

**ТЕХНОАТОМ: Продуктовая аналитика.  
А/В-тестирование**

# Лекция №1:

Основные понятия и физический смысл.

Математическое обоснование и значения для  
бизнеса



---

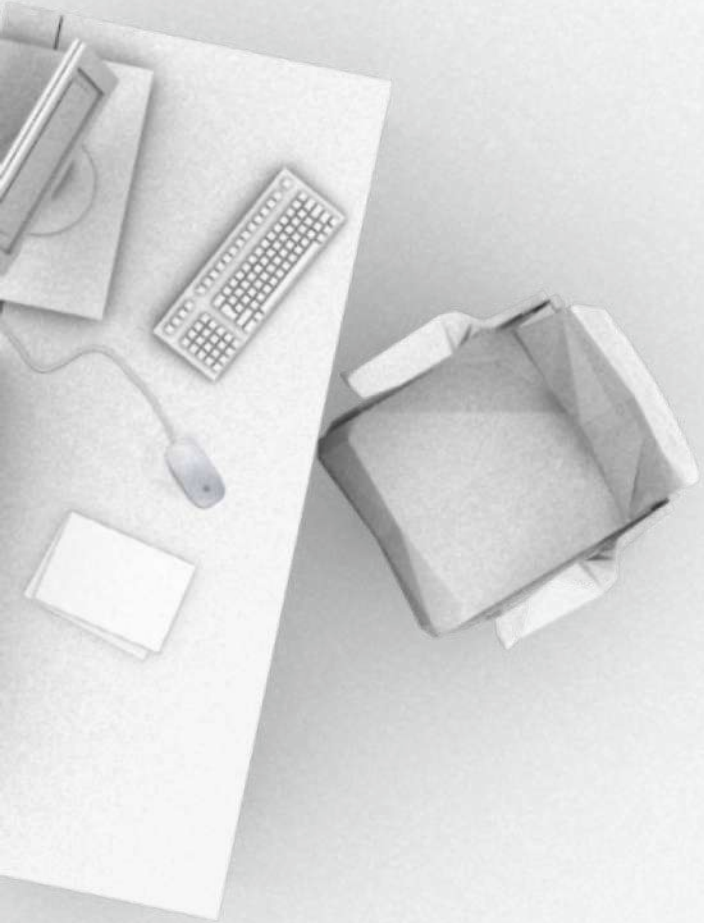
# Познакомимся?

- Ты слышал о сплит-тестировании? Как соотносятся A/B-тесты и сплит-тесты?
- Что означает фраза “рассчитать A/B-тест”? Какие методы используются для таких расчётов?
- Что ты знаешь об основных понятиях и методах математической статистики? Мини-тест: <https://forms.gle/QBcUrQp9ucs2AnzU9>
- Любишь математику и python? 😊



# План лекции

- 1. А/В-тестирование: общая задача.**  
Значение для бизнеса.  
Дизайн эксперимента.  
Контрольная и тестовые группы.  
Запуск и проверка хода эксперимента.  
Типы метрик.
- 2. Виды испытаний.** Зависимые и независимые испытания. Повторные испытания. Парные тесты. А/А-тесты. Обратные тесты.
- 3. Основы математической статистики.**  
Виды распределений. Доверительные интервалы. Ошибки первого и второго рода.
- 4. Проверка статистической значимости:**  $p$ -value,  $\alpha$ -уровень статистической значимости и  $(1-\beta)$  или мощность критерия.
- 5. Виды статистических критериев.**  
Параметрические и непараметрические критерии.



# **1. А/В-тестирование:** общая задача.

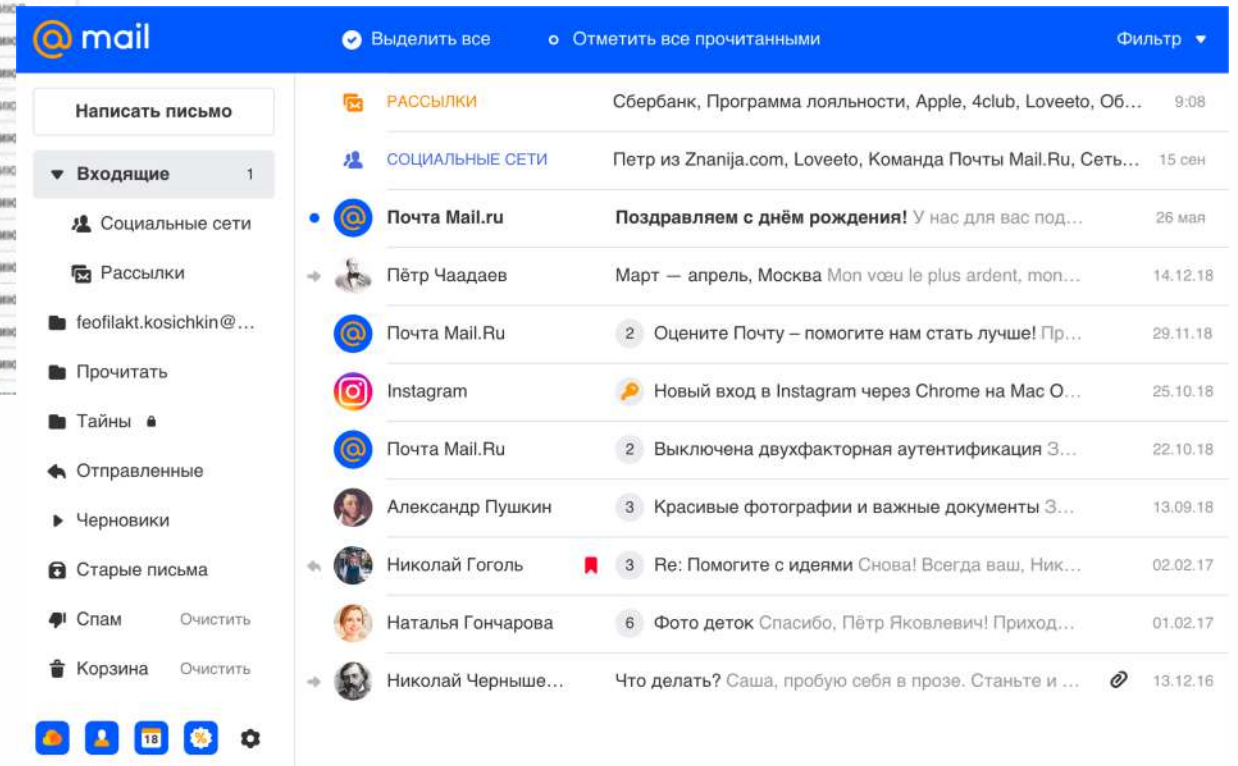
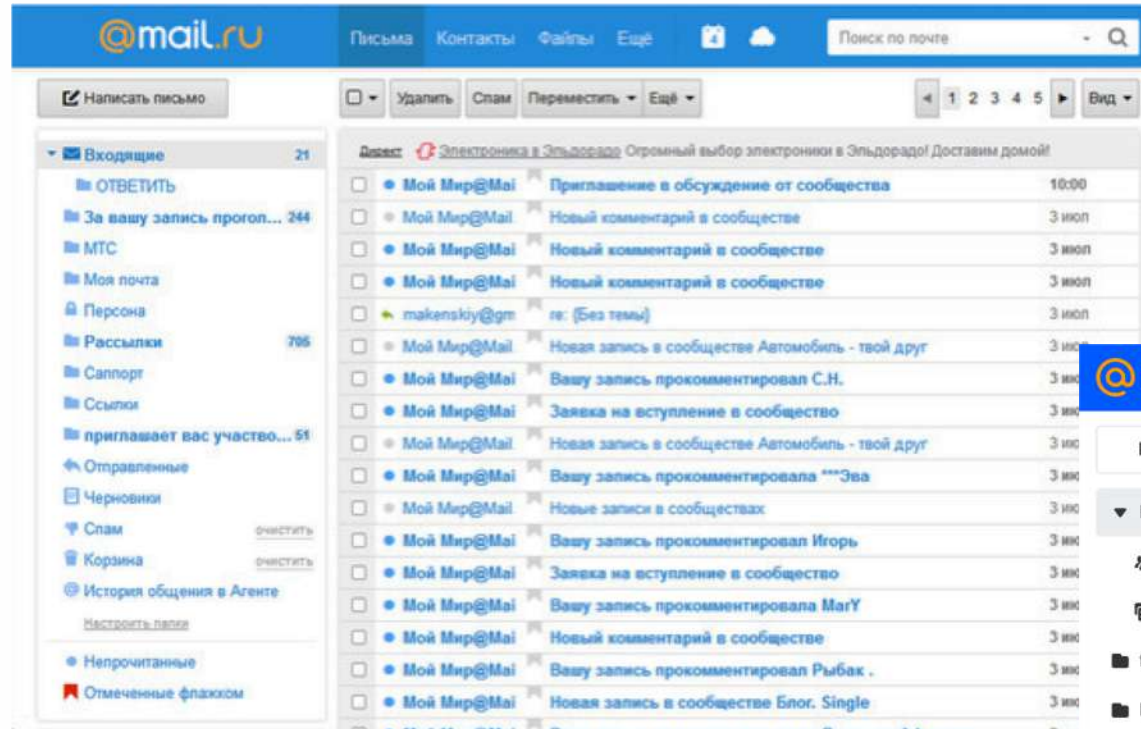
# Общая задача . Значение для бизнеса



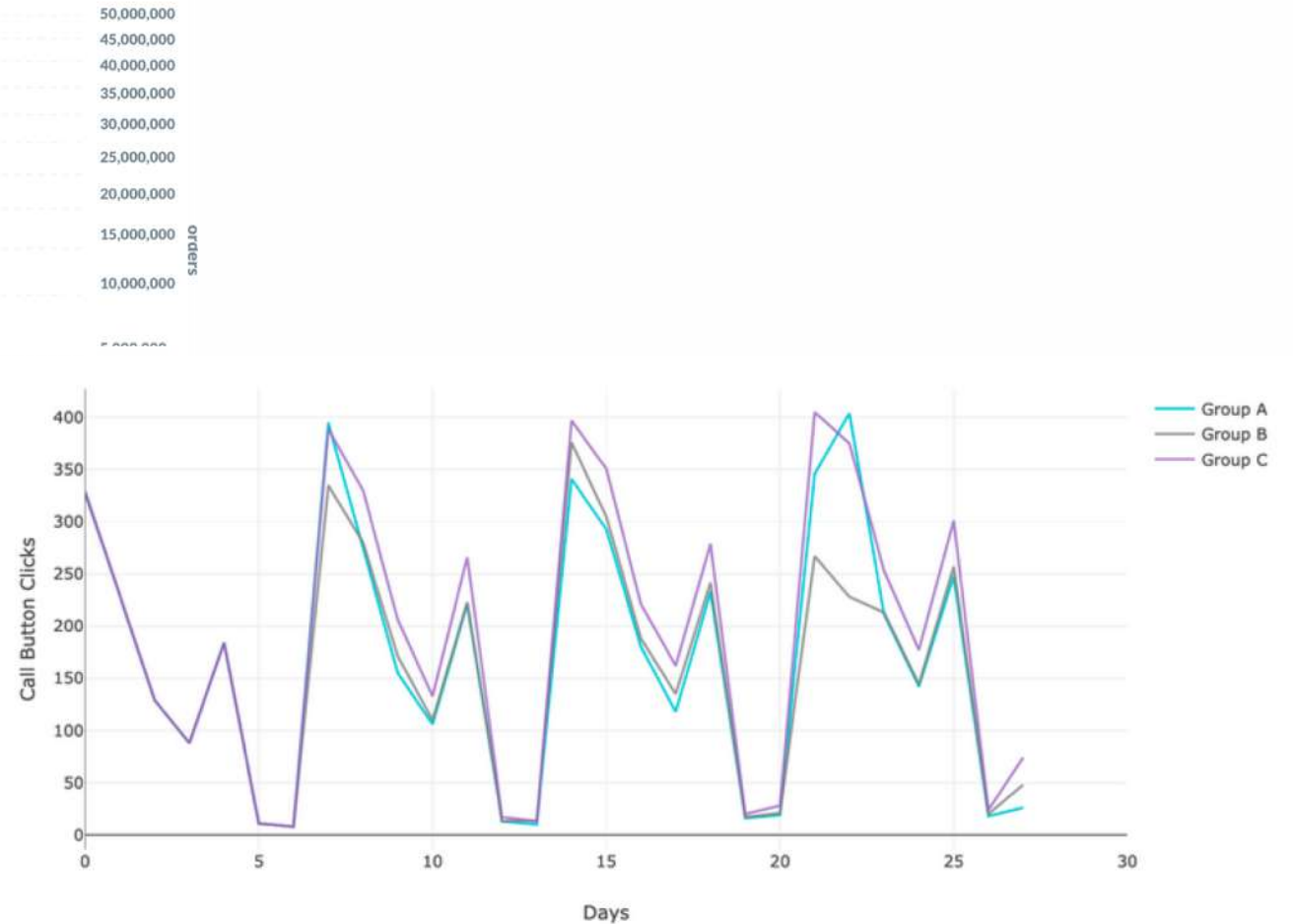
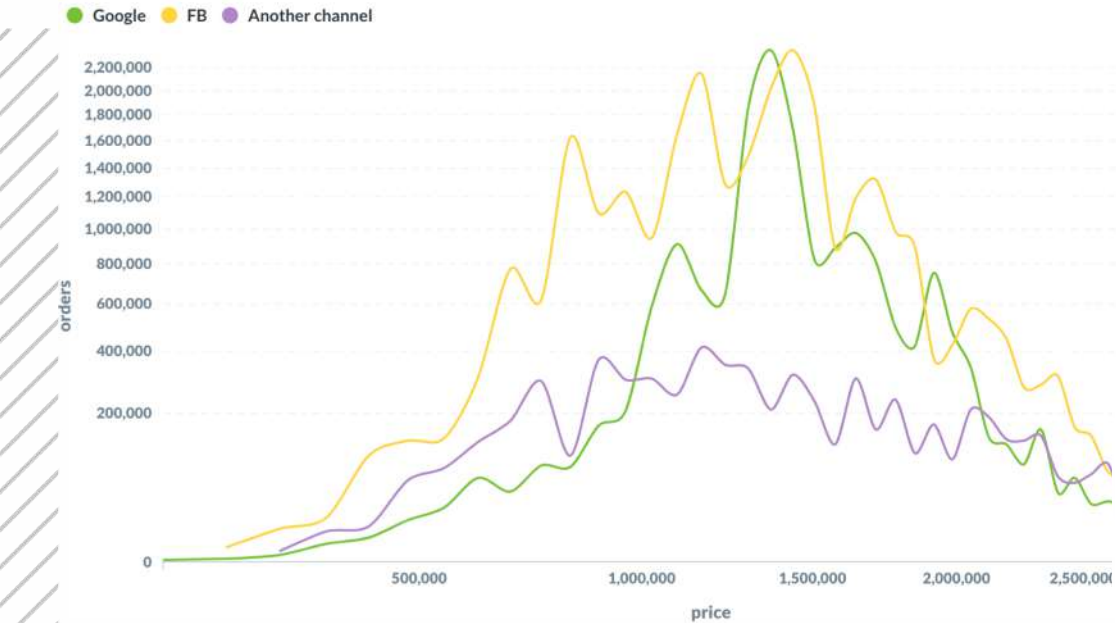
1. **Цель:** наукоёмкое тестирование, оценка запланированных или форс-мажорных изменений (оценка не "на глазок"); понять что изменилось и будет ли полезно нововведение; на какие метрики повлияет или уже повлияло, стоит ли его "катить" или "откатывать"; возможность протестировать изменение на небольшой аудитории и не просадить метрики;
2. **Метод:**
  1. Разбить самостоятельно или рассмотреть уже разбитые кем-то (природой или волею разработчиков 😊 ) группы: тестовую (что-то изменилось) и контрольную (всё по-прежнему);
  2. Проверить, что выборки в группах не смещены относительно генеральной совокупности и друг друга и достаточны по объёму для распространения выводов на всю генеральную совокупность;
  3. Продумать метрики, которые отражают суть влияния изменений и отвечают на вопросы менеджмента, рассчитать их;
  4. Применить методы расчёта стат. значимости различий для метрик в выборках.



# Общая задача . Значение для бизнеса



# Общая задача . Значение для бизнеса



---

## Значение для бизнеса: основные заказчики

- Маркетинг;
- Продукт;
- Монетизация;
- Операционный, финансовый и коммерческий отделы;
- Эксперименты с ML-моделями: ранжирование и соответствие запросу результатов в выдаче, рекомендации похожих, качество распознавания текста, голоса, изображений;
- Whatever else...




# Дизайн эксперимента: подготовка

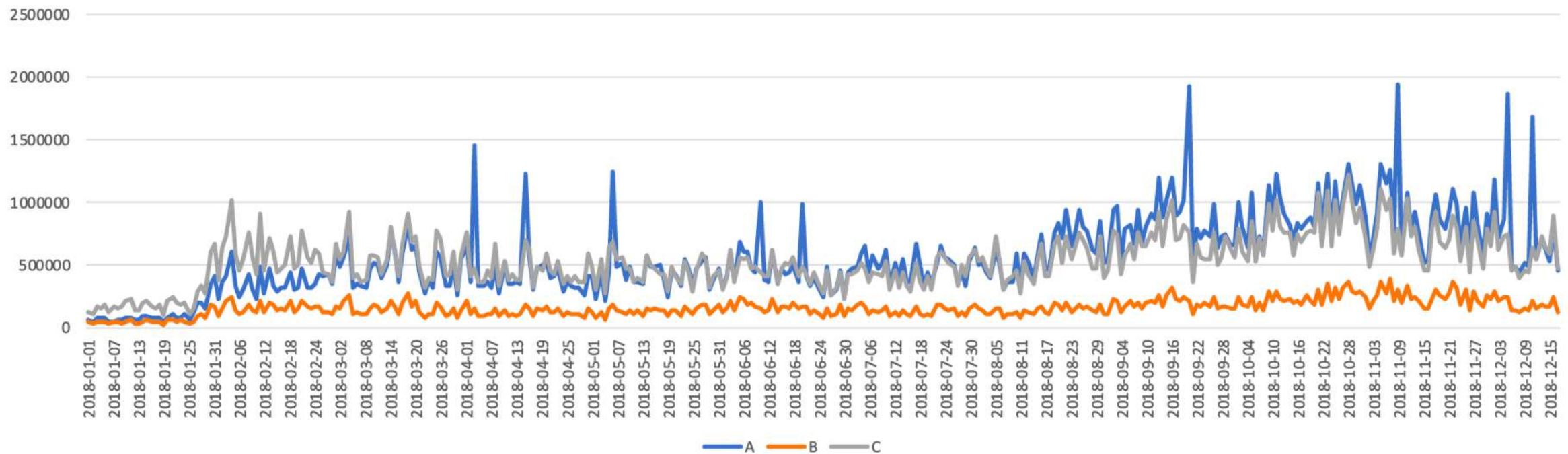
1. Этап пред-подготовки и предварительного анализа:  
Выявление бизнес-целей проводимого эксперимента;  
Формулировка цели. Объём аудитории.
2. Формулировка гипотез.  
Например, гипотеза: замена блока с иконкой сервиса «Журнал» на сниппеты (блоки с фото и кратким содержимым) статей принесёт увеличение переходов в раздел «Журнала»;
3. Оценка финансовых рисков и стоимость проведения эксперимента — преобладает ли возможный профит над фактическими затратами;
4. Исследование аудитории.
5. Пробуем оценить стоимость каждой гипотезы.



# Дизайн эксперимента

- 
1. Определение объема пользователей в эксперименте. Разбивка пользователей на контрольную и тестовую группу, оценка размеров групп и продолжительности проведения теста;
  2. Выбор метрик исходя из бизнес-задачи;
  3. Запуск и проверка хода эксперимента;
  4. Техническая сторона вопроса: формирование предварительного ТЗ разработчикам, принципы разбиения пользователей на bucket'ы, составление схемы логирования.

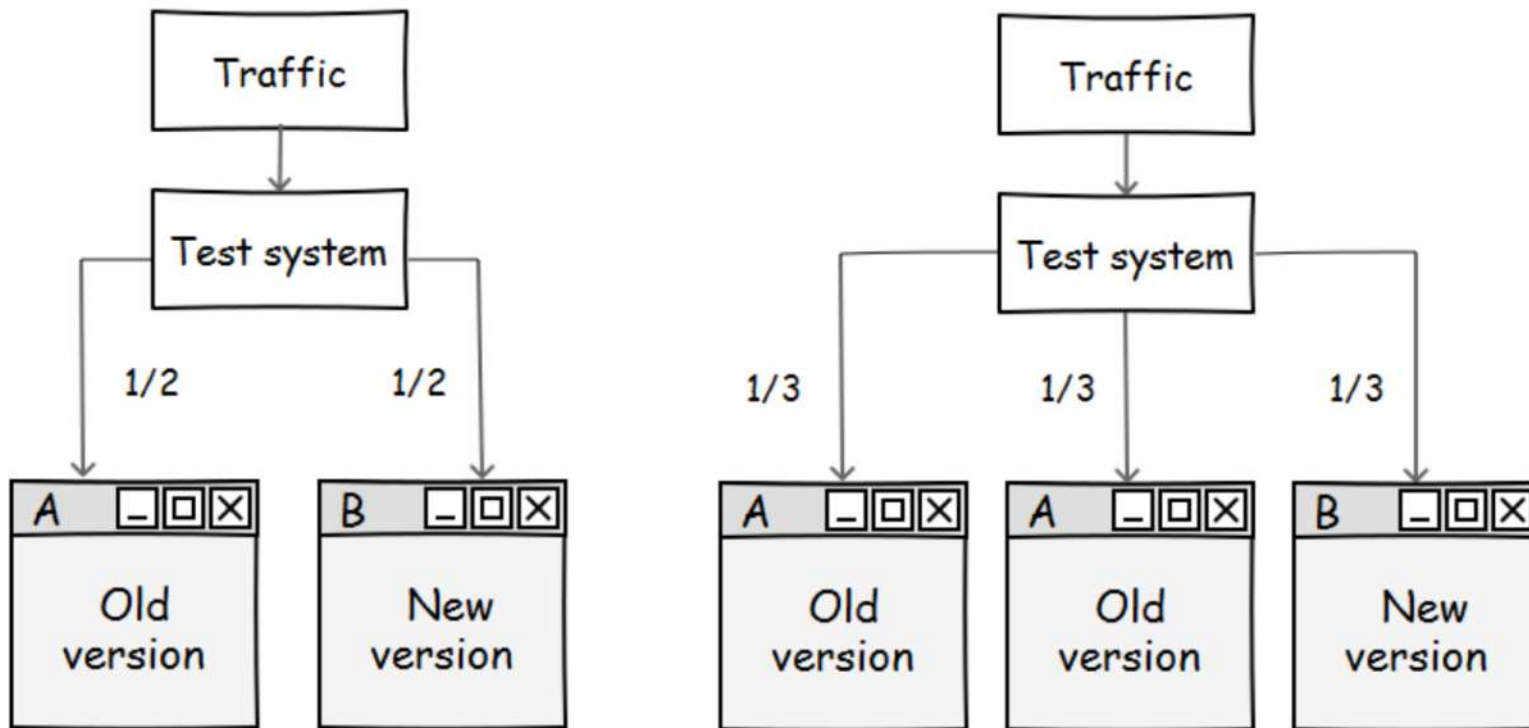
# Запуск и проверка хода эксперимента



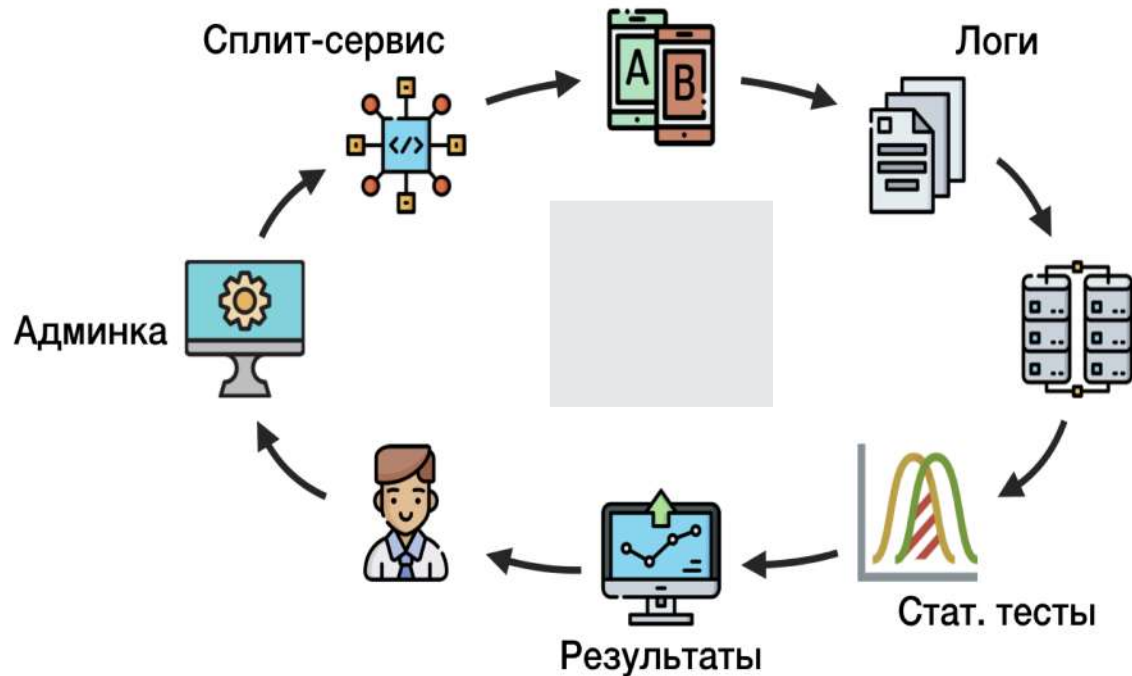


## 2. Виды испытаний.

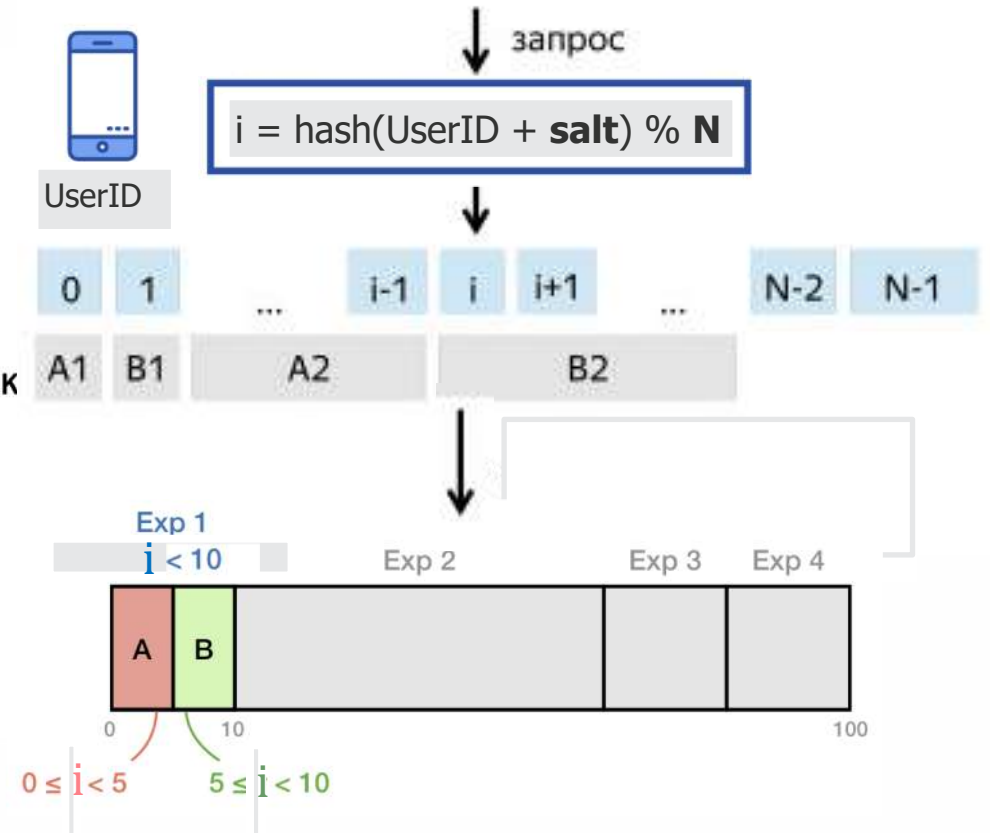
# Парные тесты. Контрольная и тестовая группы. A/A-тесты. Обратные тесты



## Тех. процесс:



## Одномерная схема разбиения:

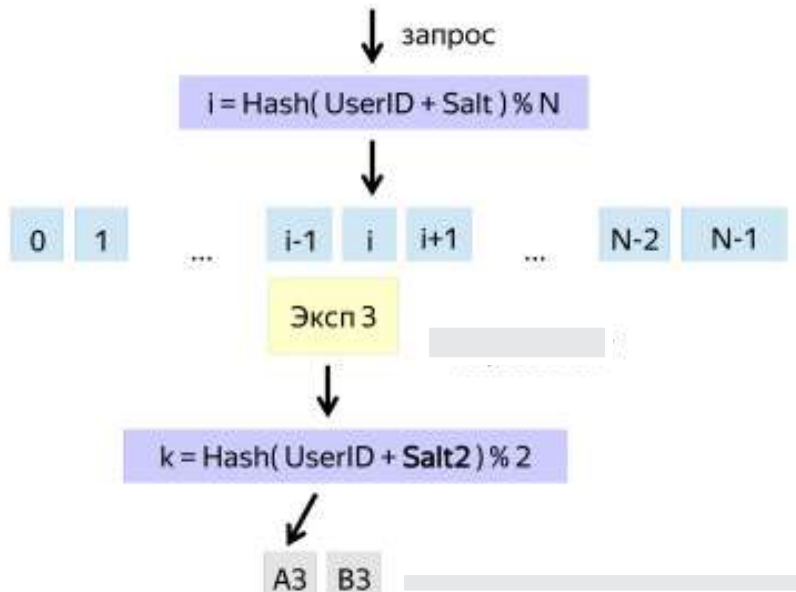


Примеры split-систем:

- Google FireBase, Google Optimize, Optimizely, VWO, внутренние разбивки AdJust;
- Собственные splitter'ы VK, YaSERP, Avito, OK, Mail.Ru;
- В nginx есть стандартный модуль для A/B тестирования:  
[nginx.org/ru/docs/http/nginx\\_http\\_split\\_clients\\_module.html](http://nginx.org/ru/docs/http/nginx_http_split_clients_module.html)



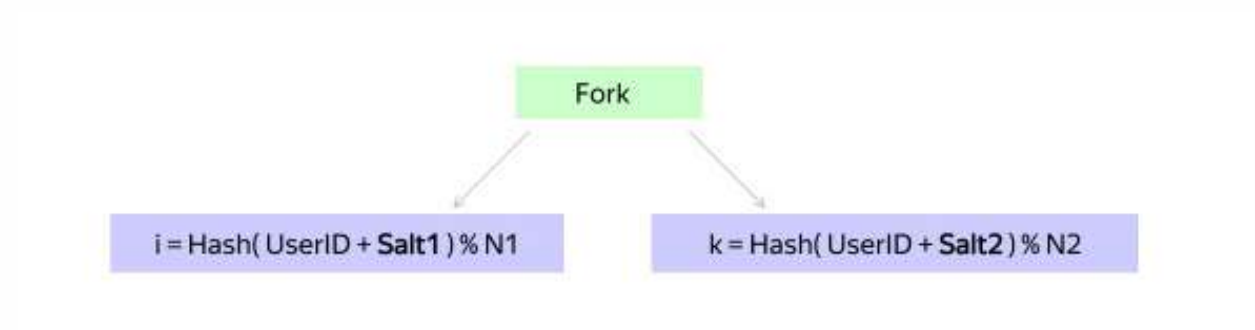
## Память аудитории:



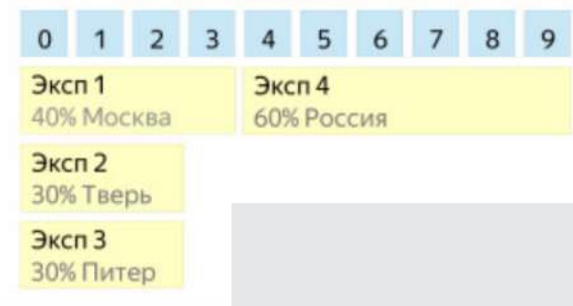
### Обратные тесты:

изменение – на всех, предыдущую версию  
– на часть

## Многомерная схема разбиения:



## Одномерная cross-схема:



# Типы метрик

- Агрегированные метрики в рамках групп:
  - DAU, WAU, goal visits, app installs, app uninstalls,
  - Оборот, выручка (GMV), прибыль (revenue), costs,
- На пользователя (user-based):
  - CAC, CPU, LTV, goals per users, orders per user
- На сессию (visit-based):
  - Goals per session, session time, page depth per session, churn rate, отказы (доля сессий < 30 секунд),
- На целевое действие:
  - CPO, CPC, CPA, CPM, CPI – по сути, ratio-метрики из 2-х агрегированных метрик:
$$\frac{\text{sum(costs)}}{\text{sum(orders||clicks||actions||1000 clicks||installs)}}$$

## Повторные испытания. Зависимые и независимые испытания

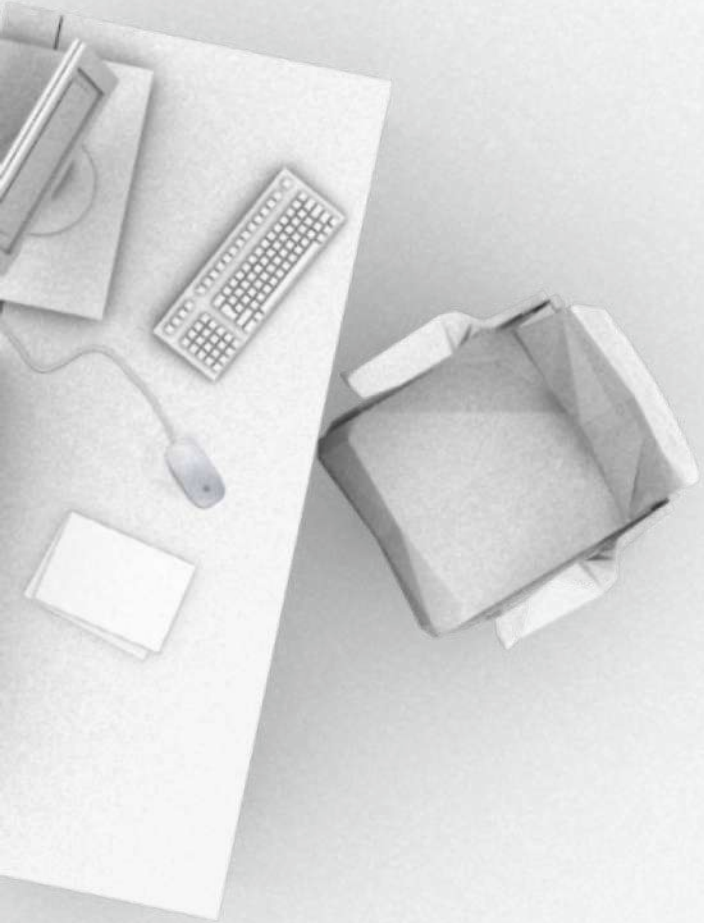
- **Зависимые (связанные) испытания** – череда событий, вероятность исхода каждого последующего из которых зависит от исхода предыдущих, или события, происходящие в связанных выборках. **Связанные (зависимые, парные) выборки** – выборки, в которых каждое наблюдение одной выборки неразрывно связано (находится в паре) с одним из наблюдений другой выборки.

*Примеры:* выборка без возвращения, попадание пользователя в обе экспериментальные группы – пользователь видит 2 разных типа выдачи в рамках одного эксперимента, зависимые выборки.

- **Независимые (независимые) испытания** – череда независимых событий, или события в выборках, в которых объекты исследования набирались независимо друг от друга.

*Формула Бернулли:* вероятность появления события определённое количество раз при любом числе независимых испытаний.

$$P_n^k = C_n^k p^k q^{n-k} = C_n^k p^k (1-p)^{n-k}, \text{ где } C_n^k = \frac{n!}{k!(n-k)!}$$



### 3. Основы математической статистики.

# Основы математической статистики

- Выборочное среднее:
- Среднеквадратичное отклонение:
  - На основании смещённой оценки дисперсии:
  - На основании несмещённой оценки:
- Математическое ожидание:
- Дисперсия генеральной совокупности:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

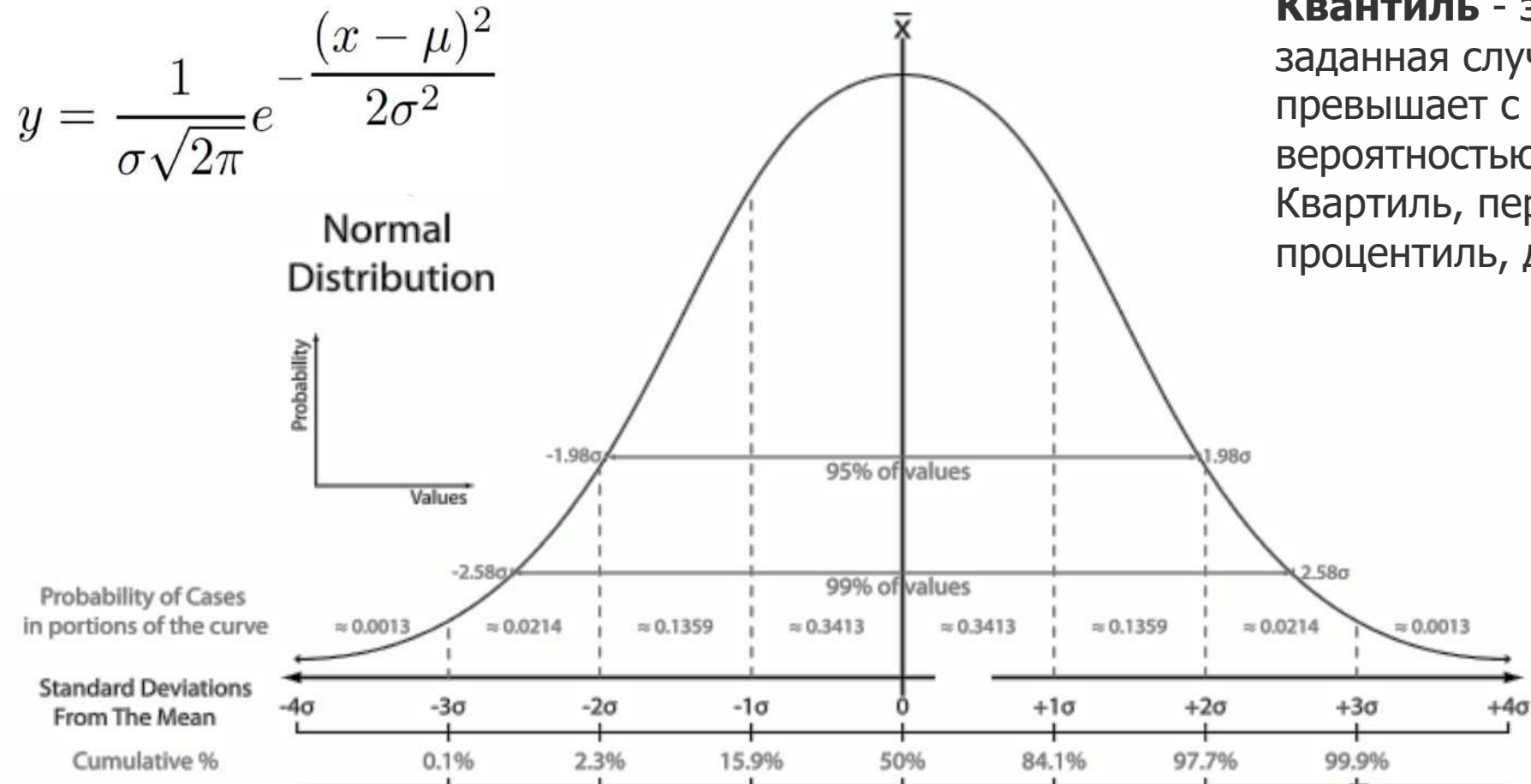
$$S_0 = \sqrt{\frac{n}{n-1} S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$M[X] = \sum_{i=1}^{\infty} x_i p_i$$

$$\sigma_X^2 = \sum_{i=1}^n p_i (x_i - M[X])^2$$

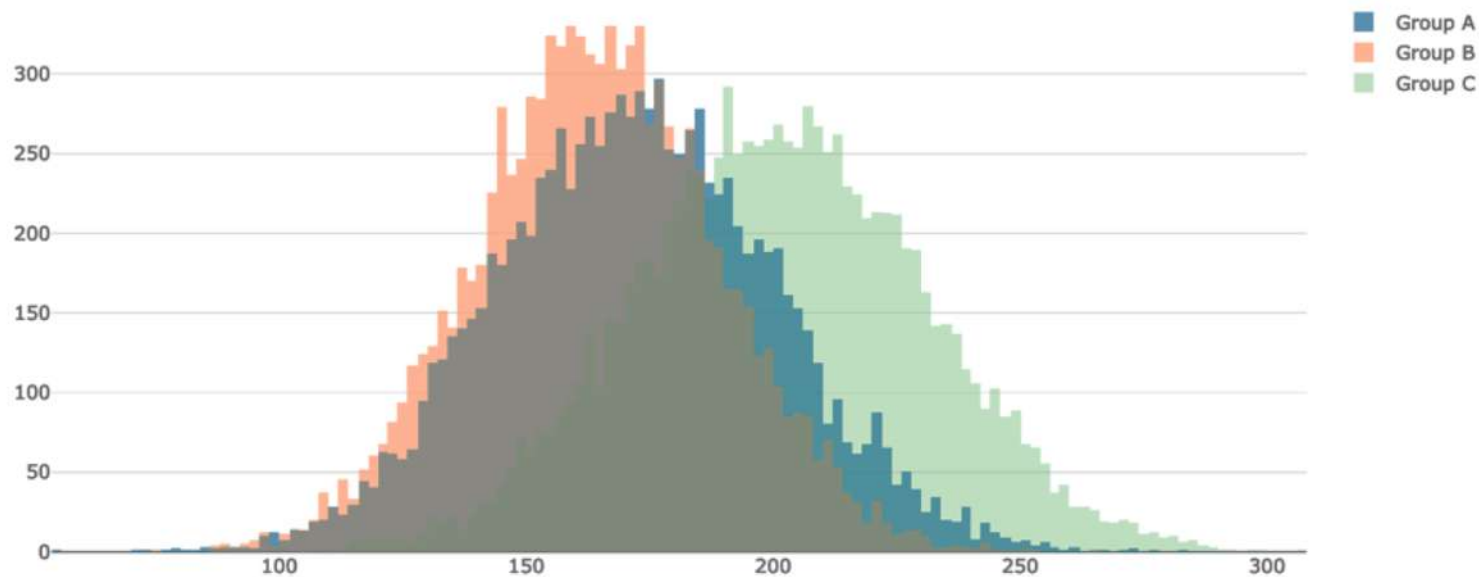
# Виды распределений. Нормальное распределение

**Квантиль** - значение, которое заданная случайная величина не превышает с фиксированной вероятностью.  
Квартиль, перцентиль или процентиль, дециль.



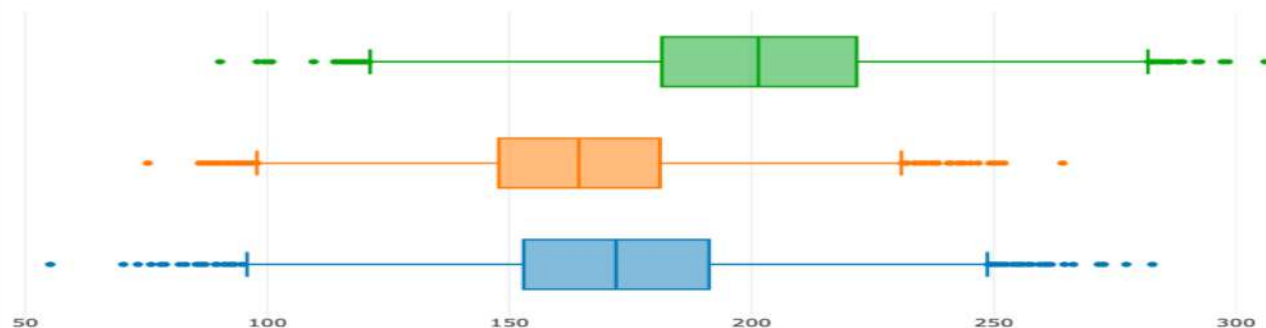


# Доверительные интервалы

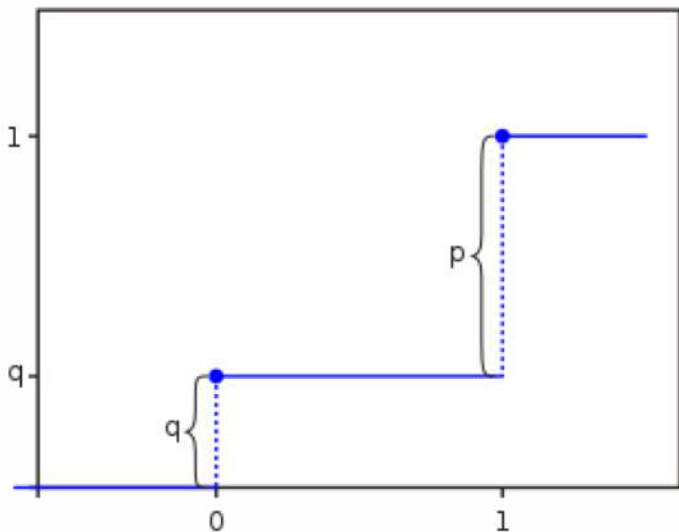


**Доверительный интервал** - интервал, в который попадают измеренные в эксперименте значения, соответствующие определенной установленной вероятности, называемой доверительной вероятностью.

Иными словами, интервальная оценка вероятности встретить значение  $x_i$  случайной величины в диапазоне от  $x_{lower\_bound}$  до  $x_{upper\_bound}$ .



# Виды распределений. Распределение Бернулли и биномиальное распределение



На практике множество продуктовых и маркетинговых метрик имеет распределение подобного вида:

- Целевое действие, просмотр телефона,
- оформление заказа, оплата покупки,
- ушёл без заказа, бросил корзину,
- установка приложения, регистрация, sign-in в рамках сессии, периода, в рамках ЖЦ пользователя;
- продолжил движение по воронке/ушёл со страницы;
- etc.

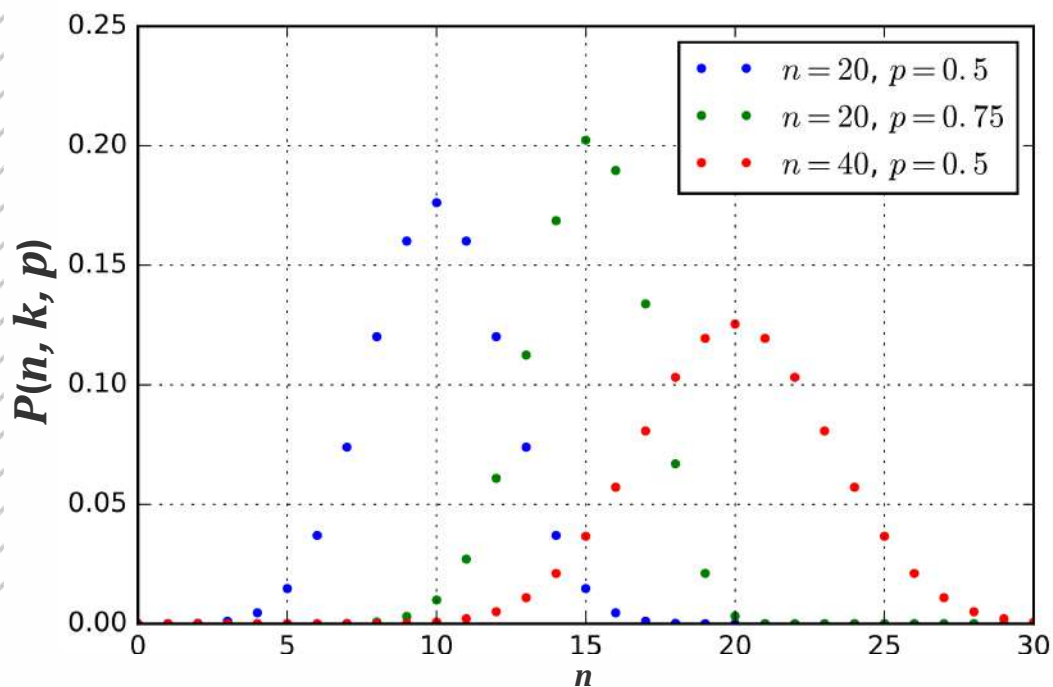
**Распределение Бернулли** – дискретное распределение вероятностей, моделирующее случайный эксперимент произвольной природы, при заранее известной вероятности успеха или неудачи.

Случайная величина  $X$  имеет распределение Бернулли, если она принимает всего два значения: 1 и 0 с вероятностями  $p$  и  $q \equiv 1 - p$  соответственно. Таким образом:

#02

$$\begin{aligned} P(X=1) &= p, \\ P(X=0) &= q. \end{aligned}$$

# Виды распределений. Распределение Бернулли и биномиальное распределение



**Биномиальное распределение** – распределение количества  $k$  «успехов» в последовательности из  $n$  независимых случайных экспериментов, таких, что вероятность «успеха» в каждом из них постоянна и равна  $p$ . Случайную величину, распределенную по закону биномиального распределения, интерпретируют как число успехов  $k$  в серии из  $n$  одинаковых независимых испытаний Бернулли с вероятностью успеха  $p$  в каждом испытании

Мат.ожидание:

$$M(X) = np,$$

Дисперсия

$$D(X) = npq,$$

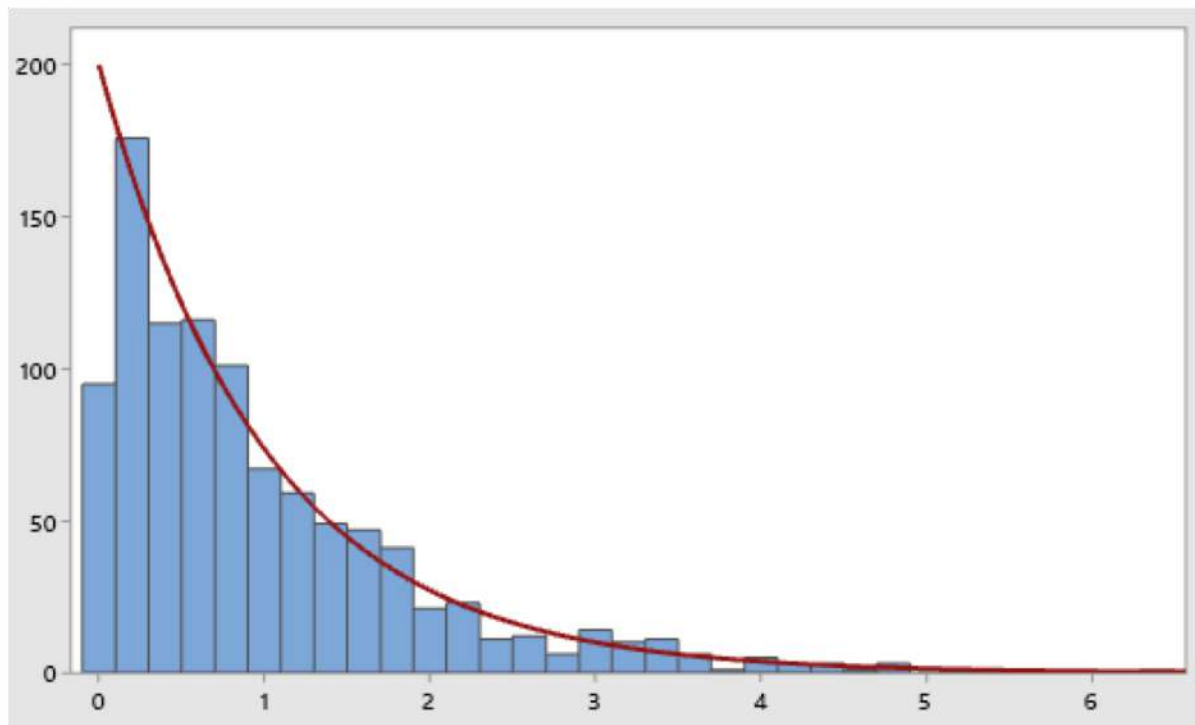
Ср.-кв. отклонение:

$$\sigma(X) = \sqrt{npq}.$$

$$P_n^k = C_n^k p^k q^{n-k} = C_n^k p^k (1-p)^{n-k}, \text{ где}$$
$$C_n^k = \frac{n!}{k!(n-k)!}$$

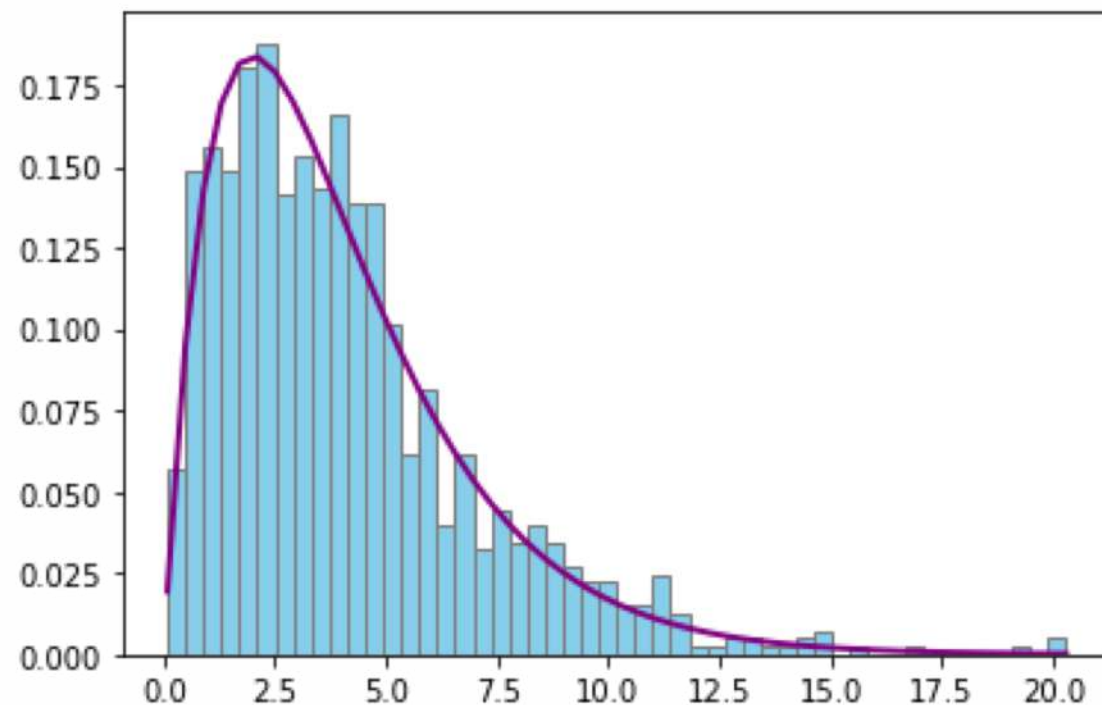
# Виды распределений. Экспоненциальное распределение и $\gamma$ -распределение

Экспоненциальное (показательное) распределение



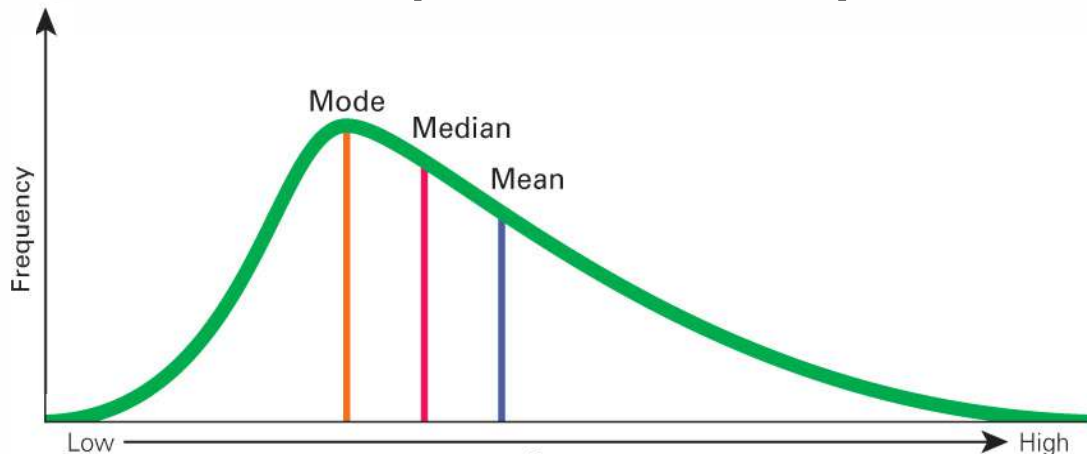
Например, retention на N-ый день после первого захода с РК; кол-во просмотров объявления на N-ый день; органический retention после установки

$\gamma$ (гамма)-распределение



Например, пользователи по частоте заказов или их кол-ву за период; кол-во установок на N-ый день начала РК

# Ассиметричные и мультимодальные распределения



Левая асимметрия:

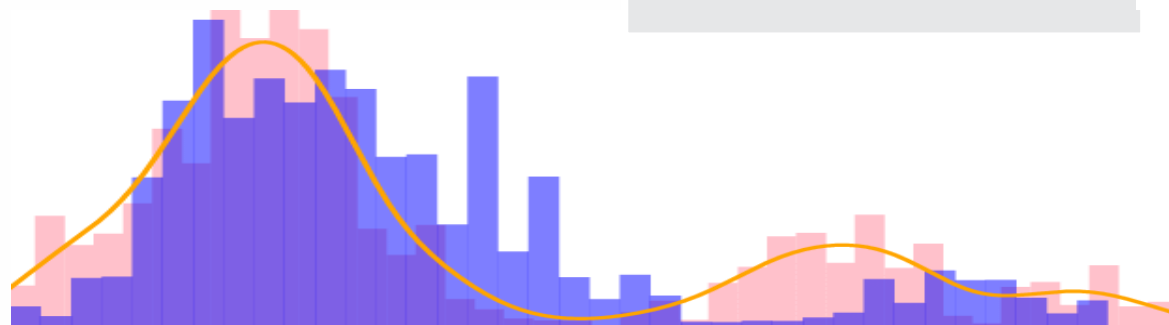
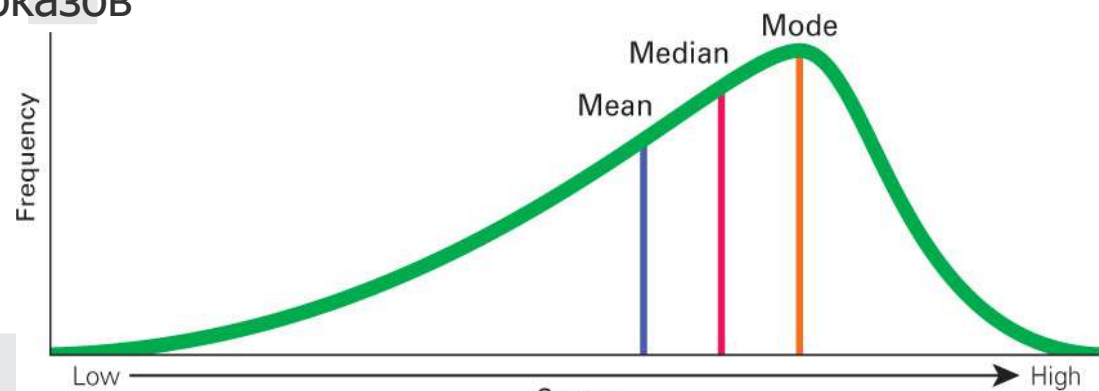
- пик левее;
- длинный хвост;
- среднее смещено к хвосту;

Примеры: аудитория, визиты, кол-во заходов при высоком churn rate в сервисе; CTR рекламного баннера во времени – затухает с большим кол-вом повторных показов

Правая асимметрия:

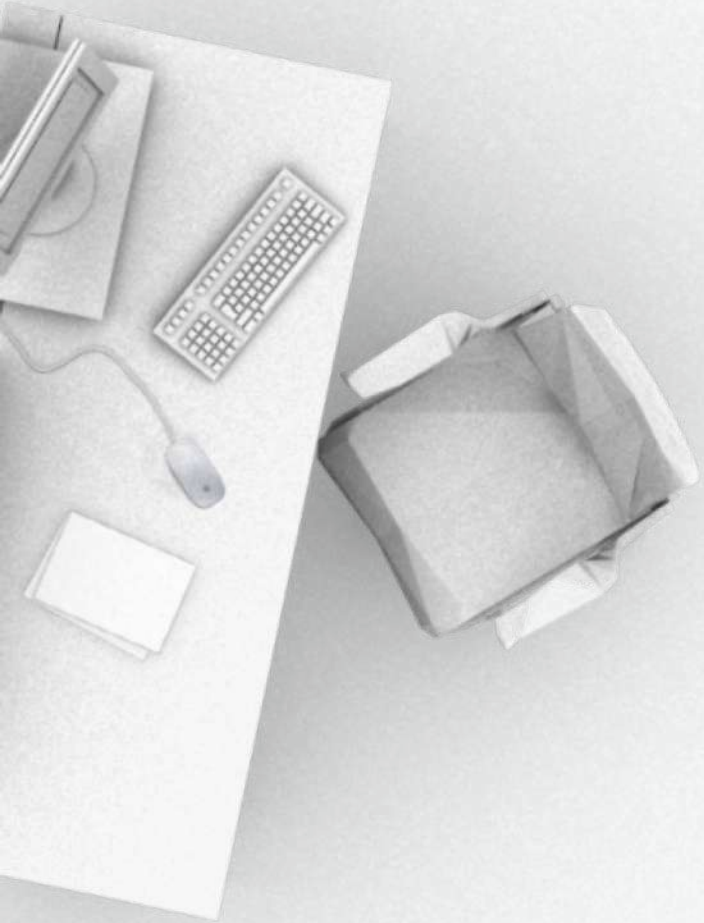
- пик правее;
- короткий хвост;
- среднее смещено к голове графика;

Примеры: запуск нового сервиса, раздела, фичи (долго раскачивается)



**Би-модальное распределение:** загруженность метро и общественного транспорта, трафик (пробки) на дорогах – утром и вечером;

**Мультимодальное распределение:** активность пользователей в социальных сетях с мобильного – утром и вечером в транспорте, днём за обедом



## 4. Проверка статистической значимости.



# Уровень значимости и мощность критерия на гистограмме

**$\alpha$ -уровень** - пороговый уровень статистической значимости; вероятность ошибочно отклонить нулевую гипотезу.

**$\beta$ -уровень** - вероятность ошибочного не отклонения нулевой гипотезы об отсутствии различий.

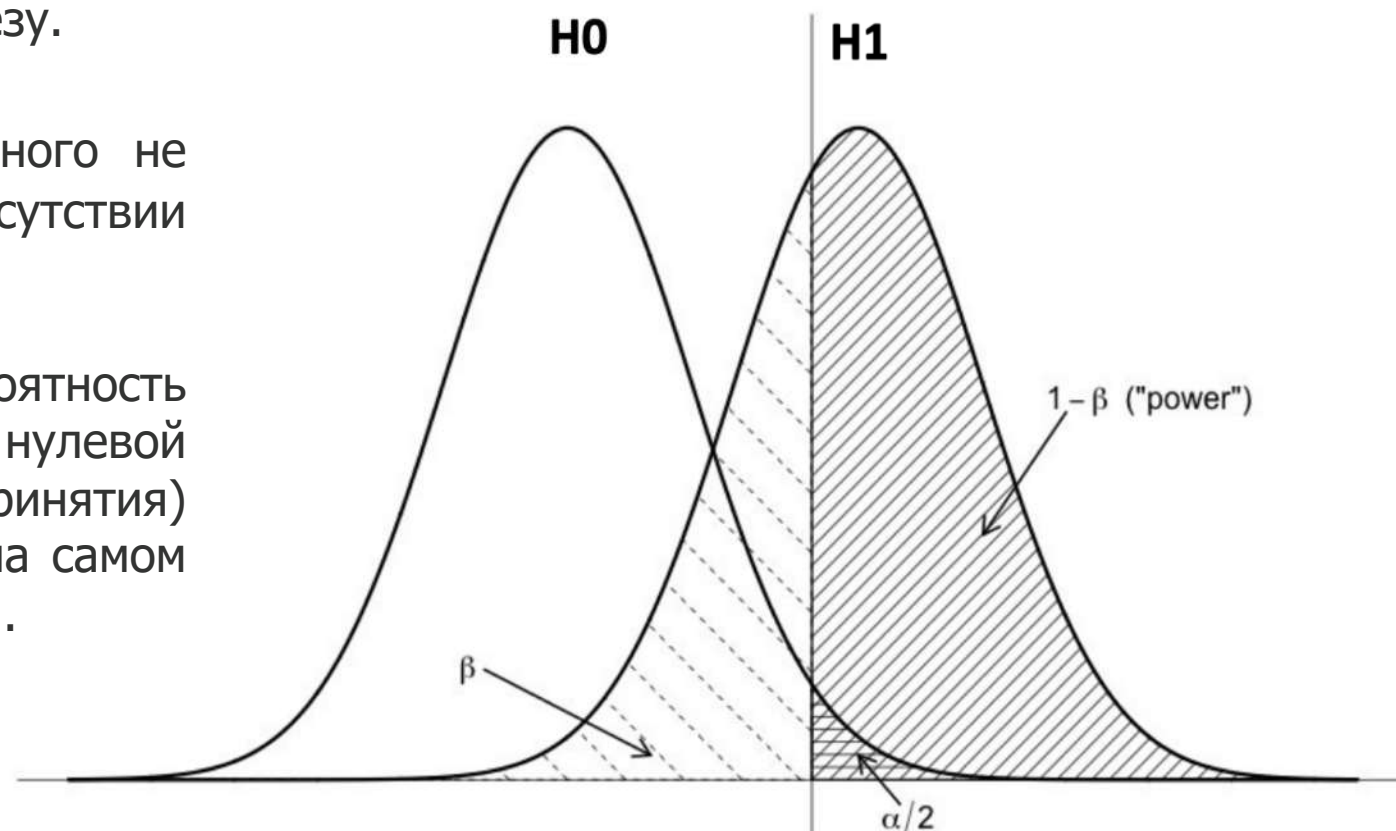
**Мощность (критерия)** - вероятность (правильного) отбрасывания нулевой гипотезы, т. е. отбрасывания (непринятия) нулевой гипотезы в случае, когда на самом деле верна альтернативная гипотеза.

- **$H_0$  – основная гипотеза** (о сходстве):

$$\mu(A) == \mu(B);$$

- **$H_1$  – альтернативная гипотеза** (о различии):

$$\mu(A) \neq \mu(B);$$



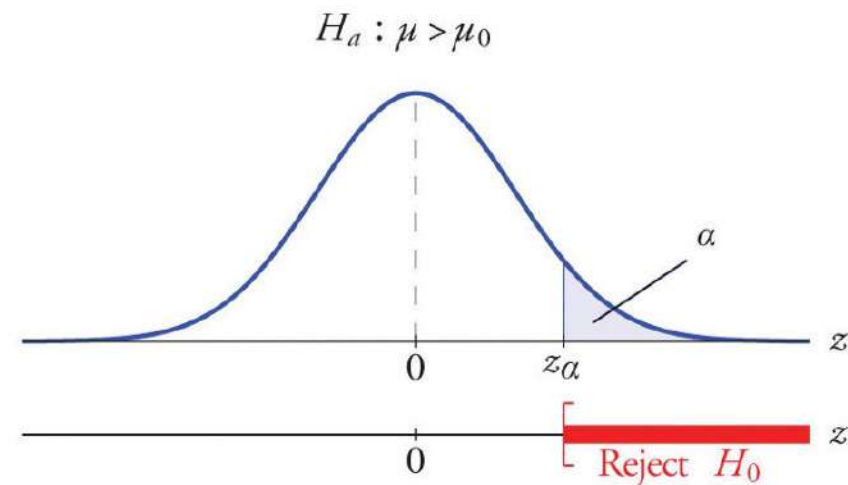
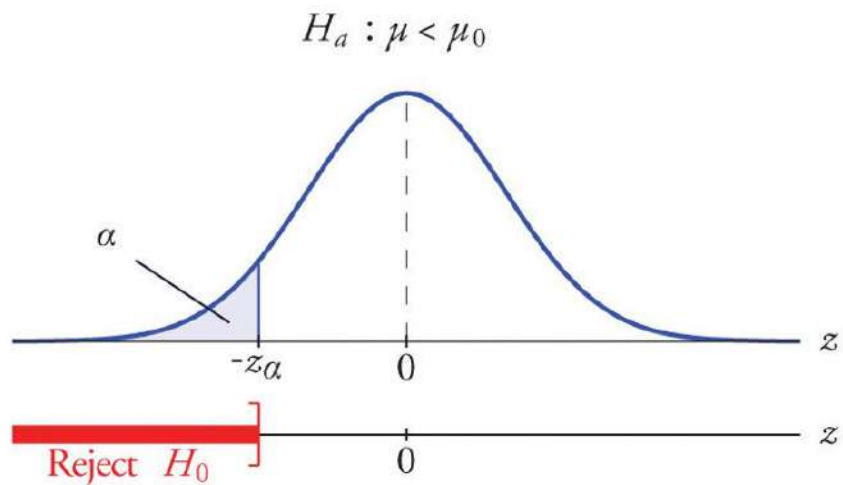
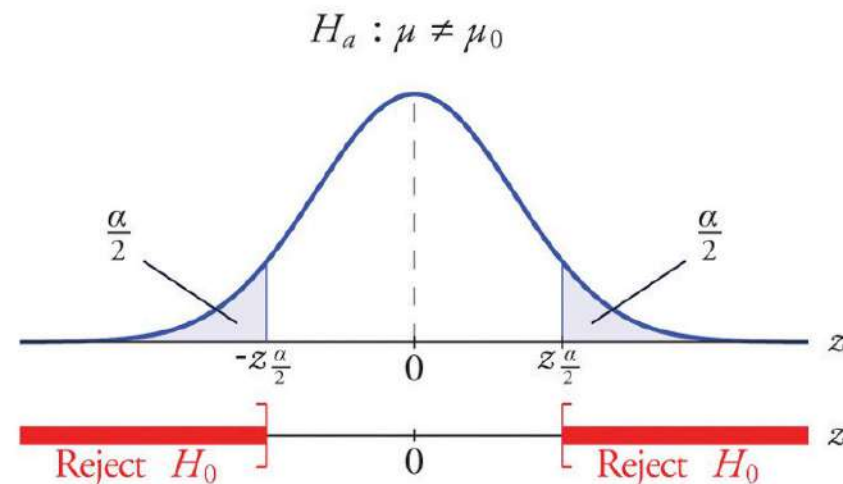
# $\alpha$ -уровень статистической значимости

**$\alpha$ -уровень** - пороговый уровень статистической значимости; вероятность ошибочно отклонить нулевую гипотезу – вероятность **ошибки I-го рода** «Ложная тревога».

Стандартный: 0,05

Высокий: 0,01

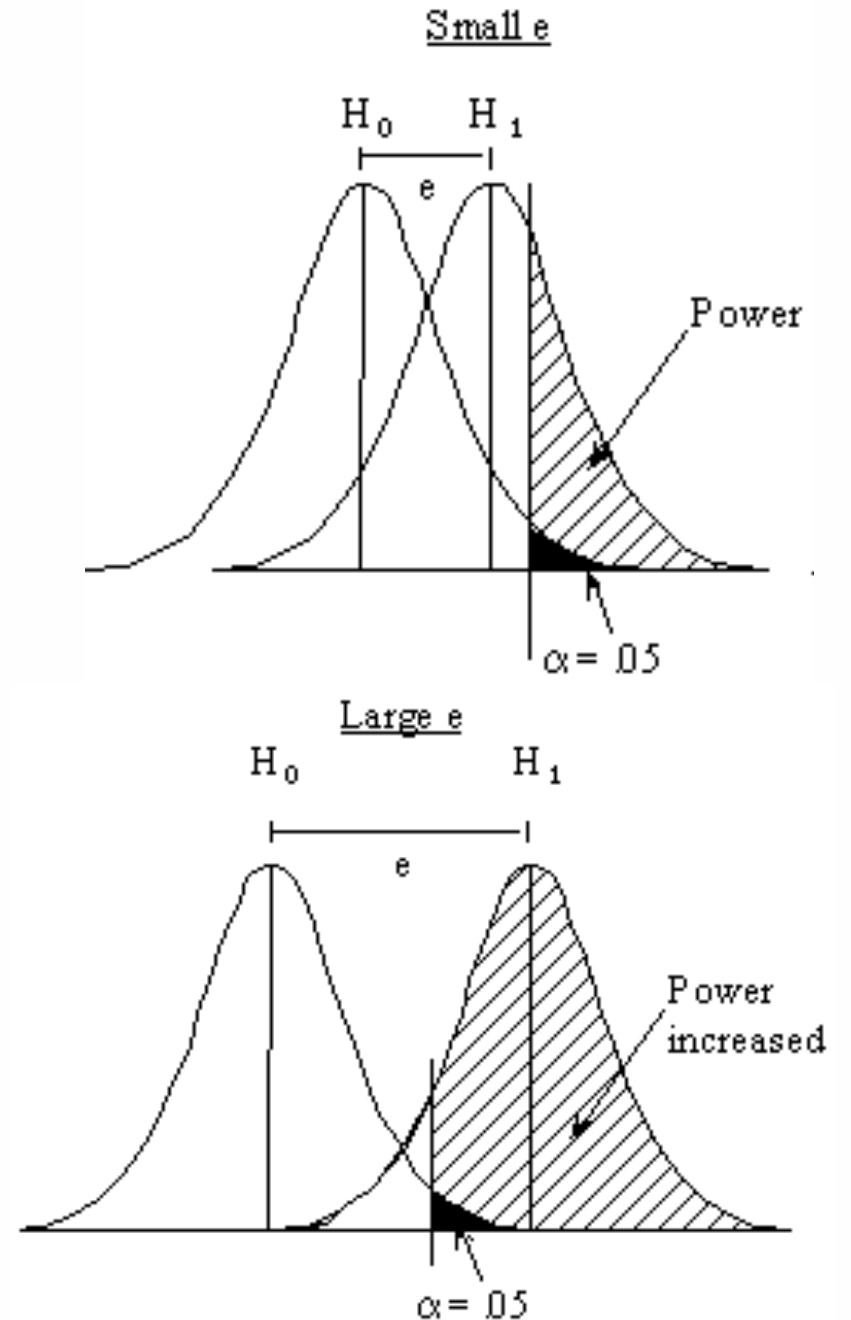
Низкий: 0,1



# Мощность критерия ( $1-\beta$ )

**$\beta$ -уровень** - вероятность ошибочного не отклонения нулевой гипотезы об отсутствии различий – вероятность **ошибки II-го рода** «Пропуск цели».

**Мощность (критерия) ( $1-\beta$ )** - вероятность (правильного) отбрасывания нулевой гипотезы, т. е. отбрасывания (непринятия) нулевой гипотезы в случае, когда на самом деле верна альтернативная гипотеза.



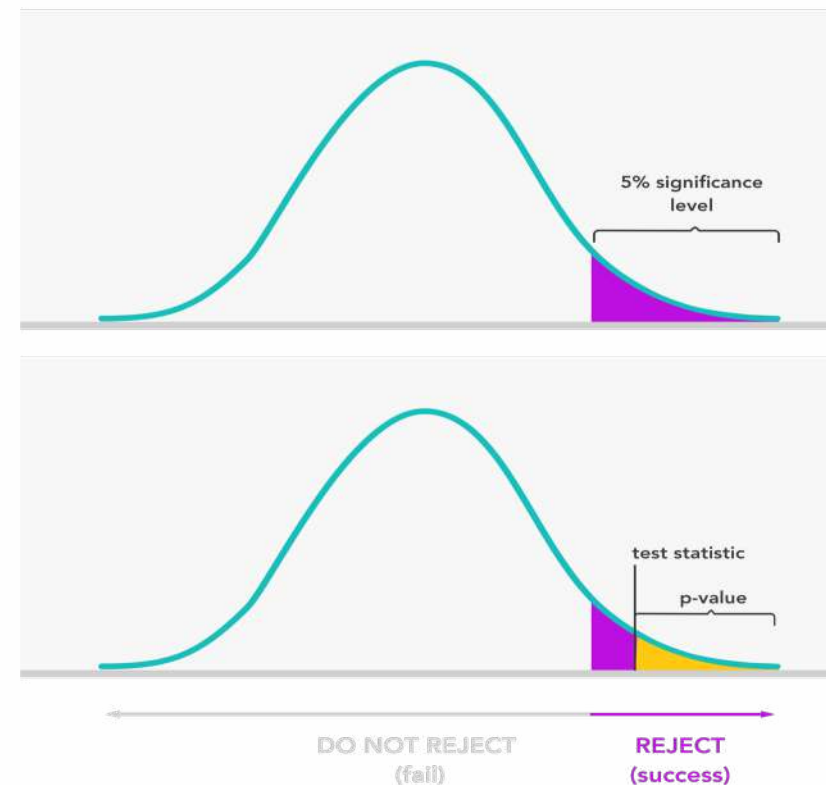
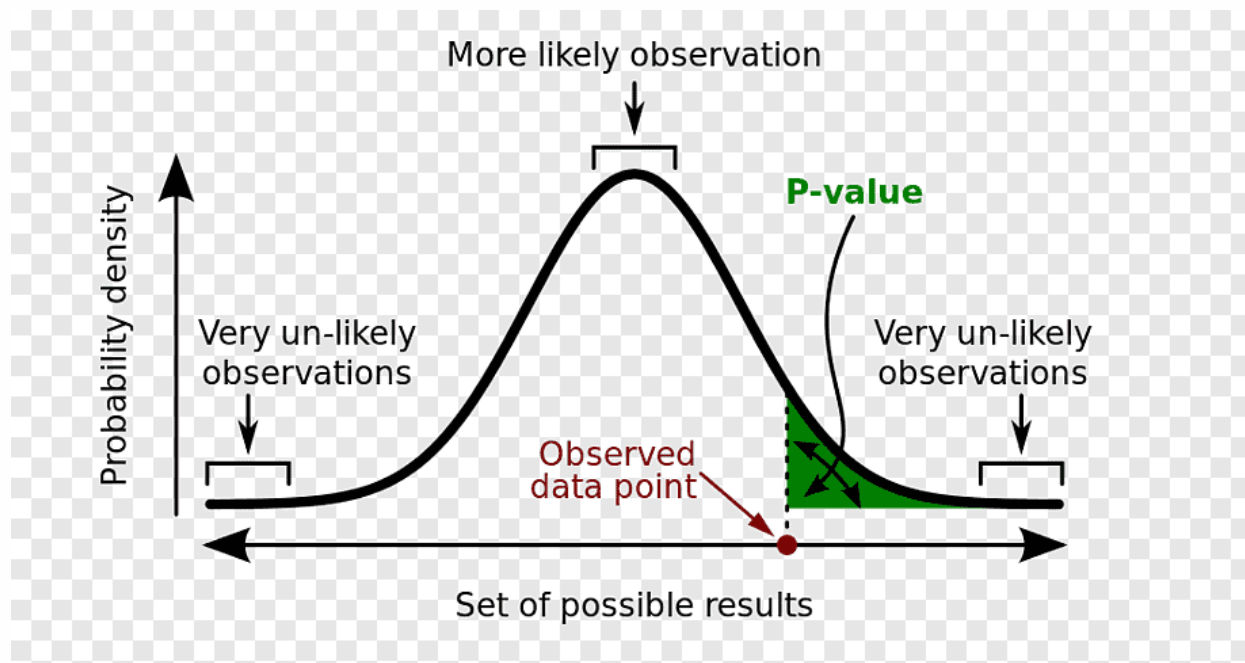
# *p*-значение

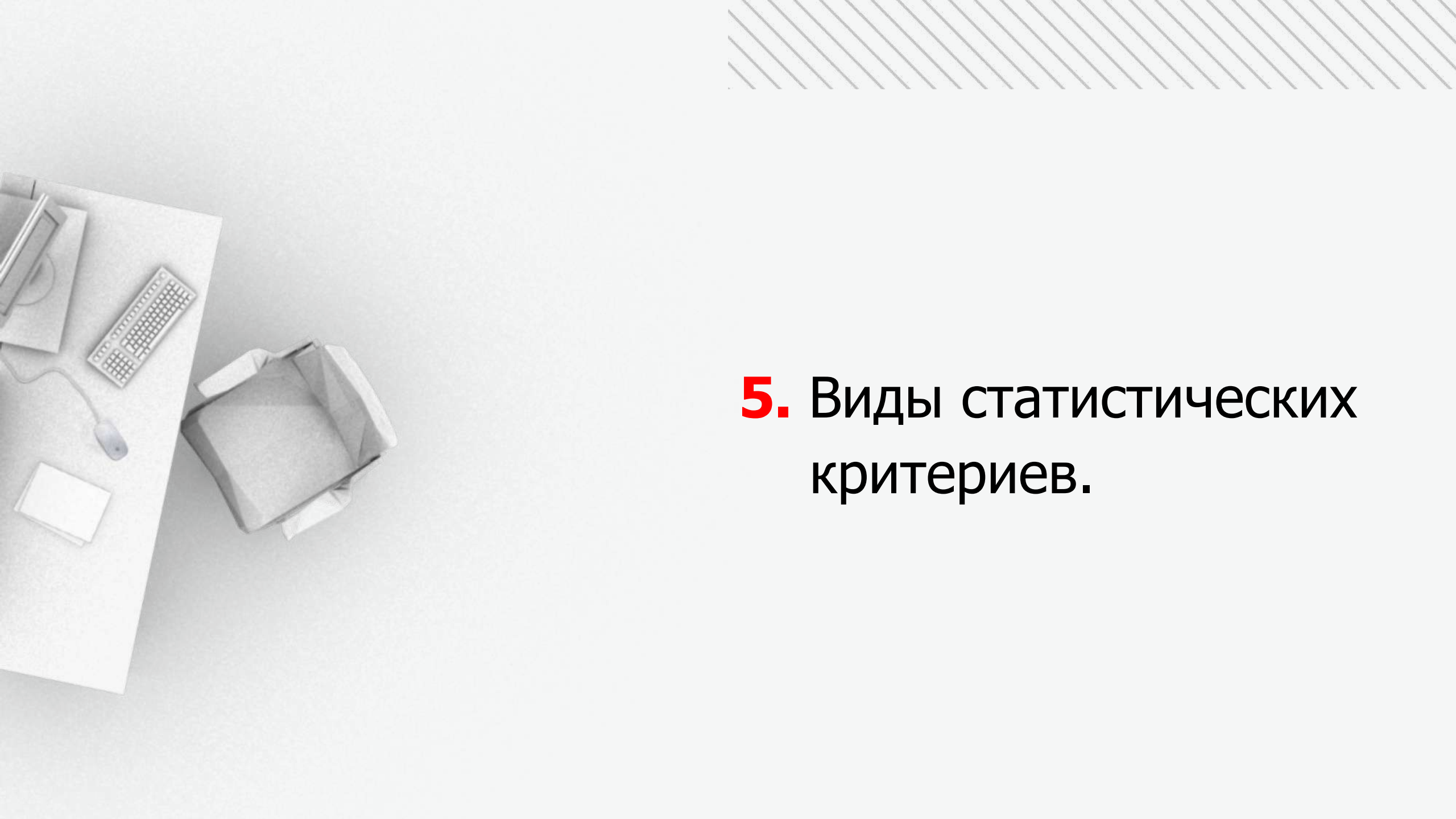
*p*-значение – рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы **H<sub>0</sub>**.

Иными словами, вероятность получить такое же или более экстремальное значение статистики (среднего арифметического, медианы и др.), по сравнению с ранее наблюдаемым, при условии, что **H<sub>0</sub>** верна.

**H<sub>0</sub>** отклонена:  $p\text{-value} < \alpha\text{-level}$

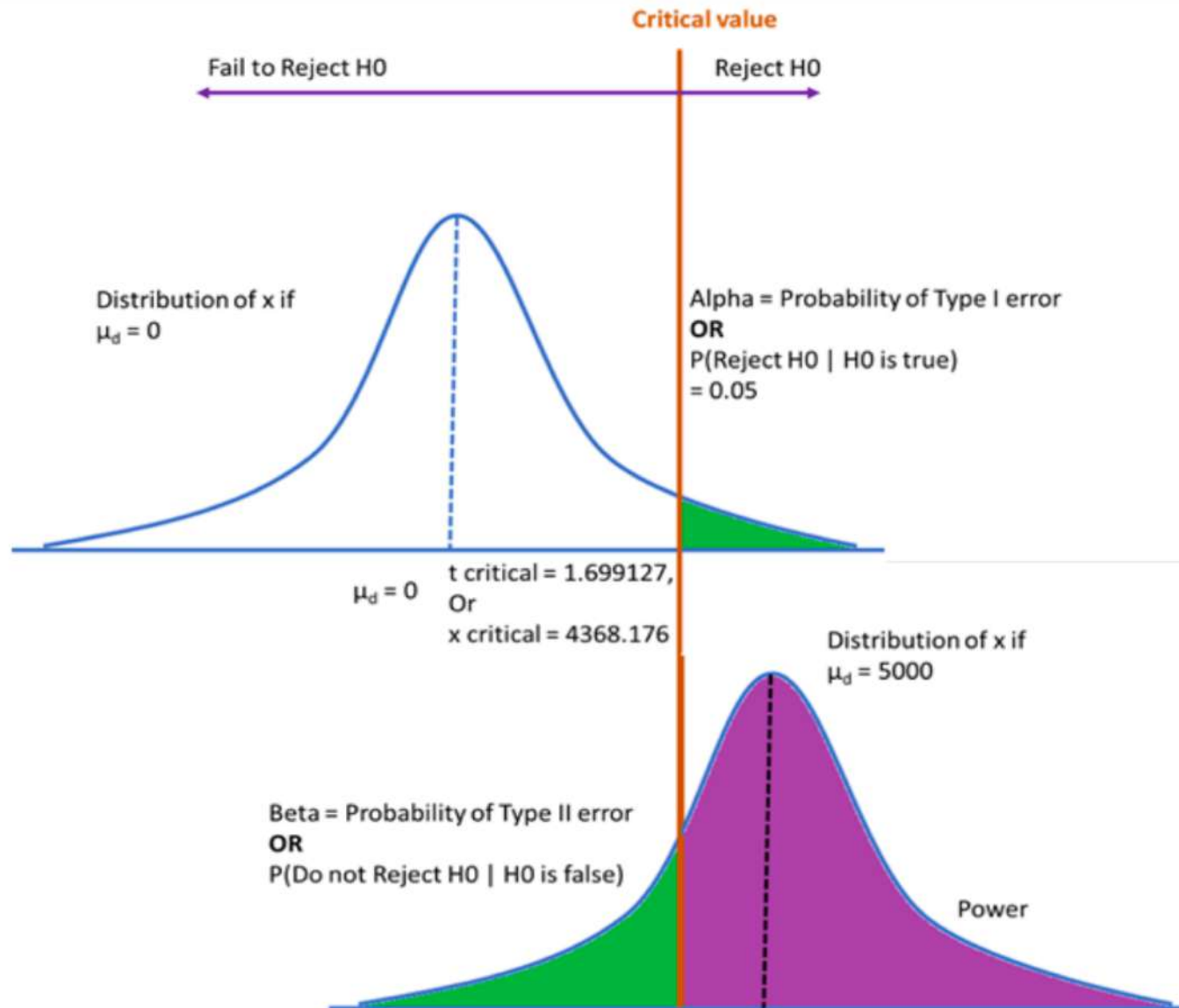
**H<sub>0</sub>** принята:  $p\text{-value} \geq \alpha\text{-level}$





## **5.** Виды статистических критериев.

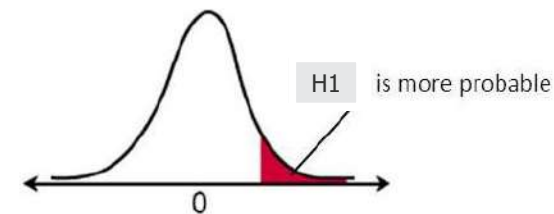
# Принцип принятия и отвержения гипотезы



**p-уровень** - рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы.

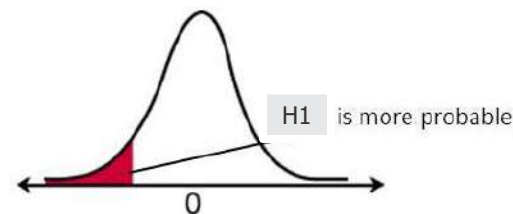
**p-level** < **a-level**  $\Rightarrow$  **H1** верна;

**p-level**  $\geq$  **a-level**  $\Rightarrow$  **H0** верна;



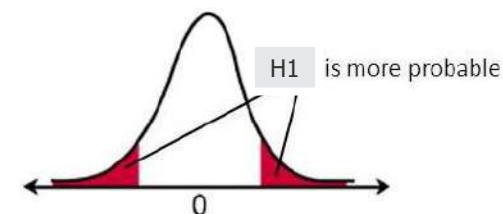
Right-tail test

**H1** :  $\mu > \text{value}$



Left-tail test

**H1** :  $\mu < \text{value}$



Two-tail test

**H1** :  $\mu \neq \text{value}$



# Виды статистических критериев по исследуемой метрике

**Критерии согласия** (проверка соответствия распределения метрики в выборке определенному виду эталонному распределению):

- 1) Критерий Колмогорова-Смирнова;
- 2)  $\chi^2$ -Критерий Пирсона (хи-квадрат);
- 3) Критерий Шапиро-Уилкса;

**Критерии сдвига** (проверка равенства групп):

- 1) Т-Критерий Стьюдента;
- 2) Т-Критерий Уилкоксона;
- 3) U-Критерий Манна-Уитни;

**Критерии однородности** (например, проверки на равенство дисперсий):

- 1) Критерий Бартлетта;
- 2) Критерий Левена

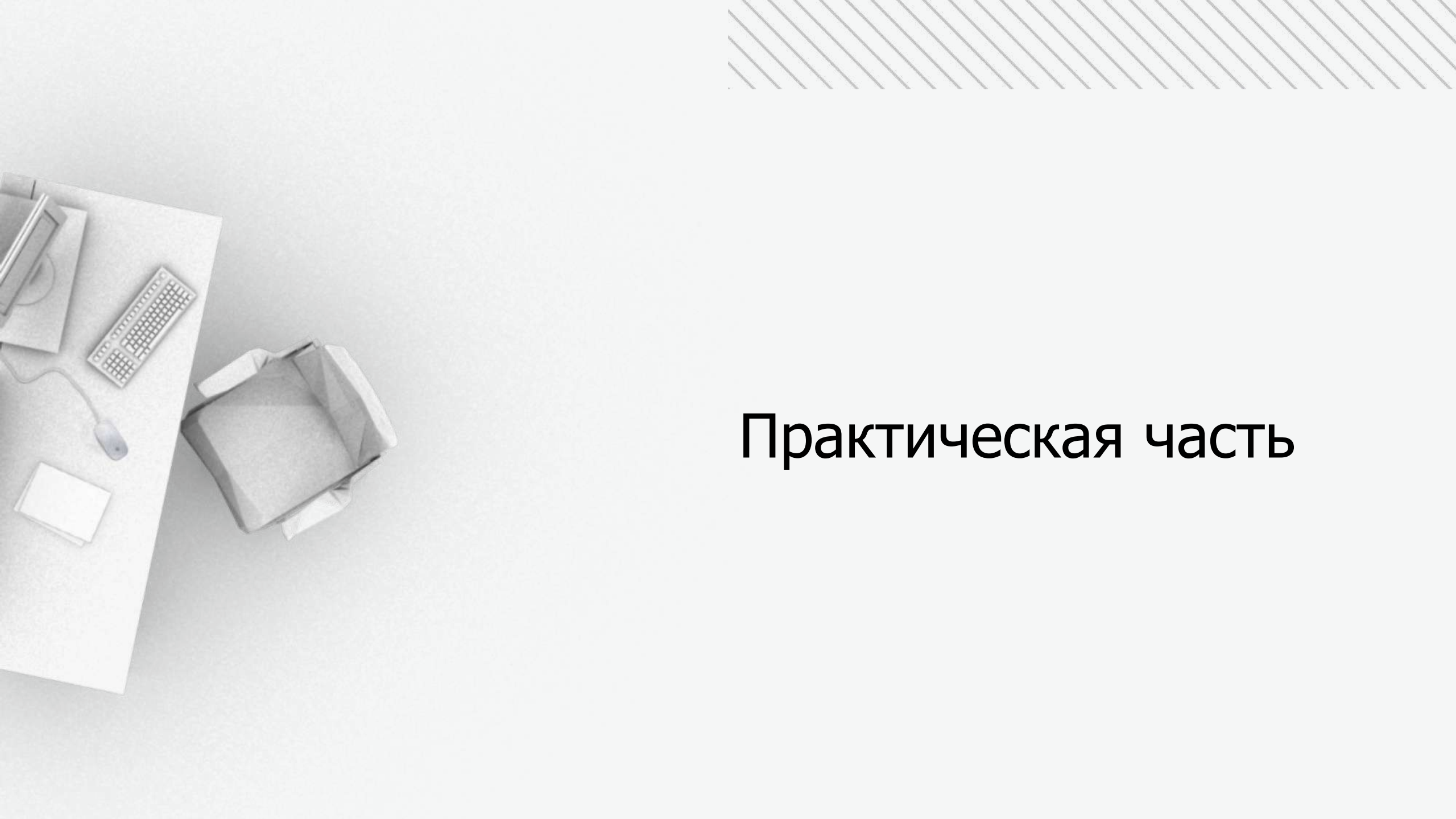
# Виды статистических критериев по применяемому алгоритму

**Параметрические** – основаны на конкретном типе распределения:

- 1) Т-Критерий Стьюдента;
- 2) Z-критерий Фишера;
- 3) F-критерий Фишера;
- 4)  $\chi^2$ -критерий Пирсона;

**Непараметрические** – не базируется на предположении о типе распределения генеральной совокупности и не использует параметры этой совокупности (ранговые критерии):

- 1) Т-Критерий Уилкоксона;
- 2) U-Критерий Манна-Уитни;
- 3) Критерий Колмогорова;
- 4) Q-Критерий Розенбаума



# Практическая часть

# Python Style-guide

```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-
"""
- *- Annotation: - *-
shebang-line and encoding derictive:
PEP 394: The "python" Command on Unix-Like Systems
https://www.python.org/dev/peps/pep-0394/
PEP 263: Defining Python Source Code Encodings
https://www.python.org/dev/peps/pep-0263/
"""
```

---

# Python Scientific Libraries

```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-

import matplotlib.pyplot as plt
import numpy as np

"""
-- Annotation: --
numpy - is a library for working with linear algebra forms:
numpy is a basis of pandas, scipy and sklearn data structures and types
"""
```

# Непрерывное распределение: определение и boxplot

```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-
from time import time
import matplotlib.pyplot as plt
import numpy as np
```

```
# 1) set random seed as current timestamp = seconds from 1970-01-01 00:00:00
```

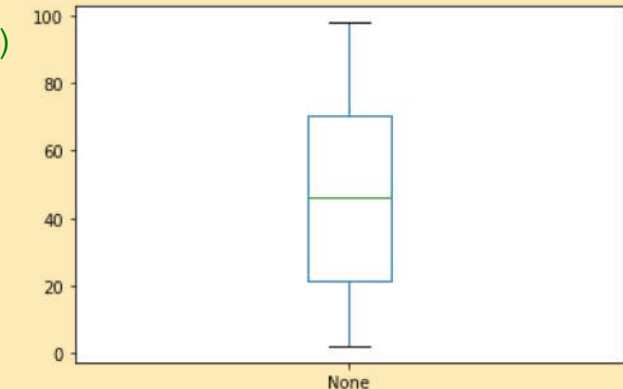
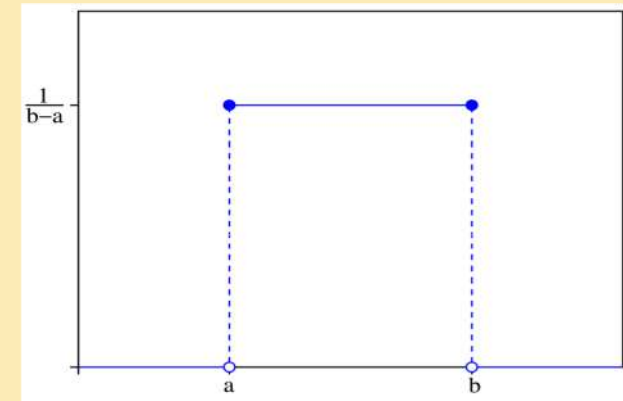
```
np.random.seed(int(time()))
```

```
# 2) np.random.rand(n) returns 1-dimensional np.array()
```

```
# of n values distributed uniformly over [0, 1)
```

```
series = (np.random.rand(50)*100).astype(int)
```

```
pd.Series(series).plot.box()
```

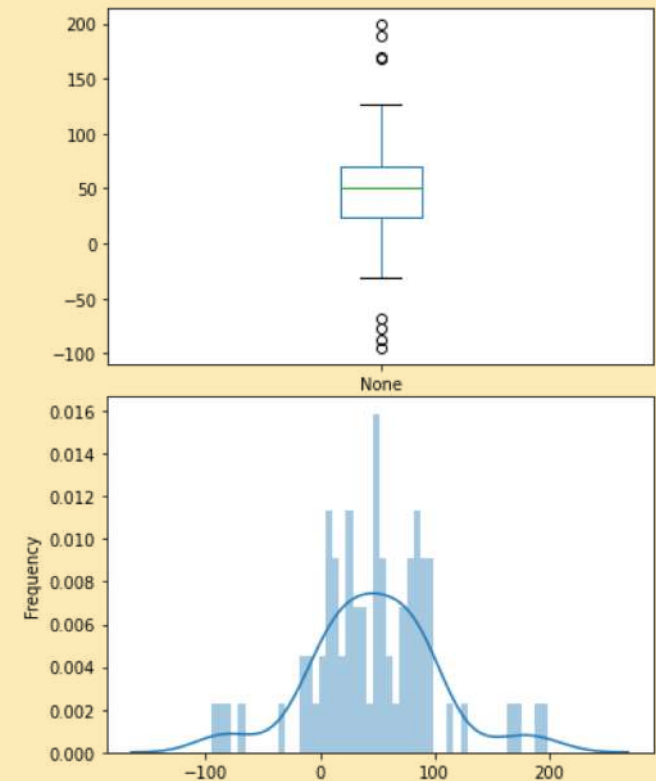


# Непрерывное распределение с выбросами: диаграмма размаха и гистограмма

```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-
from time import time
import matplotlib.pyplot as plt
import numpy as np

np.random.seed(int(time()))

distribution = (np.random.rand(50)*100).astype(int)
center = np.ones(5) * 50
high_outliers = (np.random.rand(10)*200).astype(int)
low_outliers = (np.random.rand(10)*-100).astype(int)
series = np.concatenate((distribution, center, high_outliers, low_outliers))
pd.Series(series).plot.hist(bins = 50)
```





**10**  
баллов

Срок сдачи:

**25.11.2020**

## Домашнее задание

АВТ-1

#040

Полезные ссылки и литература в помощь:

- Татьяна Мелехина: «Лекции по теории вероятностей и математической статистике»
- Wes McKinney: "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython"
- <https://pandas.pydata.org/>
- <https://matplotlib.org/>
- <https://jupyter.org/>

## Домашнее задание: АВТ. Правила сдачи

1. Оформление решения производится в **Jupyter Notebook** на сервисе **Google Colab**: <https://colab.research.google.com/>.
2. В качестве **решения** присылайте, пожалуйста, **ссылку на Notebook**, созданный на сервисе, с комментариями, вопросами и результатами работы в рамках Markdown Text Field;
3. Для доступа к сервису необходимо **зарегистрировать** Google Account на свою почту и указать **ФИО** так же, как указано у Вас в профиле на учебном портале проекта Техноатом@Mail.Ru.

## Домашнее задание: АВТ-1

1. Сгенерируйте 1000 значений в диапазоне от 5 до 105 методом `random()` и сохраните их в структуру `pandas.Series()`. Какой вид имеет распределение и почему именно такой? Что нужно сделать, чтобы распределение изменило форму?
2. Постройте гистограмму при помощи `matplotlib`. Рассчитайте основные статистические метрики для выборки вручную и в `python` (подсказка: полезно будет рассчитать частоту, с которой встречаются значения).
3. Проведите аналогичные расчёты при помощи встроенных методов класса `pandas.Series()`.
4. Результат работы пришлите в формате IPython Jupyter Notebook.

Ждём тебя на  
следующей лекции!



Spoiler: statistical significance, statistical  
significance everywhere...

