

# Instituto de Estudos Sociais e Políticos - IESP-UERJ

## Lego II - análise de dados quantitativos

Fernando Guarnieri  
fhguarnieri@iesp.uerj.br

2º semestre 2017

### 1 Ementa

O cientista político e o sociólogo, diferente do analista ou do “pitaqueiro”, estabelece relações entre fenômenos sociais com base em uma teoria, na coleta de dados que podem ajudar a testar as consequências desta teoria, na análise destes dados e na generalização destes dados para situações mais gerais. Se, por exemplo, queremos entender o efeito do Bolsa Família no voto em Lula em 2006 podemos fazer um survey onde perguntamos em quem a pessoa votou e se ela recebe ou não o benefício. Da análise destes dados podemos inferir se o PBF teve influência no voto da população como um todo.

Neste curso você aprenderá as ferramentas básicas usadas na pesquisa quantitativa. O curso foca especificamente em técnicas que permitem passar da análise de nossa amostra para conclusões mais gerais. É o que chamamos de inferência estatística. A primeira parte do curso introduz estatística univariada e o restante do curso será dedicado aos modelos de regressão linear.

O objetivo é fornecer os fundamentos para o estudo de tópicos mais avançados de metodologia quantitativa aplicada às Ciências Sociais. Nossa abordagem será teórica e computacional.

Um segundo objetivo do curso é dar o instrumental necessário para que você possa ler criticamente, interpretar e replicar artigos de colegas. Para isso, sempre que possível usaremos artigos para discutir os conceitos dados em aula.

### 2 Quem deve fazer esse curso

Alguns alunos se interessam pelo curso de regressão linear por que usarão este tipo de modelo, ou modelos mais avançados, em sua tese. Outros não fazem o curso justamente pelo motivo contrário: não usam métodos quantitativos. Mesmo se você não for usar métodos quantitativos em sua tese/dissertação o domínio dos fundamentos da inferência estatística é importante para seu desenvolvimento profissional. O mercado de trabalho para cientistas sociais cada vez mais exige este tipo de conhecimento.

### 3 Pré-requisitos

Para aproveitar ao máximo este curso é necessário ter noções introdutórias de análise de dados que são dadas em Lego I. Se você aprendeu cálculo e álgebra linear no ensino médio terá mais facilidade para compreender alguns conceitos. Se já souber usar o R, as coisas serão muito mais fáceis para você neste semestre.

### 4 Plano do curso

O curso terá como base aulas expositivas semanais. Em cada aula abordaremos a teoria, trabalharemos alguns exemplos e mostraremos como usar o R para lidar com algumas das técnicas apresentadas. Além da aula semanal faremos um laboratório por semana para resolver questões substantivas e problemas com o uso do R.

### 5 Avaliação

Aprender análise de dados quantitativos é como aprender uma nova língua. Para fixar o que aprendemos temos que nos exercitar. Portanto a avaliação dará ênfase aos exercícios. Toda semana teremos uma lista de exercícios. No meio do semestre faremos uma prova parcial. No final do semestre teremos uma prova final. A nota final será composta por:

- Exercícios semanais (40%)
- Prova parcial (10%)
- Prova Final (40%)
- Participação (10%)

Os exercícios deverão ser entregues no começo da aula. No final do mesmo dia enviaremos as respostas corretas, portanto não serão aceitos exercícios entregues fora do prazo. As respostas devem estar digitadas com tabelas e figuras incorporadas ao texto.

A prova parcial servirá para checar se todos os alunos entenderam os fundamentos da análise univariada para podermos passar para o modelo de regressão multivariada. Será aplicada quando concluirmos a primeira parte do curso (previsto para 11/10).

A prova final consistirá de análise de dados e interpretação. Será entregue na penúltima aula (13/12) e deverá ser devolvida na última aula (20/12).

### 6 Leituras

No curso usaremos o livro:

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

Outros livros que podem ser úteis:

· Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

· Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Rundel. 2015. Open-Intro Statistics. 3rd edition. <https://www.openintro.org/>

· Gelman, Andrew and Hill, Jennifer. 2007. Data Analysis Using Regression and Multi-level/Hierarchical Models. Cambridge University Press.

## 7 Software

Neste curso usaremos o programa R que pode ser baixado de graça em <http://www.r-project.org>. Sugiro fortemente que você use o R a partir do RStudio (<http://www.rstudio.com>). Para quem é completamente novo no R faça o curso online gratuito em <http://tryr.codeschool.com/>. Ajudaremos com dúvidas e problemas durante as seções de monitoramento.

## 8 Programa de Aulas

- Aula 1 (30/08) - Introdução
  - . Detalhes do curso
  - . Refrescando a memória
- Aula 2 (06/09) - Variáveis aleatórias e Distribuições de probabilidade
  - . Variáveis Aleatórias
  - . Distribuições de Probabilidade
  - . Funções de distribuição acumulada
  - . Descrevendo distribuições
  - . Famílias de distribuição
  - . Simulando variáveis aleatórias no R
- Aula 3 (13/09) - Múltiplas variáveis aleatórias
  - . Distribuições conjuntas e condicionais
  - . Covariância Correlação independência
- Aula 4 (20/09) - Médias e Teoremas do Limite
  - . Distribuição da média amostral
  - . Lei dos grandes números
  - . Teorema Central do Limite
- Aula 5 (04/10) - Estimação e Inferência Estatística
  - . População, amostra, modelo estatístico
  - . Estimação
  - . Propriedades do estimador
  - . Intervalos de confiança

- Aula 6 (11/10) - Teste de hipótese
  - . Teste de hipótese
  - . p-valor
  - . Análise de poder
- Aula 7 (18/10) - O que é regressão?
  - . Relação entre variáveis
  - . Expectância condicional
  - . Estimando a CEF
  - . CEFs lineares e projeções lineares
  - . Mínimos Quadrados
- Aula 8 (01/11) - Regressão linear simples
  - . Premissas da do Modelo de Regressão Linear
  - . Distribuição amostral do estimador MQO
  - . Variância amostral do estimador MQO
  - . Propriedades do MQO para amostras grandes
  - . Inferência para MQO
  - . Teste de hipótese e intervalo de confiança
  - . Ajuste do modelo
- Aula 9 (08/11) - Regressão com dois regressores
  - . Adicionando uma variável binária
  - . Adicionando uma variável contínua
  - . MQO com duas covariantes
  - . Premissas do MQO com duas covariantes
  - . Variável omitida e multicolinearidade
- Aula 10 (22/11) - Regressão Múltipla: usando Matrizes
  - . Álgebra matricial
  - . Operações com matrizes
  - . Modelo linear na forma de matrizes
  - . MQO com matrizes
  - . Inferência com Matrizes
- Aula 11 (29/11) - Regressão Múltipla: Interações, não linearidades, teste-F
  - . Interações entre variáveis
  - . Formas funcionais não lineares
  - . Testando múltiplas hipóteses

- Aula 12 (06/12) - Problemas, diagnóstico e soluções
  - . Ajuste do modelo, outliers e observações influentes
  - . Heteroscedasticidade
  - . Forma funcional
- Aula 13 (13/12) - Modelos em Painel
  - . Dados em Painel
  - . Métodos das primeiras diferenças
  - . Efeitos fixos
  - . Clustering
- Aula 14 (20/12) - Variável dependente binária
  - . Regressão logística
  - . Interpretando os coeficientes
  - . Avaliando checando e comparando regressões logísticas