Az R programcsomag

Az R statisztikai programcsomag letölthető az alábbi címről:

http://www.r-project.org

Az RStudio letölthető az alábbi címről:

https://www.rstudio.com/products/rstudio/download/

Alapműveletek, elemi függvények R-ben

- `+,-,*,/,^`
- gyökvonás: `sqrt(4)`
- exponenciális függvény: `exp(2)`
- *e*-alapú logaritmus: `log(3)`
- 10-es alapú logaritmus: `log10(100)`
- *a*-alapú logaritmus: `log(x,a)`
- trigonometrikus függvények: `sin, cos, tan`, pl.: `sin(pi/2)`
- %% és %/%: maradékos osztás, pl. 7%%3, 7%/%3

Értékadás

- `a<-4+5`
- `B2=sqrt(a)`
- `a-B2->mennyiez`

A változónevek:

- nem kezdődhetnek számmal
- nem tartalmazhatnak szóközt
- case sensitive

Az R objektumai

- vektorok
- tömbök (mátrixok)
- függvények
- faktorok
- listák
- adatkeretek

Vektorok

A vektorok minden eleme azonos típusú: numeric (integer vagy double), character, logical, stb.
Típus lekérdezése: `typeof(v)`

Vektorok megadása:

- `v<-c(0,1,-2.4,8)`
- `vektor=1:30` (vagy fordítva: `w=21:17`)

Vektorok kiegészítése, összefűzése:

- `X=c(v,3.15)`
- `y2=c(X,w)`

`z<-numeric()` : egy üres numerikus vektor

Vektorokhoz kapcsolódó parancsok

Legyen most `v<-y2`.

- `length(v)`: a vektor hossza
- `mean(v)`: a vektor elemeinek átlaga
- `min(v)` és `max(v)`: a vektor legkisebb és legnagyobb eleme
- `sort(v)`: a v vektor elemeit sorbarendezi
- `rev(v)`: a v vektor elemeit fordított sorrendben felsorolja
- `sample(v)`: a vektor elemeit véletlenszerű sorrendben sorolja fel
- `sum(v)`, `prod(v)`, stb.

Teszteljük, hogy mit csinál a `help(sort)` parancs!

Vektorokhoz kapcsolódó parancsok

A `rep` parancs:

- Legyen `v<-1:3`.
- `w=rep(v,times=4)`
- `z=rep(v,each=4)`

A `seq` parancs:

- `vektor=seq(from=2,to=7,by=0.5)`
- `t<-seq(from=-1,by=0.1,length=7)`
- `y<-seq(from=3,along=t)`

Műveletek vektorokkal

Az R a vektorokkal koordinátánként végzi a műveleteket!

- `v=c(1,0,0,1,5)`
- `w=3:5`
- `2*v+w`

Itt v és w nem ugyanolyan hosszúak. Az R a rövidebb vektort ciklikusan ismétli, amíg olyan hosszú nem lesz, mint a másik vektor.

Az elemi függvények is elemenként hajtódnak végre: `exp(0:3)`

Mivel egyenlő `2*1:5`?

Hivatkozás a vektorok elemeire

```
v=1:17
```

- `v[5]`
- `v[2:5]`
- `v[-1]`
- `v[-c(4,9,12)]`

Logikai vektorok

Logikai operátorok:

- `<,<=,>,>=,==,!=`
- `&` (and), | (or), `!` (not), `xor(,)`

Logikai vektorok: elemei a TRUE, a FALSE és az NA (not available)
Pl.: `a=c(-1,3,2,0)`, `B=a>1`

Logikai függvények:

- `which`: egy logikai vektor TRUE értékeinek indexeit adja
- `any`: értéke TRUE, ha egy logikai vektor valamelyik eleme TRUE
- `all`: értéke TRUE, ha egy logikai vektor mindegyik eleme TRUE

Feladatok

- Az elemek egyenkénti begépelése nélkül állítsuk elő az alábbi vektorokat!
  (1) $a = (0, 1, \ldots, 30)$
  (2) $b = (2, 4, 6, \ldots, 100)$
  (3) $c = (2, 1.9, 1.8, \ldots, 0)$
  (4) $d = (0, 3, 6, \ldots, 27, 30, -100, 30, 27, \ldots, 6, 3, 0)$
  (5) $e = \left(\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{20}\right)$
  (6) $f = \left(\frac{1}{2}, \frac{2}{3}, \ldots, \frac{19}{20}\right)$
- Legyen v egy adott 100 elemű vektor. Állítsuk elő azt a w vektort, melynek elemei
  (1) a v elemei fordított sorrendben felsorolva;
  (2) a v elemei, kivéve a 7., a 20. és a 89. elemét;
  (3) a v vektor páratlan sorszámú elemei.

Feladatok

Legyen v egy adott 20 elemű vektor. Például: `v=sample(20)`.
Állítsuk elő azt a vektort, mely

- a v vektor 3 legnagyobb elemét tartalmazza;
- a v vektor minden 4. elemét tartalmazza;
- a v vektor 15-nél nagyobb elemeit tartalmazza;
- a v vektornak a [3,10] intervallumon kívül eső elemeit tartalmazza.

Mátrixok

Ha v vektor, akkor

- `A=matrix(v,nrow=m,ncol=n)`

$m \times n$-es mátrix, a v vektorból oszlopfolytonosan elkészítve (opcionális argumentum: `byrow=F`)

Hivatkozás mátrixok elemeire:

- `A[i,j]`: a mátrix (i, j)-edik eleme
- `A[i,]`: a mátrix i-edik sora
- `A[,j]`: a mátrix j-edik oszlopa
- `A[1:3,c(2,4)]`: a mátrix 1., 2., 3. sorának és 2., 4. oszlopának elemei
- `A[-2,]`: az a mátrix, amely A-ból a 2. sor elhagyásával keletkezik

Mátrixműveletek

- `cbind(A,B)`: horizontálisan összefűz 2 (ugyanannyi sorral rendelkező) mátrixot
- `rbind(A,B)`: vertikálisan összefűz 2 (ugyanannyi oszloppal rendelkező) mátrixot
- alapműveletek: elemenként (azonos méretű mátrixokra)
- `%*%`: mátrixszorzás
- ha v vektor, A mátrix: `A*v` és `v*A` eredménye ugyanaz: az elemenkénti szorzata 2 azonos méretű mátrixnak, mégpedig A-nak, és a v oszlopfolytonos ismételgetésével kapott mátrixnak

Mátrixokhoz kapcsolódó parancsok

Ha A mátrix, akkor

- `t(A)`: az A transzponáltja
- `dim(A)`: az A mérete
- `nrow(A)`: A sorainak a száma
- `ncol(A)`: A oszlopainak a száma
- `det(A)`: A determinánsa
- `solve(A)`: A inverze
- `solve(A,b)`: az $Ax = b$ egyenletrendszer megoldása
- `colSums(A)`: az A oszlopaiban álló elemek összegei
- `rowSums(A)`: az A soraiban álló elemek összegei
- `sum(A)`: A elemeinek az összege
- `diag(A)`: A főátlójának a vektora

A `diag()` függvény egyéb használata

Példák:

- `diag(3)`: a 3×3-as egységmátrix
- `diag(2,7,3)`: egy 7×3-as diagonális mátrix, főátlójában 2-esek
- `diag(1:3,4,5)`: egy 4×5-ös diagonális mátrix, főátlójában ciklikusan az $(1, 2, 3)$ vektor
- `diag(A)<-100`: az A mátrix főátlójának minden elemét 100-ra cseréli
- `diag(A)<-v`: az A mátrix főátlójába beírja a v vektort (ha az megfelelő méretű)

Tömbök

Tömb megadása:

- `B<-array(1:12,dim=c(3,2,2))`

Ezek után a `B[3,1,2]` módon hivatkozhatunk B elemeire, `B[,i,j]` módon B soraira, stb.

Vizsgáljuk meg a beépített `Titanic` tömböt!

Az `apply` függvény végrehajt egy megadott függvényt egy mátrix, vagy tömb (esetleg több) megadott dimenziója mentén: `apply(A,v,fv)`, ahol

- A egy tömb
- v egy szám/vektor
- fv pedig a végrehajtandó függvény

Példa: `A=matrix(1:12,nrow=3)`

- `s=apply(A,1,sum)`: sorösszegek
- `s=apply(A,2,mean)`: oszlopátlagok

Feladatok

- Állítsuk elő a csupa 1-esekből álló 3×5-ös mátrixot!
- Cseréljük ki az előző mátrix főátlójának minden elemét -1-re!
- Állítsuk elő azt a 3×4-es mátrixot, amely sorfolytonosan tartalmazza az $1, 2, \ldots, 12$ számok 3-szorosait!
- Ha az előző mátrix A, mivel lesz egyenlő `rbind(A,1)`?
- Keressük meg egy adott 5×6-os mátrix minden sorában a maximális elemet!
- Egy adott A mátrix esetén konstruáljuk meg azt a B mátrixot, amelyet úgy kapunk, hogy A sorainak a végére odaírjuk az adott sorban álló elemek átlagát!
- Egy adott A 3×4-es mátrix esetén konstruáljuk meg azt a B mátrixot, amelyet úgy kapunk, hogy A második sorát kicseréljük a $(-1, 5, 0, 100)$ vektorra!
- Tekintsük a beépített `Titanic` tömböt! Az `apply` parancs segítségével döntsük el, hogy a túlélők között a nők, vagy a férfiak voltak-e többen! Milyen osztályon utazott a legtöbb túlélő?

Elágazások, ciklusok

- if elágazás
  `if(logikai kifejezés) utasítás else utasítás`
- ifelse függvény
  `ifelse(v,t,f)`: ha a *v* logikai vektor igaz, akkor t-vel, egyébként f-fel tér vissza
  Példa: `ifelse(x>0,x,-x)` eredménye az `abs(x)` vektor
- for ciklus
  `for (változó in vektor) utasítás`
  Példa: `s=0`, `for (i in 1:10) s=s+i`

Függvények

Függvény definiálása:
`fv<-function(arg1,arg2,...)utasítások`

Példa: `negyzet<-function(x) x ^ 2`
Ezek után használhatjuk a `negyzet(-5)` parancsot.

Feladatok

- Írjunk parancsot a "pozitív rész" függvényre!
- Írjunk függvényt, amely adott természetes szám esetén visszaadja $n!$ értékét!
- Írjunk egy olyan függvényt, mely kiszámolja egy vektor elemeinek az átlagát úgy, hogy a legnagyobb és legkisebb elemet nem veszi figyelembe!
- Írjunk olyan függvényt, mely egy numerikus A mátrix megadása esetén visszaadja azt a mátrixot, amely A sorait fordított sorrendben tartalmazza! (Azaz az első sora A utolsó sora, a második sora A utolsó előtti sora, stb.)

Grafikus parancsok

Magas szintű függvény: a parancs új ábrát készít

- `plot()`

Alacsony szintű függvények: meglévő ábrát módosít

- `lines()`
- `points()`
- `title()`
- `abline()`
- `legend()`

Példa:

```
plot(sin,-pi,2*pi)
abline(h=1,col="blue",lty=3)
```

Vagy:

- `plot(...,add=TRUE)` – előző plotra teszi az újat

Grafikus parancsok – példa

- `x=1:10`
- `y=(x-4)^ 2`
- `plot(x,y,pch=2,col="darkgreen",xlim=c(-1,10))`
- `abline(2,3,col="red",lty=4)`
- `title("Két függvény",`
  `xlab="függetlenváltozó",ylab="függőváltozó")`
- `legend("topleft",c("másodfokú","egyenes"),`
  `lty=c(0,4), pch=c(2,0),col=c("darkgreen","red"))`

Plot types:

- `type="o"` – pontok és vonal
- `type="p"` – csak pontok
- `type="l"` – csak vonal
- `type="h"` – hisztogram-szerű függőleges vonalak
- `type="n"` – nem csinál ábrát, "üres vászon"

Nevezetes eloszlások

Az adott eloszlás neve előtt:

- `d`: sűrűségfüggvény, vagy felvételi valószínűség
- `p`: eloszlásfüggvény
- `q`: kvantilisek
- `r`: véletlen számok

Példák nevezetes eloszlások használatára

- binomiális eloszlás:
  `dbinom(3,size=12,prob=0.2)`
- Poisson eloszlás:
  `ppois(16,lambda=12)`
- egyenletes eloszlás:
  `runif(10,min=1,max=3)`
- exponenciális eloszlás:
  `pexp(2,rate=1/3)`
- normális eloszlás:
  `dnorm(84,mean=72,sd=15)`
- χ^2-eloszlás:
  `qchisq(.95,df=7)`
- Student (t)-eloszlás:
  `qt(.95,df=5)`
- F-eloszlás:
  `qf(.95,df1=2,df2=3)`

Feladatok

- Ábrázoljuk a 10-edrendű, 0.3 paraméterű binomiális eloszlás felvételi valószínűségeit egy grafikonon. Legyen `type="h"`.
- Ábrázoljuk a 3 szabadságfokú χ^2-eloszlás sűrűségfüggvényét.
- Ábrázoljuk egy koordináta-rendszerben az $1, 2, 3, 6$ és 10 szabadsági fokú t-eloszlások sűrűségfüggvényét különböző színekkel, valamint a standard normális sűrűségfüggvényt!

Statisztika – ismérvek mérési skálái

A különböző ismérveknek a következő mérési skálái vannak:

- nominális (minőségi vagy mennyiségi)
- ordinális (mennyiségi)
- különbségi (mennyiségi)
- arány (mennyiségi)

Leíró statisztika kvalitatív adatokra

Kvalitatív adatok: nem számszerű adatok.

Próbáljuk ki az R `painters` listájának utolsó oszlopán az alábbi parancsokat:

- `library(MASS)`
- `help(painters)`, `typeof(painters)`, `dim(painters)`
- `minta=painters$School`
- `minta`: faktor, 8 szinttel, `levels(minta)`
- `table(minta)`, `cbind(table(minta))`
- `barplot(table(minta))`, `pie(table(minta))`
- Hogy kapjuk meg a mintaelemszámot?
- `rel.minta=table(minta)/length(minta)`
- Ugyanez, átláthatóbb formában:

```
eredeti=options(digits=1)
rel.minta
options(eredeti)
```

Iskolánként alkalmazott függvények

- `minta=painters$School`
- `D_iskola=minta=="D"`
  Mi a `D_iskola` vektor?
- `D_painters=painters[D_iskola,]`

Mi az átlagos `Drawing` pontszám a különböző iskolákban?

- `tapply(painters$Drawing,minta,mean)`

A `tapply` függvény működése:

`tapply(vektor,faktor,fuggveny)`

a `vektor` elemeit csoportosítja a `faktor`nak megfelelően, és a csoportokra alkalmazza a `fuggveny`t.

Leíró statisztika numerikus adatokra

Használjuk az R beépített `faithful` adathalmazát.

- Dataset: faithful
- `minta=faithful$eruptions`

Próbáljuk ki a `minta` vektoron a leíró statisztika alábbi elemeit:

- `mean, median, range, quantile, var, sd`
  (sd és var a korrigált tapasztalati szórás, ill. szórásnégyzet)
- `library(e1071)`
  `moment(minta, order=3, center=TRUE)`

A minta grafikus vizsgálata:

- `boxplot(minta,horizontal=TRUE)`
- `boxplot(minta,plot=FALSE)`
- `hist(minta)  hist(minta,20)`
  további argumentumok: `right`, `main, xlab,ylab`,
  `col=c("magenta","purple","violet")`,
  `col=heat.colors(20,0.7)`

Empirikus eloszlásfüggvény

- `minta=faithful$eruptions`
- `F=ecdf(minta)`
- `plot(F)`

Feladatok

- Generáljunk 5, 10, 100 elemű mintát 10 várható értékű, 4 szórásnégyzetű normális eloszlásból. Ábrázoljuk az elméleti eloszlásfüggvényt, valamint mindhárom minta empirikus eloszlásfüggvényét különböző színekkel.
- Demonstráljuk a Glivenko–Cantelli lemmát: generáljunk 50-szer 100 elemű mintát standard normális eloszlásból. Ábrázoljuk az elméleti eloszlásfüggvényt a [-3,3] intervallumon, valamint az 50 minta empirikus eloszlásfüggvényét citromsárgával.

Egyéb leíró statisztikai elemek

- `minta=faithful$eruptions`

További alakmutatók:

- `library(e1071)`
- `skewness(minta)`: ferdeség
- `kurtosis(minta)`: csúcsosság

Több ismérv közötti kapcsolat:

- `idotartam=faithful$eruptions`
- `varakozas=faithful$waiting`
- `plot(idotartam,varakozas)`
- `cov(idotartam,varakozas)`
- `cor(idotartam,varakozas)`

Feladatok

- Generáljunk 100 elemű mintát 3 paraméterű exponenciális eloszlásból. Számoljuk ki a minta ferdeségét, csúcsosságát. Ábrázoljuk a minta (gyakorisági) hisztogramját.
- Ábrázoljuk a sűrűséghisztogramot, azaz amely esetén a téglalapok összterülete 1-gyel egyenlő. Tegyük rá az ábrára a mintaátlaggal azonos várható értékű és tapasztalati varianciával azonos szórásnégyzetű normális eloszlás sűrűségfüggvényét.
- Töltsük le a `finance.yahoo.com` oldalról az Apple (AAPL) részvények elmúlt 1 évre vonatkozó napi adatait (nyitó, záró, ...). Importáljuk az adathalmazt R-be.