

Во всех заданиях ниже могут помочь библиотеки `numpy`, `scipy.stats`, `pandas`, `matplotlib` и `seaborn` (другие библиотеки также не воспрещается использовать).

Задание №1. Для всех признаков выведите описательные статистики.

Задание №2. Для всех признаков признаков постройте гистограмму (для каждого ресторана в отдельности). Потом постройте гистограммы для признаков с баллами в зависимости от пола и проверьте гипотезу об одинаковой распределенности. Аналогичную работу проделайте в зависимости от того проходил экзаменуемый курсы по подготовке или нет. И также в зависимости от уровня образования. Итого нужно будет проверить 9 гипотез (3 раза по 3 гипотезы) Гипотезу об одинаковой распределенности можно проверить с помощью критерия хи-квадрат.

Задание №3. Для каждого из выбранных признаков постройте `box-plot` (для каждого результата (`math`, `reading` и `writing` в отдельности)). Что можно сказать о выбросах/аномалиях в данных? Попробуйте как-то объяснить с чем связано наличие выбросов/аномалий (если они есть).

Задание №4. Для признака/признаков с большим числом выбросов вычислите и сравните между собой выборочное среднее и медиану.

Задание №5. Что можно сказать о распределениях каждого признака? Здесь нужно посмотреть отдельно по `math`, `reading` и `writing` и суммарный балл посчитать. Выдвините гипотезы о распределениях признаков и проверьте их (с помощью известных Вам критериев - хи-квадрат и Колмогорова-Смирнова).

Задание №6. Для осмысленных (на Ваш взгляд) пар признаков постройте `scatter-plot`. Сделайте выводы о зависимости этих признаков. Найдите коэффициенты корреляции Пирсона между ними. Найдите ранговые коэффициенты корреляции между ними. Для тех пар признаков в которых наблюдается (визуально) независимость - проверьте это с помощью критерия хи-квадрат;