

Причинно-следственные связи в данных

Георгий Калашнов, Ольга Сучкова

16 марта 2020 г.

План на сегодня

Еще про контрольные переменные

Плохой контроль

Эффект посредничества

Престратификация

Table of Contents

Еще про контрольные переменные

Плохой контроль

Эффект посредничества

Престратификация

Повторение

Мы узнали, что контроль нужен, чтобы

- ▶ Снизить дисперсию. Почему так происходит?
- ▶ Потому что **ковариаты** «съедают» часть дисперсии

Что еще:

- ▶ Получить **гетерогенные эффекты**. **Предикторы**
- ▶ Попробовать проанализировать неэкспериментальные данные **Смесители (confounders)**
- ▶ Разобраться в эффекте **Медиаторы**
- ▶ Испортить несмещенность **Bad Control**

Сегодня мы обо всем этом коротко поговорим, про большую часть эффектов поговорим позже

Получить гетерогенные эффекты

- ▶ Помните исследование про переселение в Чикаго?

Плохой контроль¹

Примеры:

- ▶ Контроль на место работы при исследовании влияния образования на доходы

схема

Интуиция:

- ▶ Люди, которые получили white collar job без образования сами по себе крутые – sample bias

¹Angrist и Pischke 2008, Глава 3.2.3

Плохой контроль

Примеры:

- ▶ Контролировать на явку в эксперименте с выборами
- ▶ Контроль на место работы при исследовании влияния образования на доходы

Формально:

$(Y_1, Y_0, X) \perp T$ — не выполнено

$(Y_1, Y_0, X_1, X_0) \perp T$ — предположим это

$$\begin{aligned} E(Y|X=1T=1) - E(Y|X=1T=0) &= \\ E(Y_1|X_1=1) - E(Y_0|X_0=1T=0) &= \\ E(Y_1|X_1=1) - E(Y_0|X_0=1) &= \\ E(Y_1 - Y_0|X_1=1) + (E(Y_0|X_1=1) - E(Y_0|X_0=1)) &= \\ \text{LATE} + \text{Смещение выборки} \end{aligned}$$

Table of Contents

Еще про контрольные переменные

Плохой контроль

Эффект посредничества

Престратификация

Стратификация эксперимента на Французских выборах (Pons 2018)

Precinct ID	# reg. citizens	PO	Precinct ID	# reg. citizens	PO	Stratum
1	1033	0.103	10	961	0.121	1
2	918	0.083	14	1246	0.120	1
3	1175	0.093	5	1158	0.119	1
4	1184	0.103	9	962	0.117	1
5	1158	0.119	16	1021	0.104	1
6	854	0.082	1	1033	0.103	2
7	963	0.092	4	1184	0.103	2
8	876	0.097	8	876	0.097	2
9	962	0.117	15	1098	0.096	2
10	961	0.121	3	1175	0.093	2
11	997	0.067	7	963	0.092	3
12	907	0.087	12	907	0.087	3
13	971	0.067	2	918	0.083	3
14	1246	0.120	6	854	0.082	3
15	1098	0.096	17	1218	0.076	3
16	1021	0.104	13	971	0.067	4
17	1218	0.076	11	997	0.067	4

Groups:

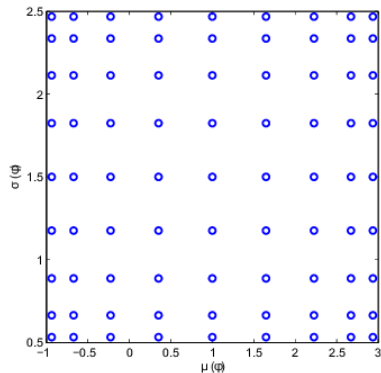
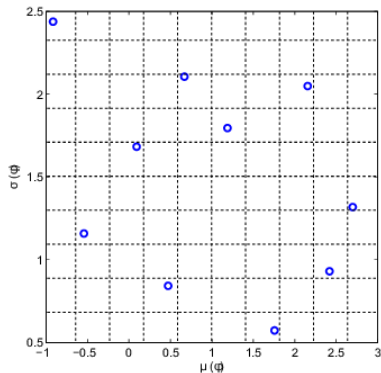
- * 4 precincts are assigned to Treatment
- * 1 precinct is assigned to Control.

Precinct ID	# reg. citizens	PO	Stratum	Treatment
10	961	0.121	1	1
14	1246	0.120	1	0
5	1158	0.119	1	1
9	962	0.117	1	1
16	1021	0.104	1	1

Second stratum an

Престарицизация непрерывной переменной

Latin Hypercube или Геренация по сетке



Литература: книжки



Angrist, Joshua D и Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*.
Princeton university press.



Pons, Vincent (2018). «Will a five-minute discussion change your mind? A countrywide experiment on voter choice in France». в: *American Economic Review* 108.6, с. 1322—63.