

Universidade de São Paulo  
Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto  
Departamento de Economia  
Programa de Pós-Graduação em Economia – Área: Economia Aplicada

Lucas Reis Correia

**Efeitos educacionais da expansão repentina de vagas em educação  
infantil: Petrolina e Juazeiro, um experimento natural**

ORIENTADOR: PROF. DANIEL DOMINGUES DOS SANTOS

Ribeirão Preto  
2018

Prof. Dr. Vahan Agopyan  
Reitor da Universidade de São Paulo

Prof. Dr. Dante Pinheiro Martinelli  
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Renato Leite Marcondes  
Chefe do Departamento de Economia

Prof. Dr. Sérgio Naruhiko Sakurai  
Coordenador do Programa de Pós-graduação em Economia – Área: Economia Aplicada

LUCAS REIS CORREIA

Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

**Efeitos educacionais da expansão repentina de vagas em educação infantil: Petrolina e Juazeiro, um experimento natural**

Dissertação de mestrado submetida ao Programa de Pós-graduação em Economia-Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, como parte dos requisitos necessários para a obtenção do título de Mestre em Ciências.

ORIENTADOR: PROF. DANIEL DOMINGUES DOS SANTOS

Ribeirão Preto

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Lucas Reis Correia

Efeitos educacionais da expansão repentina de vagas em educação infantil: Petrolina e Juazeiro, um experimento natural / Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto Programa de Pós-Graduação – Economia Aplicada; Orientador. Prof. Dr. Daniel Domingues dos Santos Ribeirão Preto, 2018. 52p.

Dissertação (Mestrado) – Universidade de São Paulo, 2018.

1. Nova Semente 2. Creches 3. Avaliação de Impacto I. Orientador: Prof. Dr. Daniel Domingues dos Santos. II. Universidade de São Paulo – Campus Ribeirão Preto. III. Faculdade de Economia Administração e Contabilidade. IV. Efeitos educacionais da expansão repentina de vagas em educação infantil: Petrolina e Juazeiro, um experimento natural

## **Resumo**

**CORREIA, L, R. Efeitos educacionais da expansão repentina de vagas em educação infantil: Petrolina e Juazeiro, um experimento natural.** 2018. Dissertação (Mestrado) – Faculdade de Economia Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo. Ribeirão Preto, 2018.

Há grande evidência internacional e nacional sobre a importância da primeira infância no desenvolvimento humano, entre as principais políticas públicas com este foco, destaca-se a criação de creches. No presente projeto, analisar-se-á o impacto do programa de criação de creches em Petrolina (PE), o Nova Semente. Mais especificamente, a partir dos dados do Censo Escolar e do Cadastro Único, deseja-se observar se a introdução de creches contribui para melhora do desempenho educacional dos alunos através de suas aprovações e frequência escolar.

**Palavras-chave:** Nova Semente; Creches; Avaliação de impacto; Educação

## **Abstract**

**CORREIA, L, R. Educational effects of a sudden expansion in early childhood school supply: Petrolina and Juazeiro, a natural experiment.**

Dissertation (Masters) – School of Economics, Business and Accounting of Ribeirão Preto, University of São Paulo. Ribeirão Preto, 2018.

There is extensive literature worldwide on the importance of early childhood on human development, childcare policies stand out as one of the main public policies with this aim. This Project intends to evaluate the impact of Nova Semente, a program in Petrolina (PE), which increases childcare supply. Specifically, educational development, through approvals and frequency, will be assessed using data from Censo Escolar and Cadastro Único.

**Key-Words:** Nova Semente; Childcare services; Impact evaluation; Education

## **Lista de Tabelas**

Tabela 1: Comparação Geral.....	18
Tabela 2: Comparação em Educação .....	18
Tabela 3: Descrição das Variáveis do Censo Escolar .....	21
Tabela 5: Efeito Médio do Nova Semente na Aprovação do Segundo Ano pelo Censo Escolar .....	28
Tabela 11: Variáveis utilizadas do Cadastro Único .....	33
Tabela 12: Balanceamento do Painel do Cadastro Único .....	33
Tabela 13: Contaminação do Programa Nova Semente .....	34
Tabela 14: Resultados pelos Modelos de Logit em Painel.....	37
Tabela 15: Efeitos na Frequência Escolar pelo Cadastro Único .....	38
Tabela 16: Médias de frequências totais e de mulheres entre 11 e 17 anos.....	39

## **Lista de Tabelas do Apêndice**

Tabela A 1: Observações de Dependência Administrativa .....	49
Tabela A 2: Observações por ano .....	49
Tabela A 3: Estatísticas Descritivas Censo Escolar .....	50
Tabela A 6: Probit do Escore de Propensão .....	50
Tabela A 7: Suporte Comum do Pareamento .....	51
Tabela A 8: Teste de Pareamento .....	51
Tabela A 12: Proporção de Beneficiários do Bolsa Família na Amostra .....	51
Tabela A 13: Estatísticas Descritivas das Variáveis do Cadastro único.....	52
Tabela A 14: Efeitos na Frequência de Mulheres entre 10 e 17 Anos.....	52

## **Lista de Figuras**

Figura 1: Petrolina e Juazeiro .....	20
Figura 2: Expansão de Creches e Pré-escolas pelo Censo Escolar.....	22
Figura 3: Paralelismo pré pareamento de variáveis dependentes .....	23
Figura 4: Anos de exposição ao programa por turmas de ensino fundamental .....	23

Figura 5: Paralelismo Pós Pareamento do Censo Escolar.....	27
Figura 6: Evolução da Frequência Escolar de Mulheres entre 11 e 17 anos .....	40



## Sumário

1. Introdução .....	10
2. Revisão da Literatura .....	12
a. Expansão de vagas em educação infantil .....	12
b. Nova Semente .....	14
c. Métodos Estatísticos .....	15
i. Modelo de escolha binária .....	15
ii. Pareamento e Diferenças em Diferenças .....	16
iii. Juazeiro como controle de Petrolina .....	18
3. Censo Escolar – Escolas .....	21
a. Base de dados - Censo Escolar .....	21
b. Metodologia e limitações .....	24
c. Resultados - Pareamento com Censo Escolar .....	26
d. Discussão .....	30
4. Cadastro Único .....	32
a. Base de dados .....	32
b. Metodologia .....	35
c. Resultados .....	37
d. Discussão .....	41
5. Conclusão .....	43
6. Referências .....	44
7. Apêndice .....	49
a. Tabelas .....	49

## 1. Introdução

A importância da educação da população, nas esferas econômica e social, levou formuladores de políticas públicas ao redor do mundo a terem esta como uma de suas principais agendas. No Brasil, políticas públicas de expansão de escolas e programas como o Bolsa Família, contribuíram para o aumento da proporção de crianças em salas de aula. (BARROS, 2002)

Porém, a expansão de vagas e presença nas escolas não foi acompanhada de uma melhora relativa no desempenho brasileiro, em exames internacionais, em esferas como Leitura, Matemática e Ciência, como se pode observar nos resultados do PISA (INEP, 2017).

Há diversas hipóteses para o baixo desempenho do aluno brasileiro, dentre elas, a falta de atenção dada para a criança na primeira infância, período no qual ela deveria desenvolver habilidades não cognitivas, que ajudam no aprendizado e desempenho futuro. Além disso, famílias com baixa renda muitas vezes não conseguem dar o apoio necessário para o desenvolvimento das crianças, desde o estímulo para o aprendizado, até condições básicas de alimentação e saneamento básico. (O COMEÇO DA VIDA)

Políticas públicas voltadas para a primeira infância, como a criação de creches, parecem ser benéficas segundo a literatura internacional, em especial segundo os autores Berlinski, Galiani e Gertler (2009). Porém, os efeitos no Brasil são mais ambíguos, conforme apontado por Santos (2015), no qual pré escola tem efeito positivo no desenvolvimento das crianças, mas creches não tem efeito. O presente projeto visa contribuir para o entendimento do fenômeno a respeito dos efeitos da política de criação de vagas em educação infantil. Trata-se de uma pesquisa empírica aplicada, de abordagem quantitativa e seus resultados são voltados para o norteamo de políticas públicas.

A partir dos dados do Censo Escolar, do Inep, e Cadastro Único, do Ministério de Desenvolvimento Social e Caixa Econômica Federal, deseja-se analisar se a introdução das creches contribui para melhora do desempenho

educacional dos alunos, em aspectos como aumento de aprovações (no nível escolas) e aumento de frequência escolar geral e específico de mulheres adolescentes (nível indivíduos).

A partir da literatura, parte-se da hipótese de que há causalidade positiva entre ir para creche e desempenho escolar, de forma que a criança será menos propensa a atrasos escolares.

Ademais, a presente pesquisa não deseja avaliar a qualidade do programa Nova Semente com relação a outras creches, apenas o efeito da expansão das vagas. Por conseguinte, há uma hipótese implícita de homogeneidade na qualidade das vagas.

Portanto, a pesquisa tem como objetivo geral investigar os impactos educacionais de uma expansão repentina de vagas em educação infantil, em especial, do caso do programa Nova Semente. Ao entender melhor estes impactos, deseja-se contribuir para a literatura sobre os benefícios de gastos públicos com creches e para literatura de avaliação do Nova Semente, como programa de criação de vagas em educação infantil.

As vantagens e desvantagens, do uso de cada base de dados e cada teste estatístico, estão explícitos nas seções de base de dados, metodologia e discussão de cada um dos exercícios realizados.

Os resultados da pesquisa apontam para alguns efeitos positivos do programa no que se refere ao desempenho escolar, conforme os exercícios com o Censo Escolar. A partir dos dados do Cadastro Único também há evidência da melhora na frequência escolar dos alunos.

## **2. Revisão da Literatura**

### **a. Expansão de vagas em educação infantil**

O trabalho de Thompson, Nelson (2001) resume literatura neurocientífica internacional acerca das fases do desenvolvimento cerebral e seus impactos no aprendizado, destacando a importância da primeira infância, de 0 a 5 anos, no aprendizado e desenvolvimento de habilidades não cognitivas. Ademais, conforme apontado por Hart e Risley (1995), a partir dos 18 meses inicia-se uma disparidade de vocabulário entre crianças de pais com ensino superior com relação a pais dependentes de assistência social, chegando a um vocabulário três vezes maior do primeiro, com relação ao segundo.

Apesar da evidência sobre a importância da primeira infância no desenvolvimento da criança, neste período as crianças estão sob o cuidado dos pais, o que dificulta o trabalho do governante. Uma política que alcança esse objetivo é o investimento em creches e pré-escolas.

Há ampla evidência internacional de que creches melhoram o desempenho de crianças, vale destacar Barnett (2011) apontando um ganho médio no desempenho escolar de crianças que passaram pelo ensino infantil. Além de Barnett, também se destaca Currie (2001), a qual aponta a importância do uso de medidas não cognitivas (habilidades socioemocionais) para avaliar impactos da educação infantil. Porém, Havnes e Mostad (2015) mostram que em nem todos os países ocorrem esse efeito positivo de creches, como apontado pelos autores na Noruega.

Resultados do efeito de creches no Brasil são ainda mais ambíguos, conforme apontado por Santos (2015), pré-escolas por outro lado continuam tendo impacto predominantemente positivo. Porém, no mesmo estudo, Santos destaca que crianças mais vulneráveis são as que mais se beneficiam da passagem pela pré-escola, com bases em resultados para Sertãozinho (SP).

Outros estudos no Brasil como de Barros e Mendonça (1999) sugerem que indivíduos que passaram pela pré-escola ganham em média 0,6 anos a mais de escolaridade e, entre os do sexo masculino, alcançam salários 6% maiores. Em Curi e Menezes-Filho (2009), a entrada na pré-escola implica cerca de 5 pontos percentuais (pp) a mais de chance de ter completado a

primeira fase do ensino fundamental quando adultos, 20 pp a mais de concluir o ensino fundamental, 25 pp a mais de terminar o ensino médio e 5 pp a mais de se formar no ensino superior. Porém para os autores, os resultados com creches foram estatisticamente inconclusivos.

Outros acadêmicos avaliaram especificamente o impacto de grandes expansões de vagas em educação infantil, conforme feito no presente projeto, entre eles é importante mencionar Duflo (2001) que analisa a criação de 61 mil escolas primárias na Indonésia e encontra que crianças de 2 a 6 anos em 1974 receberam de 0,12 a 0,19 anos a mais de estudos que crianças que não frequentaram essas escolas, o que, geraria um retorno econômico de 6,8 a 10,6%.

O trabalho de Esther Duflo foi utilizado como base para esta linha de pesquisa, cujos principais autores são Berlinski e Galiani. Em Berlinski e Galiani (2007) é utilizando o método de diferenças em diferenças e o censo familiar argentino (EPH), entre as principais conclusões, destaca-se que o programa de criação de escolas primárias na Argentina aumentou em 7,6% a proporção de crianças na educação infantil entre 1991 e 2001. Em trabalho posterior Berlinski, Galiani e Gertler (2009) destacam que esta expansão em educação infantil aumenta as notas de alunos de terceira série em 8% da média ou 23% do desvio padrão da distribuição das notas nos testes. Além disso, Para Berlinski, Galiani e Gertler, há um impacto positivo no comportamento com relação ao autocontrole desses alunos, medido pela observação de seus professores em aspectos como atenção, esforço, participação em sala e disciplina.

Em outro estudo, Berlinski, Galiani e Manacorda (2007) utilizaram o censo familiar do Uruguai (ECH), no qual há informação sobre ter frequentado pré-escola no passado, e os autores chegaram às conclusões de que pequenos ganhos da frequência em pré-escolas quando crianças (tratadas) são aumentadas conforme elas vão crescendo, de forma que até os 15 anos crianças tratadas acumulam 0,8 anos extra de educação e são 27% mais prováveis de ainda estarem nas escolas do que as crianças não tratadas, isto é, há uma redução na evasão. Além disso, os autores observam que o efeito é heterogêneo: crianças cuja mãe tem menos educação são as que mais se beneficiam do programa.

## **b. Nova Semente**

O programa Nova Semente implementado em Petrolina (PE) veio ajudar a preencher a lacuna da falta de vagas em creches em Petrolina, seu foco é justamente a primeira infância. O programa conta com parcerias entre governo, organizações não governamentais e sociedade civil organizada e tem dois eixos: O Pré-natal: atendendo mulheres neste período, parto e atenção ao bebê, incluindo uma bolsa aleitamento de R\$100, se necessário; o segundo eixo trata-se da educação infantil em si, acolhendo criança dos seis meses até o fim da pré-escola, no local da creche as crianças são fornecidas com atividades e alimentação no período da manhã e da tarde. Segundo seu edital de criação, as turmas seriam compostas por 10 alunos. (GRANJA, 2013)

Conforme apontado por Granja, entre os objetivos do programa destacam-se o aumento de índices educacionais; melhoria de índices sobre nutrição infantil; redução de mortes pós-parto entre outros. O programa teve sua primeira unidade inaugurada em outubro de 2010 e até dezembro de 2012 85 creches foram inauguradas (5100 vagas), um pouco abaixo da sua meta inicial de 100 creches (6000 vagas). O programa atende crianças de 0 a 5 anos e, chegou a ter mais de 100 unidades em 2016. (NOVA SEMENTE, 2016)

Atualmente, com a entrada de Miguel Coelho na prefeitura de Petrolina, houve fechamento de algumas unidades e o programa vem passando por uma série de reformas visando sua ampliação e redução de custos, conforme aponta Brito (2017). Ademais, o programa está sendo auditado pela Controladoria Geral da União e está passando por uma troca de gestão devido ao seu custo de manutenção.

### c. Métodos Estatísticos

#### i. Modelo de escolha binária

Conforme apresentado em Davidson e MacKinnon (2004), modelos cuja variável dependente é uma escolha binária de valores zero e um, muitas vezes provenientes de uma variável qualitativa, devem ser tratados diferentemente. O método de Mínimos Quadrados Ordinários aplicados a variáveis dependentes binárias, chamado de Modelo de Probabilidade Linear, geram dois principais problemas: heterocedasticidade e erro na estimativa das extremidades, dado que a reta linear não estará limitada a zero e um. Para corrigir essas limitações usa-se funções de distribuição acumuladas: para uma variável aleatória  $x$  variando no tempo  $t$  e uma matriz de coeficientes  $\beta$ , a função normal padrão que dá origem ao modelo Probit é dada na Equação 1 abaixo:

$$G(Xt\beta) = \int_{-\infty}^{Xt\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (1)$$

Para a função de distribuição acumulada logística, do modelo Logit, utiliza-se:

$$G(Xt\beta) = \frac{e^{Xt\beta}}{1 + e^{Xt\beta}} \quad (2)$$

Como a variável dependente ( $y$ ) deve ter valor zero ou um, assume-se que segue uma distribuição de Bernoulli e, conseqüentemente, pode ser estimada pela função de Máxima Verossimilhança com  $n$  observações:

$$L(y, \beta) = \sum_{t=1}^n (y_t \log G(Xt\beta) + (1 - y_t) \log (1 - G(Xt\beta))) \quad (3)$$

A solução do modelo se dá encontrando os estimadores  $\beta$  que maximizam a função de verossimilhança. Porém esses coeficientes não refletem o efeito da variável explicativa na dependente binária. Para tal, é necessário calcular os efeitos marginais a partir de cada função de distribuição acumulada  $G(\cdot)$ :

$$\frac{\partial P(y=1|X)}{\partial x} = G(X_i\beta)\beta \quad (4)$$

## ii. Pareamento e Diferenças em Diferenças

Avaliações de programas podem ser realizadas de forma qualitativa, quantitativa e misturando essas abordagens. Em avaliações quantitativas, foco da presente pesquisa, o objetivo maior é saber o que aconteceria com os mesmos indivíduos tratados por determinado programa caso não tivessem passado por esse tratamento, assim a diferença em determinado indicador escolhido para avaliar o programa seria o efeito exato do programa, o que é chamado Efeito Médio do Tratamento ou ATE, do inglês, Average Treatment Effect. (MENEZES 2012)

Porém, na prática, para os casos de programas sociais, envolvendo pessoas, costuma ser impossível saber o que aconteceria com os mesmos indivíduos se não tivessem sido tratados pelo programa, diversas são as ferramentas utilizadas para contornar essa dificuldade, a seguir serão explicados duas delas: o Pareamento e o método de Diferenças em Diferenças.

Uma forma de contornar a limitação descrita acima é encontrar indivíduos muito parecidos com os tratados, exceto pelo fato de não terem sido tratados, isto é, os controles. O método do Pareamento tem como objetivo encontrar bons controles para os tratados, em especial, o método de pareamento por escores de propensão utiliza variáveis relevantes à estimação dos indivíduos tratados para encontrar pares com essas características relevantes mais próximas nos indivíduos no controle.

Na prática, é estimado um modelo Logit, conforme explicado na seção anterior, para determinar o quanto cada variável independente relevante explica da variação nos indivíduos tratados. A partir desse modelo, aos indivíduos não tratados (controles) são atribuídos pesos conforme a semelhança da variação das variáveis independentes com os tratados.

Já a ideia do modelo de Diferenças em Diferenças, é contornar a limitação de estimativas ingênuas, conforme apontado em Menezes (2012).



Isto é, apenas observando a diferença entre indivíduos tratados antes e depois do tratamento, ou apenas a diferença entre tratado e controle depois do tratamento, o pesquisador estaria deixando de considerar mudanças entre os grupos no período. Assim, a proposta do modelo é medir a variação entre tratados e controles antes e depois do tratamento, conforme a Equação 5:

$$(\text{Tratado depois da intervenção} - \text{Tratado antes}) - (\text{Controle depois} - \text{Controle antes}) \quad (5)$$

Econometricamente, é necessário criar uma variável dummy  $T = 1$  para indivíduos tratados e  $T = 0$  para o controle, também se cria variável  $t = 1$  para o período depois do programa e  $t = 0$  antes do programa. Assim, o efeito do tratamento ( $\beta_1$ ) será dado pela interação das variáveis  $T$  e  $t$ , conforme a Equação 6:

$$Y = c + \beta_1 Tt + \beta_2 T + \beta_3 t + u \quad (6)$$

Onde  $Y$  é a variável dependente,  $c$  é uma constante e  $u$  o termo de erro estimado.

Os métodos de Pareamento e Diferenças em Diferenças podem ser usados separadamente ou em conjunto, mas devem respeitar duas importantes hipóteses: Suporte Comum e Tendência Comum Paralela. Na primeira é necessário que nenhuma combinação das variáveis explicativas defina com 100% de probabilidade o tratamento, isto é, deve existir variabilidade entre tratados e controles a partir das variáveis independentes. Na segunda hipótese, é necessário que a trajetória da média de indivíduos tratados antes da ocorrência do programa seja paralela à trajetória do controle para o mesmo período, isso ajuda a garantir que a única mudança relevante no período foi justamente o programa. (SCORZAFAVE, L. G.; SOARES, M. K.; DORIGAN, T. A., 2015)

### iii. Juazeiro como controle de Petrolina

Com o objetivo de analisar a expansão de vagas em educação infantil proveniente do programa Nova Semente, ocorrido em Petrolina (PE), deseja-se encontrar um bom grupo de controle (contrafatual): indivíduos bastante similares que não foram atendidos pelo programa.

A cidade Juazeiro (BA) será um bom controle, pois tem tamanho e renda próximas a Petrolina, segundo o IBGE Cidades, além disso, são cidades vizinhas, separadas apenas pelo Rio São Francisco, o que permite também atenuar diferenças geográficas relevantes (conforme a Figura 1). Ademais por não se tratar apenas de cidades, mas estados diferentes, o programa Nova Semente não pode ser utilizado pela população de Juazeiro, de fato, no exercício utilizando o Cadastro Único isso será evidenciado pelos dados.

As Tabela 1 e Tabela 2 abaixo mostram as similaridades entre as cidades no âmbito geral e educacional respectivamente. Vale mencionar também que em 2012 os prefeitos de ambas as cidades foram reeleitos, segundo o site do TSE, o que ajuda a reduzir eventuais problemas de viés político no período em análise. Além disso, programas federais relevantes no período, como o PAC2, tiveram projetos realizados em ambas as cidades.

**Tabela 1: Comparação Geral**

<b>Indicador\Cidade</b>	<b>Juazeiro</b>	<b>Petrolina</b>
População estimada 2016	220253	337683
População 2010	197965	293962
Área da unidade territorial 2015 (km <sup>2</sup> )	6500,691	4561,872
Índice de Desenvolvimento Humano Municipal (2010)	0,677	0,697
PIB per capita a preços correntes (2013)	12065,62	15334,27

Fonte: IBGE Cidades

**Tabela 2: Comparação em Educação**

<b>Indicador</b>	<b>2015</b>		<b>2009</b>	
	<b>Juazeiro</b>	<b>Petrolina</b>	<b>Juazeiro</b>	<b>Petrolina</b>
Matrícula - Ensino fundamental	33736	55138	37739	56163
Matrícula - Ensino pré-escolar	5828	10597	5606	8876
Matrícula - Ensino pré-escolar - escola pública municipal	4665	5609	3371	5628

Fonte: IBGE Cidades

Observa-se também, que será considerado tratado quem reside em Petrolina, pois apenas esses indivíduos são afetados direta ou indiretamente pelo programa, conforme será evidenciado na Seção da Base de dados do Cadastro Único (Tabela 7).

Além disso, os efeitos diretos são essencialmente o uso das novas vagas por parte da população e os efeitos indiretos são a possibilidade de usar as novas vagas se assim desejam, isto é, independente de colocarem os filhos nas creches do Nova Semente ou não, o grupo é considerado tratado pois tem essa possibilidade, o que ocorre com toda a população de Petrolina, por isso são todos tratados.

Por outro lado, pela Figura 2, na seção seguinte, observa-se que, apesar de partir de um patamar mais baixo do número de creches, Juazeiro também aproximadamente triplicou o número de creches e vagas infantil, conforme feito pelo programa Nova Semente. Isso não só enfraquece o controle, como pode apresentar efeito contrário, se o aumento de creches está sujeito a rendimentos marginais decrescentes, isto é, onde há menos creches a construção de novas implica maior impacto em indicadores educacionais. Nas variáveis dependentes, porém, de fato parece haver bom paralelismo pré tratamento, conforme a Figura 3.

**Figura 1: Petrolina e Juazeiro**



Fonte: Skyscrapercity

### 3. Censo Escolar – Escolas

#### a. Base de dados - Censo Escolar

As bases utilizadas para este exercício serão o Censo Escolar Microdados (2016) e Censo Escolar Indicadores Educacionais (2016), ambos do INEP para os anos de 2008 a 2015 no nível “Escolas”. A partir da união dessas bases, pelo software *Stata* (via *merger*), cria-se a base “Censo Escolar”. Em seguida, pelo mesmo software realiza-se o balanceamento do painel e os demais exercícios.

A Tabela 3 abaixo apresenta e explica as variáveis utilizadas nesse exercício. Encontradas no Apêndice, as Tabela A 1 e Tabela A 2 apresentam a frequência de dados para cada dependência administrativa e ano, respectivamente; e a Tabela A 3 apresenta as estatísticas descritivas. Aqui vale destacar que, com o tratamento, espera-se aumento das variáveis de aprovação e redução das variáveis de abandono.

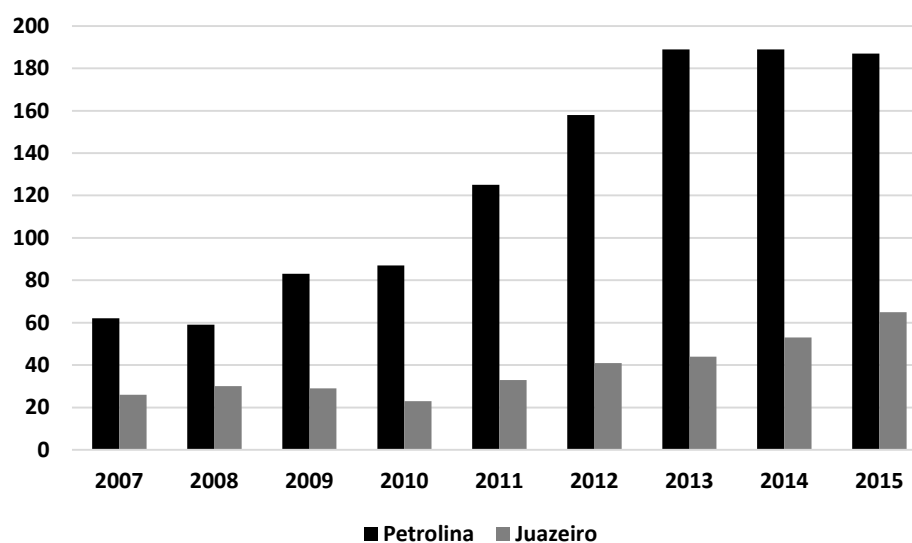
**Tabela 3: Descrição das Variáveis do Censo Escolar**

<b>Nome</b>	<b>Descrição</b>	<b>Tipo</b>
aprov_1	Taxa de Aprovação no 1º ano Fundamental	Percentual
aprov_2	Taxa de Aprovação no 2º ano Fundamental	Percentual
aband_1	Taxa de Abandono no 1º ano Fundamental	Percentual
aband_2	Taxa de Abandono no 2º ano Fundamental	Percentual
id_agua_filtrada	Água Filtrada consumida pelos alunos	0: Não; 1: Sim
id_agua_inexistente	Abastecimento de água - Inexistente	0: Não; 1: Sim
id_esgoto_rede_publica	Esgoto sanitário - Rede pública	0: Não; 1: Sim
id_lixo_coleta_periodica	Destinação do lixo - Coleta periódica	0: Não; 1: Sim
id_laboratorio_informatica	Laboratório Informática existente na escola	0: Não; 1: Sim
id_biblioteca	Biblioteca existente na escola	0: Não; 1: Sim
id_parque_infantil	Parque Infantil existente na escola	0: Não; 1: Sim
num_salas_existentes	Número de salas de aula existentes na escola	Número
num_comp_alunos	Número de computadores para alunos	Número
id_internet	Acesso à Internet	0: Não; 1: Sim
num_funcionarios	Total de funcionários da escola	Número
id_alimentacao	Alimentação escolar oferecida para os alunos	0: Não; 1: Sim
id_mod_ens_regular	Modalidade - Ensino Regular	0: Não; 1: Sim
id_reg_infantil_creche	Ensino Regular - Educação Infantil - Creche	0: Não; 1: Sim
id_reg_infantil_preescola	Ensino Regular - Educação Infantil - Pré-escola	0: Não; 1: Sim

Fonte: Censo Escolar

A Figura 2 mostra a expansão de creches conforme o período de captação dos dados do Censo Escolar, vale notar que a primeira expansão ocorreu em 2011, seguido por uma ainda maior em 2012, a qual se completou em 2013. Segundo o site do MEC, o período ideal para entrada na pré-escola é quatro anos e, no primeiro ano, seis anos. Além disso, o programa Nova Semente abrange crianças de 0 a 5 anos. Depreende-se que as primeiras crianças que passaram pelo Nova Semente ingressaram no primeiro ano em 2012, ademais, maior volume de crianças a se beneficiarem direta ou indiretamente pelo programa ingressaram no primeiro ano do ensino fundamental em 2014.

**Figura 2: Expansão de Creches e Pré-escolas pelo Censo Escolar**



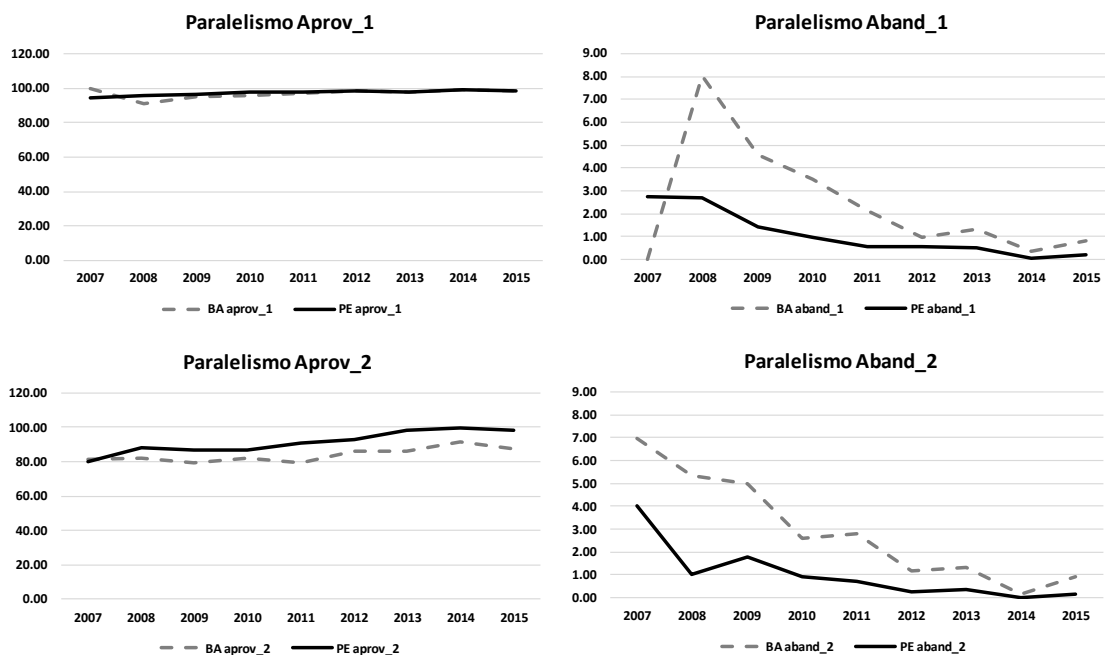
Fonte: Censo Escolar

A Figura 3 apresenta o parelismo pré pareamento, observa-se que parece haver bastante similaridade entre tratado e controle e, mesmo antes do pareamento destaca-se o diferencial aparente na variável de aprovação no segundo ano.

A Figura 4 apresenta o tempo de exposição ao programa, em anos, de cada ano escolar. Observa-se que apesar do programa ter sido criado no fim de 2010, os primeiros alunos de ensino fundamental expostos são a turma de primeiro ano em 2012, sendo que, em 2015 a exposição da turma de primeiro

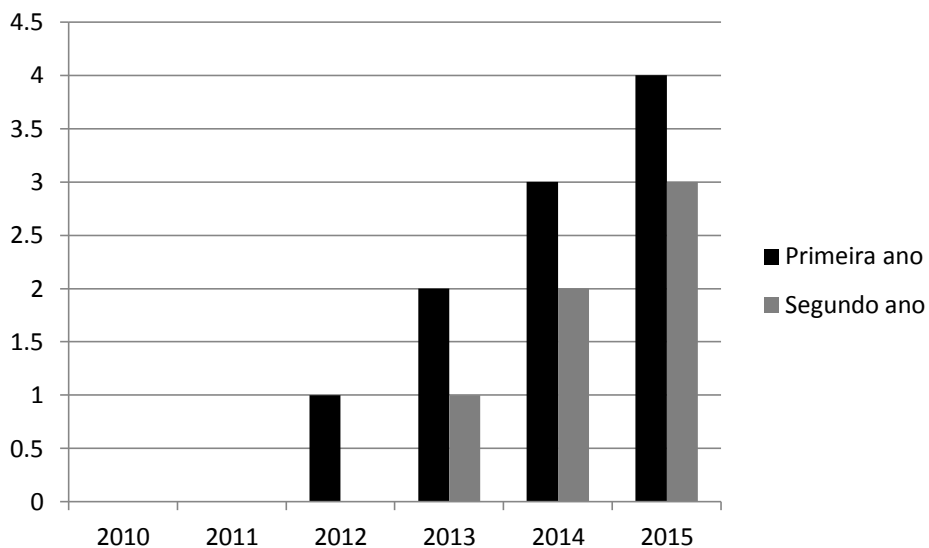
ano é de 4 anos do programa e da turma de segundo ano de 3 anos do programa.

**Figura 3: Paralelismo pré pareamento de variáveis dependentes**



Fonte: Censo Escolar

**Figura 4: Anos de exposição ao programa por turmas de ensino fundamental**



Fonte: Elaboração Própria

## **b. Metodologia e limitações**

O principal objetivo do projeto é avaliar o impacto da expansão de creches no desempenho de alunos, idealmente seria interessante observar, exatamente o que ocorreria com as crianças, caso não fossem tratadas pela criação do programa Nova Semente. Entretanto, isto não é observável, alternativamente, a literatura avalia o impacto do tratamento pelo Efeito Médio Sobre os Tratados ou ATT (*Average Treatment on Treated*), conforme apontado em Menezes Filho (2012).

Pretende-se utilizar um modelo de Diferenças em Diferenças e um pareamento entre escolas no grupo dos tratados (Petrolina) e controles (Juazeiro). Este modelo permite controlar mudanças ao longo do tempo com a expansão do programa (coortes anuais) e controla semelhanças constantes ao longo do tempo pela diferença com o par em Juazeiro. Ademais, serão realizados dois exercícios diferentes de pareamento: o primeiro com as próprias variáveis do Censo Escolar e o segundo com variáveis da Prova Brasil.

Com relação às limitações, ainda segundo Chimeli e Soares (2011) e Duflo, Bertrand e Mullainathan (2002), o modelo de diferenças em diferenças assume as hipóteses de Suporte Comum, ou seja, deve haver variabilidade das variáveis para ambos tratados e controles e Paralelismo: a variável dependente deve seguir a mesma tendência para o grupo tratado e controle antes do período de tratamento, o que é mais difícil de garantir. Assim, surgem três das principais limitações desse tipo de estimação:

I) A composição de tratado e controle pode mudar como função do tratamento – No caso da presente pesquisa este problema equivale ao problema de contaminação: até que ponto habitantes de Juazeiro conseguem se beneficiar das creches em Petrolina, tanto diretamente, isto é, crianças de Juazeiro que conseguem vagas em creches de Petrolina, quanto crianças que fizeram creche em uma cidade e foram estudar, futuramente, em escolas da outra cidade.

II) É difícil garantir que a criação de creches foi a única variável relevante para educação que mudou ao longo do tempo em Petrolina e Juazeiro.



III) O período do tempo para absorção do efeito também é difícil de escolher, períodos muito curtos tiveram poucas crianças influenciadas pelo programa e em períodos mais longos outros fatores relevantes contribuíram para a variação.

Outra dificuldade da presente pesquisa é o cruzamento dos dados das crianças que frequentaram creches com seus respectivos desempenhos, por conta disso foi escolhido utilizar o nível escolar ao invés de matrícula o que aumenta possível viés por variável omitida, mas também permite melhor captação de efeitos indiretos do programa.

No caso do presente exercício, a margem para viés de composição é relativamente grande, o que será corrigido pelo próximo exercício utilizando matrículas.

### c. Resultados - Pareamento com Censo Escolar

O pareamento entre escolas de Petrolina (tratadas) e de Juazeiro (controle) é feito por Escores de Propensão, os quais são calculados a partir de um modelo de Probit conforme a Tabela A 4. As variáveis explicativas do Probit foram escolhidas conforme sua significância estatística e bons resultados de Suporte Comum: 81 escolas não tratadas e 77 tratadas. Ficaram de fora do Suporte Comum 19 escolas do grupo de tratamento, essa proporção pode ser observada na Tabela A 5. Ademais, o ano escolhido para o pareamento foi 2010.

Em seguida, realiza-se o teste do pareamento para avaliar se houve melhora na comparação de escolas, de forma que a diferença entre as médias de tratado e controle antes do tratamento sejam mais próximas de zero, conforme a Tabela A 6 observa-se melhora nessa composição, em especial, para as variáveis *id\_agua\_filtrada*; *id\_reg\_infantil\_creche*; *id\_biblioteca*; *id lixo\_coleta\_periodica*; *num\_salas\_existentes*; *num\_comp\_alunos*, onde *N* mostra a diferença antes do pareamento e *P* mostra a diferença após o pareamento.

A partir do pareamento, realiza-se uma diferença das diferenças entre as médias de tratado e controle e antes e depois, isto é, o método Diferenças em Diferenças, conforme o procedimento descrito anteriormente. Para tal, a partir das variáveis apresentadas criam-se dummies de tempo (*t* para cada ano *i*) multiplicadas pela dummy *T* que é igual a 1 se é tratamento e 0 se for controle. O produto destas variáveis são as *li*, onde *i* indica o ano, ademais, também pode-se criar uma variável *l* geral para o efeito médio do tratamento, a partir da interação de *T* com *t*, conforme descrito anteriormente. De fato, as equações estimadas por mínimos quadrados ordinários são as seguintes:

$$Y = c + \beta_1 Tt + \beta_2 T + \beta_3 t + u \quad (6)$$

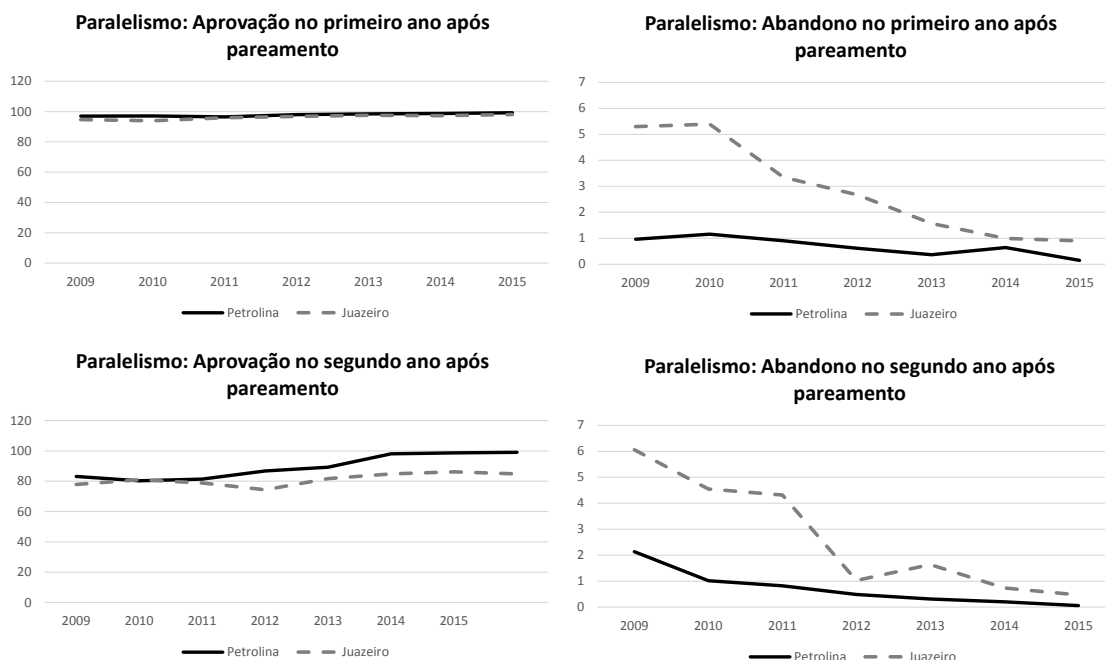
$$Y = k + \alpha_1 l + \alpha_2 T + \alpha_3 t + u \quad (7)$$

Onde  $Y$  é a variável dependente (a qual, devido ao paralelismo, explicado abaixo, foi usada aprovação no segundo ano do ensino fundamental);  $c$  e  $k$  são constantes;  $u$  e  $e$  são termos de erro.

Conforme mencionado na seção de Limitações, o método de Diferenças em Diferenças pressupõe não apenas o Suporte Comum, mas também o Paralelismo. No tratamento dos dados houve balanceamento do painel, momento no qual também se removeu o ano de 2007 por haver muitos *outliers*. Na Figura 5 a seguir tem-se o paralelismo após balanceamento e pareamento do grupo de tratados com controles.

Com relação ao período de tratamento, pode-se considerar o período de 2010 (criação do programa e abertura da primeira creche) ou 2012 em diante conforme a expansão do programa.

**Figura 5: Paralelismo Pós Pareamento do Censo Escolar**



Fonte: Censo Escolar, Elaboração própria

Observa-se que apenas a variável *aprov\_2*, que mede aprovação no segundo ano do ensino fundamental, realmente obedece a um paralelismo completo (antes do pareamento) e ainda começa a crescer depois no período de tratamento, após ser realizado pareamento (Figura 5). *Aprov\_1*, aprovação

no primeiro ano, apresenta paralelismo, mas não apresenta efeito do programa, conforme será visto em seguida.

Além disso, pelos gráficos, fica claro que as variáveis de abandono escolar (*aband*) não são boas variáveis dependentes para observar o efeito das creches, já que Juazeiro apresentava um nível muito elevado de abandono e este problema foi fortemente mitigado em período próximo à introdução do programa Nova Semente. Essa limitação na variável de abandono é bastante comum devido ao seu “efeito de piso”, isto é, as taxas já são bastante altas, o que dificulta reduzi-las ainda mais com determinado programa.

Conseqüentemente, *aprov\_2* apresenta melhores resultados no presente exercício. A Tabela 4 mostra o resultado do efeito médio da introdução do Programa Nova Semente, onde *t* é a variável que indica o período antes e depois do tratamento, no caso 2010 – criação da primeira creche do programa; *T* indica se a escola pertence ao grupo tratado (Petrolina) ou controle (Juazeiro) e *I* é a interação das variáveis (nos dá o efeito do programa). A tabela mostra que todas as variáveis foram estatisticamente significantes e, por este método de Diferenças em Diferenças com Pareamento por Escore de Propensão, a introdução do programa levou a um aumento de 6,9 p.p. no nível de aprovações no segundo ano do ensino fundamental.

**Tabela 4: Efeito Médio do Nova Semente na Aprovação do Segundo Ano pelo Censo Escolar**

Variável	Coef.	Desv. Pad.	t	P>t	[95% IC]	
t	5.747701	1.479882	3.88	0	2.8428	8.6526
T	6.457386	1.050135	6.15	0	4.3961	8.5187
I	6.914853	2.128901	3.25	0.001	2.736	11.094
_cons	79.77271	0.7361945	108.36	0	78.328	81.218

Fonte: Censo Escolar, Elaboração própria

Realizando o mesmo exercício de forma a capturar o efeito das creches para cada ano (**Error! Reference source not found.**), gerando *dummies* de ano e interação. Nos resultados observa-se um efeito similar. Porém, é necessário notar que o período no qual se inicia esse efeito é 2011, isto é, período de ingresso dos primeiros alunos do programa e, logo, estes alunos não são os mesmos obtendo melhores níveis

de aprovação no segundo ano do ensino fundamental, o que indica a presença de um efeito indireto do programa.

Vale mencionar que o mesmo exercício foi realizado para as outras variáveis dependentes e, conforme esperado, não houve resultado estatisticamente relevante, além disso, no caso da variável de abandono o efeito foi contrário decorrente da grande melhoria desta variável em Juazeiro.

Esses últimos resultados apontam que, apesar do efeito positivo na variável do segundo ano do ensino fundamental, os efeitos gerais do programa no desempenho escolar não são muito claros, sendo necessários outros exercícios.

#### d. Discussão

Os resultados no exercício utilizando Censo Escolar, inclusive no pareamento, não foram homogêneos quanto ao efeito do programa. Das possíveis variáveis indicadoras de bom desempenho dos alunos: abandono no primeiro e segundo ano e aprovações no primeiro e segundo ano, apenas a variável de aprovação no segundo ano apresentou efeito positivo e estatisticamente significativo.

No primeiro exercício, o efeito da escola estar em Petrolina (cidade do tratamento), com relação à Juazeiro (controle) após o tratamento foi de 6,9 p.p. na proporção de alunos aprovados no segundo ano em 2012, período no qual os primeiros alunos beneficiados pelo programa Nova Semente, chegaram a essa série. Porém, aprovação no primeiro ano, bem como abandono não apresentaram significância estatística – resultado esperado de abandono, devido a seu “efeito de piso”, mas não da aprovação no primeiro ano.

Corroborando o resultado mencionado acima, a partir do uso de controles sintéticos, tem-se efeitos positivos do tratamento em Petrolina nas variáveis de aprovação no segundo ano apenas para tratamento em 2012, bem como aprovação de primeiro e segundo ano e abandono de primeiro ano para tratamento em 2014, isto é, quando mais alunos ingressaram nas creches e ficam mais expostas ao programa.

Esse resultado está em linha com a literatura sobre o tema, em especial os resultados encontrados por Barros e Mendonça (1999) e trabalhos de Berlinski e Galliani, indicando efeitos positivos do programa Nova Semente.

O efeito negativo e pouco significativo na variável de abandono advém da elevada taxa apresentada por Juazeiro anteriormente, a qual foi reduzida de forma independente ao Nova Semente e pareceu provocar efeito contrário do programa. Além do “efeito de piso” de haver já pouco abandono em Petrolina.

De fato, esta metodologia sofre de dois problemas. O primeiro problema é o fraco paralelismo, sendo essa hipótese necessária para esse tipo de análise. O segundo problema decorre de um viés de composição ao se utilizar as escolas do Censo Escolar de forma agregada: pode ter havido mudanças

relevantes nas composições dos indivíduos em cada escola atribuída como tratada e como controle, o que invalidaria o exercício. Por conta dessas limitações foram realizados os exercícios a seguir.

Um terceiro problema advém do pouco tempo que o programa teve para fazer efeito na amostra, os alunos de segundo e primeiro ano fizeram, no máximo, dois anos de Nova Semente e ainda foram os anos iniciais do programa. Por fim, outras variáveis indicadoras de desempenho educacional, indisponíveis, mas interessantes, poderiam ser melhores medidas: como notas de provas, por exemplo.

## 4. Cadastro Único

### a. Base de dados

O Cadastro Único para Programas Sociais do Governo Federal (Cadastro Único) é um instrumento em parceria do Ministério do Desenvolvimento Social com a Caixa Econômica Federal, cujo objetivo é identificar e caracterizar as famílias de baixa renda no país. O instrumento é também utilizado para seleção de beneficiários em diversos programas sociais como o Bolsa Família.

Por um lado, a base do Cadastro Único é formada essencialmente por famílias de baixa renda, como se pode observar pela Tabela A 7 no Apêndice, por outro lado, possui grande riqueza de variáveis desses indivíduos e, conforme Santos (2015), há maior evidência de efeitos de creches em grupos de renda mais baixa.

Para o presente exercício foram utilizadas as variáveis da Tabela 5 para as cidades de Petrolina e Juazeiro e filtrando apenas para indivíduos entre 0 e 17 anos, determinado pela diferença entre a data de nascimento e a data da última atualização.

Uma limitação da base de dados é seu início em 2011, dado que o programa Nova Semente também começou em 2011 não é possível obter um período base sem tratamento nenhum, apenas com tratamento reduzido. Ademais, não há dados sobre aprovação escolar, apenas frequência, o que será analisado, de diferentes formas, nesse exercício. Além disso, não há cruzamento de dados entre os jovens indivíduos, nem sobre suas famílias. Portanto, esse terceiro exercício é mais complementar do que substituto dos demais na presente pesquisa.



**Tabela 5: Variáveis utilizadas do Cadastro Único**

<b>Variável</b>	<b>Descrição</b>	<b>Tipo</b>
Ano	Ano da pesquisa	Num
cd_ibge	Código Ibge do município	Num
marc_pbf	Indicador de beneficiário do Bolsa Família	Num
dta_atual_memb	Data da última atualização dos dados	Num
ind_trabalho_infantil_pessoa	Indicador se há trabalho infantil na família	Num
num_nis_pessoa_atual	Número de identificação social	Num
codsexo_pessoa	Indicador do sexo da pessoa	Num
dta_nasc_pessoa	Data de nascimento	Num
codraca_cor_pessoa	Indicador de cor de pele da pessoa	Num
cod_sabe_ler_escrever_memb	Indicador de literacidade	Num
ind_frequenta_escola_memb	Indicador de frequentar a escola	Num
nom_escola_memb	Nome da escola	Num

Fonte: Cadastro Único

A variável de acompanhamento dos indivíduos entre eles (variação natural) é a *nis*, Número de Identificação Social, a variação longitudinal é capturada pelos anos. Inicialmente, faz-se necessário balancear o painel, obtendo a Tabela 6 a seguir. Observa-se que serão utilizadas 68766 observações em comum por ano e conforme a Tabela 5 foram construídas novas variáveis como idade do indivíduo (*idade*), se frequenta alguma creche do programa Nova Semente (*ns*), se reside em Petrolina (*P*), e algumas renomeações. As estatísticas descritivas das variáveis utilizadas estão na Tabela A 8 no Apêndice.

**Tabela 6: Balanceamento do Painel do Cadastro Único**

<b>Ano</b>	<b>Frequência</b>	<b>Percentual</b>	<b>Acumulado</b>
2011	68766	20	20
2012	68766	20	40
2013	68766	20	60
2014	68766	20	80
2015	68766	20	100
Total	343830	100	

Fonte: Cadastro Único

A partir dessa base de dados foi possível testar se ocorre contaminação, isso é, indivíduos supostamente no Controle (Juazeiro) que utilizaram o programa Nova Semente. Pela Tabela 7 a seguir, observa-se que a quantidade de indivíduos nessa situação é muito pequena dentro da amostra, considerando que a base possui dados de 2011 a 2015, os 9 indivíduos se mudaram entre Petrolina e Juazeiro, ou seja, de fato apenas moradores de Petrolina se beneficiam do programa. Lembrando que são considerados tratados todos os indivíduos em Petrolina pois sofrem direta ou indiretamente os efeitos do Programa Nova Semente.

**Tabela 7: Contaminação do Programa Nova Semente**

<b>Escola</b>	<b>Petrolina</b>	<b>Juazeiro</b>	<b>Total</b>
Outro	192652	149570	342222
Nova Semente	1599	9	1608
Total	194251	149579	343830

Fonte: Cadastro Único

## b. Metodologia

No intuito de entender os efeitos do programa Nova Semente na frequência escolar, análises estatísticas no mesmo período do tempo (*cross-section*) podem ocasionar um problema de endogeneidade ou omissão de variável relevante.

Uma forma de contornar essa dificuldade é utilizando dados em painel, no qual é possível acompanhar os indivíduos ao longo do tempo. Ademais, como a variável dependente será se frequenta escola ou não, trata-se de uma variável binária. Assim, será utilizado um Logit em painel e seus devidos testes. Por fim, conforme apontado na Seção de Metodologia do Logit, é necessário calcular o Efeito Marginal para encontrar o efeito de fato do Nova Semente na frequência escolar.

Devido à baixa variabilidade de indivíduos entre tratado e controle, isto é, indivíduos que se mudaram entre Petrolina e Juazeiro, um painel terá pouca variabilidade no *cross-section*, novamente isso será corrigido estimando um logit tradicional, com variáveis de controle ancoradas no ano base e a inclusão de *dummies* de tempo e a interação dessas variáveis com o tratamento: o que captura o efeito de fato, conforme a estruturação da Equação 10 abaixo, conforme é estimada no *software* Stata. Lembrando que o mecanismo do modelo logit está explicado na seção de Revisão da Literatura.

$$\text{logit } \textit{frequenta} \textit{ branca} \textit{ mulher} \textit{ filho\_fam} \textit{ trab2011} \textit{ idade2011} \textit{ bf2011} \textit{ d2012} \textit{ d2013} \textit{ d2014} \textit{ d2015} \textit{ P} \textit{ P2012} \textit{ P2013} \textit{ P2014} \textit{ P2015} \quad (10)$$

Onde a variável *frequenta* indica se o indivíduo frequenta a escola ou não e é a variável dependente; *branca* indica se o indivíduo está registrado como branca ou não; *mulher*, se está registrada como mulher; *filho\_fam*, se na estrutura familiar o indivíduo é filho do chefe de família; *trab2011* indica se o indivíduo estava trabalhando no período base (2011); *idade2011* indica a idade do indivíduo no período base; *bf2011* indica se, em 2011, era beneficiário do programa Bolsa Família; *d2012 – d2015* são as *dummies* de tempo; *P* indica se

o indivíduo pertence ao grupo dos tratados (Petrolina);  $P_{2012} - P_{2015}$  são as interações de cada período com a variável de tratamento  $P$ .

Em seguida, para aprofundar o exercício, serão selecionadas apenas as mulheres entre 11 e 17 anos, período no qual é comum, em especial em famílias de renda mais baixa, terem que cuidar de irmãos mais novos ou outras crianças. Assim, assume-se que: mulheres entre 11 e 17 anos frequentam menos as escolas do que o resto da amostra; deseja-se testar a hipótese: O Programa Nova Semente, se bem sucedido, permite que essas mulheres frequentem mais a escola, à medida que vagas em creche se tornaram acessíveis.

### c. Resultados

Inicialmente são realizadas as estimativas dos modelos de Logit em painel, utilizando os modelos de Efeitos Aleatórios e de Efeitos Fixos. São estimados seus efeitos marginais e salvas suas estimativas, os efeitos marginais de cada modelo, bem como seus respectivos *p-valores* encontram-se na Tabela 8 a seguir. Ademais, foi realizado o teste de *Hausman* e obtido um *chi-quadrado* de  $-3428,33$ , o que indica rejeição de sua hipótese nula.

**Tabela 8: Resultados pelos Modelos de Logit em Painel**

<b>Modelo</b>	<b>Estimador do Tratamento</b>	<b>P-valor</b>
Efeitos Aleatórios	0.0487	0.00
Efeitos Fixos	0.0054	0.015

Fonte: Cadastro Único

Observa-se que em ambos os modelos o efeito é positivo, conforme a hipótese inicial do presente trabalho: indivíduos tratados pelo programa Nova Semente frequentam mais a escola do que indivíduos em Juazeiro. Ademais, como a hipótese nula do teste de *Hausman* foi rejeitada, a diferença dos estimadores não é estatisticamente significativa, de forma que apenas o estimador do modelo de Efeitos Fixos é consistente.

O modelo de Efeitos Fixos oferece um norte para o resultado, porém, conforme apontado na seção anterior, devido à baixa variabilidade entre os grupos de tratado e controle, além da variabilidade temporal, esse modelo não trará resultados adequados pois ser tratado (Petrolina) não varia no tempo. Para corrigir tal problema, é estimada modelo conforme a Equação 10. Os resultados obtidos estão apresentados na Tabela 9.

**Tabela 9: Efeitos na Frequência Escolar pelo Cadastro Único**

Variáveis	Coeficiente	Erro Padrão	z	P>z	Intervalo de Confiança	
branca	-0.18	0.02	-9.75	0.00	-0.22	-0.15
mulher	0.02	0.01	1.36	0.17	-0.01	0.04
filho_fam	0.47	0.02	22.23	0.00	0.43	0.51
trab2011	-0.27	0.12	-2.23	0.03	-0.50	-0.03
idade2011	0.73	0.00	263.03	0.00	0.73	0.74
bf2011	1.04	0.02	65.08	0.00	1.01	1.08
d2012	0.94	0.03	36.77	0.00	0.89	0.99
d2013	1.60	0.03	60.81	0.00	1.54	1.65
d2014	2.28	0.03	83.42	0.00	2.22	2.33
d2015	2.92	0.03	101.76	0.00	2.86	2.97
P	0.31	0.02	13.15	0.00	0.27	0.36
P2012	-0.10	0.03	-2.84	0.00	-0.16	-0.03
P2013	0.17	0.03	4.93	0.00	0.10	0.24
P2014	0.38	0.04	10.69	0.00	0.31	0.45
P2015	0.50	0.04	13.16	0.00	0.42	0.57
Constante	-5.37	0.04	-150.28	0.00	-5.44	-5.30

Fonte: Cadastro Único

Com relação às variáveis de controle, observa-se, pela Tabela 9, que a criança ser branca afeta negativamente na frequência escolar, a princípio esse é um resultado contrário ao esperado, mas dado que já há diversos outros controles relevantes, como renda, é plausível isso ser pouco relevante na frequência escolar. A *dummy* se é mulher ou homem não é estatisticamente significativa para essa amostra, o que pode dificultar o próximo exercício, mas é plausível pelo mesmo motivo anterior. Se a criança é filha do chefe de família há maior frequência escolar, o que é esperado, dado que isso reflete uma estrutura familiar mais equilibrada. Conforme esperado também, trabalho infantil está inversamente relacionado e a idade está positivamente relacionada à frequência escolar. Além disso ser beneficiário do programa Bolsa Família possui coeficiente positivo, indicando eficácia do programa, o qual possui frequência escolar comprovada como um dos requisitos para a transferência de renda. Além disso, as variáveis de ano se mostraram positivas e crescentes, controlando demais fatores relevantes ocorridos ao longo do tempo nas cidades para o aumento da frequência escolar.

Por fim, com relação às variáveis de interesse a *dummy P* sem interação incorpora todos os fatores constantes no tempo (pois remete ao ano base de

2011) relacionados a Petrolina que afetam no aumento da frequência escolar. Cada *dummy* de *P2012* a *P2015* reflete os efeitos do programa Nova Semente em cada ano – mantidas demais mudanças nessas cidades constantes, vale destacar que todas foram estatisticamente significantes. A variável *P2012* apresentou coeficiente negativo, diferente do esperado, o que pode indicar pouca adesão ou eficácia do programa nos seus primeiros anos ou alguma melhora na frequência escolar dos alunos em Juazeiro. Por outro lado, as demais variáveis de *P2013* a *P2015* foram positivas e crescentes, indicando que o programa foi, progressivamente, garantindo melhores resultados.

Para aprofundar o exercício foi realizada a mesma regressão da Equação 10 para mulheres entre 11 e 17 anos. Os resultados serão encontrados na Tabela A 9, conforme esperado pela pouca não significância estatística observada anteriormente, nenhuma das *dummies* de interação foram estatisticamente significantes, sinalizando não haver efeitos do programa Nova Semente nesse grupo.

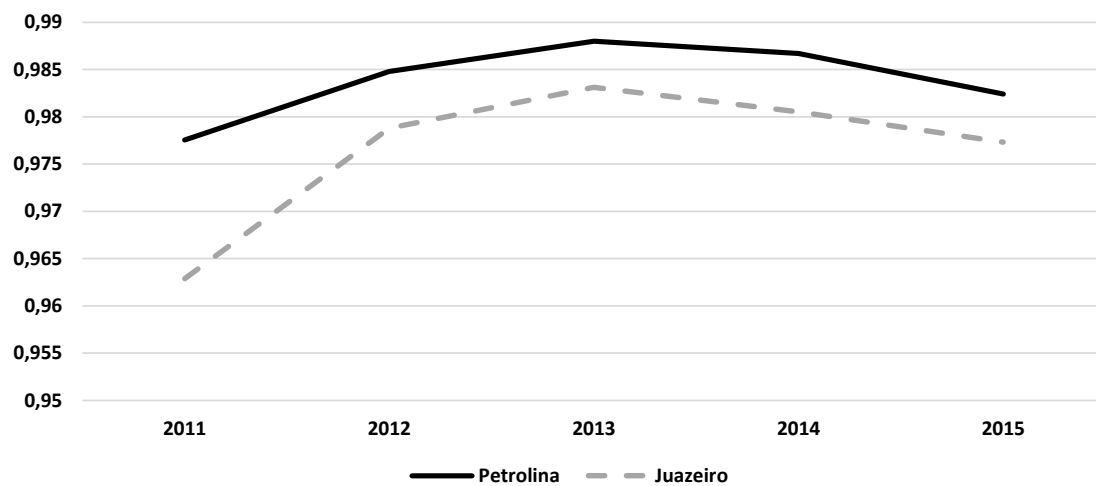
Porém, houve claro problema de “efeito de teto”, inclusive com uma média desse grupo superior ao resto da amostra, conforme a Tabela 10 e Figura 6. Foi realizado o mesmo teste controlando pelo programa Bolsa Família, que é condicional à frequência escolar, mas como há poucas observações não tratadas pelo programa Bolsa Família houve pouca variabilidade. O mesmo ocorreu quando o exercício foi realizado apenas com mulheres de 13, 14 e 15 anos respectivamente.

Esse resultado limitado também ocorreu pela falta de uma variável relacionando as crianças, pois analisando apenas as meninas que, de fato, tenham irmãos mais novos deve-se obter melhores resultados. A falta de ligação das famílias também limita saber a proporção de pessoas na família nascidas após o início do programa que também pode trazer novas conclusões.

**Tabela 10: Médias de frequências totais e de mulheres entre 11 e 17 anos**

<b>Amostra</b>	<b>Observações</b>	<b>Média</b>	<b>Desvio Padrão</b>	<b>Min</b>	<b>Max</b>
Mulheres entre 10 e 17	54004	0.98	0.13	0	1
Todos entre 0 e 17	343830	0.75	0.43	0	1

Fonte: Cadastro Único

**Figura 6: Evolução da Frequência Escolar de Mulheres entre 11 e 17 anos**

Fonte: Cadastro Único



#### d. Discussão

O exercício com o Cadastro Único trouxe duas importantes contribuições: a ausência de contaminação do programa Nova Semente e efeitos positivos do programa na frequência escolar.

Com a variável identificadora de nome da escola, foi possível estabelecer que, da amostra entre 2011 e 2015, 1599 alunos que já residiram Petrolina estavam matriculados em alguma escola do Nova Semente e apenas 9 alunos que já residiram em Juazeiro foram matriculados nessas escolas, isso implica duas coisas: não há contaminação ou essa é quase nula e não houve mobilidade relevante de mudança de cidades com benefícios ao uso do programa.

Aproveitando a riqueza da base, inicialmente, foram realizadas estimativas por métodos de Logit em painel, porém, dado a ausência de variabilidade da variável de tratamento, tal procedimento é inadequado. Portanto foi estimado um Logit com interações do tratamento com *dummies* de tempo conforme a Equação 10.

Na estimação dos efeitos na frequência escolar, a maioria dos coeficientes foram significantes e com seu sinal esperado, com relação às interações apenas o ano de 2012 apresentou variação negativa, porém os demais anos foram positivos e crescentes, indicando eficácia do programa Nova Semente no aumento da frequência escolar.

Houve a tentativa de testar se o programa aumentou a frequência escolar de alunas entre 10 e 17 anos, período no qual o grupo frequentemente precisa ficar em casa para cuidar de outras crianças mais novas. Porém, a frequência desse grupo se mostrou superior ao da amostra toda, possivelmente como consequência dos requisitos do programa Bolsa Família que abrangem cerca de 87% da amostra, conforme a Tabela A 7.

Entre as principais limitações do presente exercício destaca-se a carência de variáveis educacionais, tendo apenas frequência escolar com boa variabilidade e não havendo indicadores de aprovação, como no Censo Escolar. Além disso, nesse exercício houve foco nos alunos de 0 a 17 anos,

uma análise de toda a população, incluindo membros da família poderia enriquecer a pesquisa, tanto para analisar o mercado de trabalho da mulher como para melhorar alguns exercícios com a inclusão da possibilidade de irmão, por exemplo.

## 5. Conclusão

Com o intuito de avaliar os efeitos educacionais de uma expansão repentina em vagas de educação infantil, decorrente do programa Nova Semente, foram realizados dois principais exercícios: o primeiro utilizando um método de pareamento com diferenças em diferenças a partir do Censo Escolar, nível escola; o segundo utilizando um modelo de Logit a partir de dados do Cadastro Único. O primeiro utilizou aprovação e abandono escolar como variáveis de interesse, o segundo analisou apenas a frequência escolar.

No primeiro exercício, a partir do pareamento com as próprias variáveis da base houve evidência positiva de um aumento de 6,9 p.p. na aprovação de alunos do segundo ano, em 2012, decorrente da expansão de vagas. Porém, as demais variáveis como aprovação no primeiro ano e abandono não se mostraram estatisticamente significantes, em linha com a literatura nacional indicando ambiguidade dos efeitos da expansão, conforme apontado por Santos (2015).

O primeiro exercício apresentou uma série de limitações, em especial o viés de composição de se utilizar o nível agregado de escolas. E talvez isto possa explicar parte dos resultados alcançados. Para complementar a pesquisa, foi investigada a melhora na frequência escolar, usando a base do Cadastro Único e estimando um modelo Logit com variáveis de interação do tratamento com cada período. Apesar do sinal negativo apresentado pela variável de interação com o ano de 2012, os demais anos foram positivos e crescentes, em linha com o gradual ingresso de alunos cujo acesso a vagas (exposição ao programa) aumentou ao longo dos anos

Portanto, a partir das bases de dados utilizadas e seus respectivos exercícios concluí-se que há pouca evidência de efeitos positivos educacionais decorrentes da grande expansão de vagas em educação infantil para o caso do programa Nova Semente. Vale destaque para evidência mista em aprovação escolar, ainda que crescente ao longo dos anos, ausência de evidência significativa na redução de abandono (que sofre “efeito de teto”) e também efeitos mistos, mas crescentes com a exposição do programa no aumento da frequência escolar de indivíduos de 0 a 17 anos.

## 6. Referências

ABADIE, Alberto; DIAMOND, Alexis; HAINMUELLER, Jens. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59.2: 495-510, 2015.

BARNETT, W. Steven. "Effectiveness of early educational intervention". *Science* 333.6045: 975-978, 2011.

BARROS, Ricardo Paes de; HENRIQUES, Ricardo; MENDONÇA, Rosane. "Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil". *Instituto de Pesquisa Econômica Aplicada (Ipea)*. 2002.

BARROS, R. P.; MENDONÇA, R. "Uma avaliação dos custos e benefícios da educação pré-escolar no Brasil". mimeo (background paper para YOUNG, M.E. (ed.) "From early child development to human development: investing in our children's future" Worldbank, 2002). Republicado na série Textos para Discussão UFF/ Economia (TD 183, Novembro/ 2005). 1999.

BERLINSKI, Samuel; GALIANI, Sebastian. "The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment." *Labour Economics* 14.3: 665-680, 2007.

BERLINSKI, Samuel; GALIANI, Sebastian; GERTLER, Paul. "The effect of pre-primary education on primary school performance." *Journal of public Economics* 93.1: 219-234, 2009.

BERLINSKI, Samuel; GALIANI, Sebastian; MANACORDA, Marco. "Giving children a better start: Preschool attendance and school-age profiles." *Journal of public Economics* 92.5: 1416-1440, 2008.

BOLSA FAMÍLIA, 2018. Disponível em: <<http://mds.gov.br/assuntos/bolsa-familia/>>. Acesso em: 22/02/18.

BRITO. 2017 Disponível em: <http://www.carlosbritto.com/prefeitura-de-petrolina-e-petrape-rompem-parceria-na-gestao-do-nova-semente/>. Acesso em: 19/02/2018.

CADASTRO ÚNICO. Disponível em: <http://mds.gov.br/assuntos/cadastro-unico/>. Acesso em 19/02/2018.

CENSO DEMOGRÁFICO. Disponível em: [https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/resultados\\_gerais\\_amostra/resultados\\_gerais\\_amostra\\_tab\\_uf\\_microdados.shtm](https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/resultados_gerais_amostra/resultados_gerais_amostra_tab_uf_microdados.shtm). Acesso em: 04/12/2017.

CENSO ESCOLAR INDICADORES EDUCACIONAIS. Disponível em: <http://portal.inep.gov.br/indicadores-educacionais>. Acesso em 05/12/2016.

CENSO ESCOLAR MICRODADOS. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-acessar>. Acesso em: 05/12/2016.

CHIMELI, Ariaster and SOARES, Rodrigo R.. “The Use of Violence in Illegal Markets: Evidence from Mahogany Trade in the Brazilian Amazon”. IZA Discussion Paper No. 5923, 2011.

CURI, A. Z.; MENEZES-FILHO, N. “A Relação entre Educação Pré-primária, Salários, Escolaridade e Proficiência Escolar no Brasil”. Estudos Econômicos, São Paulo, 39, n. 4, outubro-dezembro 2009. 811-850, 2009.

CURRIE, Janet. “Early childhood education programs”. The Journal of Economic Perspectives 15.2: 213-238, 2001.

DATASUS. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=02>. Acesso em: 06/12/2016.

DAVISON, Russell; MACKINNON, James. "Econometric theory and methods". Vol. 5. New York: Oxford University Press, 2004

DUFLO, E. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". American Economic Review 91, 795–813, 2001.

DUFLO, Esther; BERTRAND, Marianne; MULLAINATHAN, Sendhil. "How much should we trust differences-in-differences estimates?". No. w8841. National Bureau of Economic Research, 2002.

ESTIMATIVA POPULACIONAL. Disponível em: <<https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?t=downloads>>. Acesso em: 12/12/17

GRANJA, Cleidiane Coelho. "Programa nova semente em perspectiva (2009-2012): uma avaliação de processo". 2013.

HART, B.; RISLEY, T. "Meaningful differences in the everyday experiences of young American children. Baltimore", MD: Brookes, 1995.

HAVNES, T.; MOGSTAD, M. "Is universal child care leveling the playing field?. Journal of Public Economics", 127, 100-114, 2015.

HAVNES, T.; MOGSTAD, M. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." American Economic Journal: Economic Policy, 3(2): 97-129, 2011.

IBGE CIDADES, Disponível em: <<http://cidades.ibge.gov.br/xtras/home.php>>. Acesso em: 14/09/2016.

INEP, Resultados do Pisa. Disponível em: <<http://portal.inep.gov.br/internacional-novo-pisa-resultados>>. Acesso em: 12/09/2016.

MEC. Disponível em: <<http://www.mec.gov.br/>>. Acesso em: 05/12/2016.

MENEZES FILHO, N. (Org.). "Avaliação econômica de projetos sociais". São Paulo: Dinâmica Gráfica e Editora, 2012.

MEYER, Paul L. "Probabilidade: aplicações à estatística." Probabilidade: aplicações à estatística. Livro Técnico, 1970.

NOVA SEMENTE, Disponível em: <<http://petrolina.pe.gov.br/novasemente/>>. Acesso em: 05/12/2016.

O COMEÇO DA VIDA. Direção: Estela Renner. [S.l.]: IMDb. 1 DVD (120 min), 2016.

PROVA BRASIL. Disponível em: <<http://dados.gov.br/dataset/microdados-prova-brasil>>. Acesso em 19/02/2018

RAIS. Disponível em: <<ftp://ftp.mtps.gov.br/pdet/microdados/RAIS/>>. Acesso em 06/12/2016.

SANTOS, Daniel Domingues dos. "Impactos do ensino infantil sobre o aprendizado: benefícios positivos, mas desiguais". Ribeirão Preto, 2015.

SCORZAFAVE, L. G.; SOARES, M. K.; DORIGAN, T. A. "Vale a pena pagar para desarmar? Uma avaliação do impacto da campanha de entrega voluntária de armas sobre as mortes com armas de fogo". Estudos Econômicos (São Paulo), 45(3), 475-497, 2015.

SKYSCRAPERCITY (Figura de Petrolina e Juazeiro). Disponível em: <<https://www.skyscrapercity.com/showthread.php?t=465755>>. Acesso em: 24.06.2018.

THOMPSON, Ross A; NELSON, Charles A. "Developmental science and the media: Early brain development". American Psychologist 56.1: 5, 2001.

TSE, Tribunal Superior Eleitoral. Disponível em: <<http://www.tse.jus.br/>>. Acesso em: 13/12/2016.



## 7. Apêndice

### a. Tabelas

**Tabela A 1: Observações de Dependência Administrativa**

<b>dep_adm</b>	<b>Freq.</b>	<b>Percent.</b>	<b>Acum.</b>
Estadual	542	24.55	24.55
Federal	8	0.36	24.91
Municipal	1,258	56.97	81.88
Particular	400	18.12	100.00
Total	2,208	100.00	

Fonte: Censo Escolar

**Tabela A 2: Observações por ano**

<b>Ano</b>	<b>Tratado (Petrolina)</b>		<b>Controle (Juazeiro)</b>	
	<b>Freq.</b>	<b>Percent</b>	<b>Freq.</b>	<b>Percent</b>
2008	154	12.50	122	12.50
2009	154	12.50	122	12.50
2010	154	12.50	122	12.50
2011	154	12.50	122	12.50
2012	154	12.50	122	12.50
2013	154	12.50	122	12.50
2014	154	12.50	122	12.50
2015	154	12.50	122	12.50
Total	1,232	100.00	976	100.00

Fonte: Censo Escolar

**Tabela A 3: Estatísticas Descritivas Censo Escolar**

Variável	Obs	Média	Desv. Pad.	Min	Max
aprov_1	2.020	97,18	6,68	0	100
aprov_2	2.284	87,77	14,51	0	100
aband_1	2.020	1,70	5,66	0	100
aband_2	2.284	1,89	5,34	0	100
id_agua_fi~a	3.247	1,44	0,59	0	2
id_agua_in~e	3.247	0,02	0,13	0	1
id_esgoto_~a	3.247	0,61	0,49	0	1
id_lixo_co~a	3.247	0,82	0,39	0	1
id_laborat~a	3.247	0,63	0,48	0	1
id_bibliot~a	3.247	0,44	0,50	0	1
id_parque_~l	3.247	0,25	0,43	0	1
num_comp_a~s	2.279	11,55	14,51	0	346
num_salas_~s	3.247	9,31	44,23	1	2.503
id_internet	3.017	0,60	0,49	0	1
num_funcio~s	3.247	31,31	29,60	1	280
id_mod_ens~r	3.247	1,00	0,00	1	1
id_aliment~o	3.247	0,72	0,45	0	1
id_reg_inf~e	3.247	0,22	0,42	0	1
id_reg_inf~a	3.247	0,61	0,49	0	1

Fonte: Censo Escolar

**Tabela A 4: Probit do Escore de Propensão**

Variáveis	Coef.	Desv. Pad	z	P>z	[95% IC]	
id_agua_filtrada	-0,835	0,316989	-2,63	0,008	-1,456	-0,213
id_esgoto_rede_publica	-0,383	0,238896	-1,6	0,109	-0,851	0,0853
id_laboratorio_informatica	0,5991	0,389979	1,54	0,125	-0,165	1,3634
id_reg_infantil_creche	1,0759	0,337897	3,18	0,001	0,4137	1,7382
id_biblioteca	0,8073	0,220267	3,67	0	0,3756	1,239
id_lixo_coleta_periodica	0,4605	0,259425	1,78	0,076	-0,048	0,969
num_salas_existentes	0,0269	0,026318	1,02	0,306	-0,025	0,0785
num_comp_alunos	-0,006	0,009541	-0,64	0,524	-0,025	0,0126
num_funcionarios	-0,01	0,005146	-1,9	0,058	-0,02	0,0003
_cons	0,7472	0,735424	1,02	0,31	-0,694	2,1886

Fonte: Censo Escolar, Elaboração própria

**Tabela A 5: Suporte Comum do Pareamento**

<b>Designação</b>	<b>Fora do Suporte</b>	<b>No Suporte</b>	<b>Total</b>
Não Tratado	0	81	81
Tratado	19	77	96
Total	19	158	177

Fonte: Censo Escolar, Elaboração própria

**Tabela A 6: Teste de Pareamento**

<b>Variável</b>	<b>Pareamento</b>	<b>Média Tratado</b>	<b>Média Controle</b>	<b>%viés</b>	<b>Redução de viés</b>	<b>t</b>	<b>p&gt;t</b>
id_agua_filtrada	N	1,8646	1,963	-35,4		-2,29	0,023
	P	1,8571	1,9481	-32,7	7,6	-1,91	0,058
id_esgoto_rede_publica	N	0,55208	0,45679	19		1,26	0,209
	P	0,46753	0,32468	28,5	-49,9	1,82	0,071
id_laboratorio_informatica	N	0,92708	0,91358	5		0,33	0,742
	P	0,90909	0,92208	-4,8	3,8	-0,29	0,774
id_reg_infantil_creche	N	0,28125	0,04938	65,3		4,22	0,000
	P	0,11688	0,11688	0	100	0	1,000
id_biblioteca	N	0,48958	0,23457	54,7		3,6	0,000
	P	0,37662	0,42857	-11,1	79,6	-0,65	0,514
id_lixo_coleta_periodica	N	0,79167	0,65432	30,9		2,06	0,041
	P	0,74026	0,7013	8,8	71,6	0,54	0,593
num_salas_existentes	N	8,8854	7,3333	27,2		1,79	0,076
	P	8,1558	8,1169	0,7	97,5	0,05	0,963
num_comp_alunos	N	8,5729	10,63	-30,8		-2,04	0,043
	P	8,2987	9,2727	-14,6	52,6	-0,95	0,346
num_funcionarios	N	30,031	30,753	-2,8		-0,18	0,856
	P	30,519	30,494	0,1	96,4	0,01	0,995

Fonte: Censo Escolar, Elaboração própria

**Tabela A 7: Proporção de Beneficiários do Bolsa Família na Amostra**

<b>Beneficiário do Bolsa Família</b>	<b>Frequência</b>	<b>Percentual</b>
0 (não)	43488	12.65
1 (sim)	300342	87.35
Total	343830	100

Fonte: Cadastro Único

**Tabela A 8: Estatísticas Descritivas das Variáveis do Cadastro único**

<b>Variável</b>	<b>Obs</b>	<b>Média</b>	<b>Desv. Pad.</b>	<b>Min</b>	<b>Max</b>
ano	343830	2013	1.414216	2011	2015
marc_pbf	343830	0.87	0.33	0	1
branca	343830	0.12	0.33	0	1
nis	343830		Confidencial		
mulher	343830	0.49	0.50	0	1
filho_fam	343830	0.93	0.25	0	1
trab_inf	343830	0.01	0.10	0	1
idade	343830	8.25	3.90	0	17
frequenta	343830	0.75	0.43	0	1
P	343830	0.56	0.50	0	1

Fonte: Cadastro Único

**Tabela A 9: Efeitos na Frequência de Mulheres entre 10 e 17 Anos**

<b>Variáveis explicativas</b>	<b>Coefficiente</b>	<b>Erro Padrão</b>	<b>z</b>	<b>P&gt;z</b>	<b>Intervalo de Confiança (95%)</b>	
branca	-0.31	0.09	-3.53	0.00	-0.48	-0.14
filho_fam	1.05	0.10	10.75	0.00	0.86	1.24
trab2011	0.63	0.31	2.06	0.04	0.03	1.23
bf2011	1.82	0.07	26.71	0.00	1.69	1.96
d2012	0.53	0.17	3.21	0.00	0.21	0.85
d2013	0.76	0.16	4.69	0.00	0.44	1.08
d2014	0.61	0.15	4.07	0.00	0.32	0.90
d2015	0.45	0.14	3.22	0.00	0.18	0.73
P	0.70	0.17	4.04	0.00	0.36	1.03
P2012	-0.17	0.24	-0.69	0.49	-0.64	0.31
P2013	-0.18	0.24	-0.76	0.45	-0.65	0.29
P2014	-0.14	0.22	-0.65	0.52	-0.58	0.29
P2015	-0.26	0.21	-1.27	0.21	-0.66	0.14
Constante	0.86	0.15	5.64	0.00	0.56	1.16

Fonte: Cadastro Único