Тема 6. Тестирование гипотез. Часть 1

Тестирование гипотез — это один из ключевых инструментов математической статистики, позволяющий на основе выборочных данных делать обоснованные выводы о характеристиках всей генеральной совокупности.

Представьте себе ситуацию:

Маша влюблена в Борю, но не уверена — взаимно ли это? Она не может прямо спросить, но может наблюдать его поведение. Например, подарит ли он ей открытку на 14 февраля? Если подарит — гипотеза о любви поддерживается. Если нет — Маша её отвергает.

Это — **простая модель проверки гипотез**.

- Мы формулируем так называемую **нулевую гипотезу** («Боря меня любит») и **альтернативную гипотезу** («Он меня не любит»).
- Затем мы производим наблюдения. (Кстати, Боря так и не подарил ей валентинку...)
- После этого мы проводим **мысленный эксперимент**: предполагаем, что **нулевая гипотеза верна**, и оцениваем **вероятность получить такие наблюдаемые данные**, если H_0 действительно верна. («Если он меня любит, насколько вероятно, что он не подарит мне валентинку?»)
- Если эта вероятность оказывается **слишком малой**, мы отвергаем нулевую гипотезу в пользу альтернативной. («Не могу поверить, что он бы так поступил, если бы действительно любил... значит, он меня не любит!»)
- Если же вероятность **не мала**, мы **не отвергаем нулевую гипотезу** и продолжаем верить в её справедливость.

Такая логика работает и в науке. Например, чиновник утверждает: «Средняя зарплата — 55 000 ₽». Мы можем проверить эту гипотезу, собрав выборку и сравнив выборочное среднее с заявленным уровнем.

Пример 1. «Завышенная зарплата»

Чувакин — губернатор одного из российских городов. Он публично заявляет, что **средняя месячная зарплата** в его городе составляет **55 000 рублей**. Это заявление вызывает сомнение: есть подозрение, что **реальная зарплата ниже**.

Мы хотим проверить это утверждение. Обозначим зарплату одного жителя через X, а **истинное среднее значение** зарплаты через μ. Предположим также, что на основе ранее собранных данных известно, что **стандартное отклонение в генеральной совокупности** равно

$$\sigma = 15\,500 \text{ рублей.}$$

Шаг 1: Формулируем гипотезы

- **Нулевая гипотеза** H_0 отражает официальное утверждение, привязанное к конкретному числу.
- **Альтернативная гипотеза** H_a выражает **наше подозрение** в сторону отклонения от этого утверждения.

В данной ситуации:

- Гипотеза Чувакина: средняя зарплата = 55 000,
- Наше подозрение: **зарплата завышена**, то есть $\mu < 55\,000$.

Итак, гипотезы будут следующими:

$$H_0 : \mu = 55\,000$$

$$H_a : \mu < 55\,000$$

Далее производятся соответствующие расчеты.

Определение: Тестирование статистических гипотез

Тестирование статистических гипотез — это метод математической статистики, предназначенный для принятия решения о правдоподобности предположения о параметрах генеральной совокупности на основе выборочных данных.

Иначе говоря:

🔍 Это процедура, при которой мы проверяем, насколько **достоверны утверждения о популяции**, опираясь на **ограниченную информацию из выборки**.

Формально:

Пусть θ — неизвестный параметр распределения (например, математическое ожидание, доля, дисперсия и т.д.).

Мы хотим проверить утверждение вида:

- $H_0: \theta = \theta_0$ — **нулевая гипотеза**
- H_a — **альтернативная гипотеза**, утверждающая, что $\theta \neq \theta_0$ $(\text{или } > \theta_0, \text{или } < \theta_0)$.

Задача — **на основе выборочных данных** решить, **принимать ли нулевую гипотезу, или отвергнуть её** в пользу альтернативной.

Интуитивно:

Представьте, что нулевая гипотеза — это **предположение, в которое мы "верим" по умолчанию**.
Альтернативная — то, что **мы хотим доказать**, но должны иметь достаточно веские основания, чтобы отвергнуть H_0.

Таким образом, **тестирование гипотез** — это **научный и формализованный способ** принимать решения в условиях неопределённости, **основываясь на вероятностях и выборочной информации**.

Вернемся к Васе Чувакину:

Он заявляет, что **средняя месячная зарплата** в его городе составляет **55 000 рублей**. Мы хотим проверить, что **реальная зарплата ниже**.

Обозначим зарплату одного жителя через X, а **истинное среднее значение** зарплаты через μ. Предположим также, что на основе ранее собранных данных известно, что **стандартное отклонение в генеральной совокупности** равно

$$\sigma=15\,500 \text{ рублей}$$

Шаг 1: Формулируем гипотезы

- **Нулевая гипотеза** H_0 отражает официальное утверждение, привязанное к конкретному числу.
- **Альтернативная гипотеза** H_a выражает **наше подозрение** в сторону отклонения от этого утверждения.

В данной ситуации:

- Гипотеза Чувакина: средняя зарплата = 55 000,
- Наше подозрение: **зарплата завышена**, то есть $\mu < 55\,000$.

Итак, гипотезы будут следующими:

$$H_0: \mu \ = \ 55\,000$$

$$H_a: \mu \ < \ 55\,000$$

Шаг 2. Уровень значимости α

Перед началом проверки гипотезы необходимо выбрать **уровень значимости** — это число, обозначаемое как α. Оно играет роль **порога**, по которому мы принимаем решение **отклонять нулевую гипотезу или нет**.

📌 Уровень значимости α — это наша **субъективная граница "маловероятных событий"**.
Если вероятность получить наблюдаемые данные при условии, что H_0 верна, **меньше α**, то результат считается **слишком редким** — и H_0 отвергается.

Как интерпретировать α?

- $\alpha - alpha$ — это **вероятность ошибки первого рода**: то есть, вероятность **ошибочно отклонить верную H_0**.
- Другими словами, если вы выбрали $\alpha = 0.05$, вы готовы **принять риск 5%** ошибочного вывода.

Типичные значения α:

Название	Значение α	Типичная область
Очень строгий тест	0.01 (1%)	Медицинские исследования, безопасность
Стандартный	0.05 (5%)	Наука, экономика
Умеренно допускающий ошибку	0.10 (10%)	Соц. науки, предварительные гипотезы

В нашем примере выберем **уровень значимости 5%**, то есть:

$$\alpha = 0.05$$

Шаг 3. Сбор и описание данных

Пусть была проведена **простая случайная выборка (SRS)** из **40 граждан** города. Их зарплаты были измерены, и получено:

- Среднее значение: $\hat{X} = 35\,000$,
- Напомним, что по условию: $\sigma=15\,500$,
- Размер выборки: $n=40$.

Так как $n>30$, можно использовать **центральную предельную теорему** и предположить, что:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Это означает, что распределение выборочного среднего **приближённо нормально**, что позволяет нам использовать **z-критерий**.

Шаг 4. Вычисление z-статистики

Переходим к **основному вычислению** — стандартной z-статистики, которая показывает, насколько далеко находится полученное значение от предполагаемого среднего из H_0.

Формула z-статистики:

$$z_{\text{ст}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Подставим значения:

$$z_{\text{ст}} = \frac{35\,000 - 55\,000}{15\,500/\sqrt{40}} = \frac{-20\,000}{2450.77} \approx \boxed{-8.16}$$

Интерпретация результата:

1. Мы видим, что полученное среднее отклоняется **на более чем 8 стандартных отклонений** от заявленного уровня 55 000.
2. Это **чрезвычайно редкое событие**, если предположить, что H_0 верна.
3. Важно: **все вычисления** ведутся **при условии**, что H_0 верна.

Шаг 5. Определение критического значения

После того как мы выбрали уровень значимости α=0.05, мы **отсекаем** соответствующую область на графике плотности нормального распределения. Так как в нашем случае альтернативная гипотеза **односторонняя (влево)**:

$$Ha: \mu < 55\,000$$

то и область отклонения гипотезы будет **слева**.

Определение Zкр:

$$P(Z < z_{\text{кр}}) = \alpha = 0.05 \quad \Rightarrow \quad z_{\text{кр}} = -z_{0.05} \approx \boxed{-1.645}$$

Сравнение:

Ранее мы нашли:

$$z_{\text{ст}} = -8.16$$

Теперь сравним:

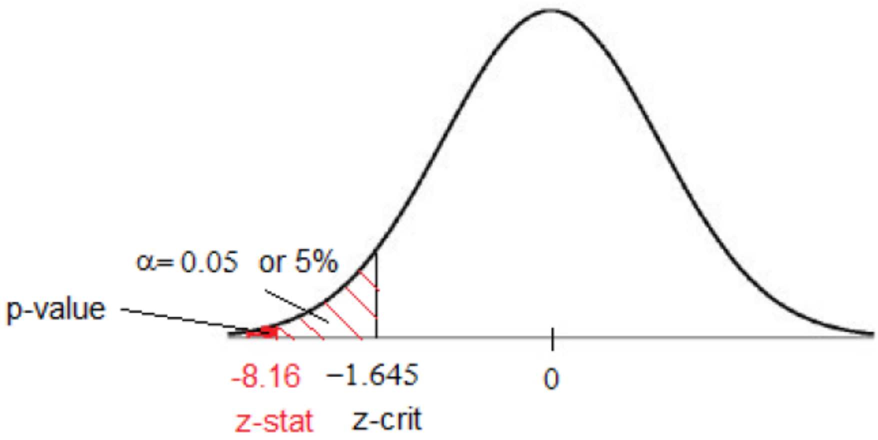
$$z_{\text{ст}} < z_{\text{кр}} \quad \text{или} \quad |z_{\text{ст}}| > |z_{\text{кр}}|$$

$\Rightarrow z_{\text{ст}}$ попадает в область отклонения.

Вывод:

Так как:

$$z_{\text{ст}} = -8.16 < -1.645 = z_{\text{кр}}$$



$$\boxed{H_0 \text{ отвергается в пользу } H_a \text{ на уровне значимости } \alpha = 0.05}$$

Вывод: Есть основания отвергнуть нулевую гипотезу на уровне значимости 5%. Губернатор Вася Чувакин – врал насчет того, что зарплата составляет 55000 рублей, в действительности они меньше.

Шаг 6. Вычисляем p-value

P-value — это **вероятность получить наблюдаемый результат или ещё более экстремальный**, при условии, что **нулевая гипотеза H_0 верна**.

Формальное определение:

$$\text{p-value — это } P(\text{данные или более экстремальные} \mid H_0 \text{ верна})$$

Важно понимать:

- p-value **не** говорит о вероятности, что H_0 верна или ложна.
- Он **не** зависит от Ha, а только от того, насколько маловероятны полученные данные **если H_0 — истина**.
- Чем **меньше** p-value, тем **меньше вероятность**, что такой результат мог получиться случайно, если H_0 верна.
- Если p-value **меньше** уровня значимости α, то H_0 **отвергается**.

Примерная интерпретация:

p-value	Интерпретация
> 0.10	Нет оснований отклонять H_0
0.05 – 0.10	Слабое свидетельство против H_0
0.01 – 0.05	Умеренное свидетельство против H_0
< 0.01	Сильное свидетельство против H_0
< 0.001	Очень сильное (почти невозможное событие под H_0)

Теперь расчёт p-value для нашего примера:

$$z_{\text{ст}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{35\,000 - 55\,000}{15\,500 / \sqrt{40}} \approx \frac{-20\,000}{2450.77} \approx \boxed{-8.16}$$

Так как альтернативная гипотеза была **левосторонняя** ($H_a : \mu < 55\,000$), то:

$$\text{p-value} = P(Z \le -8.16)$$

Значение очень малое

Для $z = -8.16$ соответствующее значение в таблице нормального распределения:

$$P(Z \le -8.16) \approx \boxed{1.7 \cdot 10^{-16}}$$

Вывод:

$$\boxed{\text{p-value} = 1.7 \cdot 10^{-16} \ll \alpha = 0.05 \Rightarrow \text{гипотеза } H_0 \text{ отвергается}}$$

Пример 2. Заниженное среднее количество бюрократов в проекте

Чувакин утверждает, что штат его сотрудников — вполне скромный. Он заявляет, что в среднем **на один городской проект приходится не более 25 человек**. Однако активист проверил информацию: взял **выборку из 20 проектов**, проведённых в прошлом году, и обнаружил, что **среднее число участников** по официальным документам составляет **37 человек на проект**.

🤔 Возможно, Чувакин снова лукавит...

Шаг 1. Формулируем гипотезы и выбираем уровень значимости

Обозначим:

- **X** — число участников в одном проекте,
- **μ** — истинное среднее число участников.

С учётом утверждения Чувакина "не более 25", нулевая гипотеза (всегда с равенством) формулируется как:

$$H_0 : \mu = 25$$
$$H_a : \mu > 25$$

Альтернатива односторонняя, направленная в сторону подозрения — что участников больше, чем он заявил.

Выбираем уровень значимости: $\alpha = 0.05$

Шаг 2. Данные из выборки

По данным активиста:

- $\bar{X} = 37$
- n=20
- σ=14 (известна по прошлым наблюдениям)

Шаг 3. Проверка условий применения нормального приближения

(а) Предполагаем, что данные были получены в результате простой случайной выборки, а объём выборки не превышает 5% от всей совокупности проектов (условие независимости).

(б) Так как n=20<30, центральная предельная теорема не гарантирует нормальность распределения выборочного среднего.

Поэтому мы должны дополнительно предположить, что исходная случайная величина

$$X \sim N(\mu, \sigma^2).$$

Тогда выборочное среднее $\hat{X}$ также будет нормально распределено.

Шаг 4. Расчёт z-статистики

Формула:

$$z_{\text{ст}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{37 - 25}{14/\sqrt{20}} = \frac{12}{3.130} \approx \boxed{3.83}$$

Шаг 5. Вычисление критического значения

Критическое значение при α=0.05 для одностороннего теста (справа):

$$z_{\text{кр}} = z_{0.05} \approx 1.645$$

Шаг 6. Вывод

Так как:

$$z_{\text{ст}} = 3.83 > 1.645 = z_{\text{кр}}$$

Вывод: Есть основания отвергнуть нулевую гипотезу на уровне значимости 5%. Губернатор Вася Чувакин – врал насчет того, что **на один городской проект приходится не более 25 человек.**

Шаг 7. Расчет p-value

Альтернативная гипотеза односторонняя (вправо), поэтому:

$$\text{p-value} = P(\bar{X} > 37 \mid H_0) = P(Z > 3.83)$$

По таблице стандартного нормального распределения:

$$\text{p-value} \approx \boxed{0.000064}$$

Это **крайне маловероятное** значение, поэтому есть основания отвергнуть нулевую гипотезу на уровне значимости 5%. Губернатор Вася Чувакин – врал насчет того, что **на один городской проект приходится не более 25 человек.**

Пример 3. Двусторонняя альтернатива

Преподаватель Руслан после одного семестра, проведенного в работе с группами ШАД и ШЦТ, решил проверить – как много балбесов слушают его лекции (или не слушают).

По полученным оценкам за контрольные, а также в ходе диалогов со студентами, наблюдением за тем, как одни из них пытались выбросить ручку от окна в окно, проливали энергетик со вкусом поп-корна в аудитории, чтобы получить атмосферу на парах как в кино (заставляет задуматься, что фамилия у Руслана Кубрик или Нолан), за тем как студенты встречали его (Руслана) с ведрами на голове, было оценено, что **среднее количество таких студентов – 12**.

Дополнительно известно, что стандартное отклонение таких студентов:

$$\sigma = 5$$

Шаг 1. Формулируем гипотезы

Пусть:

- X — уровень балбесесности одного студента,
- μ — истинное среднее количество балбесов в группе.

Согласно отчёту:

$$H_0 : \mu = 12.0$$
$$H_a : \mu \neq 12.0$$

Обратите внимание:

Здесь нет чёткого подозрения, в какую сторону может быть отклонение — больше или меньше.

*Поэтому **альтернативная гипотеза — двусторонняя**.*

Выбираем уровень значимости:

$$\alpha = 0.05$$

Шаг 2. Данные из выборки

- Объём выборки: n=25
- Выборочное среднее: $\bar{X} = 10.8$
- Стандартное отклонение (по населению): $\sigma = 5$

Шаг 3. Проверка условий применения нормального распределения

1. Предполагается, что данные получены по **простой случайной выборке (SRS)**, а объём составляет не более 5% от совокупности (независимость).
2. $n = 25 < 30$, то есть выборка **небольшая**. Поэтому нужно дополнительно предположить, что **распределение удойности нормальное**:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Тогда выборочное среднее $\bar{X}$ также будет нормально распределено.

Шаг 4. Расчёт z-статистики

$$z_{\text{ст}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{10.8 - 12.0}{5/\sqrt{25}} = \frac{-1.2}{1} = \boxed{-1.2}$$

Шаг 5. Критическое значение

Для двусторонней альтернативы мы делим уровень значимости пополам:

$$\alpha/2 = 0.025$$

По таблице:

$$z_{\text{кр}} = \pm z_{0.025} = \pm 1.96$$

Сравниваем:

$$|z_{\text{ст}}| = 1.2 < 1.96 = |z_{\text{кр}}| \Rightarrow \boxed{H_0 \text{ не отвергается}}$$

Шаг 6. Вывод

Нет оснований отвергнуть нулевую гипотезу на уровне значимости 5%. Среднее количество балбесов в группе действительно 12 студентов.

Шаг 7. Расчет p-value

Так как альтернативная гипотеза двусторонняя, нас интересует отклонение в любую сторону от значения 12.0. Поэтому используем:

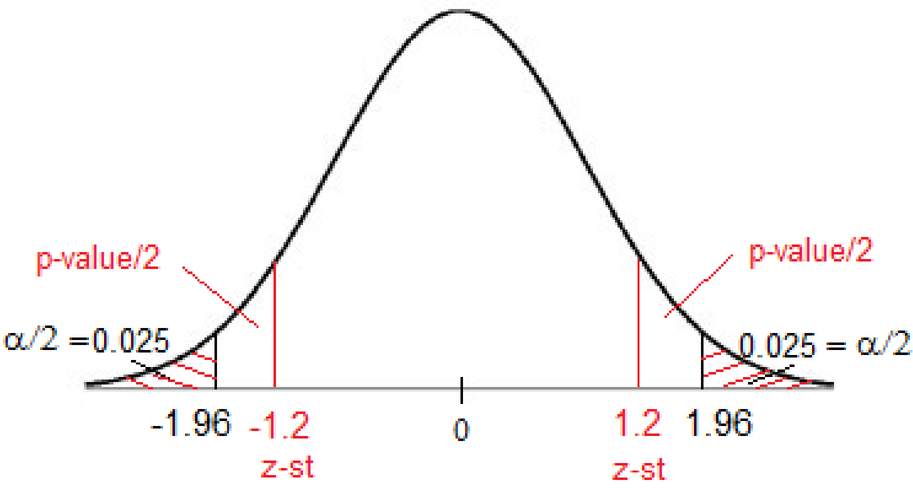
$$\text{p-value} = P(|Z| > 1.2) = 2 \cdot P(Z > 1.2)$$

По таблице нормального распределения:

$$P(Z > 1.2) \approx 0.115 \Rightarrow \boxed{\text{p-value} = 2 \cdot 0.115 = 0.23}$$

Интерпретация:

Такая вероятность означает, что **при верной H_0** мы **довольно часто** могли бы получить результат в 10.8 или ещё дальше от 12. Нулевая гипотеза принимается.



Проверка статистической гипотезы — пошаговый алгоритм

(1) **State** hypotheses and **choose** significance level:

H0: ...

Ha: ...

Let α=...

(5) Introduce a variable and **write down** given statistics

(6) **Check** requirement and state assumptions

(7) **Calculate** z-statistic

(8) **Find** p-value

(9) **Compare** p-value with α (or Zst with Zcr) and conclude about the H0.

Ошибки первого и второго рода

«Я бы не стал умирать за свои убеждения — вдруг я ошибаюсь» — Бертран Рассел

Как уже говорилось ранее, любое статистическое решение, принятое в рамках проверки гипотезы, может оказаться ошибочным. Мы всегда сталкиваемся с рискованным выбором: принять или отвергнуть гипотезу на основе ограниченных данных.

Существует два типа ошибок, связанных с тестированием гипотез:

Ошибка I рода (Type I Error)

Ошибка I рода — это ситуация, когда отвергается нулевая гипотеза, хотя на самом деле она верна.

$$\text{Тип I: отклонить } H_0, \text{ когда } H_0 \text{ истинна}$$

Пример: «Любит — не любит»

Маша хочет проверить гипотезу:

- H_0: *Боря её любит*.
- На: *Боря не любит*.

Пусть чувства Бори можно измерить и выразить через случайную величину X, которая подвержена колебаниям настроения, внешним факторам и личной тревожности. Предположим, что:

- $X \sim \mathcal{N}(\mu_X, \sigma^2)$,
- при H_0 значение $\mu_X = \heartsuit$ (условная величина "любовь").

Критерий: подарит ли Боря валентинку?

Маша считает: если Боря **не подарил валентинку**, значит, он **не любит**.

Но возможна **ошибка I рода**: даже **влюблённый Боря** может не подарить валентинку:

- забыл дату,
- потерял открытку по дороге,
- постеснялся.

Что происходит в случае ошибки I рода?

- Маша **отвергает** гипотезу любви, хотя **она верна**.
- Это — классическая **ошибка I рода**.

Вероятность ошибки I рода

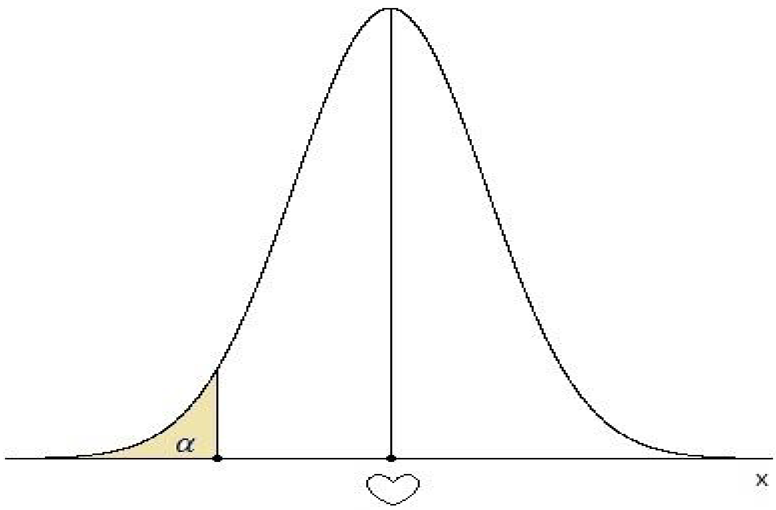
Вероятность такого события обозначается:

$$\alpha$$

- Это **площадь под графиком плотности** нормального распределения X **слева от порогового значения**, соответствующего событию "валентинка подарена".
- Именно **этот уровень значимости α** мы выбираем при проверке гипотезы.

Вывод:

- **Ошибка I рода** — отклонение правдивой гипотезы.
- Она возникает **с вероятностью α** — и именно **эту ошибку мы контролируем при построении теста**.
- Чем меньше α, тем **реже** будет происходить ошибка I рода, но тем **больше риск ошибки II рода** (о ней — в следующем блоке).



Ошибка II рода (Type II Error)

Определение:

Ошибка II рода возникает, когда мы **не отвергаем нулевую гипотезу**, хотя на самом деле она **ложна**.

$$\text{Тип II: не отвергнуть } H_0, \text{ хотя верна } H_a$$

💌 **Продолжение истории о Маше и Боре:**

Предположим, что **на самом деле Боря Машу не любит**. Это означает, что верна **альтернативная гипотеза** На, и истинное значение его чувств смещено **влево от значения μ=♡** — например, к μ=♡.

Но вот **парадокс**: даже **не влюблённый** Боря **может подарить ей валентинку**:

- может быть, он вежливый и дарит открытки всем,
- возможно, он не хочет обидеть Машу перед её подругами,
- или просто в хорошем настроении...

📌 В этом случае Маша **получает открытку** и делает вывод:

«Наверное, он меня любит» — и **не отвергает H_0**.

Но на самом деле H_0 **ложна**, и Маша совершает **ошибку II рода**.

Вероятность ошибки II рода

Обозначается:

$$\beta$$

- Это **площадь под кривой плотности распределения**, соответствующего альтернативной гипотезе На, **справа от порогового события** (например, "открытка подарена").
- Чем **больше β**, тем выше вероятность **не заметить** истинное отличие.

Баланс между ошибками I и II рода

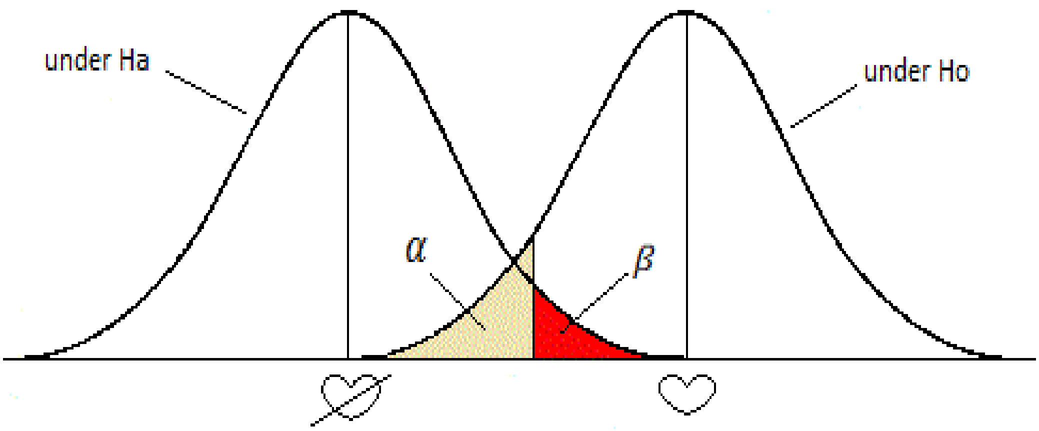
❗**Важно:**

- **Обе ошибки зависят от выбранного порога (критерия)**.
- Если Маша решит, что **открытки недостаточно**, и будет считать доказательством только **подаренные цветы**, то:
  - Порог сдвигается вправо,
  - Вероятность ошибки I рода α **возрастает** (больше «любящих» окажутся отвергнутыми),
  - Но вероятность ошибки II рода β **снижается** (труднее поверить в любовь по мелочи).

Если наоборот — Маша будет радоваться даже «лайку» в *instagram*:

- Порог сдвигается влево,
- **α** падает, но **β** растёт.

Это и есть **компромисс между двумя типами ошибок**.



А можно ли уменьшить и α, и β?

Да! Есть **только один способ**:

Увеличить объём выборки.

- Тогда стандартное отклонение статистики станет меньше,
- Распределения станут **уже**, то есть «тоньше»,
- Соответственно, **заштрихованные области ошибок уменьшатся**.

Все 4 возможных исхода проверки гипотезы:

	H_0 верна	H_1 верна
H_0 отвергнута	❌ Ошибка I рода (вер. α)	✅ Правильное решение (вер. $1-\beta$)
H_0 принята	✅ Правильное решение (вер. $1-\alpha$)	❌ Ошибка II рода (вер. β)

Обозначения:

- α — уровень значимости (вероятность ошибки I рода),
- $1-\alpha$ — доверительный уровень (насколько мы уверены в H_0),
- β — вероятность ошибки II рода,
- $1-\beta$ — мощность критерия (вероятность правильно обнаружить отличие, если оно есть).