

Домашнее задание 1

Корпусные методы, 23.11.2022

Дата сдачи: 30.11. до 23:59; расширенный дедлайн на 60% от оценки до 7.12 23:59;

после расширенного дедлайна работы не принимаются.

Присылайте работы в формате .docx или .pdf на мою почту: daschaporowa@gmail.com

1. Зависит ли район проживания москвича от его национальности? [макс. 4 балла]

Во многих городах есть кварталы, которые называются «китайский», «итальянский», «арабский». Иногда данные наименования носят исторический характер, иногда определяются преобладающим этническим составом населения квартала. Можно ли сказать про Москву, что район проживания человека не зависит от его национальности? К сожалению, данных этнического состава разных районов Москвы я не нашла, но представим такой упрощённый сценарий. (Для заинтересовавшихся: ссылка на этнический состав Москвы по переписи 2010г. http://www.statdata.ru/nacionalnyj_sostav/moskvy).

Сценарий

Предположим, что в Москве 4 округа: М, О, S, и С. Мы опросили 650 случайно выбранных жителей Москвы на предмет их национальной принадлежности (Russian, Tatar, Armenian). Есть ли зависимость между районом проживания и национальностью? Данные представлены в таблице:

	М	О	S	С	Total
Russian	90	60	104	95	349
Tatar	30	50	51	20	151
Armenian	30	40	45	35	150
Total	150	150	200	150	650

Задание:

- 0,5 балла: приведите код, который Вы использовали для того, чтобы ввести данные в R.
- 0,5 балла: сформулируйте нулевую гипотезу.
- 1 балл: какой статистический тест Вы использовали для подтверждения или опровержения гипотезы? Обоснуйте свой выбор. Приведите код.
- 1 балл: подтверждается или опровергается нулевая гипотеза? Обоснуйте свой ответ.
- 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.

2. Применение теста Вилкоксона [макс. 4 балла]

Задание:

- изучите данные по супергероям (рост приведен в см, вес в кг):

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

- 3 балла: примените к данным тест Вилкоксона:
 - выберите переменные;
 - объясните, почему к ним можно применить тест Вилкоксона;
 - сформулируйте нулевую гипотезу;
 - приведите код, который Вы использовали для применения теста (возможно, придется сначала почистить данные, убрать строки с отсутствующими значениями (- или -99), приведите релевантный код);
 - подтвердилась ли нулевая гипотеза? как Вы это определили?
- 1 балл: нарисуйте любой график, который, по Вашему мнению, хорошо иллюстрирует данные (не все данные, а те, которые Вы выбрали для применения теста). 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.

3. Применение теста Вилкоксона или критерия хи-квадрат в литературе [макс. 2 балла]

Приведите пример применения теста Вилкоксона или критерия хи-квадрат 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 p-value, α и т.п.? Сформулирована ли нулевая гипотеза? Соотнесен ли результат применения критерия с нулевой гипотезой? 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? Почему именно такой?