

Leg 2: Modelos de Regressão Linear e extensões

2º Semestre de 2024

Terças-feiras, de 9hs às 12hs.

Professores:

Rogério J Barbosa
rogerio.barbosa@iesp.uerj.br

Pedro H. G. Ferreira de Souza
rogerio.barbosa@iesp.uerj.br

Monitores:

Natália Barcelos

Tomás Paixão

1. Apresentação e objetivos

Este curso tem por objetivo fornecer um panorama de diversos Modelos de Regressão e suas aplicações nas ciências sociais. Na **Parte 1**, abordaremos o Modelo de Regressão Linear Múltipla. Revisaremos suas características básicas nas primeiras aulas, aprofundando a compreensão de seus pressupostos e propriedades matemáticas. Na **Parte 2**, abordaremos o problema da incerteza e da inferência estatística em modelos de regressão. Na **Parte 3**, abordaremos modelos não lineares e para dados categóricos ou censurados. Na **Parte 4**, discutimos a teoria subjacente à inferência causal e apresentamos modelos e estratégias que permitem aferir causalidade a partir de dados observacionais. Na **Parte 5**, abordaremos modelos para dados em painel/longitudinais (i.e. nos quais observamos unidades analíticas ao longo do tempo).

2. Pré-requisitos

Este é um curso de nível intermediário no percurso de formação em métodos quantitativos. Assumimos que você:

- Já sabe estatística básica (medidas de tendência central, dispersão e associação; noções de probabilidade; noções de testes de hipótese e inferência estatística)
- Já estudou ou aplicou, ainda que rapidamente/superficialmente, modelos de regressão linear
- Já trabalha com softwares estatísticos que são operados por linha de comando (como R, Python ou Stata)

Se você cursou Leg 1, já tem tudo o que precisa ;-)

3. Software

Utilizaremos o software R, que é livre, gratuito e aberto.

4. Logística do Curso

O curso é baseado em aulas expositivas, sessões de monitoria e realização de listas de exercício. As aulas serão ministradas presencialmente, no IESP.

5. Formas de Avaliação

- Listas de Exercício (5 listas, 15 pontos cada)
- Trabalho Final (25 pontos)

6. Horários das aulas e monitorias

As aulas ocorrerão às terças-feiras pela manhã, de 9hs às 12hs.

Haverá sessões de monitoria, com frequência e horários a serem definidos.

7. Nivelamento em Matemática para Lego 2

O Nivelamento visa familiarizar o aluno com notações e ferramentas que serão imprescindíveis para o devido acompanhamento dos conteúdos. Ele ocorrerá na primeira semana de agosto, entre os dias 5 e 9 de agosto. A frequência a esse curso é obrigatória. Haverá uma lista de exercício que renderá 15 pontos extras para a disciplina principal. Não se preocupe, não há pré-requisitos.

Nivelamento em Matemática para Lego 2

5 a 9 de agosto

Aula 1: Fundamentos de Álgebra Linear (Parte 1)

Segunda-feira, 5 de agosto, 9hs às 12hs

Tópicos:

- Vetores e Matrizes
- Combinação Linear e Produto escalar
- Normas de vetores
- Ortogonalidade

$$\begin{matrix} v_1 & v_2 & v_3 \\ \begin{bmatrix} 0 & 2 & 4 \\ 2 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix} & + & \begin{bmatrix} 0 & 0 & 5 \\ 0 & 0 & 3 \\ 5 & 3 & 0 \end{bmatrix} & = & \begin{bmatrix} 0 & 2 & 9 \\ 2 & 0 & 4 \\ 9 & 4 & 0 \end{bmatrix} \\ \begin{matrix} v_1 & v_2 & v_3 \\ \begin{matrix} 2 & 1 & 4 \\ 4 & 1 & 0 \end{matrix} \end{matrix} & + & \begin{matrix} v_1 & v_2 & v_3 \\ \begin{matrix} 5 & 3 & 0 \end{matrix} \end{matrix} & = & \begin{matrix} v_1 & v_2 & v_3 \\ \begin{matrix} 2 & 4 & 4 \\ 9 & 4 & 0 \end{matrix} \end{matrix} \end{matrix}$$

Aula 2: Fundamentos de Álgebra Linear (Parte 2)

Terça-feira, 6 de agosto, 13hs às 16hs [Extraordinariamente no período da tarde, em função da aula inaugural]

Tópicos:

- Sistemas de equação
- Solução de sistemas por eliminação: o Método de Gauss-Jordan

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$Ax = b$$

Leitura Obrigatória:

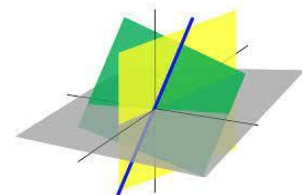
STRANG, Gilbert. *Introduction to Linear Algebra*. 5th Ed. Cap. 1.

Aula 3: Fundamentos de Álgebra Linear (Parte 3)

Quarta-feira, 7 de agosto, 9hs às 12hs

Tópicos:

- Determinantes
- Matrizes Inversas



Leitura Obrigatória:

STRANG, Gilbert. *Introduction to Linear Algebra*. 5th Ed. Cap. 2.

Aula 4: Fundamentos de Cálculo (Parte 1)

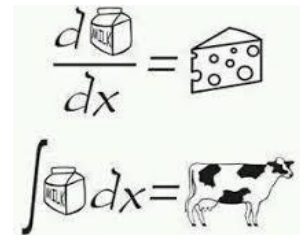
Quinta-feira, 8 de agosto, 9hs às 12hs

Tópicos:

- Funções
- Uma visão geral sobre o Cálculo: Limites, Derivadas e Integrais

Leitura Obrigatória:

A definir.



Aula 5: Fundamentos de Cálculo (Parte 2)

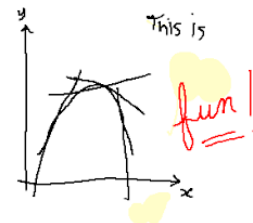
Sexta-feira, 9 de agosto, 9hs às 12hs

Tópicos:

- Algumas Regras de Derivação

Leitura Obrigatória:

A definir.

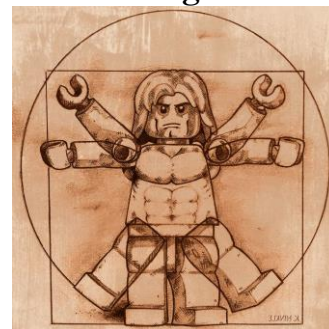


Parte 1: Regressão Linear, o básico

Aula 1: Processos geradores de dados e os pressupostos teóricos da regressão

Tópicos

- A ideia de um Processo Gerador de Dados
- Os pressupostos do modelo linear (Gauss-Markov)
- Derivando o estimador da regressão linear
- Melhor reta ou efeitos causais?
- A diferença entre *model-based* e *design-based inference*



Leitura Obrigatória:

HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022. (Cap 4. Describing Relationships. Disponível em: <https://theeffectbook.net/ch-DescribingRelationships.html>).

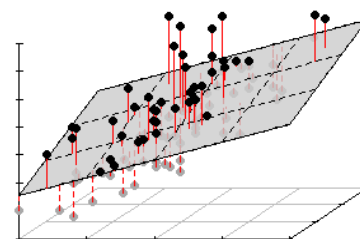
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022. (Cap 5. Identification. Disponível em: <https://theeffectbook.net/ch-Identification.html>).

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 2 – The Simple Regression Model)

Aula 2: A interpretação do modelo de regressão múltipla: o significado de “controlar” por outras variáveis

Tópicos

- Controle por variáveis e a cláusula *Ceteris Paribus*
- Regressões parciais
- Viéses de seleção e omissão de variáveis



Leitura Obrigatória:

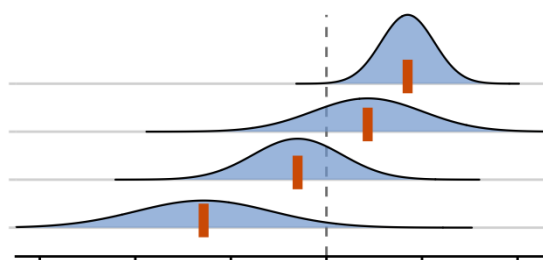
WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 3 – Multiple Regression Analysis: Estimation)

Parte 2: Incerteza, inferência e tópicos adicionais

Aula 3: Incerteza e inferência

Tópicos

- Revisão: o Teorema do Limite Central e a Lei dos Grandes Números
- Distribuição amostral dos coeficientes
- Teste de hipótese
- Derivação da matriz de variância-covariância dos coeficientes
- O pressuposto da homocedasticidade (ML.5)



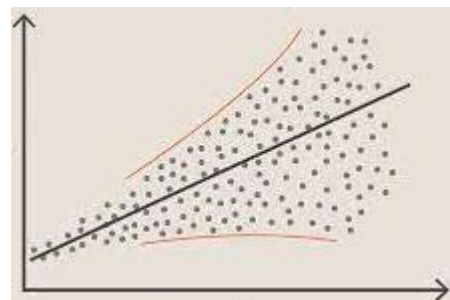
Leitura Obrigatória:

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 4 – Multiple Regression Analysis: Inference)

Aula 4: Heterocedasticidade

Tópicos

- O conceito de heterocedasticidade
- A necessidade dos supostos de homocedasticidade e normalidade dos erros em inferências com amostras pequenas (ML. 6)
- Heterocedasticidade como má especificação funcional
- Os Erros Robustos de Huber-White
- Erros clusterizados
- Implementando desenhos de Amostras Complexas



Leitura Obrigatória:

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 6 – Heterokedasticity)

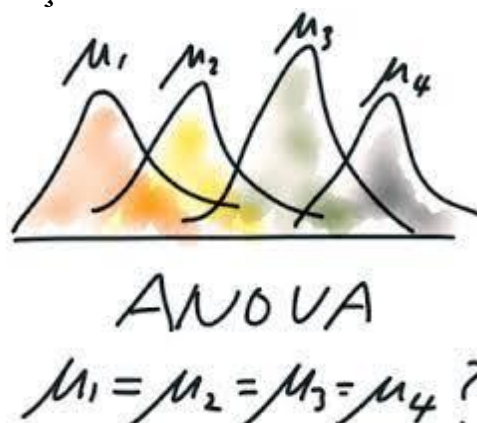
Leitura Optativa:

ABADIE, A; ATHEY, S.; IMBENS, G. W.; WOOLDRIDGE, J. “When should you adjust standard errors for clustering?” (No. w24003). National Bureau of Economic Research, 2017

Aula 5: Multicolinearidade, Análise de Variância e Seleção de Modelos

Tópicos

- O conceito de multicolinearidade
- Multicolinearidade como “micronumerosidade” (i.e. como um não problema em amostras grandes)
- Análise de variância: avaliando a diferença entre grupos
- A derivação do Teste F como teste de hipóteses conjuntas
- Teste de ajuste geral de um modelo de regressão
- Testes de inclusão/exclusão de variáveis e seleção de modelos



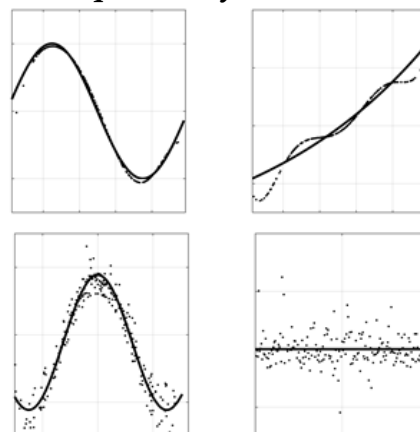
Leitura Obrigatória:

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 6 – Heterokedasticity)

Aula 6: Especificação funcional, valores preditos e *model-dependency*

Tópicos

- Quando usamos logaritmo? Modelos lin-lin, log-lin, lin-log e log-log
- Polinômios e Termos interativos
- Predições e *model dependency*
- Convex Hull e as possibilidades de construção de predições
- Bootstrap Paramétrico (Simulação de Monte Carlo) para cálculo da incerteza dos valores preditos



Leituras Obrigatórias:

BARBOSA, Rogério. "Incerteza e Valores Preditos". Material preparado para a Disciplina de Lego 2. *Mimeo*, 2022.

KING, G.; TOMZ, M.; WITTENBERG, J. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation". *American Journal of Political Science*, Vol. 44, No. 2, 2000.

Leitura Optativa:

King, G.; Zeng, L. "The Dangers of Extreme Counterfactuals". *Political Analysis* v.14, 2006.

Parte 3: Modelos Não-Lineares

Aula 7: Introdução a Modelos Lineares Generalizados

Tópicos

- Análise de tabelas de contingência
- Teste do Qui-Quadrado
- Modelos Lineares Generalizados para variáveis binárias
- Modelos Lineares Generalizados para variáveis de contagem
- Inferência estatística e comparação entre modelos

Leituras Obrigatórias:

- Agresti, A. *An Introduction to Categorical Data Analysis*. Wiley, 2007. 2nd edition. (Caps. 2 – Contingency Tables e 3 – Generalized Linear Models).
- Fernandes, A. A. T.; Figueiredo Filho, D. B.; Rocha, E. C.; Nascimento, W. S. "Leia este artigo se você quiser aprender regressão logística". *Revista de Sociologia e Política*, v. 28, n. 74, 2020.

Aula 8: Modelos para variáveis binárias: Logit, Probit e Modelo Linear de Probabilidade

Tópicos

- Problemas com modelos lineares de probabilidades
- Funções de ligação para logits e probits
- Interpretação dos coeficientes
- Testes de hipótese e ajuste dos modelos

Leituras Obrigatórias:

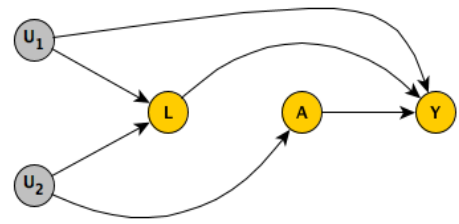
- Agresti, A. *An Introduction to Categorical Data Analysis*. Wiley, 2007. 2nd edition. (Caps. 4 – Logistic Regression e 5 – Building and Applying Logistic Regression Models).
- Breen, R.; Karlson, K.B.; Holm, A. Interpreting and understanding logits, probits, and other nonlinear probability models. *Annual Review of Sociology*, v. 44, p. 39-54, 2018.
- Long, J. S. *Regression Models for Categorical and Limited Dependent Variables*. Sage, 1997. (Capítulos 3. *Binary Outcomes: The Linear Probability, Probit, and Logit Models* e 4. *Hypothesis Testing and Goodness of Fit*)
- Mood, C. "Logistic regression: why we cannot do what we think we can do, and what we can do about it". *European Sociological Review*, v. 26, n.1, p. 67-82, 2010.

Parte 4: Inferência Causal

Aula 9: Grafos Acíclicos Dirigidos (Directed Acyclic Graphs – DAGs) - Parte 1

Tópicos

- A noção de causalidade como intervenção
- Causalidade probabilística: efeitos sobre a distribuição (não sobre indivíduos necessariamente)
- Diagramas Causais *vs.* Análises de Trajetórias
- Simplificando diagramas causais
- Mediação, confusão, colisão
- Caminhos causais
- Controlar por variáveis: uma nova perspectiva



Leituras Obrigatórias:

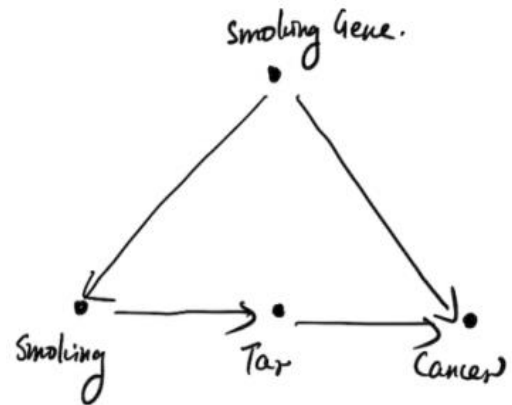
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022.

- Cap. 6. Causal Diagrams (<https://theeffectbook.net/ch-CausalDiagrams.html>)
- Cap. 7. Drawing Causal Diagrams (<https://theeffectbook.net/ch-DrawingCausalDiagrams.html>)
- Cap. 8. Causal Paths and closing back doors (<https://theeffectbook.net/ch-CausalPaths.html>)

Aula 10: Grafos Acíclicos Dirigidos (Directed Acyclic Graphs – DAGs) - Parte 2

Tópicos

- Abrindo e fechando portas com Colliders
- d-separation
- M-bias
- Implicações testáveis: independências condicionais
- Modelos alternativos para identificação
- *Frontdoor Adjustment*



Leituras Obrigatórias:

HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022.

- Cap. 9. Finding front doors. (<https://theeffectbook.net/ch-FindingFrontDoors.html>)

Aula 11: O Modelo dos Resultados Potenciais

Abordaremos, nesta aula, o famoso *Modelo dos Resultados Potenciais* (cuja elaboração levou, em 2021, a dois Prêmios Nobel).

Tópicos

- Causalidade como contrafactual?
- O problema fundamental da inferência causal
- Experimentos *vs.* Estudos Observacionais
- “Estimandos”: ATT, ATU e ATE



Leituras Obrigatórias:

CUNNINGHAM, Scott. *Causal Inference: the mixtape*. (Cap. 4: Potential Outcomes Causal Model. Disponível em: <https://mixtape.scunning.com/potential-outcomes.html>)

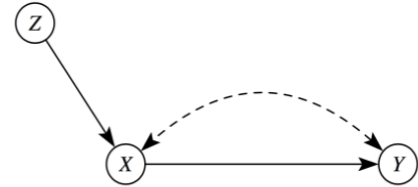
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL:

Chapman & Hall, 2022. (Cap. 10. Treatment Effects. Disponível em: <https://theeffectbook.net/ch-TreatmentEffects.html>).

Aula 12: Variáveis instrumentais

Tópicos

- A Regressão de Mínimos Quadrados em Dois Estágios (Two-Stage Least Squares)
- Instrumentos condicionais
- Instrumentos fracos
- Modelos sub-identificados, justamente identificados e sobre identificados
- O Teste de Sargan para exogeneidade
- A derivação do estimador de IV como LATE



Leitura Obrigatória:

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap.15 - Instrumental Variables Estimation and Two-Stage Least Squares, seções 15-1 a 15-6)

ANGRIST, Joshua; PISCHKE, Jörn-Steffen. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press, 2015. (Cap. 3 - Instrumental Variables)

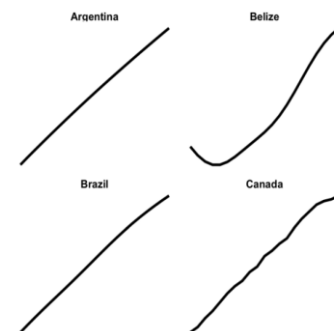
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022. (Cap. 11 - Instrumental Variables. <https://theeffectbook.net/ch-InstrumentalVariables.html>)

Parte 5: Dados em Painel

Aula 13: Uma introdução às estruturas de dados e regressões para painéis

Tópicos

- Estruturas de dados: wide e long
- Analisando e descrevendo dados temporais
- Modelo pooled



Leituras Obrigatórias:

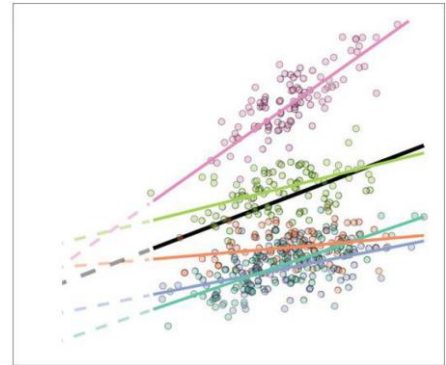
WOOLDRIDGE, J *Introductory Econometrics*. 5th Ed. (Cap.13 - Pooling Cross Sections across Time: Simple Panel Data Methods)

CUNNINGHAM, Scott. *Causal Inference: the mixtape*. (Cap. 8: Panel Data. <https://mixtape.scunning.com/panel-data.html>)

Aula 14: Especificações funcionais típicas de modelos de painel: efeitos fixos e aleatórios, variáveis com *lag* e diferenciadas

Tópicos

- A ideia de “efeitos individuais”
- Efeitos fixos e Efeitos aleatórios
- O teste de Hausman
- O estimador de primeiras diferenças
- O uso de variáveis diferenciadas
- O uso de variáveis com lag temporal



Leituras Obrigatórias:

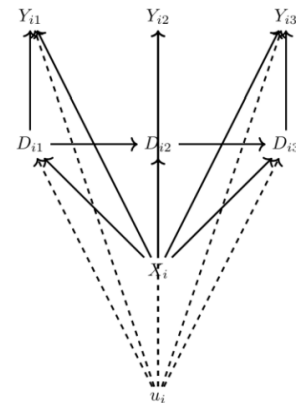
WOOLDRIDGE, J. Introductory Econometrics. 5th Ed. (Cap.14 - Advanced Panel Data Methods)

HUNTINGTON-KLEIN, Nick. The Effect: An Introduction to Research Design and Causality. Boca Raton, FL: Chapman & Hall, 2022. (Cap. 16 – Fixed Effects. <https://theeffectbook.net/ch-FixedEffects.html>)

Aula 15: Causalidade em modelos de painel

Tópicos

- Introdução aos modelos de Diferenças em Diferenças
- Modelos auto-regressivos e o risco da endogeneidade
- Variáveis instrumentais em modelos de painel



Leituras Obrigatórias:

WOOLDRIDGE, J. Introductory Econometrics. 5th Ed. (Cap.15 - Instrumental Variables Estimation and Two-Stage Least Squares, seção 15-8)

Kellstedt, Paul M.; Whitten, Guy D. The fundamentals of Political Science Research. 3rd Ed. Cambridge: Cambridge University Press, 2018. (Cap. 12 - Limited Dependent Variables and Time-Series Data, seção 12.3)