

1. Оценка преподавания родного языка

Сто школьников, изучающих хантыйский язык и язык коми, оценили качество преподавания языка по пятибалльной шкале (от 1 до 5). Данные в файле khantykomi.csv (<https://raw.githubusercontent.com/dashapopova/CorpusMethods/main/practice/khantykomi.csv>).

Определите, есть ли статистически значимая разница между рейтингом преподавания хантыйского и рейтингом преподавания коми?

Задание:

- вычислите в R среднее арифметическое и медиану для переменной khanty и для переменной komi. Приведите 4 числа и формулы, по которым Вы их находили в R. Можем ли мы по значениям медианы и среднего арифметического предложить, что переменные распределены нормально?
- нарисуйте любой график, который, по Вашему мнению, хорошо иллюстрирует данные. Приведите код и картинку. Обоснуйте свой выбор типа графика.
- нормально ли распределены переменные komi и khanty? Для каждой переменной, приведите код теста, нулевую гипотезу, результат теста, проинтерпретируйте результат.
- Сформулируйте и приведите нулевую гипотезу для ответа на вопрос задания (Определите, есть ли статистически значимая разница между рейтингом преподавания хантыйского и рейтингом преподавания коми?). Какой статистический тест Вы планируете использовать для подтверждения или опровержения гипотезы? Обоснуйте свой выбор.
- приведите код, который Вы использовали для подтверждения или опровержения нулевой гипотезы.
- подтверждается или опровергается нулевая гипотеза? Обоснуйте свой ответ.

2. Линейная регрессия: рост и вес супергероев

Задание:

https://raw.githubusercontent.com/dashapopova/Intro-to-R/main/week4/heroes_information.csv

- ➔ изучите данные по супергероям (рост в см, вес в кг)
- ➔ возможно, придется сначала почистить данные, убрать строки с отсутствующими значениями (- или -99), с выбросами, приведите релевантный

код

- приведите формулу линейной регрессии для предсказания роста супергероя по весу
- вычислите по результатам применения модели, какой модель прогнозирует рост супергероя Вашего веса
- можно ли сказать, что линейная модель подходит для моделирования отношения между этими двумя переменными? Приведите аргументы (подсказка: аргументов должно быть не меньше четырех, вспомните, на что надо обратить внимание при интерпретации результатов применения линейной регрессии; если для аргумента нужна иллюстрация, приведите её).