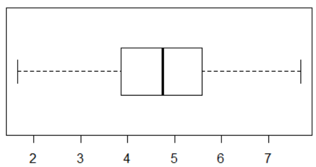
Зад.1 а) $X \in N(5, 2)$;

Генерирайте 100 случайни наблюдения над X.

```
> z = rnorm( 100, 5, sqrt(2))
```

Постройте боксплот

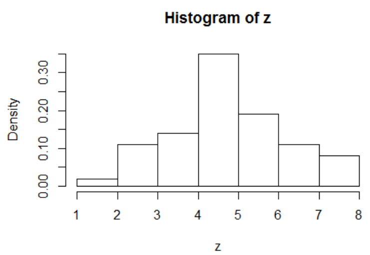
```
> boxplot(z, horizontal = T)
```



Разпределението е симетрично, без аутлайери, т.е. няма тежка опашка.

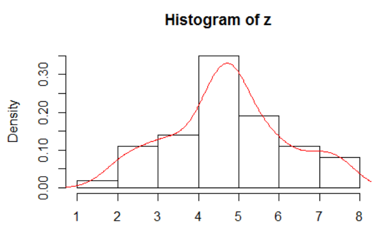
Постройте хистограма,

```
> hist(z, probability = T)
```



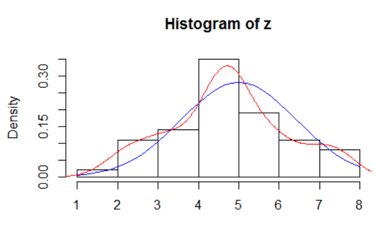
добавете емпиричната плътност,

```
> lines( density(z), col = ’red’)
```



добавете теоретичната плътност,

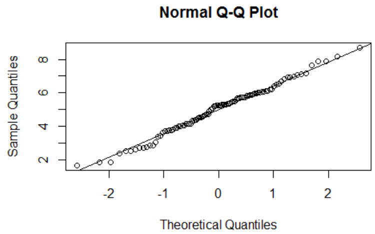
```
> x = seq( 1, 8, 0.2)
> y = dnorm( x, 5, sqrt(2))
> lines(x, y, col = ’blue’ )
```



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

Разпределението е симетрично, едномодално, няма тежка опашка. Има вид на нормално разпределение (това се вижда и от следващата графика).

```
> qqnorm(z)
> qqline(z)
```



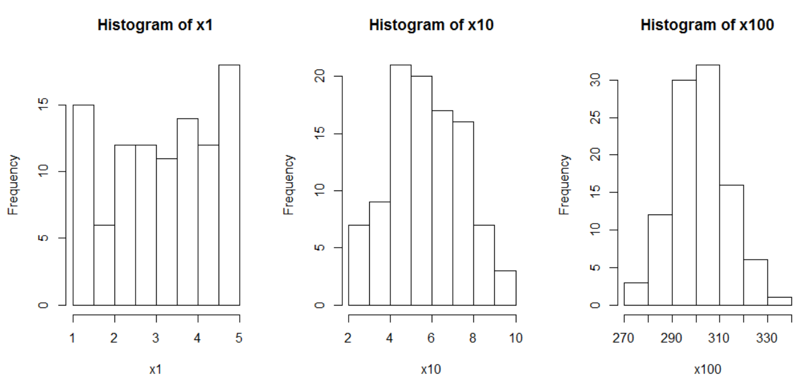
Зад.2 Нека $X_1, X_2, \dots X_n$ са независими сл.в. зададени както в Зад.1. Какво можете да кажете за разпределението на $Y = X_1 + X_2 + \dots + X_n$. Разгледайте случайте $n = 2, 10, 100$.

За да получим по-точна графика ще дефинираме $X_1, \dots X_n$, като вектори с размерност 200. Сумирането ще извършим във функция 'xsum'.

```
xsum = function( n, fn, ... ){
    s = rep( 0, 200 )
    for( i in 1 : n) s = s + fn( 200, ... )
    return(s)
}
```

Нека случайните виличини са разпределени, както в Зад.1 б), т.е. равномерно разпределени в интервала (1, 5)

```
> x1 = xsum(1, runif,1,5)
> x10 = xsum(10, runif,1,5)
> x100 = xsum(100, runif,1,5)
> split.screen(c(1,3))
> hist(x1)
> screen(2); hist(x10)
> screen(3); hist(x100)
```

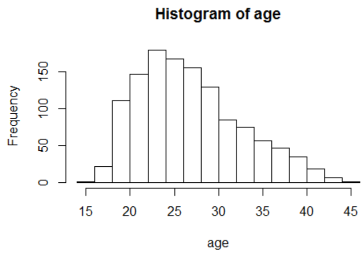


С увеличаване броя на събираемите сумата се приближава до нормално разпределение (Централна гранична теорема) независимо, че първоначалните случайни величини са нормално разпределени.

Зад.3 Определете дали са нормално разпределени:

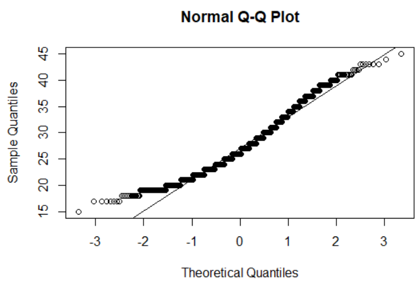
а) възрастта и теглото на майките в 'babies' от пакета UsingR;

```
> library( UsingR)
```
Премахване на 'outliers'
```
> age = babies$age[ babies$age != 99 ]
> hist( age )
```



Данните не са симетрични. Разпределението не е нормално, прилича на гама разпределение.

> qqnorm(age)
> qqline(age)



Още едно потвърждение, че данните не са нормално разпределени. Q-Q графиката има съществени отклонения от правата, по-скоро прилича на крива.

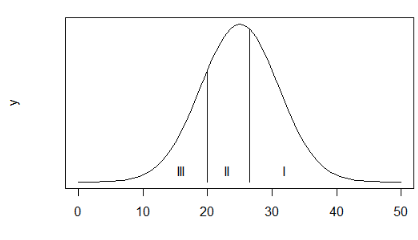
Зад.4 Размерът на пъпешите е нормално разпределена сл.в. с очакване 25 см. и дисперсия 36. Пъпешите по-малки от 20 см. са трето качество, а останалите се разделят на две равни по брой групи, като по-големите са първо качество, а по-малките второ. Каква част от пъпешите са трето качество.

> p.IIIp.II = pnorm(20, 25, 6)

```
0.2023
```

Приблизително 20% от пъпешите са трето качество.

Колко голям трябва да е пъпеш за да бъде първо качество.



Знаем че, лицето на част III е 0.2023, а лицата на I и II са равни. Като изпалзваме факта, че лицето под цялата крива е единица, лесно можем да намерим лицето на II, а след това и съответния квантил.

> p.II = (1 - p.III) / 2
> qnorm(p.I + p.II, 25, 6)

```
26.53
```

За да бъде I качество, пъпешът трябва да е по.голям от 26.53.