

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 α** , выбранного исследователем. Если *p-value* падает ниже выбранного α , 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 цифровые значения (например, рост, вес или возраст).
Категориальные	Любые переменные, в которых данные представляют группы, включая рейтинги (например, финишные места в гонке), классификации (например, марки хлопьев) и бинарные данные (например, подбрасывание монеты).

Типы статистических тестов

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

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	<code>shapiro()</code>	Шапиро-Уилка
	<code>normaltest()</code>	Д'Агостино
Тесты на сравнение выборок	<code>ttest_ind()</code>	Независимый Т-тест
	<code>f_oneway()</code>	ANOVA