

Leg 2: Modelos de Regressão Linear e extensões

2º Semestre de 2022

Terças-feiras, de 9hs às 12hs.

Professor:

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

Monitor(a):

A definir

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**, discutimos inferência causal e apresentamos modelos e estratégias que permitem aferir causalidade a partir de dados observacionais. Na **Parte 4**, 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

Como em Lego 1, utilizaremos o software R, que é livre, gratuito e aberto. Não tema!

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

- Leitura dos textos obrigatórios (10 pontos)
- Listas de Exercício Quinzenais (3 listas, 20 pontos cada)
- Trabalho Final (30 pontos)

6. Horários das aulas e monitorias

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

Haverá monitoria uma vez por semana, em horário a ser divulgado.

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, nos dias 1, 3 e 5 (segunda, quarta e sexta). A frequência a esse curso é obrigatória. Haverá uma lista de exercício que renderá 10 pontos extras para a disciplina principal. Não se preocupe, não há pré-requisitos.

Acredite: relembrar ou aprender um pouquinho de Matemática Superior facilita muito o aprendizado e o uso de Métodos Quantitativos nas Ciências Sociais! Parece conversa fiada, mas não é! Nem “tecnicismo”. Além disso, o um dos objetivos de Lego 2 é treinar vocês para serem capazes de ler e aprender coisas novas de métodos por si próprios, sem precisar de um tutor constante e sem o risco de se tornarem meros “apertadores de botão” em softwares estatísticos.

Conteúdo do Curso de Nivelamento em Matemática para Lego 2:

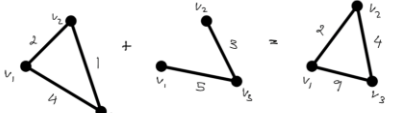
Dias 1, 3 e 5 de agosto, das 18hs às 21hs

Aula 1: Fundamentos de Álgebra Linear (Parte 1) - 1 de agosto

Matrizes encapsulam informações de modo muito eficiente. Resumem e simplificam notações, evitando longas equações. Ou seja: aprenderemos um pouquinho de Álgebra Linear para ter que fazer menos contas depois.

Tópicos:

- Vetores e Matrizes
- Produto escalar e ortogonalidade
- Matrizes Especiais: triangulares, diagonais, Identidade

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


Leitura Obrigatória:

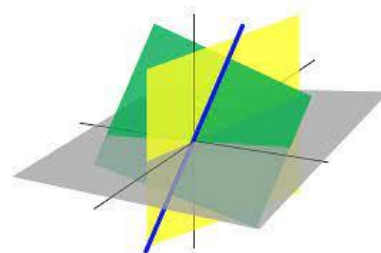
STRANG, Gilbert. *Introduction to Linear Algebra*. 5th Ed. Cap. 1.

Aula 2: Fundamentos de Álgebra Linear (Parte 2) - 3 de agosto

- Putz, sério!? Resolver sistemas?
- Sim! Mas rá! De uma forma muuuuito menos chata do que no Ensino Médio. Juro. E com aplicação prática iminente: aprender regressão para formalizar e testar teorias sociais. Não é caô.

Tópicos:

- Matrizes Inversas
- O método de Gauss-Jordan
- Sistemas de Equações Lineares



Leitura Obrigatória:

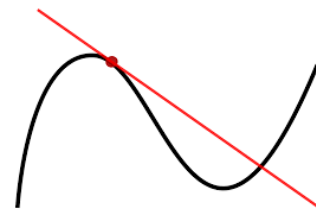
STRANG, Gilbert. *Introduction to Linear Algebra*. 5th Ed. Cap. 2.

Aula 3: Fundamentos de Cálculo - 5 de agosto

Ao infinito e além!

Tópicos:

- Funções
- Limites e Derivadas



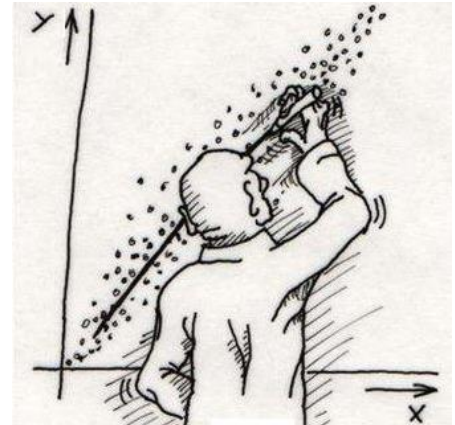
Parte 1: Regressão Linear, o básico

Aula 1: Regressão: visão geral e revisão

Na acepção mais simples, “Análise de Regressão” é ajustar uma reta ou equação aos seus dados – e, assim, diferenciar aspectos sistemáticos de idiosincrasias ou flutuações aleatórias. Nesta aula, faremos um voo panorâmico sobre o tema. Veremos aplicações em pesquisas, interpretaremos resultados, avaliaremos a adequação da técnica a contextos específicos.

Tópicos

- Correlação e regressão
- Sumarizando dados com regressões
- Interpretando coeficientes
- A qualidade do ajuste



Leituras Obrigatórias:

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>).

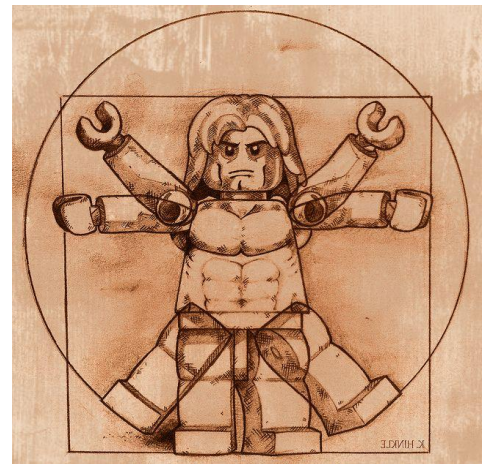
WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 2 – The Simple Regression Model)

Aula 2: Processos geradores de dados e os pressupostos teóricos da regressão

Que mecanismos sociais subjazem às informações que você analisa? Quais processos geraram os seus dados? Nesta aula, veremos que equações de regressão podem representar aspectos tanto de suas teorias sobre o mundo, como do próprio método de amostragem e obtenção dos dados. Examinaremos em detalhes os pressupostos necessários para esse empreendimento e como eles permitem até a derivação da fórmula/estimador da regressão!

Tópicos

- A ideia de um Processo Gerador de Dados
- Os pressupostos do modelo linear
- Derivando o estimador da regressão linear usando o Método dos Momentos

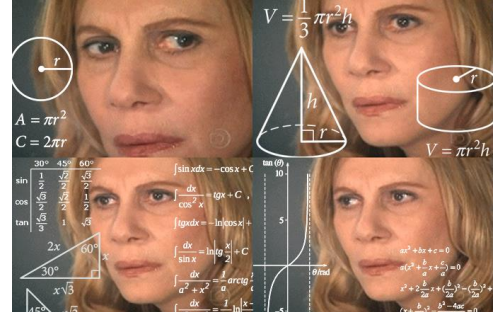


Leitura Obrigatória:

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>).

Aula 3: Regressão como “a melhor reta” (ou plano) que se ajusta aos dados: o método dos mínimos quadrados

Uma reta que melhor se ajusta aos dados não quer guerra com ninguém. Às vezes, tudo o que você precisa é resumir tendências e correlações de forma simples e eloquente – e então alguns dos pressupostos mais restritivos que conectam dados e teoria não precisam ser necessariamente mobilizados. A regressão, entendida como mera “reta (ou plano) que passa à menor distância dos pontos”, é ainda uma poderosa ferramenta de descrição e análise (ainda que não revele relações causais). Nesta aula, veremos o significado e a matemática por detrás dessa perspectiva.



Tópicos

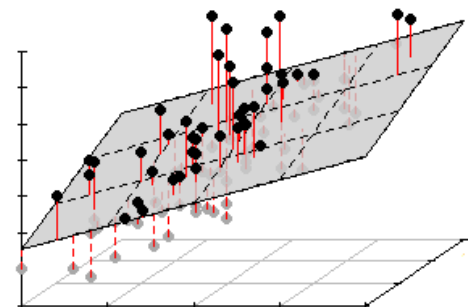
- Regressão como a reta/plano/hiperplano que passa à menor distância dos pontos
- Derivando o estimador da regressão usando o método dos mínimos quadrados
- O teorema de Gauss-Markov

Leituras Obrigatórias:

WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 3 – Multiple Regression Analysis: Estimation)
CUNNINGHAM, Scott. *Causal Inference: the mixtape*. (Cap. 2: Probability and regression). Disponível em: <https://mixtape.scunning.com/probability-and-regression.html>

Aula 4: A interpretação do modelo de regressão múltipla: o significado de “controlar” por outras variáveis

Qualquer fenômeno, social, biológico ou mesmo físico, é fruto de múltiplas determinações. Como podemos ter confiança acerca da importância de um fator explicativo particular? Como não incorrer em correlações espúrias? A resposta é: “controlar” as causas concorrentes ou alternativas! E a técnica mais conhecida para isso é a Regressão Múltipla. Nesta aula, aprenderemos como fazer isso e como selecionar a melhor especificação para seu modelo.



Tópicos

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

Leitura Obrigatória:

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

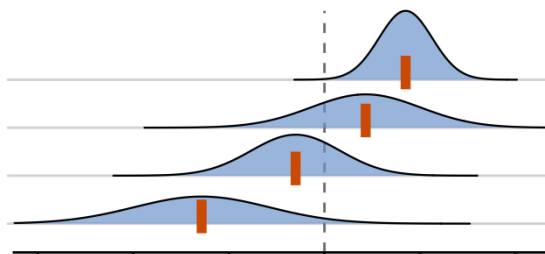
Parte 2: Incerteza, erros e inferência

Aula 5: Incerteza e inferência

Os fenômenos sociais não são determinísticos. Além disso, qualquer processo de seleção ou produção de dados é perpassado por erros, inconsistências e imprecisões. Como é possível lidar com a incerteza inerente à análise quantitativa? Nesta aula, veremos os fundamentos probabilísticos que permitem avaliar as margens de erro e a qualidade de nossa inferência.

Tópicos

- Revisão: o Teorema do Limite Central e a Lei dos Grandes Números
- Distribuição amostral dos coeficientes
- Teste de hipótese em amostras grandes e pequenas



Leitura Obrigatória:

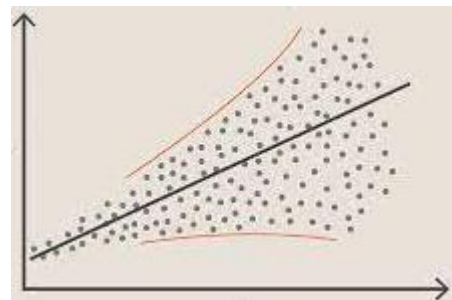
WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 4 – Multiple Regression Analysis: Inference)

Aula 6: Fontes adicionais de incerteza: heterocedasticidade, erros clusterizados e amostras complexas

Ok, a reta de regressão passa à menor distância dos pontos. Mas isso não significa que a qualidade das estimativas que produzimos é homogênea... A nuvem de pontos pode ficar cada vez mais dispersa (como no desenho ao lado), indicando que a qualidade da predição fica pior às vezes. Grupos de observações podem apresentar erros e desvios com respeito à reta sempre numa mesma direção. Sua amostra pode ter sido produzida por meio de um processo complexo, em múltiplas etapas. Para todos esses casos, há soluções. E é isso que veremos nesta aula.

Tópicos

- O conceito de heterocedasticidade
- Heterocedasticidade como má especificação funcional
- Observações correlacionadas entre si (clusterizadas)
- Os Erros Robustos de Huber-White, Erros Clusterizados
- Implementando desenhos de Amostras Complexas



Leitura Obrigatória:

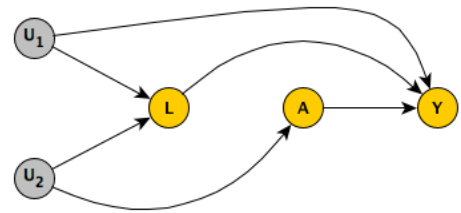
WOOLDRIDGE, J. *Introductory Econometrics*. 5th Ed. (Cap. 6 – Heterokedasticity)

Parte 3: Inferência Causal

Aula 7: A moderna linguagem causal, Parte 1: os Grafos Acíclicos Dirigidos

Nas Ciências Sociais, “causalidade” é quase um palavrão. Remete a determinismo, simplificação ou mesmo a ideais positivistas do século XIX. Mas uma literatura recente e extremamente sofisticada, que se desenvolveu nos últimos 40 anos, pensa esse tema de modo muito distinto: não-determinístico (probabilístico!), articulando-se com a complexidade e ancorada em discussões epistemológicas contemporâneas que nada lembram o velho Comte e suas aspirações por leis gerais. Nesta aula, daremos início às discussões sobre modernas abordagens causais em análises empíricas nas Ciências Sociais.

Os Grafos Acíclicos Dirigidos se parecem com um mero diagrama de setinhas, mas não são! São uma potente ferramenta para identificação causal. Têm a capacidade de simplificar e codificar complexas relações entre fenômenos, gerando implicações observáveis, fornecendo caminhos alternativos para inferência causal e lidando com a incerteza.



Tópicos

- Mediação, confusão, colisão
- “Back doors” e “Front doors”
- Implicações testáveis: independências condicionais
- Modelos alternativos para identificação

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

Aula 8: A moderna linguagem causal, Parte 2: 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 9: Variáveis instrumentais (Parte 1): conceitos básicos, a idéia de Local Average Treatment Effects (LATE), exemplos e estimação

Quando há muitos determinantes comuns entre o fenômeno a ser explicado e o fator explicativo que você deseja mobilizar, apenas controlar por variáveis não garante interpretação causal. O uso de variáveis instrumentais é uma estratégia que visa percorrer um caminho completamente diferente para contornar o problema. Passamos a pensar em efeitos que se desdobram em múltiplas etapas, em redes de determinações: Z que afeta X, que por sua vez causa Y.



Better LATE than never? Quem LATE não morde? Piadas ruins, à parte, a noção de *Local Average Treatment Effects* (Efeitos Locais Médios do Tratamento) é o que confere interpretação causal às variáveis instrumentais. E seu fundamento reside no Modelo dos Resultados Potenciais.

Tópicos

- A noção de “instrumento” e sua história
- O estimador de IV
- A derivação do estimador de IV como LATE

Leituras Obrigatórias:

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

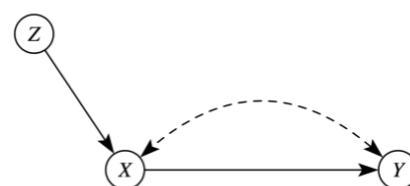
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022. (Cap. 11 – Instrumental Variables. <https://theeffectbook.net/ch-InstrumentalVariables.html>)

Aula 10: Variáveis instrumentais (Parte 2): e questões adicionais

Nesta aula, veremos sob que condições de fato podemos fazer uma leitura causal dos resultados de regressões com variáveis instrumentais. Veremos ainda o que fazer quando temos mais de uma variável instrumental (e quais as vantagens disso!).

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



Leitura Obrigatória:

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

Parte 4: Dados ao longo do tempo: modelos para painéis

Aula 11: Uma introdução às estruturas de dados e regressões para painéis

Às vezes, possuímos informações sobre os mesmos indivíduos/observações ao longo do tempo. E isso é incrível! Podemos tentar identificar o que há de fixo ou “constante” em cada um deles. Podemos identificar tendências gerais ou particulares a grupos. Podemos estar interessados nos níveis/patamares e/ou nas variações. Podemos modelar processos de retroalimentação causal. Podemos levar em conta expectativas e o impacto do futuro!

Dados em painel são aqueles em que temos um grande número de observações que se repetem no tempo, mas uma quantidade não tão grande de períodos. Hoje em dia, se tornaram muito mais comuns, pois se espalharam séries históricas de dados administrativos e pesquisas domiciliares que acompanham pessoas em várias rodadas. E há modelos de regressão específicos para isso.

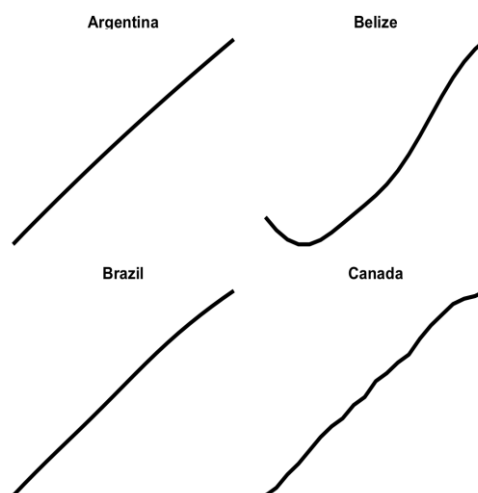
Tópicos

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

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>)

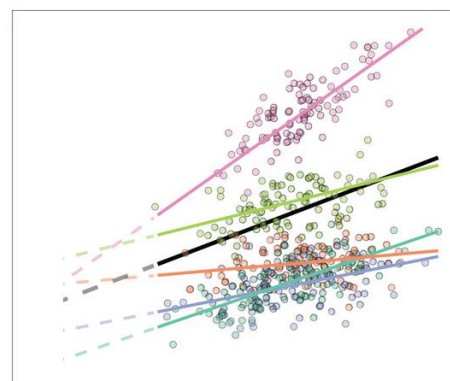


Aula 12: Efeitos fixos ou aleatórios!?

Há vários modelos para dados em painel. E eles possuem pressupostos distintos, vantagens e desvantagens. O impasse usual é escolher entre mais precisão ou menor viés.

Tópicos

- A ideia de “efeitos individuais”
- Efeitos fixos, várias formas de estimar
- Efeitos aleatórios
- O teste de Hausman



Leituras Obrigatórias:

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

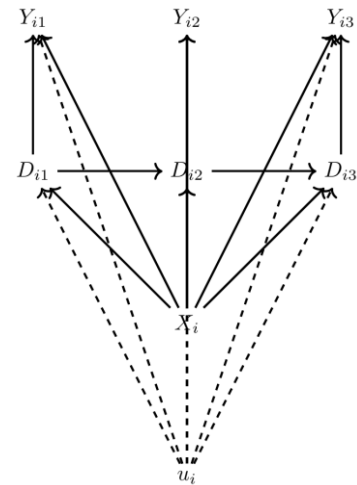
HUNTINGTON-KLEIN, Nick. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: Chapman & Hall, 2022. (Cap. 16 – Fixed Effects. <https://theeffectbook.net/ch-FixedEffects.html>)

Aula 13: Coisas que só um painel te dá: variáveis com lag temporal, variáveis diferenciadas e modelos auto-regressivos

O passado importa. O futuro também. Um evento do passado tem consequências no presente. O próprio fenômeno a ser explicado pode ter causado-se a si mesmo...! Como levar tudo isso em conta? É o que veremos nesta aula.

Tópicos

- O uso de variáveis com lag temporal
- O uso de variáveis diferenciadas
- O estimador de primeiras diferenças
- Modelos auto-regressivos e o risco da endogeneidade



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)

Parte 5: Encerramento da disciplina

Aula 14: Discussão sobre os trabalhos finais e encerramento da disciplina