



Contemporânea

Contemporary Journal

3(8): 12437-12473, 2023

ISSN: 2447-0961

Artigo

PROGRAMA BOLSA FAMÍLIA E OS INDICADORES EDUCACIONAIS NO ESTADO DO CEARÁ

BOLSA FAMILIA PROGRAM AND EDUCATIONAL INDICATORS IN THE STATE OF CEARÁ

DOI: 10.56083/RCV3N8-140

Recebimento do original: 17/07/2023

Aceitação para publicação: 16/08/2023

Kaliny Kélvia Siqueira Lima

Doutora em Desenvolvimento e Meio Ambiente pela Universidade Federal do Ceará (UFC)

Instituição: Universidade Estadual Vale do Acaraú (UVA)

Endereço: Avenida Padre Francisco Sadoc de Araújo, 850, Campus Betânia, Sobral – CE

E-mail: kaliny_pessoa@uvanet.br

Joyciane Coelho Vasconcelos

Doutora em Desenvolvimento e Meio Ambiente pela Universidade Federal do Ceará (UFC)

Instituição: Centro Universitário Inta (UNINTA)

Endereço: Rua Antônio Rodrigues Magalhães, 359, Dom Expedito, Sobral – CE, CEP: 62050-100

E-mail: joyciane.c.v@gmail.com

Gerlânia Maria Rocha Sousa

Doutora em Desenvolvimento e Meio Ambiente pela Universidade Federal do Ceará (UFC)

Instituição: Universidade do Estado do Rio Grande do Norte (UERN)

Endereço: Avenida Professor Antônio Campos, s/n, BR 110, Km 48, Presidente Costa e Silva, Mossoró – RN

E-mail: gerlaniarocha@gmail.com

Antônia Francivan Vieira Castelo Branco

Doutora em Desenvolvimento e Meio Ambiente pela Universidade Federal do Ceará (UFC)

Instituição: Escola Estadual de Ensino Médio Doutor Júlio de Carvalho

Endereço: Rua José Joaquim de Carvalho, 496, Centro, Viçosa do Ceará – CE, CEP: 62300-000

E-mail: franvcb@hotmail.com

RESUMO: Neste trabalho será abordado como o Programa Bolsa Família - PBF vem contribuindo para mudar a educação. Sendo a educação um dos principais temas de discussões e reflexões. Sua importância se deve ao fato de abranger temas sociais, econômicos, políticos e culturais. O trabalho tem como objetivo geral estudar a análise multivariada, envolvendo a relação

12437



entre aspectos socioeconômicos e variáveis educacionais do Estado do Ceará e; objetivos específicos determinar a relação quantitativa entre as variáveis e identificar os fatores representativos da educação nos municípios do Ceará, agrupando-os a partir do cálculo de um índice educacional; verificar se os indicadores socioeconômicos influenciam os indicadores educacionais nos municípios do Estado do Ceará e identificar as variáveis que mais influenciam os indicadores educacionais nos municípios cearenses. A metodologia adotada se utilizou de dados secundários extraídos do Portal Brasil, Data Social, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Instituto de Pesquisa Econômica Aplicada (IPEA). Os dados foram coletados em escala Municipal para o Estado do Ceará, tendo como referência o período o ano de 2015. O estudo mostrou como resultado principal que a taxa de aprovação do ensino médio não possui impacto sob a variável dependente, porém, com o aumento da quantidade de crianças e adolescentes de 6 a 17 anos, acompanhados na frequência escolar teve impacto negativo de 0,024 na taxa de abandono no ensino médio e, concluiu que através da metodologia da análise discriminante os municípios alcançaram notas acima da média do IDEB, tendo como variável de valor estatístico significativa, a taxa de aprovação do 1º ao 5º ano, como a que melhor que discrimina os grupos em estudo. No caso, 67,4% das notas do IDEB foram classificados corretamente, segundo o repasse do PBF, em 2015. As variáveis independentes, indicadas pelo modelo de regressão logística foram bem eficientes em prever as notas do IDEB que não seriam acima da média, percentual de 56,3% dos municípios e, mais eficientes em prever as notas acima da média dos Municípios do Estado do Ceará, chegando a 76,3% de acertos. Ou seja, houve uma melhora no percentual de acertos em relação ao modelo base, significando dizer que o valor do repasse do PBF, em 2015, contribuiu para que um percentual maior dos municípios do Estado do Ceará atingissem notas acima da média do IDEB, configurando, assim, uma melhora no ensino da educação fundamental. E, por fim, segundo a análise de correlação canônica, admite-se que existe correlação entre os grupos das variáveis, o que valida o modelo, corroborando que a análise fatorial revelou, considerou as variáveis taxa de aprovação do ensino fundamental I e as notas do IDEB como variáveis mais representativas dentro do modelo de correlação canônica; estas que já contemplam indicadores encontrados entre os três fatores mais importantes, segundo a metodologia de regressão logística e análise discriminante, endossando que as variáveis investigadas no estudo melhor explica os impactos positivos do PBF na educação de ensino fundamental.

PALAVRAS-CHAVE: Programa Bolsa Família, Ensino Fundamental, Econometria.



ABSTRACT: In this work will be addressed how the Bolsa Família Program - BFP has been contributing to change education. Education is one of the main topics for discussion and reflection. Its importance is due to the fact that it covers social, economic, political and cultural themes. The work has as general objective to study multivariate analysis, involving the relationship between socioeconomic aspects and educational variables of the State of Ceará; specific objectives to determine the quantitative relationship between the variables and identify the factors representative of education in the municipalities of Ceará, grouping them from the calculation of an educational index; to verify if the socioeconomic indicators influence the educational indicators in the municipalities of the State of Ceará and to identify the variables that most influence the educational indicators in the municipalities of Ceará. The methodology used secondary data extracted from the Portal Brasil, Data Social, Instituto Nacional de Estudos e Pesquisas Educacional Anísio Teixeira (INEP), Instituto de Pesquisa Econômica Aplicada (IPEA). The data were collected on a Municipal scale for the State of Ceará, having as reference the period the year 2015. The study showed as the main result that the approval rate of high school has no impact on the dependent variable, however, with the increase in the number of children and adolescents aged 6 to 17 years, accompanied in school attendance had a negative impact of 0.024 on the drop-out rate in high school and, concluded that through the methodology of the discriminant analysis the municipalities achieved scores above the IDEB average, having as a significant statistical value variable, the approval rate from the 1st to the 5th year, as the one that best discriminates the groups under study. In this case, 67.4% of the IDEB notes were classified correctly, according to the PBF transfer, in 2015. The independent variables, indicated by the logistic regression model, were very efficient in forecasting IDEB scores that were not above average, a percentage of 56.3% of the municipalities, and more efficient in forecasting the above average scores of the municipalities of the State of Ceará, reaching 76.3% of hits. That is, there was an improvement in the percentage of hits in relation to the base model, meaning to say that the value of the BFP transfer, in 2015, contributed to a higher percentage of the municipalities of the State of Ceará reaching marks above the average of the IDEB, thus configuring an improvement in elementary education teaching. And finally, according to the analysis of canonical correlation, it is admitted that there is correlation between the groups of the variables, which validates the model, corroborating that the factorial analysis revealed, considered the variables approval rate of elementary school I and the notes of the IDEB as more representative variables within the canonical correlation model; these that already contemplate indicators found among the three most important factors, according to the methodology of logistic regression and discriminant analysis, endorsing that the variables investigated in the study best explains the positive impacts of the BFP in elementary education.



KEYWORDS: Bolsa Família Program, Elementary School, Econometrics.



Artigo está licenciado sob forma de uma licença
Creative Commons Atribuição 4.0 Internacional.

1. Introdução

Neste trabalho será abordado como o Programa Bolsa Família - PBF vem contribuindo para mudar a educação. Sendo a educação um dos principais temas de discussões e reflexões. Sua importância deve-se ao fato de abranger temas sociais, econômicos, políticos e culturais dos diferentes países que vêm na educação o pilar das mudanças estruturais da sociedade. Para Silva e Lima (2015), a implantação de políticas de melhoria do sistema educacional é importante para avançar na superação da pobreza e indigência no semiárido brasileiro.

Nesse sentido, surgiu como uma das alternativas de mudança o Programa Bolsa Família que tem como meta vincular a transferência de renda à permanência da criança e do adolescente na escola. O intuito do Programa é elevar o grau de escolaridade e permanência das crianças para aumentar as oportunidades sociais.

Segundo Pontili (2004), o grau de escolaridade no Brasil é, em média, de 6,4 anos, considerado baixo em relação a países da América Latina. A partir dessa realidade, foram estabelecidas mudanças na forma de gestão do ensino público brasileiro, com o intuito de melhorar a qualidade do ensino e, principalmente, o interesse e permanência do aluno na escola. De acordo com Castelar, Monteiro e Lavor (2013), com a disponibilidade de ônibus escolar para alunos da zona rural é possível reduzir taxas de abandono escolar.

Logo, a contribuição do PBF na redução da desigualdade deve estar



atrelada à melhoria dos indicadores de educação. De acordo Ney (2006), as desigualdades de oportunidades educacionais geram dois problemas básicos para a redução da desigualdade de renda no Brasil. O primeiro é que elas produzem grandes diferenças na qualidade da mão-de-obra que ingressa no mercado de trabalho, tendendo a gerar futuras disparidades de rendimentos. O segundo é que a pouca chance de os jovens mais pobres chegarem ao ensino médio limita a expansão da educação justo no nível em que sua taxa de retorno é maior. O efeito da escolaridade na renda depende do valor pago pelo mercado de trabalho a cada ano adicional de estudo, que é influenciado pela escassez relativa da oferta de mão-de-obra qualificada.

A percepção de que não se reduz desigualdade sem educação está implícita, mesmo de forma pouco perceptível, nos objetivos do PBF. Ao exigir frequência escolar das crianças cuja família é beneficiada, o programa cria a expectativa de melhorar os indicadores de capital humano, mais especificamente aqueles referentes à dimensão educacional.

No entanto, o impacto do programa sobre a educação tem sido pouco explorado na literatura científica. Os trabalhos empíricos que abordam o Programa Bolsa Família se destinam, em sua maioria, à discussão do impacto desse programa sobre uma vasta gama de temáticas sociais, como o gasto familiar (consumo), o mercado de trabalho, a saúde, o trabalho infantil, a segurança alimentar.

Bouillon e Tejerina (2006) abordam 51 trabalhos que versam sobre 47 programas de proteção social na América Latina e Caribe. Todos os trabalhos foram selecionados por empregarem um processo de comparação de avaliação entre as famílias participantes e não participantes, as mudanças na variável objeto da intervenção antes e após a implementação dos programas que utiliza o estimador diferença-em-diferenças, dentre os principais resultados observados pelos autores, destacam-se maior frequência das crianças à escola, maiores gastos familiares no consumo alimentar, redução da natalidade e mortalidade infantis, e diminuição do



número de crianças ocupadas ou de horas dedicadas ao trabalho, no caso dos programas que prevêem frequência diária integral da criança na escola.

Janvry *et al.* (2007), utilizando dados de 261 municípios situados em 5 estados nordestinos, coletados junto às secretarias destes municípios, estimam que a participação no programa reduziu a taxa de abandono escolar em 7.8 pontos percentuais, assim como elevou a taxa de reprovação em 0.8 pontos percentuais. Os autores argumentam que esse último resultado pode ser explicado pelo fato de o benefício do programa ajudar a manter na escola um indivíduo com menor capacidade acadêmica ou com pouco interesse nos estudos e que, não fosse o benefício do programa, abandonaria a escola. Para obter estes resultados, utiliza regressões em painel por município, com dados entre 1999 e 2003, comparando alunos elegíveis ao recebimento do programa com alunos que efetivamente recebem o benefício.

Seguindo tal linha, Glewwe e Kassouf (2010) estimaram o impacto do PBF sobre o total de matrículas do ensino fundamental e sobre as taxas de abandono e aprovação por meio de dados do Censo Escolar entre 1998 e 2005. Por meio do Censo Escolar destes anos não era possível identificar quais e quantos alunos eram beneficiados pelo programa em cada escola, de maneira que era possível saber somente se pelo menos um aluno da escola recebia o benefício. Utilizando regressões por escola, os autores mostram que o Programa Bolsa Família aumentou as matrículas de 1^a a 4^a série em 5,5% e de 5^a a 8^a série em 6,5%, diminuiu as taxas de abandono escolar em 0,5 ponto percentual nas escolas de 1^a a 4^a série e em 0,4 por cento nas de 5^a a 8^a série e aumentou as taxas de aprovação em cerca de 0,9 ponto percentual de 1^a a 4^a série e 0,3 ponto percentual de 5^a a 8^a série.

Em relação ao impacto na alimentação, Rocha, Khan e Lima (2009), em estudo para o Ceará, apontam que o PBF tem um papel importante no bem-estar das famílias como política de curto prazo, tendo impacto positivo sobre indicadores de saúde, educação e estado nutricional das famílias beneficiadas. Eles destacam, também, que o consumo médio de bens da



cesta alimentar básica é superior nas famílias beneficiadas em relação às não beneficiadas, e para cada R\$ 1,00 transferido para as famílias R\$ 0,72 são gastos com alimentação.

Uma análise preliminar dos estudos sobre o PBF permite fazer uma tipologia dos mesmos destacando três características: I) a variável impactada (renda, segurança alimentar, educação); II) natureza dos dados (primários ou secundários) e III) metodologia empregada (análises quantitativas, qualitativas, quantiqualitativas). Os estudos apresentam como ponto comum o fato de tratar o programa como uma intervenção federal dissociada das outras esferas do governo. Essa tendência tem criado lacunas de conhecimento quanto: I) à importância do ambiente institucional, especialmente no âmbito municipal, para a efetivação de políticas públicas e II) capacidade de integração do PBF com outras intervenções governamentais no âmbito da educação.

O entendimento da contribuição do PBF para os indicadores educacionais adquire especial relevância haja vista que a educação é o principal meio de mobilidade social e é a principal variável explicativa dos diferenciais de rendimentos. A educação, isoladamente, pode não resolver os problemas da sociedade, mas promove a inclusão social e o desenvolvimento sustentável.

O impacto do bolsa família pode ir além da redução da pobreza (na perspectiva minimalista da privação de renda). Diferente da maioria dos estudos que avaliam o referido programa, o projeto busca focar o impacto na educação. Dessa forma, adquire uma dimensão interdisciplinar dado que a educação perpassa por todos os aspectos do desenvolvimento.

Ao buscar elementos que permitam avaliar de que forma o PBF cria oportunidades educacionais para a população, o artigo ganha importância e se diferencia dos demais por trazer possibilidades de análise a partir de uma visão interdisciplinar que requer a consulta a autores nas áreas de pedagogia, economia, sociologia, geografia, ciências políticas, administração



pública, entre outras.

O trabalho tem como objetivo geral estudar a análise multivariada, envolvendo a relação entre aspectos socioeconômicos e variáveis educacionais do Estado do Ceará e; objetivos específicos determinar a relação quantitativa entre as variáveis e identificar os fatores representativos da educação nos municípios do Ceará, agrupando-os a partir do cálculo de um índice educacional; verificar se os indicadores socioeconômicos influenciam os indicadores educacionais nos municípios do Estado do Ceará e identificar as variáveis que mais influenciam os indicadores educacionais nos municípios cearenses.

2. Referencial Teórico

Nesta seção será abordado como o PBF vem contribuindo para mudar a educação. Sendo a educação um dos principais temas de discussões e reflexões. Sua importância deve-se ao fato de abranger temas sociais, econômicos, políticos e culturais dos diferentes países que vêm na educação o pilar das mudanças estruturais da sociedade.

Nesse sentido, surgiu como uma das alternativas de mudança o Programa Bolsa Família que tem como meta vincular a transferência de renda à permanência da criança e adolescente na escola. O intuito do Programa é elevar o grau de escolaridade e permanência das crianças para aumentar as oportunidades sociais.

Segundo Pontili (2004), o grau de escolaridade no Brasil é, em média, de 6,4 anos, considerado baixo em relação a países da América Latina. A partir dessa realidade, foram estabelecidas mudanças na forma de gestão de ensino público brasileiro, com o intuito de melhorar a qualidade do ensino e, principalmente, o interesse e permanência do aluno na escola.

Logo, a contribuição do PBF na redução da desigualdade deve estar atrelada à melhoria dos indicadores de educação. De acordo com Ney



(2006), as desigualdades de oportunidades educacionais geram dois problemas básicos para a redução da desigualdade de renda no Brasil. O primeiro é que elas produzem grandes diferenças na qualidade da mão-de-obra que ingressa no mercado de trabalho, tendendo a gerar futuras disparidades de rendimentos. O segundo é que a pouca chance de os jovens mais pobres chegarem ao ensino médio limita a expansão da educação justo no nível em que sua taxa de retorno é maior. O efeito da escolaridade na renda depende do valor pago pelo mercado de trabalho a cada ano adicional de estudo, que é influenciado pela escassez relativa da oferta de mão-de-obra qualificada.

A percepção de que não se reduz desigualdade sem educação está implícita, mesmo de forma pouco perceptível, nos objetivos do PBF. Ao exigir frequência escolar das crianças cuja família é beneficiada, o programa cria a expectativa de melhorar os indicadores de capital humano, mais especificamente aqueles referentes à dimensão educacional.

No entanto, o impacto do programa sobre a educação tem sido pouco explorado na literatura científica. Os trabalhos empíricos que abordam o Programa Bolsa Família se destinam, em sua maioria, à discussão do impacto desse programa sobre uma vasta gama de temáticas sociais, como o gasto familiar (consumo), o mercado de trabalho, a saúde, o trabalho infantil, a segurança alimentar.

Janvry *et al.* (2007), utilizando dados de 261 municípios, situados em 5 estados nordestinos, coletados junto às secretarias destes municípios, estimam que a participação no programa reduziu a taxa de abandono escolar em 7.8 pontos percentuais, assim como elevou a taxa de reprovação em 0.8 pontos percentuais. Os autores argumentam que esse último resultado pode ser explicado pelo fato de o benefício do programa ajudar a manter na escola um indivíduo com menor capacidade acadêmica ou com pouco interesse nos estudos e que, não fosse o benefício do programa, abandonaria a escola. Para obter estes resultados, utiliza regressões em painel por município, com



dados entre 1999 e 2003, comparando alunos elegíveis ao recebimento do programa com alunos que efetivamente recebem o benefício.

Seguindo tal linha, Glewwe e Kassouf (2010) estimaram o impacto do PBF sobre o total de matrículas do ensino fundamental e sobre as taxas de abandono e aprovação por meio de dados do Censo Escolar, entre 1998 e 2005. Por meio do Censo Escolar destes anos, não era possível identificar quais e quantos alunos eram beneficiados pelo programa em cada escola, de maneira que era possível saber somente se pelo menos um aluno da escola recebia o benefício. Utilizando regressões por escola, os autores mostram que o Programa Bolsa Família aumentou as matrículas de 1^a a 4^a série em 5,5% e de 5^a a 8^a série em 6,5%, diminuiu as taxas de abandono escolar em 0,5 ponto percentual nas escolas de 1^a a 4^a série e em 0,4 por cento nas de 5^a a 8^a série e aumentou as taxas de aprovação em cerca de 0,9 ponto percentual de 1^a a 4^a série e 0,3 ponto percentual de 5^a a 8^a série.

Neste mesmo contexto, Pellegrina (2011) avalia o impacto do PBF sobre o desempenho escolar e a matrícula dos alunos do estado de São Paulo. Utilizando as notas do SARESP de 2007 e 2009, além das notas de boletins escolares neste mesmo período, o autor estimou os possíveis impactos do programa sobre diferentes variáveis associadas à educação, todas ao nível de indivíduo. Por meio de diferentes métodos de *matching* e com o uso de diferenças-em-diferenças, o autor encontrou que o programa tem efeito sobre as variáveis que estão diretamente atreladas às condições impostas pelo PBF ao recebimento, mas nenhum efeito sobre as variáveis de desempenho escolar. O autor encontrou evidências de que a participação no PBF reduz o abandono e aumenta a frequência escolar, entretanto parece não haver efeito sobre as notas em exames padronizados e sobre as notas do boletim escolar.

Em relação ao impacto na alimentação, Rocha, Khan e Lima (2009), em estudo para o Ceará, apontam que o PBF tem um papel importante no bem-estar das famílias como política de curto prazo, tendo impacto positivo



sobre indicadores de saúde, educação e estado nutricional das famílias beneficiadas. Eles destacam, também, que o consumo médio de bens da cesta alimentar básica é superior nas famílias beneficiadas em relação às não beneficiadas, e para cada R\$ 1,00 transferido para as famílias R\$ 0,72 são gastos com alimentação.

Uma análise preliminar dos estudos sobre o PBF permite fazer uma tipologia dos mesmos, destacando três características: I) a variável impactada (renda, segurança alimentar, educação); II) natureza dos dados (primários ou secundários) e III) metodologia empregada (análises quantitativas, qualitativas, quantiqualitativas). Os estudos apresentam como ponto comum o fato de tratar o programa como uma intervenção federal dissociada das outras esferas do governo. Essa tendência tem criado lacunas de conhecimento quanto: I) à importância do ambiente institucional, especialmente no âmbito municipal, para a efetivação de políticas públicas e II) capacidade de integração do PBF com outras intervenções governamentais no âmbito da educação.

O entendimento da contribuição do PBF para os indicadores educacionais adquire especial relevância haja vista a educação ser o principal meio de mobilidade social e também a principal variável explicativa dos diferenciais de rendimentos. A educação, isoladamente, pode não resolver os problemas da sociedade, mas promove a inclusão social e o desenvolvimento sustentável.

O impacto do bolsa família pode ir além da redução da pobreza (na perspectiva minimalista da privação de renda). Diferente da maioria dos estudos que avaliam o referido programa, o projeto busca focar o impacto na educação. Dessa forma, adquire uma dimensão interdisciplinar dado que a educação perpassa por todos os aspectos do desenvolvimento.

Ao buscar elementos que permitam avaliar de que forma o PBF cria oportunidades educacionais para a população, o artigo ganha importância e se diferencia dos demais por trazer possibilidades de análise a partir de uma



visão interdisciplinar que requer a consulta a autores nas áreas de pedagogia, economia, sociologia, geografia, ciências políticas, administração pública, entre outras.

Assim, o estudo é importante por se debruçar em um dos principais programas do governo federal dos últimos anos e também um dos mais estudados, porém numa perspectiva diferenciada dos demais.

3. Metodologia

3.1 Delimitação e Caracterização da Área de Estudo da Pesquisa

A área de estudo da pesquisa é o Estado Ceará que tem 184 municípios. Segundo Pinho Neto *et al.* (2011), sendo localizado na Região Nordeste, reconhecidamente a mais pobre e desigual do Brasil, o Estado do Ceará apresenta uma elevada proporção de pobres e um grau de concentração de renda relativamente elevado.

Sendo a educação um dos principais temas de discussões e reflexões para reduzir a desigualdade de renda, pois abrange temas sociais, econômicos, políticos e culturais dos diferentes países que veem na educação o pilar das mudanças estruturais da sociedade.

3.2 Origem dos Dados

A metodologia adotada utilizou dados secundários extraídos do Portal Brasil, Data Social, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Instituto de Pesquisa Econômica Aplicada (IPEA). Os dados foram coletados em escala Municipal para o Estado do Ceará, tendo como referência o período mais recente de divulgação para dados, 2015.



Quadro 1 – Variáveis da Análise Multivariada.

Quantidade de famílias beneficiárias do Programa Bolsa Família
Valor total repassado do PBF (R\$)
Total de crianças e adolescentes de 6 a 17 anos público da educação
Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar
Número de Matrícula ensino fundamental inicial parcial urbano - 2015
Número de Matrícula ensino fundamental inicial integral urbano - 2015
Número de Matrícula ensino fundamental final parcial urbano - 2015
Número de Matrícula ensino fundamental final integral urbano - 2015
Número de Matrícula ensino fundamental inicial parcial rural - 2015
Número de Matrícula ensino fundamental inicial integral rural - 2015
Número de Matrícula ensino fundamental final parcial rural - 2015
Número de Matrícula ensino fundamental final integral rural - 2015
IDEB - Projeção 2015
Taxa de aprovação (1º ao 5º ano)
Indicador de Rendimento (2013)
Taxa de aprovação 1º ao 5º ano
Taxa de Evasão ensino fundamental parcial urbano
Taxa de Evasão integral urbano
Taxa de Evasão ensino fundamental parcial rural
Taxa de Evasão ensino fundamental integral rural
População - 2015
Taxa urbanização (%) - 2010
Densidade demográfica (hab/km ²) - 2015
Ensino Médio - Aprovação (%) - 2015
Ensino Médio - Reprovação (%) - 2015
Ensino Médio - Abandono (%) - 2015
Taxa de mortalidade (por mil nascidos vivos) menores de 1 ano - 2015
Emprego formal (18 anos ou mais) - 2015
Emprego formal (Ensino Fundamental completo) - 2015
Emprego formal (Ensino Médio completo) - 2015
Emprego formal (Ensino Superior completo) - 2015
População Extremamente Pobre (%) - 2010
IDH - 2010
Taxa de cobertura de abastecimento de água - urbano (%) - 2015
Taxa de cobertura de esgotamento sanitário - urbano (%) - 2015
PIB <i>per capita</i> - 2014
Indústrias ativas - 2015

Fonte: Elaborada pelas autoras (2018).

A análise de dados por escola é dificultada pelo fato de que a variável Bolsa Família mede apenas a existência do programa, demonstrando apenas o impacto da disponibilidade do programa. Felizmente, existem dados por município com o número de famílias participantes do programa, o que permite uma estimativa do impacto de se participar do programa.



3.3 Métodos de Análise

3.3.1 Análise fatorial

A Análise Fatorial (AF) visa sintetizar informações quando se tem um grande número de variáveis correlacionadas. Esta técnica mostra a intensidade de cada variável com cada fator. Neste método, destaca-se como vantagem a abordagem complexa de um fenômeno.

A análise de agrupamento, ou análise de “cluster”, é uma técnica computacional que trabalha com subconjuntos. Este método tem o intuito de desmembrar objetos do banco de dados, em seguida reorganizá-los em novos subgrupos, com base em determinadas características. Nesta técnica, as variáveis são analisadas de forma concomitante.

A análise de regressão múltipla permite analisar a relação entre uma variável Y dependente e uma ou mais variáveis independentes (X).

O valor de KMO correspondente a 0,779 mostra que a adequação dos dados à Análise Fatorial pode ser classificada como média, segundo Fávero (2009).

O teste de esfericidade de Bartlett aponta se a matriz de correlação é uma matriz identidade com determinante igual a 1, isto é, a correlação entre as variáveis é zero.

Caso essa situação seja comprovada o modelo de análise fatorial para a análise dos dados em questão deve ser descartado.

As hipóteses do teste são:

H_0 : a matriz de correlações é uma matriz identidade;

H_1 : existem correlações significativas entre as variáveis.

Quando $p\text{-value} > 0,05$, ou seja, a hipótese nula do teste não deve ser rejeitada a um nível de significância de 5%, a análise fatorial deve ser descartada.



3.3.2 Análise de regressão múltipla

Nesta seção são apresentados conceitos sobre a regressão linear múltipla, que se referem a uma situação em que a reta ajustada não descreve bem o conjunto de dados e, com isso, podem ser levadas em consideração outras variáveis independentes que possivelmente influenciam no valor de Y, a variável dependente. Ou seja, a regressão múltipla pode ser usada no intuito de melhorar o modelo desenvolvido para explicar o comportamento das variáveis do banco de dados que estão sendo estudadas.

Em regressão múltipla, a variável determinada é aquela que tenha correlação significativa com a variável a ser prevista. A variável está no centro das análises e deve ser identificado o seu impacto coletivo, assim como a contribuição de cada variável separada para o efeito geral da variável preditora.

A regressão linear múltipla é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis estudadas (variável dependente ou resposta) e o restante das variáveis que descrevem o sistema (variáveis independentes ou explicativas), e reduzir um grande número de variáveis para poucas dimensões com o mínimo de perda de informação, permitindo a detecção dos principais padrões de similaridade, associação e correlação entre as variáveis. Sua principal aplicação, após encontrar a relação matemática, é produzir valores para a variável dependente quando se têm as variáveis independentes (cálculo dos valores preditos). Ou seja, ela pode ser usada na predição de resultados, por meio da regra estatística dos mínimos quadrados.

Quando se trabalha com análise de regressão, deve-se realizar a análise de variância com o objetivo de comparar os modelos e avaliar a significância da regressão. Considerando o modelo de regressão linear múltipla, pode-se construir a tabela ANOVA (tabela de análise de variância, corrigida pela média), dada por:



Tabela 1 – ANOVA.

Fonte de variação	Soma de quadrados	Graus de liberdade	Soma dos quadrados médios	F calculado
Regressão Erro	SQReg SQRes	p $n-p-1$	SQMReg SQMRes	SQMRes/ SQMRes
Total	SQT	$n-1$		

Fonte: Gujarato (2010).

Em que:

$$SQReg = y'Hy - n\bar{y}^2$$

$$SQRes = y'y - \hat{\beta}'X'y$$

$$SQT = y'y - n\bar{y}^2$$

$$SQMReg = SQReg/p$$

$$SQMRes = SQRes/(n-p-1)$$

Na tabela 1, as hipóteses em teste são:

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0 \text{ versus } H_1 : \text{Ao menos um } \beta_i \neq 0.$$

Assim, se a hipótese H_0 for a verdadeira, o modelo não está bem ajustado, pois os coeficientes são estatisticamente iguais a zero. Pode-se também calcular o coeficiente de determinação e o coeficiente de determinação ajustado, dados respectivamente por:

$$R^2 = SQReg/SQT \text{ e } R^2_{ajustado} = 1 - \frac{SQRes/(n-p-1)}{SQT/(n-1)}.$$

Com esta breve explicação sobre a análise de regressão múltipla, pode-se agora inseri-la no contexto de planejamento de experimentos.



3.3.3 Análise discriminante

Técnica de estatística multivariada que auxilia na identificação das variáveis que diferenciam dois ou mais grupos. É, portanto, uma técnica de classificação. Uma análise estatística das variáveis nos grupos gera uma regra matemática (função discriminante) que permite classificar qualquer elemento não pertencente a um dos grupos, em um deles.

Os principais objetivos da Análise Discriminante (AD) são: identificar e entender diferenças significativas entre variáveis em grupos previamente definidos; identificar quais variáveis melhor diferenciam grupos; e classificar indivíduos em um grupo a partir de suas características (variáveis independentes).

A análise discriminante é uma técnica de dependência adequada quando a variável dependente é dicotômica ou multicotômica, e, portanto, não métrica. As variáveis independentes devem ser métricas. O objetivo da análise é entender as diferenças entre os grupos e prever a probabilidade de um indivíduo pertencer a grupo particular com base em um conjunto de variáveis independentes.

A AD busca classificar indivíduos em grupos com a menor probabilidade de erro. Para tanto, define uma função discriminante que apresenta a seguinte forma:

$$Z_{jk} = a + \omega_1 x_{1k} + \omega_2 x_{2k} + \dots + \omega_n x_{nk}$$

Em que:

Z_{jk} = escore discriminante da função discriminante j para o elemento ou indivíduo k ; a = constante ou intercepto da função; w_i = peso ou coeficiente discriminante para a variável independente i (capacidade de uma variável discriminar os grupos); X_{jk} = valor da variável independente i para o elemento k .



A AD testa a hipótese de que as médias de um conjunto de variáveis independentes são iguais para dois ou mais grupos. Para tanto, multiplica cada variável independente por seu peso correspondente e soma todos os produtos. O resultado é um escore Z discriminante para cada indivíduo. As médias de um conjunto de variáveis independentes correspondem à média dos escores discriminantes no grupo e recebem o nome de centróide.

Portanto, o número de centróides calculados é igual ao número de grupos ou categorias da variável dependente. O centróide indica o local mais típico de um indivíduo. A comparação dos centróides mostra o quão afastados estão os grupos. Quanto mais afastados os grupos, melhor a função discriminante.

A análise discriminante é dividida em 8 estágios:

Definição dos objetivos da AD: determinar quais as principais diferenças entre dois grupos; determinar quais variáveis explicam melhor as diferenças entre grupos; estabelecer procedimentos para classificar objetos em grupos com base em escores de um conjunto de variáveis independentes.

Planejamento da pesquisa: 1º Definir o tamanho da amostra; AD é muito sensível à relação entre o tamanho da amostra e o número de variáveis independentes; O mínimo aceitável é de 1 variável para 5 observações, de preferência cada grupo deve conter pelo menos 20 observações. 2º Divisão da amostra; 3º Pré-Seleção das variáveis (devem ser selecionadas variáveis de interesse da pesquisa quanto ao seu provável poder de diferenciar os grupos).

Verificação dos pressupostos da AD: 1º Normalidade multivariada; 2º Homocedasticidade (igualdade das matrizes de variância e covariância para os grupos); 3º Ausência de multicolinearidade-as variáveis explicativas independentes; 4º Linearidade das relações entre variáveis.

Estimação da função discriminante e ajuste do modelo. Dois métodos



computacionais podem ser usados para estimar uma função discriminante: Método simultâneo (todas as variáveis independentes são consideradas simultaneamente sem levar em conta o poder discriminatório de cada uma); Método *stepwise* (cada variável é incluída na análise de cada vez, com base em seu poder discriminatório).

A escolha das variáveis independentes pré-selecionadas é realizada por meio de critérios estatísticos. Os mais comuns são: Lambda de Wilks; Traço de Hotelling; Critério de Pillai; D2 de Mahalanobis; e V de Rao. Esses critérios indicam as variáveis que apresentam poder discriminatório entre os grupos.

Lambda de Wilks: varia de 0 a 1, permite avaliar a existência de diferenças de médias entre os grupos, para cada variável. Quanto maiores os valores de Lambda de Wilks, mais semelhantes os grupos e menor a contribuição das variáveis na discriminação dos grupos. Se for igual a 1 significa que são do mesmo grupo.

Avaliação do ajuste geral do modelo: cálculo dos escores Z discriminantes para cada observação. O escore Z fornece uma maneira direta de comparar observações em cada função discriminante. Quanto mais próximos os escores Z de duas observações, mais semelhantes quanto às variáveis da função.

Interpretação dos resultados: consiste basicamente em entender a importância de cada variável independente na discriminação dos grupos. A importância de cada variável independente na discriminação dos grupos pode ser verificada de três maneiras:

- Pelos pesos (coeficientes) discriminantes (não padronizados ou padronizados).
- Pelas cargas discriminantes (correlações de estruturas).
- Pelos valores F parciais.

Validação dos resultados e significância das funções discriminantes:



validação cruzada - pode ser feita com a amostra original ou com uma amostra de teste. Dividir a amostra total em dois grupos (de análise e de teste) e estimar a função discriminante para a amostra de análise. Em seguida, aplicar na amostra de teste: diferenças de perfis de grupo - traçar o perfil das características dos grupos com base nas variáveis com maior poder discriminante (valores médios): Método U, método jackknife (uso limitado devido ao tamanho exigido para os grupos - cinco vezes o número de variáveis independentes). Classificação de elementos e é feito com base nos valores dos centróides dos grupos. O elemento é classificado no grupo cujo centróide se encontra mais próximo.

No caso de dois grupos o ponto de corte é dado por:

$$f = \frac{n_1 d_1' + n_2 d_2'}{n_1 + n_2}$$

São as médias da função discriminante nos grupos 1 e 2 (médias dos escores Z). Tamanho dos grupos 1 e 2.

Um elemento é classificado no grupo 1 se o seu *score* Z na função discriminante for maior que f, caso contrário é classificado no grupo 2. Quando há mais de dois grupos a zona de fronteira é definida para cada par de grupos, sendo definidas g regiões exclusivas. O *score* de cada elemento é calculado e localizado em um mapa territorial, sendo possível identificar a qual grupo pertence.

Esse procedimento é mais bem compreendido quando considerada a “probabilidade *a priori*” do elemento pertencer a um determinado grupo. Cada elemento é classificado no grupo onde essa probabilidade é maior.



3.3.4 Análise de relação canônica

A análise de regressão múltipla é uma técnica multivariada que pode prever o valor de uma única variável dependente (métrica) a partir de uma função linear de um conjunto de variáveis independentes.

Para alguns problemas de pesquisa, o interesse pode não se concentrar em uma única variável dependente; em vez disso, o pesquisador talvez esteja interessado em relações entre conjuntos de múltiplas variáveis dependentes e múltiplas variáveis independentes.

A análise de correlação canônica é um modelo estatístico multivariado que facilita o estudo de inter-relações entre conjuntos de múltiplas variáveis dependentes e múltiplas variáveis independentes.

Ao contrário da regressão múltipla, que prevê uma única variável dependente a partir de um conjunto de variáveis independentes múltiplas, a correlação canônica simultaneamente prevê múltiplas variáveis dependentes a partir de múltiplas variáveis independentes.

A correlação canônica apresenta o menor número de restrições sobre os tipos de dados nos quais ela opera. Como as outras técnicas impõem restrições mais rígidas, em geral crê-se que a informação obtida a partir delas é de maior qualidade e pode ser apresentada de uma maneira melhor para a interpretação.

Por essa razão, muitos pesquisadores consideram a correlação canônica como uma última alternativa, a ser usada quando todas as outras técnicas de nível mais alto forem descartadas. Mas em situações com múltiplas variáveis dependentes e independentes, a correlação canônica é a técnica multivariada mais adequada e poderosa. Ela obteve aceitação em muitas áreas e representa uma ferramenta útil para a análise multivariada, particularmente porque aumentou o interesse na consideração de múltiplas variáveis dependentes.



A análise de correlação canônica é o membro mais generalizado da família de técnicas estatísticas multivariadas. Está diretamente relacionada com diversos métodos de dependência. Semelhante à regressão, a meta da correlação canônica é quantificar a força da relação, nesse caso, entre os dois conjuntos de variáveis (independentes e dependentes). Ela corresponde à análise fatorial na criação de composições de variáveis. Também se assemelha à análise discriminante por conta de sua habilidade de determinar dimensões independentes (semelhantes às funções discriminantes) para cada conjunto de variáveis nessa situação com o objetivo de produzir a máxima correlação entre as dimensões.

Assim, a correlação canônica identifica a estrutura ou dimensionalidade ótima de cada conjunto de variáveis que maximiza a relação entre conjuntos de variáveis independentes e dependentes.

A análise de correlação canônica lida com a associação entre composições de conjuntos de múltiplas variáveis dependentes e independentes. Ao fazer isso, desenvolve diversas funções canônicas independentes que maximizam a correlação entre as composições lineares, também conhecidas como variáveis estatísticas canônicas, as quais são conjuntos de variáveis dependentes e independentes. Cada função canônica é realmente baseada na correlação entre duas variáveis estatísticas canônicas, uma para as variáveis dependentes e outra para as independentes.

Uma outra característica singular da correlação canônica é que as variáveis estatísticas são obtidas para maximizar sua correlação. Além disso, a correlação canônica não termina com a derivação de uma única relação entre os conjuntos de variáveis. Em vez disso, diversas funções canônicas (pares de variáveis estatísticas canônicas) podem ser obtidas.

A discussão de análise de correlação canônica, a seguir, é organizada em torno do processo de construção de modelo. Os passos neste processo incluem a especificação dos objetivos da correlação canônica, o



desenvolvimento do plano de análise, a avaliação das suposições inerentes à correlação canônica, a estimação do modelo canônico e a avaliação do ajuste geral do modelo, a interpretação das variáveis estatísticas canônicas e a validação do modelo.

4. Resultados e Discussões

4.1 Análise Fatorial

O resultado obtido pelo teste KMO mostrou que a análise alcançou 78% de adequação e um nível de significância associado menor que 0,01. Assim, a hipótese de que a matriz de correlação é uma matriz de identidade pode ser rejeitada, ou seja, a matriz de correlação possui correlações significativas entre pelo menos algumas das variáveis.

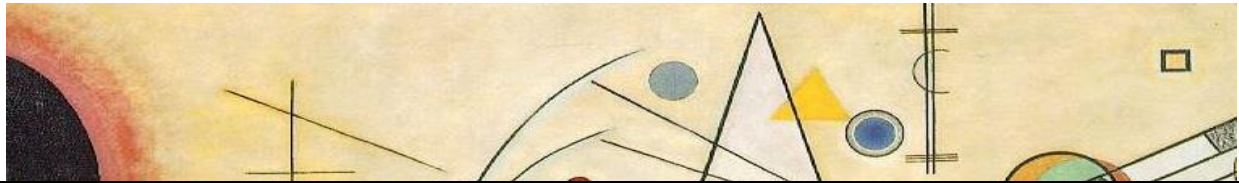
A soma das cargas ao quadrado de cada componente na variável é denominada comunalidade e representa o quanto do percentual da variação de uma variável é explicada pelo conjunto de componentes principais formados. Variáveis com comunalidades inferiores a 0,5 devem ser excluídas da análise. No caso da análise em questão, nenhuma deve ser excluída.

De acordo com os dados da variância total explicada, 9 fatores foram extraídos, de acordo com o critério de manter apenas aqueles com autovalores maiores que 1. Na tabela 2 tem-se os indicadores de prioridades a que se deve observar para se promover políticas públicas. Dos 9 indicadores apontados, hierarquicamente, os 3 aos quais se devem dar maior atenção são os indicadores voltados aos aspectos educacionais e socioeconômicos, educação básica e os que permeiam em torno de índices macroeconômicos (urbanização, pobreza, IDH, PIB e mortalidade infantil).



Tabela 2 – Identificação dos fatores.

Variáveis	Fatores	Nome
Quantidade de famílias beneficiárias do Programa Bolsa Família (PBF) - 2015 Valor total repassado do PBF (R\$) - 2015 Total de crianças e adolescentes de 6 a 17 anos público da educação - 2015 Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015 Núm. Matrícula - Ensino Fund. Inicial parcial urbano - 2015 Núm. Matrícula - Ensino Fund. Inicial integral urbano - 2015 Núm. Matrícula - Ensino Fund. final parcial urbano - 2015 Núm. Matrícula - Ensino Fund. final integral urbano - 2015 População - 2015 Dens demográfica (hab/km ²) - 2015 Emprego formal (Ensino Fundamental completo) - 2015 Emprego formal (Ensino Médio completo) - 2015 Emprego formal (Ensino Superior completo) - 2015 Indústrias ativas - 2015	F1	Aspectos educacionais e socioeconômicos
IDEB 2013 Indicador de Rendimento (2013) (taxa de aprovação 1º ao 5º ano) - 2015 Taxa de aprovação (1º ao 5º ano) - 2015 TX EVASÃO (Ens. Fund. parcial urbano) - 2015	F2	Educação Básica
Taxa urbanização (%) - 2010 Tx mortalidade (por mil nascidos vivos) menores de 1 ano - 2015 População Extremamente Pobre (%) - 2010 IDH - 2010 PIB per capita - 2014	F3	Urbanização, pobreza, mortalidade infantil, IDH e PIB
Núm. Matrícula - Ensino Fund. Inicial integral rural - 2015 Núm. Matrícula - Ensino Fund. final integral rural - 2015	F4	Matrícula - Ensino Fundamental (integral rural)
Núm. Matrícula - Ensino Fund. Inicial parcial rural - 2015 Núm. Matrícula - Ensino Fund. final parcial rural - 2015	F5	Matrícula - Ensino Fundamental (parcial rural)
Ensino Médio - Aprovação (%) - 2015 Ensino Médio - Reprovação (%) - 2015 Ensino Médio - Abandono (%) - 2015	F6	Indicadores do Ensino Médio
IDEB - Projeção 2015 TX EVASÃO (Ens. Fund. integral urbano) - 2015 Tx cobertura de abast água - urbano (%) - 2015	F7	IDEB, evasão e abastecimento de água
TX EVASÃO (Ens. Fund. integral rural) - 2015	F8	Taxa de evasão (Ensino



TX EVASÃO (Ens. Fund. parcial rural) - 2015		Fundamental integral rural)
Tx cobertura de esgotamento sanitário - urbano (%) - 2015	F9	Taxa de evasão (Ensino Fundamental parcial rural) e Taxa de esgotamento sanitário

Fonte: Elaborada pelas autoras (2018).

Os indicadores que possuem menor impacto entre o programa bolsa família e os indicadores educacionais são as taxas de evasão e taxa de esgotamento sanitário, concluindo que estes são indicadores menos importantes em relação aos outros, porém não descartáveis na análise geral.

4.2 Análise de Regressão Múltipla

Para a análise de regressão múltipla, se tem como variável dependente a quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar no ano de 2015 e; para variáveis independentes são aprovação no ensino médio (%), reprovação no ensino médio (%) e abandono no ensino médio (%), todos para o ano de 2015.

4.2.1 Testando dos pressupostos

Figura 1 – Estatística de ajuste do modelo.

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	6,6738 E3	2,4610E4	1,1805 E4	3230,22354	182
Residual	-2,04458 E4	3,37232 E5	,00000	27056,15158	182
Std. Predicted Value	-1,589	3,964	,000	1,000	182
Std. Residual	-,751	12,395	,000	,994	182

a. Dependent Variable: Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015

Fonte: Elaborada pelas autoras (2018).



A média dos resíduos é igual a zero, pressuposto não violado. Portanto, modelo ajustado.

Tabela 3 – Testando a normalidade do modelo.

Variáveis	Kolomorov-smirnov	Shapiro
Ensino Médio - Aprovação (%) - 2015	0,001	,000
Ensino Médio - Reprovação (%) - 2015	0,00	0,000
Ensino Médio - Abandono (%) - 2015	0,00	0,00

Fonte: Elaborada pelas autoras (2018).

A análise da Normalidade foi realizada de forma individual através dos testes de Kolomorov-Smirnov e Shapiro e para tais variáveis independentes descritas no modelo, esse pressuposto para o método de análise de regressão foi violado.

Figura 2 – Ausência de Autocorrelação.

Correlations					
		Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015	Ensino Médio - Aprovação (%) - 2015	Ensino Médio - Reprovação (%) - 2015	Ensino Médio - Abandono (%) - 2015
Pearson Correlation	Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015	1,000	-,083	,116	,014
	Ensino Médio - Aprovação (%) - 2015	-,083	1,000	-,822	-,795
	Ensino Médio - Reprovação (%) - 2015	,116	-,822	1,000	,308
	Ensino Médio - Abandono (%) - 2015	,014	-,795	,308	1,000
Sig. (1-tailed)	Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015		,133	,059	,424
	Ensino Médio - Aprovação (%) - 2015	,133		,000	,000
	Ensino Médio - Reprovação (%) - 2015	,059	,000		,000
	Ensino Médio - Abandono (%) - 2015	,424	,000	,000	
N	Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015	182	182	182	182
	Ensino Médio - Aprovação (%) - 2015	182	182	182	182
	Ensino Médio - Reprovação (%) - 2015	182	182	182	182
	Ensino Médio - Abandono (%) - 2015	182	182	182	182



Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	Ensino Médio - Abandono (%) - 2015, Ensino Médio - Reprovação (%) - 2015 ^a	.	Enter

a. Tolerance = ,000 limits reached.

b. Dependent Variable: Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,119 ^a	,014	,003	27206,88340	2,042

a. Predictors: (Constant), Ensino Médio - Abandono (%) - 2015, Ensino Médio - Reprovação (%) - 2015

b. Dependent Variable: Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015

Fonte: Elaborada pelas autoras (2018).

A conclusão sobre existência ou não de autocorrelação entre resíduos é feita comparando-se o valor obtido (2,042) com os valores da tabela de Durbin Watson. Costuma-se admitir que $DW < 1,5$ aponta para a autocorrelação positiva. $DW > 2,5$, aponta para a autocorrelação negativa.

- H_0 : Não existe correlação serial dos resíduos
- H_1 : Existe correlação serial dos resíduos

Figura 3 – Multicolinearidade.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	7232,130	5572,335		1,298	,196		
	Ensino Médio - Reprovação (%) - 2015	1032,949	651,399	,124	1,586	,115	,905	1,105
	Ensino Médio - Abandono (%) - 2015	-211,208	690,559	-,024	-,306	,760	,905	1,105

a. Dependent Variable: Quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar - 2015

Fonte: Elaborada pelas autoras (2018).

Regra do VIF: até 1 - sem multicolinearidade; de 1 até 10 - multicolinearidade aceitável; acima de 10 - multicolinearidade problemática.



Valores elevados de VIF (> 5) indicam colinearidade entre as variáveis da mesma forma que valores de Tolerância próximos de zero. De acordo com os resultados obtidos verifica-se que não existe problema de multicolinearidade no modelo.

$$V5_Freq_esc = 7232,17 + (0,124)V26_Ens_Médio_Reprov - (0,024)V27_Ens_Médio_Abandono \quad (1)$$

A taxa de aprovação do ensino médio não possui impacto sob a variável dependente que é a quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar – 2015, por isso a mesma foi excluída do modelo.

Porém, dado o aumento da quantidade de crianças e adolescentes de 6 a 17 anos acompanhados na frequência escolar, há um aumento de 0,124 na taxa de reprovação no ensino médio e, o aumento da quantidade de crianças e adolescentes de 6 a 17 anos, acompanhados na frequência escolar, há um impacto negativo de 0,024 na taxa de abandono no ensino médio.

4.3 Análise Discriminante

A análise discriminante foi feita com o objetivo de identificar os Municípios que alcançaram notas acima da média do IDEB dos 184 Municípios do Estado do Ceará, de acordo com o valor repassado do PBF, em 2015.

Os dados que foram trabalhados na AD, sendo a variável dependente dicotômica para os Municípios, classificados com boa nota no IDEB (nota maior e igual a 4,45 é igual a 1) e ruim nota no IDEB (nota menor que 4,45 é igual a 0), sendo 4,45 a média da projeção da nota do IDEB de 2015 que foi extraída do INEP.

A verificação dos Pressupostos: Normalidade multivariada;



Homocedasticidade; Igualdade das matrizes de covariância; Ausência de multicolinearidade; e Linearidade das relações entre variáveis.

Por meio do método simultâneo, a Figura 4, a seguir, mostra informações sobre a amostra geral. São identificados os casos selecionados para compor a função e os que foram separados ou não selecionados (amostra de teste).

Figura 4 – Analysis Case Processing Summary.

Unweighted Cases	N	Percent
Valid	184	8,3
Missing or out-of-range group codes	3	,1
At least one missing discriminating variable	0	,0
Excluded		
Both missing or out-of-range group codes and at least one missing discriminating variable	2032	91,6
Total	2035	91,7
Total	2219	100,0

Fonte: Elaborada pelas autoras (2018).

No quadro 2, observa-se o teste de igualdade de médias entre os grupos que identifica quais variáveis apresentam melhor poder de discriminação entre os grupos. Quanto menor o valor da estatística Lambda de Wilks, melhor o poder discriminatório da variável. No caso, a variável Taxa de aprovação 1º ao 5º ano é a que melhor discrimina os grupos em estudo.

Também é apresentado o teste ANOVA que corrobora a estatística Lambda de Wilks e aponta variáveis que têm poder discriminatório (Sig < 0,05).

Quadro 2 – Tests of Equality of Group Means.

Variáveis	Wilks' Lambda	F	df1	df2	Sig.
Valor total repassado do PBF (R\$)	,986	1,954	1	142	,164
Número de Matrícula ensino	,977	3,334	1	142	,070



fundamental final parcial urbano					
Número de Matrícula ensino fundamental final integral urbano	,983	2,478	1	142	,118
Número de Matrícula ensino fundamental Inicial parcial rural	,973	3,872	1	142	,051
Número de Matrícula ensino fundamental Inicial integral rural	,989	1,561	1	142	,214
Número de Matrícula ensino fundamental final parcial rural	,989	1,566	1	142	,213
Número de Matrícula ensino fundamental final integral rural	1,000	,054	1	142	,817
Número de Matrícula ensino fundamental final parcial urbano	,988	1,730	1	142	,191
Número de Matrícula ensino fundamental final integral urbano	,999	,080	1	142	,777
Taxa de aprovação 1º ao 5º ano	,842	26,709	1	142	,000
Taxa de Evasão ensino Fundamental Parcial urbano	,995	,757	1	142	,386
Taxa de Evasão integral urbano	1,000	,013	1	142	,908
Taxa de Evasão ensino fundamental Parcial rural	,998	,218	1	142	,641
Taxa de Evasão ensino fundamental integral rural	,985	2,101	1	142	,149


Fonte: Elaborado pelas autoras (2018).

4.4 Análise de Correlação Canônica

O teste de correlação canônica é aplicado para dados agrupados em dois conjuntos de variáveis, sendo cada conjunto composto por duas ou mais variáveis. A associação poderá ser positiva ou negativa. Se nula, elimina-se a possibilidade de correlação entre as variáveis.

Quadro 3 – Variáveis utilizadas para a análise de correlação canônica.

Variáveis dependentes (indicadores educacionais)	Variáveis independentes (indicadores socioeconômicos)
IDEB (2015) – V15	Índice de Desenvolvimento Humano - IDH (2010) – V34
Taxa de aprovação – Ensino Fundamental I (2015) – V17	PIB <i>per capita</i> (2014) – V37
Taxa de aprovação – Ensino Médio (2015) – V 25	População extremamente pobre (%) – 2010 – V33



	Emprego formal – Ensino Fundamental completo (2015) – V30
	Emprego formal – Ensino Médio completo (2015) – V31
	Emprego formal – Ensino Superior completo (2015) – V32
	Indústrias ativas (2015) – V38
	Densidade demográfica (2015) – V24

Fonte: Elaborado pelas autoras (2018).

A partir das variáveis selecionadas para os conjuntos de indicadores educacionais e indicadores socioeconômicos, realizando-se as análises, conformes apresentadas a seguir.

A) Teste de Wilks (λ)

O Lambda de Wilks é um teste multivariado de significância. Este teste verifica a correlação canônica, e tem como hipóteses: **H₀**: $R_c = 0$ (não há correlação entre o grupo de variáveis); **H₁**: $R_c \neq 0$ (há correlação entre o grupo de variáveis).

Na análise de variância, foi apontado o valor de Wilks (λ) igual a 0,67837, com estatística F de 2,99667 e P-valor equivalente a 0,000. Assim, a um nível de significância de 1% rejeita-se a hipótese nula. Logo, admite-se que existe correlação entre os grupos de variáveis, o que valida o modelo. Para os testes de Pillais, Hotellings e Roys, obteve-se resultado semelhante, para um nível de significância de 1%.

O tamanho do efeito (**proporção de variância compartilhada**) entre os dois grupos de variáveis, nas três funções (uma função gerada para cada variável dependente) é obtida através de $1 - \lambda$, o que resulta em 0,32163. Como o valor ideal é acima de 0,1 e, este resultado atende ao critério, conclui-se que o tamanho tem efeito moderado.

Cada função é avaliada individualmente através de autovalores.



Tabela 4 – Autovalores e correlação canônica.

Função	Autovalores	Correlação Canônica (R _c)	R _c ²
1	0,22475	0,42837	0,18351
2	0,17915	0,38979	0,15193
3	0,02075	0,14257	0,02033

Fonte: Elaborada pelas autoras (2018).

Dentre as funções geradas, na primeira tem-se que a relação é explicada em 18,35% pela função, sendo, portanto, esta função escolhida para interpretação. As demais funções têm poder de explicação inferior.

Tabela 5 – Análise de redução da dimensão.

Função	Lambda de Wilks	Estatística F	P-valor (de F)
1 a 3	0,67837	2,99667	0,000
2 a 3	0,83083	2,41349	0,003
3 a 3	0,97967	0,60513	0,726

Fonte: Elaborada pelas autoras (2018).

A tabela 5 apresenta que o primeiro conjunto de funções (1 a 3) mostra melhor nível de significância (1%) que os demais conjuntos. Considera-se também a função completa (1 a 3) mais representativa por ter menor valor de λ , o que aumenta o tamanho de efeito (proporção de variância compartilhada).

Tabela 6 – Síntese dos coeficientes.

Variável dependente	Coefficiente de correlação canônica padronizado	Coefficientes estruturais canônicos (R _s)	R _s ² (%)
V15	0,54957	0,78972	62,37
V17	0,61810	0,85028	72,30
V25	0,19667	0,20561	4,23

Fonte: Elaborada pelas autoras (2018).

Os pesos canônicos são os coeficientes canônicos padronizados. Os valores dos coeficientes indicam que as variáveis V17 (Taxa de aprovação – Ensino Fundamental I - 2015) e V15 (IDEB - 2015), respectivamente, são mais representativas no modelo do que a variável V25 (Taxa de aprovação – Ensino Médio - 2015). Estas variáveis também possuem os valores mais elevados de variância (R_s² (%)).



5. Considerações Finais

Dentre os modelos apresentados neste trabalho, podemos concluir, por meio da análise fatorial, que os fatores que apontam a observância real do impacto do programa bolsa família sob os indicadores educacionais, são os fatores socioeconômicos, fatores relacionados ao desenvolvimento da educação básica e, em seguida, os indicadores sob índices (PIB, pobreza, IDH, taxa de mortalidade infantil e urbanização).

Quanto ao modelo apresentado através da análise de regressão múltipla, onde as variáveis estimadas, não viola os pressupostos de modo a enfraquecer as análises e os resultados dos mesmos, excetuando o teste da normalidade, em que o mesmo pode ser corrigido através do método da transformação das variáveis.

A taxa de aprovação do ensino médio não possui impacto sob a variável dependente, porém, dado o aumento da quantidade de crianças e adolescentes de 6 a 17 anos, acompanhados na frequência escolar teve impacto negativo de 0,024 na taxa de abandono no ensino médio.

Segundo a análise discriminante, que tem por objetivo identificar os municípios que alcançaram notas acima da média do IDEB, dos 184 Municípios do Estado do Ceará, de acordo com o valor repassado do PBF, em 2015; o modelo mostrou-se com valor estatístico significante, a variável taxa de aprovação 1º ao 5º ano, como a que melhor discrimina os grupos em estudo.

Com isso, os modelos conseguem prever os municípios que terão notas do IDEB. No caso, 67,4% das notas do IDEB foram classificados corretamente, segundo o repasse do PBF, em 2015. As variáveis independentes, indicadas pelo modelo de regressão logística foram bem eficientes em prever as notas do IDEB que não seriam acima da média, percentual de 56,3% dos municípios e, mais eficientes em prever as notas acima da média dos Municípios do Estado do Ceará, chegando a 76,3% de



acertos. Ou seja, houve uma melhora no percentual de acertos em relação ao modelo base, significando dizer que o valor do repasse do PBF, em 2015, contribuiu para que um percentual maior dos municípios do Estado do Ceará atingissem notas acima da média do IDEB, configurando, assim, uma melhora no ensino da educação fundamental.

E, por fim, segundo a análise de correlação canônica, admite-se que existe correlação entre os grupos de variáveis, o que valida o modelo. E, esse modelo corrobora o que a análise fatorial revelou, considerando as variáveis taxa de aprovação do ensino fundamental I e as notas do IDEB como variáveis mais representativas dentro do modelo de correlação canônica; estas que já contemplam indicadores encontrados entre os três fatores mais importantes a serem analisados na pesquisa, segundo a análise factorial.



Referências

ANDRADE, T. A. Métodos estatísticos e econométricos aplicados à análise regional. In:

HADDAD, P. R. (Org.). *Economia regional: teoria e métodos de análise*. Fortaleza, CE: BNB, ETENE, 1989.

CASTELAR, Pablo Urano de Carvalho; MONTEIRO, Vitor Borges; LAVOR, Daniel Campos. Um estudo sobre as causas de abandono escolar nas escolas públicas do ensino médio do Estado do Ceará. In: BARRETO, Flávio Ataliba Flexa Daltro; MENEZES, Adriano Sarquis B. de; DANTAS, Regis Façanha; ALBUQUERQUE, Emanuel Lindemberg Silva; SOUSA, Fátima Juvenal de; GONÇALVES, Laura Carolina (Org.) **Economia do Ceará em Debate 2012**. Fortaleza: IPECE, 2013. P. 33-56.

COSTELLO, A. B; OSBORNE, J. W. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10, 7, p. 13-24, 2005.

Glewwe, P.; Kassouf, A. *O impacto do programa bolsa família no total de matrículas do ensino fundamental, taxas de abandono e aprovação*, in 'Seminário Itaú Internacional de Avaliação Econômica de Projetos Sociais', São Paulo, 2010.

FÁVERO, L.P. *et al. Análise de dados: modelagem multivariada para tomada de decisões*. Primeira Edição. Rio de Janeiro. Elsevier. 2009.

FERREIRA, C. M. de C. Métodos de regionalização. In: HADDAD, P. R. (Org.). *Economia Regional: teoria e métodos de análise*. Fortaleza, CE: BNB, ETENE, 1989.

FERREIRA, D. F. *Análise Multivariada*. Ministério da Educação e do Desporto, Universidade Federal de Lavras - Departamento de Ciências Exatas, 1996.

FREITAS, C.; PAZ, M. V.; NICOLA, D. S. Avaliando a intensidade da modernização da agropecuária gaúcha: uma aplicação de análise fatorial e cluster. *Revista Análise Econômica*. Porto Alegre, ano 25, n. 47, p. 121-149, 2007.

HAIR, JOSEPH F.; BLACK, WILLIAM C.; BABIN, BARRY J.; ANDERSON, ROLPH E.; TATHAM, RONALD L. *Análise multivariada de dados*. Trad. Adonai Schlup Sant'Anna. Rev. Maria Aparecida Gouvêa. 6. ed. Porto Alegre, Bookman, 2009, p. 688.



HÄRDLE, W.; SIMAR, L. *Applied Multivariate Statistical Analysis*. Berlin and Louvain-la-Neuve, 2003. E-book disponível em: <http://www.xplore-stat.de>.

HO, Robert. *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. CRC Press, 2006.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 3ª ed. New Jersey: Prentice Hall, 1992.

LEECH, Nancy L.; BARRETT, Karen C.; MORGAN, George A. *SPSS for Intermediate Statistics: Use and Interpretation*. LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS. Mahwah, New Jersey London. 2005.

MALHOTRA, N. K. *Marketing research*. 4 ed. New York: Prentice Hall, p. 864, 2004.

MANLY, B.F.J. *Multivariate statistical methods: a primer*. 2.ed. London: Chapman and Hall, 1998, p. 215.

MARRIOT, F.H.C. *The interpretation of multiple observations*. New York, Academic Press, 1974, p. 117.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005, p. 279.

NEY, M. G. *Educação e desigualdade de renda no meio rural brasileiro. Dissertação (Doutor em Economia Aplicada) – UNICAMP, Universidade de Campinas, Campinas, 124f., 2006.*

PONTILI, Rosângela. *A infra estrutura escolar e as características familiares influenciando a frequência e o atraso no ensino fundamental*. Piracicaba: ESALQ/USP, 2004.

REIS, E. *Estatística Multivariada*. Lisboa: Edições Silabo Ltda, 1997. 343 p.

ROCHA, L.A.; KHAN, A. S.; LIMA, P.V.P.S. Impacto do Programa Bolsa família sobre o bem estar das famílias beneficiadas no estado do Ceará. *In: Eveline Barbosa Silva Carvalho; Marcos Costa Holanda, Marcelo Ponte Barbosa. (Org.). Economia do Ceará em Debate 2008, Fortaleza, v. 1, p. 106-122, 2009.*

SAMPIERI, Roberto Hernández; COLLADO, Carlos Fernández; LUCIO, Pilar Baptista. *Metodología de la Investigación*. México: McGraw-Hill, 1997.



SILVA, Marta Maria Aguiar Sisnando; LIMA, Patrícia Verônica Pinheiro Sales. Indicadores educacionais e de renda e sua relação com pobreza e indigência no seminário brasileiro. **Rev. Econ. NE**, Fortaleza, v. 46, n. 2, p. 117-132, abr.-jun., 2015.

SOARES, J. F. Implementação de cotas na UFMG para alunos egressos de escolas públicas. In: *Universidade e democracia: experiências e alternativas para a ampliação do acesso à Universidade pública*. Belo Horizonte: Editora UFMG, 2004.

VALENTIN, J. *Ecologia numérica: uma introdução à análise multivariada de dados ecológicos*. Rio de Janeiro: Interciência, 2000, p. 117.

ZAMBRANO, C.; LIMA, J. E. *Análise Estatística Multivariada de Dados Socioeconômicos*. In: SANTOS, M. L.; VIEIRA, W. C. (Ed.). *Métodos quantitativos em economia*. Viçosa, UFV, p. 556-577, 2004.

WARD, J. *Hierarchical grouping to optimize an objective function*, Journal of the American Statistical Association, 1963, 58:236-244.