

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: **Data Analysis with SAS / SPSS / R**
Practical No. 7

Aim- Selecting and dropping variables using select () in R. import dataset.

Output-

```
> # 7. Selecting and dropping variables using select() in R
> # =====
> # 1. IMPORT DATASET
> # =====
> library(dplyr)
> # Import the csv file (using your corrected path with the space)
> cricket <- read.csv("D:/Dhairya/R/Fastest_Fifties_All_Seasons_Combine.csv")
> print("--- Original dataset (First 3 rows) ---")
[1] "--- Original dataset (First 3 rows) ---"
> print(head(cricket, 3))
  X      Player Runs BF X4s X6s Against      Venue      Match.Date
1  0      Yusuf Pathan    61  21   4   6    DEC Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
2  1 Kumar Sangakkara    94  23  13   1   MI     IS Bindra Stadium 25 April 2008
3  2 Virender Sehwag    51  23   2   5  PBKS    Arun Jaitley Stadium 17 May 2008
> # Method A: Select specific columns by name
> # Scenario: We only want the Player name, Runs scored, and the Opponent (Against)
> selected_cols <- cricket %>%
+   select(Player, Runs, Against)
> print("--- Selected specific columns ---")
[1] "--- Selected specific columns ---"
> print(head(selected_cols, 3))
  X      Player Runs Against
1  0      Yusuf Pathan    61  DEC
2  1 Kumar Sangakkara    94   MI
3  2 Virender Sehwag    51  PBKS
> # Method B: Select a range of adjacent columns
> # Scenario: Select everything from 'Player' to 'BF' (Balls Faced)
> range_cols <- cricket %>%
+   select(Player:BF)
> print("--- Selected range of columns ---")
[1] "--- Selected range of columns ---"
> print(head(range_cols, 3))
  X      Player Runs BF
1  0      Yusuf Pathan    61  21
2  1 Kumar Sangakkara    94  23
3  2 Virender Sehwag    51  23
> # Method C: Select using helper functions
> # Scenario: Select columns that start with 'R' (Runs)
> starts_with_r <- cricket %>%
+   select(starts_with("R"))
> print("--- Selected columns starting with 'R' ---")
[1] "--- Selected columns starting with 'R' ---"
> print(head(starts_with_r, 3))
  Runs
1    61
2    94
3    51
```

```
> print(head(range_cols, 3))
  X      Player Runs BF
1  0      Yusuf Pathan    61  21
2  1 Kumar Sangakkara    94  23
3  2 Virender Sehwag    51  23
> # Method C: Select using helper functions
> # Scenario: Select columns that start with 'R' (Runs)
> starts_with_r <- cricket %>%
+   select(starts_with("R"))
> print("--- Selected columns starting with 'R' ---")
[1] "--- Selected columns starting with 'R' ---"
> print(head(starts_with_r, 3))
  Runs
1    61
2    94
3    51
> # Method A: Drop a single specific column
> # Scenario: Remove the 'Venue' column
> dropped_one <- cricket %>%
+   select(-Venue)
> print("--- Dataset with 'Venue' dropped ---")
[1] "--- Dataset with 'Venue' dropped ---"
> print(names(dropped_one)) # printing names to verify it's gone
[1] "X"      "Player"    "Runs"     "BF"      "X4s"     "X6s"     "Against"  "Match.Date"
> # Method B: Drop multiple columns
> # Scenario: Remove 'BF' (Balls Faced) and 'Against'
> dropped_multiple <- cricket %>%
+   select(-BF, -Against)
> print("--- Dataset with 'BF' and 'Against' dropped ---")
[1] "--- Dataset with 'BF' and 'Against' dropped ---"
> print(names(dropped_multiple))
[1] "X"      "Player"    "Runs"     "X4s"     "X6s"     "Venue"   "Match.Date"
> # Method C: Drop a range of columns
> # Scenario: Remove everything from 'Runs' to 'BF'
> dropped_range <- cricket %>%
+   select(-(Runs:BF))
> print("--- Dataset with range 'Runs' to 'BF' dropped ---")
[1] "--- Dataset with range 'Runs' to 'BF' dropped ---"
> print(names(dropped_range))
[1] "X"      "Player"    "X4s"     "X6s"     "Against"  "Venue"   "Match.Date"
>
> |
```