

As implicações metodológicas da Revolução da Credibilidade e Inferência Bayesiana para a pesquisa qualitativa causal

Manoel Galino

2024-06-30

Abstract

Este trabalho examina as implicações da Revolução da Credibilidade para as pesquisas qualitativas causais. O objetivo é demonstrar que a suposta superioridade da metodologia quantitativa sobre a qualitativa, implícita no trabalho de KKV e na visão de muitos cientistas sociais, é sem fundamento e ignora a moderna ciência da inferência causal. Com base na distinção entre identificação causal e inferência estatística, e no uso da inferência Bayesiana, o artigo evidencia que muitas críticas aos métodos qualitativos são improcedentes, especialmente a suposta incapacidade de generalizar resultados, que também está presente em estudos quantitativos. As conclusões destacam que a Revolução da Credibilidade, combinada com a inferência Bayesiana, permite o fortalecimento da pesquisa qualitativa quando seriamente incorporada. O trabalho apresenta ilustrações de metodologias qualitativas aplicadas que apontam esses novos desenvolvimentos, como o rastreamento de processo Bayesiano e queries causais.

Introdução

Pesquisas qualitativas voltadas ao estudo da causalidade, como os que empregam métodos comparativos de casos (Ragin 2014; Rihoux and Ragin 2009) ou *process tracing* (Bennett and Checkel 2015; Collier 2011) frequentemente são forçadas a reconhecer limitações em relação a pesquisas quantitativas, em especial a da capacidade de generalização. Podemos dizer que é uma consequência do trabalho clássico de King, Keohane, and Verba (1994), conhecido como KKV. O livro pretendeu unificar a lógica da pesquisa científica em ciências sociais a partir do paradigma quantitativo em vigor na época em que o livro foi escrito. Ironicamente, ao mesmo tempo em que o livro chegava às prateleiras e era incorporado ao discurso científico das ciências sociais, uma revolução nos métodos quantitativos com fins causais se iniciava na economia.

Essa transformação, conhecida como Revolução da Credibilidade, alterou profundamente o paradigma metodológico da inferência causal não apenas na economia, mas também em todas as disciplinas sociais de maneira mais ampla. Contudo, não se notou que poderia haver impacto também sobre pesquisas qualitativas causais como a mencionadas acima.

A revolução da credibilidade enfatiza a importância de suposições críveis e transparentes para assegurar que os efeitos causais identificados sejam confiáveis, aproximando-se dos resultados de experimentos aleatórios controlados. Esta mudança de paradigma traz importantes implicações para a pesquisa qualitativa empírica com objetivos causais que não foram devidamente apreciadas. Em vez de classificar tais métodos qualitativos como intrinsecamente inferiores, o presente trabalho mostra que o entendimento produzido pela revolução da credibilidade sobre inferência causal coloca muitas das críticas às pesquisas qualitativas como infundadas e baseadas em mal-entendidos sobre o que constitui uma inferência causal válida.

Em paralelo à revolução da credibilidade, a inferência estatística, tradicionalmente dominada pelo paradigma frequencista, se abriu à escola Bayesiana com o advento dos modernos computadores pessoais. E uma característica da inferência Bayesiana é tornar supérflua a necessidade de amostras grandes para se fazer inferência.

Como mostramos no presente trabalho, esses dois desenvolvimentos (revolução da credibilidade

e inferência Bayesiana) tornam em certo sentido obsoleto o paradigma sob o qual KKV orientaram sua tentativa de unificar a ciência. Ainda assim, o eco do clássico de metodologia de King et al. (1994) ainda se faz ouvir, especialmente para pesquisadoras que empregam métodos qualitativos.

Vejam o que dizem (Amorim and Rodriguez 2016:1007) ao advogar o método (qualitativo) histórico comparativo (MHC):

Note-se que a resposta a ser dada à questão típica do MHC implicará uma generalização contingente à região e ao período em tela. Já a resposta a ser dada à pesquisa típica da MRM [Método de Regressão Multivariada, inserção minha] se valerá, muito provavelmente, de uma terminologia mais universal: quanto mais fragmentado o sistema partidário, menos duráveis são os gabinetes das democracias parlamentares.

Como se vê, mesmo em um artigo em que advogam por um método qualitativo, os autores apontam limitações na capacidade de generalização que não estariam presentes em métodos quantitativos, como preconizavam KKV. De fato, pesquisadoras quantitativistas assim enxergam os métodos quantitativos. Eis o que diz o professor Jerônimo Muniz, em entrevista:

A boa pesquisa científica também deve ser reproduzível, validada por pares e, idealmente, permitir generalizações. Todas estas características, exceto a última, são plenamente compatíveis com metodologias qualitativas. A única limitação da pesquisa qualitativa é, em função de sua operacionalização, a restrição de suas conclusões a um universo ou amostra específica. Por outro lado, ela tem a vantagem de apresentar maior validade interna e riqueza de detalhes, dificilmente captados pelos métodos quantitativos (Bachini and Chicarino 2018:277).

Entretanto, a Revolução de Credibilidade separou a identificação causal – investigação das suposições necessárias para que uma quantidade causal qualquer possa ser estimável –, da estimação estatística e técnicas de inferência estatística. Ao realizar essa separação, muitas das críticas feitas aos métodos qualitativos, em particular uma suposta incapacidade de generalização, é inválida e deriva-se de incompreensão entre a distinção de identificação causal e inferência estatística.

Feita essa distinção, mostramos também que outra grande transformação no mundo da pesquisa quantitativa, o advento da inferência Bayesiana, permite superar eventuais limitações de pesquisas com n pequeno existentes no paradigma frequentista (van de Schoot et al. 2021). Em outras palavras, mostramos como o uso da inferência Bayesiana permite que as pesquisas qualitativas com n pequeno possam enfrentar o problema da inferência com o mesmo rigor que a pesquisa quantitativa no que tange à inferência causal, não restando mais qualquer pretensão de superioridade da pesquisa empírica quantitativa sobre qualitativa.

Para não fazer apenas uma discussão teórica, apresentamos duas metodologias aplicadas que ilustram o argumento metodológico do texto, o rastreamento de processo Bayesiano de Fairfield and Charman (2022) e a abordagem de *queries* causais de Humphreys and Jacobs (2023). Ambas apontam um caminho possível pelo qual os métodos qualitativos podem se desenvolver e serem aplicados nos próximos anos.

Por fim, apresentamos também a moderna discussão sobre generalização a partir do conceito de transportabilidade e como ele pode ser usado igualmente por pesquisas qualitativas e quantitativas, ou ainda idealmente, sob métodos mistos.

Portanto, este trabalho contribui para um debate mais equilibrado sobre a metodologia nas ciências sociais, destacando como a revolução da credibilidade pode enriquecer a pesquisa qualitativa baseada em estudos de caso ou *process tracing* ao esclarecer as condições necessárias para a identificação causal. Estabelecidas essas condições, apresentamos a utilização da inferência Bayesiana como a solução “natural” para inferência estatística rigorosa com n pequeno. Desse modo, esperamos dissipar críticas infundadas e redirecionar os esforços dos pesquisadoras para questões metodológicas realmente relevantes da pesquisa aplicada em ciências sociais.

Revolução da Credibilidade

A revolução da credibilidade na economia (Angrist and Pischke 2010; Forozish 2024; Goldsmith-Pinkham 2024) impactou várias disciplinas que lidam com dados observacionais, incluindo a ciência política. Um dos trabalhos pioneiros que estabeleceram a agenda para essa revolução

é o artigo de Leamer (1983), “Let’s take the con out of econometrics”, onde o autor critica a prática padrão de especificações econométricas múltiplas para controlar vieses de variável omitida, argumentando que tais práticas não produzem resultados confiáveis.

Respondendo ao desafio de Leamer, vários autores desenvolveram desenhos de pesquisa com suposições mais críveis e transparentes (Angrist and Krueger 1991; Angrist and Pischke 2009, 2010; Card 2022; Card and Krueger 1994). Esses estudos buscaram assegurar que os efeitos causais fossem reais e se aproximassem dos resultados de experimentos aleatórios controlados, elevando o padrão de inferência causal (Rubin 1974).

A revolução da credibilidade se espalhou para outras disciplinas, como sociologia (Lundberg, Johnson, and Stewart 2021), estudos russos (Libman 2023), saúde (Glied 2021) e, claro, ciência política (Grimmer 2015; Keele 2015b, 2015a; Samii 2016). Os impactos na pesquisa quantitativa foram imediatos, com a redução do uso de regressões com muitas variáveis em favor de desenhos quase-experimentais, como diferença em diferenças, regressão descontínua e variáveis instrumentais, para identificar relações causais com maior precisão. Ensaios aleatórios controlados (RCTs) passaram a ser vistos como o padrão-ouro para inferências causais, e houve um desenvolvimento de novas ferramentas estatísticas para implementar essas metodologias.

Para nós, é importante destacar que, sob essa perspectiva, a inferência causal é uma consequência dedutiva de suposições críveis, aproximando-se da visão de causalidade de Hume, onde a causalidade não pode ser observada diretamente. O trabalho do cientista social, portanto, é desenhar estudos com suposições críveis que levem à “identificação causal”.

Identificação Causal

Identificação causal refere-se à capacidade de um desenho de pesquisa determinar um estimador específico, como o efeito causal médio (ATE - Average Treatment Effect) ou o efeito do tratamento sobre os tratados (ATT - Average Treatment on the Treated). A identificação não depende do tamanho da amostra, mas sim da estrutura metodológica do estudo e de suposições críveis.

A preocupação com identificação causal, destacada pela revolução da credibilidade, implicou

uma ênfase na validade interna dos estudos. A validade interna se refere à capacidade de um estudo identificar um efeito causal sem viés na sua amostra específica. Por outro lado, a validade externa refere-se à capacidade de generalizar os achados para outras populações e contextos (Findley, Kikuta, and Denly 2021). Essa distinção introduzida por Campbell (1957) é crucial, mas a revolução da credibilidade muitas vezes sobrevalorizou a validade interna em detrimento da validade externa (Findley et al. 2021).

Potanto, de saída podemos descartar como infundadas duas posições. A primeira é a de que haveria qualquer diferença com relação à validade interna de métodos qualitativos ou quantitativos. Ambos são igualmente capazes de atingir validade interna. Similarmente, as duas metodologias são igualmente limitadas (a princípio) em sua capacidade de possuírem validade externa. E essas duas conclusões seguem do fato de que não há relação entre validade interna e externa e tamanho amostral, se a pesquisa é em profundidade ou não, ou quaisquer das possíveis distinções empregadas para diferenciar método qualitativo de quantitativo.

Soluções Práticas para Inferência em Amostras Pequenas

Não espero, contudo, que alguém treinado na metodologia KKViana seja facilmente convencido apenas por argumentos teóricos sobre a igualdade de limitações e potencialidades das pesquisas causais de n grande e de n pequeno. Afinal, isso não significa necessariamente que a inferência causal pode ser realizada na prática para tamanhos amostrais pequenos, mesmo que um modelo seja identificado. Um cético pode argumentar que ainda será necessário um grande número de casos para estimar o efeito causal de interesse.

Nesse sentido, é fundamental apresentar soluções práticas para o problema da inferência com amostras pequenas. E a própria disciplina da estatística contém a solução para esse problema. Tradicionalmente, a inferência estatística adotou o paradigma frequencista, que depende de propriedades assintóticas dos estimadores, isto é, de quando o tamanho amostral vai para o infinito. Contudo, há desde o começo da disciplina da estatística um outro paradigma alternativo, e que vem ganhando proeminência aplicada nas últimas décadas.

Inferência Bayesiana

Na estatística Bayesiana, a quantificação da incerteza depende de uma distribuição de probabilidade a priori, que usualmente é interpretada como a incerteza da pesquisadora antes de observar evidência empírica, e uma verossimilhança, que é uma modelagem do processo de geração de dados. O teorema de Bayes, que dá nome à abordagem estatística, nos diz como combinar de maneira rigorosa a priori e a verossimilhança, para obtermos uma distribuição de probabilidade a posteriori (isto é, após observar a evidência empírica) sobre as quantidades de interesse.

A rigor, nessa abordagem, podemos quantificar incerteza até mesmo com zero observações, refletidas na priori. E com uma observação apenas já podemos computar a verossimilhança e calcular a posteriori. E não há diferença formal entre n pequeno ou grande. O impacto do tamanho amostral é apenas em reduzir a incerteza a posteriori. É possível mostrar que, sob condições bastante gerais de regularidade, a distribuição a posteriori converge para a distribuição verdadeira à medida que o tamanho amostral vai para infinito. Em outras palavras, a incerteza se reduz com n maior, mas a rigor nada impede de se estimar qualquer quantidade de interesse com um n pequeno.

Na verdade, a questão crítica para a inferência estatística não é simplesmente o tamanho da amostra (n), mas também a relação sinal-ruído; ou seja, quanto mais forte o efeito (o “sinal”) em relação ao ruído (variabilidade dos dados), menos dados são necessários para detectá-lo.

Não é por outra razão que estudos qualitativos bem desenhados geralmente se concentram em contextos onde o sinal é forte e claramente observável, permitindo inferências robustas mesmo com menos casos. Esse tipo de estratégia visa resolver problemas de inferência, mas nada dizem sobre a validade interna, que estão relacionadas às condições formais de identificação.

Por outro lado, estudos quantitativos com grandes amostras podem ser necessários para detectar efeitos menores, onde o sinal é fraco em relação ao ruído. Portanto, a crítica ao tamanho da amostra na verdade é imprecisa, pois é preciso combinar o tamanho amostral com a razão sinal-ruído, para só então avaliar qual tipo de desenho de pesquisa deve produzir menor incerteza inferencial. Sem analisar o caso concreto, nada pode ser dito sobre um tipo

de pesquisa ou outro.

E a abordagem Bayesiana é uma das formas pela qual se pode quantificar essa relação sinal-ruído e tamanho amostral. Se o sinal for muito fraco para um dado tamanho amostral, haverá reflexo na quantificação da incerteza que será muito elevada. Se por outro lado o sinal for forte o suficiente para o dado número de casos, a incerteza estimada permitirá inferências e conclusões mais assertivas.

Fundamentos da Probabilidade Bayesiana

Para facilitar a compreensão das novas metodologias que integram inferência Bayesiana com inferência causal em estudos qualitativos, é útil apresentar uma rápida introdução (um pouco) formal à maquinaria Bayesiana.

Regra do Produto O teorema de Bayes é na verdade uma mera aplicação da regra do produto de probabilidade, de forma que vale a pena começar por esta regra, para então formular o teorema de Bayes.

A probabilidade conjunta de dois eventos A e B pode ser expressa como $p(AB) = p(A)p(B|A)$ ou $p(B)p(A|B)$. Para três eventos A , B e C , a probabilidade conjunta $p(ABC)$ é decomposta em uma série de probabilidades condicionais: $p(ABC) = p(A)p(B|A)p(C|AB)$.

Teorema de Bayes O Teorema de Bayes permite combinar conhecimento a priori com dados observados para atualizar a crença sobre um parâmetro ou evento. Formalmente, seja H uma hipótese causal, e E a evidência ou conjunto de evidências coletada.

Estamos interessados em combinar quantificar a probabilidade da hipótese ser verdadeira, dada a evidência. Dado que, pela regra do produto, $p(E|H) = p(H)p(E|H) = p(E)p(H|E)$, podemos rearranjar a igualdade e obter: $p(H|E) = \frac{p(H)p(E|H)}{p(E)}$. Aqui, $p(H)$ é a probabilidade a priori da hipótese, $p(E|H)$ é a verossimilhança, isto é, uma quantificação de quanto a evidência é mais ou menos provável, dada a hipótese sobre o mecanismo causal, e $p(E)$ é chamada de constante normalizadora e mede a probabilidade a priori da evidência.

Constante Normalizadora Aplicando a regra da probabilidade total, podemos calcular $P(E)$ como: $P(E) = p(H)p(E|H) + p(\neg H)p(E|\neg H)$. Portanto, posso expressar a probabilidade não-condicional da evidência como uma série de probabilidades condicionais, ponderadas pela probabilidade das hipóteses (no exemplo, a hipótese formulada e sua negação).

Função de Verossimilhança A função de verossimilhança é a probabilidade dos dados observados tratada como uma função dos parâmetros. Pode ser escrita como $L(\theta|x_1, x_2, \dots, x_n) = p(x_1, x_2, \dots, x_n|\theta)$, onde os dados são fixos e o parâmetro θ é aleatório. No caso mais simples de uma hipótese, escrevemos $P(E|H)$. O importante é notar que, apesar de escrever P , não se trata propriamente de uma probabilidade, mas de uma verossimilhança.

Soluções Práticas para Inferência em Amostras Pequenas

Nesta seção, apresentaremos duas soluções recentes que utilizam a metodologia Bayesiana para estimar quantidades causais em pesquisas qualitativas: uma desenvolvida no livro *Integrated Inferences* de Humphreys and Jacobs (2023) (ver também Humphreys and Jacobs (2015)) e outra no livro *Social Inquiry and Bayesian Inference* de Fairfield and Charman (2022) (ver também Fairfield and Charman (2019); Fairfield and Charman (2017)).

Essas soluções deixam claro e sem sombra de dúvidas o argumento do trabalho de que o problema de identificação é completamente distinto e ortogonal ao problema da inferência estatística.

Process Tracing Bayesiano

O rastreamento de processos bayesiano é uma abordagem robusta para inferência causal em estudos qualitativos. Utilizando a probabilidade bayesiana, podemos atualizar nossas crenças sobre hipóteses causais à medida que novas evidências são coletadas. Este método é particularmente útil em contextos onde as evidências são escassas ou os processos causais são complexos.

Assim, envolve a especificação de hipóteses H_i e suas probabilidades a priori $P(H_i)$. Devemos identificar as evidências disponíveis e construir a verossimilhança $P(E|H_i)$ para cada hipótese.

As comparações de hipóteses são feitas usando as odds a posteriori, de forma a não precisa computar a constante normalizadora:

$$\frac{P(H_i|E)}{P(H_j|E)} = \frac{P(H_i)P(E|H_i)}{P(H_j)P(E|H_j)}$$

Vamos tratar como cada um dos componentes da fórmula acima (priori e verossimilhança) são formulados na prática.

Definição de Prioris

A definição de prioris pode ser feita por meio da elicitación de experts, uso de prioris não-informativas e prioris informativas definidas pela pesquisadoras.

A elicitación de experts envolve a consulta a especialistas e utilização de métodos para transformar o conhecimento deles em distribuições de probabilidade. Como tais pessoas não costumam saber traduzir seus conhecimentos em distribuição de probabilidade, a literatura dessa área tem desenvolvido métodos para extrair esse conhecimento e quantificá-lo em distribuições de probabilidade (Ohagan_2019, Albert et al. (2012), Ohagan_etal_2006, Ohagan_1998).

Prioris não-informativas são a forma mais básica e tradicional de definição de prioris (O'Hagan 2010). A ideia é escolher uma priori que torne todas as hipóteses a priori igualmente prováveis, seguindo o princípio da razão insuficiente (Kass and Wasserman 1996).

Uma terceira possibilidade é o uso de prioris informativas, que representem o conhecimento da pesquisadora, que pode advir de conhecimento substantivo ou de resultados de pesquisas prévias como meta-análises ou estudos quantitativos prévios (Gelman 2009; Gelman, Simpson, and Betancourt 2017).

Verossimilhanças

A verossimilhança avalia a plausibilidade de observar uma evidência E sob a hipótese H_i . Os autores recomendam adaptar a ideia de decibéis, que utilizamos para classificar o volume de um som, para avaliar o peso de uma evidência. Ou seja, deveríamos usar logaritmos (os

decibéis são calculados com logaritmos) para calibrar o peso das evidências:

$$10 \log_{10} \left(\frac{P(E|H_i)}{P(E|H_j)} \right)$$

Para ver a utilidade de se utilizar decibéis para ponderar o peso das evidências, os autores analisam um exemplo introduzido por Bennet (2015), no qual o autor considera um *smoking gun* se $P(E|H_j) = .05$ e $P(E|H_i) = .2$. Aplicando o logaritmo, isso dá cerca de 6db. Segundo Fairfield and Charman (2017), um volume de 6db é saliente, mas estaria longe de ser um *smoking gun*.

Evidências Na análise estatística tradicional, é em geral simples determinar quando uma observação termina e outra começa. No caso de estudos qualitativos, em especial rastreio de processo, esse passo é menos trivial.

A definição de evidências é crucial e pode incluir informações de diferentes fontes ou observações correlacionadas. No Bayesianismo, podem ser consideradas evidências de outros casos similares. Evidências devem ser suficientemente distintas para favorecer hipóteses diferentes.

Hipóteses Rivals

A formulação de hipóteses rivais ajuda a reduzir viés de confirmação (Fairfield and Charman 2019). Matematicamente, duas hipóteses rivais significam que $P(H_i) + P(H_j) = 1$. Se a pesquisa possui uma única hipótese, é fácil avaliar erroneamente se determinada evidência empírica torna a hipótese mais ou menos plausível. Digamos que estamos interessados em investigar as causas de um acidente de avião¹. Se minha única hipótese é que foi **apenas** falha mecânica, e encontramos evidências de que a caixa preta do avião registrou altitude errada e que o piloto não tinha dormido nas últimas 24 horas. A primeira evidência suporta a hipótese, enquanto a segunda é neutra. Porém, suponha que a outra hipótese (rival) é que houve uma falha mecânica e falha humana em conjunto. Nesse caso, a primeira hipótese parece menos plausível do que a segunda.

Não basta simplesmente formular uma negação lógica, pois ela inclui infinitas hipóteses

¹Agradeço à Elizabeth Balbachevsky pela sugestão do exemplo

potencialmente contraditórias entre si. Ou seja, não basta dizer que uma hipótese é que houve apenas falha mecânica e a hipótese rival é que não houve apenas falha mecânica, pois a negação aqui inclui houve falha humana e não houve falha humana.

Uma das limitações da abordagem de rastreamento de processo Bayesiana é quando a pesquisadora considera a possibilidade de múltiplas causas, pois pode ser difícil formulá-las como hipóteses rivais. Mas esse é um passo necessário para que a metodologia possa ser empregada.

Inferências Integradas

O trabalho de Humphreys and Jacobs (2023) apresenta o que eles chamam de queries causais, que consiste na utilização de Directed Acyclic Graphs (DAGs), tal como desenvolvida por Pearl, juntamente com inferência Bayesiana, para estimar quantidades causais com métodos qualitativos. O termo inferência integrada é porque, ao utilizar metodologias originariamente quantitativas, permite integrar qualit-quant, produzindo assim desenhos de pesquisa com métodos mistos. Aqui interessa para nós o aspecto qualitativo da abordagem, então será nosso foco.

Um primeiro aspecto que a metodologia de Humphreys and Jacobs (2023) deixa claro é que é necessário, para fins práticos, considerar todas as variáveis da pesquisa quali como binárias. Isso faz sentido, posto que seria inviável mensurar quantidades contínuas e relacioná-las estatisticamente com amostras pequenas. Ou seja, a variância potencial seria muito grande, tornando inviável a estimação.

Um segundo aspecto é que, como é costumeiro, é mais fácil modelar as relações causais como determinísticas, ou seja, um caso possui ou não relação causal entre variáveis, não cabendo pensar probabilisticamente. Isso, contudo, não é uma limitação muito relevante, pois a quantificação da incerteza via Bayes significa que poderemos falar de probabilidade do efeito causal existir. Ou seja, a relação é determinística, mas o nosso conhecimento sobre a relação causal é probabilístico, o que é compatível com boa parte das ontologias sociais.

A suposição de que a relação causal é determinística e binária leva a ideia de que cada caso pode ser classificado em um tipo causal. Os tipos causais podem ser classificados em quatro categorias: adverso, benéfico, crônico e destinado. Cada tipo representa uma relação

causal distinta entre tratamento e resposta. Por exemplo, indivíduos adversos melhorariam apenas se não recebessem o tratamento, enquanto os benéficos melhorariam apenas se o recebessem. Os crônicos permanecerão doentes independentemente do tratamento, e os destinados melhorarão independentemente do tratamento. Esse modelo pode ser facilmente generalizado para múltiplas causas, com uma pequena modificação na notação, a qual o leitor interessado pode encontrar no livro de Humphreys e Jacobs.

Para aplicar essa estrutura teórica, utilizamos modelos de queries causais. Esses modelos nos permitem investigar efeitos causais médios, causas ao nível de casos (case-level causal analysis) e caminhos causais. Uma query causal é um algoritmo que recebe um modelo M como entrada (especificado na forma de um DAG) e retorna uma quantidade $Q(M)$ como saída. Existem quatro queries causais chave que são fundamentais para essa análise e respondem a questões distintas: - Efeitos causais ao nível de caso (case-level). - Atribuição causal ao nível de caso. - Efeitos causais médios (ATE). - Caminhos causais.

Efeito Causal ao Nível do Caso

Perguntas de pesquisa frequentemente investigam se X tem um efeito causal em Y em casos específicos. Por exemplo, se houvesse uma crise econômica no Brasil hoje, Lula sofreria impeachment? Em termos contrafactuais, a questão é se, ao intervir exogenamente no valor de X , o valor de Y mudaria. Ou seja, qual seria o tipo causal do Brasil sob Lula III? Se não haveria impeachment com ou sem crise, e considerando impeachment como 1 e não-impeachment como 0, então seria um tipo crônico, pois o impedimento não seria observado com e sem crise econômica. Em outras palavras, poderíamos estar interessados na probabilidade de ser crônico. Ou talvez a pergunta de pesquisa seria em torno da probabilidade do país ser benéfico, isto é, o impeachment ocorreria com crise econômica, mas não em sua ausência.

Esse tipo de query nos ajuda a entender as implicações de mudanças específicas em variáveis causais. O modelo de queries causais aqui apresentado permite responder a esse tipo de pesquisa, tipicamente qualitativas, de maneira quantitativamente precisa, a partir de estudo de caso.

Atribuição Causal ao Nível do Caso

Além de investigar se X tem um efeito causal, também é importante determinar se X de fato causou Y em casos específicos. Por exemplo, a crise econômica causou o impeachment da Dilma em 2016? Em termos contrafactuais, a questão é se, dado os valores que X e Y assumiram, o valor de Y seria diferente se o valor de X fosse diferente. Este tipo de análise é crucial para entender a relação direta entre causas e efeitos em contextos específicos.

Efeito Causal Médio

Podemos também explorar efeitos causais médios em populações. Para tratamentos e respostas binárias, o efeito causal médio é a diferença entre a proporção de casos em que o efeito é benéfico e a proporção de casos em que o efeito é negativo. Esta abordagem nos permite entender o impacto geral de tratamentos em grandes grupos. Essa é um dos casos em que a abordagem qualitativa mais facilmente é integrada com estudos quantitativos.

Caminhos Causais

Em casos onde estabelecemos que X causou Y , é relevante investigar se isso ocorreu via um mediador M . Em notação de resultados potenciais, perguntamos se $Y(X = x, M = M(X = 1))$ é maior do que $Y(X = x, M = M(X = 0))$. Este tipo de query nos ajuda a mapear as vias por meio das quais os efeitos causais se manifestam.

Causas de Fato/Reais/Atuais (Actual)

A definição de causalidade contrafactual típica dos modelos de resultados potenciais pode gerar paradoxos, como no seguinte exemplo. Suponha que Suzy e Bob jogam pedras em uma garrafa. Ambos jogam a pedra no alvo, de modo que ambas atingiram a garrafa, exceto que a de Suzy quebra a garrafa e a de Bob chega um pouco depois. Nesse caso, é óbvio que a pedra de Suzy foi a causa da quebra da garrafa. Porém, na noção de causalidade contrafactual, diríamos que a pedra de Suzy não foi a causa da quebra da garrafa, pois a causalidade é definida como a diferença entre o que aconteceu de fato e o que aconteceria no contrafactual. Nesse caso, se Suzy não tivesse jogado a pedra, a garrafa quebraria do mesmo

modo. Assim, a utilização de queries causais permite tratar desse tipo de causa, algo não possível na abordagem mainstream de resultados potenciais. E para distingui-la da definição de causa contrafactual, os autores chamam de causas de fato (ou atuais) este tipo de causa, cujo efeito depende de outra variável, como a pedra de Bob não atingir a garrafa antes. Este conceito nos permite diferenciar entre causas diretas e indiretas em cenários complexos.

Comparando as Abordagens

As duas metodologias aqui apresentadas são exemplos de desenvolvimento recentes de utilização de inferência Bayesiana para estudos qualitativos. Os trabalhos de Fairfield e Charman têm como objetivo formalizar o rastreamento de processos, metodologia qualitativa que tem se desenvolvido bastante nas últimas décadas. Ela é particularmente útil quando as hipóteses são de fato rivais e permite avaliar hipóteses causais concorrentes para explicar fenômenos sociais nos quais apenas o rastreamento de processo poderá recuperar o efeito causal de interesse.

O modelo de queries causais de Humphreys e Jacobs é mais genérico, podendo ser aplicado a muitos diferentes tipos de pesquisas qualitativas, incluindo rastreamento de processo. Também utiliza inferência Bayesiana para estimação e quantificação da incerteza, mas não possui a limitação de restringir-se a situações de hipóteses rivais, admitindo hipóteses complementares. Por outro lado, é limitado a variáveis categóricas (pode ser estendido a mais de duas categorias, mas rapidamente se torna inviável em termos de implementação algorítmica à medida que o número de tipos causais aumenta). Além disso, ao não incorporar a ideia de decibéis para ponderar o peso das evidências, perde uma heurística importante do método advogado por Fairfield e Charman.

Como são metodologias novas, aperfeiçoamentos e mesmo a mistura de ambas podem vir a ser desenvolvimentos possíveis. Por hora, importa salientar que ambas já estão começando a ser aplicadas, embora requeiram treinamento longo e extenso, como os próprios autores reconhecem em seus trabalhos.

Transportabilidade

Como vimos, a inferência causal depende primeiramente de condições para a identificação causal. Superada essa questão, a inferência estatística pode ser feita por meio de métodos Bayesianos para estudos com n pequeno. Contudo, isso garante apenas a validade interna do estudo. Ou seja, nada podemos dizer, a princípio, sobre a generalização para outros contextos ou populações, no que é conhecido como validade externa.

Nesse sentido, uma suposta incapacidade de generalização devido ao número reduzido de casos analisados é sem fundamento. Estudos quanti e quali em geral sofrem dos mesmos problemas de generalização, refletidos na ausência de garantias de que há validade externa.

A distinção entre validade interna e externa foi introduzida pela primeira vez por Campbell (1957), no contexto de estudos experimentais. Validade interna, na definição original, dizia respeito à capacidade de um experimento detectar o efeito causal de interesse na amostra em questão. Hoje em dia, dizemos que um estudo (não apenas experimental) possui validade interna se seu desenho, condução (incluindo a coleta de dados) e análise permite responder às perguntas de pesquisa adequadamente para a amostra estudada (McDermott (2011)). A revolução da credibilidade formalizou as condições para o estabelecimento da validade interna de um estudo. Contudo, a ideia de validade externa continuou vaga por muito mais tempo, e apenas recentemente o conceito tem avançado em direção a uma formalização das condições nas quais a validade externa pode ser crivelmente garantida, na forma do conceito de transportabilidade.

Conforme aponta a literatura sobre transportabilidade (Pearl and Bareinboim 2011, Pearl_Bareinboim_2022), só podemos discutir validade externa se consideramos tanto a população estudada quanto a população alvo para a qual queremos generalizar os resultados do estudo.

Considere o seguinte exemplo. Suponha que uma pesquisadora quer entender como a liberação de emendas impacta a taxa de votação com o governo de deputados que tiveram emendas liberadas. A população alvo pode ser votações futuras sob o mesmo governo ou governos futuros. Esta é a população na qual a pesquisadora deseja aplicar os resultados. De fato,

na maioria dos estudos empíricos em ciências sociais, sejam qualitativos ou quantitativos, raramente explicitamos qual a população alvo da generalização. Implicitamente se considera que não há diferença entre a população efetivamente estudada e a população alvo.

A população do estudo geralmente não é observável ou bem-definida. No exemplo das relações executivo-legislativo, não é clara a população de votações ou de liberações de emendas que o estudo representa².

Às vezes a discussão sobre as diferenças entre população alvo e população estudada aparece na forma de viés de seleção amostral. No contexto da revolução da credibilidade, isso aparece quando o estimando é local, como o Local Average Treatment Effect. Nos estudos de caso, nos próprios critérios de seleção dos casos, se pela variável dependente ou não.

Nesse contexto, a questão de viés de seleção de caso não se aplica a estudos de rastreamento de processo, já que o objetivo é identificar o mecanismo causal daquele caso em específico. Se se pretende generalizar para outros casos, é um problema de validade externa.

Já para estudos de casos que pretendem testar hipóteses causais comparando casos, o viés de seleção pode ser um problema de validade interna. Abordagens como a de queries causais apresentadas no trabalho permitem estabelecer se é possível identificar o efeito causal ou não, não diferindo, portanto, de estudos quantitativos observacionais.

Assim, todo estudo interessado em generalização (no sentido de validade externa) idealmente seguiria este fluxo: definir a população alvo -> coletar uma amostra aleatória dela -> alocar aleatoriamente tratamento e controle (experimento aleatório), coincidindo populações de estudo e alvo, garantindo validade interna e externa do efeito causal (Lesko et al. (2017)).

Na prática, estudos em ciências sociais raramente seguem esse fluxo, seja por impossibilidade prática ou ética. O comum é coletar uma amostra (por exemplo, com dados administrativos no caso de métodos quantitativos, ou elites políticas a serem entrevistadas em uma análise qualitativa), criando uma população do estudo hipotética e outra população alvo diferente.

Os trabalhos de Bareinboim e Pearl Pearl and Bareinboim (2011) utilizam DAGs para demonstrar, matematicamente, as condições nas quais a transportabilidade pode ser feita.

²a literatura às vezes distingue entre população estudada e população analisada, mas aqui essa distinção não é relevante. Cf. Degtiar and Rose (2023)

Suponha por exemplo que um estudo comparativo de dois casos A e B é realizado de acordo com o método da diferença de Mill (Gisselquist 2014), com variáveis binárias (para facilitar) em que x é o tratamento, isto é, a variável que varia por caso, w_1, w_2, w_3 as variáveis que não mudam entre os casos, e y variável resposta. Esse estudo possui validade interna para os dois casos e, presumivelmente, para um terceiro caso C que compartilhe da mesma estrutura.

Digamos que um outro caso D apareça, em que w_2 possui valor distinto de A , B e C . A generalização não é permitida. Contudo, os trabalhos sobre transportabilidade Pearl and Bareinboim (2011) mostram que podemos especificar um DAG e estimar o efeito causal ajustado. Para isso, basta usar a abordagem de causal queries apresentada no trabalho³.

É preciso portanto especificar quais variáveis induzem viés na estimativa para outras populações, e ajustar para variáveis que viesam as estimativas. Mais uma vez, as condições de identificação do efeito causal de interesse em outra população é separado do problema da estimação estatística.

Como essa literatura ainda é muito recente, há pouco desenvolvimento nesse sentido na literatura quantitativa e menos ainda na qualitativa. Para os nossos propósitos, importa salientar apenas que a abordagem de DAGs são a base desse tipo de análise.

Considerações Finais

A pós o livro de KKV, consolidou-se a percepção (falsa) de que métodos qualitativos seriam mais limitados em sua capacidade de generalização, embora superiores em sua validade interna. Essa percepção, baseada na adoção da perspectiva então vigente de modelos de regressão quantitativos, dependia da ideia errônea de que inferência causal e estatística eram a mesma coisa.

A Revolução da Credibilidade, contudo, trouxe uma nova perspectiva, segundo a qual a identificação causal (garantir a validade interna de um estudo) é completamente ortogonal à inferência estatística. Como resultado, mostramos que não há nenhuma característica intrínseca à pesquisa quali ou quanti que justifique diferenças quanto à validade interna e

³cf. o capítulo 12 do livro de Humphreys and Jacobs (2023), que trata justamente de transportabilidade

externa das respectivas pesquisas.

Desfeita essa confusão, mostramos que o advento do paradigma Bayesiano permite também superar a eventual limitação de inferência estatística de estudos com n pequeno, típicas da pesquisa qualitativa causal, e, novamente, que não há relação necessária entre o tipo de pesquisa e a capacidade de estimar quantidades causais de interesse. A variável chave é o poder do estudo, que depende não apenas do tamanho amostral, mas também da relação sinal-ruído das evidências.

Além disso, demonstramos que a inferência Bayesiana oferece soluções robustas para a inferência em amostras pequenas, equiparando o rigor das pesquisas qualitativas ao das quantitativas. As ilustrações metodológicas, como o *process tracing* Bayesiano e as *queries* causais, exemplificam como essas novas abordagens podem ser aplicadas na prática. Esses métodos permitem uma análise mais detalhada e precisa, reforçando a validade dos estudos qualitativos.

Por fim, abordamos como a literatura baseada na ideia de transportabilidade pode permitir que a pesquisa qualitativa supere suas limitações quanto à validade externa, de maneira similar à pesquisa quantitativa. A distinção entre identificação causal e inferência estatística foi fundamental para dismantelar a crença na superioridade da metodologia quantitativa. As críticas frequentemente direcionadas aos métodos qualitativos, baseadas em sua suposta incapacidade de generalização, foram contestadas com sucesso.

Assim, o trabalho contribui para tornar mais claras as dificuldades e caminhos da pesquisa qualitativa baseada em casos, desfazendo qualquer pretensão de superioridade intrínseca da pesquisa quantitativa causal. Esperamos que este estudo inspire uma reavaliação das abordagens metodológicas nas ciências sociais, promovendo um debate mais equilibrado e integrador. A incorporação séria dessas novas metodologias pode fortalecer significativamente a pesquisa qualitativa, desmistificando preconceitos e encorajando a adoção de abordagens mistas. Em última análise, visamos redirecionar os esforços das pesquisadoras para questões metodológicas realmente relevantes, contribuindo para o avanço da ciência social aplicada.

References

- Albert, Isabelle, Sophie Donnet, Chantal Guihenneuc-Jouyaux, Samantha Low-Choy, Kerrie Mengersen, and Judith Rousseau. 2012. "Combining Expert Opinions in Prior Elicitation."
- Amorim, Octavio and Júlio César Cossio Rodriguez. 2016. "O Novo método Histórico-Comparativo e Seus Aportes à Ciência Política e à Administração pública." *Revista de Administração Pública* 50(6):1003–27.
- Angrist, Joshua D. and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106(4):979–1014.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24(2):3–30.
- Bachini, Natasha and Tathiana Senne Chicarino. 2018. "Os métodos Quantitativos, Por Cientistas Sociais Brasileiros: Entrevistas Com Nelson Do Valle Silva e Jerônimo Muniz." *Sociedade e Estado* 33(1):251–79.
- Bareinboim, Elias and Judea Pearl. 2016. "Causal Inference and the Data-Fusion Problem." *Proceedings of the National Academy of Sciences* 113(27):7345–52.
- Bennett, Andrew and Jeffrey T. Checkel. 2015. *Process Tracing*. Cambridge University Press.
- Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54(4).
- Card, David. 2022. "Design-Based Research in Empirical Microeconomics." *American Economic Review* 112(6):1773–81.
- Card, David and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84(4):772–93.
- Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44(4):823–30.
- Degtiar, Irina and Sherri Rose. 2023. "A Review of Generalizability and Transportability." *Annual Review of Statistics and Its Application* 10(1):501–24.
- Fairfield, Tasha and Andrew Charman. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17(1):154–67.
- Fairfield, Tasha and Andrew Charman. 2022. *Social Inquiry and Bayesian Inference*.
- Fairfield, Tasha and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3):363–80.
- Findley, Michael G., Kyosuke Kikuta, and Michael Denly. 2021. "External Validity." *Annual Review of Political Science* 24(1):365–93.
- Forozish, Ali Omar. 2024. "How the Credibility Revolution Created a Paradigm Shift." *Available at SSRN 4744474*.
- Gelman, Andrew. 2009. "Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics." *Statistical Science* 24(2):176–78.
- Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. "The Prior Can Often Only Be Understood in the Context of the Likelihood." *Entropy* 19(10):555.
- Gisselquist, Rachel M. 2014. "Paired Comparison and Theory Development: Considerations for Case Selection." *PS: Political Science & Politics* 47(2):477–84.

- Glied, Sherry. 2021. "The Credibility Revolution in Economics and How It Has Changed Health Policy." Pp. e214335–35 in *JAMA Health Forum*. Vol. 2. American Medical Association.
- Goldsmith-Pinkham, Paul. 2024. "Tracking the Credibility Revolution Across Fields."
- Grimmer, Justin. 2015. "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together." *PS: Political Science & Politics* 48(1):80–83.
- Humphreys, Macartan and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109(4):653–73.
- Humphreys, Macartan and Alan M. Jacobs. 2023. "Integrated Inferences." *Null*.
- Kass, Robert E. and Larry Wasserman. 1996. "The Selection of Prior Distributions by Formal Rules." *Journal of the American Statistical Association* 91(435):1343–70.
- Keele, Luke. 2015a. "The Discipline of Identification." *PS: Political Science & Politics* 48(1):102–6.
- Keele, Luke. 2015b. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23(3):313–35.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *The American Economic Review* 73(1):31–43.
- Lesko, Catherine R., Ashley L. Buchanan, Daniel Westreich, Jessie K. Edwards, Michael G. Hudgens, and Stephen R. Cole. 2017. "Generalizing Study Results: A Potential Outcomes Perspective." *Epidemiology* 28(4):553–61.
- Libman, Alexander. 2023. "Credibility Revolution and the Future of Russian Studies." *Post-Soviet Affairs* 39(1-2):60–69.
- Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86(3):532–65.
- McDermott, Rose. 2011. "Internal and External Validity." *Cambridge Handbook of Experimental Political Science* 27.
- O'Hagan, Anthony. 2010. *Kendall's Advanced Theory of Statistics 2B*. John Wiley & Sons.
- Pearl, Judea and Elias Bareinboim. 2011. "Transportability of Causal and Statistical Relations: A Formal Approach." Pp. 247–54 in *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 25.
- Pearl, Judea and Elias Bareinboim. 2022. "External Validity: From *Do*-Calculus to Transportability Across Populations." Pp. 451–82 in *Probabilistic and Causal Inference*, edited by H. Geffner, R. Dechter, and J. Y. Halpern. New York, NY, USA: ACM.
- Ragin, Charles C. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Univ of California Press.
- Rihoux, Benoît and Charles C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Vol. 51. Sage.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688.
- Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *The Journal of Politics* 78(3):941–55.
- van de Schoot, Rens, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet

G. Tadesse, Marina Vannucci, Andrew Gelman, Duco Veen, and Joukje Willemsen. 2021.
“Bayesian Statistics and Modelling.” *Nature Reviews Methods Primers* 1(1):1.