ITAP - Osnove uporabe R

A. Blejec

andrej.blejec@nib.si

February 8, 2009

Abstract

Ogledali si bomo uporabo R za generiranje slučajnih števil in nekatere grafične operacije.

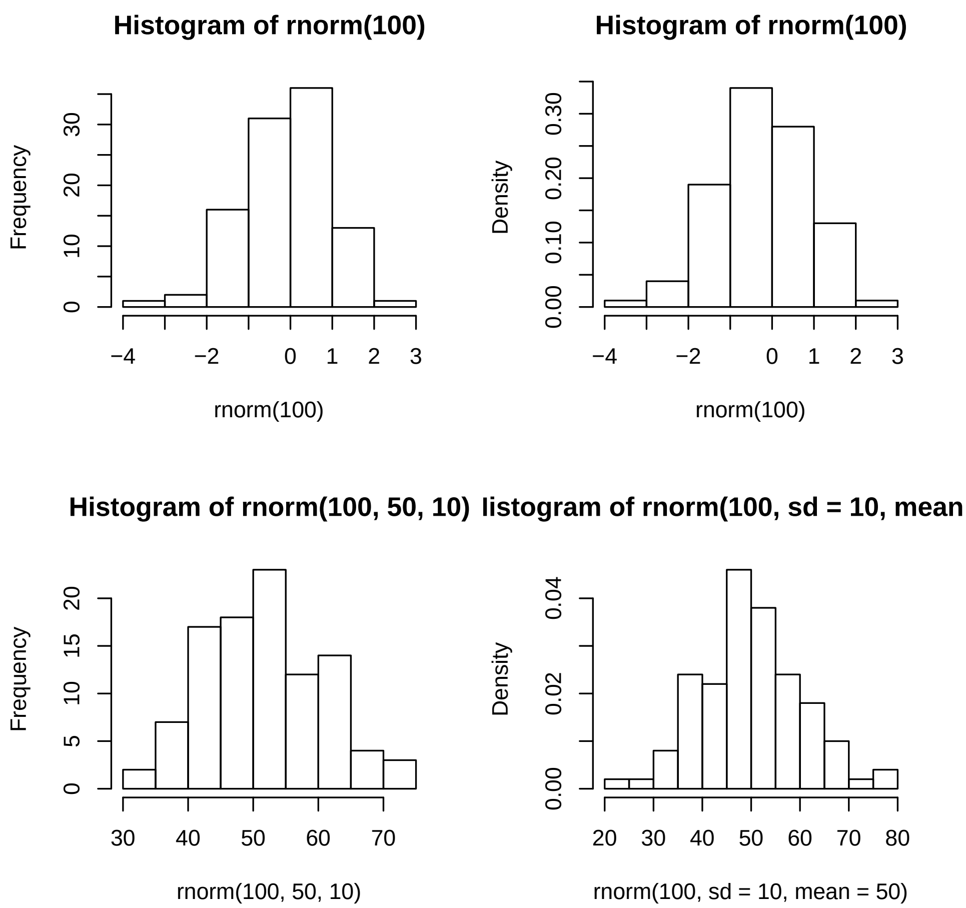
1 Verjetnostne porazdelitve

1.1

Normalna porazdelitev

```
> rnorm(5)      # 5 slučajnih števil iz N(0, 1)
[1] -1.19142464  0.54930055 -0.06240514  0.26544150
[5] -0.23459751
> args(rnorm)  # kakšni so argumenti funkcije rnorm?
function (n, mean = 0, sd = 1)
NULL

> oldpar <- par (mfrow = c(2, 2)) # pripravi štiri panele za slike
> hist(rnorm(100))                # histogram 100 slučajnih števil
> hist(rnorm(100), prob = T)      # verjetnosti na ordinatni osi
> hist(rnorm(100, 50, 10))        # argumenti po vrsti: n, mean, sd
> hist(rnorm(100, sd = 10, mean = 50), prob = T) # drugačen vrstni red argume
> par (oldpar)                    # privzemi prejšnje vrednosti grafičnih par
```



Funkcije za verjetnostne porazdelitve

Distribution	R name	additional arguments
beta	**beta**	shape1, shape2, ncp
binomial	**binom**	size, prob
Cauchy	**cauchy**	location, scale
chi-squared	**chisq**	df, ncp
exponential	**exp**	rate
F	**f**	df1, df2, ncp
gamma	**gamma**	shape, scale
geometric	**geom**	prob
hypergeometric	**hyper**	m, n, k
log-normal	**lnorm**	meanlog, sdlog
logistic	**logis**	location, scale
negative binomial	**nbinom**	size, prob
normal	**norm**	mean, sd
Poisson	**pois**	lambda
Student's t	**t**	df, ncp
uniform	**unif**	min, max
Weibull	**weibull**	shape, scale
Wilcoxon	**wilcox**	m, n

Funkcije za verjetnostne porazdelitve

Predpone

*r*andom	slučajna števila	**rnorm(n,...)**
*d*ensity	gostota verjetnosti	**dnorm(x,...)**
*p*robability	kvantilni rang	**pnorm(q,...)**
*q*antile	kvantil	**qnorm(p,...)**

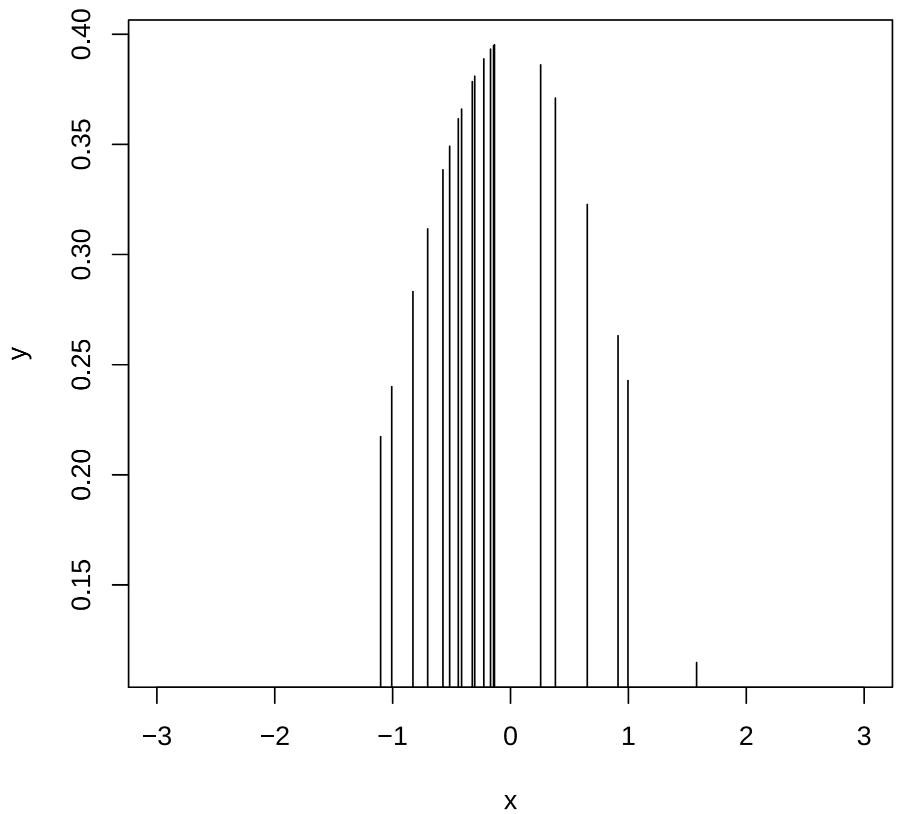
```
> oldopts <- options(digits=3)
> rnorm(5)
[1] -1.270 -0.371  0.389 -1.431 -2.503
> dnorm(seq(-3,2,1))
[1] 0.00443 0.05399 0.24197 0.39894 0.24197 0.05399
> pnorm(1.96)
[1] 0.975
> qnorm(0.975)
[1] 1.96
> options(oldopts)
```

Slika verjetij

```
> slVer <- function(n=20,m=1,xlim=c(-3,3)){
+ for(j in 1:m)
+ {
+ x <- rnorm(n)
+ y <- dnorm(x)
+ plot(x,y,type="h",xlim=xlim)
+ }
+ }
> slVer()
```

Slika verjetij

```
> slVer()
```

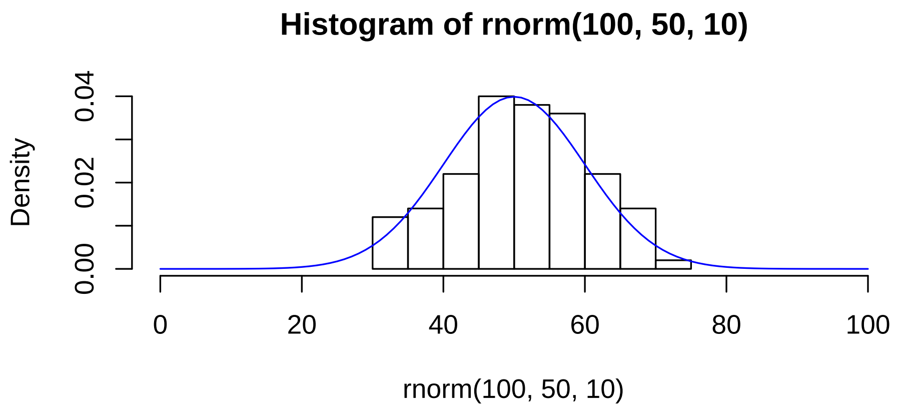


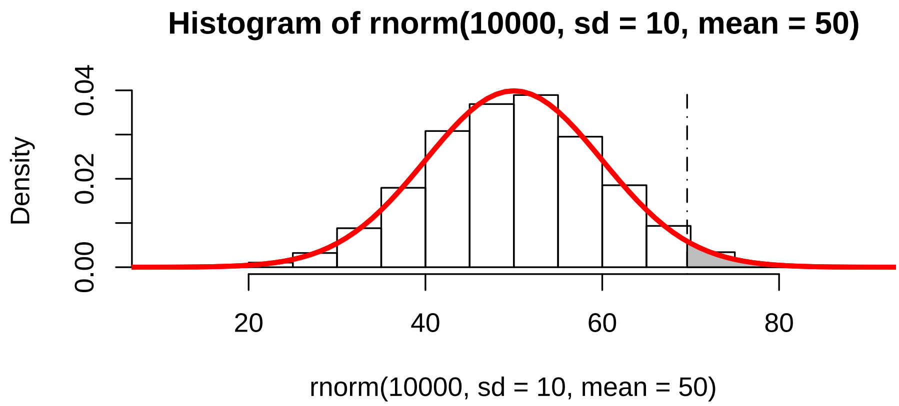
Normalna porazdelitev in pobarvane kritične vrednosti

Funkcije za verjetnostne porazdelitve

```
> oldpar <- par (mfrow = c(2, 1))
> hist(rnorm(100, 50,10), prob = T, xlim = c(0, 100))
> lines(0:100, dnorm(0:100, 50, 10), col = "blue",xpd=TRUE) # dodaj krivuljo g
> hist(rnorm(10000, sd = 10, mean = 50), prob = T) # za večji vzorec
> lines(0:100, dnorm(0:100, 50, 10), col = "red", lwd = 3) # + krivulja
> (q <- qnorm(.975, 50, 10))                          # kvantil za N(50, 10) (s
> abline (v = qnorm(.975, 50, 10), lty = "1373")  # dodaj kritične meje
> #
> # pobarvaj kritični del porazdelitve
> #
> px <- seq(q, 100)                               # pripravi koordinate polig
> py <- c(dnorm(px, 50, 10))
> polygon(c(px, q), c(py, 0), col = 8)            # pobarvaj poligon px, py
> lines(0:100, dnorm(0:100, 50, 10), col = "red", lwd = 3) # popravi krivuljo
> par (oldpar)
```

Funkcije za verjetnostne porazdelitve





2 Simulacija linearne regresije

Simulacija linerane regresije: $Y = 2X + 3 + \varepsilon \quad X \sim N(20, 5)$ **in** $\varepsilon \sim N(0, 10)$

```
> n=20
> x <- round(rnorm(n,20,5)) #
> mean(x)
[1] 20.85
> var(x)          # var računa nepristransko varianco !!
[1] 32.23947
> x
 [1] 18 25 23 26 20 15 28 24 20 25 29 17 11 15 28 15 21
[18] 12 17 28
> y <- 2*x+3           # izračunaj ustrezne y
```

Simulacija linerane regresije: $Y = 2X + 3 + \varepsilon \quad X \sim N(20, 5)$ **in** $\varepsilon \sim N(0, 10)$

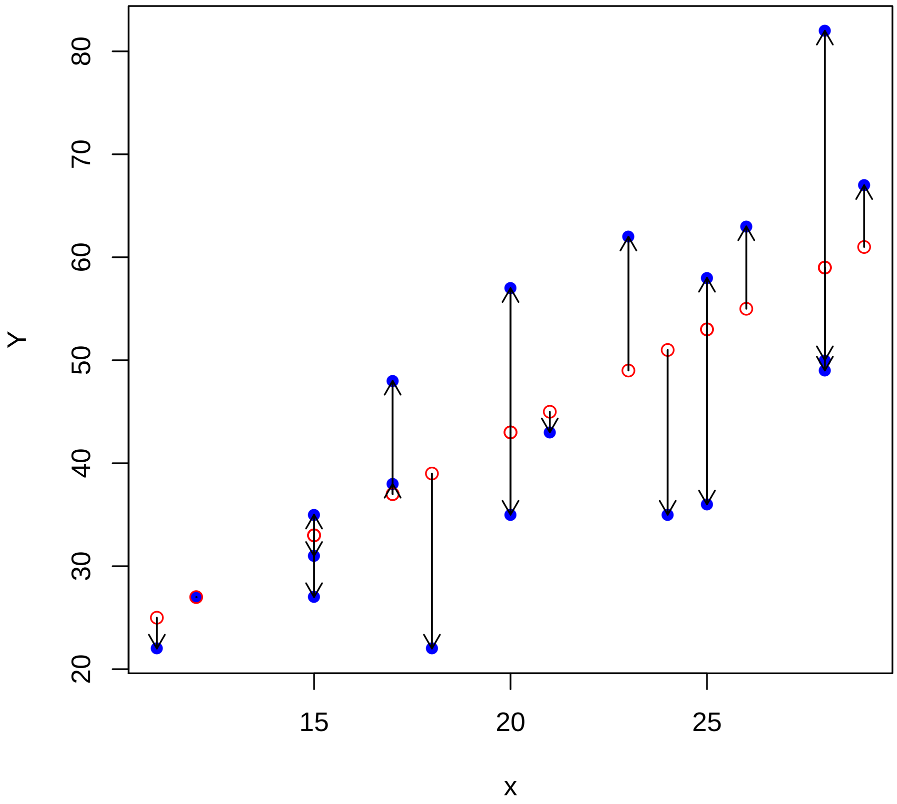
```
> mean(y)         # preveri parametre y
[1] 44.7
> var(y)
[1] 128.9579
> mean(x)*2+3
[1] 44.7
> var(x)*4
[1] 128.9579
> eps <- round(rnorm(n,sd=10)) # Pa še eps, mean=0
> mean(eps)
[1] -0.35
> var(eps)
[1] 119.7132
```

Simulacija linerane regresije: $Y = 2X + 3 + \varepsilon \quad X \sim N(20, 5)$ **in** $\varepsilon \sim N(0, 10)$

```
> Y <- y+eps       # sestavi obe slučajni spremenljivki
> Y
 [1] 22 58 62 63 35 35 82 35 57 36 67 48 22 27 50 31 43
[18] 27 38 49
> var(Y)           # Ali veljajo lastnosti za varianco vsote?
[1] 271.3974
> var(y)+var(eps) # videti je, da sta Y in eps korelirana!
[1] 248.6711
> cor(y,eps)       # !!
[1] 0.0914543
> var(y)+var(eps)+2*cov(y,eps)    # zdaj pa bo
[1] 271.3974
```

Simulacija linerane regresije: $Y = 2X + 3 + \varepsilon \quad X \sim N(20,5)$ **in** $\varepsilon \sim N(0,10)$

```
> plot(x,Y,col="blue",pch=16) # korelacijski grafikon za x in Y
> points(x,y,col="red")       # modelirane točke x,y
> segments(x,y,x,Y)           # odkloni (eps)
> arrows(x,y,x,Y,length=.1)   # podobno, a s puščicami
```



Simulacija linerane regresije: $Y = 2X + 3 + \varepsilon \quad X \sim N(20, 5)$ in $\varepsilon \sim N(0, 10)$

```
> fit<-lsfit(x,Y)          # "least squares fit"
> names(fit)
[1] "coefficients" "residuals"    "intercept"
[4] "qr"
> fit$coefficients
Intercept         X
-1.024406  2.176231
> yHat<-cbind(1,x)%*%fit$coeff
```

Linerana regresija: $Y = 2x + 3 + \varepsilon$: $x \sim N(20, 5)$ in $\varepsilon \sim N(0, 10)$

```
> plot(x,Y,col="blue",pch=16) # korelacijski grafikon za x in Y
> points(x,y,col="red")       # modelirane točke x,y
> segments(x,y,x,Y)           # odkloni (eps)
> arrows(x,y,x,Y,length=.1)   # podobno, a s puščicami
> abline(lsfit(x,Y))        # dodaj regresijsko črto
> abline(lsfit(x,Y),col=3,lwd=3)  # malo jo olepšaj
> points(x,yHat)
```