

# Выборочный метод

один из главных методов математической статистики, включающий такие основные понятия, как *генеральная совокупность* и *выборочная совокупность*.

Будем считать, что  $x_1$  — результат измерения изучаемого показателя у первого спортсмена,  $x_2$  — у второго спортсмена и т. д., всего спортсменов  $n$  (например, если спортсменов 30, то будет и  $x_{30}$ , 30 различных индивидуальных показателей)

$n$  – объем выборки

Такой ряд результатов измерений, представленный случайными числами, называется выборочной совокупностью, или выборкой

Все результаты  $x$  могут быть двух видов:

Дискретные – выражаются целым числом (напр. кол-во подтягиваний)

Непрерывные – выражаются дробным числом (время пробегания 100 м – 13.4 сек)

# Ранжирование выборки – расстановка результатов в порядке возрастания или убывания

Пример выборочных результатов ( $n = 100$ )

№ п/п	1	2	3	4	5	6	7 ... 99	100
$x$ (кг)	46	50	59	60	55	49	... 58	60
$x$ (кг) (ранжиров.)	36	36	38	38	40	40	... 70	74

числа во второй строке записаны в той последовательности, в какой проходили измерения спортсмены, то есть случайным образом.

Такие данные представляют неупорядоченную выборку.

Третья строка — выборка упорядоченная, т.е. *ранжированная*

Если мы подсчитаем количество измерений каждого показателя и выстроим их в столбцы, то получим *вариационный ряд*.

**Вариационный ряд** — это двойной столбец ранжированных чисел, где слева стоит собственно показатель — вариант  $x_i$  («икс итое»), а справа — его количество — частота  $n_i$  («эн итое»)

Сумма частот называется объемом совокупности, то есть общим числом исходных данных.

### Виды вариационных рядов:

**простой** - это ряд, в котором каждая *варианта* встречается только по одному разу (все частоты при этом равны 1);

**взвешенный** - ряд, в котором одна или несколько *вариант* встречаются неоднократно

#### Пример:

в вариационном ряду показателей систолического артериального давления, измеренного у десяти пациентов:

**110, 120, 120, 130, 130, 130, 140, 140, 160, 170;**

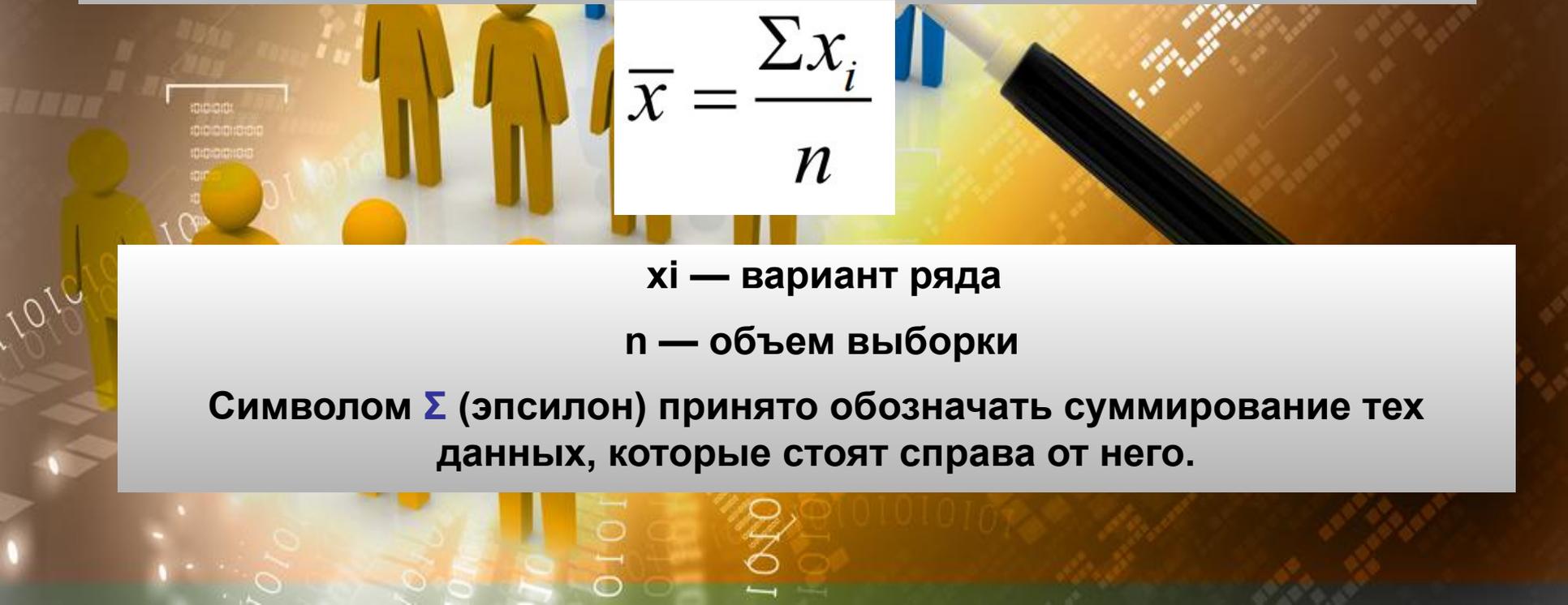
вариантами являются только 6 значений:

**110, 120, 130, 140, 160, 170.**

## Показатели вариационного ряда:

**Средняя арифметическая** - это обобщающий показатель, характеризующий размер изучаемого признака. Средняя арифметическая обозначается как ***M***, представляет собой самый распространенный вид средней.

Средняя арифметическая рассчитывается как *отношение суммы значений показателей всех единиц наблюдения к числу всех исследуемых*. Методика расчета средней арифметической различается для простого и взвешенного вариационного ряда.


$$\bar{x} = \frac{\sum x_i}{n}$$

**$x_i$**  — вариант ряда

**$n$**  — объем выборки

Символом  **$\Sigma$**  (эпсилон) принято обозначать суммирование тех данных, которые стоят справа от него.

**Мода** (обозначается символом  $M_o$ ) называют результат выборки или совокупности, наиболее часто встречающейся в этой выборке.



**Медиана** (обозначается символом  $M_e$ ) — результат измерения, который находится в середине ранжированного ряда.

Например, в некоторых видах спорта, где оценка спортсмену выставляется несколькими судьями (как в гимнастике), самые высокие и самые низкие оценки отбрасываются, и в зачет идет медиана.

Например, пять судей поставили 9,1 — 9,1 — 9,2 — 9,3 — 9,4 балла. Отбрасывая низшие и высшие оценки, получаем медиану, равную 9,2 балла. Эта оценка идет в зачет спортсмену

Все средние характеристики дают общую характеристику ряда результатов измерений.

На практике нас часто интересует, как сильно каждый результат отклоняется от среднего значения. Однако легко можно представить, что две группы результатов измерений имеют одинаковые средние, но различные значения измерений.

Например, для ряда **3, 6, 3** среднее значение = 4, а для ряда **5, 2, 5** среднее значение также = 4, несмотря на существенное различие этих рядов

Именно поэтому средние характеристики необходимо дополнять показателями *вариации*, или *колеблемости*.

Существуют определенные характеристики рассеивания: дисперсия, среднее квадратическое отклонение, коэффициент вариации, ошибка средней арифметической величины

*Дисперсия*  $\sigma^2$  указывает на варьирование, то есть рассеивание исходных данных относительно средней арифметической величины (в квадрате).

Дисперсия определяется по формуле:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}.$$

Если число измерений не более 30, то есть  $n \leq 30$ , используется формула

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}.$$

## Порядок нахождения дисперсии:

1. Определить среднюю арифметическую величину .

2. Из каждого варианта вычесть среднюю арифметическую величину:

$$x_i - \bar{x}.$$

3. Найденную разность возвести в квадрат:

$$(x_i - \bar{x})^2$$

4. Определить сумму :

$$\Sigma(x_i - \bar{x})^2$$

5. Найденную сумму разделить на объем выборки n.

Все средние характеристики дают общую характеристику ряда результатов измерений.

На практике нас часто интересует, как сильно каждый результат отклоняется от среднего значения. Однако легко можно представить, что две группы результатов измерений имеют одинаковые средние, но различные значения измерений.

Например, для ряда **3, 6, 3** среднее значение = 4, а для ряда **5, 2, 5** среднее значение также = 4, несмотря на существенное различие этих рядов

Именно поэтому средние характеристики необходимо дополнять показателями *вариации*, или *колеблемости*.

Существуют определенные характеристики рассеивания: дисперсия, среднее квадратическое отклонение, коэффициент вариации, ошибка средней арифметической величины