

Дисперсия: как измеряют разброс данных

Поймём, как числа отклоняются от среднего.

Почему одного среднего недостаточно



Два набора могут иметь одинаковое среднее, но один окажется почти ровным, а другой — сильно разбросанным. Именно поэтому в статистике рядом со средним рассматривают и дисперсию: она показывает, насколько данные удалены от центра.

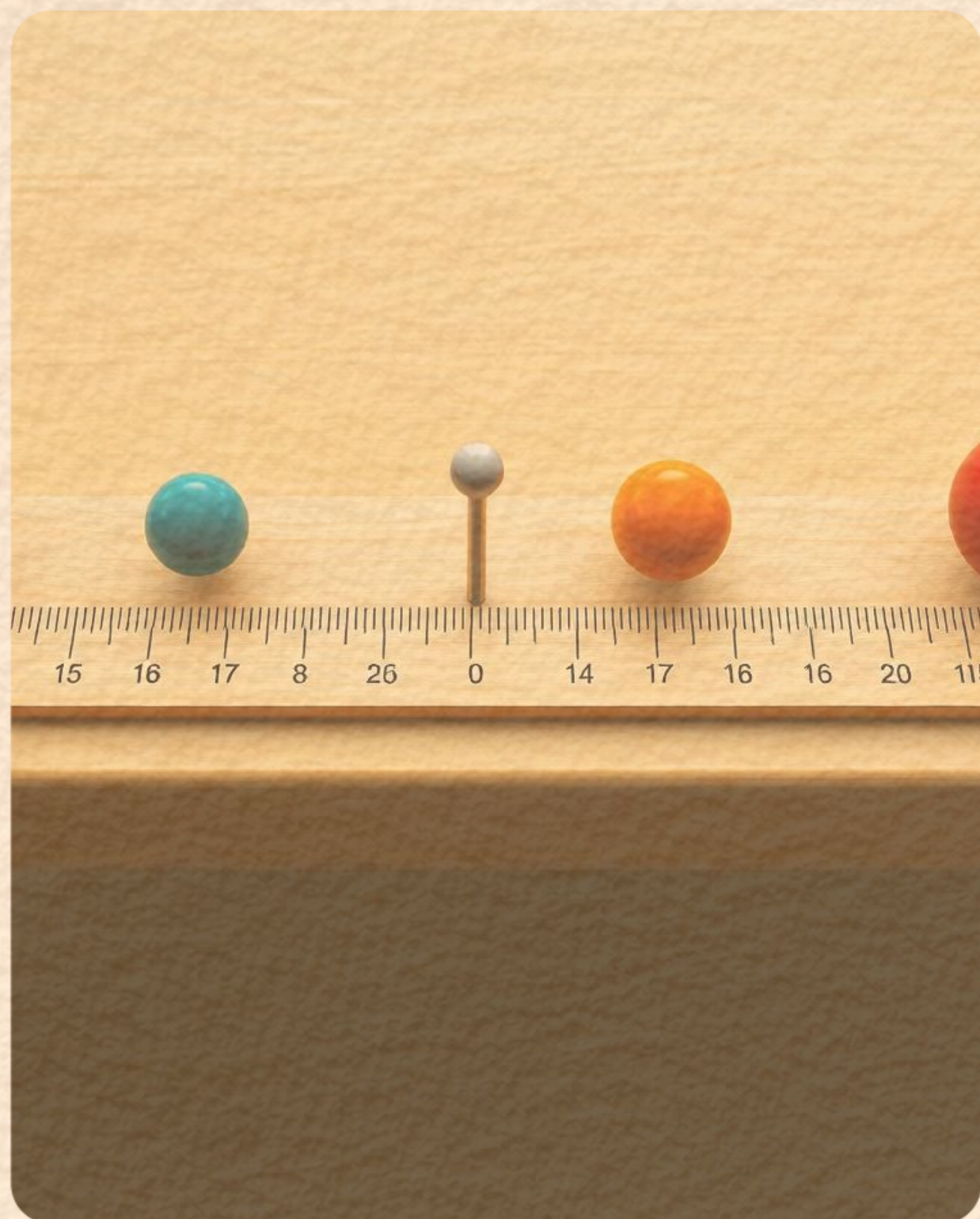
Отклонения каждого числа от этого центра показывают, как далеко оно расположено от среднего.

Учебный пример набора 2, 4, 6, 8

5

— столько равно среднее значение набора 2, 4, 6, 8.

Отклонение от среднего на простом примере



Найдём центр

Для набора 2, 4, 6, 8 среднее арифметическое равно 5. Это значение удобно считать центром, относительно которого измеряют положение каждого числа.

Посмотрим на прямую

На числовой прямой видно, что значения 4 и 6 ближе к 5, чем 2 и 8. Так отклонение можно понимать как расстояние до среднего.

Сравним расстояния

Число 2 находится от среднего на расстоянии 3, а числа 4 и 6 — на расстоянии 1. Значит, не все элементы удалены от центра одинаково.

Сделаем вывод

Если числа расположены по обе стороны от центра, их отклонения бывают положительными и отрицательными, но само расстояние до среднего остаётся понятным.

Почему отклонения возводят в квадрат

1 Проблема суммы

Если просто сложить отклонения, положительные и отрицательные значения могут взаимно уничтожиться. Тогда общая мера разброса окажется бессмысленной и не покажет различий между наборами.

2 Квадрат убирает знак

Возведение в квадрат делает все отклонения неотрицательными. После этого можно складывать их без риска, что они сократят друг друга и дадут нулевой результат.

3 Сильнее выделяются большие расстояния

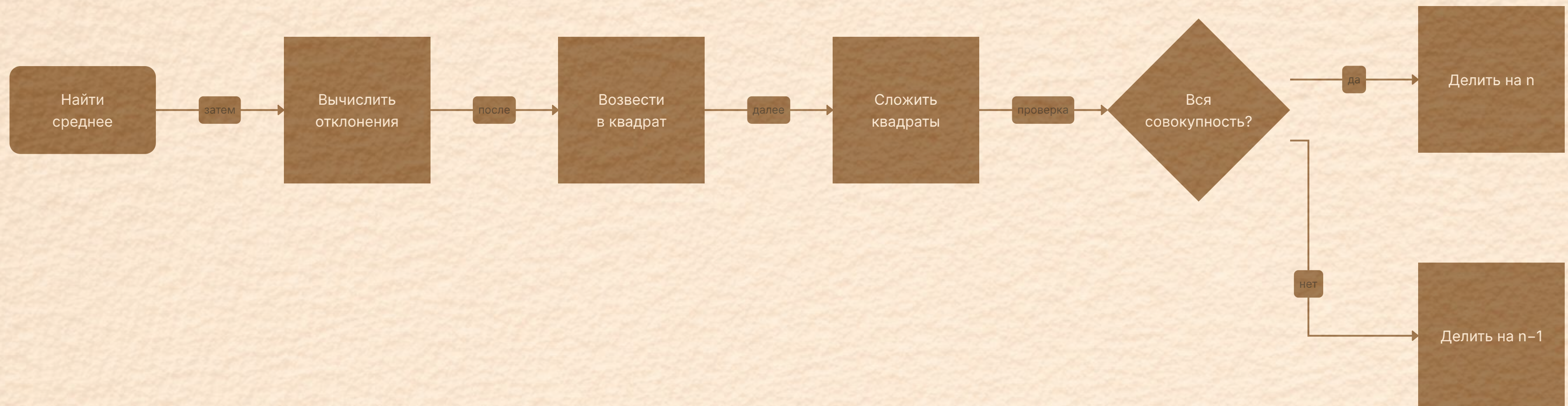
Большое отклонение после возведения в квадрат растёт быстрее, чем маленькое. Поэтому редкие далёкие значения заметнее влияют на дисперсию, чем близкие к среднему.

4 Почему это удобно

Так дисперсия показывает не только наличие разброса, но и его величину. Чем дальше числа от среднего, тем больше вклад каждого отклонения в итоговый результат.

Как вычисляют дисперсию

Последовательность действий для набора данных и выборки



Пошаговый расчет дисперсии для набора 2, 4, 6, 8

Таблица показывает, как из отклонений получается дисперсия для простого набора чисел.

Сумма квадратов равна 20, поэтому дисперсия генеральной совокупности составляет 5. Одинаковое среднее не гарантирует одинаковый разброс.

Значение	Отклонение	Квадрат
2	-3	9
4	-1	1
6	1	1
8	3	9

Учебный пример

Формула дисперсии для генеральной совокупности

1 Общий вид

Для всей совокупности дисперсию записывают как $D = (\sum(x - \bar{x})^2) / n$. Формула сразу показывает, что сначала находят отклонения, а затем усредняют их квадраты.

2 Что означает x

Буква x обозначает отдельное значение из набора данных. Каждое число сравнивают со средним, чтобы понять, насколько оно удалено от центра.

3 Роль среднего

Символ $\bar{x}$ — это среднее арифметическое. Именно от него измеряют отклонения, потому что оно задаёт точку, вокруг которой рассматривают распределение данных.

4 Зачем делят на n

В знаменателе стоит число всех наблюдений. Это делает итоговую величину средней величиной квадратов отклонений и помогает сравнивать разные наборы данных.

Выборка и деление на $n-1$



Наблюдаем часть данных

Если изучают не всю совокупность, а только её часть, такую часть называют выборкой. По ней пытаются понять свойства большого набора, не рассматривая все значения.



Поправка для оценки

При работе с выборкой часто делят на $n-1$. Такая поправка помогает получить более точную оценку разброса всей совокупности, а не только самой выборки.



Сначала определяем тип данных

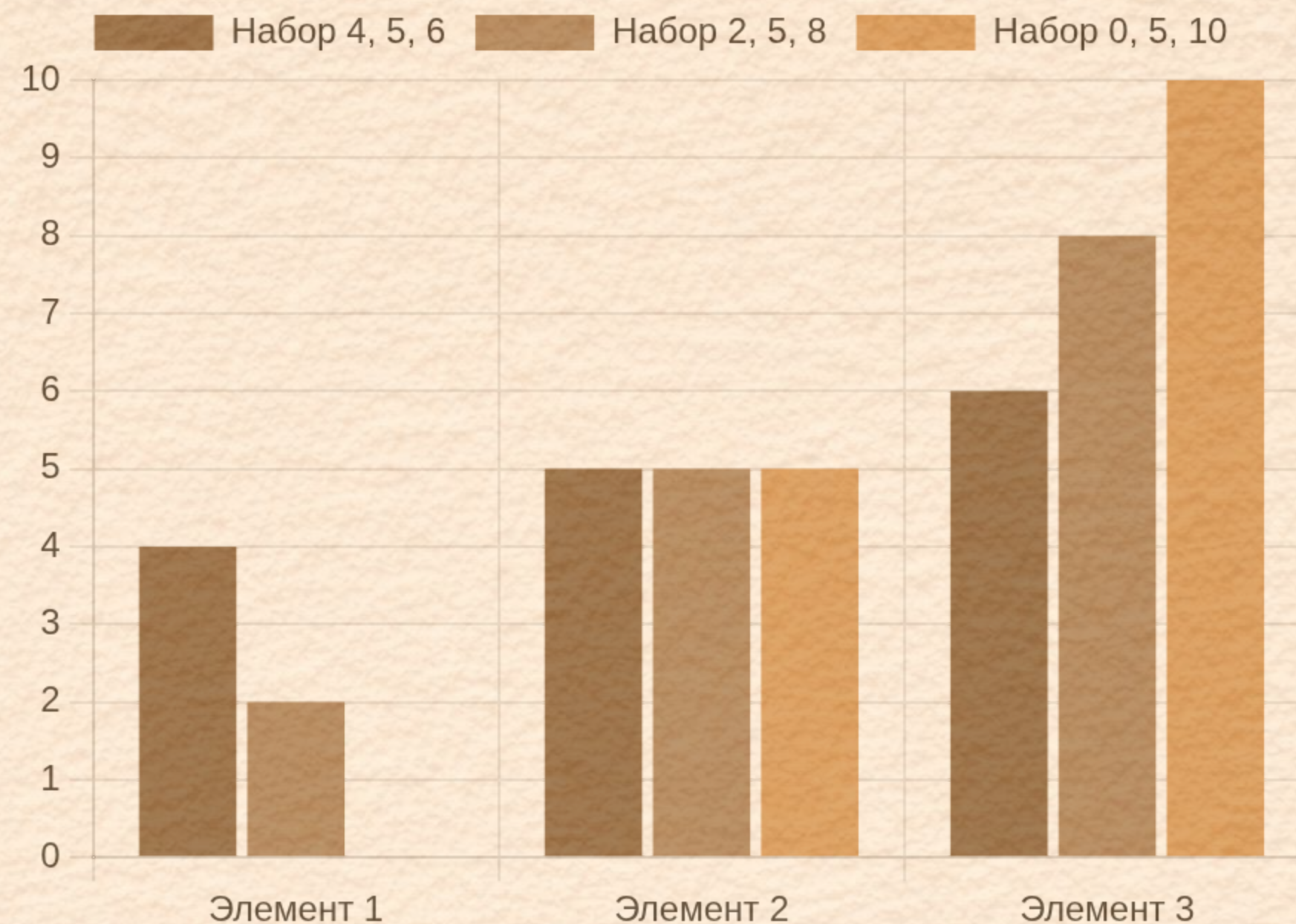
Перед вычислениями важно понять, что дано в задаче: все данные или только часть. От этого зависит формула и правильность результата.

Как меняется дисперсия при разном разбросе

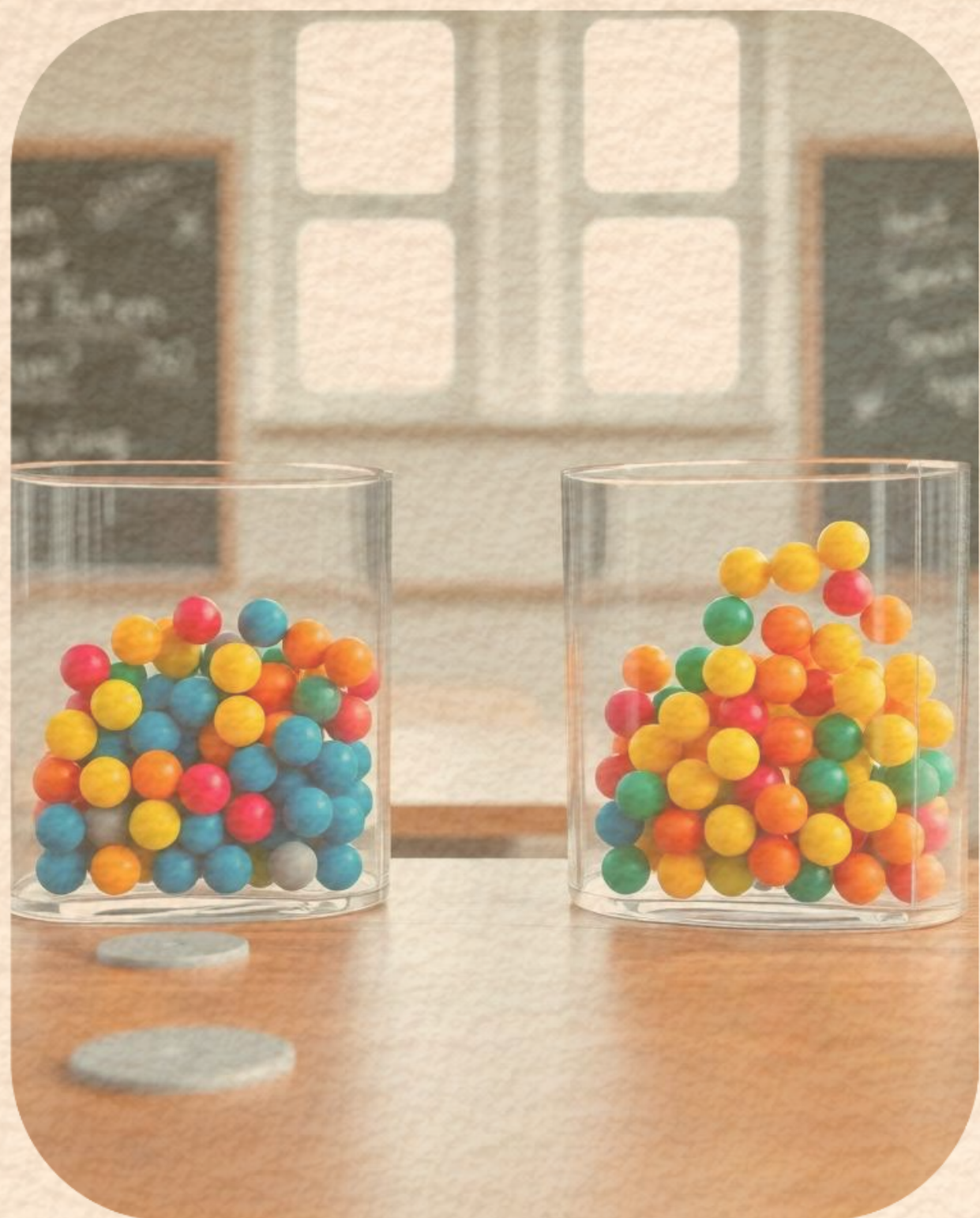
На графике видно, что все три набора имеют одинаковый центр, но различаются по удалённости чисел от него.

Чем дальше значения от среднего, тем больше разброс и тем выше дисперсия.

*Учебный пример; сравнение наборов 4, 5, 6;
2, 5, 8; 0, 5, 10*



Что означает маленькая и большая дисперсия



Значения рядом

Маленькая дисперсия означает, что почти все числа находятся недалеко от среднего. Такой набор выглядит «ровным»: значения меняются слабо и предсказуемо.

Сравнение результатов

В школьных примерах дисперсия помогает понять, где результаты стабильнее. Например, у двух учеников может быть одинаковое среднее, но у одного оценки будут почти одинаковыми, а у другого — очень разными.

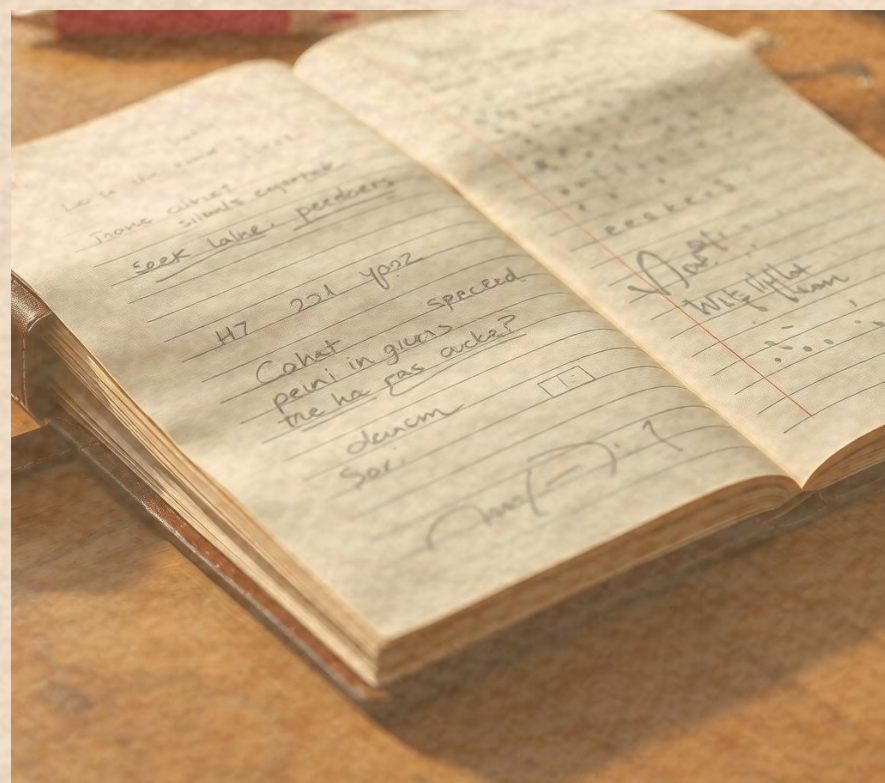
Значения разбросаны

Большая дисперсия показывает, что данные сильно расходятся от центра. В одном наборе могут быть числа около среднего, а в другом — заметно меньшие и большие значения.

Что важно заметить

Чем меньше дисперсия, тем ближе значения друг к другу. Чем больше дисперсия, тем сильнее выражены различия, и тем внимательнее нужно смотреть на отдельные числа.

Дисперсия в повседневных примерах



Оценки в дневнике

Если в дневнике почти все оценки одинаковые, разброс небольшой, и учиться получается ровно. Если рядом стоят и «пятёрки», и «тройки», и «двойки», дисперсия становится больше. По ней видно, насколько устойчивы учебные результаты.



Шаги за день

Фитнес-браслет может показывать одинаково высокую активность в течение недели, а может фиксировать резкие скачки: один день очень активный, другой почти без движения. Дисперсия помогает заметить, насколько привычка к движению стабильна.



Температура по дням

Когда температура меняется совсем мало, погода стоит ровная и спокойная. Если же один день тёплый, а следующий заметно холоднее, разброс значений увеличивается. Поэтому дисперсия удобна для сравнения погодных колебаний.



Время выполнения задания

На контрольной работе один ученик может выполнять похожие задания почти с одинаковой скоростью, а у другого время сильно меняется. Дисперсия показывает, насколько равномерно распределяется время, и помогает заметить нестабильность.

Сравнение двух классов по оценкам

Класс	Оценки	Среднее	Дисперсия
8А	3, 4, 4, 5	4	0,5
8Б	2, 4, 4, 6	4	2

Оба класса имеют одинаковое среднее значение, но различаются по разбросу оценок. Таблица показывает, что среднее не всегда достаточно для полного сравнения результатов.

Среднее у классов одинаковое, но дисперсия у 8Б больше. Значит, его результаты менее ровные и более неоднородные.

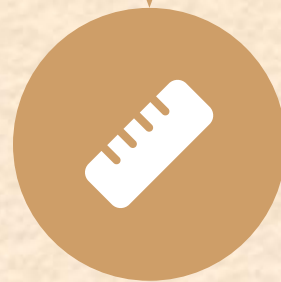
Условный учебный пример

Связь дисперсии и стандартного отклонения



Формула связи

Стандартное отклонение получают из дисперсии с помощью квадратного корня: $s = \sqrt{D}$. Это делает его удобным продолжением расчёта разброса.



Единицы измерения

Стандартное отклонение выражается в тех же единицах, что и исходные данные. Поэтому его легче сопоставлять с самими значениями, чем дисперсию.



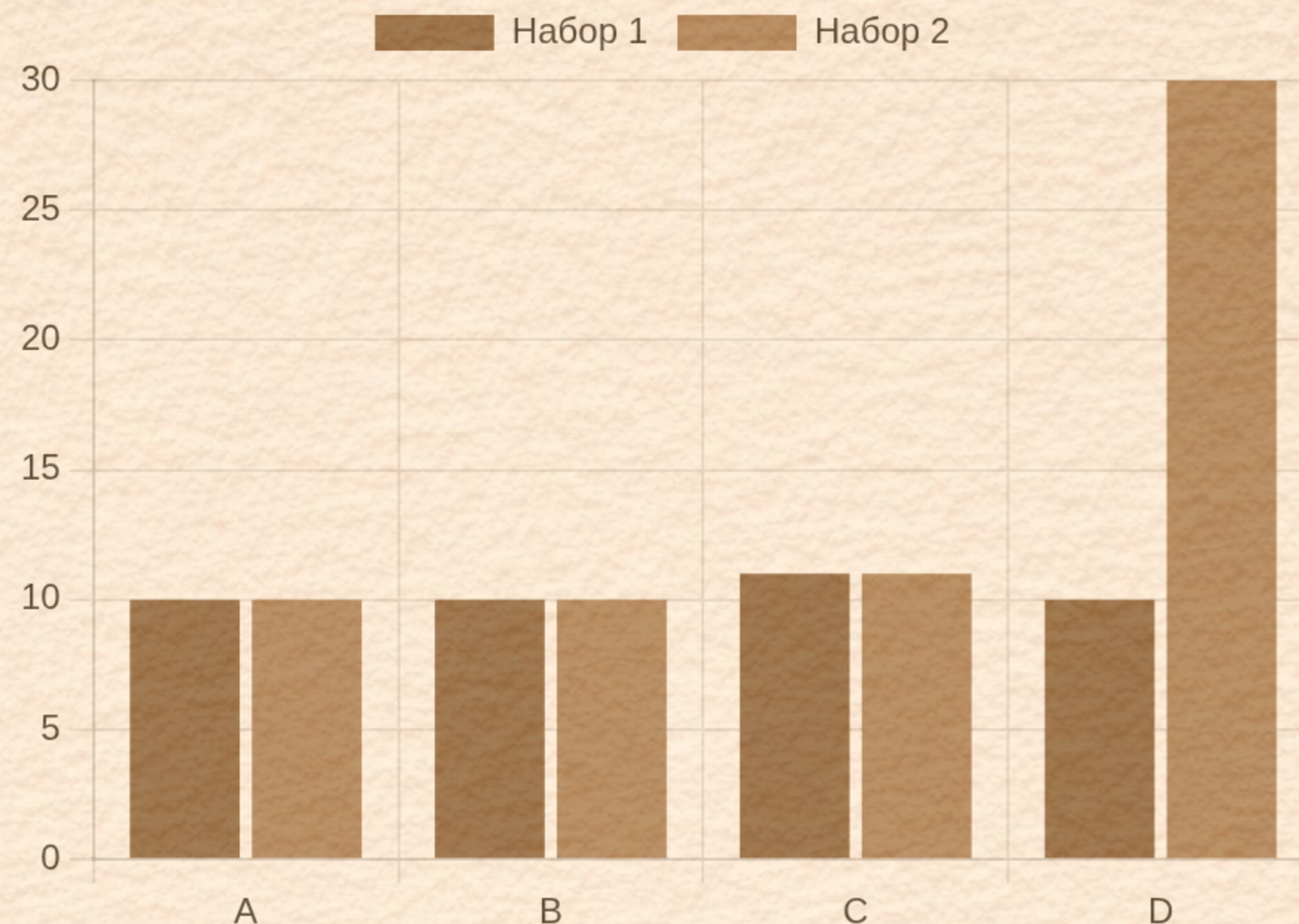
Два полезных показателя

Дисперсия удобна для вычислений, а стандартное отклонение — для наглядной интерпретации. Вместе они помогают точнее описывать разброс данных в статистике.

Как одно выбивающееся значение влияет на разброс

Обе серии начинаются почти одинаково, но одно лишнее значение резко меняет картину разброса. На графике это видно особенно отчётливо.

Выброс 30 значительно увеличивает разброс во втором наборе. Одно крайнее значение может сильно повлиять на дисперсию.



Учебный пример

Типичные ошибки при вычислении дисперсии

1 Не находят среднее

Иногда сразу начинают считать отклонения, не вычислив среднее значение. Без этого шага нельзя правильно определить, насколько каждое число удалено от центра набора.

2 Не возводят в квадрат

Если просто сложить отклонения, положительные и отрицательные значения взаимно уничтожатся. Из-за этого результат может получиться нулевым, даже когда данные сильно различаются.

3 Делят не на то число

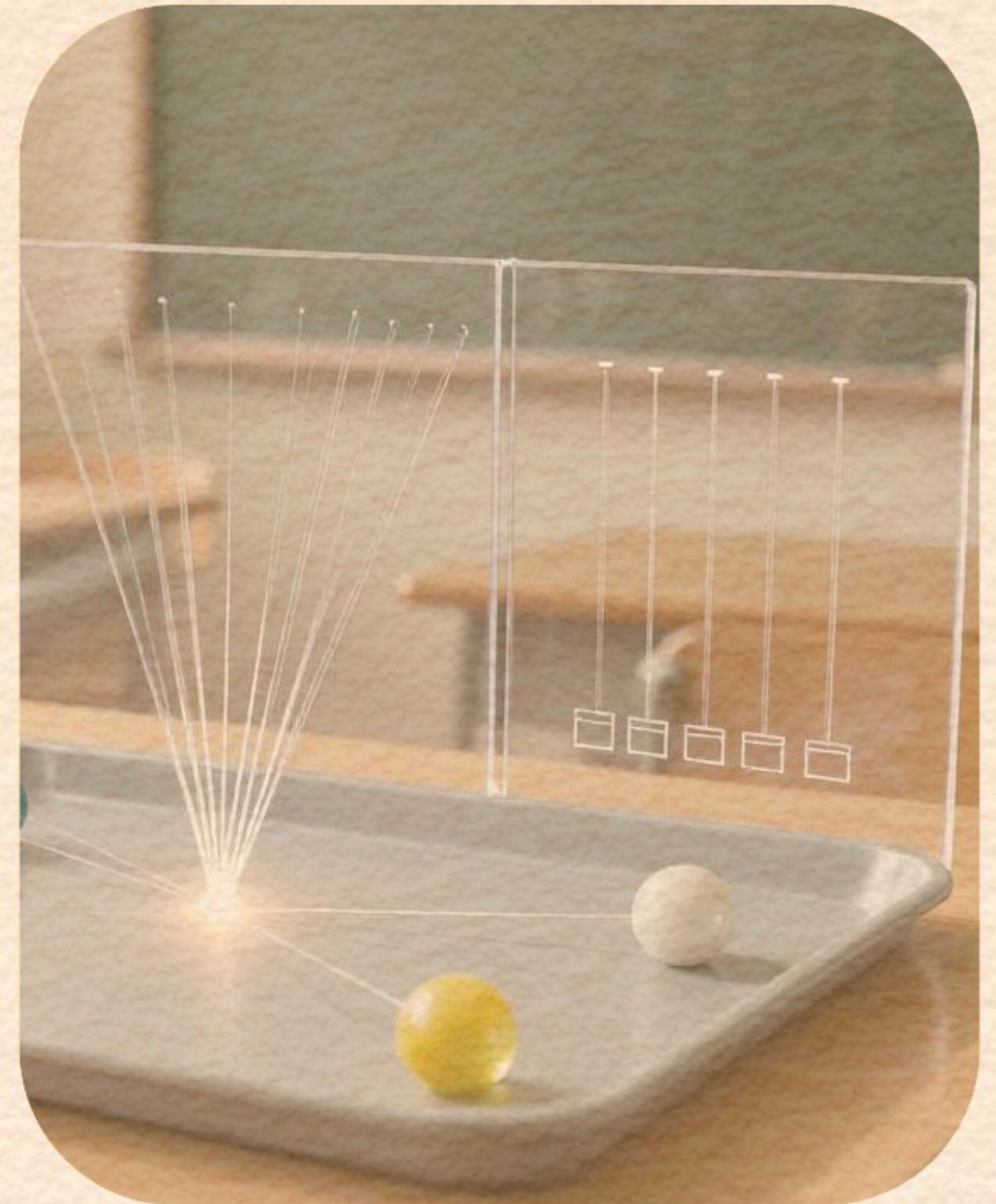
Для генеральной совокупности используют деление на n , а для выборки часто на $n-1$. Ошибка в знаменателе меняет итоговый результат и искажает вывод.

4 Путают два понятия

Дисперсия и стандартное отклонение связаны, но это не одно и то же. Дисперсия получается в квадрате единиц, а стандартное отклонение — в обычных единицах измерения.

5 Слишком рано округляют

Если округлять промежуточные значения слишком рано, можно потерять точность. Особенно это заметно в задачах с дробями, где каждый шаг влияет на конечный ответ.

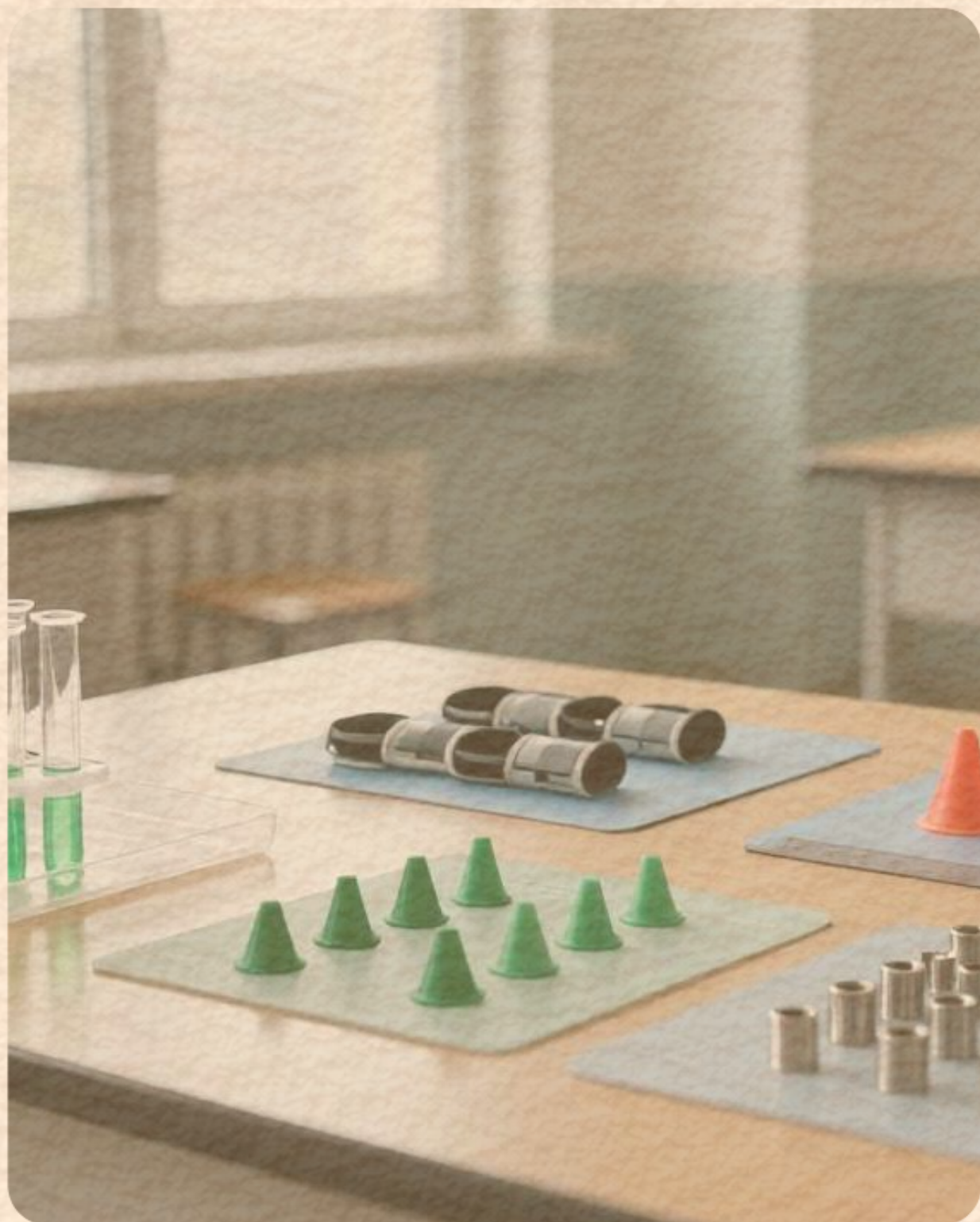


Как сравнивают наборы данных по среднему и дисперсии

Последовательность анализа показывает, как учитывать и центр данных, и их разброс.



Где используют дисперсию в жизни и науке



Школа и образование

В учебных данных дисперсия помогает понять, насколько ровно распределяются оценки и результаты тестов. Она показывает, где класс работает стабильно, а где успехи учеников заметно различаются.

Спорт и тренировки

В спорте по дисперсии оценивают стабильность результатов. Если спортсмен показывает почти одинаковые значения, его результаты более ровные, чем у того, у кого показатели сильно скачут.

Медицина и здоровье

В медицине разброс показателей важен при наблюдении за состоянием человека. Дисперсия помогает заметить, насколько сильно меняются измерения, например давление, пульс или другие данные наблюдения.

Производство и техника

На производстве проверяют, одинаковы ли детали по размеру и качеству. Небольшая дисперсия означает, что изделия получаются близкими друг к другу, а большая может указывать на нестабильность процесса.

Краткая памятка по интерпретации дисперсии

Таблица помогает быстро понять, как читать дисперсию в разных ситуациях и что означает маленький или большой разброс данных.

Малая дисперсия означает устойчивость и схожесть значений, а большая — заметные различия. Это помогает быстро оценить набор данных.

Ситуация	Малая дисперсия	Большая дисперсия
Оценки	Почти одинаковые	Сильно разные
Рост учеников	Близкие значения	Сильный разброс
Температура	Почти не меняется	Часто колеблется

Дисперсия показывает не только центр, но и характер данных

Среднее значение отвечает на вопрос, где находится центр набора, а дисперсия — насколько далеко от него расходятся числа. Вместе они дают более точное представление о данных и их изменчивости.