

Доверительные интервалы

Максим Сахаров

Senior Data Scientist в BasisSoft



Максим Сахаров

О спикере:

- старший консультант по Data Science, BasisSoft
- к.т.н., доцент МГТУ им. Н.Э. Баумана
- Автор более 40 научных работ
- Области интересов:
 - Анализ данных и машинное обучение
 - Математическая оптимизация
 - Статистическое управление процессами



План урока

1

Доверительные интервалы

2

Статистическая проверка гипотез
для несвязанных выборок



План урока

1

Доверительные интервалы

2

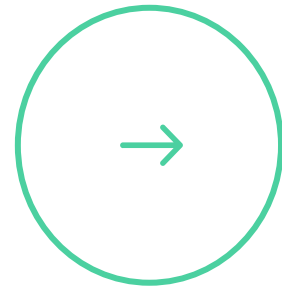
Статистическая проверка гипотез
для несвязанных выборок

Наша цель

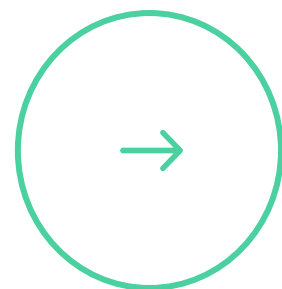
- **Доверительный интервал**
- **Уровень значимости**
- **p-value (p-значение)**
- **t-test (Критерий Стьюдента)**



Точечная оценка



Точечной оценкой называется число, которое используют для оценки параметра ГС (среднего).



Пример: Мы провели опрос/голосование с целью выявить потенциального кандидата в президенты какой-либо страны. Опрос 1000 человек показал, что за этого кандидата проголосовало 360 человек - это 36%. Это лишь показатель по малой выборке по сравнению со всей страной. Поэтому такая оценка может не все говорить о реальной картине. Что в таком случае делать?!

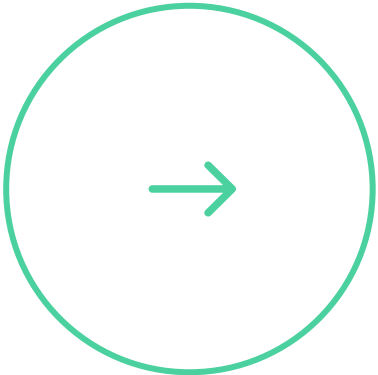
Генеральная
совокупность



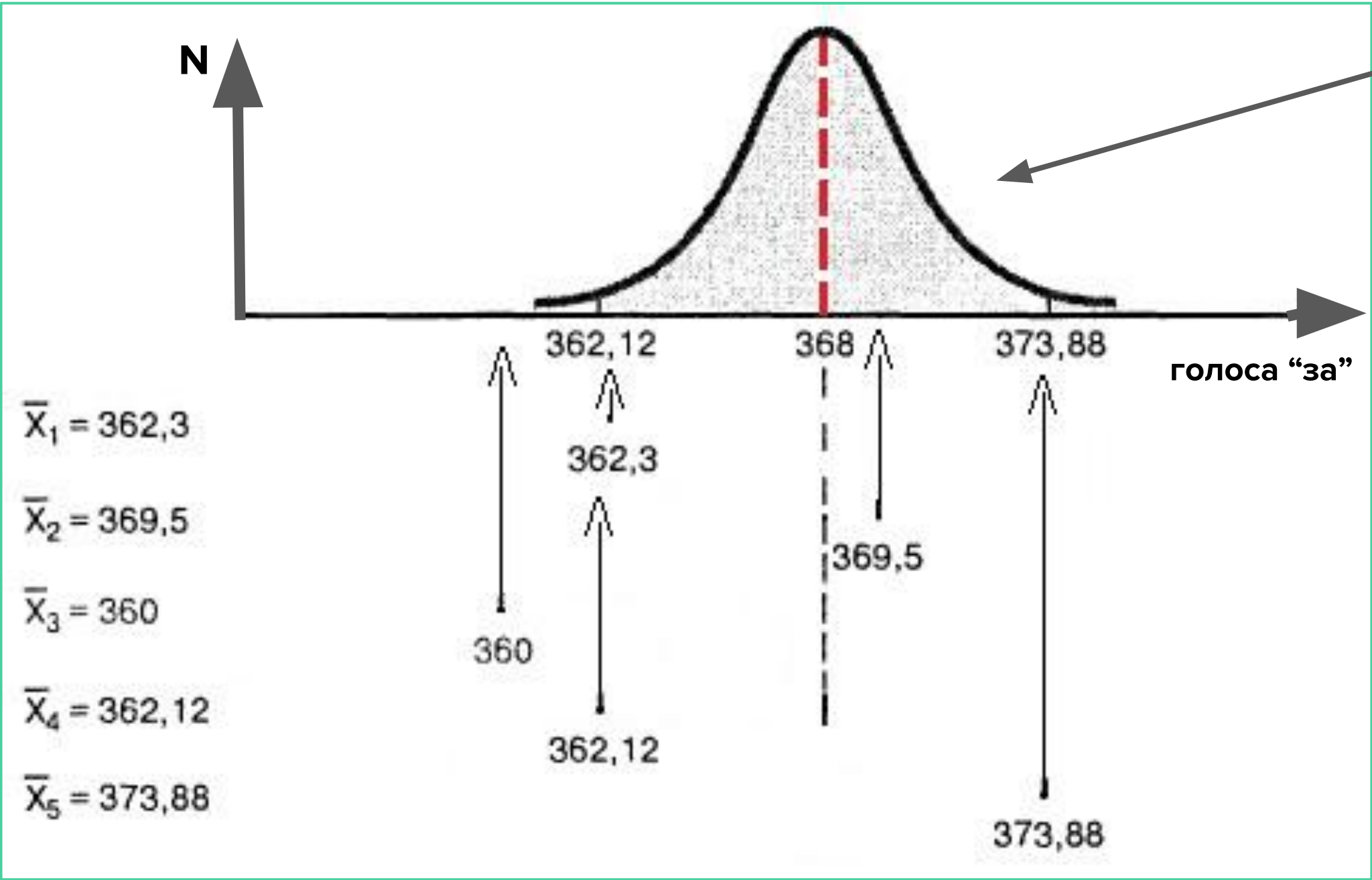
Выборочная
совокупность



Точечная оценка



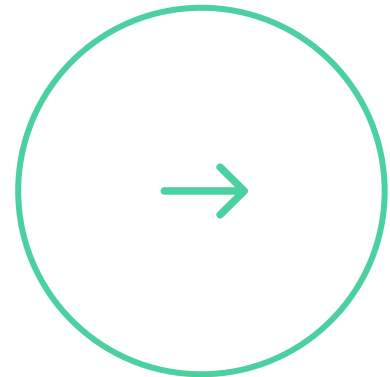
Проведем несколько таких независимых опросов на 1000 человек.



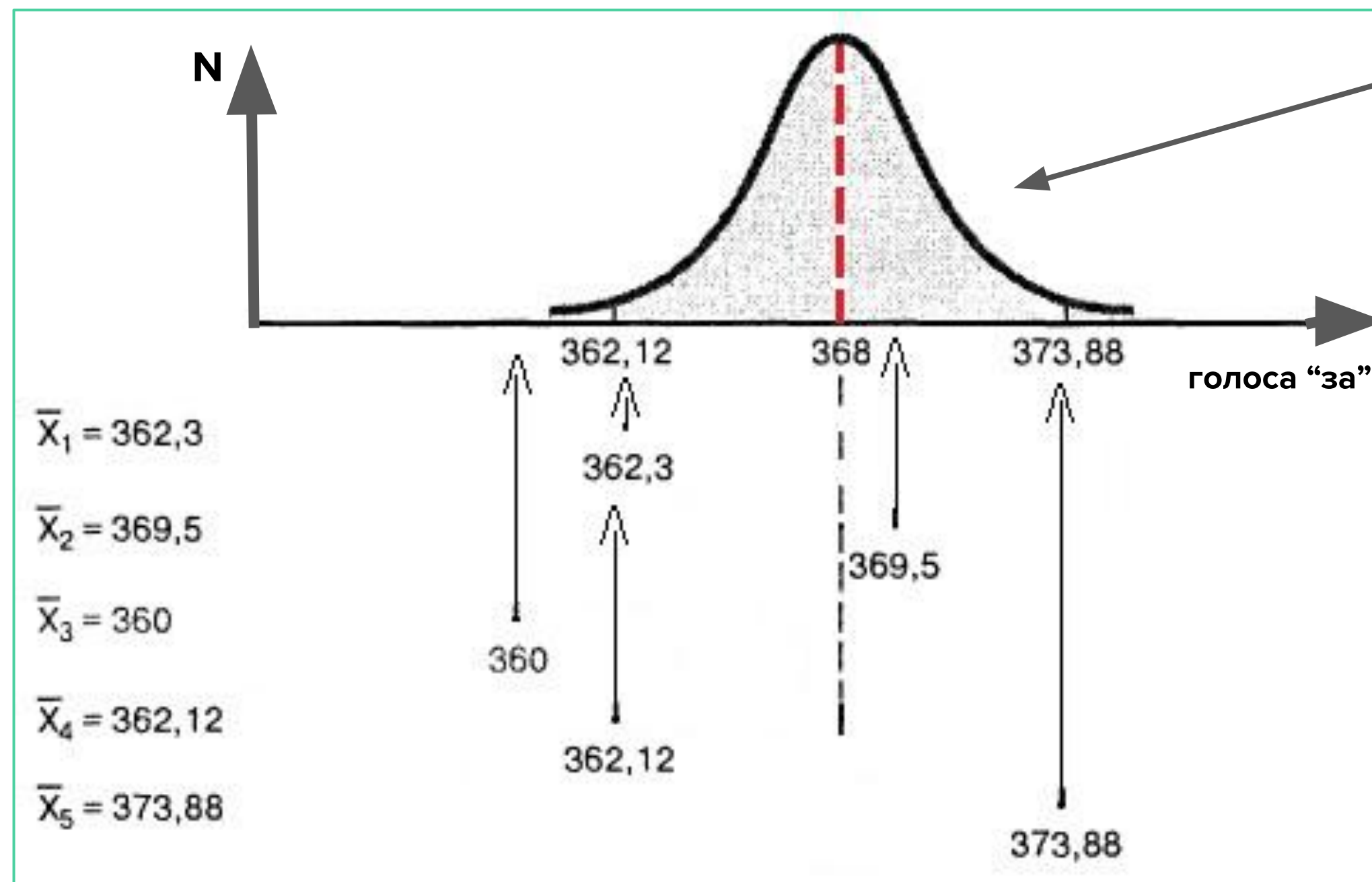
Распределение из очень большого количества разбиений генеральной совокупности на выборки



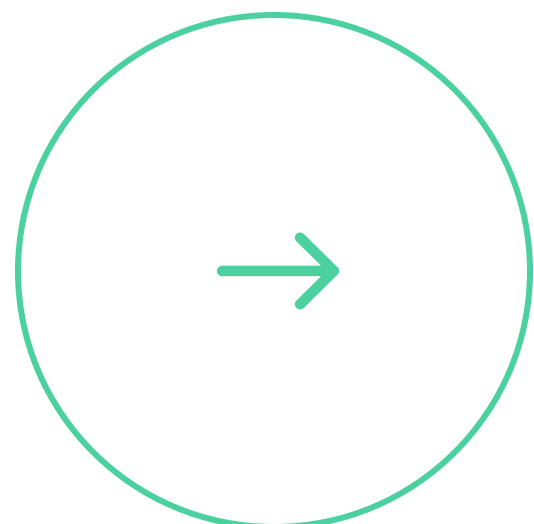
Точечная оценка



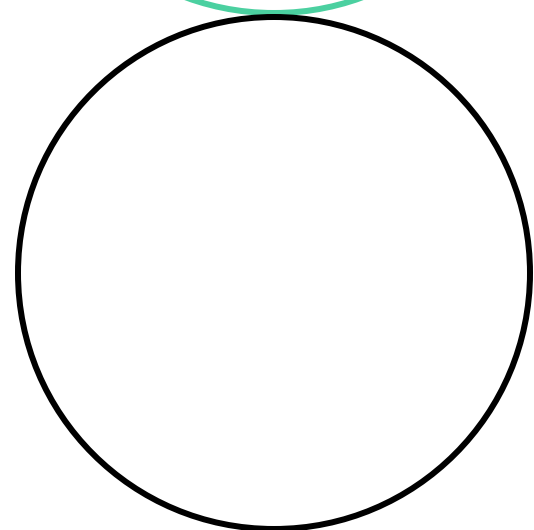
Вывод: **Точечная оценка** не так информативна, как **Доверительный интервал !!!**



Что такое доверительный интервал?

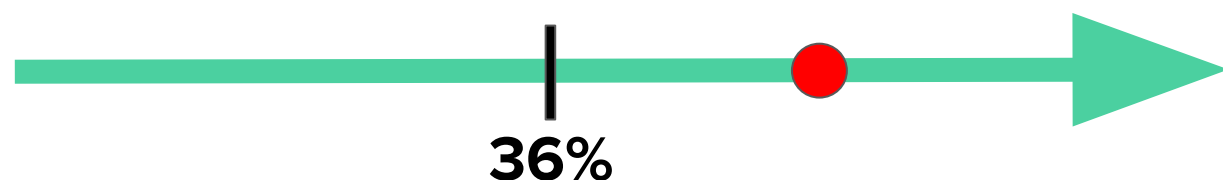


Цель: научиться оценивать параметры не просто числом, а целым интервалом, который покрывает возможные значения параметра с заданной вероятностью

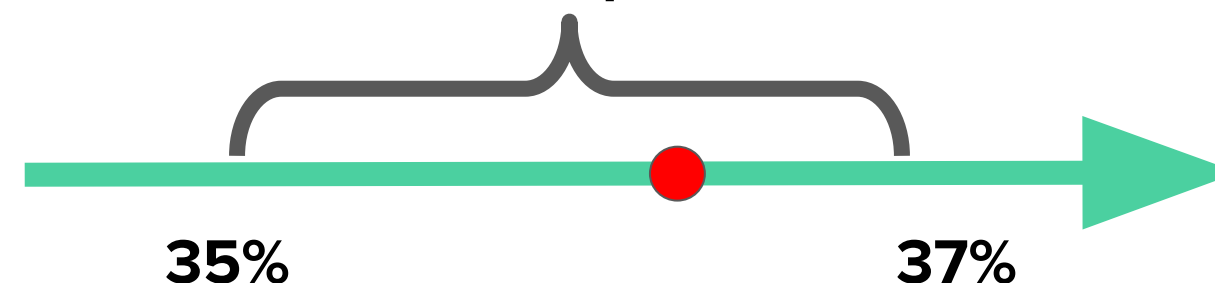


Намного лучше знать не просто оценку 36%, а интервал, в котором с большей вероятностью будет находиться реальный процент сторонников нашего президента. Скажем, с вероятностью 95% доля сторонников президента лежит в интервале от 35% до 37%

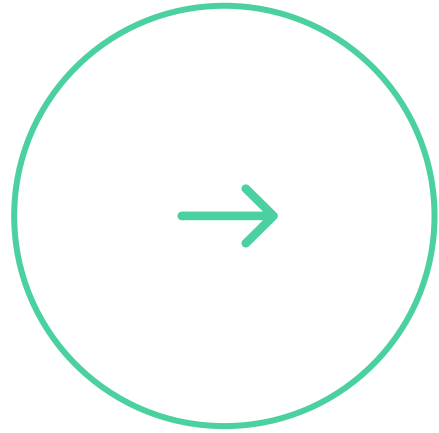
Параметр находится где-то
вокруг точки 36%



Параметр находится где-то
здесь с 95% вероятностью

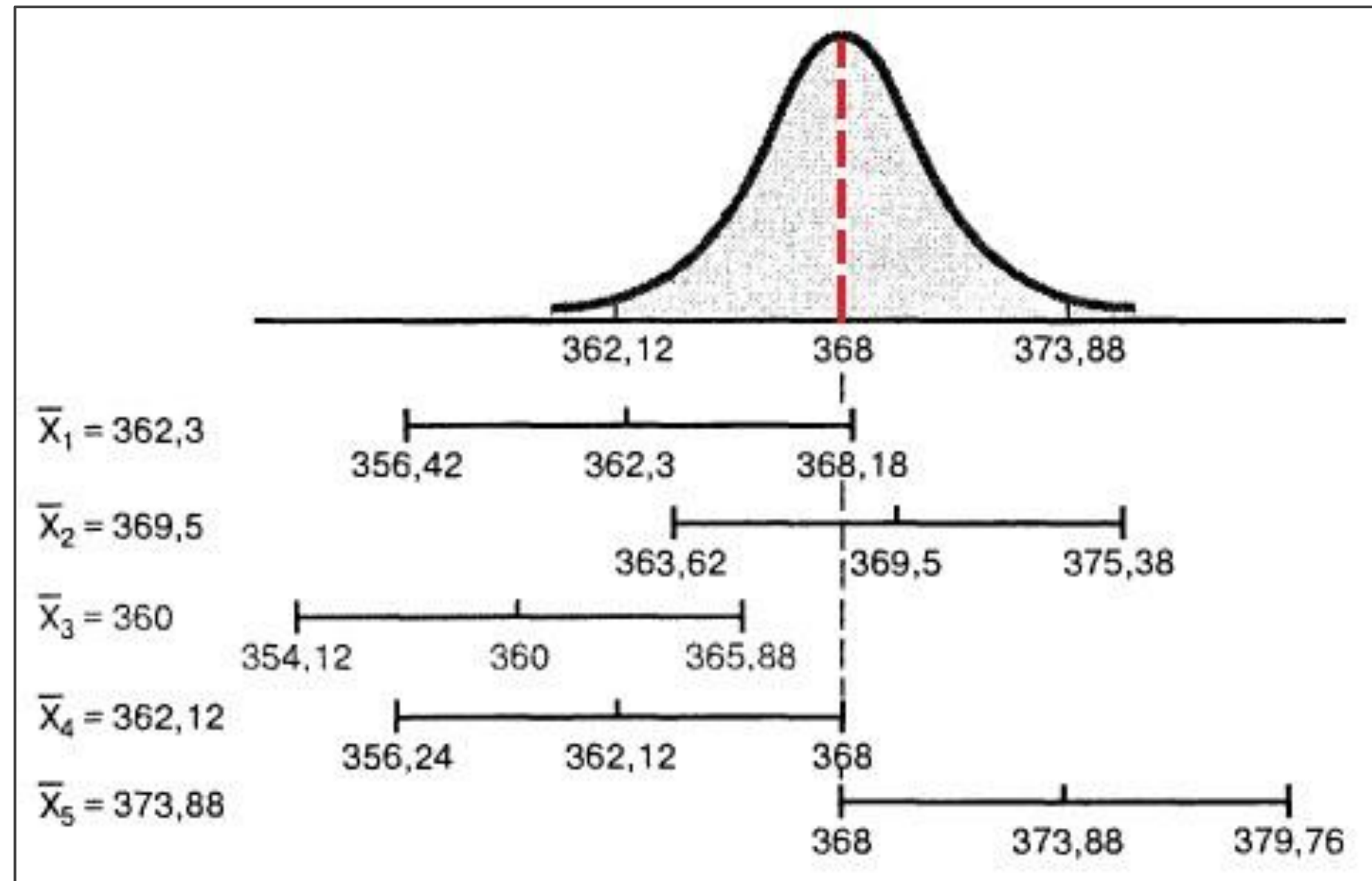


Точечная оценка VS доверительный интервал



Доверительный интервал - это интервал, который с заданной вероятностью накрывает оцениваемый нами параметр ГС.

Заметим, что **ДИ** для разных выборок одной и той же ГС могут отличаться!!!!



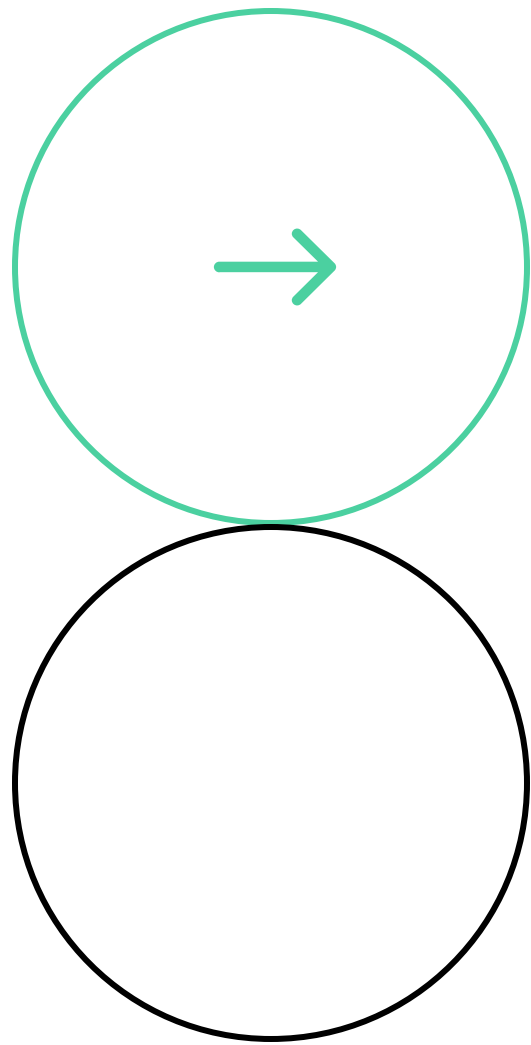
Первый шаг пройден!

Наша цель

- Доверительный интервал ✓
- Уровень значимости
- p-value (p-значение)
- t-test (Критерий Стьюдента)

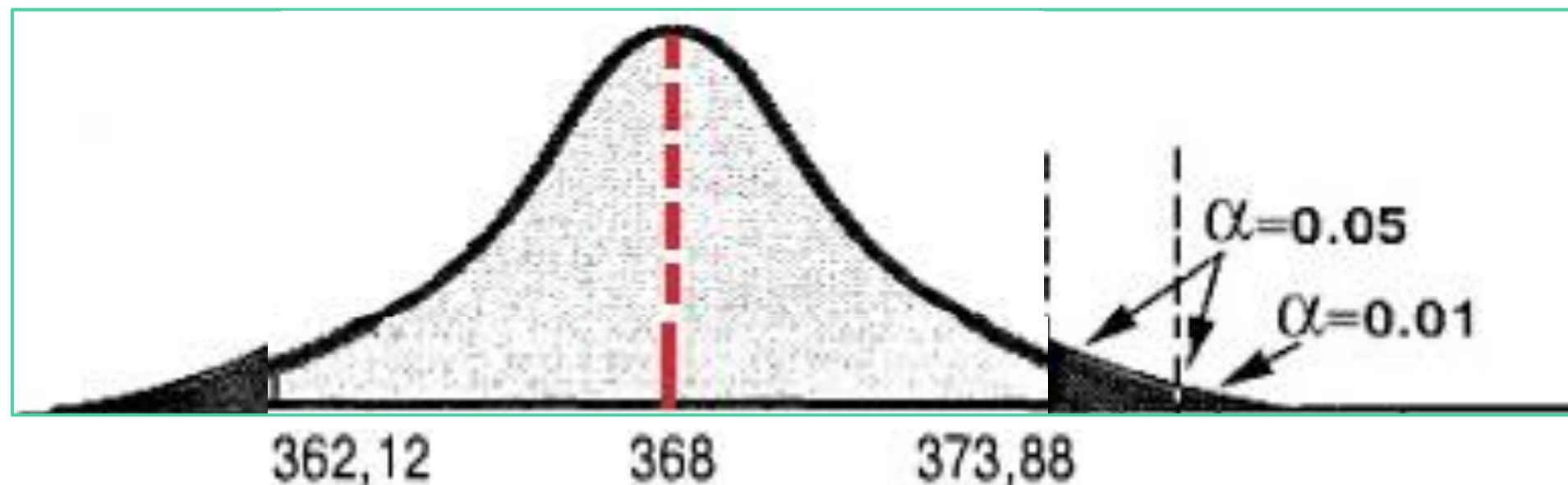


Уровень значимости и уровень доверия

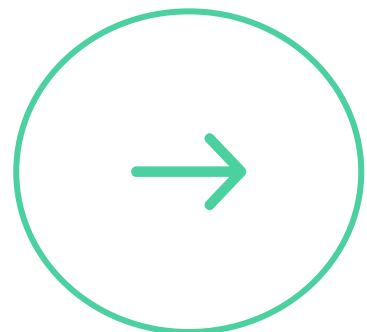


Уровень значимости α - это вероятность, с которой значение параметра не попадает в доверительный интервал.
Уровень доверия $1 - \alpha$ - это вероятность того, что доверительный интервал накрывает значение параметра

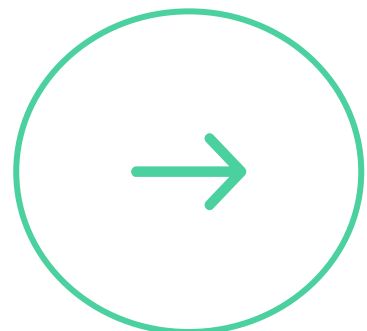
Уровень значимости 0.01 и 0.05 соответствуют уровням доверия 0.99 и 0.95 соответственно. Величины могут выражаться в процентах т.е есть уровень доверия 0.99 и 99% - это одно и то же.



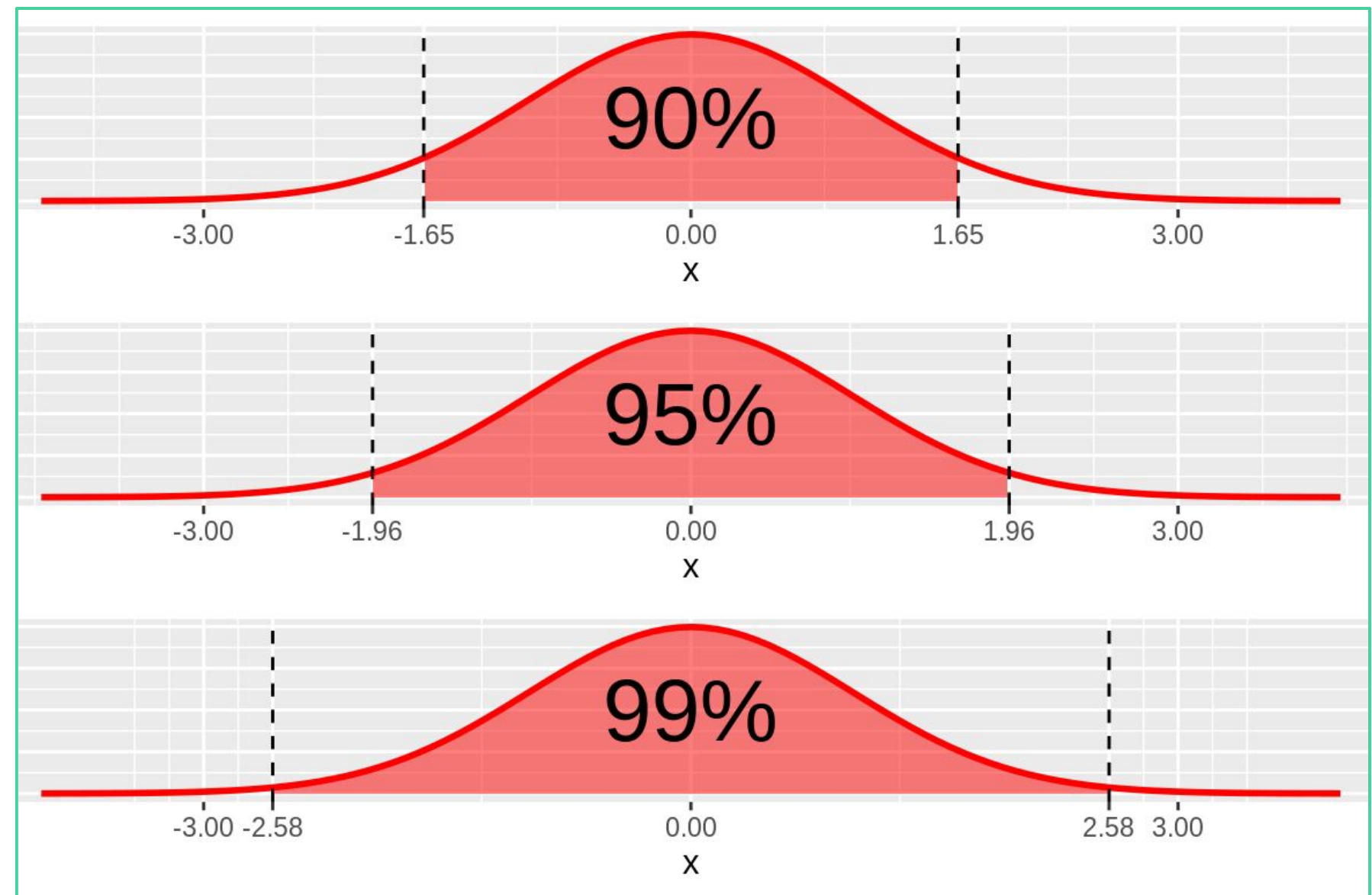
Уровень значимости и уровень доверия



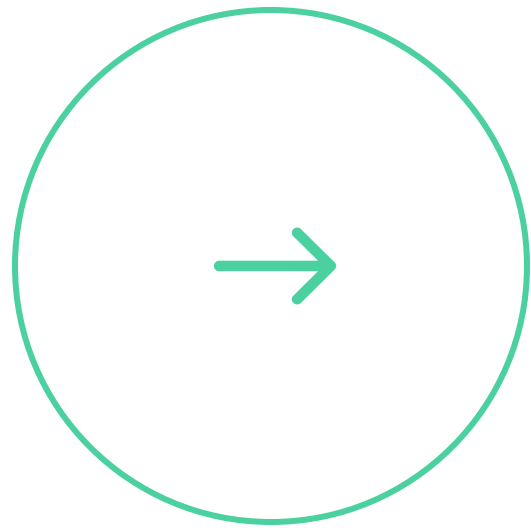
Уровень значимости 0.1, 0.05 , 0.01
соответствуют уровням доверия
0.90, 0.95 и 0.99 соответственно.



Величины могут выражаться в
процентах т.е есть уровень доверия
0.99 и 99% - это одно и то же.



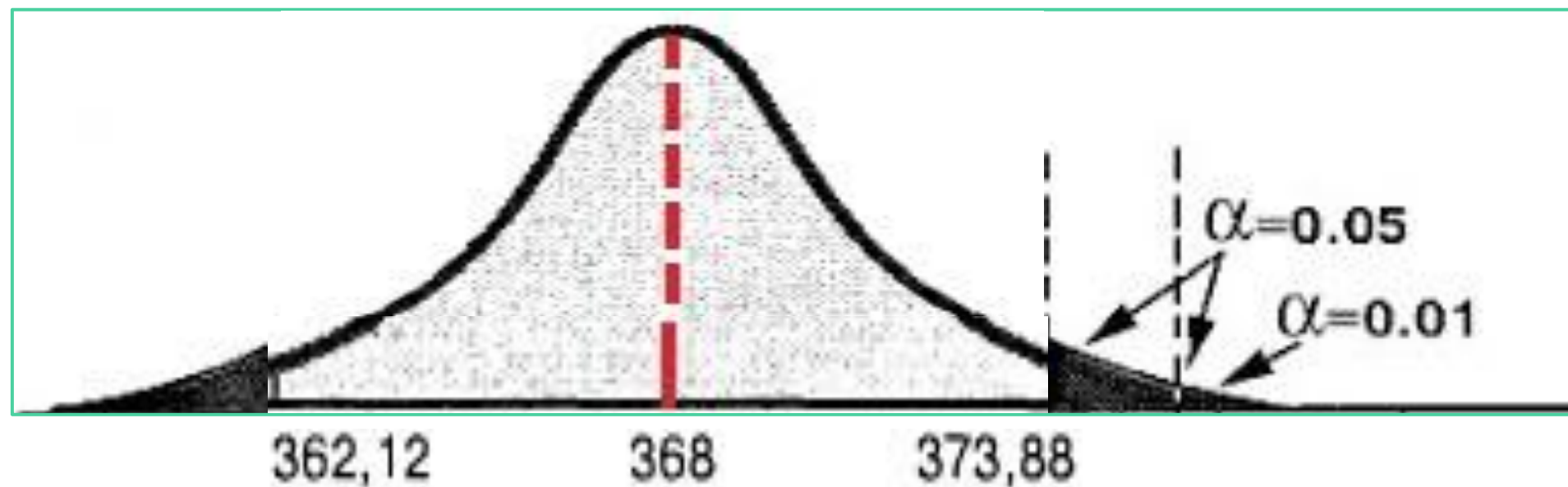
Уровень значимости и уровень доверия



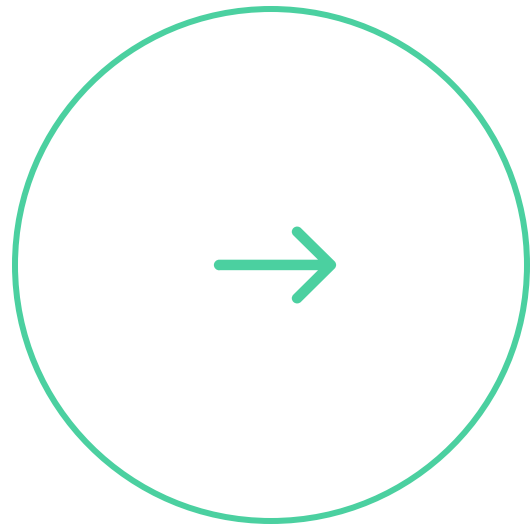
“Что-то доказано на уровне значимости 5%”

Что это значит в терминах доверительного интервала?

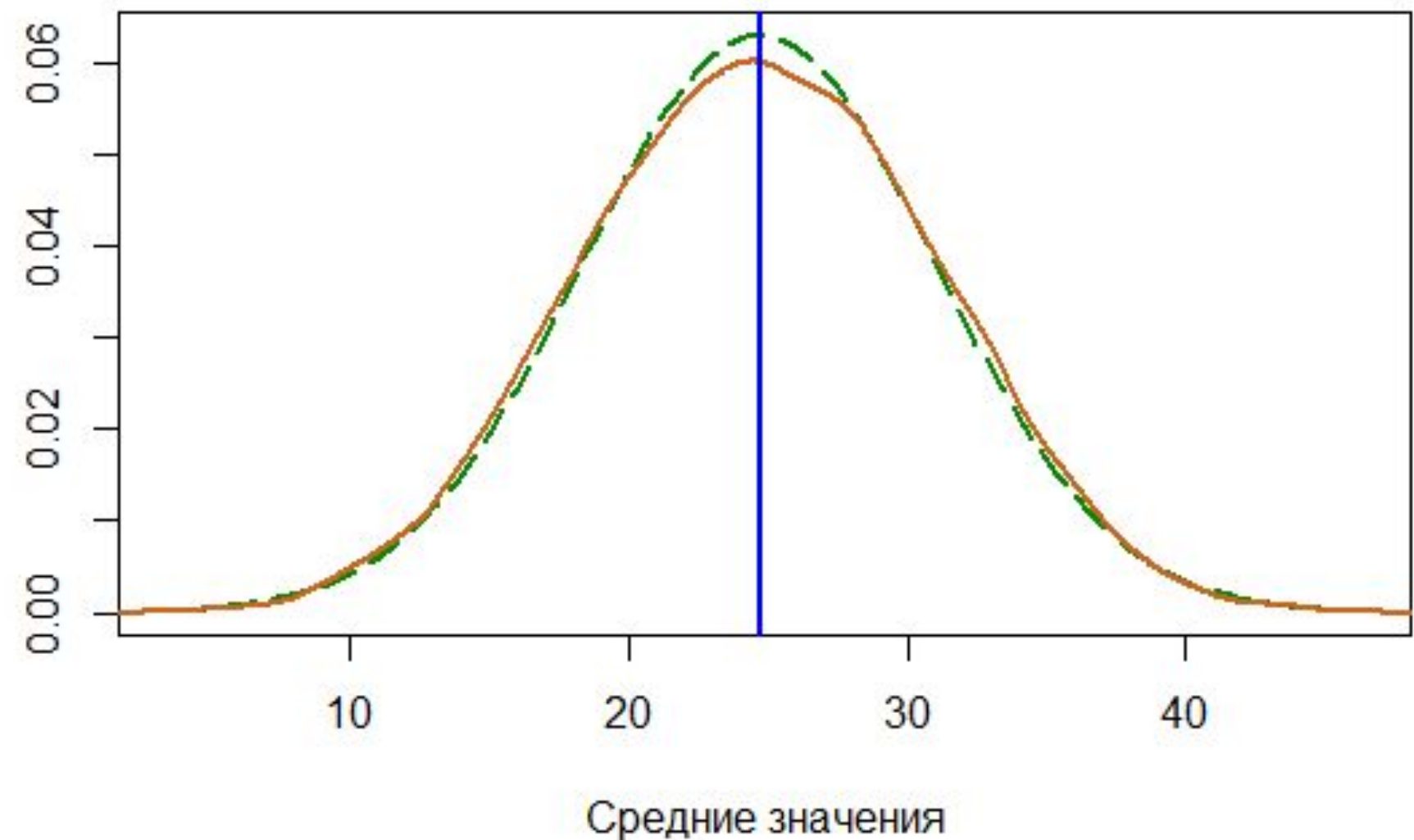
В какой области лежит наше значение?



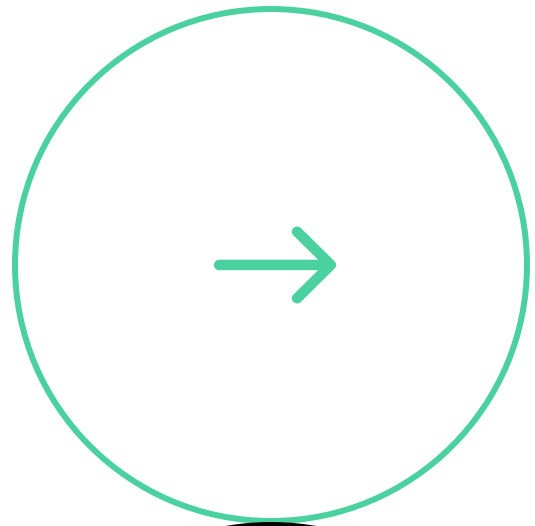
Условия для построения доверительного интервала



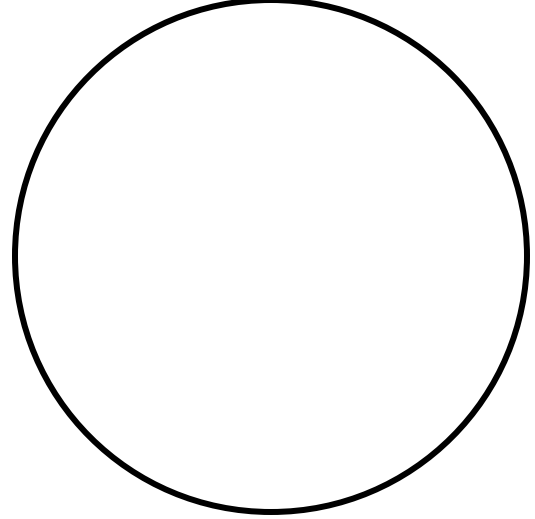
Теорема: Если распределение генеральной совокупности имеет **конечные математическое ожидание и дисперсию**, то при $n \rightarrow \infty$ основные выборочные характеристики (среднее, дисперсия, эмпирическая функция распределения) являются нормальными.
Важно! Далее мы часто будем предполагать, что генеральная совокупность имеет нормальный закон распределения.



Подсчет доверительного интервала

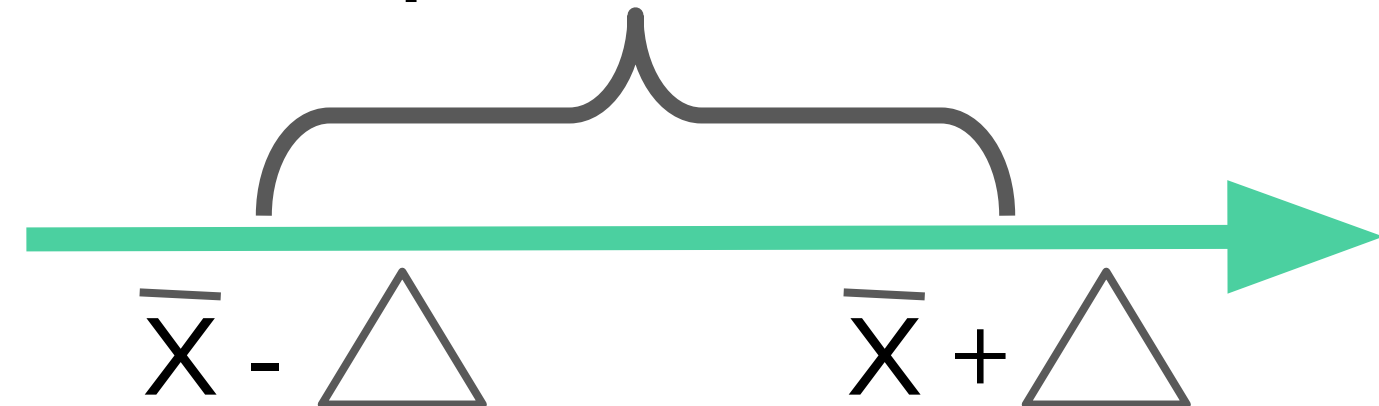


Рассмотрим случайную выборку объема n , вычислим среднее значение \bar{x} по выборке и зададим уровень значимости β .



Доверительный интервал для среднего имеет вид $(\bar{x} - \Delta; \bar{x} + \Delta)$, где Δ это точность интервальной оценки.

Параметр находится где-то
здесь с β % вероятностью



Подсчет доверительного интервала



Вычисление Δ зависит от наших знание о ГС и с какой выборкой мы имеем дело.
Допустим нам известно стандартное отклонение σ генеральной совокупности.
Тогда $\Delta = \frac{\sigma}{\sqrt{n}} Z_{\alpha}$, где Z_{α} это квантиль нормального распределения уровня $1-\alpha/2$.

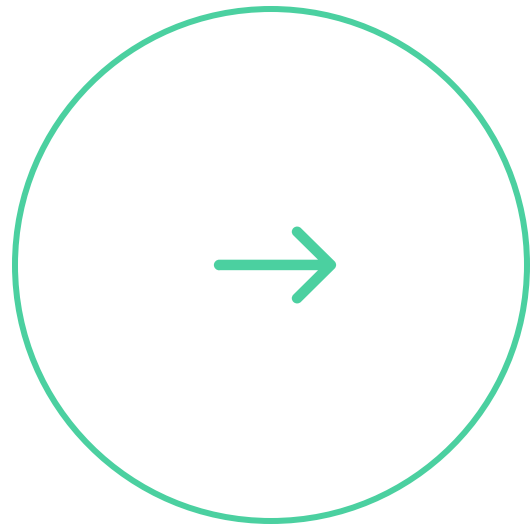
Теорема:

Доверительный интервал для среднего с известной дисперсией имеет вид:

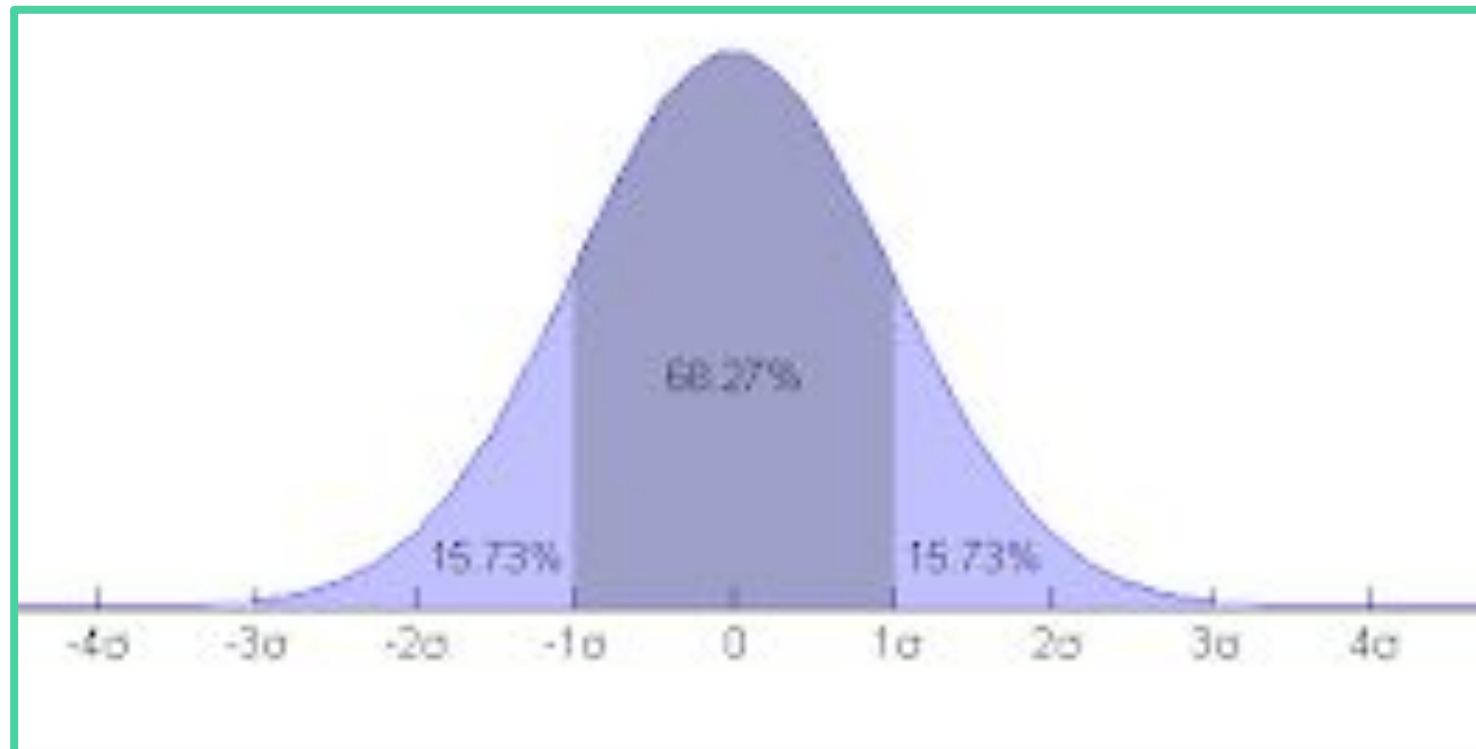
$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha}; \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha} \right)$$



Подсчет доверительного интервала



Вычисление Δ зависит от наших знаний о ГС и с какой выборкой мы имеем дело. Допустим нам известно стандартное отклонение σ генеральной совокупности. Тогда $\Delta = \frac{\sigma}{\sqrt{n}} z_{\alpha}$, где Z_{α} это квантиль нормального распределения уровня $1-\alpha/2$.



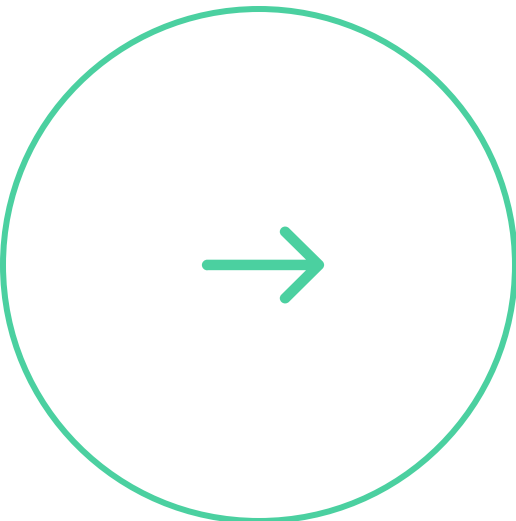
Что такое квантиль?!?!?!?

Квантиль — значение, которое наша величина не превышает с фиксированной **вероятностью**. Если вероятность задана в процентах, то квантиль называется **процентилем** или **перцентилем**.

Например, фраза «80-й процентиль времени выхода из строя станка на заводе 3 года» означает, что 80 % станков работает меньше либо равно 3 года, а 20 % станков все-таки работает больше 3 лет..



Подсчет доверительного интервала



Вычисление Δ зависит от наших знание о ГС и с какой выборкой мы имеем дело.
Допустим нам известно стандартное отклонение σ генеральной совокупности.
Тогда $\Delta = \frac{\sigma}{\sqrt{n}} z_{\alpha}$, где Z_{α} это квантиль нормального распределения уровня $1-\alpha/2$.

А где найти этот квантиль?!?!?!?

Надо взять из таблички!
(легко “гуглится”)

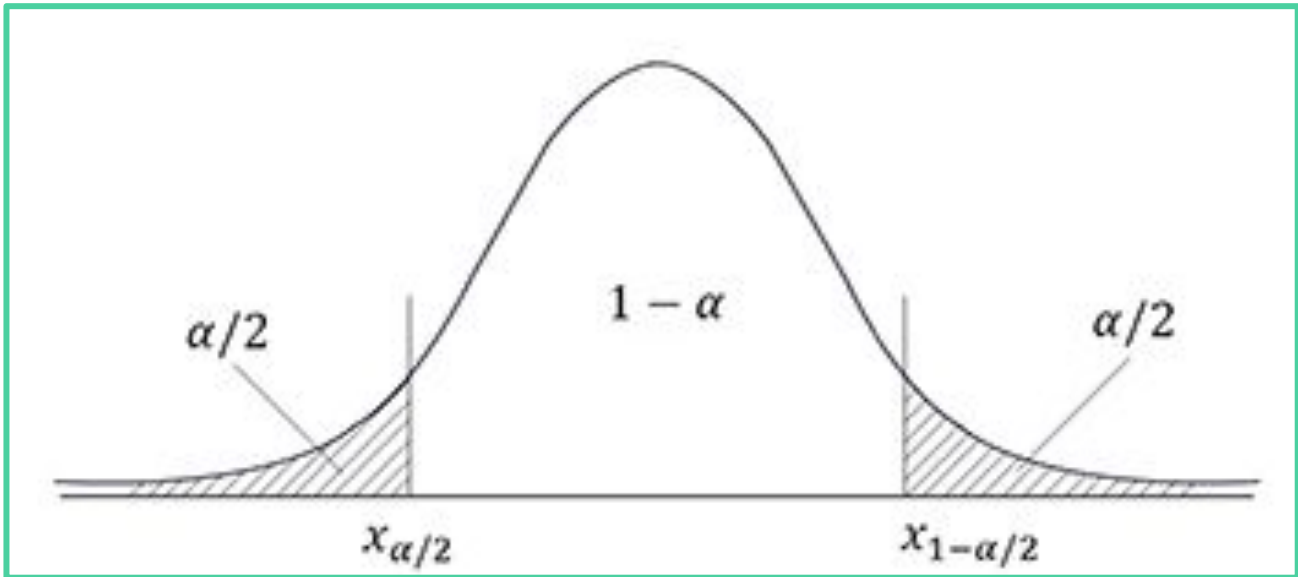
уровень	0,90	0,95	0,975	0,98	0,99	0,995	0,999	0,9995	0,9999
односторонняя	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	3,719
двусторонняя	1,645	1,960	2,241	2,326	2,576	2,807	3,291	3,481	3,891



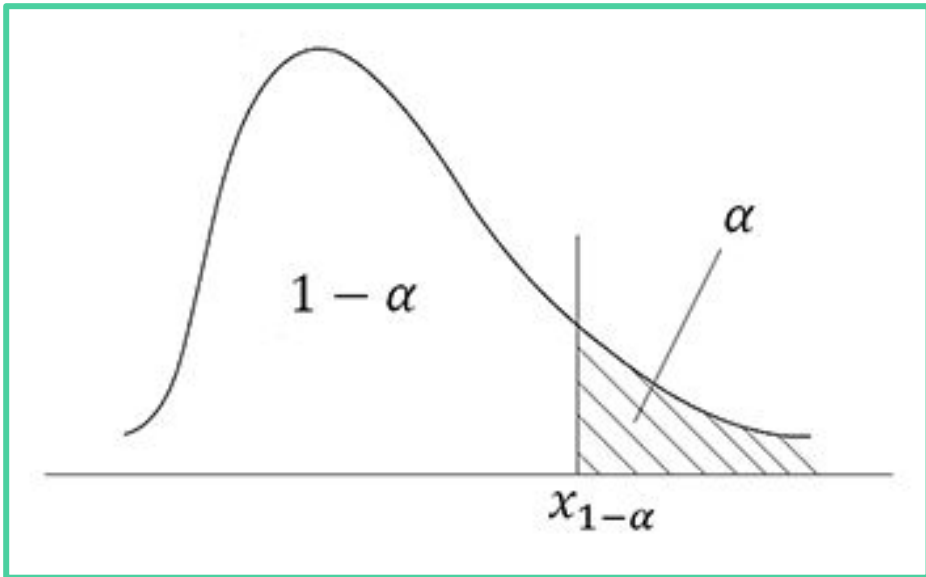
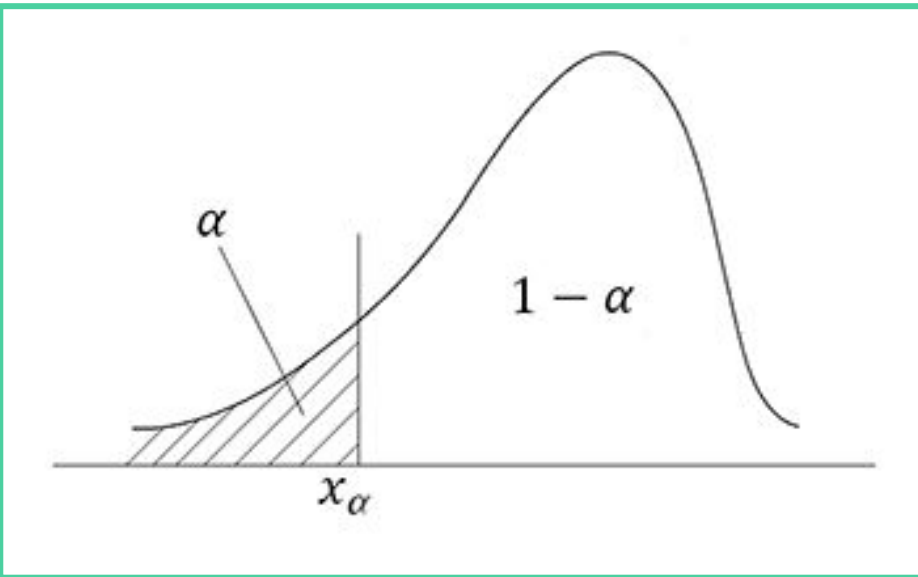
Подсчет доверительного интервала

Какой выбрать ?!?!?!

уровень	0,90	0,95	0,975	0,98	0,99	0,995	0,999	0,9995	0,9999
односторонняя	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	3,719
двусторонняя	1,645	1,960	2,241	2,326	2,576	2,807	3,291	3,481	3,891



Двусторонняя область



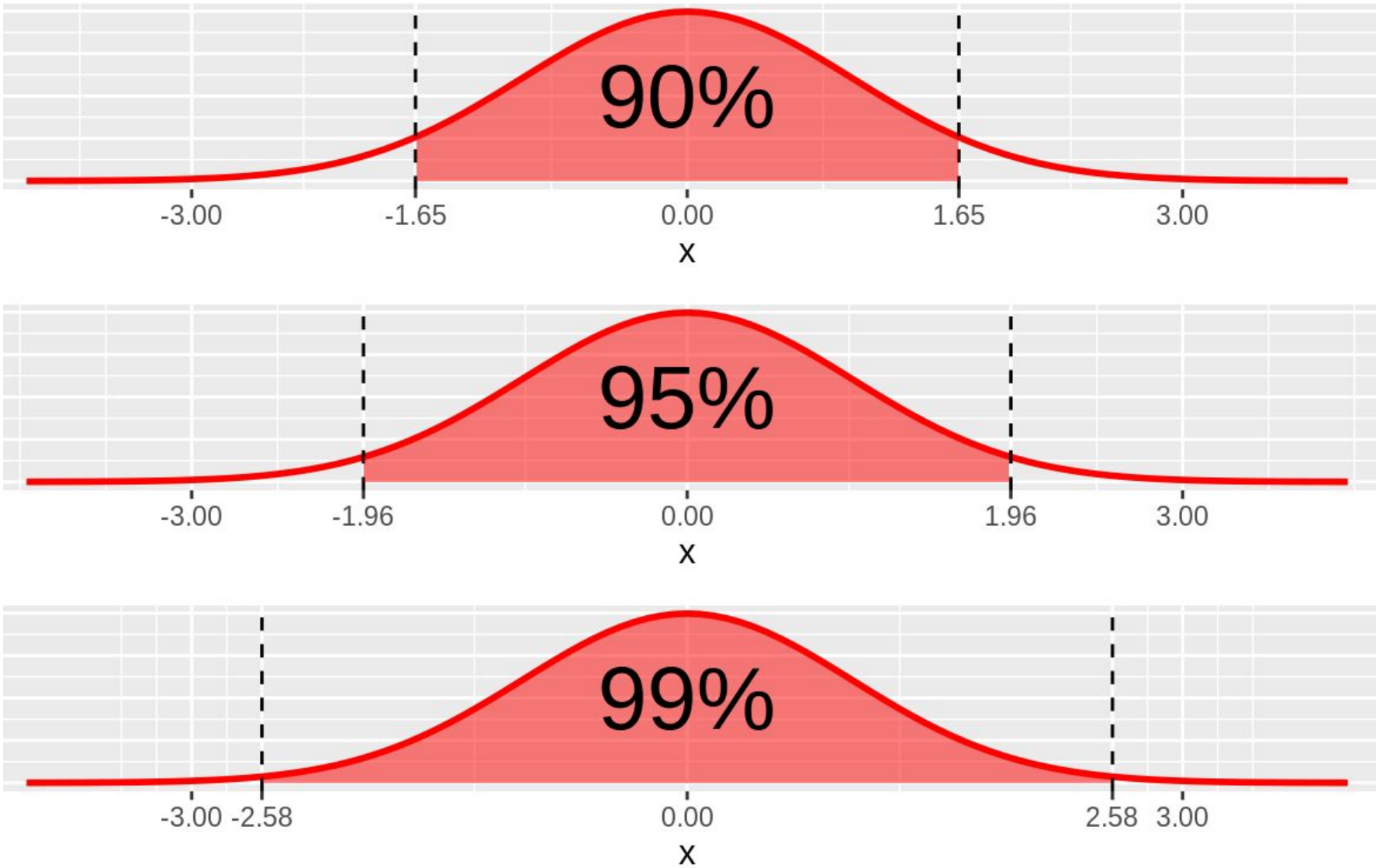
Односторонние области



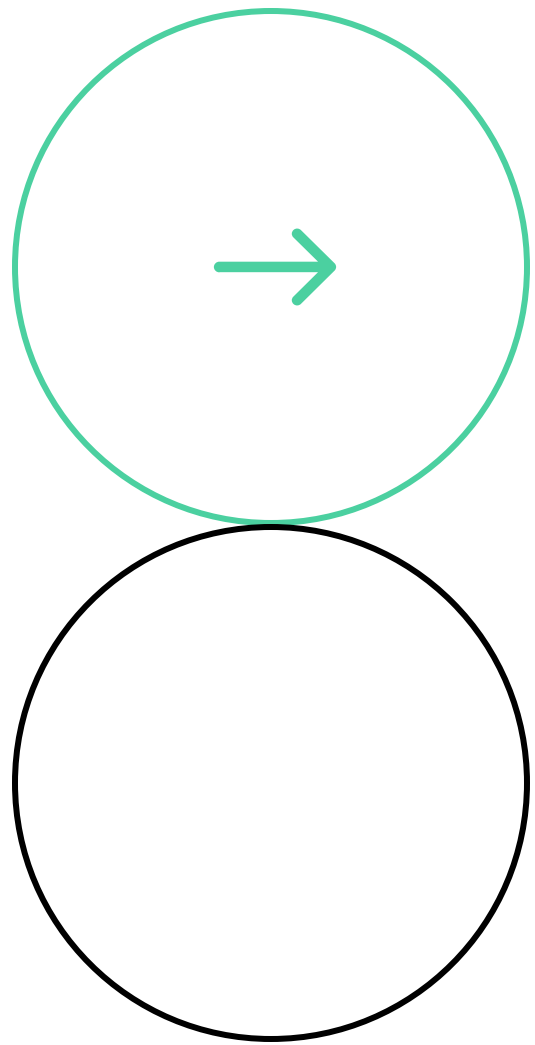
Подсчет доверительного интервала

Какой выбрать ?!?!?!

уровень	0,90	0,98	0,99
односторонняя	1,282	2,054	2,326
двусторонняя	1,645	2,326	2,576



Задача

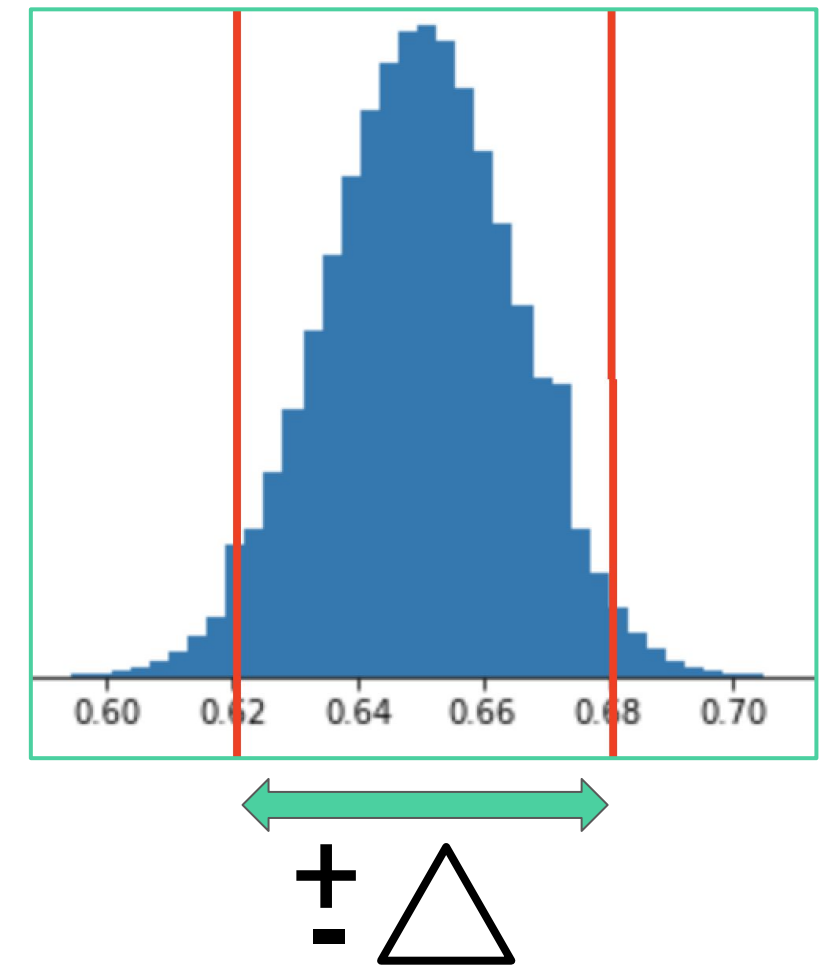


Пример:

Дана выборка 9, 5, 7, 7, 4, 10, дисперсия $\sigma^2 = 1$.
Постройте 99% доверительный интервал.

$$(\bar{x} - \Delta; \bar{x} + \Delta)$$

$$\Delta = \frac{\sigma}{\sqrt{n}} z_{\alpha}$$



Задача

уровень	0,90	0,95	0,99	0,995	0,999
односторонняя	1,282	1,645	2,326	2,576	3,090
двусторонняя	1,645	1,960	2,576	2,807	3,291

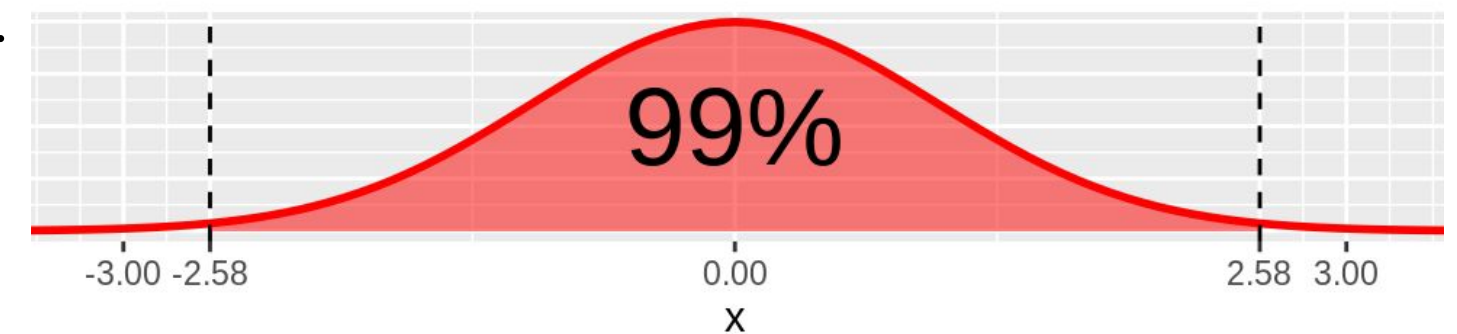


Задача:

Дана выборка 9, 5, 7, 7, 4, 10, дисперсия $\sigma^2 = 1$.
Постройте 99% доверительный интервал.

Решение:

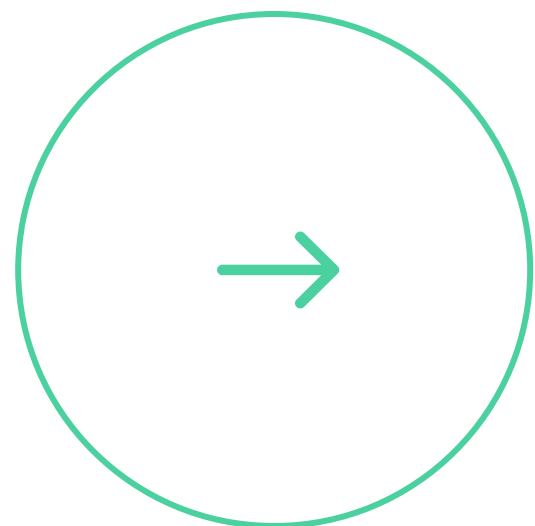
Среднее значение равно $\bar{x} = (9+5+7+7+4+10) / 6 = 7$.
Доверительный интервал имеет вид $(\bar{x} - \Delta; \bar{x} + \Delta)$.
По таблице нормального распределения находим $1 - \alpha/2 = 0.995$ и определяем квантиль $z_{\alpha} = 2.81$.



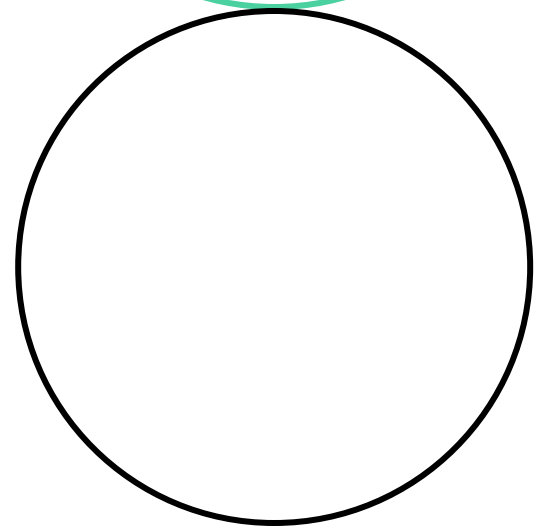
Теперь можем найти точность $\Delta = (\sigma/\sqrt{n}) * z_{\alpha} = 1/\sqrt{6} * 2.81 \approx 1.14$
(здесь мы воспользовались тем, что известна дисперсия генеральной совокупности). Искомый 99%-доверительный интервал имеет вид $(7 - 1.14; 7 + 1.14) = (5.86; 8.14)$.



ДИ для среднего при неизвестной дисперсии и маленькой выборке.



Бывают случаи, когда выборка маленькая и про её параметры еще и мало чего известно. Тогда в таком случае вместо нормального распределения используем t-распределение.



Формула: но в формуле процентыль z_{α} заменяем на $t_{\alpha}(n-1)$ - это квантиль распределения Стьюдента уровня $1-\alpha/2$ со степенью свободы = $n-1$. (это число дано в таблице распределения)



Доверительный интервал для среднего при неизвестной дисперсии.



Что делать, если дисперсия нам по каким-то причинам неизвестна? Тогда мы можем посчитать ее вручную!

Важно, чтобы выборка была больше 30, тогда вместо σ мы будем использовать выборочное стандартное отклонение

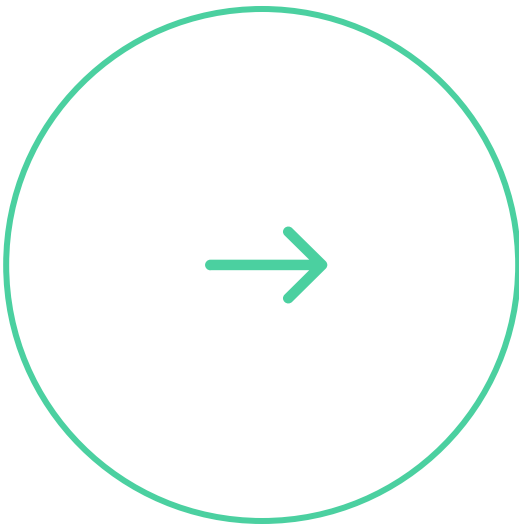
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Теорема

Доверительный интервал для среднего при неизвестной дисперсии, но большой выборке ($n > 30$), имеет вид $\bar{x} - (s/\sqrt{n}) * z_{\alpha}; \bar{x} + s/\sqrt{n} * z_{\alpha}$.



Степень свободы

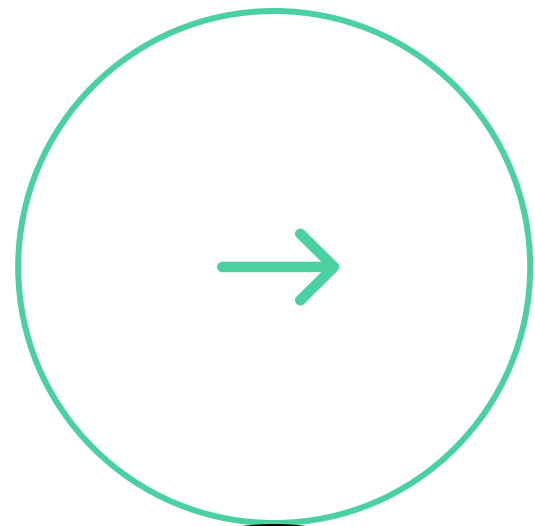


Степенью свободы = $n-1$

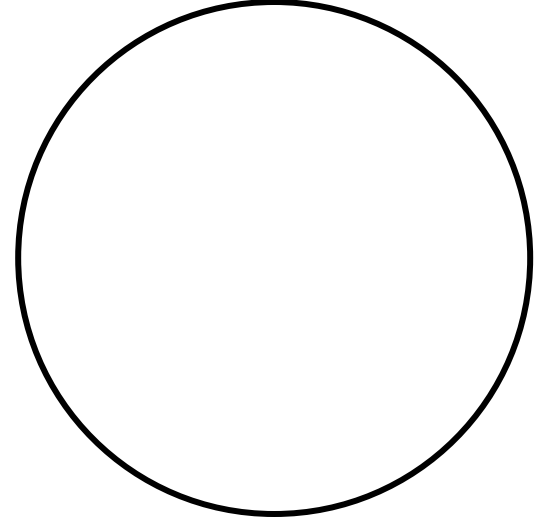
Значение t-критерия Стьюдента при уровне значимости (0,01; 0,05; 0,1; 0,15; 0,20; 0,25; 0,30)							
n	P - 0,01	P - 0,05	P - 0,1	P - 0,15	P - 0,2	P - 0,25	P - 0,3
1	63,6567412	12,7062047	6,3137515	4,1652998	3,0776835	2,4142136	1,9626105
2	9,9248432	4,3026527	2,9199856	2,2819306	1,8856181	1,6035675	1,3862066
3	5,8409093	3,1824463	2,3533634	1,9243197	1,6377444	1,4226253	1,2497781
4	4,6040949	2,7764451	2,1318468	1,7781922	1,5332063	1,3443976	1,1895669
5	4,0321430	2,5705818	2,0150484	1,6993626	1,4758840	1,3009490	1,1557673
6	3,7074280	2,4469119	1,9431803	1,6501732	1,4397557	1,2733493	1,1341569
7	3,4994833	2,3646243	1,8945786	1,6165917	1,4149239	1,2542787	1,1191591
8	3,3553873	2,3060041	1,8595480	1,5922214	1,3968153	1,2403183	1,1081454
9	3,2498355	2,2621572	1,8331129	1,5737358	1,3830287	1,2296592	1,0997162
10	3,1692727	2,2281389	1,8124611	1,5592359	1,3721836	1,2212554	1,0930581



Что за степени свободы и кто такой Стьюдент?



Замечание: Число степеней свободы зависит от того, сколько имеется связей между наблюдениями. Так как мы знаем среднее, то наблюдения связаны одним равенством и степеней свободы становится на одну меньше. То, что других связей нет, надо доказывать, но их действительно нет.



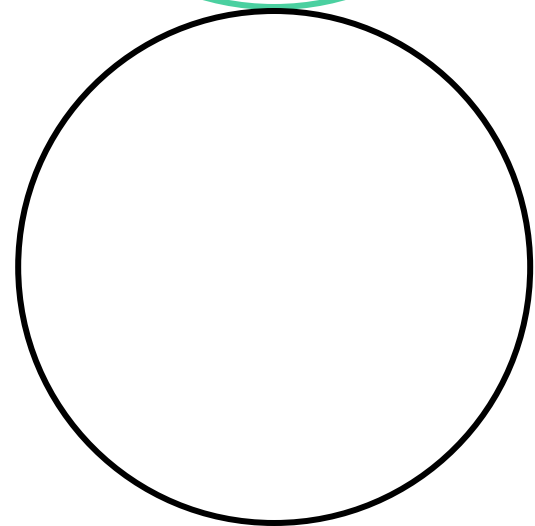
Замечание: Распределение Стьюдента было введено в 1908 году В.С. **Госсетом**, ирландским служащим пивоваренного завода, который участвовал в разработке новых технологий производства пива и **никаким студентом не был**. Придавать известности результаты исследований означало открыть корпоративную тайну, поэтому Госсет напечатал свои материалы под псевдонимом Стьюдент. Фишер ввел для него обозначение t -распределение



Пример для подсчета ДИ в случае маленькой выборки:



Задача: Пусть объем выборки $n = 16$, выборочное среднее $\bar{x} = 5$, выборочная дисперсия $s^2 = 4$. Постройте 99% доверительный интервал.



Решение: Среднее значение равно $\bar{x} = 5$, а выборочная дисперсия $s^2 = 4$. Так как неизвестна дисперсия генеральной совокупности и $n < 30$, поэтому точность интервальной оценки $\Delta = s/\sqrt{n} \cdot t_{\alpha}$. По таблице распределения Стьюдента находим $1 - \alpha/2 = 0.995$ и, так как у нас $n - 1 = 16 - 1 = 15$ степеней свободы, определяем квантиль $t_{\alpha} = 2.95$. Теперь можем найти точность $\Delta = (s/\sqrt{n}) \cdot t_{\alpha} = (2/\sqrt{16}) \cdot 2.95 \approx 1.475$. Искомый 99%-доверительный интервал имеет вид $(5 - 1.475; 5 + 1.475) = (3.525; 6.475)$.



Вернемся к нашим президентским выборам:



Сколько людей надо опросить, чтобы наши результаты давали 95% точность?

Обратная задача: раз мы знаем ДИ, то можно ли найти минимальный объем выборки голосовавших, для того, чтобы с заданной точностью и уровнем доверия найти среднее.

Важно! Для того чтобы найти минимальный необходимый объем выборки для построения доверительного интервала для среднего значения с заданной точностью Δ и уровнем значимости α , достаточно применить формулу.

$$n = \left(\frac{z_{\alpha} \sigma}{\Delta} \right)^2$$

Теперь понятно, как определить объем выборки при проведении собственных исследований!



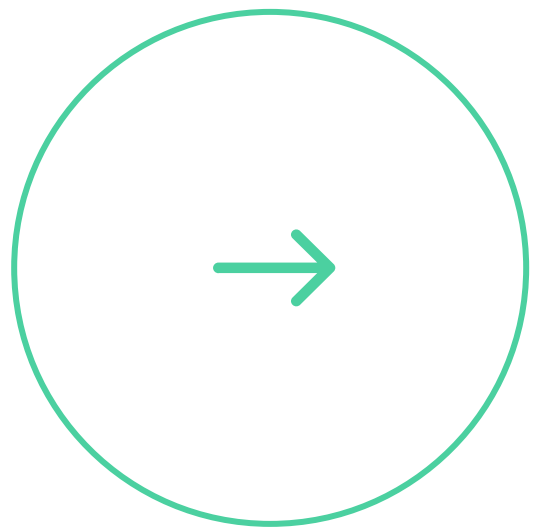
Шаг пройден!

Наша цель

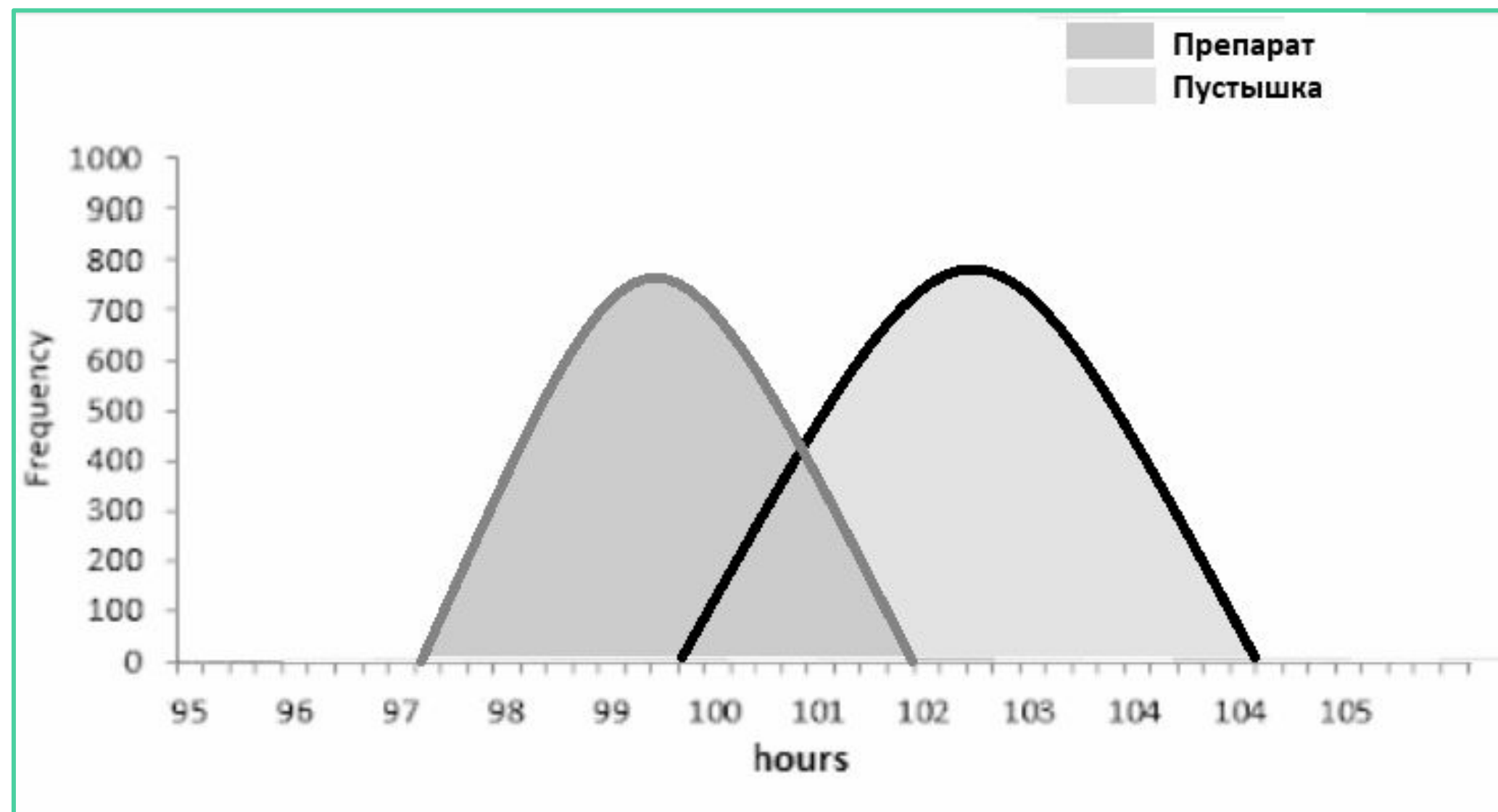
- Доверительный интервал ✓
- Уровень значимости ✓
- p-value (p-значение)
- t-test (Критерий Стьюдента)



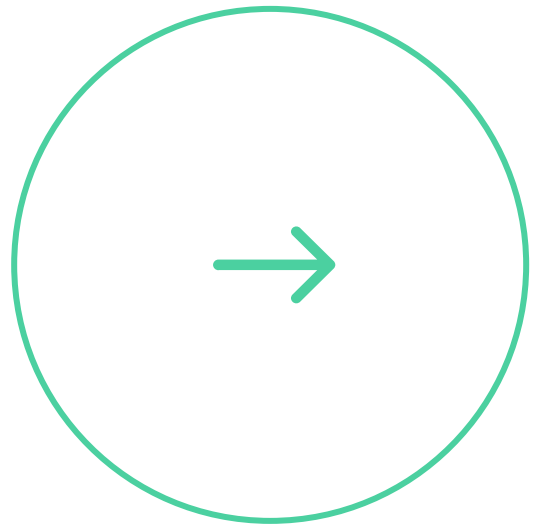
Проверка гипотез



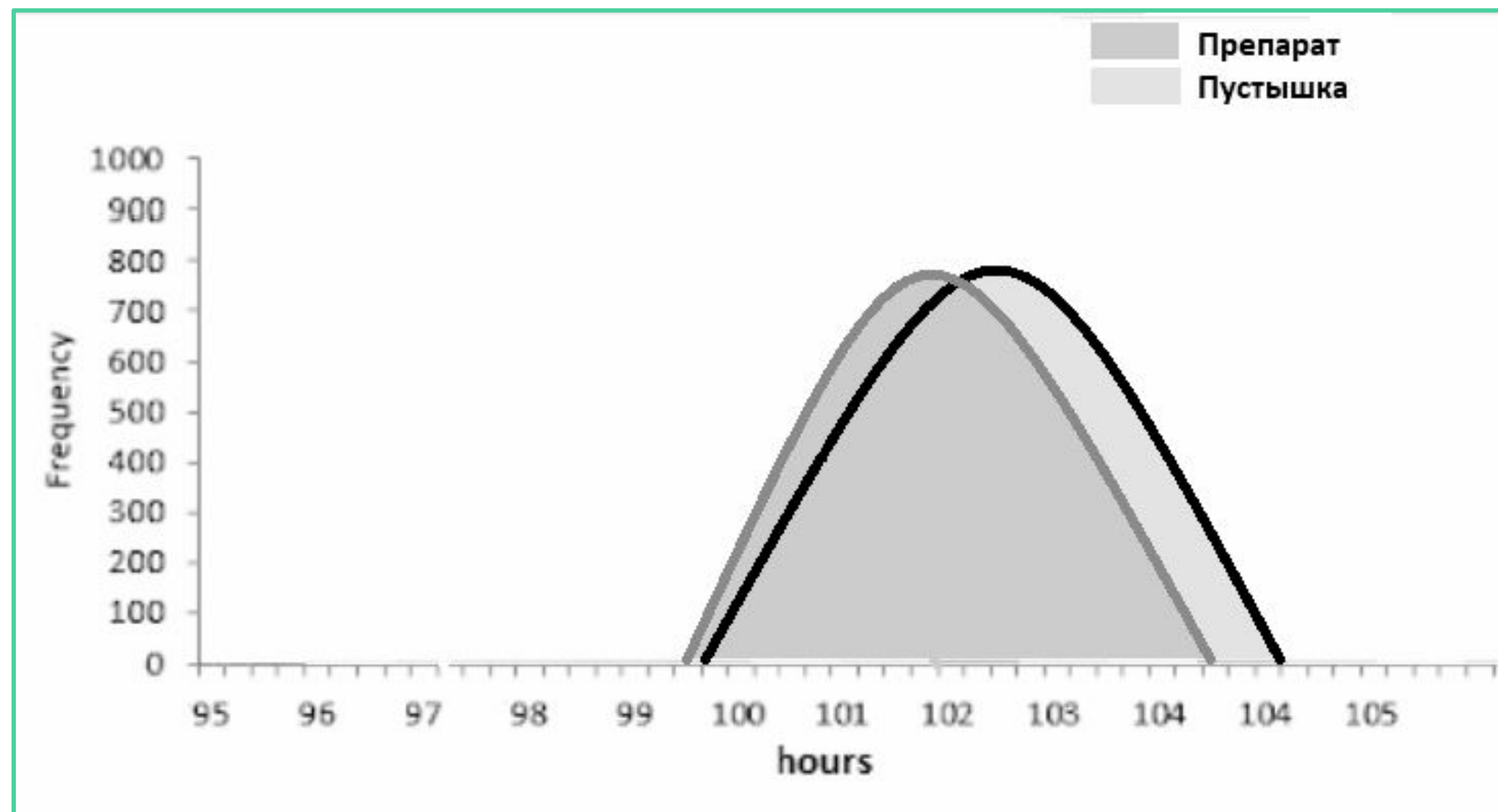
Что можно сказать о результативности препарата?



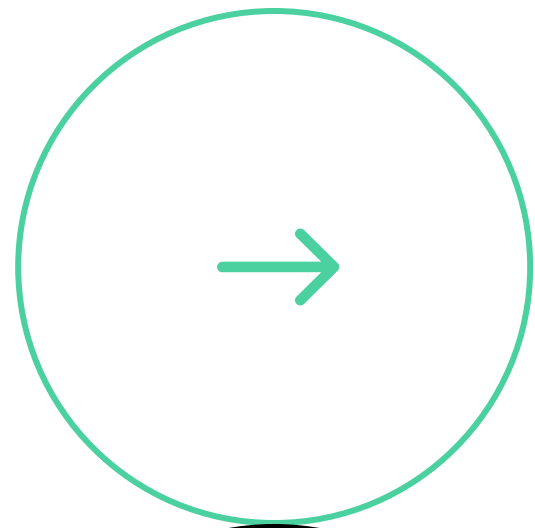
Проверка гипотез



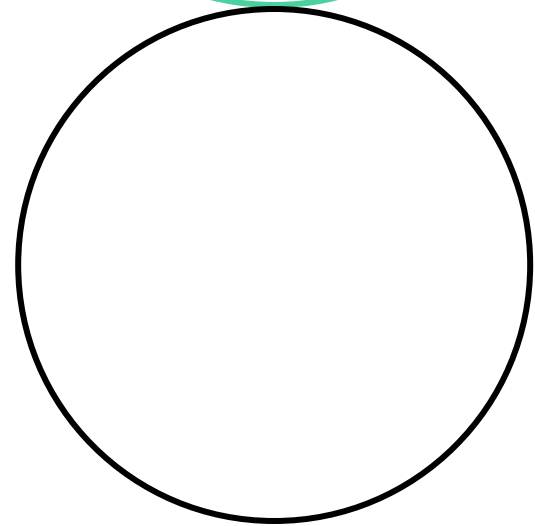
Что можно сказать о результативности препарата теперь?



Проверка гипотез

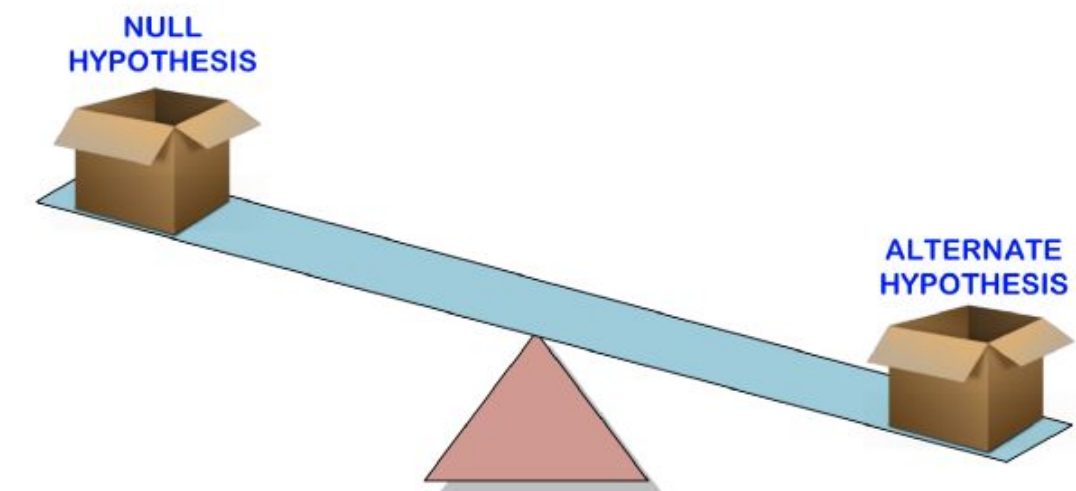


Этот процесс известен как **проверка гипотез** (проверка статистических гипотез или проверка значимости).

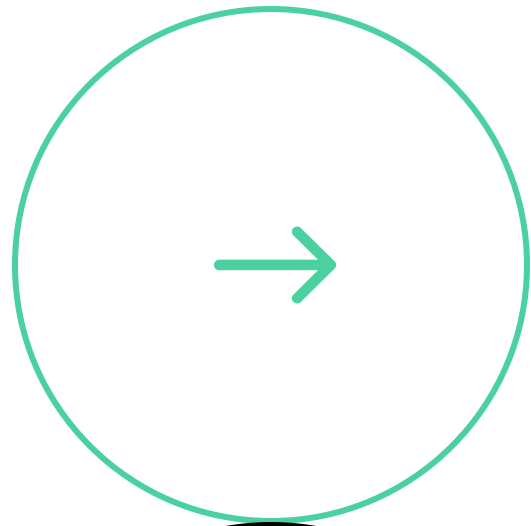


В рамках проверки гипотезы мы либо отвергаем **нулевую гипотезу** и принимаем **альтернативную**, либо **НЕ** отвергаем **нулевую гипотезу**.

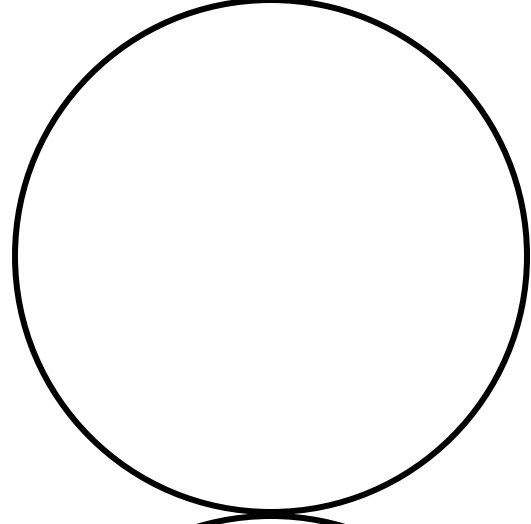
Замечание: Нулевая гипотеза всегда относится к популяции, представляющей больший интерес.



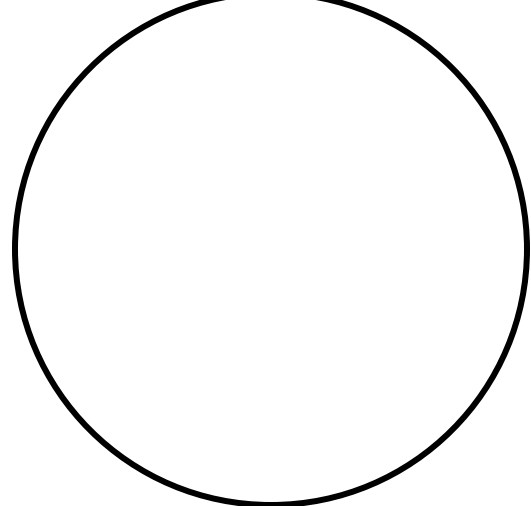
Логика проверки гипотез



1. Всегда составляют и проверяют нулевую гипотезу (H_0), которая отвергает эффект (например, эффект случайный) в генеральной совокупности. Например, при времени выздоровления пациентов - **нулевая гипотеза H_0** означала бы, что показатели времени выздоровления **не отличаются** от времени выздоровления без лекарства.



2. Затем определяют и проверяют альтернативную гипотезу (H_1), которая принимается, если нулевая гипотеза неверна. Альтернативная гипотеза больше относится к той теории, которую собираются исследовать. На примере **альтернативная гипотеза (H_1)** заключается в утверждении, что препарат помогает ускорить выздоровление.

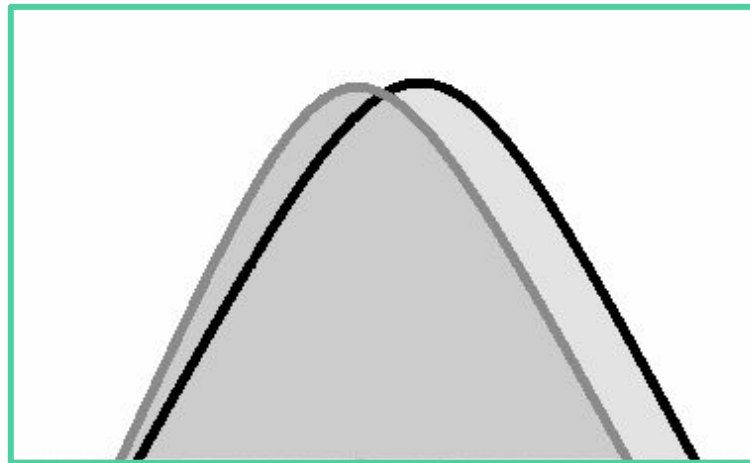
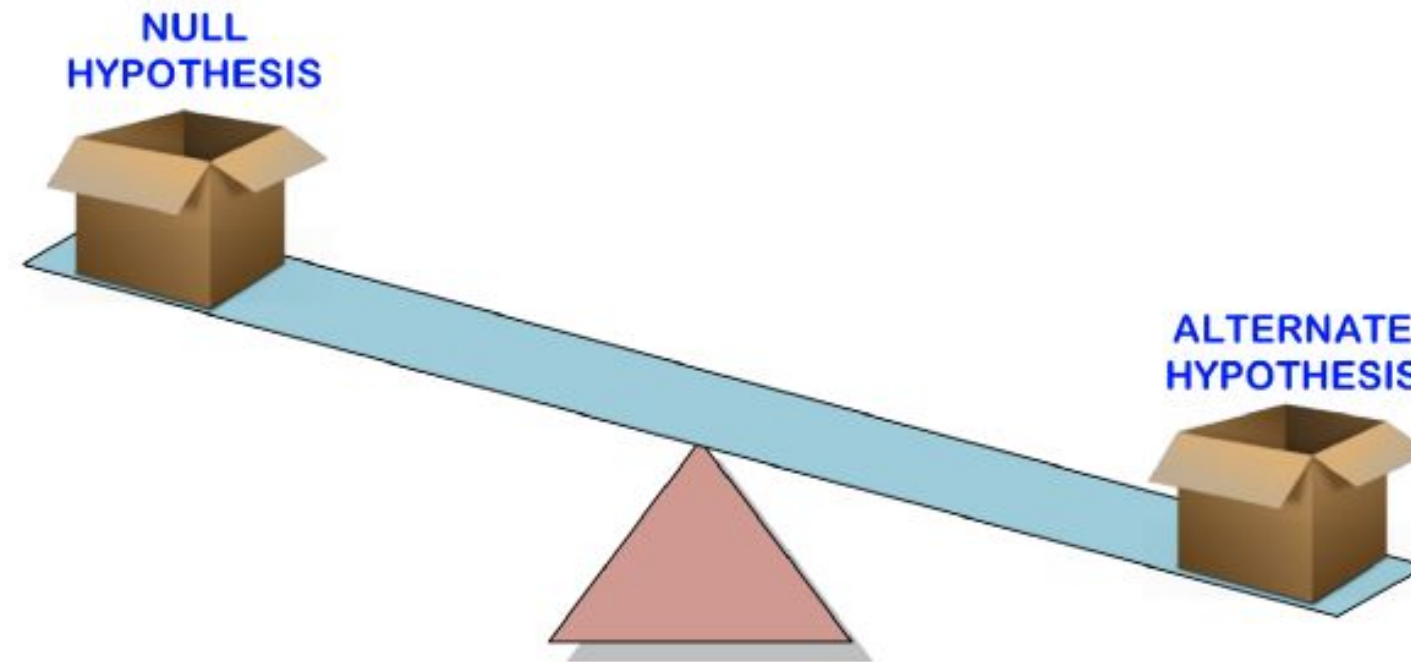


Уровень значимости. Важным этапом проверки статистических гипотез является определение уровня *статистической значимости*, т.е. максимально допускаемой исследователем вероятности ошибочного отклонения нулевой гипотезы.

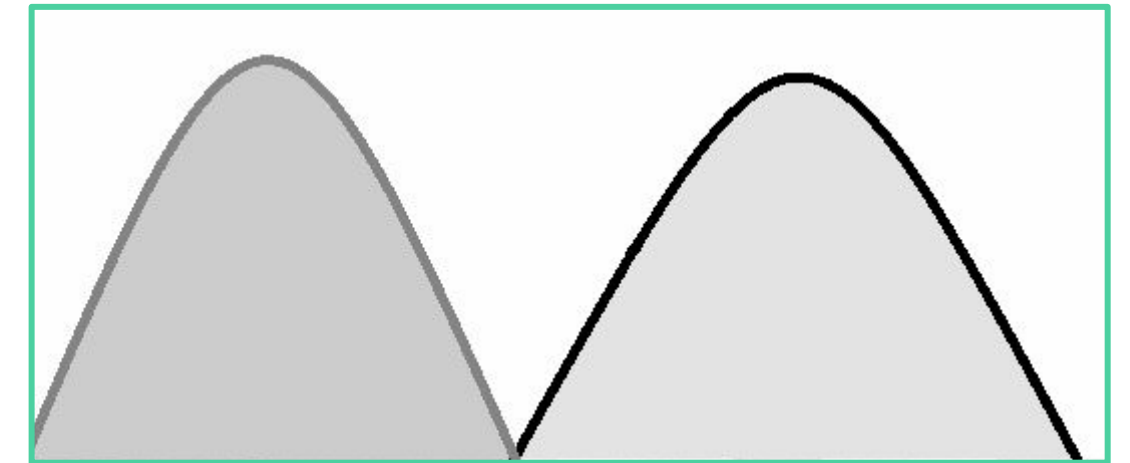
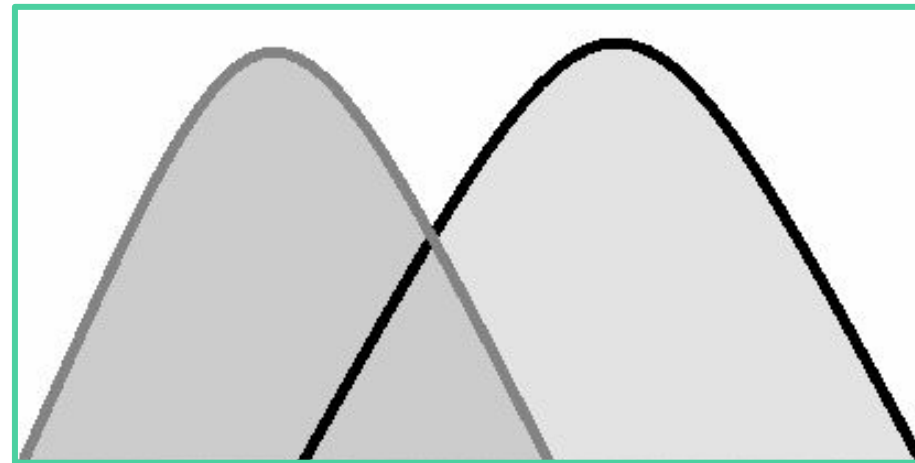
Зачастую за уровень значимости принимается 5%



Логика проверки гипотез



Выборки не отличаются.
Эффект случайный



Выборки отличаются.
Эффект не случайный



Применение значения P-value

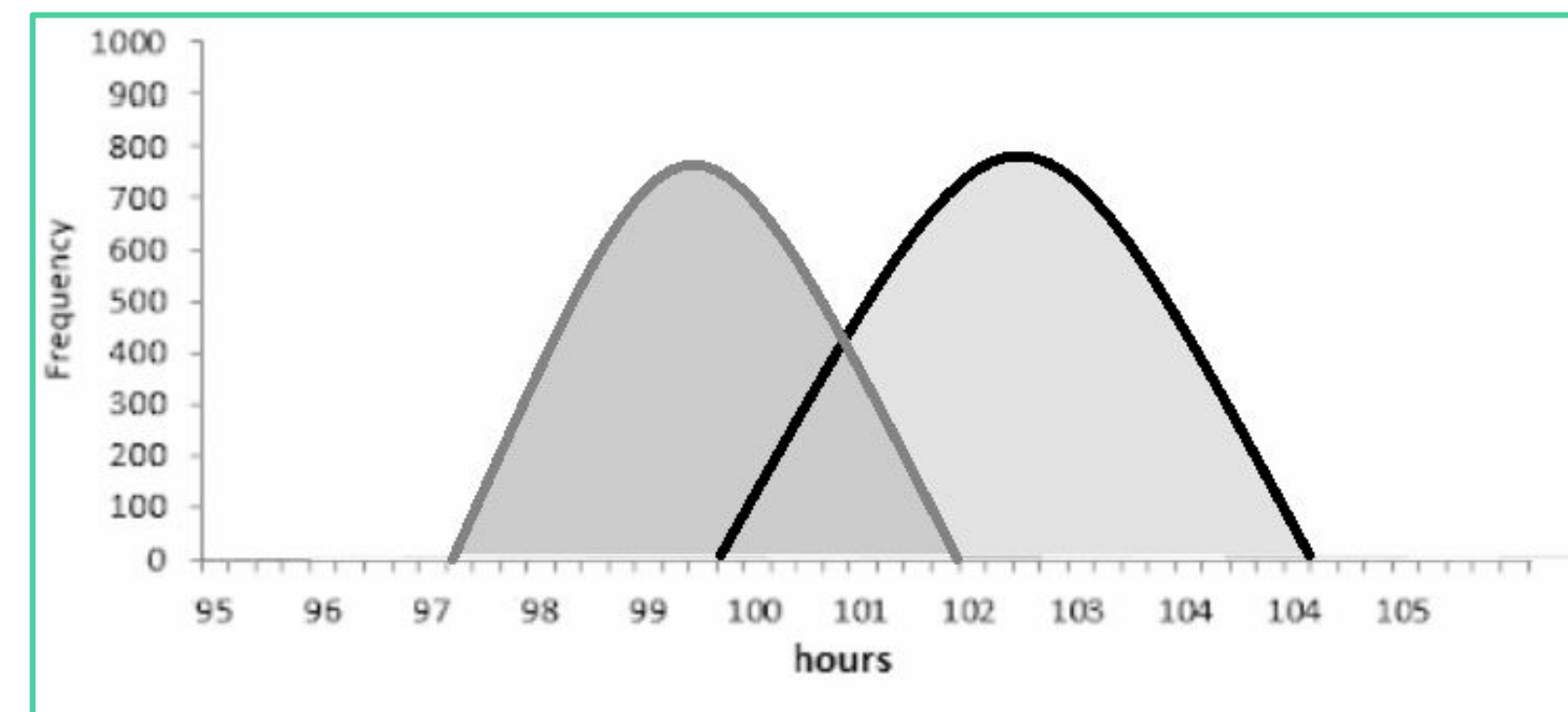
Значение статистики критерия, полученное из выборки, связывают с уже известным распределением, которому она подчиняется, чтобы получить значение p , площадь обоих "хвостов" (или одного "хвоста", в случае односторонней гипотезы) распределения вероятности.

Большинство компьютерных пакетов обеспечивают автоматическое вычисление двустороннего значения p . (Python)

Определение:

Значение p — это вероятность получения нашего вычисленного значения или его еще большего значения, если нулевая гипотеза верна.

Иными словами, p - это вероятность отвергнуть нулевую гипотезу при условии, что она верна.



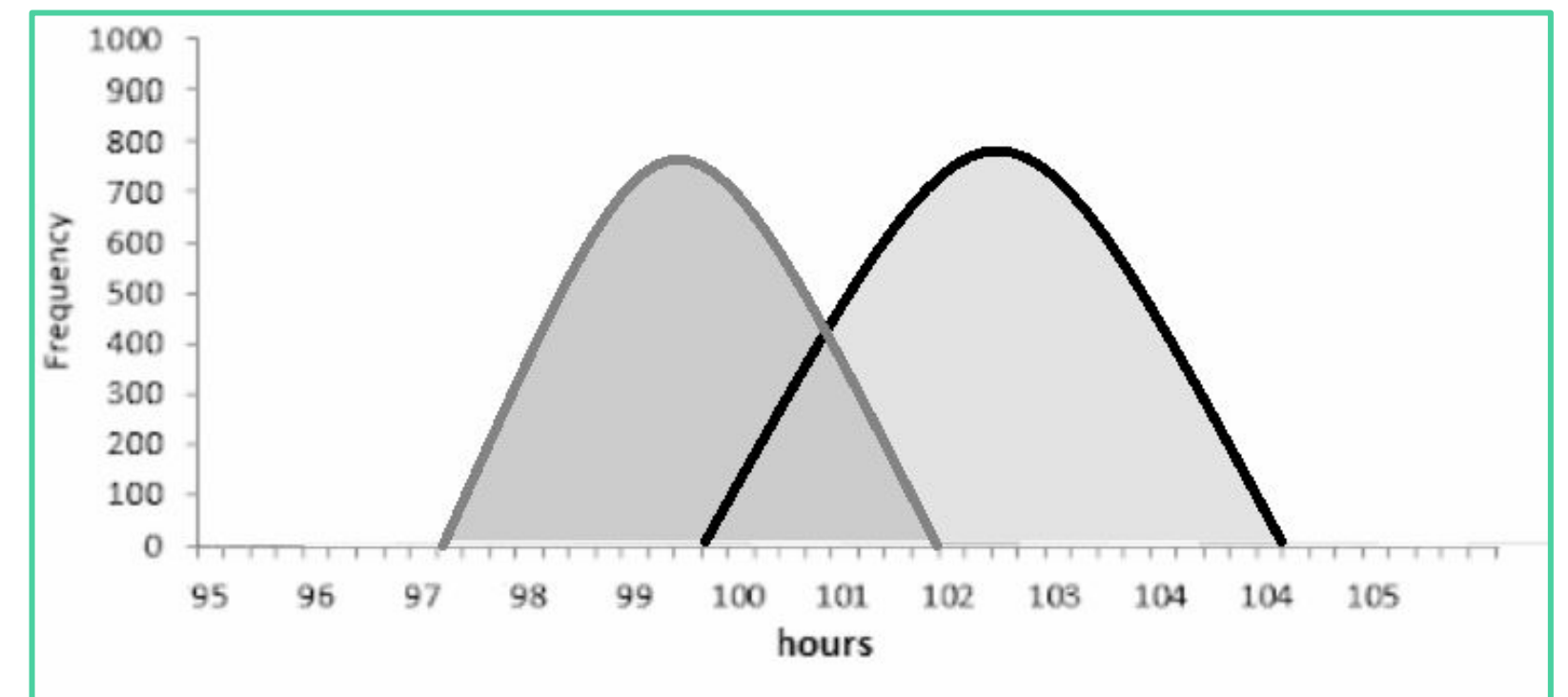
Применение значения P-value

Как по p-value определить, есть ли основания отвергнуть нулевую гипотезу? Тут важно сначала зафиксировать уровень значимости α , а потом уже делать выводы.

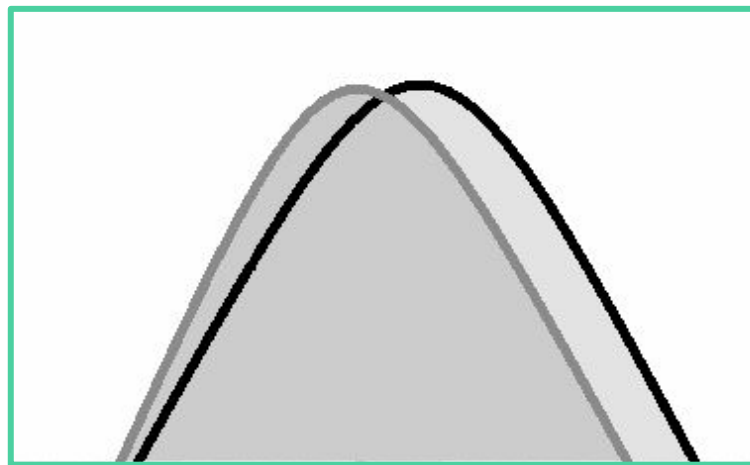
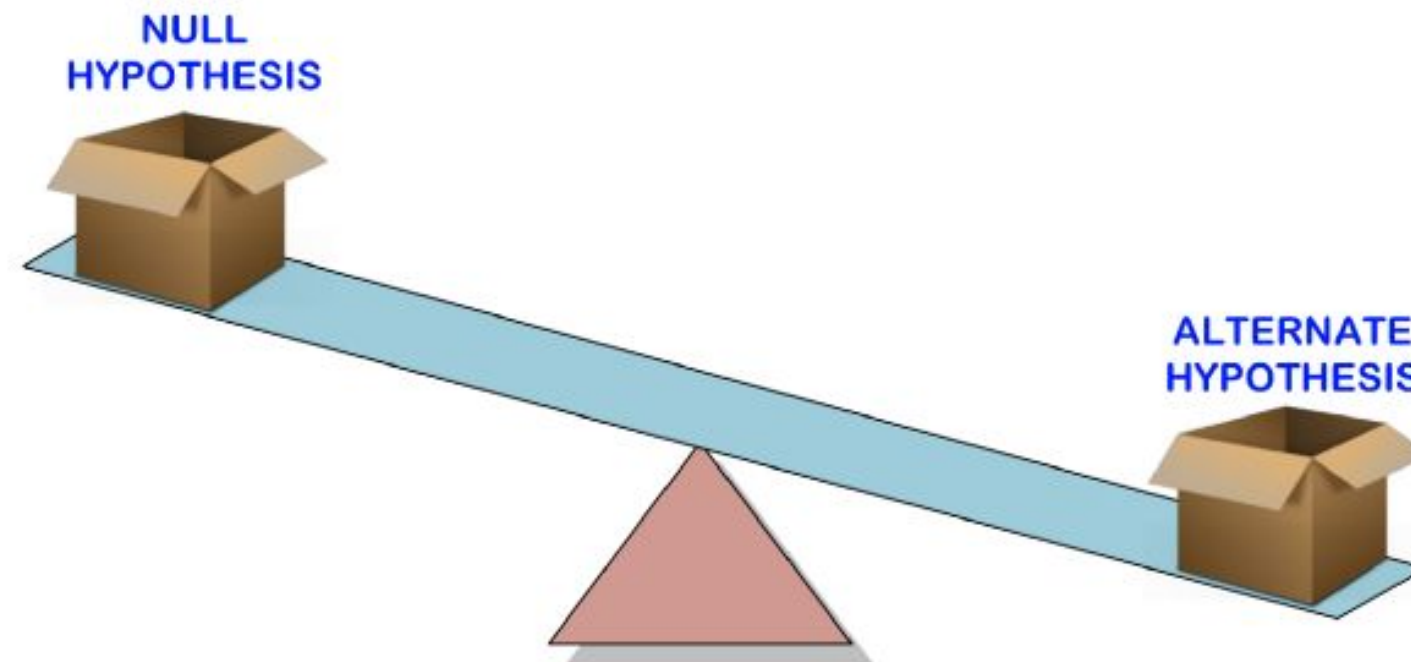
Уровень значимости α – это вероятность отвергнуть нулевую гипотезу при условии, что она верна.

P-value – это минимальный уровень значимости, на котором нулевая гипотеза может быть отвергнута.

Соответственно, если **p – value** меньше нашего фиксированного **уровня значимости α** , на котором мы проверяем гипотезу, то **нулевую гипотезу** следует отвергнуть, если больше – то отвергать нулевую гипотезу оснований нет.

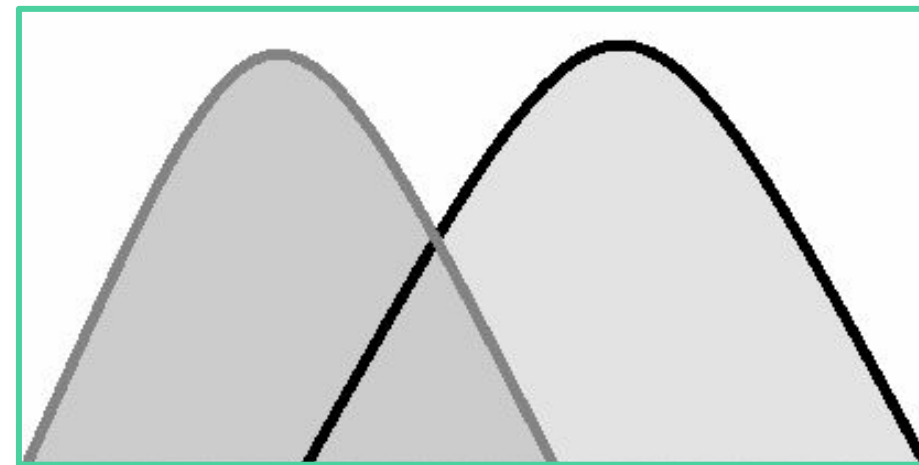


Логика проверки гипотез



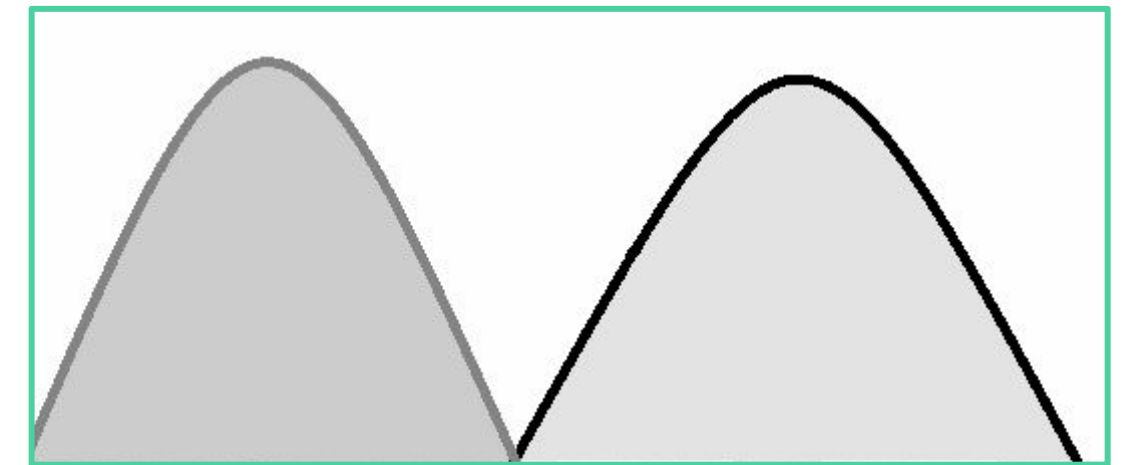
Выборки не отличаются.
Эффект случайный

$p\text{-value} > 5\%$



Неоднозначно

$p\text{-value} = 5\%$

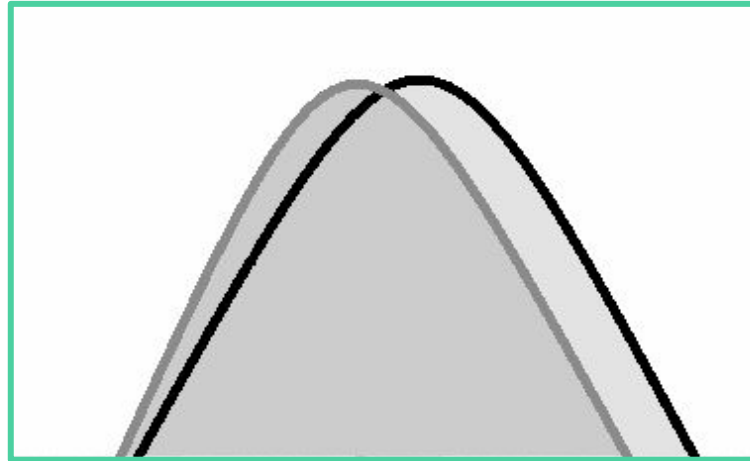


Выборки отличаются.
Эффект не случайный

$p\text{-value} < 5\%$

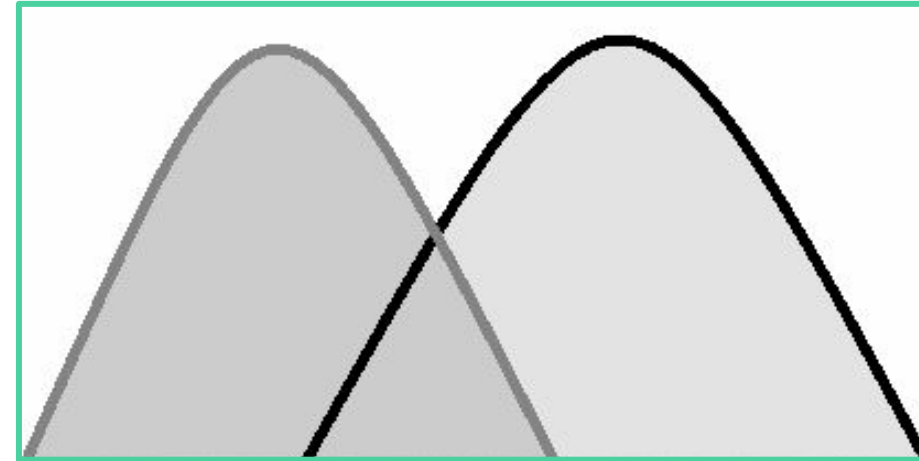


Ловушка!!!!



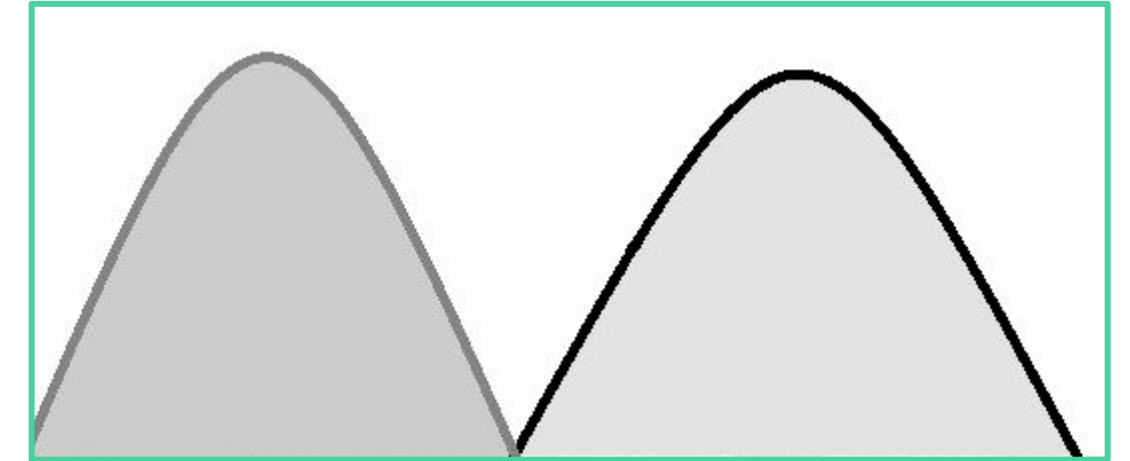
Выборки не отличаются.
Эффект случайный

$p\text{-value} > 5\%$



Неоднозначно

$p\text{-value} = 5\%$



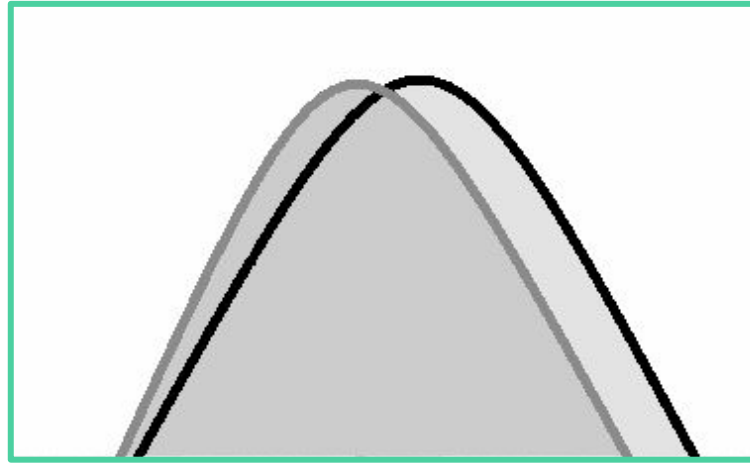
Выборки отличаются.
Эффект не случайный

$p\text{-value} < 5\%$

Какому случаю соответствует фраза:
P-value превосходит уровень значимости 5%

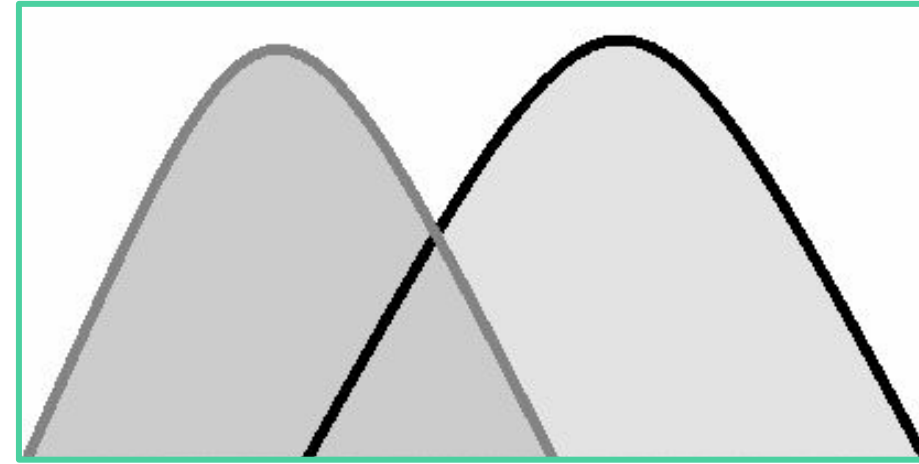


Ловушка!!!!



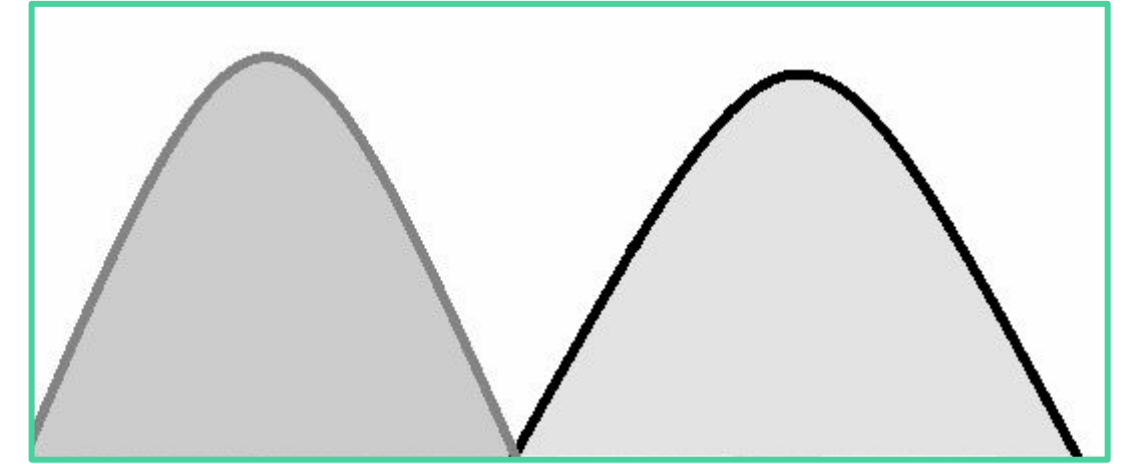
Выборки не отличаются.
Эффект случайный

$p\text{-value} > 5\%$



Неоднозначно

$p\text{-value} = 5\%$



Выборки отличаются.
Эффект не случайный

$p\text{-value} < 5\%$

Какому случаю соответствует фраза:
Р-value превосходит уровень значимости 5%



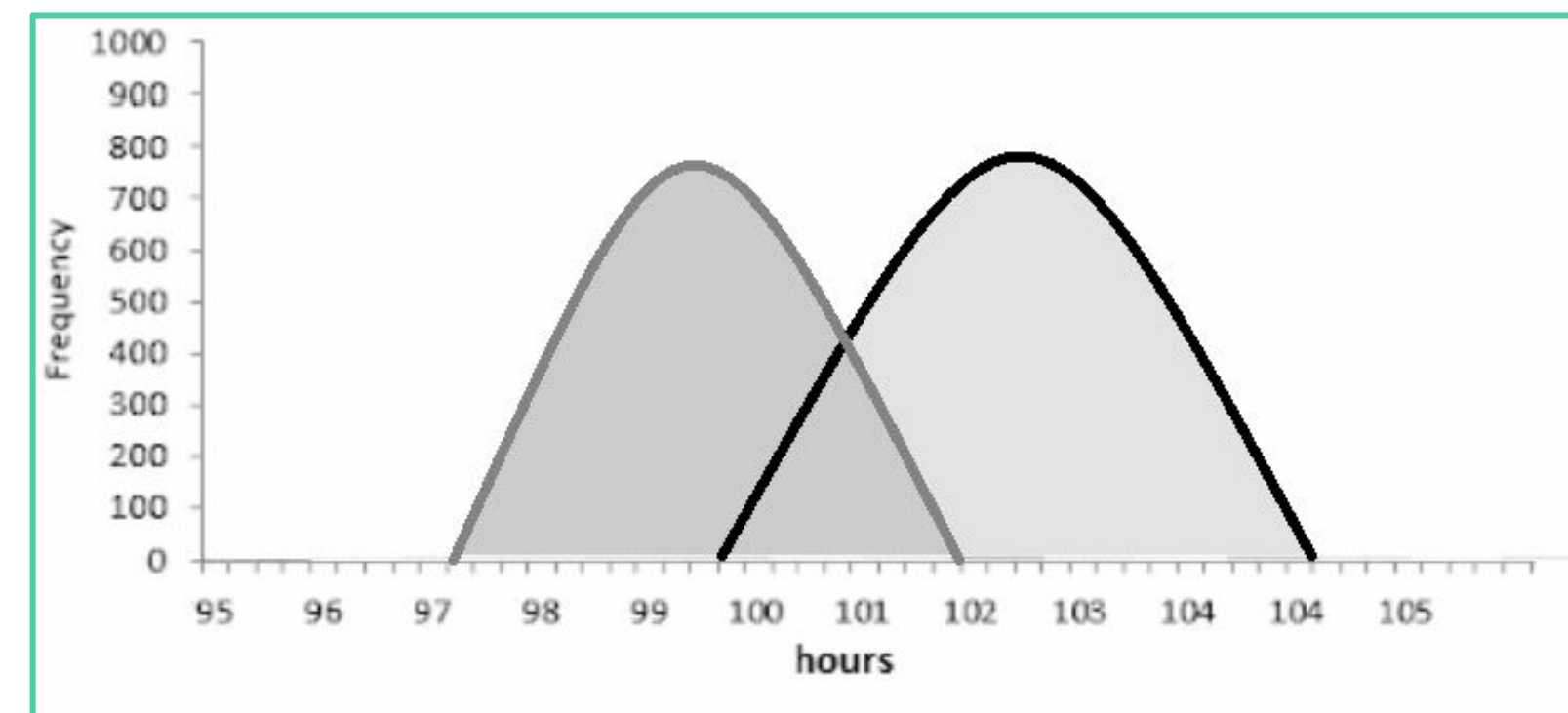
Применение значения P-value

По сути это вероятность оказаться в критической области - в области редких событий

Значит ли это, что эффект есть?

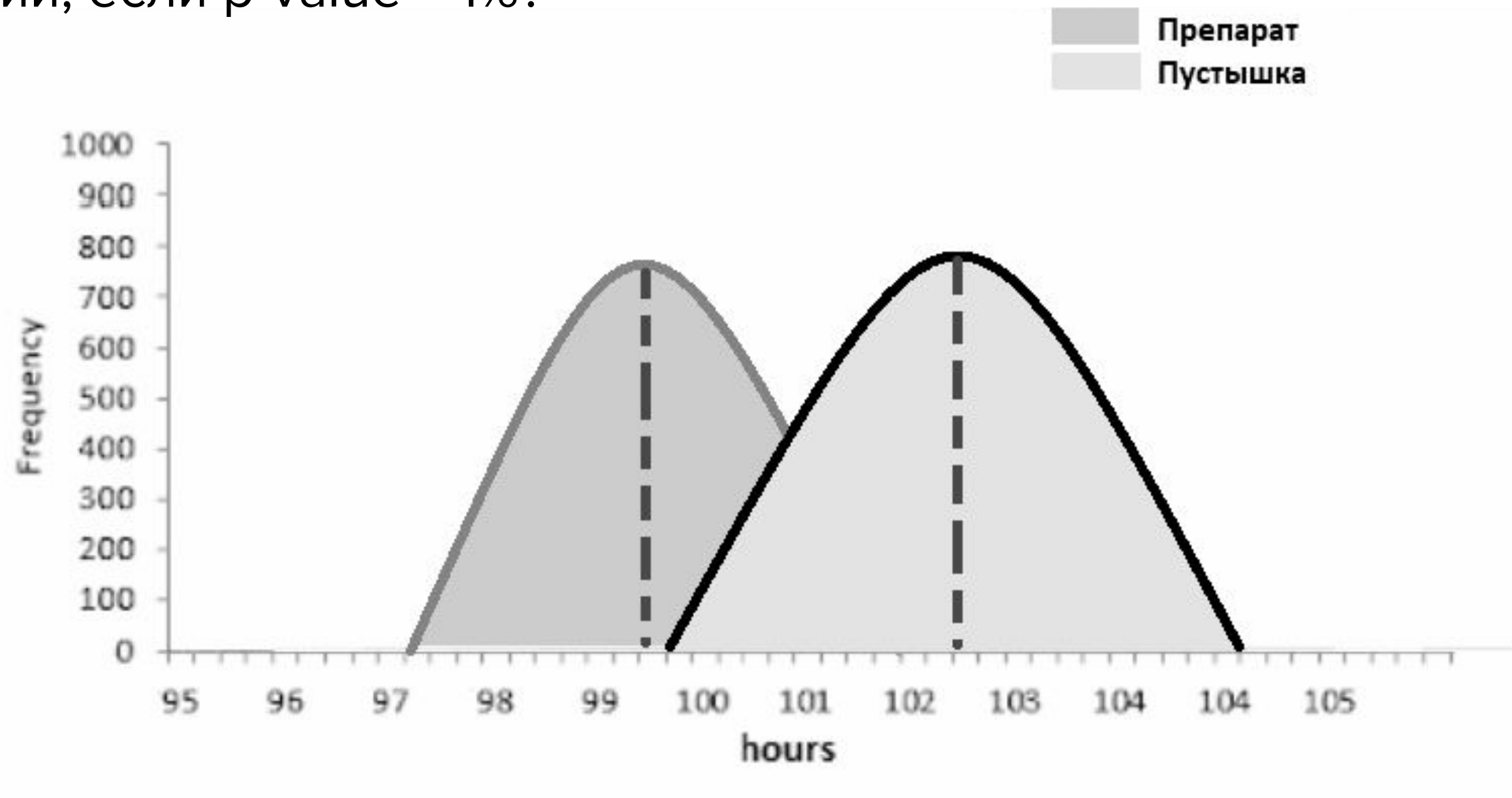
Значит ли это, что это вероятность того, что препарат помогает?

Значит ли это, что это вероятность того, что препарат **НЕ** помогает?



Применение значения P-value

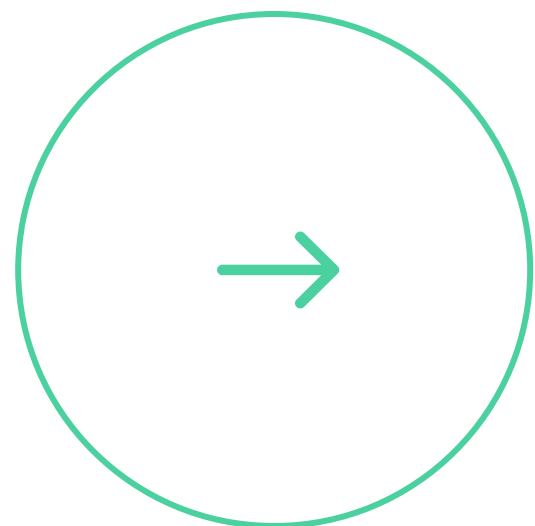
Можно ли по графику сказать, что наше лекарство помогает в лечении, если $p\text{-value} = 1\%$?



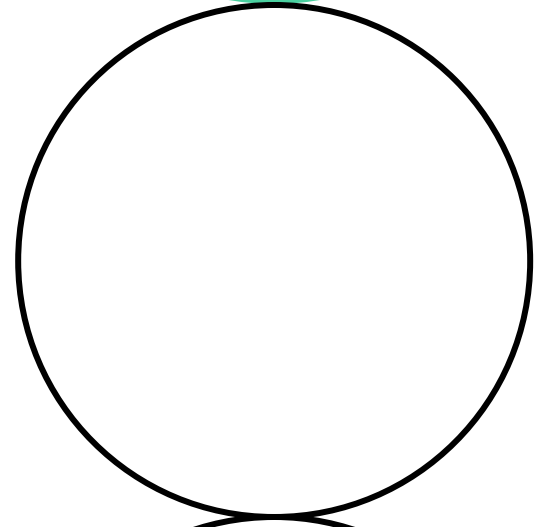
Низкие значения p - это хорошо; Они указывают на то, что ваши данные не были случайными.



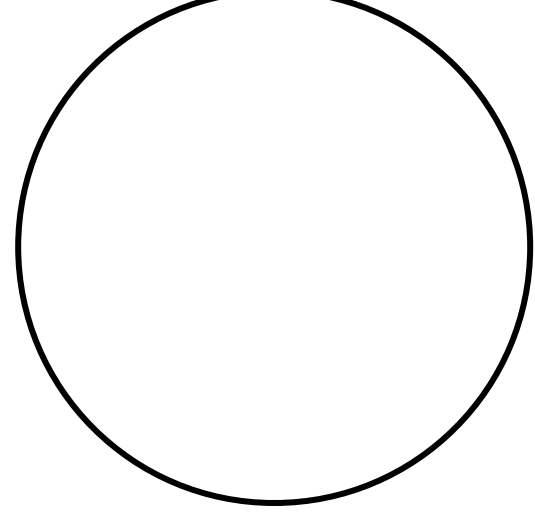
Логика применения значения P-value



Следует решить, сколько аргументов позволяют отвергнуть нулевую гипотезу в пользу альтернативной - чем меньше значение p , тем сильнее аргументы против нулевой гипотезы.



Традиционно полагают, если $p < 0,05$, (5%) то аргументов достаточно, чтобы **отвергнуть нулевую гипотезу**, хотя есть небольшой шанс против этого. Тогда можно сказать, что результаты значимы на 5% уровне.

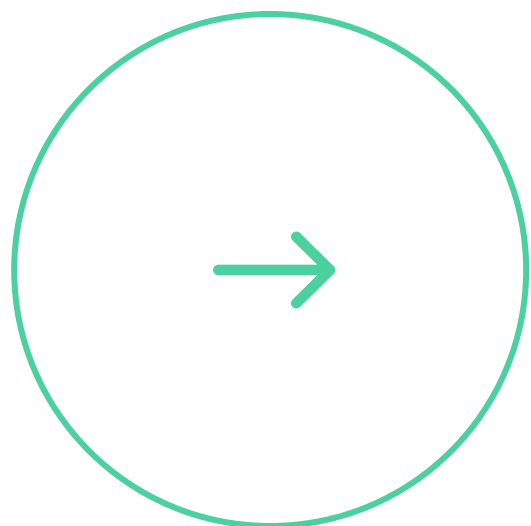


Если это может привести к серьезным последствиям, необходимо потребовать более веских аргументов, например, выбрать значение $p = 0,01$ или $0,001$ (медицинский препарат).

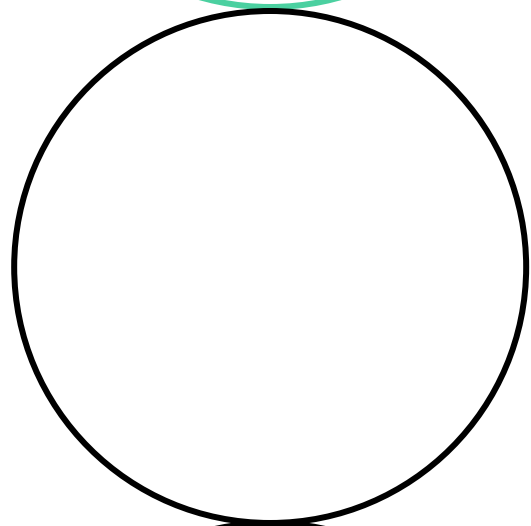
Напротив, если $p > 0,05$, то аргументов недостаточно, чтобы отвергнуть нулевую гипотезу. **Не отвергая нулевую гипотезу**, можно заявить, что результаты не значимы на 5% уровне. Данное заключение не означает, что нулевая гипотеза истинна, просто недостаточно аргументов (возможно, маленький объем выборки), чтобы ее отвергнуть.



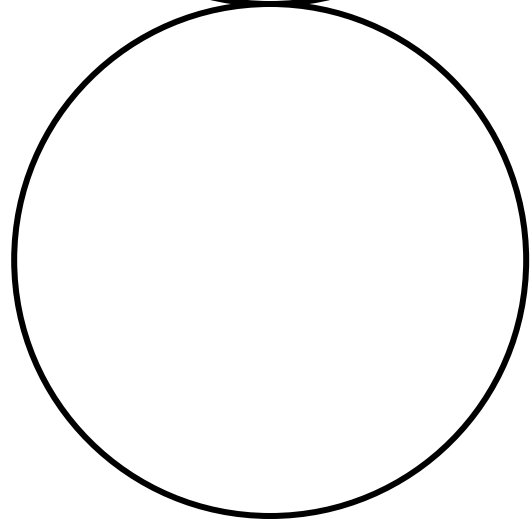
Проверка гипотез против доверительных интервалов



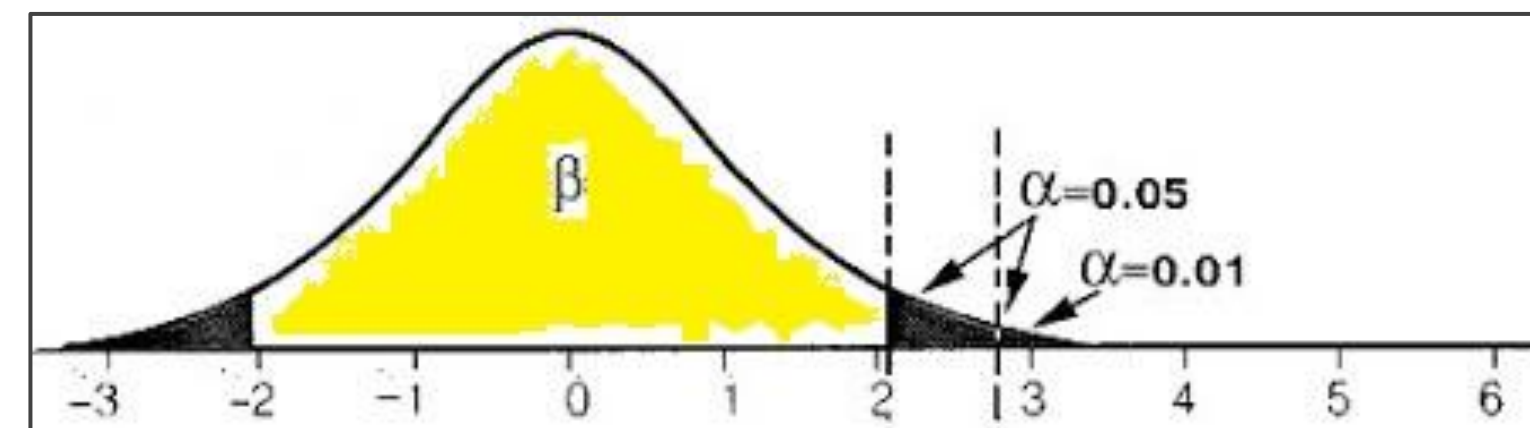
Доверительные интервалы и проверка гипотез тесно связаны. Первоначальная цель проверки гипотезы состоит в том, чтобы принять решение и предоставить точное значение p .



ДИ предоставляют интервал вероятных значений для истинного эффекта, поэтому его также можно использовать для принятия решения даже без точных значений p .



Например, если бы гипотетическое значение для данного эффекта находилось вне 95% ДИ, можно было бы счесть гипотетическое значение неправдоподобным и отвергнуть. В этом случае станет известно, что $p < 0,05$, но не станет известно его точное значение.



Проверим себя

В нашем исследовании p уровень значимости равен 0.02

- Вероятность того, что верна нулевая гипотеза (новый препарат не влияет на скорость выздоровления) также равняется всего-лишь 0,02.
- Если бы в исследовании мы получили $p=0.9$, это означало бы, что верна нулевая гипотеза, и новый препарат не влияет на скорость выздоровления.
- Чем меньше p уровень значимости, тем сильнее получаемые различия. Например, если бы p уровень значимости в нашем исследовании был бы равен 0.0001, значит новый препарат еще сильнее бы влиял на скорость выздоровления.
- Статистически значимый результат, всегда означает ценный и осмысленный результат.
- Вероятность случайно получить такие различия равняется 0.02.



3-5 минуты на подумать



Проверим себя

В нашем исследовании p уровень значимости равен 0.02

- Вероятность того, что верна нулевая гипотеза (новый препарат не влияет на скорость выздоровления) также равняется всего-лишь 0,02.
- Если бы в исследовании мы получили $p=0.5$, то верна нулевая гипотеза, и новый препарат не влияет на скорость выздоровления.
- Чем меньше p уровень значимости, тем сильнее получаемые различия. Например, если бы p уровень значимости в нашем исследовании был бы равен 0.0001, значит новый препарат еще сильнее бы влиял на скорость выздоровления.
- Статистически значимый результат, всегда означает ценный и осмысленный результат.
- Вероятность случайно получить такие различия равняется 0.02.

Внимание!
Нет верных!!!



Проверим себя

В нашем исследовании p уровень значимости равен 0.02



Вероятность того, что верна нулевая гипотеза (новый препарат не влияет на скорость выздоровления) также равняется всего-лишь 0,02.

Скорее наоборот



Если бы в исследовании мы получили $p = 0.9$, это означало бы, что верна нулевая гипотеза, и новый препарат не влияет на скорость выздоровления.

Это значит, что мы не можем отвергнуть нулевую гипотезу, но это не значит, что она верна



Чем меньше уровень значимости, тем сильнее получаемые различия. Например, если бы уровень значимости в нашем исследовании был бы равен 0.0001, значит новый препарат еще сильнее бы влиял на скорость выздоровления.

Со скоростью это никак не связано



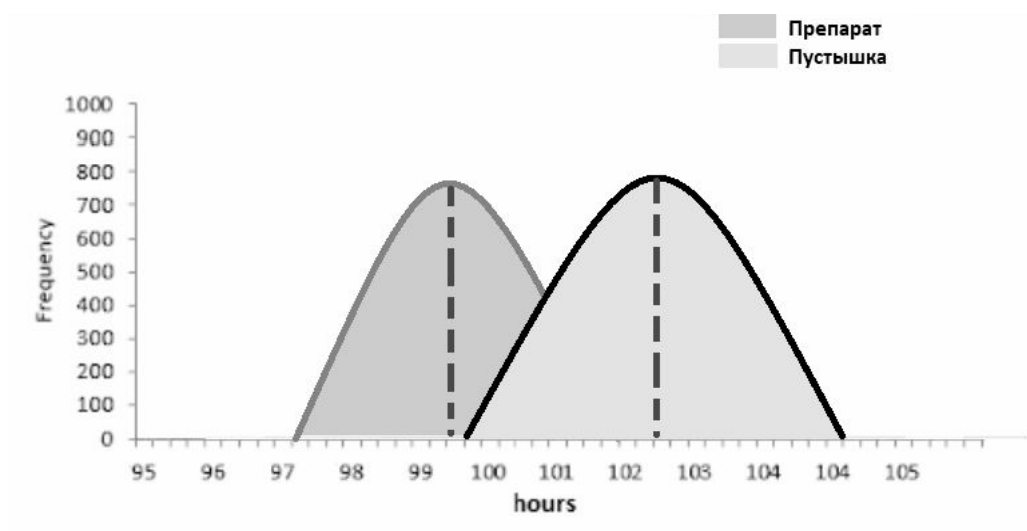
Статистически значимый результат, всегда означает ценный и осмысленный результат.

Отнюдь нет, доказывают разные примеры



Вероятность случайно получить такие различия равняется 0.02.

Близко к определению, но все равно не то.



Шаг пройден!

Наша цель

- Доверительный интервал ✓
- Уровень значимости ✓
- p-value (p-значение) ✓
- t-test (Критерий Стьюдента)



Проверка гипотез и критерии

Если наблюдаемое значение критерия (K) принадлежит критической области ($K_{кр}$), заштрихованная область на рисунке), гипотезу отвергают, если не принадлежит - не отвергают.

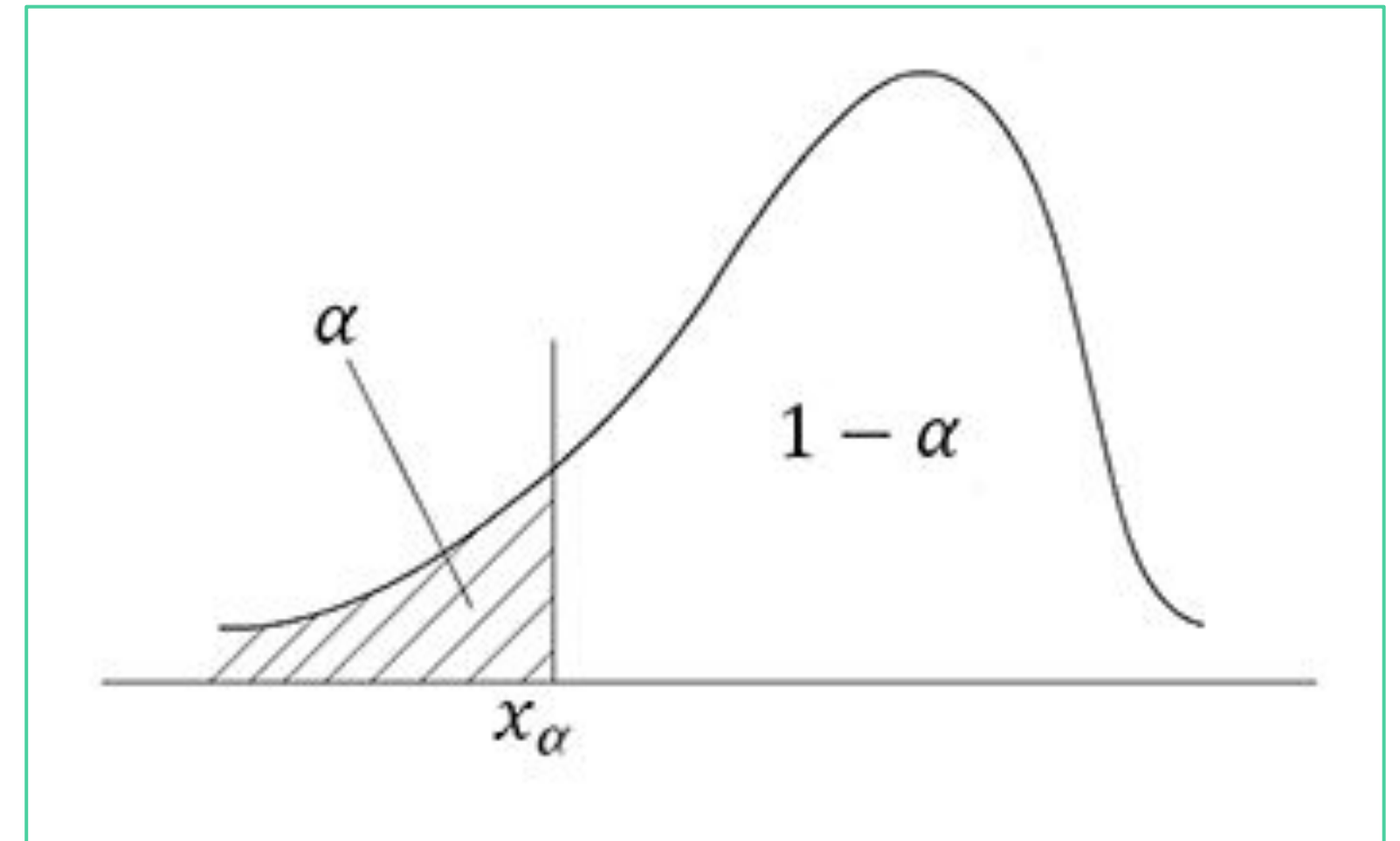
Для краткости можно записать и так:

$|K| > K_{кр}$ - отклоняем H_0

$|K| < K_{кр}$ - не отклоняем H_0

После того как данные будут собраны, значения из выборки подставляют в формулу для вычисления статистики критерия. (**об этом чуть позже**)

Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы.

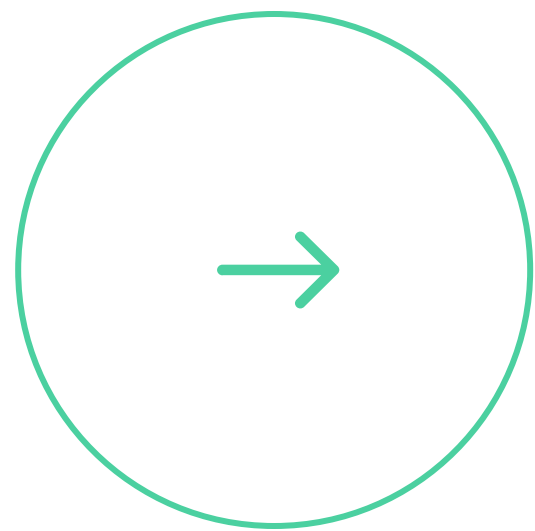


Критерием K называем величину, характеризующую отличие двух выборок

α - заданный исследователем уровень значимости.



Тесты проверки гипотез

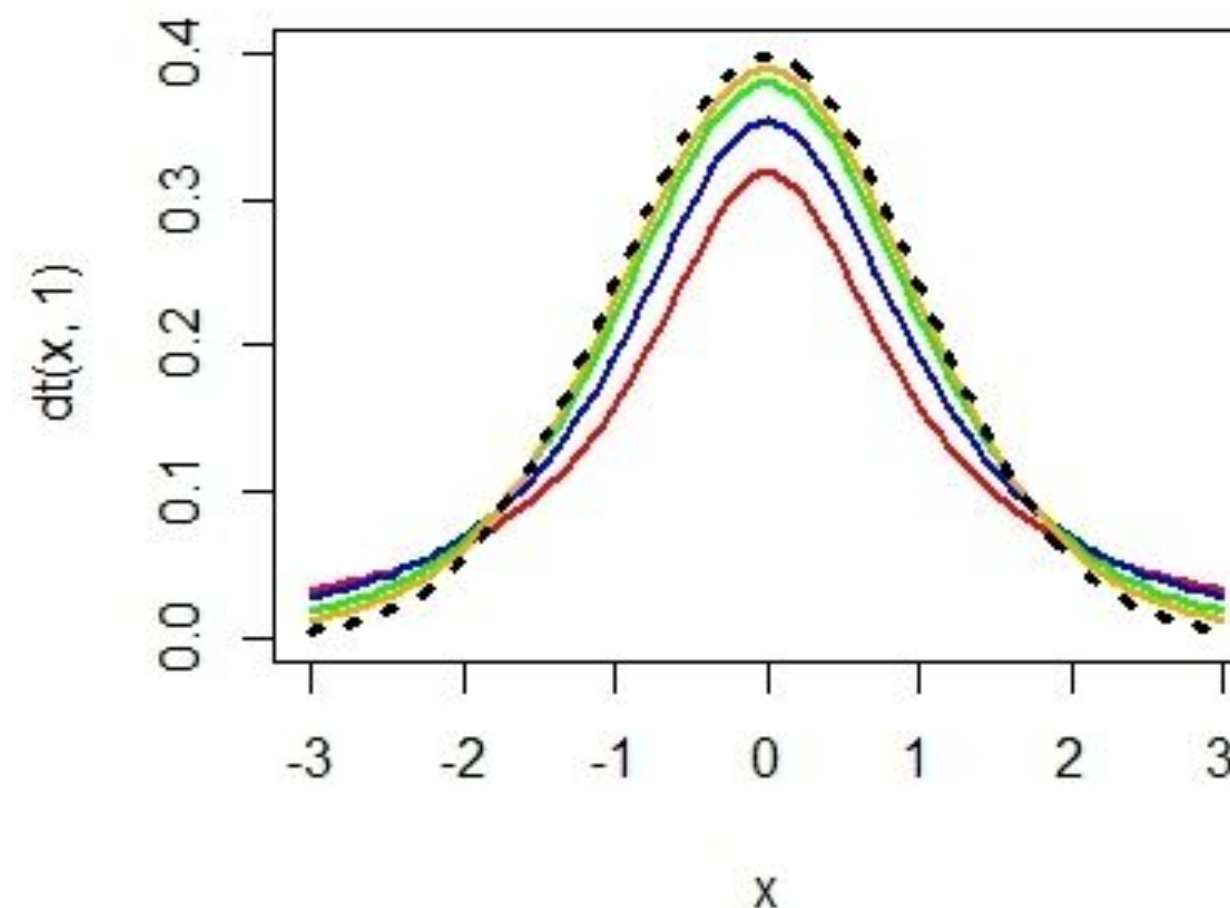


t- тест: С помощью t-теста (также называемого Т-тестом Стьюдента) сравниваются два средних значения (средних) и сообщается, отличаются ли они друг от друга. Т-тест также показывает, насколько значительны различия; Другими словами, он позволяет узнать, могли ли эти различия возникнуть случайно.

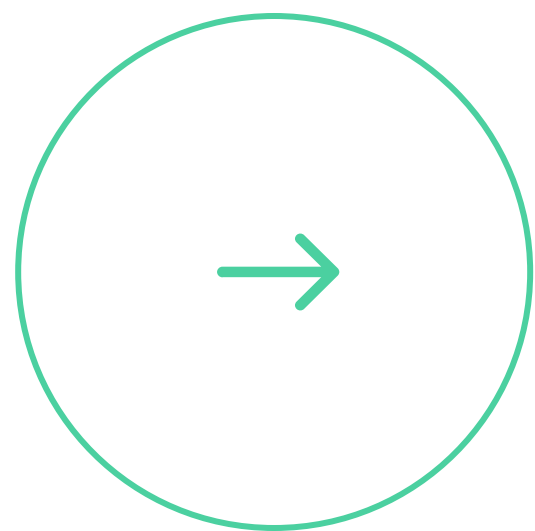
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы.

Student T distributions



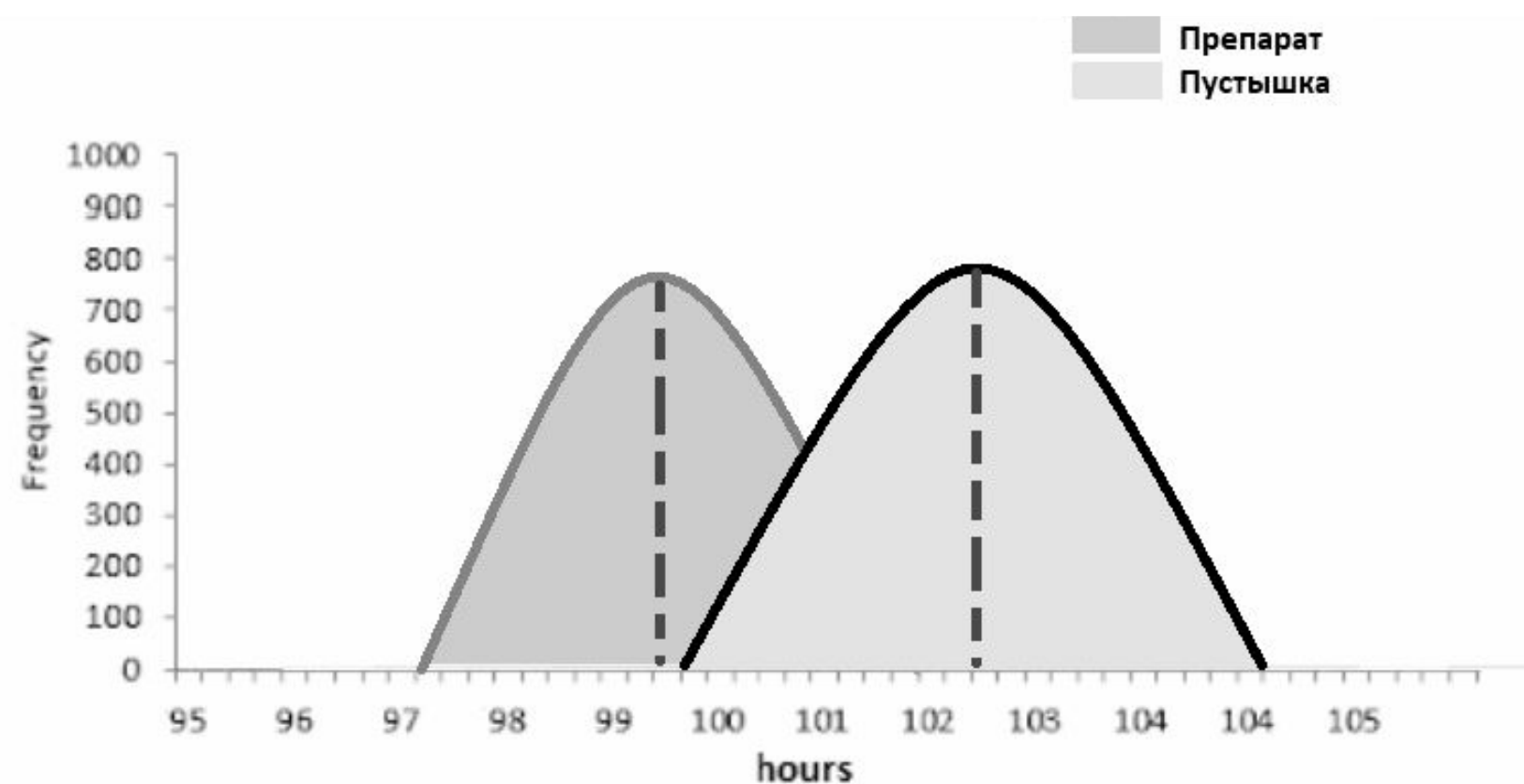
Тесты проверки гипотез



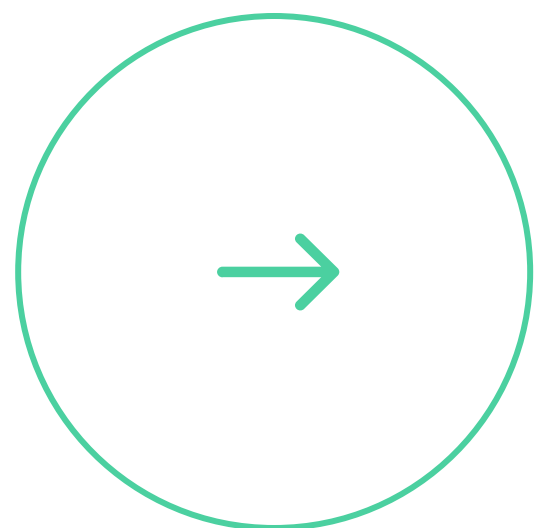
t- тест: С помощью t-теста (также называемого Т-тестом Стьюдента) сравниваются два средних значения (средних) и сообщается, отличаются ли они друг от друга. Т-тест также показывает, насколько значительны различия; Другими словами, он позволяет узнать, могли ли эти различия возникнуть случайно.

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

Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы.



Тесты проверки гипотез



t- тест: С помощью t-теста (также называемого Т-тестом Стьюдента) сравниваются два средних значения (средних) и сообщается, отличаются ли они друг от друга. Т-тест также показывает, насколько значительны различия; Другими словами, он позволяет узнать, могли ли эти различия возникнуть случайно.

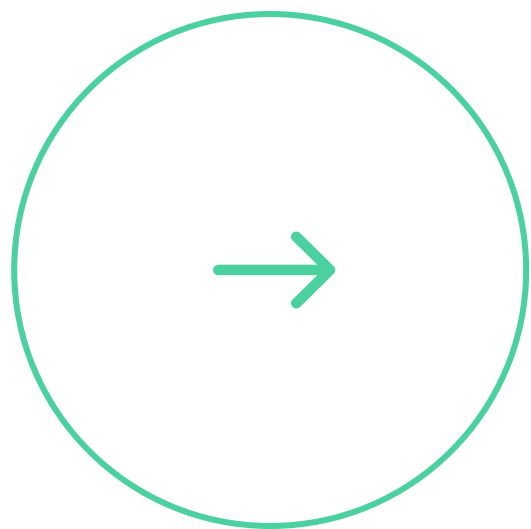
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы.

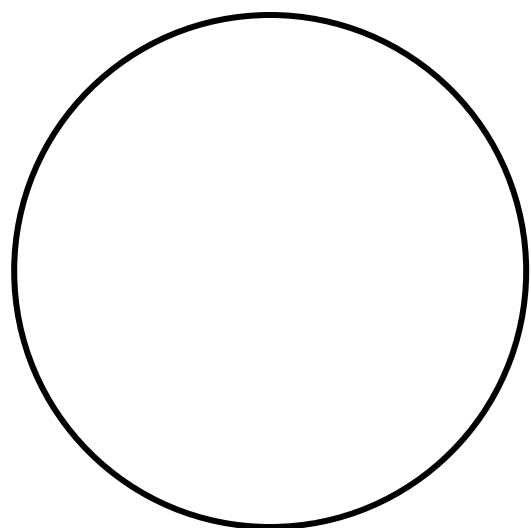
```
from scipy import stats #Подключаем библиотеку  
t, p = stats.ttest_ind(A, B) # Получаем t-value и p-value
```



Тесты проверки гипотез



Если мы хотим сравнить распределение категориальных переменных, то обычно строят таблицы сопряженности и используют критерий χ^2 .

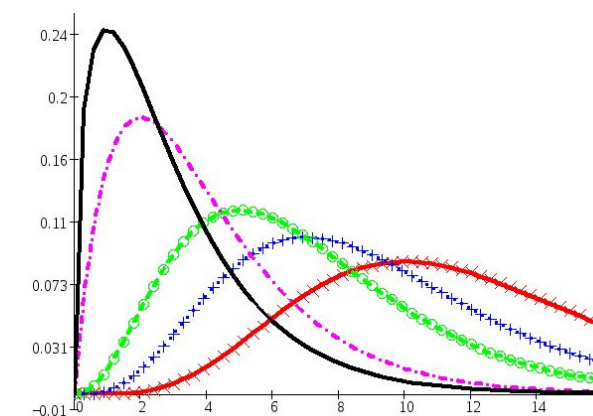


На самом деле тестов больше, но мы пока остановимся на самых важных.

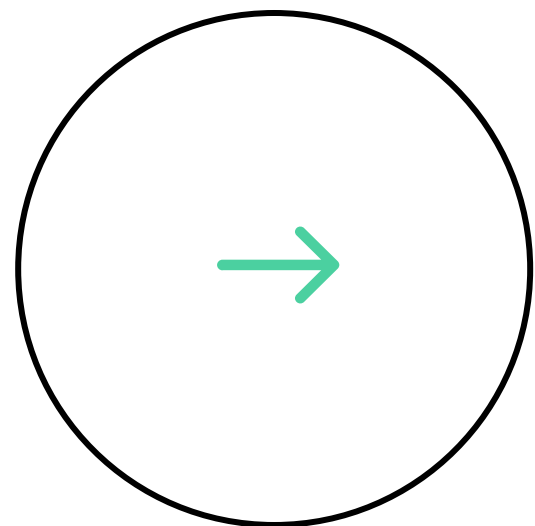
$$\chi^2 = \sum \frac{(H - O)^2}{O},$$

Если одно из ожидаемых значений меньше 5, то следует использовать тест Фишера:

107	74
93	45



Тесты проверки гипотез



Например, вы хотите проверить нет ли дискриминации пола человека в профессии Data Science?!

Вы узнали, что в какой-то компании после серии собеседований 107 мужчин взяли на работу, а отказали 93 мужчинам. А Среди женщин 74 взяли, а отказали 45

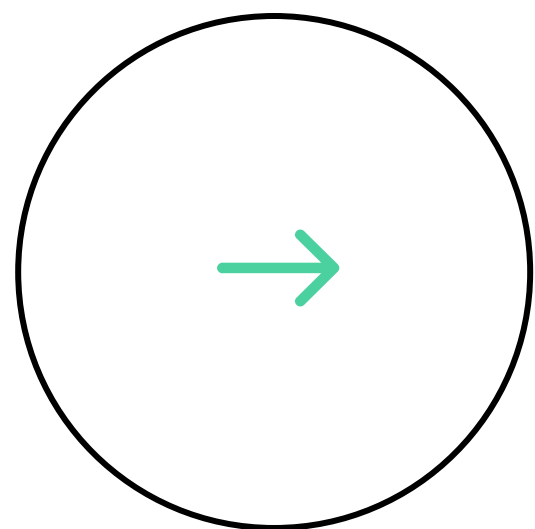
Относится ли руководство предвзято в этой компании к мужчинам или женщинам?!

$$\chi^2 = \sum \frac{(H - O)^2}{O},$$

М	Ж	
107	74	+
93	45	-



Тесты проверки гипотез



Например, вы хотите проверить нет ли дискриминации пола человека в профессии Data Science?!

Вы узнали, что в какой-то компании после серии собеседований 107 мужчин взяли на работу, а отказали 93 мужчинам. А Среди женщин 74 взяли, а отказали 45

Относится ли руководство предвзято в этой компании к мужчинам или женщинам?!

p-value = 0.10

$$\chi^2 = \sum \frac{(H - O)^2}{O},$$

М	Ж	
107	74	+
93	45	-



Шаги пройдены!

Наша цель

- Доверительный интервал ✓
- Уровень значимости ✓
- p-value (p-значение) ✓
- t-test (Критерий Стьюдента) ✓



**А как же Python?
Время практики!**



Спасибо за внимание!

Максим Сахаров

Senior Data Scientist в BasisSoft

