

Домашнее задание №3 (сдаваемое)

Дедлайн: 23:59 12 февраля

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, ДЗ3

Задание 1. Задание выполняется на уже знакомых Вам по предыдущим домашним заданиям данных hwddata.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
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. Предложите две невложенные спецификации регрессионной модели (зависимую переменную оставьте ту же самую, измените только набор предикторов) и сравните их при помощи информационных критериев AIC и BIC. Сделайте вывод
5. Предложите две вложенные спецификации регрессионной модели (зависимую переменную оставьте ту же самую) и сравните их при помощи соответствующего F-теста. Сделайте вывод

Задание 2. Ознакомьтесь со вторым разделом [данной статьи](#) («Когда МНК-оценку не спасти») и приведите самостоятельно не менее 3 примеров эндогенности (для иллюстрации трех из пяти указанных в статье случаев нарушения экзогенности: пропущенных существенных переменных, ошибок измерения, самоотбора, одновременности, автокорреляции вместе с лагированной зависимой переменной в качестве предиктора). Вы можете придумать собственные примеры или взять примеры из знакомых Вам исследований, в последнем случае поставьте ссылку на источник, опишите пример своими словами. Укажите, какой случай эндогенности иллюстрирует Ваш пример.

Примечание: Задание разрешается оформить в текстовых ячейках файла .ipynb