

# 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 shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The left sidebar has tabs for S098 M2 P1.R\*, S098 M2 P2.R, S098 M2 P3.R, S098 M2 P4.R, S098 M2 P5.R, S098 M2 P6.R, and Sleep\_Efficiency\_1. The main workspace shows R code for generating descriptive statistics:

```

1 # Practical 1: Descriptive Statistics
2
3 # Load dataset
4 df <- read.csv("StudentsPerformance.csv")
5
6 # View structure
7 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 right pane shows the Global Environment and a file browser. The file browser lists various R scripts and CSV files, including S098 M2 P3.R, S098 M2 P4.R, S098 M2 P5.R, S098 M2 P6.R, S098 P3.R, S098 P4.R, S098 P5.R, sales\_data.csv, Student\_Mental\_health.csv, student\_exam\_scores.csv, StudentsPerformance.csv, Untitled.R, xAPI-Edu-Data (1).csv, and Sleep\_Efficiency (1).csv.

# **MVLU COLLEGE**

## **Subject:-Data Analysis with SAS / SPSS /R**

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

The screenshot shows an RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Code Editor:** Displays R code for generating frequency tables from an xAPI-Edu-Data CSV file. The code includes loading the dataset, reading the CSV, and printing the structure of the data frame.
- Environment Pane:** Shows the global environment with objects like t\_test\_one, t\_test\_paired, t\_test\_two, xAPI\_Edu\_Data, young\_studen..., and Sleep\_Efficiency\_1.
- Files Pane:** Lists local files including S098 M2 P3.R, S098 M2 P4.R, S098 M2 P5.R, S098 M2 P6.R, S098 P3.R, S098 P4.R, S098 P5.R, sales\_data.csv, Student Mental health.csv, student\_exam\_scores.csv, StudentsPerformance.csv, Untitled.R, xAPI-Edu-Data (1).csv, and Sleep\_Efficiency (1).csv.
- Bottom Status Bar:** Shows the date and time as 08:05 PM 15-12-2025.

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Run, and Source.
- Code Editor:** Displays R code for generating frequency tables. The code includes:
  - Reading CSV files (S098 M2 P1.R, S098 M2 P2.R, etc.)
  - Creating a new dataset (Sleep\_Efficiency\_1)
  - Defining variables:
    - StudentAbsentDays
    - Class
  - Displaying column names (df)
  - Printing frequency tables for gender categories (cross\_tab, sku\_freq)
  - Using dplyr library and count function to generate frequency tables
  - Printing the resulting data frames (gender, sku\_count)
- Environment Tab:** Shows the global environment with objects like t\_test\_one, t\_test\_paired, t\_test\_two, xAPI\_Edu\_Data, young\_students, cross\_tab, and sku\_freq.
- Files Tab:** Shows a list of files in the current directory, including S098 M2 P3.R, S098 M2 P4.R, S098 M2 P5.R, S098 M2 P6.R, S098 P3.R, S098 P4.R, S098 P5.R, sales\_data.csv, Student Mental health.csv, student\_exam\_scores.csv, StudentsPerformance.csv, Untitled.R, xAPI-Edu-Data (1).csv, and Sleep\_Efficiency (1).csv.
- Help and Support:** Includes links for Import Dataset, 264 MB, and various help sections like Environment, History, Connections, and Tutorial.

# MVLU COLLEGE

## Subject:-Data Analysis with SAS / SPSS /R

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

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The main workspace shows an R script titled "S1 [Top Level]". The code is:

```
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 window below shows the R session 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 >
>
```

To the right of the workspace is a file browser titled "Project: (None)". It lists various files and objects:

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, 7:22 AM
student_exam_scores.csv	5.2 KB	Nov 18, 2025, 7:3 AM
StudentsPerformance.csv	70.3 KB	Dec 15, 2025, 7:01 AM
Untitled.R	39 KB	Nov 24, 2025, 5:55 AM
xAPI-Edu-Data (1).csv	37.1 KB	Nov 24, 2025, 7:01 AM
Sleep_Efficiency (1).csv	40.4 KB	Dec 15, 2025, 7:15 AM

The bottom taskbar shows the Windows Start button, a search bar, and several pinned application icons (File Explorer, Edge, Google Chrome, R, Word). The system tray indicates the date and time as 15-12-2025 and 08:10 PM.

# MVLU COLLEGE

## Subject:-Data Analysis with SAS / SPSS /R

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

The screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The main workspace shows an R script titled "Practical 4: One Sample t-test". The code performs a one-sample t-test on a dataset named "exams (1).csv" to test if the mean math score is 50. The results output shows a t-value of 33.662 and a p-value less than 2.2e-16, indicating a significant difference from the null hypothesis.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ Go to file/function Addins +
S098 M2 P1.R S098 M2 P2.R S098 M2 P3.R S098 M2 P4.R S098 M2 P5.R Sleep_Efficiency_1.R Run Source Project: (None)
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)

3:15 [Top Level] R Script
Console Terminal Background Jobs
> # 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

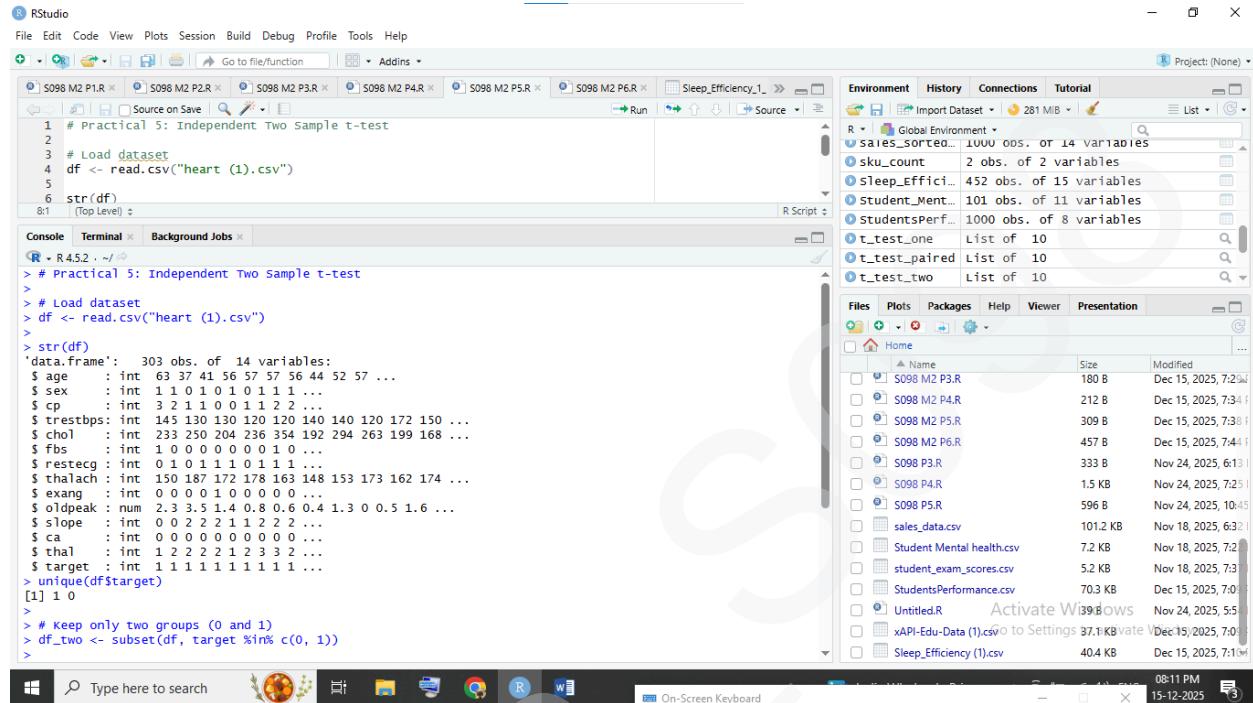
data: df$math.score
t = 33.662, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 65.44018 67.35182
sample estimates:
mean of x
 66.396
> |
```

The right sidebar displays the Global Environment, showing various objects like sales\_MULTI... (1000 obs. or 14 variables), sales\_sorted\_1 (1000 obs. of 14 variables), sales\_sorted\_2 (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), and t\_test\_one (List of 10). The Files tab shows a list of R scripts and CSV files in the current directory, including S098 M2 P3.R, S098 M2 P4.R, S098 M2 P5.R, S098 M2 P6.R, S098 P3.R, S098 P4.R, S098 P5.R, sales\_data.csv, Student Mental health.csv, student\_exam\_scores.csv, StudentsPerformance.csv, Untitled.R, xAPI-Edu-Data (1).csv, and Sleep\_Efficiency (1).csv. The status bar at the bottom shows the date and time as 08:11 PM 15-12-2025.

# MVLU COLLEGE

## Subject:-Data Analysis with SAS / SPSS /R

### 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

Source on Save | Run | Source | Environment History Connections Tutorial

```

1 # Practical 5: Independent Two Sample t-test
2
3 # Load dataset
4 df <- read.csv("heart (1).csv")
5
6 str(df)
8:1 [Top Level] R Script

Console Terminal Background Jobs
R > R 4.5.2 . ~/ ~
> # 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 0 1 0 1 0 1 1 1 1 ...
 $ cp : int 3 2 1 1 0 0 1 1 2 2 ...
 $ trestbps: int 145 130 130 120 140 140 120 172 150 ...
 $ chol : int 233 250 204 238 354 192 294 263 199 168 ...
 $ fbs : int 0 0 0 0 0 0 0 0 1 0 ...
 $ restecg : int 0 1 0 1 1 1 0 1 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] 0
>
> # Keep only two groups (0 and 1)
> df_two <- subset(df, target %in% c(0, 1))
>

```

Files Plots Packages Help Viewer Presentation

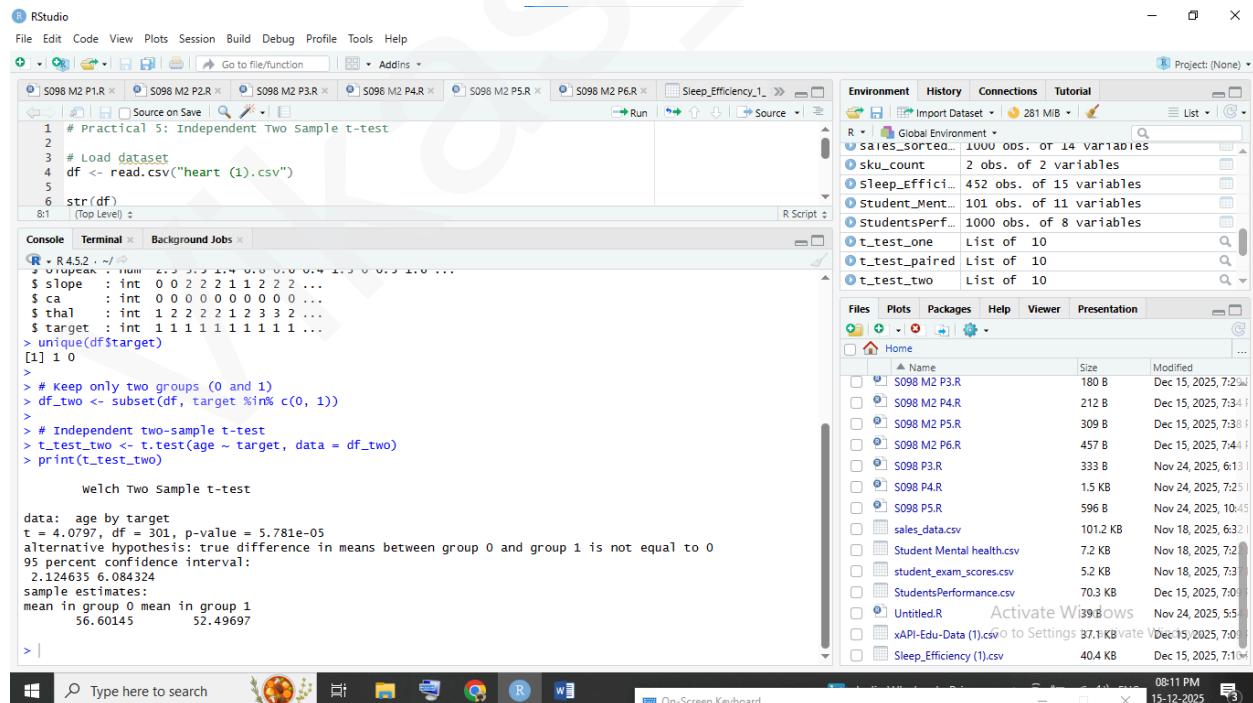
Home

Name	Size	Modified
S098 M2 P3.R	180 B	Dec 15, 2025, 7:29 PM
S098 M2 P4.R	212 B	Dec 15, 2025, 7:34 PM
S098 M2 P5.R	309 B	Dec 15, 2025, 7:38 PM
S098 M2 P6.R	457 B	Dec 15, 2025, 7:44 PM
S098 P3.R	333 B	Nov 24, 2025, 6:13 PM
S098 P4.R	1.5 KB	Nov 24, 2025, 7:25 PM
S098 P5.R	596 B	Nov 24, 2025, 10:45 PM
sales_data.csv	101.2 KB	Nov 18, 2025, 6:32 PM
Student Mental health.csv	7.2 KB	Nov 18, 2025, 7:2 PM
student_exam_scores.csv	5.2 KB	Nov 18, 2025, 7:3 PM
StudentsPerformance.csv	70.3 KB	Dec 15, 2025, 7:09 PM
Untitled.R	39 KB	Nov 24, 2025, 5:55 PM
xAPI-Edu-Data (1).csv	37.1 KB	Nov 15, 2025, 7:09 PM
Sleep_Efficiency (1).csv	40.4 KB	Dec 15, 2025, 7:10 PM

Activate Windows

On-Screen Keyboard

08:11 PM 15-12-2025



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source on Save | Run | Source | Environment History Connections Tutorial

```

1 # Practical 5: Independent Two Sample t-test
2
3 # Load dataset
4 df <- read.csv("heart (1).csv")
5
6 str(df)
8:1 [Top Level] R Script

Console Terminal Background Jobs
R > R 4.5.2 . ~/ ~
> str(df)
'data.frame': 303 obs. of 14 variables:
 $ age : num 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
 $ sex : int 0 0 2 2 2 1 1 2 2 2 ...
 $ cp : int 0 0 0 0 0 0 0 0 0 0 ...
 $ trestbps: num 145 130 130 120 140 140 120 172 150 ...
 $ chol : num 233 250 204 238 354 192 294 263 199 168 ...
 $ fbs : int 0 0 0 0 0 0 0 0 1 0 ...
 $ restecg : int 0 1 0 1 1 1 0 1 1 1 ...
$ thalach : num 150 187 172 178 163 148 153 173 162 174 ...
$ exang : num 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] 0
>
> # Keep only two groups (0 and 1)
> df_two <- subset(df, target %in% c(0, 1))
>
> # 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

```

Files Plots Packages Help Viewer Presentation

Home

Name	Size	Modified
S098 M2 P3.R	180 B	Dec 15, 2025, 7:29 PM
S098 M2 P4.R	212 B	Dec 15, 2025, 7:34 PM
S098 M2 P5.R	309 B	Dec 15, 2025, 7:38 PM
S098 M2 P6.R	457 B	Dec 15, 2025, 7:44 PM
S098 P3.R	333 B	Nov 24, 2025, 6:13 PM
S098 P4.R	1.5 KB	Nov 24, 2025, 7:25 PM
S098 P5.R	596 B	Nov 24, 2025, 10:45 PM
sales_data.csv	101.2 KB	Nov 18, 2025, 6:32 PM
Student Mental health.csv	7.2 KB	Nov 18, 2025, 7:2 PM
student_exam_scores.csv	5.2 KB	Nov 18, 2025, 7:3 PM
StudentsPerformance.csv	70.3 KB	Dec 15, 2025, 7:09 PM
Untitled.R	39 KB	Nov 24, 2025, 5:55 PM
xAPI-Edu-Data (1).csv	37.1 KB	Nov 15, 2025, 7:09 PM
Sleep_Efficiency (1).csv	40.4 KB	Dec 15, 2025, 7:10 PM

Activate Windows

On-Screen Keyboard

08:11 PM 15-12-2025

# 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

Source on Save Go to file/function Addins

Console Terminal Background Jobs

```
R > # 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" "2021-11-03 02:30:00" ...
 $ Wakeup.time  : chr "2021-03-06 07:00:00" "2021-12-05 09:00:00" "2021-05-25 05:30:00" "2021-11-03 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" "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$area_before <- df$Sleep.efficiency
df$area_after <- df$Sleep.efficiency + rnorm(nrow(df), mean = -5, sd = 2)
>
> # Paired t-test
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 = 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

Project: (None)

Files Plots Packages Help Viewer Presentation

Home

Activate Windows

On-Screen Keyboard

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source on Save Go to file/function Addins

Console Terminal Background Jobs

```
R > # 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" "2021-11-03 02:30:00" ...
 $ Wakeup.time  : chr "2021-03-06 07:00:00" "2021-12-05 09:00:00" "2021-05-25 05:30:00" "2021-11-03 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" "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$area_before <- df$Sleep.efficiency
df$area_after <- df$Sleep.efficiency + rnorm(nrow(df), mean = -5, sd = 2)
>
> # Paired t-test
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 = 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

Project: (None)

Files Plots Packages Help Viewer Presentation

Home

Activate Windows

On-Screen Keyboard