

Todo estudante de Economia deveria aprender R e por onde começar

Pedro Cavalcante

2018-12-21

Em março do ano que vem vou dar um “curso” de R na faculdade. Uma imersão rápida de uma semana nesse lindo mundo da análise de dados. Estava montando algum material para as “aulas”, procurando motivações razoáveis para que meus colegas *queiram* perder uma semana de férias programando.

Nessa breve meditação eu concluí algumas coisas e vou organizar a mente sobre elas aqui. Depois, nada mais justo que indicar para quem não teve a chance de aprender essa maravilhosa ferramenta ainda o caminho das pedras de por onde começar, o que fazer, o que esperar e esse tipo de coisa.

Ah, a realidade tão confusa, os modelos tão errados

Me ocorreu que antes de pensar por que um estudante de economia deveria aprender uma linguagem de programação voltada para análise de dados, eu deveria antes convence-lo que matemática é importante. “Matemática”, leia, algo além de dois cursos mastigados de cálculo e um de álgebra linear. Tô falando de análise, otimização e topologia.

Vamos fazer um pequeno exercício mental. Qual é a diferença entre os dois seguintes enunciados:

- Pessoas escolhem o que consomem levando em conta suas preferências e suas restrições orçamentárias

$$\text{Max } U(x, y) \text{ s.t. } P_x x + P_y y = W$$

Obviamente os dois enunciados carregam a mesma informação central. Por quê é mais interessante escrever da segunda forma? Veja bem, leitor, certamente existe quem escreva mal com matemática, mas ninguém escreve matemática mal. Equações, funções, problemas de otimização são exatamente equações, funções e problemas de otimização.

Enquanto isso posso escrever a mesma ideia central como:

- Pessoas escolhem o melhor que podem
- Consumidores dotados de faculdades cognitivas apropriadas realizam escolhas válidas para o contexto tomando em consideração preferências pessoais e restrições externas de ordem orçamentária

Eu não sei você, mas me sinto quase ofendido quando leio essas enrolações. Disso para concluir algo que depende de combinação muito específica de parâmetros como o comum é um pulo. Escrever os modelos com matemática é mais transparente, deixa clara nossa ignorância. Resultados contraintuitivos aparecem mais facilmente e induzir o leitor a concluir algo sem apresentar dados apropriados se torna um exercício de futilidade: “acredite em mim que esse parâmetro aqui é mais relevante que esse outro por favor”.

E onde entra análise e topologia aqui?

Bem, análise é onde você entra no jogo dos crescidinhos. É, em termos bem amplos, a área da matemática que fundamenta nossos conceitos usuais como funções, limites, derivadas. O jogo deixa de ser fazer conta para provar teoremas e propriedades. Isso é importante porque ver matemática (e por consequência nossos modelinhos) sob esse novo prisma te faz ver e se preocupar com coisas novas. A primeira reação de quem teve algum contato com isso quando apresentado a um resultado é “você consegue provar isso?”. Ter esse tipo de raciocínio fará mil vezes mais pelo seu pensamento crítico do que ler na fonte gente que morreu no século XIX.

A segunda reação é, tendo mais ferramentas para ponderar o quão central é uma hipótese em um modelo, questionar o papel delas nessa conclusão. No melhor dos cenários, um modelo é uma coleção de hipóteses aceitavelmente irrealistas. É, ninguém computa Multiplicadores de Lagrange no supermercado, mas quando tomates ficam mais caros pessoas compram menos. É absurdo fingir que pessoas maximizam utilidade, têm preferências racionais e outros tantos flagrantes irrealismos? Não. Existência de Equilíbrio Geral não depende de competição perfeita em todos os mercados, mas alunos de graduação são assim introduzidos a esse mundo porque é, veja você, mais simples.

Um modelo é, por definição, algo errado. Subtraímos do mundo à nossa volta peculiaridades em busca de um objeto mais tratável, matemático. Funciona? Creio que sim. Você, leitor, pode ler minhas palavras diretamente do telefone porque os teoremas de Teoria dos Leilões provados por Vickrey, Milgrom e Wilson nos anos 70-80 fundamentaram os bem-sucedidos leilões de banda larga nos anos 90. Se um parente seu precisar de doação de rim, provavelmente se beneficiará do Algoritmo de Gale-Shapley e para mercados com pareamentos e dos teoremas (de impossibilidade) demonstrados por Alvin Roth. Sim, Shapley e Roth, aqueles laureados com o Nobel de Economia. A obra de outros dois laureados com o Nobel, Roger Myerson e Bengt Holmstrom envolve muitos teoremas e tem aplicações práticas na vida de todos nós. Como um organizador determina as remunerações de uma equipe que não pode diretamente observar? Como regulamos um monopolista se não conhecemos sua curva de custo?

Matemática é importante não só para provar teoremas difíceis e levar o Nobel para casa. Eu te garanto que entender por que “*deriva e iguala a zero*” funciona fica mais fácil. Entender por que um jogo tem Equilíbrio de Nash deixa de ser questão de achar interseções de curvas e sim de entender que isso é consequência do Teorema do Ponto Fixo de Brouwer, ou mais, que consumidores têm uma cesta ótima graças ao Teorema de Bolzano-Weierstrass.

Ok, seu maluco, e o R?

Como já dizia o revolucionário, o mérito da teoria está na prática. Jogar modelos malucos no quadro é uma coisa, *eles serem bons explicadores da realidade é outra completamente diferente*. Aqui entra o R. Uma coisa que me incomoda no ambiente universitário brasileiro de economia é que nós *paramos no tempo*. Nos anos 50 eu chuto, mas talvez no século XIX.

Quais são os temas mais comuns de serem debatidos aqui na Banânia?

- Heterodoxia vs Ortodoxia
- Poupança causa Investimento ou o contrário?
- Valor-utilidade ou valor-trabalho?

Eu vou listar alguns artigos com menos de 18 anos de idade para você, leitor, ter uma ideia do que se debate lá fora, nas melhores revistas de economia:

- Incentivizing Safer Sexual Behavior: Evidence from a Lottery Experiment on HIV Prevention, de Martina Nyqvist, Lucia Corno, Damien de Walque e Jakob Svensson

Os caras montaram uma loteria para pacientes de HIV no Lesoto. A condição para ficar era participar da profilaxia. Os resultados: diminuição enorme no contágio a um custo pequeno. Saiu numa das melhores revistas de economia aplicada do mundo.

- Do Political Protests Matter? Evidence from the Tea Party Movement, de Andreas Madestam, Daniel Shoag, Stan Veuger e David Yanagizawa-Drott

Exploraram o fato de que chuva alterou a atendência a protestos coordenados nacionalmente nos EUA em cada região e depois cruzou isso com candidatos eleitos nas próximas eleições e filiações no movimento. Um trabalho realmente interessante, saiu na *melhor* revista de economia do mundo, o Quarterly Journal of Economics.

- The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth, de Justin McCrary e Heather Royer

Quem nasce logo antes de uma certa data precisa esperar um ano para entrar na escola e isso causa uma variação exógena na escolaridade de adultos que não concluíram o ensino médio. Os autores então cruzam isso com indicadores de saúde dos filhos desses pais menos escolarizados. Saiu numa das melhores revistas de economia do mundo.

Esses trabalhos são só três exemplos do que se faz hoje em dia em economia. Tem pouco ou nada a ver com o esoterismo instalado em boa parte dos departamentos de economia nacionais e sim com tratar, explorar e analisar dados. Não só isso, mas tendo cuidado e usando técnicas para procurar relações de causalidade e não só de correlação. Falo mais disso nesse post aqui se o assunto te interessar.

Aprender R é uma porta de entrada para esse mundo, o da pesquisa séria. Nesse post aqui eu mostrei como replicar um paper muito interessante sobre nepotismo usando dados ingleses do século XIX com o R. Desde importar os dados que o autor usou até rodar os modelos.

E por onde começar?

Antes de mais nada, programar na vida real não é como nos filmes de hackers. Você não vai ficar digitando sem parar por horas a fio olhando para uma tela preta e branca. A maior parte do tempo você vai passar no StackOverflow e no CrossValidated (ambos parte da rede StackExchange de fóruns) tirando dúvidas e procurando gente com o mesmo problema que você. Programar é assim, não se sinta mal com mensagens de erro.

Aqui segue uma lista não-exaustiva de fontes úteis para começar:

- Datacamp

Eles tem um curso introdutório gratuito de R e um de python. As aulas são curtas, cheias de exercícios com dados engraçadinhos (tipo bilheteria de Star Wars). O resto é pago e vale cada centavo. Existem mais de 100 cursos aí que duram entre 2 e 6 horas cada e são todos de altíssima qualidade. O maior ponto a favor é que você aprende fazendo.

- Análise Macro

Cursos pagos, mas muito bons. Você vai receber uma apostila de alta qualidade, videoaulas gravadas, 3 meses para acessar o material e mergulhar fundo em um tema específico. Desde análise de séries temporais a teoria macroeconômica, tudo com a mão na massa.

- R for Data Science

E-book gratuito de dois dos papas de R. Aqui tem praticamente tudo que você vai precisar saber na vida sobre R e de graça. Atenção: não só leia o livro, vá programando tudo que eles apresentam na hora. É fazendo que se aprende.

- R para Economistas, de Daniel Coutinho

E-book mais curto ainda e também gratuito. O Daniel é um econometrista teórico safado que faz mestrado na PUC com quem eu tenho o prazer de dividir esse blog. O livro cobre como fazer o básico de uma graduação em economia no R. De novo: é importante programar tudo.

- Swirl

A ideia é a mesma do DataCamp, aprender fazendo, só que é de graça. Cobre o básico que qualquer um precisa saber e é um ótimo ponto de partida.

- Azul

Não podia deixar de falar do meu bebê. O blog nasceu quando eu falei com o Daniel que queria ter algum jeito de me forçar a praticar R. Dito e feito. O espírito dele é um grande “como eu gostaria de ter aprendido”. A ideia é ser didático - porque ensinar é ótimo para aprender - e dar ideias para mais gente fazer mais experimentos com dados.