

X ENCONTRO DA ABCP

AT: Ensino e Pesquisa em Ciência Política e Relações Internacionais

SIMULAÇÕES DE MONTE CARLO NO ENSINO DE CIÊNCIA POLÍTICA

Fernando Meireles (UFMG)

Denisson Silva (UFMG)

Filipe Souza Corrêa (UFMG)

Belo Horizonte, MG
30 de Agosto a 2 de Setembro de 2016

SIMULAÇÕES DE MONTE CARLO NO ENSINO DE CIÊNCIA POLÍTICA

Fernando Meireles (UFMG)

Denisson Silva (UFMG)

Filipe Souza Corrêa (UFMG)

RESUMO:

Ensinar como certos mecanismos causais ou modelos estatísticos funcionam nem sempre é algo simples, especialmente quando estes não podem ser facilmente exemplificados. Este é o caso de várias aulas de metodologia, onde o professor se vê na difícil tarefa de explicar de forma clara porque certas ferramentas são adequadas em algumas situações e não em outras. Também é o caso de discussões onde variáveis de difícil mensuração, como satisfação e preferências, são objeto de análise. O objetivo deste artigo é mostrar como simulações de Monte Carlo podem ser usadas para superar dificuldades como estas. Após fazermos uma breve introdução ao método de Monte Carlo, mostramos como utilizá-lo a fim de ilustrar fenômenos de difícil mensuração ou conceitos abstratos; além disso, mostramos também como ele pode ser empregado para explicar intuitivamente a influência das violações de pressupostos sobre os resultados de alguns modelos estatísticos frequentemente empregados na Ciência Política. Discutimos sua aplicação no contexto de aulas da graduação e da pós-graduação na UFMG e oferecemos um passo-a-passo para reproduzir nossos exemplos utilizando o Software R, além de um simples aplicativo virtual (shiny app) com estes exemplos, de modo a ser adaptado para uso em sala de aula.

PALAVRAS-CHAVE: Método de Monte Carlo; Ensino; Metodologia.

INTRODUÇÃO

Ensinar como certas teorias ou modelos estatísticos funcionam nem sempre é algo simples, especialmente quando estes não podem ser facilmente exemplificados. Este é o caso de várias aulas de metodologia, onde o professor se vê na difícil tarefa de explicar de forma clara porque certas ferramentas são adequadas em algumas situações e não em outras, ou porque o uso inadequado de uma técnica produz resultados inválidos. Também é o caso de discussões onde fenômenos abstratos e variáveis de difícil mensuração, como ideologia e preferências, são objetos de análise. Embora pareçam bastante específicas, situações como estas são recorrentes em aulas de Ciências Sociais e cursos afins, ainda que não exista consenso, ou sequer literatura específica, sobre estes problemas na perspectiva do ensino de Ciência Política no Brasil.

Neste artigo, procuramos contribuir para esta ainda incipiente literatura oferecendo um método simples para lidar com problemas como os relatados acima. Nosso objetivo é mostrar como Simulações de Monte Carlo (SMC) podem ser adaptadas como recurso didático para ilustrar conteúdos complexos e/ou abstratos. O principal aspecto desta abordagem, conforme argumentamos na próxima seção, é a possibilidade que ela nos oferece de simular amostras de acordo com um processo gerador de dados (PGD) controlado, ou seja, de criar mundos artificiais totalmente manipuláveis pelos usuários. Em vez de nos levar a pensar principalmente no modelo estatístico que deve ser aplicado para cada tipo de dado, as simulações de Monte Carlo invertem esta perspectiva, colocando em primeiro lugar o exame dos mecanismos que produzem determinado fenômeno de acordo com alguma expectativa teórica. Esta inversão de perspectiva é o que torna mais intuitivo o aprendizado metodológico por meio de SMC (CARSEY; HARDEN, 2015).

Após oferecermos um breve passo a passo de como planejar uma simulação e apresentar alguns critérios para julgar sua adequação e qualidade, ilustramos o potencial da abordagem com dois exemplos adaptados para uso em sala de aula. Primeiro, apresentamos uma simulação de um processo legislativo básico, onde os votos dos legisladores em cada votação são gerados aleatoriamente e as regras de maioria necessária para aprovação podem ser modificadas previamente. Por meio deste simples exemplo, os alunos aprendem de uma maneira visual como pequenas modificações nas regras do processo legislativo podem provocar problemas de paralisia decisória. De modo geral, discutimos também como este exemplo permite ilustrar uma vantagem na aplicação didática das SMC: possibilitar uma compreensão mais concreta de aspectos normalmente difíceis de serem isolados ou observados no mundo real (MOONEY, 1997).

No segundo exemplo, mostramos como o método pode ser empregado para explicar intuitivamente a influência das violações de pressupostos sobre os resultados de um dos modelos estatísticos mais empregados na Ciência Política: a regressão linear por Mínimos Quadrados Ordinários (MQO). Como alguns alunos freqüentemente sentem dificuldade ao estudar estatística, ou se sentem intimidados pela matemática envolvida na formalização do modelos, argumentamos que uma forma de superar essas dificuldades é oferecer uma intuição do como funciona o modelo em questão em situações hipotéticas. Para isso, a título de exemplo, simulamos duas variáveis estabelecendo vários graus de correlação entre si, e criamos outra variável dependente a partir destas; em seguida, mostramos como a omissão de uma daquelas variáveis constitutivas enviesava as estimativas de modelos lineares de acordo com a correlação entre o termo omitido e o incluído (ou seja, quanto maior a correlação, maior o viés na estimativa), algo central nos debates metodológicos nas Ciências Sociais (ANGRIST; PISCHKE, 2008). Por fim, discutimos estas aplicações no contexto de aulas de graduação e de pós-graduação em Ciência Política e oferecemos um código comentado a fim de que seja possível reproduzir estes exemplos utilizando um software livre para computação estatística, o R¹, além de um aplicativo virtual de simples manuseio - *shiny app* -, que pode ser adaptado para uso em sala de aula².

Nossa expectativa com esse artigo é auxiliar os docentes na apresentação em sala de aula de conteúdos complexos e/ou abstratos de forma mais intuitiva. Adicionalmente, esperamos realçar duas das principais vantagens do uso de SMC como ferramenta didática: de um lado, o uso de simulações permite reduzir o tempo gasto pelo professor com a discussão conceitual; de outro, os alunos conseguem assimilar melhor o conhecimento requerido na área (CARSEY; HARDEN, 2015). De forma geral, portanto, pretendemos contribuir para a formação de pesquisadores em Ciência Política.

Antes de adentrar propriamente na apresentação dos aspectos técnicos da Simulação de Monte Carlo, e na sua aplicação à Ciência Política, é importante destacar que o seu uso com o objetivo de facilitar o processo de ensino-aprendizagem já é explorado por outras áreas do conhecimento técnico-científico: na Química para ensino de luminescência, cinética (WINNISCHOFER *et al.*, 2010); na Metrologia para avaliação de incerteza na medição (DONATELLI; KONRATH, 2005); simulação de eventos discretos para o ensino de Engenharia de Produção (MOLINA *et al.*, 2010); na Física para o ensino de mecanismo de transferência de calor (DIONISIO; SPALDING, 2014). Esses são alguns exemplos da variedade usos da Simulação de Monte Carlo para o ensino de teoria e conceitos abstratos,

¹O software *R* pode ser baixado de forma gratuita no site <https://www.r-project.org/>.

²O *shiny app* pode ser acessado por um navegador de internet qualquer, dispensando, portanto, o uso do R. O endereço da nossa aplicação é <http://denissonsilva.com/simulacao>.

nas próximas seções vamos explorar essa ferramenta para no ensino de Ciência Política de técnicas quantitativa de análise dados como MQO.

MONTE CARLO: UMA BREVE INTRODUÇÃO

Segundo uma definição de manual, simulação Monte Carlo é um processo onde são criadas N variáveis aleatórias com o objetivo de examinar seus resultados de forma agregada (MOONEY, 1997; PRADO, 1999). Em termos mais claros, o método consiste em simular repetidas amostras aleatórias segundo algum tipo de distribuição pré-especificada (i.e., um processo gerador de dados) para que, em seguida, seja possível analisar conjuntamente o resultado obtido a partir de cada uma destas amostras. Um simples jogo de cara ou coroa exemplifica este tipo de simulação: a distribuição teórica da variável de interesse – resultado do lançamento de uma moeda – é binomial, isto é, só podemos obter um resultado discreto de um universo de dois resultados possíveis após o lançamento de uma moeda não viciada – cara ou coroa. Neste caso, poderíamos lançar a mesma moeda 100 vezes (100 simulações) e registrar o resultado do lançamento. Depois, poderíamos calcular algum resultado de interesse, por exemplo, a probabilidade de que o resultado de um lançamento qualquer ter sido cara. Embora simples, este exercício nos ajuda a entender a principal característica de uma simulação de Monte Carlo: a geração de amostras aleatórias (em nosso caso, do lançamento de moedas) repetidas várias vezes segundo algum modelo de probabilidade e a análise posterior dos resultados dos eventos de interesse.

Em vez de coletar dados, formular uma hipótese e testá-la, portanto, as simulações nos permitem gerar dados estipulando como eles devem se comportar. Podemos, por exemplo, simular um determinado fenômeno, como os votos de um partido, ou como os resultados agregados variam conforme alteramos o processo gerador de dados (como o partido concentra seus votos, qual a probabilidade de uma moeda retornar cara, ou qual o grau de associação entre duas ou mais variáveis).

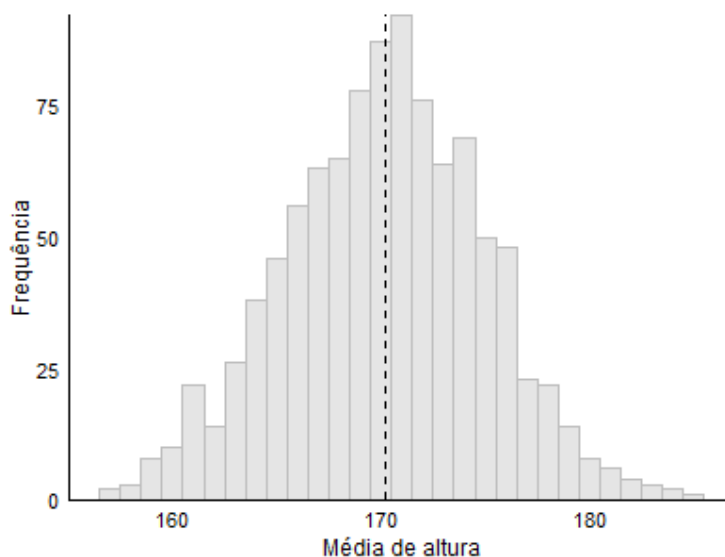
Por conta da possibilidade de controlar o número de simulações e as características das variáveis simuladas, uma das potencialidades do método é justamente o de emular um experimento em seu formato tradicional. Ao repetir o processo de geração de dados e utilizar aleatoriedade para investigar como pequenas variações alteram ou não um resultado, é possível investigar como pequenas mudanças na implementação de uma simulação afetam seus resultados (AXELROD, 1997; MOONEY, 1997). Estes experimentos ideais servem justamente para investigar questões que, de outro modo, seriam difíceis de serem respondidas empiricamente devido à necessidade de uma massiva coleta de dados

que correspondam às situações possíveis de ocorrência de determinado fenômeno, além de permitirem uma ampla gama de aplicações, como testes de modelos estatísticos, testes de hipóteses, entre outros. Como diz Axelrod (1997, p. 24) em seu artigo seminal sobre simulações nas Ciências Sociais, “*Simulation is a way of doing thought experiments. While the assumptions may be simple, the consequences may not be at all obvious*”. Ao utilizarmos SMC, portanto, podemos examinar uma série de questões que de outro modo seriam difíceis de serem examinadas somente a partir da investigação empírica.

Substantivamente, a ideia de fazer várias simulações e analisar seus resultados baseia-se num teorema bastante conhecido no campo da inferência estatística clássica: o teorema do limite central. Dito de forma condensada, este teorema propõe que se gerarmos um número suficientemente grande de amostras aleatórias e calcularmos alguma estatística a partir delas (média, mediana, etc.), o valor médio destas estatísticas indicará o valor real deste parâmetro na população. Um exemplo muito simples ajuda a clarificar este ponto: podemos sortear 10 pessoas dentro da população total do Brasil e medir a altura média da população a partir desta amostra. Basicamente, o teorema do limite central nos diz que, se repetirmos este procedimento várias vezes, a média dos resultados de todas as médias amostrais convergirá para o valor real da média de altura no Brasil – mesmo que 10 indivíduos seja, de modo geral, um número muito pequeno.

O gráfico 1 ilustra o exemplo anterior, onde estipulamos que a distribuição da altura da população brasileira segue uma distribuição normal, com média de 170 cm e desvio-padrão de 15 cm; a partir destes valores, simulamos 1000 amostras contendo 10 pessoas e calculamos a média de altura desta; por fim, fizemos um histograma desta distribuição para verificar os resultados que, como pode ser visto, estão todos ao redor da média estipulada. Como ficará claro na próxima seção, este método pode ser generalizado para uma ampla gama de situações.

GRÁFICO 1 – Simulação da média de altura da população brasileira (1000 amostras de 10 indivíduos)



Fonte: elaboração própria.

INGREDIENTES PARA CONSTRUIR UMA SIMULAÇÃO DE MONTE CARLO

Antes de nos determos nas aplicações de SMC para o ensino de Ciência Política, é preciso compreender como uma simulação é construída. De acordo com CARSEY; HARDEN (2015) e Mooney (1997), toda simulação tem três componentes³, ou ingredientes, básicos: (1) um processo gerador de dados (PGD); (2) a simulação de N amostras a partir deste processo; e (3) a análise dos resultados agregados – e, sendo o caso, a redefinição do processo gerador de dados e conseqüente repetição do processo. Discutimos estes aspectos em ordem, mostrando etapa por etapa como o exemplo anterior sobre a altura média da população brasileira foi construído.

Passo 1 – Definindo um Processo Gerador de Dados (PGD)

Um *processo gerador de dados* nada mais é do que a forma como definimos como um fenômeno será gerado. Em estudos eleitorais, por exemplo, poderíamos dizer que a votação de um candidato pode ser expressa da seguinte forma:

$$Votos_i = \beta_0 + \beta_1 Qualidades\ pessoais_i + \beta_2 Gastos\ de\ campan\ a_i + \epsilon_i$$

³Na verdade, tanto Carsey e Harden (2015) quanto Mooney (1997) reportam cinco itens, mas dois deles podem ser condensados sem perda de generalidade.

Simplificadamente, este modelo pode ser pensado como uma regressão linear por mínimos quadrados, onde as variáveis à direita são os fatores que acreditamos influir na votação de um candidato i . Numa simulação de Monte Carlo, precisamos definir previamente este processo e, a partir dele, simular as amostras. Um candidato pode ter qualidades pessoais igual a 9, num indicador sintético hipotético que varie segundo uma escala de 0 a 10, e gastar 10 mil reais em sua campanha. Essas informações consideradas de acordo com os parâmetros do modelo que definirmos, lhe renderá Y_i votos. Dependendo de como construirmos o processo gerador de dados, podemos estabelecer que cada ponto a mais no indicador de qualidades pessoais gera, em média, 1000 votos ($\beta_1 = 1.000$) e que cada real a mais gasto na campanha gera 10 votos ($\beta_2 = 10$). Assim, nosso candidato hipotético teria $(1.000 * 9) + (10 * 10.000) = 109.000$ votos. A partir disso, podemos simular diversos outros candidatos, com qualidades pessoais e gastos de campanhas gerados aleatoriamente a partir de alguma distribuição teórica de probabilidades, como a normal, por exemplo.

No caso do exemplo oferecido na seção anterior, sobre a altura média dos brasileiros, nosso processo gerador de dados é bastante simples: especificamos uma média de altura fictícia da população brasileira (170 cm) e seu desvio-padrão (15 cm). Nossos dois parâmetros de simulação, neste caso, são estas duas estatísticas que, em conjunto, indicam que estipulamos a distribuição de altura dos brasileiros como seguindo uma distribuição normal (variável contínua) em que as alturas das pessoas variam de forma simetricamente decrescente em torno da média de 170 cm - que é exatamente o que o Gráfico 1 reporta. De qualquer modo, cabe reforçar, este processo gerador de dados que estipulamos é completamente modificável, significando que podemos estipular quaisquer valores de altura média e desvio-padrão na população. Parte da força de SMC reside justamente nesta possibilidade de poder alterar completamente o processo gerador de dados de acordo com as necessidades da investigação

Uma das principais vantagens deste método, portanto, é forçar o pesquisador a pensar de modo detido em como um fenômeno é gerado, isto é, deixar de pensar no modelo estatístico usado para testar uma hipótese e, ao invés, procurar entender como a natureza produz determinado fenômeno para, aí sim, procurar a melhor ferramenta para analisá-lo (CARSEY; HARDEN, 2015). De forma semelhante, esta forma de analisar um fenômeno simplifica o entendimento dos modelos de hipóteses, já que permite entender como mínimas alterações no processo gerador de dados alteram os resultados obtidos.

Passo 2 – Simulando amostras aleatórias

O segundo passo num simulação de Monte Carlo é a geração de um número suficientemente grande de amostras a partir do processo gerador de dados definido no passo 1. Aqui, o objetivo é extrair um número de amostras aleatórias que nos permitam analisar os resultados agregados com precisão. Neste sentido, o número de amostras simuladas é o componente central desta etapa: se forem simuladas poucas amostras, teremos maiores dificuldades de distinguir o que é acerto do que é erro aleatório (poder estatístico); por outro lado, simular muitas amostras normalmente consome bastante tempo e recursos de processamento. Com os processadores que temos hoje e as facilidades de alguns *softwares* e linguagens de programação, este último aspecto normalmente não é um problema em simulações mais simples, que envolvem calcular alguma estatística básica, como a média; porém, quando se trata da estimação de modelos mais complexos, este é um aspecto que deve ser levado em consideração. Contudo, como regra geral, considera-se normalmente que algo entre 1.000 e 10.000 simulações seja suficiente para a maioria dos problemas – no exemplo utilizado na seção anterior geramos 1.000 amostras de 10 pessoas.

Passo 3 – Analisando os resultados

Nesta etapa, o objetivo é analisar as amostras geradas anteriormente de forma agregada. Enquanto examinamos o resultado de uma regressão multivariada por meio de uma série de estatísticas (coeficientes, p-valores, ajuste, resíduos), numa simulação normalmente estamos interessados numa única estatística: ou a distribuição dos coeficientes de uma das variáveis num modelo, ou a distribuição dos p-valores, ou a distribuição dos ajustes, etc. Isto significa que, em geral, o foco das simulações está em saber como determinada estatística se comporta quando é aplicada a uma amostra aleatória gerada a partir de um dado processo gerador de dados. De todo modo, nada impede que se analisem mais de uma estatística dos dados gerados, embora isto dependa, essencialmente, do objetivo a ser alcançado com a simulação.

No exemplo sobre a altura da população brasileira estávamos interessados em saber se, na média, 1.000 amostras de 10 pessoas seriam capazes de nos indicar a média real da altura da população como um todo. Conforme é possível verificar, algumas amostras apresentaram valores menores ou maiores do que o esperado – algo normal, dado que uma amostra com 10 indivíduos é pequena em relação ao tamanho da população e qualquer ocorrência de um valor extremo em relação ao parâmetro esperado afeta o resultado da

estatística; ao ampliar o número de amostras, por outro lado, o valor médio obtido de todas elas se aproxima daquele estipulado de antemão, 170 cm.

O quadro 1, abaixo, resume as etapas a formulação de uma simulação de Monte Carlo discutidas até aqui.

QUADRO 1 – Etapas na construção de uma simulação

Etapa	Descrição da etapa
Definição do processo gerador de dados (PGD)	Definição das variáveis e parâmetros necessários para simular um fenômeno de interesse
Simulação das amostras	Simulação de um número suficientemente grande de amostras a partir do processo gerador de dados definido previamente
Análise dos resultados	Análise agregada de uma ou mais estatísticas obtidas de cada amostra simulada

Fonte: Elaboração dos autores.

Programando simulações de Monte Carlo

Com todas estas etapas concluídas – definição do processo gerador de dados, estipulação do número e da estatística examinada nas amostras –, é necessário programar a simulação. Até o início dos anos 90, utilizar computadores nem sempre era a opção principal neste aspecto. Para fazer um dos estudos mais citados com uso de simulação nas Ciências Sociais até hoje, Schelling (1969) utilizou apenas moedas para gerar suas amostras aleatórias. Hoje, a principal opção recomendada pelos livros-texto sobre simulações – incluso as simulações baseadas em agentes – é a solução computacional por meio de linguagens de programação. Como o foco deste artigo está em mostrar o potencial didático de SMC, não nos deteremos na discussão sobre o melhor *software* ou linguagem de programação para este objetivo. Nossa opção neste artigo é pelo ambiente de programação *R*, que tem cada vez mais usuários na Ciência Política e é uma linguagem que apresenta uma série de vantagens para aplicação em análises estatísticas.

Assim como outras linguagens de programação, o *R* dispõe de funções de controle de fluxo e algumas funções nativas para gerar variáveis aleatórias que são indispensáveis numa simulação. Dentre as mais utilizadas, *norm* e *rbinom* são usadas para criar variáveis aleatórias contínuas com distribuição normal e binomiais, respectivamente, e permitem ao usuário especificar algumas de suas propriedades, como média, desvio-padrão e probabilidade de ocorrência de um evento. Outra característica que torna vantajoso o uso do *R* para simulações é a vetorização, que permite realizar uma série de operações de forma mais rápida do que em outras linguagens de programação. Estas e outras características do

R, bem como uma introdução à linguagem e ao ambiente de programação (em português), são apresentadas detalhadamente em Aquino (2014).

Como avaliar se uma simulação é bem-sucedida?

Tratando do método de modo mais amplo, Axelrod (1997) sugere três critérios para avaliarmos uma simulação: (1) validade, (2) usabilidade e (2) extensibilidade. O primeiro aspecto diz respeito ao que podemos entender como validade interna da simulação: ela realmente dá conta de simular o fenômeno ou amostras de interesse? Ou, por outro lado, algum erro não detectado na construção do modelo (ou erro de implementação) influíram nos resultados obtidos? Caso algum procedimento técnico na programação de uma simulação contenha erros, estes possivelmente gerarão problemas de validade.

O segundo aspecto é a usabilidade da simulação. Quando construímos um modelo, ele deve ser compreensível, fácil de ser reproduzido e adaptado por outros usuários. Inclusive, parte do uso de simulações como recurso didático passa pela replicação de alguns modelos mais simples para que, posteriormente, sejam adaptados ou reformulados pelos usuários. Portanto, na medida em que não é clara a implementação de uma simulação, ela traz prejuízos à usabilidade. Por fim, o terceiro item a ser avaliado é a extensibilidade, que diz respeito à capacidade de generalização e adaptação da simulação para aplicações em outras questões ou atividades.

EXEMPLOS DE SIMULAÇÕES DE MONTE CARLO

Nesta seção, ilustramos o potencial das SMC como recurso didático no ensino de Ciência Política por meio de dois exemplos. O primeiro é mais substantivo, tratando de como alterações em procedimentos legislativos podem afetar drasticamente a quantidade de proposições que um congresso é capaz de examinar e aprovar. O segundo exemplo, de caráter metodológico, ilustra como violações dos pressupostos de uma regressão linear por mínimos quadrados podem enviesar as estimativas obtidas por meio dele. A ideia geral é mostrar que simulações podem ser usadas para ensinar conteúdos abstratos e complexos de forma fácil e intuitiva. Ou seja, em vez de passar aos alunos um guia de ferramentas estatísticas com recomendações sobre seus usos, ou um livro-texto repleto de definições conceituais, por meio de simulações torna-se possível perceber como pequenas mudanças institucionais podem produzir enormes efeitos agregados (primeiro exemplo), bem como visualizar as consequências práticas do uso inadequado de determinada técnica (segundo exemplo).

Simulando um processo legislativo simples

Embora pareça algo não-problemático a primeira vista, aprovar uma lei num congresso moderno como os existentes em países democráticos é um processo complexo (COX, 2000). Em primeiro lugar, porque normalmente existem diversas proposições aguardando para serem apreciadas, o que, já de saída, implica em decidir coletivamente quando cada uma delas será votada. Em segundo lugar, porque os temas de cada votação devem ser discutidos pelos legisladores (ao menos em congressos minimamente democráticos), momento reservado para que se produza informação sobre suas consequências e discussões sobre seus potenciais benefícios, o que toma tempo. Por último, estas proposições são finalmente votadas; e, aqui também, a questão da maioria utilizada para aprová-la necessariamente pode influenciar o resultado final de cada votação.

Todos estes problemas coordenativos fazem parte do conhecimento comum na Ciência Política sobre como congressos funcionam e sobre como determinadas regras e instituições ajudam na superação deles. Em aulas de introdução ao tema, contudo, a discussão sobre estes pontos pode acabar não sendo tão clara, já que questões como preferências, funções de utilidade e efeitos das regras normalmente são abstratas o suficiente para produzir dúvidas em alguns alunos. Uma forma de contornar isto é utilizar o SMC para simular um processo legislativo, com um determinado número de legisladores e de proposições que são votadas por meio de diferentes regras. No caso, o exemplo que se segue trata apenas deste último aspecto, mas é fácil generalizá-lo para abarcar outros.

O primeiro passo aqui é criar nosso processo gerador de dados a fim de simular um plenário. Conforme dito anteriormente, o que é necessário para isto é estipular quais variáveis entrarão na simulação e como elas serão distribuídas. Para este exemplo, nossa unidade natural é o legislador, que, de forma simplificada, tem duas opções: votar *sim* ou *não*. Podemos definir um conjunto de legisladores (unidades) indexados por i que votam j proposições de forma binária (1 para *sim*, 0 para *não*). Deste modo, nossa variável de interesse assume uma distribuição binomial, e estabelecemos que a probabilidade de um legislador qualquer votar sim é $0,5^4$.

Para exemplificar como as maiorias requeridas para aprovação afetam a probabilidade de que uma proposição legislativa qualquer seja aceita, realizamos duas simulações: na primeira, empregamos o critério de unanimidade, segundo o qual só são aprovadas proposições endossadas por todos os legisladores no plenário; na segunda, utilizamos o

⁴Para um exemplo de simulação de processo legislativo em que os legisladores votam como se fosse um lançamento de moeda (probabilidade de votar sim igual a 0,5), ver Saiegh (2009).

critério de maioria simples, isto é, pelo qual 50% + 1 dos legisladores são suficientes para aprovar uma proposição. Em cada caso, 1.000 proposições são votadas por cinco legisladores. Abaixo, segue um resumo da simulação, de acordo com os passos discutidos anteriormente:

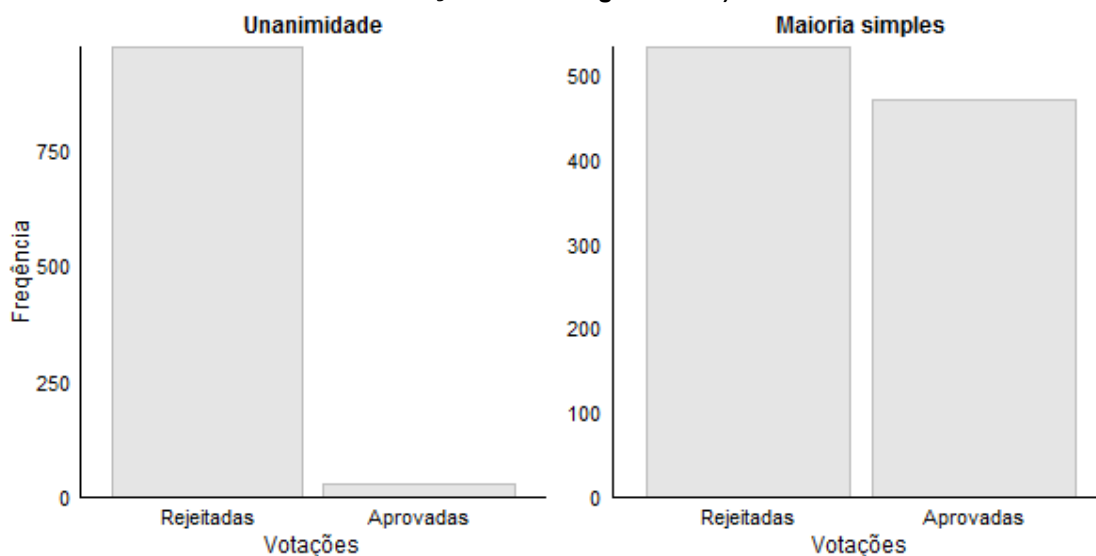
Processo gerador de dados: 5 legisladores votam proposições de forma binária (sim ou não) com probabilidade de votar sim igual a 0.5.

Número de simulações: 1.000 (para cada regra de maioria)

Estatística de interesse: proporção de proposições aprovadas em cada regra de maioria (unanimidade ou maioria simples).

Os resultados deste exercício seguem no Gráfico 2. De forma resumida, o efeito da regra de aprovação utilizada num plenário onde os deputados decidem seus votos de forma aleatória é enorme: enquanto que com regra majoritária quase metade das proposições votadas são aprovadas, num legislativo que adota regra de unanimidade apenas uma ínfima parcela das proposições o são. Intuitivamente este resultado faz sentido, já que todos os legisladores individualmente possuem poder de veto, o que reduz drasticamente o espaço possível de mudanças no *status quo* (TSEBELIS, 2002). Embora estes resultados sejam extremos, como parte da aplicação é possível utilizar outras regras de maioria, proposições, legisladores e, inclusive, alterar as preferências dos legisladores o que definiria de antemão a probabilidade de votarem de um modo ou de outro. Com as mesmas ferramentas, portanto, diversos aspectos de um processo legislativo podem ser explorados.

GRÁFICO 2 – Simulação de um processo legislativo com diferentes regras de maioria (1000 votações com 5 legisladores)



Fonte: elaboração própria.

Simulando regressões lineares com variáveis omitidas

Em nosso segundo exemplo, ilustramos como uma violação bastante comum no uso de modelos de regressão pode enviesar as estimativas obtidas por meio dele: o problema da variável omitida. Conforme alertam os manuais de metodologia (ANGRIST; PISCHKE, 2008), omitir uma variável correlacionada com o termo de erro gera diferentes graus de viés nos coeficientes de uma ou mais variáveis incluídas, embora a magnitude e o sentido deste viés normalmente não possam ser detectados de antemão. O problema com esta definição é a dificuldade de explicar a razão disto aos alunos. Variações da explicação, como dizer que outras variáveis absorvem o impacto da variável omitida ou derivar a prova formal de que a omissão tem potencial de enviesamento, nem sempre são particularmente úteis e facilmente assimiláveis. Como o foco de curso de Ciências Sociais e afins normalmente não está no treinamento formal requerido para compreender métodos quantitativos, SMC nos permitem ilustrar, na prática e de forma gráfica, o que acontece quando um ou mais pressupostos da regressão são violados, contornado dificuldades de compreensão.

Para este exemplo, nos aproveitamos de um tipo muito simples de omissão de variável de um modelo linear a fim de mostrar claramente este efeito. Basicamente, o procedimento consiste em simular duas variáveis com algum grau de correlação entre elas com o objetivo de gerar uma terceira variável, que servirá como variável dependente do modelo a ser estimado. Como as duas variáveis foram usadas para criar a dependente, necessariamente (de acordo com os pressupostos do modelo) servirão para explicar a variação desta. Contudo, ao omitirmos uma delas, as estimativas que obtivermos serão enviesadas.

O processo gerador de dados neste exemplo consiste no seguinte. Primeiro, criamos duas variáveis contínuas, X e Z , com distribuição normal e $N = 100$. Por definição, criamos estas duas variáveis com correlação de Pearson de 0,5. A partir destas duas variáveis, criamos um variável dependente (Y) também contínua, segundo o modelo:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i$$

onde β_0 representa o intercepto do modelo (o valor que Y_i deve ter quando X_i e Z_i são iguais a 0), que definimos como igual 1, β_1 e β_2 também foram estabelecidos como 1 e ϵ_i é um termo de erro com distribuição normal, média 0 e desvio-padrão de 1. Cada uma das simulações neste exercício, portanto, criará aleatoriamente três variáveis que, em conjunto, serão utilizadas para determinar o valor de Y_i . Nosso objetivo aqui, deste modo, é estimar uma regressão linear para cada uma destas simulações apenas com a variável X como variável independente, isto é, omitindo Z . Neste exemplo, rodamos 1.000 simulações. O desenho desta simulação segue abaixo:

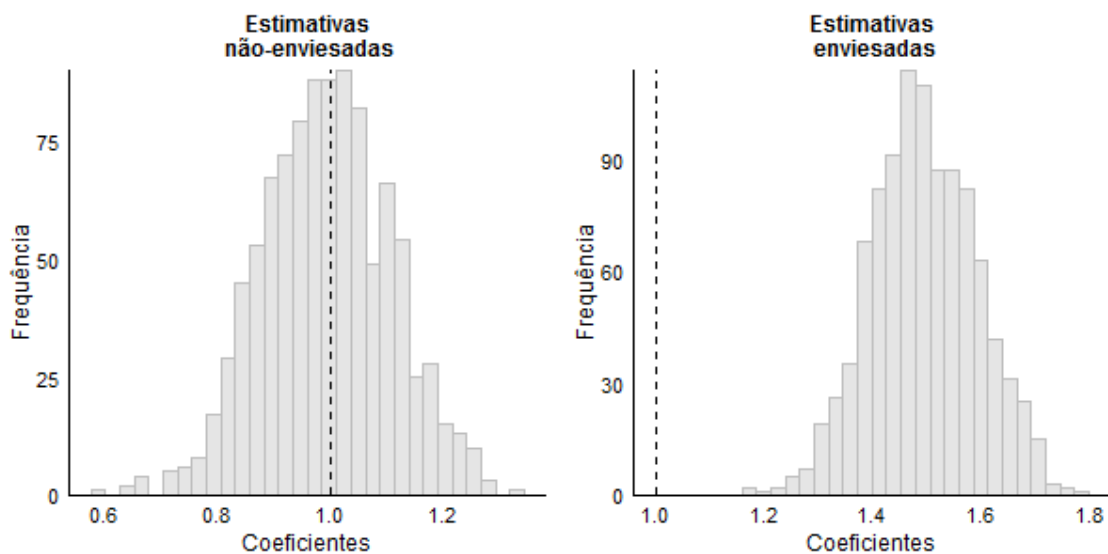
Processo gerador de dados: 3 variáveis contínuas com distribuição normal, média 0 e desvio-padrão de 1; X_i e Z_i têm correlação de 0.5 entre elas; uma quarta variável é criada a partir das outras, Y_i , e é igual a $1 + 1 * X_i + 1 * Z_i + \varepsilon_i$;

Simulações: 1.000;

Estatística de interesse: o coeficiente estimado da variável X quando Z é omitida.

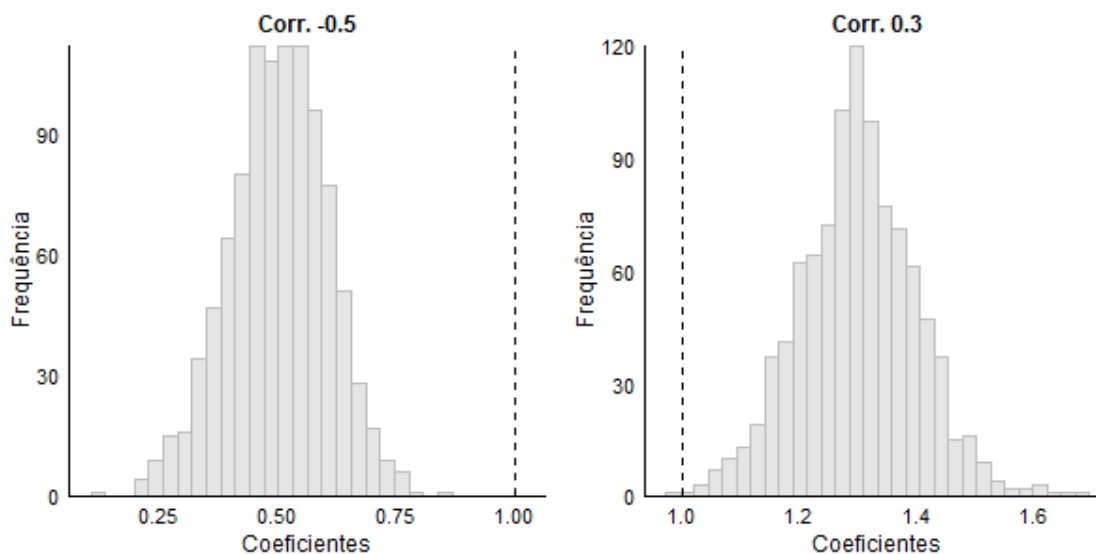
Os resultados deste segundo exercício seguem no Gráfico 3. Como é possível observar, nas regressões estimadas que não incluíram a variável Z , a estimativa média do coeficiente de X obtida ficou longe do valor estipulado para β_1 de 1. Ao contrário, incluindo Z na especificação do modelo, o coeficiente se aproxima, na média, do valor definido. Além disso, este exemplo fornece outra pista aos alunos sobre a direção e a magnitude do viés de variável omitida: tendo definido a correlação entre X e Z de 0,5, a estimativa média do efeito de X sobre Y quando omitimos Z foi 0,5 maior do que o esperado. A repetição da simulação com correlação entre X e Z de -0,5 e 0,3, como fizemos no Gráfico 4, mostra que, também nestes casos, a correlação entre o termo omitido e o incluído tem impacto direto nas estimativas que obtemos ao estimar modelos lineares. De forma geral, este exercício de simulação exemplifica de forma gráfica, sem muitos detalhes, o efeito prático que a violação de um pressuposto de um modelo linear pode ter. Assim como no exemplo anterior, esta simulação também pode ser facilmente adaptada para examinar outras questões, como a inclusão de outras variáveis, uso de outros modelos e teste de outros pressupostos.

GRÁFICO 3 – Simulação de viés de variável omitida (1.000 regressões)



Fonte: elaboração própria.

GRÁFICO 4 – Simulação de viés de variável omitida com diferentes níveis de correlação (1.000 regressões)



Fonte: elaboração própria.

Outros possíveis exemplos

Os exemplos que oferecemos neste artigo são apenas algumas das possíveis aplicações de SMC no ensino de Ciência Política. De qualquer forma, eles podem servir como modelos mais gerais para a criação de outras aplicações. No exemplo do processo legislativo, por exemplo, outros tipos de regra de maioria podem ser utilizadas para analisar seus efeitos sobre a produção legislativa; e de forma semelhante, não é difícil adaptar este exemplo para o funcionamento de comissões legislativas, onde um grupo menor de legisladores vota uma proposição antes dela ir à votação no plenário. Já no exemplo sobre variável omitida, nada impede que se testem outros pressupostos de uma regressão linear, bem como de outros tipos de modelos. Para facilitar a adoção de simulações como estas, o código para replicar nossos exemplos estão disponíveis na internet, e podem ser adaptados para abarcar outras aplicações.

CONCLUSÃO

Neste artigo, procuramos mostrar como o SMC pode ser usada como ferramenta didática em aulas de Ciência Política. Como os exemplos oferecidos procuraram mostrar, simulações tornam palpáveis conteúdos abstratos, permitindo aos alunos visualizar graficamente fenômenos complexos, como um processo legislativo com diferentes regras de maioria. Além disso, os alunos conseguem adquirir intuição sobre como funcionam modelos estatísticos e melhor compreensão sobre processos geradores de dados, como o exemplo

sobre variáveis omitidas procurou mostrar. Neste espaço final, também cabe discutir outras vantagens não abordadas anteriormente, que dizem respeito principalmente à organização das aulas.

O uso de SMC em sala de aula pode ser entendido como um esforço para tornar o aprendizado mais dinâmico. Neste sentido, estas simulações guardam semelhanças com outros tipos de simulações, como jogos em sala de aula, que vêm sendo cada vez mais adotadas na Ciência Política (FREDERKING, 2005). Exemplos destas vão desde jogos para simular transições democráticas, onde são atribuídos papéis aos alunos numa negociação hipotética, até simulações de processos eleitorais. Porém, enquanto que, de forma geral, estas simulações ativas têm o potencial de engajar alunos ativamente no aprendizado, normalmente ocupam muito tempo e demandam um planejamento de aula bastante detalhado. Neste ponto, SMC representam um meio-termo entre uma aula tradicional expositiva e estas com simulações ativas, ou seja, ao mesmo tempo em que utilizam a ideia de criar um mundo fictício totalmente manipulável, SMC podem ser visualizadas graficamente, reproduzidas e adaptadas facilmente (especialmente quando programadas em linguagens como *R* ou *Python*).

Outra vantagem de SMC é que elas podem ser usadas em aulas de metodologia, onde a simples exposição do conteúdo, desprovida de exemplos, normalmente é pouco eficaz e extremamente exaustiva. Ao invés de fazer um inventário de situações nas quais cada técnica estatística discutida deve ser aplicada, o uso de simulação de Monte Carlo põe o foco em desenvolver uma percepção visual sobre como cada uma dessas técnicas funciona e sob quais pressupostos. Exercícios complementares, também baseados em simulações, podem então ser desenvolvidos e aplicados em sala de aula, permitindo aos alunos explorar outros exemplos e refazer cada simulação inúmeras vezes.

REFERÊNCIAS

ANGRIST, Joshua D.; PISCHKE, Jörn-Steffen. **Mostly harmless econometrics: An empiricist's companion**. Princeton University Press, 2008.

AQUINO, Jakson Alves de. **R para cientistas sociais**. Ilhéus, BA: Editus, 2014.

AXELROD, Robert. Advancing the art of simulation in the social sciences. In: **Simulating social phenomena**. Springer Berlin Heidelberg, 1997. p. 21-40.

CARSEY, Thomas M.; HARDEN, Jeffrey J. Can you repeat that please?: Using Monte Carlo simulation in graduate quantitative research methods classes. **Journal of Political Science Education**, v. 11, n. 1, p. 94-107, 2015.

COX, Gary W. On the effects of legislative rules. **Legislative Studies Quarterly**, p. 169-192, 2000.

DONATELLI, Gustavo Daniel; KONRATH, Andrea Cristina. Simulação de Monte Carlo na avaliação de incertezas de medição. 2005.

DIONISIO, Guilherme; SPALDING, Luiz Eduardo Schardong. MÉTODO COMPUTACIONAL DE MONTE CARLO ADAPTADO COMO RECURSO DIDÁTICO PARA O ESTUDO DOS MECANISMOS DE TRANSFERÊNCIA DE CALOR. IV Simpósio Nacional de Ensino de Ciência e Tecnologia, 2014.

FREDERKING, Brian. Simulations and student learning. **Journal of Political Science Education**, v. 1, n. 3, p. 385-393, 2005.

MOLINA, Carlos Eduardo Corrêa; MARINS, Fernando Augusto Silva; MONTEVECHI, José Arnaldo Barra. PROPOSTA DE UTILIZAÇÃO DA SIMULAÇÃO A EVENTOS DISCRETOS NO ENSINO DA ENGENHARIA DE PRODUÇÃO. **Revista P&D em Engenharia de Produção V**, v. 8, n. 01, p. 11-15, 2010.

MOONEY, Christopher Z. **Monte Carlo Simulation**. Sage Publications, 1997.

PRADO, Darci. Teoria das Filas e da Simulação. **Belo Horizonte, MG: Editora de Desenvolvimento Gerencial**, v. 2, 1999.

SAIEGH, Sebastian M. Political prowess or “Lady Luck”? Evaluating chief executives’ legislative success rates. **The Journal of Politics**, v. 71, n. 04, p. 1342-1356, 2009.

SCHELLING, Thomas C. Models of segregation. **The American Economic Review**, v. 59, n. 2, p. 488-493, 1969.

TSEBELIS, George. **Veto players: How political institutions work**. Princeton University Press, 2002.

WINNISCHOFER, Herbert *et al.* Simulação Monte Carlo no ensino de luminescência e cinética de decaimento de estados excitados. **Quim. Nova**, v. 33, n. 1, p. 225-228, 2010.