



> Конспект > 3 урок > СТАТИСТИКА

> Оглавление

1. Построение доверительных интервалов и уровень значимости
2. Ошибки первого и второго рода

> Построение доверительных интервалов и уровень значимости

$$\mu \pm 1.96se$$

- интервал такой ширины, что при многократном повторении эксперимента в 95% из полученных интервалов будет среднее ГС

$$\mu \pm 2.58se$$

- более широкий интервал (99% что в нем лежит среднее ГС)

Помним, что

$$se = \frac{sd_x}{\sqrt{n}}$$

Пример использования:

В среднем пользователь пользуется нашим приложением 20 раз в день. Ваш коллега добавил несколько новых функций в приложение и сделал их доступными для части пользователей (N=64), а через какое-то время снова замерил среднее количество использования приложения. Теперь среднее равно 18.4 раз со стандартным отклонением 4.

Наша нулевая гипотеза H_0 - ничего не поменялось, результат случайный

Наша альтернативная гипотеза H_1 - изменение значимо, пользователи стали реже пользоваться приложением.

Если верна H_0 , то в соответствии с ЦПТ выборочные средние распределились бы нормально вокруг среднего ГС (пока что мы думаем, что оно 20) со стандартным отклонением (стандартной ошибкой среднего):

$$se = \frac{sd}{\sqrt{n}} = \frac{4}{\sqrt{64}} = 0.5$$

Теперь посчитаем, насколько наше выборочное среднее (18.4) отклонилось от среднего ГС (20) в единицах стандартного отклонения → сделаем Z-преобразование:

$$Z = \frac{\bar{x} - M}{se} = \frac{18.4 - 20}{0.5} = -3$$

Как считать доверительный интервал в python:

```
import numpy as np, scipy.stats as st
import statsmodels.stats.api as sms

a = range(100)

# первый способ
st.t.interval(0.95, len(a)-1, loc=np.mean(a), scale=st.sem(a))

# второй способ
sms.DescrStatsW(a).tconfint_mean()
```

Как это интерпретировать: если бы среднее в ГС было равно 20, то наше выборочное среднее находилось бы от него в 3 стандартных отклонениях (в левую сторону).

Теперь считаем вероятность такого (-3 или 3) или еще более сильного отклонения (>-3 или $3<$) с помощью [сайта](#). Получаем $p = 0.0027$ (p-уровень значимости).

Чем меньше p-значение, тем больше оснований отклонить нулевую гипотезу. Классические пороги для отклонения нулевой гипотезы равны 0.05 или 0.01 (более строгий вариант). Но не везде: есть области, где требуются более строгие отсечки. Например, генетика ассоциаций SNP (однонуклеотидная замена) с заболеваниями, где пороговое значение равно 10^{-7} . Поэтому важно понимать, в какой области мы проводим статистический тест, чтобы брать соответствующие пороговые p-значения.

Вывод: в нашем случае коллега статистически значимо ухудшил приложение и мы попросим его не распространять эту версию приложения на всех пользователей. То есть в нашем случае, если коллега все-таки не испортил приложение и такой результат (уменьшение использования приложения) является случайным, то вероятность этого события равна 0.27%.

Важно:

- p-уровень значимости не говорит о силе эффекта (напр., пользователи стали пользоваться приложением меньше в 3 или 1.1 раза)
- Если получили уровень значимости выше порога – это не значит, что нулевая гипотеза верна. Просто такой уровень значимости не позволяет нам её отклонить.

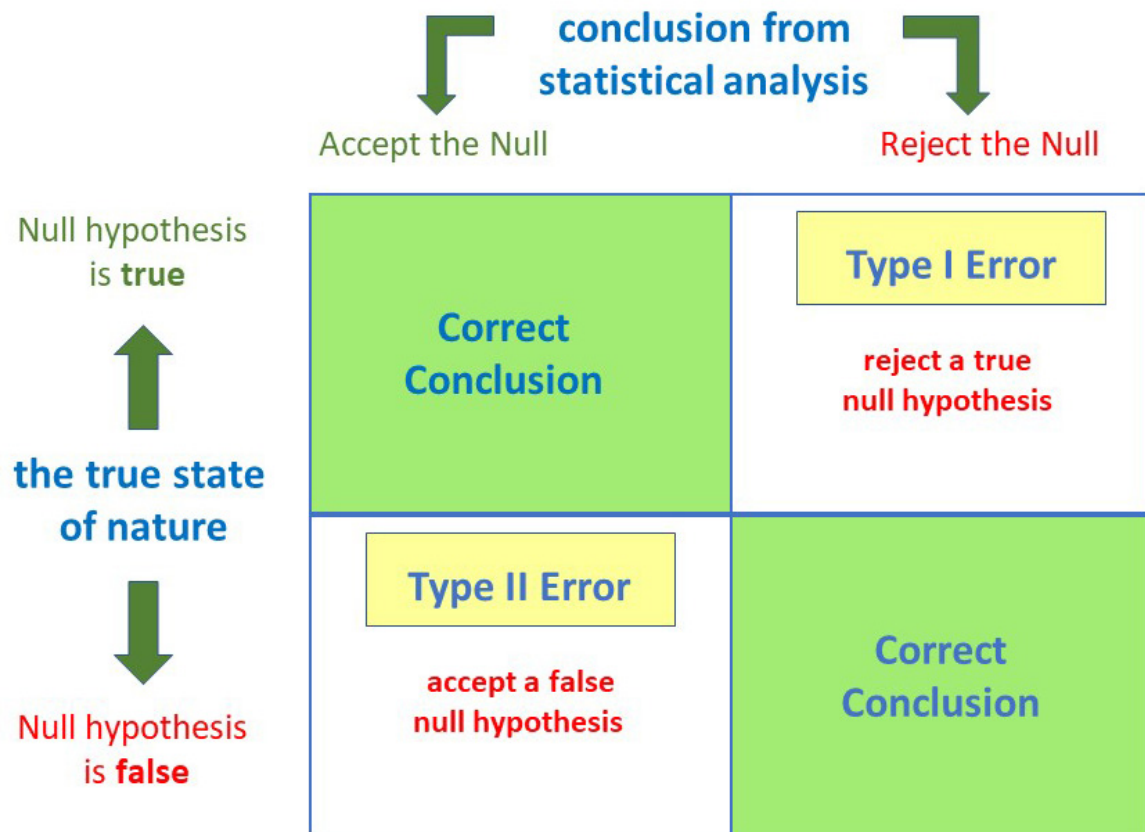
> Ошибки первого и второго рода

Ошибка 1 рода (False positive): отклонили нулевую гипотезу, хотя она была верна (сказали, что есть эффект, когда на деле его нет, false alarm)

Ошибка 2 рода (False negative): не отклонили нулевую гипотезу, хотя верна была альтернативная (говорим, что нет эффекта, когда на самом деле он есть)

Какую ошибку допустить хуже? Зависит от области, в которой вы работаете. Если вы диагностируете заболевание, то лучше скорее допустить ошибку 1 рода: вы скажете пациенту, что он болен, и отправите его на дополнительные

исследования, где подтвердится, что с ним всё в порядке. Хуже, если бы вы отправили домой больного пациента, тем самым пропустив болезнь (false negative – сказали, что пациент здоров, когда он болен).



Может помочь с запоминанием:

- False **P**ositive → "**п**" → первый → ошибка первого рода,
- False **nE**gative → **sE**cond → ошибка второго рода (либо через second, либо методом исключения)

И несколько картинок:

Type I Error



Type II Error



	Null hypothesis is TRUE	Null hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)