

Домашнее задание №3 (сдаваемое)
Дедлайн: 23:59 2 декабря

Для проверки решение ДЗ необходимо отправить на следующие почты: vebelyakova@edu.hse.ru; darysalnikova@yandex.ru

Задание выполняется на 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)? Показателем «изменение в количестве сельских школ с 1860 до 1880 гг. на душу сельского населения уезда» используется для измерения объема финансирования публичных благ. Частота крестьянских выступлений в годы, предшествующие Земской реформе, – показатель способности к коллективному действию.

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. Проинтерпретируйте полученные результаты: какой характер взаимосвязи объясняющих переменных и отклика, прокомментируйте значимость. Кроме того, прокомментируйте выбор контрольных переменных. Каким требованиям должны соответствовать контрольные переменные? Соблюдаются ли в данном случае эти требования?
2. Проинтерпретируйте значение R^2 . Чем отличается интерпретация R^2 скорректированного ($R^2_{adjusted}$)?
3. Поделите выборку на обучающую и тестовую в пропорции 70% к 30%. Обучите модель и сопоставьте качество модели на обучающей и тестовой выборках. При выполнении этого задания используйте k-блочную кросс-валидацию. Как можно улучшить качество регрессионной модели?
4. 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
5. Протестируйте, есть ли гетероскедастичность в модели:
 - Используйте визуальные способы определения гетероскедастичности. Сделайте вывод
 - Используйте формальный тест Бреуша—Пагана. Сделайте вывод
 - Есть ли основания полагать, что вариация ошибок зависит от одной из объясняющих переменных (выберите переменную самостоятельно)? Протестируйте монотонную зависимость посредством теста Goldfeld-Quandt
 - Используйте робастные стандартные ошибки — состоятельные в условиях гетероскедастичности (используйте тип ошибок HC3). Изменились ли значимо результаты?
6. 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

7. Определите влиятельные наблюдения по мере DFBETA для коэффициента при переменной `afreq`
8. Предложите самостоятельно две альтернативные спецификации регрессионной модели (к примеру, можно взять другой набор предикторов, добавить переменные взаимодействия) и сравните их при помощи информационных критериев AIC и BIC. Сделайте вывод