



**Статистическое исследование** — это научно организованный по единой программе сбор, сводка и анализ данных (фактов) о каких-то жизненных, либо других аспектах, с регистрацией их наиболее существенных признаков в учетной документации.

**Отличительными чертами (спецификой) статистического исследования являются:**

целенаправленность,  
организованность,  
массовость,  
системность (комплексность),  
сопоставимость,  
документированность,  
контролируемость,  
практичность.

## Этапы статистического исследования

1) Статистическое наблюдение - формируются первичные статистические данные, или исходная статистическая информация, которая является основой статистического исследования. Если при сборе первичных статистических данных допущена ошибка или материал оказался недоброкачественным, это повлияет на правильность и достоверность как теоретических, так и практических выводов;

2) Сводка и группировка данных - на этой стадии совокупность делится по признакам различия и объединяется по признакам сходства, подсчитываются суммарные показатели по группам и в целом. С помощью метода группировок изучаемые явления в зависимости от существенных признаков подразделяются на типы, группы и подгруппы. Метод группировок позволяет ограничивать качественно однородные в существенном отношении совокупности, что служит предпосылкой для определения и применения обобщающих показателей;

3) Обработка и анализ полученных данных, выявление закономерностей.

На этом этапе с помощью обобщающих показателей рассчитываются относительные и средние величины, дается сводная оценка вариации признаков, характеризуется динамика явлений, применяются индексы, балансовые построения, рассчитываются показатели, характеризующие тесноту связей в изменении признаков. С целью наиболее рационального и наглядного изложения цифрового материала он представляется в виде таблиц и графиков.

**Нормальное распределение** (также называемое распределением Гаусса) — распределение вероятностей, которое в одномерном случае задается функцией плотности вероятности, совпадающей с функцией Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

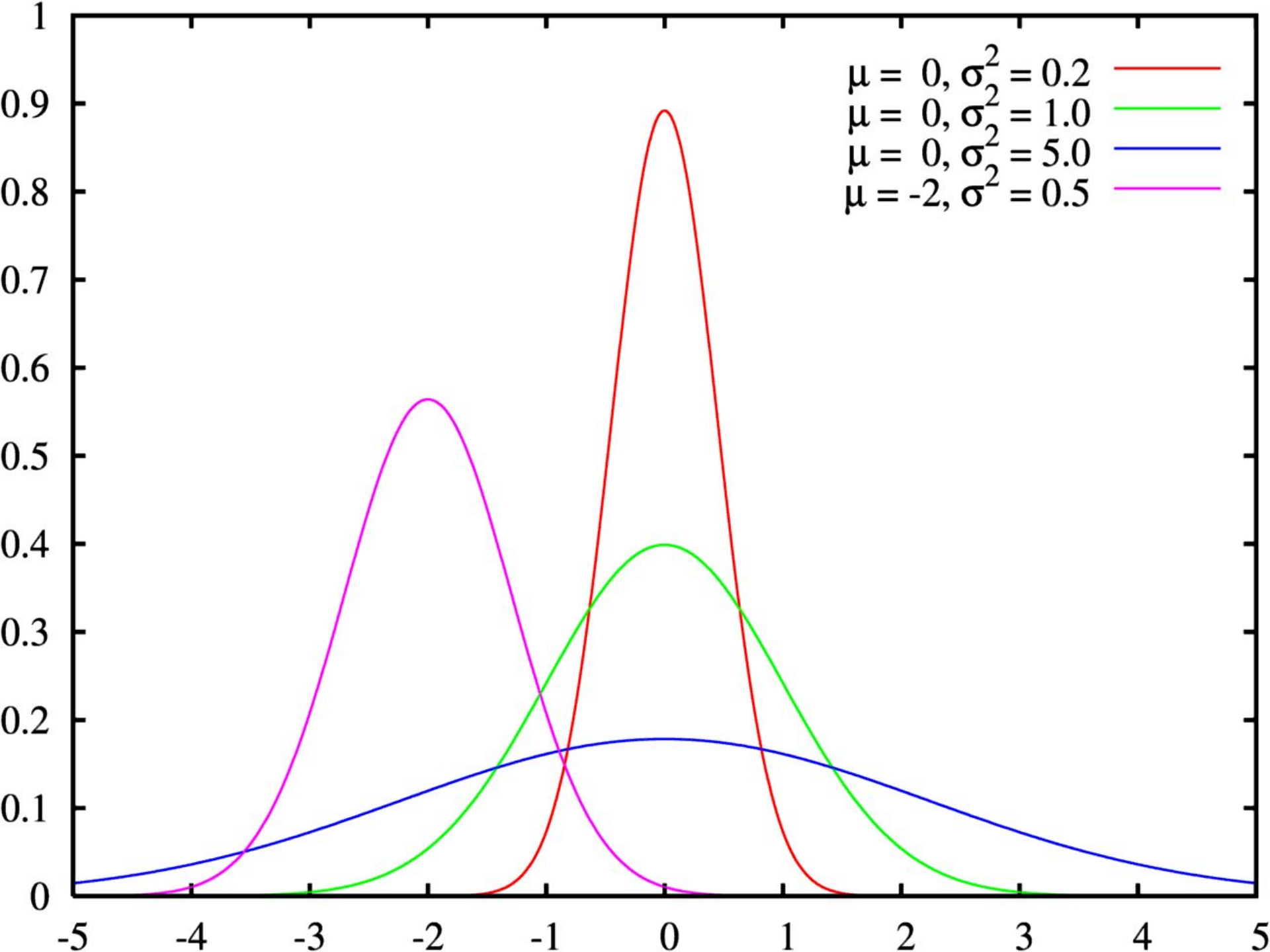
где параметр  $\mu$  — математическое ожидание (среднее значение), медиана и мода распределения, а параметр  $\sigma$  — среднеквадратическое отклонение ( $\sigma^2$  — дисперсия) распределения.

Таким образом, одномерное нормальное распределение является двухпараметрическим семейством распределений.

**Стандартным нормальным распределением** называется нормальное распределение с математическим ожиданием  $\mu = 0$  и стандартным отклонением  $\sigma = 1$ .

## Значение нормальности распределения:

Важное значение нормального распределения во многих областях науки (например, в математической статистике и статистической физике) вытекает из центральной предельной теоремы теории вероятностей. Если результат наблюдения является суммой многих случайных слабо взаимозависимых величин, каждая из которых вносит малый вклад относительно общей суммы, то при увеличении числа слагаемых распределение центрированного и нормированного результата стремится к нормальному. Этот закон теории вероятностей имеет следствием широкое распространение нормального распределения, что и стало одной из причин его наименования.



# Параметрические и непараметрические методы.

Непараметрические методы как раз и разработаны для тех ситуаций, достаточно часто возникающих на практике, когда исследователь ничего не знает о параметрах исследуемой популяции (отсюда и название методов - непараметрические).

Говоря более специальным языком, непараметрические методы не основываются на оценке параметров (таких как среднее или стандартное отклонение) при описании выборочного распределения интересующей величины. Поэтому эти методы иногда также называются свободными от параметров или свободно распределенными.

По существу, для каждого параметрического критерия имеется, по крайней мере, один непараметрический аналог.

Эти критерии можно отнести к одной из следующих групп:

- \* критерии различия между группами (независимые выборки);
- \* критерии различия между группами (зависимые выборки);
- \* критерии зависимости между переменными.

Непараметрические методы наиболее приемлемы, когда объем выборок мал. Если данных много (например,  $n > 100$ ), то не имеет смысла использовать непараметрические статистики.

**Различия между независимыми группами.** Обычно, когда имеются две выборки (например, мужчины и женщины), которые вы хотите сравнить относительно среднего значения некоторой изучаемой переменной, вы используете t-критерий для независимых выборок.

Непараметрическими альтернативами этому критерию являются: критерий серий Вальда-Вольфовица, U критерий Манна-Уитни и двухвыборочный критерий Колмогорова-Смирнова. Если вы имеете несколько групп, то можете использовать дисперсионный анализ .

**Различия между зависимыми группами.** Если вы хотите сравнить две переменные, относящиеся к одной и той же выборке (например, математические успехи студентов в начале и в конце семестра), то обычно используется t-критерий для зависимых выборок. Альтернативными непараметрическими тестами являются: критерий знаковый критерий Вилкоксона парных сравнений. Если рассматриваемые переменные по природе своей категориальны или являются категоризованными (т.е. представлены в виде частот попавших в определенные категории), то подходящим будет критерий хи-квадрат Макнемара. Если рассматривается более двух переменных, относящихся к одной и той же выборке, то обычно используется дисперсионный анализ (ANOVA) с повторными измерениями. Альтернативным непараметрическим методом является ранговый дисперсионный анализ



**Зависимости между переменными.** Для того, чтобы оценить зависимость (связь) между двумя переменными, обычно вычисляют коэффициент корреляции. Непараметрическими аналогами стандартного коэффициента корреляции Пирсона являются статистики Спирмена  $R$ , тау Кендалла и коэффициент Гамма (см. Непараметрические корреляции). Если две рассматриваемые переменные по природе своей категориальны, подходящими непараметрическими критериями для тестирования зависимости будут: Хи-квадрат, Фи коэффициент, точный критерий Фишера. Дополнительно доступен критерий зависимости между несколькими переменными так называемый коэффициент конкордации Кендалла. Этот тест часто используется для оценки согласованности мнений независимых экспертов (судей), в частности, баллов, выставленных одному и тому же субъекту.

