



# Программирование в среде R

Шевцов Василий Викторович,  
директор ДИТ РУДН, [shevtsov\\_vv@rudn.university](mailto:shevtsov_vv@rudn.university)

# Проверка статистических гипотез

# Базовые термины и идеи

- Генеральная совокупность (population) (иногда используется калька с англоязычного термина – «популяция») – все множество объектов, в отношении которых формулируется исследовательская гипотеза
- Выборка (sample) – ограниченная по численности группа объектов (респондентов), отбираемая из генеральной совокупности для изучения ее свойств
- Сплошное и выборочное исследование
- Репрезентативность выборки (representativeness of sample) – способность выборки представлять изучаемые явления достаточно полно с точки зрения их изменчивости в генеральной совокупности
- Любое исследование направлено на определение некоторой характеристики или выявление связи между признаками
- Связь может характеризоваться не только величиной (степенью связи) и направлением, но также и надежностью или статистической достоверности (statistical confidence)
- Эта характеристика связи показывает, можно ли распространить результаты, полученные на данной выборке, на всю генеральную совокупность, из которой взята эта выборка

# Базовые термины и идеи

- Статистическая гипотеза – утверждение относительно неизвестного параметра генеральной совокупности на основе выборочного исследования
- Любое заключение, полученное из статистического наблюдения / исследования / анализа, – индуктивно и строится на конечном числе наблюдений, поэтому оно не полно и может быть не достоверно
- Необходимо обоснование заключения, т.е. тестирование результатов, на которых строится гипотеза, на статистическую достоверность
- Надежность (достоверность) непосредственно связана с репрезентативностью выборки, т.е. с тем, насколько уверенно данные, полученные по выборке, позволяют судить о соответствующих параметрах генеральной совокупности
- Надежность определяется тем, насколько вероятно, что обнаруженная в выборке связь подтвердится (будет вновь обнаружена) на другой выборке той же генеральной совокупности
- Какова вероятность случайного получения результата, подтверждающего наличие связи, которой нет в генеральной совокупности?

# Статистические гипотезы

- Нулевая гипотеза (null hypothesis) – гипотеза об отсутствии различий (утверждение об отсутствии различий в значениях или об отсутствии связи в генеральной совокупности)
- Согласно нулевой гипотезе ( $H_0$ ), различие между значениями недостаточно значительно, а независимая переменная не оказывает никакого влияния  
Альтернативная гипотеза (alternative hypothesis) – гипотеза о значимости различий (утверждает наличие различий или существование связи)
- Альтернативная гипотеза ( $H_A$ ) является «рабочей» гипотезой исследования. В соответствии с этой гипотезой, различия достаточно значимы и обусловлены влиянием независимой переменной
- Ненаправленная и направленная альтернативы
  - $H_0: \mu=50$
  - $H_A: \mu \neq 50$
  - $H_A: \mu > 50$
  - $H_A: \mu < 50$
- Нулевая и альтернативная гипотезы представляют полную группу несовместных событий: отклонение одной влечет принятие другой
- Основной принцип метода проверки гипотез состоит в том, что выдвигается нулевая гипотеза  $H_0$ , с тем чтобы попытаться опровергнуть ее и тем самым подтвердить альтернативную гипотезу  $H_A$ . Если результаты статистического теста, используемого для анализа разницы между средними, окажутся таковы, что позволят отклонить  $H_0$ , это будет означать, что верна  $H_A$ , т.е. выдвинутая рабочая гипотеза подтверждается
- Не можем отклонить нулевую гипотезу - не значит «принять» альтернативную (нулевая гипотеза никогда не может быть абсолютно подтверждена!)

# Статистические ошибки при принятии решений

## Ошибки первого и второго рода

- Статистическая ошибка первого рода (Type I Error) – ошибка обнаружить различия или связи, которые на самом деле не существуют  
«Истинная нулевая гипотеза отклоняется»
- Статистическая ошибка второго рода (Type II Error) - не обнаружить различия или связи, которые на самом деле существуют  
«Ложная нулевая гипотеза не может быть отклонена»
- Более «критичной» ошибкой считается статистическая ошибка первого рода
- «Судебная» аналогия: Вердикт «Не виновен» или «Виновен»  
Ошибка первого рода - невинный обвинен  
Ошибка второго рода - виновный освобожден

# Уровни статистической значимости

- Тот или иной вывод с некоторой вероятностью может оказаться ошибочным, и обычно вероятность ошибки тем меньше, чем больше выборка. Таким образом, чем больше получено результатов, тем в большей степени по различиям между двумя выборками можно судить о том, что действительно имеет место в той генеральной совокупности, из которой взяты эти выборки
- Однако обычно используемые выборки относительно невелики, и в этих случаях вероятность ошибки может быть значительной
- *Уровень значимости* (level of significance) (уровень достоверности, уровень надежности, доверительный уровень, вероятностный порог) - это пороговая (критическая) вероятность ошибки, заключающейся в отклонении (не принятии) нулевой гипотезы, когда она верна. Другими словами, это допустимая (с точки зрения исследователя) вероятность совершения статистической ошибки первого рода – ошибки того, что различия сочтены существенными, а они на самом деле случайны
- Обычно используют уровни значимости (обозначаемые  $\alpha$ ), равные 0,05, 0,01 и 0,001
- Например, уровень значимости, равный 0,05, означает, что допускается не более чем 5%-ая вероятность ошибки. Т.е. нулевую гипотезу можно отвергнуть в пользу альтернативной гипотезы, если по результатам статистического теста вероятность ошибки, т.е. вероятность случайного возникновения обнаруженного различия (*p-уровень*) не превышает 5 из 100, т.е. имеется лишь 5 шансов из 100 ошибиться. Если же этот уровень значимости не достигается (вероятность ошибки выше 5%), считают, что разница вполне может быть случайной и поэтому нельзя отклонить нулевую гипотезу

Таким образом, *p-уровень значимости* (*p-value*) соответствует риску совершения ошибки первого рода (отклонения истинной нулевой гипотезы). Если  $p < \alpha$ , нулевая гипотеза отклоняется



# Уровни статистической значимости: содержательная интерпретация

- Вопрос о приемлемом значении  $\alpha$ , т.е. вопрос о том, при каком уровне можно отклонить  $H_0$ , не имеет однозначного ответа
- Для установленного значения  $\alpha$  вероятность ошибки второго рода уменьшается с ростом объема выборки
- При увеличении значения  $\alpha$  (например, с 0,01 до 0,05) вероятность ошибки второго рода уменьшается
- Значение  $\alpha$  устанавливается исходя из «научных конвенций» - соглашений, принятых в научном сообществе на основе практического опыта в различных областях исследования. Традиционная интерпретация различных уровней значимости исходит из  $\alpha = 0,05$  и приведена в табл. Такое значение  $\alpha$  рекомендовано для небольших выборок (когда высока вероятность ошибки второго рода). Если объемы выборок  $n \geq 100$ , то порог отклонения  $H_0$  целесообразно снизить до  $\alpha=0,01$  и принимать решение о наличии связи (различий) при  $p \leq 0,01$

Уровень значимости	Решение	Возможный статистический вывод
$p > 0,1$	$H_0$ не может быть отклонена	«Статистически достоверные различия не обнаружены»
$p \leq 0,1$	сомнения в истинности $H_0$ , неопределенность	«Различия обнаружены на уровне статистической тенденции»
$p \leq 0,05$	значимость, отклонение $H_0$	«Обнаружены статистически достоверные (значимые) различия»
$p \leq 0,01$	высокая значимость, отклонение $H_0$	«Различия обнаружены на высоком уровне статистической значимости»



# Логика проверки гипотез

- Для принятия решений о том, какую из гипотез (нулевую или альтернативную) следует принять, используют *статистические критерии*, которые включают в себя методы расчета определенного показателя, на основании которого принимается решение об отклонении или принятии гипотезы, а также правила (условия) принятия решения
- Этот показатель называется *эмпирическим значением критерия*
- Это число сравнивается с известным (например, заданным таблично) эталонным числом, называемым *критическим значением* критерия.
- Критические значения приводятся, как правило, для нескольких уровней значимости: 5% (0,05), 1% (0,01) или еще более высоких
- Если полученное исследователем эмпирическое значение критерия оказывается меньше или равно критическому, то нулевая гипотеза не может быть отклонена – считается, что на заданном уровне значимости (то есть при том значении  $\alpha$ , для которого рассчитано критическое значение критерия) характеристики распределений совпадают
- Если эмпирическое значение критерия оказывается строго больше критического, то нулевая гипотеза отвергается и принимается альтернативная гипотеза – характеристики распределений считаются различными с достоверностью различий  $1 - \alpha$ .
- Например, если  $\alpha = 0,05$  и принята альтернативная гипотеза, то достоверность различий равна 0,95 или 95%

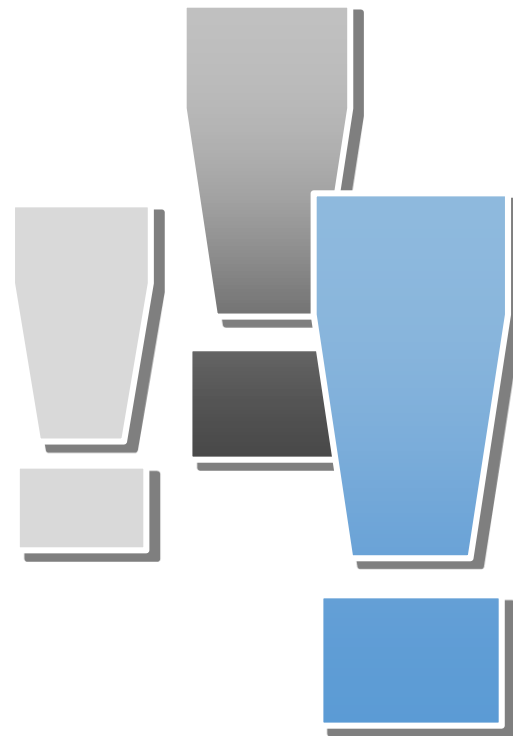
# Логика проверки гипотез

- Если эмпирическое значение критерия для данного *числа степеней свободы* ( $df=n-1$ ) оказывается ниже критического уровня, соответствующего выбранному значению  $\alpha$  (порогу вероятности), то нулевая гипотеза не может считаться опровергнутой, и это означает, что выявленная разница (или связь) недостоверна
- Чем эмпирическое значение меньше критического значения критерия, тем больше степень совпадения характеристик сравниваемых объектов
- Чем эмпирическое значение критерия больше критического значения, тем сильнее различаются характеристики сравниваемых объектов
- Если эмпирическое значение критерия оказывается меньше или равно критическому, то можно сделать вывод, что характеристики экспериментальной и контрольной групп совпадают на уровне значимости  $\alpha$
- Если эмпирическое значение критерия оказывается строго больше критического, то можно сделать вывод, что достоверность различий характеристик экспериментальной и контрольной групп равна  $\alpha$

# Процедура проверки статистической гипотезы

- Сформулировать нулевую и альтернативной гипотезы
- Выбрать соответствующий статистический тест
- Выбрать требуемый уровень значимости ( $\alpha=0.05, 0.01, 0.025, \dots$ )
- Вычислить эмпирическое значение критерия по тесту
- Сравнить с критическим значением критерия по тесту
- Принять решение (для большинства тестов приемлемо правило: если вычисленное значение больше, чем критическое, нулевая гипотеза отклоняется)

# Спасибо за внимание!



Шевцов Василий Викторович

shevtsov\_vv@rudn.university  
+7(903)144-53-57