

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce
SUBJECT NAME: Data Analysis with SAS / SPSS/R

PRACTICAL NO : 15

AIM : Generating basic summaries using str() or summary() (R).

The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code for generating basic summaries using the `str()` function.
- Environment:** Shows the global environment with various datasets like `insurance_df`, `retail_df`, and `Sample_Superstore`.
- File Explorer:** Shows files in the current directory, including CSV files and a PDF file.
- Bottom Bar:** Includes system icons for battery, signal, and date/time (08-12-2025, 12:11).

```

> # R Script: Generating Basic Summaries
> # Functions: str() and summary()
> # Dataset: Retail Product Data (Simulated)
> #=====

> # 1. SETUP: Create Sample Data
> #=====

> # we create a data frame with mixed data types (Numeric, character, Logical, NA)
> retail_df <- data.frame(
+   ID = 1:6,
+   Category = c("Electronics", "Home", "Electronics", "Clothing", "Home", "Clothing"),
+   Price = c(500.50, 45.00, 900.00, NA, 300.00, 25.00), # Note the NA
+   In_Stock = c(TRUE, TRUE, FALSE, TRUE, FALSE, TRUE),
+   Rating = c(4.5, 3.8, 4.9, 4.0, 3.5, 4.2)
+ )
> print("... Data Loaded ...")
[1] "... Data Loaded ..."
> #=====

> # 2. USING str() (Structure)
> #=====

> # Purpose: Compactly display the internal structure of the R object.
> # What it tells you:
> # - The class (data.frame)
> # - Number of observations (rows) and variables (columns)
> # - The data type of each column (num, int, chr, logi)
> # - The first few values
> print("... OUTPUT OF str() ...")
[1] "... OUTPUT OF str() ..."
> str(retail_df)
'data.frame': 6 obs. of 5 variables:
 $ ID : int 1 2 3 4 5 6
 $ Category: chr "Electronics" "Home" "Electronics" "Clothing" ...

```

The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code for generating statistical summaries using the `summary()` function.
- Environment:** Shows the global environment with various datasets.
- File Explorer:** Shows files in the current directory.
- Bottom Bar:** Includes system icons and date/time (08-12-2025, 12:11).

```

> # 3. USING summary() (Statistical Summary)
> #=====

> # Purpose: detailed summary statistics for each column.
> # What it tells you:
> # - Numeric cols: Min, 1st Qu, Median, Mean, 3rd Qu, Max, and NA count.
> # - Logical cols: Count of TRUE vs FALSE.
> print("... OUTPUT OF summary() [Before Factor Conversion] ...")
[1] "... OUTPUT OF summary() [Before Factor Conversion] ..."
> summary(retail_df)
   ID      Category       Price     In_Stock    Rating
Min.   :1.00 Length:6 Min.   :25.0 Mode :logical Min.  :3.500
1st Qu.:2.25 Class :character 1st Qu.:45.0 FALSE:2 1st Qu.:3.850
Median :3.50 Mode  :character Median :300.0 TRUE :4 Median :4.100
Mean   :3.50          Mean  :354.1          Mean  :4.150
3rd Qu.:4.75          3rd Qu.:500.5          3rd Qu.:4.425
Max.   :6.00          Max.  :900.0          Max.  :4.900
NA's   :1
```

Sheth L.U.J. College of Arts & Sir M.V. College Of Science & Commerce

SUBJECT NAME: Data Analysis with SAS / SPSS/R

The screenshot shows the RStudio environment. The left pane displays the R console with the following code and output:

```
NA's :1
> # 4. IMPROVING summary() WITH FACTORS
> # By default, R treats text as "character" and summary() just counts the length.
> # If we convert 'Category' to a 'Factor', summary() will calculate counts per category
y.
> retail_df$Category <- as.factor(retail_df$Category)
> print("--- OUTPUT OF summary() [After Factor conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor conversion] ---"
> # Notice how 'Category' now shows: Clothing:2, Electronics:2, Home:2
> summary(retail_df)
   ID      Category    Price  In_Stock Rating
Min. :1.00  Clothing :2  Min. :25.0 Mode :logical  Min. :3.500
1st Qu.:2.25 Electronics:2  1st Qu.:45.0 FALSE:2   1st Qu.:3.850
Median :3.50  Home     :2  Median :300.0 TRUE :4    Median :4.100
Mean   :3.50          Mean  :354.1          Mean  :4.150
3rd Qu.:4.75         3rd Qu.:500.5          3rd Qu.:4.425
Max.  :6.00          Max.  :900.0           Max.  :4.900
NA's  :1
```

The right pane shows the Global Environment and File Explorer. The Global Environment lists several datasets:

- insurance_df: 1338 obs. of 7 variables
- merged_data: 0 obs. of 28 variables
- my_data: 1000 obs. of 8 variables
- processed_data: 4 obs. of 11 variables
- range_cols: 1000 obs. of 4 variables
- retail_df: 6 obs. of 5 variables
- Sample_Superstore: 9994 obs. of 21 variables

The File Explorer shows the local directory structure:

- .RData
- .History
- American_Housing_Data_20231209.csv
- American_Housing_Data_20231209.csv.zip
- Bills.pdf
- Custom Office Templates
- desktop.ini
- Downloads
- Downloads - Copy
- GIS DataBase
- IISExpress
- My Music
- My Pictures
- Mv Videos

At the bottom, the taskbar shows various application icons and the system clock indicates 12:12 on 08-12-2025.