

Дисперсия (Variance)

Разделы: [Метрики](#)

Loginom: [Статистика \(визуализатор\)](#).

В статистике дисперсией называют величину, которая характеризует меру разброса значений случайной величины относительно ее математического ожидания. В русскоязычной литературе дисперсия обозначается $D[X]$, а в англоязычной $var(X)$ (от англ. variance — дисперсия).

Пусть X — случайная величина, определенная на некотором вероятностном пространстве. Тогда дисперсией называется

$$D[X] = M [(X - M[X])^2],$$

где M — математическое ожидание.

- Если случайная величина X дискретная, то: $D[X] = \sum_{i=1}^n p_i (x_i - M[X])^2$, где x_i — i -ое значение случайной величины, p_i — вероятность того, что случайная величина принимает значение x_i , n — количество значений случайной величины.
- Если случайная величина X непрерывна, то: $D[X] = \int_{-\infty}^{\infty} f(x)(x - M[X])^2 dx$, где $f(x)$ — плотность вероятности случайной величины.

Квадратный корень из дисперсии, обозначаемый σ , называется среднеквадратическим отклонением.

Свойства дисперсии:

- Дисперсия любой случайной величины неотрицательна: $D[X] \geq 0$;
- Если дисперсия случайной величины конечна, то конечно и ее математическое ожидание;
- Если случайная величина равна константе, то ее дисперсия равна нулю: $D[a] = 0$.
- Дисперсия суммы двух случайных величин равна:
 $D[X + Y] = D[X] + D[Y] + 2 \operatorname{cov}(X, Y)$, где $\operatorname{cov}(X, Y)$ — их ковариация.
- Для дисперсии произвольной линейной комбинации нескольких случайных величин имеет место равенство: $D \left[\sum_{i=1}^n c_i X_i \right] = \sum_{i=1}^n c_i^2 D[X_i] + 2 \sum_{1 \leq i < j \leq n} c_i c_j \operatorname{cov}(X_i, X_j)$, где $c_i \in \mathbb{R}$.

Дисперсия является одним из параметров нормального закона распределения. Чем больше дисперсия, тем более пологими являются «склоны» распределения и длиннее его «хвосты».

Чем выше дисперсия параметров модели (коэффициентов регрессии, значений переменных и т.д.), тем менее устойчивой она будет. Высокая дисперсия исходных данных позволяет предположить высокую значимость в них случайной компоненты, возможном наличии шума и аномальных значений.