Funkce v R, příklady předzpracování dat v R

—

B03128 – Úvod do skriptovacího jazyka R (zkrácená online verze)

Lubomír Štěpánek[1, 2]

[1]Oddělení biomedicínské statistiky
Ústav biofyziky a informatiky
1. lékařská fakulta
Univerzita Karlova v Praze

ČVUT FBMI

[2]Katedra biomedicínské informatiky
Fakulta biomedicínského inženýrství
České vysoké učení technické v Praze

Obsah

Úvod do funkcí v R

- funkce jsou důležité části kódu (a tedy programu)
- kromě funkcí má smysl odlišovat ještě procedury

```r
is.function(print)      # TRUE
is.function("print")    # FALSE
```

Komponenty funkce

- funkce v R je složena ze tří částí
  - hlavička (*head* či *formals*)
    - obsahuje argumenty funkce
    - lze volat příkazem `formals()`
  - tělo (*body*)
    - obsahuje veškerý kód uvnitř funkce
    - lze volat příkazem `body()`
  - prostředí (*environment*)
    - mapuje lokalizaci všech proměnných ve funkci
    - lze volat příkazem `environment()`

```
1  f <- function(x){x^2}
2  f                    # function(x){x^2}
3
4  formals(f)           # $x
5  body(f)              # x^2
6  environment(f)       # <environment: R_GlobalEnv>
```

Vestavěné funkce

- většina funkcionality R je dána vestavěnými funkcemi
- ty jsou optimalizované, odladěné
- je-li to možné, je vhodné je preferovat před uživatelem definovanými funkcemi

Uživatelem definované funkce

- pro unikátnější operace je možné definovat vlastní funkce
- komponenty vlastní funkce jsou *hlavička s argumenty* a *tělo*
- obecná syntaxe

```
nazevFunkce <- function(
    argument_1,
    argument_2,
    # ...
){
    # '''
    # komentář
    # '''

    procedura s argumenty
    return(výstup)
}
```

Uživatelem definované funkce

- například

```
1  sectiCtverce <- function(a, b){
2
3    # '''
4    # vrací součet čtverců čísel "a" a "b"
5    # '''
6
7    return(a ^ 2 + b ^ 2)
8
9  }
```

Pořadí argumentů v hlavičce funkce

- při volání funkce mohou být argumenty v její hlavičce specifikovány úplným názvem, částečným názvem, nebo pořadím (sestupně dle priority v tomto pořadí)

```
1  f <- function(ab, abc, b){
2      1 * ab + 2* abc + 3 * b
3  }
4  f(ab = 1, abc = 2, b = 3)  # 14
5  f(abc = 2, ab = 1, b = 3)  # 14
6  f(1, 2, 3)                 # 14
7  f(2, 1, 3)                 # 13
8  f(ab = 1, ab = 2, b = 3)
9    # formal argument "ab" matched
10   # by multiple actual arguments
11 f(a = 1, abc = 2, b = 3)   # 14
```

Iterování nad funkcemi

- předpokládejme, že chceme pro vektor hodnot `x` zjistit postupně průměr, minimum, maximum, medián, směrodatnou odchylku, varianci a vždy poslední cifru čísla

```
set.seed(1); x <- floor(runif(100) * 100)

for(my_function in c(
    "mean",
    "min",
    "max",
    "median",
    "sd",
    "var",
    function(i) i %% 10
)){
    print(do.call(my_function, list(x)))
}
```

Defaultní argument

- funkce v R mohou mít defaultní hodnoty

```
f <- function(a = 1, b = 2){c(a, b)}
f()   # c(1, 2)
```

- díky lazy evaluaci argumentů je lze definovat i závisle na sobě

```
g <- function(a = 1, b = a * 2){c(a, b)}
g()   # c(1, 2)
```

- dokonce lze definovat argumenty i pomocí proměnných, které vznikají až v těle funkce

```
h <- function(a = 1, b = d){
  d <- (a + 1) ^ 2; c(a, b)
}
h()   # c(1, 4)
```

Chybějící argument

- ve funkcích v R lze kontrolovat, zda argument chybí, pomocí příkazu `missing()`

```
1  i <- function(a, b){
2    c(missing(a), missing(b))
3  }
4  i()        # TRUE TRUE
5  i(a = 1)   # FALSE TRUE
6  i(b = 2)   # TRUE FALSE
7  i(1, 2)    # FALSE FALSE
```

Dot-dot-dot argument (...)

- umožňuje ve funkcích v R používat argumenty, které jsou unikátní jen pro některé z vnořených funkcí

```
f <- function(x, ...){
  mean(x, ...)
}
f(c(0, 1, 2, 3, 3, 3, 3))
  # 2.142857
f(c(0, 1, 2, 3, 3, 3, 3), trim = 1)
  # 3
```

Krátká případová studie na předzpracování dat

- více ve skriptu `_03_script_.R`
  - pomocí balíčku `openxlsx` můžeme nahrajeme excelová data
  - poté přetypujeme proměnné do vhodných datových typů
  - dopočítáme některé nové proměnné
  - pro některé kombinace proměnných spočítáme průměrné hodnoty
  - vykreslíme některé diagramy

Literatura

Alain F. Zuur, Elena N. Ieno und Erik Meesters. *A Beginner's Guide to R*. Springer New York, 2009. DOI: 10.1007/978-0-387-93837-0. URL: https://doi.org/10.1007/978-0-387-93837-0.

Hadley Wickham. *Advanced R*. Boca Raton, FL: CRC Press, 2015. ISBN: 978-1466586963.

Děkuji za pozornost!

lubomir.stepanek@lf1.cuni.cz
lubomir.stepanek@fbmi.cvut.cz

▸ GitHub

https://github.com/LStepanek/B03128_Uvod_do_skriptovaciho_jazyka_R