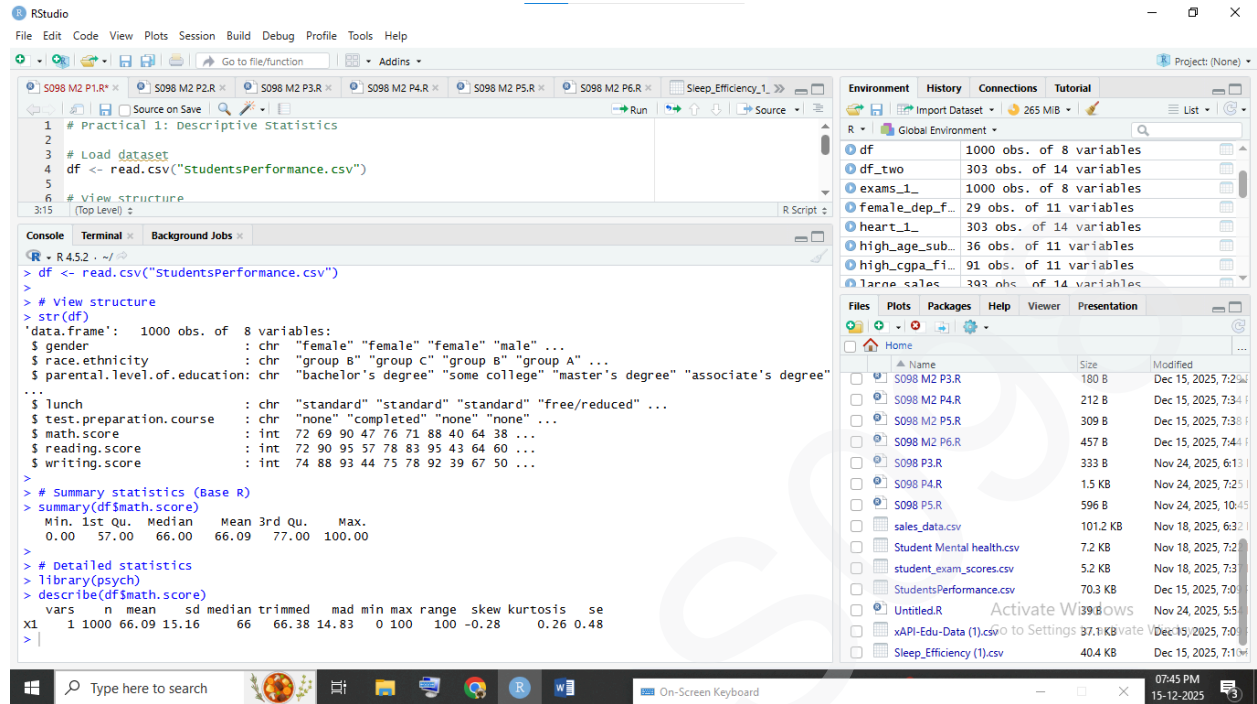


MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

Practical no.-1 to 6 (Module2)

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



The screenshot displays the RStudio interface. The script editor on the left contains the following R code:

```
# Practical 1: Descriptive Statistics
# Load dataset
df <- read.csv("StudentsPerformance.csv")
# View structure
str(df)
```

The console on the left shows the output of the executed code:

```
R > R 4.5.2 . ~/
> df <- read.csv("StudentsPerformance.csv")
> # View structure
> str(df)
'data.frame': 1000 obs. of  8 variables:
 $ gender       : chr  "female" "female" "female" "male" ...
 $ race.ethnicity : chr  "group B" "group C" "group B" "group A" ...
 $ parental.level.of.education: chr  "bachelor's degree" "some college" "master's degree" "associate's degree" ...
 $ lunch        : chr  "standard" "standard" "standard" "free/reduced" ...
 $ test.preparation.course : chr  "none" "completed" "none" "none" ...
 $ math.score    : int   72 69 90 47 76 71 88 40 64 38 ...
 $ reading.score : int   72 90 95 57 78 83 95 43 64 60 ...
 $ writing.score  : int   74 88 93 44 75 78 92 39 67 50 ...
> # Summary statistics (Base R)
> summary(df$math.score)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   57.00   66.00   66.09   77.00  100.00
> # Detailed statistics
> library(psych)
> describe(df$math.score)
  vars      n mean  sd median trimmed  mad min max range skew kurtosis  se
X1      1 1000 66.09 15.16    66   66.38 14.83   0 100   100 -0.28   0.26 0.48
>
```

The Environment pane on the right shows the loaded objects:

Object	Class	Attributes
df	data.frame	1000 obs. of 8 variables
df_two	data.frame	303 obs. of 14 variables
exams_1	data.frame	1000 obs. of 8 variables
female_dep_f	data.frame	29 obs. of 11 variables
heart_1	data.frame	303 obs. of 14 variables
high_age_sub	data.frame	36 obs. of 11 variables
high_cgpa_fi	data.frame	91 obs. of 11 variables
lame_sales	data.frame	393 obs. of 14 variables

The Files pane on the right shows the project files:

Name	Size	Modified
S098 M2 P3.R	180 B	Dec 15, 2025, 7:25 AM
S098 M2 P4.R	212 B	Dec 15, 2025, 7:34 AM
S098 M2 P5.R	309 B	Dec 15, 2025, 7:38 AM
S098 M2 P6.R	457 B	Dec 15, 2025, 7:44 AM
S098 P3.R	333 B	Nov 24, 2025, 6:13 AM
S098 P4.R	1.5 KB	Nov 24, 2025, 7:25 AM
S098 P5.R	596 B	Nov 24, 2025, 10:45 AM
sales_data.csv	101.2 KB	Nov 18, 2025, 6:32 AM
Student Mental health.csv	7.2 KB	Nov 18, 2025, 7:25 AM
student_exam_scores.csv	5.2 KB	Nov 18, 2025, 7:38 AM
StudentsPerformance.csv	70.3 KB	Dec 15, 2025, 7:00 AM
Untitled.R	39 B	Nov 24, 2025, 5:55 AM
xAPI-Edu-Data (1).csv	37.1 KB	Dec 15, 2025, 7:00 AM
Sleep_Efficiency (1).csv	40.4 KB	Dec 15, 2025, 7:10 AM

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

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

The screenshot shows the RStudio interface with the following content:

```
# Practical 2: Generating Frequency Tables
>
> # Load dataset
> df <- read.csv("XAPI-Edu-Data (1).csv")
> str(df)
'data.frame': 480 obs. of 17 variables:
 $ gender      : chr "M" "M" "M" "M" ...
 $ Nationality : chr "KW" "KW" "KW" "KW" ...
 $ PlaceOfBirth: chr "Kuwait" "Kuwait" "Kuwait" "Kuwait" ...
 $ StageID     : chr "lowerlevel" "lowerlevel" "lowerlevel" "lowerlevel" ...
 $ GradeID     : chr "G-04" "G-04" "G-04" "G-04" ...
 $ SectionID   : chr "A" "A" "A" "A" ...
 $ Topic       : chr "IT" "IT" "IT" "IT" ...
 $ Semester    : chr "F" "F" "F" "F" ...
 $ Relation    : chr "Father" "Father" "Father" "Father" ...
 $ raisedhands : int 15 20 10 30 40 42 35 50 12 70 ...
 $ VisitedResources : int 16 20 7 25 50 30 12 10 21 80 ...
 $ Announcementsview : int 2 3 0 5 12 13 0 15 16 25 ...
 $ Discussion   : int 20 25 30 35 50 70 17 22 50 70 ...
 $ ParentAnsweringSurvey : chr "Yes" "Yes" "No" "No" ...
 $ ParentschoolSatisfaction: chr "Good" "Good" "Bad" "Bad" ...
 $ StudentAbsenceDays : chr "Under-7" "Under-7" "Above-7" "Above-7" ...
 $ Class       : chr "M" "M" "L" "L" ...

> colnames(df)
[1] "gender"      "Nationality" "PlaceOfBirth"
[4] "StageID"     "GradeID"     "SectionID"
[7] "Topic"       "Semester"    "Relation"
[10] "raisedhands" "VisitedResources" "Announcementsview"
[13] "Discussion"  "ParentAnsweringSurvey" "ParentschoolSatisfaction"
[16] "StudentAbsenceDays" "Class"
```

The Environment pane on the right shows the loaded objects: `t_test_one` (List of 10), `t_test_pair` (List of 10), `t_test_two` (List of 10), `xAPI_Edu_Dat...` (480 obs. of 17 variables), and `young_studen...` (53 obs. of 11 variables).

The screenshot shows the RStudio interface with the following content:

```
# Practical 2: Generating Frequency Tables
>
> # Frequency of Gender categories
> sku_freq <- table(df$gender)
> print("Frequency Table using table():")
[1] "Frequency Table using table():"
> print(sku_freq)
 F  M
175 305

> # Load dplyr library
> library(dplyr)

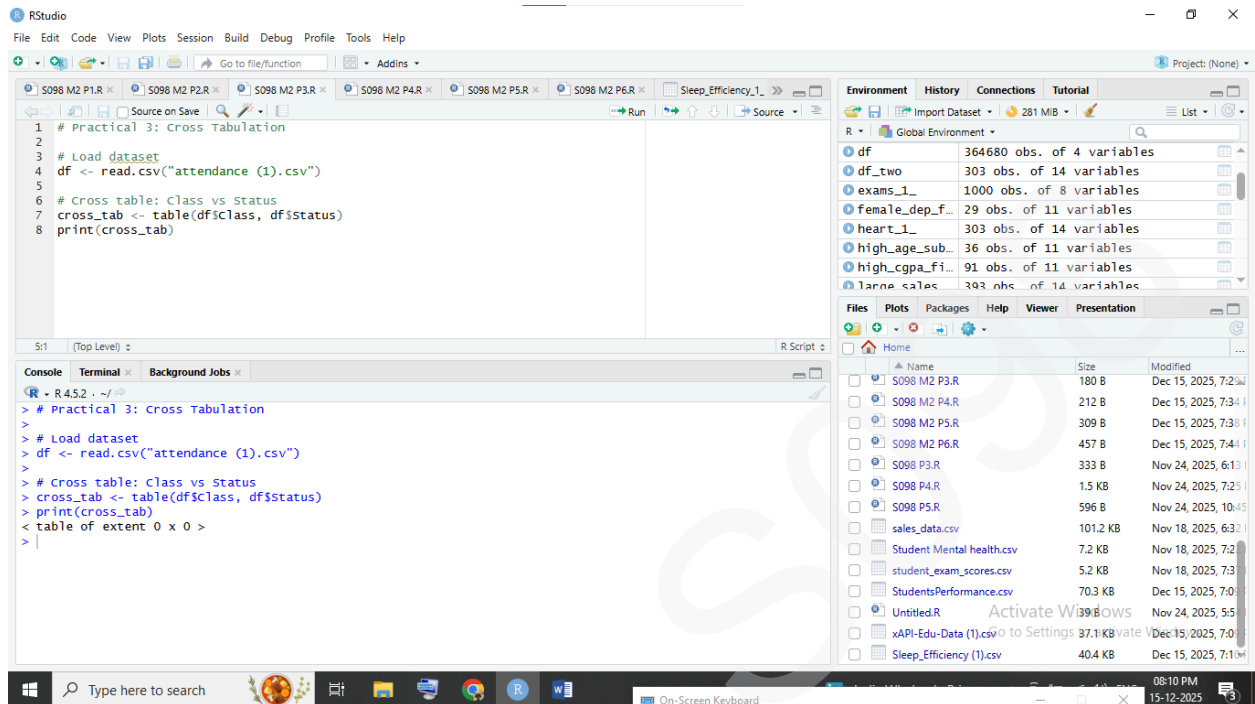
> # Frequency using count()
> sku_count <- df %>% count(gender)
> print("Frequency Table using dplyr count():")
[1] "Frequency Table using dplyr count():"
> print(sku_count)
  gender     n
1     F    175
2     M    305
```

The Environment pane on the right shows the loaded objects: `t_test_one` (List of 10), `t_test_pair` (List of 10), `t_test_two` (List of 10), `xAPI_Edu_Dat...` (480 obs. of 17 variables), and `young_studen...` (53 obs. of 11 variables).

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

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



The screenshot displays the RStudio environment. The script editor on the left contains the following R code:

```
1 # Practical 3: Cross Tabulation
2
3 # Load dataset
4 df <- read.csv("attendance (1).csv")
5
6 # Cross table: Class vs Status
7 cross_tab <- table(df$class, df$status)
8 print(cross_tab)
```

The console on the bottom left shows the execution output:

```
> # Practical 3: Cross Tabulation
>
> # Load dataset
> df <- read.csv("attendance (1).csv")
>
> # Cross table: Class vs Status
> cross_tab <- table(df$class, df$status)
> print(cross_tab)
<table of extent 0 x 0 >
> |
```

The Environment pane on the right lists the following objects in the Global Environment:

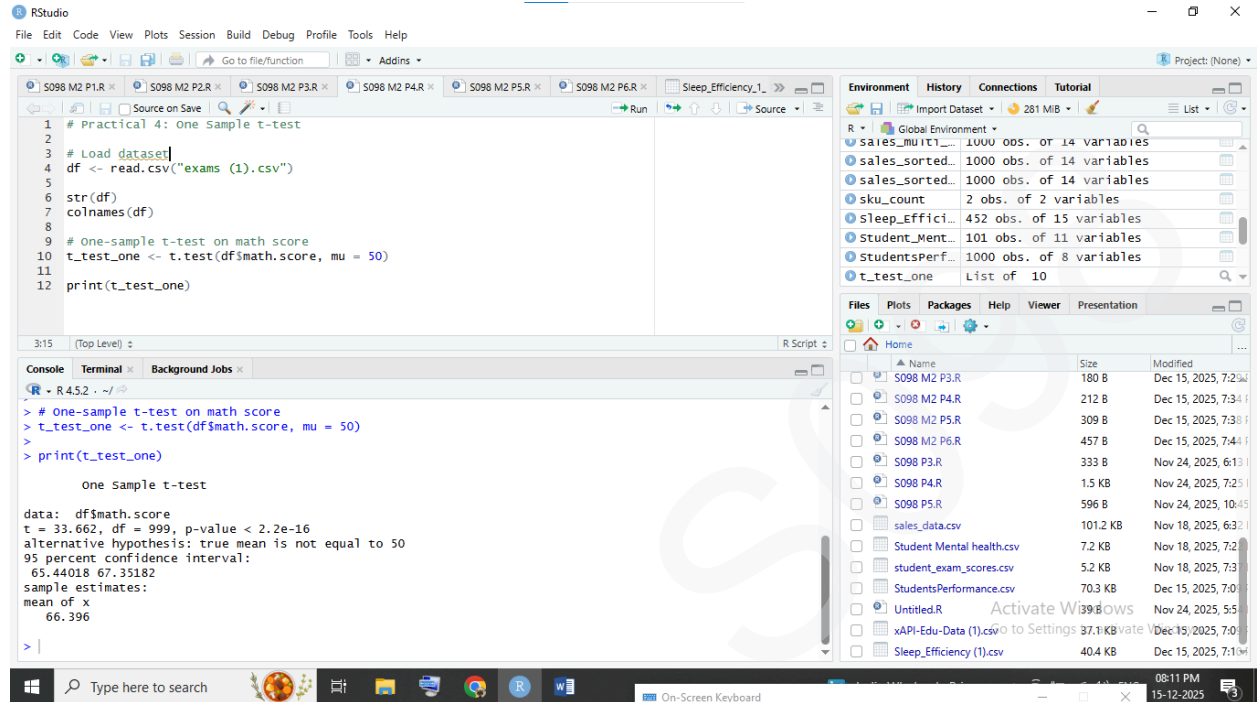
Object	Size	Modified
df	364680 obs. of 4 variables	
df_two	303 obs. of 14 variables	
exams_1	1000 obs. of 8 variables	
female_dep_f	29 obs. of 11 variables	
heart_1	303 obs. of 14 variables	
high_age_sub	36 obs. of 11 variables	
high_cgpa_fi	91 obs. of 11 variables	
lame_sales	393 obs. of 14 variables	

The Files pane on the bottom right shows a list of files in the current directory, including R scripts (e.g., S098 M2 P3.R, S098 M2 P4.R) and CSV files (e.g., sales_data.csv, Student_Mental_health.csv).

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

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



The screenshot displays the RStudio interface with a script editor, console, and environment pane. The script performs a one-sample t-test on a dataset named 'df\$math.score'.

```
1 # Practical 4: One Sample t-test
2
3 # Load dataset
4 df <- read.csv("exams (1).csv")
5
6 str(df)
7 colnames(df)
8
9 # one-sample t-test on math score
10 t_test_one <- t.test(df$math.score, mu = 50)
11
12 print(t_test_one)
```

The console output shows the results of the t-test:

```
> # one-sample t-test on math score
> t_test_one <- t.test(df$math.score, mu = 50)
> print(t_test_one)

One Sample t-test

data: df$math.score
t = 33.662, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 65.44018 67.35182
sample estimates:
mean of x
 66.396
```

The environment pane on the right lists the objects in the global environment, including 'Sales_Multitask', 'sales_sorted', 'sku_count', 'Sleep_Efficiency', 'Student_Ment', 'StudentsPerf', and 't_test_one'.

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

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

```
# Practical 5: Independent Two Sample t-test
# Load dataset
df <- read.csv("heart (1).csv")
str(df)
```

```
R - R4.5.2 ~ /
> # Practical 5: Independent Two Sample t-test
> # Load dataset
> df <- read.csv("heart (1).csv")
> str(df)
'data.frame': 303 obs. of 14 variables:
 $ age : int  63 37 41 56 57 57 56 44 52 57 ...
 $ sex : int  1 1 0 1 0 1 0 1 1 1 ...
 $ cp : int  3 2 1 1 0 0 1 1 2 2 ...
 $ trestbps: int 145 130 130 120 120 140 140 120 172 150 ...
 $ chol : int 233 250 204 236 354 192 294 263 199 168 ...
 $ fbs : int 1 0 0 0 0 0 0 0 1 0 ...
 $ restecg: int 0 1 0 1 1 1 0 0 1 1 ...
 $ thalach: int 150 187 172 178 163 148 153 173 162 174 ...
 $ exang : int 0 0 0 0 1 0 0 0 0 0 ...
 $ oldpeak: num 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
 $ slope : int 0 0 2 2 2 1 1 2 2 2 ...
 $ ca : int 0 0 0 0 0 0 0 0 0 0 ...
 $ thal : int 1 2 2 2 2 1 2 3 3 2 ...
 $ target: int 1 1 1 1 1 1 1 1 1 1 ...
> unique(df$target)
[1] 1 0
> # Keep only two groups (0 and 1)
> df_two <- subset(df, target %in% c(0, 1))
```

```
# Practical 5: Independent Two Sample t-test
# Load dataset
df <- read.csv("heart (1).csv")
str(df)

# Independent two-sample t-test
t_test_two <- t.test(age ~ target, data = df_two)
print(t_test_two)
```

```
R - R4.5.2 ~ /
> # Independent two-sample t-test
> t_test_two <- t.test(age ~ target, data = df_two)
> print(t_test_two)

Welch Two Sample t-test

data: age by target
t = 4.0797, df = 301, p-value = 5.781e-05
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 2.124635 6.084324
sample estimates:
mean in group 0 mean in group 1
 56.60145      52.49697
```

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

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

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

S098 M2 P1.R S098 M2 P2.R S098 M2 P3.R S098 M2 P4.R S098 M2 P5.R S098 M2 P6.R Sleep_Efficiency_1_
1 # Practical 6: Paired t-test
2
3 # Load dataset
10:75 (Top Level) R Script:

R - R4.5.2 ~\
> # Practical 6: Paired t-test
>
> # Load dataset
> df <- read.csv("Sleep_Efficiency (1).csv")
> str(df)
'data.frame': 452 obs. of 15 variables:
 $ ID: int 1 2 3 4 5 6 7 8 9 10 ...
 $ Age: int 65 69 40 40 57 36 27 53 41 11 ...
 $ Gender: chr "Female" "Male" "Female" "Female" ...
 $ Bedtime: chr "2021-03-06 01:00:00" "2021-12-05 02:00:00" "2021-05-25 21:30:00" "2021-11-0
3 02:30:00" ...
 $ wakeup.time: chr "2021-03-06 07:00:00" "2021-12-05 09:00:00" "2021-05-25 05:30:00" "2021-11-0
3 08:30:00" ...
 $ sleep.duration: num 6 7 8 6 8 7 5 6 10 6 9 ...
 $ sleep_efficiency: num 0.88 0.66 0.89 0.51 0.76 0.9 0.54 0.9 0.79 0.55 ...
 $ REM.sleep.percentage: int 18 19 20 23 27 23 28 28 28 18 ...
 $ Deep.sleep.percentage: int 70 28 70 25 55 60 25 52 55 37 ...
 $ Light.sleep.percentage: int 12 53 10 52 18 17 47 20 17 45 ...
 $ Awakenings: num 0 3 1 3 3 0 2 0 3 4 ...
 $ Caffeine.consumption: num 0 0 0 50 0 NA 50 50 50 0 ...
 $ Alcohol.consumption: num 0 3 0 5 3 0 0 0 0 0 ...
 $ Smoking.status: chr "Yes" "Yes" "No" "No" "Yes" ...
 $ Exercise.frequency: num 3 3 3 1 3 1 1 3 1 0 ...
> colnames(df)
[1] "ID" "Age" "Gender" "Bedtime"
[5] "wakeup.time" "sleep.duration" "sleep_efficiency" "REM.sleep.percentage"
[9] "Deep.sleep.percentage" "Light.sleep.percentage" "Awakenings" "Caffeine.consumption"
[13] "Alcohol.consumption" "Smoking.status" "Exercise.frequency"
>
> # create before and after values from existing numeric column
> df$sarea_before <- df$sleep_efficiency
> df$sarea_after <- df$sleep_efficiency + rnorm(nrow(df), mean = -.5, sd = 2)
>
> # Paired t-test
> t_test_paired <- t.test(df$sarea_before,
+ df$sarea_after,
+ paired = TRUE)
> print(t_test_paired)

Paired t-test

data: df$sarea_before and df$sarea_after
t = 50.9, df = 451, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 4.804780 5.190707
sample estimates:
mean difference
 4.997744
>

Environment History Connections Tutorial
R Global Environment
Sales_sorted_ 1000 obs. of 14 variables
sku_count 2 obs. of 2 variables
Sleep_effici... 452 obs. of 15 variables
Student_Ment... 101 obs. of 11 variables
StudentsPerf... 1000 obs. of 8 variables
t_test_one List of 10
t_test_paired List of 10
t_test_two List of 10
Files Plots Packages Help Viewer Presentation
Home
S098 M2 P3.R 180 B Dec 15, 2025, 7:25
S098 M2 P4.R 212 B Dec 15, 2025, 7:34
S098 M2 P5.R 309 B Dec 15, 2025, 7:38
S098 M2 P6.R 457 B Dec 15, 2025, 7:44
S098 P3.R 333 B Nov 24, 2025, 6:13
S098 P4.R 1.5 KB Nov 24, 2025, 7:25
S098 P5.R 596 B Nov 24, 2025, 10:45
sales_data.csv 101.2 KB Nov 18, 2025, 6:32
Student Mental health.csv 7.2 KB Nov 18, 2025, 7:2
student_exam_scores.csv 5.2 KB Nov 18, 2025, 7:3
StudentsPerformance.csv 70.3 KB Dec 15, 2025, 7:0
Untitled.R 398 B Nov 24, 2025, 5:5
xAPI-Edu-Data (1).csv 37.1 KB Dec 15, 2025, 7:0
Sleep_Efficiency (1).csv 40.4 KB Dec 15, 2025, 7:1
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

S098 M2 P1.R S098 M2 P2.R S098 M2 P3.R S098 M2 P4.R S098 M2 P5.R S098 M2 P6.R Sleep_Efficiency_1_
1 # Practical 6: Paired t-test
2
3 # Load dataset
10:75 (Top Level) R Script:

R - R4.5.2 ~\
> # Practical 6: Paired t-test
>
> # Load dataset
> df <- read.csv("Sleep_Efficiency (1).csv")
> str(df)
'data.frame': 452 obs. of 15 variables:
 $ ID: int 1 2 3 4 5 6 7 8 9 10 ...
 $ Age: int 65 69 40 40 57 36 27 53 41 11 ...
 $ Gender: chr "Female" "Male" "Female" "Female" ...
 $ Bedtime: chr "2021-03-06 01:00:00" "2021-12-05 02:00:00" "2021-05-25 21:30:00" "2021-11-0
3 02:30:00" ...
 $ wakeup.time: chr "2021-03-06 07:00:00" "2021-12-05 09:00:00" "2021-05-25 05:30:00" "2021-11-0
3 08:30:00" ...
 $ sleep.duration: num 6 7 8 6 8 7 5 6 10 6 9 ...
 $ sleep_efficiency: num 0.88 0.66 0.89 0.51 0.76 0.9 0.54 0.9 0.79 0.55 ...
 $ REM.sleep.percentage: int 18 19 20 23 27 23 28 28 28 18 ...
 $ Deep.sleep.percentage: int 70 28 70 25 55 60 25 52 55 37 ...
 $ Light.sleep.percentage: int 12 53 10 52 18 17 47 20 17 45 ...
 $ Awakenings: num 0 3 1 3 3 0 2 0 3 4 ...
 $ Caffeine.consumption: num 0 0 0 50 0 NA 50 50 50 0 ...
 $ Alcohol.consumption: num 0 3 0 5 3 0 0 0 0 0 ...
 $ Smoking.status: chr "Yes" "Yes" "No" "No" "Yes" ...
 $ Exercise.frequency: num 3 3 3 1 3 1 1 3 1 0 ...
> colnames(df)
[1] "ID" "Age" "Gender" "Bedtime"
[5] "wakeup.time" "sleep.duration" "sleep_efficiency" "REM.sleep.percentage"
[9] "Deep.sleep.percentage" "Light.sleep.percentage" "Awakenings" "Caffeine.consumption"
[13] "Alcohol.consumption" "Smoking.status" "Exercise.frequency"
>
> # create before and after values from existing numeric column
> df$sarea_before <- df$sleep_efficiency
> df$sarea_after <- df$sleep_efficiency + rnorm(nrow(df), mean = -.5, sd = 2)
>
> # Paired t-test
> t_test_paired <- t.test(df$sarea_before,
+ df$sarea_after,
+ paired = TRUE)
> print(t_test_paired)

Paired t-test

data: df$sarea_before and df$sarea_after
t = 50.9, df = 451, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 4.804780 5.190707
sample estimates:
mean difference
 4.997744
>

Environment History Connections Tutorial
R Global Environment
Sales_sorted_ 1000 obs. of 14 variables
sku_count 2 obs. of 2 variables
Sleep_effici... 452 obs. of 15 variables
Student_Ment... 101 obs. of 11 variables
StudentsPerf... 1000 obs. of 8 variables
t_test_one List of 10
t_test_paired List of 10
t_test_two List of 10
Files Plots Packages Help Viewer Presentation
Home
S098 M2 P3.R 180 B Dec 15, 2025, 7:25
S098 M2 P4.R 212 B Dec 15, 2025, 7:34
S098 M2 P5.R 309 B Dec 15, 2025, 7:38
S098 M2 P6.R 457 B Dec 15, 2025, 7:44
S098 P3.R 333 B Nov 24, 2025, 6:13
S098 P4.R 1.5 KB Nov 24, 2025, 7:25
S098 P5.R 596 B Nov 24, 2025, 10:45
sales_data.csv 101.2 KB Nov 18, 2025, 6:32
Student Mental health.csv 7.2 KB Nov 18, 2025, 7:2
student_exam_scores.csv 5.2 KB Nov 18, 2025, 7:3
StudentsPerformance.csv 70.3 KB Dec 15, 2025, 7:0
Untitled.R 398 B Nov 24, 2025, 5:5
xAPI-Edu-Data (1).csv 37.1 KB Dec 15, 2025, 7:0
Sleep_Efficiency (1).csv 40.4 KB Dec 15, 2025, 7:1
```