Зад. 1 Хвърлят се 30 зара. Каква е вероятността да се паднат по-малко от 5 шестици? Сравнете теоритичната вероятност с експериментални данни. Можем да твърдим, че с вероятност 75% ще се паднат повече от колко шестици?

```
# експериментални данни
# I-ви начин
set.seed(2020)
sol1e=function(n=1000){
   cnt=0
   for(i in 1:n){
      roll=sample(1:6, size=30, replace=T)
      sixes=sum(roll==6)
      if(sixes<5) cnt=cnt+1
   }
   cnt/n
}
sol1e()
# [1] 0.4228

# II-ри начин
# Оценете вероятността да се паднат по-малко от 5 шестици? Bi(30,1/6)
x=rbinom(n=1000, size=30, prob=1/6)
sum(x<5) / length(x)
# [1] 0.403
```

#теоритични данни:

I-ви начин: P(X<5)= $\sum_{i=0}^{4} P(X=i)$

```
sol1t=function(){
   res=0
   for(i in 0:4){
      res=res+dbinom(x=i, size=30, prob=1/6)
   }
   res
}
sol1t()
# [1] 0.4243389
```

II-ри начин: P(X<5)=P(X<=4)

```
pbinom(q=4, size=30, prob=1/6)
# [1] 0.4243389
```

III-ти начин: 1-P(X>4)

```
1-pbinom(q=4, size=30, prob=1/6, lower.tail=F)
[1] 0.4243389

quantile(rbinom(n=10000, size=30, prob=1/6), probs = c(1-.75))
# 25%
#   4
```

```
qbinom(p=0.75, size=30, prob=1/6, lower.tail=F)
# [1] 4
```

Зад. 2 Стрелец уцелва мишена с вероятност 20%. За да спечели стрелецът трябва да направи три точни попадения. Каква е вероятността за това да са му необходими:

а) точно 8 изстрела;
б) повече от 6 изстрела;
в) между 5 и 8 изстрела, включително?

```
r=3
p=.2

# а)
# 8-мия изстрел трябва да е печеливш, следователно трябва да разпределим 2 печеливши
# изстрела измежду общо 7 изстрела. Разпределението е отрицателно биномно.

dnbinom(x=8-r, size=r, prob=p)
# [1] 0.05505024

pnbinom(q=5, size = 3, prob = 0.2) - pnbinom(4, size = 3, prob = 0.2)
# [1] 0.05505024

# б)
1-pnbinom(q=6-r, size=r, prob=p)
# [1] 0.90112
pnbinom(q=6-r, size=r, prob=p, lower.tail=F)
# [1] 0.90112

# в)
pnbinom(q=8-r, size=r, prob=p)-pnbinom(q=4-r, size=r, prob=p)
# [1] 0.1758822

1-(pnbinom(q=4-r, size=r, prob=p)+pnbinom(q=8-r, size=r, prob=p, lower.tail=F))
# [1] 0.1758822

sol.c=function(){
   ans=0
   for(i in 5:8){
      ans=ans+dnbinom(x=i-r, size=3, prob=0.2)
   }
   ans
}
res=sol.c(); res
# [1] 0.1758822
```

Зад. 3 В урна има 7 бели и 6 черни топки. От урната последователно без връщане се теглят 8 топки. Нека X е броя на изтеглените бели. Направете 1000 симулации и по тях пресметнете: границите, в които се мени X, $\mathbb{E}X$ и $\mathbf{D}X$. Намерете теоритичните стойности за $\mathbb{E}X$ и $\mathbf{D}X$. Представете графично емпиричното и теоритичното разпределение на X (на една графика).

```
# симулация

balls=c(rep("W",7), rep("B",6))

set.seed(2020)
sol3sim=function(){
   white=c()
   for(i in 1:1000){
      take=sample(balls, size=8, replace=F)
      white[i]=sum(take=="W")
   }
   white
}
X=sol3sim()
summary(X)
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#  2.000   4.000   4.000   4.281   5.000   7.000

var(X)
# [1] 0.7868258

# втори начин
set.seed(2020)
sim = rhyper(nn=1000, m=7, n=6, k=8)
summary(sim)
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#  2.000   4.000   4.000   4.269   5.000   7.000

var(sim)
# [1] 0.8474865

# теоритично
sol3tEX=function(){
   EX=0
   for(i in 2:7){
     EX=EX+i*dhyper(x=i, m=7, n=6, k=8)
   }
   EX
}
sol3tEX()
# [1] 4.307692

sol3tDX=function(){
   EX=0; EX2=0
   for(i in 2:7){
      y=dhyper(x=i, m=7, n=6, k=8)
      EX=EX+i*y
      EX2=EX2+i**2*y
```

```
    }
    EX2-EX**2
}
sol3tDX()
# [1] 0.8284024

# или използвайки формулите
k=8; w=7; n=13
m = k * (w / n); m # EX
# [1] 4.307692

v = k * (w / n) * ((n - w) / n) * (n - k)/(n - 1); v # DX
# [1] 0.8284024

# графики
m=7; n=6; k=8
X=rhyper(1000, m, n, k)

# Представете графично емперичното и теоритичното разпределение на X(на една
# графика).
hist(X, breaks=c(seq(1.5,7.5,1)), probability = TRUE)
points(2:7, dhyper(2:7, m, n, k), type = "h", lwd = 3, col = "darkblue")
points(2:7, dhyper(2:7, m, n, k), type = "p", lwd = 3, col = "darkblue")

X = rhyper(nn=1000, m=7, n=6, k=8)
hist(X, breaks=6,  prob=T, col="gold")
lines(density(X, adjust=4), col="purple2", lwd=2)
```

Зад. 4 Лотария се провежда със следните правила. Всеки участник избира едно число от 1 до 2n, не е необходимо да избират различни числа. Когато броят на участниците стане n се теглят 5 печеливши числа. Каква е вероятността да се паднат точно две награди. Пресметнете при n=10, 100, 1000, 10000. С каква случайна величина ще моделирате броя на печалбите?

Имаме 5 печеливши и 2n-5 губещи. От тях теглим n с връщане. X~Bi(n,5/2n).
Каква е вероятността да се паднат точно 2 награди е същото като P(X=2). Тогава за n=10, 100, 1000, 10000 ще имаме

```
n = c(10, 100, 1000, 10000)
dbinom(2, n, 5 / (2*n))
# [1] 0.2815676 0.2587841 0.2567403 0.2565381
```

Зад. 5 Решете задача 4, ако лицата теглеха различни числа.

Имаме 5 печеливши и 2n-5 губещи. От тях теглим n без връщане. X~HG(5, 2n-5, n),
Каква е вероятността да се паднат точно 2 награди е същото като P(X=2). Тогава за n=10, 100, 1000, 10000 ще имаме

```
n = c(10, 100, 1000, 10000)
dhyper(2, 5, 2*n - 5, n)
# [1] 0.3482972 0.3156646 0.3128129 0.3125313
```

Зад. 6 За коледно парти всеки от n (n = 20) участници носи по един подарък. Подаръците се номерират и в шапка се слагат номерата от 1 до n. Всеки участник си тегли номер и получава съответния подарък. Напишете функции, които пресмятат:

а) теоретичната вероятност,

б) емперичната вероятност изчислена по 10000 опита,

никой да не получи своятподарък. Пресметнете очакването на броя хора получили своя подарък.

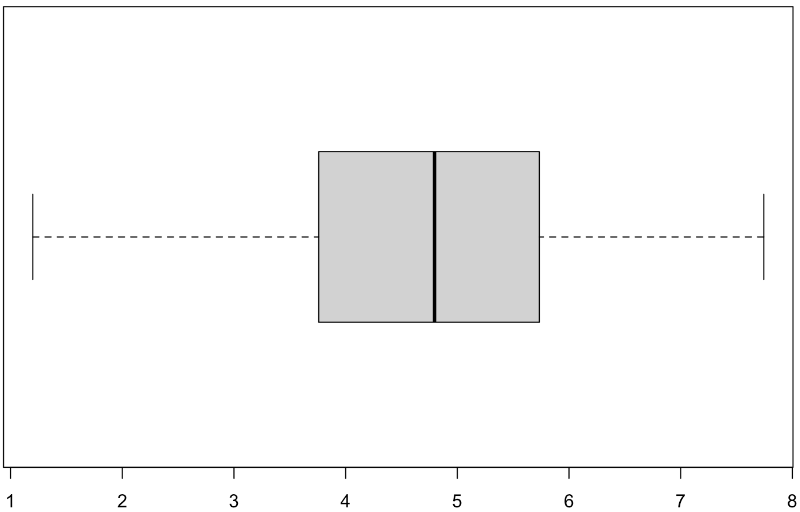
Зад. 1 Генерирайте 100 случайни наблюдения над X. Постройте боксплот и хистограма; добавете емпиричната и теоритичната плътност, ако:

а) $X \in \mathcal{N}(5,2)$
б) $X \in \mathcal{U}(1,5)$
в) $X \in Exp(3)$
г) $X \in \Gamma(5,1)$
д) $X \in \mathcal{X}^2(3)$
е) $X \in t(5)$
ж) X е смес от две разпределения $\mathcal{N}(1,2)$ и $\mathcal{N}(5,2)$ с вероятност за първото $p = 0.4$.

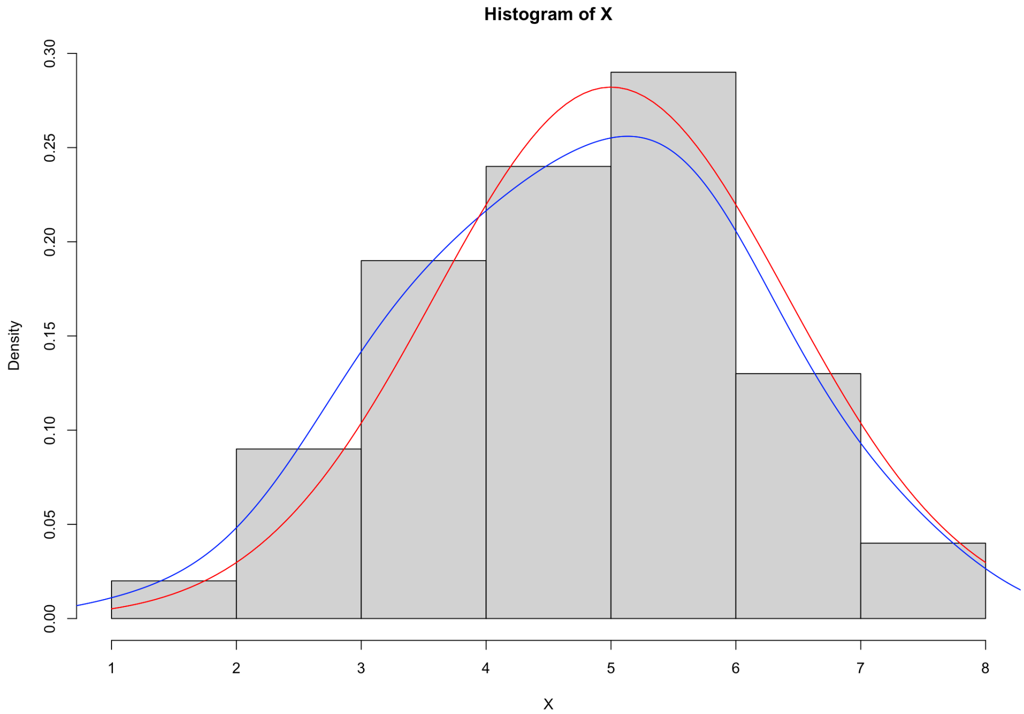
Определете вида на разпределението (симетрично или изместено, леки или тежки опашки, едномодални и т.н.)

Решение:

```
# а) X ∈ N(5,2)
X=rnorm(n=100, mean=5, sd=sqrt(2))
boxplot(X, horizontal = T)
```

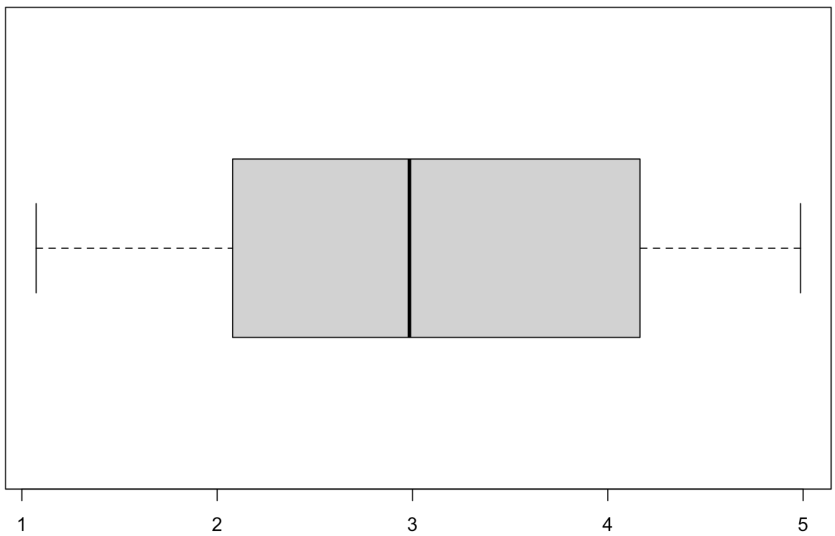



```
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dnorm(x, mean=5, sd=sqrt(2)), add=T, col="red")

install.packages("EnvStats")
library(EnvStats)

skewness(X)
# [1] -0.1115661
# Коефициента на асиметрия е приблизително 0, т.е. имаме симетрично разпределение (камбаната е леко в дясно, тъй като числото, което получаваме е <0).
kurtosis(X)
# [1] -0.3740814
# симетрично, леко изместено, унимодално, леки опашки
```

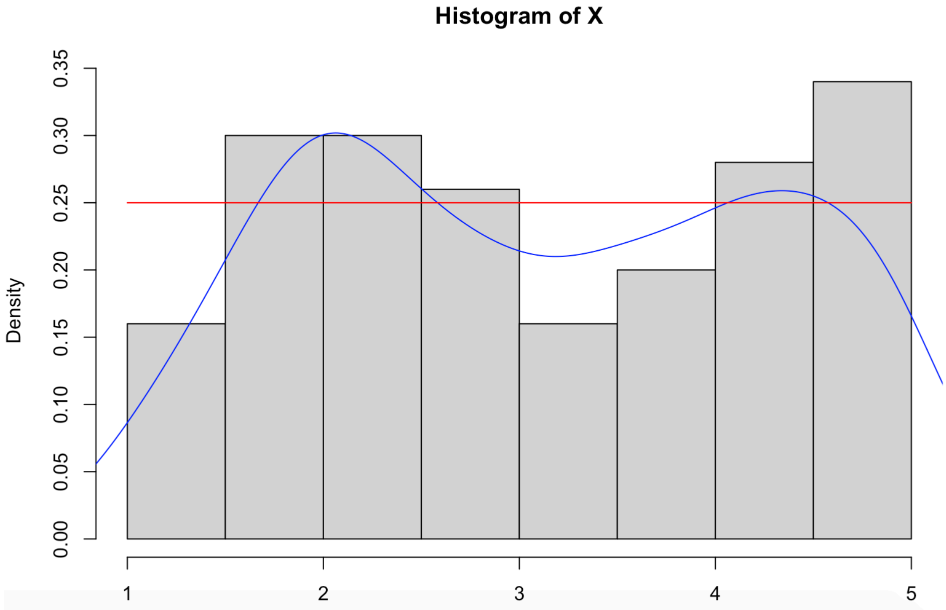


б) $X \in \mathscr{U}(1,5)$

```
X=runif(n=100, min=1, max=5)
boxplot(X, horizontal=T)
```

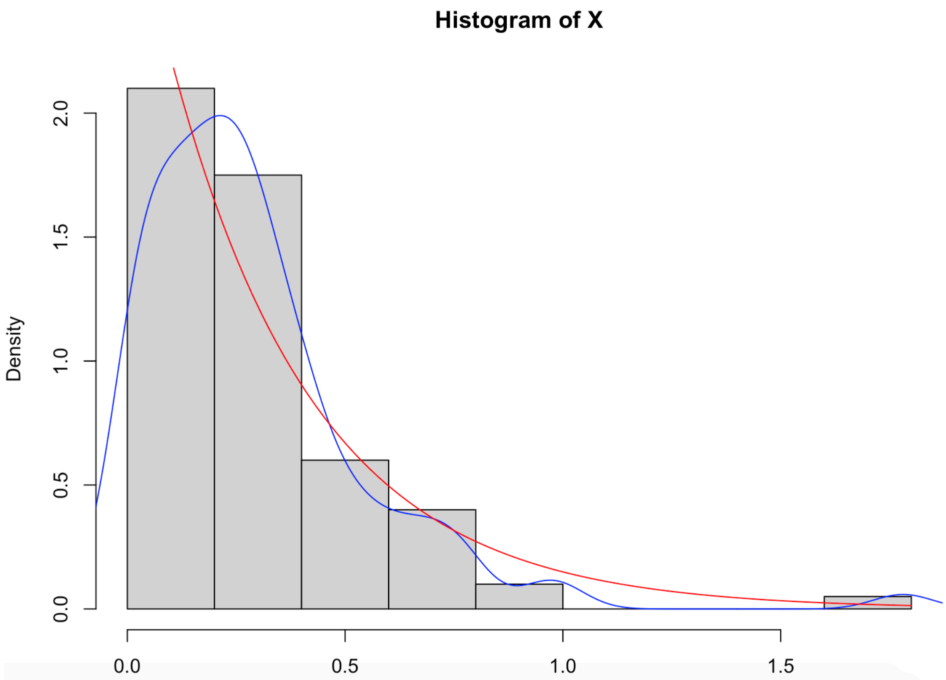



```
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dunif(x, min=1, max=5), add=T, col="red")
```



skewness(X)
# [1] 0.04045247
# коефициентът на асиметрия на нормалното е 0, а на извадката е приблизително нула - т.е. е симетрична.
kurtosis(X)
# [1] -1.290648
# симетрично унимодално, леки опашки

в) $X \in Exp(3)$




```
X=rexp(n=100, rate=3)
boxplot(X, horizontal=T)
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dexp(x, rate=3), add=T, col="red")
```

```
skewness(X)
# [1] 2.352239
kurtosis(X)
# [1] 9.51725
```

г) $X \in Gamma(5,1)$

```
X=rgamma(n=100, shape=5, rate=1)
boxplot(X, horizontal=T)
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dgamma(x, shape=5, rate=1), add=T, col="red")
skewness(X) # изместеност
# [1] 0.4098481
kurtosis(X) # изостренос
# [1] -0.4161709
```

д) $X \in \mathscr{X}^2(3)$

```
X=rchisq(n=100, df=3)
boxplot(X, horizontal=T)
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dchisq(x, df=3), add=T, col="red")
skewness(X) # изместеност
# [1] 1.906965
kurtosis(X) # изостренос
# [1] 3.988591
```

е) $X \in t(5)$

```
X=rt(n=100, df=5)
boxplot(X, horizontal=T)
hist(X, prob=T)
lines(density(X, bw="SJ"), col="blue")
curve(dt(x, df=5),add=T, col="red")
 skewness(X)
# [1] 0.7916025
kurtosis(X)
# [1] 2.295702

# ж)
i=rbinom(n=100, size=1, prob=0.4)
n1=rnorm(n=100, mean=1, sd=sqrt(2))
n2=rnorm(n=100, mean=5, sd=sqrt(2))
x=i*n1+(1-i)*n2
boxplot(x, horizontal=T)
hist(x, horizontal=T)
skewness(x)
kurtosis(x)
```

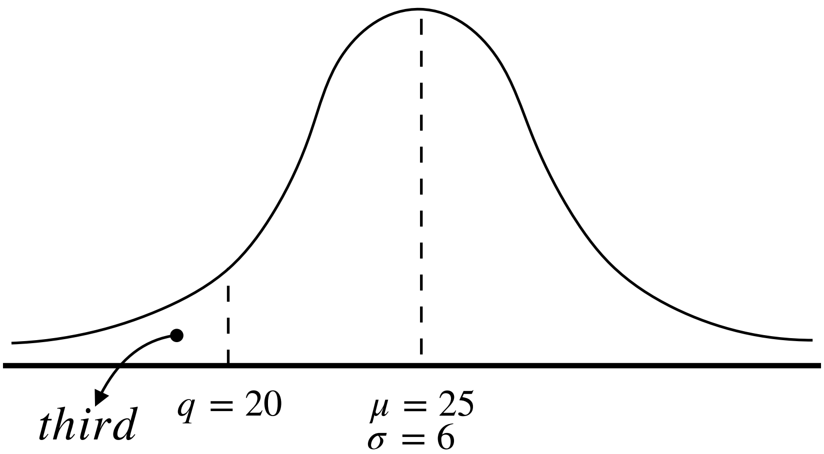
Зад. 3 Определете дали са нормално разпределени наблюденията:
а) Теглото на бебетата в babies от пакета UsingR;
б) exec.pay от пакета UsingR.

Решение:

```
# a)
library(UsingR)
attach(babies)

hist(wt, prob=T)
lines(density(wt, bw="SJ"), col="blue")
curve(dnorm(x, mean(wt), sd(wt)), add=T, col="red")

# в синьо е емпиричната плътност, а в червено е теоритичната

qqnorm(wt)
qqline(wt)

# на пръв поглед може разминаването в опашките да е незначително, но броя на
извадките може да е голям и това да ни подведе. За тази цел ще направим тест.

sh=shapiro.test(wt)
if(sh$p.value>.05) print("normal") else print("not normal")

# друг подход

qqplot.das(wt, "norm")
# при този подход може нагледно да видим дали опашките излизат от
доверителните интервали на съответното разпределение

# б)

hist(exec.pay, prob=T)
lines(density(exec.pay, bw="SJ"), col="blue")
curve(dnorm(x, mean(exec.pay), sd(exec.pay)), add=T, col="red")

qqnorm(exec.pay)
qqline(exec.pay)

sh=shapiro.test(wt)
if(sh$p.value>.05) print("normal") else print("not normal")

qqplot.das(exec.pay, "norm")
```

Зад. 4 Размерът на пъпешите е нормално разпределена сл. вел. с очакване 25 см. и дисперсия 36. Пъпешите по-малки от 20 см. са трето качество, а останалите се разделят на две равни по брой групи, като по-големите са първо качество, а по-малките са второ. Каква част от пъпешите са трето качество. Колко голям трябва да е пъпеш, за да бъде първо качество?

Решение:



m=25; s=6
third=pnorm(q=20, mean=m, sd=s)

I-ви н/н:
qnorm((1-third)/2, mean=m, sd=s, lower.tail=F)
# [1] 26.53817

II-ри н/н:
qnorm(third+(1-third)/2, mean=m, sd=s)
# [1] 26.53817

Зад. 5. Нека X гама разпределена сл. вел. с параметри 2 и 0.5. Определете:
а) P(X<1);
б) P(X>2);
в) c, така че P(X>c)=0.35;
г) Q_1, M, Q_3

Решение:

# а)
pgamma(q=1, shape=2, rate=0.5)
# [1] 0.09020401

б)
I-ви н/н
1-pgamma(q=2, shape=2, rate=0.5)
[# 1] 0.7357589

II-ри н/н
pgamma(q=2, shape=2, rate=0.5, lower.tail=F)
# [1] 0.7357589

# в)
qgamma(p=0.35, shape=2, rate=0.5, lower.tail=F)
# [1] 4.437689
# г)

```
qgamma(p=c(0.25, 0.5, 0.75), shape=2, rate=0.5)
# [1] 1.922558 3.356694 5.385269
```