Домашнее задание 2 Корпусные методы, 30.11.2022

Дата сдачи: 07.12. до 23.59; расширенный дедлайн на 60% от оценки до 14.12 23:59; после расширенного дедлайна работы не принимаются.

Присылайте работы в формате .docx, .pdf на мою почту:

daschapopowa@gmail.com

1. *Оценка преподавания родного языка [макс. 3,5 баллов]*

Сто школьников, изучающих хантыйский язык и язык коми, оценили качество преподавания языка по пятибалльной шкале (от 1 до 5). Данные в файле hw2.csv (https://raw.githubusercontent.com/dashapopova/CorpusMethods/main/HWs/hw2.csv). Определите, есть ли статистически значимая разница между рейтингом преподавания хантыйского и рейтингом преподавания коми?

Задание:

▪ 0.25 балла: 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?

▪ 1 балл: нарисуйте любой график, который, по Вашему мнению, хорошо иллюстрирует данные. Приведите код и картинку. Обоснуйте свой выбор типа графика. Балл зависит от адекватности выбора графика задаче, читабельности графика (наличия названия, подписей осей, лейблов, легенды и т.п.), а также от использования дополнительных параметров: цвета, размера подписей и т.п.

▪ 1 балл: нормально ли распределены переменные komi и khanty? Для каждой переменной, приведите код теста, нулевую гипотезу, результат теста, проинтерпретируйте результат.

▪ 0,5 балла: 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.

▪ 0.25 балла: приведите код, который Вы использовали для подтверждения или опровержения нулевой гипотезы.

▪ 0,5 балла: подтверждается или опровергается нулевая гипотеза? Обоснуйте свой ответ.

2. ***Линейная регрессия: рост и вес супергероев [макс. 5,5 балла]***

Задание:

https://raw.githubusercontent.com/dashapopova/Intro-to-R/main/HWs/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 (1 балл)
- ➔ приведите формулу линейной регрессии для предсказания роста супергероя по весу
- ➔ вычислите по результатам применения модели, какой модель прогнозирует рост супергероя Вашего веса (0,5 балла)
- ➔ 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) (4 балла).

3. ***Применение линейной регрессии, коэффициента корреляции Пирсона или критерия Стьюдента в литературе [макс. 1 балл]***

Приведите пример применения линейной регрессии, коэффициента корреляции Пирсона или критерия Стьюдента в научной статье, опубликованных тезисах конференции, книге (не в учебнике по статистике). Желательно взять работу по социолингвистике, политике, социологии, этнографии, или лингвистике. Можно взять работу на английском или на русском языках. Приведите ссылку на статью/тезисы/книгу, или скриншоты релевантных фрагментов текста. Обосновано ли, на Ваш взгляд, было применение одного из этих критериев к данным работы? Как описан результат применения теста, есть ли все необходимые данные для оценки успешности применения теста/модели? Сформулирована ли нулевая гипотеза? Соотнесен ли результат применения гипотезы с ней? Правильно ли сделаны выводы? Есть ли иллюстрация соответствующих данных? Если есть, удачно ли, на Ваш взгляд, выбран тип графика и параметры? Что бы Вы изменили? Если иллюстрации нет, то какой график Вы бы предложили?