Лекция 6. Мера склонности. Мэтчинг

Практическая эконометрика.

План на сегодня

Контрольные переменные, чтобы избежать смещения
- Примеры
- Confounders
- Предположение условной независимости

Включение confounders в регрессию

Propensity score

Причинный эффект vs Механизм

Table of Contents

Пример: эффект от вакцинации на летальность от штамма дельта. Источник 30.06.21

- Среди тех, кто получил две дозы используемых в Англии вакцин, вариантом дельта с февраля по 21 июня 2021 года заразились 7235 человек, из них умерли 50 человек. Грубая летальность (Crude case fatality ratio, CFR), то есть подтвержденные смерти в результате заражения, поделенные на выявленные случаи заражения, равна 0,7%.
- Среди тех, кто не прививался вообще, дельтой заразились 53 822 человек, но умерли только 44 человека. Это дает летальность 0,08%.
- Противники прививок делают вывод, что смертность среди привитых почти на порядок превышает смертность среди непривитых при заражении вариантом дельта.
- Что не так в рассуждениях?

Пример: эффект от вакцинации на летальность от штамма дельта. Источник 30.06.21

- В Англии сначала вакцинировали людей старше 50 и людей из группы риска, на которые приходится подавляющее большинство смертей среди заболевших ковидом. Позже началась массовая вакцинация менее возрастных граждан, которые переносят болезнь легче. (ссылка на разбор Панчина)
- Среди непривитых больше молодых, среди привитых больше пожилых, которые в принципе умирают от ковида чаще молодых (вакцина снижает эти риски, но не на 100%). Поэтому изучать следует каждую возрастную группу отдельно.

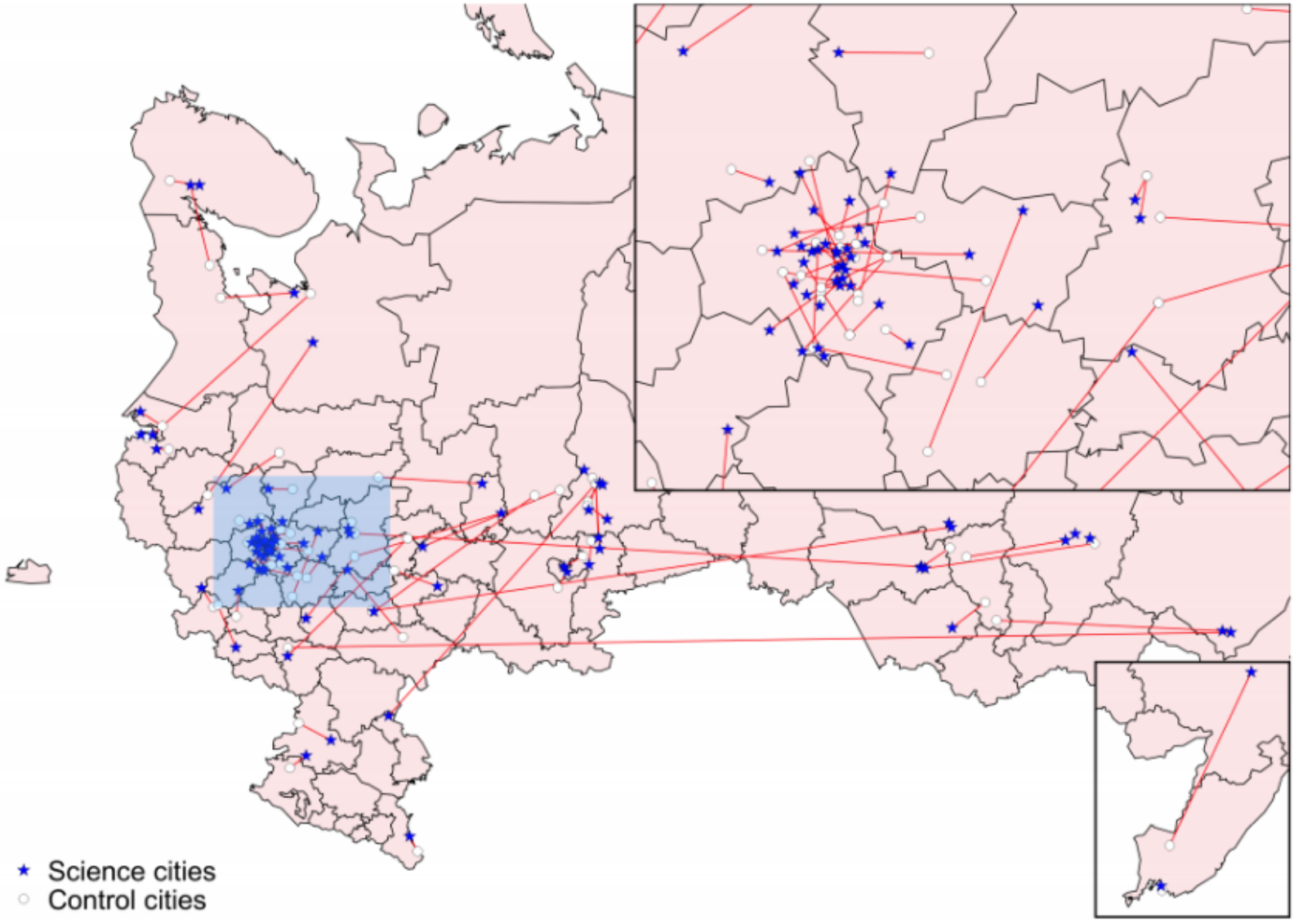
Пример: эффект от вакцинации на летальность от штамма дельта. Источник 30.06.21

- Среди вакцинированных в группе 50+ заразились вариантом дельта 3546 человек, умерли 50 человек. Среди невакцинированных их было 976 (невакцинированных людей в этой группе в принципе меньше), умерли 38 человек. Тогда CFR при вакцинации пожилых снижается в 2,76 раза ((38 / 976) $\div$ (50 / 3546) = 2,76)).
- Из вакцинированных младше 50, заразившихся вариантом дельта, не умер никто. Среди невакцинированных молодых людей, которые заразились дельтой (их было 52846 человек), 6 смертей.

Пример 2: долгосрочный эффект от R&D
Schweiger, Stepanov, Zacchia (2019)

- Сопоставление городов по климатическим условиям, географическому положению, размеру и численности населения в середине 20 века.
- Значимый эффект от долгосрочной бюджетной поддержки (вложений в R&D) на уровень развития города в 2000-е (долю высококвалифицированной рабочей силы, освещённость, число патентов...)

Пример 2: Долгосрочный эффект от R&D (Schweiger, Stepanov и Zacchia 2018)



Пример 3: Эффект от программ переобучения безработных? (Денисова, Карцева, 2009)

- Опрос клиентов служб занятости двух российских регионов
- "Характеристики участников программ переобучения и тех, кто в них не участвовал, различаются. Так, в первом регионе средний возраст участников программ составляет 34,4, а не участвовавших — 38,7 года. Доля женщин среди участников программ в обоих регионах выше, чем среди не участвовавших. Образовательный уровень участников и не-участников примерно одинаков".

Пример 3: Эффект от программ переобучения безработных? (Денисова, Карцева, 2009)

Таблица 1

Оценки общего эффекта программы переобучения *(в %)*

	Участники программы	Контрольная группа	Разность	Эффект
		Регион 1		
Вероятность иметь работу на момент проведения обследования	68,6	69,7	−1,1	−0,2
Вероятность иметь хотя бы один эпизод занятости после выбытия из реестра	84,1	85,5	−1,4	−0,6
Вероятность высокой (свыше 6 тыс. руб.) заработной платы	5,4	4,8	0,6	0,6
Длительность текущего периода безработицы, мес.	16,7	18,5	−1,8	−1,3
		Регион 2		
Вероятность иметь работу на момент проведения обследования	72,0	72,0	0,0	2,0
Вероятность иметь хотя бы один эпизод занятости после выбытия из реестра	89,0	87,3	1,7	3,0
Вероятность высокой свыше 6 тыс. руб.) заработной платы	6,0	3,7	2,3	1,8
Длительность текущего периода безработицы, мес.	15,4	16,8	−1,4	−0,2

Примечание. Все эффекты статистически не значимы.

Пример 3: эффект от программ переобучения безработных? (Денисова, Карцева, 2009)

Оценки групповых эффектов программы переобучения

Группа	Эффект			
	вероятность иметь работу на момент проведения обследования	вероятность иметь хотя бы один эпизод занятости после выбытия из реестра	вероятность высокой заработной платы (свыше 6 тыс. руб.)	длительность текущего периода безработицы, мес.
	Регион 1			
Возраст				
до 30	−10,3**	−1,7	−5,6***	−4,1
30–45	0,4	0,1	3,1**	−2,5
после 45	11,1**	−0,7	3,1^	5,3
Пол				
мужчины	1,9	−1,6	1,5	−5,1
женщины	1,0	−0,3	0,4	−0,6
Образование				
общее среднее	27,5***	13,4***	−0,1	−12,2*
начальное профессиональное	−7,1	0	3,1	1,5
среднее профессиональное	4,1	−1,5	−3,5*	0,4
высшее профессиональное	−5,4^	−3,6	1,9	−2,3
Состояние здоровья				
инвалидность	6,8	−17,7***	0,0	4,5
нет инвалидности	−1,0	1,2	0,7	−2,4
Место жительства				
город	−1,6	−0,7	1,2	−3,1
сельская местность	6,5	−0,5	−2,2	4,6

Проблема в Confounders

- Covariates – X, коррелирующие с Y
- Confounders – X, коррелирующие с Y и с T

Иллюстрация 1

	Y_1	Y_0	X
Пациент 1	-	37.8	Из Европы
Пациент 2	-	37.6	Из Европы
Пациент 3	-	40	Из Азии
Пациент 4	36.6	-	Из Европы
Пациент 5	38	-	Из Азии
Пациент 6	39.2	-	Из Азии

В чем проблема и что можно сделать?

- Нет баланса по X!
- Что с $T_i \perp (Y(1)_i, Y(0)_i, X_i)$?
- Мы можем посчитать эффект отдельно для каждой подгруппы

Иллюстрация 2

	Y_1	Y_0	X
Пациент 1	-	37.8	Эксперимент в 2019 P = 0.33
Пациент 2	-	37.6	Эксперимент в 2019 P = 0.33
Пациент 4	36.6	-	Эксперимент в 2019 P = 0.33
Пациент 3	-	40	Эксперимент в 2020 P = 0.66
Пациент 5	38	-	Эксперимент в 2020 P = 0.66
Пациент 6	39.2	-	Эксперимент в 2020 P = 0.66

- В экспериментах разные P. По чему теперь нет баланса?
- Что с $T_i \perp (Y(1)_i, Y(0)_i, X_i)$?
- Для каждой группы отдельно выполнено?

Иллюстрация 3

	Y_1	Y_0	X
Пациент 1	-	37.8	Эксперимент в 2019 P = 0
Пациент 2	-	37.6	Эксперимент в 2019 P = 0
Пациент 4	-	36.6	Эксперимент в 2019 P = 0
Пациент 3	40	-	Эксперимент в 2020 P = 1
Пациент 5	38	-	Эксперимент в 2020 P = 1
Пациент 6	39.2	-	Эксперимент в 2020 P =1

- Можем что-то сделать?

Unconfoundedness[1] и Overlap

- $T_i \perp (Y(1)_i, Y(0)_i, X_i)$ - идеальный эксперимент
- Вероятность попасть в тритмент-группу известна и одинакова для всех
- $T_i \perp (Y(1)_i, Y(0)_i)|X_i$ - unconfoundedness (CIA, conditional independence assumption). Если взять людей с одинаковыми характеристиками, то факт, что они в такой-то группе, не зависит от потенциальных исходов
- $e(X_i) = E(T_i|X_i) \in (0, 1)$ - overlap. Вероятность попадания в тритмент-группу зависит от характеристик и ненулевая для всех значений X

[1] Angrist и Pischke 2008, Раздел 3.2.1.

Пример 4: эффект от обучения на результаты по математике (Barnard и др. 2003)

- Абитуриентам из бедных семей случайным образом предлагалась грант на обучение в частной школе
- Предполагалось выдавать грант случайным образом, но
  - Детям из сильных школ давали грант с большей вероятностью
- Выполнено ли $(X, Y_1, Y_0) \perp T$?

Итого:

$$\text{ATE} = \frac{N_H}{N}\left(\frac{1}{N_{TH}}\sum_{T=1,S=H} Y - \frac{1}{N_{CH}}\sum_{T=0,S=H} Y\right) + \frac{N_L}{N}\left(\frac{1}{N_{TL}}\sum_{T=1,S=L} Y - \frac{1}{N_{CL}}\sum_{T=0,S=L} Y\right)$$

- Индекс H - сильна школа, индекс L - слабая школа.
- Что не так, если не выполнено unconfoundedness?
- Что не так, если не выполнен overlap?
- Что делать, если X принимает слишком много разных значений?

Table of Contents

Включение в регрессию

$$Y\tilde{T} + X$$

Это все равно, что сначала оценить модель только на X, а потом посмотреть на эффект на остатках (CUPED)

$$m(X) = E(Y|X)$$

$$\widehat{ATE} = \frac{1}{n_1}\sum_{T=1}(Y - m(X)) - \frac{1}{n_0}\sum_{T=0}(Y - m(X))$$

Можно построить модель отдельно для treatment и контроля

$$m_1(X) = E(Y|X, T = 1)$$
$$m_0(X) = E(Y|X, T = 0)$$

$$\widehat{ATE} = \frac{1}{n_1}\sum_{T=1}(Y - m_1(X)) - \frac{1}{n_0}\sum_{T=0}(Y - m_0(X))$$

Это все равно, что оценить регрессию вида

$$Y = X + T + X * T$$

[1] По-русски можно почитать у Ениколопов 2009.

Неверная спецификация линейной модели

- Может привести к еще большему смещению!

Лучше использовать один из методов, который мы обсудим

Машинное обучение и LOOP estimator

$$m(X) = E(Y|X)$$

В качестве модели $m(x)$ можно использовать любой метод машинного обучения, который нравится.

- Разбить выборку на K частей (folds)
- Оценивать $m(x)$ на данных, исключая один fold
- Использовать $Y - m(X)$ для оценки эффекта. (Эти шаги - кросс-валидация)
- Использовать бутстрап (спец. метод для построения доверительных интервалов) для оценки дисперсии (подробнее о бутстрапе в след. лекциях)

Table of Contents

Balancing score[2]

- Достаточная статистика

$$T_i \perp (Y(1)_i, Y(0)_i)|X_i <=> T_i \perp (Y(1)_i, Y(0)_i)|e(X_i)$$

- Propensity score: $e(X_i) = P(T_i = 1|X_i)$
- Смысл леммы: чтобы избавиться от смещения в оценке τ, вместо всех ковариатов достаточно проконтролировать на меру склонности. Доказательство у Imbens и Rubin (2015, Глава 15) и Rubin (1978)

[2]Можно почитать у Ениколопов 2009.

[2]Angrist и Pischke 2008, Раздел 3.3.

Способы применить propensity score

- Blocking
- Matching
- Regression Imputing (редко)
- Weighting

Blocking

- Вычисляем propensity score.
- Разбиваем наблюдения по блокам propensity score: например, 0.2-0.4, 0.4-0.6, 0.6-0.8
- $\widehat{ATE} =$
$$\frac{N_{0.2-0.4}}{N}\left(\frac{1}{N_{T,0.2-0.4}}\sum_{T=1,0.2-0.4} Y - \frac{1}{N_{C,0.2-0.4}}\sum_{T=0,0.2-0.4} Y\right) +$$
$$\frac{N_{0.4-0.6}}{N}\left(\frac{1}{N_{T,0.4-0.6}}\sum_{T=1,0.4-0.6} Y - \frac{1}{N_{C,0.4-0.6}}\sum_{T=0,0.4-0.6} Y\right) +$$
$$\frac{N_{0.6-0.8}}{N}\left(\frac{1}{N_{T,0.6-0.8}}\sum_{T=1,0.6-0.8} Y - \frac{1}{N_{C,0.6-0.8}}\sum_{T=0,0.6-0.8} Y\right)$$
Формула для трёх блоков
- Что плохого в пропуске данных?

Matching

- Вычисляем propensity score.
- Находим наблюдения с самыми близкими значениями propensity score. Остальные выбрасываем
- Вычисляем обычную оценку для ATE

Regression Imputing

(На доске 7.10.22)

Weighting

- Вычисляем propensity score.
- Берем наблюдения из диапазона 10-90
- $\widehat{ATE} = \frac{1}{N}\sum_{T=1}\frac{1}{e(X)}Y - \frac{1}{N}\sum_{T=0}\frac{1}{1-e(X)}Y$

Почему бы все не взять?

Все это и есть взвешивание

- Matching – веса 0/1
- Blocking: $\frac{N_H}{N_{TH}}$ и $\frac{N_L}{N_{CH}}$
- Weighting: $\frac{1}{e(X)}$ и $\frac{1}{1-e(X)}$

(Выкладки на доске и доп. файле, почему это всё - ATE!)

Double Robustness

А еще можно сделать и то и другое (выкладки на доске и в доп. файле)

Double Robustness

- $e(X) = P(T = 1|X)$ – по определению настоящая (postulated) мера склонности к попаданию в тритмент-группу.
- $m_1(X) = E[Y|T = 1, X]$, $m_0(X) = E[Y|T = 0, X]$ – истинная зависимость Y от X в двух группах.
- $ATE = E[Y_1 - Y_0]$ – по определению истинный эффект воздействия.
- $\widehat{ATE_{DR}} = \frac{1}{n}[\sum_{i=1}^{n}(\frac{T_i * Y_i}{\widehat{e(X_i)}} - \frac{(T_i - \widehat{e(X_i)}) * \widehat{m_1(X_i)})}{\widehat{e(X_i)}}] -$
  $\frac{1}{n}[\sum_{i=1}^{n}(\frac{(1-T_i) * Y_i}{(1-\widehat{e(X_i)}))} + \frac{(T_i - \widehat{e(X_i)})) * \widehat{m_0(X)})}{(1-\widehat{e(X_i)})}]$
- $\widehat{e(X)}$ – оценённая на данных вероятность попадания в тритмент-группу в зависимости от характеристик.
- $\widehat{m_1(X)}$, $\widehat{m_0(X)}$– оценённые на данных зависимости Y в тритмент- и контрольной группе от характеристик.

Double Robustness

- $\widehat{ATE_{DR}} = \frac{1}{n}[\sum_{i=1}^{n}(\frac{T_i * Y_i}{\widehat{e(X_i)}} - \frac{(T_i - \widehat{e(X_i)}) * \widehat{m_1(X_i)})}{\widehat{e(X_i)}}] -$
  $\frac{1}{n}[\sum_{i=1}^{n}(\frac{(1-T_i) * Y_i}{(1-\widehat{e(X_i)}))} + \frac{(T_i - \widehat{e(X_i)})) * \widehat{m_0(X)})}{(1-\widehat{e(X_i)})}]$
- Достаточно оценить правильно либо $e(X)$, либо $m_1(X) = E[Y|T = 1, X]$, $m_0(X) = E[Y|T = 0, X]$.