

Контрольные переменные и снижение дисперсии

Экспериментальные данные

Георгий Калашнов, Ольга Сучкова

26 февраля 2020 г.

Control

План

Пример

Контрольные переменные

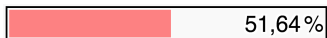
Table of Contents

Пример

Контрольные переменные

Президентские выборы во Франции 2012 (Pons, 2018)

Франсуа Олланд



Николя Саркози



Президентские выборы во Франции 2012 (Pons, 2018)

Франсуа Олланд



Николя Саркози



Эффект от агитации в 0.5 процентных пункта.

Как суметь поймать такой эффект?

2 метода снижения дисперсии:

1. Контрольные переменные (**covariates**)
2. Грамотное планирование эксперимента и престратификация (в другой раз)

Table of Contents

Пример

Контрольные переменные

Ковариаты

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

Ковариаты

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

Ковариаты X меняются вместе с Y_1 и Y_0 (коварируют или (мульти) коррелируют)

$$\text{Cov}(X, Y_1) > 0, \text{Cov}(X, Y_0) > 0$$

Маленький пример

	Y_1	Y_0	X
Пациент 1	-	36.6	Из Европы
Пациент 2	36.6	-	Из Европы
Пациент 3	35	-	Из Европы
Пациент 4	-	36	Из Европы
Пациент 5	37.3	-	Из Азии
Пациент 6	-	39.3	Из Азии
Пациент 7	37.2	-	Из Азии
Пациент 8	-	39.2	Из Азии

- *Повторение.* На глаз. Выполнена ли предпосылка экзогенности? $(Y_1, Y_0, X)_i \perp T_i$

Маленький пример

	Y_1	Y_0	X
Пациент 1	-	36.6	Из Европы
Пациент 2	36.6	-	Из Европы
Пациент 3	35	-	Из Европы
Пациент 4	-	36	Из Европы
Пациент 5	37.3	-	Из Азии
Пациент 6	-	39.3	Из Азии
Пациент 7	37.2	-	Из Азии
Пациент 8	-	39.2	Из Азии

- ▶ *Повторение.* На глаз. Выполнена ли предпосылка экзогенности? $(Y_1, Y_0, X)_i \perp T_i$
- ▶ Высока ли дисперсия с одним и тем же X : $\mathbb{V}(Y|X)$?

Маленький пример

	Y_1	Y_0	X
Пациент 1	-	36.6	Из Европы
Пациент 2	36.6	-	Из Европы
Пациент 3	35	-	Из Европы
Пациент 4	-	36	Из Европы
Пациент 5	37.3	-	Из Азии
Пациент 6	-	39.3	Из Азии
Пациент 7	37.2	-	Из Азии
Пациент 8	-	39.2	Из Азии

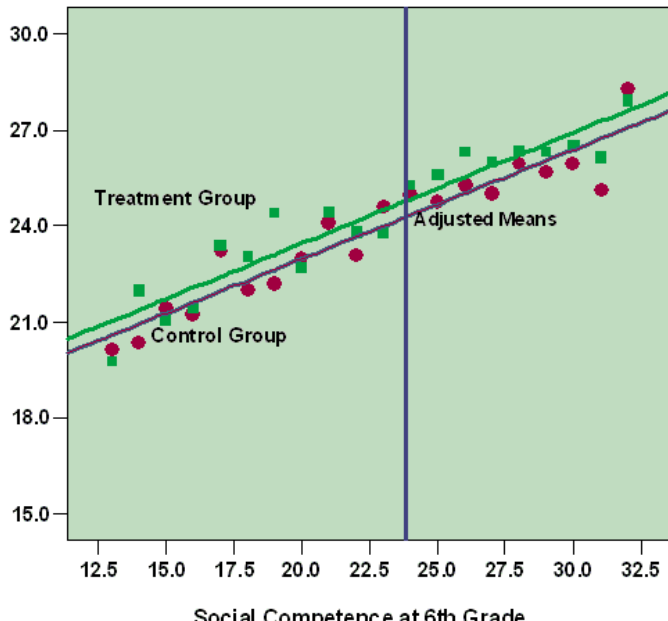
- ▶ *Повторение.* На глаз. Выполнена ли предпосылка экзогенности? $(Y_1, Y_0)_i \perp T_i$
- ▶ Высока ли дисперсия с одним и тем же X : $\mathbb{V}(Y|X)$?
- ▶ У кого в среднем выше температура? $\mathbb{Cov}(X, Y_0) - ?$

Маленький пример

	Y_1	Y_0	X
Пациент 1	-	36.6	Из Европы
Пациент 2	36.6	-	Из Европы
Пациент 3	35	-	Из Европы
Пациент 4	-	36	Из Европы
Пациент 5	37.3	-	Из Азии
Пациент 6	-	39.3	Из Азии
Пациент 7	37.2	-	Из Азии
Пациент 8	-	39.2	Из Азии

- ▶ *Повторение.* На глаз. Выполнена ли предпосылка экзогенности? $(Y_1, Y_0, X)_i \perp T_i$
- ▶ Высока ли дисперсия с одним и тем же X : $\mathbb{V}(Y|X)$?
- ▶ У кого в среднем выше температура? $\text{Cov}(X, Y_0) - ?$
- ▶ Высока ли общая дисперсия: $\mathbb{V}(Y)$?

То же, но на картинке



Контроль, снижающий дисперсию

$$\begin{aligned}\mathbb{V}(Y) &= \mathbb{E}((Y - \mathbb{E}Y)^2) = \\ &\mathbb{E}\left((Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \mathbb{E}Y)^2\right) = \\ &\mathbb{E}\left((Y - \mathbb{E}(Y|X))^2\right) + \mathbb{E}\left((\mathbb{E}(Y|X) - \mathbb{E}Y)^2\right) + \\ &\quad 2\mathbb{E}\left((\mathbb{E}(Y|X) - \mathbb{E}Y)(Y - \mathbb{E}(Y|X))\right) = \\ &\mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) + 0\end{aligned}$$

Результаты выборов во Франции

TABLE 5—IMPACT ON HOLLANDE’S VOTE SHARE

	First round			Second round			Average of first and second rounds		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Panel A. ITT estimation</i>									
Treatment	0.0063 (0.0023)	0.0050 (0.0019)	0.0044 (0.0018)	0.0048 (0.0028)	0.0053 (0.0019)	0.0046 (0.0018)	0.0056 (0.0024)	0.0049 (0.0017)	0.0043 (0.0016)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome and PO		X	X		X	X		X	X
Additional controls			X			X			X
Observations	3,390	2,660	2,660	3,390	2,660	2,660	3,390	2,660	2,660
R ²	0.003	0.516	0.528	0.001	0.632	0.645	0.002	0.645	0.655
Mean in control group	0.3157	0.2994	0.2994	0.5757	0.5597	0.5597	0.4457	0.4295	0.4295
<i>Panel B. Instrumental variable estimation: “allocated to canvassers” instrumented with “treatment”</i>									
Allocated to canvassers	0.0112 (0.0041)	0.0094 (0.0036)	0.0084 (0.0035)	0.0084 (0.0050)	0.0099 (0.0036)	0.0087 (0.0035)	0.0098 (0.0042)	0.0092 (0.0031)	0.0081 (0.0030)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome and PO		X	X		X	X		X	X
Additional controls			X			X			X
Observations	3,390	2,660	2,660	3,390	2,660	2,660	3,390	2,660	2,660

Повторение. А это зачем?

TABLE 2—SUMMARY STATISTICS

	Control group		Treatment group		<i>p</i> -value treatment = control	Number of observations
	Mean	SD	Mean	SD		
<i>Panel A. Electoral outcomes</i>						
Randomization at precinct level	0.504	0.500	0.504	0.500	0.992	3,397
Number of registered citizens	1,014.3	1,097.6	1,133.8	1,605.3	0.022	3,397
Potential to win votes, PO	0.089	0.035	0.089	0.033	0.970	3,397
Voter turnout, 2007 pres. election, first round	0.843	0.050	0.840	0.048	0.231	2,665
Voter turnout, 2007 pres. election, second round	0.837	0.045	0.836	0.045	0.675	2,665
PS vote share, 2007 pres. election, first round	0.274	0.081	0.279	0.081	0.172	2,665
PS vote share, 2007 pres. election, second round	0.515	0.103	0.516	0.101	0.743	2,665

Повторение. А это зачем?

	First round			Second round			Average of first and second rounds		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Panel A. ITT estimation</i>									
Treatment	0.0050 (0.0027)	0.0015 (0.0024)	0.0007 (0.0023)	0.0009 (0.0029)	-0.0013 (0.0024)	-0.0014 (0.0023)	0.0029 (0.0026)	-0.0002 (0.0020)	-0.0006 (0.0020)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome		X	X		X	X		X	X
Additional controls			X			X			X
Observations	2,660	2,133	2,133	2,660	2,133	2,133	2,660	2,133	2,133
R^2	0.002	0.371	0.395	0.000	0.495	0.509	0.001	0.525	0.541
Mean in control group	0.2740	0.2620	0.2620	0.5146	0.5056	0.5056	0.3943	0.3838	0.3838
<i>Panel B. Instrumental variable estimation: "allocated to canvassers" instrumented with "treatment"</i>									
Allocated to canvassers	0.0093 (0.0049)	0.0028 (0.0045)	0.0015 (0.0045)	0.0016 (0.0054)	-0.0025 (0.0046)	-0.0028 (0.0046)	0.0055 (0.0048)	-0.0004 (0.0039)	-0.0011 (0.0039)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome		X	X		X	X		X	X
Additional controls			X			X			X
Observations	2,660	2,133	2,133	2,660	2,133	2,133	2,660	2,133	2,133

- ▶ Плохой контроль: Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics, Смещенная выборка: раздел 3.2.3
- ▶ Контроль ради мощности: Imbens, G. W., Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press. (главы 7.5-7.8)
- ▶ Контроль family-wise ошибки:

Упомянутые статьи

- ▶ Плохой контроль: Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics, Смещенная выборка: раздел 3.2.3
- ▶ Контроль ради мощности: Imbens, G. W., Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press. (главы 7.5-7.8)
- ▶ Контроль family-wise ошибки: