

R Language Study Notes

Data Manipulation, Apply Functions & dplyr Package

Introduction

Data manipulation ka matlab hota hai raw data ko modify, clean, arrange aur summarize karna taake data analysis ke liye useful ban jaye. R language mein data manipulation ke liye do main approaches use hoti hain: apply functions (base R) aur dplyr package (modern R).

Data Manipulation

Data manipulation ka use data ki accuracy aur precision improve karne ke liye hota hai. Real-world data aksar incomplete ya messy hota hai, is liye analysis se pehle data manipulation zaroori hoti hai.

- Data clean karna
- Values change karna
- Naye columns banana
- Summary statistics nikalna

Apply Functions

Apply functions base R ka hissa hoti hain aur loops ke baghair repeated operations perform karne ke liye use hoti hain.

Is course mein hum teen apply functions cover karte hain:

- `apply()`
- `lapply()`
- `sapply()`

apply() Function

apply() function rows ya columns par function apply karta hai.

Syntax:

```
apply(data, MARGIN, function)
```

MARGIN = 1 ka matlab rows

MARGIN = 2 ka matlab columns

Example 1: Column-wise sum

```
m <- matrix(1:9, nrow=3)
apply(m, 2, sum)
```

Example 2: Row-wise sum

```
apply(m, 1, sum)
```

Example 3: Column-wise mean

```
apply(m, 2, mean)
```

Example 4: Row-wise maximum

```
apply(m, 1, max)
```

Example 5: Custom function

```
apply(m, 2, function(x) x * 2)
```

lapply() Function

lapply() har element par function apply karta hai aur output hamesha list hota hai.

Example 1: Square of numbers

```
x <- list(1,2,3,4)
lapply(x, function(i) i^2)
```

Example 2: Mean of columns

```
df <- data.frame(A=c(1,2,3), B=c(4,5,6))
lapply(df, mean)
```

Example 3: Length of columns

```
lapply(df, length)
```

Example 4: Convert to character

```
lapply(df, as.character)
```

Example 5: Add constant

```
lapply(x, function(i) i + 10)
```

sapply() Function

sapply() lapply() jaisa kaam karta hai lekin output ko simplify karta hai.

Example 1: Square of numbers

```
x <- c(1,2,3,4)
sapply(x, function(i) i^2)
```

Example 2: Mean of columns

```
sapply(df, mean)
```

Example 3: Length

```
sapply(df, length)
```

Example 4: Logical output

```
sapply(x, function(i) i > 2)
```

Example 5: Add constant

```
sapply(x, function(i) i + 5)
```

dplyr Package

dplyr aik modern R package hai jo tabular data ko manipulate karne ke liye simple aur readable functions provide karta hai. dplyr fast processing aur English-like commands use karta hai.

Main Features of dplyr

- Rows aur columns ke sath kaam karta hai
- Simple verbs (functions) use karta hai
- Fast processing provide karta hai

Common dplyr Functions

- `select()`
- `filter()`
- `arrange()`
- `mutate()`
- `summarize()`

Library load karna

```
library(dplyr)
```

`select()` Examples

```
select(mtcars, mpg)
select(mtcars, mpg, cyl)
select(mtcars, -hp)
select(mtcars, mpg:hp)
select(mtcars, mileage = mpg)
```

`filter()` Examples

```
filter(mtcars, mpg > 20)
filter(mtcars, cyl == 6)
filter(mtcars, mpg > 20 & cyl == 4)
filter(mtcars, mpg < 15 | hp > 150)
filter(mtcars, cyl %in% c(4,6))
```

`arrange()` Examples

```
arrange(mtcars, mpg)
arrange(mtcars, desc(mpg))
arrange(mtcars, cyl, mpg)
arrange(mtcars, hp)
arrange(mtcars, desc(hp))
```

mutate() Examples

```
mutate(mtcars, kmpl = mpg * 0.425)
mutate(mtcars, power_ratio = hp / wt)
mutate(mtcars, mpg2 = mpg * 2)
mutate(mtcars, is_fast = hp > 150)
mutate(mtcars, cyl_type = as.factor(cyl))
```

summarize() Examples

```
summarize(mtcars, avg_mpg = mean(mpg))
summarize(mtcars, max_hp = max(hp))
summarize(mtcars, min_wt = min(wt))
summarize(mtcars, count = n())

group_by(mtcars, cyl) %>%
summarize(avg_mpg = mean(mpg))
```

Conclusion

Apply functions aur dplyr package R language mein data manipulation ke liye bohat important tools hain. Apply functions loops ka efficient alternative hain jab ke dplyr package data manipulation ko readable, fast aur professional bana deta hai.