

## PRACTICAL NO 4

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

```

R - R4.2.2 - ~/R
> library(dplyr)
> library(readr)
> # Load dataset
> df <- read_csv("%/exams.csv")
Rows: 1000 Columns: 9
--- Column specification
Delimiter: ","
chr (5): gender, race/ethnicity, parental level of education, lunch, test preparation course
dbl (3): math score, reading score, writing score

# Use 'spec()' to retrieve the full column specification for this data.
# Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # Rename math score as Marks (to keep logic similar)
> df <- df %>% rename(Marks = "math score")
> # Summary statistics
> print("Summary of Marks:")
[1] "Summary of Marks:"
> summary(df$Marks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 13.0   56.0   66.5   66.4   77.0   100.0
> # Descriptive statistics (base R alternative)
> print("Descriptive statistics of Marks:")
[1] "Descriptive statistics of Marks:"
> mean(df$Marks)
[1] 66.396
> sd(df$Marks)
[1] 15.40287
> min(df$Marks)
[1] 13
> max(df$Marks)
[1] 100
> # Frequency table: Test Preparation course
> prep_counts <- table(df$test_preparation_course)
> print("Frequency Table: Test Preparation Course")
[1] "Frequency Table: Test Preparation Course"
> print(prepare_counts)
completed      none
           335      665
> # Data Frame format
> prep_df <- df %>% count("test preparation course")
> print("Test Preparation Course Frequency (Data Frame Format)")
[1] "Test Preparation Course Frequency (Data Frame Format)"
> print(prepare_df)
# A tibble: 2 x 2
  test_preparation_course     n
  <chr>             <int>
1 completed           335
2 none                665
> # Create Marks groups (Low / Medium / High)
> df$Marks_Group <- cut(
+   df$Marks,
+   breaks = 3,
+   labels = c("Low", "Medium", "High")
+ )
> print("Cross Tabulation: Test Prep Course vs Marks Group")
[1] "Cross Tabulation: Test Prep Course vs Marks Group"
> cross_tab <- table(df$test_preparation_course, df$Marks_Group)
> print(cross_tab)
           Low Medium High
completed  11  166  138
none       31  393  221
# One-sample t-test: Is mean Marks different from 50?
> print("One-sample t-test: Marks vs mu = 50")
[1] "One-sample t-test: Marks vs mu = 50"
> t.test_one <- t.test(df$Marks, mu = 50)
> print(t.test_one)

One sample t-test

data: df$Marks
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 67.33182
sample estimates:
mean of x
 66.396

# Independent t-test: Marks by test preparation course
> print("Independent t-test: Marks by Test Preparation Course")
[1] "Independent t-test: Marks by Test Preparation Course"
> t.test_two <- t.test(
+   Marks ~ test_preparation_course,
+   data = df
+ )
> print(t.test_two)

Welch Two Sample t-test

data: Marks by test preparation course
t = 4.8531, df = 670.68, p-value = 1.528e-06
alternative hypothesis: true difference in means between group completed and group none is not equal to 0
95 percent confidence interval:
 2.945390 6.931033
sample estimates:
mean in group completed      mean in group none
           69.68657              64.73833

# Paired t-test: Math vs Reading scores
> print("Paired t-test: Math vs Reading Scores")
[1] "Paired t-test: Math vs Reading Scores"
> t.test_paired <- t.test(

```

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

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation and statistical testing.
- Console:** Shows the output of the executed R code.
- Environment:** Lists loaded data frames and their dimensions.
- Files:** Shows a file explorer with CSV files.

**R Code in Source Editor:**

```
R > R452 -> />
+ table11 = c("Low", "Medium", "High")
+ )
> print("Cross tabulation: Test prep course vs Marks group")
[1] "Cross Tabulation: Test prep course vs Marks group"
> cross_tab <- table(df$test_preparation_course, df$marks_Group)
> print(cross_tab)

      Low Medium High
completed 11  166 158
none      51  393 221
> # One-sample t-test: Is mean Marks different from 50?
> print("One-sample t-test: Marks vs mu = 50")
[1] "One-sample t-test: Marks vs mu = 50"
> t_test_one <- t.test(df$marks, mu = 50)
> print(t_test_one)

One Sample t-test

data:  df$marks
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

> # Independent t-test: Marks by test preparation course
> print("Independent t-test: Marks by Test Preparation Course")
[1] "Independent t-test: Marks by Test Preparation Course"
> t_test_two <- t.test(
+   Marks ~ test_preparation_course,
+   data = df
+ )
> print(t_test_two)

Welch Two Sample t-test

data:  Marks by test preparation course
t = 4.8531, df = 670.68, p-value = 1.528e-06
alternative hypothesis: true difference in means between group completed and group none is not equal to 0
95 percent confidence interval:
 2.945390 6.951053
sample estimates:
mean in group completed      mean in group none
      69.68657              64.73855

> # Paired t-test: Math vs Reading scores
> print("Paired t-test: Math vs Reading scores")
[1] "Paired t-test: Math vs Reading scores"
> t_test_paired <- t.test(
+   df$math_score,
+   df$reading_score,
+   paired = TRUE
+ )
```

**Environment Data:**

Object	Dimensions
df	1000 obs. of 9 variables
employee_salary_dataset	50 obs. of 9 variables
exams	1000 obs. of 8 variables
prep_df	2 obs. of 2 variables
superstore	51200 obs. of 27 variables
t_test_one	List of 10
t_test_two	List of 10

**Files:**

Name	Size	Modified
employee_salary_dataset.csv	2.8 KB	Dec 15, 2025, 8:11 PM
superstore.csv	15.1 MB	Dec 15, 2025, 7:21 PM
exams.csv	70.1 KB	Dec 15, 2025, 8:29 PM