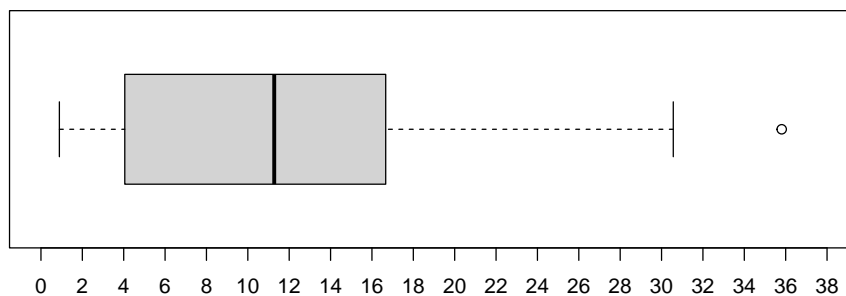


**Задача 1.** Получили сме с R следната графика за променливата  $x_2$ :



Как се нарича графиката?

Каква команда сме написали, за да я получим?

Приблизително каква е стойността на най-малкото наблюдение? А на най-голямото?

Приблизително какъв процент от наблюденията са по-малки или равни на 4?

Приблизително какъв процент от наблюденията попадат в интервала  $[4, 17]$ ?

**Задача 2.** Машина трябва да смесва фъстъци, лешници, кашу и бадеми в съотношение  $5 : 2 : 2 : 1$ . Взета е проба от 500 ядки, смесени от машината. Преброени са 267 фъстъка, 91 лешника, 103 ядки кашу и 39 бадема.

Напишете код на R, който бихте използвали, за да отговорите на въпроса: *Може ли да се приеме, че машината смесва ядките в желаното съотношение?*

Отговорете на въпроса с помощта на резултата по-долу. Обосновете отговора си. Формулирайте хипотезите и дефинирайте използваните означения.

```
data: x
```

```
X-squared = 4.476, df = 3, p-value = 0.2144
```

**Задача 3.** Извършен е експеримент, при който за сглобяването на даден елемент е използван робот. От 500 елемента, сглобени от робота, 15 се оказали дефектни. Известно е, че при сглобяване от човек дефектни са около 3.5% от елементите.

Напишете код на R, който бихте използвали, за да отговорите на въпроса: *Може ли да се твърди, че процентът дефектни елементи при използване на робот е по-малък отколкото при сглобяване от човек?*

Отговорете на въпроса с помощта на един от резултатите по-долу. Само един от резултатите е верен – посочете кой. Обосновете отговора си. Формулирайте хипотезите и дефинирайте използваните означения.

```
data: 15 out of 500, null probability 0.035
```

```
X-squared = 0.23686, df = 1, p-value = 0.3132
```

```
alternative hypothesis: true p is less than 0.035
```

```
data: 15 out of 500, null probability 0.035
```

```
X-squared = 0.23686, df = 1, p-value = 0.6868
```

```
alternative hypothesis: true p is greater than 0.035
```

**Задача 4.** За да се сравнят две марки автомобилни гуми е направен експеримент с по 48 гуми от всяка марка. За всяка от гумите е записано изминатото разстояние до износване. Данните за марка 1 са записани във вектор  $x$ , а данните за марка 2 – във вектор  $y$ .

Напишете код на R, който бихте използвали, за да отговорите на въпроса: *Може ли да се твърди, че средното изминато разстояние до износване е по-голямо при гумите от марка 1?*

Отговорете на въпроса с помощта на един от резултатите по-долу. Само един от резултатите е верен – посочете кой. Обосновете отговора си. Формулирайте хипотезите и дефинирайте използваните означения.

```
data: x and y
t = 1.7013, df = 93.933, p-value = 0.9539
alternative hypothesis: true difference in means is less than 0
mean of x mean of y
39030.99 36947.28
```

```
data: x and y
t = 1.7013, df = 93.933, p-value = 0.04609
alternative hypothesis: true difference in means is greater than 0
mean of x mean of y
39030.99 36947.28
```

**Задача 5.** Имаме таблица с оценките на 80 студенти от един поток по *Алгебра 1* и по *Алгебра 2*. Оценките по *Алгебра 1* са записани във вектор  $a1$ , а оценките по *Алгебра 2* – във вектор  $a2$ .

Напишете код на R, който бихте използвали, за да отговорите на въпроса: *Може ли да се твърди, че в средно оценките по Алгебра 1 са по-високи от тези по Алгебра 2?*

Отговорете на въпроса с помощта на един от резултатите по-долу. Само един от резултатите е верен – посочете кой. Обосновете отговора си. Формулирайте хипотезите и дефинирайте използваните означения.

Paired t-test

```
data: a1 and a2
t = 4.4441, df = 79, p-value = 1.422e-05
alternative hypothesis: true difference in means is greater than 0
sample estimates:
mean of the differences
0.2
```

Welch Two Sample t-test

```
data: a1 and a2
t = 1.1073, df = 155.37, p-value = 0.1349
alternative hypothesis: true difference in means is greater than 0
sample estimates:
mean of x mean of y
4.7625 4.5625
```

**Задача 6.** Във файла `students.csv` има данни за студентите, записали курса *Приложна статистика* в дадена година. Файлът съдържа следните променливи:

`grade`    оценка на редовната сесия;  
`study.hours`    брой часове, отделени за подготовка за изпита;  
`lectures`    брой посетени лекции през семестъра.

Използвайте следния резултат от R за построен линеен модел.

Call:

```
lm(formula = grade ~ study.hours)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.01617	0.39998	2.541	0.0134
study.hours	0.15806	0.01863	8.484	2.9e-12

Multiple R-squared: 0.5142, Adjusted R-squared: 0.507

```
> cor(students)
```

	study.hours	lectures	grade
study.hours	1.0000000	0.1441023	0.7170714
lectures	0.1441023	1.0000000	0.4236842
grade	0.7170714	0.4236842	1.0000000

- Напишете оцененото регресионно уравнение. Коя променлива е предиктор (predictor) и коя е отклик (response)?
- Интерпретирайте оценения коефициент за предиктора.
- Може ли да се твърди, че има линейна връзка между оценката и часовете, отделени за подготовка?
- Даден е и резултат от R за друг линеен модел:

Call:

```
lm(formula = grade ~ study.hours + lectures)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.54255	0.37352	1.453	0.151
study.hours	0.14767	0.01680	8.791	9.06e-13
lectures	0.09432	0.02197	4.293	5.81e-05

Multiple R-squared: 0.619, Adjusted R-squared: 0.6076

Напишете оцененото регресионно уравнение. Интерпретирайте коефициентите пред предикторите.

**Задача 7.** Данните `exams` съдържат оценките по ЕАИ и ДАА на студентите от специалности КН и Информатика в дадена учебна година. По-долу са първите четири реда от данните.

fn	spec	EAI	DAA
44005	Inf	5	5
44007	Inf	6	6
80002	KN	5	4
80003	KN	4	4

Напишете командите на R, които извеждат:

- Факултетните номера на студентите, които имат оценка 4 или повече по ЕАИ;
- Броя на студентите от специалност КН, които имат оценка 5 по ДАА;
- Средната оценка по ЕАИ на студентите от специалност КН;
- Медианата на оценка по ДАА на студентите, които има оценка 5 по ЕАИ;
- Походяща графика за успеха по ДАА в зависимост от специалността.

**Задача 8.** Хвърляме зар 4 пъти. Напишете код на R за симулация, с която се намира приближение на вероятността на събитието *„сумата от точките на първите две хвърляния е по-голяма от сумата на следващите две хвърляния“*.