

Домашнее задание (сдаваемое)

Deadline: 23:59 5 апреля

**Для проверки решение ДЗ необходимо отправить на почту УА
Варвары Мотовниковой: va.mto@yandex.ru**

Задание выполняется на данных [hwdata.dta](https://hwdata.dta.ac.cn/). 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.*

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.

Задание 1. Оцените линейную регрессионную модель, 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?

Задание 2. Протестируйте, есть ли мультиколлинеарность, насколько сильно мультиколлинеарность отражается на результатах. Используйте как визуальные диагностики, коэффициенты корреляции, так и значения VIF. Проинтерпретируйте полученные результаты и сделайте вывод.

Задание 3. Есть ли теоретические основания полагать, что данная модель страдает от гетероскедастичности? Опишите эти возможные основания (пока без статистических тестов, визуализаций, только на основании Ваших теоретических ожиданий).

Задание 4. Протестируйте, есть ли гетероскедастичность в модели.

1. Используйте визуальные способы определения гетероскедастичности. Сделайте вывод.
2. Используйте формальный тест Бреуша—Пагана. Сделайте вывод.
3. Есть ли основания полагать, что вариация ошибок зависит от одной из объясняющих переменных? Протестируйте монотонную зависимость посредством теста Goldfeld-Quandt.
4. Используйте робастные стандартные ошибки – состоятельные в условиях гетероскедастичности (используйте тип ошибок *HC3*). Изменились ли значимо результаты?

Задание 5. Бонусное задание

Покажите, что во множественной регрессии

$$se(\hat{\beta}) = \frac{\sigma \times \sqrt{VIF}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Пояснение:

se – стандартная ошибка,

$\hat{\beta}$ – оценка коэффициента при предикторе,

для оценки σ используется $\sqrt{\frac{RSS}{n-k}}$,

VIF – variance inflation factor, рассчитанный на основе вспомогательной модели для соответствующего предиктора.