

Модульная домашняя работа №2

Мягкий дедлайн: 23:59 02/05/2020.

Жесткий дедлайн: 23:59 11/05/2020.

Форма, куда отправлять работу: [Google forms](#).

Общие требования

- Для данного домашнего задания можно использовать языки программирования R, Python.
- Задания необходимо выполнять в RMarkdown либо jupyter notebook с комментариями и пояснениями, иначе работа проверяться не будет.
- Если работа будет прислана до мягкого дедлайна, то она будет проверена и вы будете иметь возможность исправить ее. Иначе ваша работа будет проверяться по мере возможности.
- После жесткого дедлайна никакие исправления нельзя будет вносить.
- Все вопросы и предложения по условию задач рассматриваются **до мягкого дедлайна**.
- **Комментируйте то, что вы делаете.**
- Если у вас возникнут вопросы, то можно их задавать в телеграм-чате или лично в день занятий.

Требования к заданиям

- Для всех используемых библиотечных функций обосновать корректность их применения. **Это нулевое задание для всех вариантов.** Комментарии «Это ж из scipy» не рассматриваются.
- Все ответы на вопросы должны быть обоснованы при помощи моделирования. По возможности необходимо предоставлять всяческие графики, а не просто выводить какие-то числа (например, для того чтобы проверить, как ведет себя ошибка первого рода, необходимо нарисовать график для всех уровней значимости α , а не вывести несколько каких-то чисел).
- В данных необходимо найти что-нибудь интересное, используя различные критерии, с которыми вы познакомились на курсе. Необходимо самостоятельно ставить гипотезы и проверять их.

Вариант 1

1. Пусть $x_1, \dots, x_n \sim N(a_1, \sigma_1^2)$ и $y_1, \dots, y_n \sim N(a_2, \sigma_2^2)$ где σ_1^2, σ_2^2 известны. Хотим проверить гипотезу $H_0 : a_1 - a_2 = 0$ против $H_1 : a_1 - a_2 \neq 0$.
 - Примените t-test и критерий Mann-Whitney для проверки данной гипотезы.
 - Почему можно сравнивать эти критерии?
 - Сравните мощности критериев. Какой из них более мощный?
2. Данные [Abalone](#).

Вариант 2

1. Пусть $x_1, \dots, x_n \sim \text{Exp}(\lambda_1)$ и $y_1, \dots, y_n \sim \text{Exp}(\lambda_2)$. Хотим проверить гипотезу $H_0 : \lambda_1 - \lambda_2 = 0$ против $H_1 : \lambda_1 - \lambda_2 \neq 0$.
 - Примените t-test и критерий Mann-Whitney для проверки данной гипотезы.
 - Корректно ли применение этих критериев для данной задачи? Почему?
 - Продемонстрируйте поведение ошибок первого и второго рода в каждом из случаев.
2. Данные Contraceptive Method Choice.

Вариант 3

1. Пусть x_1, \dots, x_n — выборка из абсолютно непрерывного распределения (существует плотность) с функцией распределения F . Хотим проверить гипотезу $H_0 : F = G$ против альтернативы сдвига.
 - Продемонстрируйте (не)состоятельность критерия Колмогорова-Смирнова против альтернативы сдвига.
 - Зависит ли ошибка второго рода от альтернативы сдвига? Продемонстрируйте это.
2. Данные Ecoli.

Вариант 4

1. Пусть x_1, \dots, x_n — выборка из геометрического распределения с параметром p . Воспользуйтесь критерием согласия χ^2 для проверки $H_0 : p = p_0$.
 - Что делать, если некоторые из ожидаемых частот (подсчитываются в ходе применения критерия) маленькие?
 - Есть ли зависимость мощности от p_1 при выборе простой $H_1 : p = p_1$? Продемонстрируйте это с помощью моделирования;
2. Данные Immunotherapy.

Вариант 5

1. Пусть $x_1, \dots, x_n \sim N(a_1, \sigma^2)$ и $y_1, \dots, y_n \sim N(a_2, \sigma^2)$, где σ^2 известна, и $\rho(x_i, y_i) = \rho > 0$ для любого $i = 1, \dots, n$. Хотим проверить гипотезу $H_0 : a_1 - a_2 = 0$ против $H_1 : a_1 - a_2 \neq 0$.
 - Примените t-test для независимых выборок для данной гипотезы.
 - Продемонстрируйте, как себя ведут ошибки первого и второго рода для $\rho > 0$ и $\rho < 0$. Критерий является радикальным/точным/консервативным?
2. Данные Liver Disorders.

Вариант 6

1. Пусть $x_1, \dots, x_n \sim N(a_1, \sigma_1^2)$ и $y_1, \dots, y_n \sim N(a_2, \sigma_2^2)$, где a_1, a_2 неизвестны. Известно, что $\rho(x_i, y_i) = \rho > 0$ для любого $i = 1, \dots, n$.

Хотим проверить гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$ против $H_1 : \sigma_1^2 \neq \sigma_2^2$.

- Примените критерий Фишера для проверки данной гипотезы.
- Корректно ли его применять? Почему?

2. Данные seeds.

Вариант 7

1. Пусть есть выборка из гамма распределения со сдвигом a (другие параметры – любые). Хотим проверить гипотезу $H_0 : a = a_0$.

- Примените для проверки данной гипотезы критерии Колмогорова-Смирнова и ω^2 .
- Какой из них более мощный против $H_1 : a = a_1 \neq a_0$ (остальные параметры те же)? Продемонстрируйте это при помощи моделирования.

2. Данные Yeast.

Вариант 8

1. Пусть x_1, \dots, x_n — выборка из распределения \mathcal{P} .

Хотим проверить гипотезу $H_0 : \mathcal{P} \in \{N(a, \sigma^2)\}$.

- Можно ли использовать критерий ω^2 для проверки принадлежности выборки к семейству нормальных? Подтвердите свой ответ с помощью моделирования.
- Подойдет ли критерий Лиллиефорса для проверки принадлежности к семейству нормальных? Продемонстрируйте это.

2. Данные Blood Transfusion Service Center.

Вариант 9

1. Пусть x_1, \dots, x_n — выборка из распределения \mathcal{P} .

Хотим проверить гипотезу $H_0 : \mathcal{P} \in \{N(a, \sigma^2)\}$.

- Рассмотрим смесь двух нормальных распределений с различными средними и дисперсиями в качестве альтернативной гипотезы. Примените критерии Колмогоров-Смирнов и ω^2 .
- Какой критерий против такой альтернативы будет более мощный: Колмогоров-Смирнов или ω^2 ?

2. Данные Cryotherapy.

Вариант 10

1. Пусть x_1, \dots, x_n — выборка из распределения \mathcal{P} .

Хотим проверить гипотезу $H_0 : \mathcal{P} \in \{N(a, \sigma^2)\}$.

- Пусть $H_1 : \mathcal{P} = Unif[0, 1]$. Примените критерии Шапиро-Уилка и Лиллиефорса.
- Какой критерий против такой альтернативы будет более мощный?

2. Данные Computer+Hardware.