



«Описательная статистика» - 10 класс

ЧТО ИЗУЧАЕТ СТАТИСТИКА?



- ▣ **Статистика это наука, которая занимается получением, обработкой и анализом количественных данных о разнообразных массовых явлениях.**

Описательная

статистик

□ Наибольшее и наименьшее значение.

□ Размах , мода

□ Отклонения

□ Дисперсия

□ Обозначения и формулы

□ Свойства среднего

арифметического и дисперсии

Наибольшее и наименьшее

Петя и Вася поспорили, кто лучше прыгает в длину с места. Чтобы избежать случайности, они решили, что будут прыгать по очереди 5 раз. Результаты своих прыжков в сантиметрах они записали в таблицу.



Результаты прыжков в длину с места. см

Номер прыжка	Петя	Вася
1	190	185
2	205	200
3	195	215
4	210	190
5	200	190

Определение

Разность между наибольшим и наименьшим числом называется

размахом набора чисел

Таблица 6. Производство пшеницы в России в 1995-2001 гг.

Год	1995	1996	1997	1998	1999	2000	2001
Млн.тонн	30,1	34,9	44,3	27,0	31,0	34,5	47,0

Размах показывает, насколько велико рассеивание значений в числовом наборе.

При изучении учебной нагрузки учащихся выделили группу из 12 семиклассников. Их попросили отметить в определенный день время (в минутах), затраченное на выполнение домашнего задания по алгебре. Получили такие данные:

23 18 25 20 25 25 32 37 34 26 34 25

$$\frac{23 + 18 + 25 + 20 + 25 + 25 + 32 + 37 + 34 + 26 + 34 + 25}{12} = \frac{324}{12} = 27$$

27 – среднее значение

Наибольшее значение – 37; наименьшее значение – 18;

Размах ряда равен $37 - 18 = 19$

При анализе сведений о времени, затраченном семиклассниками на выполнение домашнего задания по алгебре, нас могут интересовать не только среднее арифметическое и размах полученного ряда данных, но и другие показатели. Интересно, например, знать, какой расход времени является типичным для выделенной группы учащихся, то есть какое число встречается в ряду данных чаще всего.

23, 18, 25, 20, 25, 25, 32, 37, 34, 26, 34, 25.

Нетрудно заметить, что таким числом является число 25. Говорят, что число 25 – *мода* рассматриваемого ряда.

МОДА



- **Модой** ряда чисел называется число, чаще других встречающееся в данном ряду.
- Ряд чисел может иметь более одной моды или не иметь моды совсем.

Например, в ряду чисел

47, 46, 50, 52, 47, 52, 49, 45, 43, 53

две моды – это числа 47 и 52,

**а в ряду чисел 69, 68, 66, 80, 67, 65, 71, 74, 63, 73,
72**

МОДА И ЦЕНА

Моду ряда данных обычно находят тогда, когда хотят выявить некоторый типичный показатель.

Например, если изучаются данные о размерах мужских сорочек, проданных в определенный день в универмаге, то удобно воспользоваться таким показателем, как мода, который характеризует размер, пользующийся наибольшим спросом.

Находить в этом случае среднее арифметическое не имеет смысла. Мода является наиболее приемлемым показателем при выявлении,

например, цены на товар данного вида, наиболее

МОДА

- **Мода** - это то, что непременно должны учитывать производители упаковок и фасовщики.
- Если бы большинство людей покупало за один раз стакан молока, молочные пакеты не были бы



Рассмотрим еще пример. Пусть, проведя учет деталей, изготовленных за смену рабочими одной бригады, получили такой ряд данных:

36, 35, 35, 36, 37, 37, 36, 37, 38, 36, 36, 36, 39, 39, 37, 39,
~~38 38 36 30 36~~

Найдем для него среднее арифметическое, размах и моду. Для этого удобно предварительно составить из полученных данных *упорядоченный ряд чисел*, т. е. такой ряд, в котором каждое последующее число не меньше (или не больше) предыдущего. Получим:

35, 35, 36, 36, 36, 36, 36, 36, 36, 36, 37, 37, 37, 37, 38, 38, 38, 39, 39, 39,
 39.

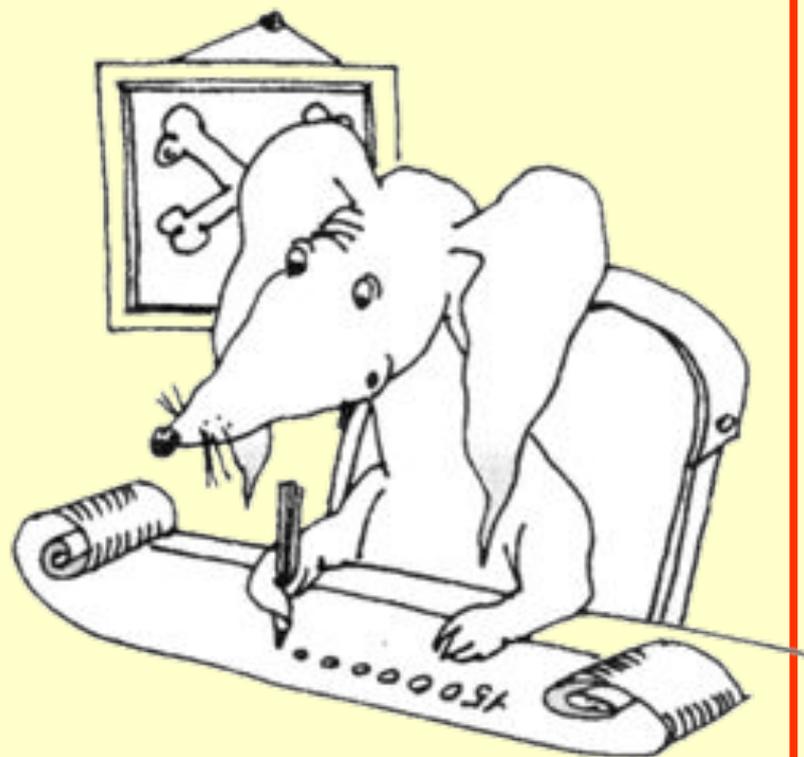
$$\frac{35 \cdot 2 + 36 \cdot 8 + 37 \cdot 4 + 38 \cdot 3 + 39 \cdot 4}{21} = \frac{776}{21} = 37$$

Вычислим среднее арифметическое

Размах ряда равен $38 - 35 = 4$. Мода данного ряда

равна 36, так как число 36 чаще всего встречается в этом ряду.

Итак, средняя выработка рабочих за смену составляет примерно 37 деталей; различие в выработке рабочих не превосходит 4 деталей; типичной является выработка, равная 36 деталям.



МЕДИАНА

- Если расставить выборку по возрастанию (или убыванию) той величины, которой мы интересуемся, то **медиана** - это то, что будет **ровно посередине "строя"**.
 - Например, если мы расположим по порядку длительности интервалы времени: **секунда, минута, час, сутки и неделя** – то медианой будет час.
-

ЧТО ПОКАЗЫВАЕТ МЕДИАНА?



19 бедняков и один миллиардер, каждый кладет на стол деньги из своего кармана. По **5 долларов** кладет каждый бедняк, а миллиардер — **\$1 млрд**. В сумме получается **$\$1\,000\,000\,095$** . Если мы **разделим** деньги равными долями на **20 человек**, то получим **$\$50\,000\,004,75$** . Это будет **среднее арифметическое**



- **Медиана** в этом случае будет равна $\$5$.
- В общем случае можно сказать, что медиана это то, сколько принес с собой *средний* человек.
- Наоборот, среднее арифметическое же совершенно неподходящая характеристика в нашем случае, поскольку выходит, что каждый, будь то бедняк или миллиардер,

Отклонени

Попробуем узнать, как числа некоторого набора расположены по отношению к своему среднему значению.

Зная только размах, разность между наибольшим и наименьшим значением, мы не можем судить о том, как расположены числа в имеющемся наборе.

Для примера возьмем набор 1, 6, 7, 9, 12. Вычислим среднее арифметическое:
 $(1+6+7+9+12):5=7.$

Найдем отклонение каждого числа от среднего:
 $1-7=-6, 6-7=-1, 7-7=0, 9-7=2, 12-7=5.$

Отклонения

Получился новый набор -6, -1, 0, 2, 5, который состоит из отклонений.

Если число меньше среднего, то его отклонение отрицательно, если число больше среднего, то его отклонение положительно. В одном случае – для числа 7, которое совпало со средним арифметическим, - отклонение равно нулю.

По набору отклонений можно судить о том, насколько разнообразны числа в наборе.

Если отклонения малы, то числа в наборе расположены близко к среднему арифметическому.

А если среди отклонений есть большие по модулю, то числа в наборе сильно разбросаны.

Отклонения

Для любого набора, если только не все числа в нем равны, часть отклонений будет положительна, а часть – отрицательна. При этом сумма всех отклонений равна 0.

Убедимся в этом на нашем примере:

$$-6+(-1)+0+2+5=0.$$

В этом состоит основное свойство отклонений: **сумма отклонений чисел от среднего арифметического этих чисел**

Дисперсия

Наиболее полной характеристикой разброса набора чисел является набор их отклонений от среднего арифметического. Но когда набор чисел велик, рассматривать набор отклонений практически неудобно. Нужно описать разнообразие чисел в наборе одной характеристикой, одним числом.

Размах – слишком грубая мера разброса чисел в наборе, поскольку учитывает только два из них – наименьшее и наибольшее. Можно попробовать взять «среднее отклонение». Но сумма отклонений всегда равна нулю, поэтому среднее

Дисперсия

Чтобы судить о разбросе, принято складывать не сами отклонения, а их квадраты. Квадраты отклонений неотрицательны, поэтому сумма квадратов отклонений зависит только от абсолютных величин отклонений, а не от их знаков. Чем больше отклонения чисел от среднего арифметического, тем больше будет сумма квадратов отклонений. Для того чтобы мера разброса чисел не зависела от их количества в наборе, в качестве такой меры берут среднее арифметическое квадратов отклонений. Эту величину называют

Дисперсия



Определе
ние.

Среднее арифметическое квадратов отклонений от среднего значения называется в статистике **дисперсией** набора чисел.

Дисперсия

Пример 1.

Обратимся к таблице
производства пшеницы
в России в 1995-2001 гг.
Среднее значение
равно 35,5 млн. тонн.
Найдем отклонение
каждого года от
среднего и возведем
его в квадрат.

Год	производство	Отклонение от среднего	Квадрат отклонения
1995	30,1		
1996	34,9		
1997	44,3		
1998	27,0		
1999	31,0		
2000	34,5		
2001	47,0		

Дисперсия

Пример 1.

Найдем квадрат отклонения от среднего

Год	производство	Отклонение от среднего	Квадрат отклонения
1995	30,1	-5,4	
1996	34,9	-0,6	
1997	44,3	8,8	
1998	27,0	-8,5	
1999	31,0	-4,5	
2000	34,5	-1,0	
2001	47,0	11,5	

Дисперс

ия

Пример 1.

**Найдем
квадрат
среднее
отклонения
и квадрато
в
отклонен**

ий

Год	производство	Отклонение от среднего	Квадрат отклонения
1995	30,1	-5,4	29,16
1996	34,9	-0,6	0,36
1997	44,3	8,8	77,44
1998	27,0	-8,5	72,25
1999	31,0	-4,5	20,25
2000	34,5	-1,0	1,00
2001	47,0	11,5	132,25

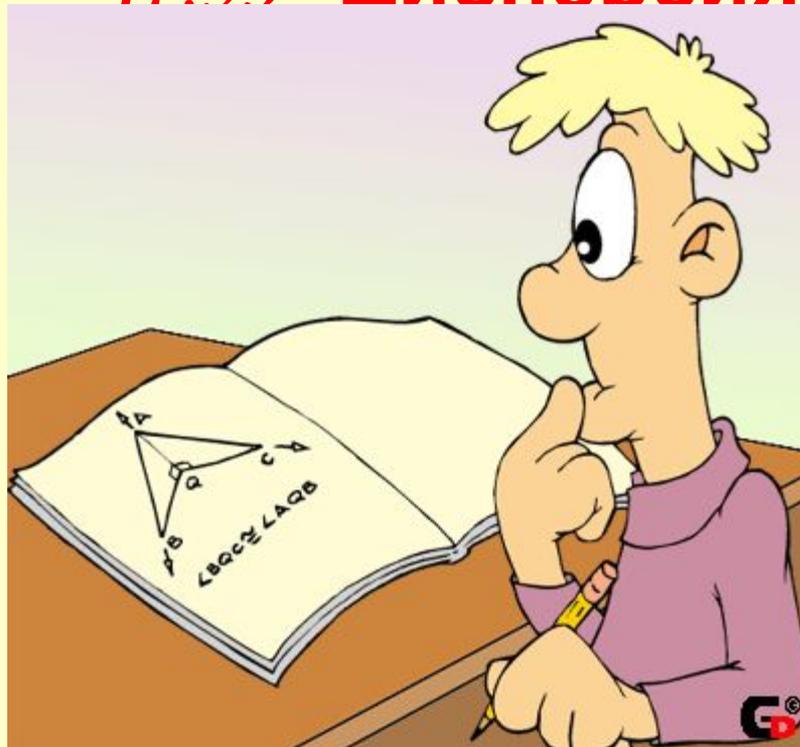
Дисперс

Пример 1.

ия

$$(29,16+0,36+77,44+72,25+20,25+1,00+132,25) \\ :7=47,53.$$

47,53 - дисперсия

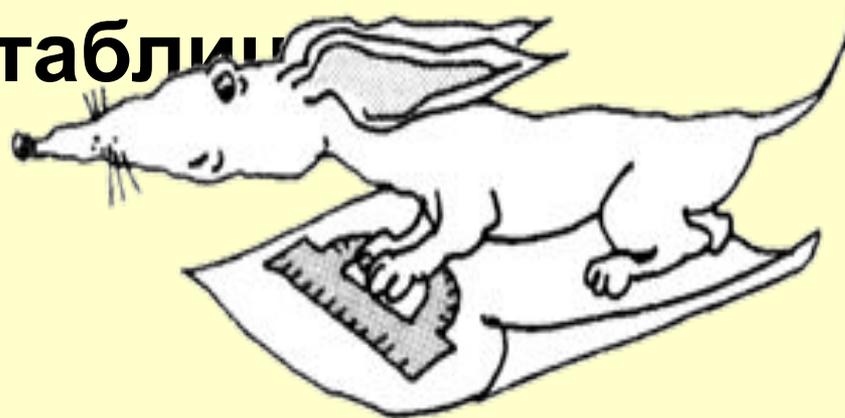


Дисперс

Пример 2.

ия

Покажем на простом примере, как дисперсия характеризует разброс отклонений. Возьмем два набора чисел $1, 2, 3$ и $0, 2, 4$. Среднее арифметическое значение обоих наборов равно 2 . Для обоих наборов вычислим отклонения и квадраты отклонений и все данные занесем в таблицу.



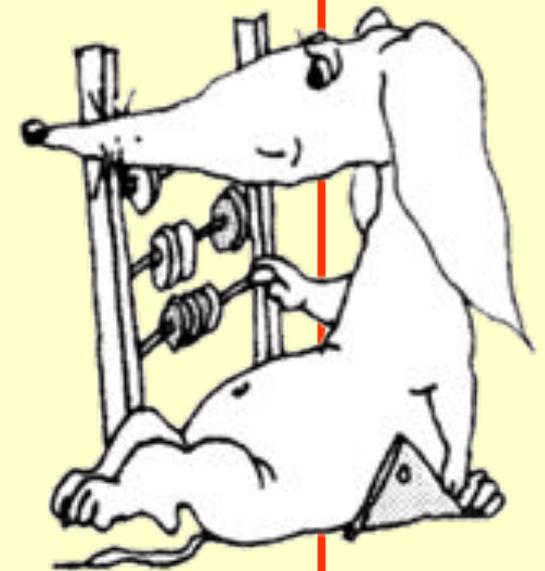
Дисперс

Пример 2.

ИЯ

1-й набор	Отклонение от среднего	Квадрат отклонения
1		
2		
3		

1-й набор	Отклонение от среднего	Квадрат отклонения
0		
2		
4		



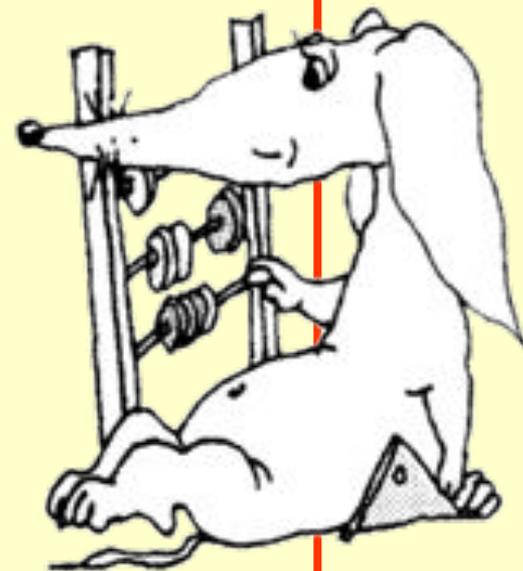
Дисперс

Пример 2.

ИЯ

1-й набор	Отклонение от среднего	Квадрат отклонения
1	-1	
2	0	
3	1	

1-й набор	Отклонение от среднего	Квадрат отклонения
0	-2	
2	0	
4	2	



Дисперс

Пример 2.

ИЯ

1-й набор	Отклонение от среднего	Квадрат отклонения
1	-1	1
2	0	0
3	1	1

1-й набор	Отклонение от среднего	Квадрат отклонения
0	-2	4
2	0	0
4	2	4



Дисперс

Пример 2.

ИЯ

1-й набор	Отклонение от среднего	Квадрат отклонения
1	-1	1
2	0	0
3	1	1

Дисперсия первого набора:

$$(1 + 0 + 1) : 3 = \frac{2}{3}$$

1-й набор	Отклонение от среднего	Квадрат отклонения
0	-2	4
2	0	0
4	2	4

Дисперсия второго набора:

$$(4 + 0 + 4) : 3 = \frac{8}{3} = 2 \frac{2}{3}$$

Дисперс

Пример 2.

ИЯ

1-й набор	Отклонение от среднего	Квадрат отклонения
1	-1	1
2	0	0
3	1	1

1-й набор	Отклонение от среднего	Квадрат отклонения
0	-2	4
2	0	0
4	2	4

Дисперсия первого набора:

$$(1 + 0 + 1) : 3 = \frac{2}{3}$$

Дисперсия второго набора:

$$(4 + 0 + 4) : 3 = \frac{8}{3} = 2 \frac{2}{3}$$

Числа в первом наборе расположены более кучно – ближе друг к другу и к своему среднему, - чем числа во втором наборе.

Поэтому дисперсия первого набора меньше

Дисперс

Пример 3.

Континентальный климат отличается от умеренного более резкими изменениями температуры в течение года. В районах с континентальным климатом жаркое лето и очень холодная зима. С помощью дисперсии различия между двумя видами климата можно выразить количественно. Сравним для примера изменение температур в течение года в Москве и Киеве, где климат умеренный, с изменением температур в Новосибирске и Хабаровске, где климат континентальный. В таблице 10 приведены средние месячные температуры за 80 лет в Москве, Киеве, Новосибирске и Хабаровске.

Дисперс

Пример 3.

Месяцы	Москва	Киев	Новосибирск	Хабаровск
1	-9,3	-5,9	-19,0	-22,3
2	-8,6	-5,2	-17,2	-17,2
3	-3,4	-0,4	-10,7	-8,5
4	5,1	7,5	-0,1	3,1
5	12,4	14,7	10,0	11,1
6	16,7	17,8	16,3	17,4
7	18,4	19,8	18,7	21,1
8	16,6	18,7	16,0	20,0
9	10,9	13,9	9,9	13,9
10	4,4	7,5	1,5	4,7
11	-2,0	1,2	-9,7	-8,1
12	-6,8	-3,5	-16,9	-18,5
Среднее за год	4,5	6,0	-0,1	-1,4
Дисперсия	98,9	86,5	185,2	228,8

Обозначения и

Среднее арифметическое чисел x_1, x_2, x_3, x_4, x_5

принято \bar{x}

обозначать через

Например, среднее арифметическое пяти

чисел запишется так:
$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

Обозначения и

Отклонения от среднего значения теперь

запишутся так: $x_1 - \bar{x}$, $x_2 - \bar{x}$, $x_3 - \bar{x}$, $x_4 - \bar{x}$, $x_5 - \bar{x}$.

Разберем на примере набора x_1, x_2, x_3, x_4 , как записывается в символьном виде дисперсия. Дисперсия равна среднему арифметическому квадратов отклонений этих чисел от среднего значения. Обозначают дисперсию обычно

через s^2 . Получается:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2}{4}.$$

Оформление практической работы по статистике.

Не забудьте посчитать все это и для второго

Для ряда **5, 4, 5, 6, 5, 6, 4** найдите:

- а) Моду
- б) Медиану
- в) Среднее арифметическое
- г) Отклонения
- д) Дисперсию
- е) Среднее квадратическое отклонение

1. Упорядочить: 4, 4, 5, 5, 5, 6, 6

2. $M = 5$, $M_e = 5$

3.
$$\bar{x} = \frac{4 \cdot 2 + 5 \cdot 3 + 6 \cdot 2}{7} = 5$$

Варианта x_i	Кратность ь	Отклонение $x_i - \bar{x}$	Квадрат отклонения $(x_i - \bar{x})^2$
4	2	-1	1
5	3	0	0
6	2	1	1
Σ	7	$-1 \cdot 2 + 0 \cdot 3 + 1 \cdot 2 = 0$	

$$D = \frac{1 \cdot 2 + 0 \cdot 3 + 1 \cdot 2}{7} = \frac{4}{7} \approx 0,57$$

$$\sigma = \sqrt{D} = \sqrt{0,57} \approx 0,75$$