

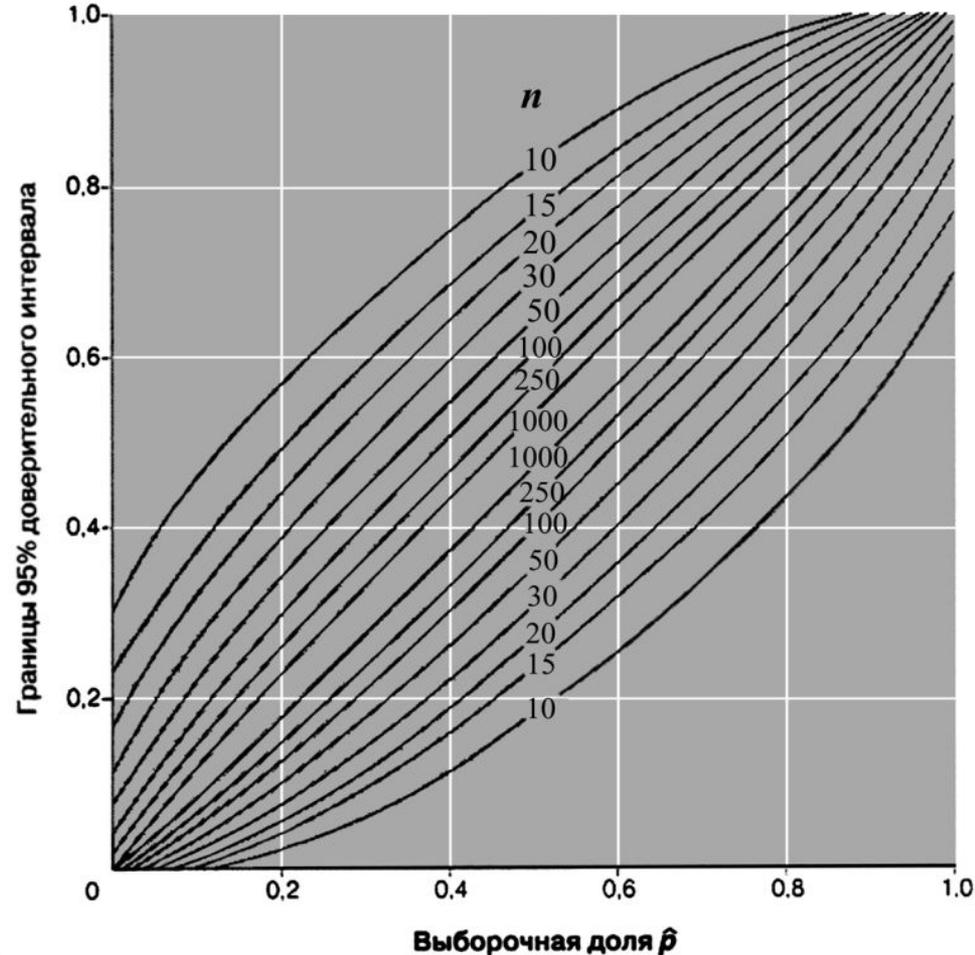
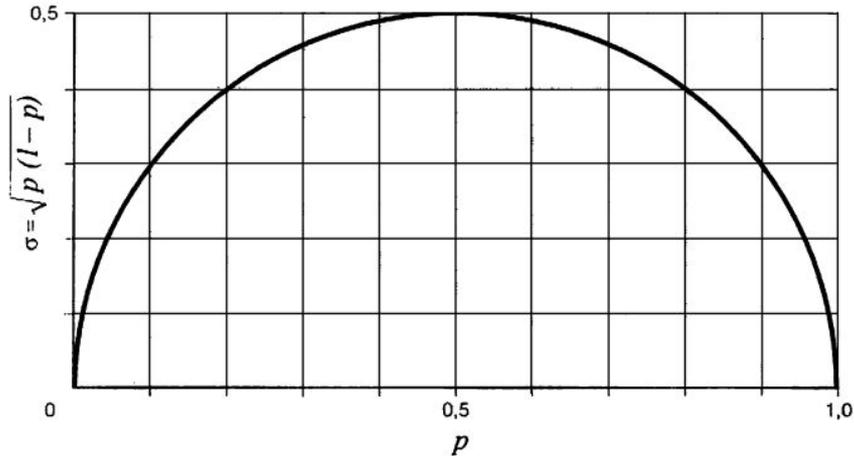
# Непараметрические критерии

Анализ качественных признаков

# Доли

Контроль на 2 из 10 чашек  
УФ на 4 из 10 чашек

Контроль  $p = 2/10 = 0,2$   
УФ  $p = 4/10 = 0,4$



```
c<-binconf(x=2, n=10)  
uv<-binconf(x=4, n=10)
```

```
> c
```

PointEst	Lower	Upper
0.2	0.05668215	0.5098375

```
> uv
```

PointEst	Lower	Upper
0.4	0.1681803	0.6873262

```
> stripchart(c, col="red", ylim=range(0:1), vertical=T, lwd=5)  
> stripchart(uv, vertical=T, lwd=5, add=T)
```

## Доли

Контроль на 2 из 10 чашек  $p = 2/10 = 0,2$   
 УФ на 4 из 10 чашек  $p = 4/10 = 0,4$

```
c<-binconf(x=2, n=10)
```

```
uv<-binconf(x=4, n=10)
```

```
> c
```

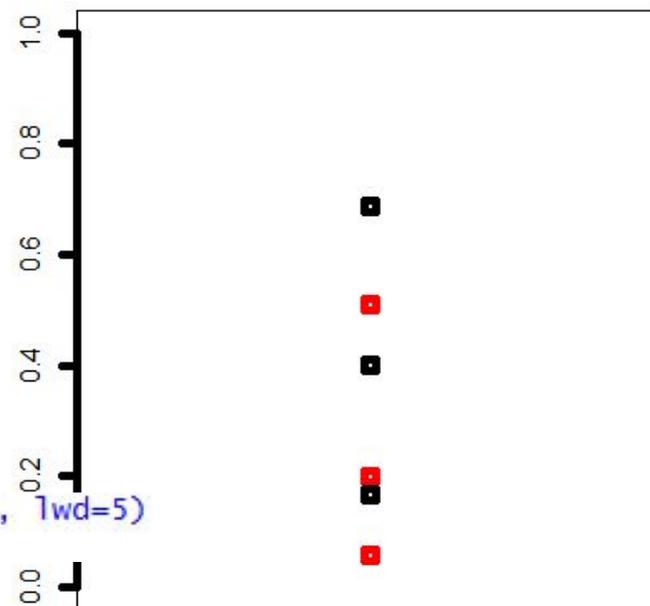
PointEst	Lower	Upper
0.2	0.05668215	0.5098375

```
> uv
```

PointEst	Lower	Upper
0.4	0.1681803	0.6873262

```
> stripchart(c, col="red", ylim=range(0:1), vertical=T, lwd=5)
```

```
> stripchart(uv, vertical=T, lwd=5, add=T)
```



Контроль на 20 из 100 чашек  $p = 20/100 = 0,2$   
 УФ на 40 из 100 чашек  $p = 40/100 = 0,4$

```
> c1<-binconf(x=20, n=100)
```

```
> uv1<-binconf(x=40, n=100)
```

```
> c1
```

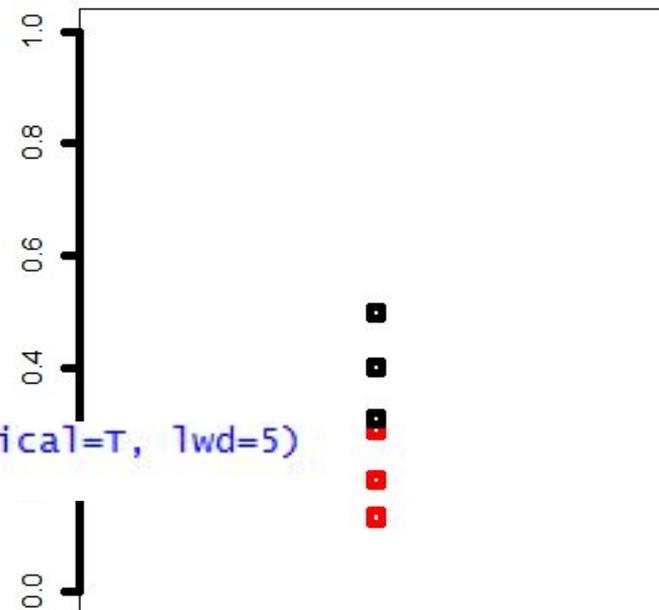
PointEst	Lower	Upper
0.2	0.1333669	0.2888292

```
> uv1
```

PointEst	Lower	Upper
0.4	0.3094013	0.4979974

```
> stripchart(c1, col="red", ylim=range(0:1), vertical=T, lwd=5)
```

```
> stripchart(uv1, vertical=T, lwd=5, add=T)
```



```
> library(Hmisc)
> prop.test(x=c(2,4), n=c(10,10))
```

2-sample test for equality of proportions with continuity correction

```
data: c(2, 4) out of c(10, 10)
X-squared = 0.2381, df = 1, p-value = 0.6256
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.6919928  0.2919928
sample estimates:
prop 1 prop 2
 0.2    0.4
```

warning message:

```
In prop.test(x = c(2, 4), n = c(10, 10)) :
  аппроксимация на основе хи-квадрат может быть неправильной
> prop.test(x=c(20,40), n=c(100,100))
```

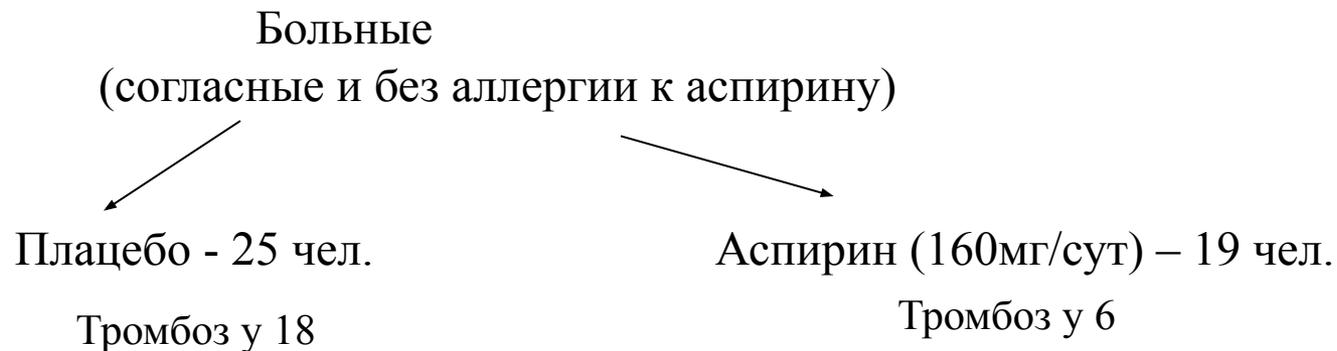
2-sample test for equality of proportions with continuity correction

```
data: c(20, 40) out of c(100, 100)
X-squared = 8.5952, df = 1, p-value = 0.00337
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.33395901 -0.06604099
sample estimates:
prop 1 prop 2
 0.2    0.4
```

## Таблицы сопряженности

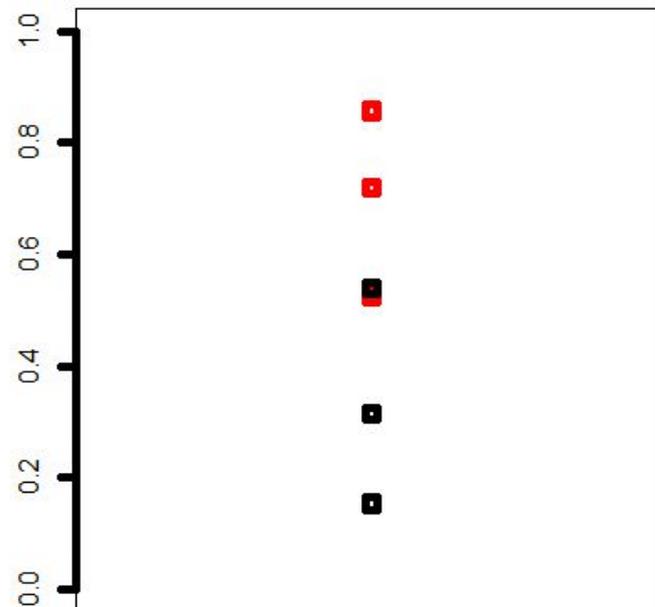
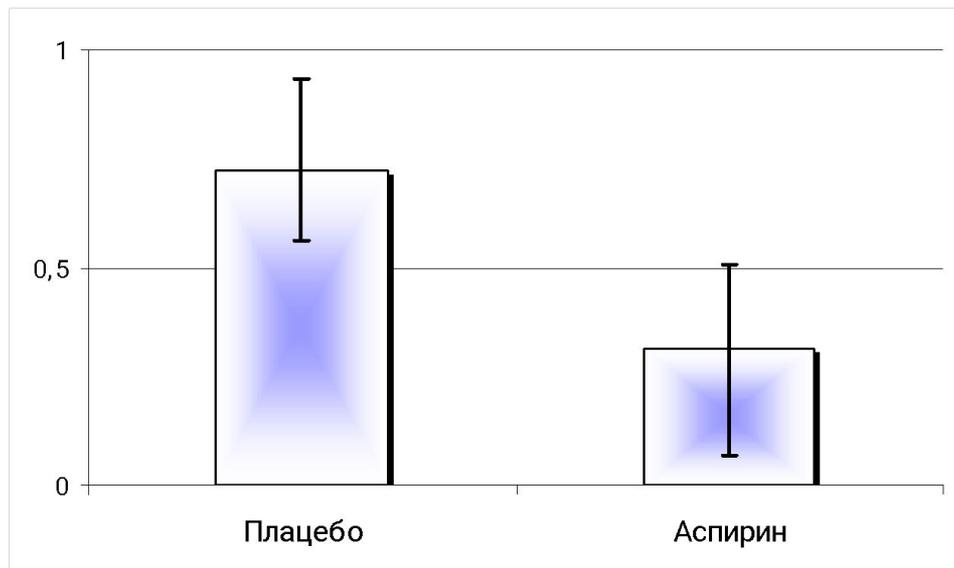
### Тромбоз шунта при гемодиализе

Hanter H.R.et al., Prevention of thrombosis in patients in hemodialysis by low-dose aspirin N.Engl.J.Med., 301:577-579, 1979



Можно ли говорить о статистически значимом различии доли больных с тромбозом, а тем самым об эффективности аспирина?

Оценим долю больных с тромбозами в каждой из групп, используя биномиальное распределение



```
> placebo<-binconf(x=18, n=25)
> aspirin<-binconf(6,19)
> stripchart(placebo, col="red", ylim=range(0:1), vertical=T, lwd=5)
> stripchart(aspirin, vertical=T, lwd=5, add=T)
```

С 95%-ной достоверностью можно утверждать, что аспирин значительно снижает риск возникновения тромба в шунте при гемодиализе

таблица наблюдаемых значений

	тромбоз +	тромбоз -
Плацебо	18	7
Аспирин	6	13

Формулируем  $H_0$  и строим таблицу ожидаемых чисел

	тромбоз +	тромбоз -	
Плацебо	13,63636	11,36364	25
Аспирин	10,36364	8,636364	19
	24	20	44

Как же рассчитать ожидаемые числа?

	тромбоз+	тромбоз-	
плацебо	18	7	25
аспирин	6	13	19
	24	20	44
	0,54545	0,45455	

Доли ожидаемых больных с тромбозами (0,5455) и без тромбозов (0,4545) при условии справедливости нулевой гипотезы

## Критерий $\chi^2$

Критерий  $\chi^2$  не требует никаких предположений относительно параметров совокупности (непараметрический)

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

где  $O$  – наблюдаемое число в клетке таблицы сопряженности,  $E$  – ожидаемое число в той же клетке. Суммирование производится по всем клеткам таблицы

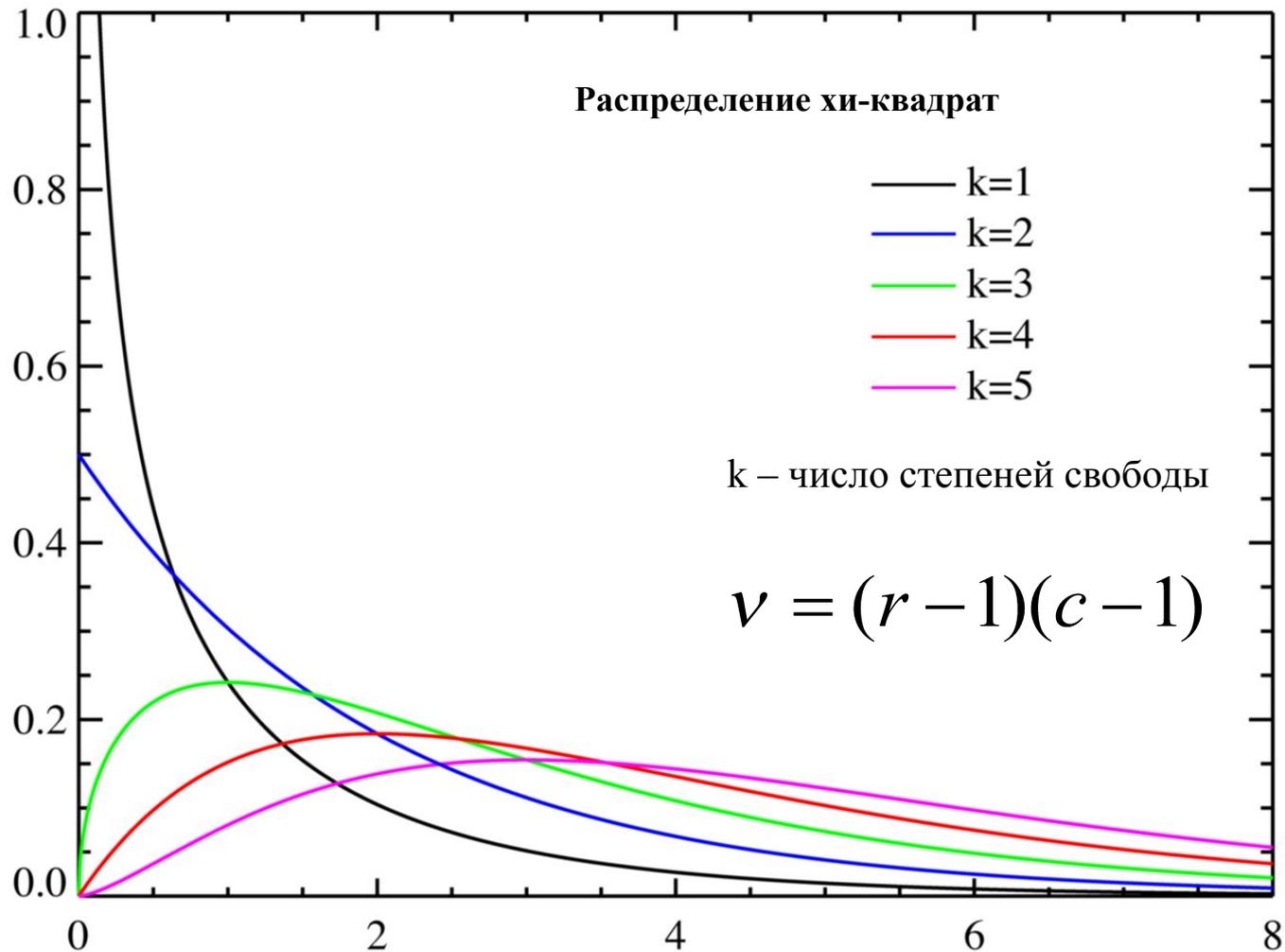
Этот критерий удовлетворяет двум требованиям:

1. измеряет различия
2. учитывает их величину относительно ожидаемых чисел

Применим критерий  $\chi^2$  для данных по тромбозам шунта.

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(18 - 13,64)^2}{13,64} + \frac{(7 - 11,36)^2}{11,36} + \frac{(6 - 10,36)^2}{10,36} + \frac{(13 - 8,64)^2}{8,64} = 7,1$$

Много это или мало?



```
> qchisq(p=0.95, df=1)  
[1] 3.841459
```

$$\chi^2 = \sum \frac{\left( |O - E| - \frac{1}{2} \right)^2}{E}$$

!!! Поправка Йейтса применяется при  $\nu=1$ , т.е.  
для таблиц  $2 \times 2$

$$\chi^2 = \frac{\left( |18 - 13,64| - \frac{1}{2} \right)^2}{13,64} + \frac{\left( |7 - 11,36| - \frac{1}{2} \right)^2}{11,36} + \frac{\left( |6 - 10,36| - \frac{1}{2} \right)^2}{10,36} + \frac{\left( |13 - 8,64| - \frac{1}{2} \right)^2}{8,64} = 5,57$$

Исправленное значение оказалось больше 5,02 – критического значения для 2,5% уровня значимости

Применим поправку для изучения связи между приемом аспирина и тромбозом шунта

$$\chi^2 = \frac{\left(|18 - 13,64| - \frac{1}{2}\right)^2}{13,64} + \frac{\left(|7 - 11,36| - \frac{1}{2}\right)^2}{11,36} + \frac{\left(|6 - 10,36| - \frac{1}{2}\right)^2}{10,36} + \frac{\left(|13 - 8,64| - \frac{1}{2}\right)^2}{8,64} = 5,57$$

Исправленное значение оказалось больше 5,02 – критического значения для 2,5% уровня значимости

```
> tromb<-matrix(c(18,7,6,13), nrow = 2, byrow = TRUE)
> tromb
      [,1] [,2]
[1,]   18   7
[2,]    6  13
> chisq.test(tromb)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: tromb
X-squared = 5.5772, df = 1, p-value = 0.0182
```

```
> chisq.test(cdc$gender,cdc$smoke100)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: cdc$gender and cdc$smoke100
X-squared = 204.6, df = 1, p-value < 2.2e-16
```

```
> chisq.test(cdc$genhlth,cdc$smoke100)
```

Pearson's Chi-squared test

```
data: cdc$genhlth and cdc$smoke100
X-squared = 333.53, df = 4, p-value < 2.2e-16
```

## Точный критерий Фишера (exact Fisher test)

$H_0$  - между лечением и исходом нет никакой связи,  
тогда вероятность получить некоторую таблицу равна

$$P = \frac{R_1!R_2!C_1!C_2!}{N! \cdot O_{11}!O_{12}!O_{21}!O_{22}!}$$

где  $R_1$  и  $R_2$  - суммы по строкам  
 $C_1$  и  $C_2$  - суммы по столбцам  
 $O_1, O_2, O_3, O_4$  - числа в клетках  
 $N$  - общее число наблюдений

$$P = P_{исх} + P_1 + P_2 + \dots + P_i$$

Обозначения, используемые в точном критерии Фишера

			Суммы по строкам
	$O_{11}$	$O_{12}$	$R_1$
	$O_{21}$	$O_{22}$	$R_2$
Суммы по столбцам	$C_1$	$C_2$	$N$

$$P_{исх} \geq P_1$$

$$P_{исх} \geq P_2$$

$$P_{исх} \geq P_i$$

где  $R_1$  и  $R_2$  - суммы по строкам ( число больных, лечившихся 1-м и 2-м способом),  $C_1$  и  $C_2$  - суммы по столбцам(число больных с первым и вторым исходом),  $O_{11}, O_{12}, O_{21}, O_{22}$  – числа в клетках,  $N$  – общее число наблюдений

		нет	есть		
В Е		<b>1</b>	<b>7</b>	<b>8</b>	
		<b>3</b>	<b>2</b>	<b>5</b>	6766,99
		<b>4</b>	<b>9</b>	<b>13</b>	60480

**0,11888112**

0	8
4	1

967680

**0,006993007**

2	6
2	3

17280

0,391608392

3	5
1	4

17280

0,391608392

4	4
0	5

69120

0,097902098

**0,118881119**

	есть рецидив		нет рецидива			
Ампициллин	20	7	27			
Цефалексин	14	2	16	4,04002E+32		
	34	9	43	2,13793E+33	0,188968689	
	19	8	27			
	15	1	16			
	34	9	43	6,41379E+33	0,062989563	
	18	9	27			
	16	0	16			
	34	9	43	4,86098E+34	0,008311123	

21	6	27				
13	3	16				
34	9	43	1,37438E +33	0,293951294		
22	5	27				
12	4	16				
34	9	43	1,55059E +33	0,260547738		
23	4	27				
11	5	16				
34	9	43	2,97196E +33	0,13593795		
24	3	27				
10	6	16				
34	9	43	9,72641E +33	0,041536596		
25	2	27				
9	7	16				
34	9	43	5,67374E +34	0,007120559		
26	1	27				
8	8	16				
34	9	43	6,55632E +35	0,000616202		
27	0	27				
7	9	16				
34	9	43	1,99148E +37	2,02865E -05	P=	0,246573

```
> table<-matrix(c(1,7,3,2), nrow = 2, byrow = TRUE)
> table
      [,1] [,2]
[1,]    1    7
[2,]    3    2
> fisher.test(table)
```

Fisher's Exact Test for Count Data

```
data: table
p-value = 0.2168
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.00160284 2.38205321
sample estimates:
odds ratio
 0.1195177
```

```
> j<-matrix(c(1,10,8,4), nrow = 2, byrow = TRUE)
> fisher.test(j)
```

Fisher's Exact Test for Count Data

```
data: j
p-value = 0.009423
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.001034782 0.656954980
sample estimates:
odds ratio
 0.05851868
```

	Рецидив	
	Есть	Нет
Ампициллин	20	7
Триметоприм/сульфаметоксазол	24	21
Цефалексин	14	2

Есть ли основания говорить о разной эффективности препаратов?

```
> A<-matrix(c(20,24,14,7,21,2), nrow = 3, byrow = TRUE)
```

```
> A
```

```
  [,1] [,2]
[1,]  20  24
[2,]  14   7
[3,]  21   2
```

```
> chisq.test(A)
```

Pearson's Chi-squared test

```
data: A
```

```
Chi-squared = 13.752, df = 2, p-value = 0.001032
```

```
> fisher.test(A)
```

Fisher's Exact Test for Count Data

```
data: A
```

```
p-value = 0.0006512
```

```
alternative hypothesis: two.sided
```

```
> ?fisher.test()
```

```
> fisher.test(A, simulate.p.value = T, B=500)
```

Fisher's Exact Test for Count Data with simulated p-value (based on 500 replicates)

```
data: A
```

```
p-value = 0.001996
```

```
alternative hypothesis: two.sided
```

The screenshot shows the R help page for the `fisher.test` function. The title bar indicates the topic is "R: Fisher's Exact Test for Count Data". The help text includes the following arguments and their descriptions:

- `control`: a list with named components for low level algorithm control. At present the only one used is "mult", a default 30 used only for larger than 2 by 2 tables. This says how many times as much space should be a keys: see file 'fexact.c' in the sources of this package.
- `or`: the hypothesized odds ratio. Only used in the 2 by 2 case.
- `alternative`: indicates the alternative hypothesis and must be one of "two.sided", "greater" or "less". You ca letter. Only used in the 2 by 2 case.
- `conf.int`: logical indicating if a confidence interval for the odds ratio in a 2 by 2 table should be computed (and return
- `conf.level`: confidence level for the returned confidence interval. Only used in the 2 by 2 case and if `conf.int = T`
- `simulate.p.value`: a logical indicating whether to compute p-values by Monte Carlo simulation, in larger than 2 by 2 tables.
- `B`: an integer specifying the number of replicates used in the Monte Carlo test.

The "Arguments" section is partially visible, showing the start of the descriptions for `x`, `y`, `workspace`, and `hybrid`.

# критерий Мак-Нимара

		р-ция на ДНХБ			
		+	-		
р-ция на кртоновое масло	+	81	48	129	77,54913 51,45087
	-	23	21	44	26,45087 17,54913
		104	69	173	
		0,601156	0,398844		

1,106911  
3,841455

		р-ция на ДНХБ			
		+	-		
р-ция на кртоновое масло	+	81	48	35,5	4,056338
	-	23	21	35,5	4,056338

8,112676  
3,841455  
6,634891

Имеем:

$$\chi^2 = \sum \frac{(|O - E| - 0,5)^2}{E} = \frac{(|23 - 35,5| - 0,5)^2}{35,5} + \frac{(|48 - 35,5| - 0,5)^2}{35,5} = 8,11$$

```
> mn<-matrix(c(81,48,23,21), nrow = 2, byrow = TRUE)
> mn
      [,1] [,2]
[1,]   81   48
[2,]   23   21
> mcnemar.test(mn)
```

McNemar's Chi-squared test with continuity correction

```
data: mn
McNemar's chi-squared = 8.1127, df = 1, p-value = 0.004396
```

```
> mcnemar.test(mn, correct = FALSE)
```

McNemar's Chi-squared test

```
data: mn
McNemar's chi-squared = 8.8028, df = 1, p-value = 0.003008
```