

Artigo Original

Deu *match*? Uma introdução às técnicas de pareamento

 Bruno Marques Schaefer¹

¹Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brasil.

E-mail: brunomschaefer@gmail.com.

 Dalson Britto Figueiredo Filho²

²Departamento de Ciência Política da Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.

E-mail: dalsonbritto@yahoo.com.br.

DOI: 10.1590/3811039/2023

Resumo

Como inferir causalidade a partir de dados observacionais? Este artigo apresenta uma introdução intuitiva ao pareamento, técnica estatística útil para identificar relações causais em desenhos de pesquisa não experimentais. Metodologicamente, apresentamos as principais características do matching a partir de três exemplos: a) o efeito das escolas militares sobre aprendizagem; b) o impacto do Bolsa Família sobre a propensão em votar no Partido dos Trabalhadores; e c) a influência do gênero sobre o desempenho eleitoral. Mostramos a implementação computacional no R e explicamos a interpretação substantiva dos resultados. Com o objetivo de aumentar o potencial didático da pesquisa, disponibilizamos todos os materiais de replicação, o que facilita que estudantes e profissionais utilizem os dados e scripts em suas atividades de estudo e trabalho. Com este artigo, esperamos difundir o uso de técnicas quase-experimentais nas Ciências Sociais e incentivar a replicabilidade como estratégia de ensino de análise de dados.

Palavras-chave: *matching*; pareamento; inferência causal; dados observacionais; técnicas quase-experimentais.

Did you match? An introduction to matching techniques

Abstract

How can we infer causality from observational data? This article provides an intuitive introduction to matching, a useful statistical technique for identifying causal relationships in non-experimental research designs. Methodologically, we present the main features of matching through three examples: a) the effect of military schools on learning; b) the impact of Bolsa Família on the propensity to vote for the Workers' Party; and c) the effect of gender in electoral results. We show the computational implementation in R and explain the substantive interpretation of the results. In order to increase the pedagogical potential of the research, we provide all replication materials, which makes it easier for students and professionals to use the data and scripts in their study and work activities. We hope to foster the use of quasi-experimental techniques in the Social Sciences with this article and encourage replicability as a teaching strategy in data analysis.

Keywords: matching; causal inference; observational data; quasi-experimental techniques.

Recebido em: 04/04/2023 | Aprovado em: 17/10/2023



Este é um artigo publicado em acesso aberto (*Open Access*) sob a licença *Creative Commons Attribution*, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

1. Introdução

O uso de experimentos na pesquisa científica se justifica por uma razão simples: se implementado corretamente, o desenho experimental garante o estabelecimento de grupos de tratamento e controle que não diferem nem em características observáveis, nem em aspectos não-observáveis (Cook et al., 2008; Gertler et al., 2018). Assim, por se tratar de uma comparação justa, a diferença média entre o grupo que sofreu a intervenção e aquele que recebeu placebo é interpretada como o efeito do tratamento. Este método é considerado o padrão ouro para avaliar desde a efetividade de uma vacina (World Health Organization, 2020) até o impacto de uma política governamental (Gertler et al., 2018).

Todavia, existem vários fenômenos sociais e políticos que não podem ser investigados por experimentos aleatórios. Em alguns casos, a aplicação do tratamento é tecnicamente impossível. Em outros, mesmo que factível, esbarra em questões éticas. Por exemplo, um pesquisador preocupado com a relação entre eleição de mulheres sobre o voto futuro em candidaturas femininas, ou uma pesquisadora interessada em analisar a relação entre abstenção eleitoral e voto em partidos de esquerda, não podem manipular as variáveis de interesse. Isto é, não podem distribuir aleatoriamente gênero entre candidatos eleitos, ou dificultar o voto de determinadas pessoas. Então, como é possível inferir causalidade em situações em que não é viável aleatorizar o tratamento ou manipular a intensidade da intervenção sobre a variável de interesse?

Batista e Domingos (2017) discutem os fundamentos de três técnicas estatísticas comumente empregadas para identificar relações causais em desenhos de pesquisa não experimentais no contexto de avaliação de políticas públicas: pareamento, diferença-em-diferença e regressão descontínua. Embora técnicas quase-experimentais ajudem a superar vários problemas tipicamente observados em dados observacionais (Sobel, 1995; Breen, 2022), o efetivo uso destes procedimentos ainda é restrito na Ciência Social brasileira (Soares, 2005; Neiva, 2015; Figueiredo et al., 2021).

Este artigo apresenta uma introdução intuitiva ao pareamento, técnica útil para identificar relações causais em desenhos de pesquisa não experimentais, com o objetivo de preencher esta lacuna. O pareamento envolve procedimentos estatísticos que permitem a criação do melhor grupo de comparação possível, levando em conta as características observáveis (Gertler et al., 2018). Para tanto, realizamos uma revisão da literatura sobre pareamento e abordamos três exemplos aplicados. Como nosso público-alvo é formado por estudantes em fases iniciais de treinamento e instrutores de cursos de análise de dados, disponibilizamos os materiais de replicação em um repositório aberto no *Open Science Framework*.¹ Assim, qualquer pessoa pode facilmente reutilizar os dados e *scripts* em suas atividades de estudo e trabalho.

Ao concluir a leitura deste artigo, a leitora deve ser capaz de identificar quando o pareamento é a ferramenta adequada para inferir causalidade a partir de dados observacionais, implementar computacionalmente no R e interpretar os principais resultados. Importante: este trabalho não substitui o consumo de obras mais técnicas, tais como: Sekhon (2009), Stuart (2010), King e Nielsen (2019), Imai et al. (2021), entre outras. No entanto, a partir de um produto especialmente desenhado com fins pedagógicos, esperamos facilitar a compreensão do *matching* e difundir o uso de técnicas quase-experimentais nas Ciências Sociais.

O restante do documento está organizado da seguinte forma: a próxima seção explica os fundamentos do pareamento, isto é, o que é e para que serve. A terceira parte apresenta dois exemplos práticos que ilustram a técnica. O primeiro deles é acerca do efeito das escolas militares na aprendizagem (Benevides e Soares, 2020) e o segundo é sobre o impacto do Bolsa Família sobre a propensão de votar no Partido dos Trabalhadores (Zucco e Power, 2013).

¹ Materiais de replicação, incluindo os dados e os *scripts* computacionais estão disponíveis em https://osf.io/9yjbe/?view_only=40ee38aba4a24b5cbe900d6289e46994, consultado em 17 de outubro de 2023.

A quarta seção descreve o passo a passo da implementação computacional a partir do R na investigação do efeito de gênero sobre o desempenho eleitoral. Por fim, resumimos as conclusões e indicamos materiais didáticos para aprofundar o conhecimento sobre o tema.

2. Matching: o que é e para que serve?

Cochran (1953), no artigo "*Matching in Analytical Studies*", formalizou o método de pareamento como forma de realizar inferências causais a partir de dados observacionais (Rubin, 2006). Fundamentalmente, o pareamento consiste em comparar observações semelhantes em relação às características observáveis. Assim como no desenho experimental, em que a comparação entre o grupo de tratamento e o grupo de controle informa o efeito da intervenção, a diferença entre as observações submetidas ao estímulo da variável independente e os casos pareados indica o efeito causal de interesse.

Em termos técnicos, o pareamento ajuda a superar os problemas causados por variáveis de confusão (*confounding variables*), como apontado por Gelman et al. (2020). A Figura 1 ilustra um contexto hipotético de pesquisa em que a variável dependente (y) é causada por três diferentes variáveis independentes (x_1 , z_1 e z_2).

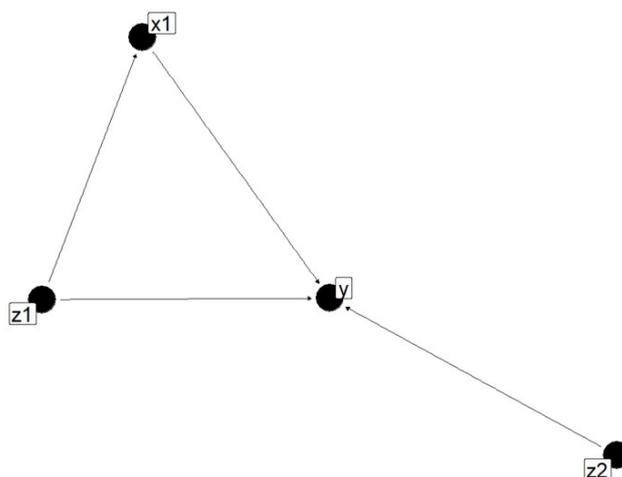


Figura 1 - Grafo direto e acíclico hipotético |
Fonte: elaboração própria, 2023.

Talvez um exemplo nos ajude a ter uma maior compreensão desta complexa estrutura de causalidade. Gelman et al. (2020) destacam um estudo hipotético que busca estimar o efeito de um tratamento sobre a melhoria da saúde de indivíduos (y). O primeiro passo é a seleção de uma amostra da população geral; o segundo é a divisão da amostra em, ao menos, dois grupos. Aquele que recebe o tratamento (T) e aquele que recebe o placebo (P). Após a intervenção e a passagem do tempo, a diferença de médias entre T e P pode ser empregada para analisar se o tratamento foi efetivo sobre a melhoria da saúde (y). O problema é que uma simples diferença de médias pode não ser um comparativo ideal, dado que outra variável importante pode estar omitida: a condição de saúde do indivíduo anteriormente ao estudo (Z). Se, por algum problema de seleção (Gelman et al., 2020), os pacientes do grupo T são mais saudáveis (Z) do que os pacientes do grupo P, o resultado da diferença entre os dois grupos pode não ser resultado causal do tratamento, mas sim da condição anterior (z). Este resultado viesado pode representar tanto uma *confounding variable* - em que a condição de saúde afeta a pessoa receber ou não o tratamento -, quanto uma variável de controle.

Da mesma forma, a avaliação do impacto de uma política pública, digamos, de recolocação profissional, beneficia um grupo de pessoas desempregadas (T) e não outro (P), que ao não considerar outras covariáveis como escolaridade (z_1), renda anterior (z_2), gênero (z_3), entre outras, pode superestimar ou subestimar os resultados finais do programa. Como apontado por Cochran (1953), técnicas de *matching* podem auxiliar nas desvantagens deste tipo de estudo. Para o autor, existem várias formas de encontrar correspondência entre T e P. A primeira e ideal seria o *pairing* (emparelhamento), ou seja, colocar lado a lado unidades de análise (indivíduos, cidades ou outra unidade) semelhantes. Dois homens, ambos com ensino superior, renda anterior de um salário mínimo, idades semelhantes, entre outras variáveis, que se diferenciam pelo fato de receberem (T) ou não (P) a política pública de recolocação profissional.

Seria, portanto, recomendável sempre encontrar um par para cada membro do grupo experimental (T), mesmo que o emparelhamento não fosse exato. Se muitas células ficaram vazias, sem seu respectivo par, há um problema que pode significar que o P não é um controle ideal. Porém, se esta etapa fosse bem-sucedida, passaria a um segundo estágio: a estimação da diferença de y - o que queremos explicar - entre os dois grupos. Cochran (1953) destaca que se poderia realizar uma regressão linear em cada um dos grupos e calcular a diferença entre os coeficientes a fim de determinar o efeito médio do tratamento. Outros métodos abordados por Cochran seriam o balanceamento: estratégia mais ampla de busca por vizinhos/pares da população T e na população P (1); e o que ele chama de pareamento estratificado. Neste caso, a covariável, ou as covariáveis, seriam divididas em estratos e, com isso, cada estrato de T seria comparado ao mesmo estrato de P (2).

Estes dois métodos seriam ideais para a pesquisadora que está trabalhando com mais casos, quando o emparelhamento exato se torna impossível ou toma muito tempo. Qualquer que fosse a escolha, no entanto, os métodos de *matching* não resolveriam problemas de substância em relação à pesquisa, afinal de contas não são uma solução mágica a problemas como o desenvolvimento da questão de pesquisa, desenho da coleta e análise de dados, entre outros. Ou seja, o estabelecimento de uma população de controle (P) comparável a população de tratamento (T) seria um passo mais importante do que a escolha do método de *matching* propriamente dito.

Outra vantagem do pareamento é a redução da dependência do modelo (King, 2018). Conforme a natureza do banco de dados trabalhado, a distribuição das variáveis de controle e das variáveis dependente e independentes, é possível que a especificação do modelo afete os resultados. Por exemplo, a escolha por um modelo linear ou quadrático pode apontar para direções diferentes, a depender da distribuição das variáveis. Para King (2018, p. 5), a vantagem de pré-processar o banco de dados, de modo a “comparar os comparáveis”, é “podar” certas unidades de análise.

O avanço de técnicas computacionais, o tamanho dos bancos de dados trabalhados e o número de covariáveis consideradas nos modelos demandaram maior sofisticação das técnicas de pareamento. Uma questão importante, principalmente quando não é possível o emparelhamento exato de unidades tratadas e de controle em todas as covariáveis, é como definir uma “boa distância” entre uma unidade e outra (Stuart, 2010). Um avanço neste sentido foi realizado por Rosenbaum e Rubin (1983), ao desenvolverem o método *Propensity Score Matching* (PSM). O PSM realiza o pareamento em duas etapas: a primeira é através do cálculo da probabilidade (através de modelos *logit*) de que cada unidade seja do grupo de tratamento e, posteriormente, com esse valor, compara unidades com probabilidades (propensões) semelhantes². Unidades na região comum de probabilidade poderiam ser comparadas (Figura 2).

² Para leitores interessados em aprofundar no assunto, sugerimos o artigo “Why Propensity Scores Should Not Be Used for Matching” de King e Nielsen (2019).

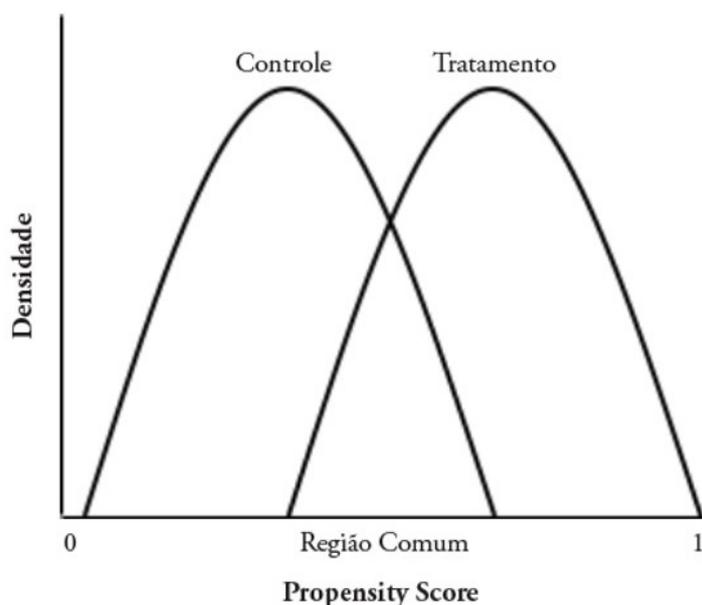


Figura 2 - PSM e a região comum |
Fonte: extraído de Batista e Domingos (2017).

O PSM, neste caso, reduz o total de informação a uma só dimensão: a probabilidade de que determinada unidade receba ou não o tratamento. Este método tem sido o mais utilizado para a realização do *matching* e pré-processamento dos dados (King, 2018). Para Iacus et al. (2019), King (2018) e King e Nielsen (2019), porém, o PSM padece de problemas que inviabilizam seu uso em condições específicas. Para destacá-los, é necessário, novamente, realizar uma diferenciação conceitual. Como apontado por King (2018), podemos pensar em dois tipos “ideais” de experimentos: (1) aqueles completamente randomizados, em que receber ou não o tratamento representaria a chance de cara ou coroa ao jogar uma moeda; e (2) os experimentos totalmente bloqueados (*fully blocked*), em que a probabilidade de receber ou não o tratamento está em conjunto com o pareamento das unidades de análise. Neste último caso, por exemplo, comparamos duas unidades de análise que se assemelham em relação às covariáveis (dois homens com ensino médio completo e com R\$ 1.500,00 de renda) e, então, jogamos a moeda para definir qual recebe o tratamento e qual recebe o placebo. Para Imai et al. (2009), experimentos completamente bloqueados seriam superiores aos completamente randomizados dado o maior balanceamento entre o grupo de controle e tratamento.

Feita esta distinção, King (2018) argumenta que o PSM reproduz (ou tenta reproduzir) experimentos completamente randomizados, o que se torna um problema. Em um cenário ideal, após a realização do primeiro passo do PSM, a aplicação do modelo *logit*, todas as unidades teriam probabilidades semelhantes de receber o Tratamento ou Controle (0,5 ou 50%). Neste caso, a exclusão das unidades seria aleatória, o que incrementa o desbalanceamento. É o que King et al. (2011) chamam de paradoxo de PSM.

Quando as probabilidades são distintas para diferentes unidades de análise não há paradoxo, mas também não há bons pares para comparação dos grupos de controle e tratamento. Iacus et al. (2011), King et al. (2011) e King (2018) encorajam o uso de outros dois métodos de *matching*: MDM (*Mahalanobis Distance Matching*) e o CEM (*Coarsened Exact Matching*). A ideia geral do MDM é calcular a distância entre unidades que recebem o tratamento e o controle, a partir da distância *Mahalanobis*, distância padronizada. Unidades de controle não são utilizadas mais de uma vez e são retiradas do banco de dados caso não haja a possibilidade de encontrar um “par”. O próximo passo, conforme King (2018), é estabelecer um *caliper* (calibre), isto é, a distância máxima entre duas unidades que será tolerada (Greifer e Stuart, 2021).

No CEM, por sua vez, se parte da definição de estratos realizados *ex ante* pelo pesquisador (por exemplo, o uso da variável renda através de faixas de rendimentos - um salário mínimo, dois salários mínimos etc., ao invés do seu uso em uma operacionalização contínua - 900 reais, 902 reais e assim por diante). Este procedimento resultaria em menor perda de informação. O *trade-off* neste caso é que pareamentos exatos podem diminuir o tamanho da amostra em um grau que impossibilite testes estatísticos, como ganho no balanceamento (comparar os comparáveis) e perda no tamanho da amostra.

Como destacado por Stuart (2010), os métodos de pareamento podem ser divididos em dois grandes grupos. Métodos mais ambiciosos (*greedy*), como o PSM, e mais exatos, como o CEM e o MDM (entre outros tantos). O ponto é que as escolhas realizadas pela pesquisadora devem ficar claras, bem como a comparação do pré-processamento dos dados a partir de diferentes métodos. Para Greifer e Stuart (2021), os critérios sobre a escolha do método devem ser julgados pelo balanço entre grupo de controle e tratamento (distância média das covariáveis), bem como o tamanho do banco de dados após o pré-processamento dos dados.

Os métodos de *matching*, como colocado anteriormente, não são salvaguardas à desenhos de pesquisa malfeitos ou bancos de dados com células vazias (*missing data*). Neste sentido, são necessárias a definição de um problema de pesquisa e a(s) hipótese(s) (Kellstedt e Whitten, 2015). A passagem dos conceitos (teoria causal) para sua operacionalização em variáveis e o teste de hipótese(s).

Em termos de estágios e procedimentos (Quadro 1), é necessário considerar, primeiramente, a definição do problema de pesquisa (Estágio 1). O que a pesquisadora quer responder/entender do mundo e se essa resposta pode ser alcançada através de métodos de pareamento. Se o objetivo, por exemplo, é entender associações, talvez não seja necessário o uso de métodos de *matching*. O próximo passo é a definição das variáveis de interesse: dependente (Y) e independente (X) (Estágio 2). No exemplo de uma política pública de recolocação profissional, Y pode ser mensurada como ser reempregado ou não após a realização dos cursos (variável dicotômica), ou a diferença de rendimentos pré e pós realização dos cursos (variável contínua); enquanto X é participar da política ($X = 1$) ou não ($X = 0$). Este é um ponto importante. De modo geral, os métodos de *matching* até aqui expostos funcionam com variáveis de tratamento dicotômicas. Apesar de existirem métodos para parear informações através de variável independente contínua (Adorno et al., 2010), estes são ainda polêmicos e demandam maior sofisticação matemática.

O próximo procedimento é a definição das covariáveis (sejam de controle e/ou confundidoras) que serão utilizadas para parear os grupos de controle e tratamento (Estágio 3). Existem variáveis observáveis e não observáveis (“coisas que você não vê, ou não pensa sobre”), conforme King (2018). Em experimentos completamente randomizados, variáveis não observáveis são controladas pelo próprio processo de randomização. Mas, com a aplicação do *matching*, este controle também é possível. Novamente, o mais importante é, a partir da teoria e da bibliografia sobre o tema, considerar quais os possíveis fatores que podem afetar Y e/ou X. No caso da política pública de recolocação profissional, se o objetivo é mensurar a saída do desemprego, variáveis como escolaridade, idade, anos de experiência e raça devem ser incluídas, por exemplo.

QUADRO 1 - estágios e procedimentos na realização do pareamento

Estágio	Procedimento
1	Definir o problema de pesquisa.
2	Definir a variável dependente (Y) e a variável de tratamento, ou independente (X).
3	Pensar “com cuidado” nas variáveis de controle e/ou confundidoras que podem estar afetando o resultado de interesse (Y).
4	Escolha do método de <i>matching</i> mais adequado ao banco de dados e os objetivos da pesquisa.
5	Análise da qualidade do processo de <i>matching</i> no banco de dados, considerando as diferenças entre o banco de dados completo (antes do pareamento) e o banco de dados balanceado (após o pareamento).
6	Caso os resultados do estágio 5 indiquem um bom balanceamento, realize a análise do efeito médio do tratamento (ATE), ou do efeito médio do tratamento sobre os tratados (ATT) sobre a variável dependente (Y).

Fonte: elaboração própria, a partir de Stuart (2010) e Greifer e Stuart (2021).

Os procedimentos posteriores levam em conta aspectos mais técnicos. É necessário escolher o método de *matching* a ser utilizado (Estágio 4), se PSM, MDM, CEM ou outro (*Optimal Matching*, *Genetic Matching* etc.). Esta escolha deve levar em conta os objetivos da pesquisa e o *trade-off* entre o balanceamento do banco de dados e a possível perda de informação relevante. O ideal será testar diferentes métodos, analisar as distâncias médias das variáveis entre os grupos de controle e tratamento e reportar estes passos na pesquisa. Além de considerar as diferenças e os *calipers* entre grupos, é necessário também observar as diferenças entre o banco de dados completo (original) e o balanceado (Estágio 5). Houve algum ganho para a resposta à pergunta de pesquisa? O objetivo, afinal de contas, é “comparar os comparáveis” e a realização de inferências causais.

Por fim, caso os resultados do Estágio 5 indiquem bom balanceamento e distâncias padronizadas aceitáveis entre os grupos (usualmente 0,05 de diferença média padronizada), passa-se à análise do efeito do tratamento sobre Y (Estágio 6). Apesar de não ser procedimento trivial à comparação de coeficientes em modelos distintos (Raudenbush et al., 1997), é desejável comparar coeficientes de regressão do banco completo e do banco balanceado. Em alguns casos, como os descritos na próxima sessão, a diferença entre ambos é de importância substantiva para a pesquisa. Por exemplo, se uma análise inicial (banco completo) de nossa política pública de recolocação profissional indica que as pessoas que participaram dos cursos tiveram 30% a mais de chances de conseguir um emprego do que aquelas que não participaram da política; e uma análise após a aplicação de um método de *matching* (banco balanceado) indica 10%, esta diferença é de interesse dos gestores, principalmente ao avaliarem o custo-benefício de tal medida.

Na próxima seção, mostramos dois exemplos de aplicação de métodos de pareamento na pesquisa empírica em Ciências Sociais.

3. Exemplos de aplicação de métodos de pareamento na pesquisa empírica em Ciências Sociais

3.1. Caso 1: Escolas Militares têm melhor desempenho que escolas civis?

A literatura sobre desempenho escolar é bastante ampla nas Ciências Sociais, mas acaba se dividindo, historicamente, em duas grandes correntes teóricas: a primeira que destaca que a escola (suas características) possui efeito sobre o desempenho dos alunos, partindo de uma perspectiva endógena (Alves e Soares, 2008); e a segunda que destaca fatores extraescolares (especialmente condições socioeconômicas) como determinantes para o sucesso ou insucesso escolar (Bourdieu e Passeron [1970], 2014), a partir de uma perspectiva exógena. Estudos recentes têm dado maior atenção a fatores endógenos, destacando que variáveis como o tipo de seleção de diretores, formação de professores, infraestrutura, disciplina, entre outras, possuem impactos sobre o aprendizado dos estudantes. Esta discussão é importante tanto em termos teóricos quanto no desenho e aplicação de políticas públicas relacionadas ao ensino. Afinal de contas, se for constatado que determinadas características das escolas podem afetar a melhoria do desempenho educacional, recursos públicos podem ser melhores empregados.

Recentemente, no Brasil, houve uma discussão sobre a gestão escolar a partir de moldes militares, inclusive com a adoção do Programa Nacional das Escolas Cívico-Militares (Decreto nº 10.004, de 2019) (Brasil, 2019).³ A militarização de escolas civis seria uma estratégia para a melhoria dos índices educacionais, baseada nos indicadores de que escolas geridas pelas Forças Armadas (Colégios Militares) e Forças Policiais (Colégios das Polícias Militares Estaduais e do Corpo de Bombeiros) teriam resultados superiores às escolas públicas civis.

³ Programa encerrado em 2023.

Entre os fatores elencados para este melhor desempenho estaria a questão da disciplina no ambiente escolar. O que esta diferença de indicadores ignora, em grande medida, são as diferentes formas de seleção e perfis dos alunos nos dois tipos de estabelecimentos. O perfil socioeconômico de estudantes em escolas civis e militares, por exemplo, é bastante diferenciado, visto que nas segundas a média de rendimentos familiares e escolaridade dos pais é maior do que nas primeiras (Alves e Toschi, 2019). Desta forma, é necessário que as comparações de desempenho levem em conta a heterogeneidade dos dados para “comparar os comparáveis”.

Benevides e Soares (2020) realizam esta comparação levando em conta dados das escolas públicas civis e militares do Ceará. Os autores consideram como variável dependente o desempenho dos estudantes em uma prova de matemática realizada pela Secretaria de Educação do Ceará (SEDUC-CE), no 2º, 5º e 9º anos do Ensino Fundamental (EF) e para a 1ª, 2ª e 3ª séries do Ensino Médio (EM). Em específico, a nota do estudante no 9º ano (Y). A variável independente, ou de tratamento, foi o tipo de escola: se militar estadual⁴ (1) ou civil (0). O pareamento, realizado através do CEM, levou em conta a nota do estudante na prova de matemática no 5º ano. Ou seja, se compara o mesmo aluno em dois pontos do tempo (5º ano e 9º ano do EF):

Se os alunos em um determinado ponto do tempo tiveram desempenho similar (5º ano), pode-se esperar que eles tenham capacidades similares, portanto, qualquer diferencial observado entre estes em séries posteriores (9º ano), salvo mudanças demográficas, pode advir do efeito escola (Benevides e Soares, 2020, p. 325).

A amostra inicial, sem pareamento, possuía 71.049 casos (estudantes) (A1), enquanto na segunda, após o pareamento, tinha 963 casos (A2). Os modelos de regressão demonstram uma superestimação da diferença entre escolas militares e civis (Benevides e Soares, 2020). Se em A1 a diferença na proficiência em matemática no 9º ano era de 26,34 pontos; em A2 esta diferença cai para 18,59. Nos dois casos é registrada significância estatística ao nível de 5%. Ou seja, por mais que se registre uma diferença entre os dois tipos de escola, ela é maior quando são comparados estudantes em condições desiguais em relação ao próprio desempenho passado.

O trabalho, no entanto, apresenta limitações ressaltadas pelos próprios autores. Benevides e Soares (2020) destacam, por exemplo, a limitação da base comparativa (são apenas duas escolas militares estaduais no Ceará); a perda de informações no cruzamento dos dados (estudantes em dois pontos do tempo); e a necessidade de não extrapolar os resultados para outras questões, como a militarização das escolas. O estudo mede a diferença entre dois tipos de escolas e não do ajuste administrativo delas ao longo do tempo. Além disso:

Outros reforços (ou não) a essa evidência inicial de efeito escola militar no Brasil poderiam advir de análises com informações de acompanhamento da fila dos alunos candidatos a vagas nessas escolas, de uma base em painel mais longa (notas do fundamental I ou do ensino médio, por exemplo), ou de escolas contrafactuais que também possuam processo de seleção competitivo e/ou rigor disciplinar sem o uso da autoridade militar (Benevides e Soares, 2020, p. 340).

Ou seja, há a necessidade de maior esforço comparativo, bem como o pareamento considerando outras variáveis, como as socioeconômicas e as demográficas. No entanto, a aplicação do CEM permite demonstrar que há uma superestimação do efeito escola. Esta superestimação é levada ao debate público como uma justificativa para a militarização das escolas.

⁴ Apenas duas escolas: Colégio da Polícia Militar do Ceará e Colégio Militar do Corpo de Bombeiros.

3.2. Caso 2: beneficiários do Bolsa Família têm maior probabilidade de votar no PT?

Um dos grandes temas da literatura sobre voto no Brasil foi a tese do realinhamento partidário ocorrido após o pleito presidencial de 2006. Se, até então, Lula (PT) possuía uma base eleitoral nas grandes cidades e entre a classe média, naquela eleição o petista teria visto um deslocamento de seu apoio para cidades menores e eleitores mais pobres (Hunter e Power, 2007; Licio et al, 2009; Simoni Jr, 2022). Uma das causas desta mudança seria a adoção, a partir de 2003, do programa de transferência de renda condicionada, o Programa Bolsa Família (PBF). Muito já foi escrito sobre o tema, então, o foco aqui é observar como o pareamento foi utilizado para medir o possível efeito PBF sobre a probabilidade de eleitores escolherem Lula como candidato.

Zucco e Power (2013), por exemplo, utilizam métodos de *matching* como forma de estimar o voto em Lula nos anos de 2002 e 2006, a partir do LAPOP (*Latin American Public Opinion Project*), de 2007. Os autores contestam os achados de Bohn (2011), para quem o PBF não teria efeito na escolha dos candidatos em 2006. A autora havia pontuado que os eleitores beneficiários do PBF já tinham maior probabilidade de votar em Lula anteriormente (2002). Ou seja, não haveria um realinhamento. Zucco e Power (2013) replicam os testes com a mesma fonte de dados e encontram que, no banco completo (A1), o fato de se declarar beneficiário do PBF aumentaria a probabilidade do voto no PT em 15%. A1 é um *survey* com 878 respondentes. Para dar mais robustez aos resultados, Zucco e Power (2013) realizam o balanceamento do banco de dados através do pareamento dos indivíduos por uma série de características socioeconômicas, demográficas e regionais. Conforme os autores⁵:

Também estimamos exatamente o mesmo modelo após pré-processar a amostra para corrigir o desbalanceamento das covariáveis. Isto é particularmente importante porque a elegibilidade para o programa Bolsa Família depende da renda familiar, que por sua vez se correlaciona com várias outras variáveis incluídas nas nossas estimativas. Os beneficiários do Bolsa Família são diferentes dos não beneficiários em muitos aspectos, e não apenas pelo benefício mensal que recebem. Idealmente, para qualquer análise conduzida no nível individual, a inferência correta de um efeito causal do Bolsa Família no voto de Lula não deveria ser baseada apenas na comparação entre beneficiários e não beneficiários, mas sim em uma comparação entre eleitores brasileiros que são semelhantes em muitos aspectos, mas que diferem apenas no recebimento do Bolsa Família (Zucco e Power, 2013, p. 8).

Para comparar indivíduos similares em tudo menos no fato de receber ou não o PBF, foram aplicadas técnicas de pareamento exato (CEM) e de vizinhos mais próximos (PSM), a partir de variáveis como região do país, faixa de renda, idade, gênero e escolaridade. Após a aplicação restaram 242 casos. A probabilidade de um beneficiário do PBF votar em Lula foi de 17% em comparação com os não-beneficiários. Ou seja, o resultado é bastante próximo ao encontrado no banco completo e coaduna com a literatura.⁶

4. Implementação computacional no R - caso 3: gênero e desempenho eleitoral

Nesta seção, apresentamos o passo a passo da implementação computacional do pareamento para a realização de inferências causais. Para tanto, devemos observar as seguintes etapas: a) definição do problema e delineamento das hipóteses de trabalho; b) identificação das variáveis de interesse (dependente e independentes); c) definição do métodos mais adequado aos propósitos do trabalho; e d) análise de qualidade do resultado do balanceamento da base de dados.

⁵ Tradução livre dos autores.

⁶ Conforme parecer recebido, amostras utilizadas em *surveys* tendem a adotar regras que tornem os grupos comparáveis. No entanto, Zucco e Power (2013) avançam por notar que mesmo com a ponderação é possível que alguns grupos não sejam diretamente comparáveis, como beneficiários e não beneficiários do PBF.

4.1. Definindo o problema de pesquisa e as hipóteses

A literatura sobre representação fornece evidências sistemáticas sobre a grande desigualdade de gênero na política (Tremblay, 2007). Mulheres possuem menos chances de acessar o campo político, mesmo controlando uma série de fatores. Variáveis que auxiliam na explicação desta diferença são apontadas pela literatura como preconceito de gênero, (sub)financiamento eleitoral, regras institucionais, cultura política, entre outras (Dhima, 2022; Spohr et al., 2016; John et al., 2018; Eymeoud e Vertier, 2023).

Neste sentido, alterações institucionais poderiam modificar as balanças de poder e aumentar a representatividade. No caso brasileiro, mudanças nesta direção são recentes. A lei nº 9.100/1995 (Brasil, 1995), que regulamentou o pleito municipal de 1996, indicava a reserva de 20% das vagas nas listas partidárias (em eleições proporcionais) para mulheres. A lei das eleições de 1997 (nº 9.097) (Brasil, 1997) estabeleceu o patamar de 30%, sem a obrigatoriedade do preenchimento de fato. Em 2009, a lei nº 12.034 (Brasil, 2009) estabeleceu a obrigatoriedade do preenchimento de, no mínimo, 30% das vagas nas listas partidárias para cada sexo.

No entanto, somente após decisão do Supremo Tribunal Federal (STF) de 2018, é que ficou assegurada a distribuição mínima de recursos para candidaturas femininas. O tribunal, ao julgar a Ação Direta de Inconstitucionalidade (ADI) 5.617, impetrada pela Procuradoria Geral da República (PGR), considerou que os partidos políticos brasileiros seriam obrigados a distribuir, ao menos, 30% dos recursos públicos recebidos às candidaturas femininas. Os valores oriundos do Fundo Partidário e do Fundo Especial de Financiamento de Campanhas (FEFC) deveriam ter uma parcela destacada às mulheres.⁷ O Fundo Partidário e o FEFC, somados, representaram em torno de R\$ 2,6 bilhões no pleito de 2018, mais de 70% do total que circulou no pleito (Schaefer, 2022). Ou seja, na esteira da proibição de doações empresariais (através de decisão do STF de 2015), os recursos públicos, destacados do orçamento público federal às direções nacionais dos partidos e depois repartidos para candidatos e diretórios subnacionais, passaram a representar a maior parte dos recursos envolvidos na eleição.

Conforme apontado por Barbieri e Ramos (2019), as alterações nas regras de financiamento eleitoral em 2018 diminuíram o *gap* das candidaturas em termos do gênero. Se em 2014 a média de recursos arrecadados por mulheres candidatas à Câmara Federal foi de 133 mil reais, em 2018 este valor foi de cerca de 151 mil, enquanto a média de recursos recebidos pelos homens caiu de 479 mil para 242 mil reais. A inserção do FEFC e as novas regras de distribuição foram positivas para que, em média, as candidatas fossem mais competitivas eleitoralmente. Em 2014, por exemplo, 51 deputadas federais foram eleitas, enquanto em 2018 este número subiu para 78.⁸ Apesar do crescimento, o percentual de representação feminina coloca o Brasil em uma das piores posições no cenário mundial no quesito mulheres na política.⁹

O objetivo aqui é avaliar o desempenho eleitoral de mulheres e homens nos pleitos para a Câmara dos Deputados de 2014 e 2018. A literatura sobre determinantes do voto, foca em variáveis que afetam o desempenho eleitoral, tais como: capital político, força do partido, ideologia, capacidade de financiamento eleitoral, entre outros. Neste sentido, a partir da seleção de casos comparáveis é possível estabelecer um possível efeito de gênero (X) sobre o percentual de votos recebidos (Y).

⁷ No momento da escrita deste texto, há uma discussão no Congresso Nacional para anistiar os partidos que não cumpriram a regra de 30%, bem como modificar este valor para patamares mais baixos.

⁸ Em 2022, houve novo crescimento, pois 91 mulheres foram eleitas para a Câmara dos Deputados. O número, porém, é ainda muito baixo, visto que representa 17,73% do Plenário.

⁹ O país fica em 140º lugar em um ranking com 190 países. Disponível em <https://www.poder360.com.br/brasil/brasil-e-140o-em-ranking-de-representacao-feminina-no-legislativo/>, consultado em 17 de fevereiro de 2023.

4.2. Definir a variável dependente (Y), a variável de tratamento ou independente (X) e tentar antecipar as covariáveis que podem estar afetando o resultado de interesse (Y)

O banco de dados utilizado aqui dá conta das candidaturas ao cargo de Deputada Federal nos pleitos de 2014 e 2018, totalizando 11.184 observações. As informações sobre as candidaturas foram extraídas do sítio eletrônico do Tribunal Superior Eleitoral (TSE), a partir do pacote do R *electionsBR* (Meireles et al., 2016). Por sua vez, os registros de prestação de contas também foram retirados do site do TSE e depois cruzados a partir do sequencial de cada candidata, também pelo R. A variável dependente (Y) é o percentual de votos recebidos por candidatos e candidatas. Ou seja, o total de votos dividido pelos votos válidos no distrito (UF).

A principal variável independente (X) é gênero (Mulher = 1; Homem = 0). É esperado que haja uma relação negativa entre “ser mulher” e “performance eleitoral”. Outras variáveis de controle são incluídas nos modelos de regressão e no pareamento. A literatura aponta como determinantes para o voto: ser político profissional, valor do patrimônio declarado, cor/raça, força do partido no distrito eleitoral (UF), ideologia, eleitorado no distrito, financiamento eleitoral e ano de 2018. É esperado que todas impactem o percentual de votos.

A variável “político profissional” é operacionalizada através da autodeclaração dos candidatos, se informam ser deputados (federais ou estaduais), senadores e governadores. Candidatos e candidatas já inseridas no campo político possuem maior probabilidade de se reeleger (Barbosa e Schaefer, 2019). A variável “patrimônio declarado” é retirada das declarações e totaliza o valor de todos os bens declarados pelo candidato, uma *proxy* do capital econômico. Espera-se que esta tenha sinal positivo em relação aos votos. “Cor/raça” é também retirada das declarações dos candidatos ao TSE e foi mensurada de maneira dicotômica (Branco = 1; Não Branco = 0). Conforme Machado et al. (2020), candidatos brancos possuem vantagens eleitorais no Brasil. Estas três variáveis mensuram características individuais dos concorrentes. Sobre as candidaturas, também incluímos a arrecadação individual. O valor total, considerada a inflação do período, foi dividido pelo eleitorado do distrito. Como fartamente documentado, há uma correlação positiva entre o financiamento e os votos conquistados (Silva e Gonçalves, 2020).

Sobre o contexto partidário e estadual, incluímos outras quatro variáveis: força do partido, ideologia, tamanho do eleitorado e ano. A força partidária é medida através da votação da legenda em eleições anteriores para cargos proporcionais (Schaefer, 2022). Esperamos que quanto mais forte o partido no distrito, mais votos para seus candidatos. Consideramos neste artigo a classificação ideológica proposta em Power e Rodrigues-Silveira (2019), que leva em conta uma escala entre -1 (Esquerda) e 1 (Direita). Esta escolha se justifica por ser uma medida já consolidada e contínua, realizada periodicamente através de questionários entre parlamentares no projeto *Brazilian Legislative Survey - BLS* (Power e Zucco Jr., 2009). Ao mesmo tempo em que esperamos que haja maior proporção de mulheres se candidatando pela esquerda, é razoável supor que quanto mais à direita mais votos, dado o contexto de disputa (em 2014 e 2018), com resultados expressivos para este campo ideológico (Lavareda e Alves, 2022). Esperamos que a relação entre votos e eleitorado no distrito seja negativa, dado que quanto mais eleitores, maior a fragmentação do percentual de votos. Por fim, consideramos o ano de 2018 também como controle. O Quadro 2 sumariza as principais características do desenho de pesquisa.

QUADRO 2 - Desenho da Pesquisa

Dimensão	Operacionalização	Fonte/Relação esperada	Relação esperada	Mensuração
População	Candidatos/as ao cargo de Deputado Federal em 2014 e 2018 (11.184)	TSE		
Técnicas	Estatística Descritiva, CEM, PSM e Regressão dos Mínimos Quadrados Ordinários (MQO)	R		
Variável Dependente	Votos conquistados (%)	TSE		Contínua
Variável Independente	Gênero	TSE	Negativo	Dicotômica: Mulher = 1; Homem = 0
Variáveis de Controle (Covariáveis)	Capital Político	TSE	Positivo	Dicotômica: Político profissional = 1; Sem capital político = 0
	Valor dos Bens	TSE	Positivo	Contínua
	Força do partido no distrito	TSE	Positivo	Contínua
	Ideologia	Power e Rodrigues-Silveira (2019)	Positivo	Contínua
	Eleitorado	TSE	Negativo	Contínua
	Recursos recebidos dividido pelo eleitorado (R\$)	TSE	Positivo	Contínua
	Ano 2018	TSE	Positivo	Dicotômica: 2018 = 1; 2014 = 0.

Fonte: elaboração dos autores, 2023.

4.3. Escolha do método de matching mais adequado ao banco de dados e os objetivos da pesquisa

Construímos um gráfico para analisar as diferenças entre mulheres e homens e os pleitos de 2014 e 2018 (Figura 3).

Candidatos homens obtiveram, em média, 0,42% dos votos na eleição de 2014 e na eleição seguinte 0,27%. Enquanto as candidatas mulheres fizeram, em média, 0,16% dos votos em 2014 e 0,13% em 2018. A diferença entre candidatos e candidatas passa de 0,26% para 0,14%. Ou seja, à primeira vista, há uma redução na diferença entre candidatas e candidatos. Bem como, é possível notar que a média de votos cai entre uma eleição e outra para ambos os tipos de candidaturas, dado o crescimento no número de candidatos (de 4.986 para 6.198).¹⁰

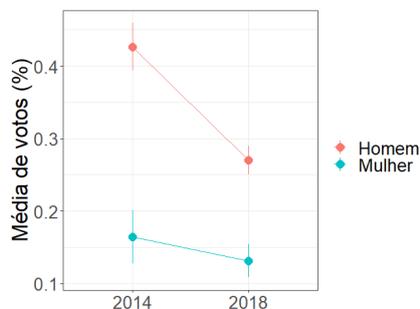


Figura 3 - Desempenho eleitoral de homens e mulheres nas eleições de 2014 e 2018 (%) |

Fonte: elaboração própria, 2023.

¹⁰Estes números consideram somente as candidaturas que tinham informação sobre prestação de contas e desempenho eleitoral.

Tratamos de aplicar métodos de *matching* para balancear o banco de dados e podar as informações (King, 2018). O objetivo aqui é considerar que (ou se) outras variáveis podem afetar o desempenho eleitoral que não o (ou em conjunto com) gênero. Variáveis como capital político, econômico, cor/raça, força do partido no distrito, ideologia e eleitorado no distrito. Aplicamos o CEM (Figura 4) e o PSM (Figura 5).

A inspeção visual das diferenças padronizadas indica, em primeiro lugar, um bom balanceamento dos dados, na medida em que as diferenças das covariáveis entre homens e mulheres são menores após a aplicação do *matching* (pontos escuros) do que no banco completo (pontos claros). Em segundo lugar, é possível perceber que o CEM reduz mais esta diferença do que o PSM. Enquanto no primeiro a diferença média padronizada entre gêneros é de no máximo -0,004 (Bens), no segundo varia até 0,024 (Ideologia).

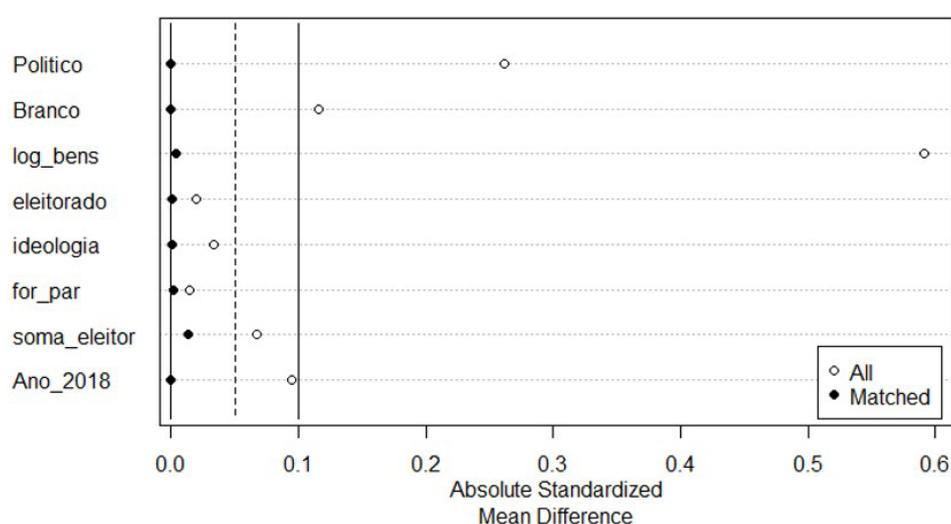


Figura 4 - Balanceamento das covariáveis grupo de tratamento e controle (CEM) |

Fonte: elaboração própria, 2023.

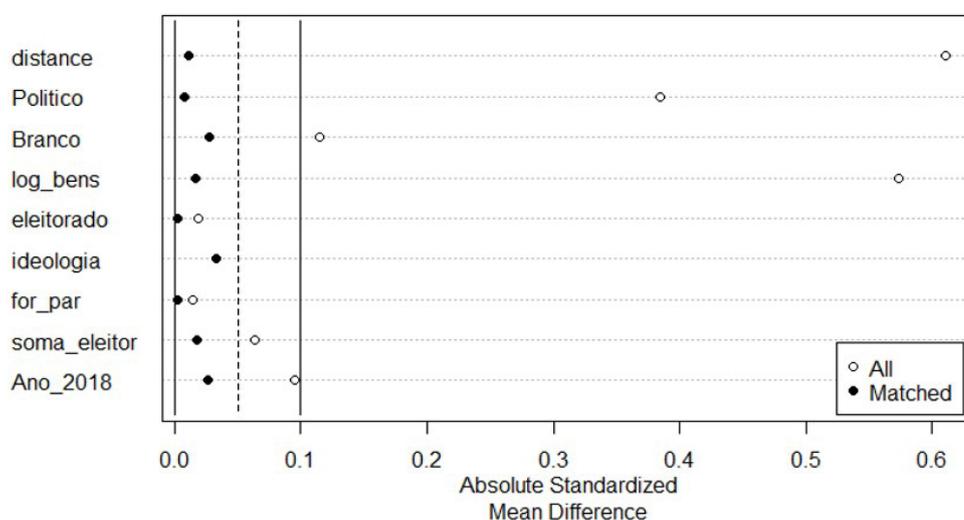


Figura 5 - Balanceamento das covariáveis grupo de tratamento e controle (PSM) |

Fonte: elaboração própria, 2023.

Na Tabela 1, comparamos as médias de distribuição das covariáveis no banco completo e nos bancos balanceados pelo método CEM e PSM (a partir do teste T de diferença de médias). Novamente, é possível perceber melhor ajuste no CEM. Outro ponto importante é que as variáveis individuais apresentam grande diferença entre homens e mulheres. Os homens são mais ricos que as mulheres (em média), têm maior capital político, são mais brancos e arrecadam mais recursos por eleitor no distrito. Se estas quatro variáveis não fossem consideradas, poder-se-ia subestimar ou superestimar as diferenças entre homens e mulheres. Outro ponto importante é que há maior concentração de mulheres candidatas em 2018 quando comparado com 2014. Importante salientar que há diferença de médias estatisticamente significativas nestas cinco variáveis quando consideramos o banco completo, mas estas diferenças se esvaem nos bancos balanceados. Ou seja, na hora de estimar o efeito do gênero sobre a variável dependente, estaremos diante de uma comparação mais justa.

TABELA 1 - **Balanceamento das covariáveis grupo de tratamento e controle**

		Político	Branco	Bens (log)	Soma Recursos	Eleitorado (log)	Ideologia	Força Partidária	Ano (2018)
Banco completo	Mulheres	0,023*	0,567*	6,035*	0,050*	15,933	0,074	0,057	0,586*
	Homens	0,080	0,624	9,543	0,070	15,912	0,090	0,042	0,540
CEM	Mulheres	0,004	0,611	6,783	0,010	16,107	0,095	-0,1373	0,565
	Homens	0,004	0,611	6,807	0,014	16,109	0,095	-0,136	0,565
	Diferença média padronizada	-0,000	0,000	-0,004	-0,001	-0,001	0,001	-0,001	0,000
PSM	Mulheres	0,023	0,567	6,035	0,050	15,912	0,075	0,074	0,589
	Homens	0,025	0,560	6,138	0,045	15,888	0,063	0,077	0,600
	Diferença média padronizada	0,008	0,014	-0,016	0,004	0,021	0,024	-0,007	-0,040

*Diferença estatisticamente significativa ao nível de 5%. |

Fonte: elaboração própria, 2023.

4.4. Para analisar a qualidade do processo de *matching* no banco de dados, deve-se considerar as diferenças entre o banco de dados completo (antes do pareamento) e balanceado (após o pareamento), e, caso os resultados do estágio anterior indiquem um bom balanceamento, realize a análise do efeito médio do tratamento sobre a variável dependente (Y).

A partir da aplicação do pareamento e da observação de seus resultados, utilizamos três bancos de dados distintos (completo, a partir do CEM e do PSM). No banco original havia 11.184 casos, sendo 3.355 do grupo de tratamento (mulheres) e 7.829 do grupo de controle (homens). Após a aplicação do CEM, restaram 6.922 casos: 1.910 mulheres e 3.982 homens; no caso do PSM restaram 6.710 casos, divididos igualmente em grupos de controle e tratamento. Estes são os bancos balanceados.

Na Tabela 2, apresentamos os resultados de uma série de modelos de regressão (Ordinary Least Squares), observando a diferença entre homens e mulheres candidatas no caso do desempenho eleitoral. Os Modelos 1 e 2 são aplicados ao banco completo, enquanto os modelos 3 a 4 são aplicados aos bancos balanceados após o pareamento. Por mais que a comparação entre coeficientes de regressão de modelos distintos não seja uma tarefa trivial, é possível observar nos modelos que a diferença entre os votos percentuais de homens e mulheres é consideravelmente menor quando “comparados

os comparáveis". Se, no Modelo 1, apenas com a variável gênero, a diferença de desempenho eleitoral entre homens e mulheres é de 0,20%; no banco balanceado, após o pareamento, esta diferença é de 0,06% (CEM); enquanto no caso da aplicação do PSM o valor é de -0,08%. A diferença no coeficiente se assemelha ao Modelo 2, quando são incluídos os controles. Estes resultados ficam mais claros na Figura 6.

TABELA 2 - Modelos de Regressão OLS

	Completo		Balanceado	
	Modelo 1	Modelo 2	Modelo 3 (CEM)	Modelo 4 (PSM)
(Intercept)	0,34***	2,15***	0,12***	0,22***
	(0,01)	(0,08)	(0,01)	(0,01)
Mulher	-0,20***	-0,08***	-0,06***	-0,08***
	(0,02)	(0,01)	(0,01)	(0,02)
Politico		0,77***		
		(0,02)		
Branca		0,05***		
		(0,01)		
Bens valor (Log)		0,01***		
		(0,00)		
Eleitorado (Log)		-0,13***		
		(0,01)		
Ideologia		0,05***		
		(0,01)		
Força Partidária		0,05***		
		(0,01)		
Soma Recursos (por eleitor)		1,16***		
		(0,02)		
Ano 2018		-0,08***		
		(0,01)		
Adj. R ²	0,01	0,48	0,01	0,01
Num. obs.	11.184	11.184	6.922	6.710

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. |

Fonte: elaboração própria, 2023.

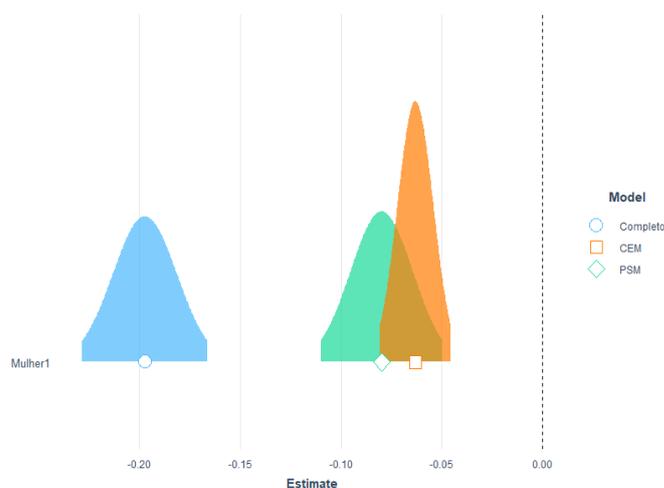


Figura 6 - Diferença entre a arrecadação de homens e mulheres nas eleições de 2014 e 2018 (banco completo x balanceado) |

Fonte: elaboração própria, 2023.

A aplicação do pareamento para a análise do desempenho eleitoral de homens e mulheres possibilita observar que as desigualdades de gênero permanecem, mesmo com o banco de dados somente com informações balanceadas. Apesar de menores, afinal de contas, nem todas as mulheres (enquanto grupo) e homens (enquanto grupo) são iguais (capital político, riqueza e capacidade de arrecadação não são distribuídos aleatoriamente na população), há evidências que apontam para um desempenho eleitoral pior de mulheres (Dhima, 2022; Spohr et al., 2016; John et al., 2018; Eymeoud e Vertier, 2023). Estes resultados devem ser discutidos com maior atenção por pesquisadores e pesquisadoras focados no tema, e aqui auxiliam na observação de como técnicas de pareamento refinam inferências causais nas Ciências Sociais.

5. Conclusão

Este artigo apresentou uma introdução intuitiva ao pareamento, técnica estatística útil para identificar relações de causalidade a partir de dados observacionais. Como a maior parte dos fenômenos de interesse em Ciências Sociais não são passíveis de aleatorização – seja por questões logísticas, seja por restrições éticas –, acreditamos que um trabalho em formato de tutorial pode contribuir com a produção de desenhos de pesquisa mais robustos.

Para ilustrar o potencial do *matching*, reproduzimos dados de Benevides e Soares (2020) e Zucco e Power (2013) para explorar o impacto das escolas militares sobre a aprendizagem e o efeito do Bolsa Família sobre a chance de apoiar eleitoralmente o Partido dos Trabalhadores, respectivamente. Com o objetivo de aumentar o potencial pedagógico da pesquisa, incluímos o passo a passo da implementação computacional no *Re* reportamos a interpretação substantiva dos resultados empíricos. Os materiais de replicação, incluindo dados originais e *scripts*, também foram disponibilizados com o intuito de facilitar o reuso das informações por estudantes e profissionais da área. Os dados apresentados no Caso 3 demonstram que, mesmo após a inclusão de controles e de técnicas diferentes de *matching* (PSM e CEM), ainda é possível encontrar desigualdades no desempenho de homens e mulheres na disputa por representação política. Isto denota a necessidade de pensarmos em quais os mecanismos que mantêm estes resultados distintos (Dhima, 2022; Spohr et al., 2016; John et al., 2018; Eymeoud e Vertier, 2023), que tem por resultado o Brasil em uma das piores posições no ranking de representação de mulheres no Legislativo.

Alertamos aos leitores e leitoras que o conteúdo deste artigo não deve substituir o consumo de materiais mais técnicos, incluindo aulas de desenho de pesquisa, análise de dados e identificação causal. A seguir, listamos algumas fontes de conteúdo potencialmente úteis ao desenvolvimento destas habilidades. Tanto o Coursera quanto o Edx oferecem cursos e especializações nas áreas de programação, estatística e *data science*. Outra opção para fortalecer a formação em métodos de pesquisa pode ser encontrada nos cursos de verão promovidos pela *International Political Science Association* (IPSA). No Brasil, este treinamento ocorre anualmente na Universidade de São Paulo. O tradicional curso de Metodologia Quantitativa (MQ), ofertado pela Universidade Federal de Minas Gerais (UFMG), também representa uma oportunidade para conhecer conteúdos que não são cobertos em disciplinas regulares da pós-graduação e/ou aprofundar temas que apenas são discutidos de forma mais superficial.

No que diz respeito especificamente à identificação de causalidade a partir de dados observacionais, recomendamos a inclusão de um módulo de técnicas quase-experimentais nas ementas dos cursos de análise de dados avançados. Sobre pareamento, em particular, o artigo clássico de Cochran (1953), o livro do Gertler et al. (2018) e o trabalho de Batista e Domingos (2017) representam excelentes opções para utilizar em cursos de

Ciências Sociais. Mais recentemente, os professores Felipe Nunes¹¹ (UFMG) e Gary King¹² (Harvard) disponibilizaram aulas em plataformas abertas e podem ser utilizadas para fixar a aprendizagem. Para trabalhos mais avançados, recomendamos verificar a referida curadoria¹³ e a obra da professora Elizabeth Stuart,¹⁴ que escreveu um dos artigos mais citados sobre o tema: "*Matching methods for causal inference: A review and a look forward*". Por fim, no que diz respeito à implementação computacional, o pareamento pode ser facilmente realizado com auxílio de pacotes estatísticos como o *Statistical Package for Social Sciences* (SPSS) e Stata. Nossa opção pelo R se justifica por ser um *software* aberto, o que facilita o uso e aumenta o potencial de colaboração.

Há cerca de 20 anos, Soares (2005) detectou uma certa hostilidade da Ciência Política nacional em relação aos métodos quantitativos e à Estatística. Tendência ainda mais forte na Sociologia e na Antropologia (Neiva, 2015). Felizmente, esta animosidade vem se dissipando e o entendimento de que a metodologia é parte indissociável do conhecimento científico parece ser ponto pacífico na maior parte dos nossos Departamentos e Programas de Pós-Graduação. Com este artigo, esperamos difundir o uso de técnicas quase-experimentais nas Ciências Sociais e incentivar a replicabilidade como estratégia de ensino no treinamento de análise de dados.

Bibliografia

- ADORNO, Valentina; BERNINI, Cristina; PELLEGRINI, Guido. (2010), "Comparing continuous treatment matching methods in policy evaluation", in F. Palumbo; C. Lauro; M. Greenacre. (eds.), *Data Analysis and Classification*. Heidelberg, Springer. DOI: https://doi.org/10.1007/978-3-642-03739-9_48.
- ALVES, Maria Teresa Gonzaga; SOARES, José Francisco. (2008), "O efeito das escolas no aprendizado dos alunos: um estudo com dados longitudinais no Ensino Fundamental". *Educação e Pesquisa*, 34, 3:527-544. DOI: <https://doi.org/10.1590/S1517-97022008000300008>.
- ALVES, Miriam Fábria; TOSCHI, Mirza Seabra. (2019), "A militarização das escolas públicas: uma análise a partir das pesquisas da área de educação no Brasil". *Revista Brasileira de Política e Administração da Educação*, 35, 3:633-647. DOI: <https://doi.org/10.21573/vol35n32019.96283>.
- BARBIERI, Catarina Helena Cortada; RAMOS, Luciana de Oliveira. (2019), *Democracia e Representação nas Eleições de 2018: campanhas eleitorais, financiamento e diversidade de gênero*. São Paulo, FGV Direito.
- BARBOSA, Tiago Alexandre Leme; SCHAEFER, Bruno Marques. (2019), "Candidatos e eleitos deputados estaduais no Brasil: para onde vai a representação política? (1998-2018)". *Revista E-Legis*, 12, 30:90-114. DOI: <https://doi.org/10.51206/e-legis.v12i30.547>.
- BATISTA, Mariana; DOMINGOS, Amanda. (2017), "Mais que boas intenções: técnicas quantitativas e qualitativas na avaliação de impacto de políticas públicas". *Revista Brasileira de Ciências Sociais*, 32, 94, e329414:1-24. DOI: <https://doi.org/10.17666/329414/2017>.
- BENEVIDES, Alessandra de Araújo; SOARES, Ricardo Brito. (2020), "Diferencial de desempenho de alunos das escolas militares: o caso das escolas públicas do Ceará". *Nova Economia*, 30, 1:317-343. DOI: <https://doi.org/10.1590/0103-6351/3929>.

¹¹ Disponível em <<https://www.youtube.com/watch?v=YplVxj07nmg&t=192s>>, consultado em 24/10/2023.

¹² Disponível em <<https://www.youtube.com/watch?v=tvMyjDi4dyg&t=3039s>>, consultado em 24/10/2023.

¹³ Disponível em <https://gking.harvard.edu/publications/term/1585>, consultado em 24/10/2023.

¹⁴ Disponível em <https://scholar.google.com/citations?user=gNBKSfEAAA&hl=en>, consultado em 24/10/2023.

- BOHN, Simone. (2011), "Social policy and vote in Brazil: Bolsa Família and the shifts in Lula's electoral base". *Latin American Research Review*, 46, 1:54-79. DOI: <https://doi.org/10.1353/lar.2011.0003>.
- BOURDIEU, Pierre; PASSERON, Jean-Claude. [1970] (2014), *Reprodução: Elementos para uma teoria do sistema de ensino*. Petrópolis, Editora Vozes.
- BRASIL. (1995), Lei nº 9.096, de 19 de setembro de 1995. Dispõe sobre partidos políticos, regulamenta os arts. 17 e 14, § 3º, inciso V, da Constituição Federal e dá outras providências. Disponível em http://www.planalto.gov.br/ccivil_03/leis/L9096.htm, consultado em 27/10/2023.
- BRASIL. (1997), Lei nº 9.504, de 30 de setembro de 1997. Estabelece normas para as eleições. Disponível em <http://www.tse.jus.br/legislacao/codigo-eleitoral/lei-daseleicoes/leidas-eleicoes-lei-nb0-9.504-de-30-de-setembro-de-1997>, consultado em 27/10/2023.
- BRASIL. (2009), Lei nº 12.034, de 29 de setembro de 2009. Altera as Leis nos 9.096, de 19 de setembro de 1995 - Lei dos Partidos Políticos, 9.504, de 30 de setembro de 1997, que estabelece normas para as eleições, e 4.737, de 15 de julho de 1965 - Código Eleitoral. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2009/Lei/L12034.htm, consultado em 27/10/2023.
- BRASIL. (2019), Decreto nº 10.004, de 5 de setembro de 2019. Institui o Programa Nacional das Escolas Cívico-Militares. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d10004.htm, consultado em 26/10/2023.
- BREEN, Richard. (2022). "Causal inference with observational data", in K. Gërkhani; N.D. De Graaf; W. Raub. (org.), *Handbook of Sociological Science*. Cheltenham; Northampton, Edward Elgar Publishing.
- COCHRAN, William. (1953), "Matching in analytical studies". *American Journal of Public Health and the Nations Health*, 43, 6:684-691.
- COOK, Thomas; SHADISH, William; WONG, Vivian. (2008), "Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons". *Journal of Policy Analysis and Management: The Journal of Policy Analysis and Management*, 27, 4:724-750. DOI: <https://doi.org/10.1002/pam.20375>.
- DHIMA, Kostanca. (2022), "Do elites discriminate against female political aspirants? Evidence from a field experiment". *Politics & Gender*, 18, 1:126-157. DOI: <https://doi.org/10.1017/S1743923X20000227>.
- EYMEUD, Jean-Benoit; VERTIER, Paul. (2023), "Gender biases: evidence from a natural experiment in French local elections". *Economic Policy*, 38, 113:3-56. DOI: <https://doi.org/10.1093/epolic/eiac067>.
- FIGUEIREDO, Dalson, FERNANDES, Antônio, BORBA, Lucas; AGUIAR, Thaís Helena. (2021), "Metodologias de pesquisa em ciência política: uma breve introdução". *BIB-Revista Brasileira de Informação Bibliográfica em Ciências Sociais*, 94:1-34.
- GELMAN, Andrew; HILL, Jennifer; VEHTARI, Aki. (2020), *Regression and other stories*. Cambridge, Cambridge University Press.
- GERTLER, Paul; MARTÍNEZ, Sebastian; PREMAM, Patrick; RAWLINGS, Laura; VERMEERSCH, Christel. (2018), *Avaliação de Impacto na Prática*. 2ª edição. São Paulo, World Bank Publications.
- GREIFER, Noah; STUART, Elizabeth. (2021), "Matching methods for confounder adjustment: an addition to the epidemiologist's toolbox". *Epidemiologic reviews*, 43, 1:118-129. DOI: <https://doi.org/10.1093/epirev/mxab003>.
- HUNTER, Wendy; POWER, Timothy. (2007), "Rewarding Lula: Executive power, social policy, and the Brazilian elections of 2006". *Latin American Politics and Society*, 49, 1:1-30. DOI: <https://doi.org/10.1111/j.1548-2456.2007.tb00372.x>.

- IACUS, Stefano; KING, Gary; PORRO, Giuseppe. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106, 493:345-361. DOI: <https://doi.org/10.1198/jasa.2011.tm09599>.
- IACUS, Stefano; KING, Gary; PORRO, Giuseppe. (2019), "A theory of statistical inference for matching methods in causal research". *Political Analysis*, 27, 1:46-68.
- IMAI, Kosuke; KING, Gary; NALL, Clayton. (2009), "The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation". *Statistical Science*, 24, 1:29-53. DOI: 10.1214/08-STS274.
- IMAI, Kosuke; KIM, In Song; WANG, Erik. (2021), "Matching methods for causal inference with time-series cross-sectional data". *American Journal of Political Science*. 67, 3:587-605. DOI: <https://doi.org/10.1111/ajps.12685>.
- JOHN, Sarah; SMITH, Haley; ZACK, Elizabeth. (2018), "The alternative vote: do changes in single-member voting systems affect descriptive representation of women and minorities?". *Electoral Studies*, 54, 90-102. DOI: <https://doi.org/10.1016/j.electstud.2018.05.009>.
- KELLSTEDT, Paul; WHITTEN, Guy. (2015), *Fundamentos da pesquisa em ciência política*. São Paulo, Editora Blucher.
- KING, Gary. (2018), "Gary King on Simplifying Matching Methods for Causal Inference". *Taiwan Journal of Political Science*, 77, 1-32. DOI:10.6166/TJPS.201809_(77).0001.
- KING, Gary; NIELSEN, Richard; COBERLEY, Carter; POPE, James; WELLS, Aaron. (2011), Comparative effectiveness of matching methods for causal inference. Disponível em <https://tinyurl.com/yyn5rt>, consultado em 30/10/2023.
- KING, Gary; NIELSEN, Richard. (2019), "Why propensity scores should not be used for matching". *Political Analysis*, 27, 4:435-454. DOI: <https://hdl.handle.net/1721.1/128459>.
- LAVAREDA, Antonio; ALVES, Vinícius. (2022), "Eleições municipais como barômetros ideológicos e a ciclicidade eleitoral da Nova República", in A. Lavareda; H. Telles. (org.), *Eleições Municipais na Pandemia*. Rio de Janeiro, FGV Editora.
- LICIO, Elaine Cristina; RENNÓ, Lucio; CASTRO, Henrique Carlos. (2009), "Bolsa Família e voto na eleição presidencial de 2006: em busca do elo perdido". *Opinião Pública*, 15, 1:31-54. DOI: <https://doi.org/10.1590/S0104-62762009000100002>.
- MACHADO, Carlos Augusto; CAMPOS, Luiz Augusto; RECCH, Filipe. (2019), "Race and Competitiveness in Brazilian Elections: Evaluating the Chances of Black and Brown Candidates through Quantile Regression Analysis of Brazil's 2014 Congressional Elections". *Brazilian Political Science Review*, 13, 3, e0005:1-31. DOI: <https://doi.org/10.1590/1981-3821201900030003>.
- MEIRELES, Fernando; SILVA, Denisson; COSTA, Beatriz. (2016), "ElectionsBR: R functions to download and clean Brazilian electoral data". Cited on, p. 3.
- NEIVA, Pedro. (2015). "Revisitando o calcanhar de Aquiles metodológico das ciências sociais no Brasil". *Sociologia, problemas e práticas*, 79:65-83.
- POWER, Timothy; RODRIGUES-SILVEIRA, Rodrigo. (2019), "Mapping ideological preferences in Brazilian elections, 1994-2018: a municipal-level study". *Brazilian Political Science Review*, 13, e0001.
- POWER, Timothy; ZUCCO JR., Cesar. (2009), "Estimating ideology of Brazilian legislative parties, 1990-2005: a research communication". *Latin American Research Review*, 44, 1:218-246.
- RAUDENBUSH, Stephen; GLEESER, Leon; HEDGES, Larry; ROWLEY, Stella; JOHNSON, Eugene; PETKOVA, Eva. (1997). Comparing regression coefficients between models: Concepts and illustrative examples. Disponível em <https://www.niss.org/research/comparing-regression-coefficients-between-models-concepts-and-illustrative-examples>, consultado em 30/10/2023.

- ROSENBAUM, Paul; RUBIN, Donald. (1983), "The central role of the propensity score in observational studies for causal effects". *Biometrika*, 70, 1:41-55. DOI: <https://doi.org/10.1093/biomet/70.1.41>.
- RUBIN, Donald. (2006), *Matched sampling for causal effects*. Cambridge, Cambridge University Press.
- SCHAEFER, Bruno Marques. (2022), "Autofinanciamento eleitoral no Brasil: regulação, causas e consequências". Tese de doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- SEKHON, Jasjeet. (2009). "Opiates for the matches: Matching methods for causal inference". *Annual Review of Political Science*, 12, 487-508. DOI: <https://doi.org/10.1146/annurev.polisci.11.060606.135444>.
- SILVA, Bruno Fernando da; GONÇALVES, Ricardo Dantas. (2019), "Pesquisas eleitorais afetam receitas de campanha: a correlação entre expectativa de vitória e financiamento de campanha em disputas ao Senado". *Revista de Sociologia e Política*, 27, 71, e005:1-17. DOI: <https://doi.org/10.1590/1678-987319277105>.
- SIMONI JR, Sérgio. (2022), "Electoral dividends from programmatic policies: a theoretical proposal based on the Brazilian case". *Brazilian Political Science Review*, 16, 1, e0006:1:42. DOI: <https://doi.org/10.1590/1981-3821202200010005>.
- SOARES, Gláucio Ary Dillon. (2005), "O calcanhar metodológico da ciência política no Brasil". *Sociologia, problemas e práticas*, 48:27-52.
- SOBEL, Michael. (1995). "Causal inference in the social and behavioral sciences", in G. Arminger; C. Clogg; M. Sobel. (eds.), *Handbook of statistical modeling for the social and behavioral sciences*. Boston, Springer Science & Business Media.
- SPOHR, Alexandre Piffero; MAGLIA, Cristiana; MACHADO, Gabriel; OLIVEIRA, Joana Oliveira de. (2016), "Participação Política de Mulheres na América Latina: o impacto de cotas e de lista fechada". *Revista Estudos Feministas*, 24, 2:417-441. DOI: <https://doi.org/10.1590/1805-9584-2016v24n2p417>.
- STUART, Elizabeth. (2010), "Matching methods for causal inference: a review and a look forward". *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25, 1:1-21. DOI: 10.1214/09-STS313.
- TREMBLAY, Manon. (2007), "Democracy, representation, and women: A comparative analysis". *Democratization*, 14, 4:533-553. DOI: <https://doi.org/10.1080/13510340701398261>.
- WORLD HEALTH ORGANIZATION. (2020), "Design of vaccine efficacy trials to be used during public health emergencies—points of considerations and key principles". Disponível em [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiG76bHp5GCaxXVJLkGHfJwA1gQFnoECBAQAQ&url=https%3A%2F%2Fwww.who.int%2Fdocs%2Fdefault-source%2Fblue-print%2Fworking-group-for-vaccine-evaluation-\(4th-consultation\)%2Fa](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiG76bHp5GCaxXVJLkGHfJwA1gQFnoECBAQAQ&url=https%3A%2F%2Fwww.who.int%2Fdocs%2Fdefault-source%2Fblue-print%2Fworking-group-for-vaccine-evaluation-(4th-consultation)%2Fa), consultado em 15/07/2023.
- ZUCCO, Cesar; POWER, Timothy. (2013), "Bolsa Família and the shift in Lula's Electoral Base, 2002-2006: a reply to Bohn". *Latin American Research Review*, 48, 2:3-24. DOI: <https://doi.org/10.1353/lar.2013.0018>.