Статистическая проверка гипотез

Статистическая гипотеза - это некоторое предположение о свойствах и характеристиках исследуемых генеральных совокупностей.

Проверка статистических гипотез – это поэтапная процедура, которая на основании данных выборки и при помощи теории вероятностей позволяет сделать вывод об обоснованности гипотезы.

Другими словами, это способ проверить, действительны ли результаты, полученные на выборке, и для генеральной совокупности.

Основные этапы:

1.Формулировка основной и альтернативной гипотезы

2.Выбор уровня значимости

3.Определение подходящего статистического критерия

4.Формулировка правила принятия решения

5.Принятие решения, формирование статистического вывода

Статистические гипотезы, так же как и научные гипотезы проверяются согласно принципу фальсифицируемости.

Фальсифицируемость — критерий научности эмпирической или иной теории, претендующей на научность (К. Поппер, 1935). Теория удовлетворяет критерию Поппера в том случае, если существует возможность её экспериментального или иного опровержения.

Принцип фальсифицируемости противоположен принципу верифицируемости: при верификации гипотезы исследователь ищет подтверждающие её примеры, при фальсифицируемости — примеры, опровергающие её.

Таким образом, гипотезы обычно формулируются следующим образом:

Нулевая гипотеза - отвергает эффект (например, разница средних равняется нулю) в популяции.

Альтернативная гипотеза - принимается, если нулевая гипотеза неверна.

Вопрос:

Нулевая гипотеза сформулирована так: средние значения параметра в исследуемой и контрольной группе равны. Альтернативная гипотеза: средние значения параметра в исследуемой группе больше чем в контрольной.

Верна ли такая формулировка?

Основные задачи:

- Изучение принадлежности «выделяющихся» единиц исследуемой выборочной совокупности генеральной совокупности;
- Изучение распределения изучаемых признаков;
- Изучение величины эффекта;
- Изучение наличия и тесноты связи между изучаемыми признаками;
- Изучение формы корреляционной связи.

Ненаправленная гипотеза – доказываем то, что выборки достоверно различаются, но не доказываем чем именно.

Направленная гипотеза – под влиянием исследуемого фактора в определенном направлении (больше или, наоборот, меньше) изменяется исследуемый признак в экспериментальной выборке

Ошибка первого рода - нулевую гипотезу отвергают, когда она истинна, и делают вывод, что имеется эффект, когда в действительности его нет.

Ошибка второго рода - не отвергают нулевую гипотезу, когда она ложна, и делают вывод, что нет эффекта, тогда как в действительности он существует.

		Верная гипотеза	
		H_0	**H_a**
Результат применения критерия	**H_0**	H0 верно принята	H0 неверно принята (Ошибка второго рода)
	H_a	H0 неверно отвергнута (Ошибка первого рода)	H0 верно отвергнута

Ошибка первого рода - обычно называетс α - ошибкой.

Ошибка второго рода - обычно называетс β - ошибкой.

Уровень значимости - это максимально допускаемая исследователем вероятность ошибочного отклонения нулевой гипотезы (α) - меньше - лучше.

Мощность - величина $= 1 - \beta$, иногда указывается в %.

Факторы влияющие на мощность:

- Объем выборки: мощность критерия увеличивается по мере увеличения объема выборки.

- Вариабельность наблюдений: мощность увеличивается по мере того, как вариабельность наблюдений уменьшается.

- Интересующий исследователя эффект: мощность критерия больше для более высоких эффектов. Критерий проверки гипотез имеет больше шансов обнаружить значительный реальный эффект, чем незначительный.

- Уровень значимости: мощность будет выше, если уровень значимости больше.

Статистический критерий - это решающее правило, обеспечивающее надежное поведение, то есть принятие истинной и отклонение ложной гипотезы с высокой вероятностью.

Критическая область Для принятия решения об отклонении или не отклонении нулевой гипотезы необходимо также определить критическую область проверки гипотезы.

- Двухсторонняя

- Левосторонняя (односторонняя)

- Правосторонняя (односторонняя)

Для поверки гипотез используется функция, называемая статистикой критерия, которая зависит от выборки

Статистикой критерия называется случайная величина, значение которой вычисляется по выборке.

С этого момента, под статистикой будем подразумевать некоторую функцию от выборки. Для каждой задачи выбирается уровень значимости и статистика критерия, по значению которой будем делать вывод о справедливости гипотезы. При справедливости основной гипотезы будет известно, с какой вероятностью какое значение принимает статистика критерия. Если эта вероятность очень маленькая, то гипотезу придётся отвергнуть.

Пример одновыборочный t-критерий:

Статистика критерия:

$$\left| \frac{\mu - \mu_0}{\sigma/\sqrt{N}} \right|$$

Критические значения:

$t_{(\alpha,\nu)}$ - критическое значение t-критерия Стьюдента.

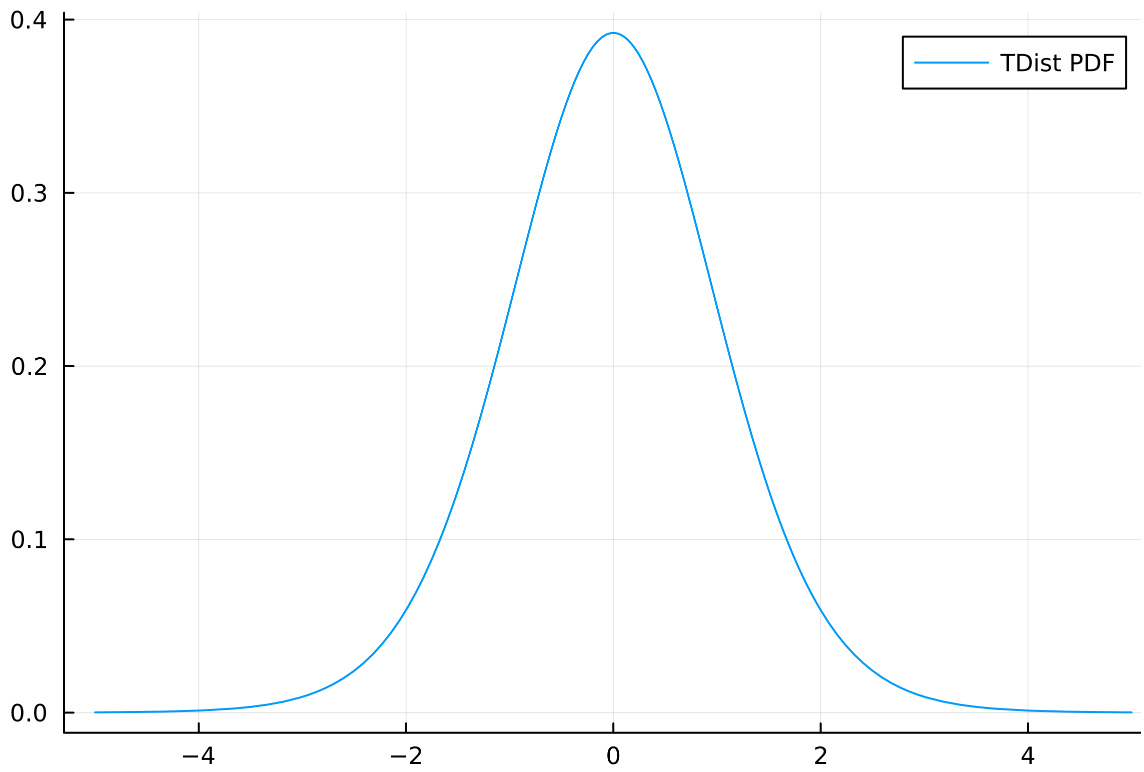
Где α - уровень значимости, ν - количество степеней свободы (в данном случае $n-1$)

In [16]:
```
using Distributions, Plots

tdist = TDist(15)

plot(x-> pdf(tdist, x), label = "TDist PDF")
```

Out[16]:

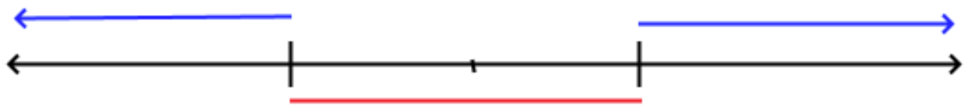


Область принятия гипотезы (ОПГ) - подмножество таких значений критерия, при которых основная гипотеза не может быть отвергнута. Область принятия гипотезы всегда включает в себя значение 0.

Критическая область - подмножество таких значений критерия, при которых основная гипотеза не может быть принята.

В случае, если используется односторонний критерий, ОПГ включает в себя подмножество положительных значений критерия. В таком случае у критерия есть только одна критическая область.

В случае, если используется двусторонний критерий, который может принимать как положительные, так и отрицательные значения, у него имеются две критические области: подмножество отрицательных и подмножество положительных значений критерия, при которых гипотеза не может быть принята.



$$\left| \frac{\mu - \mu_0}{\sigma \sqrt{N}} \right| \leq t_{(\alpha,\nu)}$$

Алгоритм выбора критерия:

- Определить тип данных (количественные или качественные)

- Определить тип распределения (нормальное или отличное от нормального)

- Определить количество групп сравнения

- Определить, связаны ли группы сравнения между собой

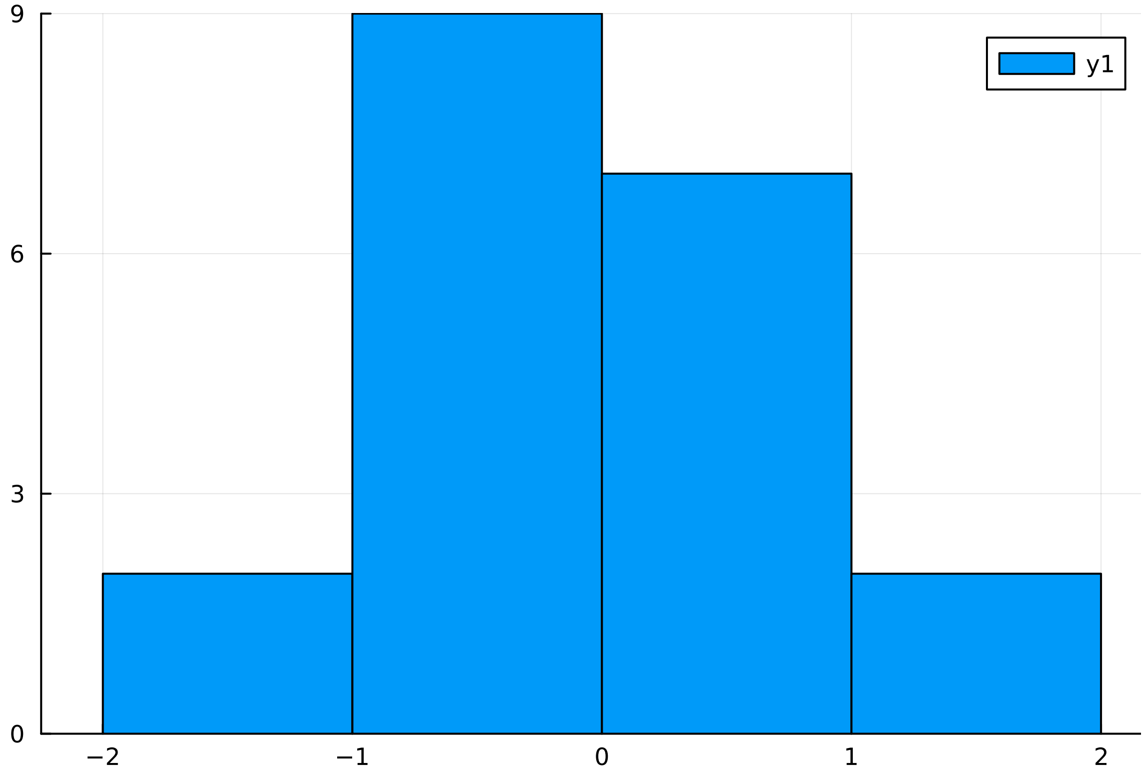
In [32]:
```
using Distributions, Plots, Random, HypothesisTests
r = rand(Normal(), 20)

println("KSTest: ", pvalue(ExactOneSampleKSTest(r, Normal(mean(r), sqrt(var(r)))
println("ADTest: ", pvalue(OneSampleADTest(r, Normal(mean(r), sqrt(var(r))))))

histogram(r)
```

```
KSTest: 0.5110250238893161
ADTest: 0.7504313814311815
```

Out[32]:



Группы критериев

По типу оценваемых параметров

- Равенство средних (T-критерий)
- Равенстов долей (χ^2 - критерий, Z-критерий)
- Равенство дисперсий (F-тест)

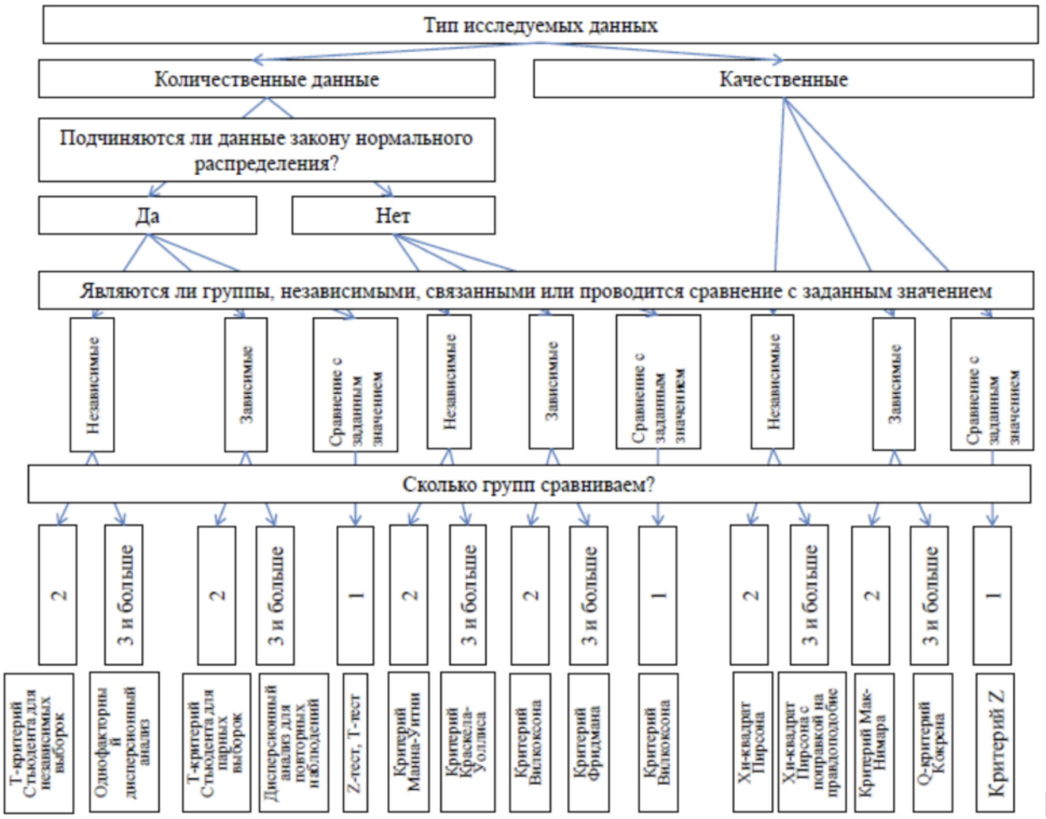
- Равенство коэффициентов корреляции (Т-критерий)
- Критерии для проверки нормальности (Шапиро-Уилка, Колмогорова-Смирнова)

Непараметрические критерии - Группа статистических критериев, которые не включают в расчёт параметры вероятностного распределения и основаны на оперировании частотами или рангами.

- Q-критерий Розенбаума (оценка различий между двумя выборками по уровню какого-либо признака, порядковая шкала)
- U-критерий Манна-Уитни (оценка различий между двумя выборками по уровню какого-либо признака)
- Критерий Уилкоксона (W-критерий Уилкоксона, связанные выборки)
- Критерий Пирсона (применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому)
- Критерий Колмогорова-Смирнова (предназначен для проверки простых гипотез о принадлежности анализируемой выборки некоторому известному закону распределения)

Параметрические критерии - Группа статистических критериев, которые включают в расчет параметры вероятностного распределения признака (средние и дисперсии).

- t-критерий Стьюдента
- Критерий Фишера
- Критерий отношения правдоподобия
- Критерий Романовского (применяется для оценки на грубую погрешность одного сомнительного значения выборки из нормально распределённой случайной величины)
- Критерий Ливиня (тест на равенство двух дисперсий)



Для проверки одной и той же гипотезы обычно существует несколько различных статистических критериев. Существуют следующие принципы подбора статистических критериев.

1. Должны выполняться предположения, обуславливающие

возможность применения рассматриваемого критерия (например, о виде распределения случайной величины и о наблюдаемых данных). 2. Критерий должен быть **состоятельным**, т.е. его мощность должна стремиться к единице с ростом объема выборки. 3. Критерий должен быть **несмещенным**, т.е. мощность должна быть больше, чем вероятность ошибки первого рода. 4. Критерий должен обладать наибольшей мощностью при заданном объеме выборки и заданном уровне значимости критерия.

Сделать статистический и содержательный вывод

Статистический вывод - обобщение информации из выборки для получения представления о свойствах генеральной совокупности.

Содержательный вывод - заключение о том, подтверждена или нет исходная научная гипотеза.

Важно всегда указывать уровень значимости , на котором проверяется гипотеза, так как без этого уточнения выводы о нулевой гипотезе не имеют большого смысла: на одном уровне значимости гипотеза может быть отвергнута, а при выборе другого уровня значимости – нет.

В случае отклонения нулевой гипотезы остается вероятность того, что H_0 верна и эта вероятность равна уровню значимости (p-value). Следовательно, нельзя утверждать, что результаты доказывают справедливость альтернативной (содержательной) гипотезы. Корректным будет более осторожный вывод о том, что получено свидетельство в пользу альтернативной гипотезы.

По результатам проверки статистической гипотезы **не делается** вывод о том, что нулевая гипотеза верна. Всё, что можно решить по итогам проверки: отвергнуть нулевую гипотезу или нет, т.е. никакого содержательного вывода сделать нельзя. Поэтому выражение «Отрицательный результат исследования – тоже результат» не следует понимать буквально– это скорее отсутствие какого бы то ни было результата.

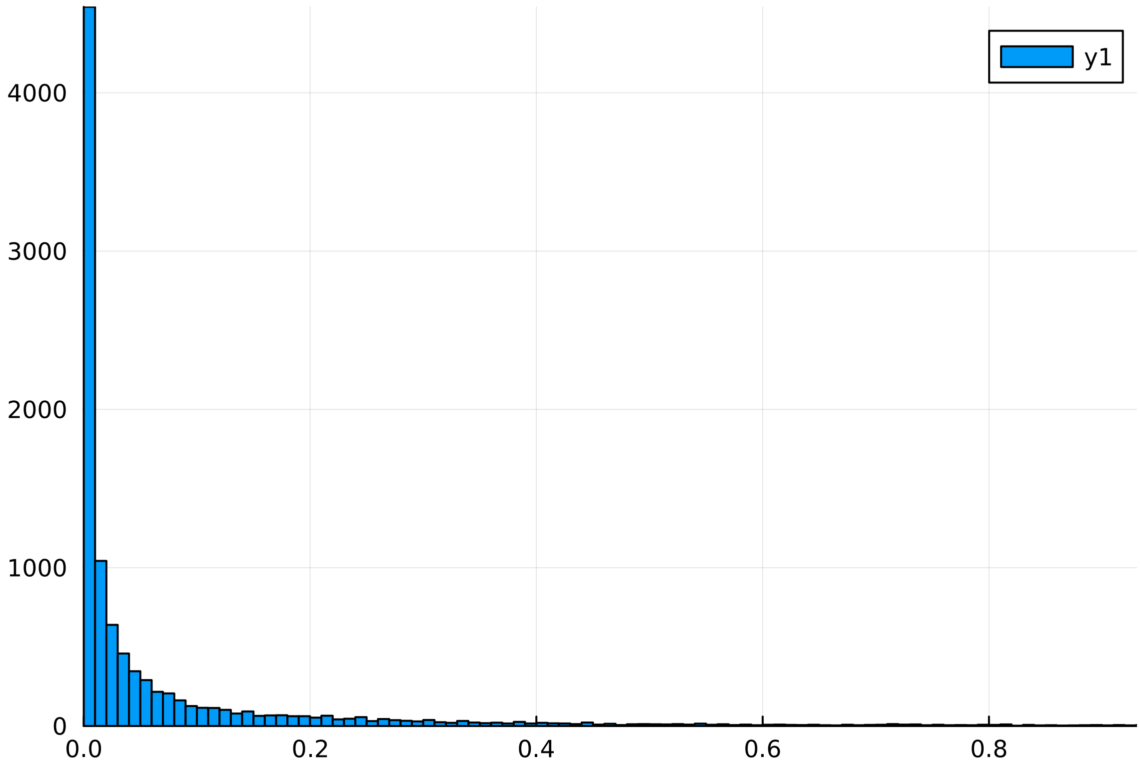
Пример статистического вывода: на имеющихся данных, на уровне значимости 5% есть основание отвергнуть нулевую гипотезу в пользу альтернативной.

Пример содержательного вывода: получено свидетельство того, что среднее значение АСТ в Группе 1 отличается от значения АСТ в Группе 2.

Проблемы p-value

In [34]:
```
using HypothesisTests
pvals = []
d1 = Normal(1, 2)
d2 = Normal(1.7, 2)
for i = 1:10000
    v1 = rand(d1, 100)
    v2 = rand(d2, 100)
    push!(pvals, pvalue(OneWayANOVATest(v1, v2)))
end
histogram(pvals, xlims = (0, 0.95))
```

Out[34]:



Некоторые гипотезы применяемые в клинических исследованиях

- Гипотеза равенства (двухсторонняя)

$H_0 : \mu_a = \mu_b$

$H_A : \mu_a \neq \mu_b$

- Гипотеза не меньшей эффективнсти (большей эффективности) (одностороняя)

$H_0 : \mu_a - \mu_b \leq \Delta$

$H_A : \mu_a - \mu_b > \Delta$

- Гипотеза эквивалентности (две односторонних)

$H_0 : |\mu_a - \mu_b| \geq \Delta$

$H_A : |\mu_a - \mu_b| < \Delta$

или

$H_0 : \mu_a - \mu_b \geq \Delta_U \cup \mu_a - \mu_b \leq \Delta_L$

$H_A : \mu_a - \mu_b < \Delta_U \cap \mu_a - \mu_b > \Delta_L$

Интервальные гипотезы

В качестве статистики используется граница(ы) доверительного интервала. Решение принимается на основании значения верзней и/или нижней границы доверительного интервала.

Доверительный интервал - интервал, который покрывает неизвестный параметр с заданной надёжностью - в случае если повторять исследование многократно, то с указанной вероятностью интервал (каждый раз новый) будет включать истинное значение параметра.

Распространенная интервальная гипоза - гипотеза биоэквивалентности.

Так гипотеза биоэквивалентнсти может быть сформулирована следующим образом:

$H_0 : \mu_a - \mu_b \geq ln(1.25) \cup \mu_a - \mu_b \leq ln(0.8)$

$H_A : \mu_a - \mu_b < ln(1.25) \cap \mu_a - \mu_b > ln(0.8)$

Фактически проводится тест двух односторонних гипотез на уровне значимости α.

$H_01 : \mu_a - \mu_b \geq ln(1.25)$

$H_02 : \mu_a - \mu_b \leq ln(0.8)$

Какой ширины доверительный интервал должен быть потроен для проверки этих гипотез?

- Фадеева Л. Н., Лебедев А. В., Теория вероятностей и

математическая статистика: учебное пособие. - 2-е изд., перераб. и доп. - М.: Эксмо, 2010. - 496 с.

- Математика. Базовый курс / Б.Ш. Гулиян, Р.Я. Хамидуллин. Москва : Синергия, 2013. - 712 с. - ISBN 978-5-4257-0109-1.

- Математическая статистика в медицине, В. А. Медик, М. С. Токмачев, 978-5-279-03195-5

- *Медико-биологическая статистика. Гланц. Пер. с англ. — М., Практика, 1998. — 459 с.*

Еще:

- *Байесовская статистика: Star Wars, LEGO, резиновые уточки и многое другое*

- *Занимательная статистика. Манга. Син Такахаси, 2009, 224с, ISBN:*

978-5-97060-179-2

```
In [ ]:
```