



Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Programiranje v R-ju

3. Vektorji

Vladimir Batagelj

Univerza v Ljubljani, FMF, Matematika

Finančna matematika
Ljubljana, februar 2009
22. oktober 2012



Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

- 1 Sestavljeni podatki
- 2 Merske lestvice
- 3 Vektorji
- 4 Vektorske operacije
- 5 Indeksi
- 6 Stavek for
- 7 Množice
- 8 Preurejanja
- 9 Imenske in urejenostne spremenljivke
- 10 Posebne funkcije



Sestavljeni podatki

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

```
vector
factor
ordered
matrix
array
list
data.frame
Date
table
function
expression
environment
pairlist
hex mode, oct mode
```



Statistični podatki

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

V statistiki najpogosteje pridemo do podatkov na naslednji način: izberemo neko množico enot in nabor lastnosti teh enot, ki nas zanimajo. Naboru vrednosti neke lastnosti na množici enot rečemo *spremenljivka*. Spremenljivke v R-ju predstavimo z vektorji, celotno podatkovje pa kot razpredelnico.



Merske lestvice

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Pri merjenju poznamo več vrst merskih lestvic, ki določajo kakovost meritev in kaj lahko s podatki počnemo. Osnovne so naslednje (Stevens, 1946):

dovoljene transformacije	vrsta lestvice	primeri
$\varphi(x) = x$ (identiteta)	<i>absolutna</i>	štetje
$\varphi(x) = a \cdot x, a > 0$ podobnost	<i>razmernostna</i>	masa temperatura (K)
$\varphi(x) = a \cdot x + b, a > 0$	<i>razmična</i>	temperatura (C,F) čas (koledar)
$x \geq y \Leftrightarrow \varphi(x) \geq \varphi(y)$ strogo naraščajoča	<i>urejenostna</i>	šolske ocene, kakovost zraka, trdost kamnin
φ je povratno enolična	<i>imenska</i>	barva las, narodnost

Absolutne, razmernostne in razmične lestvice imenujemo tudi *številске*. Številске spremenljivke lahko predstavimo s številskimi vektorji. Kako predstavimo imenske in urejenostne spremenljivke v R-ju bomo spoznali kasneje.



Vektorji

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Mnozice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

V statistiki so najpogostejša oblika podatkov spremenljivke. **Spremenljivka** vsebuje vrednosti izbrane lastnosti izmerjene na dani množici enot. Spremenljivke v R-ju predstavimo z vektorji. **Vektor** $[v_1, v_2, \dots, v_k]$ vsebuje zaporedje podatkov iste osnovne zvrsti. Pravzaprav so v R-ju tudi spremenljivke, ki smo jih uporabljali doslej, vektorji z enim samim členom.

Posamične vrednosti združimo v vektor s funkcijo `c`:

```
> (x <- c(14,5,33,18,27))
```

```
[1] 14 5 33 18 27
```

```
> c("a",2+4i,3,FALSE)
```

```
[1] "a"      "2+4i"   "3"      "FALSE"
```

Velikost (število členov) vektorja izvemo s funkcijo `length`:

```
> length(x)
```

```
[1] 5
```



Ustvarjanje vektorjev

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Izraz $\text{zvrst}(n)$ nam da vektor zvrsti zvrst dolžine n zapolnjen z 'ničelno' vrednostjo:

```
> logical(5)
```

```
[1] FALSE FALSE FALSE FALSE FALSE
```

Vektorje lahko ustvarimo tudi z zaporednostno operacijo : . Izraz $n:m$ ustvari vektor celih števil s členi od n do m :

```
> 6:14
```

```
[1] 6 7 8 9 10 11 12 13 14
```

```
> 4:-3
```

```
[1] 4 3 2 1 0 -1 -2 -3
```

Ustvarita jih tudi funkciji:

$\text{seq}(n,m,k)$ – vektor z vrednostmi od n do m s korakom k .

$\text{rep}(v,k)$ – vektor s k s ponovitvami vektorja v .

```
> seq(0.75,1.4,0.1)
```

```
[1] 0.75 0.85 0.95 1.05 1.15 1.25 1.35
```

```
> rep(c(2,3),5)
```

```
[1] 2 3 2 3 2 3 2 3 2 3
```



Vektorske operacije

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

S funkcijo `c` lahko vektorje tudi združujemo/stikamo:

```
> (u <- c(5:3,7,c(8,4)))
```

```
[1] 5 4 3 7 8 4
```

V R-ju večina operacij in funkcij deluje na vektorjih po členih:

```
> a <- 1:5
```

```
> 2*a-1
```

```
[1] 1 3 5 7 9
```

```
> sqrt(a)
```

```
[1] 1.000000 1.414214 1.732051 2.000000 2.236068
```

```
> 2^a
```

```
[1] 2 4 8 16 32
```




Slučajni vektorji

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

S funkcijo `runif(n, min, max)` ustvarimo številski vektor dolžine n s slučajnimi vrednostmi enakomerno porazdeljenimi na intervalu (min, max) . Če je $min \neq max$, meji min in max nista možni vrednosti. Privzeti vrednosti sta $min = 0$ in $max = 1$.

```
> runif(10)
```

```
[1] 0.24898854 0.06774137 0.74274699 0.84019219 0.10156272  
[6] 0.27821238 0.86360775 0.39078224 0.80653827 0.14925524
```

Vrzimo 50 krat kocko:

```
> trunc(runif(50, 1, 7))
```

```
[1] 4 6 5 5 3 6 3 5 5 6 6 3 6 6 6 2 2 3 5 5 1 1 6 2 1 3 6 2 5 2  
[31] 4 4 3 2 1 3 4 5 6 6 3 6 6 3 5 3 4 5 4 5
```

R pozna tudi drugače porazdeljene slučajne vektorje - a o tem kasneje.



Krožno dopolnjevanje

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Kaj se zgodi, če uporabimo dvočleno operacijo na vektorjih različnih dolžin? Velja pravilo:

Najprej se krajši vektor krožno dopolni do dolžine daljšega. Nato se nad njima po členih opravi operacija.

```
> 1:6 + c(3,1)
```

```
[1] 4 3 6 5 8 7
```

```
> 1:6 + c(3,1,3,1,3,1)
```

```
[1] 4 3 6 5 8 7
```

```
> 1:10 %% 3
```

```
[1] 1 2 0 1 2 0 1 2 0 1
```

```
> 1:10 %% rep(3,10)
```

```
[1] 1 2 0 1 2 0 1 2 0 1
```



Indeksi

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Posamezni člen ali izbrane člene dobimo iz vektorja x z uporabo *indeksov* (zaporednih številč členov):

$x[i]$ – i -ti člen(i) vektorja x .

Če v oglate oklepaje postavimo vektor i , ki ima za vrednosti indekse $i = [i_1, i_2, \dots, i_k]$ dobimo iz $x[i]$ vektor s členi $[x[i_1], x[i_2], \dots, x[i_k]]$.

```
> x <-(1:5)^2
```

```
> x[c(5,2,5)]
```

```
[1] 25 4 25
```

Kadar je i en sam indeks, lahko to povemo z dvojnimi oklepaji $x[[i]]$ – i -ti (posamični) člen vektorja x .

Negativni indeksi pomenijo vse člene razen teh:

```
> x[c(-1,-3)]
```

```
[1] 4 16 25
```



Indeksi

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Zamenjamo lahko tudi vrednosti izbranih členov z zaporednimi vrednostmi v vektorju na desni:

```
> x[c(2,3)] <- c(7,5)
```

```
> x
```

```
[1] 1 7 5 16 25
```

Naj bo k logični vektor. Tedaj nam $x[k]$ vrne podvektor vrednosti vektorja x , za katere ima istoležni člen v vektorju k vrednost resnično. Če je vektor k krajši od vektorja x , se najprej krožno dopolni.

```
> x <- c(14,5,33,18,27)
```

```
> x < 20
```

```
[1] TRUE TRUE FALSE TRUE FALSE
```

```
> x[x < 20]
```

```
[1] 14 5 18
```

```
> (1:10)[c(TRUE,FALSE)]
```

```
[1] 1 3 5 7 9
```



Stavek for

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Večino stvari z vektorji lahko opravimo z vektorskimi operacijami. Včasih pa pride prav stavek

```
for( v in V) izraz
```

ki zahteva, se *izraz* izračuna za vsako vrednost v iz vektorja V .

```
> x <- c("l","j","u","b","l","j","a","n","a")
> for(z in x) cat(z); cat("\n")
ljubljana
```

Pozor!

```
> kocka <- function(p=6) trunc(1+p*runif(1))
> system.time({s <- NULL; for(i in 1:100000) s <- c(s,kocka())})
  user  system elapsed
38.95   0.09   40.90
> system.time({
  s <- integer(100000); for(i in 1:100000) s[i] <- kocka()})
  user  system elapsed
 1.91   0.00   1.75
> system.time(s <- trunc(1+6*runif(100000)))
  user  system elapsed
 0.08   0.00   0.09
```



Vektorji z imeni členov

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

```
> starost <- c(24,58,3,27)
> names(starost) <-
+ c("Micka","Franca","Franček","Janez")
> starost
  Micka Franca Franček  Janez
    24    58      3     27
> names(starost)
[1] "Micka" "Franca" "Franček" "Janez"
> starost["Franček"]
Franček
      3
> starost[c("Janez","Micka")]
Janez Micka
    27    24
> for (i in names(starost)) cat(i,starost[i],"\n")
Micka 24
Franca 58
Franček 3
Janez 27
```



Operacije z množicami

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Množice so predstavljene z vektorji, ki naj bi bili sestavljeni iz različnih členov.

Nad njimi lahko opravimo običajne operacije z množicami: *unijo* `union(x, y)`, *preseki* `intersect(x, y)` in *razliko* `setdiff(x, y)` ter primerjavo *enakosti* `setequal(x, y)`. Preverimo lahko tudi ali podatek *e* *pripada* množici `x` `is.element(e,x)` ali `e %in% x`.

```
> charSet <- function(z)
+ union(substring(tolower(z),1:nchar(z),1:nchar(z)),NULL)
> (A <- charSet("Ljubljana"))
[1] "l" "j" "u" "b" "a" "n"
> (B <- charSet("Maribor"))
[1] "m" "a" "r" "i" "b" "o"
> union(A,B)
[1] "l" "j" "u" "b" "a" "n" "m" "r" "i" "o"
> intersect(A,B)
[1] "b" "a"
> setdiff(A,B)
[1] "l" "j" "u" "n"
> is.element("a",A)
[1] TRUE
```



Operacije z množicami

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Iz teorije množic vemo $A \subseteq B \Leftrightarrow A \cup B = B \Leftrightarrow A \cap B = A$.

```
> set <- function(x) union(x,NULL)
> card <- function(x) length(set(x))
> is.set <- function(x) length(x)==card(x)
> subseteq <- function(x,y){setequal(intersect(x,y),x)}
> (c <- c("m","a","m","a"))
[1] "m" "a" "m" "a"
> (C <- set(c))
[1] "m" "a"
> is.set(c)
[1] FALSE
> is.set(C)
[1] TRUE
> length(c)
[1] 4
> card(c)
[1] 2
> subseteq(C,B)
[1] TRUE
> "a" %in% C
[1] TRUE
```




Preurejanja

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

sort, rank, order, rev, which, which.min, which.max

```
> V <- c(14,6,19,24,7,12,15,3,14,7,5)
> sort(V)
[1] 3 5 6 7 7 12 14 14 15 19 24
> rank(V)
[1] 7.5 3.0 10.0 11.0 4.5 6.0 9.0 1.0 7.5 4.5 2.0
> (s <- rank(V,ties.method="first"))
[1] 7 3 10 11 4 6 9 1 8 5 2
> U <- numeric(length(V))
> for(i in 1:length(V)) U[s[i]] <- V[i]
> U
[1] 3 5 6 7 7 12 14 14 15 19 24
> (t <- order(V))
[1] 8 11 2 5 10 6 1 9 7 3 4
> (T <- V[t])
[1] 3 5 6 7 7 12 14 14 15 19 24
> rev(T)
[1] 24 19 15 14 14 12 7 7 6 5 3
> which(V<10)
[1] 2 5 8 10 11
> which.max(V)
[1] 4
```



Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Iskanje najboljšega racionalnega približka za π sedaj lahko rešimo v nekaj vrsticah:

```
> q <- 1:999
> p <- round(pi*q)
> i <- which.min(abs(p/q-pi))
> c(p[i],q[i])
[1] 355 113
> p[i]/q[i]
[1] 3.141593
```



Imenske spremenljivke v R-ju

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

```
> R <- c("LJ", "KR", "LJ", "MB", "KP", "LJ", "LJ", "KR", "CE", "GO", "NM", "LJ")
> R <- factor(R)
> R
 [1] LJ KR LJ MB KP LJ LJ KR CE GO NM LJ
Levels: CE GO KP KR LJ MB NM
> as.integer(R)
 [1] 5 4 5 6 3 5 5 4 1 2 7 5
> levels(R)
 [1] "CE" "GO" "KP" "KR" "LJ" "MB" "NM"
> levels(R)[as.integer(R)]
 [1] "LJ" "KR" "LJ" "MB" "KP" "LJ" "LJ" "KR" "CE" "GO" "NM" "LJ"
> which(R=="LJ")
 [1] 1 3 6 7 12
```



Urejenostne spremenljivke v R-ju

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

```
> k <- c("nz", "pd", "pd", "db")
> L <- c("nz", "zd", "db", "pd", "od")
> k <- factor(k, levels=L, ordered=TRUE)
> k
[1] nz pd pd db
Levels: nz < zd < db < pd < od
> as.numeric(k)
[1] 1 4 4 3
```



Posebne funkcije nad vektorji

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

$\max(v)$	$= \max_i v_i$
$\min(v)$	$= \min_i v_i$
$\text{prod}(v)$	$= \prod_i v_i$
$\text{sum}(v)$	$= \sum_i v_i$
$\text{any}(t)$	$= \bigvee_i t_i$
$\text{all}(t)$	$= \bigwedge_i t_i$
$\text{range}(v)$	$= [\min_i v_i, \max_i v_i]$
$\text{mean}(v)$	$= \frac{1}{n} \sum_i v_i$
$\text{median}(v)$	srednja po velikosti izmed vrednosti v_i
$\text{quantile}(v, p)$	kvantili
$\text{var}(v)$	varianca
$\text{sd}(v)$	standardni odklon
$\text{cor}(v, u)$	korelacijski koeficient



Posebne funkcije nad vektorji

Vektorji

V. Batagelj

Sestavljeni podatki

Merske lestvice

Vektorji

Vektorske operacije

Indeksi

Stavek for

Množice

Preurejanja

Imenske in urejenostne spremenljivke

Posebne funkcije

Funkcije `cummax(v)`, `cummin(v)`, `cumprod(v)` in `cumsum(v)` ustvarijo vektorje delnih rezultatov ustreznih operacij.

```
> (x <- runif(10,0,10))
[1] 5.1223063 0.9172606 0.9664358 7.2381382 2.0541289 8.7877618
[7] 8.6652347 6.1872376 4.5662272 3.4604274
> (x <- x/sum(x))
[1] 0.10679223 0.01912348 0.02014870 0.15090408 0.04282544
[6] 0.18321136 0.18065686 0.12899442 0.09519884 0.07214460
> cumsum(x)
[1] 0.1067922 0.1259157 0.1460644 0.2969685 0.3397939 0.5230053
[7] 0.7036621 0.8326566 0.9278554 1.0000000
```

Če vektor vsebuje vrednost NA ali NaN, bo ta tudi vrednost operacije. Če dodamo klicu funkcije argument `na.rm=TRUE`, se vrednosti NA in NaN pri izračunu ne upoštevajo.

Glejte še funkciji `pmin` in `pmax`.

```
> exp(2)
[1] 7.389056
> x <- 2; (ex <- 1+sum(cumprod(1/(1:20))*x^(1:20)))
[1] 7.389056
```