

Правило трех сигм (3-sigma rule)

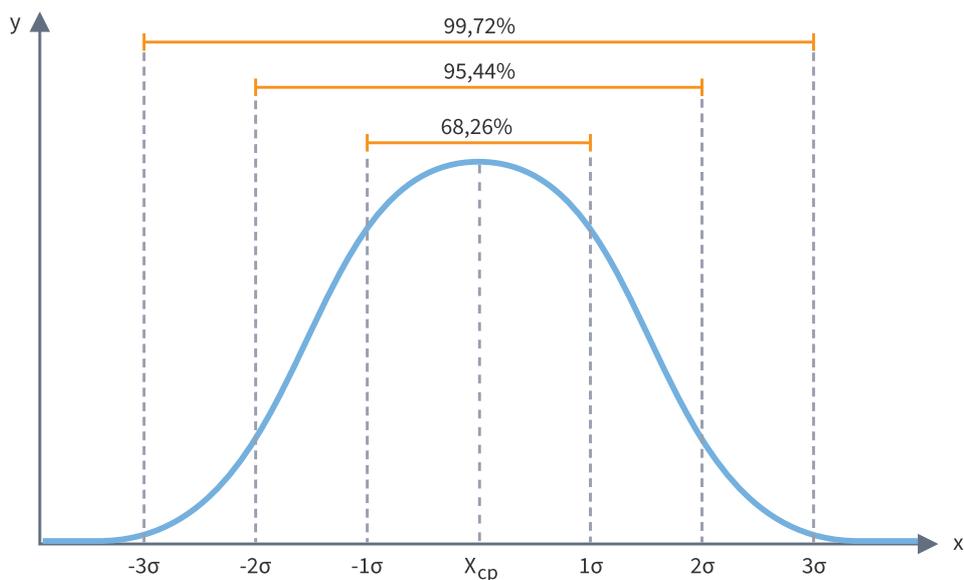
Разделы: [Алгоритмы](#)

Loginom: [Редактирование выбросов \(обработчик\)](#)

Правило, утверждающее, что для любой случайной величины ξ с конечной дисперсией вероятность того, что случайная величина отклонится от своего математического ожидания $M[\xi]$ не менее, чем на три среднеквадратических отклонения σ , не более $\frac{1}{9}$:

$$P(|\xi - M[\xi]| \geq 3\sigma) \leq \frac{1}{9}$$

Для большинства случайных величин эта вероятность меньше, например, для нормального распределения:



На рисунке видно, что в пределах одного среднеквадратического отклонения лежит 68,26% значений, принимаемых нормально распределенной случайной величиной (соответствует доли площади под кривой распределения). В пределах двух среднеквадратических отклонений — уже 95,44%, а в пределах трех — 99,72%. Это означает, что вероятность того, что случайная величина примет значение, отклоняющееся от математического ожидания больше чем на три среднеквадратических отклонения, не превышает 0,28%, т.е. пренебрежимо мала.

Например, пусть имеется выборка наблюдений за ежедневными продажами в магазине. Значения наблюдений распределены по нормальному закону со средним значением 150 000 руб. и среднеквадратическим отклонением 20 000 руб. Тогда, в соответствии с

правилом 3-х сигм, продажи ниже, чем $150\,000 - 20\,000 \times 3 = 90\,000$, и выше, чем $150\,000 + 20\,000 \times 3 = 210\,000$, являются практически невозможными событиями. Фактически это означает, что рассматривать данные объемы продаж как потенциально возможные не имеет смысла.