

INSTITUTO DE ESTUDOS SOCIAIS E POLÍTICOS (IESP-UERJ)  
**Lego 1: Métodos Quantitativos para Ciências Sociais**

*1º Semestre de 2023*

**Rogério Jerônimo Barbosa**  
[rogerio.barbosa@iesp.uerj.br](mailto:rogerio.barbosa@iesp.uerj.br)

**Fernando Meireles**  
[meirelesff@hotmail.com](mailto:meirelesff@hotmail.com)

**Monitor(a):** *A definir*

## **1. Apresentação e objetivos**

Este curso é uma introdução às perspectivas, abordagens e ferramentas analíticas para Análise Quantitativa. Habilidades para analisar grandes volumes de dados utilizando Estatística e outras estratégias quantitativas são parte do arsenal básico da pesquisa em Ciências Sociais. Temos como objetivo tanto fornecer às alunas e alunos as capacidades necessárias para realizar suas pesquisas, como para avaliarem de forma crítica resultados de pesquisas realizadas por outros. Este curso tem um aspecto eminentemente prático. Não se trata apenas de um curso de “Introdução à Estatística”, mas também de programação para análise de dados (associada ao que hoje denominamos de *Data Science*) e de um espaço para discutirmos desenhos de pesquisa e aplicações substantivas.

O curso se inicia com uma discussão sobre **causalidade**. Abordamos as diferenças entre os estudos experimentais e observacionais, e indicamos de que forma a identificação causal se configura como um parâmetro para pensar pesquisas quantitativas de forma geral (mesmo quando o objetivo não é realizar inferência causal). Tratamos, em seguida, aspectos sobre a **mensuração** e a **previsão** de fenômenos sociais de forma quantitativa. Introduzimos então aspectos sobre **probabilidade** e **inferência estatística**, destacando como a incerteza e a imprecisão afetam nossas capacidades de produzir conclusões e abordando o ferramental para quantificá-las.

## **2. “Lego”: de onde vem esse nome?**

Lego 1 é a disciplina introdutória de Análise de Dados Quantitativos para cientistas sociais, tradicionalmente oferecida pelo nosso instituto, desde 1971. Antigamente denominada

“Metodologia I”, o curso sempre teve caráter obrigatório para as turmas de ingressantes na pós-graduação. E como vocês podem pela ementa abaixo, os objetivos mais gerais permaneceram os mesmos!

111 - Metodologia I: O curso visa proporcionar os conhecimentos básicos de análise de dados em ciências sociais, e especialmente uma certa familiaridade com o uso da computação eletrônica para este fim. Todos os alunos deverão ser capazes até o final do curso, de pedir e utilizar com proveito, trabalhos de computação, dentro dos recursos técnicos proporcionados pelo Instituto. - Nelson do Vale e Simon Schwartzman

*Ementa da primeira versão da disciplina “Metodologia I”, de 1971*

Com os trabalhos pioneiros de Nelson do Valle Silva e Carlos Hasenbalg, fomos aos poucos nos tornando referência na aplicação e no ensino de métodos e técnicas de pesquisa com uso de Estatística, estando sempre afinados com pesquisas e desenvolvimentos internacionais no Campo. Formamos várias gerações de pesquisadores de excelência. Conseqüentemente, pessoas de diversos institutos e universidades nos procuram todos os anos para cursarem como ouvintes ou alunos externos. Esse é o curso do IESP mais buscado por pessoas de fora da instituição!

O nome “Lego” não é uma sigla, mas uma brincadeira: uma referência às pequenas peças que, se montadas com engenhosidade e criatividade, podem construir qualquer coisa. Assim são também as estratégias de investigação social: se armamos um bom desenho de pesquisa, ele nos permite testar qualquer hipótese. Basta ter um bom domínio do ferramental metodológico. Apenas a disciplina Lego 1 é obrigatória. Mas depois de cursá-la, se desejarem prosseguir na formação em métodos, vocês podem se matricular em Lego 2 e Lego 3, disciplinas mais avançadas que ofertamos regularmente.

### **3. A quem é dirigido esse curso?**

Lego 1 é uma disciplina obrigatória para todas as alunas e alunos ingressantes nos cursos de mestrado em Sociologia e Ciência Política do Instituto de Estudos Sociais e Políticos (IESP) da Universidade do Estado do Rio de Janeiro (UERJ). A razão para ser obrigatória é porque entendemos que um quadro analítico e ferramentas técnicas de caráter quantitativo são

componentes básicos da formação profissional, acadêmica ou não, e mesmo para aqueles que não realizarão pesquisa empírica, que saberão então julgar a qualidade, a validade e a plausibilidade das suposições em trabalhos realizados por outros.

Os requisitos matemáticos prévios são baixos. Mas assumimos que todos os alunos que realizam essa disciplina como obrigatória tenham feito o curso de **Nivelamento**, ofertado na primeira semana do semestre letivo do IESP.

#### **4. Instrutores**

Os professores responsáveis são **Rogério J. Barbosa** e **Fernando Meireles**.

#### **5. Logística do Curso**

O curso é baseado em aulas expositivas, leituras e interação on-line, sessões de monitoria e realização de listas de exercício.

##### **5.1. Aulas**

###### **[Nivelamento]**

As aulas do **Nivelamento** serão ministradas **virtualmente**, por meio da plataforma **Zoom**. Serão 5 sessões com 3 horas de duração, entre os dias 27 de fevereiro e 3 de março de 2023. Recomendamos que você assista às aulas utilizando um **computador**. Telas de celular não possuem o tamanho adequado. Você deverá baixar o aplicativo em: <https://zoom.us/client/latest/ZoomInstaller.exe> e depois criar uma conta de usuário (caso ainda não tenha). O link permanente para todas as cinco aulas do Nivelamento é:

<https://us06web.zoom.us/j/85138409578>

###### **[Lego]**

As aulas do curso principal (Lego 1) serão ministradas **presencialmente**. Serão 15 sessões (uma por semana), com 3 horas de duração.

##### **5.2. Monitorias**

As sessões de monitoria serão também realizadas **presencialmente**. Serão 15 sessões (uma por semana), com 1 hora de duração, realizadas em dia distinto da aula. A frequência à

monitoria é obrigatória e conta como presença com o mesmo peso das aulas.

A monitora pode estruturar cada sessão de monitoria conforme julgar mais adequado: ora apenas sanando dúvidas de alunos, ora revisando e apresentando conteúdos de forma expositiva (se julgar, por exemplo, que os alunos possuem dúvidas semelhantes, que podem ser endereçadas conjuntamente).

## 6. Softwares utilizados: R e RStudio

Análise de dados profissionais são feitas em programas de computador especializados, feitos para Estatística. O Microsoft Excel NÃO é um desses softwares! Há uma variedade grande. Os mais conhecidos são o SPSS, Stata, SAS e o R. Nós trabalharemos com esse último. O software R é gratuito e de código aberto. Nos últimos anos, se tornou a principal plataforma para Estatística Aplicada, aplicação de modelos de Inteligência Artificial, coleta automatizada de dados (*web scraping*), elaboração de gráficos profissionais (no crescente campo da Visualização de Dados) – enfim, para a realização das atividades e tarefas que se convencionou denominar Data Science. Nossa opção pelo R justifica-se, assim, não apenas por sua licença de livre acesso (uma razão política, que visa a democratização do acesso), como também por estar alinhado aos desenvolvimentos mais recentes do campo da Metodologia das Ciências Sociais e às formas de análise de dados também em outros campos do conhecimento.

O R, porém, não é apenas um software, mas também uma linguagem de programação. Isso significa que não se trata apenas de aprender e executar “comandos”, como usualmente se costuma pensar acerca de outros softwares estatísticos. Trata-se da aquisição de uma gramática e forma de estruturação de pensamento (e expressão). Seu aprendizado, como o de qualquer outra linguagem, depende de prática e exposição.

Uma introdução ao R será ofertada como curso de **Nivelamento**. Veja esse vídeo-tutorial abaixo com instruções sobre como instalar o R e o RStudio:

[https://www.youtube.com/watch?v=cfwn\\_aJ2o7s](https://www.youtube.com/watch?v=cfwn_aJ2o7s)

Você precisa chegar à primeira aula do Nivelamento já com os 2 softwares instalados.

## **7. Google Classroom**

Utilizaremos o mural do **Google Classroom** para nos comunicarmos (substituindo, por exemplo, um grupo de emails). Listas de exercício e outros trabalhos deverão ser submetidos por lá.

## **8. Outros Materiais**

Os slides, as listas de exercício e materiais extras serão disponibilizados em uma pasta online.

## **9. Atividades e Avaliação**

Os requisitos do curso consistem nos seguintes componentes:

- **1 Lista de exercício do Nivelamento** (10 pontos)
- **4 Listas de exercício de Lego** (60 pontos: cada lista vale 15 pontos)
- **Trabalho Final** (30 pontos)

# Nivelamento

## Estatística e R: primeiros passos

### Aula 1

- **Análise Exploratória de Dados: uma visão geral**
- **R**
  - Abrindo bancos de dados com o R
  - Visão geral sobre transformação e manuseio de dados
  - Uma introdução ao *dplyr*

*Leituras Obrigatórias:*

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #1*.

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #2*.

*Leituras Optativas:*

PENG, Roger D. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 2: “History and Overview of R”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/history-and-overview-of-r.html>

\_\_\_\_\_. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 3: “Getting started with R”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/getting-started-with-r.html>

### Aula 2

- **Estatística:**
  - Quantificando informações
  - Frequências, tabelas etc.
- **R:**
  - De volta ao “básico do básico”:
    - Operações matemáticas
    - Vetores de classe unitária
    - *Subsetting*
  - Operadores lógicos e relacionais
  - Frequências e gráficos com R-Base

*Leituras Obrigatórias:*

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #3*.

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #4*.

*Leituras Optativas:*

PENG, Roger D. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 4: “R Nuts and Bolts”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/r-nuts-and-bolts.html>

\_\_\_\_\_. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 9: “Subsetting R Objects”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/subsetting-r-objects.html>

### Aula 3

- **Estatística:**
  - Medidas de tendência central
  - Medidas de dispersão e posição
- **R:**
  - Abrindo bancos de dados em vários formatos
  - Estatísticas descritivas com o R Base e *dplyr*

#### *Leituras Obrigatórias:*

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #5*.

BARONE, Leonardo. Introdução à programação e ciência de dados para a gestão pública: *Tutorial #6*.

#### *Leitura Optativa*

PENG, Roger D. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 12: “Managing Data Frames with the *dplyr* package”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/managing-data-frames-with-the-dplyr-package.html>

WICKHAM, Hadley; GROLEMUND, Garrett. *R for Data Science: visualize, model, transform, tidy and import data*. O’Reilly, 2007. **Capítulo 5: Data Transformation**. Disponível em: <https://r4ds.had.co.nz/transform.html>

### Aula 4

- **Estatística:**
  - Representações Gráficas
- **R:**
  - Análise Exploratória de dados
  - Introdução ao *ggplot2* (I)

#### *Leitura Obrigatória:*

WICKHAM, Hadley. *ggplot2: Elegant graphics for data analysis*. 3<sup>rd</sup> Ed (on-line work-in-progress version). Disponível em: <https://ggplot2-book.org/getting-started.html>

WICKHAM, Hadley; GROLEMUND, Garrett. *R for Data Science: visualize, model, transform, tidy and import data*. O’Reilly, 2007. **Capítulo 7: Exploratory Data Analysis**. Disponível em: <https://r4ds.had.co.nz/exploratory-data-analysis.html>

### Aula 5

- **Estatística:**
  - Noções de associação estatística
- **R:**
  - De volta ao R-Base:
    - Organizando o script
    - Compreendendo mensagens de erro
    - Revendo tipos de objetos
    - Subsetting de vetores e data-frames
    - Trabalhando com valores missing

# Legu 1

## Parte 1: Introdução

### Aula 1: Fundamentos da Pesquisa Quantitativa

#### *Leituras Obrigatórias:*

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 1: “The Scientific Study of Politics”)

\_\_\_\_\_. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 2: “The art of Theory Building”)

\_\_\_\_\_. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 3: “Evaluating Causal Relationships”)

#### *Leituras Optativas:*

KING, Gary; KEOHANE, Robert; VERBA, Sidney. *Designing Social Inquiry*. Princeton: Princeton University Press, 1994. (Cap. 1: “The Science in Social Science”)

### Aula 2: Desenho de Pesquisa e Mensuração

#### *Leituras Obrigatórias:*

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 4: “Research Design”)

\_\_\_\_\_. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 5: “Measuring Concepts of Interest”)

#### *Leituras Optativas:*

BABBIE, Earl. *The Practice of Social Research*. 15<sup>th</sup> Ed. Boston: Cengage, 2021. (Cap. 4: “Research Design”)

## Parte 2: Descrevendo dados, manuseando bancos

### Aula 3: Estatística Descritiva

#### *Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 3: “Descriptive Statistics”)



KELLSTEDT, Paul; Whitten, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 6: “Getting to know your data”)

## **Aula 4: Análise de dados e programação (I): estruturas de fluxo e controle**

*Leituras Obrigatórias:*

PENG, Roger D. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 13: “Control structures”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/control-structures.html>

\_\_\_\_\_. *R programming for data science*. Victoria, BC, Canada: Leanpub, 2016. (Cap. 16: “Loop Functions”). Disponível em: <https://bookdown.org/rdpeng/rprogdatascience/loop-functions.html>

## **Aula 5: Análise de dados e programação (II): *data wrangling* com *tidyverse***

*Leituras Obrigatórias:*

WICKHAM, Hadley; GROLEMUND, Garrett. *R for Data Science: visualize, model, transform, tidy and import data*. O’Reilly, 2007. **Capítulo 12: Tidy Data**. Disponível em: <https://r4ds.had.co.nz/tidy-data.html>

## **Parte 3: Probabilidade e Incerteza**

### **Aula 6: Amostragem e desenhos de coleta de dados**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 2: “Sampling and Measurement”)

BABBIE, Earl. *The Practice of Social Research*. 15<sup>th</sup> Ed. Boston: Cengage, 2021. (Cap. 7: “The Logic of Sampling”)

### **Aula 7: Distribuições de Probabilidade**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 4: “Probability Distributions”)

### **Aula 8: Inferência Estatística (I): Estimação**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 5: “Statistical

Inference: Estimation”)

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 7: “Probability and Statistical Inference”)

## **Aula 9: Inferência Estatística (II): Testes de Significância**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 6: “Statistical Inference: Significance Tests”)

## **Parte 4: Associação, Correlação e Regressão**

### **Aula 10: Comparando Grupos: quando as diferenças estatisticamente significantes e substantivamente significantes?**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 7: “Comparison of Two Groups”)

KELLSTEDT, Paul; Whitten, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 8: “Bivariate Hypothesis Testing”)

### **Aula 11: Associação de variáveis categóricas**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 8: “Analyzing Association Between Categorical Variables”)

### **Aula 12: Regressão Linear e Correlação**

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 9: “Linear Regression and Correlation”)

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 9: “Two-Variable Regression Models”)

## Parte 5: Relações Multivariadas

### Aula 13: Relações Multivariadas

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 10: “Introduction to Multivariate Relationships”)

### Aula 14: Uma introdução à Regressão Múltipla

*Leituras Obrigatórias:*

AGRESTI, Alan. *Statistical Methods for the Social Sciences*. 5<sup>th</sup> Ed. Boston: Pearson, 2018. (Cap. 11: “Multiple regression and Correlation”)

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 10: “Multiple Regression: the basics”)

### Aula 15: Especificando modelos de Regressão Múltipla

*Leituras Obrigatórias:*

KELLSTEDT, Paul; WHITTEN, Guy. *The Fundamentals of Political Science Research*. 3<sup>rd</sup> Ed. Cambridge: Cambridge University Press, 2018. (Cap. 11: “Multiple Regression Model Specification”)