

Домашнее задание №2 (сдаваемое)

Дедлайн: 23:59 24 октября

Для проверки решение ДЗ в файле формата .ipynb необходимо
отправить на следующие почты: fi@edu.hse.ru;

ddkravchenko@edu.hse.ru; darysalnikova@yandex.ru

Тема письма в формате: Фамилия ИО студента, ДЗ2

Задание выполняется на уже знакомых Вам по первому домашнему заданию данных [hwdata.dta](#). Это данные из исследования *Dower, Paul Castaneda, Evgeny Finkel, Scott Gehlbach, and Steven Nafziger. "Collective Action and Representation in Autocracies: Evidence from Russia's Great Reforms." American Political Science Review 112, no. 1 (2018): 125–47.* Статья доступна по [ссылке](#).

Авторы обращались к вопросу, как вероятность коллективного действия народа, исключенного из процесса принятия решений, влияет на склонность элит к перераспределению благ в пользу народа (финансирования публичных благ)? Показатель «изменение в количестве сельских школ с 1860 до 1880 гг. на душу сельского населения уезда» используется для измерения объема финансирования публичных благ. Ниже представлено краткое описание переменных:

ch_schools_pc	Изменение в количестве сельских школ с 1860 до 1880 гг. на душу сельского населения уезда
afreq	Доля лет между 1851 и 1863 гг., в которые были зафиксированы крестьянские выступления
nozemstvo	Бинарная переменная: Единицей закодированы уезды тех губерний, в которых в результате реформы 1864 года земства созданы не были, 0 – в противном случае.
distance_moscow	Расстояние от Москвы до центра уезда
goodsoil	Показатель плодородности почвы
lnurban	Логарифм городского населения уезда на 1863 г.
lnpopn	Логарифм населения уезда на 1863 г.
province_capital	Бинарная переменная: принимает значение 1, если в уезде находился «столичный» город губернии, 0 – в противном случае.

1. Оцените линейную регрессионную модель (модель m1), в которой изменение в количестве сельских школ на душу населения является зависимой переменной, все остальные представленные в таблице переменные – предикторы (переменные взаимодействия включать не нужно). Протестируйте, есть ли мультиколлинеарность в модели m1, насколько сильно мультиколлинеарность отражается на результатах. Используйте как визуальные диагностики, коэффициенты корреляции, так и значения VIF. Проинтерпретируйте полученные результаты и сделайте вывод
2. Протестируйте, есть ли гетероскедастичность в модели:
 - Используйте визуальные способы диагностики гетероскедастичности. Сделайте вывод
 - Используйте формальный тест Бреуша—Пагана. Сделайте вывод
 - Есть ли основания полагать, что вариация ошибок монотонно зависит от одной из объясняющих переменных (выберите переменную самостоятельно)? Протестируйте монотонную зависимость посредством теста Голдфелда–Квандта (Goldfeld–Quandt)
 - Можно ли говорить о том, что есть существенные изменения в значимости оценок коэффициентов, если сравнить исходную регрессионную модель с классическими стандартными ошибками и эту же модель, но переоцененную с робастными стандартными ошибками — состоятельными в условиях гетероскедастичности (используйте тип ошибок HC3)? Сравните результаты
3. Выявите влиятельные наблюдения по мере Кука для модели m1. Удалите поочередно первые пять наиболее влиятельных наблюдений из массива и переоцените модель, укажите, значимым ли образом изменились результаты оценивания модели после удаления того или иного влиятельного наблюдения
4. Определите влиятельные наблюдения по мере DFBETA для коэффициента при переменной afreq