Проверка гипотез

# Введение

Целью проверки гипотез, как части математической статистики, является построение выводов об исследуемом признаке в популяции, основываясь на выборке.

Зачастую, исследуемый признак – параметр *θ* общего распределения *f(x).* Мы хотим решить "бинарную" задачу: определить принадлежит ли истинный параметр *θ* одному подмножеству всех возможных параметров *θ* ∈ Θ0 ⊂ Θ или другому *θ* ∈ Θ1 ⊂ Θ, где подмножества полностью покрывают множество параметров: Θ = Θ0 ∪ Θ1 Θ0 ∩ Θ1 = ∅ . Две гипотезы формулируются следующим образом:

*H0* : *θ* ∈ Θ0 *H1*: *θ* ∈ Θ1

где мы называем *H0* нулевой гипотезой, а *H1* - альтернативной гипотезой.

**Цель**: использовать только имеющуюся у нас выборку *X = X1,…,Xn*, чтобы выбрать между *H0* и *H1*.

Существуют 4 варианта развития событий в задаче проверки гипотез:

|  |  |  |
| --- | --- | --- |
|  | *H0* верна | *H0* неверна |
| Не отрицать *H0* | Верно | Ошибка 2-го рода |
| Отвергнуть *H0* | Ошибка 1-го рода | Верно |

На этом этапе появляется новое действующее лицо: **уровень значимости**, α. Под этим термином скрывается вероятность появления ошибки 1-го рода. Для уровня достоверности 95% значение α равно 0,05. Это означает, что существует 5% вероятность того, что мы отвергнем истинную *H0*. В задачах проверки гипотез исследователи устанавливают приемлемый для задачи уровень ошибки.

# p-значение

p-значение используется во всей статистике, от t-тестов до регрессионного анализа и древовидных моделей почти во всех моделях машинного обучения. Мы все используем p-значения для определения статистической значимости в проверке гипотезы. Несмотря на то, что p-значение очень важно, оно является скользкой концепцией, которую люди часто неправильно интерпретируют. p-значение оценивает, насколько хорошо выборка совместима с нулевой гипотезой. Насколько вероятен эффект, наблюдаемый в выборке, если нулевая гипотеза верна? Другими словами, при условии, что нулевая гипотеза верна, p-значение представляет собой вероятность получения результата, **равного** или **более экстремального**, чем результат в выборке. Высокие p-значения: вероятно, такая выборка могла быть получена с *H0*. Низкие p-значения: маловероятно, что *H0* верна, сомнительно, что параметр из *H0*. сгенерировал такую выборку.

**Пример**: Предположим, мы производим проверку гипотезы при α = 5%. *H0*: μ = 20, *H1*: μ > 20. При этом значение выборочного среднего у нас в выборке ͞x = 25, и p-значение оказалось равным 3%.

**Интерпретация**: Мы согласны, что можем отклонить нашу нулевую гипотезу 5 раз из 100, даже если она верна. Когда p-значение составляет 3%, меньше α (ниже порога совершения ошибки первого рода), это означает, что получение экстремальной выборочной статистики (͞x ≥ 25) при условии, что *H0* верна, очень маловероятно. Другими словами, мы не можем получить нашу выборочную статистику, если *H0* верна. Следовательно, мы отвергаем *H0* и принимаем *H1*.

Предположим, мы получили p-значение 6%, т. е. вероятность получения выборочной статистики более экстремальной выше, если нулевая гипотеза верна. Таким образом, сравниваясь с α, мы не можем рисковать совершить ошибку первого рода. Следовательно, мы не можем отвергнуть *H0*.

Критерий, по которому принимается решение: p -value < α . Если неравенство справедливо, тогда делаем выбор в сторону *H1*, если наоборот, тогда не отрицаем *H0*.