# §2. «Простая» статистика

В большинстве статистических расчетов приходится работать с выборками случайной величины: либо с данными эксперимента, либо с результатами моделирования методом Монте-Карло. Приведем (конспективно) формулы вычисления основных статистических характеристик случайной выборки $y_i$.

## *Среднее и дисперсия*

- Среднее значение: $m = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$ , (1)
- Дисперсия: $D = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - m)^2$ , (2)
- Среднеквадратичное отклонение: $\sigma = \sqrt{D}$ . (3)

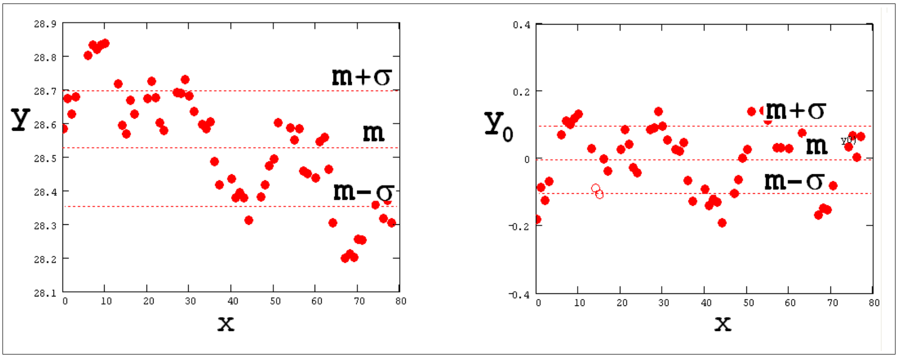


***Рис. 17.*** *К определению среднего и дисперсии*

Смысл приведенных статистических характеристик иллюстрируется рис. 17 (слева для выборки колебаний курса доллара y, как на рис. 2, а справа для $y_0$ – тех же данных, из которых вычтен линейный тренд, как на рис. 6).

## *Выборочные функции распределения*

Напомним, что, согласно определению, случайная величина принимает то или иное значение, но какое конкретно – зависит от принципиально непредсказуемых обстоятельств опыта и заранее не может быть точно определено. Можно лишь говорить о вероятности P(y) принятия случайной дискретной величиной некоторого значения y или о вероятности попадания непрерывной случайной величины в числовой интервал (y, y+Δ). Последнее означает, что вероятность непрерывной случайной величины принять значение между y и y+Δ составляет P(y)·Δ.

Вероятность принимает значения от 0 до 1. Функцию P(y) называют *плотностью вероятности*.

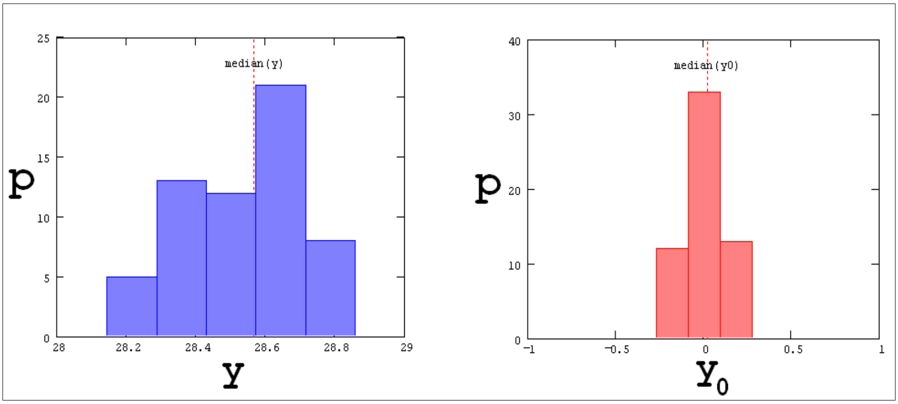


***Рис. 18.*** *Гистограммы распределения случайных величин y и $y_0$ (курсы USD и USD - тренд соответственно)*

График, аппроксимирующий по выборочным данным плотность вероятности их распределения, называется *гистограммой*. При построении гистограмм обычно сначала определяются границы интервала (a,b), содержащего внутри себя все случайные значения. Затем осуществляется его разбиение на выбранное количество (равных или неравных) элементарных сегментов. Наконец, подсчитывается процент попадания данных в каждый сегмент (рис. 18). Задание неодинаковых сегментов разбиения (с разной шириной) полезно, когда плотность вероятности сильно различается для разных значений аргумента, как например, для гауссового распределения.

Одной из статистических характеристик случайной величины является *медиана* – значение аргумента, которое делит гистограмму плотности вероятностей на две равные части (выборочное медианное значение показано на рис. 18 пунктиром).

Помимо плотности вероятности часто используется *функция распределения* F(y), равная вероятности того, что случайная величина примет значение, меньшее или равное ее аргументу y. Как следует из математического смысла, она является интегралом от плотности вероятности в пределах от -∞ до y. Функция, обратная F(x), называемая еще *квантилем распределения*, позволяет по заданному аргументу P определить значение y, причем случайная величина будет меньше или равна y с вероятностью P.

## *Равномерное и Гауссово распределение*

Особенное значение в математической статистике имеет *нормальное* (или *гауссово*) распределение. Плотность вероятности для него имеет вид:

$$p(y,m,\sigma)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left(\frac{-(y-m)^2}{2\sigma^2}\right) \quad . \qquad (4)$$

Таким образом, закон нормального распределения зависит от двух параметров: математического ожидания (среднего) m и стандартного отклонения σ (дисперсия, соответственно, равна $\sigma^2$).

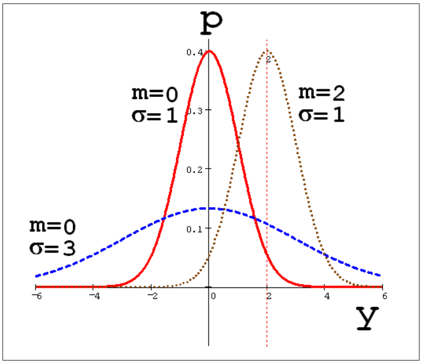


***Рис. 19.*** *Плотность вероятности нормального распределения*

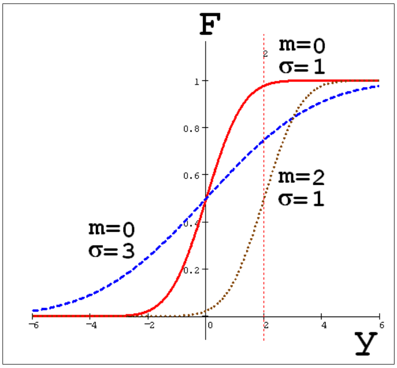


***Рис. 20.*** *Функция нормального распределения*

Плотности нормального распределения для трех пар значений параметров m и σ показана на рис. 19, а соответствующие функции распределения F(y) – на рис. 20. Напомним, что плотность распределения p(y) задает вероятность попадания случайной величины y в малый интервал от y до y+Δy. Таким образом, например, для графика с параметрами m=0 и σ=1 вероятность того, что случайная величина x примет значение в окрестности нуля, приблизительно в три раза больше, чем вероятность того, что она примет значение в окрестности x=2. А значения случайной величины, большие 5 и меньшие -5, и вовсе очень маловероятны. Как можно заметить, сравнивая три графика, значение математического ожидания влияет на их смещение вдоль оси абсцисс, а дисперсии – на их профиль (он тем шире, чем больше σ). Функция распределения F(y) задает вероятность того, что случайная величина примет значение, меньшее, чем ее аргумент y.

Особая роль, которую играет нормальное распределение, заключается в том, что в теории вероятности доказано: сумма различных независимых случайных слагаемых (с любым законом распределения) оказывается случайной величиной, распределенной согласно нормальному закону (так называемая *центральная предельная теорема*). Поэтому нормальное распределение хорошо моделирует самый широкий круг экспериментов, для которых известно, что на них влияют несколько независимых случайных факторов.

Еще одно, повсеместно использующееся распределение – равномерное – определяет случайную величину, вероятность принятия которой определенного значения на интервале (a,b) является константой. Плотность вероятности и функция распределения для равномерного закона приведена на рис. 21. Важная практическая роль равномерного распределения заключается в том, что алгоритм его генерации заложен микропрограммно практически во все вычислительные пакеты, что дает очень простую возможность генерации псевдослучайных выборок. Соответствующие примеры будут приведены в следующем параграфе.