Зад.1 Нека $X_1, \dots X_{20}$ са наблюдения над сл.в. $X \in N(3,4)$. Постройте 95% доверителен интервал за математическото очакване $EX = \mu$, ако приемете че дисперсията е:

а) известна;

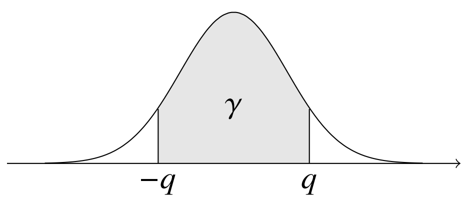
Ще генерираме наблюденията.

```
> x = rnorm( 20 ,3 ,2 )
```

Когато дисперсията е известна за построяване на доверителен интервал се използва следната централна статистика:

$$\mathrm{T} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \in N(0,1)$$

Лесно можем да построим доверителен интервал за параметъра θ. Процедурата се състои в следното. Намираме квантили $-q$ и q, такива че $\mathrm{P}(-q < \mathrm{T} < q) = \gamma = 0.95$.




```
> q = qnorm( 0.975, 0 , 1)
1.959964
```

Решаваме получените неравенства спрямо μ

$$\gamma = \mathrm{P}\left(-q < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < q\right) = \mathrm{P}\left(\overline{X} - q\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + q\frac{\sigma}{\sqrt{n}}\right)$$

Доверителния интервал, който получаваме $\mathrm{I} = \overline{X} \pm q\frac{\sigma}{\sqrt{n}}$

В нашия случай $\sigma^2 = 4$ и $n = 20$, така за границите на доверителния интервал съответно получаваме:

```
> mean( x ) - q * 2 / sqrt( 20 )
2.25 (RNG)
> mean( x ) + q * 2 / sqrt( 20 )
4.01(RNG)
```

Извод: Доверителния интервал е $I = (2.25,\ 4.01)$.

б) неизвестна.

За построяване на доверителен интервал се използва централна статиска подобна на предходната. Единствената разлика е, че заменяме неизвестната дисперсията σ^2 с оценката за нея

$$S^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2$$

В този случай статистиката T има разпределени на Стюдънт с $n-1$ степени на свобода

$$\mathrm{T} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \in t(n-1)$$

т.е. за намиране на квантила използваме:

```
> q = qt( 0.975, 19)
2.093
```

\> mean(x) - q * sd(x) / sqrt(20)
2.26 (RNG)
\> mean(x) + q * sd(x) / sqrt(20)
4.00 (RNG)
Извод: Доверителния интервал е $I = (2.26,\ \ 4.00)$.

Съществува и директен начин за намиране на доверителен интервал, ако наблюденията са нормално разпределени.

\> t.test(x)

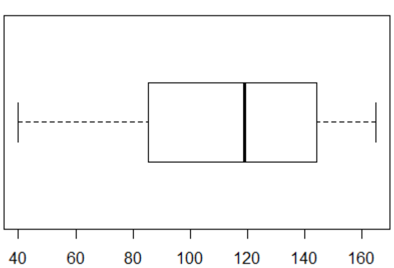
```
        One Sample t-test

data:  x
t = 7.5277, df = 19, p-value = 4.088e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2.260624 4.001870
sample estimates:
mean of x
 3.131247
```

Зад.3 Постройте 96% доверителен интервал за средната стойност по данните

a) rat

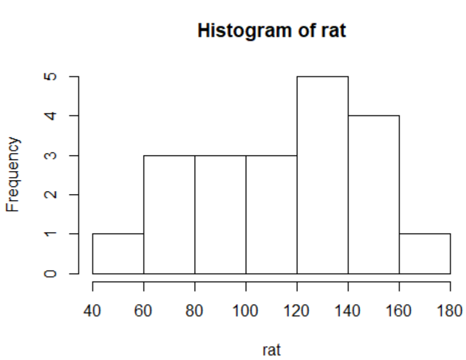
В началото ще проверим дали данните са нормално разпределени.

\> boxplot(rat, horizontal = T)



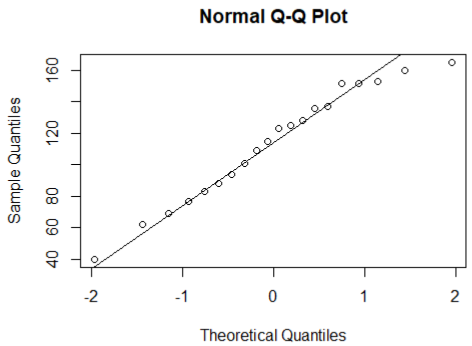
Боксплотът е приблизително симетричен няма аутлайери, възможно е да е нормално разпределение.

\> hist(rat)



Хистограмата не е идеална, но не се различава съществено от нормалните.

> qqnorm(rat)
> qqline(rat)



Данните лежат приблизително на правата, няма съществени отклонения.

Извод: Можем да приемем наблюденията за нормално разпределени.

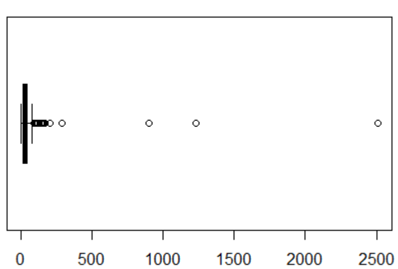
> t.test(rat, conf.level = 0.96)

```
        One Sample t-test

data:  rat
t = 14.176, df = 19, p-value = 1.48e-11
alternative hypothesis: true mean is not equal to 0
96 percent confidence interval:
  95.80624 131.09376
sample estimates:
mean of x
   113.45
```

Извод: Търсеният доверителен интервал е (95.8, 131.1).

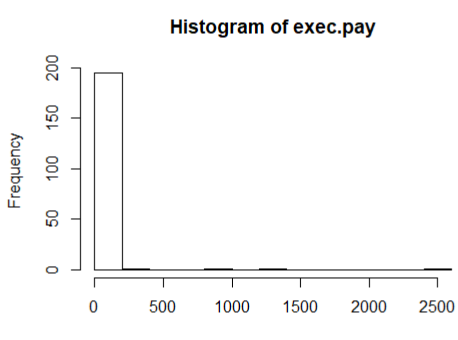
б) exec.pay Отново проверяваме данните за нормалност.

> boxplot(exec.pay, horizontal = T)



Изключително асиметричен боксплот с множество аутлайери. Няма как наблюденията да са нормално разпределени.

> hist(rat)



Хистограмата показва данни скупчени в началото.

Извод: Данните не са нормално разпределени ще използваме тест на Уилкоксън.

> wilcox.test(exec.pay, conf.int = T, conf.level = 0.96)

```
    Wilcoxon signed rank test with continuity correction

data:  exec.pay
V = 19306, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
96 percent confidence interval:
 25.99996 33.00003
sample estimates:
(pseudo)median
      29.00002
```

Извод: Има 96% вероятност, че средната компенсация получена от директорите е в интервал (26 000, 33 000). (Обърнете внимание, ние НЕ твърдим, че 96% от директорите получават компенсация между 26 000 и 33 000!)

Зад.4 При провеждане на анкета 87 от 150 анкетирани са отговорили, че са използвали даден продукт. Постройте 92% доверителен интервал за процента на хората използвали продукта.

Отговорът на всеки въпрос в анкетата е бил "Да" или "Не". Ние търсим доверителен интервал за вероятността за положителен отговор, затова използваме тест за пропорции.

> prop.test(87, 150, conf.level = 0.92)

```
    1-sample proportions test with continuity correction

data:  87 out of 150, null probability 0.5
X-squared = 3.5267, df = 1, p-value = 0.06039
alternative hypothesis: true p is not equal to 0.5
92 percent confidence interval:
 0.5051991 0.6514474
sample estimates:
   p
0.58
```

Извод: С 92% вероятност можем да твърдим, че процента на хората използвали продукта е в границите (50.5%, 65.1%).

Какъв ще бъде интервалът ако 870 от 1500 са отогоговорили, че са използвали продукта?

> prop.test(870, 1500, conf.level = 0.92)

Извод: Съотношението на положителните отговори в анкетата е същото, но поради по-големия обем интервалът става по-точен (55.7%, 60.2%).