

SHETH L.U.J AND SIR M.V COLLEGE  
PRACTICAL- M2 1,2,3,4,5,6  
SUBJECT - DATA ANALYSIS

PRAC 1

AIM- Generating descriptive statistics using summary() or describe()  
OUTPUT -

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R 4.5.2
The downloaded binary packages are in
C:\Users\itlab\AppData\Local\Temp\Rtmp4wvPhX\downloaded_packages
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> library(psych)
> print("---- 1. Descriptive Statistics ----")
[1] "---- 1. Descriptive Statistics ----"
> df <- read.csv("walmart_sales.csv")
>
> # PRE-PROCESSING: Create a grouping variable
> # Classify Holiday vs Non-Holiday weeks
> df$holiday_group <- ifelse(df$holiday_flag == 1, "Holiday", "Non-Holiday")
>
> # 1. PRACTICAL: Generating descriptive statistics using summary() or describe()
>
> print("---- 1. Descriptive Statistics ----")
[1] "---- 1. Descriptive Statistics ----"
>
> # A. using base R summary()
> print("Summary of weekly sales:")
[1] "Summary of Weekly Sales:"
> summary(df$weekly_sales)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
209986  553350  960746 1046965 1420159 3818687
>
> # B. using psych::describe()
> print("Detailed Description of Temperature:")
[1] "Detailed Description of Temperature:"
> describe(df$temperature)
  vars  n mean  sd median trimmed mad  min  max range skew kurtosis  se
X1    1 6435 60.66 18.44  62.67  61.45 20.3 -2.06 100.14 102.2 -0.34  -0.61 0.23
> |

Environment History Connections Tutorial
R - Global Environment
Data
category_pivot 6435 obs. of 46 variables
clean_exact 7 obs. of 3 variables
combined_data 10150 obs. of 2 variables
data 6435 obs. of 8 variables
data_processed 6435 obs. of 17 variables
data_science_jobs 42 obs. of 7 variables
dates_df 4 obs. of 2 variables
df 6435 obs. of 9 variables
df_clean 6435 obs. of 4 variables
duplicates_report 0 obs. of 4 variables
employee_salary_data 50 obs. of 9 variables
flower_clean 10000 obs. of 2 variables
flower_df 10000 obs. of 4 variables
iris 150 obs. of 5 variables
iris_clean 150 obs. of 2 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
orders_df 7 obs. of 3 variables
penguins 344 obs. of 9 variables

Files Plots Packages Help Viewer Presentation
Home - Find in Topic
R Resources
Learning R Online
CRAN Task Views
R on StackOverflow
Getting Help with R
RStudio
Posit Support
Posit Community Forum for the RStudio IDE
Posit Cheat Sheets
RStudio Packages
Posit Products
Manuals
```

PRAC 2

AIM -Generating frequency tables using table() or count()  
OUTPUT

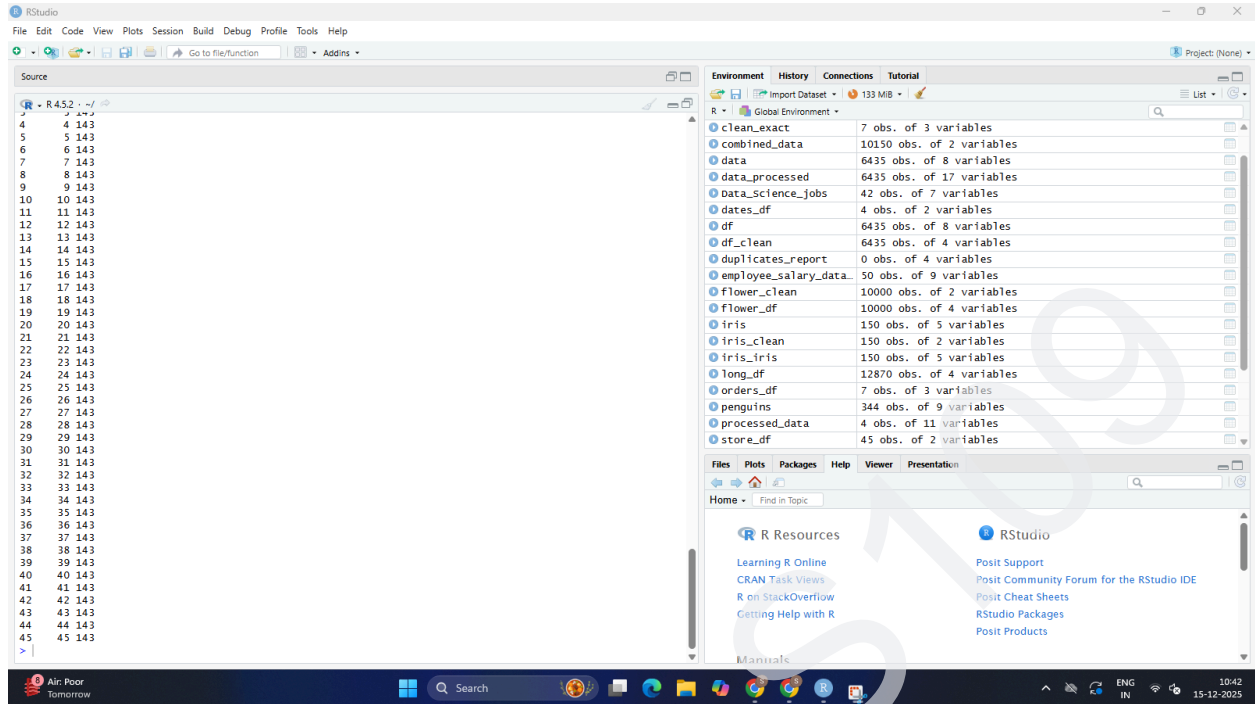
```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R 4.5.2
44 44 143
45 45 143
> df <- read.csv("walmart_sales.csv")
> # 2. PRACTICAL: Generating frequency tables using table() or count()
>
> print("---- 2. Frequency Tables ----")
[1] "---- 2. Frequency Tables ----"
>
> # A. using table()
> store_counts <- table(df$store)
> print(store_counts)
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143 143
>
> # B. using dplyr::count()
> store_df <- df %>% count(store)
> print(store_df)
  store  n
1     1 143
2     2 143
3     3 143
4     4 143
5     5 143
6     6 143
7     7 143
8     8 143
9     9 143
10    10 143
11    11 143
12    12 143
13    13 143
14    14 143
15    15 143
16    16 143
17    17 143
18    18 143
19    19 143
20    20 143
21    21 143
22    22 143
23    23 143
24    24 143
25    25 143

Environment History Connections Tutorial
R - Global Environment
clean_exact 7 obs. of 3 variables
combined_data 10150 obs. of 2 variables
data 6435 obs. of 8 variables
data_processed 6435 obs. of 17 variables
data_science_jobs 42 obs. of 7 variables
dates_df 4 obs. of 2 variables
df 6435 obs. of 9 variables
df_clean 6435 obs. of 4 variables
duplicates_report 0 obs. of 4 variables
employee_salary_data 50 obs. of 9 variables
flower_clean 10000 obs. of 2 variables
flower_df 10000 obs. of 4 variables
iris 150 obs. of 5 variables
iris_clean 150 obs. of 2 variables
iris_iris 150 obs. of 5 variables
long_df 12870 obs. of 4 variables
orders_df 7 obs. of 3 variables
penguins 344 obs. of 9 variables
processed_data 4 obs. of 11 variables
store_df 45 obs. of 2 variables

Files Plots Packages Help Viewer Presentation
Home - Find in Topic
R Resources
Learning R Online
CRAN Task Views
R on StackOverflow
Getting Help with R
RStudio
Posit Support
Posit Community Forum for the RStudio IDE
Posit Cheat Sheets
RStudio Packages
Posit Products
Manuals
```

NAME - UNNATI RATHOD  
ROLL NO - S109

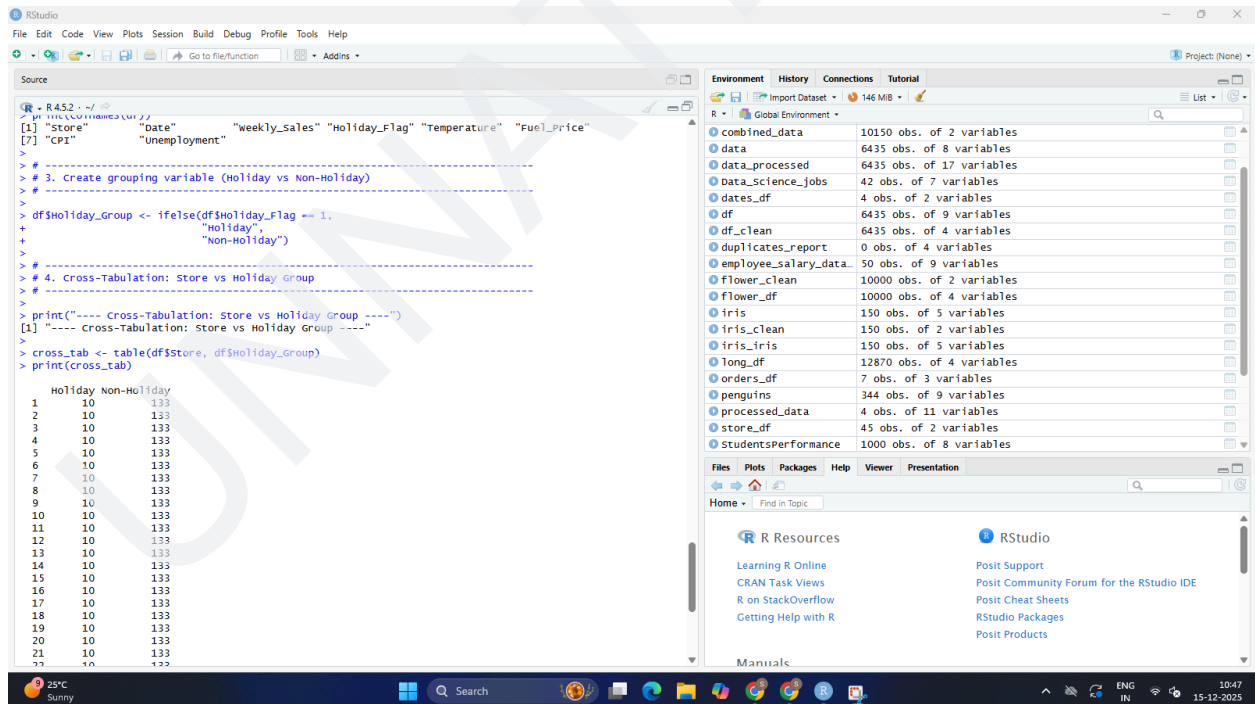
SHETH L.U.J AND SIR M.V COLLEGE  
PRACTICAL- M2 1,2,3,4,5,6  
SUBJECT - DATA ANALYSIS



PRAC3

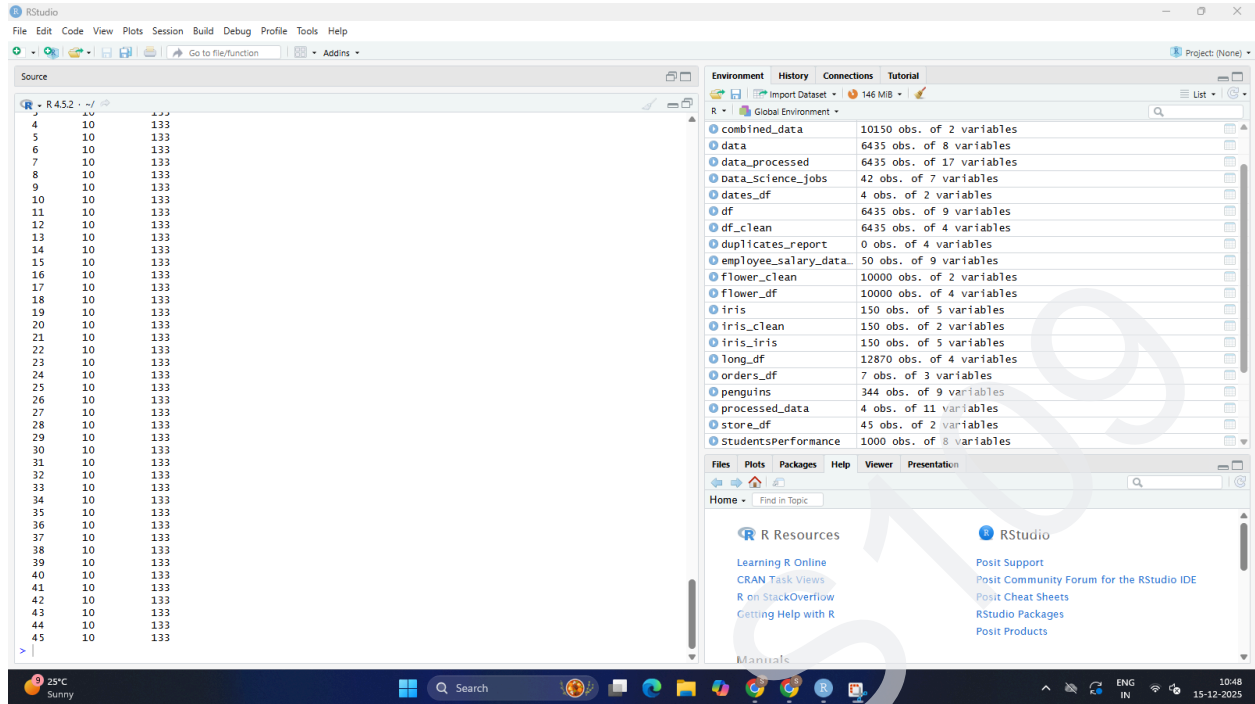
AIM - Creating cross-tabulations and two-way tables

OUTPUT



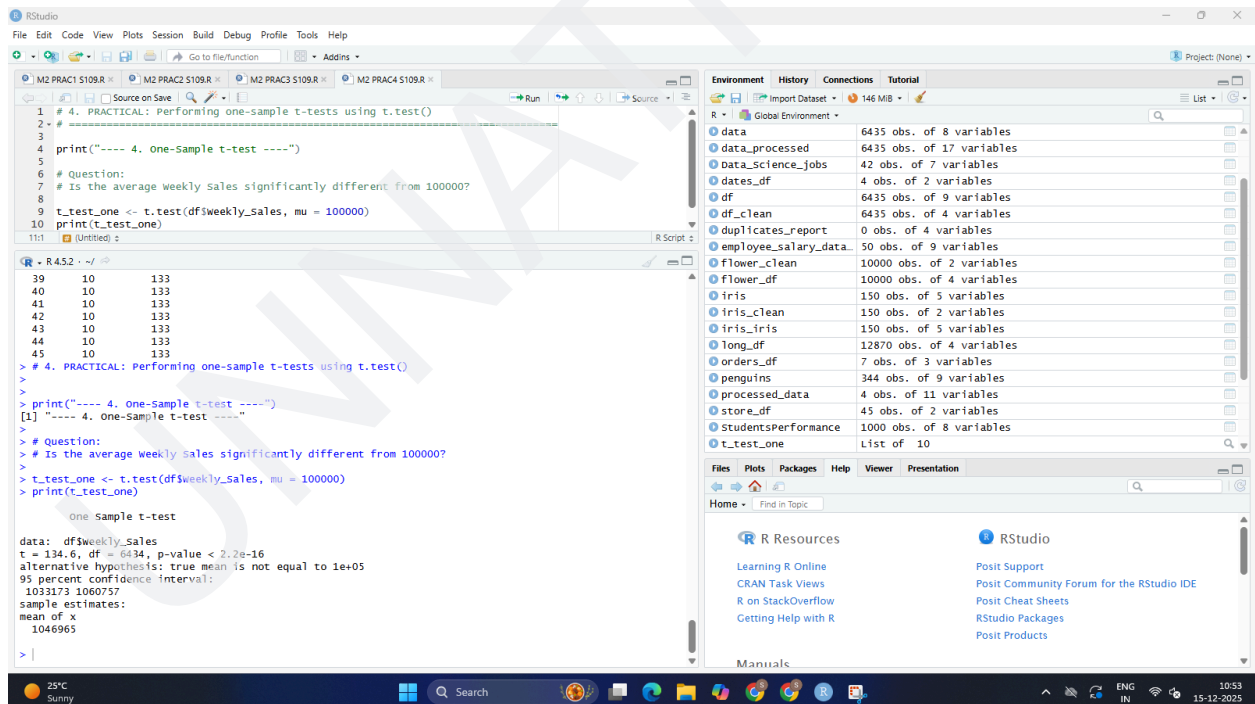
NAME - UNNATI RATHOD  
ROLL NO - S109

SHETH L.U.J AND SIR M.V COLLEGE  
PRACTICAL- M2 1,2,3,4,5,6  
SUBJECT - DATA ANALYSIS



## PRAC4

AIM - Performing one-sample t-tests using `t.test()`



## PRAC5

AIM - Independent two-sample t-test using grouping

OUTPUT -

NAME - UNNATI RATHOD

ROLL NO - S109

SHETH L.U.J AND SIR M.V COLLEGE  
PRACTICAL- M2 1,2,3,4,5,6  
SUBJECT - DATA ANALYSIS

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
M2 PRAC1 S109.R M2 PRAC2 S109.R M2 PRAC3 S109.R M2 PRAC4 S109.R M2 PRAC6 S109.R
1 # 6. PRACTICAL: Performing paired t-tests using t.test(paired = TRUE)
2 df <- read.csv("walmart_sales.csv")
3
4 print("---- 6. Paired t-test ----")
5
6 # NOTE:
7 # Dataset does not naturally contain before/after values
8 # So we simulate Fuel Price before & after for demonstration
9
10 t.test(x = 100, y = 100, paired = TRUE)
11
12 95 percent confidence interval:
13 103.173 106.0757
14 sample estimates:
15 mean of x
16 104.6965
17
18 > # 5. PRACTICAL: Independent two-sample t-test using grouping
19 >
20 > print("---- 5. Independent Two-Sample t-test ----")
21 [1] "---- 5. Independent Two-Sample t-test ----"
22
23 > # Questions:
24 > # Is there a significant difference in weekly sales
25 > # between holiday and non-holiday weeks?
26
27 > t.test_two <- t.test(weekly_sales ~ holiday_group, data = df)
28 > print(t.test_two)
29
30 Welch Two Sample t-test
31
32 data: weekly_sales by holiday_group
33 t = 2.6801, df = 504, p-value = 0.007602
34 alternative hypothesis: true difference in means between group holiday and group non-holiday is not equal to 0
35 95 percent confidence interval:
36 21789.85 141473.17
37 sample estimates:
38 mean in group holiday mean in group non-holiday
39 1122888 1041256
40
41 > # 6. PRACTICAL: Performing paired t-tests using t.test(paired = TRUE)
42 >
43 > # NOTE:
44 > # Dataset does not naturally contain before/after values
45 > # So we simulate Fuel Price before & after for demonstration
46
47 > set.seed(123)
48
49 > df$fuel_price_before <- df$fuel_price - runif(nrow(df), min = 0.1, max = 0.5)
50 > df$fuel_price_after <- df$fuel_price
51
52 > # Perform paired t-test
53 > t.test_paired <- t.test(
54 + df$fuel_price_before,
55 + df$fuel_price_after,
56 + paired = TRUE
57 + )
58 > print(t.test_paired)
59
60 Paired t-test
61
62 data: df$fuel_price_before and df$fuel_price_after
63 t = -208.86, df = 6434, p-value < 2.2e-16
64 alternative hypothesis: true mean difference is not equal to 0
65 95 percent confidence interval:
66 -0.3020415 -0.2964245
67 sample estimates:
68 mean difference
69 -0.299233
70
71 > |
```

PRAC6

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

OUTPUT-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
1 # 6. PRACTICAL: Performing paired t-tests using t.test(paired = TRUE)
2 df <- read.csv("walmart_sales.csv")
3
4 print("---- 6. Paired t-test ----")
5
6 # NOTE:
7 # Dataset does not naturally contain before/after values
8 # So we simulate Fuel Price before & after for demonstration
9
10 > set.seed(123)
11
12 > df$fuel_price_before <- df$fuel_price - runif(nrow(df), min = 0.1, max = 0.5)
13 > df$fuel_price_after <- df$fuel_price
14
15 > # Perform paired t-test
16 > t.test_paired <- t.test(
17 + df$fuel_price_before,
18 + df$fuel_price_after,
19 + paired = TRUE
20 + )
21 > print(t.test_paired)
22
23 Paired t-test
24
25 data: df$fuel_price_before and df$fuel_price_after
26 t = -208.86, df = 6434, p-value < 2.2e-16
27 alternative hypothesis: true mean difference is not equal to 0
28 95 percent confidence interval:
29 -0.3020415 -0.2964245
30 sample estimates:
31 mean difference
32 -0.299233
33
34 > |
```

NAME - UNNATI RATHOD

ROLL NO - S109