

## **Métodos de Matching na Avaliação de Políticas Públicas: Uma Comparação Empírica entre GenMatch e CEM**

Autores: Ciro Biderman e Thales Schiabel Adler

### **Resumo Simples:**

Uma das peças fundamentais do ciclo das políticas públicas é a sua avaliação. Embora o ideal metodológico seja a aleatorização dos beneficiários da política, essa estratégia nem sempre é viável. Nesses casos, uma alternativa é o pareamento (*matching*), que busca comparar unidades tratadas com outras o mais semelhantes possível, visando isolar o efeito da intervenção. Quanto melhor o grupo de comparação, mais precisa será a estimativa de impacto.

Esse trabalho contribui para essa etapa ao comparar duas técnicas consolidadas de pareamento: o Genetic Matching (GenMatch) e o Coarsened Exact Matching (CEM). Apesar da consolidação dessas abordagens, a literatura ainda carece de comparações empíricas entre ambas, especialmente no Brasil.

Aplicamos tais técnicas na política de Arranjos Produtivos Locais por meio de 616 pareamentos independentes, considerando duas dimensões: o município e o setor de atividades. O resultado foi o pareamento satisfatório em ambos os métodos. O GenMatch reduziu o desequilíbrio médio das covariáveis de forma mais eficaz (SMD médio  $< 0,01$ ), enquanto o CEM obteve SMDs médios  $< 0,09$  e sobreposição entre 70% e 80%. No entanto, o custo operacional do GenMatch se mostrou substancialmente maior do que o do CEM, evidenciando um trade-off importante na escolha do método de pareamento ideal.

### **Resumo expandido:**

Uma das peças fundamentais do ciclo das políticas públicas é a sua avaliação. Sem ela, não é possível saber se uma política está funcionando adequadamente e, portanto, se ela deve ser descontinuada, ampliada ou alterada. A “fórmula de ouro” para a avaliação de impacto ocorre quando é possível aleatorizar os beneficiários e comparar com quem não participou do programa. No entanto, nem sempre essa estratégia é viável ou desejável, especialmente quando envolveria excluir indivíduos ou regiões de programas.

Nesses casos, uma alternativa é identificar unidades de comparação que se assemelhem ao máximo às que receberam a política. Esse processo é chamado de pareamento, conhecido pelo termo em inglês *matching*. Técnicas nessa etapa visam reduzir, ao máximo, desequilíbrios entre as covariáveis dos dados observados. É construída, então, uma estrutura comparativa composta por dois grupos: o primeiro contendo as unidades tratadas, ou seja, alvo da política pública; o segundo, denominado grupo de controle, é formado por unidades tão semelhantes quanto possível às tratadas, mas que não receberam o tratamento. Quanto melhor o grupo de comparação, melhor será a estimativa, e existem diversos métodos disponíveis atualmente para a execução desse processo, sendo a questão chave saber qual é o melhor dada as circunstâncias apresentadas.

Esse trabalho contribui para essa etapa fundamental da avaliação de políticas públicas contrapondo duas técnicas distintas: o Genetic Matching (GenMatch) e o Coarsened Exact Matching (CEM). Apesar da consolidação dessas abordagens, a literatura ainda carece de comparações empíricas entre ambas, especialmente no contexto brasileiro. Nessa esteira, buscamos suprir essa ausência ao comparar ambos os métodos em um caso real, por meio da política de Arranjos Produtivos Locais (APLs).

A política foi escolhida por envolver duas dimensões, municípios e setores produtivos, o que adiciona complexidade à etapa de pareamento. Em avaliações de políticas de desenvolvimento econômico essa bidimensionalidade é comum, tornando a discussão relevante para além da política pública escolhida. Para a aplicação das técnicas de *matching*, foram escolhidas como covariáveis principais para as cidades o PIB municipal, a quantidade populacional, o valor agregado na economia e o quociente de localidade (QL).

O primeiro método, o GenMatch, é uma técnica desenvolvida por Sekhon & Mebane (1998) e faz o uso de um algoritmo evolutivo para maximizar o equilíbrio entre as covariáveis. O método é uma generalização da distância de Mahalanobis e busca a minimização de uma função de perda para convergir no melhor pareamento possível.

Isso é realizado pelo algoritmo através da atribuição de pesos às covariáveis, cujos valores compõem a primeira geração genética. O algoritmo avalia, através de testes de Kolmogorov-Smirnov (KS) e t, o conjunto de pesos que maximiza o equilíbrio geral e o usa para a construção de gerações subsequentes, finalmente convergindo em uma solução ótima.

Já o CEM atua de forma *ex-ante*, eliminando o viés antes mesmo do pareamento acontecer. Desenvolvido por Iacus, King & Porro (2011), este método estratifica as variáveis em diferentes categorias, como tercis, quartis ou decis, processo conhecido como *coarsening*. Em sequência, o algoritmo cria estratos e, apenas aqueles que contêm ao menos uma unidade tratada e uma unidade de controle são retidos. Portanto, apesar de serem metodologicamente e filosoficamente diferentes, ambos os métodos têm o potencial de auxiliar o pesquisador no campo das políticas públicas no tratamento dos dados para a posterior análise de causalidade.

Para atingir seus objetivos, esse trabalho estruturou a metodologia em pareamentos individuais no nível dos setores produtivos. A base das covariáveis foi estruturada a partir de dados governamentais, como IBGE e RAIS, e tratadas para permitir a execução do pareamento. A partir da base de APLs brasileiros, diversas combinações entre setor produtivo e APL foram formadas e pareadas uma a uma, resultando em 616 pareamentos distintos e independentes. Para executá-los, foram utilizadas três técnicas diferentes.

Para a primeira, realizada através do GenMatch, foi estruturado um pareamento 1:1 com substituição, e o algoritmo buscou minimizar a maior discrepância dentre os p- valores de cada geração. Já no CEM, duas abordagens foram testadas: uma com estratificação em quintis e outra com variáveis log-transformadas em decis, buscando reduzir a assimetria e preservar o número de unidades. Por fim, as três técnicas culminaram na criação de grupos de tratamento e de controle, permitindo a coleta e comparação de métricas de balanceamento, como diferença de médias padronizada (SMDs), L1, variância, testes-t e KS.

Os resultados mostram que as três técnicas criaram grupos controle satisfatórios. O GenMatch obteve resultados melhores nas estatísticas mensuradas em comparação com o CEM, com o valor de L1 mostrando que mais de 90% das distribuições antes e depois do pareamento se sobrepõem, além de ter um SMD médio entre as covariáveis menor que 0,01. Em contrapartida, ambos pareamentos realizados via CEM resultaram em um grau de sobreposição entre 70% e 80%, com o pareamento estratificado em decis obtendo menor valor do SMD médio do que aquele dividido em

quintis, os valores foram inferiores a 0,08 e 0,09 respectivamente. O método que menos descartou unidades foi o GenMatch, pareando todas as cidades tratadas, enquanto os outros descartaram unidades. Entretanto, a execução do GenMatch exigiu um tempo computacional significativamente maior, superando vinte horas, enquanto nos outros o pareamento foi realizado em menos de dois minutos.

Concluimos que, nesse caso, a estratégia generativa é mais precisa do que a estratificação de variáveis. O GenMatch conseguiu reduzir a assimetria das covariáveis de forma mais eficaz do que as técnicas CEM implementadas. A escolha do método de pareamento, portanto, deve ponderar não apenas o desempenho técnico, mas também as restrições de tempo e infraestrutura da avaliação. Os achados contribuem com evidências práticas para que pesquisadores e formuladores de políticas públicas façam avaliações de impacto de forma mais robusta e estruturada.

#### **Referências:**

- Iacus, S. M., King, G., & Porro, G. (2011). *Multivariate Matching Methods That Are Monotonic Imbalance Bounding*. <https://doi.org/10.1198/jasa.2011.tm09599>
- Sekhon, J. S., & Mebane, W. R. (1998). Genetic Optimization Using Derivatives. *Political Analysis*, 7, 187–210. <https://doi.org/10.1093/PAN/7.1.187>