

AVALIAÇÃO *EX-ANTE* DE IMPACTO: MODELAGEM DE MICROSSIMULAÇÃO



Avaliação *Ex-Ante* de Impacto: Modelagem de Microssimulação

Organizadores

André Portela Souza
Lycia Lima

Equipe técnica

Arthur Freitas Gomes
Caio de Souza Castro
Leon Labre
Marcos Paulo Cambrainha da Costa
Michel Szklo
Otávio Luiz Tecchio
Pedro Molina Ogeda

APRESENTAÇÃO

Este guia integra a série de publicações *Avaliação na Prática*, desenvolvida pelo FGV CLEAR com o objetivo de ampliar o acesso a conhecimentos sobre monitoramento e avaliação com foco em políticas públicas.

Fundado em 2015, o FGV CLEAR tem se dedicado ao fortalecimento da cultura de gestão orientada por evidências no Brasil e em países lusófonos. Com sede na Escola de Economia de São Paulo da Fundação Getúlio Vargas (FGV EESP), o FGV CLEAR atua como centro regional da Iniciativa CLEAR (*Centers for Learning on Evaluation and Results*).

A Iniciativa CLEAR, criada em 2010, é um programa de desenvolvimento de capacidades em monitoramento e avaliação que congrega instituições acadêmicas e parceiros doadores de modo a contribuir para o uso de evidências na tomada de decisões em países em desenvolvimento. Ao todo, são seis centros regionais CLEAR, coordenados pela *Global Evaluation Initiative (GEI)*, um programa liderado pelo Banco Mundial e pelo Programa das Nações Unidas para o Desenvolvimento (PNUD).

O FGV CLEAR atua com governos, organizações multilaterais, sociedade civil e academia, oferecendo capacitação, assistência técnica, geração e disseminação de conhecimento científico na área.

Saiba mais sobre o FGV CLEAR e acesse outras publicações em: **www.fgvclear.org**

SUMÁRIO

1 Introdução	
1.1 Uma formulação geral para o problema	3
1.2 Microsimulações.....	9
2 Escolha Discreta.....	11
2.1 Modelo de variável aleatória aditiva.....	13
2.2 Modelos multinomiais.....	15
3 Modelos dinâmicos.....	19
3.1 Cadeias de Markov.....	21
3.2 Simulação.....	33
4 Exemplos.....	42
4.1 Taxação de importados.....	42
4.2 Bolsa Família.....	44
4.3 Inclusão de Remédio no Sistema Público de Saúde.....	53
4.3.1 Matriz de transição.....	54
4.3.2 Custo.....	58
4.3.3 Utilidade.....	60
4.3.4 Algoritmo.....	63
4.3.5 Aplicação numérica.....	65
5 Conclusão.....	68

1 Introdução

Suponha que o governo do seu país decida aplicar uma taxa de 20% sobre compras internacionais de até 50 dólares. Você deseja estimar o impacto dessa medida sobre os consumidores brasileiros, especialmente quanto essa taxa representará em relação à renda da população. A primeira (e mais simples) abordagem para estimar o custo ao consumidor é calcular de que forma a tributação afetaria o valor médio das remessas internacionais abaixo de 50 dólares.

Após uma rápida busca na internet, você descobre que o valor médio das remessas internacionais entre setembro de 2023 e janeiro de 2024 é de, aproximadamente, R\$ 78.¹ Com esse dado em mãos, você calcula que o consumidor brasileiro pagaria R\$ 15,6 (20%78) por remessa. Você também calcula que, por mês, o brasileiro médio faz, aproximadamente, 0,77 compras internacionais de menos de 50 dólares. Consequentemente, o brasileiro médio gastaria pouco mais de R\$ 12 por mês, com a imposição da nova taxa. Em princípio, esse número parece pequeno em relação a, por exemplo, o rendimento domiciliar per capita brasileiro em 2023, que foi de R\$ 1.893.²

Porém, ao refletir sobre essa estimativa, você questiona tanto a abordagem por meio da qual ela foi obtida, como também o quanto ela de fato reflete a realidade. Com uma pesquisa mais aprofundada, você descobre que o Imposto sobre Circulação de Mercadorias e Prestação de Serviços (ICMS) incidirá sobre o preço da compra após a taxa de 20%. Ou seja, a tributação teria um custo adicional indireto em virtude o valor de ICMS cobrado.

Além disso, você suspeita que compras internacionais de até 50 dólares são feitas por apenas uma parcela da população brasileira, de modo que o seu exercício original ignora completamente a heterogeneidade entre os consumidores. Se esse tipo de compra estiver concentrado em um grupo específico, então parte da população fará (bem) mais do que 0,77 compras por mês. Além disso, se a maioria desses consumidores for de menor renda, o impacto da tributação representará uma fração ainda maior de sua renda. Diante dessas considerações, você começa a refletir sobre como utilizar dados que relacionem a distribuição de renda da população com o valor e o número de remessas realizadas mensalmente.

Dando continuidade à sua linha de raciocínio, você percebe que, mesmo com dados ideais, c— contendo renda mensal, número e valor de remessas para cada brasileiro —, a introdução do tributo provavelmente alteraria o comportamento dos consumidores. Isso porque o aumento do preço provocado pela taxa tenderia a reduzir a demanda por compras interna-

¹Tomazelli (2024).

²Ferrari (2024).

cionais. DPara quantificar esse efeito, seria necessário estimar a função de demanda por esse tipo de consumo.

O problema acima indica algumas direções e, sobretudo, dificuldades ao prever o efeito causal de políticas públicas. Por exemplo, ao estimar que a taxa vai gerar um custo adicional de R\$ 12 por mês, ignorando que o ICMS incide sobre esse valor, estamos ignorando como a nova política interage com as políticas existentes. Além disso, ao considerar apenas o efeito sobre a *remessa média* e sobre o *brasileiro médio*, estamos desconsiderando que a política afeta pessoas diferentes de formas diferentes, e que o seu efeito pode ser bastante heterogêneo. Adiciona-se a isso o fato de que pessoas mudam seu comportamento após a implementação de uma política, reagindo assim ao novo conjunto de incentivos.

Ao mesmo tempo, há limitações impostas pela disponibilidade de dados. Por exemplo, se quisermos estender o exercício simples aos diferentes quartis da distribuição de renda no Brasil, precisaríamos de informações mais detalhadas — como o número e o valor das compras internacionais para cada faixa de renda. No entanto, esses dados nem sempre estão disponíveis. De forma geral, quanto mais complexo o exercício, maior a necessidade de dados granulares.

Neste guia, descrevemos uma metodologia conhecida como **microsimulação**, que pode ser utilizada para prever os efeitos de políticas públicas. Em linhas gerais, essa abordagem consiste em construir um modelo que represente os aspectos relevantes da política — como a interação com outras políticas existentes e os efeitos sobre o comportamento dos agentes — e, a partir disso, simular o que ocorreria com as variáveis de interesse caso a política fosse implementada. Apresentamos uma definição mais precisa e uma descrição mais detalhada na Subseção 1.2. Antes disso, na Subseção 1.1, consideramos uma estrutura formal para o problema de previsão do efeito de uma política pública, que será a base para toda a exposição subsequente.

Em seguida, abordamos duas temáticas importantes: Escolha Discreta e Modelos Dinâmicos, nas seções 2 e 3, respectivamente. Esses dois arcabouços teóricos constituem a base da maioria dos exercícios de microsimulação. Por fim, concluímos o guia com três exemplos de aplicação da metodologia a políticas públicas, incluindo a programação utilizada para estimar os resultados.

1.1 Uma formulação geral para o problema

O modelo descrito abaixo é baseado em Figari, Paulus e Sutherland (2015).

Considere que estamos interessados em comparar duas políticas distintas e mutuamente excludentes, uma política *Base* (B) e uma política *Contrafactual* (C). Seja p o indexador para política, de forma que $p \in \{B, C\}$, e seja Y_p a variável resultado de interesse no cenário em que a política p é implementada. Nós gostaríamos de comparar Y_B com Y_C . O problema fundamental é que é impossível observar Y_B e Y_C ao mesmo tempo.³ Daqui em diante, para facilitar a exposição, vamos considerar que a política em vigor no momento da coleta dos dados é a política B , e que estamos interessados nos efeitos da política C **antes** de sua implementação. Nesse caso, somos capazes de observar Y_B , mas não Y_C .

Note aqui uma diferença fundamental entre as análises *ex-ante* e as análises *ex-post*. Nas análises *ex-post*, em geral exploramos que a política é aplicada a um subgrupo da população e que, dessa forma, observamos Y_B para as pessoas que estão sob efeito da política B e observamos Y_C para as pessoas que não estão sob efeito da política C . Sob hipóteses apropriadas, por exemplo supondo que esses dois tipos de pessoas são comparáveis, conseguimos obter comparações válidas entre Y_B e Y_C . Porém, isso não ocorre na análise *ex-ante*, em que justamente queremos comparar Y_B e Y_C antes da implementação da política C , no contexto estudado. Portanto, em essência, o que queremos é prever Y_C ou, pelo menos, características de Y_C que consideramos relevantes (sua média por exemplo).

Uma forma para avançar na previsão de Y_C é considerar um *modelo* para a variável de interesse, e usar que observamos Y_B para aprender sobre esse modelo. Formalmente, um modelo para Y_p é um trio (g_p, X_p, θ_p) em que g_p é uma função, X_p são variáveis, observáveis ou não, que determinam Y_p e/ou pelas quais a política afeta Y_p , e θ_p são os parâmetros da política p , de forma que

$$Y_p = g_p(X_p, \theta_p). \quad (1)$$

Devemos entender θ_p como as características que descrevem a política de forma que, em geral, θ_p é conhecido tanto para a política B como para a política C . Por outro lado, embora a parte observável de X_B esteja presente nos dados, não conseguimos observar X_C pelo mesmo motivo que não observamos Y_C . Por fim, g_p deve ser entendida como a estrutura que combina as variáveis X_p com os parâmetros da política θ_p para gerar o resultado Y_p . Uma política pode afetar a estrutura g_p de formas conhecidas antes de sua implementação, mas também de formas não conhecidas.

Na prática, pensar em g_p , X_p e θ_p em casos concretos consiste em escolher as variáveis e fatores relevantes para o problema em questão. Os exemplos abaixo buscam esclarecer o que cada um desses elementos significam em casos concretos.

³Para os familiarizados com métodos quantitativos de avaliação de impacto, Y_B e Y_C são o que chamamos de *resultados potenciais*.

Exemplo 1 (taxação de compras internacionais - continuação)

Primeiro, vamos voltar ao caso da tributação de compras internacionais descrito na introdução. Estávamos interessados em entender como a taxa afetaria a renda da população. Isto é, queríamos saber a diferença entre a renda nos cenários sem taxa e com taxa de 20%. Nesse caso, B indexa o cenário sem a taxa e C indexa o cenário com a taxa. A nossa variável de interesse é a renda da população **após** a taxa, de forma que Y_B e Y_C são as rendas após a incidência nos cenários B e C respectivamente.

Para computar a renda da população após a tributação, o caminho natural é descontar o valor gasto com a taxa da renda pré-taxação. Para isso, precisaríamos da renda pré-taxação, da quantidade de compras internacionais, do valor de cada compra internacional e do valor da taxa. Também, porque o ICMS incide sobre a taxa, o valor do ICMS é um fator de primeira ordem na hora de computar os custos gerados pela taxa.

Note que tanto o valor da taxa como o valor do ICMS são parâmetros da política, de forma que é natural considerar $\theta_p = (\tau_p^{CI}, \tau_p^{ICMS})$, em que τ_p^{CI} é a taxa, em %, incidente sobre as compras internacionais no cenário p e τ_p^{ICMS} é o ICMS cobrado sobre compras internacionais no cenário p , também em %. Dessa forma, temos que $\theta_B = (0, 17\%)$ e $\theta_C = (20\%, 17\%)$ neste exemplo.⁴ Note aqui uma hipótese de que o ICMS não mudaria após a introdução da taxa.

Já a renda pré-taxação, a quantidade de compras internacionais e o valor das compras internacionais são variáveis que explicam Y_p e pelas quais a política afeta Y_p , de forma que essas seriam as variáveis incluídas em X_p . Note que, a princípio, seria possível observar X_B (em um mundo de dados ideais pelo menos), mas não é possível observar X_C diretamente.

Por fim, g_p é a regra que combina as variáveis em X_p com os parâmetros da política para gerar Y_p . Nesse caso, g_p subtrai da renda pré-taxação o valor gasto devido à taxa e ao ICMS. Ou seja,

$$g_p(X_p, \theta_p) = I_p - \frac{1 + \tau_p^{CI}}{1 - \tau_p^{ICMS}} V_p, \quad (2)$$

em que I_p é a renda pré-taxação no cenário p e V_p é o valor das compras internacionais no cenário p .⁵ Note que, além de a estrutura de g_p ser conhecida, ela não depende diretamente de p . Isto é, a estrutura que combina X_p com os parâmetros da política não é afetada pela política. Toda a dependência de p na Equação (2) vem de X_p e de θ_p .

Neste caso, com dados para I_B e V_B , conseguimos facilmente computar Y_B usando a Equação (2). A questão central é como usar a Equação (2) para computar Y_C . Nós conhece-

⁴Tomazelli (2024).

⁵A alíquota do ICMS aparece subtraindo o número 1 no denominador porque é o ICMS é um imposto que incide “por dentro”. Em outras palavras, o valor cobrado pelo ICMS faz parte da base de cálculo do ICMS.

mos τ_C^{CI} e τ_C^{ICMS} , mas e I_C e V_C ? Uma abordagem possível seria supor que eles não mudam devido à taxa, de forma que $I_C = I_B$ e que $V_C = V_B$ e, portanto, usando a Equação (2) teríamos que

$$Y_C = I_B - \frac{1.20}{0.83} V_B. \quad (3)$$

Note, porém, que isso depende crucialmente da hipótese de que a política não afeta nem a renda pré-taxação, nem o valor das compras internacionais.

Em relação a V_C , tal hipótese será tão mais forte quanto mais os preços e a quantidade demandada forem afetados pela taxa. Em geral, como esperamos que os consumidores reajam demandando menos de um produto quando este é taxado, a hipótese que $V_C = V_B$ parece particularmente forte. O custo de flexibilizá-la é que precisaríamos adicionar um novo modelo, agora para explicar V_p . Isso deixa claro um trade-off entre complexidade da análise e quão bem ela se aproxima da realidade.

Por outro lado, a hipótese que $I_C = I_B$ parece mais razoável. É verdade que, a princípio, a taxa de compras internacionais pode afetar a renda pré-taxação das pessoas. Por exemplo, os consumidores podem começar a comprar mais produtos nacionais. Também é possível que haja uma expansão na demanda por trabalho, elevando os salários da população. Ou ainda, por alterar o fluxo de comércio do país com o restante do mundo, a taxa pode ter efeito no câmbio, que por sua vez pode afetar a renda. Entretanto, esses efeitos parecem ser de segunda ordem quando comparados com a incidência da taxa e o ajuste na demanda. Por isso, pode ser que $I_C = I_B$ seja de fato uma hipótese que estamos dispostos a considerar. Caso quiséssemos, por exemplo, permitir o efeito em salários devido à expansão da produção nacional, precisaríamos adicionar um modelo para o mercado de trabalho desse país. Novamente aqui aparece o trade-off entre complexidade e realismo do modelo.

A escolha de que fatores levar em consideração depende do julgamento da pessoa que está modelando o problema, e também quais simplificações a literatura especializada no assunto considera razoáveis ou não. De toda forma, dificilmente será possível (ou desejável) escrever um modelo que capte todos os detalhes da realidade. Escolher quais fatores levar em consideração e quais fatores ignorar (e justificar tais escolhas) faz parte da tarefa do analista.

Exemplo 2 (Transferência de Renda Condicional e Oferta de Trabalho)

O governo de um país está considerando aumentar o valor do principal programa de transferência de renda para reduzir ainda mais a desigualdade e melhorar o bem-estar das famílias de baixa renda. No entanto, há uma preocupação de que o aumento no valor da transferência possa reduzir a participação dos beneficiários no mercado de trabalho, ao diminuir os incentivos para busca de empregos formais ou informais. Antes de implementar essa mudança, o

governo gostaria de ter uma estimativa de como o aumento do benefício afetaria a oferta de trabalho, a participação no mercado de trabalho formal e informal, e a não participação.

Uma diferença importante entre este exemplo e o Exemplo 1 é que, neste caso, o objetivo da política é alterar o bem-estar das famílias sem desincentivar excessivamente sua participação no mercado de trabalho. Dessa forma, qualquer estimativa razoável dos efeitos deve levar em consideração o impacto que o valor da transferência tem nas decisões ocupacionais dos beneficiários. Ou seja, será necessário modelar a escolha ocupacional dos indivíduos e das famílias.

A variável de resultado em que o governo está interessado é o status ocupacional dos beneficiários. Olhando para essa variável, é possível saber tanto a probabilidade de participação no mercado de trabalho formal como a probabilidade de permanência no emprego informal ou de não participação no mercado de trabalho. Ou seja, Y_p é o status ocupacional sob a política p .

Para X_p , podemos pensar em uma série de variáveis que explicam a participação no mercado de trabalho, como escolaridade, idade e gênero. Uma variável de primeira ordem é a renda domiciliar per capita. Essa variável ajuda a entender o quanto a família depende do trabalho e/ou do Bolsa Família para compor sua renda. De forma mais geral, veremos na Seção 2 que é conveniente colocar em X_p variáveis que explicam as preferências por trabalho formal, informal ou não participação, naturalmente levando em conta a disponibilidade dos dados.

Por fim, g_p irá capturar como os beneficiários escolhem seu status ocupacional dado as variáveis em X_p . Uma forma de fazer isso é dizer que g_p compara o benefício líquido (isto é, descontando o custo) de cada status ocupacional possível e retorna o status com benefício líquido máximo. Daremos um tratamento mais detalhado desse problema na Seção 2.

Exemplo 3 (Inclusão de Remédio no Sistema Público de Saúde)

O governo do seu país está considerando incluir um novo remédio, o WZ, no sistema público de saúde, através do qual o medicamento seria distribuído gratuitamente. O objetivo do WZ é combater uma doença específica. Essa doença tem dois estágios, um menos grave, o outro mais grave. Pessoas em estágio mais grave são internadas em centros intensivos de tratamento por bastante tempo, e podem eventualmente falecer.

Estudos sobre o medicamento WZ mostram que ele é bastante efetivo em prevenir que pessoas em estado menos crítico evoluam para um quadro grave. Porém, ele não é muito efetivo em tratar pacientes que já se encontram em estágio grave da doença.

Um dos fatores que o governo está levando em consideração na decisão de prover o medicamento via o sistema público é o quanto seria economizado com a redução de internações, em virtude de um menor número de doentes em estado grave. Ou seja, o governo está

interessado em uma previsão do efeito da adoção do remédio no custo de internação nos próximos anos.

Nesse contexto, a variável de interesse Y_p é o custo de internação, que crucialmente vai depender do custo diário de internação e do número de dias que a pessoa fica internada. O número de dias que a pessoa fica internada irá variar de pessoa a pessoa, por isso essa é uma variável que entra em X_p . Por outro lado, o custo diário de internação é, basicamente, o custo para o governo em manter um leito em um centro de tratamento intensivo. Isso pode ser considerado um parâmetro da política. Porém, como o efeito do medicamento em estados avançados da doença é irrelevante, faria sentido supor que tanto o tempo de internação como o custo diário não mudam após a adoção do remédio. Mas então, como o remédio afeta Y_p ?

Embora possa fazer sentido supor que o remédio não afeta o custo por doente grave, note que o custo total para o governo depende do número de doentes graves, e nesse ponto o remédio causa um impacto. Na verdade, o medicamento afeta a probabilidade de transição do estado menos grave para o estado mais grave da doença. Por isso, g_p deve incluir um modelo para a transição entre estados, levando em conta que a probabilidade de transição do estado menos grave para o estado mais grave será alterada pela utilização do remédio por usuários do sistema público de saúde.

Como parâmetros da política, seria interessante incluir, além do custo diário do leito, o efeito do remédio na probabilidade de transição. Note que, diferente dos Exemplos 1 e ??, esse valor não é definido pela política, mas nós temos acesso a uma estimativa produzida pela literatura especializada.

Com esse modelo de transição, conseguimos simular o número de doentes graves, e com o número de doentes graves e o custo médio de internação por doente grave, conseguiríamos uma estimativa da redução dos custos de internação em razão da política.

Porém, seria possível estimar esse efeito de forma mais precisa, tornando o modelo mais complexo para capturar aspectos importantes da realidade. Por exemplo, o fato de o remédio ser ofertado gratuitamente não significa que todas as pessoas em estado menos avançado tomam o remédio. Pode ser interessante modelar essa heterogeneidade na adoção do WB. Uma forma de fazer isso é utilizar modelos para variáveis binárias (como descrito na Seção ??), considerando quais variáveis observáveis explicam a adoção heterogênea do remédio, como por exemplo gênero, idade, escolaridade e renda. Variáveis que, então, entrariam em X_p . Inclusive, o número de dias que a pessoa fica internada também podem depender dessas variáveis. Pense, por exemplo, que pessoas com idade mais avançada ficam mais tempo internadas, mas são justamente essas pessoas que mais adotam o remédio. Levar essa dimensão em conta pode ser importante para se ter uma noção mais precisa dos efeitos da política.

1.2 Microssimulações

Microssimulações é o nome dado a um conjunto de técnicas de modelagem baseadas em simulações em que a análise é feita usando micro unidades/unidades individuais (O'Donoghue 2021, capítulo 1; Figari, Paulus e Sutherland 2015). O uso de micro unidades significa que o modelo é escrito para descrever as variáveis em um nível individual, como por exemplo para pessoas, famílias ou firmas. Por isso, os dados que em geral são utilizados em microssimulações são o que chamamos de *microdados*, dados com informações a nível individual. Na verdade, Mitton, Sutherland e Weeks (2000, capítulo 1) estressam que o uso de informações a nível individual é um fator distintivo dos métodos de microssimulação.

Já o uso de simulações significa que uma sequência de regras, determinísticas ou estocásticas, é aplicado ao dados com o objetivo de simular um cenário contrafactual gerado por uma política (ou, de forma mais geral, por uma mudança de estado) (Figari, Paulus e Sutherland 2015). Os cenários contrafactuais simulados podem, então, ser comparados com os dados para obter uma estimativa dos efeitos da política (ou da mudança de estado) nas variáveis de interesse. Naturalmente, por serem simulações a um nível micro, as estimativas dos efeitos também serão a nível micro, mas que podem ser agregadas em passos subsequentes.

Na estrutura desenvolvida na subseção anterior, teríamos um modelo a nível individual para Y_p (e, potencialmente, para alguns ou todos os elementos de X_p). Usando dados a nível individual para Y_B e para a parte observável de X_B , simuláramos o modelo para gerar Y_C . Essa simulação pode envolver apenas regras determinísticas, como é o caso da regra na Equação (3). Por outro lado, o uso de regras estocásticas pode ser essencial para a construção de um modelo, especialmente em situações onde é necessário incorporar variações ou incertezas associadas ao fundamental para um modelo, como quando queremos capturar possíveis mudanças de comportamento dos indivíduos. Por exemplo, no uso de um modelo de demanda apresentado no Exemplo 1, as regras estocásticas permitem capturar alterações no comportamento dos indivíduos diante de mudanças nos preços ou na tributação., que seria estimado usando os dados para depois ser aplicado na simulação. Isso é um exemplo de regra estocástica. A noção de um modelo de microssimulação como a aplicação de sequências de regras determinísticas e estocásticas é discutida, por exemplo, em O'Donoghue (2021, capítulo 2).

Métodos e modelos de microssimulações são aplicados em diversas áreas, como transporte e infraestrutura, saúde, demografia, entre outras (veja, por exemplo, O'Donoghue 2014). Em Economia, o uso mais comum de microssimulações ocorre na simulação dos efeitos de políticas de taxaço na renda dos indivíduos e das famílias (O'Donoghue 2021, capítulo 1).

Grosso modo, os modelos de microssimulações podem ser divididos em três classes: mo-

delos estáticos, dinâmicos e comportamentais (Mitton, Sutherland e Weeks 2000, capítulo 1; Figari, Paulus e Sutherland 2015). Figari, Paulus e Sutherland (2015) descreve os modelos estáticos como aqueles em que as características dos indivíduos não mudam e apenas regras determinísticas são aplicadas aos dados, com eventuais ajustes nos dados caso necessário (como reponderação para garantir uma amostra representativa). Esse tipo de modelo é, em geral, utilizado quando o efeito de interesse é de curto prazo, e o efeito da política no comportamento da população é uma preocupação de segunda ordem. Por outro lado, quando a passagem do tempo é fundamental para que os efeitos da política sejam realizados, como é o caso de mudanças nas regras de aposentadoria ou mudanças em políticas de saúde, modelos dinâmicos devem ser considerados. A característica fundamental que diferencia modelos dinâmicos de modelos estáticos é o "envelhecimento" das micro unidades (Mitton, Sutherland e Weeks 2000, capítulo 1). Por "envelhecimento" não deve-se entender apenas que as unidades ficam mais velhas, mas que seu estado potencialmente muda com a passagem dos períodos do modelo (por exemplo, pessoas ativas no mercado de trabalho se tornam aposentadas). Regras de envelhecimento podem ser tanto determinísticas quanto estocásticas. Por fim, os modelos comportamentais são aqueles que adicionam respostas comportamentais dos indivíduos devido à política, como por exemplo mudanças na demanda devido à taxação, como discutido no Exemplo 1. Tais mudanças comportamentais são tipicamente introduzidas por modelos econométricos para preferências (Figari, Paulus e Sutherland 2015).

No final das contas, a divisão dos modelos em três classes é útil para organizar a discussão, entender a terminologia da literatura especializada e buscar referências. Na prática, porém, deve-se buscar o modelo mais adequado para a questão em análise, combinando técnicas de cada braço da literatura se necessário. Escolher levar em conta fatores dinâmicos ou comportamentais, por exemplo, é uma escolha da pessoa que conduz a análise, e deve ser guiada pelo conhecimento acumulado pela literatura e por considerações em relação a quais fatores busca-se isolar.

Este guia foca em exemplificar como métodos de microsimulações podem ser aplicados na prática, discutindo um exemplo específico, detalhando o desenvolvimento de um modelo para esse caso e a aplicação desse modelo para prever o efeito causal da política considerada. Uma referência que pode ser complementar a leitura deste guia é O'Donoghue (2014), que cobre diversos aspectos de microsimulações aplicados a diferentes contextos, como modelos para taxação, transporte e saúde. Mitton, Sutherland e Weeks (2000) apresenta um compilado de trabalhos com temas relacionados a economia, desde taxação e sistemas de pensão até migração e gastos relacionados a saúde. Mitton, Sutherland e Weeks (2000, capítulo 1) propõe reflexões interessantes sobre considerações na escolha de modelos, incerteza das estimativas, entre outros pontos. O'Donoghue (2021) discute vários métodos e modelos para

políticas relacionadas à distribuição de renda. Figari, Paulus e Sutherland (2015) foca em modelos para os efeitos de sistemas de benefícios fiscais na renda dos indivíduos.

2 Escolha Discreta

As escolhas individuais podem ser representadas por variáveis discretas ou contínuas, dependendo da natureza da decisão. No exemplo sobre taxação de compras internacionais, decisões como quanto consumir em bens importados são mais adequadamente capturadas por variáveis contínuas, como o montante em reais gasto em compras internacionais. Já no caso de Transferências de Renda Condicional e Oferta de Trabalho, a escolha é representada por uma variável categórica multinomial. Seja y uma variável aleatória que assume os valores 0, 1 e 2, onde $y = 0$ é definido como o indivíduo não trabalhar, $y = 1$ como o indivíduo trabalhar no setor informal e $y = 2$ como o indivíduo trabalhar no setor formal.

Seja Y^* uma variável aleatória latente (ou não observada), como, por exemplo, o ganho líquido de mandar o filho à escola ou o benefício de tomar ou não um remédio. Seja $X = (X_1, X_2, \dots, X_n)$ um vetor de variáveis aleatórias, chamadas covariadas, que ajudam a explicar a probabilidade de sucesso. No caso do comparecimento escolar, X_i pode representar o nível de educação dos pais. No exemplo da inclusão de um remédio no sistema público de saúde, uma possível covariável é o número de dias que uma pessoa fica internada.

Além disso, seja Y uma variável aleatória categórica que assume os valores 0, 1 e 2. Definimos $Y = 0$ se a unidade de análise não trabalha, $Y = 1$ se trabalha no setor informal, e $Y = 2$ se trabalha no setor formal. Por exemplo, $Y = 0$ pode indicar que um indivíduo opta por não trabalhar, $Y = 1$ que trabalha sem registro formal, e $Y = 2$ que possui um emprego com carteira assinada.

Sejam (Y, X) variáveis aleatórias observáveis, onde $Y \in \{0, 1, 2\}$ e $X \in \mathbb{R}^n$. O modelo de regressão natural para Y^* é o modelo de função índice:

$$Y_j^* = X' \beta_j + \epsilon_j, \quad \epsilon_j \sim F(\epsilon_j), \quad j \in \{0, 1, 2\},$$

onde Y_j^* é uma variável latente associada à utilidade da escolha j , linear em X e em um termo de erro ϵ_j , sendo este último extraído de uma distribuição simétrica F . Esse modelo não pode ser estimado diretamente, pois Y_j^* não é observado. Em vez disso, observamos:

$$Y = \begin{cases} 0 & \text{se } Y_0^* > Y_1^* \text{ e } Y_0^* > Y_2^*, \\ 1 & \text{se } Y_1^* > Y_0^* \text{ e } Y_1^* > Y_2^*, \\ 2 & \text{se } Y_2^* > Y_0^* \text{ e } Y_2^* > Y_1^*. \end{cases}$$

No contexto da oferta de trabalho, isso significa que um indivíduo opta por não trabalhar ($Y = 0$) se o ganho líquido associado a essa escolha (Y_0^*) for maior que o ganho líquido das outras opções. Similarmente, o indivíduo trabalha no setor informal ($Y = 1$) se Y_1^* for o maior, e no setor formal ($Y = 2$) se Y_2^* for o maior.

Um objeto de interesse é a probabilidade condicional de cada categoria de Y , dada X , que descreve completamente a distribuição condicional:

$$\mathbb{P}[Y = j \mid X = x] = \mathbb{P}[Y_j^* > Y_k^* \text{ para todos } k \neq j \mid X = x], \quad j \in \{0, 1, 2\}.$$

No caso do modelo logit multinomial, a probabilidade de cada categoria é modelada como:

$$\mathbb{P}[Y = j \mid X = x] = \frac{\exp(X'\beta_j)}{\sum_{k=0}^2 \exp(X'\beta_k)}, \quad j \in \{0, 1, 2\},$$

onde β_j são os coeficientes associados à categoria j .

Essa formulação permite analisar, por exemplo, como características individuais X (como idade, nível de escolaridade e número de dependentes) afetam as probabilidades de uma pessoa optar por não trabalhar, trabalhar no setor informal ou no setor formal.

A partir da maneira como se modela a cdf F , chega-se a diferentes modelos de variável categórica. Dois dos mais comuns para variáveis dependentes com múltiplas categorias são o modelo probit multinomial e o modelo logit multinomial:

Modelo Probit Multinomial: A probabilidade de $Y = j$ é dada por:

$$\mathbb{P}[Y = j \mid X = x] = \int_{-\infty}^{+\infty} \prod_{k \neq j} \Phi(x'\beta_j - x'\beta_k + \epsilon_k) \phi(\epsilon_j) d\epsilon_j,$$

onde Φ é a função de distribuição acumulada normal padrão.

Modelo Logit Multinomial: A probabilidade de $Y = j$ é modelada como:

$$\mathbb{P}[Y = j \mid X = x] = \frac{\exp(x'\beta_j)}{\sum_k \exp(x'\beta_k)},$$

onde β_j são os coeficientes associados à categoria j .

No contexto da oferta de trabalho, esses modelos permitem estimar, por exemplo, como características individuais X (como nível de escolaridade e idade) afetam a probabilidade de uma pessoa optar por não trabalhar ($Y = 0$), trabalhar no setor informal ($Y = 1$) ou no setor formal ($Y = 2$).

Seja x uma variável contínua; um parâmetro de interesse nesses modelos é o efeito marginal, que mede o impacto de uma pequena mudança em x sobre a probabilidade de $Y = j$.

Para o modelo logit multinomial, o efeito marginal é dado por:

$$\frac{\partial}{\partial x} \mathbb{P}[Y = j \mid X = x] = \mathbb{P}[Y = j \mid X = x] \left(\beta_j - \sum_k \mathbb{P}[Y = k \mid X = x] \beta_k \right).$$

Nos exemplos trabalhados até agora, esse efeito marginal poderia ser usado para medir como uma mudança no número de dependentes impacta a probabilidade de escolha de trabalhar no setor formal.

Os efeitos marginais variam com x , o que pode dificultar a interpretação dos resultados. Por isso, é comum reportar a média dos efeitos marginais para todas as observações (*Average Marginal Effect*, AME):

$$\text{AME}_j = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x} \mathbb{P}[Y = j \mid X = x_i].$$

Para estimar os parâmetros β_j , tanto o modelo probit quanto o logit multinomial são usualmente ajustados por máxima verossimilhança. A estimação é consistente e assintoticamente normal, desde que sejam satisfeitas as condições de regularidade apropriadas. Para mais detalhes teóricos sobre os estimadores, ver o Apêndice “ ”Estimação modelos variáveis categóricas e identificação.”

2.1 Modelo de variável aleatória aditiva

Nesta subseção, iremos expor uma segunda formulação do modelo de variável latente, também chamada de modelo de variável aleatória aditiva. Esse modelo servirá de base para a derivação do modelo de escolha multinomial.

Sejam $j = 0, 1$ a indexação das escolhas disponíveis para o consumidor. O tomador de decisão obteria um certo nível de utilidade de cada alternativa. A utilidade que o tomador de decisão obtém da alternativa j é U_j , $j = 0, 1$. O tomador de decisão escolhe a alternativa que fornece a maior utilidade. Ou seja:

$$j^* = \arg \max_{j \in \{0,1\}} U_j.$$

Considere o problema do pesquisador: ele não observa a utilidade do tomador de decisão, mas observa um componente determinístico das utilidades V_j e pode especificar uma função que relacione esses fatores observados à utilidade do tomador de decisão. Contudo, mesmo com essas variáveis observáveis, ainda haverá aspectos da utilidade do tomador de decisão que não são observáveis pelo pesquisador, denotados por e_j . Vamos supor que o pesquisador postule o seguinte modelo de utilidade:

$$U_j = V_j + \epsilon_j, \quad j = 0, 1.$$

Uma possibilidade é modelar o componente determinístico da seguinte maneira $V_j = x' \beta_j$.

A alternativa escolhida é aquela com a maior utilidade. Sem perda de generalidade, vamos supor que $j^* = 1$, logo temos que

$$\mathbb{P}[Y = 1] = P[U_1 > U_0] = \mathbb{P}[V_1 + \epsilon_1 > V_0 + \epsilon_0] = \mathbb{P}[\epsilon_0 - \epsilon_1 < V_1 - V_0] = F(V_1 - V_0),$$

onde F é a função de distribuição acumulada de $(\epsilon_0 - \epsilon_1)$. Isso resulta em $P[Y = 1] = F(x' \beta)$ se $V_1 - V_0 = x'(\beta_1 - \beta_0) = x' \beta$.

Se ϵ_0 e ϵ_1 seguem uma distribuição normal, então $(\epsilon_0 - \epsilon_1)$ é normalmente distribuído. A normalização da variância de $(\epsilon_0 - \epsilon_1)$ para a unidade gera o modelo probit, pois $F(\cdot)$ será a função de distribuição acumulada normal padrão.

Vamos supor que a variável aleatória ϵ tem a seguinte densidade:

$$f(\epsilon) = e^{-\epsilon} \exp(-e^{-\epsilon}), \quad -\infty < \epsilon < \infty,$$

e função de distribuição acumulada $F(\epsilon) = \frac{1}{1 + \exp(-\epsilon)}$. Ou seja, ϵ segue uma distribuição do valor extremo tipo 1.

O modelo logit surge se ϵ_0 e ϵ_1 forem assumidas como distribuídas de forma independente como valor extremo tipo 1. Nesse caso, a diferença $(\epsilon_0 - \epsilon_1)$ pode ser mostrada como logisticamente distribuída, então $F(\cdot)$ é a função de distribuição acumulada logística.

O modelo de variável aleatória aditiva proposto até agora só considerou regressores que não variam entre as alternativas. Por exemplo, com relação ao exemplo de transparência de renda condicional e comparecimento escolar, características socioeconômicas como renda e gênero não variam entre as alternativas. Há, porém, regressores em potencial, como renda, que variam entre as alternativas, como por exemplo receber ou não o Bolsa Família.

Podemos adicionar esse tipo de regressores ao modelo da seguinte maneira. Seja a seguinte componente determinística da utilidade:

$$V_j = W_j' \gamma + X' \beta_j, \quad j = 0, 1,$$

onde W_j' são regressores que tomam valores diferentes entre as duas alternativas, enquanto X são características individuais que não variam com a escolha. Então, resulta em:

$$\mathbb{P}[Y = 1] = F(W_1' \gamma - W_0' \gamma + X'(\beta_1 - \beta_0)).$$

2.2 Modelos Multinomiais

Vamos supor que, no problema da transparência de renda condicional e comparecimento escolar, há três estados possíveis: a criança não frequenta a escola, a criança vai à escola e trabalha fora de casa, ou a criança vai à escola e não trabalha fora de casa. Modelos de resposta com mais de dois valores são chamados de modelos multinomiais.

De maneira geral, seja Y uma variável aleatória multinomial que assume valores em um conjunto finito, tipicamente escrito como $Y \in \{1, 2, \dots, J\}$, ou seja, para um conjunto Y com dois elementos, voltamos ao caso do modelo binário. Seja o par (Y, X) a resposta multinomial quando Y é multinomial e $X \in \mathbb{R}^k$ são regressores. A distribuição condicional de Y dado X é resumida pela probabilidade de resposta:

$$P_j(x) = \mathbb{P}[Y = j \mid X = x].$$

Assim como no caso binário, podemos motivar um modelo de escolha multinomial a partir de um modelo de variável latente. Seja $j = 1, \dots, J$ a indexação das alternativas disponíveis. A utilidade da alternativa j é assumida como igual a:

$$U_j = X' \beta_j + \epsilon_j.$$

A alternativa escolhida é aquela com a maior utilidade, de modo que:

$$\mathbb{P}[Y = j^*] = \Pr[U_j \geq U_k, \quad \text{para todo } k \neq j]$$

$$= \mathbb{P}[U_k - U_j \leq 0, \quad \text{para todo } k \neq j]$$

$$= \mathbb{P}[\epsilon_k - \epsilon_j < X'(\beta_j - \beta_k), \quad \text{para todo } k \neq j]$$

$$= F[X'(\beta_j - \beta_k)],$$

onde F é a função de distribuição de $\epsilon_k - \epsilon_j$. Para derivar o logit multinomial, suponha que o vetor de erro $(\epsilon_1, \dots, \epsilon_J)$ tenha distribuição valor extremo tipo I. Então, as probabilidades de resposta são dadas por:

$$\begin{aligned}
P_j(X) &= \frac{\exp(U_j - U_k)}{\sum_{l=1}^J \exp(U_l)} \\
&= \frac{\exp(X'(\beta_j - \beta_k))}{\sum_{l=1}^J \exp(X'\beta_l)} \\
&= \frac{\exp(X'\beta_j^*)}{\sum_{l=1}^J \exp(X'\beta_l)}.
\end{aligned}$$

Uma característica desse modelo é que os coeficientes β_j não são identificados separadamente:

$$\begin{aligned}
P_j(X) &= \frac{\exp(X'(\beta_j + \delta) - X'(\beta_k + \delta))}{\sum_{l=1}^J \exp(X'\beta_l)} \\
&= \frac{\exp(X'(\beta_j - \beta_k))}{\sum_{l=1}^J \exp(X'\beta_l)}.
\end{aligned}$$

A identificação é obtida impondo uma normalização. Uma escolha é definir $\beta_k = 0$ para uma alternativa base k . Os coeficientes relatados β_j devem ser interpretados como diferenças em relação à alternativa base, ou seja, as diferenças entre as alternativas $\beta_j^* = \beta_j - \beta_k$ são identificadas:

$$\begin{aligned}
P_j(X) &= \frac{\exp(X'\beta_j + \delta)}{\sum_{l=0}^J \exp(X'\beta_l + \delta)} \\
&= \frac{\exp(X'\beta_j + \delta)}{\sum_{l=0}^J \exp(X'\beta_l + \delta)} \\
&= \frac{\exp\{X'\delta\} \exp\{X'\beta_j\}}{\sum_{l=0}^J \exp\{X'\delta\} \exp\{X'\beta_l\}} \\
&= \frac{\exp\{X'\delta\} \exp\{X'\beta_j\}}{\sum_{l=1}^J \exp\{X'\delta\} \exp\{X'\beta_l\} + \exp\{X'\delta\}}
\end{aligned}$$

Assim como no caso binário, também precisamos fixar a escala para estimar a variância dos termos de erro e_j . Normalmente fazemos a normalização $\sigma = 1$. Suponha que temos o modelo com erros do tipo valor extremo I:

$$U_j = X'\beta_j + \sigma e_j.$$

As probabilidades condicionais de escolha implicadas pelo logit são:

$$P_j = \frac{\exp(x_i(\beta_j/\sigma))}{\sum_{k=0}^J \exp(x_i(\beta_k/\sigma))}.$$

No caso do logit multinomial, podemos calcular os efeitos marginais pela fórmula:

$$\delta_j(x) = \frac{\partial}{\partial x} P_j(x) = P_j(x) \left(\beta_j - \sum_{\ell=1}^J \beta_\ell P_\ell(x) \right).$$

Isso é estimado por:

$$\hat{\delta}_j(x) = \hat{P}_j(x) \left(\hat{\beta}_j - \sum_{\ell=1}^J \hat{\beta}_\ell \hat{P}_\ell(x) \right).$$

Assim como no caso binário, é comum reportar o efeito marginal médio AME_j , que pode ser estimado por:

$$\widehat{\text{AME}}_j = \frac{1}{n} \sum_{i=1}^n \hat{\delta}_j(X_i).$$

Com relação ao modelo de variável aleatória aditiva, ao adicionarmos regressores que são específicos para cada alternativa, temos o seguinte modelo de utilidade:

$$U_j = W_j' \gamma + X' \beta_j + e_j.$$

Se assumirmos que os ε_j são distribuídos i.i.d. (independentemente e identicamente distribuídos) como valor extremo tipo I, temos a seguinte probabilidade de resposta:

$$P_j(w, x) = \frac{\exp(x' \beta_j + w' \gamma)}{\sum_{\ell=1}^J \exp(x' \beta_\ell + w' \gamma)}.$$

Utilizando o mesmo raciocínio do modelo anterior, os coeficientes γ e as diferenças entre coeficientes $\beta_j - \beta_\ell$ são identificados até a escala. A identificação é obtida normalizando a escala de ε_j e definindo $\beta_k = 0$ para uma alternativa base k .

Para esse modelo, podemos computar o efeito marginal em relação a W_j da seguinte maneira:

$$\delta_{jj}(w, x) = \frac{\partial}{\partial x_j} P_j(w, x) = \gamma P_j(w, x) (1 - P_j(w, x)),$$

e para $j \neq \ell$:

$$\delta_{j\ell}(w, x) = \frac{\partial}{\partial x_\ell} P_j(w, x) = -\gamma P_j(w, x) P_\ell(w, x).$$

Por exemplo, para:

- W = Renda
- j = Criança comparece à escola
- ℓ = Criança só trabalha e não comparece à escola

$\delta_{j\ell}$ é o efeito marginal de uma mudança na renda sobre a probabilidade de a criança frequentar a escola ao invés de apenas trabalhar. Os efeitos marginais médios $\text{AME}_{j\ell} = \mathbb{E}[\delta_{j\ell}(W, X)]$ podem ser estimados por médias amostrais análogas ao caso binário ou ao exemplo anterior.

A função de verossimilhança, dada uma amostra aleatória $\{Y_i, X_i\}$, é construída de modo similar ao caso binário. Seja $\theta = (\beta_1, \dots, \beta_J, \gamma)$. Dadas as observações $\{Y_i, W_i, X_i\}$, onde $X_i = \{X_{i1}, \dots, X_{iJ}\}$, a função de log-verossimilhança é:

$$\ell_n(\theta) = \sum_{i=1}^n \sum_{j=1}^J 1\{Y_i = j\} \log P_j(W_i, X_i \mid \theta).$$

O estimador de máxima verossimilhança (MLE) $\hat{\theta}$ maximiza $\ell_n(\theta)$. Não há solução algébrica, então $\hat{\theta}$ precisa ser encontrada numericamente.

Uma restrição importante desses tipos de modelos é que, para parâmetros e regressores fixos, a razão da probabilidade de duas alternativas é:

$$\frac{P_j(W, X \mid \theta)}{P_\ell(W, X \mid \theta)} = \frac{\exp(X' \beta_j + W'_j \gamma)}{\exp(X' \beta_\ell + W'_\ell \gamma)}.$$

A razão de chances (*odds ratio*) é uma função que depende exclusivamente das variáveis X_j e X_ℓ , sem ser influenciada pelas variáveis específicas das outras alternativas ou pela presença de alternativas adicionais. Essa característica é conhecida como independência das alternativas irrelevantes (IIA), o que implica que a escolha entre as opções j e ℓ é independente da existência de outras alternativas, tornando-as irrelevantes para a decisão entre as duas.

No contexto do Bolsa Família e comparecimento escolar, a razão entre a probabilidade de uma criança frequentar a escola e trabalhar (alternativa 2) e a probabilidade de não frequentar a escola (alternativa 1) deve permanecer inalterada, independentemente de existir uma terceira alternativa, como frequentar a escola sem trabalhar.

Suponha que, além do Bolsa Família, o governo implemente um novo programa que ofereça aulas de reforço gratuitas, tornando a terceira alternativa (frequentar a escola sem trabalhar) mais atraente. Como resultado, algumas famílias podem passar a preferir que seus filhos frequentem a escola sem trabalhar, em vez de frequentar a escola e trabalhar.

Segundo a suposição de IIA, essa nova atratividade da terceira alternativa não deveria alterar a razão entre as probabilidades das outras duas alternativas (frequentar a escola e trabalhar versus não frequentar a escola). No entanto, na prática, a introdução dessa terceira alternativa pode impactar diretamente a decisão entre as duas opções originais. Por exemplo, se frequentar a escola sem trabalhar se torna uma opção mais viável, a probabilidade de escolher frequentar a escola e trabalhar pode diminuir em comparação com a de não frequentar a escola.

3 Modelos Dinâmicos

Um modelo dinâmico simula o comportamento das microunidades ao longo do tempo, necessitando de um processo de "envelhecimento" para representar essas mudanças temporais. No contexto da microsimulação, o envelhecimento refere-se ao processo de atualizar uma base de dados para refletir condições atuais ou projetar cenários futuros. Esse processo pode ser classificado em dois tipos principais: estático e dinâmico (Li e O'Donoghue, 2013).

O envelhecimento estático ajusta os pesos das observações individuais para que reflitam projeções atuais ou cenários hipotéticos de variáveis de interesse. Embora seja uma técnica direta e útil para testar diferentes situações, sua eficácia diminui quando muitas variáveis precisam ser consideradas simultaneamente ou quando é necessário acompanhar transições individuais ao longo do tempo (Figari, Paulus e Sutherland, 2015).

Por exemplo, no setor de saúde, para simular um aumento na incidência de uma determinada doença, aumentam-se os pesos dos indivíduos doentes na amostra e reduzem-se os dos saudáveis. Isso permite analisar o impacto de um surto ou epidemia, mantendo constantes outras características demográficas e socioeconômicas. No entanto, essa abordagem pode atribuir as características de pacientes crônicos aos novos casos, levando a estimativas imprecisas—especialmente se os novos pacientes tiverem perfis clínicos diferentes ou se os tratamentos disponíveis tiverem evoluído. Além disso, um aumento na incidência de doenças pode afetar regiões geográficas ou grupos populacionais de maneiras distintas, algo que uma simples reponderação pode não capturar adequadamente.

O envelhecimento estático, por ser menos custoso computacionalmente, é adequado para ajustes mais simples e é frequentemente utilizado em modelos teóricos simplificados (Dekkers, 2015).

Em contraste, o envelhecimento dinâmico modifica diretamente os atributos dos indivíduos, atualizando idade e características associadas ao longo do tempo (Li e O'Donoghue, 2013). Embora seja mais exigente computacionalmente, é mais apropriado para projeções de longo prazo. Ele permite a introdução de novos perfis diferentes dos dados de base e

possibilita a modelagem detalhada de trajetórias de carreira e reformas que dependem de coorte, idade e período. Por isso, este guia focará em modelos com envelhecimento dinâmico.

Para implementar o envelhecimento dinâmico, modelos de microssimulação podem empregar duas abordagens principais para simular mudanças: modelos comportamentais estruturais e matrizes de transição.

Os modelos comportamentais estruturais, baseados na teoria econômica, capturam as alterações de comportamento dos agentes em resposta a mudanças institucionais ou de mercado. As matrizes de transição, geralmente representadas por cadeias de Markov, são usadas para modelar transições básicas, como idade, sendo abordagens mais simples, que não envolvem decisões explícitas dos agentes e menos teóricas (Li e O'Donoghue, 2013).

Enquanto matrizes de transição são comuns para simular fenômenos como mortalidade e transições no mercado de trabalho, podem não capturar adequadamente os efeitos de novas políticas ou reformas, pois não consideram respostas comportamentais a quaisquer mudanças. Em contraste, os modelos comportamentais estruturais integram parâmetros de política que influenciam diretamente o comportamento dos agentes, como o ajuste da oferta de trabalho frente a mudanças no sistema de benefícios fiscais (Li e O'Donoghue, 2013).

É importante notar que a divisão entre essas abordagens não é rígida. É possível atualizar as probabilidades de transição por meio de modelos comportamentais, integrando elementos de diferentes métodos para aprimorar a simulação.

Para ilustrar as diferenças entre essas abordagens, suponhamos que desejamos avaliar o impacto de uma transição de uma jornada de trabalho de 6x1 (seis dias trabalhados com um de descanso) para uma jornada de 5x2 (cinco dias trabalhados com dois de descanso) no *turnover* dos empregados no setor varejista. Podemos aplicar as abordagens de modelos comportamentais estruturais, matrizes de transição e uma combinação de ambas.

- **Exemplo de Modelos Comportamentais Estruturais:**

Um modelo comportamental estrutural poderia ser utilizado para capturar como a mudança no regime de trabalho influencia a probabilidade de um funcionário permanecer na empresa. Um modelo logit, por exemplo, pode estimar a probabilidade de retenção dos trabalhadores com base em fatores como horas de trabalho e características individuais (estado civil, nível de escolaridade etc.). Esse modelo permitiria observar se a alteração para a semana 5x2 reduz a chance de saída voluntária, ao aumentar o tempo destinado ao lazer.

- **Exemplo de Matrizes de Transição:**

Uma matriz de transição, por sua vez, pode ser usada para modelar as mudanças nos estados de trabalho dos funcionários ao longo do tempo. Nesse caso, cada estado de

trabalho representaria uma condição possível para o empregado, como "trabalhando", "deixando a empresa voluntariamente" e "turnover" involuntário (demissões). A matriz de transição utilizaria as taxas de *turnover* observadas antes da política para simular como a população de funcionários evolui ao longo do tempo, sem considerar a resposta direta à nova política.

- **Exemplo de Combinação das Abordagens:**

Na combinação das abordagens, uma matriz de transição seria utilizada para modelar o *turnover* antes e depois da implementação da jornada 5x2, criando duas matrizes distintas: P_{antiga} (para o regime 6x1) e P_{nova} (para o regime 5x2).

Primeiro, o modelo comportamental (como o logit) estima as novas probabilidades de retenção de acordo com a política 5x2, capturando o impacto do aumento de satisfação e equilíbrio entre vida pessoal e trabalho sobre a probabilidade de um funcionário permanecer na empresa. Com essas estimativas, a matriz P_{nova} é ajustada para refletir as probabilidades de transição sob a nova política, em comparação com P_{antiga} , que representa o regime 6x1.

A matriz P_{antiga} modela a probabilidade de um funcionário permanecer, sair voluntariamente ou ser desligado, com base nas taxas de *turnover* antes da mudança. Já P_{nova} incorpora as respostas positivas ao novo regime, mostrando uma probabilidade menor de saída voluntária. Com essas duas matrizes, a simulação consegue comparar diretamente os efeitos do regime 5x2 versus o 6x1, projetando o impacto acumulado da política na força de trabalho ao longo do tempo.

Na próxima seção, abordaremos as cadeias de Markov como um modelo básico utilizado para realizar transições entre diferentes estágios, destacando sua aplicação em matrizes de transição dentro de modelos de microsimulação.

3.1 Cadeias de Markov⁶

O modelo de cadeia de Markov é amplamente empregado para representar sistemas que evoluem de forma aleatória ao longo do tempo, consistindo em uma sequência de variáveis aleatórias, cada uma correspondendo ao estado do sistema em um instante específico. A principal característica desse modelo é que a probabilidade de transição para um estado

⁶Os teoremas e definições dessa seção são do livro "Probability and Statistics" (4th Edition) do DeGroot e Schervish, seção 3.10. As exceções são o teorema 3.3 e 3.4 que usei a formulação do livro "Introduction to Stochastic Processes with R" de Robert P. Dobrow, capítulo 2. A subseção de Simulação também é desse livro. Para a parte "Extending Markov Models: Time-Dependency" e para a prova do Teorema 3.4 foi usado o livro "Finite Markov Chains and Algorithmic Applications" de Olle Häggström, capítulo 3.

futuro depende apenas do estado atual, desconsiderando os estados anteriores — propriedade conhecida como hipótese de Markov (DeGroot e Schervish, 2012). Por exemplo, para transições no mercado de trabalho a probabilidade de transição de emprego para desemprego depende apenas do estado atual e não de todo o histórico no mercado de trabalho.

Esse tipo de modelo é particularmente útil na avaliação de estratégias de saúde pública, como programas de vacinação, ao considerar tanto os impactos de curto quanto de longo prazo. Dada a dificuldade de realizar ensaios clínicos que acompanhem efeitos prolongados, modelos de Markov oferecem uma abordagem alternativa para extrapolar dados clínicos observados em intervalos de tempo limitados. Esses modelos estimam a distribuição de uma população entre diferentes estados de saúde (como saudável, doente ou falecido) e avaliam as consequências econômicas e de saúde associadas a cada estado, possibilitando uma análise de custo-efetividade mais abrangente, tornando-se ferramentas valiosas para orientar decisões de saúde pública baseadas em evidências (Drabo e Padula, 2023).

Definição 3.1

*Uma sequência de variáveis aleatórias X_1, X_2, \dots é chamada de **processo estocástico** ou **processo aleatório com parâmetro de tempo discreto**. A primeira variável aleatória X_1 é chamada de **estado inicial do processo**; e para $n = 2, 3, \dots$, a variável aleatória X_n é chamada de **estado do processo no tempo n** .*

Cada variável aleatória em um processo estocástico possui uma distribuição marginal; ou seja, cada observação, como X_1 ou X_2 , tem sua própria distribuição marginal, que reflete a probabilidade de o paciente estar em um estado específico naquele momento.

O processo completo, representado por todas as observações (X_1, X_2, \dots, X_n) , possui uma distribuição conjunta. Esta distribuição conjunta considera as transições e as probabilidades associadas a cada sequência possível de estados ao longo do tempo, definindo assim a probabilidade de observar uma sequência específica de estados para um paciente ao longo das diferentes observações.

Definição 3.2

*Seja S um conjunto discreto. Uma **Cadeia de Markov** é uma sequência de variáveis aleatórias X_0, X_1, \dots , que assumem valores em S com a propriedade de que*

$$P(X_{n+1} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) = P(X_{n+1} = j \mid X_n = i),$$

*para todos $x_0, \dots, x_{n-1}, i, j \in S$, e $n \geq 0$. O conjunto S é o espaço de estados da cadeia de Markov. Uma cadeia de Markov é chamada **finita** se o conjunto S tem um número finito de estados.*

Uma hipótese implícita na definição anterior é que, em uma cadeia de Markov, a probabilidade de transição para um estado futuro depende apenas do estado atual, desconsiderando o histórico completo de estados anteriores.

Uma cadeia de Markov é *tempo-homogênea* se as probabilidades na definição (3.2) não dependem de n . Isto é,

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i),$$

para todo $n \geq 0$.

Esse tipo de cadeia de Markov é uma simplificação útil em várias situações. Pense em um jogo de tabuleiro como o Monopoly (ou Banco Imobiliário) com 10 casas, onde cada jogador se movimenta jogando um dado normal. A probabilidade de você ir para a décima casa dado que está na sétima é a mesma, independentemente de ser a primeira ou a enésima vez que você se encontra nessa casa.

Definição 3.3

Considere uma cadeia de Markov finita com distribuições de transição estacionárias dadas por $p_{ij} = \Pr(X_{n+1} = j \mid X_n = i)$ para todos n, i, j . A **matriz de transição** da cadeia de Markov é definida como a matriz $k \times k$ com elementos p_{ij} . Assim,

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ p_{21} & \cdots & p_{2k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}.$$

Essa matriz de transição satisfaz algumas propriedades como:

1. $p_{ij} \geq 0$ para todos i, j .

2. Para cada linha i ,

$$\sum_j p_{ij} = 1.$$

Exemplo: Estado de Saúde dos Pacientes

Considere um hospital onde o estado de saúde dos pacientes internados pode ser classificado em quatro possíveis categorias: *Saudável*, *Doente*, *Mais Doente* e *Morte*. Em termos matemáticos, nosso espaço de estados da cadeia de Markov é

$$S = \{\text{Saudável}, \text{Doente}, \text{Mais Doente}, \text{Morte}\}.$$

Em intervalos regulares de 24 horas, o estado de saúde dos pacientes é monitorado e registrado.

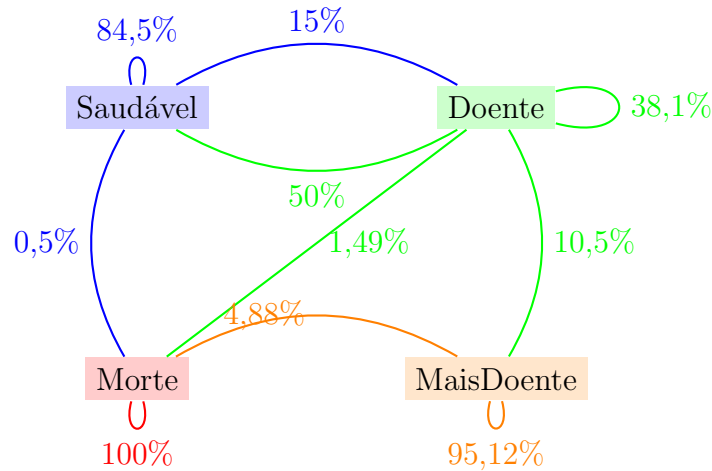
Denotemos por X_1 o estado de saúde do paciente na primeira observação, realizada no início do período de monitoramento. Da mesma forma, X_2 representa o estado de saúde na segunda observação, ocorrida 24 horas depois, e assim por diante; em termos gerais, para $n = 1, 2, \dots$, X_n indica o estado de saúde do paciente na n -ésima observação.

A matriz de transição é:

$$P = \begin{bmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Interpretamos essa matriz da seguinte forma: as linhas representam os estados atuais da pessoa, enquanto as colunas indicam os estados para os quais a pessoa poderá transitar no próximo período. Por exemplo, $p_{12} = \Pr(X_{n+1} = 2 \mid X_n = 1) = 0.15$ representa a probabilidade de uma pessoa no estado Saudável se tornar Doente.

Uma outra maneira de representar essa matriz de transição é através de um grafo que representa um modelo de transição entre diferentes estados de saúde de um paciente em um hospital. Cada nó (círculo) indica um estado de saúde possível — Saudável, Doente, Mais Doente e Morto — enquanto as setas entre eles mostram as probabilidades de transição de um estado para outro ao longo do tempo.



A Matriz de Transição para Vários Passos

Anteriormente, quando tínhamos uma matriz de transição, podíamos calcular a probabilidade de estarmos em um estado i no passo n e irmos para um estado j no passo $n + 1$. No entanto, se quisermos saber a probabilidade de chegar ao estado j no, por exemplo, passo

$n + 5$, precisamos calcular uma matriz de transição que contemple vários passos à frente, e não apenas um único passo.

Teorema 3.1

Seja P a matriz de transição de uma cadeia de Markov finita com distribuições de transição estacionárias. Para cada $m = 2, 3, \dots$, a m -ésima potência P^m da matriz P tem na linha i e coluna j a probabilidade $p_{ij}^{(m)} = \Pr(X_{m+n} = j | X_n = i)$ de que a cadeia se mova do estado i para o estado j em m passos.

Definição 3.4

Sob as condições do Teorema 3.2, a matriz P^m é chamada de **matriz de transição de m passos** da cadeia de Markov.

Ou seja, podemos ir do estado i no passo n para o estado j no passo $n + 2$ realizando uma multiplicação matricial $PP = P^2$, em que $p_{ij}^{(2)}$ será o elemento na i -ésima linha e j -ésima coluna de P^2 .

Por exemplo, considere a seguinte matriz de transição P de 2×2 :

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Para calcular a probabilidade de estar em um determinado estado após dois passos, fazemos $P^2 = P \times P$:

$$P^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \times \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$

Aqui, o elemento $p_{12}^{(2)} = 0.39$ representa a probabilidade de transitar do estado 1 para o estado 2 em dois passos.

Da mesma maneira, se consideramos a transição do estado i para o estado j em três passos, realizamos outra multiplicação matricial $PPP = P^3$. Nesse caso:

$$P^3 = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} \times \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.583 & 0.417 \\ 0.556 & 0.444 \end{bmatrix}$$

Assim, o elemento $p_{12}^{(3)} = 0.417$ em P^3 indica a probabilidade de estar no estado 1 no passo n e transitar para o estado 2 no passo $n + 3$. Esses cálculos ilustram como as probabilidades de transição para múltiplos passos podem ser obtidas através de sucessivas multiplicações da matriz de transição.

Exemplo: Matriz de transição para 2 passos e 4 estados

Um médico monitora o estado de saúde de um paciente internado a cada 24 horas, verificando se ele está em um dos estados possíveis: *Saudável* (1), *Doente* (2), *Mais Doente* (3), ou *Morto* (4). Ele modela a saúde do paciente como uma cadeia de Markov.

O médico percebe que, no dia seguinte, terá que se ausentar por 48 horas e, portanto, perderá duas verificações consecutivas do estado do paciente. Ele deseja calcular a distribuição condicional do estado do paciente para dois períodos de tempo à frente, dado cada um dos estados possíveis. O raciocínio é o seguinte: se o paciente estiver no estado *Doente* no momento $X_n = 2$, por exemplo, o estado no tempo $n + 2$ poderá ser qualquer um dos quatro estados, mas com probabilidades diferentes. Para calcular a distribuição condicional, ele pode somar as probabilidades da matriz de transição para dois períodos adiante, considerando o estado inicial em $X_n = 2$. Ou seja, o médico quer calcular a seguinte probabilidade $P(X_{n+2} = j | X_n = \text{Doente})$. Retornando ao exemplo, suponha que o médico deseje saber se, após 48 horas, o paciente estará saudável. Para obter a distribuição condicional de X_{n+2} dado $X_n = \text{Doente}$, somamos sobre os valores possíveis de X_{n+1} .

$$\begin{aligned}
 PP &= \begin{bmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 0.789 & 0.183 & 0.0157 & 0.0114 \\ 0.613 & 0.220 & 0.139 & 0.028 \\ 0 & 0 & 0.904 & 0.095 \\ 0 & 0 & 0 & 1 \end{bmatrix}
 \end{aligned}$$

Assim, segue que $\Pr(X_{n+2} = \text{Saudável} | X_n = \text{Doente}) = 0,4225 + 0,1905 + 0 = 0,613$

Vetor de Probabilidade/Distribuição Inicial

Um médico em uma unidade de terapia intensiva verifica o estado de saúde de um paciente e estima que há uma probabilidade de 0,5 de que o paciente esteja no estado *Saudável* no momento da primeira observação. Além disso, há uma probabilidade de 0,5 de que o paciente esteja *Doente*, uma probabilidade de 0 de que esteja *Mais Doente* e uma probabilidade de 0 de que o paciente esteja *Morto*. Esses valores especificam a distribuição marginal do estado do paciente no momento 0, X_0 . Podemos representar essa distribuição pelo vetor $\alpha^0 = (0.5, 0.5, 0, 0)$, que fornece as probabilidades dos quatro estados no momento inicial, na mesma ordem em que aparecem na matriz de transição. O vetor que fornece a distribuição marginal de X_1 neste exemplo tem um significado especial, pois representa as condições

iniciais do paciente em relação aos estados de saúde possíveis.

Um vetor composto por números não negativos que somam 1 é chamado de vetor de probabilidade. Um vetor de probabilidade cujas coordenadas especificam as probabilidades de que uma cadeia de Markov estará em cada um de seus estados no tempo 0 é chamado de **distribuição inicial da cadeia** ou **vetor de probabilidade inicial**.

De maneira similar, para $n > 0$, vamos denotar os vetores linha $\alpha^{(1)}, \alpha^{(2)}, \dots$ como as distribuições da cadeia de Markov nos tempos 1, 2, \dots , de modo que

$$\alpha^{(n)} = \left(\alpha_1^{(n)}, \alpha_2^{(n)}, \dots, \alpha_k^{(n)} \right) = (\mathbb{P}(X_n = x_1), \mathbb{P}(X_n = x_2), \dots, \mathbb{P}(X_n = x_k)).$$

Em que para $j \in \{1, \dots, k\}$, $x_j \in S$.

Com a distribuição inicial e a matriz de transição, podemos calcular todas as distribuições marginais $\alpha^{(1)}, \alpha^{(2)}, \dots$ da cadeia de Markov.

Teorema 3.2

Seja X_0, X_1, \dots uma cadeia de Markov com matriz de transição P e distribuição inicial α^0 . Para todo $n \geq 0$, a distribuição de X_n é αP^n . Ou seja,

$$\alpha^n = \alpha^0 P^n$$

7

Exemplo. Considere um paciente que pode estar em um dos quatro estados de saúde: *Saudável* (1), *Doente* (2), *Mais Doente* (3), ou *Morto* (4). Suponha que a distribuição inicial do paciente seja de 50% no estado *Saudável* e 50% no estado *Doente*. Queremos encontrar a probabilidade de o paciente estar *Saudável* 2 dias depois.

A matriz de transição é:

$$P = \begin{pmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Calculamos P^2 (a matriz de transição em dois passos):

$$P^2 = \begin{pmatrix} 0.789 & 0.183 & 0.015 & 0.011 \\ 0.613 & 0.220 & 0.139 & 0.028 \\ 0 & 0 & 0.904 & 0.095 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

⁷A prova desse teorema se encontra no Apêndice

Com a distribuição inicial $\alpha = (0.5, 0.5, 0, 0)$, temos:

$$\alpha P^2 = (0.5, 0.5, 0, 0) \begin{pmatrix} 0.789 & 0.183 & 0.015 & 0.011 \\ 0.613 & 0.220 & 0.139 & 0.028 \\ 0 & 0 & 0.904 & 0.095 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (0.672, 0.265, 0.052, 0.009).$$

Portanto, a probabilidade de o paciente estar *Saudável* após 2 dias é $\Pr(X_2 = \text{Saudável}) = (\alpha P^2)_{\text{Saudável}} = 0.672$.

Além da distribuição marginal, com uma distribuição inicial podemos chegar à distribuição conjunta:

Teorema 3.3

Seja X_0, X_1, \dots uma cadeia de Markov com matriz de transição P e distribuição inicial α^0 .

Para todos $0 \leq n_1 < n_2 < \dots < n_{k-1} < n_k$ e estados $i_1, i_2, \dots, i_{k-1}, i_k$,

$$\begin{aligned} \Pr(X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_{k-1}} = i_{k-1}, X_{n_k} = i_k) \\ = (\alpha^0 P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} \dots (P^{n_k - n_{k-1}})_{i_{k-1} i_k}. \end{aligned}$$

8

Exemplo. Considere uma cadeia de Markov com quatro estados que representam a saúde dos pacientes: "Saudável", "Doente", "Mais Doente" e "Morte". A matriz de transição P é:

$$P = \begin{pmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Assumimos a seguinte distribuição inicial para a cadeia de Markov:

$$\alpha^0 = [0.5, 0.5, 0, 0]$$

onde:

- 50% dos pacientes começam no estado "Saudável",
- 50% começam no estado "Doente",

⁸A prova desse teorema se encontra no Apêndice

- 0% começam no estado “Mais Doente”,
- 0% começam no estado “Morte”.

Queremos calcular a probabilidade conjunta de que:

- 1. O paciente esteja no estado “Saudável” no tempo 2 ($X_2 = \text{Saudável}$)
- 2. Esteja no estado “Doente” no tempo 4 ($X_4 = \text{Doente}$),
- 3. E volte ao estado “Saudável” no tempo 6 ($X_6 = \text{Saudável}$).

Usamos a fórmula:

$$P(X_2 = \text{Saudável}, X_4 = \text{Doente}, X_6 = \text{Saudável}) = (\alpha^0 P^2)_{\text{Saudável}} \cdot (P^2)_{\text{Saudável, Doente}} \cdot (P^2)_{\text{Doente, Saudável}}$$

- Passo 1: Cálculo de P^2

Calculamos $P^2 = P \times P$:

$$P^2 = \begin{pmatrix} 0.789 & 0.183 & 0.015 & 0.011 \\ 0.613 & 0.220 & 0.139 & 0.028 \\ 0 & 0 & 0.904 & 0.095 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Passo 2: Calcular $\alpha^0 P^2$ para o tempo 2

Multiplicamos α^0 por P^2 para obter a distribuição no tempo 2:

$$\alpha^0 P^2 = [0.5, 0.5, 0, 0] \times P^2 = [0.672, 0.265, 0.052, 0.009]$$

Portanto, a probabilidade de estar no estado “Saudável” no tempo 2 é:

$$(\alpha^0 P^2)_{\text{Saudável}} = 0.672$$

- Passo 3: Cálculo de $(P^2)_{\text{Saudável, Doente}}$ e $(P^2)_{\text{Doente, Saudável}}$

A probabilidade de transição de “Saudável” para “Doente” em dois passos é:

$$(P^2)_{\text{Saudável, Doente}} = 0.183$$

A probabilidade de transição de "Doente" para "Saudável" em dois passos é:

$$(P^2)_{\text{Doente, Saudável}} = 0.613$$

- Passo 4: Cálculo da Probabilidade Conjunta

Finalmente, a probabilidade conjunta é dada por:

$$P(X_2 = \text{Saudável}, X_4 = \text{Doente}, X_6 = \text{Saudável}) = 0.672 \times 0.1839 \times 0.613 = 0.075$$

Portanto, a probabilidade de que o paciente esteja "Saudável" no tempo 2, "Doente" no tempo 4 e volte a estar "Saudável" no tempo 6 é aproximadamente 0.075 ou 7,5%.

Estendendo modelos de Markov: dependência temporal

Assumir a homogeneidade no tempo nem sempre é adequado porque, em muitos contextos, as probabilidades de transição entre estados podem variar ao longo do tempo. Por exemplo, em problemas de saúde, a evolução do estado de saúde de um paciente é influenciada pelo tempo. No início da internação, pode haver uma maior probabilidade de recuperação (transição para "Melhorando"), enquanto, com o passar do tempo, a probabilidade de piora pode aumentar devido à progressão da doença ou complicações associadas. A suposição de tempo-homogeneidade não captura essa dinâmica temporal.

Para acomodar situações onde as probabilidades de transição dependem do tempo, estendemos os modelos de Markov para considerar a dependência temporal.

Definição 2.2 Seja $P^{(1)}, P^{(2)}, \dots$ uma sequência de matrizes $k \times k$, cada uma das quais satisfaz os pontos 1 e 2 da definição 3.3. Um processo estocástico (X_0, X_1, \dots) com espaço de estados finito $S = \{s_1, \dots, s_k\}$ é dito ser uma cadeia de Markov não homogênea com matrizes de transição $P^{(1)}, P^{(2)}, \dots$, se para todo n , todos $i, j \in \{1, \dots, k\}$ e todos $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ temos

$$\begin{aligned} \mathbb{P}(X_{n+1} = s_j \mid X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) \\ = \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i) \\ = p_{i,j}^{(n+1)}. \end{aligned}$$

Exemplo: Progressão de uma Doença com Aumento de Gravidade e Custo

Doenças crônicas, como a diabetes tipo 2 ou a hipertensão arterial, tendem a se agravar ao longo do tempo se não forem gerenciadas adequadamente, resultando em sintomas mais

severos e custos de tratamento mais elevados. Vamos modelar a evolução da saúde de uma pessoa em três estados:

- s_1 : Saudável – sem sinais ou sintomas da doença.
- s_2 : Doente Inicial – sintomas leves, controláveis com tratamento básico.
- s_3 : Doente Avançado – sintomas graves, requerendo intervenções complexas e dispendiosas.

A probabilidade de transição entre esses estados depende do tempo que o paciente permanece no estado “Doença Inicial”, refletindo o agravamento progressivo associado à falta de controle efetivo da condição. Considere as seguintes matrizes de transição para dois cenários: Curto Prazo (início da doença) e Longo Prazo (progressão prolongada sem controle adequado):

$$P_{\text{curto prazo}} = \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.2 & 0.75 & 0.05 \\ 0.05 & 0.15 & 0.8 \end{bmatrix} \quad \text{e} \quad P_{\text{longo prazo}} = \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0.9 \end{bmatrix}$$

• **Matriz de Curto Prazo ($P_{\text{curto prazo}}$):**

- A probabilidade de um indivíduo permanecer “Saudável”(s_1) é alta (95%), com uma chance de 5% de desenvolver a “Doença Inicial”(s_2).
- No estado “Doença Inicial”, há 20% de chance de recuperação para o estado “Saudável”, 75% de permanecer no mesmo estado e 5% de progressão para “Doença Avançada”(s_3).
- No estado “Doença Avançada”, o paciente tem 80% de chance de permanecer nesse estado, com 15% de possibilidade de regressão para “Doença Inicial” e 5% de recuperação completa.

• **Matriz de Longo Prazo ($P_{\text{longo prazo}}$):**

- A probabilidade de permanecer “Saudável”(s_1) continua alta (95%).
- Para pacientes no estado “Doença Inicial”, o risco de progressão para “Doença Avançada” aumenta significativamente para 30%, refletindo o agravamento ao longo do tempo sem intervenção adequada.
- No estado “Doença Avançada”, a probabilidade de permanecer é muito alta (90%), indicando a dificuldade de reversão nesta fase avançada da doença.

Este modelo é útil para representar a progressão de doenças crônicas, onde o estado "Doente Avançado" implica um cuidado intensivo e custos elevados, enquanto o "Doente Inicial" oferece melhores chances de recuperação com intervenção rápida. O modelo ilustra como a duração da doença afeta a gravidade e os custos de tratamento.

A extensão da distribuição marginal para o caso de modelos markovianos para casos não homogêneos se dá pelo seguinte teorema:

Teorema 3.4

Suponha que (X_0, X_1, \dots) seja uma cadeia de Markov não-homogênea com espaço de estados $\{s_1, \dots, s_k\}$, distribuição inicial $\alpha^{(0)}$ e matrizes de transição $P^{(1)}, P^{(2)}, \dots$. Para qualquer n , temos então que

$$\alpha^{(n)} = \alpha^{(0)} P^{(1)} P^{(2)} \dots P^{(n)}.$$

Isso indica que a distribuição no tempo n ($\alpha^{(n)}$) é obtida multiplicando a distribuição inicial pelas matrizes de transição até o tempo n em sequência.

Exemplo: Aplicação do Teorema ao Modelo de Progressão da Doença

Assumimos a distribuição inicial ($n = 0$), onde todos os indivíduos estão saudáveis:

$$\alpha^{(0)} = [1, 0, 0]$$

Para o primeiro período ($n = 1$), usamos a matriz de curto prazo $P_{\text{curto prazo}}$. A partir do segundo período ($n = 2$), aplicamos a matriz de longo prazo $P_{\text{longo prazo}}$.

- Período 1 ($n = 1$)

$$\alpha^{(1)} = \alpha^{(0)} P_{\text{curto prazo}} = [1, 0, 0] \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.2 & 0.75 & 0.05 \\ 0.05 & 0.15 & 0.8 \end{bmatrix} = [0.95, 0.05, 0]$$

Após o primeiro período, 95% permanecem saudáveis, enquanto 5% progridem para o estado "Doente Inicial".

- Período 2 ($n = 2$)

Para o segundo período, aplicamos a matriz de longo prazo $P_{\text{longo prazo}}$:

$$\alpha^{(2)} = \alpha^{(1)} P_{\text{longo prazo}} = [0.95, 0.05, 0] \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0.9 \end{bmatrix}$$

Portanto,

$$\alpha^{(2)} = [0.9075, 0.0775, 0.015]$$

Após dois períodos, cerca de 90.75% permanecem saudáveis, 7.75% estão no estado "Doente Inicial" e 1.5% no estado "Doente Avançado".

3.2 Simulação

Muitas vezes, expressões analíticas para as matrizes de transição em cadeias de Markov podem se tornar complexas, especialmente quando o número de estados é grande ou quando se deseja calcular a distribuição após muitos passos. Nesse contexto, a simulação se apresenta como uma ferramenta poderosa, permitindo analisar o comportamento de cadeias de Markov de forma prática e eficiente. O algoritmo abaixo descreve como simular uma cadeia de Markov a partir de uma distribuição inicial, uma matriz de transição e o número desejado de passos.

Entrada: (i) distribuição inicial α , (ii) matriz de transição P , (iii) número de passos n .

Saída: X_0, X_1, \dots, X_n

Algoritmo:

- Gere X_0 de acordo com α .
- **PARA** $i = 1, \dots, n$
 - Assuma que $X_{i-1} = j$.
 - Defina p como a j -ésima linha de P .
 - Gere X_i de acordo com p .
- **FIM**

A implementação em R deste algoritmo utiliza a função `sample` para simular as transições de estados na Cadeia de Markov. A função `sample` permite escolher um próximo estado aleatoriamente a partir de um conjunto de probabilidades.

Para exemplificar, considere uma cadeia com três estados: "Saudável", "Doente" e "Hospitalizado". Suponha que a matriz de transição define a probabilidade de um paciente passar de um estado para outro, como "Saudável" para "Doente" com probabilidade de 10%, e "Doente" para "Hospitalizado" com probabilidade de 30%:

- **Início da Simulação:** O primeiro estado, X_0 , é gerado com base na distribuição inicial α , que representa a probabilidade de o processo começar em cada estado possível. Para

isso, usamos a função `sample` para selecionar um estado inicial de acordo com essas probabilidades.

Início da Simulação

```
# Definindo os estados com seus nomes
states <- c("Saudável", "Doente", "Hospitalizado")

# Distribuição inicial (probabilidades para cada estado)
alpha <- c(Saudável = 0.6, Doente = 0.3, Hospitalizado =
  0.1)

# Matriz de transição
P <- matrix(c(
  0.8, 0.1, 0.1,    # Probabilidades de transição a partir
    de Saudável
  0.2, 0.5, 0.3,    # Probabilidades de transição a partir
    de Doente
  0.1, 0.3, 0.6     # Probabilidades de transição a partir
    de Hospitalizado
), nrow = 3, byrow = TRUE)

# Nomear as linhas e colunas da matriz de transição para
  facilitar a leitura
rownames(P) <- states
colnames(P) <- states

# Número de passos (vamos fazer manualmente, então começ
  amos com apenas o estado inicial)
n <- 5

# Inicializando a simulação
set.seed(123) # Para resultados reproduzíveis
X <- character(n + 1)

# Passo 0: Estado inicial baseado na distribuição alpha
X[1] <- sample(states, 1, prob = alpha)
cat("Passo 0 (Estado inicial):", X[1], "\n")
Passo 0 (Estado inicial): Saudável
```

- Transições no Loop: Em cada passo i da simulação, a partir do estado atual X_{i-1} , a

matriz de transição P define as probabilidades para transições entre estados. A função seleciona o próximo estado X_i de acordo com a linha correspondente ao estado atual X_{i-1} na matriz P , permitindo uma escolha probabilística do próximo estado conforme a estrutura da cadeia de Markov.

Transições no Loop

```
# Passo 1: Primeira transição
current_state <- X[1]
transition_prob <- P[current_state, ]
X[2] <- sample(states, 1, prob = transition_prob)
cat("Passo 1:", X[2], "\n")
Passo 1: Saudável

# Passo 2: Segunda transição
current_state <- X[2]
transition_prob <- P[current_state, ]
X[3] <- sample(states, 1, prob = transition_prob)
cat("Passo 2:", X[3], "\n")
Passo 2: Saudável

# Passo 3: Terceira transição
current_state <- X[3]
transition_prob <- P[current_state, ]
X[4] <- sample(states, 1, prob = transition_prob)
cat("Passo 3:", X[4], "\n")
Passo 3: Hospitalizado

# Passo 4: Quarta transição
current_state <- X[4]
transition_prob <- P[current_state, ]
X[5] <- sample(states, 1, prob = transition_prob)
cat("Passo 4:", X[5], "\n")
Passo 4: Saudável
```

Transições no Loop(Código compacto)

```
# Simulando os passos da cadeia de Markov
for (i in 2:(n + 1)) {
  current_state <- X[i - 1]
  transition_prob <- P[current_state, ]
  X[i] <- sample(states, 1, prob = transition_prob)
}

# Exibindo o resultado final
print(X)
[1] "Saudável"      "Saudável"      "Saudável"      "
     Hospitalizado" "Saudável"
```

Uma cadeia de Markov não homogênea pode ser simulada com uma adaptação no algoritmo original, permitindo que as probabilidades de transição variem a cada passo. Em vez de uma matriz de transição fixa P , temos uma sequência de matrizes P_i , cada uma específica para o passo i , refletindo a natureza não homogênea da cadeia.

Entrada: (i) distribuição inicial α , (ii) sequência de matrizes de transição $\{P_1, P_2, \dots, P_n\}$, onde cada matriz P_i corresponde ao passo i , (iii) número de passos n .

Saída: X_0, X_1, \dots, X_n

Algoritmo:

- Gere X_0 de acordo com α .
- **PARA** $i = 1, \dots, n$
 - Assuma que $X_{i-1} = j$.
 - Defina p como a j -ésima linha da matriz P_i , que corresponde ao passo atual i .
 - Gere X_i de acordo com p .
- **FIM**

A implementação em R deste algoritmo não homogêneo utiliza novamente a função `sample`, que permite simular transições de estados variáveis ao longo do tempo, baseando-se em uma lista de matrizes de transição que mudam em cada estado.

- **Início da Simulação:** O primeiro estado, X_0 , é gerado com base na distribuição inicial α , assim como em uma cadeia homogênea. A função `sample` é usada para selecionar este estado de forma probabilística, com cada estado possível sendo ponderado pela probabilidade em α .

Início da Simulação

```
# Definindo os estados com seus nomes
states <- c("Saudável", "Doente", "Hospitalizado")

# Distribuição inicial (probabilidades para cada estado)
alpha <- c(Saudável = 0.6, Doente = 0.3, Hospitalizado =
  0.1)

# Definindo múltiplas matrizes de transição (não homogêneas
  ) - uma para cada passo
P1 <- matrix(c(
  0.8, 0.1, 0.1, # Transições a partir de Saudável
  0.2, 0.5, 0.3, # Transições a partir de Doente
  0.1, 0.3, 0.6  # Transições a partir de Hospitalizado
), nrow = 3, byrow = TRUE, dimnames = list(states, states))

P2 <- matrix(c(
  0.7, 0.2, 0.1, # Transições a partir de Saudável
  0.3, 0.4, 0.3, # Transições a partir de Doente
  0.2, 0.3, 0.5  # Transições a partir de Hospitalizado
), nrow = 3, byrow = TRUE, dimnames = list(states, states))

P3 <- matrix(c(
  0.6, 0.3, 0.1, # Transições a partir de Saudável
  0.4, 0.3, 0.3, # Transições a partir de Doente
  0.3, 0.2, 0.5  # Transições a partir de Hospitalizado
), nrow = 3, byrow = TRUE, dimnames = list(states, states))
```

Início da Simulação(continuação)

```
P4 <- matrix(c(
  0.5, 0.3, 0.2, # Transições a partir de Saudável
  0.5, 0.2, 0.3, # Transições a partir de Doente
  0.4, 0.2, 0.4  # Transições a partir de Hospitalizado
), nrow = 3, byrow = TRUE, dimnames = list(states, states))

P5 <- matrix(c(
  0.4, 0.4, 0.2, # Transições a partir de Saudável
  0.6, 0.2, 0.2, # Transições a partir de Doente
  0.5, 0.1, 0.4  # Transições a partir de Hospitalizado
), nrow = 3, byrow = TRUE, dimnames = list(states, states))

# Lista de matrizes de transição, uma para cada passo
P_list <- list(P1, P2, P3, P4, P5)

# Número de passos
n <- 5

# Inicializando a simulação
set.seed(123) # Para resultados reproduzíveis
X <- character(n + 1)

# Passo 0: Estado inicial baseado na distribuição alpha
X[1] <- sample(states, 1, prob = alpha)
cat("Passo 0 (Estado inicial):", X[1], "\n")
Passo 0 (Estado inicial): Saudável
```

- **Transições Não Homogêneas no Loop:** Em cada passo i , é utilizada uma nova matriz de transição P_i , cuja estrutura varia ao longo do tempo para capturar mudanças específicas de uma cadeia de Markov não homogênea. Cada linha de P_i representa as probabilidades de transição de um estado para os demais. A seleção do próximo estado X_i é baseada na linha correspondente ao estado atual X_{i-1} em P_i , permitindo uma escolha probabilística e adaptativa do próximo estado de acordo com as transições variáveis de cada passo.

Transições Não Homogêneas no Loop:

```
# Passo 1: Transição usando a matriz P1
current_state <- X[1]
transition_prob <- P1[current_state, ]
X[2] <- sample(states, 1, prob = transition_prob)
cat("Passo 1 (usando P1):", X[2], "\n")
Passo 1 (usando P1): Saudável

# Passo 2: Transição usando a matriz P2
current_state <- X[2]
transition_prob <- P2[current_state, ]
X[3] <- sample(states, 1, prob = transition_prob)
cat("Passo 2 (usando P2):", X[3], "\n")
Passo 2 (usando P2): Saudável

# Passo 3: Transição usando a matriz P3
current_state <- X[3]
transition_prob <- P3[current_state, ]
X[4] <- sample(states, 1, prob = transition_prob)
cat("Passo 3 (usando P3):", X[4], "\n")
Passo 3 (usando P3): Doente

# Passo 4: Transição usando a matriz P4
current_state <- X[4]
transition_prob <- P4[current_state, ]
X[5] <- sample(states, 1, prob = transition_prob)
cat("Passo 4 (usando P4):", X[5], "\n")
Passo 4 (usando P4): Doente
```

Transições Não Homogêneas no Loop:

```
# Passo 5: Transição usando a matriz P5
current_state <- X[5]
transition_prob <- P5[current_state, ]
X[6] <- sample(states, 1, prob = transition_prob)
cat("Passo 5 (usando P5):", X[6], "\n")
Passo 5 (usando P5): Saudável

# Resultado final após todas as transições
cat("Sequência final de estados:", X, "\n")
Sequência final de estados: Saudável Saudável Saudável
Doente Doente Saudável
```

Transições Não Homogêneas no Loop(Código Compacto):

```
# Passos 1 a n: Transições não homogêneas
for (i in 1:n) {
  current_state <- X[i]
  transition_prob <- P_list[[i]][current_state, ]
  X[i + 1] <- sample(states, 1, prob = transition_prob)
  cat("Passo", i, "com matriz P_", i, ":", X[i + 1], "\n")
}
Passo 1 com matriz P_ 1 : Saudável
Passo 2 com matriz P_ 2 : Saudável
Passo 3 com matriz P_ 3 : Doente
Passo 4 com matriz P_ 4 : Doente
Passo 5 com matriz P_ 5 : Saudável
>

# Resultado final após todas as transições
cat("Sequência final de estados:", X, "\n")
Sequência final de estados: Saudável Saudável Saudável
Doente Doente Saudável
```

4 Exemplos

Nas últimas seções, discutimos ferramentas úteis para realizar microssimulações, como os modelos de escolha discreta e as transições com cadeias de Markov. Agora, nesta seção, vamos retomar os exemplos que foram brevemente citados e aplicar as microssimulações na prática com os diferentes métodos que foram explicados.

4.1 Taxação de importados

Utilizando dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) e o software R, realizamos uma microssimulação para estimar o impacto da introdução de uma taxa de 20% sobre as compras internacionais de até 50 dólares na renda dos brasileiros. Adotamos um modelo estático para capturar o impacto imediato da taxa, ou seja, antes que os consumidores ajustem seus comportamentos em resposta à nova política tributária.

Os dados utilizados são do quarto trimestre de 2023 da PNADc, que fornece uma amostra representativa da distribuição de renda no Brasil. Estima-se que, em um trimestre, os brasileiros realizem cerca de 13 milhões de compras internacionais nesse valor, com um gasto médio de R\$79,30 por compra.⁹ A taxa de 20% sobre esse valor representa um custo adicional médio de R\$15,86 por compra. Além disso, o ICMS de 17% incide sobre o valor total (compra + taxa), elevando o impacto total da nova taxa para R\$18,56 por compra¹⁰.

Sabemos que nem todos os brasileiros fazem compras internacionais. Aproximadamente 30% da população não participa desse tipo de consumo. Além disso, o padrão de consumo varia entre as classes sociais: as classes A e D compram proporcionalmente à sua participação populacional, enquanto as classes B e C compram o dobro da média, e a classe E compra menos¹¹. Para refletir esses padrões, ajustamos o modelo para considerar diferentes probabilidades de compra entre as faixas de renda.

Para simular os impactos, utilizamos os pesos amostrais da PNADc, que indicam a representatividade de cada indivíduo na população brasileira. Esses pesos nos permitem distribuir as 13 milhões de compras de forma proporcional à população representada, ajustando a quantidade de compras atribuída a cada pessoa de acordo com sua renda e peso amostral.

Com esses dados, calculamos o impacto da taxa de importação em termos da proporção da renda que seria comprometida com o pagamento da taxa. Para o “brasileiro médio”, ou seja, uma pessoa representativa da média ponderada da população, o impacto estimado da taxa é de 3,63% da renda mensal. Analisamos também o efeito da taxa em diferentes faixas

⁹Tomazelli (2024)

¹⁰Pavan (2024)

¹¹*Qual classe social mais compra na Shein?* (2023)

de renda, como mostrado na Tabela 1. Observa-se que, quanto maior a renda, menor é o impacto proporcional da taxa. Além disso, verificamos os impactos por gênero e cor/raça, apresentados nas Tabelas 2 e 3, onde constatamos que mulheres são mais afetadas do que homens, e que pardos e indígenas enfrentam os maiores impactos.

Por fim, simulamos o efeito de uma taxa de 30%. Nesse cenário, o impacto no “brasileiro médio” seria de 5,45% da renda mensal, um aumento de 67% em relação à taxa de 20%. Esse aumento afetaria de maneira proporcional todas as faixas de renda, gêneros e grupos raciais, já que a taxa é aplicada uniformemente a todos.

Entretanto, é importante reafirmar as limitações desta metodologia. Racionalmente, espera-se que indivíduos reajam a mudanças como essa. Ou seja, se os preços de um grupo de bens aumentam por conta da incrementação de uma tarifa, é muito provável que os indivíduos passem a consumir menos esses produtos. Isso é especialmente relevante no contexto de diferenciar os efeitos por grupos de pessoas com características comuns, porque é possível, por exemplo, que pessoas negras reduzam mais o seu consumo do que pessoas brancas. Para uma análise mais precisa dos efeitos dessa política, é necessário fazer um modelo que leve em conta essas mudanças comportamentais.

Faixa de Renda	Impacto Médio da Taxa (%)
Classe E (até 1 s.m.)	4,40
Classe D (de 1 à 3 s.m.)	0,78
Classe C (de 3 à 5 s.m.)	3,55
Classe B (de 5 à 15 s.m.)	2,93
Classe A (mais de 15 s.m.)	2,82

Tabela 1: Impacto médio da taxa de importação de 20% sobre a renda, por faixa de renda

Gênero	Impacto Médio da Taxa (%)
Homem	3,32
Mulher	4,02

Tabela 2: Impacto médio da taxa de importação de 20% sobre a renda, por gênero

Cor/Raça	Impacto Médio da Taxa (%)
Branca	3,74
Preta	3,62
Amarela	4,13
Parda	3,51
Indígena	4,09

Tabela 3: Impacto médio da taxa de importação de 20% sobre a renda, por faixa de renda

4.2 Bolsa Família

Há um debate na literatura sobre os possíveis efeitos de programas de transferência de renda na oferta de trabalho dos beneficiários. No caso brasileiro, o Bolsa Família é frequentemente alvo dessas discussões, com alguns especialistas argumentando que, em seu formato atual, o programa pode penalizar beneficiários que conseguem um emprego (Mendes 2024). Nesse contexto, em 2023 o Bolsa Família passou a incluir uma regra de proteção aos beneficiários que conseguissem um emprego. A regra é disparada quando um beneficiário encontra um emprego formal e determina que, caso o salário advindo deste emprego faça com que a renda domiciliar per capita seja de até meio salário mínimo (R\$ 706,00), o beneficiário continuará recebendo 50% do valor do Bolsa Família por até dois anos. Até julho de 2024 havia 2,8 milhões de beneficiários na regra de proteção (Mali 2024).

Apesar disso, evidências mais robustas sobre o efeito do Bolsa Família na oferta de trabalho são limitadas. Por exemplo, em uma avaliação de um modelo anterior do benefício, Tavares (2010) encontrou que o efeito líquido sobre a oferta de trabalho de mães beneficiadas era positivo. Neste contexto, utilizamos os dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) e o software R para realizar uma microsimulação. Nosso objetivo é estimar como o aumento de R\$200,00 no valor mínimo do Bolsa Família impactaria a oferta de trabalho dos indivíduos. Para isso, empregamos um modelo comportamental projetado para analisar como reajustes na transferência de renda afetam a decisão de trabalhar ou não.

A tabela 4 apresenta estatísticas descritivas das variáveis que serão utilizadas no modelo. Entre os resultados destacados, chama atenção a elevada variância no valor do Bolsa Família recebido pelos domicílios. No entanto, essa aparente inconsistência pode ser explicada pela estrutura do programa. O Bolsa Família estabelece um valor mínimo fixo de R\$600,00, mas também inclui adicionais variáveis, como R\$150,00 por criança de até 6 anos, R\$50,00 para gestantes ou crianças/adolescentes de 7 a 17 anos, e R\$50,00 para bebês de até seis meses. Esses critérios, ao flexibilizar o valor final recebido, resultam em maior dispersão nos valores transferidos.

Tabela 4: Estatísticas descritivas da Base de Dados

Variável	Mínimo	Máximo	Média	Mediana	Desvio Padrão
Recebe Bolsa Família	0	1,00	0,09	0	0,28
Valor do Bolsa Família	45	2574,00	639,58	600	149,74
Renda Domiciliar per Capita	0	250.000,00	1763,93	1204	2591,21
Contribuição das Crianças para a Renda	0	1,14	0,00	0	0,00
Idade	14	115,00	44,27	43	18,86
Homem	0	1,00	0,48	0	0,50
Branco	0	1,00	0,40	0	0,49
Urbano	0	1,00	0,74	1	0,44
Anos de estudo	0	16,00	9,34	11	4,70

Fonte: PNADc, quinta entrevista de 2023 com 331.953 observações.

O aumento no valor de uma transferência de renda pode influenciar a oferta de trabalho por meio de um mecanismo principal: o efeito renda (Kaufman e Hotchkiss 2006). O efeito renda ocorre quando indivíduos, ao receberem uma renda maior, decidem trabalhar menos, aproveitando o aumento da renda disponível para dedicar mais tempo a outras atividades, como lazer ou cuidados familiares, conforme predito por modelos tradicionais de escolha entre consumo e lazer. É possível, também, que os indivíduos decidam trabalhar menos por receio de que um aumento em sua renda advinda do trabalho o faça perder o benefício financeiro, já que há um critério de renda para receber o Bolsa Família (ter renda domiciliar per capita igual ou menor à R\$ 218,00). Esse mecanismo é bastante particular para programas de transferência de renda, porque neles a renda adicional está desvinculada do trabalho.

Para modelar a decisão individual sobre participar do mercado de trabalho, utilizamos um modelo logit binomial. No modelo binomial, o indivíduo escolhe entre trabalhar ou não trabalhar com base no que maximiza sua utilidade esperada. A utilidade esperada para o indivíduo i do domicílio j que decide se trabalha ou não (onde k pode ser igual a 0 ou 1) é expressa por:

$$U_{ij}(w_{ij} = k) = \beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot D_j + \beta_3 \cdot X_i + \varepsilon_{ij},$$

onde:

- T_i representa o valor da transferência do Bolsa Família recebido pelo indivíduo i ;
- D_j são as características do domicílio j (por exemplo, zona urbana ou rural, renda domiciliar per capita e contribuição do trabalho infantil para a renda domiciliar);

- X_i são características individuais observáveis do trabalhador i (idade, gênero, cor e anos de estudo);
- ε_{ij} é o erro aleatório associado à decisão de trabalho.

A probabilidade de um indivíduo i do domicílio j optar por trabalhar ($w_{ij} = 1$) ou não trabalhar ($w_{ij} = 0$) é dada pela função de distribuição logística:

$$P(w_{ij} = 1|T_i, D_j, X_i) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot D_j + \beta_3 \cdot X_i))},$$

onde a probabilidade de não trabalhar ($w_{ij} = 0$) é dada por:

$$P(w_{ij} = 0|T_i, D_j, X_i) = 1 - P(w_{ij} = 1|T_i, D_j, X_i).$$

Alternativamente, o modelo probit pode ser utilizado como uma variação do modelo logit, substituindo a função logística pela distribuição normal cumulativa (Φ). No modelo probit, a decisão do indivíduo também é baseada na maximização da sua utilidade esperada. A equação para a probabilidade de um indivíduo i do domicílio j optar por trabalhar ($w_{ij} = 1$) ou não trabalhar ($w_{ij} = 0$) é dada pela função de distribuição normal cumulativa:

$$P(w_{ij} = 1|T_i, D_j, X_i) = \Phi(\beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot D_j + \beta_3 \cdot X_i),$$

onde Φ é a função de distribuição acumulada da normal padrão. A probabilidade de não trabalhar ($w_{ij} = 0$) é dada por:

$$P(w_{ij} = 0|T_i, D_j, X_i) = 1 - P(w_{ij} = 1|T_i, D_j, X_i).$$

Neste caso, a interpretação dos coeficientes β_i também é semelhante à do modelo logit, mas as estimativas são baseadas na normal cumulativa, o que resulta em uma modelagem ligeiramente diferente, especialmente em termos da escala da magnitude dos efeitos.

Para facilitar a interpretação dos resultados dos modelos logit e probit, utilizamos o Efeito Marginal Médio (ou *Average Marginal Effect* - AME). Essa medida é amplamente empregada para compreender os coeficientes em modelos não lineares, como logit e probit. Em modelos lineares, os coeficientes indicam diretamente o impacto marginal das variáveis explicativas sobre a variável dependente. Porém, em modelos não lineares, o impacto marginal depende dos valores das covariáveis, variando para diferentes indivíduos ou observações. O AME, portanto, calcula a média desses efeitos marginais, oferecendo uma interpretação mais intuitiva e geral do impacto das variáveis explicativas.

O AME calcula a média dos efeitos marginais individuais, avaliados em todas as observações da amostra. Isso resulta em uma interpretação prática do impacto médio que uma pequena variação em uma variável explicativa tem sobre as probabilidades previstas de um evento. Para o modelo logit binomial, o efeito marginal para a variável X_k é dado por:

$$\frac{\partial P(Y = 1)}{\partial X_k} = P(Y = 1) \cdot [1 - P(Y = 1)] \cdot \beta_k,$$

onde $P(Y = 1)$ é a probabilidade prevista de sucesso para uma observação, e β_k é o coeficiente associado à variável explicativa X_k .

No modelo probit, o cálculo do AME segue uma lógica similar, mas com base na função de distribuição normal cumulativa (Φ), ao invés da função logística. O efeito marginal no modelo ****probit**** para a variável X_k é dado por:

$$\frac{\partial P(Y = 1)}{\partial X_k} = \phi(\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k) \cdot \beta_k,$$

onde ϕ é a densidade de probabilidade da distribuição normal padrão. Assim, para o modelo ****probit****, o AME também pode ser calculado como a média dos efeitos marginais, avaliados em todas as observações da amostra:

$$\text{AME}_k = \frac{1}{N} \sum_{i=1}^N [\phi(\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k) \cdot \beta_k],$$

onde N é o número total de observações na amostra. O cálculo do AME para o modelo probit oferece uma interpretação similar à do modelo logit, mas com a diferenciação devido à função de distribuição normal.

Adicionalmente, podemos estimar um modelo multinomial. Neste caso, o modelo logit é ampliado para permitir mais de duas categorias na variável dependente, representando diferentes estados do mercado de trabalho: $w_{ij} \in \{0, 1, 2\}$, onde 0 indica que o indivíduo não trabalha, 1 que trabalha no setor informal, e 2 que trabalha no setor formal. A utilidade esperada para o indivíduo i do domicílio j ao escolher a alternativa k ($k = 0, 1, 2$) é dada por:

$$U_{ij}(w_{ij} = k) = \beta_{k0} + \beta_{k1} \cdot T_i + \beta_{k2} \cdot D_j + \beta_{k3} \cdot X_i + \varepsilon_{ijk}.$$

A probabilidade de o indivíduo i escolher a alternativa k é então modelada pela função de distribuição logística multinomial:

$$P(w_{ij} = k | T_i, D_j, X_i) = \frac{\exp(\beta_{k0} + \beta_{k1} \cdot T_i + \beta_{k2} \cdot D_j + \beta_{k3} \cdot X_i)}{\sum_{m=0}^2 \exp(\beta_{m0} + \beta_{m1} \cdot T_i + \beta_{m2} \cdot D_j + \beta_{m3} \cdot X_i)}.$$

O modelo logit multinomial permite estimar os efeitos diferenciados das variáveis explicativas sobre as probabilidades de escolha de cada alternativa, mantendo a independência das alternativas irrelevantes (IAI). Assim, os coeficientes β_{km} podem ser interpretados como os efeitos relativos das variáveis explicativas sobre a probabilidade de escolher a alternativa k em relação a uma categoria de referência (por exemplo, $w_{ij} = 0$).

Por fim, para facilitar a interpretação dos resultados, também calculamos os Efeitos Marginais Médios (*Average Marginal Effects* - AME), adaptados ao contexto multinomial.

A Tabela 5 apresenta os resultados estimados dos modelos logit e probit binomiais, além do Efeito Marginal Médio (AME) para ambos os casos. Como esperado, o AME do valor do Bolsa Família na decisão de oferta de trabalho é o mesmo nas duas especificações, o que reforça a consistência dos resultados. O valor de -0,001 do AME indica que, para cada aumento de R\$1 no valor do benefício do Bolsa Família, a probabilidade de o indivíduo decidir trabalhar diminui em 0,01% em média. Isso sugere que o aumento do benefício tem um efeito negativo sobre a oferta de trabalho. Entretanto, a Tabela 6 apresenta o resultado do logit multinomial, que vai na contramão dos modelos binomiais. O AME do Bolsa Família para o trabalho formal é de 0,0001 e para o trabalho informal é de 0,0007, o que significa que um aumento de R\$1 na transferência de renda aumenta a probabilidade do indivíduo trabalhar no mercado de trabalho formal em 0,01% e no informal em 0,07%.

Tabela 5: Resultados da Regressão Logística Binomial e Probabilística Binomial e Efeitos Marginais Médios (AME)

Variável	Logit		Probit	
	Coeficiente	AME	Coeficiente	AME
Bolsa Família (valor)	-0,0003*** (0,00002)	-0,0001 (0,0000)	-0,0002*** (0,00001)	-0,0001 (0,0000)
Urbano (zona do domicílio)	0,0001*** (0,00000)	-0,0049 (0,0020)	-0,009 (0,006)	-0,0033 (0,0020)
Renda domiciliar per capita	0,998 (0,897)	0,0000 (0,0000)	0,0001*** (0,00000)	0,0000 (0,0000)
Contribuição das crianças	-0,012*** (0,0002)	0,2127 (0,1912)	-0,598 (0,540)	0,2100 (0,1896)
Idade	0,935*** (0,008)	-0,0026 (0,0000)	-0,007*** (0,0001)	-0,0025 (0,0000)
Homem	-0,061*** (0,008)	0,1993 (0,0016)	0,570*** (0,005)	0,2003 (0,0016)
Branco	-0,023** (0,009)	-0,0130 (0,0018)	-0,031** (0,005)	-0,0108 (0,0018)
Anos de estudo	0,114*** (0,001)	0,0243 (0,0002)	0,072*** (0,001)	0,0251 (0,0002)
Constante	-1,133*** (0,018)	— —	0,697*** (0,011)	— —
Observações	311.841			

Nota:

*p<0,1; **p<0,05; ***p<0,01

Tabela 6: Resultados da Regressão Logística Multinomial e Efeitos Marginais Médios (AME)

Variável	Trabalha Formal		Trabalha Informal	
	Coeficiente	AME	Coeficiente	AME
Bolsa Família (valor)	-0,0029*** (0,0001)	0,0001 (0,0000)	0,0004 (0,0000)	0,0007 (0,0000)
Urbano (zona do domicílio)	0,4730*** (0,0007)	-0,0514 (0,0021)	-0,2372*** (0,0010)	-0,0679 (0,0021)
Renda domiciliar per capita	0,0001 (0,0000)	0,0000 (0,0000)	0,0001*** (0,0000)	-0,0001 (0,0000)
Contribuição das crianças	-0,3630*** (0,0001)	0,2755 (0,0000)	1,2728*** (0,0000)	-0,2622 (0,0000)
Idade	-0,0160*** (0,0002)	-0,0022 (0,0000)	-0,0102*** (0,0002)	0,0075 (0,0002)
Homem	0,8262*** (0,0038)	0,2193 (0,0016)	1,0136*** (0,0052)	-0,5299 (0,0052)
Branco	-0,0690*** (0,0045)	-0,0123 (0,0019)	-0,0571 (0,0061)	0,0363 (0,0061)
Anos de estudo	0,1863*** (0,0009)	0,0162 (0,0000)	0,0752*** (0,0008)	-0,0753 (0,0008)
Constante	-2,8988*** (0,0009)	— —	-1,2929*** (0,0013)	— —
Observações	311.841			
Categoria de comparação	Não trabalha			

Nota: *p<0,1; **p<0,05; ***p<0,01

Um aspecto interessante deste trabalho é a possibilidade de realizar uma avaliação ex-ante. Para isso, calculamos o impacto de um aumento de R\$200 no valor do Bolsa Família sobre a decisão de trabalhar. Primeiramente, geramos as previsões das probabilidades de trabalho com os dados originais, sem nenhuma modificação no valor do benefício. Em seguida, criamos uma versão contrafactual dos dados, onde o valor do Bolsa Família foi incrementado em R\$200 para os beneficiários existentes. A partir dessa nova base de dados, geramos pre-

visões considerando o aumento no benefício. A diferença entre as médias das probabilidades de trabalho nas duas bases (original e contrafactual) fornece uma medida direta do efeito do aumento no valor do benefício sobre a decisão de trabalhar.

O código a seguir foi utilizado para calcular o impacto do aumento do Bolsa Família em R\$200:

```
## Criar uma cópia da base para o cenário contrafactual
Base_contrafactual <- Base

# Aumentar o bf_valor em 200 apenas para quem tem bf == 1
Base_contrafactual$bf_valor <- ifelse(Base$bf == 1, Base$bf_
  valor + 200, Base$bf_valor)

# Prever probabilidades para o cenário atual
Base$prob_atual <- predict(logit_binomial, newdata = Base, type
  = "response")

# Prever probabilidades para o cenário contrafactual
Base$prob_contrafactual <- predict(logit_binomial, newdata =
  Base_contrafactual, type = "response")

# Calcular a diferença média das probabilidades (apenas para bf
  == 1)
efeito_medio <- mean(
  na.omit(Base$prob_contrafactual[Base$bf == 1] - Base$prob_
    atual[Base$bf == 1])
)
```

Nesse código, criamos uma cópia da base de dados original e aplicamos um aumento de R\$200,00 no valor do Bolsa Família apenas para aqueles que já recebem o benefício (`bf == 1`). Em seguida, calculamos as probabilidades de trabalhar para o cenário atual (sem o aumento) e para o cenário contrafactual (com o aumento). A diferença média das probabilidades entre os dois cenários é então calculada para quantificar o efeito do aumento do Bolsa Família na decisão de trabalhar.

No modelo logit binomial, a diferença média foi de -1,15 pontos percentuais, e no modelo probit binomial, de -1,20 pontos percentuais, indicando que o aumento do benefício leva a uma leve diminuição na probabilidade de os indivíduos optarem por trabalhar.

No modelo logit multinomial, que distingue entre as categorias “não trabalha”, “trabalha informalmente” e “trabalha formalmente”, os resultados indicam que o aumento do valor do benefício diminui a probabilidade de “não trabalhar” em 0,83 pontos percentuais, reduz a probabilidade de “trabalhar informalmente” em 1,30 pontos percentuais e aumenta a probabilidade de “trabalhar formalmente” em 2,12 pontos percentuais. Esses resultados sugerem que o aumento do valor do benefício pode estar associado a um leve deslocamento na distribuição das formas de ocupação, com uma tendência de migração do trabalho informal para o trabalho formal, além de uma pequena redução na probabilidade de inatividade.

A literatura sobre os efeitos do Bolsa Família no mercado de trabalho apresenta resultados contraditórios. De Brauw et al. (2015) não encontraram impactos estatisticamente significativos do programa na participação individual na força de trabalho ou no número de horas de trabalho doméstico. Oliveira e Soares (2012) fazem uma revisão sobre os efeitos das transferências de renda sobre a oferta de trabalho e também chegam à conclusão de que são praticamente nulos. Por outro lado, Teixeira (2010) observaram uma redução na participação no mercado de trabalho, com uma diminuição de 0,56 horas por semana para homens (1,3%) e de 1,118 horas por semana para mulheres (4,1%). No entanto, o efeito médio do Bolsa Família na probabilidade de trabalhar não foi significativo. Já Tavares (2010) encontra que a presença de um beneficiário do programa na família eleva a participação das mães no mercado de trabalho em cerca de 5,6%. Na mesma linha, **mariane2024** encontra que o Bolsa Família aumenta em 7,6% a oferta de trabalho de mulheres, mas não a de homens.

Assim, os resultados obtidos neste estudo podem ser interpretados à luz da literatura existente sobre os efeitos do Bolsa Família no mercado de trabalho. Em primeiro lugar, as reduções nas probabilidades de não trabalhar e de trabalhar informalmente, acompanhadas por um aumento na probabilidade de trabalhar formalmente, dialogam com a literatura que aponta para efeitos heterogêneos do programa. Por exemplo, Teixeira (2010) destacam que o Bolsa Família pode impactar subgrupos específicos de beneficiários, enquanto Tavares (2010) e **mariane2024** identificaram efeitos positivos na participação no mercado de trabalho para mulheres e mães.

Além disso, os resultados obtidos no modelo logit multinomial oferecem uma perspectiva complementar aos estudos que argumentam que o programa não altera significativamente a participação geral na força de trabalho, como observado por De Brauw et al. (2015) e Oliveira e Soares (2012). Embora os impactos médios sejam pequenos, as mudanças na composição da ocupação sugerem que o programa pode incentivar a transição para formas mais estáveis de emprego. Isso é consistente com a ideia de que transferências de renda podem reduzir a necessidade de empregos informais precários e permitir que os beneficiários busquem ocupações formais com melhores condições de trabalho.

Essa abordagem de modelos comportamentais com uso de logit e o probit também oferecem a vantagem de possibilitar a captura de efeitos heterogêneos para diferentes subgrupos de beneficiários. Ou seja, é possível explorar como mudanças no valor do Bolsa Família impactam categorias específicas da população, como mulheres, em comparação a homens, pessoas brancas versus pretas, ou indivíduos com diferentes níveis de escolaridade. Essa característica pode melhorar a qualidade da aplicação da política pública ao identificar os grupos que são mais ou menos beneficiados.

Por outro lado, é importante destacar as limitações inerentes a esse tipo de exercício. Primeiramente, os modelos comportamentais dependem de hipóteses fortes sobre o comportamento dos indivíduos, como a suposição de que eles maximizam sua utilidade com base nos incentivos financeiros simulados. Essas hipóteses podem não capturar completamente outros fatores contextuais, como normas sociais ou barreiras de acesso ao mercado de trabalho formal. Além disso, os resultados de microsimulações são altamente sensíveis às escolhas metodológicas, como a especificação do modelo, a seleção das variáveis explicativas e a qualidade dos dados utilizados. Por fim, ao utilizar modelos baseados em probabilidades, assume-se que as relações estimadas no passado continuarão válidas para simular cenários futuros, o que pode não refletir mudanças estruturais ou choques inesperados na economia.

Portanto, os resultados obtidos não apenas dialogam com a literatura existente sobre os efeitos do Bolsa Família no mercado de trabalho, mas também ilustram como a microsimulação pode ser uma ferramenta poderosa para capturar nuances nas respostas dos beneficiários, desde que suas limitações sejam devidamente consideradas.

4.3 Inclusão de Remédio no Sistema Público de Saúde¹²

Retomando o nosso exemplo sobre saúde pública, vamos agora usar os conhecimentos sobre cadeias de Markov para construir um modelo dinâmico que avalie o impacto ex-ante da introdução de um medicamento no sistema público de saúde. A análise considera tanto os benefícios financeiros como a redução nos custos de internação, além dos ganhos em qualidade de vida para os pacientes, medidos por métricas como QALYs (Quality-Adjusted Life Years).

O problema envolve uma doença com progressão dinâmica, dividida em quatro estados de saúde: saudável (H), doente ($S1$), doente grave ($S2$) e óbito (D). A introdução do remédio WZ tem como objetivo prevenir a progressão de $S1$ para $S2$, melhorando a qualidade de vida de indivíduos em $S1$, sem impacto direto em $S2$. Por outro lado, o remédio influencia indiretamente os custos ao reduzir o número de casos graves ($S2$), que demandam internações prolongadas e onerosas.

¹²Esta seção é baseada no artigo "Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial" de Krijkamp et al. (2018).

Para modelar a trajetória de saúde de cada indivíduo ao longo do tempo, usamos cadeias de Markov não homogêneas com probabilidades de transição ajustadas para refletir os efeitos do remédio. Essas probabilidades são dinâmicas, incorporando fatores como a duração dos estados $S1$ e $S2$, que aumenta o risco de mortalidade com o tempo. Adicionalmente, o modelo captura o impacto econômico e os benefícios de saúde pública através da integração de custos descontados e ganhos de utilidade ajustados pela qualidade de vida.

Por meio dessa abordagem, será possível estimar os custos totais e os QALYs médios por indivíduo, de modo a avaliar a viabilidade econômica e os benefícios sociais da implementação dessa política de saúde.

Esse exemplo será construído em cinco partes. A primeira irá expor a matriz de transição de probabilidades para os diferentes ciclos temporais. A segunda parte irá descrever os custos do modelo. A terceira irá expor como os agentes julgam os benefícios da nova política. A quarta irá expor o algoritmo do modelo, e a quinta irá realizar um exercício quantitativo.

4.3.1 Matriz de transição

Seja $i \in I$ um indivíduo pertencente ao conjunto I de todos os indivíduos. Para cada indivíduo i , definimos o vetor de estados de saúde ao longo do tempo como:

$$M_i = (M_{i0}, M_{i1}, M_{i2}, \dots, M_{in}),$$

onde:

1. M_{it} representa o estado de saúde do indivíduo i em um instante temporal t , com $t \in \{0, 1, 2, \dots, n\}$.
2. $n \in \mathbb{N}$ é o número total de ciclos temporais considerados no modelo.

Cada elemento M_{it} pertence ao conjunto de possíveis estados de saúde S , definido como:

$$M_{it} \in S, \quad S = \{H, S1, S2, D\},$$

onde:

- H representa o estado de saúde “Saudável”;
- $S1$ representa o estado de saúde “Doente”
- $S2$ representa o estado de saúde “Doente em estado grave”
- D representa o estado de “Óbito”.

Adicionalmente, o vetor M_i descreve a trajetória dos estados de saúde do indivíduo i ao longo do tempo t , fornecendo uma representação completa da dinâmica de saúde desse indivíduo dentro do período analisado.

A sequência de estados de saúde de um indivíduo i ao longo do tempo $(M_{i0}, M_{i1}, M_{i2}, \dots, M_{in})$ é um processo estocástico, dado que a ocorrência de um determinado estado em um instante temporal t é governada pelas probabilidades de transição do modelo. Essas probabilidades refletem as incertezas inerente à progressão dos estados de saúde ao longo do tempo.

Vamos supor que nosso processo estocástico $(M_{i0}, M_{i1}, M_{i2}, \dots, M_{in})$ é uma cadeia de markov não homogênea. O modelo acompanha esses indivíduos ao longo de 30 anos ($n = 30$), em ciclos anuais, mantendo uma duração constante para cada ciclo.

Inicialmente, todos os indivíduos são considerados saudáveis (H). Ou seja a $P(M_{i0} = H) = 1; \forall i$. Em outras palavras, a distribuição inicial da cadeia é $\alpha^{(0)} = (p_H, p_{S1}, p_{S2}, p_D) = (1, 0, 0, 0)$. Com o tempo, eles podem desenvolver a doença e progredir para o estado S1. Nesse estado, os indivíduos têm quatro possíveis trajetórias: podem se recuperar, retornando ao estado saudável (H); permanecer em S1; piorar, avançando para o estado mais grave, S2; ou falecer (D). Por outro lado, indivíduos em S2 não têm possibilidade de recuperação, ou seja, não podem retornar para S1 ou H.

Com relação a mortalidade, indivíduos saudáveis (H) possuem uma taxa de mortalidade fixa, mas aqueles nos estados S1 e S2 enfrentam um risco elevado, que aumenta conforme o tempo que permanecem nesses estados. A cada ciclo adicional nesses estados, a taxa de mortalidade aumenta em 20%.

Duas estratégias de tratamento são consideradas: sem tratamento e com tratamento. O tratamento consiste na introdução de um medicamento que melhora a qualidade de vida de indivíduos no estado S1, mas não tem efeito sobre aqueles no estado S2. No entanto, devido à natureza da doença, não é possível distinguir entre os estados S1 e S2. Por isso, todos os indivíduos nesses estados recebem o medicamento até se recuperarem (retornando a H) ou falecerem (D).

A tendência de um paciente mudar de um estado de saúde para outro é frequentemente expressa em termos de **taxa de transição** (r), que representa o número de ocorrências de um evento (como adoecer ou falecer) para um determinado número de pacientes por unidade de tempo (ou ciclo temporal). As taxas de transição podem assumir valores entre 0 e ∞ . (Sonnenberg e Beck, 1993).

Por outro lado, a **probabilidade de transição** descreve a chance de um evento ocorrer em um intervalo de tempo específico, com valores variando entre 0 e 1. Para converter uma taxa de transição em probabilidade, utiliza-se a **distribuição exponencial**, uma distribuição amplamente empregada para modelar o tempo até a ocorrência de um

evento.(Sonnenberg e Beck, 1993)

A distribuição exponencial é definida pela seguinte função densidade de probabilidade (FDP):

$$f(t) = r e^{-rt}, \quad t \geq 0,$$

onde:

- $f(t)$ é a probabilidade de o evento ocorrer no instante exato t ,
- r é a taxa de transição (constante no tempo),
- t é o tempo.

A probabilidade acumulada de o evento ocorrer até o tempo t é dada pela função de distribuição acumulada (FDA) da distribuição exponencial:

$$F(t) = P(T \leq t) = 1 - e^{-rt},$$

onde $P(T \leq t)$ é a probabilidade de o evento ocorrer em algum momento dentro do intervalo de tempo $[0, t]$.

Para modelos discretos de tempo, como em cadeias de Markov, essa fórmula é utilizada para calcular a probabilidade de transição entre estados. Se assumirmos que a taxa de transição r é constante, a probabilidade de transição $P(t)$ dentro de um intervalo de tempo t é:

$$P(t) = 1 - e^{-rt}.$$

Seguindo o modelo contido em Krijkamp et al(2018), os valores dos parâmetros serão os seguintes:

- Probabilidade de morte quando saudável: $p_{HD} = 0.005$
- Probabilidade de permanecer doente quando saudável: $p_{HS1} = 0.15$
- Probabilidade de se recuperar de doente para saudável: $p_{S1H} = 0.5$
- Probabilidade de piorar de doente para mais doente: $p_{S1S2} = 0.105$
- Razão de taxas de morte quando doente em relação a saudável: $rr_{S1} = 3$
- Razão de taxas de morte quando mais doente em relação a saudável: $rr_{S2} = 10$

- Aumento da taxa de mortalidade com cada ano adicional no estado doente: $rp_{S1S2} = 0.2$

A taxa de morte quando saudável é calculada como:

$$r_{HD} = -\ln(1 - p_{HD}) = -\ln(1 - 0.005) \approx 0.0050125$$

A taxa de morte quando doente é ajustada pelo fator de risco relativo:

$$r_{S1D} = rr_{S1} \cdot r_{HD} = 3 \cdot 0.0050125 \approx 0.0150375$$

A taxa de morte quando mais doente é ajustada de forma similar:

$$r_{S2D} = rr_{S2} \cdot r_{HD} = 10 \cdot 0.0050125 \approx 0.050125$$

As probabilidades correspondentes são então calculadas como:

$$p_{S1D} = 1 - e^{-r_{S1D}} = 1 - e^{-0.0150375} \approx 0.014925$$

$$p_{S2D} = 1 - e^{-r_{S2D}} = 1 - e^{-0.050125} \approx 0.048890$$

Com base nessas informações, podemos calcular nossa matriz de transição para o primeiro ciclo $t = 1$.

$$P^{(1)} = \begin{pmatrix} 1 - p_{HS1} - p_{HD} & p_{HS1} & 0 & p_{HD} \\ p_{S1H} & 1 - p_{S1H} - p_{S1S2} - p_{S1D} & p_{S1S2} & p_{S1D} \\ 0 & 0 & 1 - p_{S2D} & p_{S2D} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Dados os valores dos parâmetros, a nossa matriz fica da seguinte maneira:

$$P^{(1)} = \begin{pmatrix} 0.845 & 0.15 & 0 & 0.005 \\ 0.5 & 0.381 & 0.105 & 0.0149 \\ 0 & 0 & 0.9512 & 0.0488 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Para definirmos a matriz de transição para $t > 1$, precisamos da seguinte definição – Um paciente pode permanecer Doente ou Doente em estágio grave ao longo de alguns ciclos temporais, ou seja, sua condição pode se manter estável *por* um determinado período. Assim, podemos definir formalmente essa duração (*dur*) da seguinte forma: Para cada indivíduo i

no ciclo t :

$$dur_i(t) = \begin{cases} dur_i(t-1) + 1, & \text{se } M_{i,t} \in \{S1, S2\} \\ 0, & \text{se } M_{i,t} \notin \{S1, S2\} \end{cases}$$

Onde:

- $M_{i,t}$: Estado de saúde do indivíduo i no ciclo t .
- $dur_i(t)$: Duração acumulada nos estados doentes até o ciclo t .

A matriz de transição P^n é praticamente idêntica com relação aos parâmetros da P^1 . A única e importante exceção é que as probabilidades de morte nos estados $S1$ e $S2$ dependem da duração dur e aumentam conforme o indivíduo permanece mais tempo nos estados doentes. São definidas pelas seguintes equações:

- **Probabilidade de Morte em $S1$:**

$$p_{S1D}(dur) = 1 - e^{-r_{S1D} \times (1 + dur \times r_{pS1S2})}$$

- **Probabilidade de Morte em $S2$:**

$$p_{S2D}(dur) = 1 - e^{-r_{S2D} \times (1 + dur \times r_{pS1S2})}$$

Em que

- r_{pS1S2} : Aumento proporcional da taxa de mortalidade por ciclo adicional doente.
- dur : Duração (em ciclos) que o indivíduo passou nos estados $S1$ ou $S2$.

Logo a matriz de transição para $t > 1$ é:

$$P(dur) = \begin{pmatrix} 1 - p_{HS1} - p_{HD} & p_{HS1} & 0 & p_{HD} \\ p_{S1H} & 1 - p_{S1H} - p_{S1S2} - p_{S1D}(dur) & p_{S1S2} & p_{S1D}(dur) \\ 0 & 0 & 1 - p_{S2D}(dur) & p_{S2D}(dur) \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4.3.2 Custo

Um gestor público responsável pelo orçamento de uma unidade hospitalar enfrenta o desafio de calcular os custos associados a cada paciente, considerando o estado de saúde individual

de cada um. Esses custos podem variar significativamente de acordo com a condição clínica do paciente. Para um paciente saudável, o custo pode se restringir ao tempo necessário para uma avaliação médica de rotina. No caso de um paciente com algum problema de saúde, os custos podem incluir despesas com medicamentos ou atendimentos primários, como consultas ou procedimentos simples. Já para pacientes em estado grave, os custos podem ser muito mais elevados, envolvendo, por exemplo, internações prolongadas em unidades de terapia intensiva (UTIs), além de tratamentos especializados e equipamentos de suporte à vida.

Por hipótese, vamos supor que para cada indivíduo i em cada ciclo t o custo de cada estágio (c_{it}) é:

$$c_{it} = \begin{cases} c_H & \text{se } M_{it} = H \\ c_{S1} + c_{Trt} \times Trt & \text{se } M_{it} = S1 \\ c_{S2} + c_{Trt} \times Trt & \text{se } M_{it} = S2 \\ c_D & \text{se } M_{it} = D \end{cases}$$

Em que

- c_H é o custo de estar no estado “Saudável”.
- c_{S1} é o custo de estar no estado “Doente”.
- c_{S2} é o custo de estar no estado “Doente em estado grave”
- c_{Trt} é o custo do tratamento.
- Trt é uma variável com valor igual a 1 se o indivíduo i recebe o tratamento e 0 caso o contrário.
- c_D é o custo de estar no estado “Óbito”.

Para nosso exemplo iremos postular os seguintes custos:

- $c_H = 2000$
- $c_{S1} = 4000$
- $c_{S2} = 15000$
- $c_{Trt} = 12000$
- $c_D = 0$

Para cada ciclo temporal t , esses custos são computados para cada paciente. Porém, é necessário trazê-los a valor presente porque o dinheiro no futuro tem um valor diferente do de hoje, devido, por exemplo, ao custo de oportunidade. Esse ajuste permite comparar de forma consistente os custos incorridos ao longo do tempo. A formula do fator de desconto $v_c(t)$ é:

$$v_c(t) = \frac{1}{(1 + d_c)^t}$$

Em que d_c é a taxa de desconto. Para nosso exemplo, iremos postular que essa taxa é igual a 3%. Logo o custo descontado em cada ciclo é:

$$c_{it}^{desc} = c_{it} \times v_c(t)$$

Se somarmos o custo descontado para um paciente i em todos os ciclos temporais de seu histórico hospitalar, chegamos ao custo total descontado por indivíduo:

$$TC_i = \sum_{t=0}^{n_t} c_{it}^{desc}$$

Se um gestor estiver interessado em saber qual foi o custo médio de cada paciente, levando em conta todos os ciclos temporais,, isso pode ser feito pelo custo médio na coorte:

$$\bar{C} = \frac{1}{n_i} \sum_{i=1}^{n_t} TC_i$$

Onde:

- n_i : Número total de indivíduos.
- n_t : Número total de ciclos.

4.3.3 Utilidade

Suponha que um medicamento ofereça a um paciente uma sobrevida de 10 anos, mas exija que ele permaneça hospitalizado durante todo esse período. Outro medicamento, por sua vez, pode proporcionar uma sobrevida de 5 anos, permitindo que o paciente passe esse tempo em casa. Esse exemplo mostra que há duas margens benéficas ao bem estar individual, como aumento na expectativa de vida e da qualidade da vida. Para avaliar e comparar a utilidade dessas intervenções, a literatura médica utiliza o Quality-Adjusted Life Year (QALY), ou Ano de Vida Ajustado pela Qualidade. Essa métrica quantifica o benefício acumulado das intervenções em saúde, ponderando não apenas a extensão da vida proporcionada, mas

também a qualidade de vida experimentada em cada período, servindo como uma base consistente para decisões em políticas de saúde.

O QALY é composto por dois elementos fundamentais. O primeiro é a utilidade instantânea, um valor entre 0 e 1 que reflete a preferência do indivíduo por estar em determinado estado de saúde em um momento específico t , com 1 representando o estado “Saudável” e 0 indicando o estado de “Óbito”. O segundo elemento é a quantidade de ciclos temporais que o indivíduo já passou até o momento t .

Para cada indivíduo i em cada ciclo t :

Utilidade Instantânea:

- Sem Tratamento ($Trt = 0$):

$$u_{it} = \begin{cases} u_H & \text{se } M_{it} = H \\ u_{S1} & \text{se } M_{it} = S1 \\ u_{S2} & \text{se } M_{it} = S2 \\ u_D & \text{se } M_{it} = D \end{cases}$$

- Com Tratamento ($Trt = 1$):

$$u_{it} = \begin{cases} u_H & \text{se } M_{it} = H \\ \gamma_{it} \times (u_{Trt} - dur \times ru_{S1S2}) & \text{se } M_{it} = S1 \\ u_{S2} & \text{se } M_{it} = S2 \\ 0 & \text{se } M_{it} = D \end{cases}$$

Em que,

- u_H é Utilidade Instantânea do estado “Saudável”.
- u_{S1} é Utilidade Instantânea do estado “Doente”.
- u_{S2} é Utilidade Instantânea do estado “Doente em estado grave”.
- ru_{S1S2} é uma redução na utilidade por ciclo adicional no estado doente durante o tratamento.
- γ_i é um variável de Heterogeneidade Individual do tratamento
- u_D é Utilidade Instantânea do estado “Óbito”.

Há duas hipóteses nessa estrutura de utilidade. A primeira é que os benefícios do tratamento diminuem ao longo do tempo, com a utilidade dos indivíduos no estado $S1$ em tratamento reduzindo em 0,03 a cada ano que permanecem em $S1$. E a segunda é que a melhora na qualidade de vida proporcionada pelo tratamento varia entre os indivíduos devido a uma característica que atua como um modificador do efeito do tratamento modelada aqui como uma distribuição uniforme, com valor mínimo de 0.95 e máximo de 1,05. Para nosso exemplo iremos postular as seguintes utilidades instantâneas :

- $u_H = 1$
- $u_{S1} = 0.75$
- $u_{S2} = 0.5$
- $ru_{S1S2} = 0.03$
- $u_D = 0$
- $\gamma_i \sim U(0.95, 1.05)$

Assim, por definição o Cálculo dos QALYs no Ciclo é:

$$QALY_{it} = u_{it} \times cl$$

Em que,

- cl = Número de ciclos até o tempo t

Novamente, assim como no cálculo dos custos, precisamos levar a valor presente o calculo QALYs:

Fator de Desconto:

$$v_e(t) = \frac{1}{(1 + d_e)^t}$$

Em que d_e é a taxa de desconto para a QALYs. Para nosso exemplo iremos postular que essa taxa é igual a 3%. Logo o benefício descontado em cada ciclo é:

$$QALY_{it}^{desc} = QALY_{it} \times v_{dwc}(t)$$

Somando esse benefício ao longo de todo os ciclos de um indivíduo temos :

$$TE_i = \sum_{t=0}^{n_t} QALY_{it}^{desc}$$

E se tivermos interesse em saber qual é a media dos QALYs para uma coorte de indivíduos temos que

$$\bar{E} = \frac{1}{n_i} \sum_{i=1}^{n_i} TE_i$$

4.3.4 Algoritmo

Dado o exposto nas partes de matriz de transição, custos e utilidade, o seguinte algoritmo descreve a simulação dinâmica de transições de saúde, custos e benefícios acumulados ao longo do tempo para uma coorte de indivíduos.

A simulação é estruturada em três etapas principais:

- Inicialização dos parâmetros gerais do modelo, incluindo a definição dos estados de saúde, probabilidades de transição, custos, utilidades e taxas de desconto.
- Execução da simulação para cada indivíduo, com a evolução dos estados de saúde ao longo dos ciclos, cálculo dos custos e QALYs por ciclo, e acumulação dos resultados ao longo do tempo.
- Cálculo dos resultados médios da coorte, permitindo a análise consolidada dos custos totais e QALYs acumulados por indivíduo.

A seguir, o algoritmo é detalhado passo a passo:

1. Inicialização dos Parâmetros Gerais do Modelo

- a) Defina o conjunto de estados de saúde possíveis:

$$S = \{H, S1, S2, D\},$$

onde:

- H representa o estado “Saudável”;
 - $S1$ representa o estado “Doente”;
 - $S2$ representa o estado “Doente Grave”;
 - D representa o estado “Óbito”.
- b) Configure as probabilidades de transição iniciais e as regras de atualização para os parâmetros dependentes do tempo t e da duração dur nos estados $S1$ ou $S2$.

c) Defina os custos associados a cada estado:

$$c_H, \quad c_{S1}, \quad c_{S2}, \quad c_D,$$

e o custo do tratamento:

$$c_{\text{Trt}}.$$

d) Estabeleça as utilidades associadas a cada estado:

$$u_H, \quad u_{S1}, \quad u_{S2}, \quad u_D,$$

bem como a redução na utilidade por ciclos adicionais nos estados doentes durante o tratamento:

$$ru_{S1S2}.$$

e) Especifique as taxas de desconto para custos (d_c) e utilidades (d_e).

2. Simulação para Cada Indivíduo

a) Para cada indivíduo $i = 1, \dots, n_i$, execute:

i) Determine se o indivíduo receberá tratamento, definindo:

$$Trt_i \in \{0, 1\}.$$

ii) Gere a variável de heterogeneidade individual:

$$\gamma_i \sim U(0,95, 1,05).$$

iii) Inicialize:

- Estado inicial de saúde:

$$M_{i0} = H.$$

- Custos acumulados:

$$TC_i = 0.$$

- QALYs acumulados:

$$TE_i = 0.$$

- Duração nos estados doentes:

$$dur_i(0) = 0.$$

iv) **Para cada ciclo** $t = 1, \dots, n_t$, proceda:

a) **Cálculo das Probabilidades de Transição:**

- Com base em $M_{i,t-1}$, $dur_i(t-1)$ e nos parâmetros do modelo, calcule as probabilidades de transição para os possíveis estados em t .

b) **Atualização do Estado de Saúde:**

- Sorteie o novo estado M_{it} de acordo com as probabilidades de transição calculadas.

c) **Cálculo dos Custos e QALYs:**

- Determine o custo imediato c_{it} e o benefício $QALY_{it}$, considerando o estado M_{it} , o tratamento Trt_i e a heterogeneidade γ_i .
- Aplique os fatores de desconto:

$$c_{it}^{\text{desc}} = c_{it} \times \frac{1}{(1 + d_c)^t}, \quad QALY_{it}^{\text{desc}} = QALY_{it} \times \frac{1}{(1 + d_e)^t}.$$

d) **Atualização dos Acumulados:**

- Atualize os valores acumulados até o ciclo n :

$$TC_i = \sum_{t=0}^n c_i^{\text{desc}}, \quad TE_i = \sum_{t=0}^n QALY_{it}^{\text{desc}}.$$

e) **Atualização da Duração:**

- Atualize a duração nos estados doentes:

$$dur_i(t) = \begin{cases} dur_i(t-1) + 1, & \text{se } M_{it} \in \{S1, S2\} \\ 0, & \text{caso contrário.} \end{cases}$$

3. Cálculo dos Resultados Médios da Coorte

a) Após a simulação para todos os indivíduos e ciclos, calcule os resultados médios:

$$\bar{C} = \frac{1}{n_i} \sum_{i=1}^{n_i} TC_i, \quad \bar{E} = \frac{1}{n_i} \sum_{i=1}^{n_i} TE_i.$$

4.3.5 Aplicação Numérica

Os resultados apresentados na Tabela 7 sintetizam os custos e benefícios médios associados ao tratamento em comparação com a ausência de tratamento.

Os custos médios por paciente foram R\$62.667 na ausência de tratamento e R\$117.455 quando há tratamento, resultando em um custo incremental médio de tratamento de R\$54.787. Por outro lado, os QALYs médios foram 15,279 na ausência de tratamento e 15,786 quando há tratamento, indicando um ganho incremental médio de 0,507 QALYs por paciente ao se introduzir o tratamento.

Com esses valores, o ICER (Incremental Cost-Effectiveness Ratio), calculado como a razão entre os custos e os benefícios incrementais, foi estimado em 108.047 reais por QALY. Ou seja, o custo médio de ganho de um ano adicional ajustado pela qualidade de vida ao implementar o tratamento é de R\$108.047.

Contudo, é possível que haja uma incerteza sobre os custos do tratamento. É interessante, portanto, considerar a possibilidade deste exceder os R\$12.000 originalmente considerados. A Tabela 8 revela que, caso o custo do tratamento seja de R\$15.000 ou de R\$20.000, o custo incremental médio de oferecer tratamento por pessoa é de R\$64.484 e R\$91.312, respectivamente. Isso representa 112% e 166% do custo incremental originalmente calculado. Evidentemente, o ICER também aumenta, passando para R\$134.983 e R\$179.977 o custo médio de ganho de um ano adicional, ajustado pela qualidade de vida ao implementar-se o tratamento.

Tabela 7: Resultados da análise incremental de custo-efetividade

Intervenção	Custo (R\$)	QALYs
Sem Tratamento	62.667	15,279
	(120)	(0,017)
Com Tratamento	117.455	15,786
	(231)	(0,017)
Custo Incremental	54.787	
	(117)	
QALYs Ganhos	0,507	
	(0,001)	
ICER	108.047	

Tabela 8: Resultados da análise incremental de custo-efetividade

	$c_{Trt} = 15.000$		$c_{Trt} = 20.000$	
Intervenção	Custo (R\$)	QALYs	Custo (R\$)	QALYs
Sem Tratamento	62.667	15,279	62.667	15,279
	(120)	(0,017)	(120)	(0,017)
Com Tratamento	131.152	15,786	153.980	15,786
	(260)	(0,017)	(308)	(0,017)
Custo Incremental	68.484		91.312	
	(146)		(195)	
QALYs Ganhos	0,507		0,507	
	(0,001)		(0,001)	
ICER	134.983		179.977	

Além de verificar os resultados do tratamento, podemos identificar a quantidade de indivíduos em cada estado de saúde do modelo por ciclo. Ou seja, quantas pessoas se encontram no estado “saudável” (H), “doente”(S1), “doente grave”(S2) e em “óbito”(D). Estas estimativas estão presentes na Tabela 9 abaixo.

Tabela 9: Simulação *Cohort*

Ciclo	H	S1	S2	D
0	100000	0	0	0
1	84446	15042	0	512
2	78936	18290	1583	1191
3	75939	18589	3413	2059
4	73416	18383	5154	3047
5	71362	17813	6606	4219
6	69112	17472	7961	5455
7	67164	16892	9097	6847
8	65226	16352	10048	8374
9	63306	15894	10812	9988
10	61409	15409	11476	11706
11	59527	15069	11925	13479
12	57731	14639	12218	15412
13	56131	14164	12420	17285
14	54363	13876	12531	19230
15	52756	13493	12538	21213
16	51395	12866	12544	23195
17	50040	12398	12396	25166
18	48489	12100	12187	27224
19	46695	12148	11966	29191
20	45673	11451	11699	31177
21	44420	11064	11438	33078
22	43080	10828	11143	34949
23	41885	10464	10881	36770
24	40596	10186	10661	38557
25	39290	10052	10352	40306
26	38183	9688	10047	42082
27	37154	9307	9754	43785
28	36054	9000	9535	45411
29	34951	8822	9252	46975
30	33981	8538	8963	48518

5 Conclusão

Ao longo deste guia, discutimos a metodologia de microsimulação como uma ferramenta para prever os impactos de políticas públicas. Exploramos sua base teórica, composta por arcabouços como Escolha Discreta e Modelos Dinâmicos, e apresentamos exemplos práticos que ilustram sua aplicabilidade em cenários reais. Essa abordagem permite capturar interações complexas entre políticas existentes, heterogeneidade nos comportamentos indivi-

duais e ajustes dinâmicos dos agentes econômicos às mudanças de incentivos.

A microssimulação é particularmente útil em problemas onde os efeitos heterogêneos são importantes, como políticas fiscais e de transferência de renda. Sua flexibilidade permite incorporar características institucionais específicas e simular diversos cenários contrafactuais, fornecendo estimativas detalhadas de impactos potenciais. No entanto, sua implementação requer escolhas cuidadosas por parte dos analistas, incluindo a seleção das variáveis relevantes, especificações adequadas para os modelos e o uso de dados detalhados.

Adicionalmente, algumas limitações inerentes devem ser reconhecidas. A precisão das previsões depende da qualidade dos dados disponíveis e das hipóteses feitas no modelo. Dados insuficientes ou hipóteses muito rígidas podem levar a resultados enviesados ou pouco informativos.

Apesar dessas limitações, a microssimulação se destaca como uma ferramenta poderosa para auxiliar na formulação e avaliação de políticas públicas. Sua capacidade de integrar análises teóricas e empíricas a torna indispensável em problemas que envolvem múltiplas camadas de interação e complexidade. Esperamos que este guia sirva como um ponto de partida para analistas e formuladores de políticas que desejam explorar as possibilidades oferecidas por essa metodologia.

6 Apêndice

6.1 Estimação modelos variáveis binárias

Se Y é uma variável de Bernoulli, tal que $\mathbb{P}[Y = 1] = p$ e $\mathbb{P}[Y = 0] = 1 - p$, então Y tem a seguinte função de massa de probabilidade:

$$\pi(y) = p^y(1 - p)^{1-y}, \quad y = 0, 1.$$

No modelo de índice $\mathbb{P}[Y = 1 | X] = F(X'\beta)$, Y é condicionalmente Bernoulli; portanto, sua função massa de probabilidade condicional é¹³:

$$\begin{aligned} \pi(Y | X) &= F(X'\beta)^Y [1 - F(X'\beta)]^{1-Y} \\ &= F(X'\beta)^Y F(-X'\beta)^{1-Y}. \end{aligned}$$

Tomando logaritmos e somando as observações, obtemos a função log-verossimilhança:

¹³Como F é simétrica por hipótese, temos que $F(-u) = 1 - F(u)$. Essa propriedade da simetria é utilizada na simplificação da expressão abaixo, permitindo a reescrita de $1 - F(X'\beta)$ como $F(-X'\beta)$.

$$\ell_n(\beta) = \sum_{i=1}^n [\log F(X_i'\beta)^Y + \log F(-X_i'\beta)^{(1-Y)}] .$$

Para os modelos probit e logit, isso se torna:

$$\ell_n^{\text{probit}}(\beta) = \sum_{i=1}^n [\log \Phi(X_i'\beta)^Y + \log \Phi(-X_i'\beta)^{1-Y}] ,$$

$$\ell_n^{\text{logit}}(\beta) = \sum_{i=1}^n [\log \Lambda(X_i'\beta)^Y + \log \Lambda(-X_i'\beta)^{1-Y}] .$$

O estimador de máxima verossimilhança (MLE) é o valor que maximiza $\ell_n(\beta)$ ¹⁴. Escrevemos isso como:

$$\hat{\beta}^{\text{probit}} = \underset{\beta}{\operatorname{argmax}} \ell_n^{\text{probit}}(\beta),$$

$$\hat{\beta}^{\text{logit}} = \underset{\beta}{\operatorname{argmax}} \ell_n^{\text{logit}}(\beta).$$

Como as funções de log-verossimilhança do probit e do logit são globalmente côncavas, os estimadores $\hat{\beta}^{\text{probit}}$ e $\hat{\beta}^{\text{logit}}$ são únicos. No entanto, devido à ausência de uma solução explícita, esses estimadores precisam ser obtidos por métodos numéricos. Tanto a estimação por probit quanto por logit é consistente e assintoticamente normal, desde que sejam satisfeitas as condições de regularidade apropriadas.

Uma hipótese implícita ao usar o estimador de máxima verossimilhança e que faz diferença na interpretação do modelo é que para construir a verossimilhança, precisamos que a distribuição do nosso parâmetro de interesse seja identificado. Dizemos que um parâmetro β_0 é identificado se para qualquer $\beta \neq \beta_0$, $\beta \in \Theta$ implica que $f(y|X; \beta) \neq f(y|X; \beta_0)$. Essa característica é importante por causa do seguinte teorema:

Desigualdade de Informação: Se β_0 é identificado e

$$\mathbb{E}[\log f(y|X; \beta)] < \infty$$

para todo β , então

$$\ell(\beta) = \mathbb{E}[\log f(y|X; \beta)]$$

tem um máximo único em β_0 . Ou seja, se β_0 não é identificado não teríamos uma estimando

¹⁴HANSEN, Bruce. *Probability and Statistics for Economists*. Princeton: Princeton University Press, 2022. Capítulo 10.

único a partir do método de Verosimilhança.

Um problema é que β não pode ser identificado no modelo de variável latente, veja, por exemplo o caso da distribuição logística(logit)

15

$$F(y|X; \beta, \gamma, \tau) = \frac{1}{1 + \exp\left(\frac{X'\beta - \gamma}{\tau}\right)}$$

$$F(y|X; c\beta, c\gamma, c\tau) = \frac{1}{1 + \exp\left(\frac{X'c\beta - c\gamma}{c\tau}\right)}$$

Logo,

$$\begin{aligned} f(Y | X; \beta, \gamma, \tau) &= F(X'\beta)^Y (1 - F(X'\beta))^{1-Y} \\ &= F(X'\beta)^Y F(-X'\beta)^{1-Y} = f(Y | X; c\beta, c\gamma, c\tau) \end{aligned}$$

Assim, a função objetivo da amostra pode ser maximizada para um contínuo de valores de c . β é identificado apenas até um fator de escala.

Uma solução para esse problema é fixar a escala onde o erro ϵ é uma normal padrão ou

¹⁵A função de distribuição acumulada (CDF) é:

$$F(x|\mu, s) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

Onde:

- μ é o parâmetro de localização, que define o ponto médio da distribuição (o valor em torno do qual a distribuição é centrada).
- s é o parâmetro de escala, que controla a dispersão dos dados. Quanto maior s , mais achatada será a distribuição.

A Média e variância da distribuição logística são dadas por:

$$\mathbb{E}(X) = \mu$$

$$\text{Var}(X) = \frac{\pi^2}{3} s^2$$

Onde:

- $\mathbb{E}(X)$ é a média da distribuição.
- $\text{Var}(X)$ é a variância da distribuição.

uma logística padrão. A seguir, mostramos um exemplo de identificação usando essa solução no modelo logit (o raciocínio para o modelo probit é similar):

$$\begin{aligned}
F(y|X; \frac{\beta}{\sigma}, 0, \tau) &= \frac{1}{1 + \exp\left(\frac{X' \frac{\beta}{\sigma} - 0}{\tau}\right)} \\
F(y|X; \frac{\beta}{\sigma}, 0, \tau) &= \frac{1}{1 + \exp\left(\frac{X' \frac{\beta}{\frac{\pi}{\sqrt{3}}\tau}}{\tau}\right)} \\
F(y|X; c\frac{\beta}{\sigma}, 0, c\tau) &= \frac{1}{1 + \exp\left(\frac{X' \frac{c\beta}{\frac{\pi}{\sqrt{3}}c\tau}}{c\tau}\right)} \\
&\neq F(y|X; \frac{\beta}{\sigma}, 0, \tau)
\end{aligned}$$

Apenas o tamanho relativo de $\beta^* = \beta/\sigma = c\beta/c\sigma$ importa para a probabilidade de resposta e, portanto, para a verossimilhança. Portanto, β e σ não podem ser identificados separadamente – apenas o tamanho relativo β/σ .

Resumindo, a abordagem padrão é normalizar σ para um valor conveniente: os modelos probit e logit utilizam as normalizações $\sigma = 1$ e $\sigma = \frac{\pi}{\sqrt{3}} \approx 1.8$, respectivamente. E a média μ para 0 nos dois modelos. Embora o coeficiente β não seja identificado, os seguintes parâmetros são:

1. Coeficientes escalados: $\beta^* = \frac{\beta}{\sigma}$.
2. Razão dos coeficientes: $\frac{\beta_1}{\beta_2} = \frac{\beta_1^*}{\beta_2^*}$.
3. Efeitos marginais: $\frac{\partial}{\partial x} P(x) = \frac{\beta}{\sigma} f\left(\frac{x'\beta}{\sigma}\right) = \beta^* f(x'\beta^*)$.

Esses parâmetros dependem de β^* e, portanto, são identificados.

6.2 Prova teorema 3.1

Para provar o teorema 3.1 precisamos de dois teoremas preliminares:

Teorema 6.1

Regra de Multiplicação para Probabilidades Condicionais: Suponha que A_1, A_2, \dots, A_n são eventos tais que $\Pr(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Então

$$\Pr(A_1 \cap A_2 \cap \dots \cap A_n) = \Pr(A_1) \Pr(A_2|A_1) \Pr(A_3|A_1 \cap A_2) \dots \Pr(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Teorema 6.2

Versão Condicional da Lei da Probabilidade Total: Suponha que os eventos B_1, \dots, B_k formam uma partição do espaço S e que $\Pr(B_j \cap C) > 0$ para $j = 1, \dots, k$. A lei da probabilidade total possui um análogo condicional a um outro evento C , a saber,

$$\Pr(A|C) = \sum_{j=1}^k \Pr(B_j|C) \Pr(A|B_j \cap C).$$

Prova Teorema 3.1 Considere uma cadeia de Markov com k estados possíveis $1, \dots, k$ e a matriz de transição P associada a esses estados. Suponha que a cadeia esteja no estado i no tempo n . Queremos, então, determinar a probabilidade de que a cadeia estará no estado j no tempo $n+m$. Em outras palavras, calcularemos a probabilidade condicional de $X_{n+m} = j$ dado $X_n = i$, denotada por $p_{ij}^{(m)}$. Seja um estado qualquer $r \in \{1, \dots, k\}$ e um passo s tal que $0 < s < n$:

$$\begin{aligned} p_{ij}^{(m)} &= \Pr(X_m = j | X_0 = i) \\ &= \sum_{r=1}^k \Pr(X_s = r, X_m = j | X_0 = i) \\ &= \sum_{r=1}^k \Pr(X_s = r | X_0 = i) \Pr(X_m = j | X_s = r, X_0 = i) \\ &= \sum_{r=1}^k \Pr(X_s = r | X_0 = i) \Pr(X_m = j | X_s = r) \\ &= \sum_{r=1}^k p_{ir}^s p_{rj}^{m-s}, \end{aligned}$$

Em que na segunda equação usamos a versão condicional da lei da probabilidade total, terceira equação usamos a regra de multiplicação das probabilidades condicionais e na quarta utilizamos a definição de uma Cadeia de Markov.

6.3 Prova teorema 3.4

Prova Considere primeiramente o caso $n = 1$. Obtemos, para $j = 1, \dots, k$, que

$$\alpha_j^{(1)} = \mathbb{P}(X_1 = s_j) = \sum_{i=1}^k \mathbb{P}(X_0 = s_i, X_1 = s_j)$$

$$\begin{aligned}
&= \sum_{i=1}^k \mathbb{P}(X_0 = s_i) \mathbb{P}(X_1 = s_j \mid X_0 = s_i) \\
&= \sum_{i=1}^k \alpha_i^{(0)} P_{i,j} = (\alpha^{(0)} P)_j
\end{aligned}$$

onde $(\alpha^{(0)} P)_j$ denota o j -ésimo elemento do vetor linha $\alpha^{(0)} P$. Portanto, $\alpha^{(1)} = \alpha^{(0)} P$.

Para o caso geral, usamos indução. Fixe m e suponha que o teorema 3.4 seja válido para $n = m$. Para $n = m + 1$, obtemos

$$\begin{aligned}
\alpha_j^{(m+1)} &= \mathbb{P}(X_{m+1} = s_j) = \sum_{i=1}^k \mathbb{P}(X_m = s_i, X_{m+1} = s_j) \\
&= \sum_{i=1}^k \mathbb{P}(X_m = s_i) \mathbb{P}(X_{m+1} = s_j \mid X_m = s_i) \\
&= \sum_{i=1}^k \alpha_i^{(m)} p_{i,j} = (\alpha^{(m)} P)_j
\end{aligned}$$

de modo que $\alpha^{(m+1)} = \alpha^{(m)} P$. Mas $\alpha^{(m)} = \alpha^{(0)} P^m$ pela hipótese de indução, então

$$\alpha^{(m+1)} = \alpha^{(m)} P = \alpha^{(0)} P^m P = \alpha^{(0)} P^{m+1}$$

e a prova está completa.

6.4 Prova teorema 3.5

Prova:

$$\begin{aligned}
&\Pr(X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_{k-1}} = i_{k-1}, X_{n_k} = i_k) \\
&= \Pr(X_{n_1} = i_1) \Pr(X_{n_2} = i_2 \mid X_{n_1} = i_1) \\
&\quad \times \Pr(X_{n_3} = i_3 \mid X_{n_1} = i_1, X_{n_2} = i_2) \dots \\
&\quad \times \Pr(X_{n_k} = i_k \mid X_{n_1} = i_1, \dots, X_{n_{k-1}} = i_{k-1}) \\
&= \Pr(X_{n_1} = i_1) \Pr(X_{n_2} = i_2 \mid X_{n_1} = i_1) \\
&\quad \times \Pr(X_{n_3} = i_3 \mid X_{n_2} = i_2) \dots \\
&\quad \times \Pr(X_{n_k} = i_k \mid X_{n_{k-1}} = i_{k-1}) \\
&= (\alpha^0 P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} \dots (P^{n_k - n_{k-1}})_{i_{k-1} i_k}
\end{aligned}$$

Em que no primeiro passo utilizamos a regra de multiplicação das probabilidades condicionais, no segundo passo utilizamos a definição de cadeia de Markov, e no terceiro passo utilizamos o teorema 3.4 e o Teorema 3.1.

Referências

- De Brauw, Alan, Daniel O Gilligan, John Hoddinott e Shalini Roy (2015). “Bolsa Família and household labor supply”. *Economic Development and Cultural Change* 63.3, pp. 423–457.
- Ferrari, Hamilton (2024). *Renda per capita do Brasil foi de R\$ 1.893 em 2023*. Acessado em: 11/06/2024. URL: <https://www.poder360.com.br/economia/renda-per-capita-do-brasil-foi-de>
- Figari, Francesco, Alari Paulus e Holly Sutherland (2015). “Chapter 24 - Microsimulation and Policy Analysis”. Em: *Handbook of Income Distribution*. Ed. por Anthony B. Atkinson e François Bourguignon. Vol. 2. Handbook of Income Distribution. Elsevier, pp. 2141–2221.
- Kaufman, Bruce E. e Julie L. Hotchkiss (2006). *The economics of Labour Markets*. Thomson South-Western.
- Mali, Tiago (2024). *56% das vagas formais ficam com beneficiários do Bolsa Família*. Acessado em: 26/11/2024. URL: <https://www.poder360.com.br/poder-economia/56-das-vagas-formais>
- Mendes, Vinícius (2024). ‘Bolsa Família deveria premiar quem consegue trabalho. Hoje, penaliza’. Acessado em: 21/11/2024. URL: <https://www.bbc.com/portuguese/articles/ced0365vwq0o>.
- Mitton, Lavinia, Holly Sutherland e Melvyn Weeks (2000). *Microsimulation modelling for policy analysis: challenges and innovations*. Cambridge University Press.
- O’Donoghue, Cathal (2014). *Handbook of microsimulation modelling*. Emerald Group Publishing.
- (2021). *Practical Microsimulation Modelling*. Oxford University Press.
- Oliveira, Luis Felipe Batista de e Sergei Suarez Dillon Soares (2012). “O que se sabe sobre os efeitos das transferências de renda sobre a oferta de trabalho”.
- Pavan, Bruno (2024). *Compras internacionais de US\$50 : calculadoradog1mostracomoficamospregos*. Acessado em: 22/10/2024. URL: <https://g1.globo.com/economia/noticia/2024/06/06/compras-internacionais-de-us-50-calculadoradog1mostracomoficamospregos>
- Qual classe social mais compra na Shein?* (2023). Acessado em: 22/10/2024. URL: <https://www.portalinsider.com.br/qual-classe-social-mais-compra-na-shein/>
- Tavares, Priscilla Albuquerque (2010). “Efeito do Programa Bolsa Família sobre a oferta de trabalho das mães”. *Economia e sociedade* 19, pp. 613–635.
- Teixeira, Clarissa Gondim (2010). *A heterogeneity analysis of the Bolsa Família Programme effect on men and women’s work supply*. Rel. téc. Working paper.
- Tomazelli, Idiana (2024). *Brasileiros importam por mês 13 milhões de mercadorias de até US\$ 50; valor chega a R\$ 1 bi*. Acessado em: 11/06/2024. URL: <https://www1.folha.uol.com.br/mercado/brasil/2024/06/brasileiros-importam-por-mes-13-milhoes-de-mercadorias-de-ate-us-50-valor-chega-a-r-1-bi/>