

# Причинно-следственные связи в данных

Георгий Калашнов, Ольга Сучкова

26 февраля 2020 г.

Вот почему попугаи за номером один, два и три были так похожи друг на друга: они были просто одним и тем же попугаем. (А. и Б. Стругацкие)

# План на сегодня

Немного о курсе

Причинная модель Рубина

Оценка эффекта воздействия

# Table of Contents

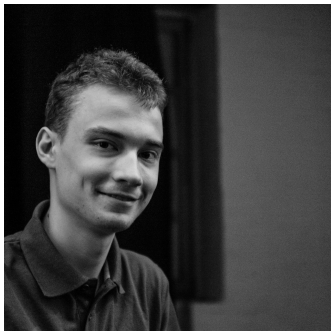
Немного о курсе

Причинная модель Рубина

Оценка эффекта воздействия

# Преподаватели

Георгий Калашнов  
go9513@gmail.com



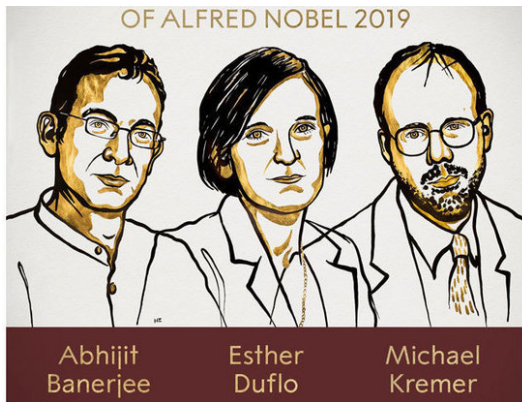
Ольга Сучкова  
suchkovaolga.91@mail.ru



# На какие вопросы можно ответить?

- ▶ Как изменится занятость в результате принятия закона о минимальной заработной плате?
- ▶ Как размер школьного класса влияет на эффективность обучения школьников?
- ▶ Как изменится заболеваемость ВИЧ среди школьниц в Кении после проведительских лекций и/или раздачи бесплатных средств защиты?

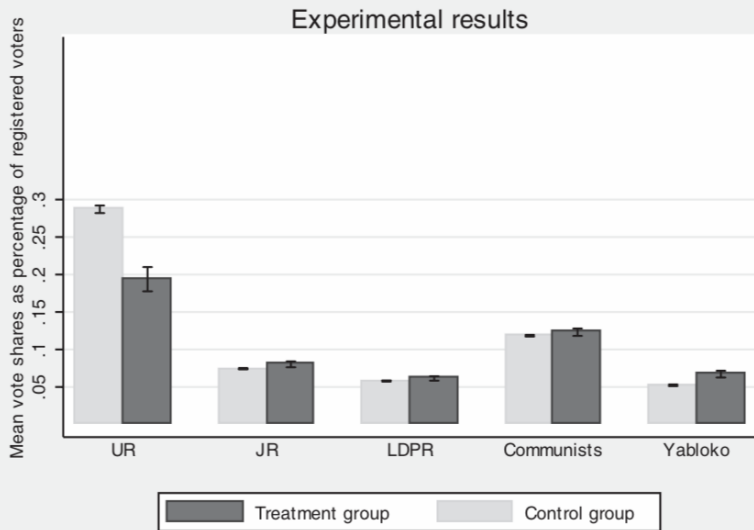
# За "экспериментальный подход к борьбе с глобальной бедностью"



## Пример - жульничество на выборах: Enikolopov, R., Korovkin, V., Petrova, M., Sonin, K. (2012)

- ▶ Рандомизация независимых наблюдателей по 156 участкам из 3164 в Москве.
- ▶ Рандомизация – не «монеткой»/ «кубиком», а по порядковому номеру участка.
- ▶ Однородные участки (исключены больницы и военные части).

# Результат эксперимента



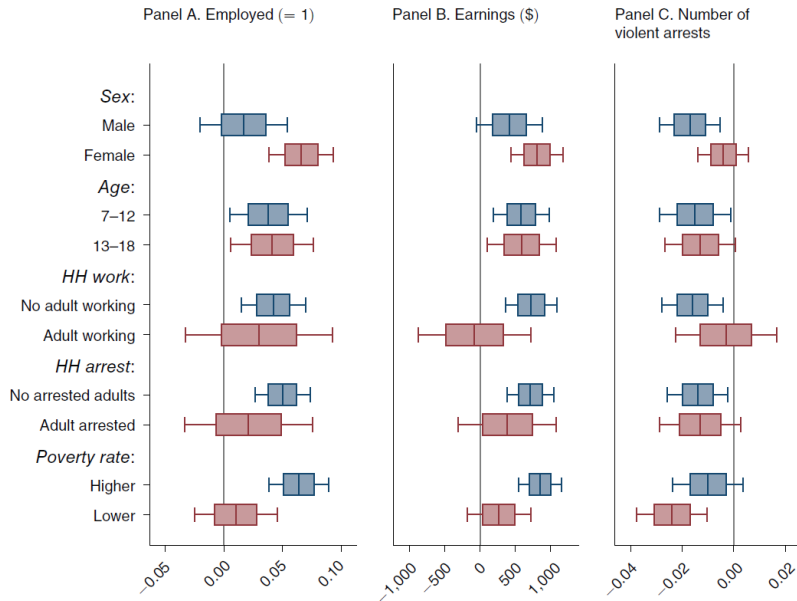


# Пример гетерогенности - программа переселения из ветхого жилья в Чикаго (статья Chyn, 2018)

- ▶ Различия в судьбе детей из переселённых семей и оставшихся жить в неблагополучном районе.



# Эффекты для разных групп (Chyn, 2018)



## Примеры: что если нет эксперимента, но...

- ▶ Мы понимаем, от чего зависит назначение таблетки
- ▶ Кандидат побеждает на выборах с 50.1 процентом голосов
- ▶ Не каждый, кому выдадут таблетку, ее примет

# Про зачёт по курсу

В качестве зачетной работы будет предложено разобрать эмпирическое исследование и ответить на вопросы по нему. Пример - статья Vincent 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.

- ▶ Какой исследовательский вопрос интересует авторов? Что именно они хотят измерить?
- ▶ Какие данные используют исследователи для ответа на этот вопрос? Какой метод используют и почему?
- ▶ Какой из полученных результатов отвечает на главный исследовательский вопрос? Проинтерпретируйте оценки.
- ▶ В уравнениях регрессии в качестве контрольных переменных используются результаты прошлых выборов. Является ли это «плохим контролем»?

# Table of Contents

Немного о курсе

Причинная модель Рубина

Оценка эффекта воздействия

# Потенциальные исходы

**Потенциальные исходы** – температура пациента, если он принял таблетку и если не принял

	$Y_1$	$Y_0$	X
Пациент 1	36.6	36.8	Из Европы
Пациент 2	37	36.6	Из Европы
Пациент 3	38	37.3	Из Азии
Пациент 4	39.2	39.1	Из Азии
Пациент 5	35.3	35	Из Европы

Зная потенциальные исходы, можно оценить средний эффект воздействия:

$$\frac{1}{N_1} \sum Y_1 - \frac{1}{N_0} \sum Y_0$$

# Причинная модель Рубина

Вероятностная модель:

- ▶  $Y_1, Y_0$  – потенциальные исходы (**potential outcomes**)
- ▶  $T = 1$ , если наблюдение в эксперименте и 0 иначе (**treatment variable**)
- ▶  $X$  – Независимые переменные (**covariates**)

Мы хотим оценить распределение эффекта воздействия (**treatment effect**):  $\tau = Y_1 - Y_0$

А скорее средний эффект воздействия (**average treatment effect**):  $ATE = \mathbb{E}_T$

$$\frac{1}{N_1} \sum Y_1 - \frac{1}{N_0} \sum Y_0 \xrightarrow{p} \mathbb{E}_T$$

# Table of Contents

Немного о курсе

Причинная модель Рубина

Оценка эффекта воздействия



# Фундаментальная проблема причинного вывода

	$Y_1$	$Y_0$	$X$
Пациент 1	-	36.8	Из Европы
Пациент 2	-	36.6	Из Европы
Пациент 3	38	-	Из Азии
Пациент 4	39.2	-	Из Азии
Пациент 5	35.3	-	Из Европы

**Fundamental problem of causal inference:** для каждого элемента выборки мы наблюдаем либо  $Y_1$ , либо  $Y_0$

- ▶ Исходные данные:  $(Y_1, Y_0, T, X)$
- ▶ Мы наблюдаем только  $(Y, T, X)$ , где
$$Y = TY_1 + (1 - T)Y_0 - \text{observed outcomes}$$

Можем ли мы оценить эффект воздействия?

	$Y_1$	$Y_0$	X
Пациент 1	36.6	36.8	Из Европы
Пациент 2	37	36.6	Из Европы
Пациент 3	38	37.3	Из Азии
Пациент 4	39.2	39.1	Из Азии
Пациент 5	35.3	35	Из Европы

Средний эффект положительный

	$Y_1$	$Y_0$	X
Пациент 1	-	36.8	Из Европы
Пациент 2	-	36.6	Из Европы
Пациент 3	38	-	Из Азии
Пациент 4	39.2	-	Из Азии
Пациент 5	-	35	Из Европы

Средний эффект отрицательный

Почему?

# Немного определений

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

$$ATE = \mathbb{E}(\tau) = \mathbb{E}(Y_1 - Y_0)$$

- ▶ Средний эффект воздействия на задействованных (**average treatment on the treated**):

$$ATT = \mathbb{E}(\tau | T = 1) = \mathbb{E}(Y_1 - Y_0 | T = 1)$$

- ▶ Средний эффект воздействия на незадействованных (**average treatment on the non-treated**):

$$ATnT = \mathbb{E}(\tau | T = 0) = \mathbb{E}(Y_1 - Y_0 | T = 0)$$

# Смещенность выборки

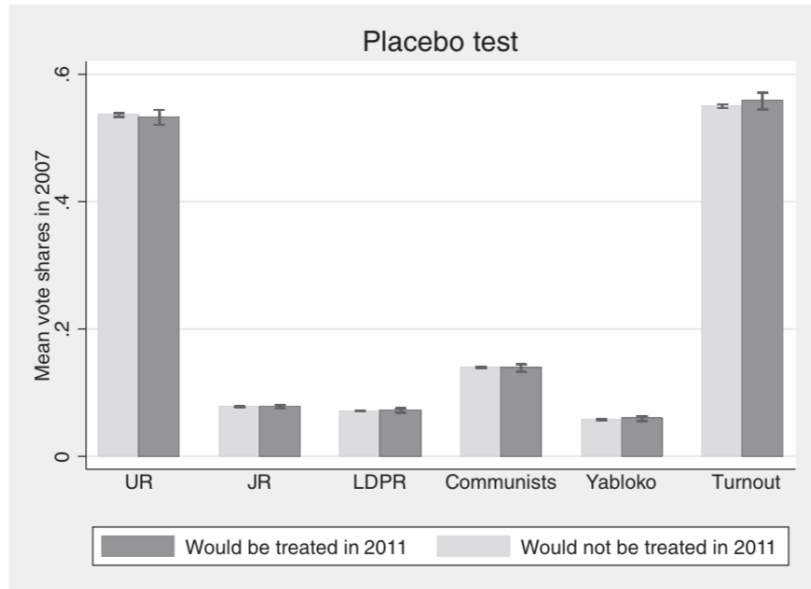
$$\begin{aligned}\frac{1}{N_1} \sum Y_1 - \frac{1}{N_0} \sum Y_0 &\longrightarrow \mathbb{E}(Y_1|T=1) - \mathbb{E}(Y_0|T=0) = \\ \mathbb{E}(Y_1|T=1) - \mathbb{E}(Y_0|T=1) + \mathbb{E}(Y_0|T=1) - \mathbb{E}(Y_0|T=0) &= \\ \mathbb{E}(Y_1 - Y_0|T=1) + \mathbb{E}(Y_0|T=1) - \mathbb{E}(Y_0|T=0) &= \\ = \text{ATT} + \text{Sample Bias} \neq \text{ATT}\end{aligned}$$

$$\begin{aligned}\frac{1}{N_1} \sum Y_1 - \frac{1}{N_0} \sum Y_0 &\longrightarrow \mathbb{E}(Y_1|T=1) - \mathbb{E}(Y_0|T=0) = \\ \mathbb{E}(Y_1|T=1) - \mathbb{E}(Y_1|T=0) + \mathbb{E}(Y_1|T=0) - \mathbb{E}(Y_0|T=0) &= \\ \mathbb{E}(Y_1 - Y_0|T=0) + \mathbb{E}(Y_1|T=1) - \mathbb{E}(Y_1|T=0) &= \\ = \text{ATnT} + \text{Sample Bias} \neq \text{ATnT}\end{aligned}$$

## Что нужно, чтобы не было смещения

- ▶ Экзогенность воздействия:  $(Y_1, Y_0, X)_i \perp T_i$ . Таблетка назначается случайным образом и не связана с потенциальными исходами и другими характеристиками. **Это можно проверить!**

# Баланс ковариатов и плацебо-тест



# Что надо запомнить

- ▶ Наша основная задача: оценить эффект от бинарного воздействия. Например от приема таблетки.
- ▶ Эффект воздействия может отличаться на разных подвыборках
- ▶ Чтобы получить несмещенную оценки нужно назначать таблетку случайно.
- ▶ Можно проверить, случайно таблетка назначалась, или нет

## Литература: книжки

- ▶ Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics, Смещенная выборка: раздел 2, Простой вариант экзогенности воздействия (равенство условных матожиданий): раздел 3.2.1
- ▶ Imbens, G. W., & Rubin, D. B. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, Потенциальные исходы, SUTVA: глава 1, Полная экзогенность: глава 3, парная регрессия для treatment эффектов: глава 7.4



## Литература: статьи

- ▶ Enikolopov, R., Korovkin, V., Petrova, M., Sonin, K., & Zakharov, A. (2013). Field experiment estimate of electoral fraud in Russian parliamentary elections. *Proceedings of the National Academy of Sciences*, 110(2), 448-452.
- ▶ Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34-58.
- ▶ Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, 108(10), 3028-56.