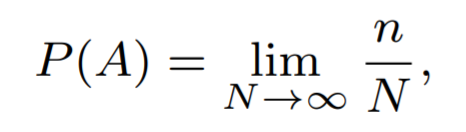
Статистики

* Каждый закон распределения – это некоторая функция, полностью описывающая случайную величину с вероятностной точки зрения.

На практике о распределении вероятностей случайной величины *Х* - по результатам испытаний – по эмпирической функции распределения.

* Относительной частотой случайного события А называется отношение числа появлений *n* этого события к общему числу *N* проведенных испытаний.
* При неограниченном числе испытаний n\* относительная частота события А сходится к вероятности этого события



* Эмпирическая функция распределения (функция распределения выборки) - это функция *P*\*(*x*), которая определяет для каждого значения *xi*относительную частоту события *X* < *x*.
* Эмпирические частоты получают в результате опыта (наблюдения).

# **Оценка распределения по выборке**

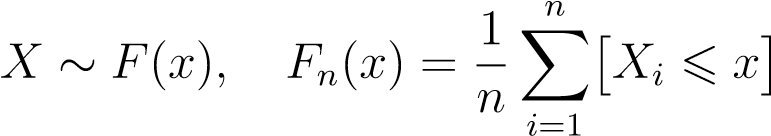
Рассматривается выборка из случайной величины *X*:

*,*

где *n* — объем выборки. Величины *X*1*, X*2*,...,Xn* — независимые одинаково распределенные случайные величины .

Статистикой называется любая функция от данной выборки.

Если *непрерывная случайная величина* задается с помощью функции распределения, то ее можно оценить с помощью эмпирической функции распределения:

*.*

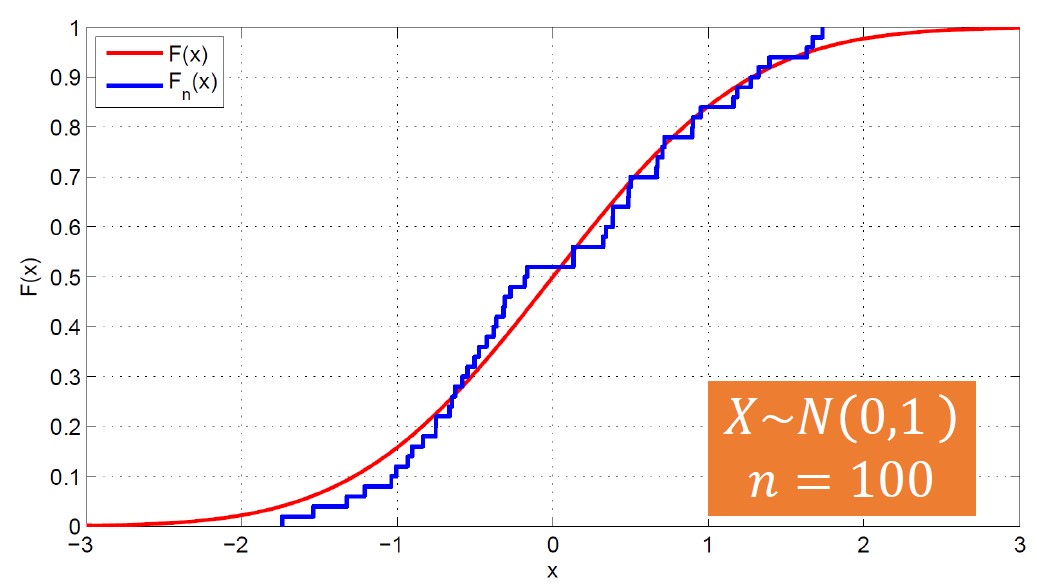
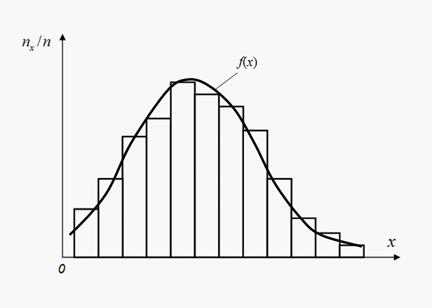


Рис. 1.

На рис. 1 красная линия соответствует теоретической функции стандартного нормального распределения (нормальное распределение со средним, равным нулю, и с дисперсией, равной 1). Синяя линия соответствует эмпирической функции распределения, построенной по выборке объема 100.

Непрерывные случайные величины также могут задаваться с помощью плотностей. Для оценки плотности можно разбить область определения случайной величины на интервалы одинаковой длины. Количество объектов выборки в каждом интервале будет пропорционально среднему значению плотности на нем. Именно так устроена гистограмма.



Для построения гистограммы по оси абсцисс откладывают интервалы равной длины, на которые разбивается весь диапазон возможных значений случайной величины *Х*, а по оси ординат откладывают частоты *nx*/ *n*. Тогда высота каждого столбика гистограммы равна соответствующей частоте. Таким образом, получается приближенное представление закона распределения вероятностей для случайной величины *Х* в виде ступенчатой функции, аппроксимация (выравнивание) которой некоторой кривой *f*(*x*) даст плотность распределения.

* **Отличие эмпирической функции от теоретической состоит том, что теоретическая функция определяет вероятность события.**
* При этом мы принимаем, что относительные частоты случайных событий близки к их вероятностям. Это тем более верно, чем больше число проведенных опытов. При этом частоты, как и вероятности, следует относить не к отдельным значениям случайной величины, а к интервалам.

На гистограмме, приведенной на рис. 2, изображена продолжительность жизни крыс на строгой диете (в днях).

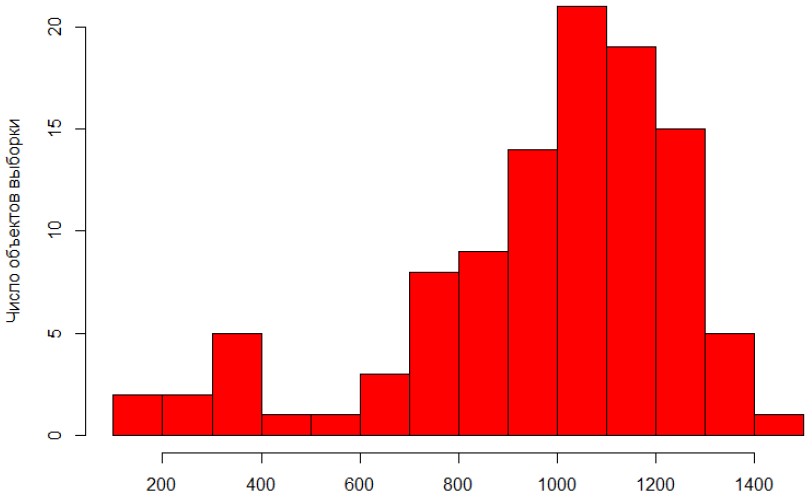


Рис. 2.

По такой гистограмме хорошо видны все особенности распределения данных: основной пик приходится примерно на 1000 дней, но есть крысы, которые живут существенно меньше.

Важным аспектом работы с гистограммами является правильный выбор числа интервалов.

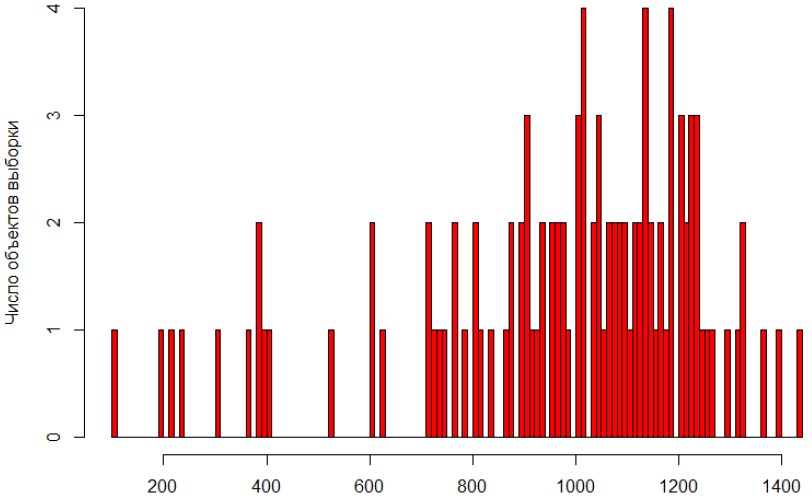
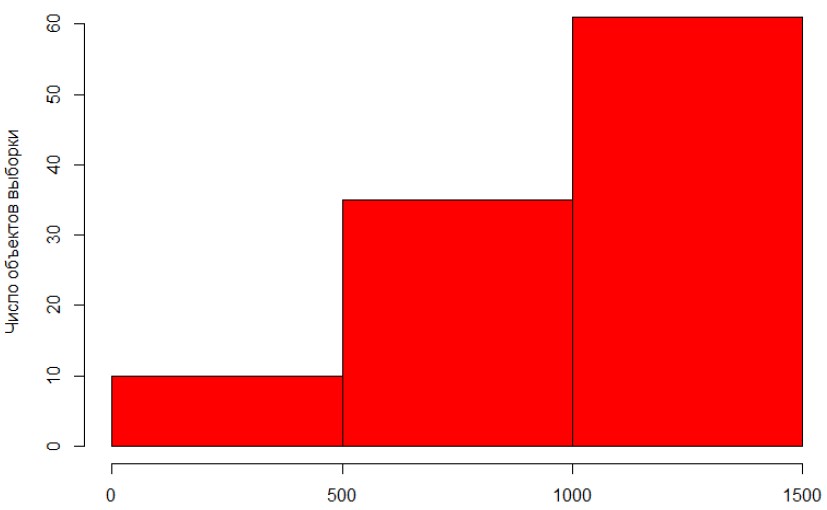
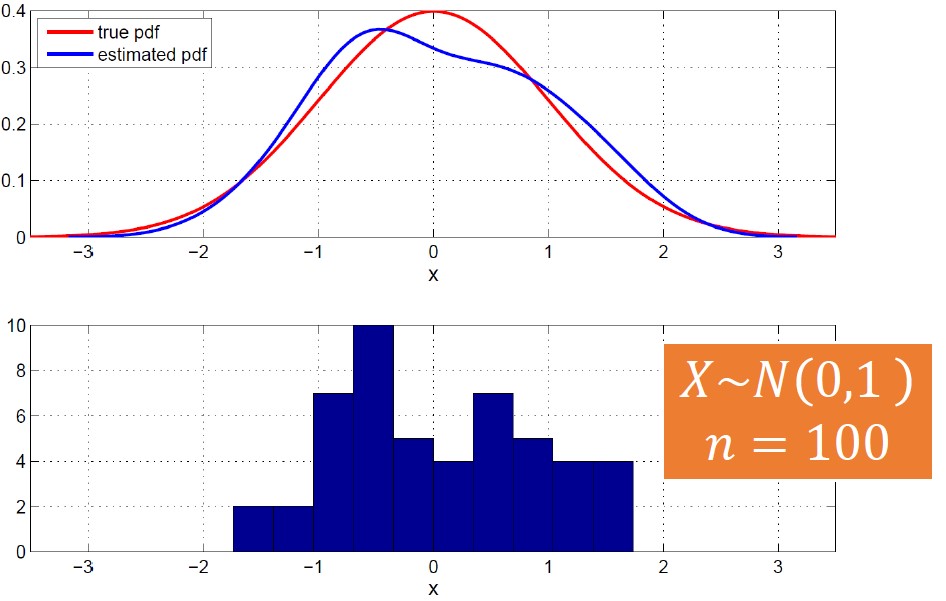


Рис. 3. Рис. 4.

Если рассмотреть слишком мало интервалов, то они будут слишком большими, в результате гистограмма получится грубой (см. рис. 3). Аналогично в случае слишком большого количества интервалов — в большую часть из них не попадет ни одного объекта выборки (см. рис. 4). В обоих случаях построенные гистограммы не являются информативными.



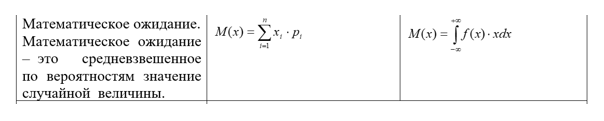
**Эмпирические частоты** получают в результате наблюдения. **Теоретические частоты** рассчитывают по формулам.

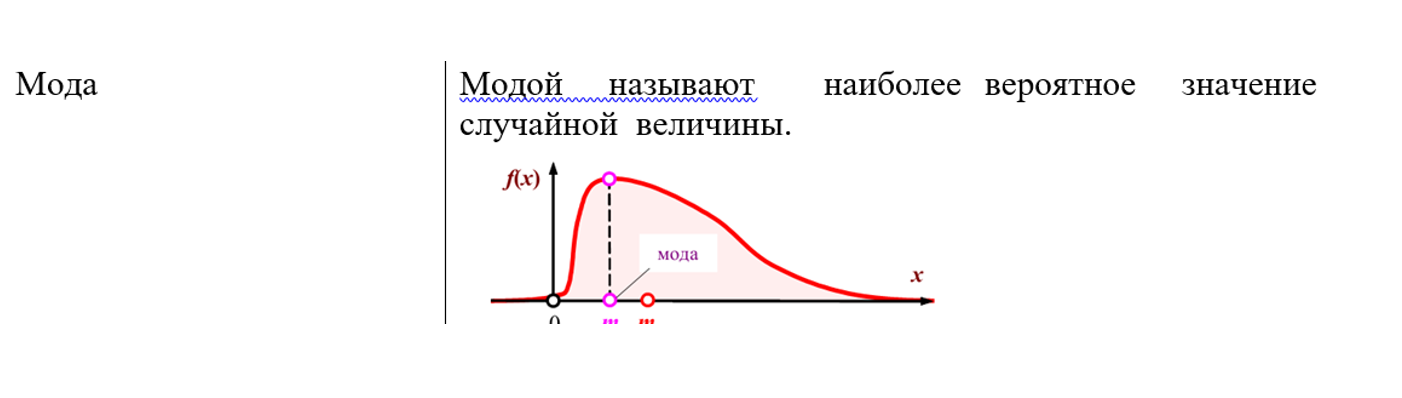
Необходимо отметить, представленный способ оценки плотностей не является идеальным.

Поэтому часто оценивают характеризующие основные свойства распределения (мат ожидание, медиана, дисперсия…)

# **Важные характеристики распределений**

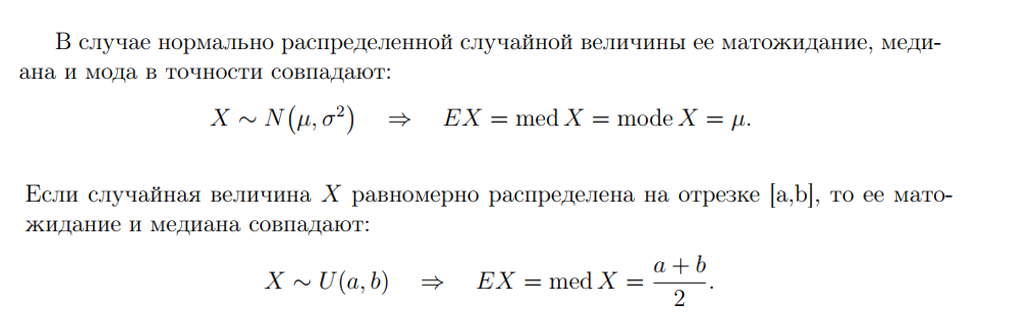
Часто возникает необходимость оценить не всю функцию распределения, а некоторые ее параметры. Самым важным классом параметров распределения являются средние. Нестрогое определение можно сформулировать следующим образом: среднее — это значение, вокруг которого группируются все остальные.





**Медиана случайной величины** — такое ее значение, для которого одинаково вероятно, окажется ли случайная **величина** больше или меньше **медианы**





мода

медиана

матожидание

**Мера разброса значений случайной величины относительно её математического ожидания**

