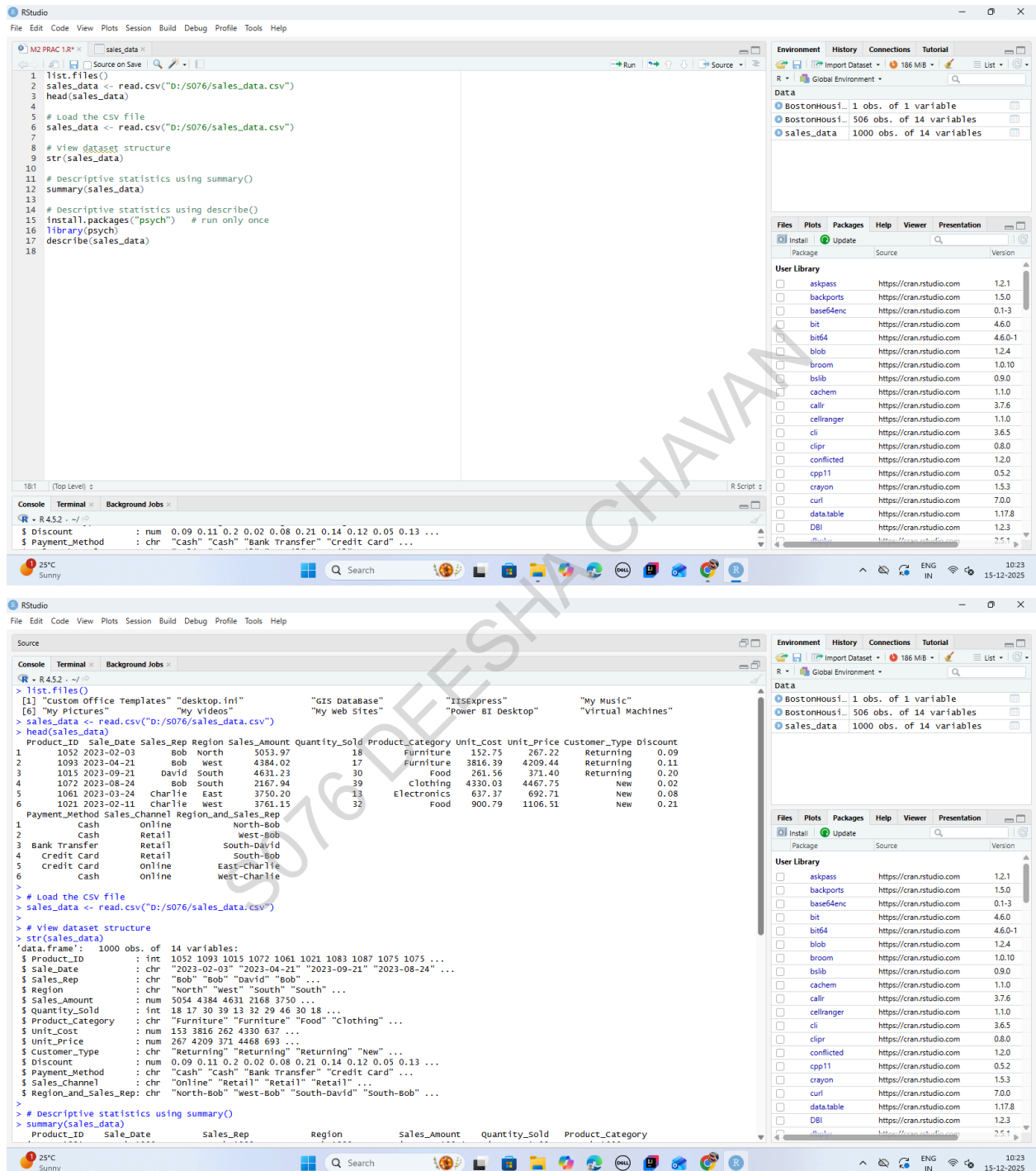


PRACTICAL NO.: M2 1 - 6

1. Generating descriptive statistics using `summary()` or `describe()` (R).



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

SUBJECT: DATA ANALYSIS WITH R

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

```
R > R452 ~./
$ Discount : num 0.09 0.11 0.2 0.02 0.08 0.21 0.14 0.12 0.05 0.13 ...
$ Payment_Method : chr "Cash" "Cash" "Bank Transfer" "Credit Card" ...
$ sales_channel : chr "online" "Retail" "Retail" "Retail" ...
$ Region_and_Sales_Rep: chr "North-Bob" "West-Bob" "South-David" "South-Bob" ...

> # Descriptive statistics using summary()
> summary(sales_data)
  Product_ID   Sale_Date   Sales_Rep   Region   Sales_Amount   Quantity_Sold   Product_Category
Min.   :1001   Length:1000   Length:1000   Length:1000   Min.   : 100.1   Min.   : 1.00   Length:1000
1st Qu.:1024   Class :character   Class :character   Class :character   1st Qu.:2550.3   1st Qu.:13.00   Class :character
Median :1051   Mode  :character   Mode  :character   Mode  :character   Median :5019.3   Median :25.00   Mode  :character
Mean   :1050                                     Mean   :5019.3   Mean   :25.36
3rd Qu.:1075                                     3rd Qu.:7507.4   3rd Qu.:38.00
Max.   :1100                                     Max.   :9989.0   Max.   :49.00
Unit_Cost   Unit_Price   Customer_Type   Discount   Payment_Method   sales_channel   Region_and_Sales_Rep
Min.   : 60.28   Min.   :167.1   Length:1000   Min.   :0.0000   Length:1000   Length:1000   Length:1000
1st Qu.:1238.38   1st Qu.:1509.1   Class :character   1st Qu.:0.0800   Class :character   Class :character   Class :character
Median :2467.24   Median :2696.4   Mode  :character   Median :0.1500   Mode  :character   Mode  :character   Mode  :character
Mean   :2475.30   Mean   :2728.4   Mean   :0.1324   Mean   :0.1324   Mean   :0.1324   Mean   :0.1324
3rd Qu.:3702.86   3rd Qu.:3958.0   3rd Qu.:0.2300   3rd Qu.:0.2300   3rd Qu.:0.2300   3rd Qu.:0.2300
Max.   :4995.30   Max.   :5442.1   Max.   :0.3000   Max.   :0.3000   Max.   :0.3000   Max.   :0.3000

> # Descriptive statistics using describe()
> install.packages("psych") # run only once

Restarting R session...

> install.packages("psych")

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools
before proceeding:

https://cran.rstudio.com/bin/windows/rtools/
Installing package into 'C:/Users/itlab/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/psych_2.5.6.zip'
Content type 'application/zip' length 3594552 bytes (3.4 MB)
downloaded 3.4 MB

package 'psych' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/itlab/AppData/Local/Temp/Rtmpacut4C/downloaded_packages
> |
```

The User Library pane on the right shows a list of installed packages:

Package	Source	Version
askpass	https://cran.rstudio.com	1.2.1
backports	https://cran.rstudio.com	1.5.0
base64enc	https://cran.rstudio.com	0.1-3
bit	https://cran.rstudio.com	4.6.0
bit64	https://cran.rstudio.com	4.6.0-1
blob	https://cran.rstudio.com	1.2.4
broom	https://cran.rstudio.com	1.0.10
bslib	https://cran.rstudio.com	0.9.0
cachem	https://cran.rstudio.com	1.1.0
callr	https://cran.rstudio.com	3.7.6
cellranger	https://cran.rstudio.com	1.1.0
cli	https://cran.rstudio.com	3.6.5
clipr	https://cran.rstudio.com	0.8.0
conflicted	https://cran.rstudio.com	1.2.0
cpp11	https://cran.rstudio.com	0.5.2
crayon	https://cran.rstudio.com	1.5.3
curl	https://cran.rstudio.com	7.0.0
data.table	https://cran.rstudio.com	1.17.8
DBI	https://cran.rstudio.com	1.2.3

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

The screenshot shows the RStudio interface with the following R code in the script editor:

```
1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Frequency table using table()
5 table(sales_data$Region)
6
7 # Frequency table using count()
8 library(dplyr)
9 sales_data %>% count(Region)
10
```

The console shows the output of the code:

```
10:1 (Top Level) >
R Script >
```

The User Library pane on the right shows the same list of installed packages as in the first screenshot.

NAME: DEESHA CHAVAN
ROLL NO.: S076

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

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

```
R - R 4.5.2 ~ /> # Load the CSV file
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
> # Frequency table using table()
> table(sales_data$Region)

East North South West
263  267  226  244
> # Frequency table using count()
> library(dplyr)
> sales_data %>% count(Region)

  Region    n
1  East  263
2  North 267
3  South 226
4  West  244
> |
```

The Environment pane on the right shows the loaded data objects:

- BostonHousi...: 1 obs. of 1 variable
- BostonHousi...: 506 obs. of 14 variables
- sales_data: 1000 obs. of 14 variables

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

The screenshot shows the RStudio interface with a script file named "M2 PRAC 1.R". The console window displays the following R code and its output:

```
R - R 4.5.2 ~ /> # Load the CSV file
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
> # Create two-way table using table()
> table(sales_data$Region, sales_data$Sales_Channel)

      online Retail
East   121   142
North  130   137
South  112   114
West   125   119
> |
```

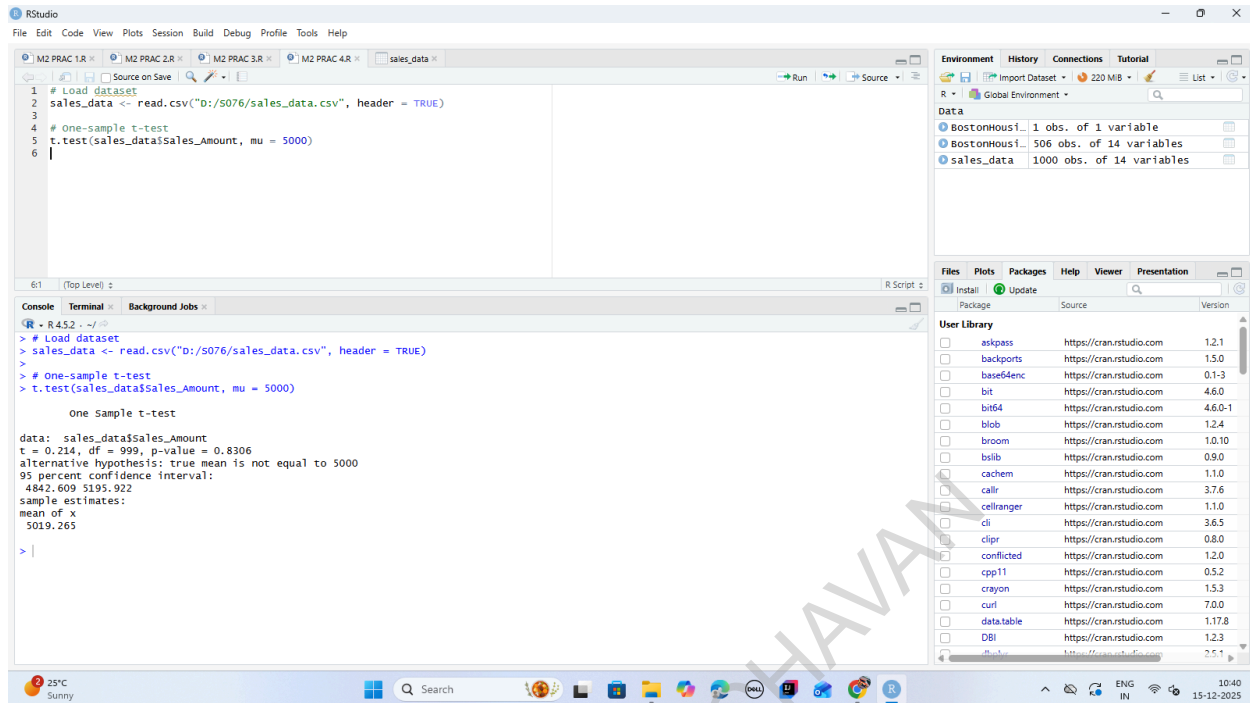
The Environment pane on the right shows the loaded data objects:

- BostonHousi...: 1 obs. of 1 variable
- BostonHousi...: 506 obs. of 14 variables
- sales_data: 1000 obs. of 14 variables

NAME: DEESHA CHAVAN
ROLL NO.: S076

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

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

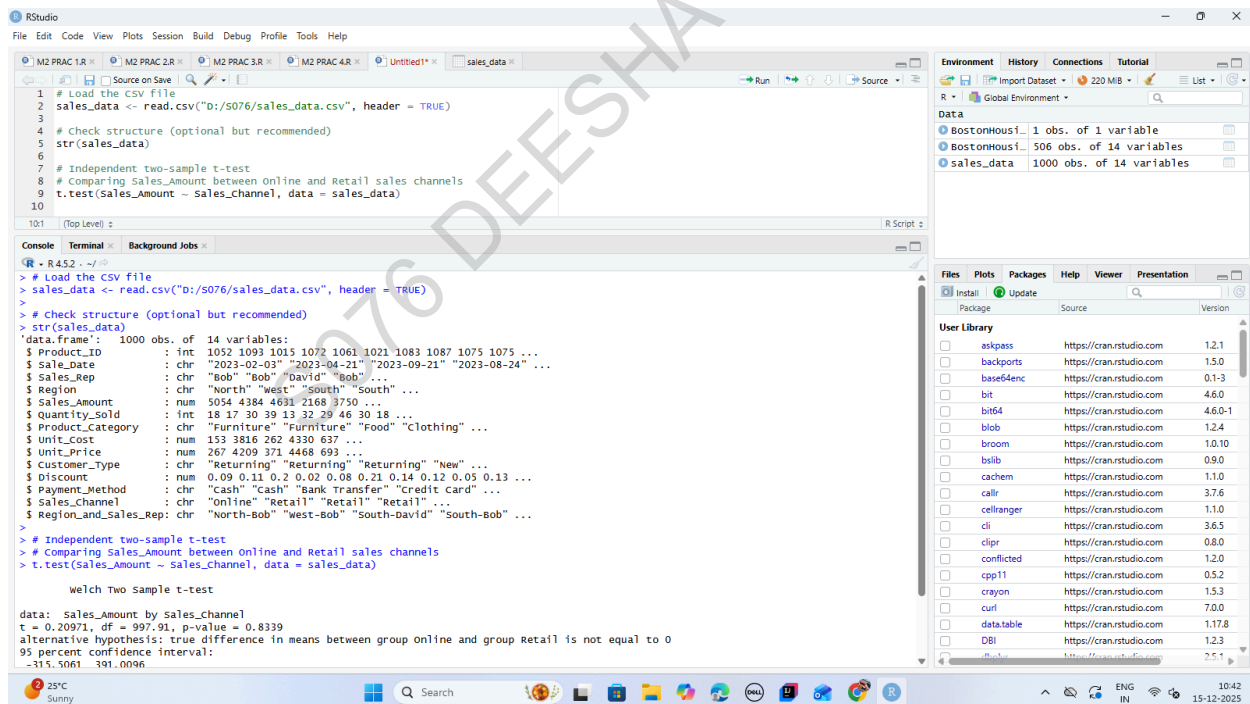


```
1 # Load dataset
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # One-sample t-test
5 t.test(sales_data$Sales_Amount, mu = 5000)
6
```

One Sample t-test

data: sales_data\$Sales_Amount
t = 0.214, df = 999, p-value = 0.8306
alternative hypothesis: true mean is not equal to 5000
95 percent confidence interval:
4842.609 5195.922
sample estimates:
mean of x
5019.265

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



```
1 # Load the CSV file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Check structure (optional but recommended)
5 str(sales_data)
6
7 # Independent two-sample t-test
8 # Comparing Sales_Amount between Online and Retail sales channels
9 t.test(Sales_Amount ~ Sales_Channel, data = sales_data)
10
```

welch two Sample t-test

data: Sales_Amount by Sales_Channel
t = 0.20971, df = 997.91, p-value = 0.8339
alternative hypothesis: true difference in means between group online and group Retail is not equal to 0
95 percent confidence interval:
-415.5061 491.0096

SHETH L.U.J. AND SIR M.V. COLLEGE
SUBJECT: DATA ANALYSIS WITH R

```
1 # Load the csv file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Check structure (optional but recommended)
5 str(sales_data)
6
7 # Independent two-sample t-test
8 # Comparing Sales_Amount between Online and Retail sales channels
9 t.test(Sales_Amount ~ Sales_Channel, data = sales_data)
10
```

Console Output:

```
R - R 4.5.2 - ~/R
> # Load the csv file
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
>
> # Check structure (optional but recommended)
> str(sales_data)
data: Sales_Amount by Sales_Channel
t = 0.20971, df = 997.91, p-value = 0.8339
alternative hypothesis: true difference in means between group Online and group Retail is not equal to 0
95 percent confidence interval:
 -315.5061 391.0096
sample estimates:
mean in group Online mean in group Retail
      5038.594      5000.842
```

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

```
1 # Load the csv file
2 sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
3
4 # Select first 100 observations to form pairs
5 unit_cost <- sales_data$unit_cost[1:100]
6 unit_price <- sales_data$unit_price[1:100]
7
8 # Perform paired t-test
9 t.test(unit_cost, unit_price, paired = TRUE)
10
```

Console Output:

```
R - R 4.5.2 - ~/R
> # Load the csv file
> sales_data <- read.csv("D:/S076/sales_data.csv", header = TRUE)
>
> # Select first 100 observations to form pairs
> unit_cost <- sales_data$unit_cost[1:100]
> unit_price <- sales_data$unit_price[1:100]
>
> # Perform paired t-test
> t.test(unit_cost, unit_price, paired = TRUE)
Paired t-test
data: unit_cost and unit_price
t = -17.179, df = 99, p-value = 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -272.3805 -215.9735
sample estimates:
mean difference
      -244.177
```

NAME: DEESHA CHAVAN
ROLL NO.: S076