

Лекция №5 Тестирование гипотез (параметрические тесты)

Цель лекции:

- ✓ Научиться формулировать гипотезы
- ✓ Изучить параметрические тесты и условия их применимости
- ✓ Провести тестирование гипотез с использованием t и Z критерия
- ✓ Научиться правильно интерпретировать результат с использованием величины p -value
- ✓ Сделать проверку на нормальность с помощью графического метода и с помощью теста Шапиро-Уилка

Материал прошлого урока:

Последним, что мы изучили, были нормальное распределение и центральную предельную теорему. Сегодня будем знакомиться с параметрическими критериями, используемыми для тестирования гипотез. Мы будем сегодня не раз обращаться к изученному материалу, потому что применение параметрических тестов подразумевает нормальность генеральной совокупности, из которой мы брали выборки, а центральная предельная теорема легла в основу тестирования гипотез.

План урока:

1. Алгоритм проведения тестирования гипотез
2. Нулевая гипотеза H_0 и альтернативная гипотеза H_1
3. Выбор уровня статистической значимости α
4. Выбор статистического критерия и вывод
5. Ошибки при тестировании гипотез.
6. Критерий Стьюдента.
7. P -value
8. Виды статистических гипотез
9. Проверка на нормальность (Шапиро - тест и QQ-график)

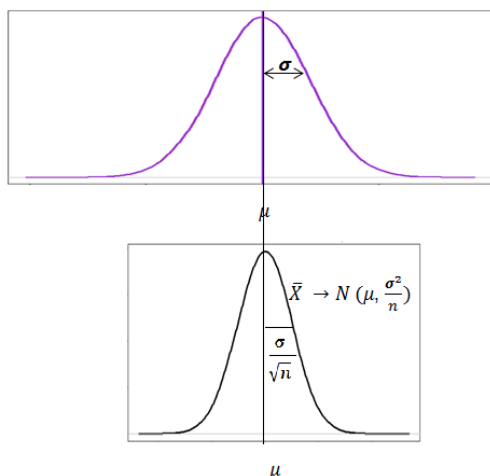
Алгоритм проведения тестирования гипотез

На предыдущих лекциях я упоминала, что обычно у нас нет доступа ко всей генеральной совокупности по ряду причин (дорого, долго, после измерения объект нельзя вернуть обратно в совокупность). И поэтому в реальной жизни часто прибегают к выборочному контролю. Предположим, у нас есть случайная величина (далее СВ) «Рост». Мы берем первую выборку, измеряем по ней среднее арифметическое и получаем некоторую точечную оценку генеральной совокупности, затем можем взять еще одну выборку и получить по ней новую точечную оценку генеральной совокупности, потом взять третью выборку и снова получить какую-то новую точечную оценку генеральной совокупности и т.д. Т.е. точечная оценка будет меняться от выборки к выборке. Т.е. средняя по выборке есть случайная величина. (Мы это уже рассматривали, когда изучали центральную предельную теорему). Значит, мы не можем взять какую-то одну точечную оценку и сказать, что вот именно эта точечная оценка и будет истинным средним арифметическим всей генеральной совокупности. Не можем, потому что мы этого не знаем наверняка. И вот здесь нам как раз и помогает такой метод, как тестирование гипотез. Он позволяет нам сравнивать

параметры распределений. Т.е. с помощью тестирования гипотез мы хотим ответить на вопрос, а видим ли мы статистически значимые различия, например, между 2мя средними арифметическими. Что это значит?

Давайте сразу приведем пример. По аутсорсингу закупаются детали у поставщика. Поставщик заявляет, что он выпускает детали с диаметром 10 мм. Прежде, чем делать большую закупку, проведем контроль: проверим, правда ли детали выпускаются с размером 10 мм. Для этого взяли выборку из 16 деталей и получили средний диаметр по этой выборке 10.3 мм. И вот вопрос, а является ли разница в 0.3 мм статистически значимой или нет.

Еще раз вспомним центральную предельную теорему, которая гласит, что среднее выборочное стремится к нормальному распределению с тем же, как у генеральной совокупности и дисперсией, равной дисперсии генеральной совокупности, деленной на объем выборки. Т.е. средняя выборочная – это случайная величина, т.е. меняется от выборки к выборке, иными словами, это не фиксированное одно и то же значение, которое мы будем получать во всех выборках. (Еще раз посмотрите на знакомый уже рисунок, где изображено распределение среднего выборочного значения)



А значит, мы хотим понять, а являются ли различия случайными или, все-таки, наше значение по выборке принадлежит другой генеральной совокупности, т.е. с другим размером диаметра деталей. Т.о. если различия случайны, то мы считаем, что различий между тем, что мы измерили и тем, что заявлено поставщиком о генеральной совокупности, не являются статистически значимыми. Но если различие велико, то мы обнаруживаем статистически значимое различие.

При тестировании гипотез формулируется 2 гипотезы. Одна из них нулевая, которая гласит, что различий нет, и противоположная гипотеза, альтернативная, которая заявляет, что есть статистически значимые различия.

Метод тестирования гипотез имеет определенный алгоритм.

1. Формулирование нулевой H_0 и альтернативной гипотез H_1
2. Выбор уровня статистической значимости α
3. Выбор статистического критерия
4. Расчет наблюдаемого критерия

5. Сравнение табличного и наблюдаемого значений критерия
6. Вывод

Сейчас мы пройдемся по каждому из этих этапов.

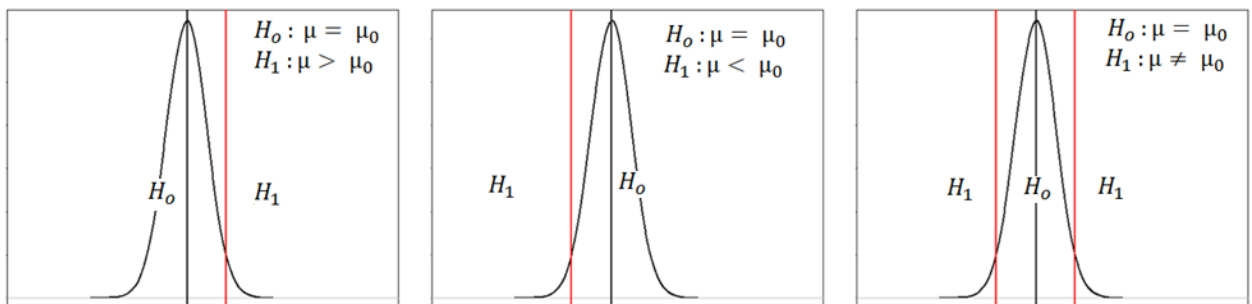
Нулевая гипотеза H_0 и альтернативная гипотез H_1

Нулевая и альтернативная гипотезы – это две противоположные гипотезы. Если принимается нулевая гипотеза, то альтернативная гипотеза отвергается, и наоборот, если принимается альтернативная, то нулевая отвергается.

Нулевая гипотеза H_0 всегда звучит едино. Различий нет. Математически мы можем записать это, как $\mu = \mu_0$. Пока не будет доказано, что верна альтернативная гипотеза, верной считается H_0 .

В альтернативную гипотезу H_1 мы вкладываем наше заявление о статистически значимых различиях. Она может быть сформулирована тремя способами, все зависит от поставленной задачи.

- 1) $\mu > \mu_0$ (вариант для нашего примера, т.к. мы получили $\mu = 10.3$ больше, чем заявлял производитель)
- 2) $\mu < \mu_0$
- 3) $\mu \neq \mu_0$ (двусторонний тест)



Выбор уровня статистической значимости α

Выбор уровня статистической значимости α делается на усмотрение того, кто проводит тест. Обычно это 0.01 (1%), 0.05 (5%) или 0.1 (10%). Если нет никаких предпосылок, то смело берите 0.05. Отсечка α разделяет график на область принятия нулевой гипотезы H_0 и альтернативной гипотезы H_1 . На рисунке выше эта отсечка представлена красной линией. Хочу заметить, что на рисунках изображены распределения не генеральных совокупностей, а распределения средних выборочных. Т.е. если мы выбираем $\alpha = 0.05$, это означает, что мы отсекаем 5% значений средних по выборкам у распределения, соответствующего гипотезе H_0 , причем эти 5% значений будут попадать теперь в область принятия гипотезы H_1 . Вернемся к этому, когда будем говорить об ошибках.

Важно помнить, что уровень статистической значимости всегда выбирается заранее.

Выбор статистического критерия и вывод

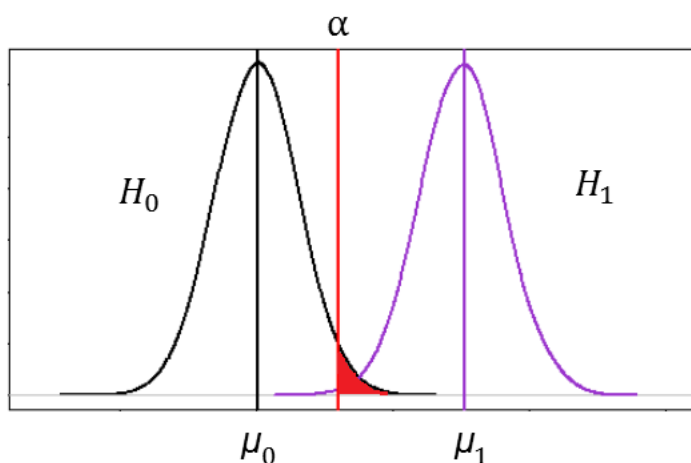
После того, как вы определились с гипотезами, перед вами встает задача правильно определиться с критерием. Критерии делятся на параметрические и непараметрические. Параметрические критерии – это Z, t (критерий Стьюдента) для сравнения двух выборок. Они предполагают, что выборка взята из нормально распределенной генеральной совокупности, более того предполагают равенство дисперсий в группах. Если эти условия не выполняются то лучше воспользоваться непараметрическим тестом, основанным на сумме рангов, например, критерием суммы рангов Манна—Уитни. Но сегодня речь пойдет о наиболее используемых параметрических тестах: Z и t критериях.

Вся идея теста сводится к тому, чтобы найти расчетное значение и сравнить его с табличным значением, соответствующим выбранному уровню значимости α .

Если у нас условия применимости параметрических тестов соблюдены, то в случае известной сигмы генеральной совокупности более предпочтителен Z- критерий, в противном случае - критерий Стьюдента t. Студенты часто спрашивают, а откуда может быть известна сигма генеральной совокупности. Ответ: из предыдущих исследований или, если мы выпускаем нашу продукцию давно и у нас накоплено очень много данных о ее некотором параметре, то рассчитанную на практике сигму мы можем принять за истинное значение.

Вернемся к задаче, которую мы будем использовать в качестве примера на протяжении всего этого урока.

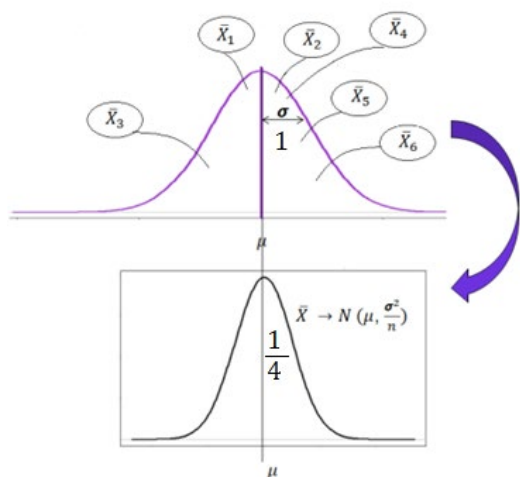
Например, для нашей задачи о деталях с диаметром 10 мм график будет выглядеть следующим образом:



Мы знаем, что диаметры следуют нормальному распределению и пусть у нас известно среднее квадратичное отклонение генеральной совокупности, равное 1 мм. Значит, нам следует использовать критерий Z.

Мы знаем, что производитель заявляет среднее арифметическое 10 мм, но мы получили 10.3 мм и к тому же мы знаем сигму генеральной совокупности. Давайте измерим расстояние в универсальной единице измерения в нормальном распределении, а именно посмотрим, в скольких сигмах лежит 10.3 мм от 10 мм. $10.3 - 10 = 0.3$ мм. Сколько же сигм лежит в этом отрезке 0.3 мм.

Чему равна сигма? Помните, я выше писала, что мы на графиках изображаем распределение средних выборочных. В условии задачи нам дана сигма генеральной совокупности 1 мм. Значит, сигма для изображенного ниже распределения на рисунке, будет являться стандартной ошибкой среднего (SE). Согласно центральной предельной теореме $SE = \sigma/\sqrt{n}$. Поскольку выборка была из 16 деталей, то получаем 1/4



Осталось посчитать в скольких SE 10.3 лежит от 10 мм.

$$(10.3-10)/(1/4) = 1.2$$

1.2 – это наш расчетный Z критерий по выборке.

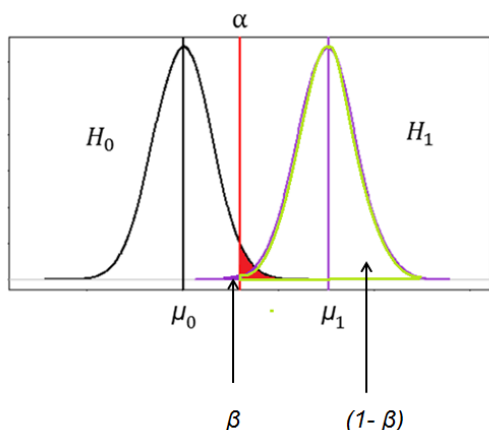
А вот табличный критерий надо найти для $\alpha = 5\%$. Поскольку на рисунке α отсекает 0.05 справа, то ниже отсечки лежат 0.95 или 95% всех значений стандартного нормального распределения. По таблице Z-значений сначала ищем долю 0.95 (обвела в рамочку на рисунке ниже). А потом смотрим, на пересечении каких составных частей Z лежит эта доля. Т.е. мы делаем обратное действие тому, которое изучили на прошлом уроке. Приблизительное Z табличное 1.645

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756

Получается, что расчетное значение 1.2 попадает в область принятия гипотезы H_0 . Оно меньше 1.645. Т.е. делаем вывод, что верна нулевая гипотеза о том, что размер деталей на самом деле и есть 10 мм на уровне значимости 5%.

Ошибки при тестировании гипотез

Так зачем мы упоминаем в выводе уровень статистической значимости? При тестировании гипотез нельзя полностью избежать ошибок. Обратите внимание на графике с распределениями, где α закрашена красным цветом (площадь 0.05), эта площадь является частью распределения H_0 , но попадает в область принятия гипотезы H_1 . Это 5% значений, которые будут лежать слишком далеко от среднего арифметического. И если мы попадем расчетным критерием в эту область, то мы ошибочно примем альтернативную гипотезу H_1 , когда на самом деле была верна нулевая гипотеза H_0 . Поскольку мы заранее устанавливаем уровень статистической значимости α , то иными словами можно сказать, что мы в 5% случаев разрешаем ошибиться, а именно принять альтернативную гипотезу, когда на самом деле верна нулевая гипотеза. Это и есть ошибка I рода, а α – это вероятность ошибки первого рода. Есть 5% вероятность, что мы найдем отличия, когда их на самом деле нет.



Так же на графике мы видим β – это вероятность ошибки II рода. По графику понятно, что если, расчетное значение критерия попадет в этот регион, то будет ошибочно принята нулевая гипотеза, когда на самом деле верна альтернативная. Т.е. не найдем различия там, где они на самом деле были.

И еще одна площадь на графике - это $(1 - \beta)$ – это мощность теста.

Мощность теста - это вероятность найти отличия, когда они на самом деле есть. Мощность теста должна быть не менее 80%. Следовательно, вероятность ошибки II рода не должна превышать 20%. Если мы будем уменьшать α , то будет расти β . Посмотрите на график и мысленно сдвиньте красную линию вправо, уменьшая красную площадь, тогда фиолетовая площадь (β) будет расти.

Итак, до начала теста нужно определиться с α и мощностью теста. Исходя из этого, высчитывается нужный объем выборки, который обеспечит нам надежность результата. Мы не будем затрагивать расчеты выборок. Но упомяну, что для расчета выборки также нужно определиться с размером эффекта. Размер эффекта - это то отличие, которое вы хотите искать, ради которого начинаете исследование. Размер эффекта обосновывается с практической или научной точки зрения. Например, детали с размером 10 мм установятся в какой-то прибор, а вот с

размером 11 мм и выше нам не подойдут, потому что мы не сможем их установить в прибор. Тут размер эффекта 1 мм.

Критерий Стьюдента

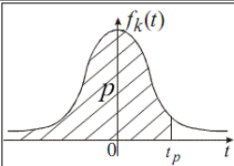
Мы уже в начале обсуждали, что если неизвестна сигма генеральной совокупности и соблюдаются остальные условия применимости параметрических тестов, то следует использовать критерий Стьюдента. Как и любой тест гипотезы, идея та же. Найти расчетное значение критерия и сравнить с табличным. Формула для нахождения расчетного критерия та же.

$$t_n = \frac{\bar{X} - \mu}{\sigma_n / \sqrt{n}}$$

Только сигму мы рассчитываем по выборке, причем используем формулу для несмещенного стандартного отклонения. Для нахождения табличного значения используем таблицу распределения Стьюдента.

Распределение Стьюдента оно зависит от степеней свободы, не буду утомлять Вас определением, а проще говоря, степени свободы здесь будут находиться как разность объема выборки и 1. Т.е. распределение Стьюдента зависит от объема выборки. С увеличением объема выборки оно стремится к нормальному стандартному распределению.

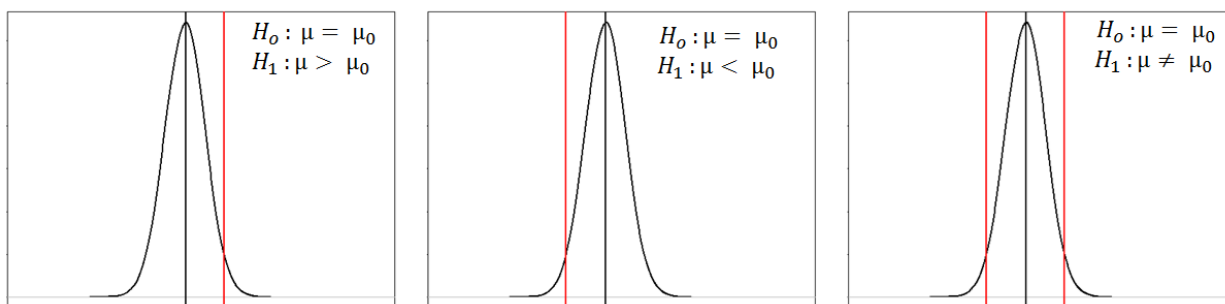
Например, по таблице ниже найдем табличное значение для $\alpha = 5\%$. На графике заштрихована площадь p , а α не заштрихована. Значит, нам надо найти $p = 95\%$. И предположим, что объем выборки был 10. Т.е. 9 степеней свободы. Табличное значение будет равно 1,833.



Число степеней свободы k	Вероятность, p					
	0,9	0,95	0,975	0,99	0,995	0,9995
1	2	3	4	5	6	7
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
12	1,356	1,782	2,179	2,681	3,055	4,318
14	1,345	1,761	2,145	2,625	2,977	4,140

Таблица Стьюдента https://fsd.multiurok.ru/viewImage.php?image=http://nice-diplom.ru/templates/blue/images/img/form_69.gif

Чтобы сделать вывод, мы схематично рисуем графики, как для тестирования гипотезы с использованием Z критерия (помним, что распределение Стьюдента с увеличением выборки стремится к стандартному нормальному распределению).



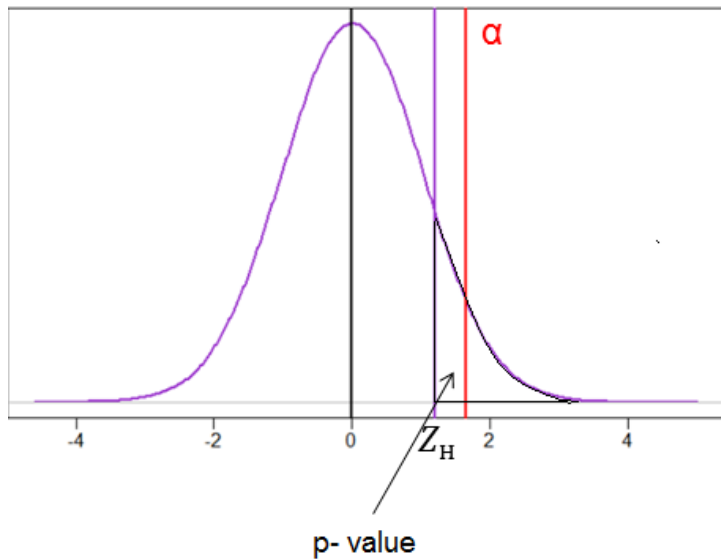
Т.е. такие схематичные рисунки универсальны для критерия Z и t, только пользуемся разными таблицами для нахождения табличного критерия.

P-value

Когда мы проводим тестирование гипотезы мы обращаемся к такому термину как p-value.

Ему дают разные определения, но наиболее корректное – **это вероятность увидеть такое же или более экстремальное значение, чем наблюдаемое, при условии, что нулевая гипотеза верна.**

Вспомним задачу с деталями и смотрим на график ниже. Мы получили расчетное значение $Z_{\text{н}} = 1.2$. На графике p-value - это та площадь, которая лежит за наблюдаемым (расчетным значением). И эта площадь больше, чем площадь за отсечкой α , и при этом Z расчетное попадает в область принятия нулевой гипотезы H_0 . Т.е. если pvalue будет больше, чем заранее установленный уровень статистической значимости, то это означает, что верна нулевая гипотеза.



А теперь представим, что $Z_{\text{н}}$ расчетное получилось не 1.2, а, например, 2, т.е. больше, чем табличное значение 1.645. Тогда $Z_{\text{н}}$ упадет в область принятия альтернативной гипотезы H_1 .

И тогда p-value (та площадь, что находится за $Z_{\text{н}}$) будет меньше площади α и это означает, что верна альтернативная гипотеза. Т.е. когда p-value меньше заранее установленного уровня статистической значимости α , то верна альтернативная гипотеза.

Давайте еще рассмотрим одну задачу в Python

У нас дана некоторая выборка x (те же детали), только объем будет 12 и сигма генеральной совокупности нам не известна. И производитель заверяет, что он изготавливает детали размером 10мм.

Мы помним, что расчетное значение критерия Стьюдента находится по формуле:

$$t_n = \frac{\bar{X} - \mu}{\sigma_n / \sqrt{n}}$$

Произведем эти действия в Python и мы видим, что $t \approx 0.814$, а табличное значение критерия Стьюдента около 1.8 (для 11 степеней свободы и $p = 0.95$)

```
import numpy as np
from scipy import stats
x=np.array([10.50,9.94,10.42,10.47,10.4,9.93,9.17, 9.26,10.11, 10.15, 10.5, 10.47])

np.mean(x)
10.110000000000001

np.std(x,ddof=1)
0.4683238972412927

len(x)
12

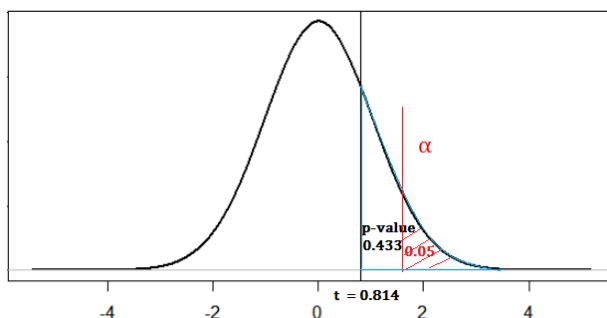
t= (10.11-10)/(0.468/np.sqrt(12))
t
0.8142119180879295
```

А теперь давайте воспользуемся готовой функцией.

```
stats.ttest_1samp(x,10)
Ttest_1sampResult(statistic=0.8136488014166606, pvalue=0.43310585815519953)
```

И мы получили statistic (расчетное значение критерия) около 0.814. Такое же значение мы получили при расчете по формуле. А $pvalue \approx 0.433$

Давайте изобразим результаты на графике



Голубым изображена площадь, соответствующая p-value (0.433), а красным – площадь отсечки α (0.05). Эта площадь соответствует табличному значению критерия Стьюдента 1.8.

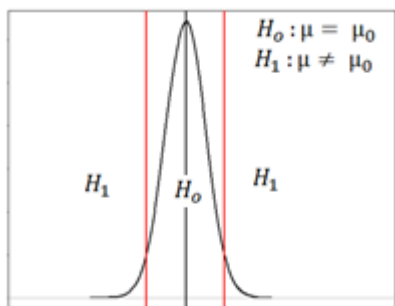
И по графику видим, что у нас верна нулевая гипотеза, $p\text{-value} > \alpha$. Т.о. чтобы Вам интерпретировать результат функции необходимо просто посмотреть на pvalue и сравнить его с заранее установленным уровнем статистической значимости α .

Запомните! Если $p\text{value} < \alpha$, то верна альтернативная гипотеза H_1 .

Виды статистических гипотез

Давайте теперь рассмотрим, какие бывают виды статистических гипотез. Мы уже обсудили ранее, что тест может быть односторонним и двусторонним. Примеры с односторонним тестом мы рассмотрели.

Если же у нас тест двусторонний, т.е. альтернативная гипотеза сформулирована, как показано на графике, то уровень статистической значимости α делится пополам. И у нас появляется две области принятия альтернативной гипотезы. Этот схематичный рисунок подойдет и для Z теста и для t теста. Если, например, альфа 5%, используем Z тест, то табличные значения будут ± 1.96 , соответствуют красным осечкам ($\alpha/2$).



Тест может быть одновыборочными и двухвыборочными.

Примеры, которые мы рассмотрели, это и были примеры одновыборочных тестов. Когда мы с такими тестами можем столкнуться? Например, мы хотим сделать большую закупку какой-то продукции и нам нужно проверить соответствие ее определенному размеру. Как в задаче поставщик заявлял размер деталей 10 мм, а мы из его генеральной совокупности взяли выборку и получили среднее по выборке 10.3. И мы сравниваем среднее по выборке со средним генеральной совокупности.

Двухвыборочные тесты – это тесты, где данные представляют разные совокупности. Например. Мы хотим сравнить вес людей на диете и вес людей на обычном питании. Сравниваем средние арифметические двух выборок. Или сравниваем среднее давление в выборке до приема лекарства со средним давлением после приема лекарства. До приема лекарства – это были люди, не принимающие лекарство, а после приема эти люди представляют другую генеральную совокупность, ту, где люди принимают это лекарство.

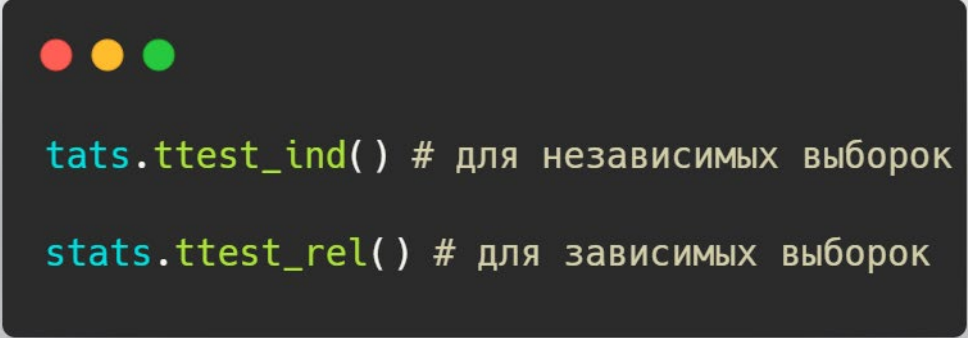
Только что я привела 2 примера с весом и с давлением. Для выполнения теста в Python вам понадобятся разные функции, потому что первый случай с весом- это независимые выборки, а второй пример с давлением- это зависимые выборки.

Т.е. двухвыборочные тесты в свою очередь бывают с зависимыми и независимыми выборками.

С независимыми выборками, думаю, все понятно, это , когда между выборками нет никакой связи: есть представители одной популяции, что придерживается диеты, и есть представители другой популяции, которые не следуют никаким диетам.

А вот приведу примеры зависимых выборок. Это могут быть одни и те же люди, но сначала измерения сняты до приема лекарства, а потом - после. Если мы изменим эксперимент с диетой и вычислим средний вес до диеты и после , то выборки станут зависимыми. Или другой пример, сравниваем рост матерей и рост их взрослых дочерей. Это тоже зависимые выборки. Или измеряем какой-то параметр у близнецов. Тогда это тоже зависимые выборки. Или пример: зрение правого и левого глаза. Это все примеры зависимых выборок.

Ниже приведены функции для использования на языке Python.



```
stats.ttest_ind() # для независимых выборок  
stats.ttest_rel() # для зависимых выборок
```

Проверка на нормальность

Ну и заключительное. Мы сказали, что параметрические тесты надо использовать, когда у нас есть уверенность, что выборка взята из нормально распределенной совокупности ну или хотя бы приблизительно соблюдаются условия нормальности.

Мы можем проверить нашу выборку на нормальность с помощью разных тестов на нормальность. Например, тест Шапиро- Уилка. Где нулевая гипотеза о том, что нет отличий от нормального распределения и альтернативная гипотеза, что распределение отлично от нормального. Помним, в альтернативную гипотезу мы вкладываем наличие различий!

Но проблема в том, что подобные тесты очень чувствительны к объему выборки. И если вы используете большие объемы выборок (сотни, тысячи единиц в выборке), то такой тест может показать вам очень маленькое значение *pvalue*, что будет свидетельствовать об отклонениях от нормальности. Т.е. с увеличением выборок тест будет обнаруживать даже малейшие отклонения от нормальности. Поэтому, если Вы столкнулись с большими объемами выборок, лучше использовать графический метод оценки на нормальность QQ- плот (квантиль-квантиль график).

Сгенерируем выборку *S* объемом 50 из нормально распределенной генеральной совокупности со средним арифметическим 0 и стандартным отклонением 1.

```
import pylab
import scipy.stats as stats
s = np.random.normal(0, 1, 50)
s
array([ 0.46511423,  1.331116 , -0.36073098,  0.85568697, -1.39618853,
        1.61312159,  0.662726 , -0.08169337, -0.5891192 ,  0.48822306,
        0.12044938,  0.37597382, -0.9044236 ,  0.61183822,  1.70055474,
        0.97044029,  0.39987742,  0.18285462, -0.67213515, -1.31172427,
        0.40915827, -1.1307501 ,  1.06527913,  0.86669938,  0.48236301,
        1.27709691, -0.75910924,  0.07983363, -0.2478608 , -0.44095061,
        0.24926677, -0.0937671 ,  0.83121422, -0.38080475,  0.71657152,
        0.63920215, -0.04543604, -0.14885646,  0.51548896, -0.26782449,
       -0.38287857, -0.02960498,  2.24098621, -0.15812507,  0.14214584,
        0.03127712,  1.03201889, -0.62408648,  0.03897425, -0.9698134 ])
```

```
stats.shapiro(s)
(0.9925397038459778, 0.9872482419013977)
```

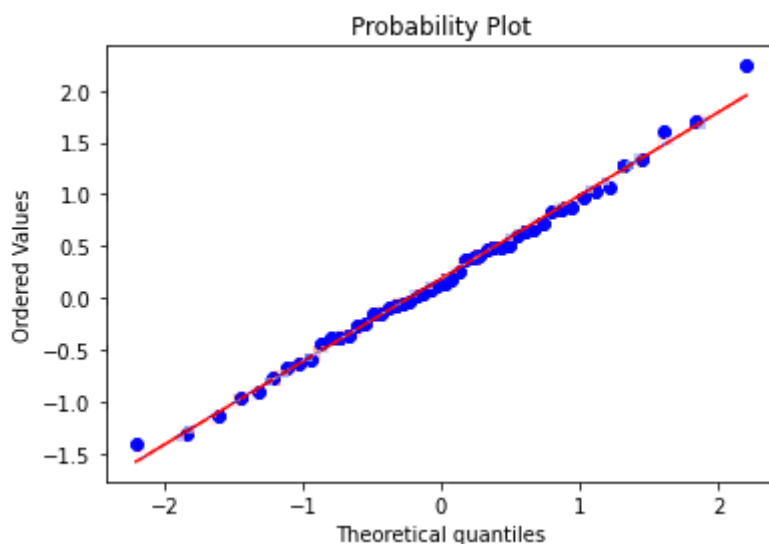
Применим Шапиро-тест

Статистик получился 0.9925 , а *pvalue* \approx 0.987, что свидетельствует в пользу нулевой гипотезы, что нет отличий от нормального распределения.

А теперь рассмотрим QQ- график

```
import pylab
import scipy.stats as stats

stats.probplot(s, dist= "norm", plot= pylab)
pylab.show()
```



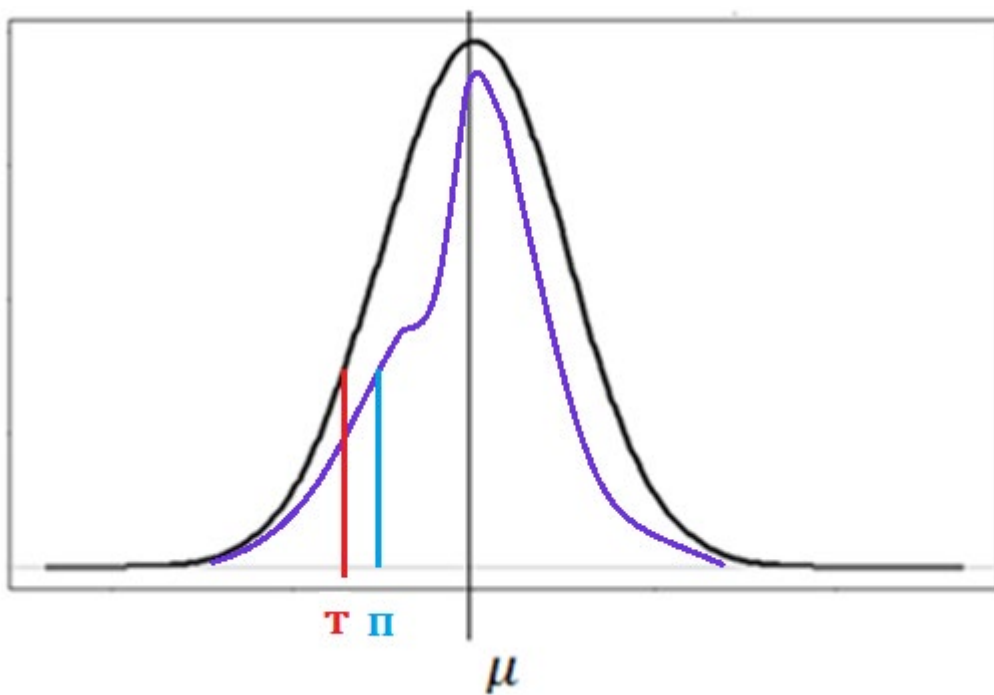
Красная линия на этом графике символизирует нормальное распределение, а синие точки - это 50 значений выборки S . Чем ближе к красной прямой лежат точки, тем лучше аппроксимация нормальному распределению. Здесь практически все точки лежат на красной прямой, мы смело говорим, что выборка следует нормальному распределению.

Идея графика заключается в следующем:

по оси x теоретические квантили, а по оси y упорядоченные значения. Теоретические квантили – это квантили нормального стандартного распределения. Т.е. это z значения. Помним таблицу z -значений? Мы искали, например, какая доля лежит ниже 1.96 и получали 97.5% значений. Вот 1.96 – это и есть квантиль.

А теперь нашу выборку S упорядочим по возрастанию. Мы можем узнать, какому перцентилю соответствует каждое значение этой выборки. Вспоминаем описательную статистику на 3й лекции. Мы искали 1 квартиль, помните? Т.е. мы искали, а ниже какого значения в выборке, лежат 25% значений. А у нас в выборке 50 значений. И мы можем находить, какому перцентилю соответствует каждое из этих значений. Но эти же перцентили мы можем взять и для нормального стандартного распределения и найти для них квартили.

Посмотрите, пожалуйста, на рисунок ниже. Черным схематично обозначено нормальное стандартное распределение, а фиолетовым **схематично** обозначено распределение 50 значений выборки S . Отсечем 30% выборки S и получим «практическое» значение (π), а затем отсечем точно такую же долю у нормального распределения и получим «теоретический» квантиль (τ). Так вот это значение, которое соответствует 30-му перцентилю в выборке как бы имеет две координаты на QQ-графике (τ, π) и соответственно откладывается по оси x и y (теоретический квантиль и упорядоченное значение в выборке). Но если бы наше распределение точно бы повторяло форму нормального распределения, то мы бы видели только одну черную колоколообразную кривую и квантили теоретические совпадали бы с практическими. Вот именно так и строится красная линия на QQ - графике. Поэтому, чем ближе к красной линии, тем надежнее наше предположение о нормальности распределения.



Сегодня мы изучили тестирование гипотез, узнали, какие бывают виды тестирования гипотез, познакомились с параметрическими тестами, научились делать проверку на нормальность. На следующем уроке будем знакомиться с доверительными интервалами, в основе которых также лежит центральная предельная теорема.