

SHEETH L.U.J. AND SIR M.V. COLLEGE  
SUBJECT: DATA ANALYSIS WITH R  
PRACTICAL 1 TO 6 MOD 2

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

The screenshot shows the RStudio interface. The Source pane on the left contains the following R code:

```
R 4.5.2 - D:\S095 Aashka\
[4] "Iris.csv"
[7] "prac2.ipynb"
[10] "S095 R Prac 11.R"
[11] "S095 R Prac 14.R"
[16] "S095 R Prac 6.R"
[19] "S095 R Prac 9.R"
[22] "Student Stress Factors.csv"
      "myfirst.js"
      "Retail Product.csv"
      "S095 R Prac 10.R"
      "S095 R Prac 12.R"
      "S095 R Prac 13.R"
      "S095 R Prac 15.R"
      "S095 R Prac 3.R"
      "S095 R Prac 7.R"
      "S095 R Prac 8.R"
      "Student Mental health.csv"
      "sales_data.csv"
      "student_exam_scores.csv"
> # Load required libraries
> library(psych) # for describe()
> library(readr) # for reading csv
>
> # Read the dataset
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
--- Column specification ---
Delimiter: ","
chr (1): student_id
dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score
i use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
>
> # view dataset
> view(df)
>
> # Descriptive statistics
> summary(df)
  student_id      hours_studied      sleep_hours      attendance_percent      previous_scores      exam_score
Length:200      Min.      :1.000      Min.      :4.000      Min.      :50.30      Min.      :40.0      Min.      :17.10
Class :character 1st Qu.: 3.500      1st Qu.: 5.300      1st Qu.: 62.20      1st Qu.: 54.0      1st Qu.: 29.50
Mode :character  Median : 6.150      Median : 6.700      Median : 75.25      Median : 67.5      Median : 34.05
      Mean : 6.325      Mean : 6.622      Mean : 74.83      Mean : 66.8      Mean : 33.95
      3rd Qu.: 9.000      3rd Qu.: 8.025      3rd Qu.: 87.42      3rd Qu.: 80.0      3rd Qu.: 38.75
      Max. :12.000      Max. : 9.000      Max. :100.00      Max. : 95.0      Max. : 51.30
>
> describe(df)
  vars  n  mean  sd median trimmed  mad min  max range skew kurtosis  se
student_id* 1 200 100.50 57.88 100.50 100.50 74.13 1.0 200.0 199.0 0.00 -1.22 4.09
hours_studied 2 200 6.33 3.23 6.15 6.28 4.08 1.0 12.0 11.0 0.12 -1.24 0.23
sleep_hours 3 200 6.62 1.50 6.70 6.64 2.08 4.0 9.0 5.0 -0.04 -1.31 0.11
attendance_percent 4 200 74.83 14.25 75.25 74.94 18.46 50.3 100.0 49.7 -0.07 -1.22 1.01
previous_scores 5 200 66.80 15.66 67.50 66.69 20.02 40.0 95.0 55.0 0.06 -1.24 1.11
exam_score 6 200 33.95 6.79 34.05 33.94 6.75 17.1 51.3 34.2 0.03 -0.39 0.48
> |
```

The Environment pane on the right shows the following objects:

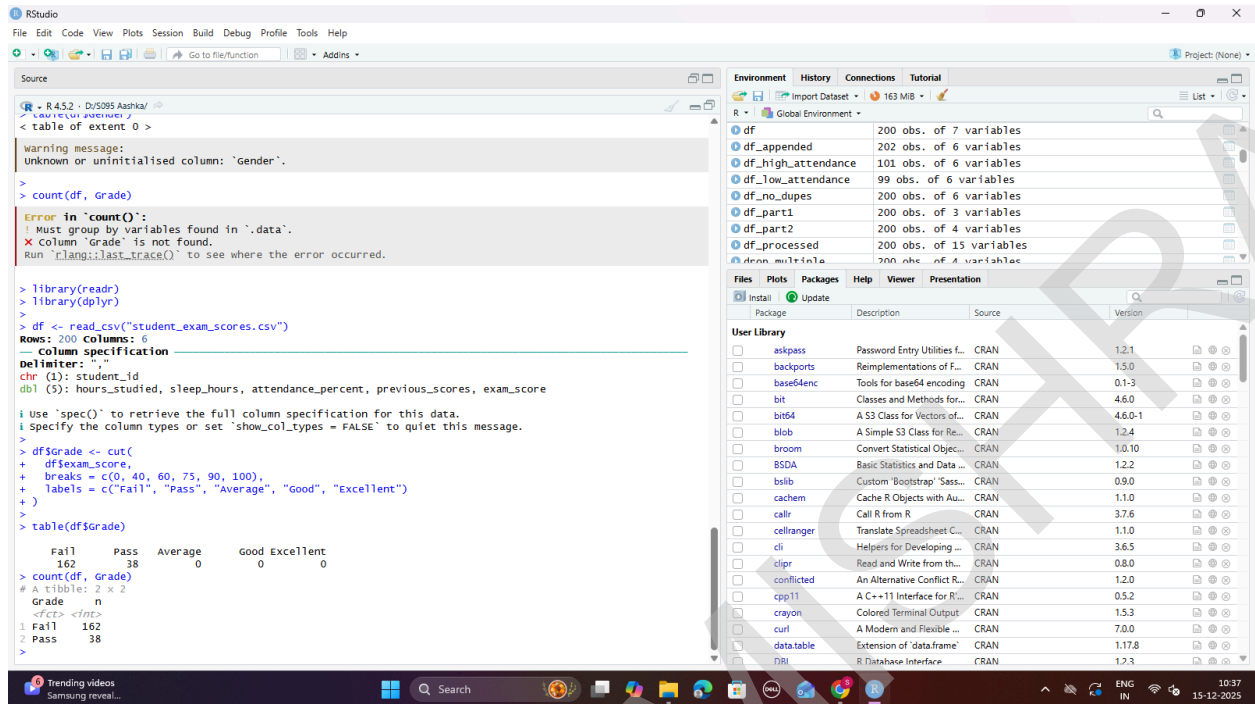
Object	Obs	Vars
df	200	6
df_appended	202	6
df_high_attendance	101	6
df_low_attendance	99	6
df_no_dups	200	6
df_part1	200	3
df_part2	200	4
df_processed	200	15
dfnn_multifila	200	15

The User Library pane shows a list of installed packages, including askpass, backports, base64enc, bit, bit64, blob, broom, BSDA, bslib, cachem, callr, cellranger, cli, clipr, conflicted, cpp11, crayon, curl, data.table, and dplyr.

NAME: AASHKA MISHRA  
ROLL NO: S095

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

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



The screenshot shows the RStudio interface. The Source pane on the left contains the following R code:

```
> library(readr)
> library(dplyr)
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
column specification
  delimiter: ","
chr (1): student_id
dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score

i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> df$grade <- cut(
+   df$exam_score,
+   breaks = c(0, 40, 60, 75, 90, 100),
+   labels = c("Fail", "Pass", "Average", "Good", "Excellent")
+ )
> table(df$grade)

  Fail    Pass  Average    Good Excellent 
   162     38         0         0         0 
> count(df, grade)
# A tibble: 2 x 2
  grade     n
<fct> <int>
1 Fail    162
2 Pass     38
```

The Environment pane on the right shows the Global Environment with the following objects:

Object	Size
df	200 obs. of 7 variables
df_appended	202 obs. of 6 variables
df_high_attendance	101 obs. of 6 variables
df_low_attendance	99 obs. of 6 variables
df_no_dupes	200 obs. of 6 variables
df_part1	200 obs. of 3 variables
df_part2	200 obs. of 4 variables
df_processed	200 obs. of 15 variables
df_res_multinla	200 obs. of 4 variables

The User Library pane shows a list of installed packages, including askpass, backports, base64enc, bit, bit64, blob, broom, BSDA, bslib, cachem, callr, cellranger, cli, clipr, conflicted, cpp11, crayon, curl, data.table, and dplyr.

NAME: AASHKA MISHRA  
ROLL NO: S095

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

## SUBJECT: DATA ANALYSIS WITH R

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

```
R - R 4.5.2 - D:\S095 Aashka
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R - R 4.5.2 - D:\S095 Aashka
> df$Attendance_Level <- ifelse(df$Attendance_percent >= 75,
+                               "High Attendance",
+                               "Low Attendance")
> table(df$Grade, df$Attendance_Level)

      High Attendance Low Attendance
Fail           77           85
Pass           24           14
Average         0            0
Good            0            0
Excellent       0            0
> library(readr)
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
Column specification
Delimiter: ","
chr (1): student_id
dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score
I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> df$Grade <- cut(
+   df$exam_score,
+   breaks = c(0, 40, 60, 75, 90, 100),
+   labels = c("Fail", "Pass", "Average", "Good", "Excellent")
+ )
> df$Attendance_Level <- ifelse(df$Attendance_percent >= 75,
+                               "High Attendance",
+                               "Low Attendance")
> table(df$Grade, df$Attendance_Level)

      High Attendance Low Attendance
Fail           77           85
Pass           24           14
Average         0            0
Good            0            0
Excellent       0            0
>

Environment History Connections Tutorial
R - Global Environment
df 200 obs. of 8 variables
df_appended 202 obs. of 6 variables
df_high_attendance 101 obs. of 6 variables
df_low_attendance 99 obs. of 6 variables
df_no_dups 200 obs. of 6 variables
df_part1 200 obs. of 3 variables
df_part2 200 obs. of 4 variables
df_processed 200 obs. of 15 variables
df_res_multinla 200 obs. of 4 variables

Files Plots Packages Help Viewer Presentation
Package Update
askpass Password Entry Utilities f... CRAN 1.2.1
backports Reimplementations of F... CRAN 1.5.0
base64enc Tools for base64 encoding CRAN 0.1-3
bit Classes and Methods for... CRAN 4.6.0
bit64 A 64 Bit Class for Vectors of... CRAN 4.6.0-1
blob A Simple S3 Class for Re... CRAN 1.2.4
broom Convert Statistical Objec... CRAN 1.0.10
BSDA Basic Statistics and Data... CRAN 1.2.2
bslib Custom "Bootstrap" Sess... CRAN 0.9.0
cachem Cache R Objects with Au... CRAN 1.1.0
callr Call R from R CRAN 3.7.6
cellranger Translate Spreadsheet C... CRAN 1.1.0
cli Helpers for Developing ... CRAN 3.6.5
clipr Read and Write from th... CRAN 0.8.0
conflicted An Alternative Conflict R... CRAN 1.2.0
cpp11 A C++11 Interface for R... CRAN 0.5.2
crayon Colored Terminal Output... CRAN 1.5.3
curl A Modern and Flexible ... CRAN 7.0.0
data.table Extension of 'data.frame' CRAN 1.17.8
furl R Database Interface CRAN 1.2.3
```

NAME: AASHKA MISHRA  
ROLL NO: S095

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

## SUBJECT: DATA ANALYSIS WITH R

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

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
R - R4.5.2 - D:/S095 Aashka/
Average 0 0
Good 0 0
Excellent 0 0
> library(readr)
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
-- column specification --
delimiter: ","
chr (1): student_id
dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score
I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> t.test(df$exam_score, mu = 50)

One Sample t-test

data: df$exam_score
t = -33.421, df = 199, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 33.00828 34.90172
sample estimates:
mean of x
 33.955

> t.test(df$exam_score, mu = 50, alternative = "greater")

One Sample t-test

data: df$exam_score
t = -33.421, df = 199, p-value = 1
alternative hypothesis: true mean is greater than 50
95 percent confidence interval:
 33.16162      Inf
sample estimates:
mean of x
 33.955

> t.test(df$exam_score, mu = 50, alternative = "less")

One Sample t-test

data: df$exam_score
t = -33.421, df = 199, p-value < 2.2e-16
alternative hypothesis: true mean is less than 50
95 percent confidence interval:
 -Inf 34.74838
sample estimates:
mean of x
 33.955

> |

Environment History Connections Tutorial
R - Global Environment
df 200 obs. of 6 variables
df_appended 202 obs. of 6 variables
df_high_attendance 101 obs. of 6 variables
df_low_attendance 99 obs. of 6 variables
df_no_dups 200 obs. of 6 variables
df_part1 200 obs. of 3 variables
df_part2 200 obs. of 4 variables
df_processed 200 obs. of 15 variables
df_spec_multisla 200 obs. of 4 variables

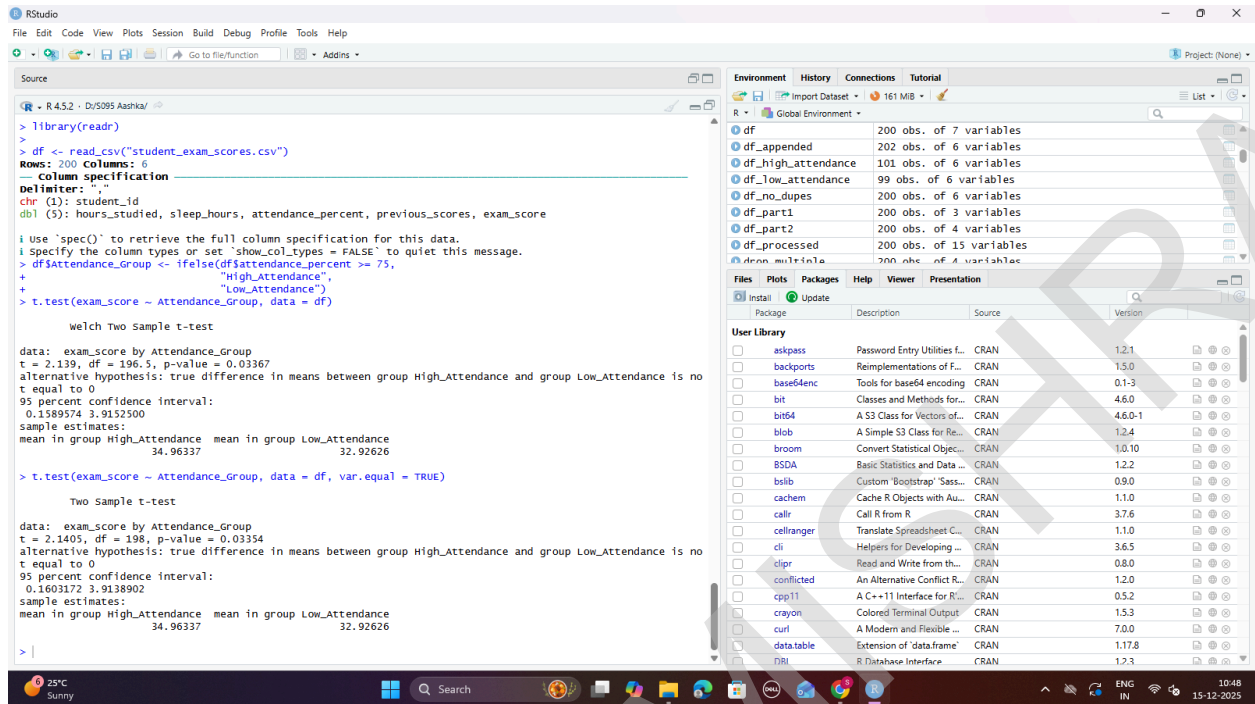
Files Plots Packages Help Viewer Presentation
Package Description Source Version
askpass Password Entry Utilities f... CRAN 1.2.1
backports Reimplementations of F... CRAN 1.5.0
base64enc Tools for base64 encoding CRAN 0.1-3
bit Classes and Methods for... CRAN 4.6.0
bit64 A 63 Class for Vectors of... CRAN 4.6.0-1
blob A Simple S3 Class for Re... CRAN 1.2.4
broom Convert Statistical Objec... CRAN 1.0.10
BSDA Basic Statistics and Data... CRAN 1.2.2
bslib Custom 'Bootstrap' Sass... CRAN 0.9.0
cachem Cache R Objects with Au... CRAN 1.1.0
callr Call R from R CRAN 3.7.6
cellranger Translate Spreadsheet C... CRAN 1.1.0
cli Helpers for Developing ... CRAN 3.6.5
clipr Read and Write from th... CRAN 0.8.0
conflicted An Alternative Conflict R... CRAN 1.2.0
cpp11 A C++11 Interface for R... CRAN 0.5.2
crayon Colored Terminal Output... CRAN 1.5.3
curl A Modern and Flexible ... CRAN 7.0.0
data.table Extension of 'data.frame' CRAN 1.17.8
furl R Database Interface CRAN 1.2.3
```

NAME: AASHKA MISHRA  
ROLL NO: S095

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

## SUBJECT: DATA ANALYSIS WITH R

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



```
> library(readr)
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
Column specification:
  delimiter: ","
  chr (1): student_id
  dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score

i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> df$Attendance_Group <- ifelse(df$attendance_percent >= 75,
+                               "High_Attendance",
+                               "Low_Attendance")
> t.test(exam_score ~ Attendance_Group, data = df)

Welch Two Sample t-test

data: exam_score by Attendance_Group
t = 2.139, df = 196.5, p-value = 0.03367
alternative hypothesis: true difference in means between group High_Attendance and group Low_Attendance is not
equal to 0
95 percent confidence interval:
 0.1589574 3.9152500
sample estimates:
mean in group High_Attendance mean in group Low_Attendance
          34.96337              32.92626

> t.test(exam_score ~ Attendance_Group, data = df, var.equal = TRUE)

Two Sample t-test

data: exam_score by Attendance_Group
t = 2.1405, df = 198, p-value = 0.03354
alternative hypothesis: true difference in means between group High_Attendance and group Low_Attendance is not
equal to 0
95 percent confidence interval:
 0.1603172 3.9138902
sample estimates:
mean in group High_Attendance mean in group Low_Attendance
          34.96337              32.92626

> |
```

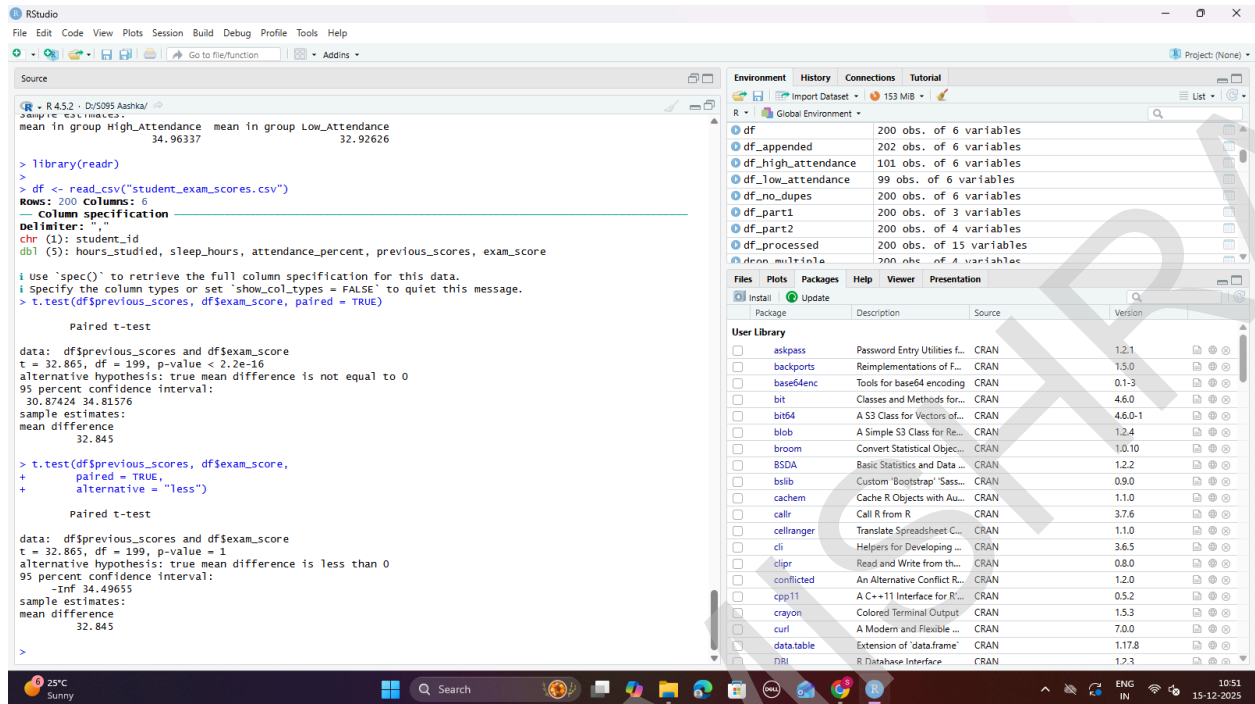
Package	Description	Source	Version
askpass	Password Entry Utilities f...	CRAN	1.2.1
backports	Reimplementations of F...	CRAN	1.5.0
base64enc	Tools for base64 encoding	CRAN	0.1-3
bit	Classes and Methods for...	CRAN	4.6.0
bit64	A 63 Class for Vectors of...	CRAN	4.6.0-1
blob	A Simple S3 Class for Re...	CRAN	1.2.4
broom	Convert Statistical Objec...	CRAN	1.0.10
BSDA	Basic Statistics and Data...	CRAN	1.2.2
bslib	Custom 'Bootstrap' Sass...	CRAN	0.9.0
cachem	Cache R Objects with Au...	CRAN	1.1.0
callr	Call R from R	CRAN	3.7.6
cellranger	Translate Spreadsheet C...	CRAN	1.1.0
cli	Helpers for Developing ...	CRAN	3.6.5
clipr	Read and Write from th...	CRAN	0.8.0
conflicted	An Alternative Conflict R...	CRAN	1.2.0
cxx11	A C++11 Interface for R...	CRAN	0.5.2
crayon	Colored Terminal Output	CRAN	1.5.3
curl	A Modern and Flexible ...	CRAN	7.0.0
data.table	Extension of 'data.frame'	CRAN	1.17.8
DBI	R Database Interface	CRAN	1.2.3

NAME: AASHKA MISHRA  
ROLL NO: S095

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

## SUBJECT: DATA ANALYSIS WITH R

AIM: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
Source
R 4.3.2 - D:/S095 Aashka/
D:/S095 Aashka/
mean in group High_Attendance mean in group Low_Attendance
34.96337 32.92626

> library(readr)
>
> df <- read_csv("student_exam_scores.csv")
Rows: 200 Columns: 6
Column specification
Delimiter: ","
chr (1): student_id
dbl (5): hours_studied, sleep_hours, attendance_percent, previous_scores, exam_score

I use 'spec()' to retrieve the full column specification for this data.
I specify the column types or set 'show_col_types = FALSE' to quiet this message.
> t.test(df$previous_scores, df$exam_score, paired = TRUE)

Paired t-test

data: df$previous_scores and df$exam_score
t = 32.865, df = 199, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 30.87424 34.81576
sample estimates:
mean difference
 32.845

> t.test(df$previous_scores, df$exam_score,
+       paired = TRUE,
+       alternative = "less")

Paired t-test

data: df$previous_scores and df$exam_score
t = 32.865, df = 199, p-value = 1
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
 -Inf 34.49655
sample estimates:
mean difference
 32.845

>

Environment History Connections Tutorial
R - Global Environment
df 200 obs. of 6 variables
df_appended 202 obs. of 6 variables
df_high_attendance 101 obs. of 6 variables
df_low_attendance 99 obs. of 6 variables
df_no_dupes 200 obs. of 6 variables
df_part1 200 obs. of 3 variables
df_part2 200 obs. of 4 variables
df_processed 200 obs. of 15 variables
df_spec_multisets 200 obs. of 4 variables

Files Plots Packages Help Viewer Presentation
Package Install Update Description Source Version
User Library
askpass Password Entry Utilities f... CRAN 1.2.1
backports Reimplementations of F... CRAN 1.5.0
base64enc Tools for base64 encoding CRAN 0.1-3
bit Classes and Methods for... CRAN 4.6.0
bit64 A 63 Class for Vectors of... CRAN 4.6.0-1
blob A Simple S3 Class for Re... CRAN 1.2.4
broom Convert Statistical Objec... CRAN 1.0.10
BSDA Basic Statistics and Data... CRAN 1.2.2
bslib Custom 'Bootstrap' Sass... CRAN 0.9.0
cachem Cache R Objects with Au... CRAN 1.1.0
callr Call R from R CRAN 3.7.6
cellranger Translate Spreadsheet C... CRAN 1.1.0
cli Helpers for Developing ... CRAN 3.6.5
clipr Read and Write from th... CRAN 0.8.0
conflicted An Alternative Conflict R... CRAN 1.2.0
cpp11 A C++11 Interface for R... CRAN 0.5.2
crayon Colored Terminal Output... CRAN 1.5.3
curl A Modern and Flexible ... CRAN 7.0.0
data.table Extension of 'data.frame' CRAN 1.17.8
furl R Database Interface CRAN 1.2.3
```

NAME: AASHKA MISHRA  
ROLL NO: S095