

Introdução à Inferência Causal

Fernando Meireles

Website: www.fmeireles.com

E-mail: fmeireles@ufmg.br

Apresentação

"Without an experiment, natural experiment, a discontinuity, or some other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive", — Sekhon (2009)

Este curso oferece uma introdução à inferência causal para as Ciências Sociais. Particularmente, ele fornece os fundamentos para avaliar e criticar os desenhos de pesquisa quantitativos que economistas, cientistas políticos e sociólogos cada vez mais utilizam para fazer afirmações causais – do tipo $X \rightarrow Y$.

Na prática, isolar o efeito de um fator sobre outro é algo difícil. Na primeira parte do curso, aprenderemos o básico sobre um *framework* que simplifica enormemente essa tarefa: o dos *potential outcomes*. Partindo disso, discutiremos identificação de efeitos causais e veremos algumas ferramentas analíticas que nos ajudarão a distinguir desenhos válidos de outros inválidos. Na segunda parte do curso, nos deteremos principalmente sobre algumas estratégias de identificação canônicas: a experimental (e suas variantes, como o *field experiment* e o *survey experiment*); o baseado em experimentos naturais (como *natural loteries* e políticas governamentais por sorteio); a regressão descontínua; e o que depende da seleção em observáveis.

O curso é estruturado em cinco aulas. Misturando exposições do conteúdo, discussão de trabalhos recentes e exercícios práticos, espera-se que, ao final, as(os) alunas(os) tenham as habilidades necessárias para consumir, e potencialmente realizar, pesquisas baseadas em estratégias de identificação.

Logística do curso

As aulas serão precedidas de exposição do conteúdo por meio de *slides* e baseadas em pequenos tutoriais e discussões de exemplos recentes de pesquisa. Para acompanhar devidamente o curso, é esperado que leituras e exercícios sejam realizadas previamente. Como a parte principal do curso será orientada, auxílio e esclarecimento de dúvidas serão feitos principalmente em aula (outras questões podem ser discutidas antes ou depois destas, ou por e-mail). Na pequena atividade final que teremos (ver a seguir), trabalhos em duplas são permitidos.

Local: Sede IESP-UERJ

Horários: Das 19h às 22h

Carga horária total: 15h

Materiais

Todos os materiais de apoio deste curso, incluindo os *slides* utilizados na parte expositiva das aulas, ficarão disponíveis em um *website* cujo endereço é:

- fmeireles.com/causal

Atividades

Para além da realização de exercícios durante e após as aulas, também teremos uma pequena tarefa ao final do curso. Basicamente, ela consistirá na replicação de um *paper* pré-selecionado em, no máximo, 5 páginas. Com esse exercício, a ideia de fundo é reproduzir desenhos de pesquisa simples, eventualmente sugerindo modificações e aprimoramentos. A execução da replicação é opcional, mas é extremamente recomendada para praticar o que veremos ao longo do curso.

Como sugestão, os *papers* poderão ser escolhidos para o exercício (todos contêm materiais de replicação públicos):

- 1) *Seleção em observáveis:*

[Carnes and Lupu \(2016\)](#). *What good is a college degree? Education and leader quality reconsidered*. The Journal of Politics, 78(1), 35–49.

- 2) *Field experiment:*

[Broockman \(2013\)](#). *Black politicians are more intrinsically motivated to advance blacks' interests: A field experiment manipulating political incentives*. American Journal of Political Science, 57(3), 521–536.

3) Regressão descontínua:

De Magalhaes (2015). *Incumbency effects in a comparative perspective: Evidence from Brazilian mayoral elections*. *Political Analysis*, 1 (23), 113–126.

Informações Específicas

Pré-Requisitos

Embora este curso não tenha pré-requisitos – veremos apenas conteúdos introdutórios –, conhecimentos prévios sobre estatística e programação em R são desejáveis. Para quem carece de bases nesses tópicos, o *website* do curso (fmeireles.com/causal) contém alguns materiais de apoio que devem ser estudados para ajudar a suprir essas lacunas – o que nos ajudará a economizar tempo de aula para trabalharmos com conteúdos substantivos.

Leituras

Cursos de metodologia também requerem leitura. Neste, leremos textos de dois tipos: trechos de manuais, que cobrirão os principais tópicos e conceitos; e vários artigos e *workings papers*, para exemplificar aplicações de estratégias de identificação.

Usaremos principalmente três livros. Primeiro, o mais acessível *Mastering Metrics*, de Angrist and Pischke (2014). Segundo, o clássico de Dunning (2012), *Natural Experiments in the Social Sciences*. Por fim, o *Causal Inference: The Mixtape*, disponibilizado gratuitamente pelo seu autor, Scott Cunningham. Para uma introdução sucinta e não-técnica a várias conceitos, sugiro a leitura de Keele (2015b).

Para quem deseja se aprofundar, ou ainda acompanhar abordagens mais formais à inferência causal, ver Angrist and Pischke (2008); Morgan and Winship (2014); Pearl (2009). Para uma retrospectiva histórica sobre como o uso de inferência causal se tornou prevalente na Economia, e eventualmente nas demais Ciências Sociais, ver Angrist and Pischke (2010) e Druckman et al. (2006).

Usando R

Ao longo deste curso, utilizaremos em vários momentos o ambiente de programação R. Não só por ser a língua franca da pesquisa social quantitativa moderna, o R tem outras vantagens que são úteis para os propósitos desse curso: ele dispõe nativamente de um amplo conjunto de funções para implementarmos aleatorização, testes paramétricos e não-paramétricos simples e para exportarmos resultados de forma simples e reprodutível.

Para quem não tem experiência alguma com o R, um guia introdutório que cobre tudo o que precisaremos, e que pode ser estudado em cerca de uma hora, pode ser visto em [R: Um Guia Prático](#). Uma introdução mais detalhada, direcionada a cientistas sociais, pode ser vista em [Imai \(2017\)](#). Fora esses, *R for Data Science*, de [Wickham and Grolemund \(2016\)](#), ainda é um dos melhores lugares para começar.

Para instalar o R, basta ir ao site do CRAN (*Comprehensive R Archive Network*), que é a rede de fundadores e administradores do *core* da linguagem R, e baixar o *setup* indicado para o seu sistema operacional:

- cran.r-project.org

Feito isto, já é possível usar o R – mas só via *console*, o que não é tão fácil. É por isso que recomendo o uso de uma IDE (i.e. Ambiente de Desenvolvimento Integrado) neste curso: especificamente, o *RStudio*. Para baixá-lo, basta entrar no seguinte site e escolher a opção mais adequada para o seu sistema operacional:

- www.rstudio.com

Atendimento a Necessidades Especiais

Alunas(os) com quaisquer necessidades ou solicitações individuais não devem exitar em procurar auxílio, tanto por e-mail quanto ou pessoalmente.

Política de Gênero

Em cursos de metodologia, homens frequentemente monopolizam a participação. Por conta disso, seguiremos um protocolo muito simples neste aqui: para intervir, é necessário estender a mão; quando mulheres falam, colegas não as interrompem.

Estrutura do Curso

Parte I Preliminares

1.1 O problema fundamental da inferência causal; 1.2 *Potential outcomes*; 1.3 Grafos acíclicos diretos; 1.4 Randomização e experimentos; 1.5 Quantias de interesse; 1.6 Estimação de efeitos e inferência.

Parte II Identificação

2.1 Estratégias de identificação; 2.2 Experimentos naturais; 2.3 Descontinuidades e *kinks*; 2.4 Inferência causal baseada em modelos (seleção em observáveis); 2.5 Outras estratégias de identificação; 2.6 Testes de placebo e *sensitivity analysis*; 2.7 Auxílio na avaliação.

Plano das Aulas

Parte I

Aula 1 – Introdução à inferência causal. Discutiremos as dificuldades envolvidas em fazer afirmações causais. Veremos o *framework* de resultados potenciais de [Rubin \(2005\)](#) e o problema fundamental da inferência causal. Analisaremos o exemplo do texto clássico de [Ross \(2001\)](#) sobre o efeito do petróleo sobre a democracia. Finalmente, como nivelamento, teremos um micro-curso aplicado de R.

Leituras obrigatórias:

- Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press, *Introduction*;

Leituras recomendadas:

- Keele, L. (2015b). The statistics of causal inference: A view from political methodology. *Political Analysis*, 23(3):313–335;
- Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12:487–508;

Aula 2 – Modelos causais e randomização. Aprenderemos a usar grafos acíclicos diretos (DAGS) para modelar processos causais. Também definiremos algumas quantias – efeito causal médio, efeito causal sobre os tratados, etc. Finalmente, veremos como aleatorização nos permite identificar algumas delas.

Leituras obrigatórias:

- Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press, *Randomized Trials*;
- [The Mixtape](#), DAGS;

Exemplos de pesquisa:

- Broockman, D. E. (2013). Black politicians are more intrinsically motivated to advance blacks' interests: A field experiment manipulating political incentives. *American Journal of Political Science*, 57(3):521–536;
- Gerber, A. S., Green, D. P., and Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. *American Journal of Political Science*, 47(3):540–550;

- Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013;
- Hidalgo, F. D., Canello, J., and Lima-de Oliveira, R. (2016). Can politicians police themselves? natural experimental evidence from brazil’s audit courts. *Comparative Political Studies*, 49(13):1739–1773;
- Carreras, M. and Vera, S. (2018). Do corrupt politicians mobilize or demobilize voters? a vignette experiment in colombia. *Latin American Politics and Society*, 60(3):77–95;
- Boas, T. C., Hidalgo, F. D., and Melo, M. A. (2019). Norms versus action: Why voters fail to sanction malfeasance in brazil. *American Journal of Political Science*, 63(2):385–400;

Leituras recomendadas:

- [Directed Acyclic Graph \(Wiki\)](#);

Parte II

Aula 3 – Estratégias de identificação I. Trabalharemos com aproximações da randomização com dados observacionais, como loterias e eventos aleatórios. Cobriremos algumas aplicações do que é normalmente chamado de *natural experiments* e praticaremos algumas coisas no R para explorar exemplos de pesquisas que os usam para estimar quantias de interesse.

Leituras obrigatórias:

- Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press;

Exemplos de pesquisa:

- Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311;
- Bhavnani, R. R. (2009). Do electoral quotas work after they are withdrawn? evidence from a natural experiment in india. *American Political Science Review*, 103(1):23–35;

Leituras recomendadas:

- Keele, L. (2015a). The discipline of identification. *PS: Political Science and Politics*, 48(1):102–6;
- Imai, K. (2017). *Quantitative social science: An introduction*. Princeton University Press, 2.4 Randomized Controlled Trials;

Aula 4 – Estratégias de identificação II. Abordaremos como, sob determinados pressupostos, inferir causalidade explorando descontinuidades, *kinks* e outras regras de atribuição de tratamentos. Além disso, também veremos como estimar efeitos causais condicionando o efeito de uma variável em outra – como em modelos de regressão. Isso, como ficará claro, é extremamente difícil.

Leituras obrigatórias:

- [The Mixtape](#), pp. 35–66, 245–262;

Exemplos de pesquisa:

- Mattingly, D. C. (2017). Colonial legacies and state institutions in china: Evidence from a natural experiment. *Comparative Political Studies*, 50(4):434–463;
- Keele, L. J., Shah, P. R., White, I., and Kay, K. (2017). Black candidates and black turnout: A study of viability in louisiana mayoral elections. *The Journal of Politics*, 79(3):780–791;
- Mazumder, S. (2018). The persistent effect of us civil rights protests on political attitudes. *American Journal of Political Science*, 62(4):922–935;
- Gingerich, D. W. et al. (2014). Brokered politics in brazil: An empirical analysis. *Quarterly Journal of Political Science*, 9(3):269–300;
- Bormann, N.-C., Cederman, L.-E., Gates, S., Graham, B. A., Hug, S., Strøm, K. W., and Wucherpfennig, J. (2019). Power sharing: Institutions, behavior, and peace. *American Journal of Political Science*, 63(1):84–100;

Leituras recomendadas:

- [Introduction to R Markdown](#);

Aula 5 – Validação. Estudaremos as principais formas de por à prova os pressupostos de identificação mais comuns por meio de métodos como testes de placebo e *sensitivity analysis*. Iniciaremos a avaliação final do curso.

Leituras obrigatórias:

- [The Mixtape](#), pp. 35–66, 245–262;

Exemplos de pesquisa:

- Keele, L. J., Shah, P. R., White, I., and Kay, K. (2017). Black candidates and black turnout: A study of viability in louisiana mayoral elections. *The Journal of Politics*, 79(3):780–791;
- Mazumder, S. (2018). The persistent effect of us civil rights protests on political attitudes. *American Journal of Political Science*, 62(4):922–935;

Leituras recomendadas:

- [Introduction to R Markdown](#);

Referências

- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.
- Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.
- Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013.
- Bhavnani, R. R. (2009). Do electoral quotas work after they are withdrawn? evidence from a natural experiment in india. *American Political Science Review*, 103(1):23–35.
- Boas, T. C., Hidalgo, F. D., and Melo, M. A. (2019). Norms versus action: Why voters fail to sanction malfeasance in brazil. *American Journal of Political Science*, 63(2):385–400.
- Bormann, N.-C., Cederman, L.-E., Gates, S., Graham, B. A., Hug, S., Strøm, K. W., and Wucherpfennig, J. (2019). Power sharing: Institutions, behavior, and peace. *American Journal of Political Science*, 63(1):84–100.
- Broockman, D. E. (2013). Black politicians are more intrinsically motivated to advance blacks' interests: A field experiment manipulating political incentives. *American Journal of Political Science*, 57(3):521–536.

- Carnes, N. and Lupu, N. (2016). What good is a college degree? education and leader quality reconsidered. *The Journal of Politics*, 78(1):35–49.
- Carreras, M. and Vera, S. (2018). Do corrupt politicians mobilize or demobilize voters? a vignette experiment in colombia. *Latin American Politics and Society*, 60(3):77–95.
- De Magalhaes, L. (2015). Incumbency effects in a comparative perspective: Evidence from brazilian mayoral elections. *Political Analysis*, 23(1):113–126.
- Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. (2006). The growth and development of experimental research in political science. *American Political Science Review*, 100(4):627–635.
- Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.
- Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311.
- Gerber, A. S., Green, D. P., and Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. *American Journal of Political Science*, 47(3):540–550.
- Gingerich, D. W. et al. (2014). Brokered politics in brazil: An empirical analysis. *Quarterly Journal of Political Science*, 9(3):269–300.
- Hidalgo, F. D., Canello, J., and Lima-de Oliveira, R. (2016). Can politicians police themselves? natural experimental evidence from brazil’s audit courts. *Comparative Political Studies*, 49(13):1739–1773.
- Imai, K. (2017). *Quantitative social science: An introduction*. Princeton University Press.
- Keele, L. (2015a). The discipline of identification. *PS: Political Science and Politics*, 48(1):102–6.
- Keele, L. (2015b). The statistics of causal inference: A view from political methodology. *Political Analysis*, 23(3):313–335.
- Keele, L. J., Shah, P. R., White, I., and Kay, K. (2017). Black candidates and black turnout: A study of viability in louisiana mayoral elections. *The Journal of Politics*, 79(3):780–791.
- Mattingly, D. C. (2017). Colonial legacies and state institutions in china: Evidence from a natural experiment. *Comparative Political Studies*, 50(4):434–463.

- Mazumder, S. (2018). The persistent effect of us civil rights protests on political attitudes. *American Journal of Political Science*, 62(4):922–935.
- Morgan, S. L. and Winship, C. (2014). *Counterfactuals and causal inference*. Cambridge University Press.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Ross, M. L. (2001). Does oil hinder democracy? *World politics*, 53(3):325–361.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.
- Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12:487–508.
- Wantchekon, L. (2003). Clientelism and voting behavior: Evidence from a field experiment in benin. *World politics*, 55(3):399–422.
- Wickham, H. and Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. "O'Reilly Media, Inc."