

MAT02010 - Tópicos Avançados em Estatística II

Introdução à Inferência Causal: Apresentações

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2019

O professor

Olá!



Olá!



Graduação

- ▶ **Estatística - UFRGS**

- ▶ 2002 - 2007



- ▶ **Orientador:** Prof. Alberto Cargnelutti

- ▶ **“Métodos de agrupamento para o estudo da divergência genética em cultivares de feijão”**



Mestrado

- ▶ **Estatística - UFMG**

- ▶ 2008 - 2010



- ▶ **Orientadores:** Prof. Enrico A. Colosimo e Prof.^ª Maria do Carmo P. Nunes
- ▶ **“Técnicas estatísticas para avaliação de novos marcadores de risco”**



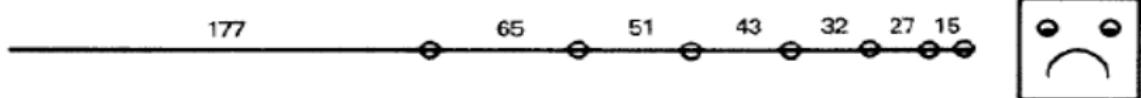
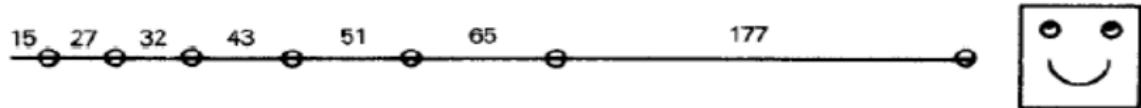
- ▶ **PPG em Infectologia e Medicina Tropical**
- ▶ Cardiopatia em pacientes com doença de Chagas

Doutorado

- ▶ Estatística - UFMG
 - ▶ 2010 - 2014



- ▶ Orientadores: Prof. Enrico A. Colosimo e Prof. Gustavo L. Gilardoni
- ▶ “Análise hierárquica de múltiplos sistemas reparáveis”



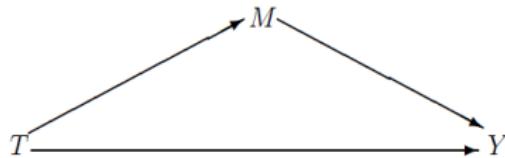
- ▶ ELSA - Brasil

Pós-Doutorado

- ▶ Saúde Pública - UFMG
 - ▶ 2015 - 2017



- ▶ Supervisora: Prof.^a Sandhi M. Barreto
- ▶ “Análise de mediação causal”



- ▶ ++ ELSA - Brasil

Um caminho potencial



Atividades atuais

- ▶ **Departamento de Estatística da UFRGS**
- ▶ **Pós-Graduação em Epidemiologia da UFRGS**
 - ▶ Gabriel Dalla Costa. Predição do tempo de permanência hospitalar através de métodos de aprendizagem de máquina (Orientação de mestrado em conjunto com a Prof.^ª Patrícia Ziegelmann).
 - ▶ Letícia Menegotto. Análise de mediação em modelos de curva de crescimento (Orientação de mestrado em conjunto com a Prof.^ª Vanessa Leotti)
 - ▶ Luísa Back. Efeito mediador da trajetória de índice de massa corporal na relação peso ao nascer e desenvolvimento de diabetes na vida adulta: o Estudo Longitudinal de Saúde do Adulto (Orientação de mestrado em conjunto com a Prof. Alvaro Vigo)
- ▶ +++++++ **ELSA - Brasil**

ELSA - Brasil



Estudo Longitudinal de Saúde do Adulto



- ▶ O **Estudo Longitudinal de Saúde do Adulto** é uma investigação multicêntrica de coorte composta por 15 mil funcionários de seis instituições públicas de ensino superior e pesquisa das regiões Nordeste, Sul e Sudeste do Brasil.
- ▶ A pesquisa tem o propósito de investigar a **incidência** e os **fatores de risco** para doenças crônicas, em particular, as cardiovasculares e o **diabetes**.
- ▶ Em cada centro integrante do estudo, os sujeitos da pesquisa — com idade entre 35 e 74 anos — fazem exames e entrevistas nas quais são avaliados aspectos como condições de vida, diferenças sociais, relação com o trabalho, gênero e especificidades da dieta da população brasileira.

ELSA - Brasil

Centro de investigação Minas Gerais (Hospital Borges da Costa)



ELSA - Brasil

Centro de investigação Rio Grande do Sul (Centro de Pesquisas Clínicas)



ELSA - Brasil

Uma propaganda: mural de bolsas da UFRGS

0

Curtiram

DE: ALVARO VIGO

 alvaro.vigo@ufrgs.br

07/08/2019



Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil) - Estatística

Exatas e Tecnológicas

Descrição: PRÉ-REQUISITOS: Ser aluno regularmente matriculado no curso de Bacharelado em Estatística, a partir do 2º semestre.

ATIVIDADES A SEREM DESENVOLVIDAS PELO BOLSISTA: As principais atividades estão relacionadas ao manuseio de banco de dados utilizando Softwares Estatísticos, extração de dados, desenvolvimento e modificação de sintaxes, análises descritivas e elaboração de tabelas e gráficos. Os interessados deverão encaminhar cópia completa do histórico escolar para o contato estatisticaelsa@ufrgs.br até o dia 18/08/2019 (domingo), preenchendo o campo assunto com o texto "Bolsista Equipe Estatística". Maiores informações podem ser obtidas pelo mesmo endereço. Local da bolsa: Ramiro Barcelos, 2600. Bairro Santa Cecília. Prédio Psicologia UFRGS, 5º andar.

Última atualização em: 07/08/2019



20h/semana



manhã/tarde

Valor: R\$ 400,00

A disciplina

Objetivos

- ▶ Apresentar os fundamentos da inferência causal em estudos aleatorizados e não-aleatorizados
 - ▶ Definição dos efeitos causais
 - ▶ Identificação dos efeitos causais
 - ▶ Estimação e inferência de efeitos causais

Organização

- ▶ **Disciplina:** Tópicos Avançados em Estatística II: Introdução à Inferência Causal
- ▶ **Turma:** U
- ▶ **Modalidade:** Presencial com apoio online (Moodle)
- ▶ **Professor:** Rodrigo Citton Padilha dos Reis
 - ▶ e-mail: rodrigocpdosreis@gmail.com ou citton.padilha@ufrgs.br
 - ▶ Sala: B215 do Instituto de Matemática e Estatística
- ▶ Sala A114 do IME - Campus do Vale - 43111
- ▶ Horário: Quartas, das 8 hs 30 min às 10 hs 10 min

Aulas e material didático

- ▶ **Aulas** (teóricas e práticas)
 - ▶ Exposição e **discussão** dos conteúdos
 - ▶ **Faremos leituras semanais de artigos e capítulos de livros**
 - ▶ Exemplos
- ▶ **Notas de aula**
- ▶ **Canais de comunicação:**
 - ▶ Durante as aulas
 - ▶ Moodle: aulas, materiais, listas de exercícios
 - ▶ Sala de aula virtual: notas das avaliações
 - ▶ e-mail do professor

Aulas e material didático



- ▶ Exemplos com dados reais e dados artificiais (**simulação!**) e exercícios com o apoio do computador:
 - ▶ R e RStudio
 - ▶ DAGitty
 - ▶ ... ?!

Conteúdo programático

1. Experimentos aleatórios

- ▶ Um ensaio aleatorizado
- ▶ Uma estrutura para inferência causal
- ▶ Inferência causal em experimentos aleatorizados

2. Estudos observacionais

- ▶ Entre estudos observacionais e experiências
- ▶ O papel do escore de propensão na estimativa de efeitos causais
- ▶ Diagramas causais (grafos acíclicos dirigidos - DAGs)
- ▶ Sensibilidade
- ▶ Técnicas de estimativa
 - ▶ estratificação
 - ▶ ajuste/regressão
 - ▶ pareamento
 - ▶ escore de propensão: estratificação, ajuste, pareamento, estimativa ponderada pelo inverso da probabilidade de tratamento

3. Tópicos avançados

- ▶ Análise de mediação
- ▶ Análise de interação
- ▶ O aprendizado de máquina na inferência causal

Avaliação

- ▶ Serão realizadas trabalhos teóricos e práticos em sala de aula (MT é a média dos trabalhos) e um seminário (S ; escrito + apresentação) como atividades de avaliação
 - ▶ O seminário e os trabalhos práticos podem ser feitos em duplas
- ▶ Cada atividade de avaliação vale 10 pontos
- ▶ Será realizado um trabalho escrito como atividade de recuperação (TR)
 - ▶ Para os alunos que não atingirem o conceito mínimo e com frequência mínima de 75%

Avaliação

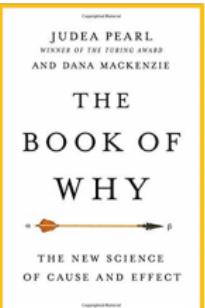
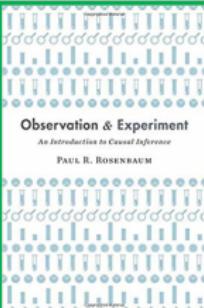
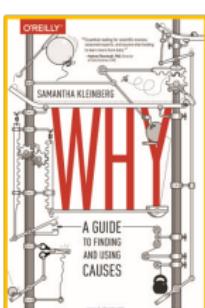
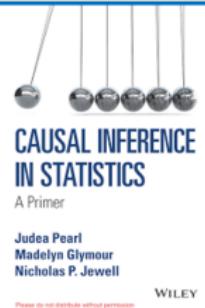
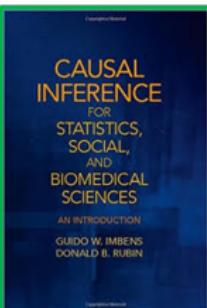
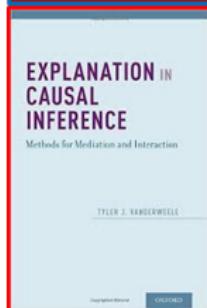
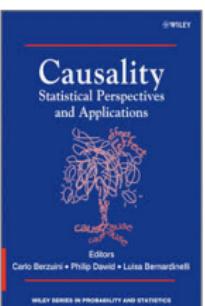
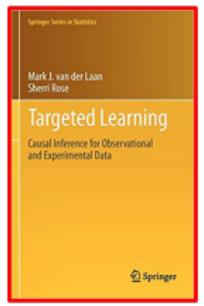
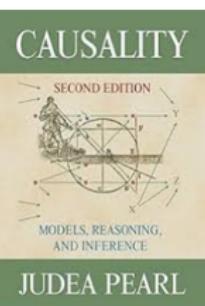
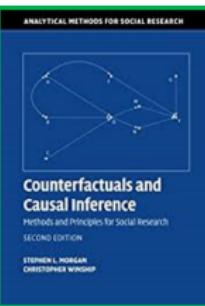
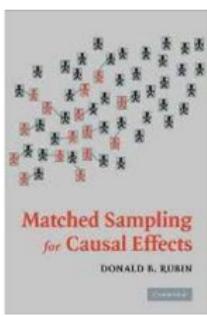
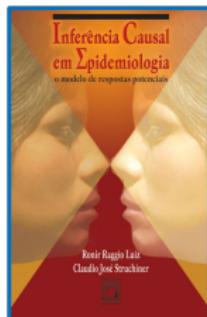
$$NF = 0,5 \times S + 0,5 \times MT$$

- ▶ **A:** $9 \leq NF \leq 10$
- ▶ **B:** $7,5 \leq NF < 9$
- ▶ **C:** $6 \leq NF < 7,5$
- ▶ Se $NF < 6$ e frequência $\geq 75\%$, o aluno poderá realizar o trabalho de recuperação e neste caso

$$NF' = NF \times 0,4 + PR \times 0,6$$

- ▶ **C:** $NF' \geq 6$
- ▶ **D:** $NF' < 6$

Referências bibliográficas



► Ainda: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

Inferência causal

A primeira pergunta é fácil ...



O que é
causa(lidade)?

Uma pergunta mais fácil!

O que é inferência causal?

O termo “inferência causal” passou a denotar certas maneiras especiais de abordar os aspectos causais da análise estatística.^a

^aAalen, O. O., Røysland, K., Gran, J. M. Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society, Series A*, 175:831-861, 2012.

Um comentário



No entanto, a teoria estatística tem sido relativamente **silenciosa** em questões de causalidade. Muitos livros didáticos, especialmente os mais antigos, evitam qualquer menção ao termo, **exceto** em configurações de experimentos aleatórios. Alguns mencionam principalmente para enfatizar que a **correlação** ou **associação** **não é o mesmo que causação**, e alguns até advertem seus leitores para evitar o uso de linguagem causal na estatística. No entanto, para muitos usuários de métodos estatísticos, as declarações causais são exatamente o que elas buscam.¹

¹Imbens, G. W., Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015

Uma terceira pergunta!

Por que estudar inferência causal?

Uber Engineering

Blog ▾ Research ▾ Engineering Offices ▾ Q

[Uber Data](#)



Using Causal Inference to Improve the Uber User Experience

Totti Harinen and Bonnie Li

June 19, 2019





Sign up for Uber Engineering updates:

Your email address

Subscribe

Popular Articles


Uber's Big Data Platform: 100+ Petabytes with Minute Latency
October 17, 2018


Meet Michelangelo: Uber's Machine Learning Platform
September 5, 2017


Introducing Ludwig, a Code-Free Deep Learning Toolbox
February 11, 2019

Por que estudar inferência causal?

Questões causais

- ▶ As escolas *charter* aumentam os escores dos testes dos alunos do ensino fundamental?
 - ▶ Se sim, quão grandes são os ganhos em comparação com aqueles que poderiam ser realizados pela implementação de reformas educacionais alternativas?
- ▶ A obtenção de um diploma universitário aumenta o salário de um indivíduo no mercado de trabalho?
 - ▶ Em caso afirmativo, esse efeito específico é grande em relação aos ganhos salariais que poderiam ser obtidos apenas através de treinamento no trabalho?
- ▶ O uso de uma cédula de borboleta em alguns condados da Flórida na eleição presidencial de 2000 custou votos de Al Gore?
 - ▶ Em caso afirmativo, o número de votos errados seria suficientemente grande para alterar o resultado eleitoral?

Por que estudar inferência causal?

- ▶ Em essência, esses tipos de perguntas são simples questões de causa e efeito da forma X causa Y ?
 - ▶ Se X causa Y , quão grande é o efeito de X em Y ?
 - ▶ O tamanho desse efeito é grande em relação aos efeitos de outras causas de Y ?
- ▶ Uma implicação prática: **ação** (ou manipulação, tratamento, ou intervenção).
 - ▶ Saber “a causa” nos permite tomar uma ação a fim de prevenir (no caso de um efeito nocivo) ou de estimular (no caso de um efeito benéfico).

Por que estudar inferência causal?

- ▶ Simples questões de **causa** e **efeito** são a motivação para muitas pesquisas nas **ciências sociais**, **demográficas** e da **saúde**, embora respostas definitivas para questões de causa e efeito nem sempre sejam possíveis de serem formuladas, dadas as restrições que os pesquisadores enfrentam na coleta de dados e avaliar explicações alternativas.
- ▶ Mesmo assim, há motivos para **otimismo** em relação às nossas habilidades atuais e futuras para tratar efetivamente de questões de causa e efeito.
- ▶ Nas últimas quatro décadas, um **modelo de causalidade** contrafactual **foi desenvolvido** e **aperfeiçoado** e, como resultado, uma estrutura unificada para o julgamento de questões causais está agora disponível.

Modelos estatísticos

Breiman, 2001² sugere que existem dois objetivos na análise de dados:

- ▶ **Predição:** ser capaz de prever o que as respostas vão ser para futuras variáveis de entrada.
- ▶ **Informação:** extrair algumas informações sobre como a natureza está associando as variáveis de resposta às variáveis de entrada.

²Breiman, L. Statistical modeling: the two cultures. *Statistical Science*, 16:199-231, 2001.

Modelos estatísticos

*To Explain or to Predict?*³

- ▶ **Modelagem explicativa:** aplicação de modelos estatísticos aos dados para testar hipóteses causais.
- ▶ **Modelagem preditiva:** aplicação de modelos estatísticos (**mineração de dados, aprendizado estatístico/máquina**) para predição/classificação de novas ou futuras observações.
- ▶ **Modelagem descriptiva:** aplicação de modelos estatísticos para representar de maneira compacta a estrutura dos dados; captura a associação entre as variáveis dependente e independentes (**ausência de hipóteses causais**).

As abordagens explicativa e preditiva são diferentes

- ▶ No entanto, não são necessariamente inconsistentes ou incompatíveis.

³Shmueli, G. To explain or to predict. *Statistical Science*, 25:289-310, 2010.

Próximas aulas

- ▶ Costuma-se dizer: “Não se pode provar causalidade com estatística”.
- ▶ Vamos conhecer argumentos contrários, e assim, passar a afirmarmos que “Só é possível provar causalidade com estatística”!

Por hoje é só!

caffeine causality loop

