

И. ВЫСОЦКИЙ,  
г. МоскваПродолжение.  
Начало см. № 4–5, 7, 10 за 2023 г.,  
№ 1, 2 за 2024 г.

# ТЕМА:

## «ПРЕДСТАВЛЕНИЕ ДАННЫХ»

### СЦЕНАРИИ УРОКОВ

#### Урок: «Медиана числового набора»

**Цель урока:** познакомить учащихся с понятием медианы и научить вычислять медиану числового набора в несложных случаях. Сформировать представление о медиане числового набора как об устойчивой характеристике, подвергающейся меньшей изменчивости, чем среднее арифметическое.

**Оборудование:** калькулятор.

#### Повторение

1. В каком случае для представления данных используется круговая диаграмма?

2. На круговой диаграмме слева показано распределение площади земной суши по всем шести материкам.

*Обсудите с учащимися некоторые вопросы.*

а) Верно ли, что Африка и обе Америки вместе занимают примерно половину всей суши?

б) Верно ли, что Африка занимает больше четверти всей суши?

в) Во сколько примерно раз Антарктида больше Австралии по площади территории?

**Ответы:** а) да; б) нет, чуть меньше; в) примерно в полтора раза.

#### Изучение нового материала

**Пример 1.** (Пример 5, с. 51 учебника, с обновленными данными) В 2018 году в России было всего 15 городов с числом жителей больше 1 млн человек. Данные о населении этих городов приведены в таблице.

| Город           | Население на 01.01.2018 (тыс. чел) | Город          | Население на 01.01.2018 (тыс. чел) |
|-----------------|------------------------------------|----------------|------------------------------------|
| Москва          | 12 506                             | Самара         | 1163                               |
| Санкт-Петербург | 5352                               | Ростов-на-Дону | 1130                               |
| Новосибирск     | 1613                               | Уфа            | 1121                               |
| Екатеринбург    | 1469                               | Красноярск     | 1091                               |
| Нижний Новгород | 1259                               | Пермь          | 1052                               |
| Казань          | 1244                               | Воронеж        | 1048                               |
| Челябинск       | 1202                               | Волгоград      | 1014                               |
| Омск            | 1172                               |                |                                    |

Найдем среднее значение численности жителей этих городов в 2018 году. Оно равно приблизительно 2229,1 тыс. чел. Обратите внимание: в таблице нет города, население которого было бы близко к этой величине. Почти во всех городах население лишь немного превышает 1 млн человек. Исключение составляют Москва и Санкт-Петербург.

☁ Есть дополнительные материалы на сайте [raum.math.ru](http://raum.math.ru).

Обсудите с учащимися некоторые вопросы.

– Является ли среднее арифметическое хорошим описательным параметром величины «население города-миллионера в России». Если какой-нибудь иностранец узнает, что крупные города в России имеют среднее население 2,2 млн человек, верное ли представление о типичном крупном городе создастся у этого иностранца?

– Как исправить положение? Нужно подобрать более подходящий описательный параметр.

*Желательный результат обсуждения*

Мы уже сталкивались с таким свойством среднего арифметического на уроке по теме «Среднее арифметическое числового ряда» (см. № 10, 2023), когда обсуждали население крупнейших городов Московской области. Численность населения города Балашихи оказалась «выбросом», который «оттащил» среднее от наиболее типичных значений. Здесь такими выбросами являются население Москвы и Санкт-Петербурга.

Мы, конечно, помним, что среднее арифметическое гораздо более устойчиво, чем каждое отдельное значение в массиве данных, но в некоторых случаях этой устойчивости недостаточно. Выбросы оказывают большое влияние на сумму, и тогда среднее арифметическое не является хорошим описательным параметром.

В данном случае поступить можно очень просто: удалить выбросы. Если исключить Москву и Санкт-Петербург, останется 13 городов со средним населением 1198,3 тыс. человек. Это значение вполне удовлетворительно описывает массив численности населения крупных городов. Города, ближайšie к среднему по населению, — Челябинск и Омск. Население этих городов отличается от среднего менее чем на 30 тыс. чел. (меньше 3%). Москву и Санкт-Петербург придется описывать и учитывать отдельно.

Можно ли придумать такое среднее, которое не реагирует на выбросы? Описательную меру центра, которая устойчивее среднего арифметического? Такая мера центра есть. Она называется *медиана*. Медиана числового набора делит его на две равные по численности части.

**Пример 2.** (Пример 1, с. 48) Возьмем набор чисел: 1, 4, 7, 9, 11. Удобно, что числа записаны по возрастанию. Подберем число  $m$  так, чтобы в наборе оказалось поровну чисел, которые не больше и которые не меньше, чем  $m$ . Очевидно,  $m = 7$ . Три числа не больше 7 и три числа не меньше 7.

**Пример 3.** (Пример 2, с. 48) Рассмотрим набор чисел: 1, 3, 6, 11. Числа тоже записаны по возрастанию, но их четыре, поэтому среди них нет числа, стоящего точно посередине. Любое число от 3

до 6 можно считать медианой этого набора (и даже числа 3 и 6). Обычно в качестве конкретного значения берут среднее арифметическое двух «серединных» чисел. В данном примере это будет число

$$\frac{3+6}{2} = 4,5.$$

Как быть, если в наборе встречаются повторяющиеся числа, а сам набор не упорядочен по возрастанию?

**Пример 4.** (Пример 3, с. 49) Найдем медиану набора чисел: 12, 2, 11, 3, 7, 10, 3.

Сначала расположим числа по возрастанию. Получится упорядоченный по возрастанию ряд. Его называют *вариационным рядом*<sup>1</sup>:

$$2, 3, 3, \underline{7}, 10, 11, 12.$$

Чисел в этом ряду семь, поэтому медианой служит четвертое по счету число. Это число 7, оно подчеркнуто. Четыре числа не больше 7 (числа 2, 3, 3 и само число 7) и четыре числа не меньше 7 (числа 7, 10, 11 и 12). Число 7 входит в обе группы.

В этом наборе нечетное число чисел, поэтому медиану было найти очень просто. Чуть сложнее дело обстоит, если чисел четное количество.

**Пример 5.** Найдем медиану набора:

$$12, 2, 11, 3, 7, 10, 3, 15.$$

Снова упорядочим числа:

$$2, 3, 3, \underline{7}, \underline{10}, 11, 12, 15.$$

Теперь посередине ряда стоят два числа: 7 и 10 (подчеркнуты). Медианой можно считать любое число от 7 до 10. Чаще всего эти срединные числа усредняют:

$$\frac{7+10}{2} = 8,5.$$

*Комментарий и результаты обсуждения*

Строгое определение медианы можно не давать (это будет сделано в следующем уроке), но можно описать правило или алгоритм построения медианы. Чтобы найти медиану числового набора, нужно:

1. Упорядочить набор по возрастанию (получить вариационный ряд).

2. Если в наборе нечетное количество чисел, то медианой будет число, стоящее посередине упорядоченного набора.

3. Если в наборе четное количество чисел, то медианой будет среднее арифметическое двух чисел, стоящих посередине<sup>2</sup>.

<sup>1</sup> Вариационный ряд можно получить и упорядочиванием по убыванию. Так обычно поступают в тех случаях, когда данные естественным образом упорядочиваются по убыванию (например, число баллов у спортсменов на соревнованиях и т.п.).

<sup>2</sup> В общем случае — любое число, принадлежащее отрезку, ограниченному двумя срединными числами. Если посередине числа  $a$  и  $b$ , то медианой будет любое число  $m$ , если  $a \leq m \leq b$ .

**Пример 6.** (Пример 4, с. 50) Найдём среднее арифметическое и медиану данных о производстве (в млн тонн) пшеницы в России.

|              |      |      |      |      |      |
|--------------|------|------|------|------|------|
| Год          | 2008 | 2009 | 2010 | 2011 | 2012 |
| Производство | 63,8 | 61,7 | 41,5 | 56,2 | 37,7 |

|              |      |      |      |      |      |
|--------------|------|------|------|------|------|
| Год          | 2013 | 2014 | 2015 | 2016 | 2017 |
| Производство | 52,1 | 59,7 | 61,6 | 73,3 | 85,8 |

Среднее арифметическое 59,3 млн тонн.

Чтобы найти медиану, упорядочим числа:

37,7; 41,5; 52,1; 56,2; 59,7; 61,6; 61,7; 63,8;  
73,3; 85,8.

Чисел в ряду 10. Медианой будет среднее между числами 59,7 и 61,6:

$$\frac{59,7 + 61,6}{2} = 60,65.$$

Видно, что в этом примере среднее арифметическое и медиана близки: разность между ними  $60,65 - 59,3 = 1,35$  — мала по сравнению с самими числами. Как вы думаете, чем это может объясняться?

*Желательный результат обсуждения*

В данных о производстве пшеницы нет выбросов. Все значения сгруппированы довольно кучно, поэтому среднее и медиана близки.

**Пример 7.** В таблице даны сведения о площади водной поверхности десяти крупнейших озер и бессточных озер (морей) мира. Посмотрите на таблицу и без вычислений скажите, какое среднее будет лучше описывать площадь типичного крупного озера: среднее арифметическое или медиана.

| Озеро            | Площадь (км <sup>2</sup> ) |
|------------------|----------------------------|
| Каспийское море  | 376 000                    |
| Верхнее          | 82 100                     |
| Виктория         | 68 100                     |
| Гурон            | 60 000                     |
| Мичиган          | 57 800                     |
| Аральское море   | 51 100                     |
| Танганьика       | 32 900                     |
| Байкал           | 31 500                     |
| Большое Медвежье | 31 326                     |
| Ньяса            | 30 800                     |

Проверьте свое предположение с помощью вычислений. Можете ли вы указать озеро, чья площадь близка к среднему арифметическому? К медиане (медианного представителя)?

*Желательный результат обсуждения*

Средняя площадь равна 82 162 км<sup>2</sup>, медиана данных равна 54 450 км<sup>2</sup>. Представитель среднего — озеро Верхнее, медианные представители — Мичиган и Аральское море. Лучше описывает площади крупнейших озер, видимо, медиана, поскольку среднее арифметическое сильно сдвинуто от медианы за счет Каспийско-

го моря: только его площадь больше среднего значения.

*Выводы.* Медиана — средний показатель числового набора, который часто лучше среднего арифметического описывает типичный элемент набора. Особенно преимущества медианы проявляются при наличии выбросов в данных. Для вычисления медианы сначала следует упорядочить числа в наборе, а затем выбрать либо единственное срединное число (если чисел нечетное количество), либо взять среднее арифметическое двух срединных чисел (если в наборе четное количество чисел).

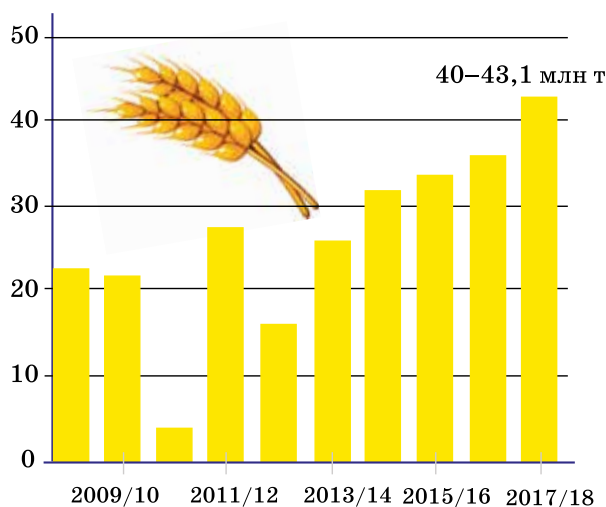
*Рекомендуемое домашнее задание:* с. 52, упр.: 1 (а, в); 4 (а); 5 (б); 8.

## Урок: «Определение и свойства медианы»

*Цель урока:* познакомить учащихся с теоретическим определением медианы, научить проверять определение для разных чисел, продемонстрировать устойчивость медианы на примерах и научиться находить медианных представителей.

### Повторение

На столбиковой диаграмме показаны данные об экспорте зерна из России с 2008 года (для 2018 года сделаны прогнозы 40–43,1 млн т). Рассмотрите диаграмму и выполните задания (устно).



1. Как вы думаете, почему объем экспорта считается не по календарным годам, а по концу года / началу следующего (как учебные годы). Что, например, означает экспорт в 2009/10 гг.?

2. Чем может объясняться резкое падение экспорта зерна в 2010/11 гг.?

3. Какая тенденция наблюдается в последние 6 лет? Придумайте правдоподобное объяснение.

4. Можно ли оценить эту тенденцию количественно. Каков примерный средний ежегодный прирост экспорта за последние 6 лет?

*Желательный результат обсуждения*

Экспорт (продажа за рубеж) пшеницы производится из запасов собранного урожая. Урожай собирают осенью. Поэтому экспорт этого урожая продолжается в конце текущего года и в начале следующего — до следующего урожая. В 2010 году на хлебобродных территориях европейской части России почти все лето стояла жестокая засуха. Большая часть урожая сгорела. Запасов собранного зерна оказалось меньше, чем обычно, поэтому экспорт резко сократился.

Последние шесть лет экспорт устойчиво растет. Вероятно, этому способствует снижение курса рубля в 2014 году — российское зерно на мировых рынках оказалось дешевле зерна из других стран. Спрос на него возрос, прибыль увеличилась. Дополнительный доход был направлен на оснащение сельского хозяйства, что привело к дальнейшему росту производства и увеличению экспорта. За шесть лет экспорт вырос примерно на  $40 - 16 = 24$  млн т, то есть рос в среднем на 4 млн т ежегодно.

**Изучение нового материала**

На предыдущем уроке мы познакомились со способом вычисления медианы. Напомним: если чисел в наборе нечетное количество, то медиана — срединное число в упорядоченном ряду чисел. Если чисел четное количество, то медиана — любое из двух срединных чисел или любое число между ними. Эти два случая можно объединить в один следующим определением.

**Определение.** Число  $m$  называется медианой числового набора, если в этом наборе хотя бы половина чисел *не больше* числа  $m$  и хотя бы половина чисел *не меньше* числа  $m$ .

**Пример 1.** Дан числовой набор:

1, 7, 1, 3, 1, 4, -3, 0, 1, -2, 5, -1.

Проверьте, пользуясь определением, является ли медианой этого набора число: а) 1; б) 0?

**Решение.** а) Всего в наборе 12 чисел. Чтобы число 1 было медианой, нужно, чтобы в наборе было хотя бы шесть чисел, которые не больше числа 1 (то есть меньше или равны числу 1), и хотя бы шесть чисел, которые не меньше числа 1 (то есть больше или равны числу 1). Обведем кружком числа, которые не больше числа 1, и подчеркнем числа, которые не меньше 1.

$\underline{\textcircled{1}}$ , 7,  $\textcircled{1}$ , 3,  $\textcircled{1}$ , 4,  $\textcircled{-3}$ ,  $\textcircled{0}$ ,  $\textcircled{1}$ ,  $\textcircled{-2}$ , 5,  $\textcircled{-1}$ .

Заметим, что число 1 попадает в обе группы. Поэтому его нужно и обводить, и подчеркивать

всякий раз, когда оно встречается. Всего обведено 8 чисел и подчеркнуто тоже 8 чисел.

Таким образом, в данном наборе восемь чисел, которые не больше числа 1, и восемь чисел, которые не меньше числа 1. Определение медианы выполнено. Значит, 1 — медиана данного набора.

б) Сделаем то же самое, приняв за границу число 0. Обведем числа, которые не больше нуля, и подчеркнем те, которые не меньше нуля.

$\underline{1}$ ,  $\underline{7}$ ,  $\underline{1}$ ,  $\underline{3}$ ,  $\underline{1}$ ,  $\underline{4}$ ,  $\textcircled{-3}$ ,  $\textcircled{0}$ ,  $\underline{1}$ ,  $\textcircled{-2}$ ,  $\underline{5}$ ,  $\textcircled{-1}$ .

В наборе четыре числа, которые не больше числа 0, и девять чисел, которые не меньше числа 0. Чисел, которые не больше числа 0, слишком мало:  $4 < 6$ . Поэтому 0 не является медианой данного набора по определению.

**Пример 2.** Дан числовой набор:

3, 2, 2, 7, 4, 1, 5, 4.

Проверьте, пользуясь определением, является ли медианой этого набора число: а) 2; б) 3; в) 4; г) 3,22.

**Решение.** а) В наборе 8 чисел. Не превосходят число 2 три числа (2, 2, 1). А нужно, чтобы таких чисел было не меньше четырех. Значит, 2 не является медианой.

б) Не превосходят число 3 четыре числа (3, 2, 2, 1). Не меньше, чем число 3, пять чисел (3, 7, 4, 5, 4). Значит, число 3 — медиана.

Аналогично показываем, что числа 4 и 3,22 являются медианами.

*Желательный результат обсуждения*

Этот пример показывает нам, что иногда медиана определяется неоднозначно. Вычисляя медиану на практике, мы будем придерживаться правила: если чисел в наборе четное число, медианой является среднее арифметическое двух срединных чисел.

Медиана так же, как и среднее арифметическое, «двигается» и «сжимается» вместе со всем набором данных.

**Пример 3.** На протяжении суток каждые 3 часа записывали результат измерения температуры воздуха (в градусах Цельсия) в таблицу.

| Время | Температура (°C) |
|-------|------------------|
| 00:00 | 18               |
| 03:00 | 14               |
| 06:00 | 12               |
| 09:00 | 14               |
| 12:00 | 19               |
| 15:00 | 23               |
| 18:00 | 22               |
| 21:00 | 27               |
| 24:00 | 16               |

а) Найдите медиану этих данных.

б) Предположим, для публикации в канадском журнале эти данные нужно перевести в шкалу Фаренгейта. Для перевода температуры в градусах Цельсия в градусы Фаренгейта используется формула

$$T^{\circ}\text{F} = 1,8 \cdot T^{\circ}\text{C} + 32.$$

Можно ли найти медиану в шкале Фаренгейта, не переводя в шкалу Фаренгейта все девять измерений? Найдите медиану в  $^{\circ}\text{F}$ .

Ответ: а)  $18^{\circ}\text{C}$ ; б)  $64,4^{\circ}\text{F}$ .

### Устойчивость медианы к выбросам<sup>3</sup>

Предположим, что в числовом наборе 10 чисел. Мы помним, что если увеличивать или уменьшать одно число набора (двигать его), то среднее арифметическое будет двигаться в ту же сторону, но в 10 раз медленнее. А как поведет себя медиана? Проще всего разобраться в этом на примере.

**Пример 4.** Рассмотрим набор из первых 10 натуральных чисел:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Среднее арифметическое и медиана совпадают. Они равны 5,5.

а) Увеличим последнее число на 10, а потом еще на 100. Как изменятся среднее и медиана?

б) Увеличим первое число на 10, а потом еще на 100. Как изменятся среднее и медиана в этом случае?

**Решение.** а) Если увеличить последнее число на 10, то среднее арифметическое увеличится на 1 и станет равно 6,5. А медиана останется прежней (5,5): она зависит только от двух срединных чисел, 5 и 6, которые не изменялись. Если увеличить последнее число еще на 100, то среднее арифметическое вырастет еще на 10 и теперь будет 16,5. Медиана и в этом случае не изменится.

б) Увеличим теперь первое число на 10. Получится 11. Среднее вырастет на 1 и станет равно 6,5. Так как вариационный ряд теперь изменился, медиана изменилась тоже. Срединными числами нового набора:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11,

являются не числа 5 и 6, как прежде, а числа 6 и 7. Следовательно, медиана теперь равна 6,5. Но дальнейшее увеличение первого числа уже не изменит медиану.

Если теперь первое число увеличить еще на 100, получится набор:

2, 3, 4, 5, 6, 7, 8, 9, 10, 111,

<sup>3</sup> До сих пор мы не обсуждали, как определить, является ли значение выбросом или нет. На этом этапе обучения понятие выброса остается интуитивным. Впоследствии будет предложено некоторое правило.

где среднее равно  $6,5 + 10 = 16,5$ , а медиана уже больше не изменилась. Она равна 6,5.

### Желательный результат обсуждения

На простом примере мы увидели, что один выброс может повлиять на медиану, но влияние это слабое. Сильно «сдвинуть» медиану с места одно «гуляющее» значение набора не может. Это свойство называют *устойчивостью медианы по отношению к выбросам*. Именно это свойство делает медиану очень полезной мерой центра в современной статистике.

Рассмотрим пример 7 (см. предыдущий урок) про площади озер. Нужно было найти медиану данных о площади поверхности десяти крупнейших озер мира. Медиана равна  $54\,450\text{ км}^2$ . Мы назвали медианными представителями озеро Мичиган и Аральское море, поскольку их площади близки к медиане.

Если даны числовые характеристики каких-то объектов (озер, городов, стран и т.п.), то объекты, у которых характеристика близка к медиане, называют *медианными представителями этой характеристики*. Медианных представителей может быть один или несколько. Возникает вопрос — что считать значением «близким к медиане». Однозначного ответа нет. Все зависит от конкретной задачи и целей исследования.

**Пример 5.** В таблице даны высоты крупнейших вулканов Срединного вулканического пояса Камчатки. Найдите медиану данных о высотах и медианного представителя (или медианных представителей). Подумайте, как можно разумно сформулировать правило или правила для выбора медианного представителя по высоте в данном случае.

| Вулкан                   | Высота<br>(м над уровнем моря) |
|--------------------------|--------------------------------|
| Ичинская Сопка           | 3621                           |
| Хувхойтун                | 2618                           |
| Алней                    | 2581                           |
| Острая Сопка             | 2539                           |
| Чашаконджа               | 2526                           |
| Шишель                   | 2525                           |
| Иктунул                  | 2300                           |
| Снежный                  | 2211                           |
| Снеговой                 | 2172                           |
| Спокойный                | 2171                           |
| Кевезйтунул              | 2133                           |
| Белый                    | 2080                           |
| Вулкан Горного института | 2024                           |
| Новограбленова           | 2000                           |
| Хангар                   | 2000                           |
| Малый Паяллан            | 1980                           |

|                    |      |
|--------------------|------|
| Южный Черпук       | 1962 |
| Лагерный           | 1961 |
| Ульванэй           | 1954 |
| Северный           | 1936 |
| Чинейнейн          | 1922 |
| Большой Паяллан    | 1906 |
| Апнгей             | 1856 |
| Анау               | 1828 |
| Паяллан            | 1811 |
| Леякина            | 1770 |
| Атласова           | 1764 |
| Сергеева           | 1759 |
| Каменистый         | 1758 |
| Айнелькан          | 1725 |
| Кайленэй           | 1680 |
| Северный Черпук    | 1679 |
| Гречишкина         | 1651 |
| Ука                | 1643 |
| Междусопочный      | 1641 |
| Титила             | 1559 |
| Уксичан            | 1550 |
| Бонгапчи           | 1550 |
| Лангуткин          | 1534 |
| Кэбеней            | 1529 |
| Большая Кытэпана   | 1502 |
| Еловский           | 1381 |
| Иэттунул           | 1340 |
| Большой Чекчебонай | 1338 |
| Леутонгей          | 1333 |
| Мутный             | 1315 |
| Большой            | 1301 |
| Теклетунул         | 1290 |
| Тильмыг            | 1265 |
| Малый Чекчебонай   | 1261 |
| Плоский            | 1255 |
| Малая Кытэпана     | 1230 |
| Воямпольский       | 1225 |
| Кахтана            | 1217 |
| Тунупилянум        | 1200 |
| Ламутский          | 1198 |
| Слюнина            | 1175 |
| Озерной            | 1021 |
| Шлен               | 1001 |
| Федотыч            | 965  |
| Терпук             | 765  |
| Киненин            | 583  |

**Желательный результат обсуждения**

Найти медиану легко без вычислений. Вулканы в таблице упорядочены по убыванию высоты. Всего их 62, деление в таблице проходит точно посередине. Медиана равна 1679,5 м. Ближайшими по высоте являются вулканы Кайленэй и Северный Черпук. Они и являются медианными представителями.

**Выводы.** Медиана набора числовых данных иногда определяется неоднозначно. Но на практике в таких случаях в качестве медианы следует брать среднее арифметическое двух срединных чисел.

При одинаковом изменении всех данных медиана меняется вместе с ними таким же образом. Однако медиана не меняется или мало меняется (устойчива) при изменении одного-двух элементов массива. Это дает свойство устойчивости относительно выбросов.

**Рекомендуемое домашнее задание**

1. Юный статистик записывал в электронную таблицу результаты взвешивания шоколадных батончиков массой 50 г. Потом он посмотрел на среднее арифметическое (345 г) и очень удивился.



- а) Почему среднее арифметическое получилось таким большим? Есть ли в данных выброс?
- б) Как вы думаете, в чем может быть причина появления этого выброса? Как исправить ситуацию?
- в) Найдите медиану данных до и после исправления ошибки.
- г) Можно ли считать, что медиана ошибочных данных существенно отличается от медианы данных с исправленной ошибкой?
- д) Вместо того чтобы исправлять ошибку, удалите ошибочное число из расчетов. Останется 15 значений. Найдите теперь среднее арифметическое и медиану. Сильно ли они отличаются, по вашему мнению?

|   | A                               | B        | C    | D     |
|---|---------------------------------|----------|------|-------|
| 1 | <b>Масса шок. батончиков, г</b> |          |      |       |
| 2 | 49,8                            | 53,1     | 47,8 | 52,4  |
| 3 | 50,2                            | 50,4     | 49,3 | 4902  |
| 4 | 51,2                            | 47,9     | 48,3 | 49,9  |
| 5 | 52,8                            | 59,2     | 49,2 | 50,1  |
| 6 |                                 |          |      |       |
| 7 |                                 | Среднее: |      | 354   |
| 8 |                                 |          |      | 50,15 |

- 2. В таблице дана плотность населения стран Южной Европы.
- а) Найдите медиану и среднее арифметическое данных. Какое среднее лучше отражает плотность населения страны Южной Европы?
- б) В чем причина сильного отличия значений найденных средних?
- в) Укажите медианного представителя (или представителей).

| Страна               | Плотность населения (чел./км <sup>2</sup> ) |
|----------------------|---|
| Черногория           | 46  |
| Босния и Герцеговина | 76  |
| Хорватия             | 79  |
| Македония            | 83  |
| Греция               | 86  |
| Испания              | 95  |
| Словения             | 98  |
| Сербия               | 103   |
| Албания              | 105   |
| Португалия           | 117   |
| Андорра              | 167   |
| Италия               | 205   |
| Сан-Марино           | 350   |
| Мальта               | 1315  |
| Ватикан              | 1818  |
| Монако               | 16 500                                      |

**Ответы:** 1. а) Есть выброс 4902 г. Это неправдоподобно много. б) Причиной выброса послужила, скорее всего, ошибка в записи. Вероятно, юный статистик случайно вместо запятой нажал клавишу с цифрой 0. в) Медианы равны 50,15 г и 50 г соответственно. г) Разность 0,15 г вряд ли можно считать существенной. д) Если удалить значение 4902, то среднее будет 50,77 г, а медиана 50,1 г. Разность составляет 0,67 г, что очень мало. 2. Среднее арифметическое 1327,7 чел./км<sup>2</sup>. Медиана 104 чел./км<sup>2</sup>. Такая разница из-за выброса 16 500 (Монако). Возможно, Мальту и Ватикан тоже следует считать странами с особенно высокой плотностью. Медиана намного лучше описывает плотность населения типичной южно-европейской страны.

## Урок-практикум: «Медиана»

**Цель урока:** учащиеся должны научиться использовать стандартную функцию MS Excel для вычисления медианы. Умение использовать функции электронных таблиц для поиска медианного представителя массива и урезанного среднего следует считать неосновной целью.

**Оборудование:** персональный компьютер или планшет с установленным процессором электронных таблиц и загруженным файлом Les25.xlsx.

### Повторение (устная работа, 5 мин)

Учитывая природу данных и цель описания, подберите подходящий описательный параметр набора данных.

1. **Эксперимент:** соревнования по метанию копья. **Данные:** дальность броска для каждого спортсмена (в метрах). Цель — выявить победителя соревнований.

2. **Эксперимент:** сбор данных об изменении биржевых цен на никель за несколько дней (см. табл.). **Данные:** отношение цены в конкретный день к цене предыдущего дня (в правой графе). Цель — найти среднее дневное изменение цены.

| Дата       | Цена (USD за т) | Изменение (%) | Множитель |
|------------|-----------------|---------------|-----------|
| 09.11.2018 | 11 415          | -2,97         | 0,9703    |
| 12.11.2018 | 11 345          | -0,61         | 0,9939    |
| 13.11.2018 | 11 325          | -0,18         | 0,9982    |
| 14.11.2018 | 11 300          | -0,22         | 0,9978    |
| 15.11.2018 | 11 325          | 0,22          | 1,0022    |
| 16.11.2018 | 11 395          | 0,62          | 1,0062    |
| 19.11.2018 | 11 190          | -1,80         | 0,9820    |
| 20.11.2018 | 11 085          | -0,94         | 0,9906    |
| 21.11.2018 | 11 130          | 0,41          | 1,0041    |

3. **Эксперимент:** сбор данных о заработной плате рабочих на промышленных производствах в автомобильной промышленности. **Данные:** месячные зарплаты постоянных рабочих из случайной выборки объемом 1500 чел. Цель — определение типичной, средней зарплаты рабочего по отрасли.

4. **Эксперимент:** сбор данных о заработной плате рабочих на промышленных производствах в текстильной промышленности. **Данные:** месячные зарплаты сотрудников (не только рабочих) текстильных производств из случайной выборки объемом 1000 чел. Цель — определение типичной, средней зарплаты рабочего текстильной промышленности.

### Желательные результаты обсуждения

1. Наибольшее значение, поскольку победитель — тот, кто бросил копьё дальше всех.

2. Отношение цен показывает, во сколько раз цена увеличилась (уменьшилась) сегодня по сравнению со вчерашним днем. Собранные данные — множители. Подходящий описательный параметр — среднее геометрическое.

3. В выборке участвуют только постоянные рабочие — в этой выборке не должно быть выбросов. Должно подойти среднее арифметическое.

4. Сказано, что опрашиваются не только рабочие. В выборку могло попасть несколько очень высокооплачиваемых или низкооплачиваемых сотрудников. Поэтому нужен более устойчивый показатель, чем среднее арифметическое. Вероятно, медиана подойдет лучше.

Для вычисления медианы числового массива разработчики электронных таблиц предусмотрели специальную функцию МЕДИАНА (массив данных).

<sup>4</sup> Вопрос можно обсуждать, если изучалась тема «Среднее геометрическое» (см. № 1, 2024).

**Пример 1.** В таблице (см. файл Les25.xlsx, лист «Города Подмосковья») показано население всех городов Московской области на 2018 год. Мы знаем, что среднее арифметическое — не самый лучший показатель для населения типичного города из-за того, что в данных присутствует выброс — население города Балашихи. Медиана лучше описывает население типичного города. Найдите медиану данных и медианного представителя.

**Решение.** Откроем файл Les25.xlsx и ознакомимся с таблицей (рис. 1). Города упорядочены по алфавиту. Можно упорядочить города по населению. Тогда медианным будет 37-й город, поскольку всего городов 73. Но лучше использовать функцию МЕДИАНА. Где-нибудь в удобном месте, например в ячейке E3, напишем слово «Медиана» (лучше сразу привыкнуть аккуратно оформлять таблицы). А в ячейку F3 впишем имя функции =МЕДИАНА.

| № | Город       | Население, тыс. чел. |
|---|-------------|----------------------|
| 1 | АПРЕЛЕВКА   | 27,9                 |
| 2 | БАЛАШИХА    | 468,2                |
| 3 | БРОННИЦЫ    | 22,6                 |
| 4 | ВЕРЕЯ       | 5,1                  |
| 5 | ВИДНОЕ      | 66,2                 |
| 6 | ВОЛОКОЛАМСК | 19,8                 |

Рис. 1

Теперь нужно указать массив данных: =МЕДИАНА(С3:С75).

Осталось нажать клавишу *Enter*. В ячейке F3 появится результат 51,9. Теперь нужно найти медианного представителя — город, где население 51,9 тыс. человек. Это можно сделать, просматривая таблицу. Но можно использовать функцию поиска позиции в MS Excel. В ячейку E4 впишем название: «Медианный представитель» (столбец придется расши-

рить), а в ячейке F4 напишем функцию поиска =ПОИСКПОЗ(F3;С3:С75;0).

Здесь F3 — значение, которое мы ищем (медиана, равная 51,9), массив С3:С75 — это массив, в котором мы ищем. Аргумент 0 нужен, чтобы функция нашла точное совпадение (с другими аргументами функция работает только на упорядоченных массивах). Результат представлен на рисунке 2.

| № | Город     | Население, тыс. чел. |
|---|-----------|----------------------|
| 1 | АПРЕЛЕВКА | 27,9                 |
| 2 | БАЛАШИХА  | 468,2                |
| 3 | БРОННИЦЫ  | 22,6                 |
| 4 | ВЕРЕЯ     | 5,1                  |

Рис. 2

Но все же это не то, что мы ищем. Мы теперь знаем номер города в нашем списке (легко видеть, что это Солнечногорск). Как же сделать, чтобы MS Excel дал название города? Для этого применим еще одну функцию: ВПР. Эта функция, зная номер города (59) в первом столбце, найдет его название во втором столбце. Используем ячейку G4 (рис. 3): =ВПР(F4;A3:В75;2).

Здесь F4 — ячейка, где хранится номер города; A3:В75 — массив номеров и названий; 2 — столбец, в котором мы ищем название города.

**Примечание.** Если бы городов было четное количество, то в таблице могло бы не быть города, у которого население в точности равно медиане. В этом случае поиск медианных представителей в MS Excel сложнее, но все же возможен. Предложите желающим подумать: как найти ячейку, в которой находится наибольшее число, не превосходящее медиану. Или наоборот — наименьшее число, которое больше медианы.

Мы уже знаем, что среднее арифметическое в равной мере учитывает все значения в массиве и позволяет судить о сумме значений — это его главные достоинства. Достоинство медианы —

| № | Город     | Население, тыс. чел. |
|---|-----------|----------------------|
| 1 | АПРЕЛЕВКА | 27,9                 |
| 2 | БАЛАШИХА  | 468,2                |
| 3 | БРОННИЦЫ  | 22,6                 |
| 4 | ВЕРЕЯ     | 5,1                  |

Рис. 3



устойчивость. Хочется объединить полезные свойства среднего арифметического и медианы. Для этого придумано *урезанное среднее*. Урезанное среднее массива данных находят в два приема.

1. Из массива удаляют часть самых больших значений и такую же часть самых маленьких значений. Если в массиве есть выбросы, то они будут удалены. Например, можно договориться, что удаляется 50% значений: 25% самых больших и 25% самых маленьких.

2. Находят среднее арифметическое оставшихся данных. Оставшиеся данные — *регулярные*. Они не являются выбросами. Регулярные значения наиболее надежные и образуют большинство. Вот эти данные и нужно усреднить с помощью обычного среднего арифметического.

**Пример 2.** Найдем урезанное среднее набора чисел:

1, 3, 2, -18, 3, 4, 54, 4, 3, 2,

удалив 40% значений.

**Решение.** 1. Выстроим числа по возрастанию так же, как мы это делаем для поиска медианы:

-18, 1, 2, 2, 3, 3, 3, 4, 4, 54.

Удалить нужно всего 40% чисел, то есть 20% наименьших чисел (с левой стороны ряда) и 20% наибольших — с правой стороны. Всего чисел 10. Значит, 20% — это два числа. Удаляем два числа слева и два числа справа:

2, 2, 3, 3, 3, 4.

В результате сильно выделяющиеся значения, -18 и 54, оказались удалены.

2. Находим среднее арифметическое оставшихся восьми чисел:

$$\frac{2+2+3+3+3+4}{8} = \frac{17}{8} = 2,125.$$

Для поиска урезанного среднего арифметического в MS Excel предусмотрена специальная функция =УРЕЗСРЕДНЕЕ(массив; удаляемая\_часть).

**Пример 3.** Найдем урезанное среднее данных о городах Подмосковья, отбрасывая 50% массива — четверть самых больших и четверть самых маленьких значений. Обратите внимание: удаляемую часть нужно указать в процентах (50%) или десятичной дробью (0,5). Если указать 50 без знака «%», функция сообщит об ошибке.

Не будем описывать последовательность действий подробно. Результат виден на рисунке 4.

**Выводы.** Медиана как статистический описательный параметр используется очень широко. Поэтому в электронных таблицах заложена специальная функция. Урезанное среднее является специальным видом среднего, которое применяется для автоматического удаления выбросов из массива неочищенных данных. Оно соединяет в себе преимущества среднего арифметического и медианы, но, конечно, обладает и недостатками.

При умелом использовании MS Excel можно искать не только медиану числовых характеристик объектов, но и медианных представителей.

*Рекомендуемое домашнее задание*

В файле Les25.xlsx на листе «Эмиссия углерода» собраны данные о выбросе углекислого газа в атмосферу по странам мира (тонн на душу населения). Данные неоднородны и содержат выбросы.

а) Рассмотрите таблицу. Подумайте, в чем причина того, что в некоторых странах на одного человека приходится очень много выбросов углекислого газа, а в других — совсем мало (например, сравните расположенные близко друг к другу Австралию и Восточный Тимор).

б) Найдите медиану данных о выбросах и медианную страну.

в) Найдите урезанное среднее данных, удаляя 10% и 50% наблюдений.

| F7 |             | fx =УРЕЗСРЕДНЕЕ(C3:C75;F6) |                                |
|----|-------------|----------------------------|--------------------------------|
| A  | B           | C                          | F                              |
| №  | Город       | Население, тыс.чел.        |                                |
| 1  | АПРЕЛЕВКА   | 27,9                       | Медиана 51,9                   |
| 2  | БАЛАШИХА    | 468,2                      | Медианный представитель 59 СОГ |
| 3  | БРОННИЦЫ    | 22,6                       |                                |
| 4  | ВЕРЕЯ       | 5,1                        | Удаляемая часть значений 50%   |
| 5  | ВИДНОЕ      | 66,2                       | Урезанное среднее 53,39459     |
| 6  | ВОЛОКОЛАМСК | 19,8                       |                                |

Рис. 4

