

Домашнее задание 3 Deadline: 10 февраля 2021

Задание 1. Это задание выполняется на основе данных `reg_hw.txt`. В представленном файле – оценки студентов по ряду предметов.

1. Оставьте в массиве данных только следующие переменные: оценки по микроэкономике (`micro`) и математическому анализу (`math`). Создайте новую переменную, показывающую принадлежность студента к учебной группе, если первые 25 студентов в указанном массиве – студенты группы А, последующие 35 – студенты группы В, последующие 35 обучаются в группе С, остальные – в группе D.
2. На основании однофакторного дисперсионного анализа протестируйте, различается ли средняя успеваемость по математическому анализу в учебных группах. Сформулируйте нулевую гипотезу. Сделайте статистический и содержательный вывод, пояснив на какую информацию в выдаче R Вы опирались.
3. Проверьте аналогичную гипотезу на основе оценок по микроэкономике.

Задание 2. В R оцените линейную регрессию, в которой откликом является переменная «оценки по микроэкономике», предиктором – «оценки по математическому анализу». Используйте тот же массив, что и в предыдущем задании, но без разделения на учебные группы.

1. Запишите спецификацию указанной регрессионной модели с использованием полученных оценок.
2. Проинтерпретируйте оценки коэффициентов (константу и коэффициент при предикторе).
3. Являются ли оценки коэффициентов значимыми? Поясните свой ответ. Покажите в R, как на основе наблюдаемого значения t-статистики были рассчитаны значения p-value (рассчитайте, используя необходимую функцию распределения).
4. Постройте в R 95%-ые доверительные интервалы для коэффициентов (константы и коэффициента при предикторе), на основе них сделайте вывод о (не)значимости коэффициентов. Проинтерпретируйте доверительные интервалы в логике многократного сэмплинга.
5. Проинтерпретируйте полученное значение коэффициента детерминации.
6. Сделайте вывод о (не)значимости коэффициента детерминации. Поясните свой ответ: на какую информацию в выдаче R Вы опирались?