

χ^2 -тесты

Критерий согласия Пирсона рассчитывается по формуле:

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

Здесь:

- O — наблюдаемое значение;
- E — ожидаемое значение.

$$\chi_n^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Таким образом, КСП — это разница между наблюдаемым и ожидаемым значением, возведенная в квадрат (нам важно не направление отличий, а только факт их наличия) и нормированная с помощью деления на ожидаемое значение (чтобы слишком быстро не росла).

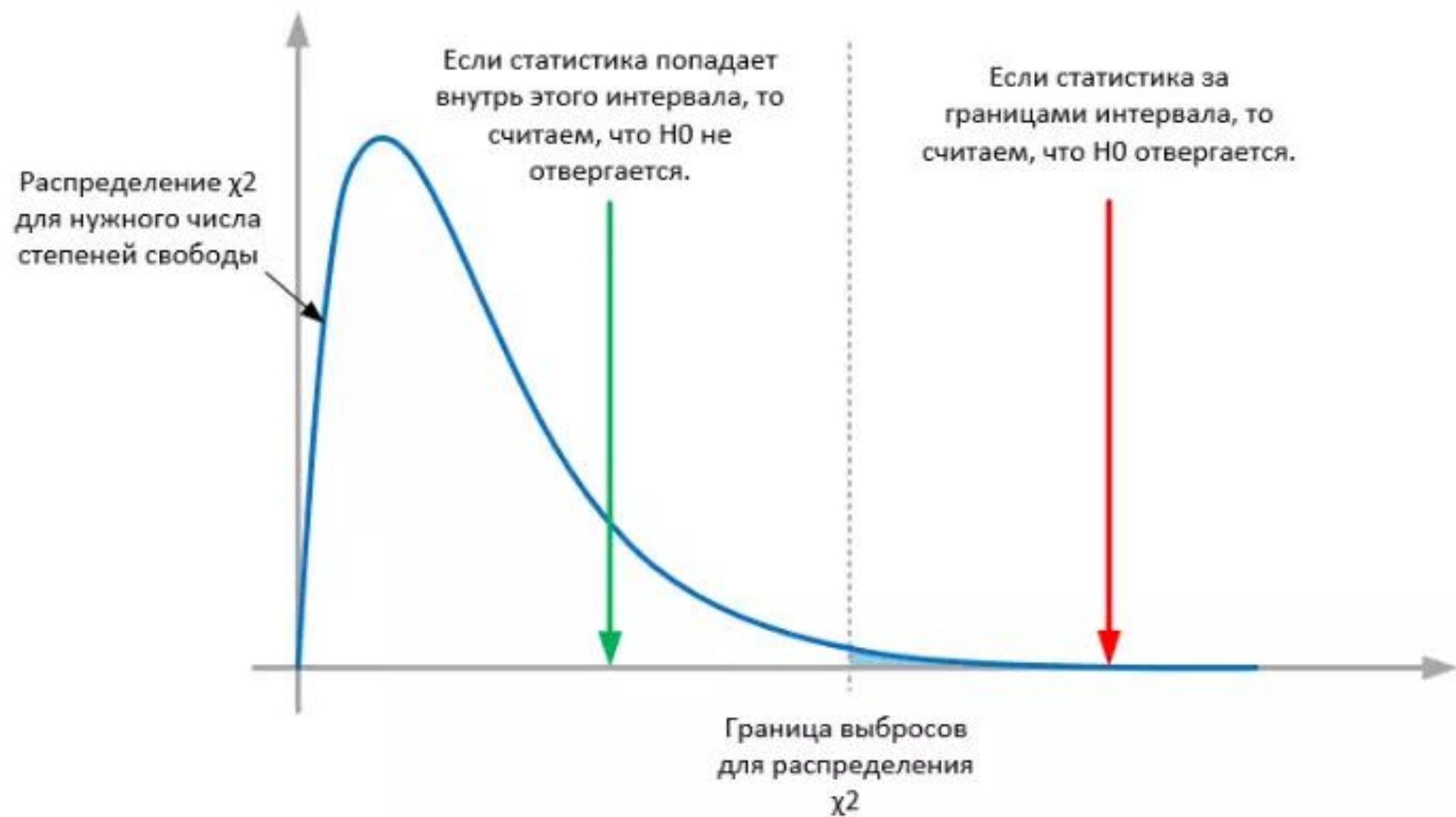
• Тест на независимость (он же test of independence/association) — непараметрический, одновыборочный тест, который проверяет наличие связи между двумя категориальными переменными. В Python реализован функцией `scipy.stats.chi2_contingency`.

Получается, что мы можем провести статистический тест с гипотезами:

H0: между наблюдаемым распределением и эталонным распределением нет различий;

H1: между наблюдаемым распределением и эталонным распределением есть различия.

Если различий нет, то КСП будет стремиться к нулю. В противном случае она окажется за пределами интервала наиболее вероятных значений:



Тесты с КСП, таким образом, непараметрические (поскольку не оценивают никакие из параметров распределений) и односторонние (так как статистика всегда положительная за счет возведения в квадрат).

Поскольку тест непараметрический, **мы можем напрямую сравнивать полученную статистику с эталонным распределением**

<https://www.real-statistics.com/chi-square-and-f-distributions/goodness-of-fit/>

		Observed Right foot longer	Observed Left foot longer	Observed Both feet same	Total	Probability
Observed	Right hand longer	11	3	8	22	22%
Observed	Left hand longer	2	9	14	25	25%
Observed	Both hands same	12	13	28	53	53%
Total		25	25	50	100	
Probability		25%	25%	50%		

		Observed Right foot longer	Observed Left foot longer	Observed Both feet same	Total	Probability
Observed	Right hand longer	11	3	8	22	22%
Expected		$= 25\% * 22\% * 100 = 5.5$	5.5	11		
Observed	Left hand longer	2	9	14	25	25%
Expected		6.25	6.25	12.5		
Observed	Both hands same	12	13	28	53	53%
Expected		13.25	13.25	26.5		
Total		25	25	50	100	
Probability		25%	25%	50%		

	Right foot longer	Left foot longer	Both feet same
Right hand longer	$= (11 - 5.5)^2 / 5.5 = 5.5$	1.136	0.818
Left hand longer	2.890	1.210	0.180
Both hands same	0.118	0.005	0.085
Sum	8.508	2.351	1.083
Statistic	11.942		

```
: import pandas as pd
  from scipy.stats import chi2_contingency

: contingency_table = pd.DataFrame([[11, 3, 8], [2, 9, 14], [12, 13, 28]])

: chi2_contingency(contingency_table)[0:3]

: (11.942092624356777, 0.0177871146098607, 4)
```

H0: между наблюдаемым распределением и эталонным распределением нет различий;

H1: между наблюдаемым распределением и эталонным распределением есть различия.