

Регрессионный анализ: панельные данные и каузальность

Введение: план действий и правила игры

5 сентября 2023

К чему стремимся: в каком направлении мы движемся

- выйти в два измерения: как пространственное, так и временное (привет панельным данным!)
- выйти хорошо вооруженными (подходящими методами для работы с панельными данными)
- практиковать разборчивость (понимать, когда и за какой инструмент – инструмент в широком смысле – хвататься)
- переосмыслить свои работы (в рамках курса Вас ждет творческая работа – эссе) в терминах каузального вывода

В связи с этим нужно быть готовым к:

- моделям с самыми разными эффектами (фиксированными, случайными, смешанными)
- спагетти (первый график-спагетти уже сегодня) и как сделать такие графики удобоваримыми
- подаче под «соусом» каузального вывода
- не статистикой единой: обсуждению дизайну исследования

Наш план действий

- ❶ Фейерверк эффектов для анализа панельных данных:
 - фиксированные эффекты (**FE**)
 - случайные эффекты (**RE**)
 - смешанные эффекты (**ME**)
 - кроме того, модели с лагированными переменными (**FDL, LDV, ADL**)
- ❷ Рисуем идеальную картинку для каузального вывода: рандомизированный эксперимент, считаем «голыми руками»
- ❸ Каузальный вывод: ключевые понятия, логика, ограничения. Сопровождение: картинки (**DAG**).
Квазиэкспериментальный дизайн
- ❹ Подключаем тяжелую артиллерию: инструменты **IV**
- ❺ **DiD** + мэтчинг
- ❻ Заключение: Презентуем эссе и готовимся к экзамену

Если Вы чувствуете, что FE смешался с RE,
то Вам помогут:

Я – тот же человек, что ведет занятия

Сальникова Дарья Вячеславовна

(e-mail: dsalnikova@hse.ru; darysalnikova@yandex.ru;
web-page: <https://www.hse.ru/staff/salnikova>)

Вы получите ответ в течение 24 часов.

Наш ассистент

Информация появится чуть позже

Материалы по курсу

Материалы по курсу доступны на [странице](#)

Что Вас ожидает

Формат занятий

Семинары:

- новый материал
- дискуссии
- практические занятия: решение задач и анализ данных (RStudio)

Типы заданий

- Домашние задания
- Проверочные работы ($\times 2$)
- Контрольная работа ($\times 1$)
- Активность – работа на семинарах
- Эссе ($\times 1$)

Правила прохождения дистанции (1)

Заветная формула

Итог = $0.15 \times \text{ДЗ} + 0.05 \times \text{Проверочная1} + 0.1 \times \text{Проверочная2} + 0.15 \times \text{KP} + 0.15 \times \text{ЭССЕ} + 0.1 \times \text{Активность} + 0.3 \times \text{Экзамен}$

Домашние задания

Итоговая оценка за домашние задания – результат деления суммы баллов, набранной студентом за выполненные им домашние задания, на максимально возможное количество баллов за все домашние задания по курсу.

Опоздание в пределах одного дня (24 часа) – штраф 1 балл. При опоздании более чем на 2 дня (48 ч. после дедлайна) за данное домашнее задание выставляется оценка 0. Но для получения содержательных комментариев можно сдать ВСЕГДА (конечно, не позднее декабря 2023 г.).

Выполнение домашних заданий

Объем НЕ имеет значения и НЕ влияет на оценку,
НО влияет ПОЛНОТА и КОРРЕКТНОСТЬ ответов

Выполнение домашних заданий

Объем НЕ имеет значения и НЕ влияет на оценку,
НО влияет ПОЛНОТА и КОРРЕКТНОСТЬ ответов

- Результаты (в табличной форме)
- Описание результатов текстом (графики при необходимости) и ответы на поставленные вопросы
- Напишите небольшое заключение по результатам анализа: основные выводы, ограничения
- Код в R должен быть реплицируемым. Можно также выполнять задания в Python

Правила прохождения дистанции (2)

Проверочные работы

Closed-book. В случае пропуска по уважительной подтвержденной причине вес проверочной переносится на экзамен, в противном случае при пропуске выставляется оценка 0.

Контрольная работа

Closed-book. В случае пропуска по уважительной подтвержденной причине вес переносится на экзамен, в противном случае при пропуске выставляется оценка 0.

Правила прохождения дистанции (3)

Активность

Оценивается выполнение заданий на семинарах, решение задач у доски и выступление с презентацией, представляющей наработки / результаты эссе. Для получения оценки «отлично» (8, 9, 10 баллов) за данную форму контроля количество баллов, набранное за ответы на семинарах, должно составлять не менее 75% от общего количества занятий (2 семинара будем считать за 1 занятие) по курсу. Семинарская активность вне зависимости от причины пропуска занятий не возмещается!

Правила прохождения дистанции (4)

Экзамен

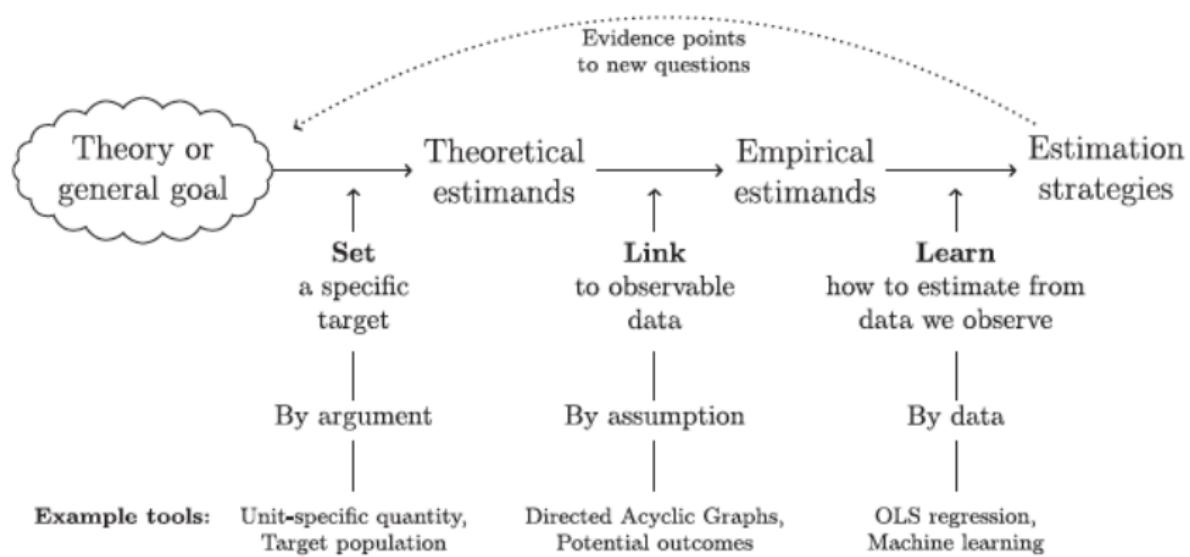
Экзамен проходит в формате closed-book. В экзаменационной работе могут встретиться задания повышенной сложности или задания, только частично разобранные на семинарах, верное решение которых предполагается для возможности получения оценок 9 и 10. Процент таких заданий не превышает 20%.

Эссе: когда сдавать и что сдавать?

Deadline: 23.59 15 декабря 2023

- Файл с текстом эссе (примерно 15000 – 25000 знаков с пробелами). Лучше всего в .pdf
- R-код или Python-notebook (проверьте на реплицируемость без ошибок)
- Массив данных

Схема из Lundberg et al., 2021



Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Abstract (кратко о чем Ваша работа, основные результаты, можно упомянуть о данных и методах)

Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Abstract (кратко о чем Ваша работа, основные результаты, можно упомянуть о данных и методах)
- Введение
 - ▶ Ключевой вопрос. Определить theoretical estimand
 - ▶ Кто еще задавался таким вопросом? Кратко – об их находках.
 - ▶ Чем ответы предыдущих исследователей Вам не угодили? (неполные / есть ошибки – неточности)

Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Abstract (кратко о чем Ваша работа, основные результаты, можно упомянуть о данных и методах)
- Введение
 - ▶ Ключевой вопрос. Определить theoretical estimand
 - ▶ Кто еще задавался таким вопросом? Кратко – об их находках.
 - ▶ Чем ответы предыдущих исследователей Вам не угодили? (неполные / есть ошибки – неточности)
- Гипотезы (формулируются эмпирически тестируемые предположения, НЕ ПУТАТЬ со статистическими гипотезами, множить гипотезы не стоит – не более 3)

Как организовать свой творческий порыв (2)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

Как организовать свой творческий порыв (2)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Данные и методы

- Откуда данные? Опишите источники
- Опишите данные: структура данных, размер выборки, описательные статистики, как измерены Ваши переменные и погрузите в контекст исследования: какой концепт они отражают, выделите empirical estimand
- Обоснование методов, обсуждение альтернативных вариантов
- Спецификация моделей, какую задачу решает оценивание той или иной модели

Как организовать свой творческий порыв (3)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

Как организовать свой творческий порыв (3)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Результаты

- В табличной форме, соответствующей зарубежным публикационным стандартам
- Описание результатов в «сухой» форме (просто то, что получилось, без критической оценки – оставьте это до раздела Discussion)
- Robustness checks (проверки результатов на устойчивость)

Как организовать свой творческий порыв (3)

ЭССЕ (примерно 15000 – 25000 знаков с пробелами)

- Результаты

- В табличной форме, соответствующей зарубежным публикационным стандартам
- Описание результатов в «сухой» форме (просто то, что получилось, без критической оценки – оставьте это до раздела Discussion)
- Robustness checks (проверки результатов на устойчивость)

- Обсуждение результатов и заключение

- Резюме полученных основных результатов и проверок на устойчивость
- Насколько соответствуют результаты предшествующей литературе, если есть расхождения, то почему
- Ограничения Вашего исследования

