

Lego 1: Introdução aos Métodos Quantitativos e *Data Science* para Ciências Sociais

1º Semestre de 2021

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

Carlos Antônio Costa Ribeiro
carloscr@iesp.uerj.br

Monitor: Matheus Cavalcanti Pestana

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. Em todas as aulas abordaremos pesquisas empíricas exemplares que fizeram uso de técnicas abordadas no programa do curso e discutiremos seus resultados.

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). Abordamos, 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. 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 os cursos de **Nivelamento**, ofertados no IESP nas semanas que antecedem o início do semestre letivo da UERJ.

3. Logística do Curso

O curso é baseado em aulas expositivas, sessões de monitoria e realização de listas de exercício semanais.

3.1. Aulas

As aulas serão ministradas virtualmente, por meio da plataforma Google Meet. Serão 15 sessões (uma por semana), com 3 horas de duração. Cada aula é composta das seguintes partes:

- 2 horas de exposição do professor, baseada em conteúdos do livro-texto (ver mais abaixo).
- 10 minutos de seminário realizado por aluno, com base na leitura obrigatória da aula. Por isso, a pessoa expositora do seminário do dia será sorteada aleatoriamente para essa atividade durante a aula. Por esta razão, espera-se que todos os alunos tenham realizado a leitura e estejam igualmente capacitados para realizar a exposição.
- 50 minutos de discussões.

3.2. Monitorias

As sessões de monitoria serão também realizadas virtualmente, por meio da plataforma *Google Meet*. 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.

O monitor 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).

4. Atividades e Avaliação

Os requisitos do curso consistem nos seguintes quatro componentes:

- **Participação (10%):** Espera-se que todos os alunos participem ativamente das aulas – fazendo comentários ou trazendo questões e dúvidas pelo menos uma vez por sessão.

- **Seminários sobre as leituras obrigatórias (20%):** Em todas as sessões de aula do curso, uma aluna ou aluno será sorteada aleatoriamente para realizar uma exposição verbal (não é necessário uso de slides ou outros instrumentos auxiliares para apresentação) sobre a leitura obrigatória do dia. Assume-se que todos os alunos tenham realizado a leitura e estejam igualmente capacitados para apresentar. O sorteio será realizado com reposição – o que implica que uma mesma pessoa pode ser sorteada mais de uma vez e que, inclusive, existe a probabilidade de que venha a apresentar em aulas consecutivas. A pessoa sorteada terá o direito de passar seu turno/vez em uma única ocasião. Nesse caso, um novo sorteio é feito imediatamente e a nova pessoa designada não poderá passar sua vez. Existe a probabilidade de algumas pessoas não serem sorteada em nenhuma das aulas. Nesse caso, ao fim do semestre, os não sorteados receberão a totalidade dos pontos de forma automática.
- **Listas de exercício sobre R (20%):** Haverá listas de exercício semanais. Elas não serão corrigidas, mas as entregas devem ser feitas no prazo. Alunas e alunos são estimulados a trabalhar em conjunto e colaborar uns com os outros, mas as listas são individuais.
- **Listas de exercício sobre estatística e temas substantivos (20%):** Haverá quatro listas de exercício ao longo do semestre. Elas devem ser entregues no prazo e serão corrigidas e pontuadas. Colaboração entre alunos é estimulada, mas as entregas devem ser individuais.
- **Provas realizadas em casa (30%):** Haverá duas provas: uma no meio do semestre e outra no fim. Elas serão entregues com um dias de antecedência e devem ser entregues em dia e horário definidos.

5. Livro-texto e leituras obrigatórias

Este curso usa um livro-texto como guia para os conteúdos fundamentais

das aulas:

Imai, Kosuke. Quantitative social science: An introduction. Princeton University Press, 2018.

A cada aula, os alunos devem realizar também a leitura de um texto obrigatório. A lista completa de leituras será apresentada na versão final do Programa de Curso.

6. Softwares utilizados: R e RStudio

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** duas semanas antes do início de Lego 1. A frequência ao nivelamento é obrigatória para todos os alunos regulares do IESP.

7. Plano de aulas

Parte 1: Causalidade

Semana 1: Causalidade: uma visão geral, epistemológica e operacional

Semana 2: Experimentos em Ciências Sociais

- Livro-texto: Capítulo 2 (Seções 2.1-2.4)

Semana 3: Inferência causal em estudos observacionais

- Livro-texto: Capítulo 2 (Seções 2.5-2.7)

Parte 2: Mensuração

Semana 4: Seleção de casos e amostragem: traçando inferências a partir de um número limitado de casos

- Livro-texto: Capítulo 3 (Seções 3.1-3.4)

Semana 5: Mensuração de variáveis latentes: como podemos lidar empiricamente com aspectos não diretamente observáveis?

- Livro-texto: Capítulo 3 (Seções 3.5-3.7)

Parte 3: Predição

Semana 6: Usando estatística para previsão de resultados

- Livro-texto: Capítulo 4 (Seção 4.1)

Semana 7: Modelos de Regressão para previsão

- Livro-texto: Capítulo 4 (Seções 4.2 e 4.3)

Semana 8: Prevendo um resultado que não aconteceu (ou sobre como modelos de Regressão podem ser usados para Inferência Causal, sob determinados pressupostos)

Parte 4: Probabilidade

Semana 9: Probabilidade e probabilidade condicional

- Livro-texto: Capítulo 6 (Seções 6.1-6.3)

Semana 10: Distribuições de probabilidade

- Livro-texto: Capítulo 6 (Seções 6.4-6.5)

Semana 11: Simulação de processos geradores: uma concepção probabilística dos fenômenos sociais

Parte 5: Incerteza e Inferência Estatística

Semana 12: Estimativas e testes (estatísticos!) de hipótese

- Livro-texto: Capítulo 7 (Seções 7.1-7.2)

Semana 13: Regressão com incerteza (I)

- Livro-texto: Capítulo 7 (Seção 7.3)

Semana 14: Regressão com incerteza (II)

- Livro-texto: Capítulo 7 (Seção 7.3)

Parte 6: Fechamento do Curso

Semana 15: O que vem depois de Lego 1? *Big Data*, *Data Science* e os avanços técnicos recentes da pesquisa quantitativa