

# Медиана (Median)

Синонимы: Медиана статистическая, Непараметрическая средняя

Разделы: Метрики

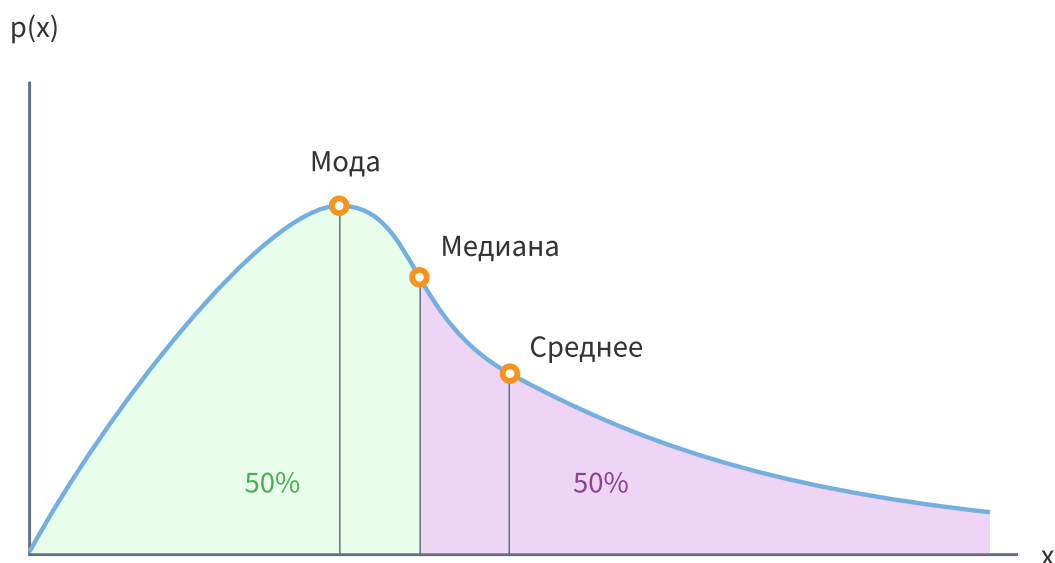
Медиана (от лат. *mediāna* — середина, срединный) — статистическая характеристика распределения случайной величины. Наряду с модой и математическим ожиданием медиана определяет центр распределения, т.е. область, в которой его значения наиболее вероятны.

На практике это означает, что если признак некоторого объекта принимает значение, близкое к медианному, то оно является типичным для данной выборки. Если же значение оказывается далеким от медианы (как говорят, расположено в «хвосте» распределения), то у аналитика возникает повод задуматься о том, какими закономерностями исследуемого процесса вызвано такое отклонение и не является ли оно выбросом или аномальным значением.

Если определять медиану в контексте распределения случайной величины, то она представляет собой значение, которое делит площадь под кривой функции распределения на две равные части. Это означает, что суммарные вероятности значений слева и справа от медианы равны, т.е.

$$P(x < M_e) = P(x > M_e),$$

где  $x$  — случайная величина,  $M_e$  — медиана.



Если распределение симметричное, то мода, математическое ожидание и медиана совпадают. Если симметрия распределения нарушается, то медиана смещается от моды в сторону более длинного «хвоста». При этом медиана всегда меньше математического ожидания.

Медиана относится к классу непараметрических статистик, поскольку для ее определения не требуется каких-либо предположений о распределении случайной величины, для которой она вычисляется. Медиана имеет смысл только для упорядочиваемых данных, значения которых могут быть ранжированы.

Если случайная величина  $x$  задана не функцией распределения вероятности  $P(x)$ , а набором выборочных значений признака (дискретным рядом), то медиана вычисляется следующим образом:

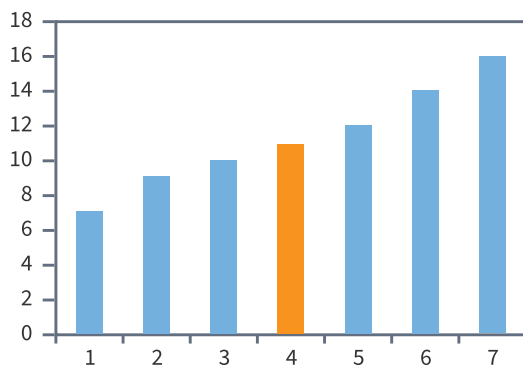
1. Упорядочить (расположить в порядке неубывания) значения набора данных:

$$x_1 \leq x_2 \leq \dots \leq x_{(n-1)} \leq x_n, \text{ где } n \text{ — количество элементов выборки.}$$

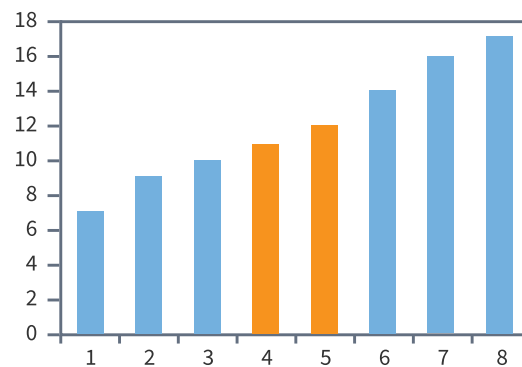
2. Определить номер медианного элемента выборки по формуле  $N_{ME} = (n + 1)/2$ .

Полученное значение будет точно показывать, где находится элемент ряда, значение которого является медианным.

Например, пусть задана последовательность [7, 14, 10, 12, 16, 9, 11]. Для вычисления медианы упорядочим ее: [7, 9, 10, 11, 12, 14, 16]. Тогда номер медианного элемента  $N_{ME} = (7 + 1)/2 = 4$ , следовательно, медиана равна 11.



$n=7, N_e=4, Me=11$



$n=8, N_e=4.5,$   
 $Me=(11+12)/2=11.5$

Если  $n$  четное, то номер медианного элемента оказывается дробным. Например, при  $n = 8$  номер медианного элемента окажется  $N_{ME} = (8 + 1)/2 = 4.5$ . В этом случае берут значения 4-го и 5-го элементов и вычисляют их среднее. Т.е. медиана в случае нечетного размера выборки определяется как среднее двух срединных значений ряда ее ранжированных элементов. Легко увидеть, что номер первого усредняемого значения при четном  $n$  будет  $N_{ME} + 0.5$ , а второго  $N_{ME} - 0.5$ .

В анализе данных медиана может использоваться как альтернатива среднего значения, устойчивая к выбросам и аномальным значениям. Действительно, из-за ранжирования аномально большие и аномально низкие значения всегда будут оказываться в начале или конце ряда и никогда не станут медианными.

Медиана является порядковой статистикой распределения и может быть выражена через другие порядковые статистики (т.е. получаемые при ранжировании). Она является 50-м перцентилем, 0,5-квантилем или вторым квартилем выборки или распределения.

В Logiном существует специализированный обработчик группировка, который позволяет объединять записи избранных полей в группы, а для оставшихся полей вычислять статистические показатели (медиану, сумму, среднее, минимум и т.д.). А также визуализаторы статистика и качество данных, в перечень показателей которых включена медиана.