

**Зад.1** Броят на починалите от корона вирус в България в последните две седмици на ноември, разпределени по дни е следния:

Ден от седмицата	П	В	С	Ч	П	С	Н
Починали	125	410	310	300	318	298	148

Една и съща ли е смъртността във всеки ден от седмицата?

С  $p_i$  ще означим вероятността човек да почине в  $i$ -тия ден от седмицата. Проверяваме хипотезата:

$$H_0 : p_1 = p_2 = \dots = p_7 = 1/7$$

$$H_1 : \text{поне едно } p_i \text{ е различно}$$

Нека  $x$  е векторът с броя на починалите. Проверката, че е равномерно разпределен се извършва с Хи-квадрат тест.

```
> chisq.test( x )
```

Chi-squared test for given probabilities

```
data: x
```

```
X-squared = 223.84, df = 6, p-value < 2.2e-16
```

Извод: Полученото p-val е пренебрежимо малко, тогава отхвърляме хипотеза  $H_0$  и приемаме алтернативата  $H_1$ . Отчетената смъртност не е една и съща във всеки ден от седмицата.

**Зад.2** В променливата `pi2000` са първи 2000 цифри на числото  $\pi$ . Разгледайте първите 200 цифри. Можем ли да приемем, че всяка цифра се среща с една и съща вероятност?

В началото ще изброим цифрите в първите двеста знака на  $\pi$

```
> t = table( pi2000[ 1 : 200 ] )
```

```
 0  1  2  3  4  5  6  7  8  9
19 20 24 20 22 20 15 12 25 23
```

Проверяваме хипотеза

$$H_0 : p_0 = p_1 = \dots = p_9 = 1/10$$

$$H_1 : \exists k : p_k \neq 1/10$$

```
> chisq.test( t )
```

Chi-squared test for given probabilities

```
data: t
```

```
X-squared = 7.2, df = 9, p-value = 0.6163
```

Извод: При  $p\text{-val} = 0.6$  нямаме основание за отхвърляне на хипотеза  $H_0$ . Вероятността за всяка цифра е една и съща.

**Зад.3** Честотата на срещането на буквите в английски език е както следва:

Буква	<i>E</i>	<i>T</i>	<i>A</i>	<i>O</i>	<i>I</i>	<i>N</i>	...
Честота	12,7%	9.56%	8.17%	7.51%	6.97%	6.75	48.34%

Анализиран е текст от 1036 букви и се оказва, че в него броят срещания на тези букви е съответно

Буква	<i>E</i>	<i>T</i>	<i>A</i>	<i>O</i>	<i>I</i>	<i>N</i>
Честота	102	108	90	95	82	40

На английски език ли е текстът?

Знаем честотите на срещане на основните букви в английския език. Това са теоретичните вероятности, ще ги запишем във вектор 'prob'. В  $x$  са честотите в конкретния текст.

```
> prob
```

```
[1] 0.1270 0.0956 0.0817 0.0751 0.0697 0.0675 0.4834
```

```
> x
```

```
102 108 90 95 82 40 519
```

Проверяваме хипотеза, че теоретичните и емперичните вероятности съвпадат, т.е :

$H_0 : prob[k] = x[k]/1036$

$H_1 : \neq$

```
> chisq.test(x, p = prob)
```

Chi-squared test for given probabilities

data: x

X-squared = 26.396, df = 6, p-value = 0.0001878

Извод: p-val е малко отхвърляме хипотезата. Текстът не е на английски, тъй като честотите на буквите не съвпадат.

**Зад.4** В таблицата са дадени данни за пострадалите при катастрофи пътници в зависимост от поставянето на предпазен колан.

	Наранявания			
	Без	Леки	Средни	Тежки
С колан	12813	647	359	42
Без колан	65963	4000	2642	303

Можем ли да твърдим, че предпазните колани намаляват вероятността за поражения на водачите.

Трябва да проверим хипотеза, че носенето на колан и нараняванията са независими, т.е.

$H_0$  : Случайните величини са независими

$H_1$  : Зависими са

Когато данните, който се подават на функцията “chisq.test” са в матрица, тя автоматично извършва тест за независимост на случайните величини по ред и по стълб.

```
> x = scan( )
[1] 12813  647  359  42  65963  4000  2642  303

> m = matrix( x, nrow = 2, byrow = T)

> chisq.test(m)

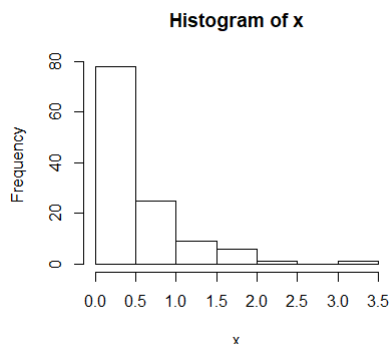
Pearson's Chi-squared test
```

```
data: m
X-squared = 59.224, df = 3, p-value = 8.61e-13
```

Извод: Отхвърляме хипотезата, не са независими, т.е. поставянето на колан влияе на нараняванията при произшествие.

**Зад.6** Разгледайте данните във файл ‘data.txt’. Можем ли да приемем, че те са експоненциално разпределени с параметър  $\lambda = 2$ ?

```
> data = read.csv('data.txt')
> x = data $ x
> hist( x )
```



Хистограмата има вид на експоненциално разпределена. За да проверим хипотезата

$$H_0 : X \in Exp(2)$$

$$H_1 : \quad \notin$$

ще трябва да разделим реалната права на интервали, да пресметнем теоретичната вероятност за попадане в тези интервали, ако е изпълнена хипотезата  $H_0$ . Също така да намерим броя на наблюденията попадащи във всеки интервал и да го сравним с теоретичния, чрез Хи-квадрат критерий.

Данните са сгупчени в началото затова ще направим интервалите там по-къси. Ще използваме командата “cut” за разделяне.

```
> t = table(cut( x, breaks = c( 0, 1/8, 1/4, 1/2, 1, 5)))

(0,0.125]  (0.125,0.25]  (0.25,0.5]  (0.5,1]  (1,5]
      23           22           33           25           17
```

Това са броя наблюдения попадащи в съответния интервал.

За да пресметнем теоретичната вероятност за попадане в интервала ще използваме функцията “pexp” с параметър  $\text{rate} = \lambda = 2$ .

```
> prob[1] = pexp( 1/8, 2) - pexp( 0, 2)
```

```
> prob[2] = pexp( 1/4, 2) - pexp( 1/8, 2)
```

```
> prob[3] = pexp( 1/2, 2) - pexp( 1/4, 2)
```

```
> prob[4] = pexp( 1, 2) - pexp( 1/2, 2)
```

```
> prob[5] = 1 - pexp( 1, 2)
```

```
[1] 0.221 0.172 0.238 0.232 0.135
```

```
> chisq.test( t, p = prob)
```

```
Chi-squared test for given probabilities
```

```
data: t
```

```
X-squared = 1.5608, df = 4, p-value = 0.8158
```

Извод: Приемаме хипотезата. Данните наистина са експоненциално разпределени с параметър  $\lambda = 2$ .