

Laborator 07

Petculescu Mihai-Silviu

Laborator 07

Petculescu Mihai-Silviu

Familia de funcții apply

lapply()

sapply()

apply()

tapply()

split()

Familia de funcții apply

lapply()

- Evaluează o funcție pentru fiecare element al unei liste

```
> set.seed(222)
> x = list(a = 1:5, b = rnorm(10), c = rnorm(20, 1), d = rnorm(100, 5))
> x
$a [1] 1 2 3 4 5
$b [1] 1.487757090 -0.001891901 1.381020790 -0.380213631 0.184136230 [...]
$c [1] 2.0524585 -0.3050636 0.3073924 1.6026489 0.8022469 -0.1858745 [...]
$d [1] 4.182571 5.675893 4.784519 4.885350 4.797735 5.406493 5.656772 5.106191
[...]

> lapply(x, mean)
$a [1] 3
$b [1] 0.1996044
$c [1] 0.7881026
$d [1] 5.064188

> x = 1:4
> lapply(x, runif, min = 0, max = 3)
[[1]] [1] 0.03443616
[[2]] [1] 1.267361 1.365441
[[3]] [1] 1.8084700 2.1902665 0.4139585
[[4]] [1] 1.5924650 0.7355067 2.1483841 1.6082945
```

sapply()

- La fel ca `lapply` numai că încearcă să simplifice rezultatul

```
> set.seed(222)
> x = list(a = 1:4, b = rnorm(10), c = rnorm(20, 1), d = rnorm(100, 5))
> sapply(x, mean)
      a      b      c      d
2.5000000 0.1996044 0.7881026 5.0641876
```

apply()

- Aplică o funcție după fiecare dimensiune a unui array
- Este echivalentă cu viteza unui `for`

```
> str(apply)
function (X, MARGIN, FUN, ...)
# X - un tablou multidimensional
# MARGIN - vector numeric care dimensiunile dupa care se va aplica
# FUN - functia ce urmeaza a fi aplicata (... - argumentele functiei)

> x = matrix(rnorm(200), 20, 10)
> apply(x, 2, mean) # media fiecărei coloane
[1] 0.28587446 0.05948248 0.21493204 -0.08170742 -0.15856742 -0.21369909
[7] 0.16799575 0.24737917 -0.04896986 -0.21706983

> apply(x, 1, sum) # media fiecărei linii
[1] 3.3527297 -4.4692479 2.6544048 6.3557239 7.5393424 -4.0228744
[7] 2.8193347 -0.6993723 -4.4275708 3.9586768 0.1403420 1.9774544
[13] -3.1413268 -2.7606268 0.6038099 -0.5686469 -0.1749793 2.7262661
[19] -5.3589594 -1.3914744
```

tapply()

- Aplică o funcție pe submulțimi ale unui vector
- Se comportă ca o combinație între `split()` și `sapply()`

```
> str(tapply)
function (X, INDEX, FUN = NULL, ..., default = NA, simplify = TRUE)
# X - vector
# INDEX - factor / lista de factori
# FUN - functia ce urmeaza a fi aplicata (... - argumentele functiei)
# simplify - simplificarea rezultatului

> x = c(rnorm(10), runif(10), rnorm(10, 1))
> f = gl(3, 10)
> f
[1] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
Levels: 1 2 3

> tapply(x, f, mean)
      1      2      3
-0.4350688 0.5007445 1.0403945

> tapply(x, f, range)
$`1` [1] -1.3347193 0.9701611
$`2` [1] 0.08515255 0.96974967
$`3` [1] -0.2190898 2.7965309
```

split()

- Împarte un vector în grupuri definite de o variabilă de tip factor.

```
> str(split)
function (x, f, drop = FALSE, ...)
# x - vector
# f - un factor sau o listă de factori

> x = c(rnorm(10), runif(10), rnorm(10, 1))
> f = gl(3, 10)
> split(x, f)
$`1` [1] -2.27414224 -0.11266780  0.61308167  0.07733545  0.57137727
      0.11672493
      [7] -0.95685256 -1.90008460 -1.48972089  0.55925676
$`2` [1]  0.91159086  0.03291829  0.78368939  0.11852882  0.64443831
      0.78790988
      [7]  0.82451477  0.05642366  0.65075027  0.95426854
$`3` [1]  2.6666242  2.6634334  1.8106280 -0.7837308  1.6575684  0.1546575
      [7]  0.4930056 -0.9031544  2.4042311  1.4106863

# lapply - pentru a aplica o functie FUN pe grupuri de date
> lapply(split(x, f), mean)
$`1` [1] 0.4907709
$`2` [1] 0.6276443
$`3` [1] 0.8362917
```