

Efeito do Programa Bolsa Família sobre a Distorção idade-série no Rio Grande do Sul: Uma Análise a partir do Cadastro Único

Fábio Antonio Rasche Júnior¹

Sabino da Silva Pôrto Júnior²

RESUMO: O objetivo deste trabalho é avaliar o impacto do Programa Bolsa Família sobre o indicador educacional de distorção idade-série de crianças, adolescentes ou jovens estudantes de escolas públicas que estão cursando o ensino fundamental “novo” (9 anos) por meio do método de Regressão Descontínua (RD). Para isso utilizaremos os microdados das famílias presentes no Cadastro Único de 2024 para o estado do Rio Grande do Sul. O resultado da estimação da regressão descontínua *Fuzzy* sugere efeito negativo e significativo do programa sobre o indicador educacional de defasagem escolar. Conforme o efeito médio local do tratamento, crianças ou adolescentes de escolas públicas do ensino fundamental novo presentes no Cadastro Único, em torno do ponto de corte, possuem 0,234 ano a menos de distorção idade-série se estiverem em famílias beneficiárias do programa.

PALAVRAS-CHAVE: Distorção idade-série; Programa Bolsa Família; Regressão Descontínua.

ABSTRACT: The objective of this study is to evaluate the impact of the Bolsa Família Program on the educational indicator of age-grade distortion of children, adolescents, or young students in public schools who are attending the “new” elementary school (9 years) using the Regression Discontinuity (RD) method. To do this, we will use microdata from families present in the 2024 “Cadastro Único” for the state of Rio Grande do Sul. The result of the Fuzzy regression discontinuity estimation suggests a negative and significant effect of the program on the educational indicator of school gap. According to the local average treatment effect, children or adolescents from new elementary school in public schools present in the “Cadastro Único”, around the cutoff, have 0.234 years less age-grade distortion if they are in families that benefit from the program.

KEY WORDS: Distortion age-grade; Bolsa Família Program; Regression Discontinuity.

ÁREA: Desigualdade e pobreza e exclusão social.

CLASSIFICAÇÃO JEL: I24; I38; J18.

1. INTRODUÇÃO

A presença do estudante na idade adequada conforme o ano que ele está cursando é uma das formas de medir o progresso educacional. A Lei de Diretrizes e Bases da Educação (LDB) estabelece que o intervalo de idade da escolarização obrigatória vai dos 4 aos 17 anos no Brasil. Mais especificamente, aos 4 anos a criança deve ingressar na pré-escola, aos 6 anos no ensino fundamental e, aos 15 anos no ensino médio (BRASIL, 1996). Contudo, essa trajetória linear nem sempre é a realidade de todas as crianças e quando não acontece é comumente referida como uma situação de distorção idade-série.

¹Doutorando em Economia Aplicada pela Universidade Federal do Rio Grande do Sul. E-mail: junierrasche@live.com

²Professor titular do Departamento de Economia e do Programa de Pós-Graduação em Economia da Universidade Federal do Rio Grande do Sul. E-mail: sabino@ppge.ufrgs.br

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) calcula a taxa de distorção idade-série para as escolas. Na sua definição, o estudante encontra-se na situação de distorção idade-série quando a diferença entre a idade dele e a idade prevista para a série for de 2 anos ou mais³ (BRASIL, 2004). Segundo dados do INEP para o Brasil, a proporção de estudantes com atraso escolar tem diminuído ao longo dos anos, mas ainda é elevada. No ano de 2024, 12,4% dos estudantes do ensino fundamental de escolas públicas estavam em situação de defasagem escolar. Para o ensino médio de escolas públicas, essa proporção aumenta para 19,7% dos estudantes. Especificamente sobre o Rio Grande do Sul, a situação é pior, a taxa de distorção idade-série é 13,1% no ensino fundamental de escolas públicas e 23,6% para o ensino médio de mesma dependência administrativa (BRASIL, 2025a).

À medida que a criança ou adolescente encontra-se em uma situação de atraso escolar, ele está sob o risco de não desenvolvimento adequado de suas habilidades. O aprendizado de habilidades cognitivas e não cognitivas não ocorre de forma perfeitamente substituível entre os anos. Sob uma perspectiva multiestágio, a infância é composta por períodos sensíveis e períodos críticos para a produção das diferentes habilidades. Os estágios sensíveis dizem respeito ao retorno mais elevado do investimento no aprendizado de determinada habilidade, e o crítico é o único período em que a habilidade pode ser aprendida. Sendo assim, o incentivo à produção das habilidades deve ocorrer no momento adequado e de forma continuada (CUNHA & HECKMAN, 2007).

Garantir que os indivíduos ingressem no sistema educacional na idade adequada e se mantenham ao longo da trajetória escolar é um dos grandes desafios que programas de transferência de renda condicionada, como o Programa Bolsa Família (PBF), tentam incidir. Dentre os seus objetivos, o PBF busca "(...) promover o desenvolvimento e a proteção social das famílias, especialmente das crianças, dos adolescentes e dos jovens em situação de pobreza" (BRASIL, 2023b. Lei Nº 14.601/2023, inciso III, art. 3º). Como a situação de atraso escolar pode ser decorrente de fatores como a entrada tardia na escola, repetência ou evasão/abandono escolar, o Programa Bolsa Família pode mitigar o problema de defasagem à medida que incentiva o ingresso das crianças na escola conforme idade estabelecida por lei e possui como uma de suas condicionalidades a frequência escolar (RIBEIRO & CACCIAMALI, 2012).

O objetivo deste trabalho é avaliar o impacto do Programa Bolsa Família sobre o indicador educacional de distorção idade-série de crianças, adolescentes ou jovens estudantes de escolas públicas que estão cursando o ensino fundamental “novo” (9 anos) por meio do desenho de Regressão Descontínua (RD). Para isso utilizaremos os microdados das famílias presentes no Cadastro Único de 2024 para o estado do Rio Grande do Sul. Observaram-se 2.724.460 indivíduos com cadastros atualizados até 24 meses no estado.

Alguns trabalhos já têm explorado esse tema. Ribeiro e Cacciamali (2012) investigaram o efeito do Bolsa Família sobre a defasagem idade-série e a frequência escolar por meio do método de *Propensity Score Matching* (PSM). Os autores utilizaram como base de dados a Pesquisa Nacional por Amostra de Domicílios (PNAD) do ano de 2006, porém não encontraram resultados significativos do programa sobre esses indicadores.

Araújo, Frio e Alves (2021) também analisaram o impacto do PBF sobre a distorção idade-série utilizando a metodologia de pareamento por escore de propensão. A partir dos microdados do Suplemento Educação da PNAD Contínua, referente ao 2º semestre dos anos de 2016 a 2019, os autores obtiveram resultados significativos do Bolsa Família sobre a defasagem escolar, exceto para a região Sul do Brasil. Além disso, a magnitude do efeito foi maior para estudantes do sexo masculino.

³ Consideremos t o ano da base de dados, k a série de ensino, e i a idade adequada para essa mesma série. A justificativa para utilizar 2 anos ou mais como critério é pelo fato de que os estudantes que “(...) nasceram em $t-[i+1]$, completam $i+1$ anos no ano t e, portanto, em algum momento deste ano (de 1º de janeiro a 31 de dezembro) ainda permaneciam com i anos e, por isso, o critério adotado considera estes alunos como tendo idade adequada para esta série. Os que nasceram depois de $t-[i+1]$ completam, no ano t , i anos ou menos” (BRASIL, 2004, p. 17).

Hermeto e Romero (2011), analisaram o efeito do Programa Bolsa Família sobre indicadores educacionais utilizando a base do Cadastro Único de 2005 e o método RD Sharp. No entanto, a distorção idade-série não foi explicitamente objeto de análise. Os autores avaliaram, especificamente, a evasão ou abandono escolar, a aprovação, repetência e alocação entre trabalho e estudo.

Dessa forma, apesar da proximidade com os trabalhos citados, buscamos contribuir para a discussão com uma base de dados mais recente e ainda não explorada para a análise do problema da distorção idade-série.

Os resultados encontrados, por meio da estimação da regressão descontínua *Fuzzy*, apontam que o Programa Bolsa Família possui um efeito negativo e significativo sobre a distorção idade-série. O efeito médio local do tratamento para os *compliers* sugere que crianças ou adolescentes de escolas públicas do ensino fundamental novo presentes no Cadastro Único, em torno do ponto de corte, possuem 0,234 ano a menos de defasagem escolar se estiverem em famílias beneficiárias do programa. Ademais, conforme o método de regressão logística binária, crianças ou adolescentes do grupo de beneficiários têm 42,5% menos chances de estarem no grupo daqueles com distorção idade-série.

Além desta introdução, o presente trabalho conta com cinco seções. Na seção dois, é discutido o referencial empírico. Na seção três apresentaremos a base de dados do Cadastro Único e estatísticas descritivas. Em seguida, na seção quatro, é apresentado a estratégia empírica utilizada para a inferência causal. Na seção cinco apresentamos os resultados e os testes de robustez. Por fim, são apresentadas as considerações finais.

2. REFERENCIAL EMPÍRICO

A distorção idade-série pode estar diretamente associada a fatores como a entrada tardia na escola, repetência ou evasão (RIBEIRO & CACCIAMALI, 2012). Esses aspectos, no entanto, estão relacionados a fatores principalmente socioeconômicos dos estudantes.

A partir da base de dados da PNAD de 1996 para o Brasil, Machado & Gonzaga (2007) utilizaram o método de variáveis instrumentais para identificar o efeito da renda per capita e do nível educacional dos pais sobre a probabilidade de as crianças terem distorção idade-série. Os resultados encontrados pelos autores apontam para um impacto negativo e significativo da renda e da educação dos pais na defasagem escolar, sendo o primeiro de maior magnitude. Eles também destacam que a situação de distorção tem maior chance de ocorrer em crianças do sexo masculino, não brancos e amarelos, e de áreas metropolitanas.

Conforme Neto et al (2019), em sua investigação para o Rio Grande do Sul com dados do Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul (SAERS) de 2016, a proporção de estudantes com defasagem escolar é maior em escolas que possuem uma maior taxa de abandono e uma menor média do indicador socioeconômico. Além disso, essa situação de atraso escolar é maior entre os estudantes do sexo masculino que vivem em áreas urbanas.

Sendo a renda um fator importante para a determinação da distorção idade-série, faz sentido pensar que um programa de transferência de renda condicionada pode atuar sobre esse indicador educacional. No caso brasileiro, o Programa Bolsa Família, enquanto combate a insuficiência de renda por meio da transferência direta, pode contribuir para o ingresso das crianças no sistema educacional na idade adequada e mitigar a evasão por meio da condicionalidade de frequência escolar (RIBEIRO & CACCIAMALI, 2012).

O efeito do Programa Bolsa Família sobre indicadores educacionais é frequentemente objeto de análise e seus resultados podem variar dependendo da base de dados utilizada ou do método adotado (RIBEIRO, SHIKIDA & HILLBRECHT, 2017).

Por meio do método de regressão descontínua e da base do Cadastro Único de 2005, Hermeto e Romero (2011) identificaram que o Programa Bolsa Família possui efeitos significativos sobre a

evasão escolar, aprovação e repetência. Eles observaram que o programa de transferência de renda diminuiu a evasão para as regiões do Norte/Centro-Oeste e do Nordeste, assim como é menor para o sexo feminino. No que se refere à aprovação, o efeito foi maior para os meninos da região Nordeste. Além disso, os resultados demonstraram que o programa diminuiu a repetência para as crianças do sexo masculino no Nordeste.

Um dos primeiros estudos com o intuito de investigar especificamente o efeito do Programa Bolsa Família sobre a distorção idade-série foi realizado por Ribeiro e Cacciamali (2012). Os autores utilizaram o método de *Propensity Score Matching* para a estimação e os dados foram provenientes dos questionários suplementares da PNAD do ano de 2006. A amostra foi delimitada para famílias com até 3 crianças e/ou jovens de 7 a 15 anos e considerou apenas o ensino fundamental “antigo” (de 8 anos). Os resultados obtidos mostraram que não houve impacto significativo do programa sobre a defasagem idade-série.

Araújo, Frio e Alves (2021) buscaram atualizar o trabalho de Ribeiro e Cacciamali (2012) ao utilizarem os microdados do Suplemento Educação da PNAD Contínua referentes ao 2º semestre dos anos de 2016 a 2019. O método utilizado para estimar o efeito causal também foi o pareamento por escore de propensão. A fim de buscar heterogeneidade nos resultados, os autores realizaram estimações separadas por gênero, região e ano da pesquisa domiciliar (PNAD). Diferentemente do trabalho anterior, os resultados identificaram que o impacto do PBF sobre a distorção idade-série foi negativo e significativo, variando entre 2,2 e 6,02 pontos percentuais. Além disso, a redução observada foi maior para meninos, variando entre 4,4 e 7,2 pontos percentuais, enquanto para meninas a variação foi entre 2,7 e 4,5 pontos percentuais. Regionalmente, observou-se redução significativa em todas as regiões, exceto no Sul.

Dessa maneira, buscamos contribuir para a discussão ao analisar o efeito do Programa Bolsa Família sobre a distorção idade-série dos estudantes do ensino fundamental novo (9 anos) especificamente para o estado do Rio Grande do Sul por meio da base de dados do Cadastro Único de 2024. No próximo tópico apresentaremos a base utilizada e as estatísticas descritivas relacionadas.

3. BASE DE DADOS

A base de dados utilizada para a análise do impacto do Programa Bolsa Família sobre a distorção idade-série será o Cadastro Único de janeiro de 2024 para o estado do Rio Grande do Sul.

3.1 Cadastro Único

O Cadastro Único (CadÚnico), instituído pelo art. 6º-F da Lei nº 8.742, de 7 de dezembro de 1993 (BRASIL, 1993), é um instrumento de identificação e caracterização socioeconômica de famílias brasileiras que apresentam algum grau de vulnerabilidade social. Regulamentado pelo Decreto nº 11.016, de 2022, o CadÚnico é considerado como a porta de entrada para programas sociais, visto que seus dados cadastrais podem ser utilizados para a seleção de beneficiários. Um dos principais exemplos é o Programa Bolsa Família, que possui como um dos critérios para o seu recebimento a inscrição no Cadastro Único e a sua atualização recorrente.

Além do seu caráter instrumental para a implementação de programas sociais, o CadÚnico destaca-se pela sua abrangência quase censitária da população brasileira em situação de vulnerabilidade social e, principalmente, pela variedade de informações a respeito das condições desses indivíduos e dessas famílias, como perfil e composição familiar, condições habitacionais, condições educacionais, trabalho e renda, entre outros (BARROS et al 2009). Desse modo, os dados presentes no Cadastro Único nos permitem identificar a situação educacional específica de distorção idade-série dos indivíduos cadastrados, assim como a participação em programas sociais como o Bolsa Família.

O número de indivíduos cadastrados para o mês de janeiro de 2024 foi de 3.457.882. No entanto, consideraremos apenas os casos cuja atualização cadastral foi igual ou inferior a 24 meses⁴. Nesse caso, observaram-se 2.724.460 indivíduos cadastrados no estado do Rio Grande do Sul. Isso representa um total de 1.211.057 famílias com cadastros atualizados.

3.1.1 Variáveis e Estatísticas Descritivas

Foram selecionadas variáveis na base do Cadastro Único que nos permitissem analisar a relação existente entre a participação dos indivíduos no PBF e a situação de defasagem escolar deles.

A amostra selecionada foi composta por indivíduos com cadastros atualizados até 24 meses cuja renda familiar per capita é igual ou inferior a meio salário-mínimo (R\$ 706,00⁵) e diferente de zero. Ainda, restringimos a amostra às pessoas com idade maior que 5 anos e menor ou igual a 25 anos e que declararam estarem cursando o ensino fundamental “novo” (9 anos) em escola pública. Além disso, foram considerados apenas aquelas pessoas que não estavam trabalhando na semana anterior ao preenchimento do cadastro e cuja espécie de domicílio é classificada como particular permanente. Dessa forma, o número final de observações foi de 279.073 indivíduos.

Conforme a tabela de estatísticas descritivas abaixo, podemos perceber que a nossa amostra é composta majoritariamente por indivíduos brancos ou amarelos que vivem em áreas urbanas, 77% e 86%, respectivamente. Ainda, pode-se observar que 71,5% dos indivíduos são beneficiários do Programa Bolsa Família. No que se refere à educação dos responsáveis familiares dos estudantes, destaca-se que 61,5% não alcançaram o ensino médio. Em relação ao domicílio, 24,2% vivem em casas cujo esgotamento sanitário é inadequado⁶ e 17,3% não têm acesso à rede geral de distribuição de água⁷.

Quanto a proporção de estudantes em distorção idade-série, a base do CadÚnico não possui uma variável específica para a identificação dessa situação. Não obstante, podemos construí-la a partir da idade da pessoa em comparação com a idade adequada para o ano que ela está cursando na escola. O critério de cálculo do INEP, e normalmente utilizado, considera os estudantes em distorção idade-série aqueles que possuem 2 anos ou mais de diferença para a idade adequada. A partir dessa métrica, observamos na Tabela 1 que na nossa amostra de janeiro de 2024 para o Rio Grande do Sul foram identificados 16% dos estudantes nessa situação (HERMETO & ROMERO, 2011; RIBEIRO & CACCIAMALI, 2012; ARAÚJO, FRIJO & ALVES, 2021). Conforme o Censo da Educação Básica de 2024, o estado do RS registrou, para toda população de estudantes do ensino fundamental de escolas públicas, uma taxa de distorção idade-série de 13,1% (BRASIL, 2025a). Isto é, nossa amostra apresenta uma proporção maior de estudantes na situação de defasagem.

Tabela 1 - Estatísticas Descritivas

Variável	Média	Desvio Padrão	Mínimo	Máximo
Recebe PBF	0.715	0.451	0.000	1.000
Renda Familiar per Capita	204.409	186.619	1.000	706.000
Em Distorção Idade-Série	0.160	0.367	0.000	1.000
Distorção em anos	1.141	1.524	-8.979	17.956
Feminino	0.485	0.500	0.000	1.000
Branco ou Amarelo	0.770	0.421	0.000	1.000
Idade	11.015	3.136	5.002	24.999
Responsável sem Instrução	0.028	0.164	0.000	1.000

⁴ Esse critério é normalmente adotado visto que os cadastrados com tempo de atualização maior que 24 meses passam para averiguação cadastral do Ministério do Desenvolvimento Social.

⁵ Valor referente ao salário-mínimo do ano de 2024.

⁶ Esgotamento sanitário inadequado foi classificado como a situação em que não existe banheiro no domicílio ou quando existe, porém, a forma de escoamento sanitário ocorre por meio de fossa rudimentar, vala a céu aberto, direto em um rio, lago ou mar, ou de outra forma.

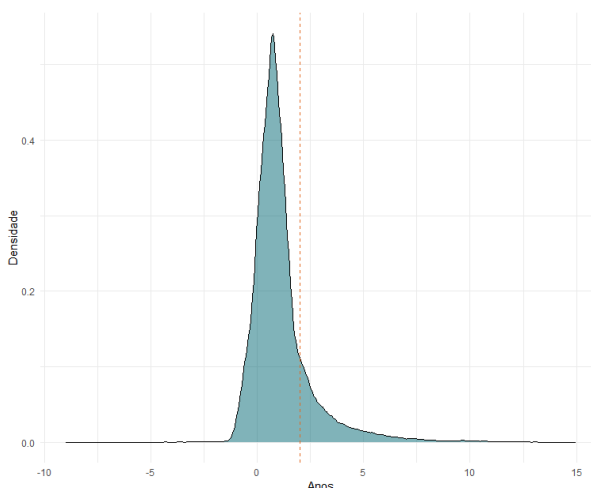
⁷ Ausência de rede geral de distribuição de água foi classificada como situação em que a forma de abastecimento de água ocorre por meio de poço ou nascente, cisterna, ou outra forma.

Variável	Média	Desvio Padrão	Mínimo	Máximo
Resp. Fundamental Incompleto	0.422	0.494	0.000	1.000
Resp. Fundamental Completo	0.165	0.371	0.000	1.000
Resp. Médio Incompleto	0.129	0.335	0.000	1.000
Resp. Médio Completo	0.231	0.421	0.000	1.000
Resp. Superior Incompleto ou mais	0.026	0.158	0.000	1.000
Urbano	0.856	0.352	0.000	1.000
Qtde Pessoas Domicílio	3.730	1.522	0.000	70.000
Sem Rede Geral Dist. Água	0.173	0.378	0.000	1.000
Esgotamento Sanitário Inadequado	0.242	0.428	0.000	1.000

Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Outra forma de visualizarmos a situação de distorção é a partir da distribuição da quantidade de anos de defasagem que os estudantes possuem em relação à idade adequada do ano que estão cursando. Se essa diferença for negativa, significa que os estudantes estão antecipados em relação a idade para aquele ano escolar. Caso contrário, a diferença positiva representa a quantidade de anos que o estudante está defasado em relação a idade adequada. Conforme Figura 1, podemos observar que a distribuição da quantidade de anos de defasagem dos estudantes para a nossa amostra está concentrada em torno de 1 ano com uma assimetria positiva, em que o valor da mediana é de 0,8241 ano e da média é de 1,1408 ano de defasagem escolar. A linha tracejada vermelha refere-se ao critério de cálculo do INEP de 2 anos.

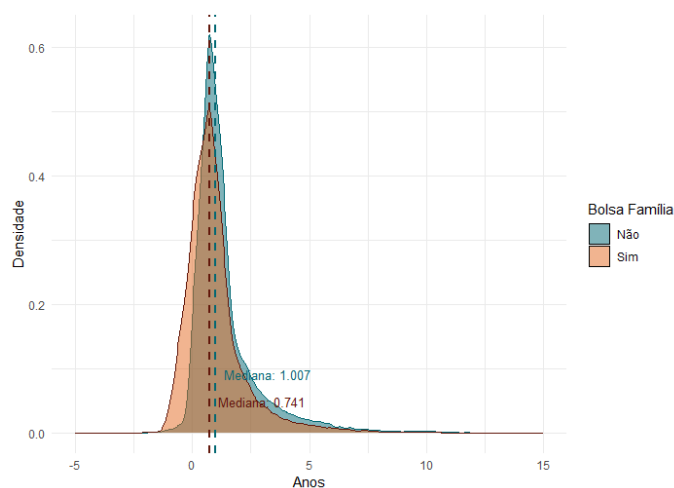
Figura 1 - Distribuição dos Anos de Defasagem dos Indivíduos



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Se analisarmos a distribuição dos anos de distorção separados conforme recebimento ou não do Programa Bolsa Família, notamos que a mediana de anos de defasagem de quem é beneficiário do programa (0,741) é menor do que a mediana do grupo de não beneficiários (1,010).

Figura 2 - Distribuição dos Anos de Defasagem dos Indivíduos Conforme Bolsa Família



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

A Tabela 2 nos mostra que, na amostra em questão, o percentual de pessoas em distorção idade-série que são beneficiários do Programa Bolsa Família é de 14,3%. Ao passo que esse percentual é de 20,3% para os não beneficiários.

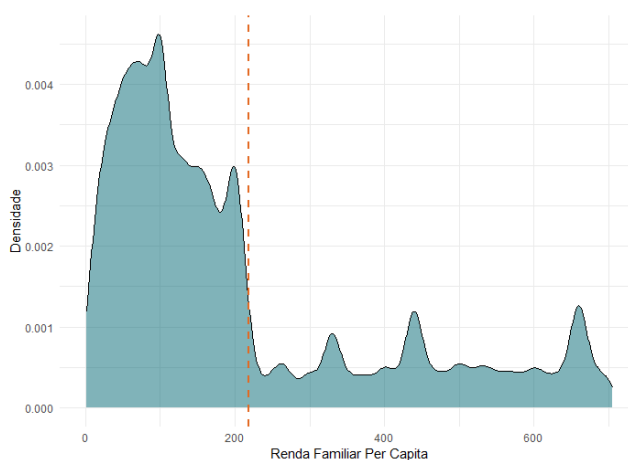
Tabela 2 - Quantidade de Pessoas em Distorção Idade-Série conforme PBF

Recebe PBF	Sem Distorção	Com Distorção
Sim	170.899	28.614
Não	63.436	16.124

Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Em relação à renda familiar per capita, a média observada para nossa amostra foi de R\$ 204,41. No gráfico abaixo apresentamos a distribuição dela. Verificamos que existe uma grande concentração de indivíduos com renda abaixo de R\$ 200,00, possivelmente por se tratar de uma base cujo objetivo é identificar pessoas em situação de vulnerabilidade social. A linha tracejada refere-se ao ponto de corte utilizado pelo PBF de R\$ 218,00.

Figura 3 - Distribuição da Renda Familiar Per Capita dos Indivíduos



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Se classificarmos a renda familiar per capita em faixas, pela Tabela 3, observamos que 73% das pessoas da amostra possuem renda inferior a R\$ 218,00, ou seja, encontra-se nas chamadas faixas de pobreza ou extrema pobreza.

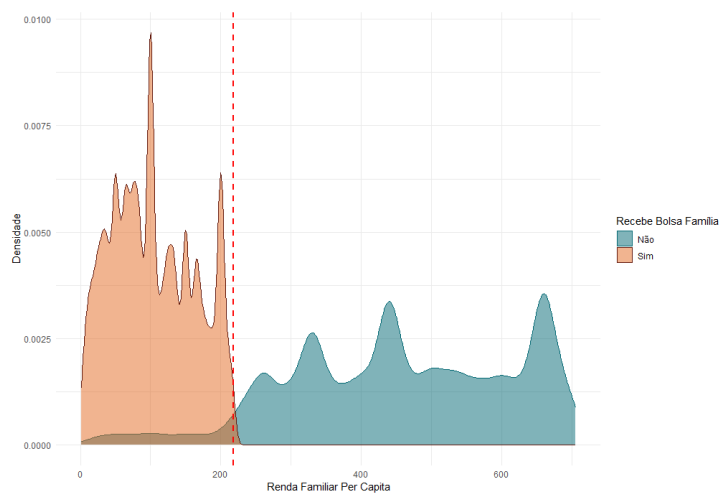
Tabela 3 - Quantidade de Pessoas por Faixa de Renda

Faixas de renda	Quantidade	Percentual (%)
R\$0,00 a R\$109,00	116.996	41,92
R\$109,01 a R\$218,00	86.746	31,08
R\$218,01 a R\$706,00	75.331	26,99

Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Assim como no caso da distorção idade-série, podemos analisar a distribuição da renda familiar per capita conforme participação no Programa Bolsa Família (Figura 4). É possível perceber que os indivíduos que participam do programa estão localizados abaixo de R\$ 218,00 (linha vertical tracejada). Isso se deve ao fato de que o critério de elegibilidade para a família receber o benefício, além da inscrição no Cadastro Único, é possuir renda per capita mensal inferior a esse limite (BRASIL, 2023b). Entretanto, não necessariamente quem está registrado no CadÚnico e possui renda inferior a R\$ 218,00 será beneficiário. No nosso caso, 1,5% das pessoas da nossa amostra encontram-se abaixo desse limite e não fazem parte das famílias beneficiárias. Como a participação é voluntária, é esperado que a adesão não seja total. Na figura abaixo, podemos visualizar esse comportamento pela área sombreada abaixo da linha tracejada.

Figura 4 - Distribuição da Renda Familiar Per Capita Conforme PBF



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Vale destacar que a base de dados do Cadastro Único apresenta registros de famílias com renda per capita mensal superior a R\$ 218,00 e que são beneficiárias do PBF. Contudo, esses casos acontecem em virtude das regras de transição. Quando a renda familiar per capita ultrapassa o critério de elegibilidade, mas não é superior a meio salário-mínimo, ela passa a fazer parte da chamada “Regra de Proteção”, na qual 50% do benefício é mantido pelo período de 24 meses (BRASIL, 2023b. Lei Nº 14.601/2023, art. 6º). Para nossa amostra, retiramos essas famílias, por isso só observamos não beneficiários acima da linha tracejada na Figura 4.

A variável da renda familiar per capita mensal, no Cadastro Único, decorre da renda total declarada pelo responsável familiar no momento da entrevista ou atualização cadastral e do número de pessoas na família (BRASIL, 2023b. Lei Nº 14.601/2023, inciso III, art. 4º). Apesar do caráter autodeclaratório das informações, as famílias passam pelo processo de “averiguação cadastral”, que se trata da verificação das informações declaradas no CadÚnico com outros registros administrativos

do governo federal a fim de identificar possíveis inconsistências, principalmente no que diz respeito à renda dos integrantes da família (BRASIL, 2025b).

Isto posto, apresentaremos na próxima seção a estratégia empírica utilizada para analisar o efeito do Programa Bolsa Família sobre a distorção idade-série dos estudantes do ensino fundamental novo do estado do Rio Grande do Sul a partir da base de 2024 do Cadastro Único.

4. ESTRATÉGIA EMPÍRICA

Vimos anteriormente que o Programa Bolsa Família possui como critério de elegibilidade que a família esteja inscrita no Cadastro Único e possua renda per capita mensal igual ou inferior a R\$ 218,00 (BRASIL, 2023b. Lei Nº 14.601/2023, Art. 5º). Isto é, a escolha dos beneficiários não acontece de forma aleatória, mas com base em uma variável observável e um ponto de corte que determinará o tratamento. Assim sendo, para estimarmos o efeito causal do PBF sobre a distorção idade-série, lançaremos mão do desenho de regressão descontínua (RD).

4.1 Regressão Descontínua

Sob o arcabouço de resultados potenciais, podemos escrever $Y_i(0)$ como o resultado potencial do indivíduo i quando não recebe o tratamento, independentemente de ter sido tratado ou não; e $Y_i(1)$ o resultado potencial do mesmo indivíduo quando recebe o tratamento, independentemente de ter recebido ou não o tratamento. Nesse caso, o efeito causal do tratamento para o indivíduo i seria dado pela diferença $Y_i(1) - Y_i(0)$. Contudo, não podemos observar uma pessoa em dois estados diferentes simultaneamente. Para o par de resultados potenciais $(Y_i(1), Y_i(0))$ e T_i a variável binária que recebe valor 1 caso o indivíduo for tratado e 0, caso contrário, o resultado que observamos efetivamente para o indivíduo i é apenas uma das duas situações:

$$Y_i = Y_i(0) + T_i(Y_i(1) - Y_i(0)) \quad (1)$$

Sendo assim, precisamos calcular a diferença entre a média de $Y_i(1)$ e a média de $Y_i(0)$ sobre os grupos de tratamento e controle. A grande dificuldade que surge a partir disso é a construção de um grupo controle que seja capaz de representar o grupo de tratamento caso esse não fosse tratado, isto é, a formação do contrafactual. O método de aleatorização, conhecido como “padrão-ouro” da avaliação, utiliza o processo aleatório de atribuição do tratamento para construir os dois grupos de comparação. Esse procedimento garante que os grupos de tratamento e controle sejam idênticos tanto em características observáveis como não observáveis. Assim, esse método contorna o problema do viés de seleção e possibilita o cálculo da diferença de médias da variável resultado entre os dois grupos (MENEZES FILHO; PINTO, 2017).

No entanto, no caso de programas sociais como o Programa Bolsa Família, não é possível atribuir tratamento às famílias de forma aleatória, pois seu desenho possui um critério que estabelece quem é elegível a ser beneficiário ou não. Dessa maneira, o método de regressão descontínua é uma alternativa utilizada para mitigar o problema do viés de seleção nessas situações.

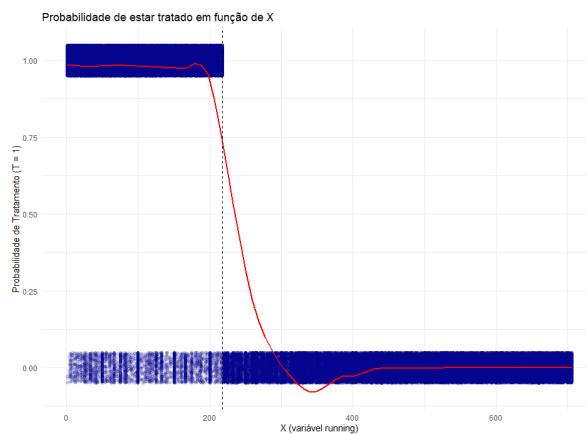
O desenho de regressão descontínua é uma das abordagens comumente utilizadas para avaliação de impacto de programas em que a probabilidade de ser tratado varia de maneira descontínua com base em uma variável observável, chamada de *running variable*, cuja representação será X e um ponto de corte c . Intuitivamente, os indivíduos que possuem o valor da variável observável (X) logo acima do corte ou logo abaixo são muito parecidos, exceto pelo fato de que uns tiveram a “sorte” de serem tratados e outros não. Essa aleatoriedade em torno do ponto de corte é o argumento que define a capacidade do método em contornar o problema de viés de seleção. Assim, se compararmos a média da variável de resultado dos indivíduos que estão logo acima de c com os que estão logo abaixo, encontraremos o efeito médio do tratamento em torno do ponto de corte, ou

seja, o efeito médio local do tratamento (EML) (CUNNINGHAM, 2021; HUNTINGTON-KLEIN, 2021; MENEZES FILHO; PINTO, 2017).

4.1.1 Definição do desenho

Para o Programa Bolsa Família, a renda familiar per capita mensal é considerada a *running variable* X do modelo, visto que determina a participação ou não no tratamento. Para a família ser elegível ao programa ela deve possuir renda per capita mensal igual ou inferior ao ponto de corte c de R\$ 218,00. Como observamos na seção anterior, a participação no programa não é determinística, ou seja, existem famílias que se enquadram nos critérios de elegibilidade, porém, por algum motivo, não participam. Na figura abaixo, também demonstramos essa relação entre a probabilidade de a família ser beneficiária do PBF e a variável de renda familiar per capita. Nota-se, pela curva “vermelha” desenhada no gráfico, que a probabilidade de ser tratado não “salta” de 0 para 1 ao cruzar o ponto de corte, mas se altera à medida que se aproxima dele. Dessa forma, temos um desenho de regressão descontínua comumente chamada de *Fuzzy*, isto é, a atribuição ao tratamento não acontece de forma perfeita no ponto de corte.

Figura 5 - Probabilidade de Tratamento e Renda Familiar Per Capita



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Formalmente, esta atribuição probabilística ao tratamento pode ser escrita como:

$$\lim_{X_i \rightarrow c^+} Pr[T_i | X_i = c] \neq \lim_{X_i \rightarrow c^-} Pr[T_i | X_i = c] \quad (2)$$

Ou seja, a probabilidade condicional é descontínua conforme a *running variable* se aproxima de c no limite (CUNNINGHAM, 2021).

Visto que a adesão ao programa é imperfeita, pode-se classificar os indivíduos em quatro grupos: *compliers*, *always-takers*, *never-takers* e *defiers*. No nosso caso específico, os *compliers* são aqueles que recebem o tratamento se o valor de X for igual ou abaixo de c , e não o recebem se o valor de X estiver acima. Os *always-takers* são os indivíduos que sempre aceitam o tratamento, independentemente do valor de X . Já os *never-takers*, são os indivíduos que sempre recusam o tratamento, seja qual for o valor de X . E os *defiers* são aqueles que recebem o tratamento se o valor de X for maior que c , e não recebem se o valor for menor que c .

Sob as hipóteses de monotonicidade - em que $T_i(X)$ é uma função não decrescente em X -, de exclusão - em que X ao cruzar o ponto de corte não pode impactar Y exceto por meio do recebimento do tratamento -, e a hipótese de continuidade nos resultados potenciais, o efeito médio do tratamento a ser identificado pelo desenho *Fuzzy* será apenas para os *compliers*. Muito semelhante ao caso de variáveis instrumentais, estimaremos um efeito médio local de tratamento (EML). Formalmente podemos escrever:

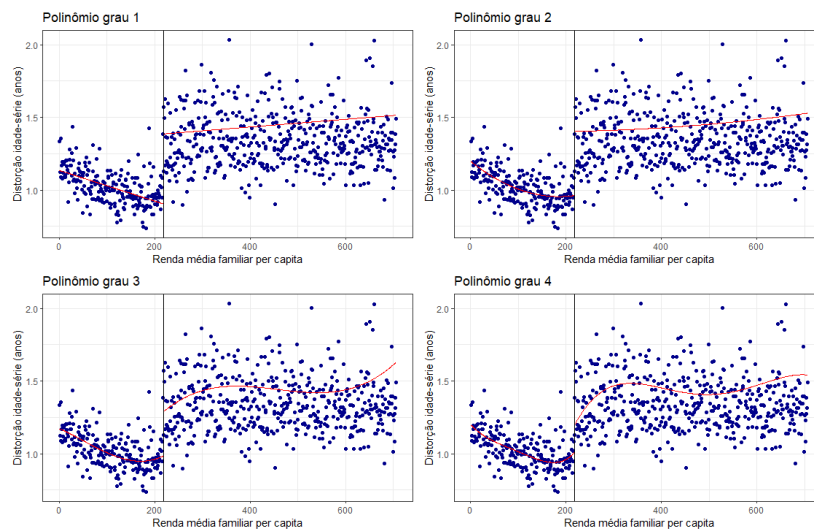
$$EML(c) = \frac{\lim_{X \rightarrow c^+} E[Y|X = c] - \lim_{X \rightarrow c^-} E[Y|X = c]}{\lim_{X \rightarrow c^+} E[T|X = c] - \lim_{X \rightarrow c^-} E[T|X = c]} \quad (3)$$

Isto é, o efeito médio local do tratamento é calculado em torno do ponto de corte como a diferença na variável de resultado à medida que X se aproxima de c , dividido pela diferença na probabilidade de tratamento conforme X se aproxima de c , pela direita e pela esquerda.

4.1.2 Verificação da Descontinuidade

Nossa variável de resultado Y será a quantidade de anos de distorção idade-série que o estudante do ensino fundamental novo possui. Diferentemente de outros trabalhos que utilizaram a defasagem escolar de forma binária, no nosso caso usaremos a variável contínua. Dessa forma, vamos verificar a existência de descontinuidade da variável de resultado ao longo da *running variable* antes de estimarmos os modelos.

Figura 6 - Verificação Descontinuidade Anos de Defasagem vs Renda Per Capita



Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Na figura acima, apresentamos quatro gráficos que fazem a relação entre anos de defasagem escolar e renda familiar per capita. Foram utilizadas regressões polinomiais locais para a estimação das curvas⁸, em que variamos o grau do polinômio a fim de analisarmos o ajuste aos dados. Visualmente podemos perceber uma descontinuidade na variável Y , anos de defasagem escolar, em torno do ponto de corte de R\$ 218,00. Embora útil como evidência gráfica, os saltos observados não podem ser interpretados como o efeito definitivo, pois não substituem as estimativas formais (CATTANEO e TITIUNIK, 2022).

A partir da informação de que o desenho de regressão descontínua é do tipo *Fuzzy* - visto que a probabilidade de participar do Programa Bolsa Família não é determinística em relação à renda familiar per capita - e da sugestão gráfica da existência de uma descontinuidade entre a quantidade de anos de defasagem escolar e a renda per capita, vamos estimar o efeito médio local do tratamento na próxima seção. Para isso, empregaremos o pacote *rdrobust* desenvolvido por Calonico et al (2015a) para o *software* R. O algoritmo desenvolvido pelos autores nos permite estimar o EML por meio de regressão polinomial local além da aplicação de erros-padrão robustos à heterocedasticidade.

⁸ A estimação foi executada no software R por meio do pacote *rdrobust* desenvolvido por Calonico et al (2015a). Para a construção dos gráficos, as *bins* foram selecionadas por meio do método de *mimicking variance evenly-spaced* usando regressão polinomial e a função Kernel selecionada foi uniforme (CALONICO ET AL 2015b).

5. RESULTADOS

O objetivo deste trabalho é verificar se o Programa Bolsa Família possui algum efeito significativo no indicador educacional de distorção idade-série. Estudos anteriores específicos sobre o tema, que foram realizados para o caso brasileiro, mostraram resultados distintos. O primeiro deles, feito por Ribeiro e Cacciamali (2012), não demonstrou efeito significativo do programa sobre o indicador educacional. Os autores utilizaram dados provenientes da PNAD do ano de 2006. Possivelmente, a vigência do programa ainda não tinha refletido impacto sobre a defasagem escolar. Questionando esse resultado, Araújo, Frio e Alves (2021) demonstraram, por meio de dados da PNAD de 2016 a 2019, efeitos negativos e significativos do PBF sobre a distorção idade-série. Contudo, eles não encontraram efeitos significativos para a região Sul. Nos tópicos a seguir, apresentaremos os resultados estimados a partir do desenho de regressão descontínua *Fuzzy*, acompanhados dos testes de robustez. Ainda, estimamos um modelo de regressão logística binária para complementar a inferência causal.

5.1. Análise Contrafactual

Utilizaremos um método não-paramétrico, por meio de regressões polinomiais locais, para a estimação do desenho de regressão descontínua do tipo *Fuzzy*. Essa abordagem é preferível a métodos polinomiais globais pois possui boas propriedades de fronteira e evita problemas como sobreajuste e falta de robustez (CALONICO et al, 2015a; CATTANEO e TITIUNIK, 2022).

5.1.1 Estimação do efeito

A estimação por regressão polinomial local permite capturar curvaturas na relação entre X e Y dentro de uma janela local em torno do ponto de corte, com a possibilidade de ponderação das observações na proximidade do corte. Nesse sentido, o método exige três escolhas principais: a definição da ordem p do polinômio; a função kernel $K(\cdot)$, que ditará o peso das observações; e a escolha da *bandwidth* h , que determinará a vizinhança em torno do ponto de corte c . Para isso, lançaremos mão do pacote *rdrobust* do *software* R desenvolvido por Calonico et al (2015a). A função *rdrobust(.)* desse pacote permite a implementação de estimadores polinomiais locais com a seleção de *bandwidths* que otimizam o erro quadrático médio (EQM) das estimações, além da possibilidade de construção de intervalos de confiança corrigidos por viés⁹.

Executamos a rotina fazendo simulações com polinômios de ordem $p = \{1, 2, 3, 4\}$, combinados com diferentes funções kernel (Triangular, Uniforme e *Epanechnikov*) e com dois procedimentos de escolha de *bandwidth* EQM-ótima (uma comum em ambos os lados do ponto de corte e outra com duas janelas distintas). Os resultados das 24 simulações podem ser encontrados no Apêndice. A primeira observação que podemos fazer a respeito dos resultados dos modelos é que todos foram negativos. Se considerarmos os modelos significativos a 1%, temos estimativas no intervalo de -0,234 a -0,417. Ou seja, sugere-se que os estudantes pertencentes a famílias beneficiárias do Programa Bolsa Família possuem entre 0,23 e 0,42 ano a menos de defasagem escolar.

A fim de selecionar um modelo para seguirmos nossa análise, Cattaneo e Titiunik (2022) sugerem que a ordem do polinômio seja baixa para evitar algum problema de sobreajuste. Quanto à escolha da função kernel, eles apontam que as funções do tipo triangular e uniforme são as mais utilizadas¹⁰. Porém, quando a função triangular é acompanhada de uma escolha de *bandwidth* EQM-ótima, o estimador pontual do desenho de regressão descontínua será consistente e EQM-ótimo (CATTANEO et al, 2020a). Assim, selecionamos o modelo que chamamos de *m112*, cujo polinômio é de primeira ordem, a função kernel é triangular e a *bandwidth* é EQM-ótima com bandas de larguras

⁹ Para uma comparação entre os pacotes *rdrobust*, *rdd* e *rddtools* desenvolvidos para o *software* R para análise de desenhos de regressão descontínua, conferir o trabalho de THOEMMES et al (2017).

¹⁰ Lee e Lemieux (p. 319, 2010) argumentam que a escolha do kernel não teria tanto impacto na prática nas estimações.

diferentes. Conforme Tabela 4, as *bandwidths* escolhidas foram 45,139 para o lado esquerdo do ponto de corte e 56,385 para o lado direito.

Tabela 4 - Observações Segmentadas Modelo

Atributo	Esquerda do corte	Direita do corte
Nº Observações	203.437	75.636
Nº Efetivo de Obs.	33.806	7.113
Ordem est. (p)	1	1
Ordem viés (q)	2	2
BW est. (h)	45,139	56,385
BW viés (b)	117,171	241,759
Rho (h/b)	0,385	0,233
Observações Únicas	217	489

Fonte: Elaboração própria.

A Tabela 5, a seguir, apresenta os resultados das estimativas do primeiro-estágio executadas pela função *rdrobust*. Nessa etapa é estimado a relação entre a *running variable* - renda familiar per capita - e a variável de beneficiário do PBF. Pode-se notar que existe uma relação negativa, isto é, quanto menor a renda maior a chance de participar do programa, como foi visto também na Figura 5 na seção anterior.

Tabela 5 - Estimativas Primeiro-Estágio

Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-0,734	0,004	-198,765	0.000	[-0,741; -0,726]
Correção-viés	-0,731	0,004	-198,157	0.000	[-0,739; -0,724]
Robusto	-0,731	0,004	-192,161	0.000	[-0,739; -0,724]

Fonte: Elaboração própria.

Na Tabela 6, abaixo, temos os resultados do segundo estágio, isto é, as estimativas do efeito médio local do tratamento para os *compliers* no nosso desenho de regressão descontínua do tipo *Fuzzy*. Observamos que o coeficiente encontrado é estatisticamente significativo e foi de -0,234, ou seja, crianças ou adolescentes do ensino fundamental novo de escolas públicas presentes no CadÚnico, em torno do ponto de corte, possuem 0,234 ano a menos de defasagem escolar se forem beneficiárias do Programa Bolsa Família.

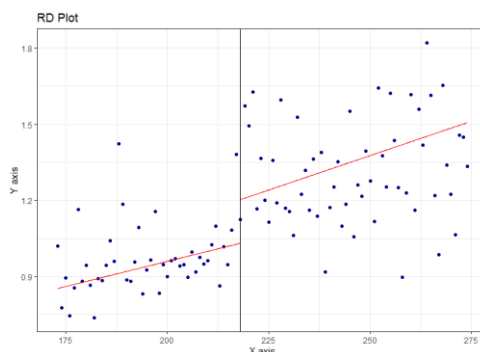
Tabela 6 – Estimativas do efeito do tratamento

Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-0,234	0,057	-4,077	0.000	[-0,346; -0,121]
Correção-viés	-0,221	0,057	-3,856	0.000	[-0,334; -0,109]
Robusto	-0,221	0,059	-3,775	0.000	[-0,336; -0,106]

Fonte: Elaboração própria.

Ilustramos na Figura 7 essa descontinuidade presente nos anos de defasagem escolar em torno do ponto de corte de R\$ 218,00 utilizando os parâmetros estabelecidos na estimação.

Figura 7 - Efeito do Programa Bolsa Família Sobre a Distorção Idade-Série



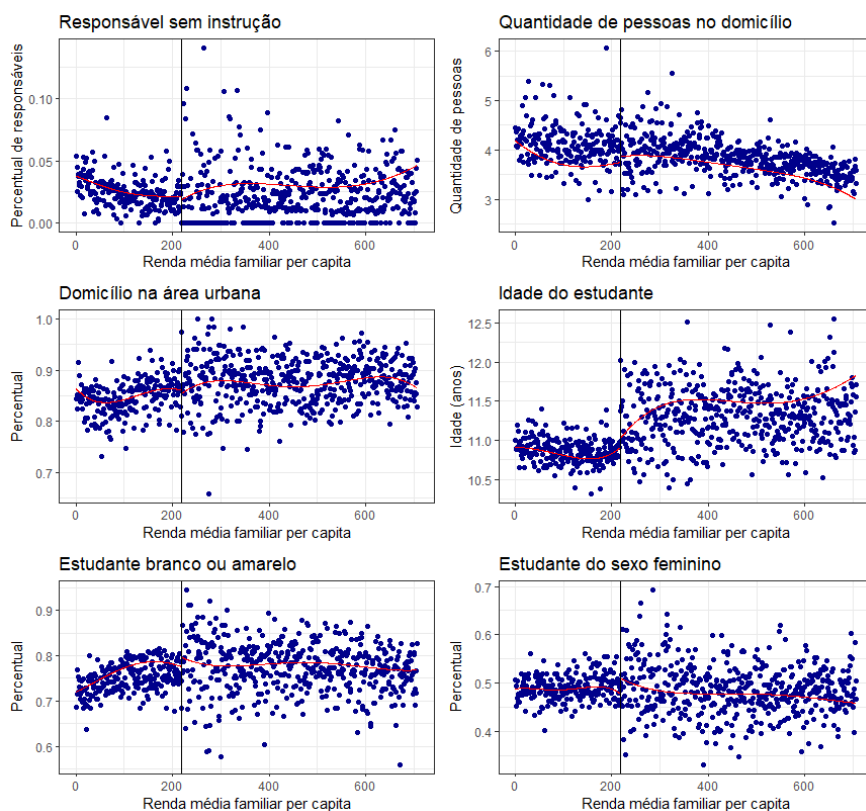
Fonte: Elaboração própria.

Percebe-se o efeito negativo e significativo do programa sobre a distorção idade-série dos estudantes do ensino fundamental novo de escolas públicas que fazem parte do Cadastro Único no Rio Grande do Sul. O resultado difere do encontrado pelo trabalho de Araújo, Frio e Alves (2021), que não identificou efeito significativo para região Sul, e de Ribeiro e Cacciamali (2012), cujo impacto do programa não foi observado.

5.1.2 Análise de Covariáveis

Após a estimação do efeito do tratamento precisamos realizar alguns testes para validar o resultado encontrado (LEE e LEMIEUX, 2010). O primeiro deles diz respeito à análise de possíveis “saltos” em covariáveis. O objetivo é verificar se existe alguma outra variável que apresente descontinuidade no ponto de corte. Para isso, realizaremos uma análise gráfica e uma estimação para cada covariável, semelhante aos exercícios anteriores. Vamos utilizar a função *rdrobust* empregada anteriormente, mantendo os argumentos como *default* da própria rotina: polinômio global de ordem 4 para o gráfico; e polinômio de ordem 1 para a estimação do modelo. Na figura abaixo, podemos visualizar os gráficos para cada uma das variáveis.

Figura 8 - Covariáveis vs Renda per capita



Fonte: Elaboração própria.

Visualmente as covariáveis parecem não apresentar descontinuidade em torno do ponto de corte de renda de R\$ 218,00. Para confirmar, a Tabela 7 expõe os resultados das estimações. Notamos que para nenhuma das covariáveis as estimativas foram significativas a 5%.

Tabela 7 - Estimativas das covariáveis

Covariável	Polinômio	Estimativa	Erro Padrão	Estatística z	p-valor
Responsável sem instrução	1	-0.012	0.009	-1.438	0.150
Qtde pessoas na família	1	-0.090	0.131	-0.686	0.492
Urbano	1	0.065	0.037	1.754	0.079
Idade	1	-0.395	0.255	-1.546	0.122
Cor ou raça	1	0.025	0.035	0.726	0.468
Feminino	1	-0.028	0.041	-0.665	0.506

Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

5.1.3 Teste Placebo *Cutoff*

Assim como não deveria haver efeito no ponto de corte sobre as covariáveis, devemos esperar que não haja efeito na nossa variável resultado de interesse em um ponto de corte escolhido arbitrariamente. Conforme Imbens e Lemieux (2008), uma forma de testar esse comportamento é olharmos para um lado da descontinuidade, calcular a mediana da *running variable* e utilizar o resultado encontrado como ponto de corte fictício c' . No nosso caso, calculamos a mediana de renda familiar per capita abaixo do ponto de corte de R\$ 218,00 e encontramos um c' igual a R\$ 100,00. O resultado encontrado, conforme tabela abaixo, sugere um efeito não significativo, ou seja, não é possível observar uma descontinuidade dos anos de defasagem no ponto de corte de R\$ 100,00.

Tabela 8 - Teste Placebo Cutoff – Observações segmentadas

Atributo	Esquerda do corte	Direita do corte
Nº Observações	94.785	184.288
Nº Efetivo de Obs.	25.209	66.665
Ordem est. (p)	1	1
Ordem viés (q)	2	2
BW est. (h)	25,691	64,127
BW viés (b)	46,211	70,253
Rho (h/b)	0,556	0,913
Observações Únicas	99	607

Tabela 9 - Teste Placebo Cutoff - Estimativas

Estimativas Primeiro Estágio					
Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-0,002	0,003	-0,534	0,594	[-0,007; 0,004]
Estimativas Efeito do Tratamento					
Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-40,046	75,833	-0,528	0,597	[-188,677; 208,584]

Fonte: Elaboração própria.

5.1.4 Análise de Sensibilidade à *Bandwidth*

O teste de sensibilidade à *bandwidth* será realizado a partir do modelo *m112* estimado anterior. Para tal, utilizaremos as duas *bandwidths* estimadas, 45,14 para esquerda e 56,39 para direita. A partir disso, primeiro simulamos a estimação dividindo por 2 as *bandwidths* ótimas e depois multiplicando por 2. Para o primeiro caso, a estimativa encontrada foi -0,260 ano, e no segundo caso, obtivemos uma estimativa de -0,361 ano, ambas significativas a 5% (o resultado completo das estimações pode ser encontrado no Apêndice). Esse resultado indica robustez nos achados principais.

Tabela 10 – Sensibilidade à Bandwidth

Dividindo por 2					
Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-0,260	0,102	-2,545	0,011	[-0,446; -0,060]
Multiplicando por 2					
Método	Coefficiente	Erro Padrão	z	P > z	[95% I.C.]
Convencional	-0,361	0,035	-10,229	0,000	[-0,431; -0,292]

Fonte: Elaboração própria

5.1.5 Análise de Manipulação

Uma das ameaças ao desenho de regressão descontínua é a possibilidade de os indivíduos ajustarem suas situações em relação a *running variable* a fim de se enquadrarem para o recebimento ou não do tratamento (CATTANEO e TITIUNIK, 2022). Para verificar esse comportamento, McCrary (2008) sugere a estimação de funções densidade da *running variable* em cada lado do ponto de corte. Caso haja evidência de descontinuidade na densidade, o teste sugere uma possível manipulação na variável X . Para nosso caso, utilizaremos uma versão mais atualizada desenvolvida por CATTANEO et al (2020b) em que os autores propõem um teste de manipulação por meio de estimadores de

densidade polinomial local. A aplicação desse teste foi realizada por meio da função *rddensity(.)* do pacote *rddensity*¹¹ desenvolvido pelos mesmos autores para o *software* R. Conforme os resultados apresentados na tabela a seguir, rejeitamos a hipótese nula de que a distribuição da renda familiar per capita é contínua no ponto de corte. Isso sugere que os indivíduos possam ajustar sua renda a fim de participarem do tratamento.

Tabela 11 - Análise de manipulação

Atributo	Esquerda do corte	Direita do corte
Nº Observações	203.437	75.636
Nº Efetivo de Obs.	20.906	2.378
Ordem est. (p)	2	2
Ordem viés (q)	3	3
BW est. (h)	23	23
Método	T	P > T
Robusto	-49,6877	0,00

Fonte: Elaboração própria

Apesar do resultado sugerido pelo teste acima, a averiguação cadastral realizada periodicamente no Cadastro Único (descrita na seção 3.1.1) tem por objetivo mitigar possíveis inconsistências, principalmente no que se refere a renda declarada, e evitar algum tipo de alocação indevida (BRASIL, 2025b). Nesse sentido, as famílias teriam capacidade reduzida para manipular de forma perfeita a sua renda familiar. De qualquer forma, tentamos solucionar a descontinuidade apontada pelo teste por meio da abordagem sugerida pelo trabalho de Gerard et al (2020). No entanto, a estimativa ainda está pendente devido a problemas na execução da rotina computacional (pacote *rdbounds* para o *software* R).

5.2. Análise Complementar

A fim de complementar nosso resultado encontrado pela regressão descontínua *Fuzzy*, iremos estimar um modelo de regressão logística binária. Esse método é comumente utilizado quando possuímos uma variável de resultado binária.

A variável dependente será a variável dicotômica de distorção idade-série que utiliza a métrica determinada pelo INEP de dois anos ou mais. As variáveis independentes que selecionamos foram relacionadas no tópico de estatísticas descritivas. Além da variável que indica se o indivíduo é beneficiário do PBF (variável Bolsa Família), selecionamos características pessoais do indivíduo, como sexo e cor ou raça (Feminino; Branco ou Amarelo), local do domicílio (Urbano), renda familiar per capita, escolaridade do responsável familiar (Responsável Fundamental Incompleto, Fundamental Completo, Médio Incompleto, Médio Completo, Superior Incompleto ou Mais)¹², informação sobre acesso à rede geral de distribuição de água (Sem Rede Geral Distribuição de Água) e sobre esgotamento sanitário (Esgotamento Sanitário Inadequado).

Na tabela 7, observamos que o resultado da estimação do modelo de regressão logística binária aponta para uma relação negativa e estatisticamente significativa entre pertencer a uma família beneficiária do Bolsa Família e estar em uma situação de distorção idade-série.

Tabela 12 - Resultados do Modelo Logit

	Variável Dependente:
	Em Distorção Idade-série
Bolsa Família	-0,553*** (0,021)
Feminino	-0,266*** (0,011)

¹¹ O pacote *rddensity* é sugerido por CUNNINGHAM (2021).

¹² A referência é a variável “Responsável Sem Instrução”.

Branco ou Amarelo	-0,234*** (0,012)
Urbano	0,201*** (0,018)
Renda Familiar Per Capita	-0,0001* (0,0001)
Responsável Fundamental Incompleto	-0,351*** (0,026)
Responsável Fundamental Completo	-0,901*** (0,029)
Responsável Médio Incompleto	-1,164*** (0,030)
Responsável Médio Completo	-1,415*** (0,029)
Responsável Superior Incompleto ou Mais	-1,527*** (0,049)
Sem Rede Geral Distribuição de Água	0,001 (0,016)
Esgotamento Sanitário Inadequado	0,027** (0,013)
Constante	-0,320*** (0,040)
Observações	279.073
Log Likelihood	-117.711,20
Nota:	*p < 0,1; **p < 0,05; ***p < 0,01
Fonte:	CadÚnico, Janeiro de 2024. Elaboração própria.

Como a estimação realizada apresenta os resultados em escala logarítmica, precisamos transformá-los para facilitar a interpretação, isto é, precisamos fazer a exponenciação das variáveis do modelo. Dessa forma, obtemos a Razão de Chances (OR – Odds Ratio, em inglês) das variáveis independentes e seus respectivos intervalos de confiança.

Tabela 13 - Razão de Chances (OR)

	OR	2.5 %	97.5 %
(Intercepto)	0.674	0.623	0.729
Bolsa Família	0.575	0.552	0.599
Feminino	0.766	0.751	0.783
Branco ou Amarelo	0.791	0.773	0.810
Urbano	1.223	1.181	1.266
Renda Familiar Per Capita	1.000	1.000	1.000
Responsável Fundamental Incompleto	0.704	0.669	0.741
Responsável Fundamental Completo	0.406	0.384	0.430
Responsável Médio Incompleto	0.312	0.294	0.331
Responsável Médio Completo	0.243	0.230	0.257
Responsável Superior Incompleto ou Mais	0.217	0.198	0.239
Sem Rede Geral Distribuição de Água	1.001	0.970	1.032
Esgotamento Sanitário Inadequado	1.027	1.002	1.053

Fonte: CadÚnico, Janeiro de 2024. Elaboração própria.

Conforme a tabela 8 acima, podemos inferir que os estudantes pertencentes a famílias beneficiárias do Programa Bolsa Família possuem 42,5% $((0,575 - 1) * 100)$ menos chances de estarem em uma situação de distorção idade-série. Esse resultado corrobora o efeito negativo do programa sobre a defasagem escolar encontrado pela regressão descontínua *Fuzzy*.

6. CONSIDERAÇÕES FINAIS

O presente estudo buscou avaliar o efeito do Programa Bolsa Família sobre o indicador educacional de distorção idade-série para crianças e adolescentes do ensino fundamental novo de escolas públicas presentes na base de dados do Cadastro Único no estado do Rio Grande do Sul. Para

tanto, utilizou-se o desenho de Regressão Descontínua *Fuzzy* a fim de mitigar o possível viés de seleção.

O resultado obtido aponta para um efeito significativo do PBF sobre a distorção idade-série. Conforme a estimação do método, o efeito médio local do tratamento para os *compliers* foi de -0,234, ou seja, crianças ou adolescentes do ensino fundamental novo de escolas públicas presentes no CadÚnico, em torno do ponto de corte, possuem 0,23 ano a menos de defasagem escolar se forem beneficiárias do Programa Bolsa Família. Nossa análise apresenta resultado distinto do trabalho de Ribeiro e Cacciamali (2012), que não observou efeito significativo do programa sobre o indicador educacional de defasagem, e do trabalho de Araújo, Frio e Alves (2021), que não apresentou resultados significativos para a região Sul do Brasil.

Dessa forma, nosso trabalho contribui para a discussão a respeito da eficácia do PBF sobre indicadores educacionais ao utilizar a base de dados do Cadastro Único ainda não explorada para a investigação específica sobre distorção idade-série no estado do Rio Grande do Sul.

REFERÊNCIAS

- ARAÚJO, Jevuks Matheus; FRIO, Gustavo Saraiva; ALVES, Pedro Jorge Holanda. O efeito do Bolsa Família sobre a distorção idade-série. *Estudos Econômicos* (São Paulo), v. 51, n. 02, p. 343-371, 2021.
- BARROS, Ricardo Paes de; CARVALHO, Mirela de; MENDONÇA, Rosane Silva Pinto de. Sobre as utilidades do Cadastro Único. 2009.
- BRASIL. Lei nº 8.742, de 7 de dezembro de 1993. Dispõe sobre a organização da assistência social e dá outras providências. Brasília, DF: 1993.
- _____. Ministério de Educação e Cultura. LDB - Lei nº 9394/96, de 20 de dezembro de 1996. Estabelece as Diretrizes e Bases da Educação Nacional. Brasília: MEC, 1996.
- _____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Dicionário de indicadores educacionais: fórmulas de cálculo. Brasília: Inep, 2004. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/dicionario_de_indicadores_educacionais_formulas_de_calculo.pdf>. Acesso em 10 jul. 2024.
- _____. Decreto nº 11.016, de 29 de março de 2022. Regulamenta o Cadastro Único para Programas Sociais do Governo Federal. Brasília, DF: 2022.
- _____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Censo Escolar da Educação Básica 2024: Resumo Técnico. Brasília, 2025a. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/resumo_tecnico_censo_escolar_2024.pdf>. Acesso em 29 mai. 2025.
- _____. Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome. Lei nº 14601/23, de 19 de junho de 2023. Institui o Programa Bolsa Família. Brasília: MEC, 2023b.
- _____. Ministério do Desenvolvimento Social. Instrução Normativa Conjunta nº 1 SAGICAD/SENARC/SNAS/MDS, de 26 de fevereiro de 2025. Diário Oficial da União: seção 1, Brasília, DF. 2025b.
- CALONICO, Sebastian; CATTANEO, Matias D.; TITIUNIK, Rocio. Rdrobust: an R package for robust nonparametric inference in regression-discontinuity designs. *R J.*, v. 7, n. 1, p. 38, 2015a.
- CALONICO, Sebastian; CATTANEO, Matias D.; TITIUNIK, Rocio. Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, v. 110, n. 512, p. 1753-1769, 2015b.
- CATTANEO, M. D., IDROBO, N., TITIUNIK, R.. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge: Cambridge University Press, 2020a.
- CATTANEO, Matias D.; JANSSON, Michael; MA, Xinwei. Simple local polynomial density estimators. *Journal of the American Statistical Association*, v. 115, n. 531, p. 1449-1455, 2020b.

- CATTANEO, Matias D.; TITIUNIK, Rocio. Regression discontinuity designs. *Annual Review of Economics*, v. 14, n. 1, p. 821-851, 2022.
- CUNHA, F. E HECKMAN, J. The Technology of Skill Formation. *AEA Papers and Proceedings*, vol. 97, n. 2, pp. 31-47. 2007.
- CUNNINGHAM, Scott. *Causal inference: The mixtape*. Yale University Press, 2021.
- GERARD, François; ROKKANEN, Miikka; ROTHE, Christoph. Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quantitative Economics*, v. 11, n. 3, p. 839-870, 2020.
- HERMETO, Ana Maria; ROMERO, Julio Alfredo Racchumi. Avaliação de Impacto do Programa Bolsa Família sobre Indicadores Educacionais: uma abordagem de regressão descontínua. Em: Encontro Nacional de Economia, ANPEC, 37, Foz do Iguaçu, 2009. Anais do XXXVII Encontro Nacional de Economia. Disponível em: <<http://www.anpec.org.br/encontro2009/inscricao.on/arquivos/000-c1356554aa37dfe2e7daa6708e30ba7d.doc>>. Acesso em: 22 de abril de 2011.
- HUNTINGTON-KLEIN, Nick. *The effect: An introduction to research design and causality*. Chapman and Hall/CRC, 2021.
- LEE, David S.; LEMIEUX, Thomas. Regression discontinuity designs in economics. *Journal of economic literature*, v. 48, n. 2, p. 281-355, 2010.
- MACHADO, Danielle Carusi; GONZAGA, Gustavo. O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil. *Revista Brasileira de Economia*, v. 61, p. 449-476, 2007.
- MCCRARY, Justin. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, v. 142, n. 2, p. 698-714, 2008.
- MENEZES FILHO, Naercio Aquino et al. *Avaliação econômica de projetos sociais*. Fundação Itaú Social, 2017.
- NETO, Robinson Mattos; URANGA, Paulo Ricardo Ricco; BAGOLIN, Izete Pengo. Determinantes de defasagem idade-série: Uma análise espacial para o Rio Grande do Sul com base nos dados do SAERS para o ano de 2016. *Anais ANPEC SUL 2019, Brasil.*, 2019.
- RIBEIRO, Rosana; CACCIAMALI, Maria Cristina. Impactos do Programa Bolsa Família sobre os indicadores educacionais. *Revista Economia*, v. 13, n. 2, p. 415-446, 2012.
- RIBEIRO, Felipe Garcia; SHIKIDA, Claudio; HILLBRECHT, Ronald Otto. Bolsa Família: Um survey sobre os efeitos do programa de transferência de renda condicionada do Brasil. *Estudos Econômicos (São Paulo)*, v. 47, n. 4, p. 805-862, 2017.
- THOEMMES, Felix; LIAO, Wang; JIN, Ze. The analysis of the regression-discontinuity design in R. *Journal of Educational and Behavioral Statistics*, v. 42, n. 3, p. 341-360, 2017.

APÊNDICE A - MODELOS ESTIMADOS

Modelo	Polinômio	Kernel	Seleção Bandwidth	Bandwidth (esquerda/direita)	Estimativa	Erro Padrão	Estatística z	p-valor
m111	1	Triangular	mserd	26.55 / 26.55	-0.285	0.105	-2.713	0.007
m112	1	Triangular	msetwo	45.14 / 56.39	-0.234	0.057	-4.077	0.000
m121	1	Epanechnikov	mserd	19.78 / 19.78	-0.294	0.114	-2.577	0.009
m122	1	Epanechnikov	msetwo	41.63 / 34.98	-0.369	0.071	-5.229	0.000
m131	1	Uniform	mserd	15.53 / 15.53	-0.241	0.115	-2.098	0.036
m132	1	Uniform	msetwo	25.84 / 17.18	-0.342	0.106	-3.226	0.001
m211	2	Triangular	mserd	36.67 / 37.67	-0.224	0.139	-1.617	0.106
m212	2	Triangular	msetwo	51.06 / 70.07	-0.195	0.087	-2.230	0.026
m221	2	Epanechnikov	mserd	33.06 / 33.06	-0.178	0.138	-1.293	0.196
m222	2	Epanechnikov	msetwo	45.52 / 57.24	-0.292	0.093	-3.139	0.002

m231	2	Uniform	mserd	33.65 / 33.65	-0.189	0.125	-1.521	0.128
m232	2	Uniform	msetwo	42.93 / 44.96	-0.365	0.102	-3.588	0.000
m311	3	Triangular	mserd	75.66 / 75.66	-0.384	0.120	-3.209	0.001
m312	3	Triangular	msetwo	91.99 / 111.83	-0.184	0.088	-2.098	0.036
m321	3	Epanechnikov	mserd	67.46 / 67.46	-0.417	0.123	-3.381	0.001
m322	3	Epanechnikov	msetwo	82.43 / 96.72	-0.246	0.096	-2.567	0.010
m331	3	Uniform	mserd	67.9 / 67.9	-0.414	0.113	-3.655	0.000
m332	3	Uniform	msetwo	85.6 / 79.31	-0.295	0.100	-2.943	0.003
m411	4	Triangular	mserd	95.32 / 95.32	-0.407	0.141	-2.872	0.004
m412	4	Triangular	msetwo	96.59 / 130.27	-0.197	0.112	-1.767	0.077
m421	4	Epanechnikov	mserd	99.9 / 99.9	-0.406	0.128	-3.180	0.001
m422	4	Epanechnikov	msetwo	103.55 / 119.62	-0.292	0.111	-2.615	0.009
m431	4	Uniform	mserd	99.06 / 99.06	-0.356	0.119	-2.994	0.003
m432	4	Uniform	msetwo	87.6 / 106.44	-0.321	0.116	-2.765	0.006

Fonte: Elaboração própria.

mserd = bandwidth EQM-ótima igual em ambos os lados.

msetwo = bandwidth EQM-ótima diferente para cada lado.

APÊNDICE B - ANÁLISE DE SENSIBILIDADE À *BANDWIDTH*

Dividindo bandwidth por 2

Estimates – Dividing by 2

Attribute	Left	Right			
Number of Observations	203.437	75.636			
Effective Number of Obs.	20.434	2.666			
Order est. (p)	1	1			
Order bias (q)	2	2			
BW est. (h)	22.570	28.193			
Estimates of treatment effect					
Method	Coefficient	Standard Error	z	P > z	[95% I.C.]
Conventional	-0,260	0,102	-2,545	0,011	[-0,446; -0,060]

Multiplicando bandwidth por 2

Estimates – Multiplying by 2

Attribute	Left	Right			
Number of Observations	203.437	75.636			
Effective Number of Obs.	70.606	16.493			
Order est. (p)	1	1			
Order bias (q)	2	2			
BW est. (h)	90.279	112.770			
Estimates of treatment effect					
Method	Coefficient	Standard Error	z	P > z	[95% I.C.]
Conventional	-0,361	0,035	-10,229	0,000	[-0,431; -0,292]