

# Теория вероятностей и математическая статистика

## Начала проверки гипотез

## Вопрос 1

Прежде чем что-то проверять, нужно сформулировать это предположение на статистическом языке. Основное предположение называется нулевой гипотезой. Она должна быть простой. Что это значит?

## Вопрос 1

Прежде чем что-то проверять, нужно сформулировать это предположение на статистическом языке. Основное предположение называется нулевой гипотезой. Она должна быть простой. Что это значит?

## Ответ

Простой гипотезой называется такая гипотеза, для которой можно однозначно задать распределение. Примеры таких гипотез:

- 1 параметр  $E X = 5$  (равенство среднего конкретному значению)
- 2 предполагаю, что монета правильная:  $p(\text{орел}) = 0.5$  (равенство вероятности конкретному значению)

## Вопрос 2

На следующем шаге обязательно формулируем альтернативную гипотезу, против которой проверяется нулевая. Обязательно ли альтернативные гипотезы должны быть простыми?

## Вопрос 2

На следующем шаге обязательно формулируем альтернативную гипотезу, против которой проверяется нулевая. Обязательно ли альтернативные гипотезы должны быть простыми?

## Ответ

Совершенно необязательно. Примеры альтернатив:

- 1 параметр  $EX > 5$
- 2 параметр  $EX < 5$
- 3  $p(\text{орел}) \neq 0.5$

## Напоминания

- 1 Статистическая гипотеза формулируется о ГЕНЕРАЛЬНОМ ПАРАМЕТРЕ, а не о его оценке.

## Напоминания

- 1 Статистическая гипотеза формулируется о ГЕНЕРАЛЬНОМ ПАРАМЕТРЕ, а не о его оценке.
- 2 Чаще всего используются двусторонние альтернативы. Если Вы все же решили воспользоваться односторонней альтернативой, предварительно посмотрите на оценки необходимых параметров. К примеру, если проверяете гипотезу о равенстве средних, сравните средние в двух выборках, чтобы правильно определиться с лево- или правосторонней альтернативой.

## Вопрос 3

После того, как гипотезы сформулированы, нужно выбрать статистику для проверки. Что такое статистика в схеме проверки гипотез и зачем она нужна?



### Вопрос 3

После того, как гипотезы сформулированы, нужно выбрать статистику для проверки. Что такое статистика в схеме проверки гипотез и зачем она нужна?

### Ответ

Статистика – функция от выборки со своим распределением, которая используется для того, чтобы принять решение об отвержении или неотвержении нулевой гипотезой.

Распределение статистики позволяет нам рассчитать, в частности,  $p$ -value (минимальный уровень значимости).

Пример такой статистики: случайная величина «количество орлов / решек, выпавших при эксперименте из 100 бросков монеты» для проверки гипотезы о правильной монете

## Вопрос 4

Далее на основе наблюдаемого значения статистики мы рассчитываем значение минимального уровня значимости. Что такое минимальный уровень значимости (p-value)?

## Вопрос 4

Далее на основе наблюдаемого значения статистики мы рассчитываем значение минимального уровня значимости. Что такое минимальный уровень значимости (p-value)?

## Ответ

Вероятность того, что статистика примет наблюдаемое значение (включительно) или более «экстремальные» значения (свидетельствующие в большей мере в пользу альтернативы, так как для нулевой гипотезы эти значения статистики маловероятны) при условии того, что нулевая гипотеза ВЕРНА. То есть, при расчете p-value мы фокусируемся только на распределении для нулевой гипотезы!

## Вопрос 5

Как сделать статистический вывод на основании p-value?

## Вопрос 5

Как сделать статистический вывод на основании  $p$ -value?

## Ответ

Если  $p$ -value мало, значит наблюдаемое значение статистики ближе к хвостам распределения («экстремальным» значениям), следовательно, на основании имеющихся данных мы отвергаем нулевую гипотезу в пользу альтернативы.

## Вопрос 5

Как сделать статистический вывод на основании  $p$ -value?

## Ответ

Если  $p$ -value мало, значит наблюдаемое значение статистики ближе к хвостам распределения («экстремальным» значениям), следовательно, на основании имеющихся данных мы отвергаем нулевую гипотезу в пользу альтернативы.

И наоборот, если  $p$ -value достаточно велико, значит наблюдаемое значение статистики ближе к центру распределения (характерным значениям), следовательно, на основании имеющихся данных мы не можем отвергать нулевую гипотезу в пользу альтернативы.



Продолжение следует...  
ВАШИ ВОПРОСЫ?