

# Гипотезы и критерии

- Статистической гипотезой—называется предположение о свойстве (каком-либо параметре, форме распределения...) генеральной совокупности, которое можно проверить, опираясь на данные выборки.
- **Статистическая гипотеза** (statistical hypothesis) — это определённое предположение о распределении вероятностей, лежащем в основе наблюдаемой [выборки данных](#).
- **Проверка статистической гипотезы** (testing statistical hypotheses) — это процесс принятия решения о том, противоречит ли рассматриваемая статистическая гипотеза наблюдаемой [выборке данных](#).
- **Статистический тест или статистический критерий** — строгое математическое правило, по которому принимается или отвергается статистическая гипотеза!

# Тестирование гипотез в статистике

- Гипотеза формулируется о свойствах генеральной совокупности.
- Формулируем ДВЕ взаимоисключающие гипотезы:
- **H0**(нулевая гипотеза, null hypothesis) – её мы собираемся опровергать; обычно говорит, что нет различий, нет эффекта, нет изменений...
- **H1**(альтернативная гипотеза, alternative hypothesis)–её мы примем, если удастся отвергнуть H0.

- **Пример1.** Нулевая гипотеза записывается следующим образом:
- $H_0: \mu_1 = \mu_2$  (нулевая гипотеза заключается в том, что генеральное среднее одной совокупности равно генеральному среднему другой совокупности).
- **Альтернативная гипотеза ( $H_1$ )** – предположение о том, что между параметрами генеральных совокупностей есть достоверные различия.
- **Пример 2.** Альтернативные гипотезы записываются следующим образом:
- $H_1: \mu_1 \neq \mu_2$  (гипотеза заключается в том, что генеральное среднее одной совокупности не равно генеральному среднему другой совокупности).
- $H_1: \mu_1 > \mu_2$  ( гипотеза заключается в том, что генеральное среднее одной совокупности больше генерального среднего другой совокупности).
- $H_1: \mu_1 < \mu_2$  (нулевая гипотеза заключается в том, что генеральное среднее одной совокупности меньше генерального среднего другой совокупности).
-

- **$H_0$ : гипотеза математические ожидания равны, вторая гипотеза  $H_1$  - не равны.**
- **$H_0$ : Средние продаж до рекламной акции и после рекламной акции равны.**
- **$H_1$  не равны, но нас интересует , что продажи после рекламной акции возросли, лучше  $H_1$  возросли.**

**Мы выбираем между статистическими гипотезами.**

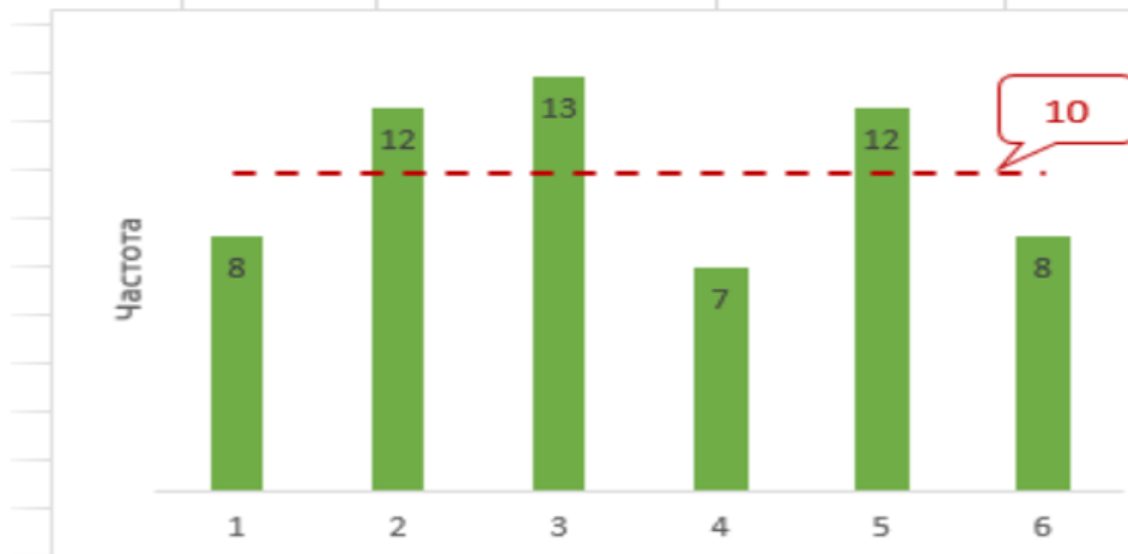
**$H_0$ - основная,  $H_1$ -альтернативная.**

- Теоретические и эмпирические частоты на примере бросания игральной кости

Очки	Частота	
	Наблюдаемая O	Ожидаемая E
1	8	10
2	12	10
3	13	10
4	7	10
5	12	10
6	8	10
Итого	60	60

Критерий согласия Пирсона  $\chi^2$  (Хи-квадрат)

значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот.



## ОШИБКИ I И II РОДА

	$H_0$ верна	$H_0$ неверна
$H_0$ принимается	$H_0$ верно принята	Ошибка II рода
$H_0$ отвергается	Ошибка I рода	$H_0$ верно отвергнута

» Ошибки I и II рода не равнозначны!

- Решение о том, принять или отвергнуть гипотезу принимается на основе **статистики критерия** (test statistic).



## *Классификация статистических критериев:*

### *По видам:*

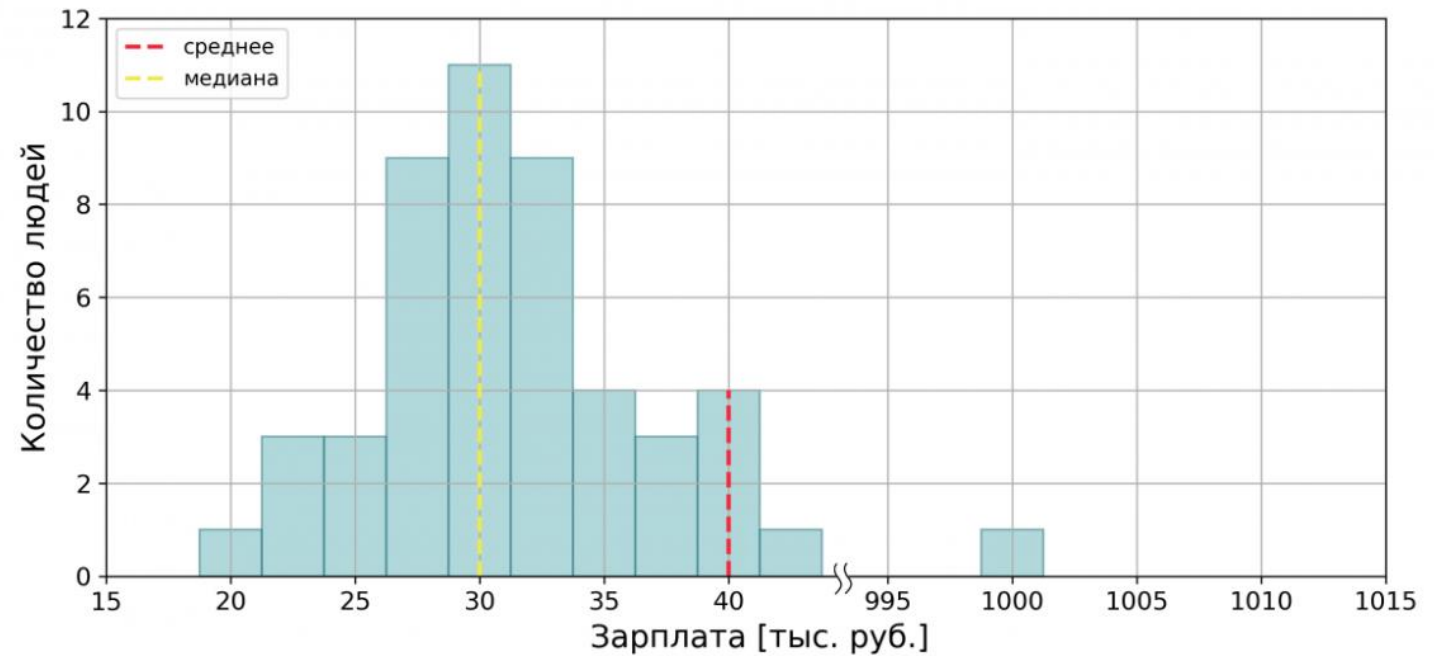
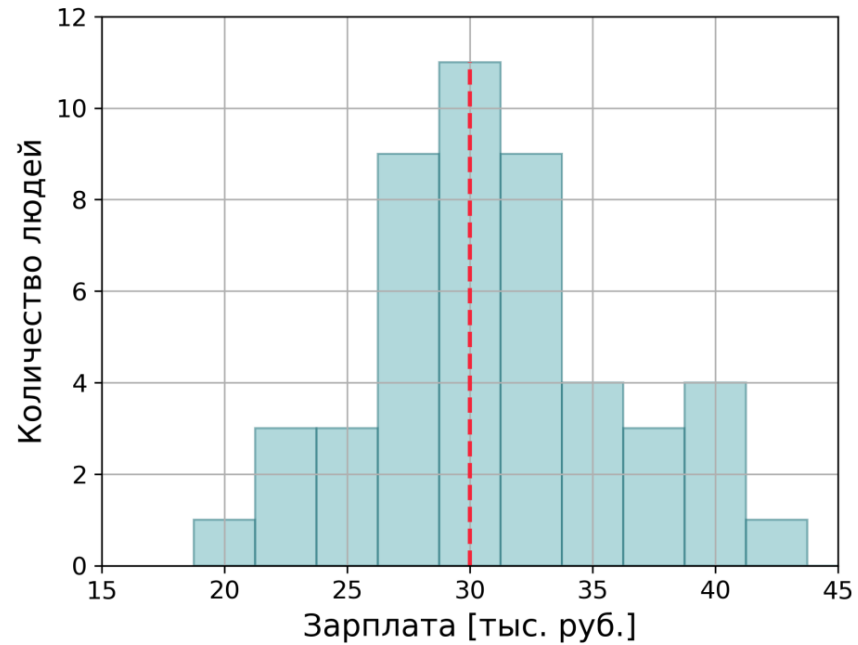
- *Критерии согласия.* Проверка предположения о том, что исследуемая случайная величина подчиняется предполагаемому закону распределения.
- *Критерии значимости.* Проверка гипотезы о численных значениях параметров известного закона распределения.
- *Критерии на однородность.* При проверке на однородность случайные величины исследуются на факт взаимного соответствия их законов распределения (подчиняются ли эти величины одному и тому же закону). Используются в факторном (дисперсионном) анализе для определения наличия зависимостей.

1. Параметрические критерии(требуют знание закона распределения);
2. Непараметрические критерии(не зависят от закона распределения).

# Сравнение центров распределений

- *Центр распределения* - то одно единственное число, которое описывало, характеризовало бы выборку.
- В качестве центра чаще всего используют среднее арифметическое или медиану.

# Среднее или медиана?



# Среднее арифметическое или медиана?

- Если распределение хотя бы одной из выборок существенно отличается от нормального, в качестве центра предлагается использовать медиану используем **критерий Манна – Уитни-Вилкоксона или (редко) Mood's median критерий** .
- В остальных случаях, то есть если распределение каждой выборки можно считать нормальным или несущественно отличающимся от нормального, в качестве центра предлагается использовать среднее арифметическое - используем одну из версий **критерия Стьюдента**.

# На практике

- Применить оба теста.  
Если выводы совпадают, ответ есть
- Если выводы различны, начинаем разбираться.

# **Гипотеза о равенстве средних двух совокупностей.**

## **$t$ – критерий однородности Стьюдента**

Достаточно распространенной статистической задачей является задача сравнения средних значений каких-либо количественных показателей. Например, необходимо сопоставить величину средней заработной платы в группах респондентов, опрошенных в разных типах населенных пунктов, либо сравнить средний возраст людей, проголосовавших за разных кандидатов на выборах, и т.п.

# Сравнение средних значений нормально распределенных выборок – параметрический t-критерий Стьюдента

- **H0**: Математические ожидания равны.

- **H1**: Математические ожидания различны.

Традиционный метод проверки однородности (критерий Стьюдента) позволяет найти вероятность того, что оба средних значения в выборках относятся к одной и той же совокупности



Классические условия применимости критерия Стьюдента. Согласно математической теории статистики должны быть выполнены два условия применимости критерия Стьюдента, основанного на использовании статистики  $t$ :

- а. результаты наблюдений имеют нормальные распределения с математическими ожиданиями  $\mu_1$  и  $\mu_2$  и дисперсиями  $\sigma_1^2$  и  $\sigma_2^2$  в первой и во второй выборках соответственно;
- б. дисперсии результатов наблюдений в первой и второй выборках совпадают  $\sigma_1^2 = \sigma_2^2$ .

# Парные и независимые выборки

## 1. Двухвыборочный t-критерий.

В случае *независимых выборок* каждое наблюдение соответствует отдельному объекту, т.е. измеряются разные объекты.

independent samples

2. **Парный t-критерий.** В случае парных выборок имеются пары наблюдений (измерений) одного и того же объекта.

dependent samples, linked samples.

# Пример расчета парного t-критерия Стьюдента

N пациента	Уровень глюкозы в крови, ммоль/л		Разность значений (d)
	до приема препарата	после приема препарата	
1	9.6	5.7	3.9
2	8.1	5.4	2.7
3	8.8	6.4	2.4
4	7.9	5.5	2.4
5	9.2	5.3	3.9
6	8.0	5.2	2.8
7	8.4	5.1	3.3
8	10.1	6.9	3.2
9	7.8	7.5	2.3
10	8.1	5.0	3.1

2. Найдем среднюю арифметическую разностей по формуле:

$$M_d = \frac{\sum d}{n} = \frac{28}{10} = 2.8$$

3. Найдем среднее квадратическое отклонение разностей от средней по формуле:

$$\sigma_d = \sqrt{\frac{\sum (M_d - d)^2}{n - 1}} = \sqrt{\frac{9.5}{9}} \approx 1.03$$

4. Рассчитаем парный t-критерий Стьюдента:

$$t = \frac{M_d}{\sigma/\sqrt{n}} = \frac{2.8}{1.03/\sqrt{10}} = 8.6$$

5. Сравним полученное значение t-критерия Стьюдента 8.6 с табличным значением, которое при числе степеней свободы *f* равном 10 - 1 = 9 и уровне значимости *p*=0.05 составляет 2.262. Так как полученное значение больше критического, делаем вывод о наличии статистически значимых различий содержания глюкозы в крови до и после приема нового препарата.

Число степеней свободы, f	Значение t-критерия Стьюдента при <i>p</i> <0,05
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262

«табличные» значения критерия Стьюдента для сверки -  
искать в математических таблицах в Приложениях в конце книг по статистике:

### Значения критерия Стьюдента

Число степеней свободы, $df$	Доверительная вероятность ( $P$ ) Уровень значимости ( $\alpha$ )		
	$P=0.095$ $\alpha=0.05$	$P=0.099$ $\alpha=0.01$	$P=0.0999$ $\alpha=0.001$
2	4.303	9.925	31.598
3	3.182	5.841	12.941
4	2.776	4.604	8.610
5	2.571	4.032	6.859
6	2.447	3.707	5.959
7	2.365	3.499	5.405
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850

22	2.074	2.819	3.792
25	2.060	2.787	3.725
30	2.042	2.750	3.646
35	2.030	2.724	3.591
40	2.021	2.704	3.551
45	2.014	2.690	3.520
50	2.008	2.678	3.496
55	2.004	2.669	3.476
60	2.000	2.660	3.460
70	1.994	2.648	3.435
80	1.989	2.638	3.416
90	1.986	2.631	3.402
100	1.982	2.625	3.390
120	1.980	2.617	3.373
>120	1.960	2.5758	3.2905

если получившееся число **больше** табличного – «нулевая» гипотеза  
**отвергается** и можно говорить об имеющихся различиях

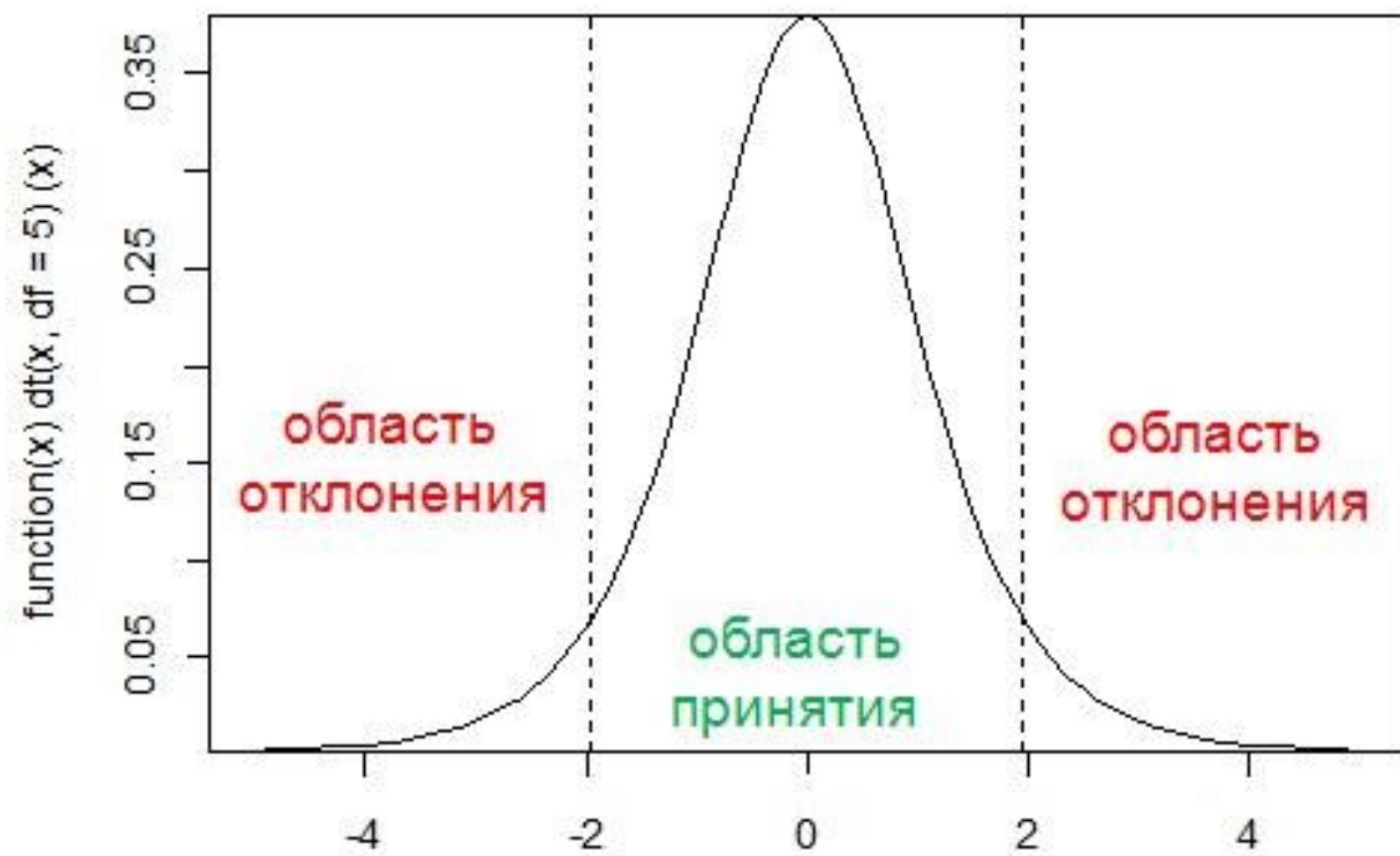
# **Критерий Стьюдента для независимых выборок равных дисперсий**

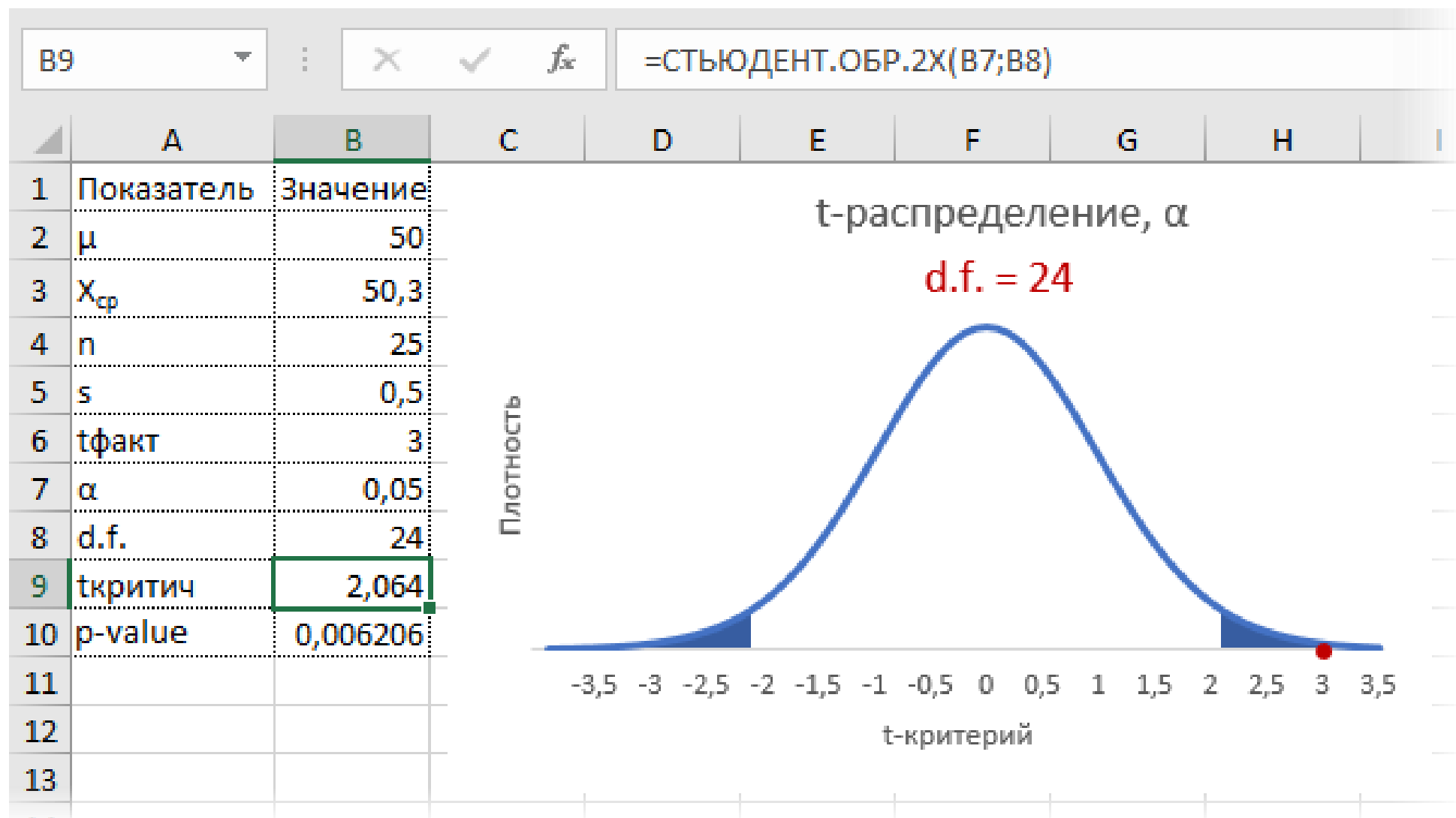


параметр	выборка $x_i$ (объемом $n_x$ )	выборка $y_i$ (объемом $n_y$ )
выборочное среднее-арифметическое	$\bar{x} = \frac{1}{n_x} \sum_i x_i$	$\bar{y} = \frac{1}{n_y} \sum_i y_i$
выборочная дисперсия	$\sigma_x^2 = \frac{1}{n_x - 1} \sum_i (x_i - \bar{x})^2$	$\sigma_y^2 = \frac{1}{n_y - 1} \sum_i (y_i - \bar{y})^2$
статистика Стьюдента $t$ , на основе которой и принимается решение по гипотезе $H_0$ :	$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n_x - 1)\sigma_x^2 + (n_y - 1)\sigma_y^2}} \sqrt{\frac{n_x n_y}{n_x + n_y} df}$	

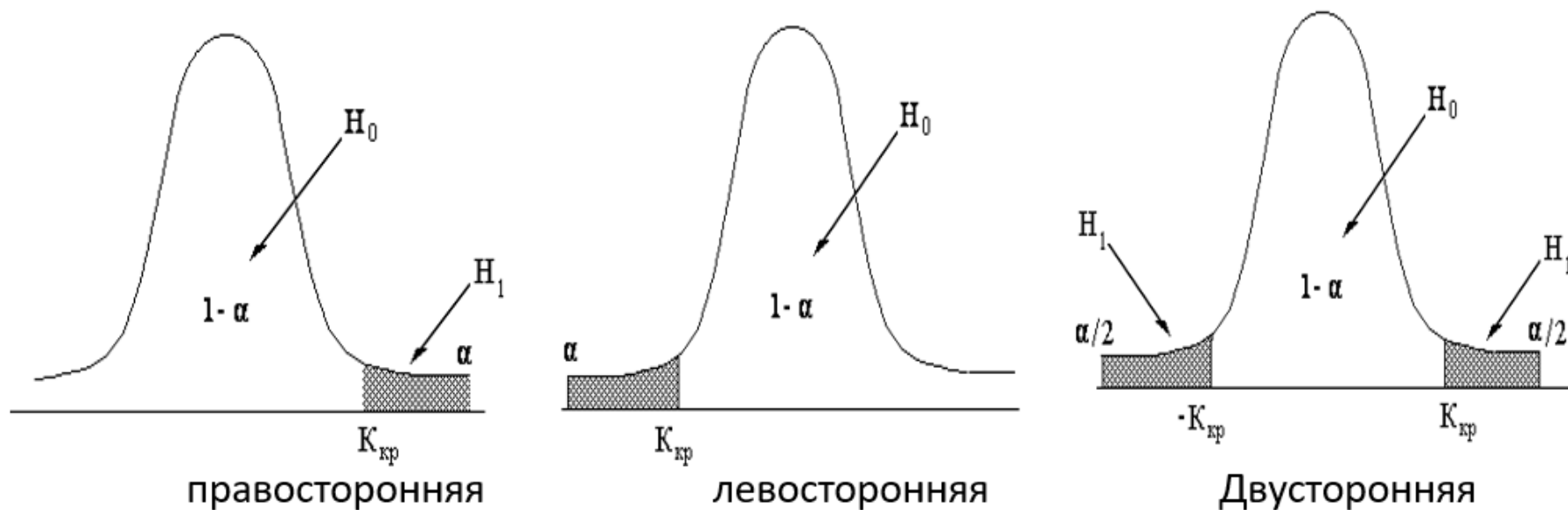
По заданному уровню значимости  $\alpha$  и числу степеней свободы  $df = (n_x + n_y - 1)$  из распределения Стьюдента находят критическое значение  $t_{\text{крит}}$ .

Если  $|t| > t_{\text{крит}}$ , то гипотезу однородности средних (отсутствия статистически значимого различия) отклоняют, если же  $|t| < t_{\text{крит}}$ , то принимают.









Критические области бывают трех видов в зависимости от вида гипотезы  $H_1$ .

Если  $H_1: \theta > \theta_0$ , то критическая область будет **правосторонней**.

Если  $H_1: \theta < \theta_0$ , то критическая область будет **левосторонней**.

Критическая точка:  $P(K_{набл} < K_{кр}) = \alpha$ .

Если  $H_1: \theta \neq \theta_0$ , то критическая область будет **двусторонней**.

## Этапы проверки гипотез

1. Формулируются нулевая и альтернативная гипотезы.
2. Задаётся некоторая функция  $T$  от выборки, для которой в условиях справедливости нулевой гипотезы  $H_0$  известна функция распределения  $F_T(x) = P(T < x)$ .
3. Фиксируется уровень значимости  $\alpha$  — допустимая для данной задачи вероятность *ошибки первого рода* (чаще всего 0.01, 0.05 или 0.1).
4. Определяется критическая область  $\Omega_\alpha$ , такая, что  $P(T \in \Omega_\alpha | H_0) = \alpha$ .
5. Проводится \_статистический тест\_: для конкретной выборки  $X$  считается значение  $T(X)$ , и если оно принадлежит  $\Omega_\alpha$ , то заключаем, что данные противоречат гипотезе  $H_0$ , и принимается гипотеза  $H_1$ .

- Если выборки парные, используется *ttest\_rel()*
- Если выборки независимые, используется *ttest\_ind()*

# Критерий Стьюдента для независимых выборок неравных дисперсий

## Критерий Беренса-Фишера

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\Omega_x + \Omega_y}}, \text{ где } \Omega = \frac{\sigma^2}{n}.$$

Известно, однако, что это распределение близко к распределению Стьюдента с числом степеней свободы, равным

$$df = \frac{(\Omega_x + \Omega_y)^2}{\frac{\Omega_x^2}{n_x - 1} + \frac{\Omega_y^2}{n_y - 1}}.$$

Проверим гипотезу о равенстве средних используя t-критерий Стьюдента.  
t-критерий для независимых выборок.

Нулевая гипотеза здесь следующая:  
математические ожидания равны.

```
from scipy.stats import *  
y = [16.8, 18, 20.3, 17.3, 18.9, 18]  
z = [18.5, 17.5, 18.5, 19.8, 17.9, 19.4]  
data=[y,z]  
ttest_ind(y, z)
```

#p-value.здесь заметно больше разумного уровня значимости  
#(например 0.05) поэтому оснований для отклонения нулевой  
#гипотезы нет: математические ожидания генеральных совокупностей равны

# Критерий Манна-Уитни - Непараметрический аналог критерия Стьюдента для парных выборок

- Mann-Whitney U-test
- Mann–Whitney–Wilcoxon,
- Wilcoxon rank-sum test,
- Wilcoxon–Mann–Whitney test

# Непараметрический U-критерий Манна-Уитни

- Критерий Манна-Уитни проверяет не равенство медиан
- Имеются две выборки наблюдений случайных величин  $X$  и  $Y$ .
- Гипотеза:  $P\{X > Y\} = P\{X < Y\}$ .
- Альтернативная гипотеза:  $P\{X > Y\} \neq P\{X < Y\}$ .

U-критерий Манна-Уитни - это непараметрический критерий статистической значимости для определения того, были ли взяты две независимые выборки из популяции с одинаковым распределением.

Для корректной работы теста требуется выполнение следующих условий:

- Распределения  $X$  и  $Y$  имеют одинаковую форму, единственным возможным отличием является их расположение (т.е. медиана);
- Число элементов в каждой выборке не менее 5 ( $\geq 3, \geq 5$ );

Данным методом определяется насколько перекрещиваются (совпадают) значения между двумя выборками (1-я выборка – ряд, в котором, по предварительной оценке значения выше, а 2-я выборка – ряд, где они предположительно ниже).

Чем меньше перекрещивающихся значений (чем меньше  $U$ ), тем более вероятно, что различия достоверны: т.е.

если  $U < U_{кр}$ , то нулевая гипотеза отвергается.



$$U = (n_x \cdot n_y) + \frac{n_*(n_* + 1)}{2} - T_*,$$

где  $n_x, n_y$  – количество вариантов в выборках  $X$  и  $Y$ ;

$T_*$  – большая из двух ранговых сумм;

$n_*$  – количество вариантов в группе с большей суммой рангов.

ФИО Испытуемого	Баллы IQ	ФИО Испытуемого	Баллы IQ
КТИ	112	БРИ	121
ВСИ	105	ДРО	120
МНИ	109	РНА	134
АНМ	90	ВРА	119
УРА	130	ГРА	115
ВФЫ	117	ДЖА	106
РКИ	117	ВЦК	107
ТРИ	125	ЮЕР	101
ТРК	134	ЖЕН	97
ТНК	109	КОР	117

ФИО Испытуемого	Баллы IQ (2.1)	Ранг (2.2)
РНА	134	(1+2)/2=1,5
ТРК		
УРА	130	3
ТРИ	125	4
БРИ	121	5
ДРО	120	6
ВРА	119	7
ВФЫ	117	(8+9+10)/3=9
РКИ		
КОР		
ГРА	115	11
КТИ	112	12
ТНК	109	(13+14)/2=13,5
МНИ		
ВЦК	107	15
ДЖА	106	16
ВСИ	105	17
ЮЕР	101	18
ЖЕН	97	19
АНМ	90	20

### Шаг 3. Суммировать ранги значений в группе 7 «а» и в группе 7 «б»

Ранги 7 «а» = 1,5+3+4+9+9+12+13,5+13,5+17+20 = 102,5

Ранги 7 «б» = 1,5+5+6+7+9+11+15+16+18+19 = 107,5

**Шаг 4. определить какая из ранговых сумм бОльшая.**

Ранг 7 «а» < Ранг 7 «б» => ранговая сумма 7"б» больше

**Шаг 5. Определить эмпирические значения критерия U Манна-Уитни по формуле:**

$$U_{emp} = (n_1 * n_2) + \frac{n_x * (n_x + 1)}{2} - T_x$$

,где  $n_1$  — количество испытуемых в 1 группе;  $n_2$  — количество испытуемых во 2 группе;  $T_x$  — большая из двух ранговых сумм;  $n_x$  — количество испытуемых в группе с бОльшей ранговой суммой.

$$U_{emp} = (10 * 10) + \frac{10 * (10 + 1)}{2} - 107,5 = 100 + 55 - 107,5 = 47,5$$

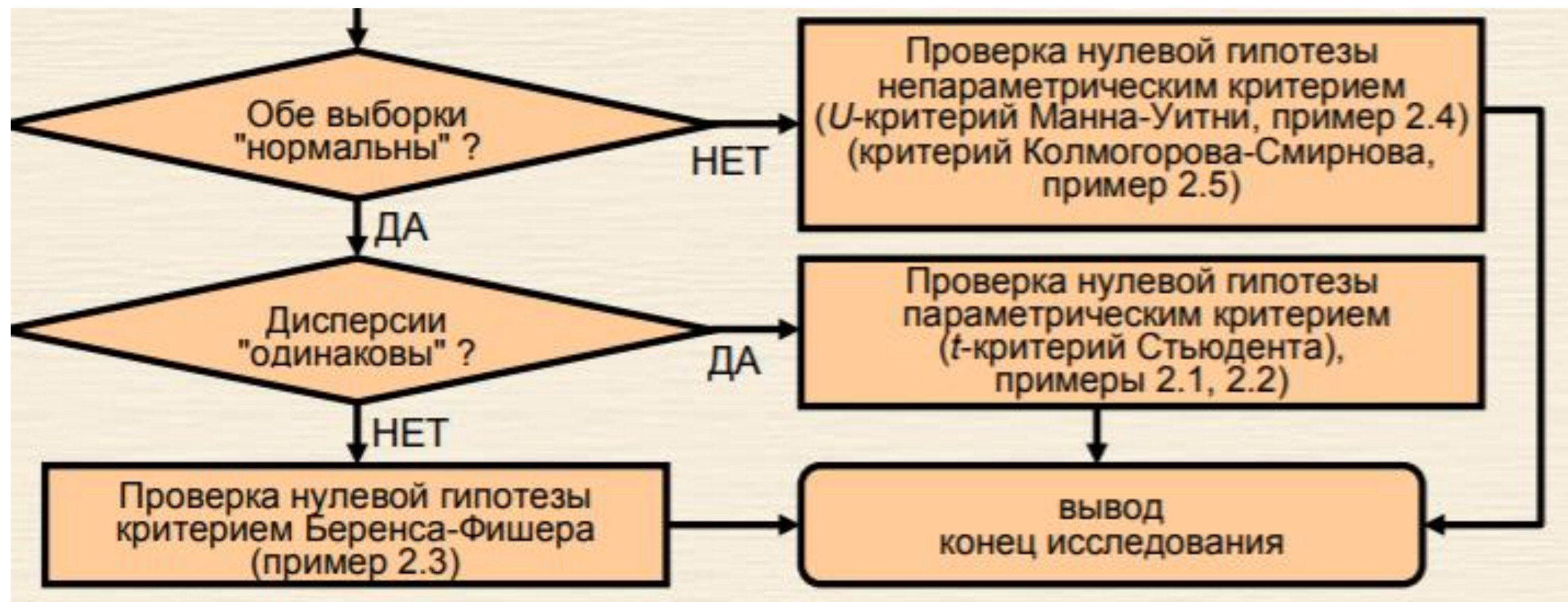
**Шаг 6. По таблице определить критические значения критерия U Манна-Уитни.**

$$U_{kr} = 23$$

**Шаг 7. Сравнить критические значения с эмпирическими.**

$$23 < 47,5 \Rightarrow U_{kr} < U_{emp}$$

**Шаг 8. Сделать выводы.**



## Параметрические критерии

### t-критерий Стьюдента

Дисперсии равны  $\sigma_1^2 = \sigma_2^2$

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s = \sqrt{\frac{df_1 \cdot s_1^2 + df_2 \cdot s_2^2}{df_1 + df_2}}$$

Сравнить с  $t_{\text{крит.}}$  для

$$df = n_1 + n_2 - 2$$

Если  $t < t_{\text{крит}}$  для  $p < 0,01$ , то гипотеза  $H_0$  верна

### t-критерий Стьюдента

Дисперсии неравны

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Найти  $df$  по формуле:

$$\frac{1}{df} = \frac{c^2}{n_1 - 1} + \frac{(1 - c)^2}{n_2 - 1}$$

Где  $c = \frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$  и сравнить

## Задание:

### 1. Вычислить пошагово Т- критерий Стьюдента для выборок

```
from scipy import stats
```

```
a = [742,148,423,424,122,432,-1,232,243,332,213]
```

```
b = [-1,3,4,2,1,3,2,4,1,2]
```

```
print (stats.ttest_ind(a,b))
```