Правильные ответы на мини-тест

- Задача про открытки: 9/49, доля верно ответивших 68,2%
- Задача на комбинаторику: 3628800,3628800, доля верно ответивших 64,5%
- Ссылка: https://forms.gle/vqRd6HSVnfWkFzNo7

#01

FeedBack по занятию ABT-1:

- Что понравилось:
    - Много времени уделяется ответам на вопросы;
    - Новый практический материал из реальных проектов;
    - Узнал, зачем вообще матстат в реальной жизни;
- Что хотелось бы улучшить:
    - Слишком быстро, не были понятны некоторые моменты;
    - Больше кода для практики, перешли к теории;
    - Только одна неделя на ДЗ;
- Доходчивость и простота: среднее - 7,45, max - 50% ответов – 8;
- Полнота: 8,45, max – 30% ответов – 7;
- Новизна: 5,23, max – 30% ответов – 5.

ТЕХНОАТОМ: Продуктовая аналитика.
A/B-тестирование

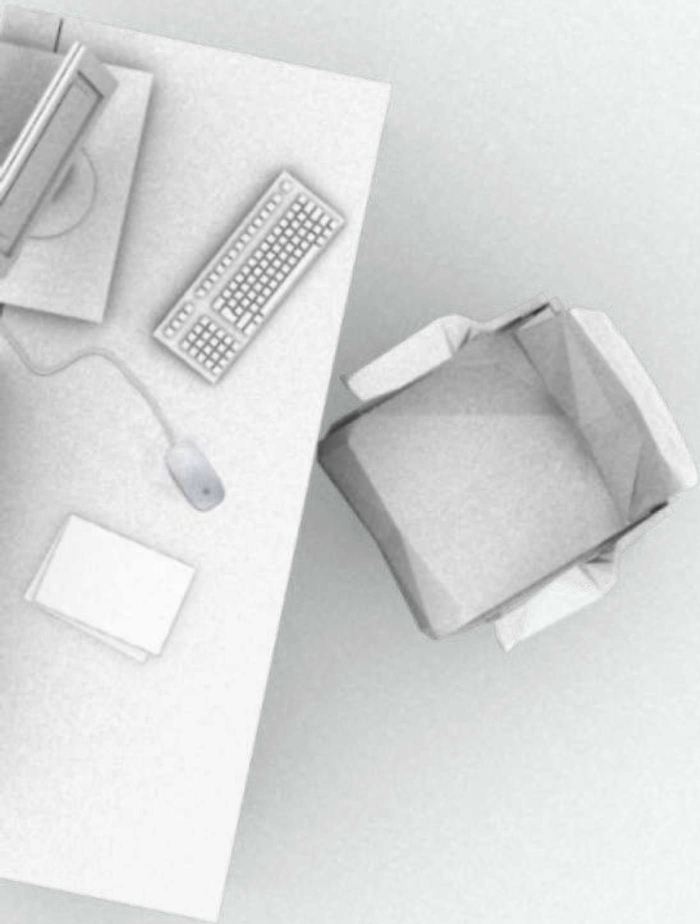
Лекция №2:

Методы оценки статистической значимости

Классические подходы

План лекции

1. **Виды статистических критериев**. Параметрические и непараметрические критерии.
2. **Оценка дисперсии**: критерии равенства дисперсий.
3. **Критерии нормальности**. Q-Q plot.
4. **Проверка статистической значимости**: t- и z-статистики, p-value, alpha-уровень и статистическая значимость, beta-уровень и мощность критерия. Классические методы оценки стат.значимости.
5. **Генеральная совокупность и ожидаемый эффект**. Расчёт размера выборки для достижения статистически значимых результатов. Оценка времени проведения теста.



1. Виды статистических критериев

Основы математической статистики

- Выборочное среднее:

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

- Среднеквадратичное отклонение:
- На основании смещённой оценки дисперсии:

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

- На основании несмещённой оценки:

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

- Математическое ожидание:

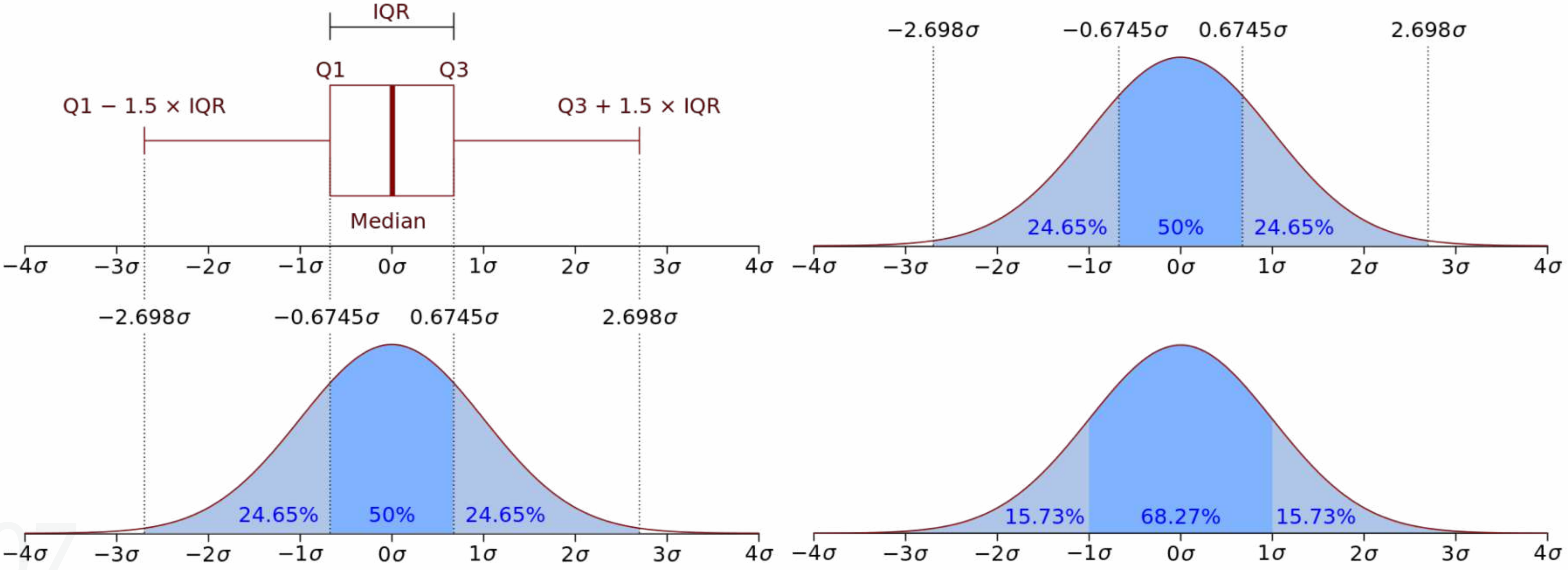
$$M[X] = \sum^{\infty} x_i\, p_i$$

- Дисперсия генеральной совокупности:

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

Основы математической статистики

Квантиль Q - значение, которое заданная случайная величина не превышает с фиксированной вероятностью;
IQR – межквартильный размах (**I**nter**Q**uartile **R**ange)



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;

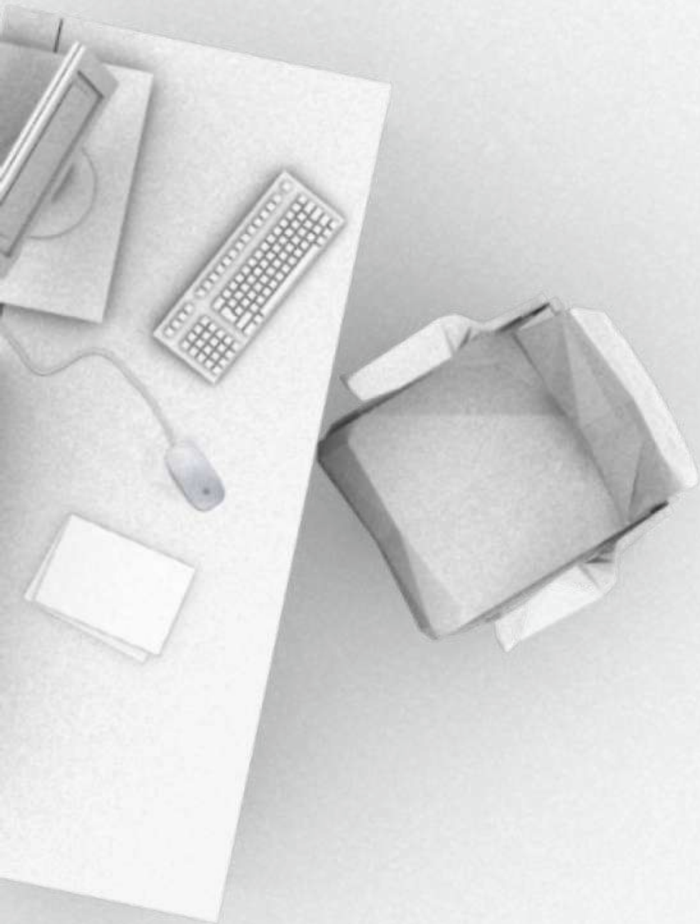
Критерии однородности (например, проверка равенства дисперсий):

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) χ2-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



2. Оценка дисперсии: теория

Оценка дисперсии: критерии равенства дисперсий

ANOVA – ANalysis Of Variance, дисперсионный анализ:

- F-критерий Фишера для 2-х выборок;
- Критерий Бартлетта для выборок, имеющих вид нормального распределения;
- Критерий Левена;
- Критерий или среднее Тьюки на основе внутригрупповой дисперсии;
- Критерий Хартли для выборок равного объема;
- G-критерий Кохрена для выборок равного объема;
- Dozens of them…

F-критерий Фишера

- Равенство (гомогенность) дисперсий 2-х выборок;
- Параметрический, основан на предположении о равенстве выборок нормальному закону;
- Критерий Фишера проверяет гипотезу о том, что дисперсии 2-х выборок одинаковы:

$$H_0: \sigma_1 = \sigma_2$$

- Альтернативная гипотеза: $H_1: \sigma_1 \neq \sigma_2$
- Статистика критерия:
    - одномерный случай: $F = \frac{S_{10}}{S_{20}} = \frac{\sigma_1}{\sigma_2}$
    - случай для нескольких групп: $F = \frac{MS_b}{MS_w}$, где

    MS_w – средний квадрат суммы внутригрупповых отклонений,

    MS_b – средний квадрат суммы межгрупповых отклонений;
- $H0$ отклоняется, если $F > F_{cr}$, где F_{cr} − верхнее критическое значение F-распределения, с соответствующими степенями свободы числителя и знаменателя, соответственно, при выбранном уровне стат.значимости.

Критерий Бартлетта

- Равенство дисперсий 2х и более выборок;
- Параметрический, основан на предположении о равенстве выборок нормальному закону;
- Объём каждой выборки должен быть больше трёх;
- Объемы выборок могут отличаться;
- Критерий Бартлетта проверяет гипотезу о том, что дисперсии всех выборок одинаковы:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

- Альтернативная гипотеза: существует, по крайней мере, две выборки i, j, где $i \neq j$ с несовпадающими дисперсиями:

$$\exists\, i, j : H_1 : \sigma_i \neq \sigma_j$$

Критерий Бартлетта

- Статистика: $$T = \frac{(N-k)\cdot\ln(s_p^2) - \sum^{k}(n_i - 1)\cdot\ln(s_i^2)}{1 + \frac{1}{3\cdot(k-1)}\cdot\left(\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{(N-k)}\right)}$$, где

$N = \sum_{i=1}^{k} n_i$ – суммарный объём всех групп,

$s_p^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)\cdot s_i^2$ – суммарная оценка дисперсий,

k – количество групп,

- Условие отвержения H0 на уровне стат.значимости α:

$T > \chi^2_{k-1,\alpha}$, где $\chi^2_{k-1,\alpha}$- α-квантиль распределения Пирсона

с $k-1$ степенью свободы

Критерий Левена

- Равенство дисперсий 2х и более выборок;
- Менее чувствителен к отклонению распределения от нормального;
- Уступает критерию Бартлетта по уровню мощности (выше вероятность принять H_0 ложно);
- Тест выполняется над преобразованными данными;
- Критерий Левена проверяет гипотезу о том, что дисперсии всех выборок одинаковы:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

- Альтернативная гипотеза: существует, по крайней мере, две выборки i, j, где $i \neq j$ с несовпадающими дисперсиями:

$$\exists\ i, j : H_1 : \sigma_i \neq \sigma_j$$

Критерий Левена

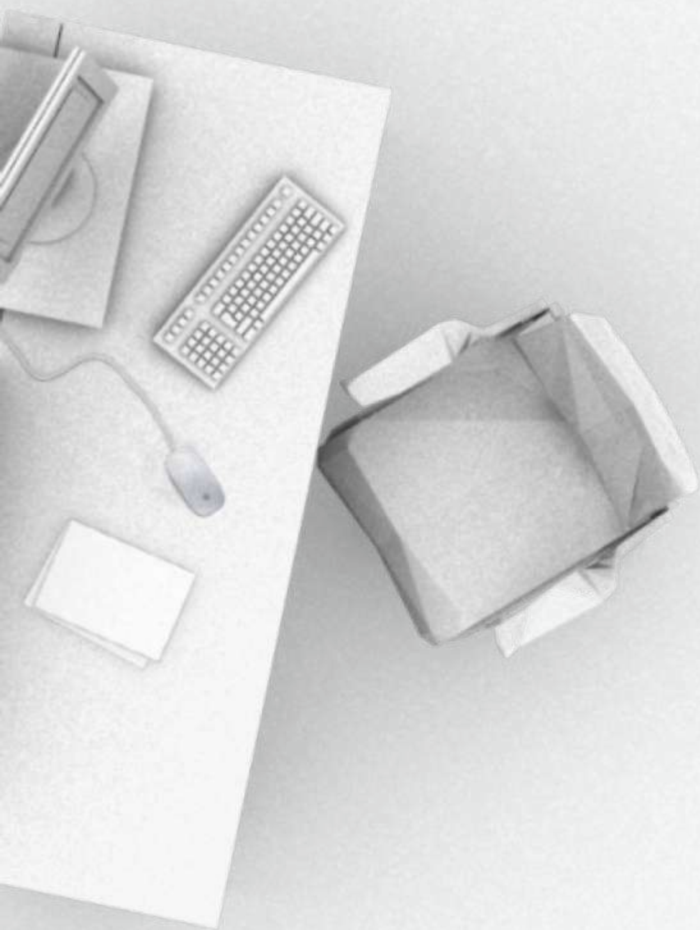
- Статистика: $W = \frac{(N-k)\sum_{i=1}^{k}(\bar{Z}_i - \bar{Z})}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Zi_j - \bar{Z}_i)}$,

где n_i – объём i-й выборки, $N = \sum_{i=1}^{k} n_i$,

X_{ij} – j-е наблюдение в i-й выборке,

Z_{ij} = $|X_{ij} - \bar{X}_i|$ или Z_{ij} = $|X_{ij} - median(X_i)|$

- $H0$ отклоняется, если $W > F$, где F – верхнее критическое значение F-распределения, с $m-1$ и $N-m$ степенями свободы числителя и знаменателя, соответственно, при выбранном уровне стат.значимости.



2. Оценка дисперсии: Практическая часть

F-критерий Фишера: scipy.stats.f_oneway

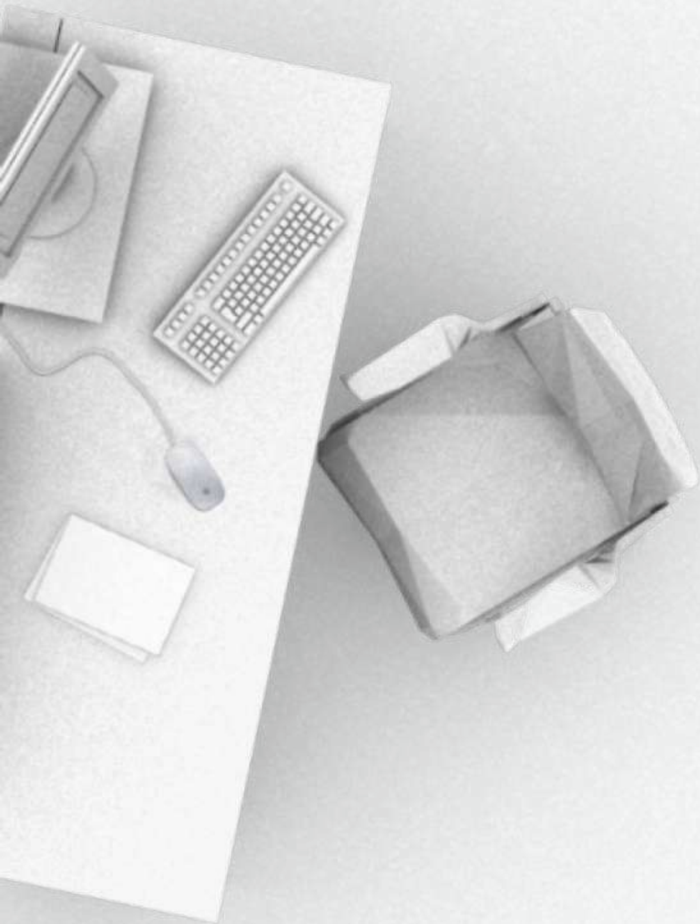
```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-
from time import time
import numpy as np
from scipy.stats import f_oneway

np.random.seed(int(time()))
series_1 = (np.random.rand(50)*100).astype(int)
series_2 = (np.random.normal(size=50)*100).astype(int)
alpha = 0.05
p_value = f_oneway(series_1, series_2)[1]
print('{} the H0: Var(X1) == Var(X2)\n'.format(
{True:'Reject', False:'Confirm'}[p_value > alpha]))
```

Критерий Бартлетта: scipy.stats.bartlett

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

import numpy as np
from scipy.stats import bartlett

alpha = 0.05

def bartlett_test(df1, df2, df3, p_value = alpha):
        st = bartlett(df1, df2, df3)
        print('Variances of distributions is {}equal\n'.format( {True:'not ',
False:''}[st[1] < p_value]))

series_1 = (np.random.rand(50)*100).astype(int)
series_2 = (np.random.normal(size=50)*100).astype(int)
shape, scale = 2., 2. # mean=4, std=2*sqrt(2)
series_3 = (np.random.gamma(shape, scale, 100)).astype(int)

bartlett_test(series_1, series_2, series_3)
```

Критерий Левена: `scipy.stats.levene`

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

import numpy as np
from scipy.stats import levene

alpha = 0.05

def levene_test(df1, df2, df3, p_value = alpha):
       st = levene(df1, df2, df3)
       print('Variances of distributions is {}equal\n'.format( {True:'not ',
False:''}[st[1] < p_value]))

series_1 = (np.random.rand(50)*100).astype(int)
series_2 = (np.random.normal(size=50)*100).astype(int)
shape, scale = 2., 2. # mean=4, std=2*sqrt(2)
series_3 = (np.random.gamma(shape, scale, 100)).astype(int)

levene_test(series_1, series_2, series_3)
```



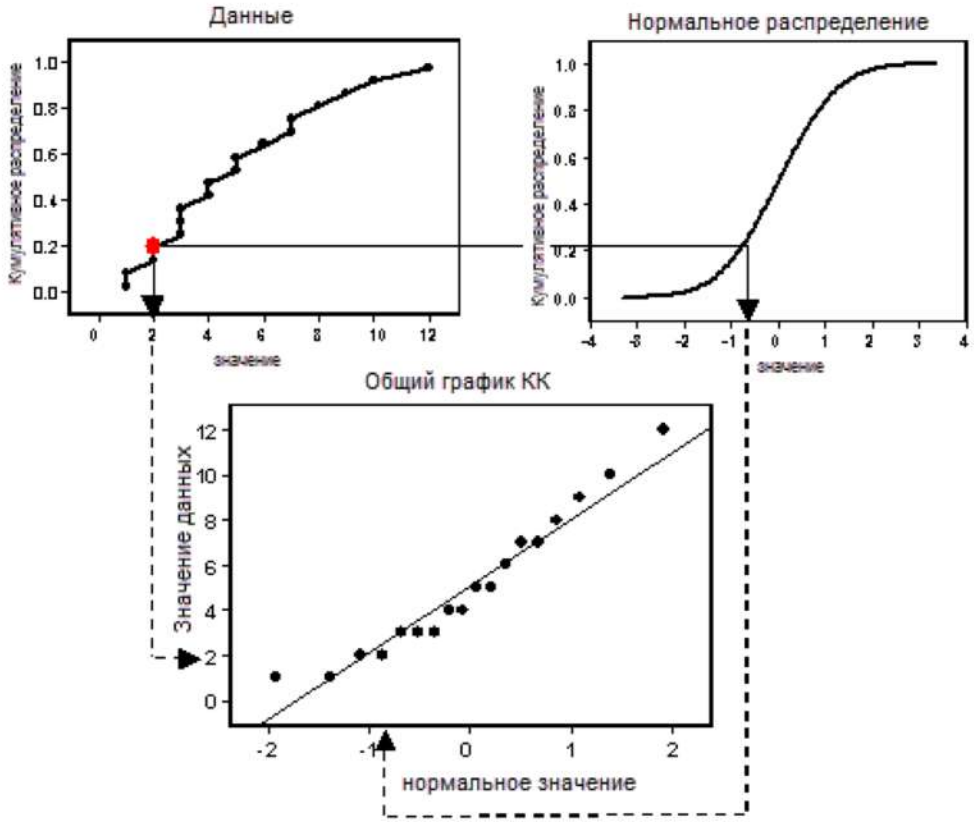
3. Критерии нормальности: теория

Критерии нормальности

- Критерий Шапиро-Уилка;
- Критерий Колмогорова-Смирнова;
- Критерий ассиметрии и эксцесса;
- Критерий Пирсона χ^2;
- Dozens of them…

Визуализация: Q-Q plot

- По оси Y – квантили (кумулятивным итогом начиная с наименьшего значения);
- По оси X – значения (отсортированные в порядке начиная с наименьшего)



Критерий Колмогорова-Смирнова

- Непараметрический критерий согласия;
- Проверяет соответствие закона распределения заданному:

H0: $F_n \sim F$

- Статистика – на основе максимума разности между кумулятивным распределением выборки $F_n(x)$ и предполагаемым кумулятивным распределением эмпирического распределения $F(x)$:

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

- Если статистика $\sqrt{n}D_n$ превышает квантиль K_α заданного уровня распределения Колмогорова, то H0 отвергается.

Критерий Шапиро-Уилка

- Обладает большей мощностью;
- Проверяет соответствие закона распределения нормальному:

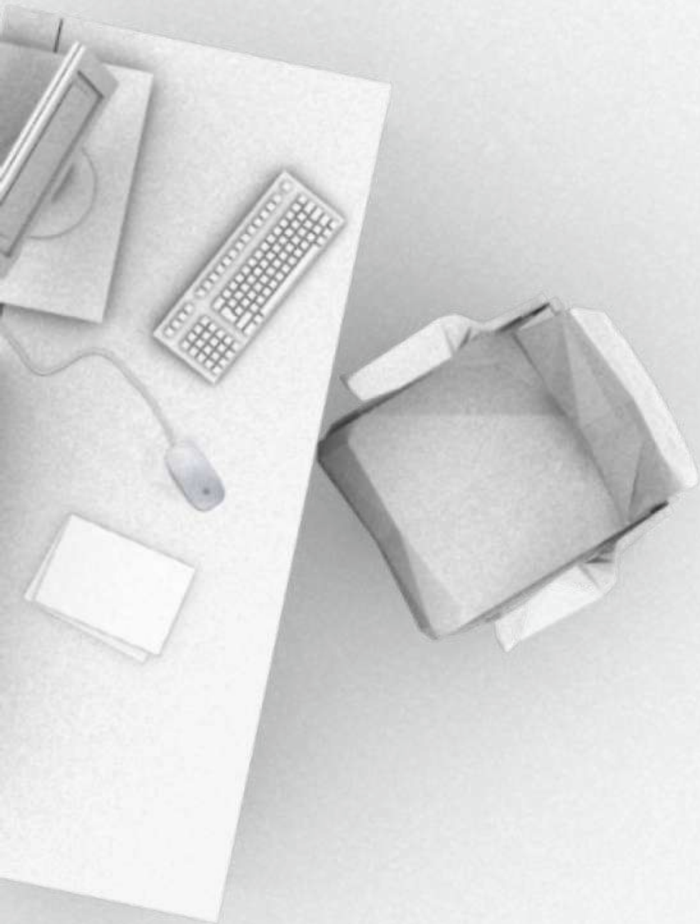
$$H0: F_n \sim N(\mu, \sigma^2)$$

- Статистика критерия:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2},$$

где a_i берутся из таблицы значений по α и i.

- Рассчитанная статистика сравнивается с критическим табличным значением $W(\alpha)$.



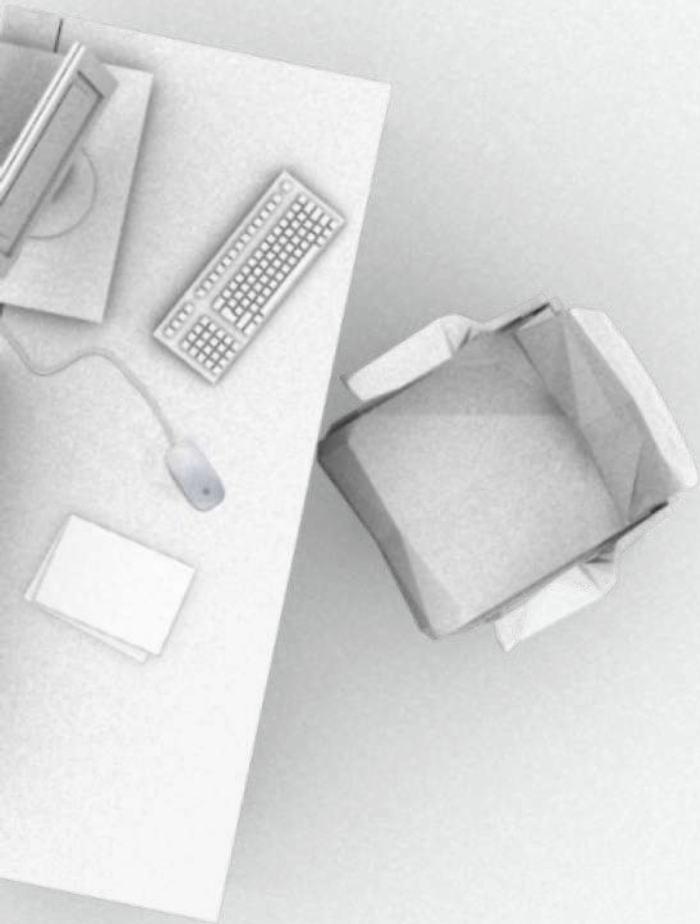
3. Критерии нормальности: Практическая часть

Критерий Колмогорова-Смирнова: scipy.stats.kstest

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

import numpy as np
import pandas as pd
from scipy.stats import kstest

alpha = 0.05

series = (np.random.normal(size=50)*100).astype(int)
st = kstest(pd.Series(series), 'norm')

print('Distributions is {}normal\n'.format( {True:'not ',
False:''}[st[1] < alpha]))
```

Критерий Шапиро: scipy.stats.shapiro

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

import numpy as np
import pandas as pd
from scipy.stats import shapiro

alpha = 0.05

series = (np.random.normal(size=50)*100).astype(int)
st = shapiro(pd.Series(series))

print('Distributions is {}normal\n'.format( {True:'not ',
False:''}[st[1] < alpha]))
```



4. Проверка статистической значимости

Ошибки первого и второго рода

H0, **H_o**, или **основная** (гипотеза о сходстве): **μ(A) == μ(B)**
H1, **H_a**, **альтернативная** (гипотеза о сходстве): **μ(A) != μ(B)**

		Верная гипотеза	
Результат применения критерия		H0	H1
	H0	H0 **верно принята** (true positive)	H0 **неверно принята** (ошибка второго рода – «пропуск цели», false positive)
	H1	H0 **неверно отвергнута** (ошибка первого рода – «ложная тревога», false negative)	H0 **верно отвергнута** (true negative)

Уровень значимости и мощность критерия на гистограмме

α-уровень - пороговый уровень статистической значимости; вероятность ошибочно отклонить нулевую гипотезу.

β-уровень - вероятность ошибочного не отклонения нулевой гипотезы об отсутствии различий.

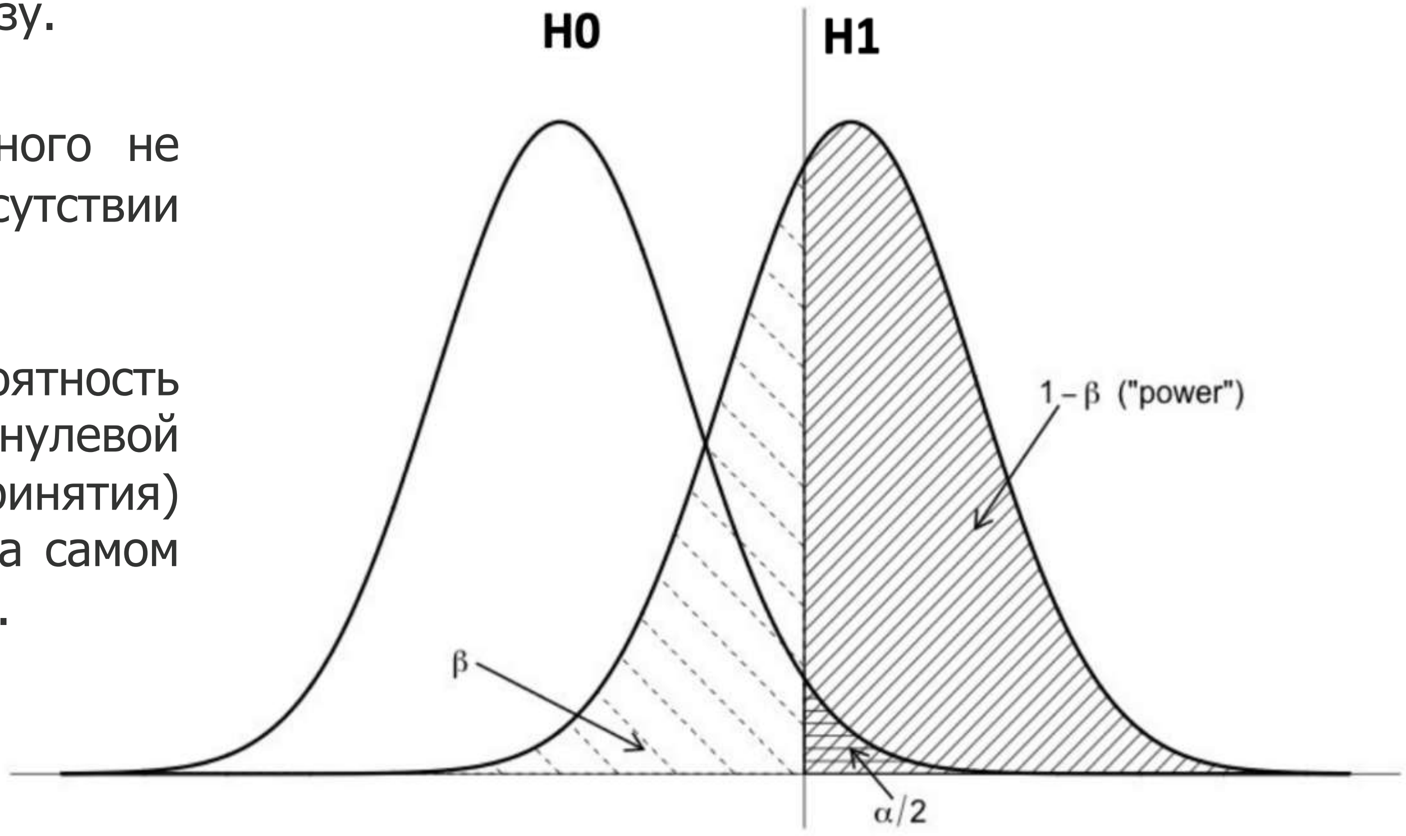
Мощность (критерия) - вероятность (правильного) отбрасывания нулевой гипотезы, т. е. отбрасывания (непринятия) нулевой гипотезы в случае, когда на самом деле верна альтернативная гипотеза.

- **H0 – основная** гипотеза (о сходстве):

  μ(A) == μ(B);

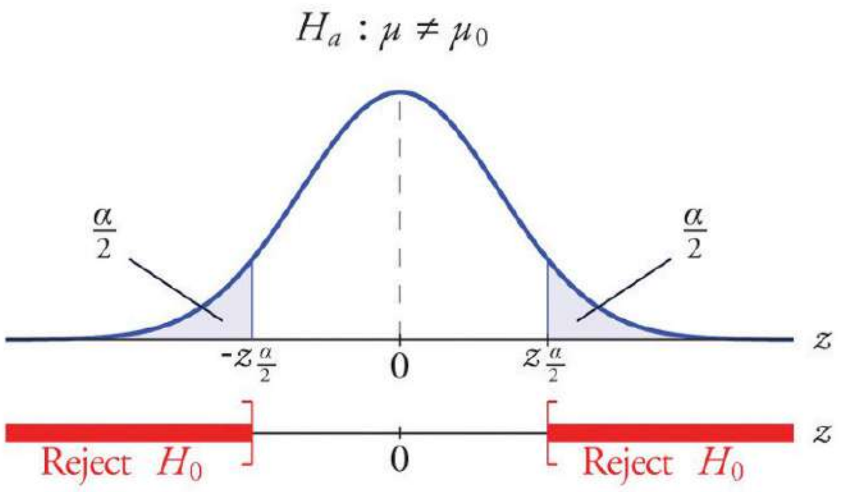
- **H1 – альтернативная** гипотеза (о различии):

  μ(A) != μ(B);

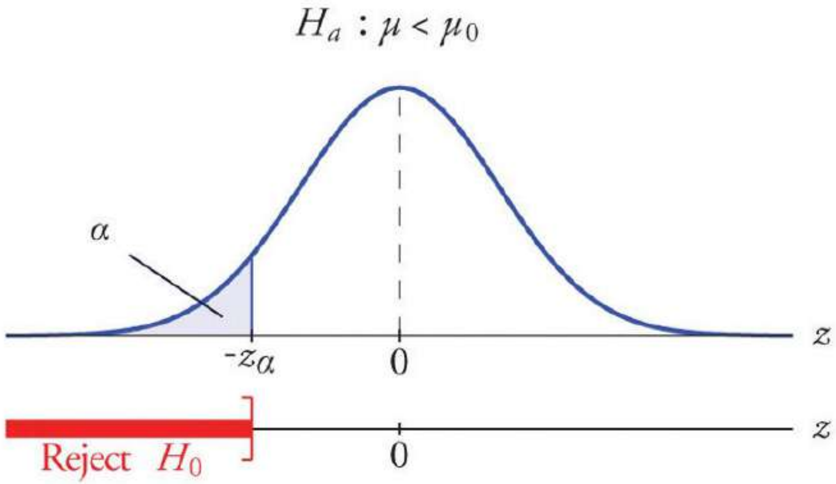


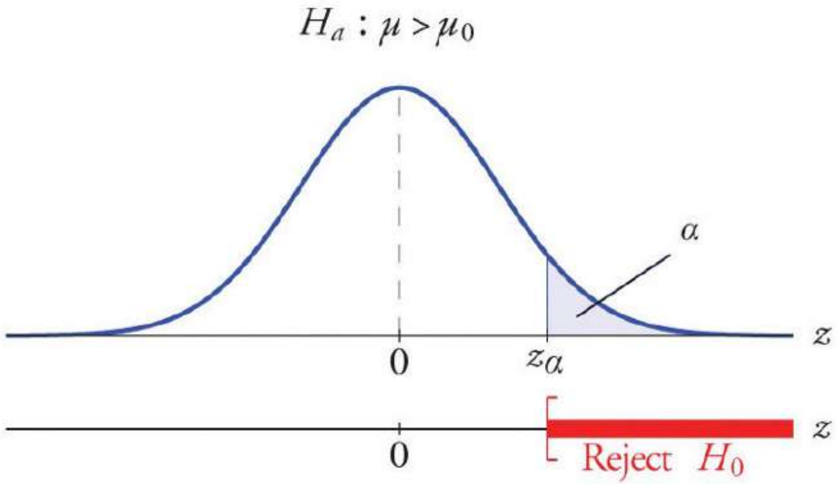
α-уровень статистической значимости

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

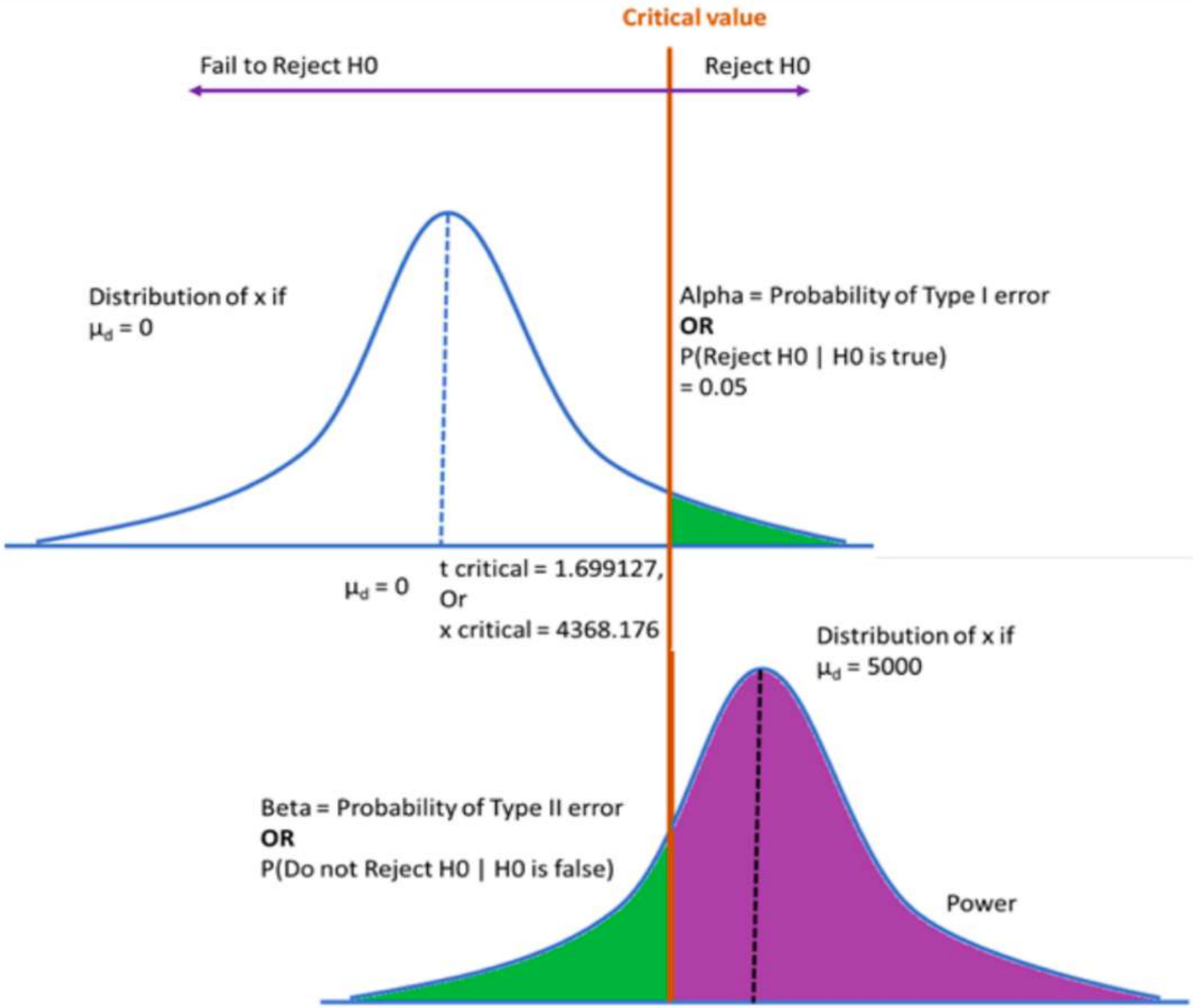


Стандартный: 0,01
Высокий: 0,05
Низкий: 0,1





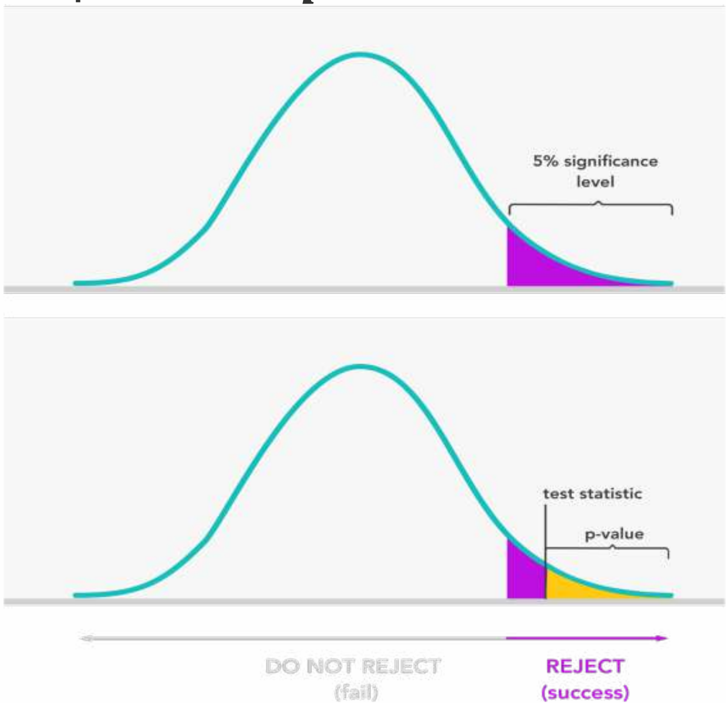
Принцип принятия и отвержения гипотезы



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

H0 отклонена: p-value < α-level

H0 принята: p-value $\geq$ α-level

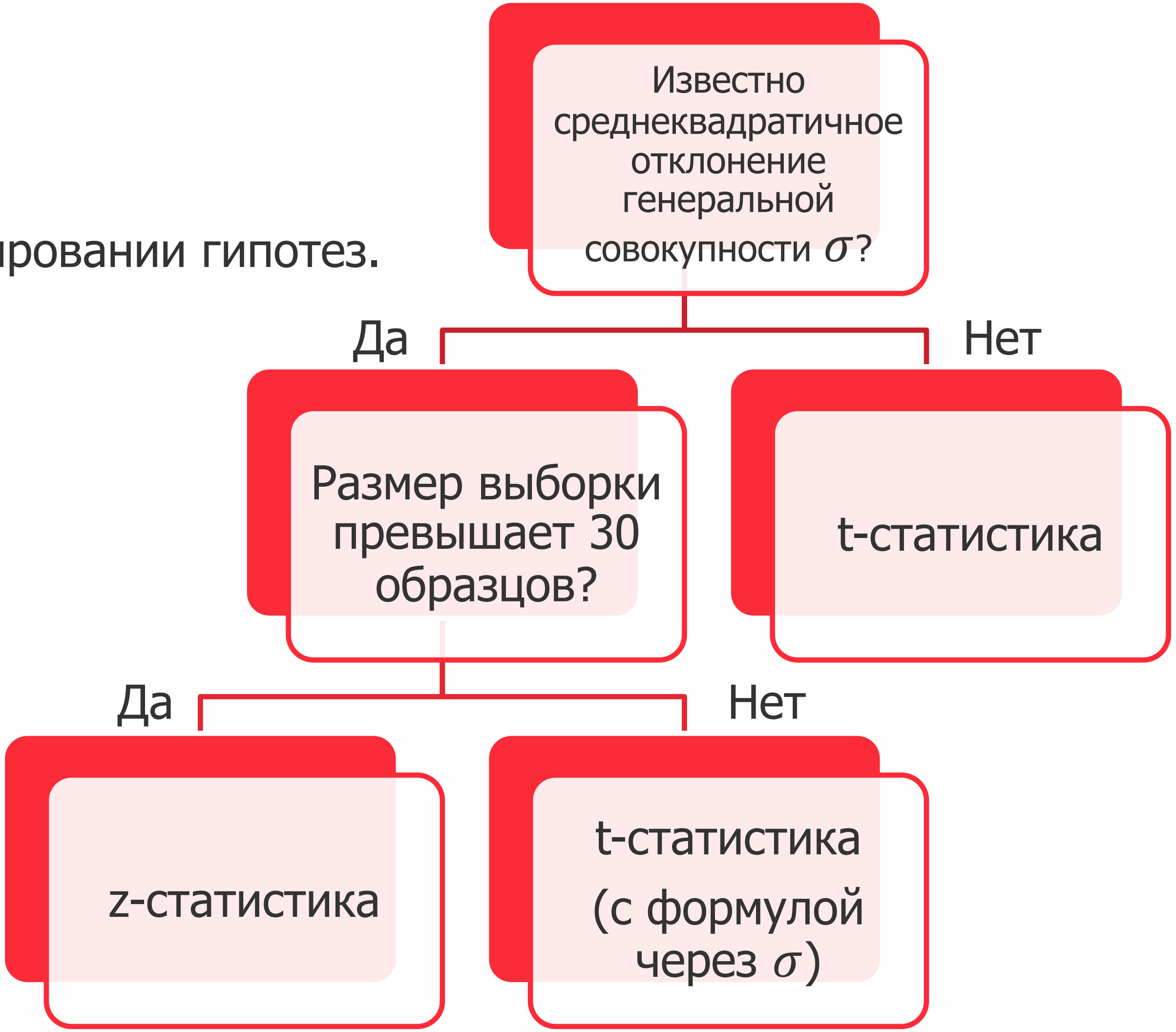


t- и z-статистики

Базовые статистики, используемые при тестировании гипотез.

Z-статистика: $z_x = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

t-статистика: $t_x = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$



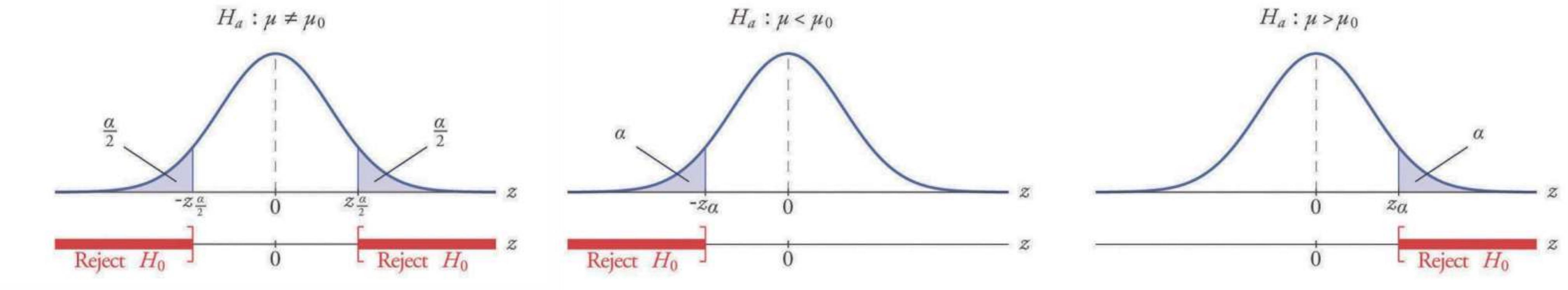
Одновыборочный Z-тест Фишера

Одновыборочный тест исследует равенство мат.ожидания выборки заданному значению мат.ожидания;

Z-статистика: $z_x = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$,

$\bar{x}$ – выборочное среднее;
μ_0 – математическое ожидание;
$\sigma/\sqrt{n}$ – стандартная ошибка среднего, где:
σ – заведомо известная величина среднеквадратичного отклонения генеральной совокупности,
n – размер выборки;

По таблице для заданного α-уровня выбирается критическое значение z, например,
при α = 0,05 при привышении $|z_x|$ > 1,96 отвергается H0 и отличие считается стат.значимой:



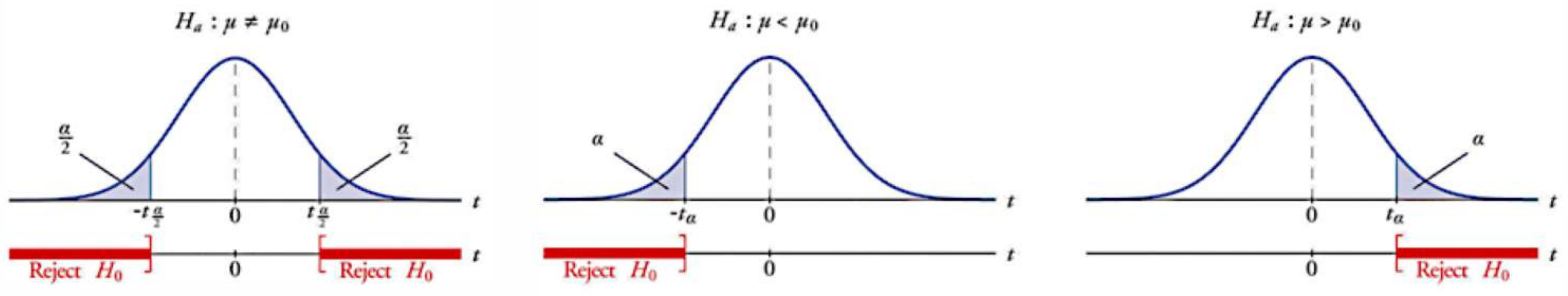
Одновыборочный критерий Стьюдента

Одновыборочный тест исследует равенство мат.ожидания выборки заданному значению мат.ожидания;

t-статистика: $t_x = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$,

$\bar{x}$ – выборочное среднее;
μ_0 – математическое ожидание;
$s_x/\sqrt{n}$ – стандартная ошибка среднего, где:
s_x – величина стандартного отклонения по выборке, $s_x =$
n – размер выборки;

По таблице для заданного α-уровня выбирается критическое значение t, при привышении которого отвергается H0:



Двухвыборочный критерий Стьюдента

t-статистика: $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$,

$\bar{X}_i$ – выборочное среднее i –й выборки;
μ_0 – математическое ожидание для генеральной совокупности;
s_i – стандартная ошибка среднего, где:
n_i – размер выборки;

По таблице для заданного α-уровня выбирается критическое значение t, при привышении которого отвергается H0

Парный критерий Стьюдента для зависимых выборок

Переход от двух групп со связанными наблюдениями к выборке попарных разностей этих наблюдений и применения к ним одновыборочного t-критерия:

t-статистика: $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \rightarrow t_x = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$,

$\bar{x}$ – выборочное среднее в выборке попарных наблюдений;

μ_0 – математическое ожидание;

$s_x/\sqrt{n}$ – стандартная ошибка среднего, где:

s_x – заведомо известная величина стандартного отклонения по выборке,

n – размер выборки;

U-критерий Манна-Уитни

- Непараметрический ранговый тест для независимых выборок;
- Нет требований нормальности и однородности дисперсий;
- Ограничения:

1. В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти.
2. В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало (до 10).

U-критерий Манна-Уитни

- Алгоритм:

1. Составить единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг r_i (при наличии повторяющихся элементов в выборке использовать средний ранг). Общее количество рангов получится равным $N = n_1 + n_2$, где n_1, n_2 – количество элементов в 1-й и 2-й выборках, соответственно;

2. Разделить единый ранжированный ряд на два, состоящие соответственно из рангов первой и второй выборок. Подсчитать отдельно суммы рангов, пришедшихся на долю элементов первой и второй выборок: $R_1 = \sum_{i=1}^{n_1} r_{1i}$ и $R_2 = \sum_{i=1}^{n_2} r_{2i}$

3. Вычислить U-статистику $U = min\{U_1, U_2\}$: $U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1,$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2,$$ где $U_1 + U_2 = n_1 \cdot n_2$

4. Полученная U-статистика сравнивается с критическим табличным значением по n_1, n_2 для заданного уровня стат.значимости, если значение больше, то H0 отклоняется.

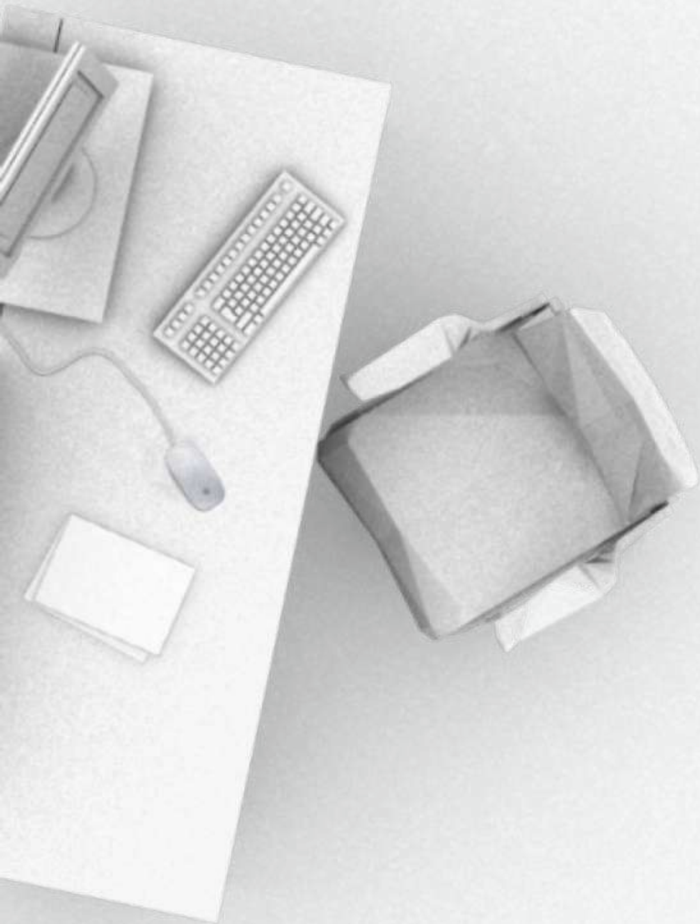
Прочие критерии

1. T-критерий Уилкоксона:

- Непараметрический ранговый критерий для проверки парных выборок;
- Нет требований нормальности и однородности дисперсий;
- Ранжируются сдвиги попарных величин, рассчитывается статистика критерия T и вычисляется граница критической области $T(\alpha)$;
- Применяется так же к сериям из двух опытов (W-статистика).

2. H-критерий Краскела-Уоллиса:

- Непараметрический ранговый критерий для проверки независимых выборок;
- Равенство медиан нескольких выборок;
- Можно сравнивать более 2-х выборок.



4. Проверка статистической значимости: практическая часть

t-критерий Стьюдента: scipy.stats.ttest_ind

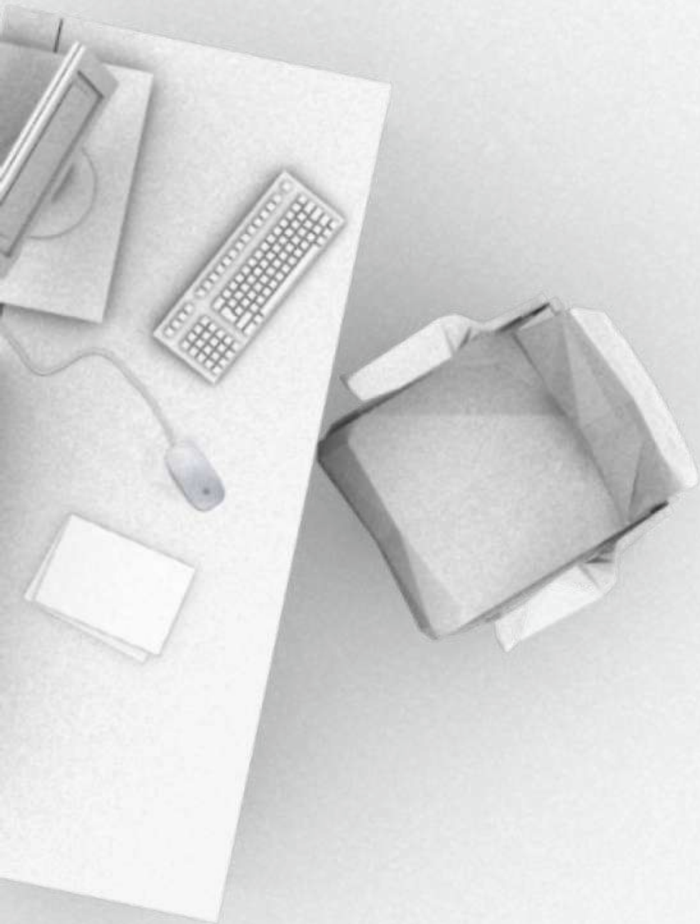
```
#!/usr/local/bin/python3
# -*- coding: utf-8 -*-

import numpy as np
import pandas as pd
from scipy.stats import ttest_ind

alpha = 0.05

series1 = (np.random.normal(size=50)*100).astype(int)
series2 = (np.random.normal(size=50)*100).astype(int)
st = ttest_ind(pd.Series(series1), pd.Series(series2))

print(‘Diffs is {}significant\n'.format( {True:'not ',
False:''}[st[1] < alpha]))
```

Парный t-критерий Стьюдента: scipy.stats.ttest_rel

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

import numpy as np
import pandas as pd
from scipy.stats import ttest_rel

alpha = 0.05

series1 = (np.random.normal(size=50)*100).astype(int)
series2 = (np.random.normal(size=50)*100).astype(int)
st = ttest_rel(pd.Series(series1), pd.Series(series2))

print('Diffs is {}significant\n'.format( {True:'not ',
False:''}[st[1] < alpha]))
```

U-критерий Манна-Уитни: scipy.stats.mannwhitneyu

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

import numpy as np
import pandas as pd
from scipy.stats import mannwhitneyu

alpha = 0.05

series1 = (np.random.gamma(size=50)*100).astype(int)
series2 = (np.random.gamma(size=50)*100).astype(int)
st = mannwhitneyu(pd.Series(series1), pd.Series(series2))

print(‘Diffs is {}significant\n'.format( {True:'not ',
False:''}[st[1] < alpha]))
```



5. Размер выборки и ожидаемый эффект

Расчёт необходимого объёма выборки

- При заданной мощности и известной функции распределения Φ (cumulative distribution function, CDF):

$$n \geq \left(\frac{z_\alpha + \Phi^{-1}(1-\beta)}{\mu\ /\sigma} \right)^2$$

- При заданных ожидаемом эффекте Δ, α-уровне и мощности:

$$n = \frac{2(Za + Z_{1-\beta})^2 \sigma^2}{\Delta^2}$$

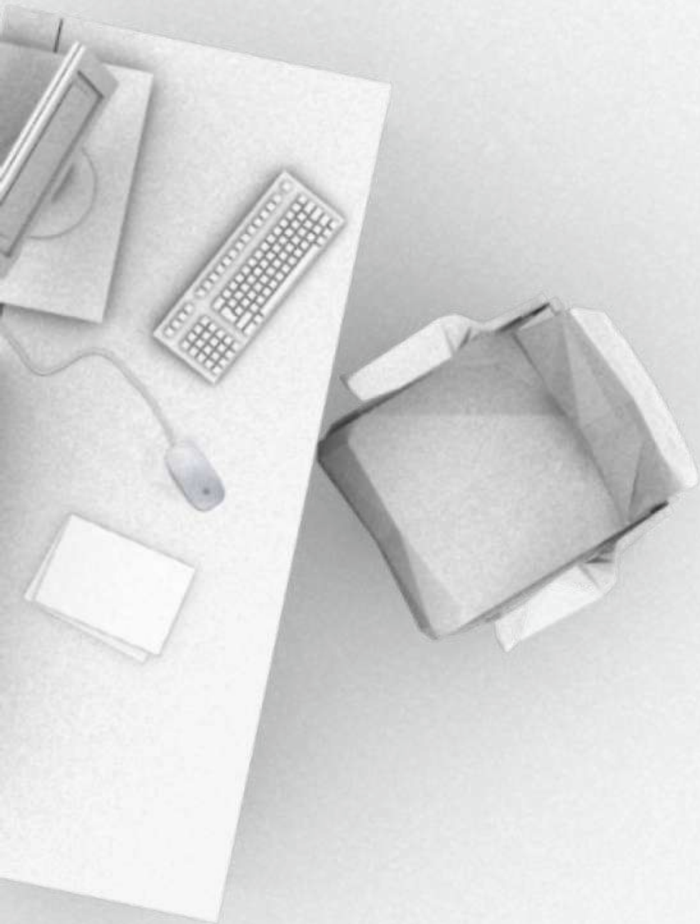
Расчёт необходимого времени проведения теста

- Обычно от 2-х недель, но что делать, если не стат.значимо?
- Оценить необходимое время для накопления данных, достаточных для получения стат.значимых различий:

  0) по опыту – набору экспериментов;

  1) аналитик принимает решение: либо по прокраске смежных метрик, либо по иным показателям, в соответствии с тем, чего добивается данным экспериментом бизнес и менеджеры, катить или откатывать, или продолжать эксперимент (и если продолжать, то сколько);

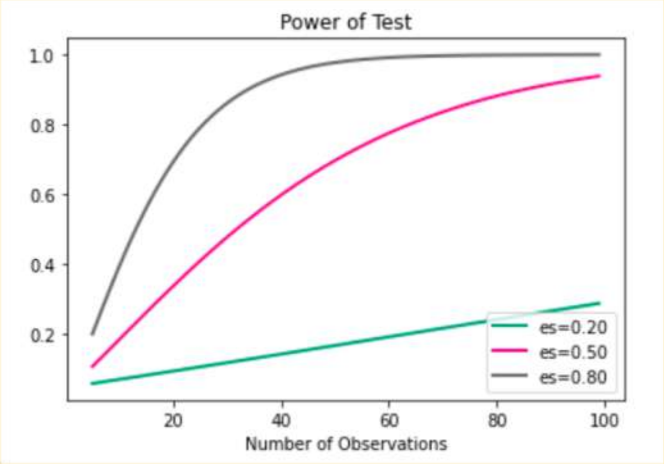
  2) посмотреть, за какой период времени накапливаются данные и дать приблизительную оценку.



5. Размер выборки и ожидаемый эффект:
Практическая часть

Расчёт необходимого объёма выборки для **t-критерия Стьюдента**

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

import matplotlib.pyplot as plt
import numpy as np
from statsmodels.stats.power import TTestIndPower

# calculate power curves
# for varying sample and effect size
analysis = TTestIndPower()
analysis.plot_power(nobs=np.arange(5, 100),
      effect_size=np.array([0.2, 0.5, 0.8]))
```



Расчёт необходимого объёма выборки для **t-критерия Стьюдента**

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

import numpy as np
from statsmodels.stats.power import TTestIndPower

# parameters for power analysis
effect = 0.08
alpha = 0.05
power = 0.8
# perform power analysis
analysis = TTestIndPower()
result = analysis.solve_power(effect, power=power,
nobs1=None, ratio=1.0, alpha=alpha)
print('Sample Size: %.3f' % result)
```

Домашнее задание

АВТ-2

#051

15 баллов

Срок сдачи:
02.12.2020

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-2. Описание case’а

1. На сайте проводится некоторый эксперимент. Время проведения эксперимента – 2 недели. Всего на сайте DAU около 500 000 и WAU около 850 000. Всего за время работы на сайте образовалась аудитория из около 1 500 000 не уходящих в отток посетителей.

2. В файле https://cloud.mail.ru – данные о результатах эксперимента. В колонке group_id вы видите bucket’ы с номерами от 1 до 16. К группе ‘A’, контрольной, относятся bucket’ы 1..8, к группе ‘B’, экспериментальной – 9..16. Колонки:
   metric_value – сумма целевых действий пользователей в рамках bucket’а – например, просмотров рекламного баннера, телефона на сайте classified-сервиса или покупок в retail/FMCG;
   users – кол-во пользователей, совершивших целевое действие, visits – кол-во целевых сессий,
   churn_users и *churn_visits* – кол-во пользователей в рамках bucket’а, отказавшихся сделать заказ, и сессий, не закончившихся успешным целевым действием.

3. Менеджера продукта интересует, какие позитивные и негативные эффекты вызвало нововведение. Продумайте метрики, которые стоит исследовать в рамках имеющихся данных для ответа на вопрос менеджера.

Домашнее задание: ABT-2

1. Загрузите данные из файла в структуру pandas.DataFrame().

2. Проведите оценку равенства дисперсий исследуемых метрик в группах и исследуйте распределение на нормальность. Какими критериями вы воспользовались и почему?

3. Оцените статистическую значимость различий исследуемых метрик в выборках. Какой критерий вы выбрали для оценки и почему?

4. Размер генеральной совокупности составляет 1 500 000. Достаточен ли размер выборок для представления достоверных выводов? Если нет, сколько ещё по времени следует проводить тест и/или на какую долю пользователей его раскатывать, чтобы результаты можно было считать достоверными?

5. Изобразите гистограммы и диаграммы размаха.

6. Результат работы пришлите в формате IPython Jupyter Notebook на сервисе Google Colab.

Ждём тебя на следующей лекции! ☺

Spoiler: normalize or die, modern algorithms of statistical analysis…