

Пакет **sampling** у R-у

Токаи Мелинда

writesample(n,N)

- Ова команда даје матрицу чије су врсте вектори (са елементима 0 или 1) свих узорака фиксног обима, узорак без понављања
- n обим узорка
- N обим популације

```
> M=writesample(2,6)
> t(apply(M,1,function(x) (1:ncol(M))[x==1]))
```

```
  [,1] [,2]
[1,]  5  6
[2,]  4  6
[3,]  4  5
[4,]  3  6
[5,]  3  5
[6,]  3  4
[7,]  2  6
[8,]  2  5
[9,]  2  4
[10,] 2  3
[11,] 1  6
[12,] 1  5
[13,] 1  4
[14,] 1  3
[15,] 1  2
> |
```

```
> M
  [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  0  0  0  0  1  1
[2,]  0  0  0  1  0  1
[3,]  0  0  0  1  1  0
[4,]  0  0  1  0  0  1
[5,]  0  0  1  0  1  0
[6,]  0  0  1  1  0  0
[7,]  0  1  0  0  0  1
[8,]  0  1  0  0  1  0
[9,]  0  1  0  1  0  0
[10,] 0  1  1  0  0  0
[11,] 1  0  0  0  0  1
[12,] 1  0  0  0  1  0
[13,] 1  0  0  1  0  0
[14,] 1  0  1  0  0  0
[15,] 1  1  0  0  0  0
> |
```

srswor(n,N)

- Ова комада даје SRSWOR узорак (узорак фиксне дужине, једнаких вероватноћа, без понављања).
- Враћа вектор дужине N са елементима 0 односно 1 у зависности да ли је k -та јединица популације изабрана у узорак (ако је на k -тој позицији 1 јединица је изабрана, односно није изабран ако је 0)

```
> u=srswor(3,8)
> u
[1] 0 0 1 1 0 0 0 1
> uzorak=(1:8)[u==1]
> uzorak
[1] 3 4 8
> |
```

```
>
>
> ##baza koja sadrzi podatke o populaciji u Belgiji
> #podaci iz jula 2004. god
> #podaci iz jula 2003. god
> #vrsi se poredjenje medju njima
> #sadrzi i neke podatke o finansijama gradova na kraju 2001.
> data(belgianmunicipalities)
>
> total=belgianmunicipalities$Tot04
> #Tot04 broj stanovnika po gradovima jula2004.
>
> #Commune imena gradova
> ime=belgianmunicipalities$Commune
>
> n=200 #obim uzorka
> uz=srswor(n, length(total)) #izaberemo uzorak
>
> #sa komandom as.vector(ime[uz==1])
> #dobijamo vektor sa imenima gradova koji su u uzorku
>
```

```
> # za "laksi" prikaz uzeemo n=10
> n=10
> uz=srswor(n, length(total))
> as.vector(ime[uz==1])
[1] "Hove" "Oud-Turnhout" "Lennik" "Rotselaar" "Zwevegem"
[6] "Moorslede" "Wichelen" "Heron" "Gedinne" "Fernelmont"
>
```

srswor1(n,N)

- Ради скоро исто што и претходна ф-ја али ова користи selection-rejection методу

- Selection-rejection procedure for SRS/SRSWOR

Definition k, j : Integer;

$j = 0$

For $k = 1, 2, \dots, N$ do

са вероватноћом $\frac{n-j}{N-(k-1)}$ одабери јединицу k ; $j = j + 1$

```
>  
> s=srswor1(3,10)  
> #uzorak je  
> (1:10)[s==1]  
[1] 1 2 5  
>
```

srswr(n,N)

- Ова комада даје SRSWR узорак (узорак фиксне дужине, једнаких вероватноћа, са понављањем). Враћа вектор дужине N , при чему елементи овог вектора представљају број понављања јединице k у узорку.

```
> s=srswr(6,15)
> #izabrane jedinice su
> (1:15)[s!=0]
[1] 5 8 10 13
> #sa brojem ponavljanja
> s[s!=0]
[1] 2 1 2 1
>
```

getdata(data, m)

”Извлачи” сам узорак из базе. Ова ф-ја се користи након што добијемо узорак тј. вектор који нам представља узорак.

```
> getdata(belgianmunicipalities, uz)
```

	ID_unit	Commune	INS	Province	Arrondiss	Men04	Women04	Tot04	Men03	Women03	Tot03	Diffmen	
11	11	Hove	11021		1	11	3979	4170	8149	3987	4183	8170	-8
62	62	oud-Turnhout	13031		1	13	6285	6245	12530	6227	6182	12409	58
123	123	Lennik	23104		2	23	4319	4413	8732	4328	4408	8736	-9
146	146	Rotseelaar	24094		2	24	7423	7517	14940	7366	7430	14796	57
215	215	Zwevegem	34042		3	34	11648	11933	23581	11583	11822	23405	65
229	229	Moorslede	36012		3	36	5279	5356	10635	5278	5344	10622	1
264	264	wichelen	42026		4	42	5405	5676	11081	5365	5659	11024	40
385	385	Heron	61028		6	61	2156	2229	4385	2122	2221	4343	34
557	557	Gedinne	91054		9	91	2108	2282	4390	2078	2252	4330	30
579	579	Fernelmont	92138		9	92	3269	3383	6652	3240	3355	6595	29
	Diffwom	DiffTOT	TaxableIncome	Totaltaxation	averageincome	medianincome							
11	-13	-21	132669412	43230689		35549		23678					
62	63	121	165508600	44215801		29093		21752					
123	5	-4	128811484	39468852		31069		22594					
146	87	144	208830522	61405448		29388		21202					
215	111	176	286243776	74587242		25013		19515					
229	12	13	110035596	26842210		22895		18912					
264	17	57	137211259	37095577		25860		20825					
385	8	42	47186894	12265221		24666		19846					
557	30	60	39888308	8725768		21138		17510					
579	28	57	73030853	20480526		27200		20986					

```
>
```


UPmultinomial(pik)

- Користи *Hansen-Hurwitz*-ову методу за избор узорка (узорак фиксне дужине, неједнаких вероватноћа).
- Враћа вектор дужине N са елементима 0 односно 1 у зависности да ли је k -та јединица популације изабрана у узорак (ако је на k -тој позицији 1 јединица је изабрана, односно није изабран ако је 0)

```
>  
> x=c(0.2,0.7,0.8,0.5,0.4,0.4)  
> #uzimamo uзорak  
> uz=UPmultinomial(x)  
> #uzorak je  
> (1:length(x))[uz==1]  
[1] 3 5 6  
>
```

UPpoisson(pik)

- Извлачи Пуасонов узорак користећи вектор вероватноћа укључивања првог реда (узорак произвољне дужине, неједнаких вероватноћа, без понављања).
- Враћа вектор дужине N са елементима 0 односно 1 у зависности да ли је k -та јединица популације изабрана у узорак (ако је на k -тој позицији 1 јединица је изабрана, односно није изабран ако је 0)

```
>
> # definisemo vektor verovatnoca ukljucivanja
> x=c(1/3,1/3,1/3)
> # izaberomo uzorak
> s=UPpoisson(x)
> #uzorak je
> (1:length(x))[s==1]
[1] 2 3
> # izaberemo drugi uzorak
> s=UPpoisson(x)
> #uzorak je
> (1:length(x))[s==1]
integer(0)
> #dobili smo uzorak obima 0
```

```

>
> total=belgianmunicipalities$Tot04
> ime=belgianmunicipalities$Commune
> n=20
> x=inclusionprobabilities(total,n)
> # izaberemo uzorak
> s=UPpoisson(x)
> #uzorak je
> getdata(ime,s)

```

	ID_unit	data
1	2	Anvers
2	16	Mortsel
3	42	sint-katelijne-waver
4	76	Evere
5	77	Forest
6	91	Beersel
7	93	Dilbeek
8	94	Gammerages
9	113	Vilvorde
10	124	Affligem
11	143	Louvain
12	271	Zelzate
13	320	Charleroi
14	375	Rumes
15	376	Tournai
16	402	Chaudfontaine
17	409	Liege
18	471	Herck-la-Ville

```
> |
```

- Пуасонов узорак – на сваку јединицу популације се примењује Бернулијев експеримент (независно) на основу чега ће се одредити да ли та јединица улази у узорак или не

inclusionprobabilities(a,n)

- Рачуна вероватноће укључивања прве врсте користећи вектор позитивних бројева (за план узорковања са вероватноћама пропорционалним величини)

```
>
> # vector pozitivnih brojeva
> a=1:20
> # racunanje verovatnoca ukljucivanja
> # za uzorak obima n=12
> x=inclusionprobabilities(a,12)
> x
 [1] 0.05882353 0.11764706 0.17647059 0.23529412
0.29411765 0.35294118 0.41176471 0.47058824
 [9] 0.52941176 0.58823529 0.64705882 0.70588235
0.76470588 0.82352941 0.88235294 0.94117647
[17] 1.00000000 1.00000000 1.00000000 1.00000000
 .. . . . .
```

inclusionprobastrata(strata,nh)

- Рачуна вероватноће укључивања за стратификовани план узорковања
- *strata* вектор који дефинише стратуме
- *nh* вектор са бројевима јединица колико желимо да узмемо у сваком од стратума

```
> # stratum
> strata=c(1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3)
> # obim uzorka u svakom od stratuma
> nh=c(2,3,3)
> inclusionprobastrata(strata,nh)
 [1] 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.5000000 0.5000000
 [8] 0.5000000 0.5000000 0.5000000 0.5000000 0.4285714 0.4285714 0.4285714
[15] 0.4285714 0.4285714 0.4285714 0.4285714
>
```

strata(data, stratanames, size, method, pik, description=F)

- Овом ф-јом извлачимо стратификован узорак
- *stratanames* је облежје у бази по којој се врши стратификација (има неколико категорија)
- *size* вектор који представља обим стратума који се извлачи из сваке од категорија
- *method* може бити: "*srswor*", "*srswr*", "*poisson*" или "*systematic*"
- *pik* вектор вероватноћа укључивања или помоћних информација, овај аргумент се користи код узорака неједнаких вероватноћа ("*poisson*" или "*systematic*")

```
> # Koristimo bazu 'iris' kao populaciju za uzimanje uzorka
> data(iris)
> # oblezje 'species' ima 3 kategorija
> # ovo cemo koristiti za stratifikaciju
> # racuna stratume populacije
> table(iris$species)

      setosa versicolor  virginica
      50         50         50
>
> # velicine stratuma uzorka su date sa size=c(6,4,3)
> # koristimo metodu SRSWOR (jednake verovatnoce, bez ponavljanja)
> st=strata(iris,stratanames=c("Species"),
+          size=c(6,4,3),
+          method="srswor")
> # izvlaci sam uzorak iz baze
> getdata(iris, st)
  sepal.Length sepal.width Petal.Length Petal.width  Species ID_unit Prob Stratum
11           5.4         3.7         1.5         0.2   setosa    11 0.12      1
19           5.7         3.8         1.7         0.3   setosa    19 0.12      1
23           4.6         3.6         1.0         0.2   setosa    23 0.12      1
26           5.0         3.0         1.6         0.2   setosa    26 0.12      1
33           5.2         4.1         1.5         0.1   setosa    33 0.12      1
42           4.5         2.3         1.3         0.3   setosa    42 0.12      1
53           6.9         3.1         4.9         1.5 versicolor  53 0.08      2
59           6.6         2.9         4.6         1.3 versicolor  59 0.08      2
71           5.9         3.2         4.8         1.8 versicolor  71 0.08      2
79           6.0         2.9         4.5         1.5 versicolor  79 0.08      2
130          7.2         3.0         5.8         1.6  virginica  130 0.06      3
133          6.4         2.8         5.6         2.2  virginica  133 0.06      3
142          6.9         3.1         5.1         2.3  virginica  142 0.06      3
>
> # prikazivanje rezultat pomocu tabele kontigencije
> table(st$species)

      setosa versicolor  virginica
      6         4         3
>
```

UPsystematic(pik,eps=1e-6)

- Узима узорак користећи систематски метод (као што је узорак са вероватноћом пропорционалном величини)
- Враћа вектор дужине N са елементима 0 односно 1 у зависности да ли је k -та јединица популације изабрана у узорак (ако је на k -тој позицији 1 јединица је изабрана, односно није изабран ако је 0)
- pik вектор вероватноћа укључивања
- eps контролна вредност, по дефолту $1e-6$

```
>  
> #definiseo verovatnoce ukljucivanja  
> pik=c(0.2,0.7,0.8,0.5,0.4,0.4)  
> #izaberemo uzorak  
> s=UPsystematic(pik)  
> #uzorak je  
> (1:length(pik))[s==1]  
[1] 2 3 6  
>
```


UPsystematicpi2(pik)

- Рачуна заједничку вероватноћу укључивања (друге врсте) за систематски узорак
- Враћа матрицу $N \times N$ при чему се на главној дијагонали налазе вероватноће укључивања прве врсте, а на позицијама (k, l) су заједничке вероватноће укључивања
- pik вектор вероватноћа укључивања прве врсте

```
>
> #definiseemo verovatnoce ukljucivanja
> pik=c(0.2,0.7,0.8,0.5,0.4,0.4)
> #matrica verovatnoca druge vrste
> UPsystematicpi2(pik)
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  0.2  0.0  0.2  0.2  0.0  0.0
[2,]  0.0  0.7  0.5  0.2  0.4  0.3
[3,]  0.2  0.5  0.8  0.3  0.4  0.2
[4,]  0.2  0.2  0.3  0.5  0.0  0.3
[5,]  0.0  0.4  0.4  0.0  0.4  0.0
[6,]  0.0  0.3  0.2  0.3  0.0  0.4
>
```

HTestimator(y, pik)

- Рачуна Horvitz-Thompson-ову оцену за тотал популације
- y вектор обележја коју испитујемо њена дужина је једнака обиму узорка
- pik вектор вероватноћа укључивања прве врсте (дужина као обим узорка)

```
>
>
> data(belgianmunicipalities)
> attach(belgianmunicipalities)
> # racunamo verovatnoce ukljucivanja
> pik=inclusionprobabilities(Tot04,200)
>
>
> # definiesemo velicinu koju ispitujemo
> y=TaxableIncome
> # Uzimamo Puasonov uzorak
> s=UPpoisson(pik)
> # racunamo Horvitz-Thompson-ovu ocenu
> HTestimator(y[s==1],pik[s==1])
                [,1]
[1,] 114635195908
>
>
> |
```

HTstrata($y, pik, strata, description=F$)

- Рачуна Horvitz-Thompson-ову оцену за тотал популације при стратификованом плану узорковања
- y вектор обележја који испитујемо њена дужина је једнака обиму узорка n
- pik вектор вероватноћа укључивања прве врсте (дужине n)
- $strata$ вектор дужине n , са јединицама популације који се налазе у стартумима
- $description$ ако је $TRUE$ оцена се штампа за сваки стратум

```

>
> # Koristimo bazu 'iris' kao populaciju za uzimanje uzorka
> data(iris)
> # obelezje 'species' ima 3 kategorija
> # ovo cemo koristiti za stratifikaciju
> # racuna stratume populacije
> table(iris$species)

      setosa versicolor  virginica
      50         50         50

>
> # velicine stratuma uzorka su date sa size=c(6,4,3)
> # koristimo metodu SRSWOR (jednake verovatnoce, bez ponavljanja)
> st=strata(iris,stratanames=c("Species"),
+          size=c(6,4,3),
+          method="srswor")
> # izvlaci sam uzorak iz baze
> x=getdata(iris, st)
>
> # racunanje HT ocene za obelezje Petal.Length
> HTstrata(x$Petal.Length,x$Prob,x$Stratum,description=TRUE)
For stratum 1 ,the Horvitz-Thompson estimator is: 79.16667
For stratum 2 ,the Horvitz-Thompson estimator is: 225
For stratum 3 ,the Horvitz-Thompson estimator is: 253.3333
The Horvitz-Thompson estimator is:
[1] 557.5
> |

```

balancedtwostage(X , selection, m , n , PU ,
comment=TRUE, method=1)

- Даје балансирани двофазни узорак
- Φ -ја враћа вектор са следећим колонама: јединице изабране у другој фази (0-није изабрана, 1-изабрана), коначне вероватноће укључивања, јединице изабране у првој фази, вероватноће укључивања у првој фази, вероватноће укључивања у другој фази
- X матрица помоћних променљивих на основу којих се узорак балансира
- *selection*: 1 – за SRSWOR у свакој фази
2 – са self-weighting двофазни узорак
- m обим узорка у првој фази
- n обим узорка у другој фази
- PU вектор који дефинише узорак у првој фази
- *comment* пише се коментар извршавања ако је *TRUE*

```
>data(MU284)
>X=cbind(MU284$P75,MU284$CS82,
+        MU284$SS82,MU284$ME84)
>N=dim(X)[1]
>PU=MU284$CL
>m=20
>n=60
>res=balancedtwostage(X,1,m,n,PU,TRUE)
#uzorak i verovatnoce ukljucivanja
>res
```

poststrata(data, postnames)

- Ова команда врши постстратификацију на основу неколико критеријума. Даје објекат који садржи коначну базу са новом колоном '*poststratum*' (јединице постстратума) и број постстратума
- *data* база података, број врста је једнак n (обим узорка)
- *postnames* вектор обележја по којем се врши постстратификација


```
> data=rbind(matrix(rep("nc",165),165,1,byrow=TRUE),
+             matrix(rep("sc",70),70,1,byrow=TRUE))
> data=cbind.data.frame(data,c(rep(1,100), rep(2,50), rep(3,15),
+                               rep(1,30),rep(2,40)),1000*runif(235))
> names(data)=c("state","region","income")
> table(data$region,data$state)
```

	nc	sc
1	100	30
2	50	40
3	15	0

```
> # postratifikacija koriscenjem kriterija 'state' i 'region'
> poststrata(data,postnames=c("state","region"))
```

postest(data, y, pik, NG, description=FALSE)

- Рачуна оцену за тотал популације при постстратификације
- *data* база података, број врста је једнак n (обим узорка)
- *y* вектор обележја који испитујемо њена дужина је једнака обиму узорка n
- *pik* вектор вероватноћа укључивања прве врсте (дужине n)
- *NG* вектор фреквенције популације у свакој групи G ; за стратификовани узорак са постстратификацијом NG је матрица фреквенција популације у сваком пољу GN
- *description* ако је *TRUE* штампа се оцена за сваки постстратум

```

>
> data(swissmunicipalities)
> attach(swissmunicipalities)
> table(swissmunicipalities$REG)

 1    2    3    4    5    6    7
589 913 321 171 471 186 245
> st=strata(swissmunicipalities, stratanames=c("REG"),
+          size=c(30,20,45,15,20,11,44), method="srswor")
> x=getdata(swissmunicipalities, st)
> px=poststrata(x, "REG")
> ct=unique(px$data$REG)
> yy=numeric(length(ct))
> for(i in 1:length(ct))
+ {xx=swissmunicipalities[REG==ct[i],]
+  yy[i]=nrow(xx)
+ }
> yy
[1] 171 589 321 913 471 186 245
> postest(px$data, y=px$data$Pop020, pik=px$data$Prob,
+         NG=diag(yy), description=FALSE)
[1] 1534257
> HTstrata(x$Pop020, x$Prob, x$Stratum)
[1] 1534257
>

```

ratioest(y, x, T_x, p_{ik})

- Рачуна количничку оцену за тотал популације
- y вектор обележја који испитујемо њена дужина је једнака обиму узорка n
- x вектор помоћних података дужине n
- T_x тотал популације за x
- p_{ik} вектор вероватноћа укључивања прве врсте (дужине n)

```
> data(MU284)
> # postoji 3 outliera koji se brisu iz populacije
> MU281=MU284[MU284$RMT85<=3000,]
> attach(MU281)
> # racuna verov ukljuc koristeći oblezje P85; obim uzorka 120
> pik=inclusionprobabilities(P85,120)
> # definisemo oblezje koje posmatramo
> y=RMT85
> # definisemo popocni podatak
> x=CS82
> # izvlati sistematicni uzorak obima 120
> s=UPsystematic(pik)
> # racuna kolicnicku ocenu
> ratioest(y[s==1],x[s==1],sum(x),pik[s==1])
[1] 53330.48
> |
```

ratioest_strata(y,x,TX_strata,pik,strata, description=FALSE)

- Рачуна количничку оцену за тотал популације за стратификован план узорковања
- y вектор обележја који испитујемо њена дужина је једнака обиму узорка n
- x вектор помоћних података дужине n
- TX_strata вектор који садржи тотале популације x на сваком стратуму
- pik вектор вероватноћа укључивања прве врсте (дужине n)
- $strata$ вектор дужине n , са јединицама популације који се налазе у стартумима
- $description$ ако је $TRUE$ штампа се оцена за сваки стратум

```

> data(MU284)
> # postoji 3 outliers koji se brisu iz populacije
> MU281=MU284[MU284$RMT85<=3000,]
> attach(MU281)
> # racuna verov ukljuc koristeci oblezje P85; obim uzorka 120
> pik=inclusionprobabilities(P85,120)
> # definisemo oblezje koje posmatramo
> y=RMT85
> # definisemo popocni podatak
> x=CS82
> table(MU281$REG)

 1  2  3  4  5  6  7  8
24 48 32 37 55 41 15 29
> # izvlati sistematicni uzorak obima 120
> s=strata(MU281,c("REG"),size=c(4,10,8,4,6,4,6,7),
+         method="systematic",pik=P85)
> #izdvojimo uzorak
> MU281sample=getdata(MU281,s)
> # racuna x total populacije u svakom stratumu
> TX_strata=as.vector(tapply(CS82,list(REG),FUN=sum))
> # racuna kolicnicku ocenu
> ratioest_strata(MU281sample$RMT85,MU281sample$CS82,TX_strata,
+                MU281sample$Prob,MU281sample$Stratum)
[1] 56501.81
>

```

```
regest(formula,Tx,weights,pikl,n,  
       sigma=rep(1,length(weights)))
```

- Рачуна регресиону оцену тотала популације
- *formula* регресиони модел ($y \sim x$)
- *Tx* тотал популације за обележје x
- *weights* вектор тежина дужине n
- *pikl* матрица вероватноћа укључивања друге врсте узорка
- *sigma* вектор позитивних бројева који се односи на *heteroscedasticity* (то је када подпопулација има нека другачија обележја него што има цела популација)

Ф-ја враћа листу која садржи следеће аргументе

- *regest* вредност регресионе оцене
- *coefficients* вектор бета коефицијената
- *std_error* стандардна грешка коефицијената
- *t_value* t-вредност која је придружена коефицијентима
- *p_value* p-вредност која је придружена коефицијентима
- *cov_mat* коваријациона матрица коефицијената
- *weights* назначене тежине.
- *y* резултирајућа променљива
- *x* матрица модела

```

>
> data(MU284)
> # ima 3 outliera koje izbacujemo iz populacije
> MU281=MU284[MU284$RMT85<=3000,]
> attach(MU281)
> # racuna verov ukljuc koristeći obelezje P85; obim uzorka 40
> pik=inclusionprobabilities(P85,40)
> # verov ukljuc druge vrste za sistematski uzorak
> pik1=UPsystematicpi2(pik)
> # izvlaci uzorak obima 40
> s=UPsystematic(pik)
> # definisemo obelezje od vaznosti
> y=RMT85[s==1]
> # definisemo pomocne podatke
> x1=CS82[s==1]
> x2=SS82[s==1]
> # verov ukljuc druge vrste za s
> pik1s=pik1[s==1,s==1]
> # verov ukljuc prve vrste za s
> piks=pik[s==1]
> # racuna regresionu ocenu za model  $y \sim x_1 + x_2 - 1$ 
> r=regest(formula=y~x1+x2-1,
+           Tx=c(sum(CS82),sum(SS82)),weights=1/piks,pik1=pik1s,n=40)
> |

```

```
> r
$call
regest(formula = y ~ x1 + x2 - 1, Tx = c(sum(CS82), sum(SS82)),
        weights = 1/piks, pik1 = pik1s, n = 10)
```

```
$formula
y ~ x1 + x2 - 1
```

```
$x
  x1 x2
1  24  8
2  14 33
3  19 38
4  15 39
5  13 30
6  12 30
7  11 19
8  15 35
9   5 24
10  4 19
attr(,"assign")
[1] 1 2
```

```
$y
  1  2  3  4  5  6  7  8  9 10
241 488 807 431 275 396 208 965 249 54
```

```
$weights
[1] 25.117857 10.820000  6.572897 11.721667 20.094286 15.289130
[7] 21.978125  5.960169 18.507895 78.144444
```

```
$regest
[1] 64660.81
```

```
$coefficients
      xx1      xx2
10.893502  7.073114
```

```
$std_error
[1] 0.4362238 0.6255010
```

```
$t_value
      xx1      xx2
24.97228 11.30792
```

```
$p_value
      xx1      xx2
1.271414e-09 1.275042e-06
```

```
$cov_matrix
      [,1] [,2]
[1,] 0.1902912 -0.1037101
[2,] -0.1037101  0.3912516
```

```
attr(,"class")
[1] "regest"
```

```
> |
```

```
regest_strata(formula,weights,Tx_strata,strata,pikl,  
sigma=rep(1,length(weights)),description=FALSE)
```

- Рачуна регресиону оцену тотала популације за стратификовани узорак, исти регресиони модел се користи за сваки стратум
- *formula* регресиони модел ($y \sim x$)
- *weights* вектор тежина дужине n
- *Tx_strata* тотал популације за обележје x
- *strata* вектор идентификатора страума (vector of stratum identifier)
- *pikl* матрица вероватноћа укључивања друге врсте узорка
- *sigma* вектор позитивних бројева који се односи на *heteroscedasticity*
- *description* ако је TRUE за сваки страум се штампа : Horvitz-Thompson-ова оцена, бета коефицијенти, њихове стандардне грешке, t вредности, p вредности и ковариациона матрица
- Ф-ја враћа регресиону оцену као суму оцена по стратумима

```

>
>
> # dajemo neke nase podatke
> y=rgamma(10,3)
> x=y+rnorm(10)
> stratum=c(1,1,2,2,2,3,3,3,3,3)
> # obim populacije
> N=200
> # obim uzorka
> n=10
> # proporcionalna raspodela, nh/Nh=n/N
> pikl=matrix(0,n,n)
> for(i in 1:n)
+ {for(j in 1:n)
+   if(i!=j)
+     pikl[i,j]=pikl[j,i]=n*(n-1)/(N*(N-1))
+   pikl[i,i]=n/N
+ }
> regist_strata(formula=y~x-1,weights=rep(N/n,n),
+               Tx_strata=c(50,30,40),
+               strata=Stratum,pikl,description=TRUE)

```

ИЗЛАЗ

```
Stratum 1 , the regression estimator is: 48.16086
Number of units: 2
Beta coefficient(s): 0.9781678
Std. error: 0.01336873
t-value: 73.16832
p_value: 0.008700215
cov_matrix:
      [,1]
[1,] 0.000178723
Stratum 2 , the regression estimator is: 27.39036
Number of units: 3
Beta coefficient(s): 0.7903391
Std. error: 0.05614442
t-value: 14.0769
p_value: 0.005008571
cov_matrix:
      [,1]
[1,] 0.003152196
Stratum 3 , the regression estimator is: 35.62239
Number of units: 5
Beta coefficient(s): 0.8107618
Std. error: 0.1126522
t-value: 7.197034
p_value: 0.001975194
cov_matrix:
      [,1]
[1,] 0.01269052
The regression estimator is:
[1] 111.1736
> |
```

Нájек-ова оцена

- Нека се врши узорковање са вероватноћама пропорционалним величин, нека су π_i вероватноће укључивања i -те јединице. Тада је Нájек-ова оцена средине популације:

$$\hat{Y}_{\text{Нај}} = \frac{\sum_{i=1}^n \frac{Y_i}{\pi_i}}{\sum_{i=1}^n \frac{1}{\pi_i}}$$

- Слично, оцена за тотал популације је

$$\bar{Y}_{\text{Нај}} = N \frac{\sum_{i=1}^n \frac{Y_i}{\pi_i}}{\sum_{i=1}^n \frac{1}{\pi_i}}$$

Најкекеstimator($y, pik, N=NULL,$
 $type=c("total", "mean")$)

- Рачуна Најек-ову оцену за тотал популације или средиште популације
- y вектор обележја чију оцену тражимо, њена дужина је једнака обиму узорка n
- pik вектор вероватноћа укључивања прве врсте (дужине n)
- N обим популације, овај аргумент је дат само када се рачуна оцена за тотал, иначе је $NULL$
- $type$ тип оцене: " $total$ " или " $mean$ "


```
>
> data(belgianmunicipalities)
> # verovatnoce ukljucivanja
> pik=inclusionprobabilities(belgianmunicipalities$Tot04,200)
> N=length(pik)
> n=sum(pik)
> # obelezje za koju racunamo ocenu
> y=belgianmunicipalities$TaxableIncome
> # uzima Puasonov uzorak obima 200
> s=UPpoisson(pik)
> # Hajek-ova ocena za sredinu populacije
> Hajkestimator(y[s==1],pik[s==1],type="mean")
      [,1]
[1,] 207251898
Warning message:
In Hajkestimator(y[s == 1], pik[s == 1], type = "mean") :
  by default the mean estimator is computed
> # Hajek-ova ocena za total populacije
> Hajkestimator(y[s==1],pik[s==1],N=N,type="total")
      [,1]
[1,] 122071367886
> |
```