

Seleção de Grupo de Controle por Pareamento por Escore de Propensão: Ensaio com Método para Avaliação de Política Pública em Bairros do Espírito Santo



NOTA
TÉCNICA

69

Março de 2023

Instituto Jones dos Santos Neves

Seleção de Grupo de Controle por Pareamento por Escore de Propensão: Ensaio com Método para Avaliação de Política Pública em Bairros do Espírito Santo.

Vitória, ES, 2023. 18 p.; il. tab. (Nota Técnica | 69)

1. Pareamento por Escore de Propensão. 2. Grupo de Controle.
3. Avaliação de Impacto. 4. Programa Estado Presente. 5. Bairros do Espírito Santo.

I. Andrade, Bárbara Caballero de. II. Neto, Darcy Ramos da Silva. III. Cayres, Domitila Costa. IV. Título.

As opiniões emitidas são exclusivas e de inteira responsabilidade do(s) autor(es), não exprimindo necessariamente, o ponto de vista do Instituto Jones dos Santos Neves ou da Secretária de Estado de Economia e Planejamento do governo do Estado do Espírito Santo.

GOVERNO DO ESTADO DO ESPÍRITO SANTO

José Renato Casagrande

VICE-GOVERNADORIA

Ricardo Ferraço

SECRETARIA DE ECONOMIA E PLANEJAMENTO – SEP

Álvaro Rogério Duboc Fajardo

INSTITUTO JONES DOS SANTOS NEVES – IJSN

Diretor Presidente

Pablo Silva Lira

Diretoria de Estudos e Pesquisas

Pablo Medeiros Jabor

Diretoria de Integração e Projetos Especiais

Latussa Laranja Monteiro

Diretoria de Gestão Administrativa

Katia Cesconeto de Paula

Elaboração

Darcy Ramos da Silva Neto

Bárbara Caballero de Andrade

Domitila Costa Cayres

Bibliotecário

Jair Rosário Filho



Sumário

1. Introdução	5
2. O Programa Estado Presente em Defesa da Vida	6
3. Pareamento por Escore de Propensão.....	7
4. Dados criminais e socioeconômicos.....	10
5. Resultados para o Programa Estado Presente	12
6. Considerações Finais	18
7. Referências Bibliográficas.....	19
Apêndice: Bairros dos grupos de tratamento e controle.....	20

1. Introdução

Avaliar o impacto de políticas públicas sobre um determinado indicador de interesse (seja na área da saúde, educação ou segurança, por exemplo) é exercício complexo e nada trivial. Seu ponto de partida é a existência de dados antes e após o evento que se quer analisar, que permita a seleção de um grupo de tratamento e de um grupo de controle.

Na avaliação de impacto, o grupo de tratamento refere-se àqueles indivíduos que recebem a intervenção de uma determinada política que se quer investigar seus efeitos. Já o grupo de controle são aqueles indivíduos que possuem características similares ao grupo de tratamento no período anterior à intervenção, com a diferença que não receberam as ações da política. Tal grupo é responsável por garantir que a comparação entre grupos está sendo feita entre homogêneos, sem viés de seleção.

Os métodos mais rigorosos, segundo a literatura, sugerem que a formação desses grupos se dê aleatoriamente antes do início da implementação do programa. Contudo, muitas ações são desenhadas sem a intenção prévia de serem avaliadas posteriormente e sem a seleção aleatória dos grupos – o que não impede sua avaliação. Há métodos quase-experimentais ou não experimentais que podem ser usados para encontrar o melhor grupo de comparação (isto é, o grupo controle).

Nesta Nota Técnica, o método quase-experimental de seleção de grupo de controle por Pareamento por Escore de Propensão (ou *Propensity Score Matching*, em inglês) é aplicado a uma hipotética avaliação de impacto do Programa Estado Presente em Defesa da Vida, importante política pública implementada no Espírito Santo a partir de 2011. O objetivo é mostrar, com dados e métodos econométricos, como a escolha do grupo de controle para uma avaliação do Programa Estado Presente pode ser feita ao nível de bairros quando esses não foram selecionados aleatoriamente.

Para isso, este trabalho está dividido em mais seis seções além desta introdução: a próxima apresenta o Programa e a seguinte detalha o método. Na sequência, os dados utilizados para a seleção do grupo de controle são apresentados para, na seção seguinte,

apresentar os resultados estimados, os bairros selecionados como grupo de controle e as respectivas análises. Por fim, as seções seis e sete trazem, nessa ordem, as considerações finais e as referências bibliográficas.

2. O Programa Estado Presente em Defesa da Vida

O Programa Estado Presente é uma iniciativa do Governo do Estado do Espírito Santo que tem por objetivo promover a segurança pública por meio de estratégias integradas e articuladas entre as secretarias e os órgãos estaduais para o enfrentamento à violência, proteção e defesa social (IJSN, 2019; SESP, 2019). As ações foram divididas entre os eixos de Proteção Policial, sob gerência da Secretaria de Estado da Segurança Pública e Defesa Social (SESP); e de Proteção Social, sob gerência da Secretaria de Estado de Direitos Humanos (SEDH).

A abrangência do programa é estadual com priorização estratégica de algumas regiões. A focalização territorial considerou similaridades no padrão de urbanização, as características socioeconômicas e demográficas e, principalmente, os dados de violência letal (SESP, 2019). Na época da implementação do programa, os dados de homicídios apontavam que cerca de 75% dos homicídios registrados no estado estavam concentrados em dez municípios - cinco na Região Metropolitana da Grande Vitória (Serra, Cariacica, Vila Velha, Vitória e Guarapari) e outros cinco no interior (São Mateus, Linhares, Colatina, Aracruz e Cachoeiro de Itapemirim).

A partir disso, a priorização territorial tomou como ponto de partida o indicador de homicídio que foi complementado com indicadores de vulnerabilidade social, escolaridade e renda (IJSN, 2019), apontando para 43 territórios nos 10 municípios. Cada território abrange um grupo de bairros adjacentes, totalizando 140 bairros focalizados, conforme tabelas A1 e A2 do Apêndice.

Assim, considerando que no desenho do programa a escolha dos bairros alvos das ações não foi aleatória, e sim a partir de características observáveis, nossa proposta é mostrar a aplicabilidade do método de pareamento por score de propensão na

seleção de um grupo de controle a partir de um exercício empírico. Da próxima seção em diante, esses 140 bairros serão chamados de “grupo de tratamento” do Programa Estado Presente.

3. Pareamento por Escore de Propensão

Embora existam diversos métodos para estimar o contrafactual de uma avaliação de impacto, como por “Diferenças em Diferenças” ou “Regressão por descontinuidade”, por exemplo, a literatura apresenta a seleção do grupo de controle por Pareamento por Escore de Propensão como uma das formas mais adequadas quando a escolha do grupo de tratamento não é aleatória e quando as variáveis de interesse (características que serão comparadas entre os grupos) não são medidas antes da intervenção. O método escolhido garante uma seleção de grupo de controle mantendo o rigor técnico exigido mesmo nessas circunstâncias (ROSEMBAUM e RUBIN, 1983; FRASER e GUO, 2015; COWLING, 2017).

Como já foi dito, nesse método busca-se encontrar um grupo de controle adequado para cada unidade tratada (no caso do Programa Estado Presente, bairros), admitindo-se que apenas suas características observáveis (ou seja, as que possuem dados disponíveis na agregação territorial) são responsáveis por tornar “tratados” e “controles” diferentes. Assim, olhando dois bairros (um de cada grupo), o único critério que irá diferenciá-los é a informação de quem é tratamento e de quem é controle, dado que as características observáveis são, em média, as mais próximas possíveis. Em resumo, o pareamento será criado de acordo com as variáveis utilizadas para se obter um escore de propensão, ou seja, uma probabilidade de o bairro do grupo de controle pertencer ao grupo de tratamento, onde o único fator de diferença entre eles é a informação de qual é de um grupo e qual é do outro.

Formalmente, tem-se que o objetivo do método é encontrar um conjunto de unidades não tratadas que sejam muito semelhantes aos participantes da política em um vetor X de todas as características pré-tratamento observáveis (CALIENDO e

KOPEINIG, 2008). A primeira condição necessária para a estimação do efeito médio sobre os tratados é chamada de hipótese de independência condicional:

$$Y^0 \perp D|X \quad (1)$$

Ou seja, os resultados potenciais dos bairros não tratados, Y^0 , condicionados em X , são independentes de sua atribuição de tratamento, D . Assim, o resultado de um indivíduo (bairro) no grupo de controle torna-se um bom preditor do resultado potencial de um indivíduo (bairro) na ausência do tratamento, desde que compartilhem o mesmo vetor X de características observáveis. A segunda hipótese exigida é a de sobreposição (ou suporte comum):

$$0 < \Pr[D = 1|X] < 1 \quad (2)$$

Isto é, faz-se necessário que, para cada valor de X , existam indivíduos tratados (bairros selecionados para receber o Programa) e não tratados (bairros não selecionados para receber o Programa), de forma que seja possível que cada unidade tratada tenha uma contraparte similar no grupo de controle que venha a representar seu contrafactual¹ (CALIENDO e KOPEINIG, 2008).

É mostrado que, se a hipótese de independência condicional vale para o vetor X , então ela também será verdadeira para funções de X . O escore de propensão é uma dessas funções, sendo, então, a probabilidade de participar do tratamento condicional em X :

$$P(x) = \Pr[D = 1|X = x] \quad (3)$$

¹ O exercício contrafactual é estimar como os bairros selecionados para o programa estariam hoje, em termos de violência, se o programa nunca tivesse sido aplicado. Ou seja, serve para medir exatamente o efeito da intervenção sobre o indicador de violência.

Assim, a hipótese de independência condicional pode ser reescrita em termos do *Propensity Score* (ROSEMBAUM e RUBIN, 1983):

$$Y^0 \perp D | P(x) \quad (4)$$

A introdução do escore de propensão conseguiu contornar o problema da dimensão do vetor X , que tornaria cada vez mais difícil encontrar bairros tratados e não tratados com a mesma característica em cada variável. Além disso, tornou-se possível testar a condição de balanceamento:

$$Y^0 \perp x | p(x) \quad (5)$$

Ou seja, uma vez que a atribuição do tratamento condicional no escore de propensão é aleatória, não deve haver diferenças significativas nas características observáveis de tratados e não tratados.

Para melhor ajuste do método, foram testados os processos que buscam o pareamento de bairros “sem reposição” e “com reposição”. Assim, dentre os 1.294 bairros do estado do Espírito Santo que não foram priorizados pelo Estado Presente, temos que selecionar até 140 bairros como grupo de controle. Se a seleção é feita “sem reposição”, significa que o grupo de controle terá exatamente o mesmo número de bairros do tratamento. Assim, se o bairro Z do grupo de controle é pareado com o bairro A no grupo de tratamento ele é descartado da amostra como candidato a controle de outro(s) bairro(s) no grupo de tratamento.

Entretanto, o método “com reposição” é sugerido, pela própria literatura, como mais adequado ao modelo, podendo, conseqüentemente, o grupo de controle ter um número menor de bairros que o grupo de tratamento. Por exemplo: um mesmo bairro Z pode ser, ao mesmo tempo, controle dos bairros de tratamento A e B , já que está disponível para ser sorteado mais de uma vez (por isso “com reposição”), caso ele tenha

as características bem parecidas com esses dois bairros (*A* e *B*) do grupo de tratamento. Às vezes, um bairro pode ser o mais parecido em termos de variáveis com outro, mas, sem reposição, ele já teria sido usado “como par” de outro, não podendo selecioná-lo novamente.

4. Dados criminais e socioeconômicos

Como posto, a escolha dos territórios do Programa Estado Presente não foi aleatória e sim orientada por indicadores criminais e socioeconômicos. Para selecionar as unidades que comporão o grupo de controle, o primeiro passo foi lastrear esses dados por bairro em data anterior ao início da intervenção, isto é, antes do ano de 2011.

Uma primeira dificuldade foi a inconsistência nos dados de criminalidade devido à falta de informação e/ou mau preenchimento dos dados. As informações disponíveis apresentavam apenas números agregados por município, não sendo possível levantar as ocorrências de homicídio por bairro no ano de 2010 para todos os municípios do estado. Cabe frisar que, para o desenho do Programa Estado Presente, foi feito um levantamento específico nos 10 municípios com o maior número de homicídios para a identificação e priorização dos bairros alvos. Ademais, novos bairros surgiram após 2010, não sendo possível atribuir corretamente o dado de homicídio a essas localidades. Assim, apesar da centralidade do indicador de homicídios na seleção dos municípios e bairros priorizados pelo programa, não foi possível a utilização desses dados para a seleção das características de pareamento entre o grupo de tratamento e o grupo de controle.

No que tange aos indicadores socioeconômicos, o desafio foi compatibilizar os dados por bairro a partir das informações disponibilizados em setor censitário pelo Censo Demográfico do IBGE (2011). Como em 2010 apenas 12 dos 78 municípios do Espírito Santo possuíam delimitação oficial de seus bairros, a solução foi utilizar os dados do Censo por setor censitário e agregá-los a partir da demarcação geográfica utilizada atualmente pelo Instituto Jones dos Santos Neves.

Contudo, como esperado, os limites do setor censitário e dos bairros não eram equivalentes e foram estabelecidos critérios para a compatibilização. Assim, quando pelo menos 95% da área de um setor censitário estava dentro de um bairro, todas as estatísticas foram atribuídas àquele bairro. Do contrário, os dados foram repartidos entre os bairros². Nos casos em que o setor possuía uma parte em um bairro urbano e outra parte em área rural ou em água, as estatísticas do setor foram atribuídas em sua totalidade ao bairro urbano. Foram ainda excluídos os setores em que 100% do seu território estava em área rural.

A agregação dos setores censitários em bairros permitiu que apenas estivessem disponíveis os dados socioeconômicos provenientes do “Questionário Universo” do Censo, que apresenta um rol menor de variáveis quando comparadas àquelas do “Questionário Amostra” do Censo – que contém uma série de informações socioeconômicas sobre os domicílios e seus residentes. Assim, para selecionar, por exemplo, um dado sobre educação, só foi possível usar a taxa de alfabetização dos moradores, e não outras mais específicas, como taxas de abandono, taxa de matrícula ou nível de escolaridade, que estão disponíveis somente no “Questionário Amostra”.

Assim, para identificar as características dos bairros que seriam utilizadas na comparação entre o grupo de controle e de tratamento, foram testados vários indicadores criados a partir dos dados disponíveis no Censo³. As variáveis que deram melhor resposta estatística ao modelo foram o percentual de homens negros (pretos e

² Por exemplo, quando um setor possuía 60% do território no bairro *A* e 40% no bairro *B*, as estatísticas foram divididas com base nesse percentual: se no setor habitavam 100 pessoas, foram contabilizadas 60 pessoas para o bairro *A* e 40 pessoas para o bairro *B*. Esse procedimento foi adotado para todas as características socioeconômicas selecionadas, mesmo sabendo-se que as características não são distribuídas uniformemente pelo território.

³ Foram utilizados, por exemplo, o total da população residente no bairro e totais e percentuais da população de: pessoas responsáveis pelo domicílio do sexo feminino/masculino; pessoas responsáveis pelo domicílio alfabetizadas do sexo feminino/masculino; pessoas alfabetizadas do sexo feminino/masculino em diversos recortes etários; pessoas jovens (diversos recortes etários); homens negros (pretos e pardos) jovens (diversos recortes etários); entre outros

pardos) de 15 a 29 anos⁴ sobre o total da população, o percentual de alfabetizados de 10 a 17 anos⁵ sobre o total da população, e a renda domiciliar *per capita*.

5. Resultados para o Programa Estado Presente

Nosso exemplo empírico para demonstrar a utilização desse método é uma hipotética avaliação do Programa Estado Presente, que aplicou ações prioritárias em 140 bairros do estado do Espírito Santo (ver Tabela A1 do Apêndice). A partir disso, é necessário encontrar um grupo de controle compatível em relação às variáveis, ou seja, um grupo de bairros que tenha características similares aos bairros do grupo de tratamento antes do momento da intervenção. Segue o modelo empírico de análise desta Nota Técnica:

$$probit\{\Pr(Y_{i2010}) = 1\} = \alpha + \beta X_i + \epsilon_i \quad (6)$$

Em que X_i são as variáveis utilizadas na modelagem, para cada bairro i (renda domiciliar *per capita*, percentual de homens pretos e pardos de 15 a 29 anos sobre o total da população e percentual de pessoas alfabetizadas de 10 a 17 anos sobre o total da população)⁶, e ϵ_i o termo de erro.

Para esta aplicação, foram utilizados 137 dos 140 bairros⁷ do Programa Estado Presente como grupo de tratamento. Também foram feitos dois testes em relação às

⁴ Esse corte etário é utilizado na literatura de crime, que afirma que, nessa idade, os jovens são mais propensos a estarem envolvidos em atividades criminais (HIGGINSON *et al.*, 2016). É ainda o recorte utilizado em um dos indicadores do Programa Estado Presente na focalização territorial (IJSN, 2019, p. 13-15).

⁵ Considera-se que é nessa fase que os jovens estão mais suscetíveis a evasão escolar e que, segundo prevê a legislação brasileira, há obrigatoriedade do ensino para jovens até os 17 anos.

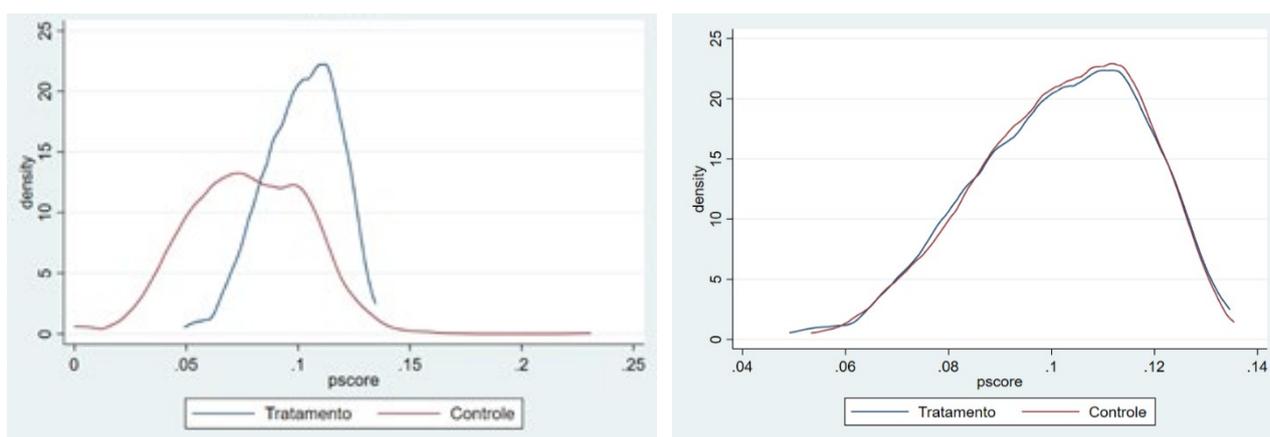
⁶ Foram feitos testes com diversas variáveis, como citado anteriormente, que possuem correlação com homicídios e demais tipos de crime, segundo a literatura. As três variáveis utilizadas foram as únicas que fizeram um ajuste adequado no pareamento por escore de propensão.

⁷ No Programa são priorizados 140 bairros. Entretanto, três bairros foram excluídos por não terem delimitação oficial ou por estarem em sua maior parte em área rural (Nova Rosa da Penha II, em Cariacica; Jabaeté, em Vila Velha; e Capixaba, em Vitória).

características a serem comparadas para verificarmos qual ajuste ficaria mais adequado, dado que existe o *trade-off* do método em relação ao número de variáveis utilizadas: uma única variável não possui poder de predição suficiente para um ajuste adequado (não é suficiente para comparar os bairros); enquanto a inclusão de muitas variáveis pode desajustar o pareamento, sem conseguir bairros muito parecidos entre si.

No primeiro teste com reposição (Figura 1 e Tabela 1), foi criado um escore de propensão com duas variáveis – percentual de homens pretos e pardos de 15 a 29 anos e renda *per capita* domiciliar.

Figura 1 – Densidade do escore de propensão por grupo de tratamento ou controle, antes (esquerda) e depois (direita) do pareamento com duas variáveis



Fonte: Elaboração própria a partir da base de dados.

O estado do Espírito Santo possui, segundo levantamento do Instituto Jones dos Santos Neves, 1.431 bairros em área urbana. Como 137 foram escolhidos para receber o Programa, restaram 1.294 bairros candidatos ao grupo de controle. O teste 1 encontrou 126 bairros como grupo de controle para os 137 de tratamento, considerando duas características de comparação. É possível notar, pelo gráfico da direita na Figura 1, que o pareamento funcionou bem para esse modelo, encontrando os bairros com as características observáveis muito parecidas com as do grupo de

tratamento. Com isso, é possível ter um grupo adequado de comparação com o grupo de tratamento para que se possa fazer uma avaliação de impacto com rigor técnico.

O modelo faz um pareamento com as variáveis observáveis e calcula um escore de propensão, onde o esperado é que as curvas de tratamento e controle estejam o mais coincidentes possível pós-pareamento, indicando que o método fez a combinação entre bairros com características iguais. Pelos gráficos pode-se perceber as diferenças das curvas antes e depois do pareamento. Antes (gráfico da esquerda na Figura 1), os dados dos bairros apareciam dispersos, sem qualquer relação de proximidades ou características. Pós-pareamento, os dados foram organizados a partir das variáveis de interesse indicando proximidade de características e um bom pareamento dos dados (observado no gráfico da direita da Figura 1).

Tabela 1 – Probabilidade do bairro do grupo de controle pertencer ao grupo de tratamento, usando duas variáveis de comparação

<i>Probit</i>	Coefficiente (Erro padrão)
Renda <i>per capita</i> domiciliar	0,0001 (0,0001)
Percentual de homens pretos e pardos de 15 a 29 anos sobre o total da população	21,695*** (2,444)
Percentual de alfabetizados de 10 a 17 anos sobre o total da população	-0,7765 (0,0002)
Nº de observações	1.410
<i>Log Likelihood</i>	-396,303
Pseudo R2	0,119

Nota 1: 21 bairros foram excluídos por falta de informação na base do Censo sobre as variáveis utilizadas.

Nota 2: Significância estatística de 1%***, 5%** e 10%*.

Fonte: Elaboração própria a partir da base de dados.

A Tabela 1 mostra o *probit* no qual o método descreve que, se não houvesse a indicação de quem é grupo de controle e de tratamento, não seria possível descobrir em que grupo cada unidade estaria. De outra forma, o *probit* apresenta a probabilidade de um bairro do grupo de controle pertencer ao grupo de tratamento dadas as variáveis

observáveis: sobre o percentual de homens pretos e pardos de 15 a 29 anos, a razão de chance⁸ de ser factível é 21,6 vezes maior.

O percentual de homens pretos e pardos de 15 a 29 anos sobre o total da população apresenta um estimador significativo estatisticamente a 1%, sendo, então, a variável mais relevante para esse pareamento. Ou seja, a principal característica considerada para selecionar os bairros de controle mais parecidos com os de tratamento é o percentual de homens jovens negros.

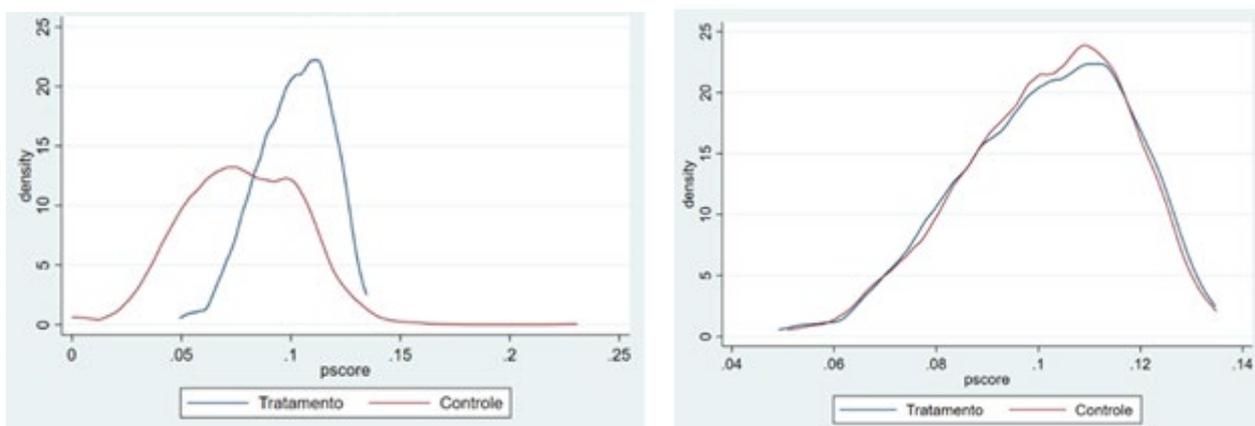
O pseudo-R2 é uma estatística, medida em percentual, que traduz o quanto o modelo consegue explicar o pareamento de acordo com as variáveis que foram utilizadas. O poder de explicação dos dados desse modelo, dada as variáveis utilizadas, foi de 11,9%, que é um percentual baixo, decorrente do pequeno número de variáveis disponíveis.

O *Log Likelihood* (LL) é, de maneira geral, a estimação de diferentes valores que maximiza a probabilidade de ocorrência dos dados observados. Via de regra, um número alto (em módulo) de LL (como o encontrado de -396,303) condiz com a distribuição Normal dessa amostra, ou seja, quanto maior for o valor do LL, diz-se que melhor o modelo se ajustou a esse conjunto de dados.

Para o segundo teste (Figura 2 e Tabela 2), além das duas variáveis utilizadas, foi adicionada uma terceira variável, que é o percentual de pessoas alfabetizadas de 10 a 17 anos sobre o total da população. Esse teste resultou um grupo de controle com 120 bairros pareados aos 137 bairros do tratamento.

⁸ Razão de chance (ou razão de possibilidades) é a razão entre a chance de um evento ocorrer em um grupo e a chance do mesmo evento ocorrer em outro grupo. Uma alta probabilidade daquele bairro pertencer ao grupo de tratamento (mesmo sabendo que ele está no grupo de controle), significa que ele se enquadra, a partir desses dados, nos requisitos.

Figura 2 – Densidade do escore de propensão por grupo de tratamento ou controle, antes (esquerda) e depois (direita) do pareamento com três variáveis



Fonte: Elaboração própria a partir da base de dados.

Pela aproximação das curvas pós-pareamento (gráficos da direita) das Figuras 1 e 2, é possível notar que elas se ajustaram mais no teste 1 do que no 2. Entretanto, o segundo teste trouxe mais variáveis que o teste 1, elevando, ainda que timidamente, o poder de explicação do modelo (Pseudo R²). No mais, o *probit* do teste 2 teve um ajuste com mais significância estatística do que no teste 1, como mostra a Tabela 2.

Tabela 2 – Probabilidade de o bairro do grupo de controle pertencer ao grupo de tratamento, usando três variáveis de comparação

<i>Probit</i>	Coefficiente (Erro padrão)
Renda <i>per capita</i> domiciliar	0,0005** (0,0002)
Percentual de homens pretos e pardos de 15 a 29 anos sobre o total da população	21,841*** (2,943)
Percentual de alfabetizados de 10 a 17 anos sobre o total da população	-0,7765 (0,0002)
Nº de observações	1.365
<i>Log Likelihood</i>	-395,142
Pseudo R ²	0,122

Nota 1: 66 bairros foram excluídos por falta de informação na base do Censo sobre as variáveis utilizadas.

Nota 2: Significância estatística de 1%***, 5%** e 10%*.

Fonte: Elaboração própria a partir da base de dados.

Com a inclusão da variável do percentual de alfabetizados de 10 a 17 anos sobre o total da população, a variável de renda *per capita* domiciliar se mostrou significativa estatisticamente a 5%. Embora parecidos, os valores de LL e Pseudo R2 foram levemente melhores em relação ao primeiro teste, indicando que este segundo teste, que utilizou mais variáveis que o modelo anterior, se mostrou mais adequado para a escolha do grupo de pareamento. Mesmo que umas das premissas do método, conforme já mencionado, seja o *trade-off* entre o uso de mais variáveis e o “desequilíbrio” desse modelo, o teste 2 mostrou que é, de acordo com os dados disponíveis, o mais aderente para a escolha de um grupo de bairros controle para o grupo de bairros tratamento do Estado Presente.

Em relação às probabilidades, no que diz respeito ao percentual de homens pretos e pardos de 15 a 29 anos na população, o resultado está bem próximo ao teste anterior. O que mudou foi que a variável de renda domiciliar *per capita*, mesmo que apontando um valor baixo, mostrou resultado estatisticamente significativo a 5%. A razão de chance de os bairros do grupo de controle pertencerem ao grupo de tratamento olhando para essa variável é 0,0005 maior.

Portanto, o grupo mais adequado a essa modelagem é o resultante do teste 2, pois teve mais variáveis significativas ao modelo e, visualmente, o pareamento também ajustou de forma adequada, mesmo que o Pseudo R² de ambos tenha apresentado valores muito próximos.

Em outros termos, o grupo de controle para o grupo de tratamento (com características observáveis parecidas) antes da intervenção do Programa Estado Presente, o qual pode ser usado como um grupo de comparação e servir de base para uma eventual avaliação de impacto, são os do teste 2, que utilizou três características para pareamento e que estão listados na Tabela B1 do Apêndice. Nessa tabela é possível perceber que, enquanto os 140 bairros do grupo de tratamento se concentravam em somente dez municípios, os bairros do grupo de controle foram selecionados de maneira bem mais dispersa (126 bairros em 35 municípios) e bem mais presente no interior do estado (dos 126, 55 pertenciam aos sete municípios da RMGV e 71 ao interior do

estado). No Programa, os 140 bairros priorizados estavam em sua maioria na RMGV (99) e somente 41 no interior.

Os municípios com maior número de bairros no grupo de controle foram Cariacica, com 17 (e 19 de tratamento) e Serra, com 11 (e 22 de tratamento). Chama a atenção os municípios que não foram identificados como prioritários no Estado Presente, mas que em 2010 eram mais parecidos com os selecionados para o Programa: Conceição da Barra, com 8 bairros selecionados para controle; Viana, na RMGV, com 6; Anchieta, Barra de São Francisco e Pinheiros, com 4 cada.

Vila Velha e Vitória, os municípios com mais bairros atendidos pelo Estado Presente (31 e 24, respectivamente), tiveram apenas 5 e 6 bairros selecionados para o grupo de controle, não necessariamente adjacentes aos do grupo de tratamento.

6. Considerações Finais

Esta Nota Técnica cumpriu a função de encontrar o grupo de controle mais adequado para o grupo de tratamento do Programa Estado Presente, ao nível de bairro, utilizando o método quase-experimental de Pareamento por Escore de Propensão, apesar dos desafios colocados: a não aleatorização prévia para seleção do grupo de tratamento; a ausência de dados dos indicadores de interesse ao nível de bairro em um período pré-tratamento; e o agrupamento dos setores censitários em bairros.

O exercício feito apresenta os caminhos que podem ser trilhados quando nos deparamos com problemas parecidos em relação à disponibilidade de dados anteriores a um evento e também à agregação de setores censitários em bairros. Demostramos como é possível fazer uma comparação dos bairros do Programa Estado Presente com um grupo de bairros de controle, fornecendo às secretarias e órgãos estaduais engajados nas ações do Programa um instrumento para avaliar suas iniciativas específicas.

Igualmente, demonstramos que a vantagem desse método é a sua replicabilidade, visto que existem outros casos em que é preciso escolher um grupo de controle adequado a um grupo de tratamento que sofreu algum tipo de intervenção sem aleatorização prévia – o que pode facilitar outras pesquisas que possuem as mesmas condições.

7. Referências Bibliográficas

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of Propensity score Matching. **Journal of Economic Surveys**, Volume 22, Issue 1, 2008.

COWLING, B. **Propensity Score Analysis**. The University of Hong Kong, 2017.

FRASER, M. W.; GUO, S. **Propensity Score Analysis: statistical methods and applications**. 2nd edition. SAGE Publications, 2015.

HIGGINSON, A. *et al.* Youth Membership and violence in low-and middle-income countries: A systematic review. **International Initiative for Impact Evaluation**, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Base de Informações do Censo Demográfico 2010: Resultados do Universo por Setor Censitário**. Rio de Janeiro: IBGE, 2011.

INSTITUTO JONES DOS SANTOS NEVES (IJSN). **Eixo Proteção Social**. 1ª ed. Vitória: IJSN, 2019. Disponível em: <<http://www.ijsn.es.gov.br/artigos/5662-eixo-protecao-social-programa-estado-presente>>. Acessado em: 02 jan. 2023.

ROSEMBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, 70:41-55, 1983.

SECRETARIA DE ESTADO DE SEGURANÇA PÚBLICA E DEFESA SOCIAL (SESP). **Eixo Proteção Policial: Manual Básico**. 1ª ed. Vitória: Programa Estado Presente, 2019. Disponível em: <<http://www.ijsn.es.gov.br/artigos/5403-programa-estado-presente>>. Acessado em: 02 jan. 2023.

Apêndice: Bairros dos grupos de tratamento e controle

Tabela A1 – Grupo de Tratamento: Bairros priorizados pelo Programa Estado Presente, por município, e se pertence à RMGV

Município	Bairro	RMGV
Aracruz	Barra do Riacho, Bela Vista, Fátima, Guaraná, Jacupemba, São Clemente, Segato, Vila do Riacho	Não
Cachoeiro De Itapemirim	Alto Novo Parque, Gilson Carone, Monte Cristo, Novo Parque, Village da Luz e Zumbi	Não
Cariacica	Alzira Ramos, Bandeirantes, Castelo Branco, Flexal I, Flexal II, Ipiranga, Jardim de Alah, Loteamento Cordovil, Nova Esperança, Nova Rosa Da Penha I, Nova Rosa Da Penha II, Padre Gabriel, Parque Residencial Maracanã, Rio Marinho, São Bernardo, São Rafael, Vale do Marinho, Vila Isabel e Vista Mar	Sim
Colatina	Alto São Vicente, Ayrton Senna, Bela Vista, Jardim Planalto, Operário, Pôr do Sol, Santa Cecília, Santo Antônio, São Judas Tadeu, São Pedro, São Vicente, Vicente Suella	Não
Guarapari	Adalberto Simão Nader, Coroado, Jabarai, Kubitschek	Sim
Linhares	Aviso, Interlagos, Jocafe I, Jocafe II, Nova Esperança, Planalto, Santa Cruz	Não
São Mateus	Bom Sucesso, Cacique, Guriri, Litorâneo, Vila Nova, Vitória	Não
Serra	André Carloni, Bairro das Laranjeiras, Campinho da Serra I, Campinho da Serra II, Carapina, Carapina Grande, Central de Carapina, Conjunto Feu Rosa, Costa Dourada, Enseada de Jacaraípe, Estância Monazítica, Jardim Atlântico, Jardim Carapina, Lagoa de Jacaraípe, Novo Horizonte, Parque Jacaraípe, Planalto Serrano, Residencial Jacaraípe, São Francisco, São Patrício, Sítio Esmeralda (Ourimar), Vila Nova de Colares	Sim
Vila Velha	23 de Maio, Alecrim, Argolas, Boa Vista I, Boa Vista II, Chácara do Conde, Cidade da Barra, Cobi de Baixo, Cobi de Cima, Cristóvão Colombo, Divino Espírito Santo, Estrela (Barramares), Ilha Da Conceição, Ilha Dos Aires, Jabaeté, João Goulart, Morada da Barra, Norminia da Cunha Azeredo, Pedra dos Búzios, Primeiro de Maio, Riviera da Barra, Sagrada Família, Santa Rita, São Conrado, São Torquato, Soteco, Terra Vermelha, Ulisses Guimarães, Vila Garrido, Vista da Penha, Zumbi dos Palmares	Sim
Vitória	Bairro da Penha, Bela Vista, Bonfim, Capixaba, Condusa, Conquista, Consolação, Gurigica, Ilha das Caieiras, Ilha do Príncipe, Inhanguetá, Itararé, Morro da Fonte Grande, Morro da Piedade, Morro de São Benedito, Morro do Moscoso, Nova Palestina, Redenção, Resistência, Santo André, Santo Antônio, Santos Reis, São José, São Pedro, Vila Rubim	Sim

Tabela A2 – Grupo de Tratamento: Número de bairros priorizados pelo Programa Estado Presente, por município, e se pertence à RMGV

Município	Nº de bairros	RMGV
Vila Velha	31	Sim
Vitória	25	Sim
Serra	22	Sim
Cariacica	19	Sim
Colatina	12	Não
Aracruz	8	Não
Linhares	7	Não
Cachoeiro de Itapemirim	6	Não
São Mateus	6	Não
Guarapari	4	Sim
10 municípios	140 bairros	5

Tabela B1 – Grupo de Controle: Bairros selecionados pelo Pareamento por Escore de Propensão, por município, e se pertence à RMGV

Município	Bairro	RMGV
Afonso Cláudio	Jandira Alves Giesta	Não
Alegre	Chácara da Serra, São Manoel	Não
Anchieta	Cantagalo, Justiça II, Ponta dos Castelhanos, Praia de Parati	Não
Aracruz	Morobá, Nova Colatina, Santa Marta, São Marcos, Sauaçu	Não
Atílio Vivácqua	Área de Expansão Urbana	Não
Baixo Guandu	Alto Guandu, Rosário I, Rosário II	Não
Barra de São Francisco	Colina, Vila Gonçalves, Vila Landinha, Vila Luciene	Não
Boa Esperança	Boa Vista	Não
Cachoeiro de Itapemirim	Alto Independência, Bom Pastor, Marbrasa, São Lucas	Não
Cariacica	Bela Vista, Caçaroca, Cruzeiro do Sul, Expedito, Formate, Foz do Rio Bubú, Itapemirim, Jardim Botânico, Parque Gramado, Piranema, Porto se Cariacica, Presidente Médici, Santo Antônio, São Gonçalo, Tabjara, Tucum, Vale dos Reis	Sim
Castelo	Niterói	Não
Colatina	Aeroporto, Santa Margarida, São Marcos, Vila Lenira	Não
Conceição da Barra	Floresta, Aluísio, Antônio Lopez, Campo Verde II, Cobrace, Nossa Senhora Aparecida, Nova Bethânia, São Jorge	Não
Ecoporanga	Benedita Monteiro, Centro, Santa Rita	Não
Fundão	Balneário Costa Azul, Loteamento Direção	Sim
Guarapari	Camurugi, Comunidade Urbana de Machinda, Comunidade Urbana de Palmeiras, Comunidade Urbana de Taquara do Reino, Condados, Fátima Cidade Jardim, Itapebussu, Santa Margarida	Sim
Ibiraçu	Parque Industrial José Luiz Fiorotti, Pedra Branca	Não
Linhares	Movelar	Não
Mantenópolis	Antur Artur, Ipiranga	Não
Mimoso Do Sul	Itapoá	Não
Montanha	Decão, Esperança	Não
Mucurici	Planalto, Sostene Ferreira	Não
Nova Venécia	Altoé, Rúbia	Não
Pedro Canário	Boa Vista, Camata	Não
Pinheiros	Canário, Galileia, Loteamento Amaranto, Lulina	Não
Piúma	Boa Vista, Céu Azul	Não
Ponto Belo	Alvorada, Centro, Morumbi	Não
Rio Novo do Sul	Nossa Senhora da Penha	Não
São Mateus	Alvorada, Aviação, Centro	Não
Serra	Barro Branco, Cascata, Diamantina, Lagoa de Carapebus, Laranjeiras Velha, Nova Carapina II, Ourima, Polo Industrial Tubarão, Porto Canoas, Solar de Anchieta, Taquara II	Sim
Sooretama	Centro, Sayonara	Não
Venda Nova do Imigrante	Bela Vista	Não
Viana	Araçatiba, Bom Pastor, Canaã, Industrial, Morada Bethânia, Soteco	Sim
Vila Velha	Brisamar, Garoto, Planalto, Pontal das Garças, Ulisses Guimarães	Sim
Vitória	Andorinhas, Morro do Cabral, Morro do Quadro, Grande Vitória, Jesus de Nazareth, Santa Tereza	Sim

Tabela B2 – Grupo de Controle: Número de bairros selecionados pelo Pareamento por Escore de Propensão, por município, e se pertence à RMGV

Município	Nº bairros	RMGV
Cariacica	17	Sim
Serra	11	Sim
Guarapari	8	Sim
Conceição da Barra	8	Não
Viana	6	Sim
Vitória	6	Sim
Aracruz	5	Não
Vila Velha	5	Sim
Anchieta	4	Não
Barra de São Francisco	4	Não
Cachoeiro de Itapemirim	4	Não
Colatina	4	Não
Pinheiros	4	Não
Baixo Guandu	3	Não
Ecoporanga	3	Não
Ponto Belo	3	Não
São Mateus	3	Não
Alegre	2	Não
Fundão	2	Sim
Ibiraçu	2	Não
Mantenópolis	2	Não
Montanha	2	Não
Mucurici	2	Não
Nova Venécia	2	Não
Pedro Canário	2	Não
Piúma	2	Não
Sooretama	2	Não
Afonso Cláudio	1	Não
Atílio Vivácqua	1	Não
Boa Esperança	1	Não
Castelo	1	Não
Linhares	1	Não
Mimoso do Sul	1	Não
Rio Novo do Sul	1	Não
Venda Nova do Imigrante	1	Não
35 municípios	126	7