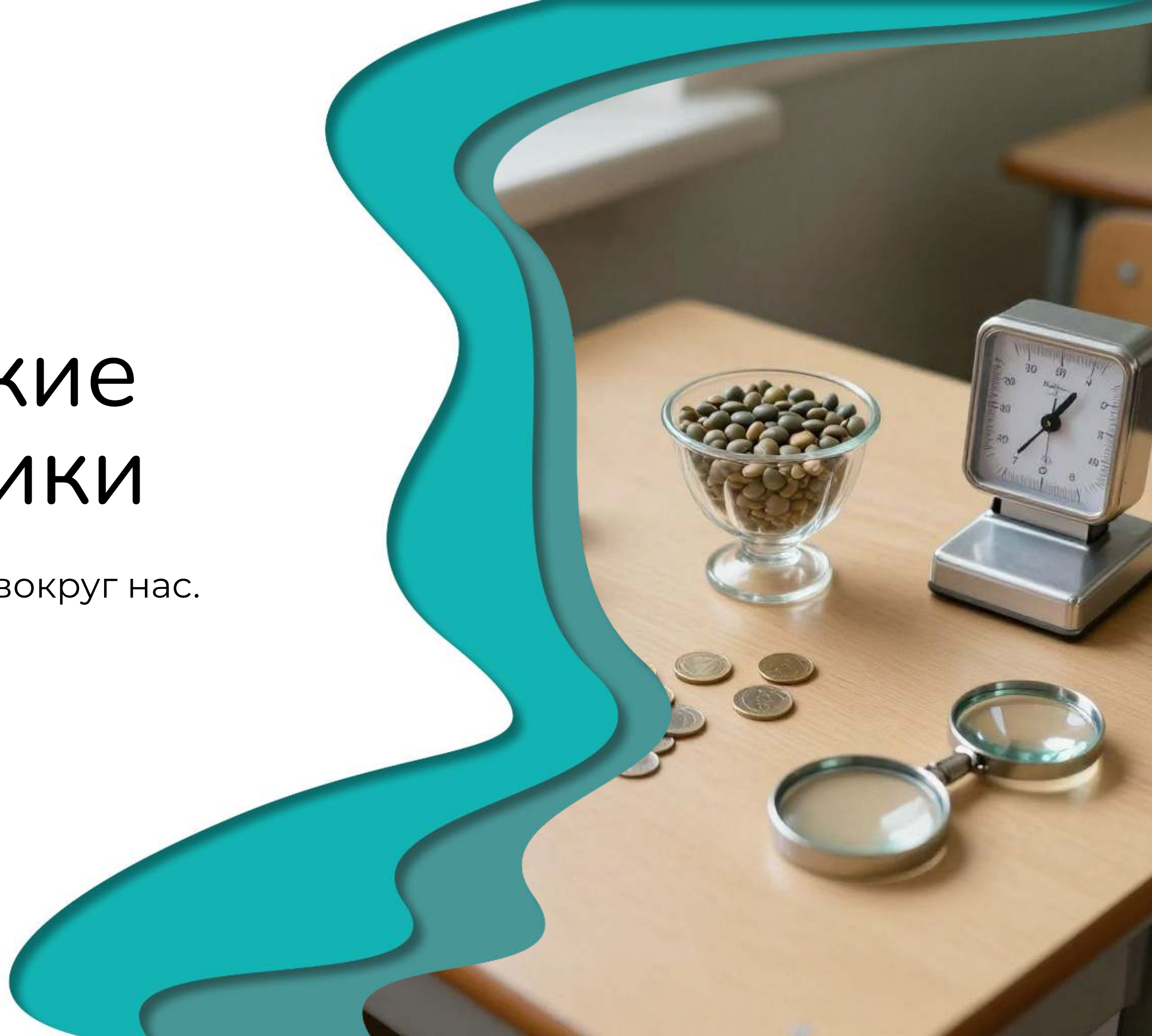


Статистические характеристики

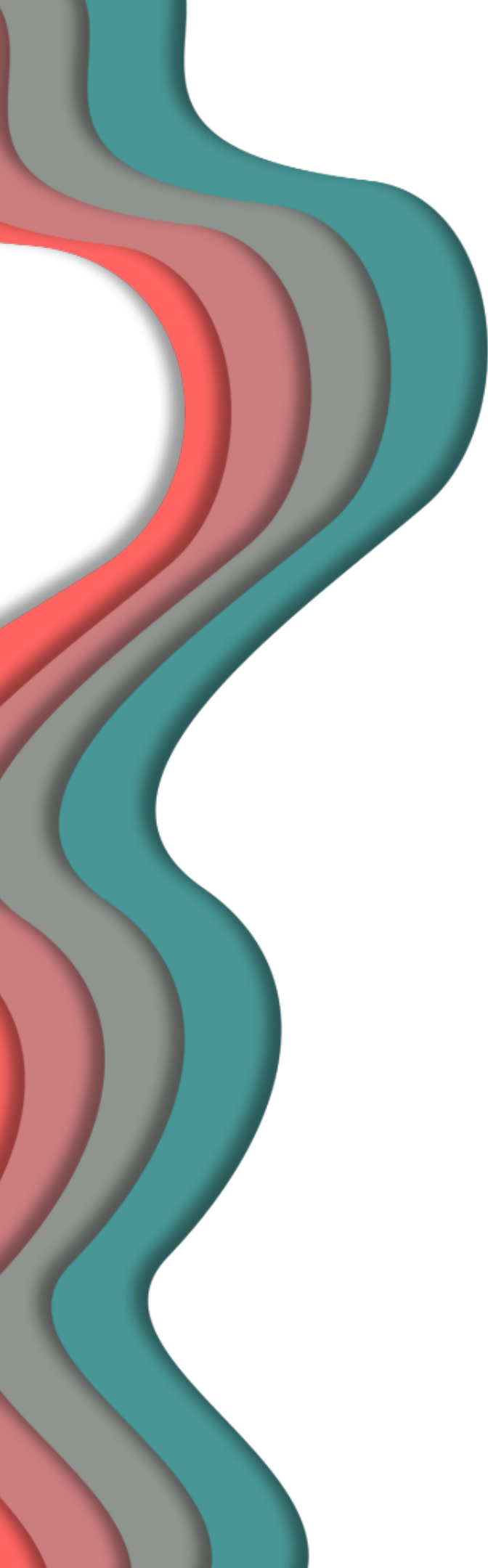
Они помогают понять данные вокруг нас.



Статистика в повседневной жизни

Статистика изучает данные из жизни и помогает делать выводы о большой группе по её части. Для этого важны точность, порядок записи и правильный выбор характеристики.





Зачем нужны статистические характеристики

01

Данные становятся понятнее

Оценки, рост, время и температура — это разные по смыслу числа. Статистические характеристики помогают превратить их в ясную картину, а не просто хранить как список значений.

03

Замечаем необычные значения

Если в списке есть слишком большое или слишком маленькое число, характеристики помогают заметить возможную ошибку или редкое событие, которое требует проверки.

02

Ищем центр и сходство

Они показывают, где находится центр набора, и насколько значения похожи между собой. Благодаря этому легче сравнивать результаты одного класса, команды или недели наблюдений.

04

Сравниваем группы правильно

Один и тот же набор можно сравнивать по-разному: по типичному значению, по разбросу или по самой частой величине. Выбор зависит от задачи.

Статистика и статистическая характеристика

Что делает статистика

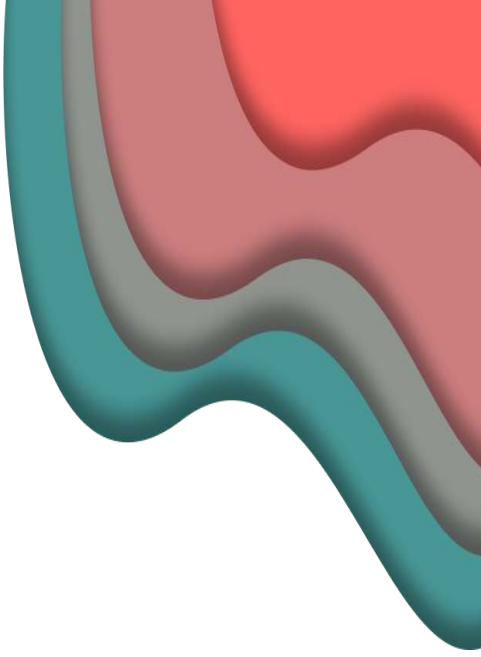
Статистика изучает данные, находит в них закономерности и помогает делать выводы, если чисел много и сразу увидеть картину трудно.

Что такое характеристика

Статистическая характеристика — это одно число, которое кратко описывает набор данных. Оно позволяет заменить длинный список более понятным итогом.

Два важных взгляда

Одни характеристики показывают типичное значение, а другие — насколько сильно данные отличаются друг от друга. Вместе они дают более полное описание.



Числовые данные и частоты

01 С числами удобно работать

Числовые данные можно упорядочить, сравнить и посчитать. Это удобно, когда нужно понять, какие значения встречаются чаще, а какие почти не повторяются.

02 Считаем повторения

Частота показывает, сколько раз одно и то же значение появляется в наборе. Она помогает увидеть повторяющиеся результаты и сделать таблицу данных нагляднее.

03 Простой пример частот

В наборе 3, 4, 4, 5 значение 4 встречается два раза, а остальные — по одному. Такие частоты сразу показывают, что именно повторяется.

Как подготовить данные к анализу

Записываем данные

Сначала значения записывают аккуратно и без пропусков, чтобы ничего не потерять и не перепутать при дальнейшем анализе.

Ставим по порядку

Затем данные упорядочивают, чаще всего по возрастанию. Так сразу видно, где находятся самые маленькие и самые большие значения.

Находим нужные характеристики

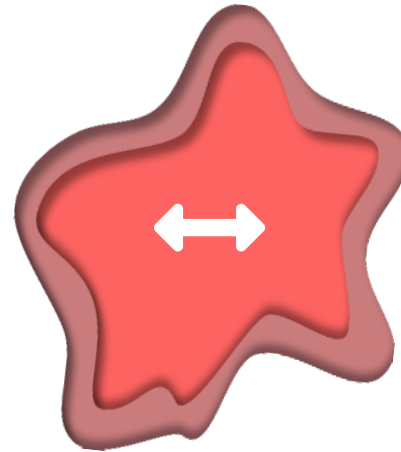
После этого легче находить минимум, максимум, медиану и квартиль, а значит — меньше риск ошибки в вычислениях.

Три меры центра



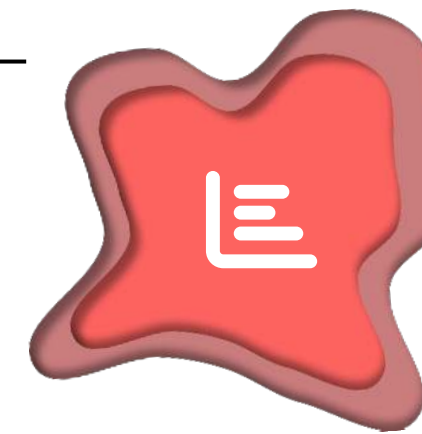
Среднее

Среднее подходит для наборов, где значения расположены близко друг к другу. Оно показывает общий уровень данных, но может заметно измениться из-за слишком большого числа.



Медиана

Медиана особенно полезна, когда в наборе есть выбросы. Она делит упорядоченные данные на две равные части и лучше сохраняет представление о центре.



Мода

Мода помогает найти самое частое значение. Она удобна, когда важно понять, какой ответ, размер или оценка встречаются чаще всего.

Среднее арифметическое показывает общий уровень набора, но одно большое значение может заметно сдвинуть результат вверх.

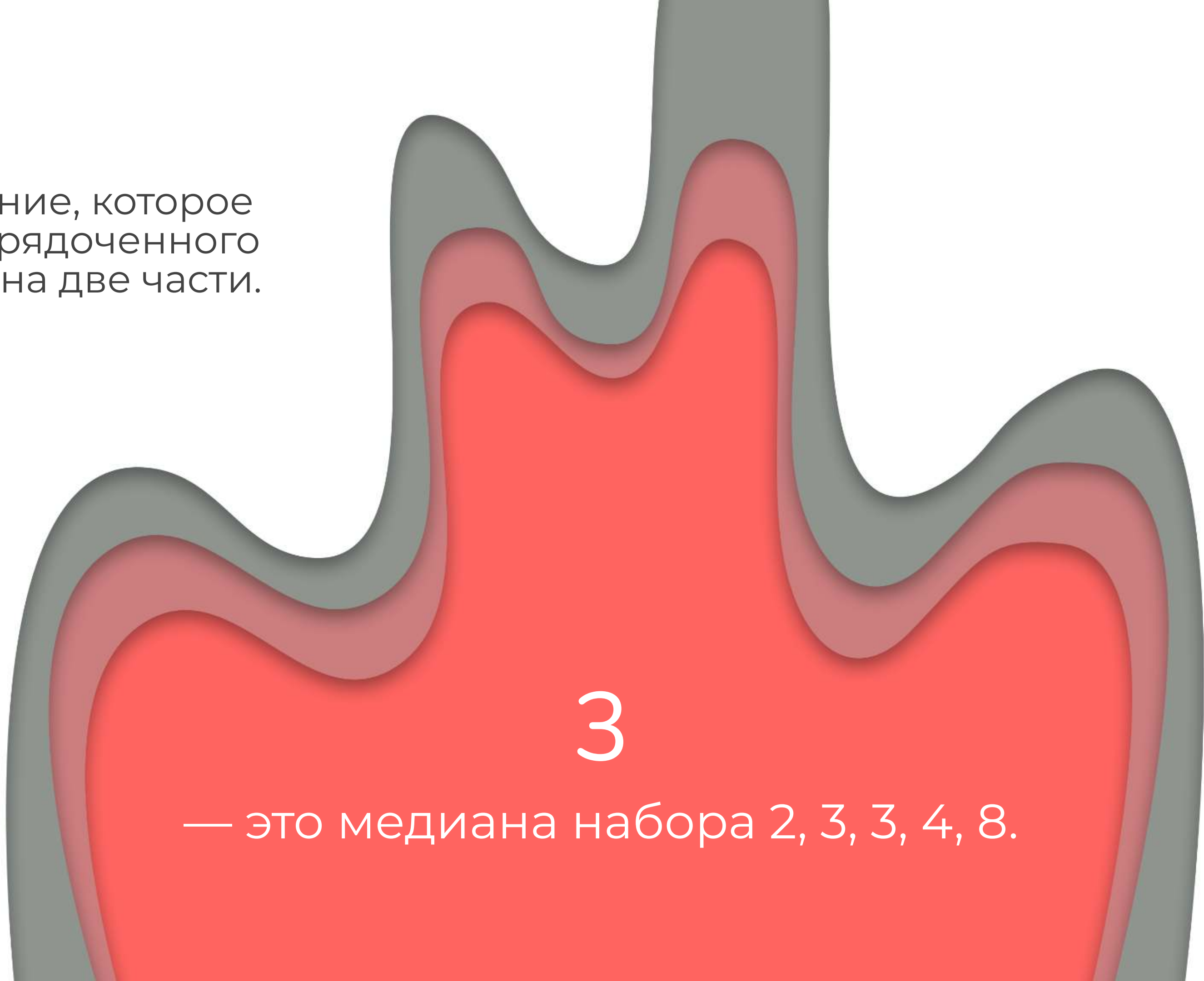
Пример из условия

4

получается как сумма значений,
делённая на их количество.

Медиана — это значение, которое стоит в середине упорядоченного ряда и делит данные на две части.

Пример из условия



3

— это медиана набора 2, 3, 3, 4, 8.

В этом ряду именно 5 встречается чаще остальных значений, поэтому оно лучше всего показывает самый популярный результат.

Пример из условия



5

— это мода данного набора.

Два способа увидеть центр данных



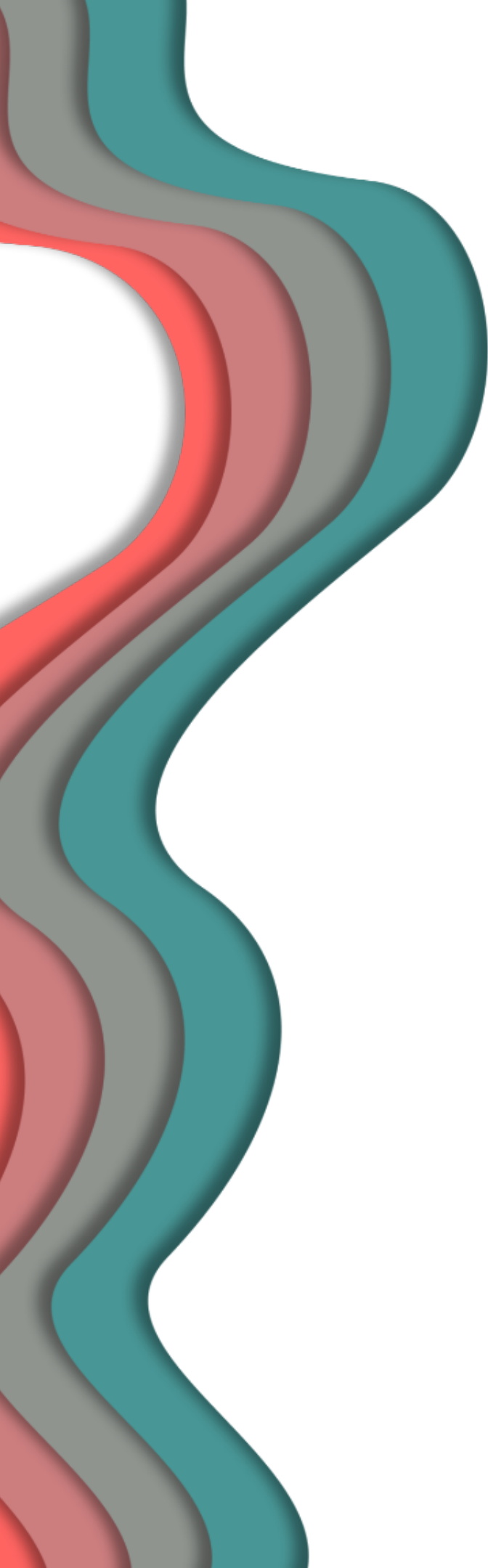
Когда одно число выделяется

В классе собрали оценки: 2, 3, 3, 4 и 8. Большинство чисел держится рядом с тремя и четырьмя, но одна восьмёрка заметно выбивается. Из-за неё среднее выглядит выше, чем основная часть данных.



Почему медиана устойчивее

Если убрать крайнее значение, картина становится понятнее: центр группы расположен ближе к трём. Медиана не «тянется» за большим числом и поэтому лучше показывает типичную оценку в таком наборе.



Три способа оценить разброс

01

Сравнить недостаточно

Разброс показывает, насколько значения похожи друг на друга. Если среднее одинаковое, но один класс получил почти одинаковые оценки, а другой — очень разные, без меры разброса это не заметить.

03

Устойчивый взгляд на середину

Квартильный размах описывает среднюю часть данных и меньше зависит от крайних значений. Поэтому он лучше подходит, когда в наборе есть редкие слишком большие или слишком маленькие числа.

02

Самая простая мера

Размах быстро сообщает границы набора: от минимума до максимума. Он полезен для первого взгляда, хотя не показывает, как распределены остальные значения внутри этих границ.

04

Точный показатель разброса

Стандартное отклонение показывает, насколько обычно значения отдаются от среднего. Чем оно больше, тем менее собран набор и тем сильнее данные расходятся вокруг центра.

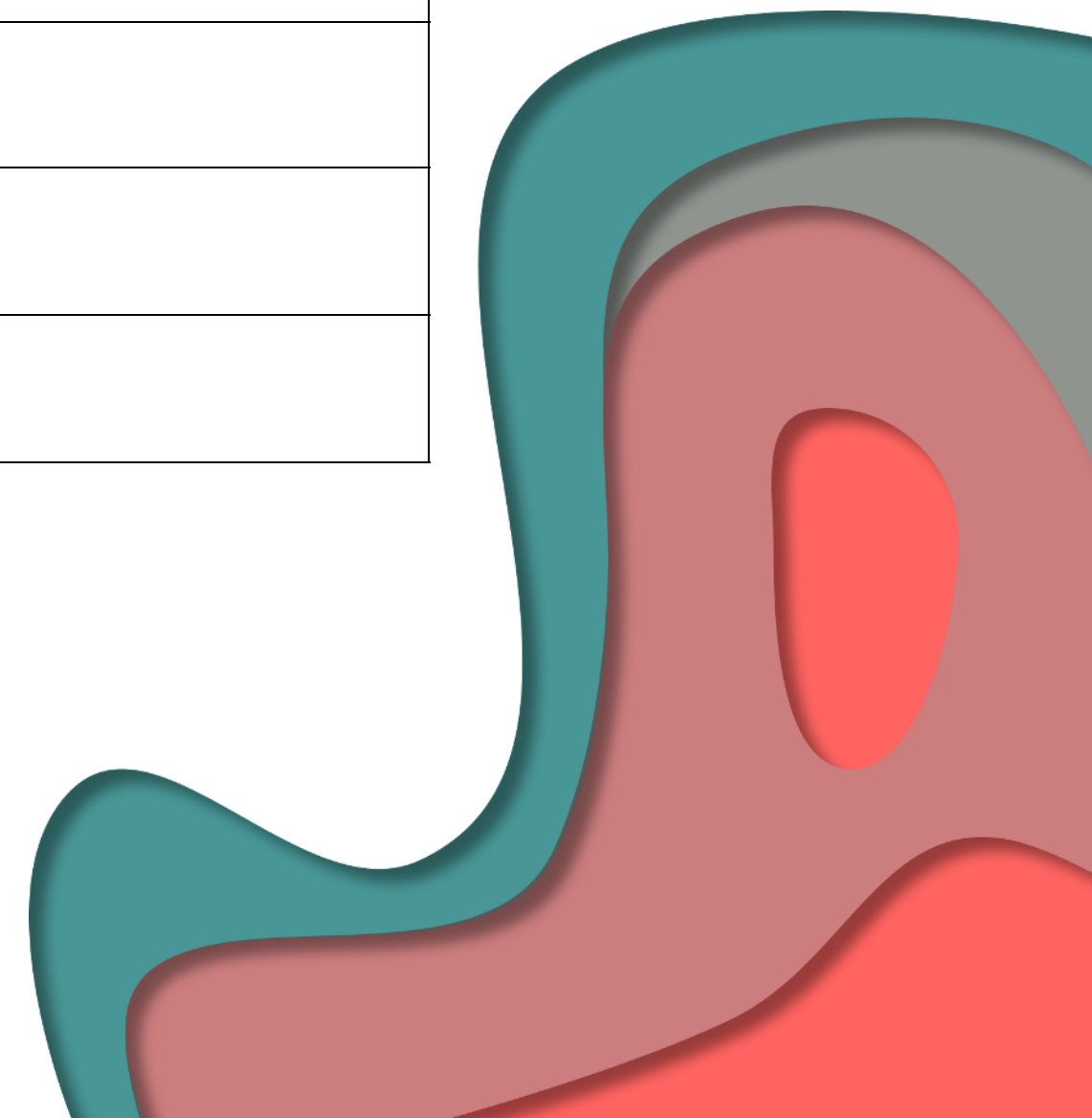
Размах времени бега

Минимум и максимум сразу показывают границы набора. По ним легко понять, насколько широким оказался диапазон значений, но этого недостаточно для полного описания данных.

Размах равен разнице между максимальным и минимальным значениями. Он быстро описывает границы, но не показывает, как распределены остальные результаты внутри диапазона.

Учебный пример по времени бега, без внешнего источника

Показатель	Значение
Минимум	12 с
Максимум	18 с
Размах	6 с



Как работают квартили

Упорядочить значения

Сначала данные упорядочивают по возрастанию. Так легче увидеть середину набора и разделить его на равные части без путаницы в порядке чисел.

Найти квартили

Затем находят нижний и верхний квартили. Они отделяют средние 50% данных от нижней и верхней четвертей и помогают описать основную группу значений.

Оценить центральный разброс

После этого вычисляют межквартильный размах как разницу между квартилями. Он показывает ширину центральной части и почти не зависит от крайних значений.

Показатели разброса вокруг среднего

01 Что измеряет дисперсия

Дисперсия показывает, насколько значения в среднем удалены от среднего арифметического. Если она мала, набор выглядит ровным; если велика, данные сильно расходятся.

02 Почему его удобно читать

Стандартное отклонение связано с дисперсией и выражается в тех же единицах, что и сами данные. Поэтому его удобнее воспринимать и сравнивать в привычных числах.

03 Где это применяют

Обе характеристики особенно полезны при сравнении классов, результатов контрольной или спортивных попыток. Они помогают понять не только центр, но и степень «собранности» данных.

04 Как делать вывод

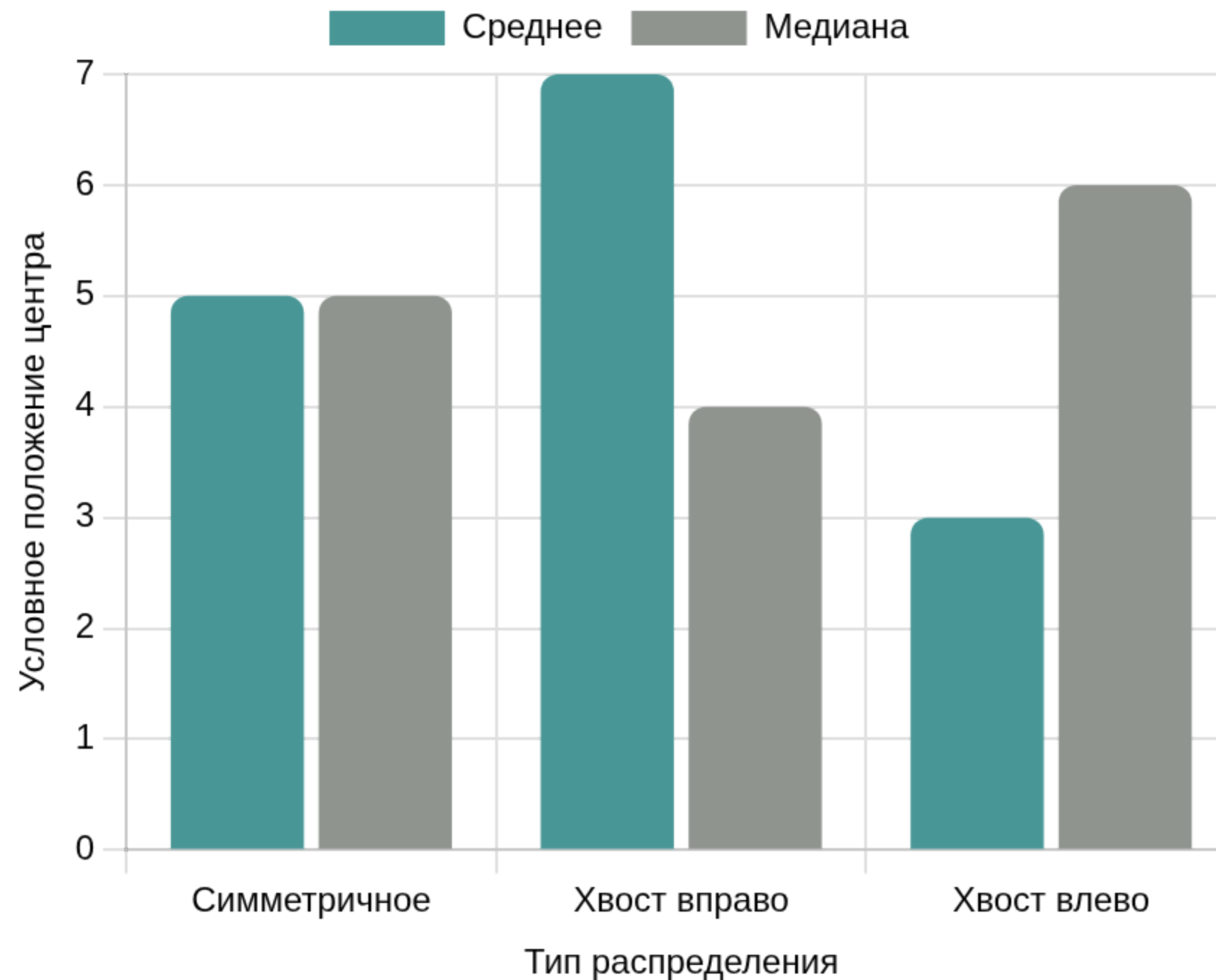
Важно смотреть на смысл, а не только на число. Чем больше разброс, тем труднее говорить, что все значения похожи друг на друга.

Как форма распределения влияет на выбор

На графике показаны три формы распределения. В симметричном случае центр почти совпадает, а при длинном хвосте одна из характеристик заметно смещается.

При симметрии среднее и медиана близки. При асимметрии медиана чаще лучше описывает центр, потому что меньше зависит от длинного хвоста.

Схематический учебный пример, без внешнего источника



Пошаговая проверка необычных значений

- **Сравнить с остальными**

Сначала смотрят на общий рисунок данных: если одно значение сильно отделяется от остальных, оно требует проверки. Важно понять, это ошибка записи, неверное измерение или редкое, но реальное событие.

- **Проверить запись**

Затем проверяют источник числа: лист наблюдений, прибор, дату и условие измерения. Один лишний ноль или опечатка могут превратить обычное значение в ложный выброс.

- **Принять решение**

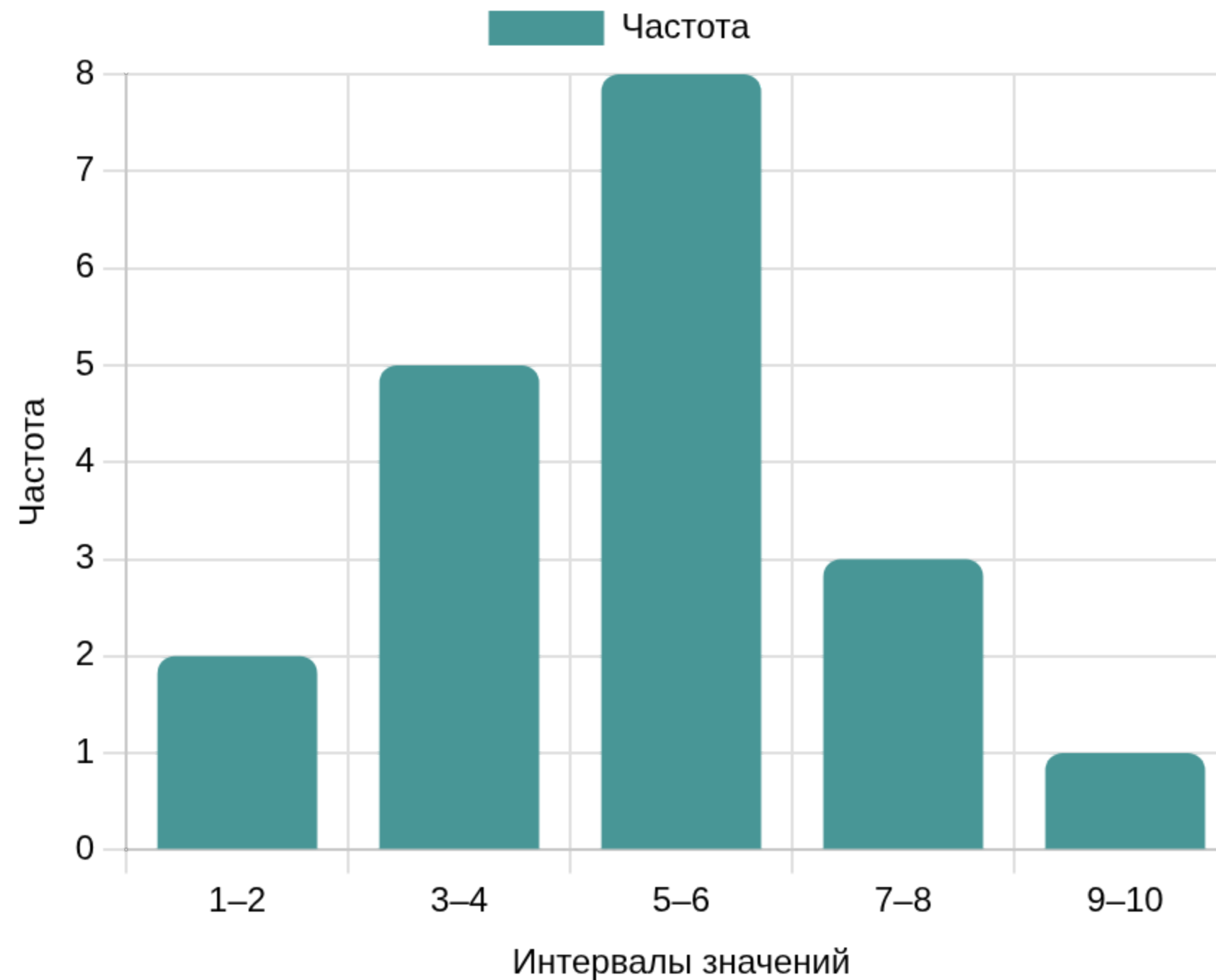
После проверки решают, что делать дальше. Иногда значение исправляют, иногда оставляют, но отдельно объясняют его влияние, чтобы среднее, размах и графики не ввели в заблуждение.

Гистограмма частот

Гистограмма помогает увидеть не отдельные числа, а форму всего набора. По высоте столбцов легко сравнить, какие интервалы встречаются чаще, а какие почти пустые.

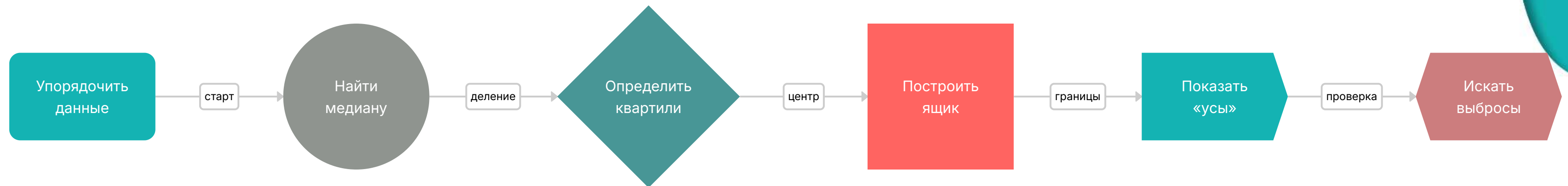
Наибольшая группа значений сосредоточена в середине, а по краям частота снижается. Это указывает на один основной пик и небольшой хвост справа.

Схематический учебный пример, без внешнего источника



От списка чисел к ящику с усами

Схема построения по упорядоченным значениям выборки



Главное о данных

Статистические характеристики превращают длинный список чисел в понятную картину. Если помнить, что центр и разброс отвечают на разные вопросы, легче выбирать правильный показатель и делать аккуратные выводы.

