr library  
> library(dplyr)  
>  
> # Frequency using count()  
> sku\_count <- df %>% count(sku\_Category)  
> print("Frequency Table using dplyr count():")  
[1] "Frequency Table using dplyr count():"  
> print(sku\_count)

SKU_Category	n
OIF	2158
OEF	209
OHZ	4500
OKX	460
Ort	673
10Y	52
144	26
1E0	1120
1L6	630
1R3	18
1TS	730
1VL	1868
206	20
29A	3033
2J0	1
2K4	42
2LA	211
2ML	1724
200	418

Environment History Connections Tutorial

Global Environment StudentPerformance tidy\_data

Values avg\_sales NA\_real\_ category\_freq 'table' int[0 (1d)] product\_freq 'table' int[0 (1d)] sku\_freq 'table' int [1:187(1d)] 2158 209 4500 ...

Files Plot Packages Help Viewer Presentation

Home

Name	Size	Modified
my music		
My Pictures		
My Videos		
OpenIV		
Prostate Cancer.csv	4.3 KB	Nov 13, 2025, 9:19 AM
Retail Product.csv	104.1 KB	Dec 1, 2025, 12:20 PM
Rockstar Games		
sales_data.csv	101.2 KB	Nov 18, 2025, 12:48 PM
Sas_1.txt	66 B	Nov 24, 2025, 11:47 AM
scanner_clean_omit.csv	367.8 KB	Dec 1, 2025, 12:35 PM
scanner_clean_replace.csv	367.8 KB	Dec 1, 2025, 12:35 PM
scanner_data.csv	5.9 MB	May 19, 2021, 9:08 AM
Student Mental health.csv	7.2 KB	Nov 13, 2025, 8:33 AM
student_exam_scores.csv	5.2 KB	Nov 13, 2025, 9:33 AM
Superstore.csv	2.2 MB	Aug 18, 2023, 7:01 PM
Walmart_clean_omit.csv	367.8 KB	Dec 1, 2025, 12:36 PM
Walmart_clean_replace.csv	367.8 KB	Dec 1, 2025, 12:36 PM
Walmart_Sales.csv	355.2 KB	Nov 18, 2025, 9:21 AM
StudentPerformance.csv	171 KB	Dec 15, 2025, 10:44 AM

27°C Sunny ENG IN WiFi 11:15 AM 15-12-2025

The screenshot shows the RStudio interface with several tabs open:

- SAS/SPSS.R Programming Pract...
- S12 SAS/SPSS.R Programming Pract...
- S12 R Programming P1 Module 2.R
- S12 R Programming P2 Module 2.R\*

In the top right, there's a "Project: (None)" indicator.

The main area includes:

- A "Console" tab showing R output:

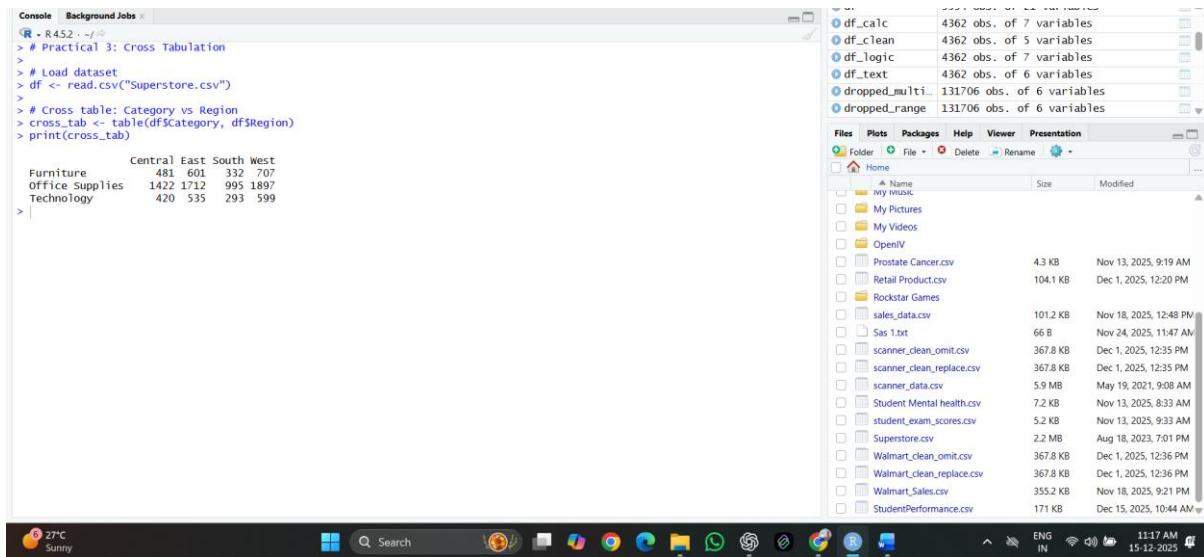
```
152   RU6  393
153   S30  14
154   SFC  1381
155   SIM  328
156   SJS  1899
157   SKG  97
158   T80  383
159   TEU  170
160   TIF  108
161   TVL  547
162   TW8  1913
163   T2B  1241
164   U3N  2
165   USP  4570
166   UCR  22
167   UZI  12
168   V9H  575
169   VD0  9
170   VQD  483
171   W7T  50
172   W41  100
173   WHL  94
174   WMW  9
175   XS2  1852
176   XAE  284
177   XG4  1995
178   XVK  196
179   XXL  90
180   XZC  30
181   YMJ  1406
182   Z23  606
183   Z40  708
184   Z99  48
185   ZJG  28
186   ZX5  139
187   ZYU  131
```
- A "Background Jobs" tab showing "R 4.5.2 - ~" with 20.1 items.
- An "Environment" tab showing global variables:

Name	Type	Value
avg_sales	NA_real_	
category_freq	"table"	int[0:1d]
product_freq	"table"	int[0:1d]
sku_freq	"table"	int[1:187(1d)] 2158 209 4500

- A "Files" tab showing a file tree with various CSV files and their details.

**Sheth L.U.J & Sir M.V College**  
**SAS?SPSS?R Programming**  
**Practical No:1 to 6 Module 2**

**Aim 3: Creating cross-tabulations and two-way tables using table() (R).**



The screenshot shows the RStudio interface. The console window displays R code for creating a cross-tabulation table:

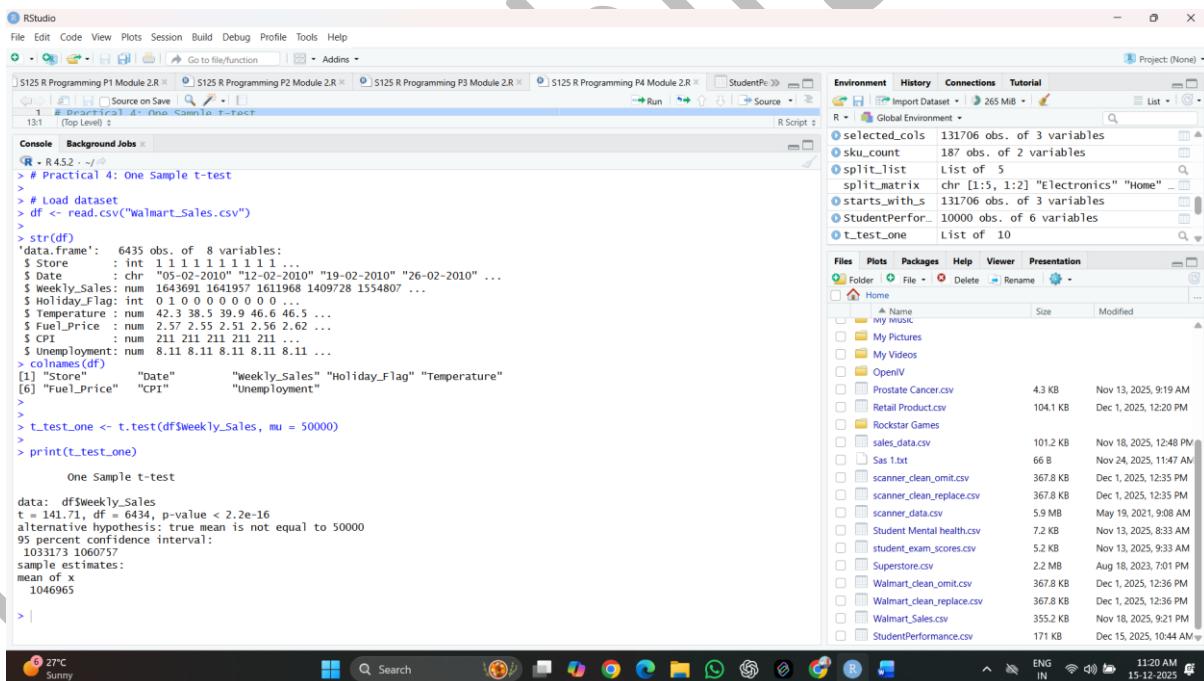
```
R > # Practical 3: Cross Tabulation
> 
> # Load dataset
> df <- read.csv("Superstore.csv")
> 
> # Cross table: Category vs Region
> cross_tab <- table(df$category, df$Region)
> print(cross_tab)
```

The output shows a table:

	Central	East	South	West
Furniture	481	601	332	707
Office Supplies	1422	1712	995	1897
Technology	420	535	293	599

The file browser on the right lists various CSV files in the working directory.

**Aim 4: Performing one-sample t-tests using t.test() (R).**



The screenshot shows the RStudio interface. The console window displays R code for performing a one-sample t-test on weekly sales data:

```
R > # Practical 4: One Sample t-test
> 
> # Load dataset
> df <- read.csv("walmart_Sales.csv")
> 
> str(df)
'data.frame': 6435 obs. of 8 variables:
 $ Store      : int 1 1 1 1 1 1 1 1 ...
 $ Date       : chr "05-02-2010" "12-02-2010" "19-02-2010" "26-02-2010" ...
 $ Weekly_Sales: num 1832 1641937 161968 1409728 1554807 ...
 $ Holiday_Flag: num 0 1 0 0 1 0 0 0 0 ...
 $ Temperature: num 42.3 38.5 39.9 46.6 46.5 ...
 $ Fuel_Price  : num 2.57 2.55 2.51 2.56 2.62 ...
 $ CPI        : num 211.211 211.211 211.211 ...
 $ Unemployment: num 8.11 8.11 8.11 8.11 8.11 ...
> colnames(df)
[1] "Store"          "Date"           "Weekly_Sales"   "Holiday_Flag"   "Temperature"
[6] "Fuel_Price"     "CPI"            "Unemployment"
```

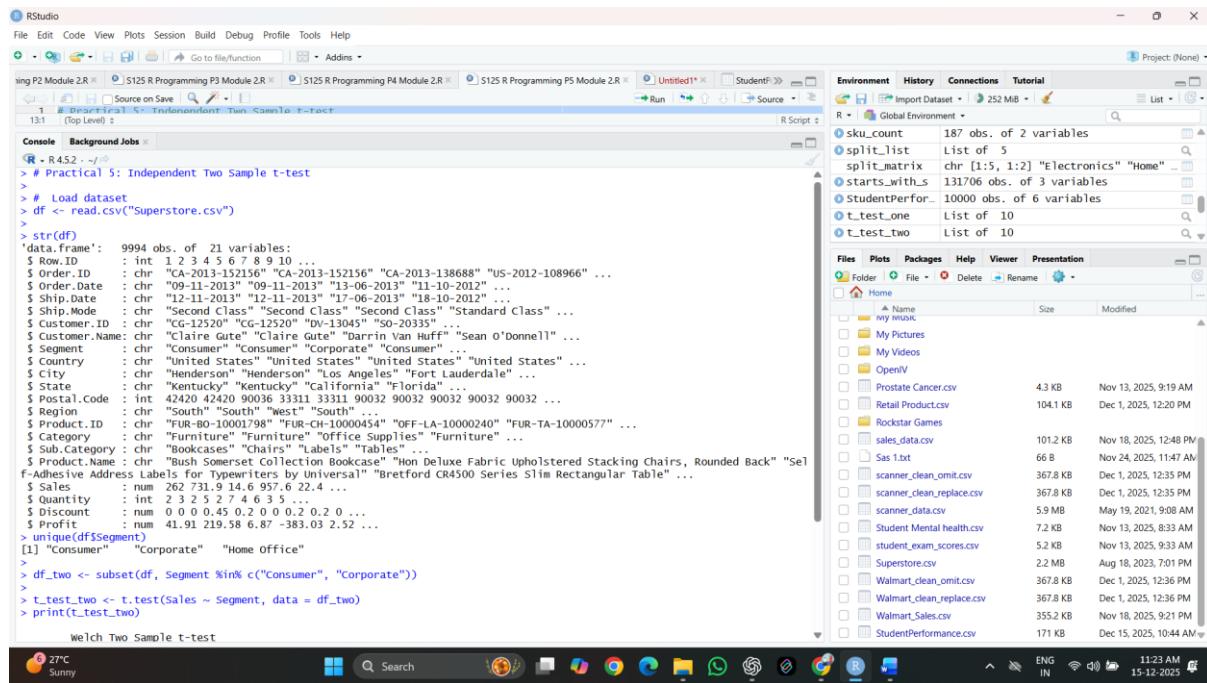
The output shows the results of the t-test:

```
One Sample t-test

data: df$Weekly_Sales
t = 141.71, df = 6434, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50000
95 percent confidence interval:
 1033173 1060757
sample estimates:
mean of x
1046965
```

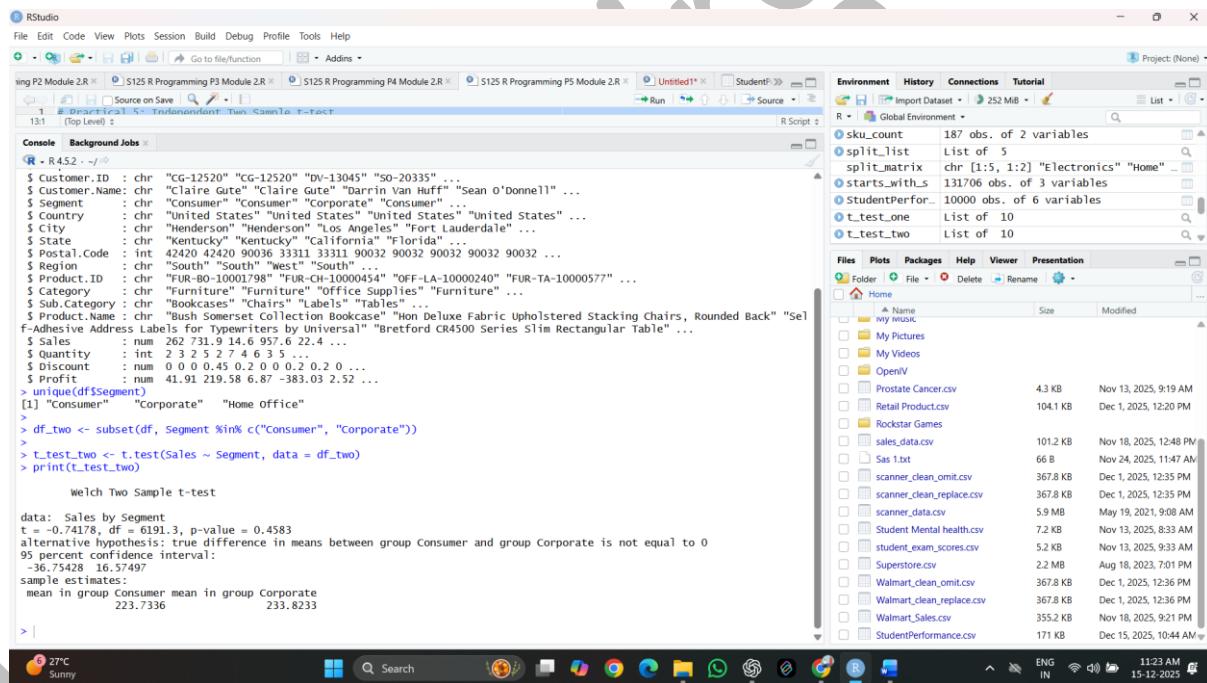
**Sheth L.U.J & Sir M.V College**  
**SAS?SPSS?R Programming**  
**Practical No:1 to 6 Module 2**

**Aim 5: Performing independent two-sample t-tests using t.test() with grouping (R).**



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Source On Save Go to file/folder Addins ...
131 (Top Level) 
R Script t
Environment History Connections Tutorial
Project: (None)
R Global Environment
  sku_count 187 obs. of 2 variables
  split_list List of 5
  split_matrix chr [1:5, 1:2] "Electronics" "Home" ...
  starts_with_s 131706 obs. of 3 variables
  StudentPerfor_ 10000 obs. of 6 variables
  t_test_one List of 10
  t_test_two List of 10
  Files Plots Packages Help Viewer Presentation
  Folder File Delete Rename ...
  Home
    - Name Size Modified
      my music
      My Pictures
      My Videos
      OpenV
        Prostate Cancer.csv 4.3 KB Nov 13, 2025, 9:19 AM
        Retail Product.csv 104.1 KB Dec 1, 2025, 12:20 PM
        Rockstar Games
        sales_data.csv 101.2 KB Nov 18, 2025, 12:48 PM
        Sas 1.txt 66 B Nov 24, 2025, 11:47 AM
        scanner_clean omit.csv 367.8 KB Dec 1, 2025, 12:35 PM
        scanner_clean replace.csv 367.8 KB Dec 1, 2025, 12:35 PM
        scanner_data.csv 5.9 MB May 19, 2021, 9:08 AM
        Student Mental health.csv 7.2 KB Nov 13, 2025, 8:33 AM
        student_exam_scores.csv 5.2 KB Nov 13, 2025, 9:33 AM
        Superstore.csv 2.2 MB Aug 18, 2023, 7:01 PM
        Walmart_clean omit.csv 367.8 KB Dec 1, 2025, 12:36 PM
        Walmart_clean replace.csv 367.8 KB Dec 1, 2025, 12:36 PM
        Walmart_sales.csv 355.2 KB Nov 18, 2025, 9:21 PM
        StudentPerformance.csv 171 KB Dec 15, 2025, 10:44 AM ...
  27°C Sunny 11:23 AM IN 15-12-2025 ENG
```

R> # Practical 5: Independent Two Sample t-test  
> # Load dataset  
> df <- read.csv("Superstore.csv")  
>  
> str(df)  
'data.frame': 9994 obs. of 21 variables:  
\$ Row.ID : int 1 2 3 4 5 6 7 8 9 10 ...  
\$ Order.ID : chr "CA-2013-152156" "CA-2013-152156" "CA-2013-138688" "US-2012-108966" ...  
\$ Order.Date : chr "09-11-2013" "09-11-2013" "13-06-2013" "11-10-2012" ...  
\$ Ship.Date : chr "12-11-2013" "12-11-2013" "17-06-2013" "18-10-2012" ...  
\$ Ship.Mode : chr "Second Class" "Second Class" "Second Class" "Standard Class" ...  
\$ Customer.ID : chr "CG-12520" "CG-12520" "DV-13045" "SO-20335" ...  
\$ Customer.Name: chr "Claire Gute" "Claire Gute" "Darrin Van Huff" "Sean O'Donnell" ...  
\$ Segment : chr "Consumer" "Consumer" "Corporate" "Consumer" ...  
\$ Country : chr "United States" "United States" "United States" "United States" ...  
\$ City : chr "Henderson" "Henderson" "Los Angeles" "Fort Lauderdale" ...  
\$ State : chr "Kentucky" "Kentucky" "California" "Florida" ...  
\$ Postal.Code : int 42420 42420 90032 33311 33311 90032 90032 90032 90032 ...  
\$ Region : chr "South" "South" "West" "South" ...  
\$ Product.ID : chr "FUR-B0-1000198" "FUR-CH-1000045" "OFF-LA-10000240" "FUR-TA-10000577" ...  
\$ Category : chr "Furniture" "Furniture" "Office Supplies" "Furniture" ...  
\$ Sub.Category: chr "Bookcases" "Chairs" "Labels" "Tables" ...  
\$ Product.Name : chr "Bush Somerset Collection Bookcase" "Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back" "Sel F-Adhesive Address Labels for Typewriters by Universal" "Bretford CR4500 Series Slim Rectangular Table" ...  
F-Adhesive Address Labels for Typewriters by Universal" "Bretford CR4500 Series Slim Rectangular Table" ...  
 \$ Sales : num 262 731.9 14.6 957.6 22.4 ...  
 \$ Quantity : int 2 3 2 5 2 7 4 6 3 5 ...  
 \$ Discount : num 0 0 0.45 0.2 0 0.2 0.2 0 ...  
 \$ Profit : num 41.91 219.58 6.87 -383.03 2.52 ...  
> unique(df\$Segment)  
[1] "Consumer" "Corporate" "Home Office"  
>  
> df\_two <- subset(df, Segment %in% c("Consumer", "Corporate"))  
>  
> t\_test\_two <- t.test(Sales ~ Segment, data = df\_two)  
> print(t\_test\_two)

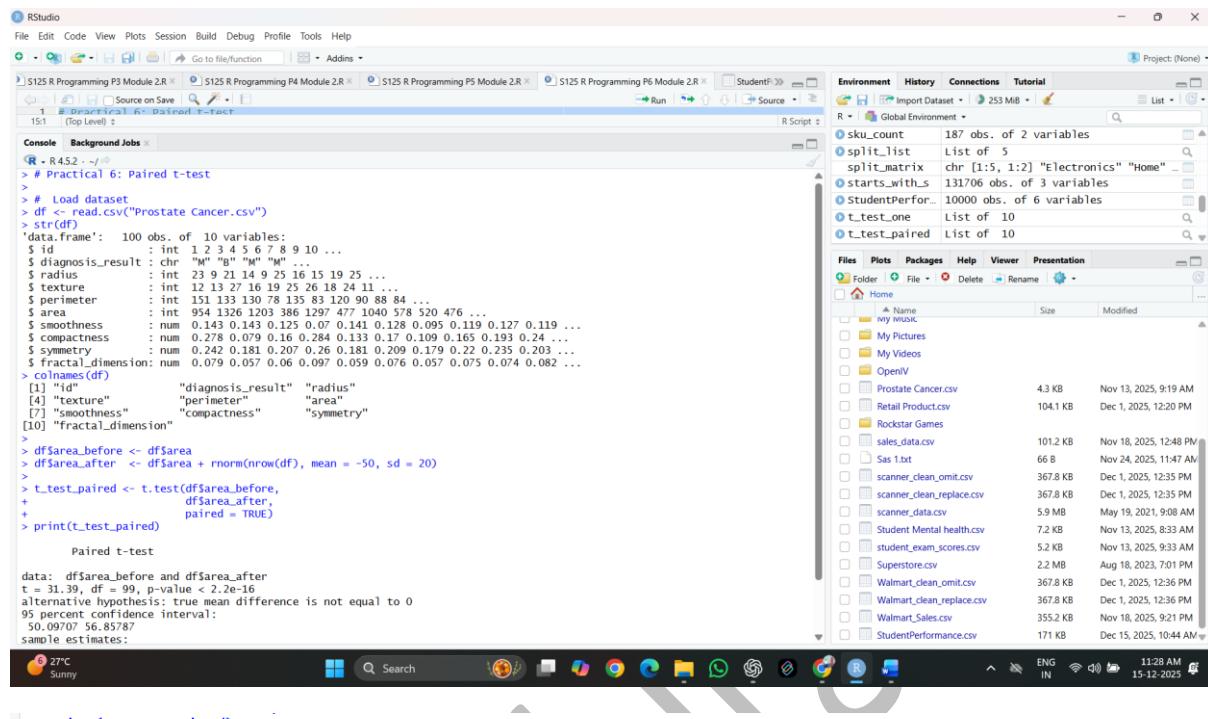


```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Source On Save Go to file/folder Addins ...
131 (Top Level) 
R Script t
Environment History Connections Tutorial
Project: (None)
R Global Environment
  sku_count 187 obs. of 2 variables
  split_list List of 5
  split_matrix chr [1:5, 1:2] "Electronics" "Home" ...
  starts_with_s 131706 obs. of 3 variables
  StudentPerfor_ 10000 obs. of 6 variables
  t_test_one List of 10
  t_test_two List of 10
  Files Plots Packages Help Viewer Presentation
  Folder File Delete Rename ...
  Home
    - Name Size Modified
      my music
      My Pictures
      My Videos
      OpenV
        Prostate Cancer.csv 4.3 KB Nov 13, 2025, 9:19 AM
        Retail Product.csv 104.1 KB Dec 1, 2025, 12:20 PM
        Rockstar Games
        sales_data.csv 101.2 KB Nov 18, 2025, 12:48 PM
        Sas 1.txt 66 B Nov 24, 2025, 11:47 AM
        scanner_clean omit.csv 367.8 KB Dec 1, 2025, 12:35 PM
        scanner_clean replace.csv 367.8 KB Dec 1, 2025, 12:35 PM
        scanner_data.csv 5.9 MB May 19, 2021, 9:08 AM
        Student Mental health.csv 7.2 KB Nov 13, 2025, 8:33 AM
        student_exam_scores.csv 5.2 KB Nov 13, 2025, 9:33 AM
        Superstore.csv 2.2 MB Aug 18, 2023, 7:01 PM
        Walmart_clean omit.csv 367.8 KB Dec 1, 2025, 12:36 PM
        Walmart_clean replace.csv 367.8 KB Dec 1, 2025, 12:36 PM
        Walmart_sales.csv 355.2 KB Nov 18, 2025, 9:21 PM
        StudentPerformance.csv 171 KB Dec 15, 2025, 10:44 AM ...
  27°C Sunny 11:23 AM IN 15-12-2025 ENG
```

R> Welch Two Sample t-test  
  
 data: Sales by Segment  
 t = -0.74178, df = 6191.3, p-value = 0.4583  
 alternative hypothesis: true difference in means between group Consumer and group Corporate is not equal to 0  
 95 percent confidence interval:  
 -36.75428 16.57497  
 sample estimates:  
 mean in group Consumer mean in group Corporate  
 223.7336 233.8233  
> |

**Sheth L.U.J & Sir M.V College**  
**SAS?SPSS?R Programming**  
**Practical No:1 to 6 Module 2**

**Aim 6: Performing paired t-tests using t.test(paired=TRUE) (R).**



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Source Go file/function Addins
S125 R Programming P3 Module 2.R S125 R Programming P4 Module 2.R S125 R Programming P5 Module 2.R S125 R Programming P6 Module 2.R StudentF...
R Script
Console Background Jobs
R - R 4.5.2 - ~/...
> # Practical 6: Paired t-test
>
> # Load dataset
> df <- read.csv("Prostate Cancer.csv")
> str(df)
'data.frame': 100 obs. of 10 variables:
 $ id : int 1 2 3 4 5 6 7 8 9 10 ...
 $ diagnosis_result : chr "M" "B" "M" "M" ...
 $ radius : int 23 9 21 14 9 25 16 15 19 25 ...
 $ texture : int 12 13 27 16 19 25 26 18 24 11 ...
 $ perimeter : int 151 133 128 78 135 83 120 96 104 84 ...
 $ area : int 1490 1064 1205 388 1158 1040 978 520 476 ...
 $ smoothness : num 0.143 0.143 0.125 0.07 0.141 0.128 0.095 0.119 0.127 0.119 ...
 $ compactness : num 0.278 0.079 0.16 0.284 0.133 0.17 0.109 0.165 0.193 0.24 ...
 $ symmetry : num 0.242 0.181 0.207 0.26 0.181 0.209 0.179 0.22 0.235 0.203 ...
 $ fractal_dimension: num 0.079 0.057 0.06 0.097 0.059 0.076 0.057 0.075 0.074 0.082 ...
> colnames(df)
[1] "id"           "diagnosis_result" "radius"
[4] "texture"       "perimeter"      "area"
[7] "smoothness"    "compactness"    "symmetry"
[10] "fractal_dimension"
>
> df$area_before <- df$area
> df$area_after <- df$area + rnorm(nrow(df), mean = -50, sd = 20)
>
> t_test_paired <- t.test(df$area_before,
+                         df$area_after,
+                         paired = TRUE)
> print(t_test_paired)

Paired t-test

data: df$area_before and df$area_after
t = 31.39, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
50.09707 56.85787
sample estimates:
mean difference
53.47747

> |
```