

Математическая статистика

Статистика -

- наука, которая занимается получением, обработкой и анализом количественных данных о разнообразных массовых явлениях, происходящих в природе и обществе.

Основная задача статистики

- Создание методов сбора и обработки статистических данных для получения научных и практических выводов.

- Генеральная совокупность – совокупность объектов, из которых производится выборка.
- Выборочная совокупность – совокупность случайно отобранных объектов.
- Объём совокупности – число объектов этой совокупности.

- Выборка с возвращением – выборка, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность.
- Выборка без возвращения – выборка, при которой отобранный объект не возвращается в генеральную совокупность.

Статистические характеристики:

- характеристики среднего, описывающие положение всего числового ряда в целом на числовой прямой;
- характеристики разброса, показывающие, насколько значения ряда различаются между собой, как сильно они разбросаны, рассеяны вокруг средних.

К характеристикам среднего относятся:

- среднее арифметическое;
- мода;
- медиана.

К характеристикам разброса относятся:

- размах;
- дисперсия;
- стандартное отклонение.

- Средним арифметическим ряда чисел называется частное от деления суммы этих чисел на их количество.
- Модой называют число ряда, которое встречается в этом ряду наиболее часто.
- Моды у ряда может вообще не быть или быть несколько мод (полимодалный ряд).
- Размах – это разность наибольшего и наименьшего значений ряда данных.

- **Медианой** ряда, состоящего из нечетного количества чисел, называется число данного ряда, которое окажется посередине, если этот ряд упорядочить. Медианой ряда, состоящего из четного количества чисел, называется среднее арифметическое двух стоящих посередине чисел этого ряда.

Дисперсия

- Дисперсия в статистике очень важный показатель, который активно используется в других видах статистического анализа (проверка гипотез, анализ причинно-следственных связей и др.). Как и среднее линейное отклонение, дисперсия также отражает меру разброса данных вокруг средней величины.
- Формула для расчета дисперсии выглядит так:

$$D = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n}$$

- где D – дисперсия, x – анализируемый показатель, с черточкой сверху – среднее значение показателя, n – количество значений в анализируемой совокупности данных.

Стандартное отклонение

■ Чтобы использовать дисперсию, из нее извлекают квадратный корень. Получается так называемое **стандартное отклонение**. В статистике этот показатель еще называют среднеквадратическим отклонением, но первое название более короткое и распространенное. Формула стандартного отклонения имеет вид:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}}$$

- Рассмотрим ряд данных:
1; 3;3;2;1;6;4 и определим его статистические характеристики:
- Среднее арифметическое: 2,86
- Мода: 1 и 3 – полимодальный ряд
- Медиана: 3 (1;1;2;3;3;4;6)
- Размах:5
- Дисперсия: $((1-2,86)^2 + (1-2,86)^2 + (2-2,86)^2 + (3-2,86)^2 + (3-2,86)^2 + (4-2,86)^2 + (6-2,86)^2) : 7 = 2,69$
- Стандартное отклонение: $\sqrt{2,69} = 1,64$