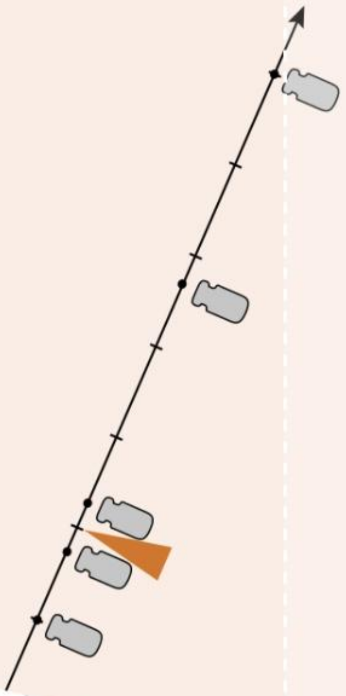
II Описательная статистика

В этой главе речь пойдёт о том, как одним-двумя числами описать важные свойства большого массива данных. Отсюда название главы — «Описательная статистика».

Существует множество описательных показателей, по которым можно судить о средних значениях, рассеивании, симметричности и характере изменения статистических данных.

В статистике широко используются среднее арифметическое и медиана.



В этой главе речь пойдёт о массивах или наборах чисел и о том, как описать одним-двумя числами важные свойства набора, в котором много чисел. Раздел статистики, изучающий методы описания данных, называется **описательной статистикой**. Чтобы одним числом охарактеризовать весь числовой массив, используют различные средние значения. В статистике широко используются среднее арифметическое, медиана. Иногда требуются и другие средние.

Какое именно среднее лучше выбрать для описания того или иного набора данных, зависит от природы данных, целей исследования и сложившихся традиций.

7 Среднее арифметическое

В таблице 22 показаны цены одного из популярных смартфонов в 10 разных интернет-магазинах на 13 января 2022 г.

Таблица 22. **Цена смартфона, р.**

Магазин	Цена	Магазин	Цена
1	8050	6	7790
2	8480	7	8290
3	8590	8	7890
4	8340	9	7970
5	8190	10	7910

Цены в разных магазинах разные, это связано с многими факторами, но в целом можно считать, что данные **однородны** — речь идёт об одном и том же смартфоне, и резко выделяющихся значений нет. Чтобы одним числом охарактеризовать цену смартфона в указанный день, разумно найти среднюю цену. Сложим все цены и разделим на количество слагаемых, то есть на 10. Получается

$$\frac{8050+8480+8590+8340+8190+7790+8290+7890+7970+7910}{10}=8150 \text{ (р.)}.$$

Для простоты можно считать, что массив состоит не из цен в рублях, а из чисел. Поэтому мы не писали единицы измерения у каждого слагаемого, а указали их один раз в конце в скобках.

Найденное значение называется **средним арифметическим**. На это значение можно ориентироваться при принятии решения о покупке. Цена намного выше средней, скорее всего, не устроит покупателя, а цена намного ниже выглядит подозрительно.

Среднее арифметическое — самая употребительная центральная мера. Поэтому иногда среднее арифметическое называют просто **средним** или **средним значением**.

Средним арифметическим числового массива называется отношение суммы всех чисел массива к их количеству.

Нанесём все величины на числовую ось точками, а среднее значение укажем стрелкой (рис. 4).

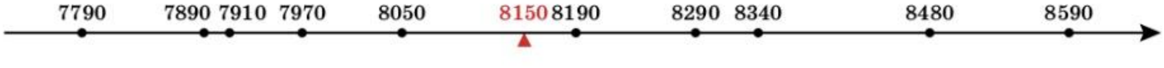


Рисунок 4. **Цены на смартфон в 10 магазинах, р.**

Пять чисел меньше среднего и пять больше. При этом значение 8190 довольно близко к среднему.

Среднее арифметическое является центром набора чисел. Поясним, что здесь означает слово «центр». Представим, что числовая ось является стержнем, на который подвешены одинаковые гири в точках, соответствующих отмеченным числам (рис. 5).

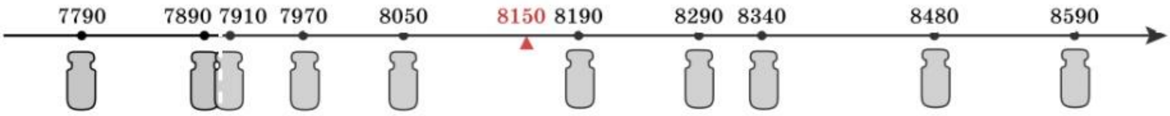


Рисунок 5. **Среднее арифметическое — точка равновесия**

На стержне существует точка равновесия. На эту точку можно «опереть» стержень с гирями так, что стержень окажется в равновесии. Этой точкой оказывается среднее арифметическое. В физике точку равновесия называют **центром масс**.

Пример с гирями иллюстрирует главное свойство среднего арифметического: оно одинаково зависит от всех чисел набора. И большие, и малые числа входят в среднее арифметическое «с одинаковым весом». В некоторых случаях это достоинство оборачивается недостатком: если в наборе по ошибке или в силу каких-то особенных причин есть очень большое или очень малое значение, то оно одно очень сильно влияет на среднее арифметическое, и тогда среднее не очень хорошо описывает весь набор в целом.

Среднее арифметическое хорошо описывает однородные массивы данных, то есть массивы, в которых величины имеют один и тот же смысл, и нет значений, которые сильно отличаются от большинства.

Иногда среднее арифметическое используют для описания данных просто в силу сложившейся традиции. Хороший пример — школьные оценки. Чтобы вывести четвертную оценку, учителя используют среднее арифметическое, хотя оценки не являются числами, а поэтому сумма оценок не имеет математического смысла. Ведь если сложить двойку и тройку, то оценка «пять» не получится. Более того, если школьник вначале получал двойки, но позже разобрался в материале и стал заслуженно получать отличные оценки, среднее арифметическое всё равно будет низким: средняя оценка не учитывает, какие оценки получены раньше, а какие — позже. Средняя арифметическая четвертная и годовая оценки — дань традиции.

Вопросы

1. Дайте определение среднего арифметического числового набора.
2. Чему равно среднее арифметическое числового набора, все числа в котором одинаковы и равны 5,6?
3. Как можно описать среднее арифметическое с точки зрения физики?
4. Может ли среднее арифметическое числового набора быть больше, чем наибольшее значение в наборе; меньше, чем наименьшее?

В редакторах электронных таблиц на персональных компьютерах для вычисления среднего арифметического предусмотрена специальная функция

СРЗНАЧ()

На рисунке показан пример вычисления среднего значения массива из четырёх чисел.

fx =СРЗНАЧ(C1:C4)

C	D	E
1		
9		
5		
6	Среднее	5,25

Задачи

41 Найдите среднее арифметическое чисел:

а) 8 и 10; б) 8, 9 и 10.

42 Найдите среднее арифметическое чисел:

а) 3, 9 и 27; б) 6, 10, 16 и 20.

Сколько из данных чисел меньше среднего значения; больше среднего значения?

43 Придумайте какие-нибудь четыре разных числа, среднее арифметическое которых равно:

а) второму по величине числу;

б) третьему по величине числу;

в) полусумме второго и третьего по величине из этих чисел.

44 Придумайте какие-нибудь пять разных чисел, у которых среднее значение:

а) больше четырёх чисел, но меньше пятого;

б) больше первого числа, но меньше остальных четырёх.

45 Найдите среднее значение набора чисел, не вычисляя их сумму:

а) 13, 14, 15, 16, 17; г) 20, 25, 30, 35, 40;

б) 16, 17, 18, 19, 20; д) 22, 24, 26, 28, 30;

в) 21, 22, 23, 24, 25; е) 102, 104, 106, 108, 110.

46 Сергею нужно было найти среднее арифметическое всех натуральных чисел от 1 до 100. Он вычислил его так:

$$\frac{1+100}{2}=50{,}5.$$

а) Верно ли Сергей нашёл среднее арифметическое?

б) Приведите пример другого набора чисел, среднее арифметическое которого равно полусумме наименьшего и наибольшего чисел.

в) Приведите пример набора, для которого такой способ вычисления среднего даёт неверный результат.

47 Отметьте числа и их среднее арифметическое на числовой прямой:

а) 1, 2, 3, 4; в) 3, 4, 5, 6;

б) 2, 3, 4, 5; г) 10, 11, 12, 13.

48 Вычислите среднее арифметическое числового набора:

а) 2, 4, 7, 8, 9;

б) 20, 40, 70, 80, 90;

в) 200, 400, 700, 800, 900.

Числовые наборы б) и в) получены из набора чисел а) умножением всех чисел на 10 и на 100. Как средние значения наборов б) и в) можно получить из среднего значения набора а)?

49 Вычислите среднее арифметическое числового набора:

а) 2, 4, 7, 8, 9;

б) 10, 20, 35, 40, 45;

в) 50, 100, 175, 200, 225.

Числовые наборы б) и в) получены из набора чисел а) умножением всех чисел на 5 и на 25. Как средние значения наборов б) и в) можно получить из среднего значения набора а)?

50 В таблице 23 дана урожайность зерновых культур[1] в России за несколько лет.

Таблица 23. **Урожайность зерновых культур в России**

Год	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Урожайность зерновых, ц/га	22,7	18,3	22,4	18,3	22,0	24,1	23,7	26,2	29,2	27,2

а) Пользуясь таблицей 23, найдите среднюю урожайность зерновых культур в России за пять лет: с 2009 по 2013 г.

б) Найдите среднюю урожайность зерновых культур в России за пять лет: с 2014 по 2018 г.

в) Сравните среднюю урожайность за первые пять лет (2009—2013) и за следующие пять лет (2014—2018). Сильно ли различаются между собой эти средние значения, с вашей точки зрения?

51 а) Найдите среднюю урожайность зерновых культур в России за весь период с 2009 по 2018 г. по данным таблицы 23.

б) Назовём год из данного периода (см. табл. 23) **годом средней урожайности**, если урожайность зерновых в этот год отличается от средней менее чем на 5%. Какие годы были годами средней урожайности?

в) Назовём год из данного периода (см. табл. 23) **низкоурожайным**, если урожайность зерновых в этот год ниже средней больше чем на 10%. Какие годы были низкоурожайными? Каковы, по вашему мнению, причины низкой урожайности в эти годы?

52 В таблице 24 показано число жителей шести крупнейших городов Московской области (с населением более 200 тыс. чел.) в разные годы. Города перечислены в алфавитном порядке.

Таблица 24. **Население крупнейших городов Московской области, тыс. чел.**

Город \ Год	1959	1970	1979	2002	2010	2019
Балашиха	58,6	92,3	117,9	147,9	215,5	490,0
Королёв	41,4	105,9	133,5	142,6	183,4	224,5
Люберцы	93,3	139,4	159,6	156,7	172,6	207,3
Мытищи	98,7	118,7	140,7	159,9	173,2	222,7
Подольск	129,4	168,7	201,8	181,0	188,0	304,2
Химки	47,8	85,0	118,0	141,0	207,4	254,8

Найдите среднее число жителей крупнейших городов Московской области:

а) в 1959 г.; б) в 1970 г.; в) в 2010 г.; г) в 2019 г.

53 Назовём подмосковный город из таблицы 24 очень крупным, если численность населения в нём выше средней численности населения шести крупнейших городов области. Какие подмосковные города были очень крупными:

а) в 1959 г.; б) в 2019 г.?

Чем, по вашему мнению, можно объяснить изменение списка очень крупных городов?

[1] Урожайность зерновых культур — масса зерна, полученная с одного гектара посевов.

8 Медиана

В предыдущем пункте мы сказали, что среднее арифметическое хорошо описывает массивы однородных данных. Что делать, если в числовом наборе, который мы изучаем, встречаются **выбросы**, то есть одно или несколько чисел, которые намного больше или намного меньше всех остальных? В таких случаях в качестве центральной меры часто используют **медиану**.

ПРИМЕР 1. Возьмём какой-нибудь набор различных чисел, например: 4, 9, 1, 7, 11. Сначала упорядочим набор по возрастанию: 1, 4, 7, 9, 11. Упорядоченный набор называется **вариационным рядом**. Теперь найдём число, которое стоит посередине. Это число 7. Число 7 — медиана этого набора.

В этом примере набор состоял из пяти чисел. Медианой в этом случае оказывается число, стоящее в точности посередине.

ПРИМЕР 2. Рассмотрим набор 1, 3, 6, 11. Числа уже упорядочены, но их четыре, поэтому среди них нет числа, стоящего точно посередине. Возьмём два числа, стоящих посередине. Это числа 3 и 6. Любое из них, а также любое число между ними можно взять в качестве медианы. Чаще всего в качестве медианы берут среднее арифметическое двух центральных чисел: $\frac{3+6}{2}=4{,}5$. Обобщим эти два примера и сформулируем общее правило.

Чтобы найти медиану числового массива, нужно выполнить следующие действия.
1. Упорядочить массив по возрастанию. Получится вариационный ряд.
2. Если в массиве нечётное количество чисел, то медианой является число, стоящее посередине вариационного ряда.
3. Если в массиве чётное количество чисел, то медианой обычно считают среднее арифметическое двух чисел, стоящих посередине.

В пункт 3 следует внести поправку: если в массиве чётное количество чисел, то медиан у такого массива много — два срединных числа и все числа, заключённые между ними. Дадим теперь точное определение медианы.

Медианой числового массива называют такое число m, что хотя бы половина чисел массива не больше числа m и хотя бы половина чисел массива не меньше числа m.

ПРИМЕР 3. С помощью определения покажем, что число 6 является медианой числового набора

$$1,\ 6,\ 3,\ 2,\ 0,\ 4,\ 9,\ 12,\ 8,\ 6.$$

Всего в наборе 10 чисел, поэтому число 6 будет медианой, если в наборе найдутся 5 (или больше) чисел, которые не больше числа 6, а также найдутся 5 (или больше) чисел, которые не меньше числа 6.

Не будем упорядочивать числа, а просто подчеркнём все числа, которые не больше числа 6, а над всеми числами, которые не меньше числа 6, поставим черту сверху (надчеркнём их):

$$\underline{1},\ \overline{\underline{6}},\ \underline{3},\ \underline{2},\ \underline{0},\ \underline{4},\ \overline{9},\ \overline{12},\ \overline{8},\ \overline{\underline{6}}.$$

Обратите внимание: число 6 входит и в одно, и в другое множество — оно и подчёркнуто, и надчёркнуто одновременно.

Всего мы сделали 7 подчёркиваний и 5 надчёркиваний. Значит, в этом наборе 7 чисел, не бо́льших чем 6, и 5 чисел, которые не меньше чем 6. Требование определения выполнено, поэтому число 6 является медианой.

Определение точно говорит, что такое медиана, но использовать его для поиска медианы неудобно. Чтобы найти медиану, нужно действовать уже известным нам способом: упорядочить числа и найти одно или два числа, стоящие посередине вариационного ряда.

Пусть всего в ряду n чисел.

— Если n нечётно, то медианой будет число с порядковым номером $\frac{n+1}{2}$.

— Если n чётно, то медианой будет любое из чисел с номерами $\frac{n}{2}$ и $\frac{n}{2}+1$ или любое число между ними (чаще всего в качестве медианы берут среднее арифметическое этих чисел).

Мы сказали, что медиана часто используется тогда, когда нарушена однородность данных, то есть в массиве имеются выбросы. Чем же хороша медиана в таких случаях? Рассмотрим пример.

ПРИМЕР 4. В 2021 г. в России было 15 городов с числом жителей более 1 млн человек. Данные о населении этих городов приведены в таблице 25.

Таблица 25. Численность населения городов-миллионеров в России, тыс. чел.

Год / Город	2010	2021
Волгоград	1021	1004
Воронеж	890	1050
Екатеринбург	1350	1495
Казань	1144	1257
Красноярск	974	1092
Москва	11 504	12 655
Нижний Новгород	1251	1244
Новосибирск	1474	1620
Омск	1154	1139
Пермь	991	1049
Ростов-на-Дону	1089	1137
Самара	1165	1144
Санкт-Петербург	4880	5384
Уфа	1062	1125
Челябинск	1130	1187
Итого	31 079	33 582

Предположим, что мы хотим описать население российского города-миллионера одним числом. Найдём среднее арифметическое:

$$\frac{33\,582}{15} \approx 2239 \text{ тыс. чел.}$$

Обратите внимание: в таблице нет города, население которого было бы близко к получившемуся среднему. В большинстве городов население лишь немного превышает 1 млн человек. Исключение составляют Москва и Санкт-Петербург. Данные о населении Москвы и Санкт-Петербурга в этом массиве можно рассматривать как выбросы. Из-за этих двух значений среднее арифметическое оказалось намного больше, чем население типичного города-миллионера. Значит, среднее арифметическое не даёт верного представления о населении типичного крупного города России. Лучше использовать медиану.

Упорядочим значения:

1004 1049 1050 1092 1125 1137 1139 1144 1187 1244 1257 1495 1620 5384 12 655

Медианой является восьмое по порядку значение (выделено): 1144 тыс. чел. Это население г. Самары. Можно сказать, что Самара — медианный по численности город-миллионер в 2021 г., или **медианный представитель** данного набора.

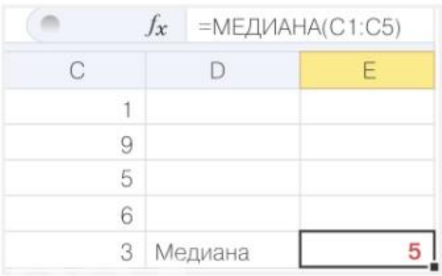
Мы увидели главное достоинство медианы — **устойчивость относительно выбросов**, то есть отдельных сильно выделяющихся значений. Если в наборе есть выбросы, то среднее арифметическое может не очень хорошо описывать большинство чисел. Медиана в таких случаях лучше описывает набор, чем среднее арифметическое.

Медиана имеет не только достоинства, но и недостатки. И дело не только в том, что для поиска медианы приходится упорядочивать набор. Гораздо существеннее, что приходится отбрасывать все значения, кроме одного-двух. Таким образом, теряется много полезной информации о массиве данных.

Чтобы найти медиану массива данных в электронных таблицах, удобно использовать функцию

МЕДИАНА()

На рисунке показан пример нахождения медианы.



Вопросы

1. В каких случаях среднее арифметическое не очень хорошо описывает большинство значений в числовом наборе? Приведите пример таких данных.
2. Дайте определение медианы числового набора.
3. В упорядоченном по возрастанию числовом наборе 19 чисел. Каким по счёту числом является медиана в этом наборе?
4. В упорядоченном по возрастанию числовом наборе 24 числа. Между какими двумя числами (по счёту) заключена медиана этого набора?
5. Что такое выброс?
6. В чём главное достоинство медианы как центральной меры?

Задачи

54 Найдите медиану и среднее арифметическое чисел:

а) 1, 3, 5, 7, 9; в) 1, 3, 5, 7, 9, 11;

б) 1, 3, 5, 7, 14; г) 1, 3, 5, 7, 9, 17.

55 Отметьте числа наборов и их медианы на числовой прямой:

а) 8, 11, 3; б) 7, 4, 8, 1, 5; в) 10, 3, 9, 8, 4, 5, 7.

56 Отметьте числа наборов и их медианы на числовой прямой:

а) 9, 11, 3, 17; б) 7, 4, 8, 1, 5, 6; в) 11, 3, 9, 8, 13, 4, 5, 7.

57 Найдите медиану набора чисел:

а) 11, 3, 21, 4, 17;

б) 25, 17, 19, 28, 18;

в) 25, 50, 25, 29, 27, 40, 28.

58 Найдите медиану набора чисел:

а) 9, 2, 8, 4;

б) 8, 9, 5, 7, 1, 3;

в) 12, 11, 18, 10, 22, 17, 11, 14.

59 Пользуясь таблицей 4 (с. 9), найдите медиану величины «время забега на 100 м» и медианных представителей, то есть бегунов, которые показали время, наиболее близкое к медианному значению.

60 Пользуясь таблицей 19 (с. 24), найдите медиану величины «площадь поверхности океана» и медианного представителя.

61 Пользуясь таблицей 24 (с. 35), ответьте на вопросы.

а) На сколько изменилось среднее число жителей крупнейших городов России к 2021 г. по сравнению с 2010 г.? Можно ли считать, что средняя численность населения выросла за эти одиннадцать лет?

б) Найдите медиану числа жителей городов в 2010 г. Сравните её с медианой, вычисленной для 2021 г. Найдите медианных представителей в эти годы.

62 Рассмотрите данные о числе жителей крупнейших городов России (см. табл. 25), исключив Москву и Санкт-Петербург.

а) Вычислите среднее значение числа жителей этих городов в 2021 г.

б) Вычислите медиану числа жителей этих городов в 2021 г.

в) Сильно ли, с вашей точки зрения, различаются медиана и среднее значение?

63 В таблице 23 (с. 35) даны сведения об урожайности зерновых культур в России в 2009—2018 гг. Найдите медиану урожайности и среднюю урожайность зерновых культур в России за период:

а) 2009—2018 гг.; б) 2009—2013 гг.; в) 2014—2018 гг.

Сравните между собой медиану и среднее за каждый период. Значительно ли, с вашей точки зрения, они отличаются друг от друга?

64 Средний рост учащихся в классе — 165 см. Медиана роста равна 168 см.

а) Обязательно ли не меньше половины учеников выше 165 см?

б) Обязательно ли не меньше половины учеников выше 168 см?

в) Обязательно ли найдётся в этом классе ученик ростом больше 165, но меньше 168 см?

г) Обязательно ли найдётся в этом классе ученик, рост которого ровно 168 см?

9 Наименьшее и наибольшее значения. Размах

Наименьшее и наибольшее значения

Иногда нужны не только среднее арифметическое или медиана, но и другие значения, характеризующие набор данных, например **наибольшее** и **наименьшее значения**.

Если мы хотим узнать, кто победил в соревнованиях по прыжкам в длину, то выберем того, кто прыгнул дальше всех, то есть выберем наибольший результат. Напротив, в соревнованиях по бегу победителем считается тот, кто пробежал быстрее всех, то есть пробежал дистанцию за наименьшее время.

Нам всегда интересно, какова наименьшая цена на нужный товар. Увидев новый автомобиль, мы интересуемся, какова его максимальная скорость. Иногда стремление к рекордам возникает в самых неожиданных ситуациях. В Книге рекордов Гиннесса можно найти и забавные достижения: кто дольше всех простоял на одной ноге, кто выпустил больше всего мыльных пузырей, у кого самый длинный нос и т. п.

ПРИМЕР 1. Во время подготовки к соревнованиям четыре спортсмена устроили мини-турнир по прыжкам в длину с места. Каждый из них сделал по пять попыток. Все результаты занесены в таблицу 26. Кроме того, в ней указаны средние результаты, а также наилучший и наихудший прыжки каждого спортсмена.

Таблица 26. **Результаты прыжков в длину с места, см**

Номер прыжка	Пётр	Иван	Алексей	Сергей
1	215	197	203	208
2	228	205	212	234
3	208	212	227	240
4	236	241	205	212
5	205	233	215	203
Среднее значение	218,4	217,6	212,4	219,4
Наибольшее значение	236	241	227	240
Наименьшее значение	205	197	203	203

Как вы думаете, почему результаты в разных попытках у одного и того же спортсмена разные? Какие из следующих факторов могут влиять на результат: удача; техника прыжка; рост; масса; тренированность; настроение; плотный обед; усталость; ветер; обувь? Какие ещё факторы могут повлиять на дальность прыжка?

Самое большое среднее значение прыжка у Сергея. По этому показателю Иван лишь третий. Но рекордсменом всё же стал Иван: в одной из попыток он прыгнул дальше всех.

Во многих спортивных дисциплинах принято учитывать только лучший показатель.

Тем не менее тренеру есть над чем задуматься: если упорядочить прыгунов по лучшему прыжку и по среднему значению, то получаются два разных упорядочивания. Сегодня лучшим был Иван, но Сергей уступает лишь немного, а в среднем он прыгает лучше. Кроме того, худший результат сегодня показал тоже Иван. Не исключено, что на соревнованиях лучше прыгнет Сергей.

Наибольшие и наименьшие значения важны не только в спорте.

ПРИМЕР 2. Каждая судоходная река имеет **фарватер** (судовой ход) — часть русла, где возможно движение судов. Течение несёт с собой песок и ил, глубины на разных участках фарватера меняются, поэтому их приходится промерять несколько раз в год. Каждый раз составляется список глубин, которые наносятся на карту. Важно знать наименьшую глубину, чтобы понять, может ли судно пройти этот участок реки.

ПРИМЕР 3. В Санкт-Петербурге за всё время наблюдений с 1706 г. случилось более 300 наводнений. Самый высокий уровень воды (421 см выше **ординара**[1]) зарегистрирован 7 ноября 1824 г. Это наводнение описано в поэме А. С. Пушкина «Медный всадник».

В Санкт-Петербурге наблюдения за уровнем воды в Неве ведутся всё время. Наивысший известный уровень воды учитывается при строительстве набережных, дамб, мостов не только в Санкт-Петербурге, но и в любом другом месте, где существует опасность наводнения.

Во многих случаях наибольшее и наименьшее значения — неудачные характеристики. Они часто не являются типичными (вспомните пример с населением городов-миллионеров, где Москва и Санкт-Петербург сильно выбиваются из общего ряда). Иногда наибольшее и наименьшее значения оказываются попросту ошибочными. Ошибка, например, может возникнуть при вводе данных в компьютер.

ПРИМЕР 4. В таблице 27 собраны данные о росте учащихся класса.

Наибольший рост у Евсеевой — 1154 см! Больше чем 11 метров! Ясно, что это измерение — выброс, получившийся по ошибке. Скорее всего, кто-то при вводе данных в компьютер случайно нажал лишнюю единицу. Лучше не гадать, как получилось такое значение, а исключить его вовсе. Тогда наибольшее значение станет равно 167 см.

Таблица 27. **Рост школьников**

Фамилия	Рост, см	Фамилия	Рост, см	Фамилия	Рост, см	Фамилия	Рост, см
Алексеев	156	Вольский	158	Евсеева	1154	Коваль	154
Андреева	159	Гетманов	161	Железов	167	Петровская	149
Борисов	162	Добромыслов	156	Завидов	163	Юсуфов	165

Посмотрим на медиану. До исключения ошибочной записи медиана равнялась 160 см, а после исключения стала равна 159 см (проверьте).

Этот пример показывает неустойчивость наибольшего и наименьшего значений и ещё раз иллюстрирует устойчивость медианы.

Измерение рассеивания данных с помощью размаха

Часто нужно знать не только среднее значение в наборе данных, но и иметь представление о том, как сильно значения разбросаны, рассеяны. Самой простой характеристикой, описывающей **рассеивание** данных, является **размах**.

[1] Ординар — средний многолетний уровень воды в водоёме.

Низкий заряд
Заряда остается 20 %.
Режим энергосбережения
Закрыть

Размах числового массива — это разность между наибольшим и наименьшим значениями.

ПРИМЕР 5. Вспомним таблицу 22 с ценами на один и тот же смартфон в разных магазинах. Наибольшая цена 8590 р., а наименьшая — 7790 р. Размах цен в этом массиве данных равен 8590 – 7790 = 800 р., то есть меньше, чем 10% средней цены, которая равна 8150 р.

Размах — очень простая и наиболее употребительная мера рассеивания. Но для вычисления размаха используются только наименьшее и наибольшее значения, которые неустойчивы. Поэтому и размах — неустойчивая мера.

ПРИМЕР 6. Вернёмся к примеру 4, где у ученицы Евсеевой рост 1154 см. До удаления этого ошибочного значения размах был равен 1005 см, а после удаления он стал 18 см (проверьте).

Наибольшее и наименьшее значения нередко попадают в набор данных по ошибке. Начиная работать с данными, полезно обратить внимание на наибольшее и наименьшее значения — правдоподобны ли они?

Для поиска наименьшего и наибольшего значений в электронных таблицах есть функции

МИН() и МАКС()

Чтобы найти размах, нужно вычислить разность:

= МАКС() – МИН()

fx =МАКС(C1:C5)-МИН(C1:C5)

C	D	E	F
1			
9			
5	Наим.	1	
6	Наиб.	9	
3	Размах	8	

Вопросы

1 Приведите примеры данных, для описания которых лучше использовать наибольшее значение, чем среднее или медиану.

2 Приведите примеры данных, для описания которых лучше использовать наименьшее значение, чем среднее или медиану.

3 Что такое размах числового набора?

Задачи

65 Найдите наибольшее и наименьшее значения, размах, среднее значение и медиану набора чисел:

а) 12, 7, 25, 3, 19, 15; б) 17, 19, 5, 41, 47, 13, 19.

66 Пользуясь таблицей 2 (с. 8), укажите:

а) самый большой по числу жителей в 2021 г. город России;

б) второй по населению город в России в 2021 г.;

в) третий и четвёртый по числу жителей города в России в 2010 г.

67 В таблице 28 приведены данные о производстве зерновых культур в России в 2011—2020 гг.

Таблица 28. **Производство зерновых культур в России**

Год	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Производство зерновых, млн т	94,2	70,9	92,4	105,2	104,7	120,7	135,5	113,2	120,6	133,0
Урожайность зерновых, ц/га	22,4	18,3	22,0	24,1	23,7	26,2	29,2	27,2	26,6	28,6

Найдите наибольшее, наименьшее значения и размах:

а) производства зерновых культур;

б) урожайности зерновых культур.

68 При сборке автомобильного двигателя нужно добиться того, чтобы все поршни двигателя имели одинаковую массу (размах должен быть не более 0,1 г). Увеличить массу поршня нельзя, зато её можно уменьшить, высверливая углубления в специальных местах поршня. В таблице 29 показаны массы 8 поршней для одного двигателя.

а) Определите размах масс поршней.

б) Какой поршень не требует доработки?

Таблица 29. **Массы поршней**

Поршень	1	2	3	4	5	6	7	8
Масса, г	124,4	124,8	125,2	123,9	124,1	125,4	125,2	124,8

69 В таблице 30 даны результаты измерения температуры тела пациента в больнице.

Таблица 30. **Измерение температуры**

Время	7 ч	9 ч	11 ч	13 ч	15 ч	17 ч	19 ч	21 ч
t, °C	38,3	39,2	39,2	39,4	39,1	38,7	381	38,2

а) Найдите наибольшее значение температуры, размах, среднее арифметическое и медиану температуры.

б) Найдите явно ошибочное значение. Как оно могло получиться?

в) Исключите ошибочное значение и найдите наибольшее значение температуры, размах, среднее арифметическое и медиану температуры после исключения ошибки.

г) На сколько градусов изменился размах после исключения ошибки?

д) На сколько изменилось среднее значение после исключения ошибки?

е) На сколько градусов изменилась медиана после исключения ошибки?

70 Как изменится размах числового набора, если:

а) к каждому числу набора прибавить 5;

б) от каждого числа набора отнять 3?

71 Как изменится размах числового набора, если:

а) наименьшее число набора уменьшить на 50;

б) наибольшее число набора увеличить на 100?

72 Все числа в наборе положительны. Как изменится размах числового набора, если:

а) каждое число набора разделить на 5;

б) каждое число умножить на 3?

73 Все числа в наборе отрицательны. Как изменится размах числового набора, если:

а) каждое число набора разделить на 5;

б) каждое число умножить на 3?

10* Обозначения в статистике. Свойства среднего арифметического

Обозначения

Данные в статистических массивах часто приходится обозначать буквами. Использовать отдельную букву для каждого числа невозможно. Поэтому для чисел одного массива обычно используют одну и ту же букву с **индексами** — номерами. Можно рассмотреть набор X, в котором пять чисел x_1, x_2, x_3, x_4, x_5, или набор Y, состоящий из чисел y_1, y_2, y_3, y_4, y_5, y_6, и т. п. Чтобы подчеркнуть, что числа образуют один набор, иногда записывают значения в фигурных скобках.

ПРИМЕР. Набор чисел 1, 8, 3, 0, −3 можно обозначить X и записать: $X = \{1, 8, 3, 0, -3\}$. В этом наборе $x_1 = 1$, $x_2 = 8$ и так далее до $x_5 = -3$.

Если чисел в наборе много, то вместо слов «и так далее до» используют многоточие. Например, набор X, в котором 100 чисел, записывают так:

$$X = \{x_1, x_2, x_3, \ldots, x_{100}\}.$$

Среднее арифметическое чисел набора принято обозначать $\overline{x}$. Например, среднее арифметическое набора, в котором пять чисел x_1, x_2, x_3, x_4, x_5, запишется так:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5},$$

а среднее арифметическое набора X, состоящего из n чисел, можно записать, используя многоточие:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Для наименьшего значения набора обычно используется обозначение min, а для наибольшего — max (от лат. *minimum* и *maximum*). Таким образом, можно записать:

$$\min\{2, 4, 1, 5\} = 1, \ \max\{2, 4, 1, 5\} = 5.$$

Если набор {2, 4, 1, 5} обозначить буквой X, то можно записать короче:

$$\min X = 1, \ \max X = 5.$$

Обозначения не только укорачивают запись, но и делают её ясной, однозначной и строгой.

Для некоторых описательных характеристик нет общепринятых обозначений. Например, нет общепринятого обозначения для медианы числового массива.

Свойства среднего арифметического

Буквенные обозначения позволяют записать и доказать свойства среднего арифметического числового массива.

Свойство 1. Если каждое число набора увеличить (уменьшить) на одно и то же число a, то среднее арифметическое набора увеличится (уменьшится) на это же число a.

Аналогичное свойство верно для умножения.

Свойство 2. Если каждое число набора умножить на одно и то же число b, то среднее арифметическое набора также умножится на число b.

Докажем эти свойства на примере массива из четырёх чисел:

$$X = \{x_1,\ x_2,\ x_3,\ x_4\}.$$

Прибавим к каждому значению число a. Получится набор

$$\{x_1 + a,\ x_2 + a,\ x_3 + a,\ x_4 + a\}.$$

Нужно показать, что среднее арифметическое нового массива равно $\overline{x} + a$. Найдём среднее арифметическое:

$$\frac{(x_1 + a) + (x_2 + a) + (x_3 + a) + (x_4 + a)}{4} = \frac{x_1 + x_2 + x_3 + x_4 + 4a}{4}.$$

Разобьём дробь на два слагаемых:

$$\frac{x_1 + x_2 + x_3 + x_4}{4} + \frac{4a}{4}.$$

Первое слагаемое равно $\overline{x}$ — среднему арифметическому первоначального набора X. Второе слагаемое равно a. Получается, что среднее арифметическое нового массива равно $\overline{x} + a$.

Покажем теперь, что верно свойство 2. Умножим каждое число набора X на число b. Нужно показать, что среднее арифметическое нового набора bx_1, bx_2, bx_3, bx_4 равно $b\overline{x}$.

Вычислим среднее арифметическое нового набора:

$$\frac{bx_1 + bx_2 + bx_3 + bx_4}{4} = \frac{b(x_1 + x_2 + x_3 + x_4)}{4} = b \cdot \frac{x_1 + x_2 + x_3 + x_4}{4} = b\overline{x}.$$

Доказательство окончено.

Вопросы

1 Как называются номера в обозначениях x_5, y_2 и т. п.?

2 Как обозначается среднее арифметическое, наименьшее и наибольшее значения?

3 Сформулируйте свойство 1 среднего арифметического.

4 Сформулируйте свойство 2 среднего арифметического.

Задачи

74 Дан числовой набор X, состоящий из чисел 17, 3, 6, 21, 15.

а) Чему равно x_2 в этом наборе? б) Чему равно x_5?

75 Среднее арифметическое числового набора X равняется 5. Найдите среднее арифметическое числового набора, который получится, если:

а) ко всем числам набора X прибавить число 4;

б) из всех чисел набора X вычесть число 12;

в) ко всем числам набора X прибавить число 8;

г) из всех чисел набора X вычесть число 3.

76 Среднее арифметическое числового набора Y равняется 8. Найдите среднее арифметическое числового набора, который получится, если все числа набора Y:

а) умножить на число 2; в) умножить на −3;

б) разделить на 4; г) разделить на −3.

77 Сначала ко всем числам числового набора X прибавили число 8, а затем все полученные числа умножили на 3. Найдите среднее арифметическое получившегося набора, если среднее арифметическое набора X равно:

а) 2; б) −4; в) 5,2; г) −9,1.

78 Сначала все числа числового набора X умножили на 3, а затем к каждому полученному числу прибавили 8. Найдите среднее арифметическое получившегося набора, если среднее арифметическое набора X равно:

а) 2; б) −4; в) 5,2; г) −9,1.

79 Средняя цена нефти на протяжении месяца составляла 65 долл. за один **баррель** (мера объёма, равная 158,9873 л.). Найдите среднюю цену одного литра нефти за этот период, считая, что в одном барреле 159 л.

80 Средняя зарплата на предприятии составляла 56 000 р. С нового года зарплату всем сотрудникам проиндексировали (повысили) на 4%. Других изменений не было. Найдите среднюю зарплату после индексации.

81 В наборе n чисел. На сколько увеличится среднее арифметическое этого набора, если одно число в этом наборе увеличить на 1?

82 В наборе было 100 чисел, а сумма всех чисел равнялась 134. К набору добавили ещё одно число, при этом среднее арифметическое не изменилось. Какое число добавили?

83 Каждое число набора увеличили на одно и то же число a. Как изменились наименьшее и наибольшее значения набора?

84 Наименьшее значение набора равно m, наибольшее значение равно M. Каждое число набора умножили на одно и то же число b. Найдите наименьшее и наибольшее значения получившегося набора, если: а) $b > 0$; б) $b < 0$.

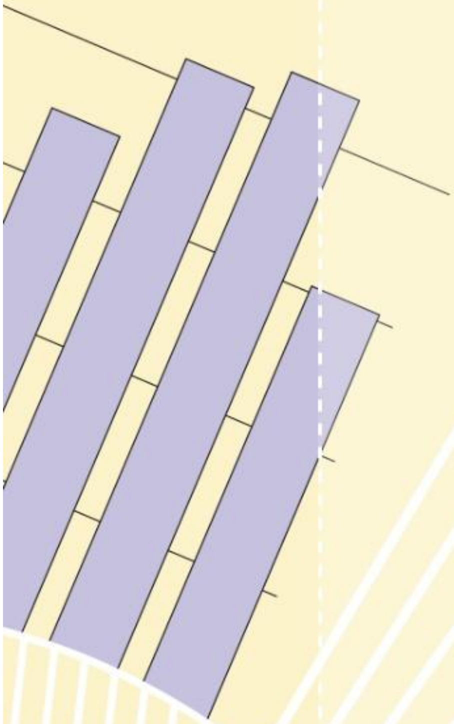
85 В наборе было n чисел, их среднее арифметическое равнялось $\overline{x}$. К набору добавили число a. Запишите выражение для среднего арифметического получившегося набора.

86 Студент в течение семестра выполнил несколько контрольных работ. Когда он узнал результаты всех работ, кроме последней, он заметил, что если за последнюю контрольную работу он получит 90 баллов, то средний балл за все работы будет равен 84, а если за последнюю работу он получит 72 балла, то средний балл за все работы составит 81. Сколько всего контрольных работ должен выполнить студент?

III Случайная изменчивость

Неизменные величины в жизни встречаются крайне редко. Даже те величины, которые считаются постоянными, обычно подвержены изменчивости. Помимо закономерной изменчивости почти всегда присутствует разнонаправленная случайная изменчивость, причины которой известны частично, а порой неизвестны вовсе.

В этой главе мы обсудим несколько важных примеров изменчивости, которую приходится учитывать в повседневной жизни. Кроме того, мы увидим, как разные виды изменчивости отражаются на диаграммах.



11 Примеры случайной изменчивости

В природе неизменные величины встречаются очень редко. Большинство величин подвержены **случайной изменчивости**. Иногда мы можем указать причины изменений. Иногда причины изменчивости известны частично, а порой неизвестны вовсе.

Колебания напряжения в электрической сети

В таблице 31 даны результаты 25 измерений напряжения в бытовой электросети. Все измерения сделаны днём, в случайно выбранные моменты.

Таблица 31. **Измерения напряжения в бытовой сети, В**

225	225	227	225	228
228	218	217	218	223
225	216	222	220	218
221	220	214	219	231
228	227	220	224	216

В России **номинальное**[1] напряжение в бытовых сетях 220 В (вольт). Напряжение в сети редко равно в точности 220 В. Оно колеблется. Напряжение понижается, если в вашей квартире или у соседей включаются электроприборы. Моменты включения и выключения электрических приборов случайные, и потому колебания напряжения тоже случайные.

Кроме того, производитель электроэнергии не может обеспечить напряжение в точности 220 В, а при транспортировке электричества по линиям электропередач неизбежны потери, которые тоже непостоянны.

Электрические приборы в России рассчитаны на колебания напряжения в определённых пределах. Если вы посмотрите на заднюю панель микроволновой печи или холодильника, вы найдёте табличку, где написан интервал рабочего напряжения. Например, от 190 до 250 В. Если напряжение выходит за эти пределы, прибор может выйти из строя. Поэтому в некоторых случаях люди используют стабилизаторы напряжения, которые сглаживают колебания напряжения.

Урожайность зерновых культур

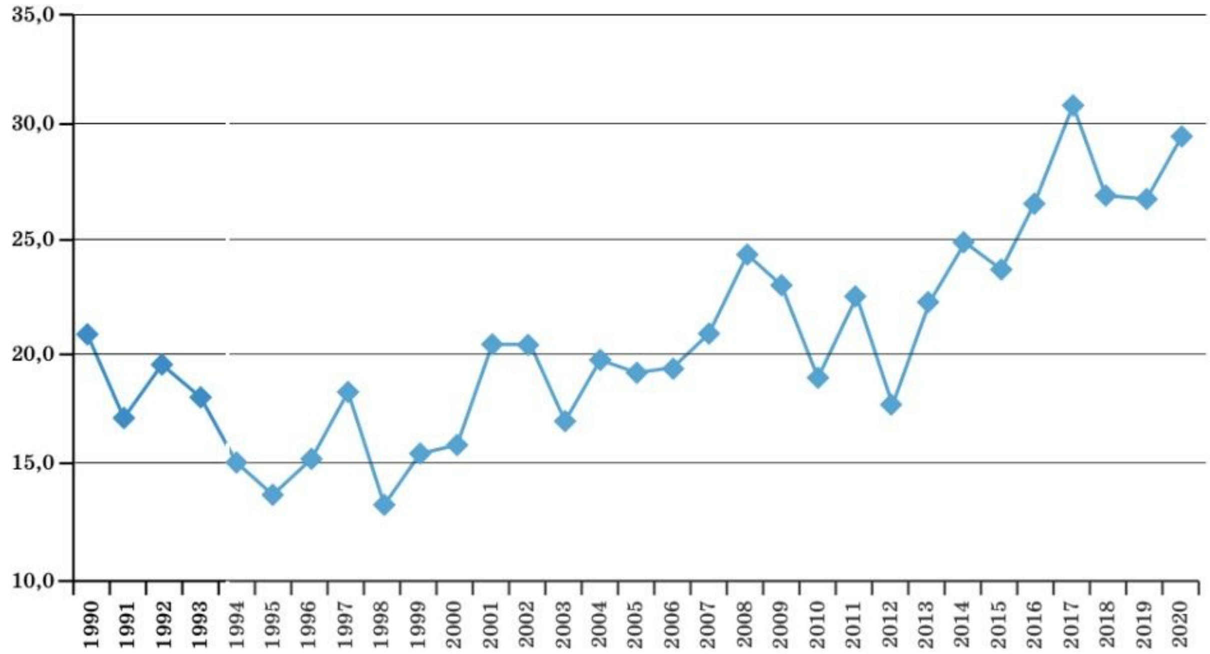
Урожайность зерновых культур — это средняя масса зерна, полученного с одного засеянного гектара. Это важный экономический показатель, который отражает эффективность сельского хозяйства.

На диаграмме 16 видно, что урожайность — очень изменчивая величина. Большое влияние оказывают погодные условия, которые год от года разные.

Но кроме погоды, урожайность зависит от усилий людей, качества посевного зерна, удобрений, состояния уборочной техники и многих других факторов. Усилия людей по повышению урожайности сказываются постепенно. За счёт улучшения хозяйствования средняя урожайность к нашему времени выросла практически в 2 раза по сравнению с серединой 1990-х гг. и в 3 раза по сравнению с 1950-ми гг.

[1] Номинальное — заявленное, стандартное, то, которое должно быть.

Диаграмма 16. **Урожайность зерновых культур в России, ц/га**



Массовое производство

На обёртке шоколадного батончика написано, что его масса 50 г. Это номинальная масса. В таблице 32 даны массы двадцати купленных одинаковых батончиков.

Таблица 32. **Масса шоколадных батончиков, г**

49,1	50,0	49,7	50,5	48,1	50,3	49,7	51,6	49,8	50,1
49,7	48,8	51,4	49,1	49,6	50,9	48,5	52,0	50,7	50,6

Наибольшее значение — 52,0 г, а наименьшее — 48,1 г. Размах — 3,9 г. И только один батончик весит в точности 50 г. Но средняя масса всех двадцати батончиков равна 50,01 г, то есть практически не отличается от номинальной.

Такая ситуация часто встречается при массовом производстве. Если отклонение размера, массы мало отличается от заданного стандарта, то есть находится в пределах **допустимой погрешности**, то изделие считается годным. Оно поступает в продажу или дальнейшее производство. Если отклонение превышает допуск, то изделие считается бракованным.

Невозможно и не нужно контролировать массу каждого отдельного изделия. Достаточно проверять контрольные партии. Большие или систематические отклонения говорят об износе оборудования или о низком качестве сырья. В этих случаях приходится останавливать производство и устранять причину.

Для разных изделий допуски разные. Для продуктов питания допуски могут достигать нескольких граммов, для деталей мебели — 1—2 мм, а для точных изделий, например подшипников, допуски не превышают сотых долей миллиметра.

Вопросы

1 Приведите примеры изменчивых величин, помимо тех, что описаны в тексте учебника.

2 Какие факторы влияют на значение напряжения в электросети в случайный момент времени? Назовите ещё один-два фактора, помимо тех, которые приведены в тексте учебника.

3 Какие факторы влияют на изменчивость урожайности зерновых культур? Назовите ещё один-два фактора, кроме тех, что приведены в тексте учебника.

4 Какие факторы могут влиять на изменчивость массы шоколадного батончика с арахисом?

5 Что больше подвержено изменчивости — масса каждого отдельного шоколадного батончика или средняя масса в партии батончиков?

Задачи

87 Рассмотрите таблицу 31. Найдите наибольшее и наименьшее значения напряжения.

88 Найдите среднее значение и медиану напряжения по данным таблицы 31. Сильно ли эти характеристики, по вашему мнению, отличаются от 220 В?

89 Сколько значений из таблицы 31 превышают номинальное и сколько меньше номинального? Можно ли считать, что шансы событий «напряжение в случайный момент выше 220 В» и «напряжение в случайный момент ниже 220 В» примерно одинаковы?

90 По данным диаграммы 16 найдите средние урожайности зерновых за первые три года и за последние три года. Сравните результаты. На сколько центнеров с гектара средняя урожайность в последние три года превышает среднюю урожайность в первые три года?

91 Рассмотрите таблицу 32. Сколько в купленной партии батончиков массой более 50 г? Какую долю и какой процент они составляют?

92 Масса купленного шоколадного батончика может быть больше или меньше номинальной. Можно ли считать, что шансы этих событий равны, если судить по данным из таблицы 32?

12 Точность и погрешность измерений

Можно точно узнать, сколько учеников в классе, или подсчитать деревья в сквере. Но чаще всего измерения или подсчёты невозможно выполнить абсолютно точно. Возникают неизбежные **погрешности**, то есть случайные отклонения от истинного значения. Главный источник погрешности — изменчивость самой измеряемой величины. Кроме того, измерительный инструмент также обладает погрешностью. Обсудим это подробнее.

Число жителей города

Численность населения крупных городов измеряется обычно в тысячах человек (см. табл. 2 на с. 8). Мы понимаем, что число жителей городов дано приближённо, что истинное число жителей, скорее всего, не равно в точности целому числу тысяч.

Но что такое «истинное число жителей города»? Люди всё время приезжают и уезжают, кто-то умирает, кто-то появляется на свет. Поэтому число жителей города постоянно меняется. Если подсчитывать людей ежедневно в маленьких городах, то различия составят десятки человек, в средних городах — сотни. Но число тысяч людей, населяющих город среднего размера, за день, скорее всего, не изменится. Поэтому данные о населении крупных городов обычно даются в тысячах человек. Более высокая точность не имеет смысла[1].

Рост человека

Когда мы говорим о росте человека, мы округляем данные до сантиметра. Так измеряют рост в России и в континентальной Европе. В США, Великобритании, Австралии и некоторых других странах рост измеряют в **дюймах**. Один дюйм (1") — это примерно 2,54 см.

Рост человека редко равен целому числу сантиметров или дюймов. Казалось бы, можно измерить рост с большой точностью. Но это не нужно, измерение с точностью до сантиметра или дюйма достаточно для практических нужд: пошива одежды, определения размеров мебели, сидений в автомобилях и т. п.

Кроме того, рост человека не остаётся постоянным в течение суток: утром рост человека чуть больше, чем вечером. За день под влиянием нагрузки хрящи в суставах несколько сжимаются, а во время сна вновь расправляются. Поэтому само понятие «рост человека» не очень точно определено.

Расстояние между городами

Мы понимаем, как измерить расстояние между двумя точками вдоль некоторой линии. Но город — это не точка, а обширная территория. К тому же граница города не всегда отчётлива. Поэтому расстояние между городами точно определить нельзя.

В России за расстояние между городами принимают расстояние между их центральными почтовыми отделениями (почтамтами) вдоль главных дорог. Этот способ определения расстояния появился одновременно с регулярной почтовой службой.

Например, протяжённость автомобильной дороги М10 «Россия» между Москвой и Санкт-Петербургом равна 664 км. А длина автомагистрали М11 «Нева» между этими же городами равна 669 км. Железнодорожники считают[2], что между Москвой и Санкт-Петербургом 650 км. Пилоты регулярных рейсов измеряют расстояние от Москвы до Санкт-Петербурга в соответствии со своим полётным планом, который может меняться даже в ходе полёта[3].

Выбирать точность измерения изменчивых величин нужно так, чтобы погрешность не влияла на последующие выводы.
Слишком высокая точность измерения не нужна, а иногда даже вредна. Излишне точные измерения отнимают время, силы и даже могут порождать ошибки.

[1] В данных Российской службы статистики можно встретить измерения численности населения городов с точностью до одного человека. Это объясняется тем, что данные не округляются. При этом обычно указывается метод, с помощью которого проведён пересчёт.

[2] Фактически расстояние составляет 645 км с 2001 г. после спрямления участка около п. Веребье в Маловишерском районе Новгородской области. Официально протяжённость трассы решили не менять. И теперь 204-й км железной дороги граничит с 211-м км.

[3] Схема взлёта и захода на посадку зависит от обстановки в воздухе и направления ветра. Диспетчер может сообщить пилоту об изменениях в любой момент.

Тем не менее измерительные приборы должны обеспечивать необходимую точность. Допустимая погрешность измерительного прибора определяется стандартом.

ПРИМЕР 1. Допустимая погрешность весов зависит от **класса точности** весов. Чем выше класс точности, тем меньше погрешность и тем дороже весы.

Не очень точные весы (III класс точности), которые рассчитаны на груз от 20 до 3000 кг, в соответствии с ГОСТ[1] имеют разные допустимые погрешности при разной нагрузке (табл. 33).

Таблица 33. Допустимая погрешность весов при эксплуатации

Нагрузка, кг	Допустимая погрешность, кг
От 20 до 200	±1
От 200 до 2000	±2
От 2000 до 3000	±3

Знак ± означает, что отклонение возможно в любую сторону — в меньшую или в бóльшую. Например, если человек массой ровно 65 кг взвешивается на напольных весах, то они могут показать от 64 до 66 кг. Наоборот, если напольные весы показали, что человек имеет массу 65 кг, то истинная масса человека m находится в пределах от 64 до 66 кг. Это удобно записать двойным неравенством

$$64 \text{ кг} \leqslant m \leqslant 66 \text{ кг}.$$

У лабораторных весов допустимые погрешности намного меньше. Например, точные лабораторные весы могут иметь допустимую погрешность всего 0,001 г.

Таблица 34. Допустимая погрешность автомобильного спидометра

Диапазон скорости, км/ч	Допустимая погрешность при $T = 20 \pm 5$ °C, км/ч
0—60	+4
60—80	+5
80—100	+6
100—120	+7
120—140	+8

ПРИМЕР 2. Согласно ГОСТу автомобильный спидометр при температуре окружающего воздуха 20 ± 5 °C не должен иметь погрешности в меньшую сторону, а допустимые погрешности в большую сторону зависят от скорости и указаны в таблице 34. Обратите внимание: погрешность в бóльшую сторону допускается, а в меньшую — нет. Это сделано для того, чтобы ошибка спидометра не приводила к превышению скорости.

До сих пор мы приводили пример **абсолютных погрешностей**. Иногда погрешности измеряются в процентах самой величины. Это **относительные погрешности**. Например, допустимая погрешность длины верёвки или шнура в мотке обычно даётся в процентах.

Рисунок 6

На рисунке 6 — катушка, на которую намотан шнур. На ярлыке указана длина шнура: 250 м ± 10%. Это значит, что истинная длина шнура находится в пределах от $250 \cdot 0{,}9 = 225$ (м) до $250 \cdot 1{,}1 = 275$ (м). Это можно выразить двойным неравенством

$$225 \text{ м} \leqslant l \leqslant 275 \text{ м}.$$

Здесь l — истинная длина шнура.

[1] ГОСТ (государственный стандарт) — совокупность требований и норм, определяющая свойства и качество продукции.

Вопросы

1 Можно ли совершенно точно определить понятие «численность населения страны»?

2 Приведите несколько факторов, влияющих на изменчивость числа жителей страны. Как вы думаете, зачем нужно знать численность населения страны?

3 Почему размеры готовой мебели обычно указывают в сантиметрах, а не в миллиметрах?

4 Приведите примеры измерений длины, когда точность ±1 мм недостаточна.

5 С какой точностью, на ваш взгляд, следует измерять напряжение в бытовой электросети?

6 Укажите подходящую единицу измерения массы морского судна; железнодорожного вагона; шоколадки; человека.

7 Температуру в России и в большинстве других стран измеряют в градусах Цельсия[1]. Укажите подходящую точность измерения для температуры воздуха в комнате или на улице; тела человека; пламени в газовой плите; температуры на поверхности звезды.

Задачи

93 Рассмотрите таблицу 33. В ней даны допустимые погрешности весов в разных диапазонах измерения. Предположим, что весы исправны и погрешность измерения не выходит за пределы допустимой. Запишите с помощью двойного неравенства границы, в которых находится истинное значение массы m некоторого груза, если при взвешивании этого груза весы показывают:

а) 38 кг; в) 2956 кг;
б) 129 кг; г) 2543 кг.

94 Рассмотрите таблицу 34. Предположим, что автомобильный спидометр исправен. Укажите с помощью двойного неравенства границы истинной скорости v автомобиля, если спидометр показывает:

а) 28 км/ч; в) 96 км/ч;
б) 69 км/ч; г) 127 км/ч.

95 На мотке верёвки указано, что длина верёвки составляет 30 м ± 5%. В каких пределах может быть заключена истинная длина верёвки?

13 Тенденции и случайные отклонения

В п. 11 мы обсуждали напряжение в электрической сети, которое редко бывает равно номинальному значению 220 В. Похожие явления встречаются повсюду.

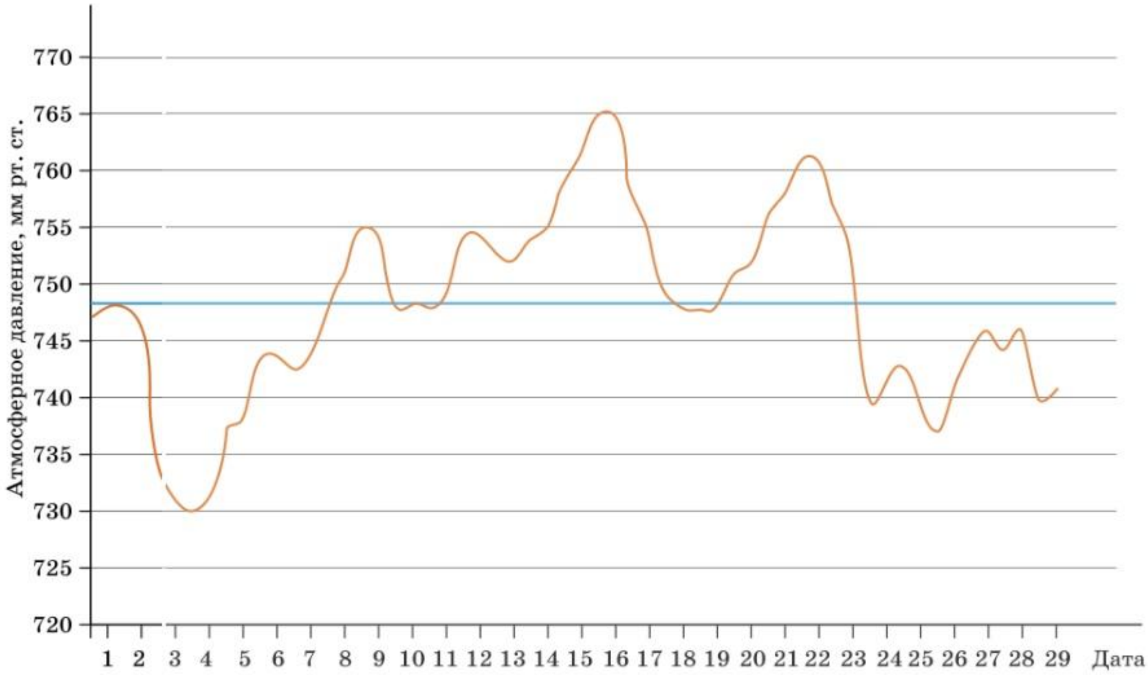
ПРИМЕР 1. Нормальное атмосферное давление[2] в каждой местности своё: оно зависит от широты, высоты над уровнем моря, от температуры и влажности воздуха и ещё — от случайности.

[1] В США, Белизе и в нескольких других странах температуру обычно измеряют в градусах Фаренгейта.

[2] На уровне моря в местности с географической широтой 45° при температуре воздуха вблизи земной поверхности 0 °С нормальным считается давление 760 мм рт. ст.

На диаграмме 17 оранжевая линия показывает изменение атмосферного давления на протяжении февраля 2020 г. в г. Ижевске. Давление меняется плавно, непрерывно, поэтому график — непрерывная линия.

Диаграмма 17. Атмосферное давление в Ижевске. Февраль 2020 г.



Синей прямой показано среднее давление 748,4 мм рт. ст. График давления колеблется около среднего значения. При этом само среднее давление случается только в редкие и короткие моменты времени. Можно говорить лишь о бóльших или меньших отклонениях давления от среднего.

Глядя на причудливую форму графика, можно подумать, что давление воздуха крайне неустойчиво и всё время сильно меняется. Но давайте посмотрим, каков размах данных по отношению к среднему. Наименьшее давление — 730 мм рт. ст., а наибольшее — 765 мм рт. ст. Размах равен 35 мм рт. ст., что составляет $35 : 748{,}4 \cdot 100\% \approx 4{,}7\%$. Получается, что на самом деле давление колеблется вблизи постоянного значения, и изменчивость здесь не очень высокая.

Иногда случайная величина в целом растёт или убывает со временем, при этом испытывая случайные колебания. Говорят, что изменчивость складывается из **тенденции** и случайных колебаний.

Тенденция (тренд) — характерное, устойчивое изменение. Как правило, тенденция обусловлена долгосрочными факторами, которые заставляют величину расти или убывать.

В примере с атмосферным давлением можно сказать, что тенденция в феврале отсутствовала или была нулевой.

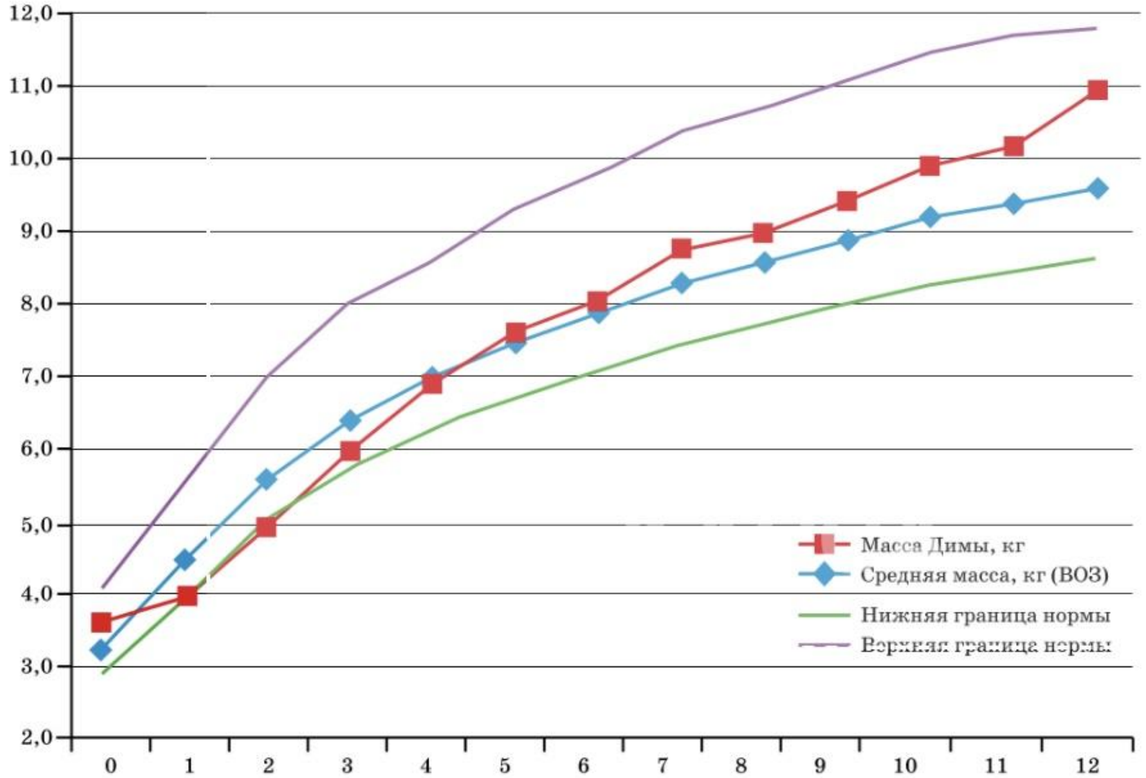
В отличие от тенденции, случайные колебания непредсказуемы, поскольку вызваны кратковременными случайными и разнонаправленными факторами.

ПРИМЕР 2. На диаграмме 18 показано, как растёт масса мальчиков от рождения до года. Эти данные медицинской статистики получены ВОЗ[1] в России за много лет наблюдений. Средняя ожидаемая масса увеличивается плавно — тенденция показана синей линией. Красная линия — масса одного конкретного мальчика по имени Дима по данным ежемесячного взвешивания в районной поликлинике. Дима родился немного тяжелее среднего, но за первый месяц набрал совсем немного и «ушёл ниже среднего». Затем Дима нагнал сверстников, а в конце первого года жизни вдруг резко прибавил почти килограмм. Можно ли считать, что в развитии Димы на первом году жизни были отклонения от нормы?

Каждый ребёнок развивается индивидуально, ежемесячный привес зависит от множества факторов, и практически никогда не следует средним значениям. Границы нормы очень условны, и ни у родителей, ни у врачей опасений по поводу питания или здоровья маленького Димы на первом году жизни не было.

В случае с массой **возрастающая тенденция** определена природой — ребёнок растёт, и масса его тела увеличивается. Но каждый конкретный ребёнок может отставать от тенденции или обгонять её. Эти случайные колебания вызваны индивидуальными особенностями, наследственностью, образом жизни, составом питания и многими другими факторами, часть из которых неизвестна.

Диаграмма 18. **Масса мальчиков от рождения до года, кг**

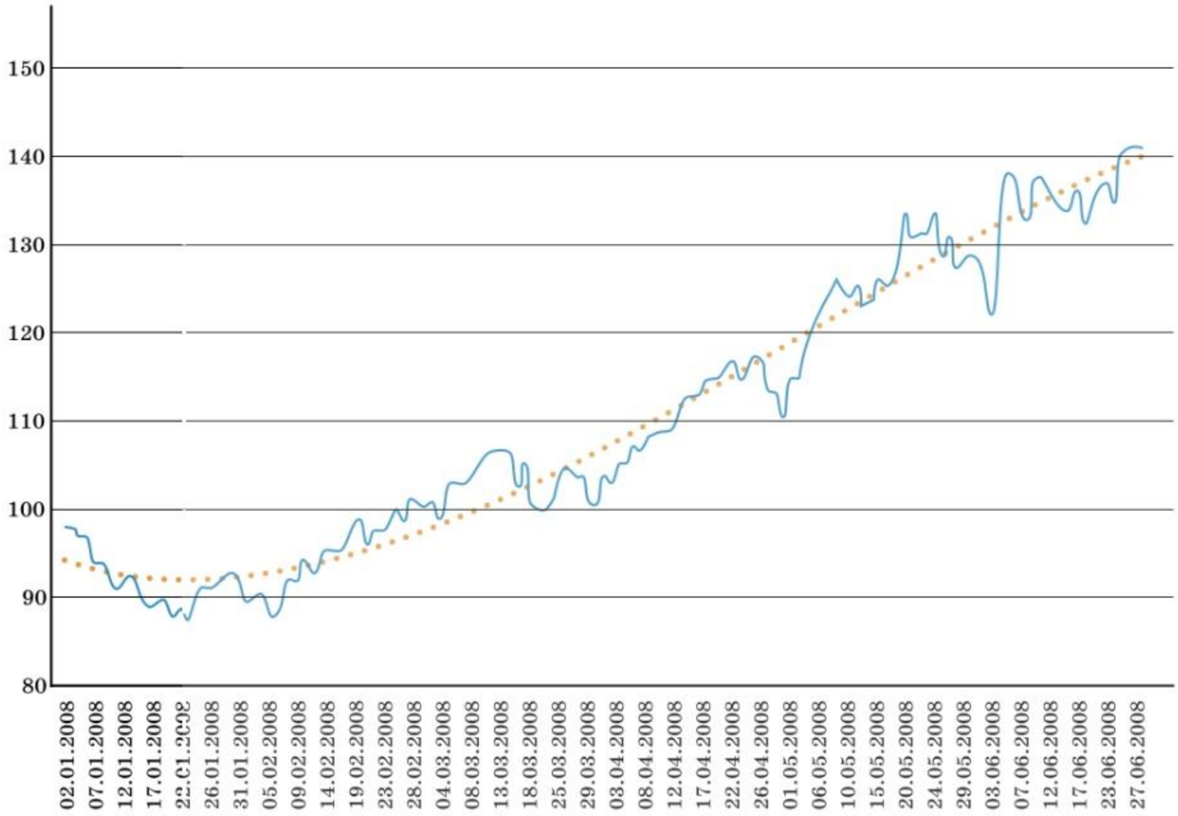


Среднюю массу ребёнка в каждом возрасте можно определить, взвешивая и измеряя тысячи мальчиков и девочек в школах, детских садах и поликлиниках. Поэтому общие тенденции в развитии мальчиков и девочек хорошо известны.

---
[1] ВОЗ — Всемирная организация здравоохранения.

Однако бывает так, что изменчивая величина не имеет аналогов ни в прошлом, ни в настоящем, а потому невозможно сказать, насколько она соответствует норме. Таких уникальных, изменчивых явлений много в экономике, где условия всё время меняются, и настоящее нельзя сравнивать с прошлым. Поэтому прогнозы в экономике делать намного труднее, чем предсказывать погоду.

Диаграмма 19. **Цена барреля нефти в первом полугодии 2008 г.**



ПРИМЕР 3. Цена нефти на мировых биржах традиционно измеряется в долларах США за баррель. Мы выбрали два периода — первое полугодие 2008 г., когда цены на нефть росли, и второе полугодие 2014 г., когда они снижались (диагр. 19 и 20).

Оранжевыми пунктирными линиями на обеих диаграммах показаны тенденции. На диаграмме 19 — по большей части **возрастающая** тенденция. На диаграмме 20 — **убывающая**. Но эти тенденции не были известны заранее, а были определены позже на основе уже сформировавшихся цен.

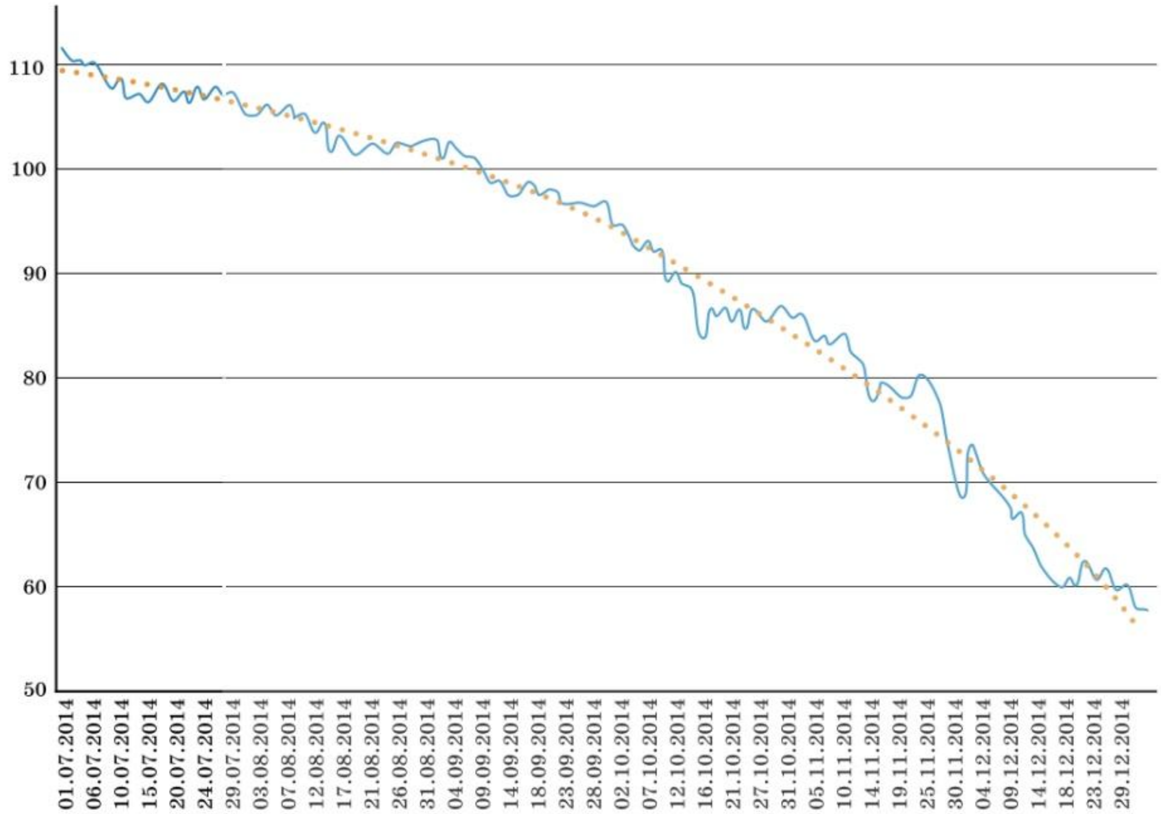
Мы видим, что в январе — июне 2008 г. цена на нефть в целом росла. Но были дни, когда случались падения: синяя линия, показывающая реальные цены, сильно колеблется, прыгает вверх и вниз.

Точно так же на диаграмме 20 при общем снижении цен часто наблюдаются незначительные, но резкие краткосрочные повышения.

Общее повышение не всегда означает постоянное повышение. Общее понижение не всегда означает постоянное понижение.

Какие факторы формируют тенденции в изменениях цены барреля нефти? В начале 2008 г. бурно развивалось строительство, крупные банки охотно давали кредиты, экономика быстро росла, и ей требовалось всё больше нефти. Это определило тенденцию: повышение спроса привело к постоянному росту цен. В 2014 г. сказался мировой кризис: экономическая ситуация в большинстве стран ухудшалась, спрос на нефть падал, поэтому она постепенно дешевела.

Диаграмма 20. Цена барреля нефти во втором полугодии 2014 г.



А какие факторы обусловливают кратковременные скачки цен? Внезапное решение какой-нибудь крупной компании о покупке нефти или отказ от ожидаемой покупки, задержка нефтеналивного танкера в океане, заявление влиятельного политика или экономиста, пожар на морской нефтедобывающей платформе, открытие нового месторождения и даже непроверенные слухи могут вызвать краткосрочное изменение цены в ту или иную сторону.

Часто изменчивость в природных, экономических и социальных явлениях состоит из двух составляющих. Первая составляющая — тенденция, которая обусловлена серьёзными и долгосрочными факторами. Вторая составляющая — случайные отклонения, вызванные разнонаправленными краткосрочными действиями, которые зачастую невозможно предвидеть.

Одна из важных задач статистики — научиться выделять тенденции в изменчивых явлениях и отличать их от незначительных случайных колебаний.

Вопросы

1 Приведите свой пример величины, имеющей постоянное среднее значение, вокруг которого наблюдаются случайные колебания.

2 Что такое тенденция?

3 Приведите свой пример величины, имеющей возрастающую тенденцию.

4 Какие ещё факторы, помимо перечисленных в тексте, по вашему мнению, могут влиять на изменение массы растущего младенца?

14 Частоты значений в массивах данных

ПРИМЕР 1. Рассмотрим числовой набор 1, 3, 1, 2, 1, 5, 3, 2, 1, 1.

В этом наборе 10 чисел, но различных чисел только четыре: 1, 2, 3 и 5. Единица встречается 5 раз. Можно сказать, что **частота** числа 1 в этом наборе равна $\frac{5}{10}$, то есть 0,5. Аналогично частоты значений 2 и 3 равны 0,2, а частота значения 5 равна 0,1. Заметим, что сумма частот равна 1:

$$0{,}5 + 0{,}2 + 0{,}2 + 0{,}1 = 1.$$

Пусть в наборе всего N чисел, и значения, равные a, встречаются N_a раз. **Частотой** значения[1] a называется отношение $\frac{N_a}{N}$.

Чтобы в электронной таблице подсчитать, сколько раз в наборе встречается определённое значение, используйте функцию

СЧЁТЕСЛИ()

На рисунке подсчитано, что в данном наборе число 1 встречается 5 раз.

fx =СЧЁТЕСЛИ(B1:D5;1)

B	C	D	E
0	1	3	5
1	2	2	
1	2	1	
1	3	2	
2	4	4	

Частоты значений можно подсчитывать не только в числовых наборах, но и там, где значения не являются числами.

ПРИМЕР 2. Обратимся к таблице 8 (с. 12) с результатами подсчёта домашних животных у школьников одного класса. Различными значениями в нашем наборе являются виды животных, а также значение «Никого». Найдём их частоты. Для этого подсчитаем общее количество значений: 9 + 11 + 7 + 3 + 2 + 1 = 33.

Значение «Собака» встречается 9 раз. Поэтому частота этого значения равна $\frac{9}{33} \approx 0{,}273$. В статистике принято обыкновенные дроби превращать в десятичные с округлением при необходимости. Десятичные дроби удобнее писать в строку, с ними легче выполнять действия и их легче сравнивать, чем обыкновенные дроби.

[1] Частота бывает не только у значения в наборе. Позже мы познакомимся с частотой события.

Аналогично найдём частоты прочих значений и занесём их в таблицу 35.

Таблица 35. **Частоты значений**

Животное	Всего	Частота
Собака	9	0,273
Кошка	11	0,333
Никого	7	0,212
Рыбка	3	0,091
Птица	2	0,061
Черепаха	1	0,030
Сумма	33	1

Обратите внимание на последнюю строку. Сумма частот снова равна единице.

Свойство частот. В любом наборе сумма частот значений равна единице.

Докажем это свойство для набора, в котором четыре различных значения a, b, c и d, причём каждое может встречаться несколько раз. Предположим, значение a встречается N_a раз, а значения b, c и d встречаются N_b, N_c и N_d раз соответственно.

Тогда всего в наборе $N = N_a + N_b + N_c + N_d$ значений. Частота значения a равна $\frac{N_a}{N}$. Частота значения b равна $\frac{N_b}{N}$ и т. д. Найдём сумму частот:

$$\frac{N_a}{N} + \frac{N_b}{N} + \frac{N_c}{N} + \frac{N_d}{N} = \frac{N_a + N_b + N_c + N_d}{N} = \frac{N}{N} = 1.$$

Аналогично свойство доказывается для набора, в котором произвольное количество значений. Пользуясь тем, что сумма частот всегда равна единице, удобно делать промежуточную проверку в вычислениях.

ПРИМЕР 3. Рассмотрим набор, в котором есть одинаковые значения. Например, оценки по математике, которые некоторый школьник получил в течение четверти:

$$3,\ 4,\ 3,\ 5,\ 4,\ 3,\ 4,\ 4,\ 4,\ 2,\ 3,\ 5,\ 3,\ 3,\ 4,\ 5,\ 2,\ 4,\ 4,\ 4.$$

Всего 20 оценок. Чтобы вывести четвертную оценку, учитель находит среднее арифметическое: делит сумму всех чисел на 20. Можно поступить иначе. Запишем в таблицу 36 не все числа, а только различные значения и их частоты.

Таблица 36. **Частоты оценок**

Значение	2	3	4	5
Частота	0,1	0,3	0,45	0,15

Полезно проверить, равна ли сумма частот единице: $0,1 + 0,3 + 0,45 + 0,15 = 1$.

Теперь найдём среднее арифметическое: умножим значения на их частоты и сложим произведения:

$$2 \cdot 0,1 + 3 \cdot 0,3 + 4 \cdot 0,45 + 5 \cdot 0,15 = 0,2 + 0,9 + 1,8 + 0,75 = 3,65.$$

Убедитесь в том, что обычный способ даст то же самое среднее.

Связь между частотами значений и средним арифметическим*

ТЕОРЕМА. **Пусть в наборе N чисел, но среди них только k различных значений $x_1, x_2, \dots, x_k$. Пусть значение x_1 встречается N_1 раз, значение x_2 встречается N_2 раз, и т. д.: значение x_k встречается ровно N_k раз. Тогда среднее арифметическое набора равно**

$$\bar{x} = x_1 \cdot \frac{N_1}{N} + x_2 \cdot \frac{N_2}{N} + \dots + x_k \cdot \frac{N_k}{N}.$$

Теорема утверждает, что среднее арифметическое числового массива равно сумме произведений значений и их частот. Для доказательства составим сумму из произведений различных чисел и их частот:

$$x_1 \cdot \frac{N_1}{N} + x_2 \cdot \frac{N_2}{N} + \dots + x_k \cdot \frac{N_k}{N} = \frac{x_1 \cdot N_1 + x_2 \cdot N_2 + x_3 \cdot N_3 + \dots + x_k \cdot N_k}{N} = \bar{x}.$$

Получилось среднее арифметическое. Теорема доказана.

Вопросы

1. Чему равны частоты значений в наборе, где 10 различных значений, но каждое встречается ровно один раз?
2. Сформулируйте определение частоты значения.
3. Сформулируйте свойство частот.
4. Как найти среднее значение числового набора, зная различные значения и их частоты?

В редакторах электронных таблиц можно быстро вычислить сумму произведений значений и их частот. Для этого есть функция

СУММПРОИЗВ()

На рисунке показано решение задачи из примера 3.

fx =СУММПРОИЗВ(D2:G2;D3:G3)

C	D	E	F	G
Значение	2	3	4	5
Частота	0,1	0,3	0,45	0,15
			Среднее	3,65

Задачи

96 Дан числовой набор 5, 4, 8, 1, 1, 3, 4, 5, 8, 1. Найдите частоту:
а) значения 1; б) значения 4.

97 Дана последовательность букв: ФАВЫДОВЛЫДЯЮФЛЧЛИОЫДСОЫЖФ.
Найдите в этой последовательности частоту:
а) буквы Д; в) буквы Ы;
б) буквы Ф; г) буквы С.

98 В числовом наборе 5 значений. Частоты четырёх значений известны: 0,35, 0,2, 0,1 и 0,05. Найдите частоту пятого значения.

99 В таблице 14 (с. 16) даны четвертные оценки учащихся класса.
а) Найдите частоты различных четвертных оценок по математике. Составьте таблицу значений и частот.
б) Вычислите среднюю оценку по математике за четверть.