



ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

МГУ имени М. В. Ломоносова

Эксперименты Семинар 1

Ставнийчук Анна

annastavnychuk@gmail.com

Практическая эконометрика

5 сентября 2023 г.

Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать

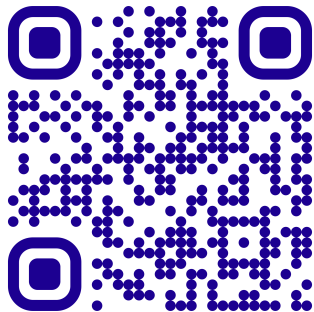


Преподаватели

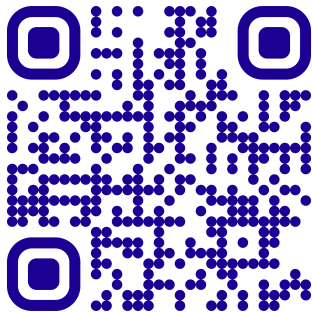
- **Сучкова Ольга Владимировна**
 - лекции (по четвергам на 5 паре)
 - suchkovaolga.91@mail.ru
- **Ставнийчук Анна Юрьевна**
 - семинары в группах 401-403 (по вторникам на 3 паре)
 - annastavnychuk@gmail.com
- **Замниус Алексей Васильевич**
 - семинары в группах 404-411 (по вторникам на 6 паре)
 - a.zamnius@me.com



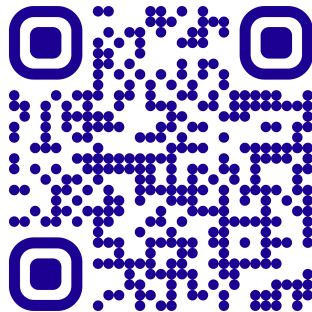
Важные ссылки



(a) Чат курса



(b) Чат семинарской группы



(c) Входной опрос



- **Экзамен** (40%) – письменная индивидуальная работа, январь
- **Домашние задания** (30%) – практическая работа в парах, конец сентября, конец октября, начало декабря
- **Контрольная работа** (20%) – письменная индивидуальная работа, середина октября
- **Коллоквиум**¹ (10%) – устная индивидуальная работа, конец ноября
- ★ **Бонусы** (до 15%)

Ориентиры критериев: 40–65–85, отдельный критерий по коллоквиуму

¹ для получения «хор» и «отл» необходимо сдать коллоквиум хотя бы на 4 из 10



№	Неделя	Тема лекции	Важное
1	4 - 10 сентября	Эксперименты	
2	11 - 17 сентября	Эксперименты	
3	18 - 24 сентября	Мощность, множественное тестирование гипотез	
4	25 сентября - 1 октября	Снижение дисперсии	
5	2 - 8 октября	Мэтчинг	
6	9 - 15 октября	Мэтчинг	Сдача ДЗ 1
7	16 - 22 октября	Контрольная работа	Контрольная работа
8	23 - 29 октября	Большая размерность	
9	30 октября - 5 ноября	Гетерогенные эффекты	
10	6 - 12 ноября	Разрывная регрессия	
11	13 - 19 ноября	Разрывная регрессия	Сдача ДЗ 2
12	20 - 26 ноября	Коллоквиум	Коллоквиум
13	27 ноября - 3 декабря	Разность разностей, синтетический контроль	
14	4 - 10 декабря	Событийный анализ	
15	11 - 17 декабря	Ступенчатая разность разностей	
16	18 - 24 декабря	Обобщение методов	Сдача ДЗ 3
17	25 - 31 декабря	Резерв	



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать



Разминочный пример

Есть 2 лекарства А и В от гипотетической болезни. Больные поступают на лечение в разной степени тяжести состояния. В таблице находится процент неудачного лечения (чем он меньше, тем лучше):

	Средней тяжести	Тяжелые	Итого
А	15% (210/1400)	30% (20/100)	16% (240/1500)
В	10% (5/50)	20% (100/500)	19% (105/550)

Какое лекарство эффективнее?



Разминочный пример

Есть 2 лекарства А и В от гипотетической болезни. Больные поступают на лечение в разной степени тяжести состояния. В таблице находится процент неудачного лечения (чем он меньше, тем лучше):

	Средней тяжести	Тяжелые	Итого
А	15% (210/1400)	30% (20/100)	16% (240/1500)
В	10% (5/50)	20% (100/500)	19% (105/550)

Какое лекарство эффективнее?

А 1400 из 1500 имели среднюю тяжесть заболевания

В 500 из 550 были в тяжелом состоянии



Разминочный пример

Есть 2 лекарства А и В от гипотетической болезни. Больные поступают на лечение в разной степени тяжести состояния. В таблице находится процент неудачного лечения (чем он меньше, тем лучше):

	Средней тяжести	Тяжелые	Итого
А	15% (210/1400)	30% (20/100)	16% (240/1500)
В	10% (5/50)	20% (100/500)	19% (105/550)

Какое лекарство эффективнее?

А 1400 из 1500 имели среднюю тяжесть заболевания

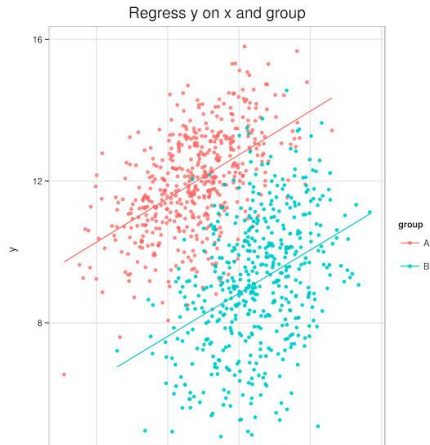
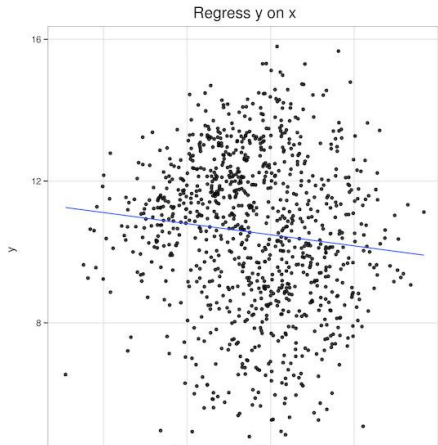
В 500 из 550 были в тяжелом состоянии

- Правильный выбор зависит от структуры ваших данных



Разминочный пример

Парадокс Симпсона: явление в статистике, когда при наличии двух групп данных, в каждой из которых наблюдается одинаково направленная зависимость, при объединении этих групп направление зависимости меняется на противоположное.



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать

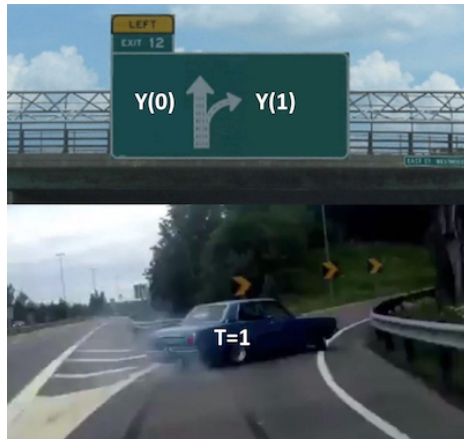


Условные обозначения

- X_i – независимые переменные (covariates)
- T_i – бинарная переменная воздействия (treatment variable):

$$T_i = \begin{cases} 1, & \text{воздействие на объект } i \text{ оказано} \\ 0, & \text{воздействие на объект } i \text{ не оказано} \end{cases}$$

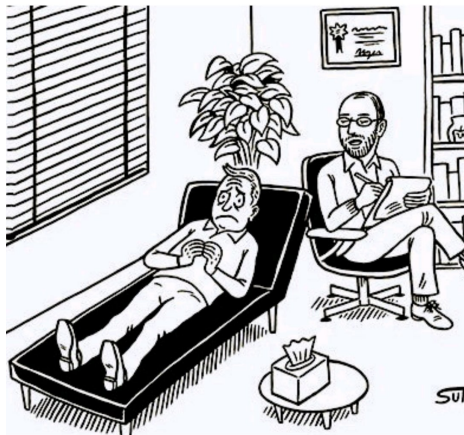
- Y_{i1}, Y_{i0} – потенциальные исходы (potential outcomes) [factual and counterfactual]



Условные обозначения

- **Наблюдаемые исходы** Y_i отличаются от **потенциальных исходов**
 - Потенциальные исходы являются **гипотетическими** случайными величинами
 - Наблюдаемые исходы являются **фактическими** случайными величинами
- Наблюдаемые исходы являются функцией от потенциальных исходов:

$$Y_i = T_i \cdot Y_{i1} + (1 - T_i) \cdot Y_{i0}$$



And are the potential outcomes in the room with us now?



Условные обозначения

- **Наблюдаемые исходы** Y_i отличаются от **потенциальных исходов**
 - Потенциальные исходы являются **гипотетическими** случайными величинами
 - Наблюдаемые исходы являются **фактическими** случайными величинами
- Наблюдаемые исходы являются функцией от потенциальных исходов:

$$Y_i = T_i \cdot Y_{i1} + (1 - T_i) \cdot Y_{i0}$$

- Тогда эффект воздействия для **конкретного наблюдения** равен разнице между двумя состояниями мира для этого наблюдения (потенциальными исходами):

$$\tau_i = Y_{i1} - Y_{i0}$$

- В чем тут проблема?



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать



Фундаментальная проблема причинного вывода

- Чтобы оценить эффект воздействия для конкретного индивида, мы должны знать потенциальные исходы сразу для двух его состояний мира
- Реально мы наблюдаем только одно из них – либо, если индивид подвергся воздействию, либо, если он ему не подвергался
- Оценка индивидуального эффекта требует доступа к данным, которых у нас физически не может быть
- Если с распределением индивидуального эффекта воздействия (treatment effect) работать не получается, будем довольствоваться средними величинами



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты**
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать



Средние эффекты

- **Средний эффект воздействия** (average treatment effect)

$$ATE = \mathbb{E}[\tau_i] = \mathbb{E}[Y_{i1} - Y_{i0}] = \mathbb{E}[Y_{i1}] - \mathbb{E}[Y_{i0}] \xrightarrow{p} \frac{1}{N_1} \sum_{i=1}^{N_1} Y_{i1} - \frac{1}{N_0} \sum_{i=1}^{N_0} Y_{i0}$$

- **Средний эффект воздействия на подвергшихся воздействию** (average treatment effect for the treatment group)

$$ATT = \mathbb{E}[\tau_i | T_i = 1] = \mathbb{E}[Y_{i1} - Y_{i0} | T_i = 1] = \mathbb{E}[Y_{i1} | T_i = 1] - \mathbb{E}[Y_{i0} | T_i = 1]$$

- **Средний эффект воздействия на не подвергшихся воздействию** (average treatment on the non-treated)

$$ATnT = \mathbb{E}[\tau_i | T_i = 0] = \mathbb{E}[Y_{i1} - Y_{i0} | T_i = 0] = \mathbb{E}[Y_{i1} | T_i = 0] - \mathbb{E}[Y_{i0} | T_i = 0]$$

- И средние, и индивидуальный эффект воздействия нельзя напрямую рассчитать, но мы будем пробовать их оценить



Средние эффекты

- И средние, и индивидуальный эффект воздействия нельзя напрямую рассчитать, но мы будем пробовать их оценить
- Самая простая идея для оценки АТЕ, которая всем придет в голову, взять простую разницу в средних:

$$\mathbb{E}[Y_1|T = 1] - \mathbb{E}[Y_0|T = 0]$$

- Но тут всё не так просто, после небольших преобразований мы получим следующее (доказательство тут):

$$\begin{aligned} \mathbb{E}[Y_1|T = 1] - \mathbb{E}[Y_0|T = 0] = \\ = \underbrace{\mathbb{E}[Y_1] - \mathbb{E}[Y_0]}_{\text{ATE}} + \underbrace{\mathbb{E}[Y_0|T = 1] - \mathbb{E}[Y_0|T = 0]}_{\text{Selection Bias}} + \underbrace{(1 - \pi)(ATT - ATnT)}_{\text{Heterogeneous treatment effect bias}} \end{aligned}$$



Средние эффекты

$$\begin{aligned} \mathbb{E}[Y_1|T = 1] - \mathbb{E}[Y_0|T = 0] = \\ = \underbrace{\mathbb{E}[Y_1] - \mathbb{E}[Y_0]}_{\text{ATE}} + \underbrace{\mathbb{E}[Y_0|T = 1] - \mathbb{E}[Y_0|T = 0]}_{\text{Selection Bias}} + \underbrace{(1 - \pi)(ATT - ATnT)}_{\text{Heterogeneous treatment effect bias}} \end{aligned}$$

- **ATE** – интересующий нас эффект
- **Selection Bias** – смещение, возникающее из-за того, что контрольная группа и группа воздействия различались, даже если бы на них не было оказано воздействие, то есть имеет место некоторый дисбаланс
- **Heterogeneous treatment effect bias** – различие в интенсивности эффекта для тритмент и контрольной группы, взвешенное на долю выборки $(1 - \pi)$, которая попала в контрольную группу



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки**
- 7 Симуляция в R
- 8 Что почитать



Предпосылки

$$\begin{aligned} \mathbb{E}[Y_1|T=1] - \mathbb{E}[Y_0|T=0] = \\ = \underbrace{\mathbb{E}[Y_1] - \mathbb{E}[Y_0]}_{ATE} + \underbrace{\mathbb{E}[Y_0|T=1] - \mathbb{E}[Y_0|T=0]}_{\text{Selection Bias}} + \underbrace{(1-\pi)(ATT - ATnT)}_{\text{Heterogeneous treatment effect bias}} \end{aligned}$$

- **Экзогенность воздействия (Independence assumption)** – распределение объекта в тритмент или контрольную группы осуществляется случайно и независимо от его изначальных характеристик $(T_1, Y_0, X)_i \perp T_i$

- $\mathbb{E}[Y_0|T=1] - \mathbb{E}[Y_0|T=0] = 0 \Rightarrow \text{Selection Bias} = 0$
- $\mathbb{E}[Y_1|T=1] - \mathbb{E}[Y_1|T=0] = 0$
 - $(1-\pi)(ATT - ATnT) =$
 $(1-\pi)[(\mathbb{E}[Y_1|T=1] - \mathbb{E}[Y_0|T=1]) - (\mathbb{E}[Y_1|T=0] - \mathbb{E}[Y_0|T=0])] = 0 \Rightarrow$
 $\text{Heterogeneous treatment effect bias} = 0$

- Хорошая рандомизация, а следовательно, и выполнение предпосылок, позволяет нам очистить эффект воздействия от двух типов смещения:

$$ATE = \mathbb{E}[Y_1] - \mathbb{E}[Y_0] = \mathbb{E}[Y_1|T=1] - \mathbb{E}[Y_0|T=0] \xrightarrow{p} \frac{1}{N_1} \sum^{N_1} Y_{i1} - \frac{1}{N_0} \sum^{N_0} Y_{i0}$$



Предпосылки

- Отсутствие «внешних эффектов» воздействия (**SUTVA** – Stable unit treatment value assumption)
 - ① воздействие оказывается только на один объект и внешние эффекты у него отсутствуют
 - ② воздействие гомогенно – существует только один тип тритмента



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R**
- 8 Что почитать



Симуляция в R

- Симуляции – это «игрушечные» примеры
- За ними не стоят реальные данные. Данные для симуляций мы будем специальным образом заранее моделировать
- Это удобно, когда идеально подходящих данных нет. К тому же, живые данные часто могут быть зашумлены из-за других факторов, на которые нам не всегда будет удобно отвлекаться



Симуляция в R

Смоделируем гипотетическую ситуацию. Мы хотим оценить величину **эффекта от использования сайта с расписанием sasc.ws на свободное время студента**

- Предположим, что наша экспериментальная выборка состоит из $N = 1000$ человек
- Для простоты у них будет всего две характеристики (X – возраст и Z – время в пути от дома до ЭФ), имеющих влияние на потенциальный исход (Y_0 и Y_1 – свободное время), причем $X \sim U[18, 25]$, $Z \sim N[60, 20]$
- При этом предположим, что **реальный эффект воздействия** равен $\tau = 15$ минутам
- Будем считать, что **реальная зависимость потенциального исхода от ковариатов и тритмента** выглядит следующим образом:
 - $Y_0 = 240 - 3 \cdot X - Z + 15 \cdot \underbrace{T}_{=0} + \varepsilon = 120 - 3 \cdot X - Z + \varepsilon$ – не было воздействия
 - $Y_1 = 240 - 3 \cdot X - Z + 15 \cdot \underbrace{T}_{=1} + \varepsilon = 120 - 3 \cdot X - Z + 15 + \varepsilon$ – было воздействие



Содержание

- 1 Организационная информация
- 2 Разминочный пример
- 3 Условные обозначения
- 4 Фундаментальная проблема причинного вывода
- 5 Средние эффекты
- 6 Предпосылки
- 7 Симуляция в R
- 8 Что почитать



Что почитать

- Causal Inference: The Mixtape (Scott Cunningham, 2021). Chapter 4. Potential Outcomes Causal Model
- Causal Inference for The Brave and True (Matheus Facure). Chapter 1. Introduction To Causality
- Introduction to Causal Inference (Brady Neal, 2020). Chapter 1. Motivation: Why You Might Care, Chapter 2. Potential Outcomes

