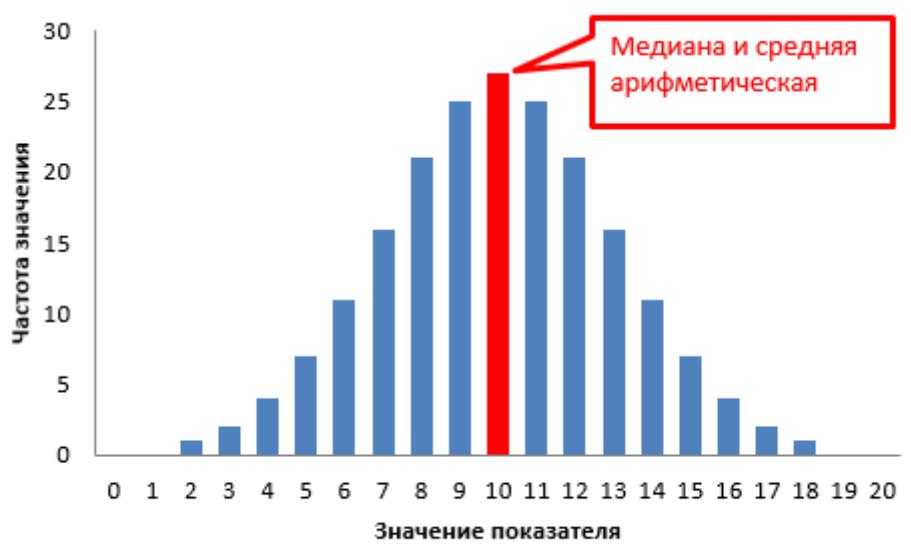


Медиана в статистике

Описание данных

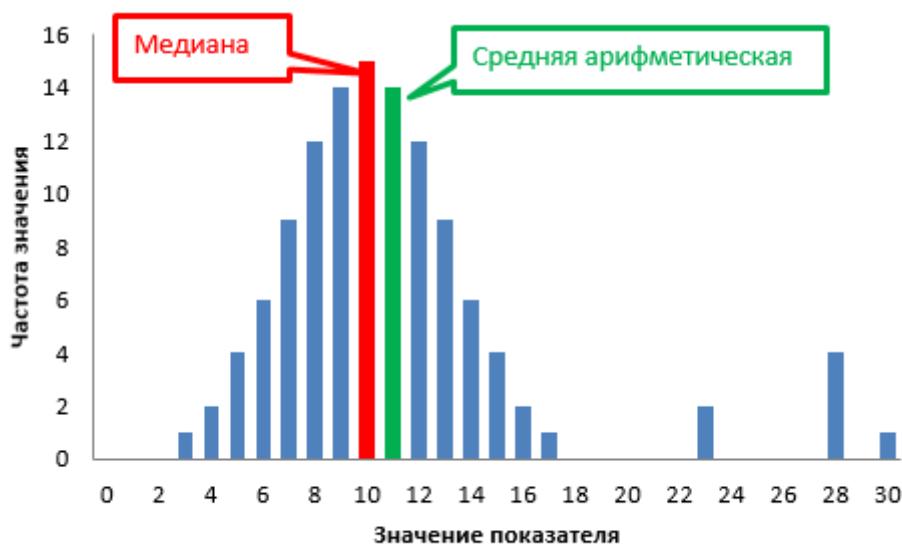
Центральную тенденцию данных можно рассматривать не только, как значение с нулевым суммарным отклонением (среднее арифметическое) или максимальную частоту (мода), но и как некоторую отметку (значение в совокупности), делящую ранжированные данные (отсортированные по возрастанию или убыванию) на две равные части. Половина исходных данных меньше этой отметки, а половина – больше. Это и есть **медиана**.

Итак, медиана в статистике – это уровень показателя, который делит набор данных на две равные половины. Значения в одной половине меньше, а в другой больше медианы. В качестве примера обратимся к набору [нормально распределенных](#) случайных чисел.



Очевидно, что при симметричном распределении середина, делящая совокупность пополам, будет находиться в самом центре – там же, где средняя арифметическая (и мода). Это, так сказать, идеальная ситуация, когда мода, медиана и средняя арифметическая совпадают и все их свойства приходятся на одну точку – максимальная частота, деление пополам, нулевая сумма отклонений – все в одном месте. Однако, жизнь не так симметрична, как нормальное распределение.

Допустим, мы имеем дело с техническими замерами отклонений от ожидаемой величины чего-нибудь (содержания элементов, расстояния, уровня, массы и т.д. и т.п.). Если все ОК, то отклонения, скорее всего, будут распределены по закону, близкому к нормальному, примерно, как на рисунке выше. Но



Медиана выборки – это альтернатива средней арифметической, т.к. она устойчива к аномальным отклонениям (выбросам).

Математическим **свойством медианы** является то, что сумма абсолютных (по модулю) отклонений от медианного значения дает минимально возможное значение, если сравнивать с отклонениями от любой другой величины. Даже меньше, чем от средней арифметической, о как! Данный факт находит свое применение, например, при решении транспортных задач, когда нужно рассчитать место строительства объектов около дороги таким образом, чтобы суммарная длина рейсов до него из разных мест была минимальной (остановки, заправки, склады и т.д. и т.п.).

Формула медианы

Формула медианы в статистике для дискретных данных чем-то напоминает формулу моды. А именно тем, что формулы как таковой нет. Медианное значение выбирают из имеющихся данных и только, если это невозможно, проводят несложный расчет.

Первым делом данные ранжируют (сортируют по убыванию). Далее есть два варианта. Если количество значений нечетно, то медиана будет соответствовать центральному значению ряда, номер которого можно определить по формуле:

$$N_{Me} = \frac{N + 1}{2}$$

где

N_{Me} – номер значения, соответствующего медиане,



$$Me = \frac{x_{\frac{N+1}{2}}}{2}$$

Это первый вариант, когда в данных есть одно центральное значение. Второй вариант наступает тогда, когда количество данных четно, то есть вместо одного есть два центральных значения. Выход прост: берется средняя арифметическая из двух центральных значений:

$$Me = \frac{\frac{x_N}{2} + \frac{x_{\frac{N}{2}+1}}{2}}{2}$$

В *интервальных* данных выбрать конкретное значение не представляется возможным. Медиану рассчитывают по определенному правилу.

Для начала (после ранжирования данных) находят **медианный интервал**. Это такой интервал, через который проходит искомое медианное значение. Определяется с помощью накопленной доли ранжированных интервалов. Где накопленная доля впервые перевалила через 50% всех значений, там и медианный интервал.

Не знаю, кто придумал формулу медианы, но исходили явно из того предположения, что распределение данных внутри медианного интервала равномерное (т.е. 30% ширины интервала – это 30% значений, 80% ширины – 80% значений и т.д.). Отсюда, зная количество значений от начала медианного интервала до 50% всех значений совокупности (разница между половиной количества всех значений и накопленной частотой предмедианного интервала), можно найти, какую долю они занимают во всем медианном интервале. Вот эта доля аккурат переносится на ширину медианного интервала, указывая на конкретное значение, именуемое впоследствии медианой.

Обратимся к наглядной схеме.





Немного громоздко получилось, но теперь, надеюсь, все наглядно и понятно. Чтобы при расчете каждый раз не рисовать такой график, можно воспользоваться готовой формулой. Формула медианы имеет следующий вид:



где x_{Me} — нижняя граница медианного интервала;



$S_{(Me-1)}$ – суммарное количество наблюдений, которое было накоплено до начала медианного интервала, т.е. накопленная частота предмедианного интервала;

f_{Me} – число наблюдений в медианном интервале.

Как нетрудно заметить, формула медианы состоит из двух слагаемых: 1 – значение начала медианного интервала и 2 – та самая часть, которая пропорциональна недостающей накопленной доли до 50%.

Для примера рассчитаем медиану по следующим данным.



Требуется найти медианную цену, то есть ту цену, дешевле которой по половине количества товаров. Для начала произведем вспомогательные расчеты накопленной частоты, накопленной доли, общего количества товаров.



По последней колонке «Накопленная доля» определяем медианный интервал – 300-400 руб (накопленная доля впервые более 50%). Ширина интервала – 100 руб. Теперь остается подставить данные в приведенную выше формулу и рассчитать медиану.



То есть у одной половины товаров цена ниже, чем 350 руб., у другой половины – выше. Все просто. Средняя арифметическая, рассчитанная по этим же данным, равна 355 руб. Отличие не значитель но оно есть.



МЕДИАНА. Другое дело интервальные данные. Соответствующей функции в Excel нет. Поэтому нужно задействовать приведенную выше формулу. Что поделаешь? Но это не очень трагично, так как расчет медианы по интервальным данным – редкий случай. Можно и на калькуляторе разок посчитать.

Напоследок предлагаю задачку. Имеется набор данных. 15, 5, 20, 5, 10. Каково среднее значение?

Четыре варианта:

а) 11;

б) 5;

в) 10;

г) 5, 10, 11.

Мода, медиана и среднее значение выборки – это разный способ определить центральную тенденцию в выборке.

Ниже видеоролик о том, как рассчитать медиану в Excel.





Онлайн курс

Статистика в MS Excel



Корпоративный тренинг

Статистика в MS Excel

Поделиться в социальных сетях:

3



5
Поделились

← Предыдущая Запись

Следующая Запись →

★★★★★ 5 / 5

CACKLE



Оставьте свой комментарий...



Новые (11) ▾

✉ Подписаться

🔗 Поделиться





Йож 2017.05.02 17:40

ответ Г. 5 - мода, 10 - медиана, 11 - среднее арифметическое. все три показателя являются средними величинами

Ответить



Анастасия 2017.04.05 15:25

в ответ 10

Ответить



Дмитрий Езепов 2016.09.10 16:37

Случайную выборку? Так для этого нужно знать распределение.

Ответить



Денис 2016.09.10 15:42

Как построить выборку имея данные моду=8;медиану=1;среднее=10

Ответить



Вовк ↗ Денис 2019.07.17 12:16

-2, -1, 0, 1, 8, 8, 56

Ответить



Дмитрий Езепов 2016.07.07 18:51

СреднЕ АрифметическоE значениE.
Или выборочное среднее.

Ответить



о 2016.07.07 13:40

Что у вас с родом? "СреднЕ арифметическоE"

Ответить



Дмитрий Езепов 2016.03.18 11:02

Подвох в том, что под средней могут иметь виду среднюю арифметическую, медиану и моду.
Поэтому правильный ответ - г)

Ответить



Николай 2016.03.17 15:01

Средняя арифметическая 11, медиана 10, мода 5, максимум 20 минимум 5

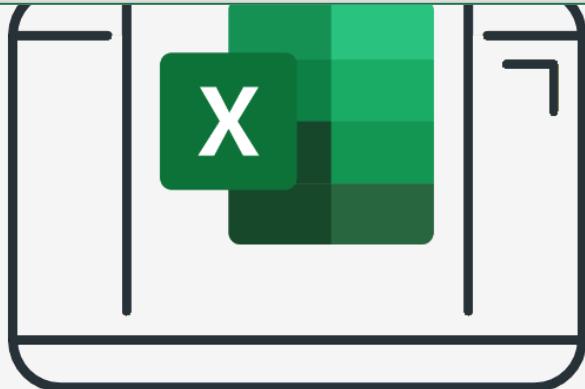
Ответить



Мбрья 2016.03.16 07:54

б





Продвинутый уровень Excel

От новичка до эксперта за 4 недели

[Онлайн курс](#)

Разное

[Шаблоны](#)

[Донатить](#)

[Ссылки](#)

ИП Езепов Дмитрий Александрович, УНП 192504505, Минский горисполком 10.07.2015, г.Минск,
ул. Киреенко, 3-61. Тел. +375296677292. Время работы: Пн-Вс с 9-00 до 18-00.

[Оферта](#)

[Политика конфиденциальности](#)



