

Домашнее задание №1

Практическая эконометрика

13 октября 2020 г.

1 Общие требования

Задание выполняется 1-2 студентами, сдаётся на онэкон в виде архива с 4 файлами - оформленного связного текста с ответами на вопросы, таблицами и графиками (если они необходимы по заданию) в формате pdf, и 1 файл с кодом в R. Они должны называться именами студентов в формате `kalashnov_g_suchkova_o.pdf` и `kalashnov_g_suchkova_o.R`. Файлы R должны запускаться с помощью команды `source('HA1code_sec2.R')` и отработывать без ошибок. В коде в первой строчке в качестве комментария напишите, пожалуйста, имена автора(ов). Часть вопросов факультативные, ответы на них оцениваются бонусными баллами (которые выходят за пределы 60 баллов на домашние работы)

2 срока сдачи задания: 17 октября 23:59 - промежуточный дедлайн, 24 октября 23:59 - окончательный. Можно сделать 2 или больше любых пункта к первому дедлайну только еще 2 пункта после первого дедлайна. Можно сразу сдать все 17 октября. Если первый дедлайн пропущен, то ко второму можно сдать не более 2 пунктов.

2 Задание

В рамках задания мы попробуем реплицировать часть статьи¹

Данные содержат непосредственно данные Data и Stata код (Do-files). Мы будем реплицировать с помощью R, но Stata код полезно почитать, если вы застряли. Единственные 2 файла с данными, которые вам понадобятся – это `Analysis/analysis.dta` и `Intermediate/base_randomization_v2.dta`

¹Vincent Pons "Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France" American Economic Review 2018, 108(6): 1322–1363

Установите в R пакет `readstata13`, чтобы прочитать набор данных. В некоторых пунктах указаны рекомендованные пакеты. Используйте их или любые другие пакеты R на ваше усмотрение.

В рамках этой домашней работы вы можете целиком пропустить подпунктов В, С, D раздела II. А также пропускать все Panel В таблиц с результатами.

Вопросы этой работы разбиты на 4 логических блока, которые содержат 2 типа вопросов: поработать с данными или ответить на содержательный вопрос. Напротив каждого пункта есть пояснения о том, как он будет оцениваться

2.1 General (2 балла + 1 бонус)

1. Какой исследовательский вопрос интересует авторов? Что именно они хотят измерить? (1 балл за 1-2 предложения)
2. Почему этот вопрос интересен с точки зрения авторов? А именно: какой вклад он вносит в своей области (политические предпочтения)? С какими другими исследовательскими вопросами связано это исследование? (1 балл за описание по пунктам)
3. Знаете ли вы, почему ещё этот вопрос может быть интересен (со ссылками на литературу)? (1 бонусный балл в случае интересной и релевантной ссылки на литературу)

2.2 Рандомизация (5 баллов + 2 бонуса)

В этом пункте мы попытаемся реплицировать рандомизацию из статьи. Пожалуйста, пользуйтесь функциями-заготовками, которые мы создали для вас в шаблоне. Так на прощ будет проверять

1. Для начала заполните пропуски в тех переменных, которые мы будем использовать
`ifelse(level_randomization==1, prop_leftabstention, prop_leftabstentionmun)`
`ifelse(level_randomization==1, nb_registered_prim, nb_registered_mun)`
2. Сделайте функцию `generate_stratum`, генерирующую на данных номер `stratum`. Для этого отсортируйте данные по убыванию `prop_leftabstention`, предварительно сгруппировав по `territory, departement_code`. Присвойте каждому блоку из 5 строчек одно и то же целое число. Сделать это можно например так: получить вектор от 1 до общего количества строк и поделить на 5, игнорируя остаток. (2 балла при условии прохождения теста)

3. Сделайте функцию `generate_treatment`, которая для заданной колонки `stratum` создает колонку `treatment` таким образом, как это описано в статье на странице. Если вы чувствуете, что не справляетесь с этим пунктом, просто назначайте `treatment = 1` с вероятностью 80% (1 балл в случае просто 80% + 1 бонусный балл в случае «честной» репликации 1 балл при условии прохождения теста)
4. Сделайте функцию `allocate_canvassers`, которая имея колонки `stratum` и `treatment`, отталкиваясь от количества зарегистрированных граждан на этом избирательном участке (`nb_registered_prim`) и целевых чисел по объему эксперимента (`target_ter`) возвращает `data.frame` с колонками (1 бонусный балл)
 - `stratum` в неизменном виде
 - `treatment` заменяет на NA если эта `stratum` не должна войти в эксперимент
 - создает колонку `allocated`, которая равна 1, если по правилам эксперимента туда следует отправить агитаторов
5. Почему авторы делают рандомизацию по стратам, а не попросту назначают в `treatment` группу избирательный участок с фиксированной вероятностью? Выпишите, какую цель они преследуют (отталкиваясь от знаний с лекций и семинаров и другую, которую сами авторы подчеркивают. Какие детали рандомизации позволяют авторам добиться своей цели (по пунктам)? (1 балл)
6. Что необычного отмечают сами авторы статьи в своём способе рандомизировать? (выпишите в виде пунктов) (1 балл)

2.3 Balance on covariates

(3 балла + 1 бонус) В этом пункте вы будете реплицировать таблицу Table 2—Summary Statistics. Добавьте эту таблицу в ваш письменный отчет и выводите ее как `text` в результате выполнения кода

1. Постройте таблицу с описательной статистикой. Сгруппируйте данные по `treatment` и внутри группы посчитайте средние. В качестве результата вы должны иметь функцию `summary_table`, которая принимает на вход исходный `data.frame` и производит `data.frame` с колонками `treatment`, `control` и названиями переменных в `row.names`. Вам может пригодиться пакет `dplyr` (функции `group_by`, `t`, `summarise` or `summarise_all`). Вам поможет этот материал: <https://dplyr.tidyverse.org/> (2 балла)
2. Посчитайте тем же способом стандартные отклонения и проведите t-test. В качестве результата вы должны иметь функцию `balance_on_covariates`, которая принимает на

вход исходный `data.frame` и производит `data.frame` с колонкой `t_test` и названиями переменных в `row.names`. (1 бонусный балл).

3. С какой целью авторы приводят эту таблицу? (1 балл)

2.4 Результаты (5 баллов + 4 бонус)

1. С помощью функции `read.dta13` импортируйте `analysis.dta`
2. отфильтруйте данные по `territory_in = 1`
3. Зачем нужна эта фильтрация? (1 бонусный балл)
4. Создайте функцию `make_models`, которая принимает на вход название переменной `y`, название переменной, отвечающей за предыдущие результаты, вектор названий прочих контрольных переменных и данные. Возвращает 3 оцененные модели. Вам могут пригодиться функции `update`, `reformulate`. (1 балл)
5. С помощью пакета `stargazer` реплицируйте одну таблицу из основных результатов и один плацебо тест (Panel A). Вы можете использовать параметр `type='html'` для экспорта в word или `type='latex'` для экспорта в Latex. Переменная `out` содержит имя файла, в который необходимо записать результат. Со остальными параметрами разберитесь сами по мере необходимости. Ваш код должен в точности воспроизводить те таблицы, которые содержатся в работе (править руками их нельзя. (2 балла если из таблиц можно получить коэффициент при `treatment` для каждой модификации модели + 1 бонусный балл в случае удачного оформления: в таблице нет ничего лишнего, строчки имеют человекочитаемые названия и из каждой колонки понятно, о какой модели идет речь).
6. Какой из полученных вами результатов (какой коэффициент в какой таблице) отвечает на главный исследовательский вопрос? Проинтерпретируйте его. (0.5 баллов)
7. Объясните, почему несмотря на то, что авторы провели эксперимент, они включают контрольные переменные? Сравните оценки коэффициентов и их стандартные ошибки. Почему авторы называют оценку `within estimator`? Как бы они оценивали не `within estimator`? Зачем делать именно `within estimator`? (0.5 баллов)
8. В регрессиях в качестве контрольных переменных используются результаты прошлых выборов. Является ли это "плохим контролем"? (0.5 баллов)
9. Зачем нужен плацебо тест, если авторы уже проводили `balance of covariates`? (0.5 баллов)

10. Можете ли вы сами задать ещё какие-нибудь содержательные вопросы к результатам? Может авторы в каком-то месте неверно интерпретируют коэффициент. Может на их данных можно еще что-то важное посчитать? (1 бонусный балл в случае интересного вопроса)
11. Ваши стандартные ошибки отличаются от тех, которые были в статье. Почему? Сделайте так, чтобы стандартные ошибки совпадали? (1 бонусный балл)