



Instituto de Estudos Sociais e Políticos (IESP-UERJ)

Lego II - Modelo de Regressão Linear

Carlos Costa Ribeiro
carloscr@iesp.uerj.br

Thiago Moreira da Silva
thiagomoreira@iesp.uerj.br

Introdução

Este curso tem por objetivo fornecer um panorama do modelo de regressão linear e suas aplicações nas ciências sociais. Serão debatidos exemplos práticos, relacionados a temas atinentes à ciência política e à sociologia, bem como técnicas que autorizam o pesquisador a fazer previsões com bases amostrais. De início, os alunos terão contato com uma revisão de conhecimentos estatísticos básicos, como a probabilidade, o teste de hipóteses e os intervalos de confiança. Em seguida, forneceremos as bases teóricas para a compreensão da modelagem estatística e aprofundaremos o conteúdo em torno de mecanismos mais sofisticados para a análise da regressão linear.

Pré-requisitos

Alunos que tiveram algum contato anterior com noções básicas de estatística e dominam alguma técnica de programação, especialmente em R, aproveitarão melhor o curso. No entanto, essas condições não são necessárias aos interessados. Isso posto, nas primeiras aulas faremos uma revisão dos preceitos estatísticos mais básicos e, ao longo do curso, serão realizados laboratórios para facilitar os primeiros passos no R.

Software

Neste curso usaremos o software R, a ferramenta mais popular de análise de dados para cientistas sociais atualmente. O software pode ser baixado gratuitamente em <https://cran.r-project.org/bin/windows/base/>. Recomendamos que os alunos façam o download da interface gráfica Rstudio, que facilita a operacionalização do software. O download pode ser feito em <https://www.rstudio.com/products/rstudio/download/>.

Plano de curso

O curso terá como base aulas expositivas semanais. Em cada aula abordaremos a teoria e trabalharemos alguns exemplos no R. Além das aulas semanais, faremos laboratórios esporádicos para solucionar possíveis dúvidas em relação ao funcionamento do software.

Avaliação

Com o objetivo de fixar os conhecimentos transmitidos, nossa avaliação será composta por listas de exercícios, enviadas aos alunos quinzenalmente.

Leituras

No curso, usaremos os seguintes livros:

Angrist, Joshua D. e Pischke, Jorn-Steffen, 2014. *Mastering Metrics: The Path from Cause to Effects*. The Princeton University Press.

Fox, John (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. NY: Sage.

Fox, John (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

Gelman, Andrew e Hill, Jennifer. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gujarati, Damodar e Porter, Dawn (2009). *Basic Econometrics*. The McGraw-Hill.

Monogan, J. E. (2015). *Political analysis using R*. Springer.

Morgan, Stephen e Winship, Christopher (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. The Cambridge University Press.

Wheelan, Charles (2013). *Naked Statistics: Stripping the Dread from the Data*. W.W. Norton.

Wooldridge, Jeffrey M (2015). *Introductory Econometrics*. New York: South-Western. 5th edition.

Programas das Aulas

**O cronograma pode ser modificado ao longo do curso.*

- Aula 1 - Introdução
 - Apresentação do curso

Leitura obrigatória:

John Fox (2007): *Statistics and Social Sciences*, cap. 1.

- Aula 2 - Fundamentos da Probabilidade

- Distribuições de probabilidade
- Distribuições amostrais
- Famílias de distribuição
- Funções de distribuição acumulada

Leitura obrigatória:

Wooldridge, Apêndice B.

Agresti, caps. 6 e 7.

Leitura complementar:

Wheelan, caps. 2 e 3

- Aula 3 - Métodos de estimação

- População, amostra, modelo estatístico
- Estimação
- Propriedades do estimador
- Intervalos de confiança

Leitura obrigatória:

Wooldridge, Apêndice C.1 - C.3, C.5

Agresti, cap. 8.

Leitura complementar:

Wheelan, cap. 9

- Aula 4 - Testes de hipóteses

- Testes de hipóteses
- P-valor
- Análise de poder

Leitura obrigatória:

Wooldridge, apêndice C.6.

Agresti, cap. 9.

- Aula 5 - Regressão linear simples
 - Associação entre variáveis
 - Expectância condicional
 - Mínimos quadrados ordinários

Leitura obrigatória:

Wooldridge, cap. 1.

Agresti, cap. 12.

Leitura complementar:

Angrist e Pischke: capítulo 2

Wheelan, cap. 11.

- Aula 6 - Pressupostos básicos do modelos linear
 - Premissas do modelo de regressão linear
 - Distribuição amostral do estimador MQO
 - Variância amostral do estimador MQO
 - Propriedades do MQO para amostras grandes
 - Inferência para o MQO
 - Ajuste do modelo

Leitura obrigatória:

Wooldrige, capítulo 2.

Gujarati, caps. 3 e 4.

Leitura complementar:

Agresti, cap. 12

- Aula 7 - Regressão múltipla
 - Adição de variáveis categóricas (binárias e multinomiais)
 - Adição de variável contínua
 - Variável omitida e multicolinearidade

Leitura obrigatória:

Wooldridge, cap. 3

Fox, cap. 7, pp. 128-141.

Abramowitz, Alan. Forecasting the 2008 Presidential Election with the Time-for-Change Model. *PS: Political Science and Politics* 41: 691-695. URL: <http://goo.gl/DLckXv>.

Leitura complementar:

Gujarati and Porter (Essentials of Econometrics), pp. 102-104; 113-116, pp. 116-131, 160-163.

Wheelan, cap. 12 e 13.

- Aula 8 - Estimação com dados deficientes

- Multicolinearidade
- Heteroscedasticidade
- Erros de mensuração
- *Outliers* e observações influentes
- Imputação múltipla
- Diagnósticos

Leitura obrigatória:

Wooldridge, cap. 3. Fox (2015), caps.: 11, 12 e 13

Lall, Ranjit. "How multiple imputation makes a difference." *Political Analysis* 24.4 (2016): 414-433.

Arel-Bundock, Vincent, and Krzysztof J. Pelc. "When Can Multiple Imputation Improve Regression Estimates?." *Political Analysis* 26.2 (2018): 240-245.

- Aula 9 - Interações

- Interações entre variáveis
- Formas funcionais não-lineares

Leitura obrigatória:

Wooldridge, cap. 6.2, 8 e 9.5

Fox (2015), cap. 7, pp. 140-150.

- Aula 10 - ANOVA

- *One-way ANOVA*
- Estimção de diferenças entre grupos
- *Two-way ANOVA*

Leitura obrigatória:

Agresti, cap. 14.

Gujarati, cap. 9.

- Aula 11 - Variáveis instrumentais
 - Utilização de variáveis instrumentais
 - Endogeneidade

Leitura obrigatória:

Angrist and Pischke, Cap. 4.

Gujarati, cap. 9.

Gujarati and Porter, cap. 13, seção 13.5; Cap. 20.

Acemoglu, Johnson, and Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review* 91(5): 1369-1401. URL: <http://goo.gl/57CUYJ>.

- Aula 12 - Modelos em painel
 - Dados em painel
 - Métodos das primeiras diferenças
 - Efeitos fixos
 - *Clustering*

Leitura obrigatória:

Wooldridge, caps. 13 e 14.

Gujarati, cap. 16

- Aula 13 - Modelos generalizados
 - Regressão logística
 - Interpretação dos coeficientes
 - Diagnósticos e comparações

Leitura obrigatória:

Wooldridge, cap. 17.

Fox (2015), cap. 14