

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

PRACTICAL NO 1

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

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to File/Function Adding
Source Console Terminal Background Jobs
R - R452
> library(psych)
> df <- read_csv("R/student_performance_analysis.csv")
Rows: 50 Columns: 6
# Column specification
Delimiter: ","
chr (1): gender
dbl (5): student_id, hours_studied, attendance_percent, assignments_completed, test_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.
> print("structure of the dataset:")
[1] "Structure of the dataset:"
> str(df)
spec_tbl_ [50 x 6] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ student_id      : num [1:50] 1 2 3 4 5 6 7 8 9 10 ...
 $ gender          : chr [1:50] "female" "Male" "female" "Male" ...
 $ hours_studied   : num [1:50] 6.3 7.6 8.9 10.2 11.5 12.8 14.1 15.4 16.7 5 ...
 $ attendance_percent : num [1:50] 61.62 63.64 65.66 67.68 69.70 ...
 $ assignments_completed: num [1:50] 6 7 8 9 10 11 12 13 14 5 ...
 $ test_score      : num [1:50] 50.9 51.8 52.7 53.6 54.5 55.4 56.3 57.2 58.1 59 ...
- attr(*, "spec")=
.. cols(
..   student_id = col_double(),
..   gender = col_character(),
..   hours_studied = col_double(),
..   attendance_percent = col_double(),
..   assignments_completed = col_double(),
..   test_score = col_double()
.. )
- attr(*, "problems")=externalptr
> print("column names of the dataset:")
[1] "Column names of the dataset:"
> colnames(df)
[1] "student_id" "gender" "hours_studied" "attendance_percent"
[5] "assignments_completed" "test_score"
> if ("Score" %in% colnames(df)) {
+   print("Converting Score column to numeric")
+   df$Score <- as.numeric(as.character(df$Score))
+ } else {
+   print("score column not found in dataset")
+ }
[1] "score column not found in dataset"
> if ("Practice_Hours" %in% colnames(df)) {
+   print("Converting Practice_Hours column to numeric")
+   df$Practice_Hours <- as.numeric(as.character(df$Practice_Hours))
+ } else {
+   print("Practice_Hours column not found in dataset")
+ }
[1] "Practice_Hours column not found in dataset"
> if ("Experience_Years" %in% colnames(df)) {
+   print("Converting Experience_Years column to numeric")
+   df$Experience_Years <- as.numeric(as.character(df$Experience_Years))
+ } else {
+   print("Experience_Years column not found in dataset")
+ }
[1] "Experience_Years column not found in dataset"
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to File/Function Adding
Source Console Terminal Background Jobs
R - R452
> print("dataset after converting columns:")
[1] "Dataset after converting columns:"
> str(df)
spec_tbl_ [50 x 6] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ student_id      : num [1:50] 1 2 3 4 5 6 7 8 9 10 ...
 $ gender          : chr [1:50] "female" "Male" "female" "Male" ...
 $ hours_studied   : num [1:50] 6.3 7.6 8.9 10.2 11.5 12.8 14.1 15.4 16.7 5 ...
 $ attendance_percent : num [1:50] 61.62 63.64 65.66 67.68 69.70 ...
 $ assignments_completed: num [1:50] 6 7 8 9 10 11 12 13 14 5 ...
 $ test_score      : num [1:50] 50.9 51.8 52.7 53.6 54.5 55.4 56.3 57.2 58.1 59 ...
- attr(*, "spec")=
.. cols(
..   student_id = col_double(),
..   gender = col_character(),
..   hours_studied = col_double(),
..   attendance_percent = col_double(),
..   assignments_completed = col_double(),
..   test_score = col_double()
.. )
- attr(*, "problems")=externalptr
> print("Descriptive statistics using describe() function:")
[1] "Descriptive statistics using describe() function:"
> describe(df)
vars n mean sd median trimmed mad min max range skew kurtosis se
student_id 1 50 25.50 14.58 25.50 25.50 18.53 1 50.0 49.0 0.00 -1.27 2.06
gender 2 50 1.50 0.51 1.50 1.50 0.74 1 2.0 1.0 0.00 -2.04 0.07
hours_studied 3 50 10.85 1.77 10.85 10.85 4.82 5 16.7 11.7 0.00 -1.29 0.53
attendance_percent 4 50 76.10 11.94 74.50 76.10 14.08 60 99.0 39.0 0.34 -1.27 1.69
assignments_completed 5 50 9.50 2.90 9.50 9.50 3.71 5 14.0 9.0 0.00 -1.29 0.41
test_score 6 50 72.05 13.12 72.05 72.05 16.68 50 94.1 44.1 0.00 -1.27 1.86
```