

МАТЕМАТИЧЕСКАЯ ВЕРТИКАЛЬ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И СТАТИСТИКА. 7 КЛАСС (2 ч./нед.)

Уроки 52–53. Итоговое повторение. Описательная статистика и случайная изменчивость

Документ содержит задачи для итогового повторения материала по статистике за 7 класс. Учитель может на свое усмотрение использовать материалы целиком или частично наряду с собственными разработками и материалами учебника¹. Авторы будут благодарны за замечания и предложения по структуре и содержанию сценариев.

Глава 3. Описательная статистика

Содержание. Наибольшее и наименьшее значение, размах. Меры центральной тенденции: среднее арифметическое, медиана, среднее геометрическое, среднее гармоническое. Отклонения и рассеивание данных. Дисперсия.

Требования. В результате изучения этой группы тем школьник должен понимать следующее.

Числовые данные сначала нужно подходящим способом представить, а затем — описать. Однородные данные удобно описывать одним числом — центральной мерой (или мерой центральной тенденции). Для этого нужны такие показатели, как наибольшее и наименьшее значение, середина отрезка данных, размах данных, медиана, среднее арифметическое, среднее геометрическое, среднее гармоническое и другие. Выбор среднего совершают исходя из цели исследования, природы данных и зачастую из сложившихся традиций.

Наименьшее и наибольшее значения в некоторых случаях являются естественными описательными параметрами (спортивные рекорды и т.п.), но зачастую они оказываются наименее надежными показателями, поскольку подвержены сильной случайной изменчивости.

Размах — наиболее простая мера рассеивания данных. Он показывает, как далеко отстоят друг от друга наибольшее и наименьшее значения. Размах часто применяют в повседневной жизни для описания интервала значений. Однако размах также подвержен случайной изменчивости. Более полной характеристикой разброса данных является весь набор отклонений значений от некоторого среднего.

¹ Математика 7–9 класс. Теория вероятностей и статистика / Ю. Н. Тюрин, А. А. Макаров, И. Р. Высоцкий, И. В. Яценко. — 3-е изд., стереотипное. — М.: МЦНМО: ОАО «Московские учебники», 2011. — 256 с.: ил.

Среднее арифметическое имеет ясный физический смысл — точка равновесия массива данных. Его разумно применять для суммируемых величин (объем производства, доходы, расходы, численность населения).

Однако среднее арифметическое не подходит для прогнозов в случае, когда величина сильно меняется во времени (имеет тенденцию роста или падения). Кроме того, оно неустойчиво к выбросам, так как зависит от всех значений массива данных. Более устойчивой к выбросам характеристикой является медиана.

Среднее геометрическое используется, когда нужно усреднить коэффициент пропорционального изменения величины. Это касается в первую очередь данных о процентном изменении величины.

Среднее гармоническое используется, когда нужно найти среднюю скорость прохождения участка пути, трудоёмкость (например, среднее время на создание единицы продукции) и в других подобных случаях.

Дисперсия — мера рассеивания данных, равная среднему квадрату отклонения от среднего арифметического. Она обладает рядом свойств, которые делают дисперсию важным и удобным инструментом.

Задачи

1. Дан числовой набор:

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

Найдите среднее арифметическое, медиану, наибольшее и наименьшее значение, середину интервала значений, размах, среднее геометрическое, среднее гармоническое.

2. Какая мера центральной тенденции является естественной для следующих данных:

- а) дальность прыжка (прыжки в длину);
- б) время передачи файла;
- в) время забега на 100 м;
- г) коэффициенты изменения цен на нефть или другое сырьё?

3. В штате отдела 20 сотрудников. Средняя зарплата — 30 тысяч рублей. Чему будет равна средняя зарплата, если:

- а) Одному из сотрудников поднять зарплату на 2 тыс. руб.;
- б) Всем сотрудникам увеличить зарплату на 2 тыс. руб.;
- в) Одному сотруднику увеличить зарплату на 5 тыс. руб., а другому — уменьшить на эту же сумму;
- г) Принять на работу нового сотрудника на зарплату 51 тыс. руб.

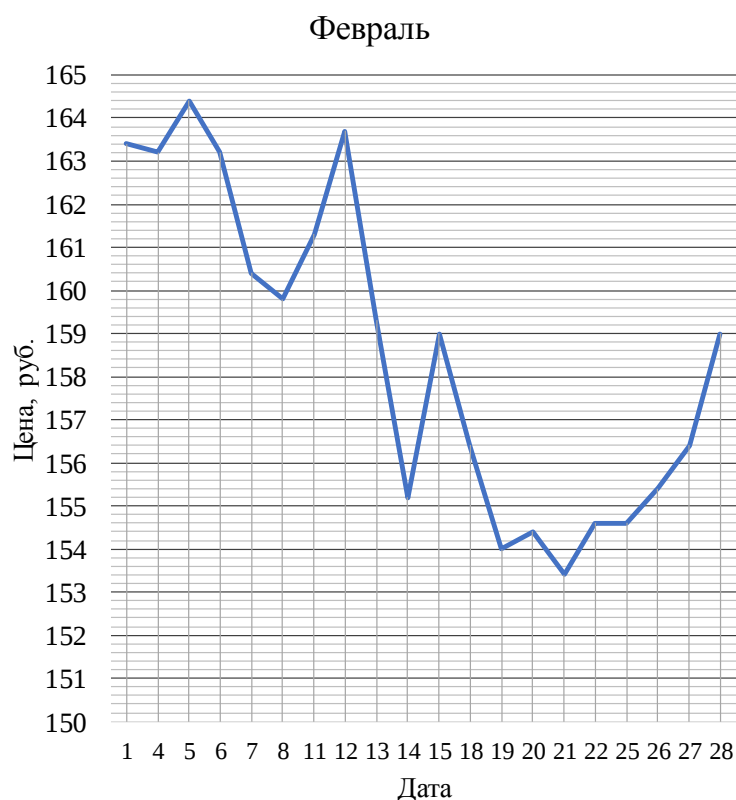
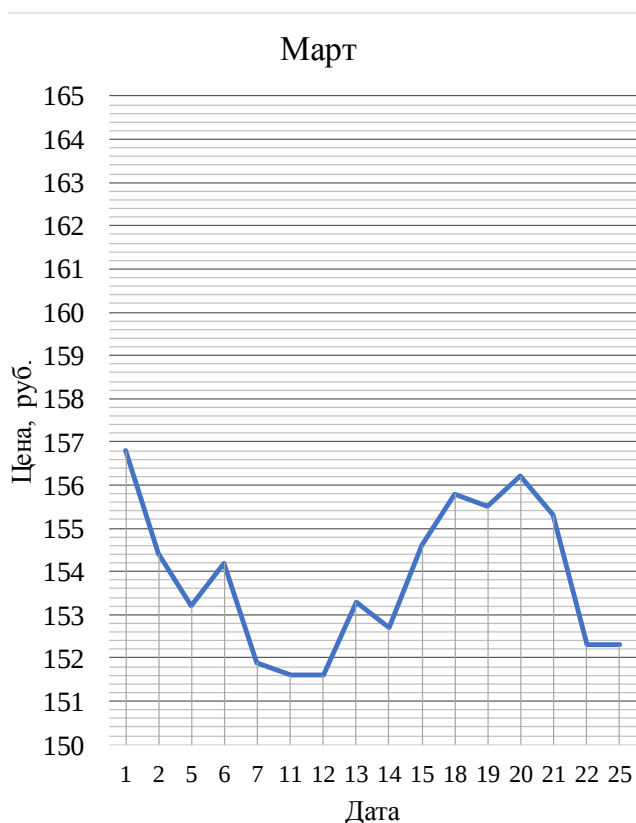
4. В городе восемь школ. В таблице приведена численность учащихся в этих школах на начало учебного года. Найдите среднее арифметическое и медиану данных. Какой из этих показателей лучше отражает среднюю численность учащихся школы этого города? Свое мнение обоснуйте.

Школа	Кол-во учащихся
№1	843
№2	726
№3	915
№4	123
№5	779
№6	811
№7	77
№8	896

5. В таблице содержатся данные о количестве проданных пар обуви каждого размера в обувном магазине за день. Найдите среднее арифметическое размеров проданной обуви.

Размер	35	36	37	38	39	40	41	42
Кол-во	7	28	37	44	34	21	17	12

6. На диаграммах представлены данные о курсе акций компании Газпром (GAZP) за февраль и март 2019 года.



а) В каком месяце размах цены больше?

б) Клиент биржи продал 1000 акций GAZP по самой высокой цене в феврале, а потом купил 1000 акций — но уже по самой низкой цене в феврале. Какую прибыль он получил?

7. В таблице приведены данные о средней температуре в каждом месяце года в Воркуте. Данные собраны в результате многолетних наблюдений.

Месяц	Средняя температура, °С
Январь	–19,4
Февраль	–19,7
Март	–13,8
Апрель	–10
Май	–1,9
Июнь	7,6
Июль	13,1
Август	9,7
Сентябрь	4,4
Октябрь	–3,2
Ноябрь	–13
Декабрь	–17,4

- Найдите среднюю температуру за год.
- Составьте таблицу отклонений температуры от средней температуры.
- Чему равна сумма отклонений температуры от среднего?
- Найдите дисперсию данных. Результат округлите до сотых.

8. В России 15 городов-миллионников (с числом жителей свыше 1 млн. чел.). В таблице собраны данные о населении и площади этих городов России.

Город	Население, тыс. чел.	Площадь, кв.км	Плотность населения, тыс. чел./кв.км
Москва	12330	2561,5	
Санкт-Петербург	5225	1403	3,72
Новосибирск	1603	505,6	3,17
Екатеринбург	1456	468	3,11
Нижний Новгород	1262	460	2,74
Казань	1232	515,8	
Уфа	1121	707,9	1,58
Челябинск	1199	530	2,26
Омск	1178	573	2,06
Самара	1170	541,4	2,16
Ростов-на-Дону	1125	354	3,18
Красноярск	1083	348	3,11
Пермь	1048	803	1,31
Воронеж	1040	596,5	1,74
Волгоград	1013	859,4	1,18

- Средняя площадь города-миллионника равна приблизительно 748,5 кв.км. Найдите медиану площади.
- Найдите медианного представителя — город, площадь которого ближе всего к медиане.

в) Какой из найденных показателей — среднее арифметическое или медиана — лучше характеризует площадь типичного российского города-миллионника? Кратко обоснуйте свое мнение.

г) В последнем столбце таблицы содержатся данные о плотности населения в этих городах. Найдите плотность населения в Москве и в Казани. Результаты округлите до сотых.

д) Будем считать город-миллионник густонаселенным, если его плотность населения выше 2,5 тыс. чел. на кв. км. Какие города-миллионники можно считать густонаселенными в соответствии с этим правилом?

9. Таблица содержит данные о средней по России цене на дизельное топливо на момент начала года в период с 2015 по 2019 годы.

Год	2015	2016	2017	2018	2019
Цена за литр, р.	34	35	36,7	40,5	47,5
Динамика, %		+2,9	+4,9	+10,4	+17,3

Найдите среднее годовое изменение цены на дизельное топливо в процентах. Ответ округлите до десятых.

10. Компания занимается сборкой и продажей ПК. В сборочном цехе работают пять человек. В таблице представлены данные о результатах работы цеха сборки за месяц. В столбце «Рабочие часы» указано суммарное отработанное время за месяц, в столбце «Единицы продукции» содержатся данные о количестве собранных компьютеров.

Сотрудник	1	2	3	4	5
Рабочие часы	156	145	126	128	144
Единицы продукции	52	58	45	40	60
Трудоемкость					

Заполните строку «Трудоемкость» и найдите среднюю трудоемкость (в часах на единицу продукции) за этот месяц. Результаты округлите до десятых.

11. Швейцарские часы испытывают на точность с помощью специального теста. В ходе теста определяется ошибка измерения времени (в секундах на протяжении суток) при разной температуре, влажности и в разных положениях механизма. Часы получают сертификат точности, если размах ошибки меньше 5 секунд за сутки, а дисперсия меньше 3.

Если ошибка в ту или иную сторону превышает 2 секунды, то часы нуждаются в регулировке. В таблице даны результаты пяти испытаний отдельного часового механизма.

Номер испытания	1	2	3	4	5
Ошибка, с	−1,1	−2,7	−0,8	−5,5	−2,9

а) Найдите среднюю ошибку, размах и дисперсию ошибки.

б) Определите, получают ли эти часы сертификат точности и нуждаются ли они в регулировке.

12. Среднее значение набора чисел равно 4, а дисперсия равна 18. Каждое число набора изменили на противоположное. Найдите среднее значение и дисперсию нового набора.

13. В лаборатории производится анализ крови. Содержание гемоглобина в крови вычисляется как среднее арифметическое результатов нескольких измерений. Таблица содержит результаты пяти измерений гемоглобина (г/л) в одной пробе крови пациентки.

Номер измерения	1	2	3	4	5
Содержание гемоглобина, г/л	130	140	110	50	120

а) Найдите среднее арифметическое и дисперсию результатов измерений.

Выбрано правило: если квадрат отклонения некоторого значения от среднего арифметического превышает дисперсию более чем в 3,5 раза, то это значение считается ненадежным (выбросом) и в дальнейшем не учитывается.

б) Определите, есть ли среди данных измерений ненадежные.

в) Найдите среднее арифметическое надежных значений.

г) Нормальное содержание гемоглобина в крови у женщин 120 – 150 г/л. Можно ли считать, что у данной пациентки нормальное содержание гемоглобина?

Глава 4. Случайная изменчивость

Содержание. Примеры случайной изменчивости, точность измерений. Гистограммы (диаграммы частот).

Требования. В результате изучения этой группы тем школьник должен понимать следующее.

В жизни в основном мы имеем дело с изменчивыми величинами. Чаще всего причины изменчивости известны лишь частично. Поэтому говорят о случайной изменчивости.

В производстве или при проведении измерений изменчивость чаще всего является следствием ошибок: систематических и случайных.

Систематическая ошибка не создает разброса данных, но приводит к значительному отличию среднего значения от номинального (массы, объема, напряжения в сети и т.п.).

Случайные ошибки разнонаправлены — они дают отклонения то в меньшую, то в большую сторону, компенсируя друг друга. Поэтому случайные ошибки создают рассеивание значений, но мало влияют на среднее.

Также на изменчивость величины влияет точность измерений. Точность измерения для разных величин должна быть разной. Слишком большая точность может быть экономически невыгодна, и даже вредна. Измерения бывают непосредственными и косвенными.

В природе и в обществе встречаются самые разные закономерности в изменчивых величинах. Один из способов их изучения — гистограммы (диаграммы частот). По форме гистограммы можно не только строить прогнозы, но иногда удаётся делать предположения о родственности разных изменчивых величин, вовсе непохожих с первого взгляда.

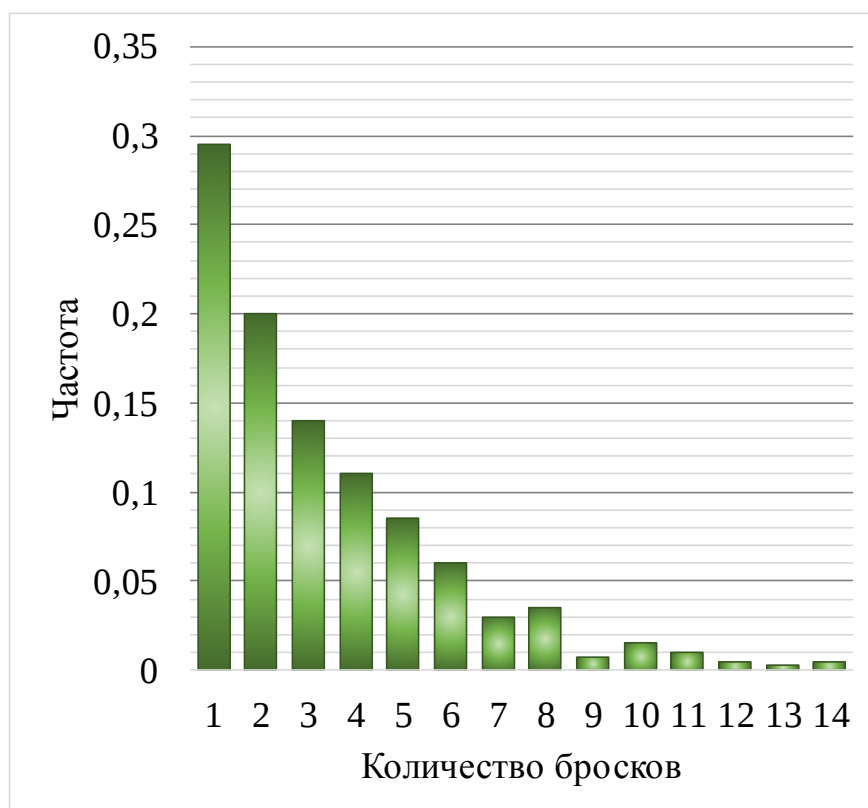
Задачи

14. На шоколадной фабрике стоят два одинаковых дозирующих автомата. Каждый автомат должен отмерять порции шоколада массой 100 г. На фабрике производится проверка оборудования. Для этого инженеры взвесили с большой точностью по 10 порций, отмеренных каждым автоматом. Результаты измерений представлены в таблице.

Автомат	Номер пробы	1	2	3	4	5	6	7	8	9	10
1	Масса, г	102,1	109,3	98,4	99,2	105,6	103,7	97,1	101,7	102,4	95,7
2	Масса, г	96,7	95,1	97,1	96,8	97,3	96,5	96,8	97,2	95,6	97,3

- С какой точностью измерили массу порции?
- Попробуйте сделать выводы о состоянии дозирующих автоматов. Нуждается какой-то из этих автоматов в ремонте или в регулировке?

15. Катя проводила серию экспериментов. Она бросала игральный кубик до тех пор, пока не выпадет 3 или 6 очков. В таблицу она записывала, на каком по счету броске это случилось. Так она сделала много раз. На рисунке представлена гистограмма, построенная по результатам эксперимента. По горизонтали отмечено число понадобившихся бросков, а по вертикали — частота этого события.



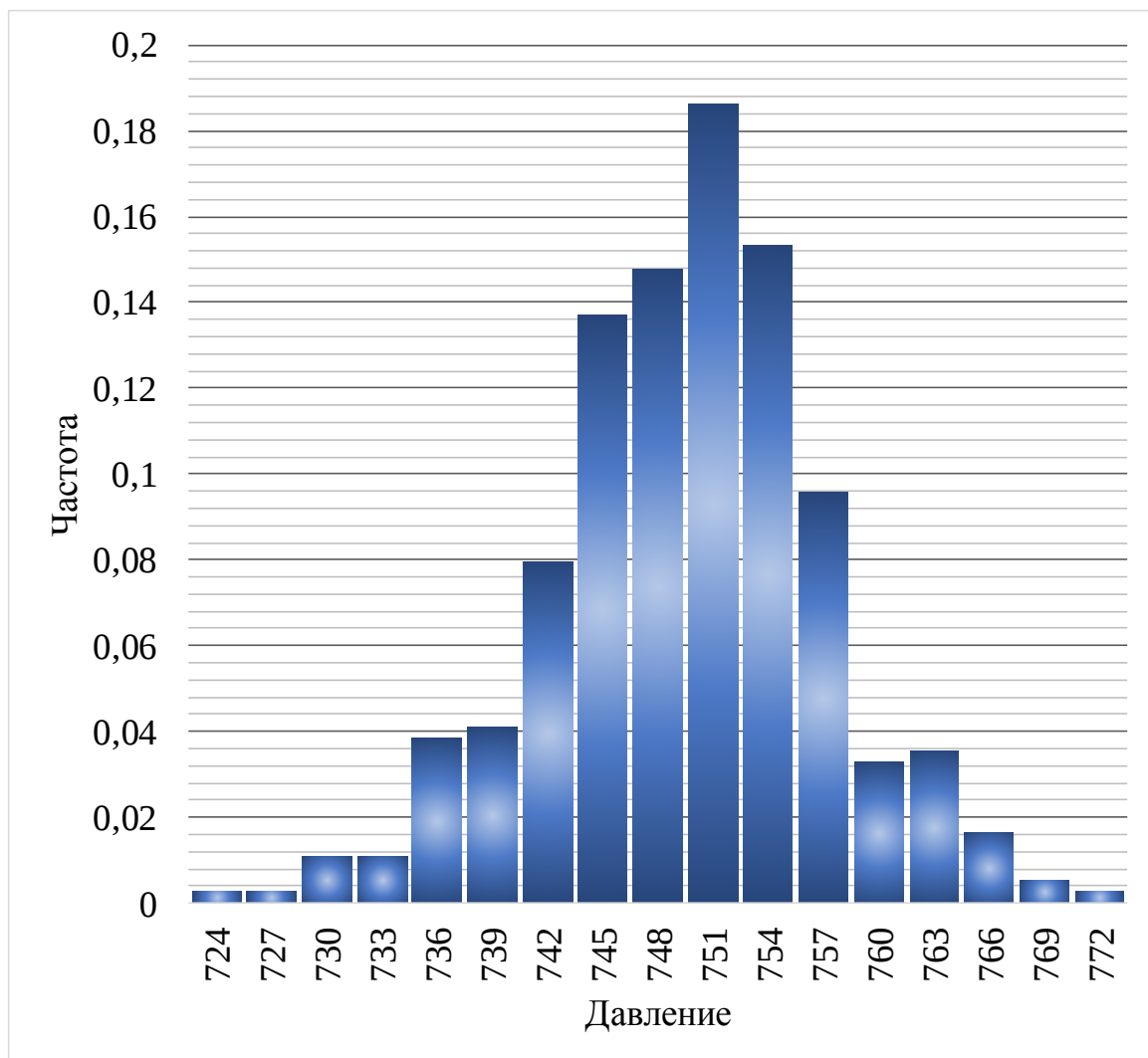
- Какова доля случаев, когда 3 или 6 очков выпадало с первого или со второго раза?
- Пользуясь диаграммой, оцените примерно, сколько будет случаев из 1000, когда 3 или 6 очков выпадет при четвертом броске (не раньше).
- Какие из известных вам величин имеют внешне похожий закон распределения?

16. Метеостанция ежедневно в 12:00 измеряет атмосферное давление в Москве. На рисунке представлена гистограмма, построенная по результатам измерений в течение года (с марта 2018 г. по февраль 2019 г.). По горизонтали отмечено атмосферное давление в миллиметрах ртутного столба (интервал группировки 3 мм.рт.ст), а по вертикали — частота этого события.

- Какие из известных вам величин имеют внешне похожий закон распределения?
- В каком интервале находится среднее атмосферное давление в Москве за этот год?

в) Определите по гистограмме, сколько приблизительно дней за год исследования (в году 365 дней) метеостанция фиксировала давление от 739 до 742 мм.рт.ст.

г) Известно, что нормальное давление на уровне моря составляет 760 мм.рт.ст. В среднем давление падает на 1 мм.рт.ст. при подъеме на каждые 12 метров. Зная среднее давление за год в Москве, оцените, на какой высоте над уровнем моря находится метеостанция, где проводились наблюдения.



Ответы и решения некоторых задач

1.

Ср. арифм.	Медиана	Наим.	Наиб.	Середина	Размах	Ср. геом.	Ср. гарм.
1,34	1,3	1,1	1,7	1,4	0,6	1,325	1,31

2. а) наибольшее значение; б) среднее гармоническое; в) наименьшее значение; г) среднее геометрическое.

3. а) $30 + \frac{2}{20} = 30,1$ тыс. руб. б) $30 + 2 = 32$ тыс. руб. в) не изменится

г) $\frac{30 \cdot 20 + 51}{21} = 31$ тыс. руб.

4. Среднее арифметическое равно 646,25, медиана равна 795.



На числовой прямой отметим значения, их среднее арифметическое (красная точка) и медиану (черная). Видно, что медиана лучше отражает среднюю численность учащихся школы. Численность учащихся в школах №4 и №7 сильно отличается от остальных (это выбросы), это сильно удалило среднее от медианы.

5. Среднее арифметическое группированных данных равно $\frac{7661}{200} = 38,305$.

Можно сказать, что средний размер проданной обуви – 38-й.

6. а) Без расчетов по диаграммам видно, что размах в феврале больше, чем в марте.

б) Судя по февральской диаграмме, размах цены одной акции за февраль равен 11 рублям. Поэтому прибыль бизнесмена составит 11 тыс.руб.

7. а) Средняя температура за год $-5,3^{\circ}\text{C}$.

б)

Месяц	Средняя температура, $^{\circ}\text{C}$	Отклонения от среднего, $^{\circ}\text{C}$	Квадрат отклонений, кв. $^{\circ}\text{C}$
Январь	$-19,4$	$-14,1$	198,81
Февраль	$-19,7$	$-14,4$	207,36
Март	$-13,8$	$-8,5$	72,25
Апрель	-10	$-4,7$	22,09
Май	$-1,9$	3,4	11,56
Июнь	7,6	12,9	166,41
Июль	13,1	18,4	338,56
Август	9,7	15	225
Сентябрь	4,4	9,7	94,09
Октябрь	$-3,2$	2,1	4,41
Ноябрь	-13	$-7,7$	59,29
Декабрь	$-17,4$	$-12,1$	146,41

в) Сумма отклонений от среднего всегда равна нулю.

г) Дисперсия равна примерно 128,85 кв. гр. Цельсия.

8. а) 541,4; б) Самара; в) медиана;

г) Москва — 4,81 тыс. чел./кв.км, Казань — 2,39 тыс. чел./кв.км;

д) Москва, Санкт-Петербург, Новосибирск, Екатеринбург, Нижний Новгород, Ростов-на-Дону, Красноярск.

9. 8,7%

10. Заполненная таблица:

Сотрудник	1	2	3	4	5
Рабочие часы	156	145	126	128	144
Единицы продукции	52	58	45	40	60
Трудоемкость	3	2,5	2,8	3,2	2,4

Средняя трудоемкость примерно равна 2,7.

11. а) Средняя ошибка равна $-2,6$; размах равен 4,7; дисперсия ошибки равна 2,8. б) Часы получают сертификат точности, но нуждаются в регулировке.

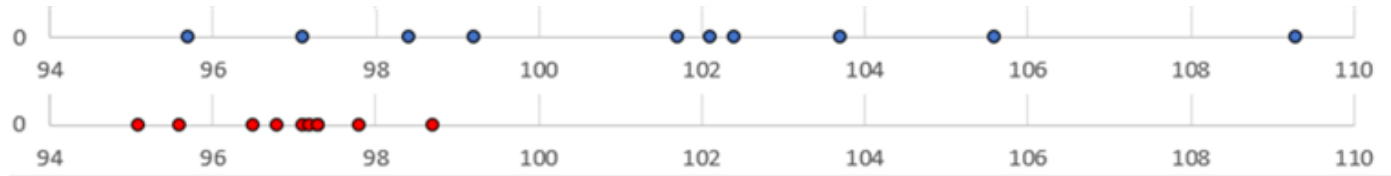
12. Среднее арифметическое поменяло знак (стало -4), а дисперсия не изменилась (осталась равна 18).

13. а) Среднее арифметическое равно 110, дисперсия равна 1000.

б) Значение 50 ненадежно. в) 125; г) да

14. а) Точность до десятых долей грамма. б) Найдем средние значения порции для каждого автомата. У первого автомата среднее равно 101,52 г (отличается

от номинала на 1,52 г), а у второго 96,64 г (отличается от номинала на 3,36 г). Казалось бы, с первым автоматом все в порядке — он дает небольшую ошибку в среднем. Но, чтобы разобраться в ситуации, представим измерения наглядно, отметив для каждого автомата полученные значения точками на числовой прямой:



Видно, что порции шоколада, выдаваемые первым автоматом, сильно разбросаны по массе. Иными словами, случайная величина «масса порции» у первого автомата имеет большое рассеивание. Такая случайная изменчивость скорее всего вызвана износом деталей, так что первый автомат определенно требует ремонта. У второго автомата разброс данных небольшой (это говорит о хорошем состоянии механизма в целом), такая систематическая ошибка исправляется небольшой регулировкой.

15. а) примерно 50%; б) примерно 110 случаев; в) продолжительность телефонного разговора или время, затраченное на устранение неполадок в работе сервера (примеры с уроков).

16. а) Например, рост человека или число выпавших орлов (примеры из уроков). б) от 748 до 751 мм.рт.ст. в) $0,08 \cdot 365 \approx 29$ дней г) от 108 до 144 м над у.м.