

Анализ категориальных данных (АКД)

Введение: план действий и правила игры

11 января 2022

Наши задачи – изучить и научиться применять методы, подходящие для работы с:

- бинарной зависимой переменной
- порядковой зависимой переменной
- номинальной зависимой переменной (более, чем 2 наименования)

Содержание курса

Программа-минимум:

- модели бинарного выбора: логит- и пробит-модели
- модели множественного упорядоченного выбора
- модели множественного неупорядоченного выбора

Содержание курса

Программа-минимум:

- модели бинарного выбора: логит- и пробит-модели
- модели множественного упорядоченного выбора
- модели множественного неупорядоченного выбора

Дополнительные сюжеты:

- регрессионная модель Пуассона
- отрицательная биномиальная регрессионная модель

Кто будет помогать в освоении материала:

Преподаватель

Дарья Сальникова

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

Ассистент

Иоанн Довгополый

Email: iadovgopolyu@edu.hse.ru

Что Вас ожидает

Формат занятий

Семинары (новый материал + дискуссии) и практические занятия (решение задач и отработка в RStudio)

Типы заданий

- Домашние задания
- Проверочные работы (×3)
- Контрольная работа (×1)
- Эссе (×1)
- Экзамен (×1)

Правила прохождения дистанции (1)

Заветная формула

$$\text{Итог} = 0.15 \times \text{ДЗ} + 0.15 \times \text{Проверочные} + 0.15 \times \text{КР} + 0.25 \times \text{ЭССЕ} + 0.3 \times \text{Экзамен}$$

Домашние задания

Опоздание в пределах одного дня (24 часа) – штраф 1 балл.
Опоздание в пределах двух дней (48 часов) – штраф 2 балла.
При опоздании более чем на 2 дня (более 48 ч. после дедлайна) за данное домашнее задание выставляется оценка 0.
НО для получения содержательных комментариев можно сдать ВСЕГДА (конечно, не позднее июня 2022). Такие же правила действуют для эссе.

Правила прохождения дистанции (2)

Проверочные работы

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

Контрольная работа

Контрольная работа может включать практическую часть в RStudio. В случае пропуска по уважительной подтвержденной причине вес переносится на экзамен, в противном случае при пропуске выставляется оценка 0.

Дополнительные баллы

Активность на занятиях + презентация по эссе

Правила прохождения дистанции (3)

Экзамен

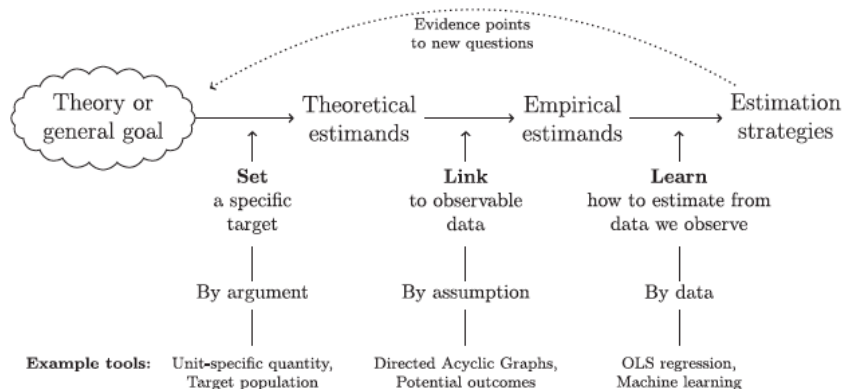
Экзамен проходит в формате closed-book. В экзаменационной работе могут встретиться задания повышенной сложности или задания, только частично разобранные на семинарах, верное решение которых предполагается для возможности получения оценок 9 и 10. Процент таких заданий не превышает 20%.

Эссе: когда сдавать и что сдавать?

Deadline: 23.59 1 июня 2022

- Файл с текстом эссе (примерно 15000 – 20000 знаков с пробелами). Лучше всего в .pdf
- R-код (проверьте на реплицируемость без ошибок)
- Массив данных

Схема из Lundberg et al., 2021



Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 20000 знаков с пробелами)

Как организовать свой творческий порыв (1)

ЭССЕ (примерно 15000 – 20000 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 – 20000 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 – 20000 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 – 20000 знаков с пробелами)

Как организовать свой творческий порыв (2)

ЭССЕ (примерно 15000 – 20000 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 – 20000 знаков с пробелами)

Как организовать свой творческий порыв (3)

ЭССЕ (примерно 15000 – 20000 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 – 20000 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