

Медиана (статистика)

Материал из Википедии — свободной энциклопедии

Меди́ана (от лат. *mediāna* — середина) в математической статистике — число, характеризующее выборку (например, набор чисел). Если все элементы выборки различны, то медиана — это такое число выборки, что ровно половина из элементов выборки больше него, а другая половина меньше него. В более общем случае медиану можно найти, упорядочив элементы выборки по возрастанию или убыванию и взяв средний элемент. Например, выборка {11, 9, 3, 5, 5} после упорядочивания превращается в {3, 5, 5, 9, 11} и её медианой является число 5. Если в выборке чётное число элементов, медиана может быть не определена однозначно: для числовых данных чаще всего используют полусумму двух соседних значений (то есть медиану набора {1, 3, 5, 7} принимают равной 4), подробнее см. ниже.

Также медиану можно определить для случайных величин: в этом случае она делит пополам распределение. Грубо говоря, медианой случайной величины является такое число, что вероятность получить значение случайной величины справа от него равна вероятности получить значение слева от него (и они обе равны 1/2); более точное определение см. ниже.

Можно также сказать, что медиана является 50-м персентилем, 0,5-квантилем или вторым квартилем выборки или распределения.

Содержание

Свойства медианы для случайных величин

Пример использования

Неуникальность значения

Примечания

Свойства медианы для случайных величин

Если распределение непрерывно, то медиана является одним из решений уравнения

$$F(x) = 0.5$$

Если распределение является непрерывной строго возрастающей функцией, то решение уравнения однозначно. Если распределение имеет разрывы, то медиана может совпадать с минимальным или максимальным (крайним) возможным значением случайной величины, что противоречит «геометрическому» пониманию этого термина.

Медиана является важной характеристикой распределения случайной величины и, так же как математическое ожидание, может быть использована для центрирования распределения. Поскольку оценки медианы более робастны, её оценивание может быть более предпочтительным для распределений с т. н. *тяжёлыми хвостами*. Однако о преимуществах

оценивания медианы по сравнению с математическим ожиданием можно говорить только в случае, если эти характеристики у распределения совпадают, в частности, для симметричных функций плотности распределения вероятностей.

Медиана определяется для всех распределений, а в случае неоднозначности, естественным образом доопределяется, в то время как математическое ожидание может быть не определено (например, у распределения Коши).

Пример использования

Предположим, что в одной комнате оказалось 19 бедняков и один миллионер. У каждого бедняка есть 5 ₽, а у миллионера — 1 млн ₽ (10⁶). В сумме получается 1 000 095 ₽. Если мы разделим деньги равными долями на 20 человек, то получим 50 004,75 ₽. Это будет среднее арифметическое значения суммы денег, которая была у всех 20 человек в этой комнате.

Медиана в этом случае будет равна 5 ₽ (полусумма десятого и одиннадцатого, *срединных* значений ранжированного ряда). Можно интерпретировать это следующим образом. Разделив всю компанию на две равные группы по 10 человек, мы можем утверждать, что в первой группе у каждого не больше 5 ₽, во второй же не меньше 5 ₽. В общем случае можно сказать, что медиана это то, сколько принёс с собой «средний» человек. Наоборот, среднее арифметическое — неподходящая характеристика, так как оно значительно превышает сумму наличных, имеющуюся у среднего человека.

Неуникальность значения

Если имеется чётное количество случаев и два средних значения различаются, то медианой, по определению, может служить любое число между ними (например, в выборке {1, 3, 5, 7} медианой может служить любое число из интервала (3,5)). На практике в этом случае чаще всего используют среднее арифметическое двух средних значений (в примере выше это число (3+5)/2=4). Для выборок с чётным числом элементов можно также ввести понятие «нижней медианы» (элемент с номером $n/2$ в упорядоченном ряду из n элементов; в примере выше это число 3) и «верхней медианы» (элемент с номером $(n+2)/2$; в примере выше это число 5)^[1]. Эти понятия определены не только для числовых данных, но и для любой порядковой шкалы.

Примечания

- ↑ *Кормен, Томас Х., Лейзерсон, Чарльз И., Ривест Рональ Л., Штайн, Клиффорд*. Алгоритмы. Построение и анализ (неопр.). — 2-е издание. — М.: Издательский дом «Вильямс», 2005. — С. 240. — 1296 с.

Источник — https://ru.wikipedia.org/w/index.php?title=Медиана_(статистика)&oldid=104229213

Эта страница в последний раз была отредактирована 29 декабря 2019 в 12:56.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.