

# Проверка гипотез

## Проверка гипотезы о равенстве среднего генеральной совокупности некоторому значению

```
In from scipy import stats as st

results = st.ttest_1samp(array, interested_value)
# array – выборка
# interested_value – предполагаемое среднее, на равенство которому мы делаем тест

print('p-значение: ', results.pvalue)
```

## Проверка гипотезы о равенстве среднего двух генеральных совокупностей по взятым из них выборкам

```
In from scipy import stats as st

sample_1 = [...] # sample_1 – выборка из первой генеральной совокупности
sample_2 = [...] # sample_2 – выборка из второй генеральной совокупности

results = st.ttest_ind(sample_1, sample_2, equal_var = True)
# equal_var – считать ли равными дисперсии выборок, по умолчанию имеет значение True

print('p-значение: ', results.pvalue)
```

## Проверка гипотезы о равенстве средних двух генеральных совокупностей для зависимых (парных) выборок

```
In from scipy import stats as st

results = st.ttest_rel(before, after)
# pair_1 – первая парная выборка
# pair_2 – вторая парная выборка

print('p-значение: ', results.pvalue)
```

# Словарь

## Генеральная совокупность

большой объём данных для статистических исследований

## Выборка

часть генеральной совокупности

## Репрезентативная выборка

часть данных, представляющая всю генеральную совокупность

## Случайная выборка

часть данных из генеральной совокупности, выбираемая случайным образом

## Страты

группы в генеральной совокупности, объединённые общим признаком

## Стратифицированная выборка

выборка, состоящая из пропорциональных выборок из разных страт

## Выборочное среднее

среднее значение, посчитанное на выборке

## Выборочная дисперсия

дисперсия, посчитанная на выборке

## Стандартная ошибка

стандартное отклонение выборочного среднего от настоящего среднего генеральной совокупности ( $S$  — оценка стандартного отклонения генеральной совокупности;  $n$  — размер выборки):

$$E.S.E. = \frac{S}{\sqrt{n}}$$

## Нулевая гипотеза $H_0$

гипотеза, которая проверяется на выборке

## Альтернативная гипотеза $H_1$

гипотеза, противоположная по смыслу нулевой гипотезе

## Уровень значимости

это суммарная вероятность того, что измеренное эмпирически значение окажется далеко от среднего

## Статистика разности

это количество стандартных отклонений между сравниваемыми значениями, если оба распределения привести к стандартному нормальному распределению со средним 0 и стандартным отклонением 1

## p-value

уровень значимости, на котором находятся друг от друга среднее значение и то, с которым производится сравнение

## Парная выборка

выборка, используемая для измерения некоторой переменной для одних и тех же единиц