

## Основные понятия модуля

**Проверка гипотез** — это формальная процедура исследования наших представлений о мире с использованием статистики.

**Гипотеза** формулирует прогнозы о том, что будет обнаружено в результате исследования. Она должна основываться на существующих теориях и знаниях, а также поддаваться проверке с помощью методов научных исследований (таких как эксперименты, наблюдения и статистический анализ данных).

При проверке статистических гипотез **нулевая гипотеза** предсказывает отсутствие связи между переменными, в то время как **альтернативная гипотеза** утверждает, что исследовательский прогноз предполагает значимую связь.

**P-value** — это число, описывающее вероятность того, что данные были получены при нулевой гипотезе статистического теста.

**Статистическая значимость** используется для доказательства малой вероятности того, что наблюдения могли произойти при нулевой гипотезе статистического теста.

Статистическая значимость произвольна — она зависит от **порога  $\alpha$** , выбранного исследователем. Если *p-value* падает ниже выбранного  $\alpha$ , мы говорим, что результат теста статистически значим.

## Ошибки проверки гипотез

- **Ошибка I рода.** Отклонение нулевой гипотезы, когда она на самом деле верна.
- **Ошибка II рода.** Нулевую гипотезу не отвергают, когда она на самом деле ложна.

## Типы данных

Тип	Описание
Количественные	Любые переменные, в которых данные представляют <b>цифровые</b> значения (например, рост, вес или возраст).
Категориальные	Любые переменные, в которых данные представляют группы, включая <b>рейтинги</b> (например, финишные места в гонке), <b>классификации</b> (например, марки хлопьев) и <b>бинарные данные</b> (например, подбрасывание монеты).

## Типы статистических тестов

Тип	Требования
Параметрические	<ul style="list-style-type: none"> <li>→ Данные распределены нормально</li> <li>→ Зависимая переменная является количественной</li> </ul>
Непараметрические	<ul style="list-style-type: none"> <li>→ Данные необязательно распределены нормально</li> <li>→ Зависимая переменная может быть категориальной</li> </ul>

## Тесты на сравнение выборок

Параметрический	Непараметрический аналог
Парный Т-тест	Знаковый ранговый критерий Уилкоксона
Независимый Т-тест	Сумма рангов Уилкоксона
ANOVA	Крускал-Уоллис
MANOVA	ANOSIM

## Популярные статистические тесты в Python (модуль stats библиотеки SciPy)

	Тест	Значение
Тесты на нормальность	shapiro()	Шапиро-Уилка
	normaltest()	Д'Агостино
Тесты на сравнение выборок	ttest_ind()	Независимый T-тест
	f_oneway()	ANOVA