Дисперсия числового набора.

Цели:

- познакомиться с понятиями **отклонение** и **дисперсия** и их применением в реальных практических ситуациях;
- учиться **вычислять характеристики числовых наборов: находить отклонение** и **дисперсию.**

Повторение

Характеристики числового набора

- **Размах**
- **Медиана**
- **Среднее арифметическое**
- **Среднее геометрическое**
- **Среднее гармоническое**
- **Отклонение от среднего арифметического**
- **Дисперсия числового набора.**

Дисперсия числового набора.

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 арифметическое отклонений тоже равно нулю и его нельзя использовать как меру разброса.

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 ***дисперсией***.

Среднее числового набора

Среднее арифметическое квадратов отклонений от среднего значения называют **дисперсией числового набора**

Дисперсия — мера рассеивания данных, равная среднему квадрату отклонения от среднего арифметического.

Дисперсия показывает величину отклонений наблюдений от среднего. Чем она больше, тем более пологой выглядит колоколообразная кривая. При маленькой дисперсии колоколообразная кривая, наоборот, становится островерхой. Все наблюдения при этом лежат близко к своему среднему значению

Алгоритм вычисления дисперсии

Для нахождения дисперсии D данных $x_1, x_2, \ldots, x_n$ измерения следует вычислить:

1) среднее значение $M = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$;

2) отклонения данных от M, т. е. $x_1 - M$, $x_2 - M$, ..., $x_{n-1} - M$, $x_n - M$;

3) квадраты $(x_i - M)^2$ отклонений $(x_i - M)$ $(i = 1, 2, 3, \ldots, n)$, найденных на предыдущем шаге;

4) среднее значение всех квадратов отклонений — это и есть дисперсия D:

$$D = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_{n-1} - M)^2 + (x_n - M)^2}{n};$$

$\sigma = \sqrt{D}$ — среднее квадратическое отклонение.

Пример 1. Рассмотрим таблицу производства пшеницы в России.
Найдем среднее производство пшеницы

1) ср/ар = 35,5 млн. тон в год

период 1995—2001 гг. составило 35,5 млн. тонн в год. Вычислим дисперсию. Составим таблицу, разместив данные по производству не в строке, а в столбце. Вычислим отклонения от среднего и их квадраты. Полученные числа занесем в два новых столбца.

Таблица 8. Производство пшеницы в России в 1995—2001 гг., млн. тонн

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

3. Найдем дисперсию.

Дисперсия = квадрат отклонений разделить на количество чисел

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

Пример 2. Покажем на простом примере, как дисперсия характеризует разброс наблюдений. 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 9.

Таблица 9

1-й набор	Отклонение от среднего	Квадрат отклонения	2-й набор	Отклонение от среднего	Квадрат отклонения
1	−1	1	0	−2	4
2	0	0	2	0	0
3	1	1	4	2	4

Дисперсия первого набора: $(1+0+1):3=\frac{2}{3}$.

Дисперсия второго набора: $(4+0+4):3=2\frac{2}{3}$.

Числа в первом наборе расположены более кучно – ближе друг к другу и к своему среднему, – чем числа во втором наборе. Поэтому дисперсия первого набора получилась меньше, чем второго.

Упражнения

1. Для данных чисел вычислите среднее значение. Составьте таблицу отклонений от среднего и квадратов отклонений от среднего и вычислите дисперсию:

а) −1, 0, 4; б) 2, 3, 7; в) −3, 1, 2, 4; г) 2, 6, 7, 5;
д) −2, −1, 1, 2, 5; е) −1, −3, −2, 3, 3.

Решение

1. Найдем среднее арифметическое.
2. Отклонение = число из набора – ср/арифметическое

а) ср/ар = (-1+0+4) / 3 =3:3 =1

Число из набора	Отклонение от среднего	Отклонение в квадрате
-1	-1 – 1 = -2	$(-2)^2 = 4$
0	0 – 1 = -1	$(-1)^2 = 1$
4	4 - 1 = 3	$3^2 = 9$

3. Найдем дисперсию.

Дисперсия = ср/ар отклонений в квадрате

(4+1+9) : 3 = 14:3 = =4,3

№1-в) дан ряд: -3, 1, 2, 4

Решение

1. Найдем среднее арифметическое.
2. Отклонение = число из набора – ср/арифметическое

в) ср/ар = (-3+1+2+4) / 4 =4:4 =1

Число из набора	Отклонение от среднего	Отклонение в квадрате
-3	-3 -1 = -4	$(-4)^2 = 16$
1	1 - 1 = 0	$0^2 = 0$
2	2 – 1 = 1	$1^2 = 1$
4	4 – 1 = 3	$3^2 = 9$

3. Найдем дисперсию.

Дисперсия = ср/ар отклонений в квадрате

(16+0++1+9) : 4 =
= 26:4 = 26/4 =6,5

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.

Таблица 10. Средние месячные температуры, °C

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				
7				
8				
9				
10				
11				
12				
Среднее за год	4,5	6,0	−0,1	−1,4
Дисперсия	98,9	86,5	185,2	228,8

Дисперсии этих четырех рядов чисел различны. Для Москвы и Киева это 98,9 и 86,5, для Новосибирска и Хабаровска это 185,2 и 228,8.

Вы видите, что дисперсии для рядов месячных температур в умеренном и континентальном климате значительно различаются.

Выводы и итоги урока.

Мы познакомились еще с одним показателем рассеивания набора – с дисперсией.

Отклонения и дисперсия числового набора

- ***Отклонение*** – отличие члена ряда от среднего арифметического.
- ***Дисперсия ряда чисел*** (D) – среднее арифметическое квадратов их отклонений от среднего арифметического этого ряда.
- Дисперсия является ***мерой разброса*** чисел в ряду.
- ***Среднее квадратичное отклонение числового ряда*** (σ) – квадратный корень из дисперсии этого ряда.

Пример: ряд данных 7, 5, 10, 6, 5, 15.

Среднее арифметическое равно $\overline{X} = \frac{7+5+10+6+5+15}{6} = 8.$

Отклонение каждого члена ряда от среднего арифметического:

$7 - 8 = -1; 5 - 8 = -3; 10 - 8 = 2; 6 - 8 = -2; 5 - 8 = -3; 15 - 8 = 7.$

Дисперсия равна $D = \frac{(-1)^2+(-3)^2+2^2+(-2)^2+(-3)^2+7^2}{6} = 12\frac{2}{3} \approx 13.$

Среднее квадратичное отклонение числового ряда $\sigma = \sqrt{13}.$

Выводы и итоги урока.

- **Стабильность** можно оценивать с помощью **отклонений элементов числового набора от среднего значения (отклонение – это разность между числом из данного набора и средним арифметическим этого набора)**
- **Когда набор чисел велик, рассматривать отклонения практически неудобно**, нужно описать разнообразие чисел в наборе одним числом.
- Сумма отклонений всегда равна 0.

Следовательно сумма отклонений не может нести информацию о разбросе.

- **Квадраты отклонений** (они всегда неотрицательны).
- **Чем меньше сумма квадратов отклонений, тем меньше разброс чисел относительно среднего значения, тем более стабилен набор.**
- При сравнении нескольких числовых наборов **с различным количеством чисел в наборе** в качестве меры сравнения берут **дисперсии** наборов.
- Стабильность каждого из числовых наборов можно оценить по величине **среднего арифметического квадратов отклонений от среднего значения – дисперсии.**

Вопросы

- Всегда ли средние характеристики числового ряда могут дать точную информацию о нём?

- Что такое **отклонение**?

- В каком случае для сравнения числовых наборов можно использовать **суммы квадратов отклонений**?

- В каком случае для сравнения числовых наборов предпочтительно вычислить их **дисперсии**?

Домашнее задание

Упражнения

1. Для данных чисел вычислите среднее значение. Составьте таблицу отклонений от среднего и квадратов отклонений от среднего и вычислите дисперсию:

а) $-1, 0, 4$; б) $2, 3, 7$; в) $-3, 1, 2, 4$; г) $2, 6, 7, 5$;
д) $-2, -1, 1, 2, 5$; е) $-1, -3, -2, 3, 3$.