

Instituto de Estudos Sociais e Políticos - IESP-UERJ

LEGO III (em construção)

Fernando Guarnieri, Fernando Meireles, Bruno Schaefer

1^o semestre 2023

1 Ementa

Este curso parte dos fundamentos trabalhados em LEGO I e II para abordar algumas técnicas estatísticas utilizadas nas Ciências Sociais. Serão abordados modelos não-lineares, técnicas de inferência causal, modelos multinível, dados em painel e séries temporais. O curso tem um caráter prático permitindo ao aluno não só compreender os conceitos como se apropriar das técnicas por meio do uso de softwares de computação estatística (leia-se R).

2 Plano de Trabalho

Nas primeiras aulas trataremos de modelos onde a variável dependente é discreta (binária, categórica, contagem etc). A relação entre apoio no congresso e a queda de uma presidente, a relação entre a cor da pessoa e a probabilidade de ser parada pela polícia, os fatores que fazem com que mulheres trabalhem fora em tempo integral, parcial ou que não trabalhem, são alguns exemplos destes modelos. Neste ponto conceitos como máxima verossimilhança e modelos lineares generalizados passarão a ser familiares. Em seguida veremos como a computação pode ser utilizada para estimar modelos de qualquer tipo. Falaremos de coisas que parecem herméticas como Bootstrap, Monte Carlo e Gibbs Sampler. Técnicas computacionais também permitem o uso de modelos de aprendizado estatístico ou machine learning. Dentre as varias técnicas de machine learning nos dedicaremos às árvores

de decisão, random forrest e boosting. Até este momento as técnicas utilizadas nos permitiam usar certos inputs para prever certos outputs. Passaremos então à inferência causal onde estimaremos o que aconteceria com determinado output sob diferentes tratamentos. Trataremos de técnicas como matching, variáveis instrumentais e diferenças-em-diferenças. Essas técnicas permitem estimar de maneira precisa relações de causa e efeito permitindo explicar, prever e avaliar certas intervenções na realidade como o efeito da adoção da urna eletrônica nos votos válidos ou o efeito da escolaridade na renda. Em seguida abordaremos modelos hierárquicos e multiníveis (como alunos em escolas, projetos de lei e parlamentares etc.) e onde nos interessa tanto a variação inter-unidades quanto intra-unidades. Falaremos então de efeitos fixos, aleatórios e mistos. Esses modelos permitem estimar os efeitos do FUNDEB no desempenho dos alunos de diferentes cidades, a opinião nos estados com base em uma pesquisa de opinião nacional ou a relação entre renda e voto em diferentes contextos, entre muitas outras coisas. Por fim serão apresentados modelos onde as variáveis variam no tempo, como dados em painéis e séries temporal.

3 Avaliação

A avaliação será baseada em três listas de exercício (60%) e um trabalho final (40%). No trabalho o aluno, individualmente ou em grupo, receberá uma base de dados e um problema de pesquisa. Utilizando uma das técnicas examinadas nas aulas 1 a 10 o aluno analisará os dados apresentando os resultados de sua análise na última aula do curso.

4 Leituras

FOX, John. Applied regression analysis, linear models, and related methods. Sage Publications, Inc, 1997.

GELMAN, Andrew; HILL, Jennifer. Data analysis using regression and multi-level/hierarchical models. Cambridge university press, 2006.

MCELREATH, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press, 2020.

JAMES, Gareth et al. An introduction to statistical learning. New York: springer, 2013.

ANGRIST, Joshua D.; PISCHKE, Jörn-Steffen. Mostly harmless econometrics: An empiricist's companion. Princeton university press, 2008.

WOOLDRIDGE, Jeffrey M. Introductory econometrics: A modern approach. Nelson Education, 2016.

GREENE, William H. Econometric analysis. Pearson Education India, 2003.

KING, Gary. Unifying political methodology: The likelihood theory of statistical inference. Cambridge University Press, 1989.

5 Programa de Aulas e leituras

1. Apresentação do curso

MCELREATH (2)

KING (1-2)

2. Variáveis binárias (Logit e Probit)

FOX (15.1)

GELMAN (5)

3. Outras variáveis discretas (Modelos Lineares Generalizados)

FOX (15.2 15.4)

GELMAN (6)

KING (3-4)

Primeira lista de exercícios

4. Simulando Incerteza (Bootstrapping e Cross-Validation)

FOX (16)

GELMAN (7)

5. Simulando estimadores (Monte Carlo Markov Chain e Gibbs Sampler)

GREENE (17)

MCELREATH (8)

6. Aprendizado Estatístico (Decision Trees, Random Forrest e Boosting)

JAMES (8)

Segunda lista de exercícios

7. Inferência Causal (contrafactuais, aleatorização, experimentos e matching)
ANGRIST (3)
8. Inferência Causal (variáveis instrumentais, Regression Discontinuity Design)
ANGRIST (4)
9. Inferência Causal (Dados em Painel, Diferenças-em-Diferenças)
ANGRIST (5)
WOOLDRIDGE (13)
10. Modelos Multinível
GELMAN (12)
WOOLDRIDGE (14)
Terceira lista de exercícios
11. Texto como Dado
GRIMMER, Justin; STEWART, Brandon M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, v. 21, n. 3, p. 267-297, 2013.
HOPKINS, Daniel et al. ReadMe: Software for automated content analysis. Institute for Quantitative Social Science, 2010.
CAMPOS, Luiz Augusto; FERES JÚNIOR, João; GUARNIERI, Fernando. 50 anos da Revista DADOS: uma análise bibliométrica do seu perfil disciplinar e temático. *Dados*, v. 60, n. 3, p. 623-661, 2017.
12. Análise de Redes Sociais
CALVO, Ernesto. Anatomía política de Twitter en Argentina. *Tuiteando# Nisman*. Buenos Aires: Capital Intelectual, 2015.
13. Experimentos (laboratório, campo e survey)
DRUCKMAN, James N. et al. Experimentation in political science. *Cambridge handbook of experimental political science*, p. 3-11, 2011.
14. (preparação de trabalhos)
15. Apresentação de trabalhos