Основные понятия модуля

A/B-тестирование — это метод, который заключается в сравнении текущей версии продукта (*версии A*) с изменённой версией (*версией B*) на основании данных, полученных до введения обновления в продукт и после него. Метод основан на проверке статистической значимости результатов эксперимента и позволяет заранее задать границу уверенности в результатах исследования (уровень надёжности).

Контрольная версия — текущая версия продукта.
Тестовая версия — новая версия продукта.

Механизм A/B-тестирования

Чтобы протестировать какую-либо гипотезу при помощи A/B-теста, аудиторию разделяют на две части:

- **Группа A** продолжает использовать (видеть) старую версию продукта.
- **Группа B** видит новую версию.
- В реальном времени собирается информация об обеих группах теста (A и B).
- Проводятся замеры важных показателей.
- Проводится сравнение этих показателей.
- Принимается решение об эффективности влияния гипотезы на показатели продукта.

Принципы A/B-тестирования

- исключать влияние извне
- использовать большое количество данных
- применять правильные инструменты для анализа

Алгоритм A/B-тестирования

Конверсия — отношение числа посетителей сайта, выполнивших на нем какие-либо целевые действия, к общему числу посетителей сайта, выраженное в долях или процентах. Под целевым действием можно подразумевать покупку товара, лайк или репост поста в Instagram, просмотр фильма на Кинопоиске и многое другое.

Кумулятивные метрики

> **Кумулятивная метрика** — это отображение целевой метрики, когда вы отслеживаете её поведение за каждый день — накопленным итогом по дням.

Принимать какие-либо решения по результатам A/B-теста стоит только после того, как метрика стабилизируется!

Кумулятивная метрика считается **стабилизированной**, когда на графике прекращаются резкие пики и спады показателя, и линия постепенно выравнивается в прямую.

Пример вычисления кумулятивной метрики

```
daily_data_a.loc[:, 'cum_users_count'] =
daily_data_a['users_count'].cumsum()
```

Кумулятивную сумму можно записать в виде рекурсивной формулы:

$$S_t = S_{t-1} + x_t$$

x_t — значение показателя в день t

S_t — значение суммы в день t

Расчет кумулятивной конверсии в процентах

```
daily_data['cum_conversion'] =
daily_data['cum_converted']/daily_data['cum_users_count'] * 100
```

Статистические тесты

Для проверки гипотезы равенства пропорций мы можем воспользоваться уже знакомым нам **z-критерием** для пропорций. Этот критерий является наиболее популярным для задачи определения статистической значимости изменения конверсии.

Нулевая гипотеза нашего теста всегда будет звучать следующим образом:

$$H_0: p_a = p_b$$

А вот альтернативных может быть несколько:

$$H_1 \text{(двусторонняя)}: p_a \neq p_b$$
$$H_1 \text{(левосторонняя)}: p_a < p_b$$
$$H_1 \text{(правосторонняя)}: p_a > p_b$$

```
from statsmodels.stats.proportion import proportions_ztest
H0 = 'Конверсии в группах A и B равны'
H1 = 'Конверсия в группе A выше чем конверсия в группе B'
alpha = 0.05
_, p_value = proportions_ztest(
    count=converted_piv['sum'],
    nobs=converted_piv['count'],
    alternative='larger',
)
print('p-value: ', round(p_value, 2))
```

Помимо конверсии нужно рассматривать и другие метрики, например **средний чек**.

Любой статистический метод имеет свою область применения, которая зависит от задачи и распределения данных. Например, время проведенное на сайте, часто распределено нормально, и тогда мы можем использовать T-тест для средних.

```
from scipy.stats import ttest_ind
H0 = 'Среднее время в группах одинаково'
H1 = 'Среднее время в группе A меньше, чем в группе B'
alpha = 0.05
results = ttest_ind(
    a=time_data['time(A)'],
    b=time_data['time(B)'],
    alternative='less'
)
print('p-value:', round(results.pvalue, 2))
```

Денежные метрики, такие как средний чек, часто (но не всегда) напоминают **логнормальное распределение**. Для их исследования используются **непараметрические тесты**: критерий Манна-Уитни, ANOVA-тест и другие.

Доверительные интервалы

Интервальные оценки — один из способов оценки параметров генеральной совокупности, при использовании которого ответ даётся не в виде одного числа, а в виде интервала.

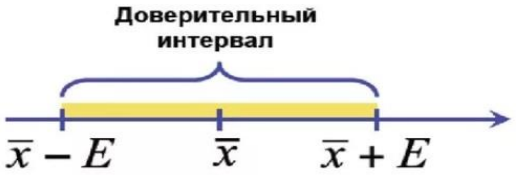
Доверительный интервал — интервал, который с заданной надёжностью покрывает значение неизвестного параметра.

Виды доверительных интервалов:

- ➔ двусторонние
- ➔ левосторонние
- ➔ правосторонние

Любой **двусторонний** доверительный интервал обладает следующей структурой:

Параметр = Выборочная оценка ± Предел погрешности



Доверительный интервал для **истинного среднего при известном стандартном отклонении**:

$$\mu = X_{mean} \pm z_{\text{крит}} \times \frac{\sigma}{\sqrt{n}}$$

X_{mean} – выборочное среднее
σ – истинное стандартное отклонение
n – выборочное среднее
$z_{\text{крит}} = z_{(1-\gamma)/2} = z_{\alpha/2}$ – значение, которое отсекает критическую область нормального распределения при надежности, равной γ

Под уровнем надежности γ понимается вероятность того, что истинное значение параметра окажется в построенном интервале. А **под уровнем значимости** α — вероятность того, что построенный доверительный интервал «промахнется» и не захватит истинное значение параметра.

```
from scipy.stats import norm
def z_mean_conf_interval(n, x_mean, x_std, gamma=0.95):
    alpha = 1 - gamma
    z_crit = -norm.ppf(alpha/2)
    eps = z_crit * x_std / n**0.5
    lower_bound = x_mean - eps
    upper_bound = x_mean + eps
    return lower_bound, upper_bound
```

Доверительный интервал для **истинного среднего при неизвестном стандартном отклонении**:

$$\mu = X_{mean} \pm t_{\text{крит}} \times \frac{X_{std}}{\sqrt{n}}$$

X_{std} — выборочное стандартное отклонение

$t_{\text{крит}}(k) = t_{(1-\gamma)/2}(k) = t_{\alpha/2}(k)$ — значение, которое отсекает критическую область распределения Стьюдента при надежности, равной γ

k — число степеней свободы

```
from scipy.stats import t
def t_mean_conf_interval(n, x_mean, x_std, gamma=0.95):
    k = n - 1
    alpha = 1 - gamma
    t_crit = -t.ppf(alpha/2, k)
    eps = t_crit * x_std / n**0.5
    lower_bound = x_mean - eps
    upper_bound = x_mean + eps
    return lower_bound, upper_bound
```

Доверительный интервал для **истинной пропорции**:

$$p = \mu = X_p \pm z_{\text{крит}} \times \sqrt{\frac{X_p(1 - X_p)}{n}}$$

X_p — выборочная пропорция

```
from scipy.stats import norm
def proportions_conf_interval(n, x_p, gamma=0.95):
    alpha = 1 - gamma
    z_crit = -norm.ppf(alpha/2)
    eps = z_crit * (x_p * (1 - x_p) / n) ** 0.5
    lower_bound = x_p - eps
    upper_bound = x_p + eps
    return lower_bound, upper_bound
```

Доверительный интервал **разницы пропорций:**

$$\Delta p = \Delta X_p \pm z_{\text{крит}} \times \sqrt{\frac{X_{p_a}(1 - X_{p_a})}{n_a} + \frac{X_{p_b}(1 - X_{p_b})}{n_b}}$$

Индексы a и b обозначают принадлежность параметра группе A или B соответственно.

$\Delta p = p_b - p_a$ — истинная разница пропорций групп B и A,

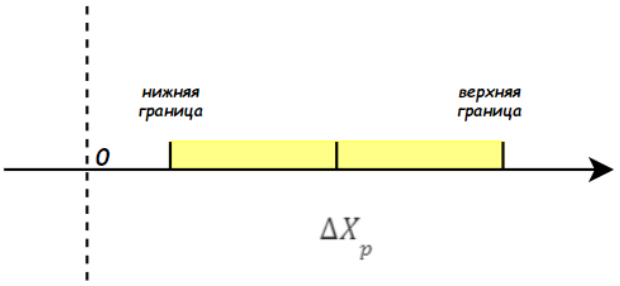
$\Delta X_p = X_{p_b} - X_{p_a}$ — выборочная разница пропорций групп B и A

```
from scipy.stats import norm
def diff_proportions_conf_interval(n, xp, gamma=0.95):
    alpha = 1 - gamma
    diff = xp[1] - xp[0]
    z_crit = -norm.ppf(alpha/2)
    eps = z_crit * (xp[0] * (1 - xp[0])/n[0] + xp[1] * (1 - xp[1])/n[1]) ** 0.5
    lower_bound = diff - eps
    upper_bound = diff + eps
    return lower_bound, upper_bound
```

Три случая доверительного интервала для разницы пропорций

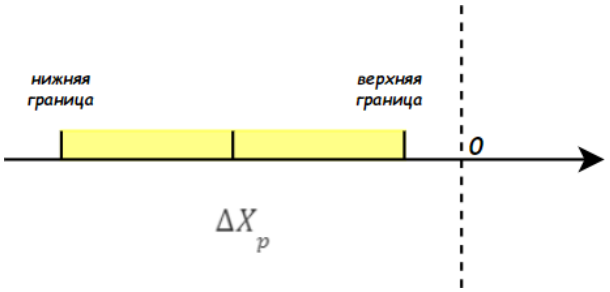
Обе границы доверительного интервала являются **положительными** (больше 0). То есть истинная разница в пропорциях $\Delta p = p_b - p_a$ положительная.

Пропорция A < пропорции B.



Обе границы доверительного интервала являются **отрицательными** (меньше 0). То есть истинная разница в пропорциях $\Delta p = p_b - p_a$ отрицательна.

Пропорция A > пропорции B.



Интервал охватывает точку 0. Левая граница доверительного интервала отрицательная, а правая — положительная. То есть истинная разница в пропорциях $\Delta p = p_b - p_a$ может быть как положительной, так и отрицательной.

Тогда это будет значить, что пропорция A равна пропорции B.