



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE

AVALIAÇÃO DA EFICIÊNCIA NA ATENÇÃO BÁSICA À SAÚDE NO BRASIL E REGIÃO NORDESTE

Camila Ribeiro da Silva

João Pessoa/PB

2017

CAMILA RIBEIRO DA SILVA

**AVALIAÇÃO DA EFICIÊNCIA NA ATENÇÃO BÁSICA À SAÚDE
NO BRASIL E REGIÃO NORDESTE**

Dissertação apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde – Nível Mestrado do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como requisito regulamentar para a obtenção do título de Mestre.

Linha de pesquisa: Modelos de Decisão

Orientadores:

Profa. Dra. Tatiene Correia de Souza

Profa. Dra. Caliandra Maria Bezerra Luna Lima

João Pessoa/PB

2017

Catálogo na publicação
Seção de Catalogação e Classificação

S586a Silva, Camila Ribeiro da.
Avaliação da eficiência na atenção básica à saúde no Brasil e região
Nordeste / Camila Ribeiro da Silva. - João Pessoa, 2017.
76 f. : il.

Orientadora: Tatiene Correia de Souza.
Orientadora: Caliandra Maria Bezerra Luna Lima.
Dissertação (Mestrado) - UFPB/CCEN

1. Modelos de decisão - Saúde. 2. Eficiência técnica. 3. Atenção básica -
Brasil. I. Título.

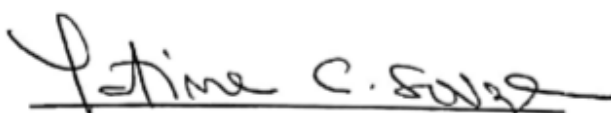
UFPB/BC

CAMILA RIBEIRO DA SILVA

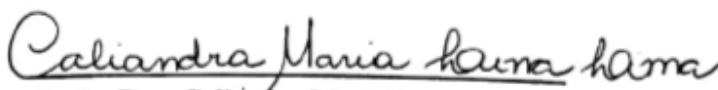
**AVALIAÇÃO DA EFICIÊNCIA NA ATENÇÃO BÁSICA À SAÚDE
NO BRASIL E REGIÃO NORDESTE**

João Pessoa, 17 de novembro de 2017.

BANCA EXAMINADORA



Profa. Dra. Tatiene Correia de Souza
Orientador (DE/UFPB)



Profa. Dra. Caliandra Maria B. Luna Lima
Orientador (DFP/UFPB)

Profa. Dra. Tarciana Liberal Pereira
Membro Interno (DE/UFPB)

Profa. Dra. Maria Eliane Moreira Freire
Membro Externo (PPGENF/UFPB)

*Este trabalho é dedicado aos meus pais, José e Rizete, e a minha
querida amiga Lusinete (in memoriam), com amor.*

AGRADECIMENTOS

Agradeço a Deus por estar sempre iluminando meu caminho...

À minha família pelo amor, apoio e compreensão ao longo da minha jornada. Especialmente aos meus pais, por sempre incentivarem meus estudos e por celebrarem todas as minhas conquistas. Aos meus irmãos, Carol e Gabriel, e ao meu pequeno Heitor, anjo de nossas vidas, que faz tudo valer a pena e torna meus dias mais felizes.

Ao meu amor Alisson, por sua amizade, respeito e amor singular dedicados a mim. Obrigada por me proporcionar crescer e amadurecer do seu lado.

À professora Tatiene pela orientação constante, amizade e convivência enriquecedora. Por todo aprendizado, pelos conselhos e lições de vida. Hoje posso dizer que tenho um pouco da personagem "Pollyanna" (história da menina que via um lado bom em tudo que acontecia), graças a você. Obrigada!

À professora Caliandra, sempre gentil e atenciosa comigo. Obrigada pela confiança, orientação e direcionamentos para a conclusão desta dissertação.

Ao professor Hemílio, pelo incentivo e ensinamentos desde a graduação e pela presteza em todos os momentos que o recorri.

As professoras Maria Conceição Sampaio e Tarciana Liberal, pela gentileza na disponibilização da base de dados para a construção deste trabalho.

Aos amigos: Ramon, Pedro, Jodavid, Luana, Marina, Marília, Geisislane e Saul pelos momentos de amizade, convivência enriquecedora e aprendizado constante. Aos colegas que fiz ao longo do mestrado: Luciano Bellas, Rafaela Prima, Kaline, Juliana, Emanuelle Cyntia e Rayanne, meu conhecimento "interdisciplinar" ficou mais abrangente com a participação de vocês.

A Banca Examinadora pelas contribuições para a melhoria deste trabalho.

A todos os Professores do DE-UFPB, por contribuírem para minha formação acadêmica.

Aos funcionários do Departamento de Estatística e do Programa de Pós-Graduação em Modelos de Decisão e Saúde.

A CAPES, pelo apoio financeiro.

A persistência é o menor caminho do êxito.
(Charles Chaplin)

RESUMO

Esta dissertação teve por objetivo identificar os fatores que influenciam a eficiência técnica na aplicação dos recursos destinados à atenção básica no Brasil, região nordeste e no estado da Paraíba. Para isso foi utilizado o modelo de regressão beta inflacionado, uma vez que a variável resposta apresenta assimetria e está restrita ao intervalo $(0, 1]$. A amostra utilizada foi de dados secundários obtidos através de fontes de domínio público. O número de municípios observados foram, respectivamente, 3.409, 1.255 e 154, para o Brasil, nordeste e Paraíba. Após a análise descritiva dos dados verificamos que, do total de municípios da amostra (3.409), apenas 15,6% são considerados plenamente eficientes na aplicação dos recursos destinados à atenção básica. As regiões norte e nordeste apresentaram, em geral, maior proporção de crianças com menos de 5 anos de idade, quando comparadas às demais regiões. Observou-se que apenas 0,12% dos municípios observados possui cobertura completa quanto aos serviços de saneamento básico, sendo eles Espírito Santo do Pinhal, Ituverava, Junqueirópolis e São José dos Campos, todos localizados no estado de São Paulo. Após o ajuste dos modelos de regressão verificou-se que, no modelo que considera dados para o Brasil, as variáveis índice Firjan de desenvolvimento municipal na saúde, a população municipal e o percentual de famílias com serviços de saneamento básico, exercem efeito positivo na eficiência média do município. Ao passo que na região nordeste, as variáveis gasto per capita em atenção básica, população municipal e índice Firjan de desenvolvimento na educação exercem efeito positivo na eficiência média da atenção básica. Para o modelo com dados da Paraíba, as variáveis número de consultas por estabelecimento e média de médicos por estabelecimento da atenção básica influenciam de forma negativa a eficiência média do município.

Palavras-chave: Atenção básica. Eficiência. Regressão beta inflacionada.

ABSTRACT

This dissertation aimed to identify the factors that influence the technical efficiency in the application of resources destined to primary care in Brazil, in the northeastern region and in the state of Paraíba. For this, the inflated beta regression model was used, since the response variable presents asymmetry and is restricted to the range $(0, 1]$. The sample used was from secondary data obtained from public domain sources. After the descriptive analysis of the data we verified that, of the total of municipalities of the sample (3,409), only 15.6% are considered fully efficient in the application of the resources destined to the basic attention. The northern and northeastern regions presented, in general, a greater proportion of children under 5 years of age when compared to the other regions. It was observed that only 0.12 % of the municipalities observed have complete coverage regarding basic sanitation services, being Espírito Santo do Pinhal, Ituverava, Junqueirópolis and São José dos Campos, all located in the state of São Paulo. After adjusting the regression models, it was verified that, in the model that considers data for Brazil, the Firjan index of municipal development in health, the municipal population and the percentage of families with basic sanitation services exert a positive effect on efficiency average of the municipality. While in the Northeast, the variables per capita expenditure on basic care, municipal population and Firjan index of development in education exert a positive effect on the average efficiency of basic care. For the model with data from Paraíba, the variables number of visits per establishment and average number of doctors per establishment of primary care negatively influence the average efficiency of the municipality.

Keywords: *Basic attention. Efficiency. Inflation beta regression.*

LISTA DE ILUSTRAÇÕES

Figura 1 – Histograma de frequências e <i>box-plot</i> da variável Eficiência.	32
Figura 2 – Histograma de frequência da variável Eficiência para os dados da região nordeste.	34
Figura 3 – Histograma de Frequência e <i>box-plot</i> da variável Eficiência para os dados da Paraíba.	36
Figura 4 – Gráfico dos resíduos quantis aleatorizados e normal de probabilidade para o modelo de regressão obtido com dados do Brasil.	41
Figura 5 – Gráficos dos resíduos versus as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) do modelo para o Brasil.	42
Figura 6 – Distância de Cook versus ordem das observações para a parte discreta e contínua do modelo de regressão obtido a partir dos dados do Brasil.	43
Figura 7 – Gráfico dos resíduos quantis aleatorizados e normal de probabilidade para o modelo de regressão obtido a partir dos dados da região nordeste.	47
Figura 8 – Gráficos dos resíduos para as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) do modelo de regressão obtido a partir dos dados da região nordeste.	47
Figura 9 – Distância de Cook versus ordem das observações para o ajuste da região nordeste.	48
Figura 10 – Gráficos de resíduos quantis aleatorizados versus ordem das observações e de probabilidade normal com envelopes simulados para o modelo de regressão obtido a partir dos dados da Paraíba.	52
Figura 11 – Gráficos dos resíduos versus as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) para o modelo de regressão obtido a partir dos dados da Paraíba.	53
Figura 12 – Distância de Cook versus ordem das observações para o ajuste da Paraíba.	53
Figura 13 – Mapa das eficiências estimadas para os dados da Paraíba.	55
Figura 14 – Mapa das eficiências estimadas para os dados da região nordeste.	56
Figura 15 – Mapa das eficiências estimadas para os dados do Brasil.	57

LISTA DE TABELAS

Tabela 1 – Descrição das variáveis utilizadas no ajuste dos modelos de regressão beta inflacionados.	27
Tabela 2 – Descrição das variáveis contínuas e frequência das variáveis dummy para os dados do Brasil.	30
Tabela 3 – Matriz de correlação para as variáveis contínuas	31
Tabela 4 – Descrição das variáveis contínuas e frequência das variáveis dummy para a região nordeste.	33
Tabela 5 – Descrição das variáveis contínuas e frequência das variáveis dummy para os dados da Paraíba.	36
Tabela 6 – Modelo de regressão beta inflacionado com dispersão variável para os dados do Brasil.	40
Tabela 7 – Variação percentual das estimativas dos parâmetros sem as observações 1171, 3106 e 1407, que correspondem aos municípios de Aracaju-SE, Porto Alegre-RS e Madre de Deus-BA.	44
Tabela 8 – Modelo de regressão beta inflacionado com dispersão variável ajustado para a região nordeste.	46
Tabela 9 – Variação percentual das estimativas dos parâmetros para o modelo com dados da região nordeste sem as observações 252, 840 e 1120, que representam os municípios de Teresina-PI, Maceió-AL e Madre de Deus-BA, respectivamente.	49
Tabela 10 – Modelo de regressão beta inflacionado com dispersão variável ajustado para o estado da Paraíba.	51
Tabela 11 – Variação percentual das estimativas dos parâmetros para o modelo com dados da Paraíba sem as observações 6, 154 e 75, que referem-se aos municípios de Algodão de Jandaíra, Zabelê e Malta.	54

SUMÁRIO

1	INTRODUÇÃO	12
1.1	EFICIÊNCIA NOS SERVIÇOS DE SAÚDE	13
2	OBJETIVO GERAL	16
2.1	OBJETIVOS ESPECÍFICOS	16
3	REFERENCIAL TEÓRICO	17
3.1	MODELOS DE REGRESSÃO	17
3.1.1	Modelo de Regressão Beta Inflacionado em Zero ou Um	18
3.2	ADEQUABILIDADE DO MODELO	21
3.3	ANÁLISE DE DIAGNÓSTICO	21
3.3.1	Resíduos Quantis Aleatorizados	22
3.3.2	Medidas de Influência	22
3.3.2.1	Distância de Cook	22
4	METODOLOGIA DE PESQUISA	25
4.1	TIPOLOGIA DA PESQUISA	25
4.2	POPULAÇÃO E AMOSTRA	25
4.3	OBTENÇÃO DOS DADOS	25
5	ANÁLISE DE DADOS	27
5.1	ANÁLISE DESCRITIVA PARA OS DADOS DO BRASIL	27
5.2	ANÁLISE DESCRITIVA PARA OS DADOS DA REGIÃO NORDESTE	30
5.3	ANÁLISE DESCRITIVA PARA OS DADOS DA PARAÍBA	34
6	AJUSTES DOS MODELOS DE REGRESSÃO BETA INFLACIONADOS	37
6.1	ESPECIFICAÇÃO DO MODELO PARA OS DADOS DO BRASIL	37
6.2	ESPECIFICAÇÃO DO MODELO PARA OS DADOS DA REGIÃO NORDESTE	44
6.3	ESPECIFICAÇÃO DO MODELO CONSIDERANDO OS MUNICÍPIOS PARA O ESTADO DA PARAÍBA	49
6.4	MAPAS DAS EFICIÊNCIAS ESTIMADAS ATRAVÉS DOS MODELOS DE REGRESSÃO BETA INFLACIONADOS	54
6.5	ANÁLISE COMPARATIVA DOS RESULTADOS	58
7	CONSIDERAÇÕES FINAIS	59
7.1	TRABALHOS FUTUROS	60

REFERÊNCIAS	62
APÊNDICE A – <i>SCRIPT</i> UTILIZADO NO <i>SOFTWARE</i> R	68

1 INTRODUÇÃO

A Atenção Básica (AB) constitui o primeiro nível de assistência sanitária, responsável pelo conjunto de serviços de saúde direcionados às necessidades do indivíduo e a integração dos níveis de atenção à saúde. Segundo Starfield (1998) a AB fundamenta-se no conjunto das seguintes características: acessibilidade ao serviço diante de uma primeira necessidade (caracterizando o primeiro contato), o cuidado sendo regular e associado à relação pessoal entre profissionais e pacientes (longitudinalidade), o acesso a toda forma de atenção à saúde (integralidade) e a continuidade da assistência (coordenação), além de ter como foco a família e orientação comunitária, sendo estas duas últimas características, em específico, do Programa de Saúde da Família.

Nos últimos anos a AB tem ganhado abrangência expressiva no país. No final de 2013, haviam 36 mil equipes de Estratégia de Saúde da Família, 300 mil agentes comunitários de saúde e 23 mil equipes de saúde bucal (RODRIGUES et al., 2014). A importância dada a esse nível de atenção pode ser explicada pelo fato de que um sistema eficiente em cuidados primários de saúde, proporciona melhorias substanciais de saúde a custos relativamente pequenos, principalmente quando comparados aos níveis secundário e terciário, que compreendem a assistência prestada pelas unidades de pronto atendimento e hospitais escolas, e a assistência de hospitais de grande porte, respectivamente. Pode-se comprovar, através de ampla literatura, os inúmeros benefícios que um sistema de atenção básica pode proporcionar, estando bem estruturado e interligado com os níveis superiores (ATUN, 2004; ENGSTROM; FOLDEVI; BORQUIST, 2001).

Entre esses benefícios estão incluídos melhores resultados de saúde (FORREST; STARFIELD, 1996; STARFIELD, 1998) e atenção maior na prevenção e na detecção precoce de problemas de saúde (SHEA et al., 1992). Além disso, o uso de forma regular dos serviços de AB proporciona redução de mortalidade por todas as causas (SHI, 1992; VILLALBI et al., 1999), reduz as re-hospitalizações (WEINBERGER; ODDONE; HENDERSON, 1996), as consultas com especialistas (WOODWARD et al., 2004), induz um menor uso de serviços de emergência e melhora a detecção de fatores adversos de intervenções médicas (KRAVITZ; DUAN, 2004). Os impactos benéficos da AB, se mantêm mesmo quando controladas outras características da população, tais como renda, sexo, etnia e adoção de comportamento de risco (SHI; STARFIELD, 2000).

Estudos afirmam, ainda, que os sistemas de referência, onde o primeiro contato do paciente se faz nas unidades de AB, conduzem à tratamentos mais apropriados, efetivos e mais baratos (STARFIELD, 1998). Segundo Shi et al. (2003), evidência empírica disponível revela que o acesso ampliado aos serviços de atenção básica, ao contrário dos mais especializados, leva a uma distribuição mais igualitária das condições de saúde auxiliando, inclusive, para reduzir os

efeitos adversos da concentração de renda sobre essas condições.

Apesar de estratégias e iniciativas para aprimorar o financiamento da AB, sabe-se que os municípios apresentam diversos tipos de dificuldades, em especial os municípios menores, que não contam com recursos próprios suficientes para financiar as despesas não cobertas pelo governo federal (MENDES; MARQUES, 2014).

Neste sentido, faz-se necessário à avaliação do desempenho municipal na provisão dos serviços de saúde no âmbito da AB. Conforme Fadel et al. (2009), as tentativas são várias e constantes, não somente no contexto de potencializar os serviços públicos em saúde, mas também de garantir o acesso dos usuários a todos os níveis de atenção. Entretanto, segundo opinião dessas autoras, como cada município possui suas características e particularidades como as desigualdades geográficas, culturais e socioeconômicas, a questão de saúde pública torna-se ainda mais complexa.

1.1 EFICIÊNCIA NOS SERVIÇOS DE SAÚDE

No contexto de temas relacionados à políticas públicas, é importante abordar os aspectos da qualidade dos serviços que são prestados à população. Fonseca e Ferreira (2009) acreditam que entre os pesquisadores da área da saúde e entre os chefes de governo existe uma preocupação em avaliar a qualidade e eficácia da prestação dos serviços de saúde no setor público. Os autores entendem ainda que, é cada vez mais necessário o desenvolvimento e utilização de métodos que possibilitem a avaliação da eficiência do setor saúde, de modo a tornar possível a identificação do nível máximo de serviços que podem ser oferecidos considerando os recursos disponíveis no âmbito da gestão pública.

A avaliação sistemática, contínua e eficaz pode ser um instrumento fundamental para melhorar os resultados dos investimentos governamentais (COSTA; CASTANHAR, 2003). Ademais, também pode fornecer dados mais consistentes para os gestores e formuladores de políticas públicas. Os autores sugerem que os gestores públicos utilizem critérios de avaliação para comparação dos resultados obtidos em um programa.

Conforme Ala-Harja e Helgason (2000), a avaliação de programas deve ser vista como mecanismo de melhoria no processo de tomada de decisão e prestação de contas. Através do uso de informações mais confiáveis, os governantes podem obter um conhecimento mais relevante dos resultados sobre as políticas e programas públicos melhorando sua concepção, bem como sua implementação. Para Varela (2008) a avaliação de desempenho no Sistema Único de Saúde (SUS) se apresenta de forma complexa, devido a necessidade em conhecer a dinâmica do setor, de transformá-la em indicadores sociais e estabelecer relações entre eles. Por essa razão, para que a avaliação seja completa, ela deve usar informações que expressem o ponto de vista dos principais envolvidos com esse sistema.

Conforme Viacava et al. (2012), eficiência está relacionada a relação entre o custo e o impacto dos serviços sobre a saúde da população, mantendo-se um nível de qualidade determinado. A eficiência é definida por Peña (2008) como sendo a relação entre os insumos e métodos necessários (*inputs*) no processo produtivo de modo que gerem o máximo de produto (*output*). Desse modo, a eficiência no âmbito da saúde pública consiste na alocação mais acertada dos recursos públicos, para oferecer a sociedade a maior quantidade possível de produtos e serviços.

No Brasil, alguns estudos avaliaram a eficiência na aplicação de recursos públicos municipais e buscaram identificar os impactos sobre os indicadores de atenção à saúde (VARELA; MARTINS; FÁVERO, 2012; FARIA; JANNUZZI; SILVA, 2008; MARINHO, 2003). Conforme Varela, Martins e Fávero (2012), que realizaram um estudo com municípios paulistas, seria possível aumentar, expressivamente, a quantidade de serviços de saúde prestados à população sem a necessidade de novos custos, na maioria dos municípios. Santos (2008) comparou a eficiência de 3.370 municípios na alocação de recursos públicos da saúde, no período de 1997 a 2000, e concluiu que a intensidade de utilização do capital mostrou-se negativamente correlacionada com a ineficiência. Isto é, quanto maior a intensidade de utilização do capital, menos ineficiente é o município. Varela (2008) e Dias (2010) fizeram uso de modelo de regressão linear para avaliar a influência de determinadas variáveis na eficiência da atenção básica.

Um estudo divulgado pelo jornal Folha de São Paulo em parceria com o Datafolha, apresentou resultados de eficiência para 5.281 municípios brasileiros (Folha de São Paulo, 2017). Com base nos resultados obtidos foi possível traçar um ranking de eficiência dos municípios e gerar informações que auxiliam no direcionamento de iniciativas para políticas públicas. A eficiência obtida no estudo levou em conta os indicadores nos aspectos de saúde, educação e saneamento básico. Segundo o critério adotado no estudo, o município com valor de eficiência acima de 0,5 foi considerado eficiente. Sendo apenas 24% das cidades observadas.

Com relação ao indicador de saúde verificou-se grandes discrepâncias entre alguns municípios, por exemplo, as cidades de Cachoeira da Prata-MG, Presidente Lucena-RS e Altaneira-CE, apresentaram indicadores em saúde de 0,687; 0,786 e 0,586, respectivamente, e foram considerados eficientes de acordo com o ranking. Por outro lado, os municípios de Boa Viagem-CE e Campo Alegre-AL, registraram indicadores de saúde bem inferiores quanto aos municípios citados anteriormente, sendo de 0,283 e 0,387, respectivamente, e foram classificados como ineficientes. Na Paraíba, João Pessoa obteve indicador de saúde igual a 0,52, enquanto cidades menores como Santa Luzia, Araruna e Cabedelo, apresentaram indicadores de 0,56; 0,57 e 0,66, respectivamente. Entretanto, apesar dos resultados relevantes do respectivo estudo, salientamos que a eficiência mensurada foi baseada em três aspectos distintos (saúde, educação e saneamento).

Com relação aos trabalhos realizados na Paraíba, destaca-se o de Andrade (2009), que avaliou a eficiência técnica do SUS no estado da Paraíba utilizando as microrregiões como

unidades de análise. Costa e Filho (2013), que usaram o método de fronteira estocástica para avaliar a eficiência técnica do sistema hospitalar do SUS no estado da Paraíba, através dos municípios sedes das Gerências Regionais de Saúde. Os resultados indicaram para a ineficiência do sistema hospitalar do SUS no estado.

Desse modo, torna-se necessário a viabilização de ferramentas que possam nortear a aplicação eficiente dos recursos disponíveis, com ênfase nos gastos em AB. Neste quesito, destaca-se a relevância do presente estudo, tendo em vista a possibilidade de obtenção de modelos capazes de descrever a relação da eficiência na aplicação dos recursos destinados à AB com outros fatores, viabilizando a resolução de problemas práticos. Além disso, o estudo é inovador ao utilizar uma metodologia de análise que permite acomodar de forma flexível a estrutura da variável de interesse (eficiência técnica na aplicação dos recursos destinados á AB).

A presente dissertação encontra-se dividida em sete capítulos, incluindo esta introdução. O Capítulo 2 apresenta os objetivos gerais e específicos definidos para o estudo. No Capítulo 3 é apresentado o modelo de regressão beta inflacionado, bem como formas para verificar a adequabilidade do modelo ajustado e análise de diagnóstico. No Capítulo 4 encontram-se as informações acerca dos procedimentos metodológicos. A descrição completa dos dados é apresentada no Capítulo 5. No Capítulo 6 encontram-se os resultados dos ajustes dos modelos de regressão beta inflacionados. Por fim, no capítulo 7, apresentamos as considerações finais.

2 OBJETIVO GERAL

O objetivo geral desta dissertação é modelar o escore de eficiência técnica no gerenciamento de recursos destinados à Atenção Básica no Brasil e na região nordeste, via modelo de regressão beta inflacionado. Como a variável de interesse possui valores limitado ao intervalo $(0,1]$, os modelos lineares usuais tornam-se inadequados, visto que são necessárias transformações na variável original. Dessa forma, foi utilizado o modelo de regressão beta inflacionado proposto por Ospina e Ferrari (2012) para a modelagem dos escores de eficiência na Atenção Básica.

2.1 OBJETIVOS ESPECÍFICOS

- Obter um modelo de regressão beta inflacionado para descrever a eficiência na atenção básica no estado da Paraíba, bem como a comparação dos resultados com os obtidos para o Brasil e região nordeste.
- Construir mapas para o Brasil, Nordeste e Paraíba baseados nos escores de eficiência.

3 REFERENCIAL TEÓRICO

3.1 MODELOS DE REGRESSÃO

A análise de regressão pode ser compreendida como uma área da estatística destinada a estudar relações entre variáveis. Essa relação podendo ser representada através de modelos matemáticos capazes de descrever o comportamento de determinadas variáveis. A ideia consiste em investigar como uma variável resposta ou dependente (Y), esta relacionada a um conjunto de variáveis explicativas ($X_1, \dots, X_n$), as quais acredita-se influenciar a variabilidade da variável dependente (MONTGOMERY; PECK; VINING, 2012). Conforme Paula (2004), durante muito tempo os modelos normais lineares foram utilizados para descrever a maioria dos fenômenos aleatórios. Em diversas situações práticas utiliza-se o modelo de regressão normal linear (SLATER; MARCHIONI; VOCI, 2007; COELHO-BARROS et al., 2008; RIBEIRO, 2012; BASTOS; GUIMARÃES; SEVERO, 2015), mesmo nos casos em que a suposição de normalidade para a variável resposta não é adequada, aplica-se uma transformação na variável com o objetivo de alcançar a respectiva suposição.

A família de transformações mais conhecida foi proposta por Box e Cox (1964). O objetivo da transformação quando aplicada a um conjunto de dados, é de tornar a variável transformada aproximadamente normal, linear e com variância constante em relação a um conjunto de variáveis explicativas $X_1, \dots, X_n$. Porém, esse enfoque possui a desvantagem de que a interpretação dos parâmetros não pode ser realizada de forma direta em termos da resposta original (PAOLINO, 2001), dependendo da transformação. Uma das transformações mais usuais neste contexto é a logarítmica. Ela apresenta boas propriedades matemáticas...

Com o avanço tecnológico, a geração e o acúmulo de informações, tornou-se conveniente o desenvolvimento de modelos mais gerais e flexíveis, quanto as suposições de normalidade e variância constante para o erro aleatório, e a adoção de uma forma funcional linear entre regressando e regressores, presentes no modelo normal linear. Desse modo, foram surgindo ao longo dos anos modelos que buscaram acomodar de forma mais adequada o tipo ou natureza da variável a ser estudada, com o propósito de obter modelos mais próximos da realidade. Entre essas classes pode-se destacar os Modelos Lineares Generalizados (*Generalized Linear Models* - *GLM*), os Modelos Aditivos Generalizados (*Generalized Additive Models* - *GAM*), os quais ocupam lugar de destaque na literatura de Nelder e Wedderburn (1972) e Hastie e Tibshirani (1990), os Modelos Lineares Mistos (efeitos aleatórios), entre outros. Entretanto, esses modelos apesar de serem mais flexíveis, em alguns aspectos, quando comparados ao modelo normal linear, não apresentam uma estrutura que permita a modelagem de uma variável que esteja restrita ao intervalo unitário, ou seja, $y \in (0, 1)$.

3.1.1 Modelo de Regressão Beta Inflacionado em Zero ou Um

A distribuição beta é bastante utilizada para modelar dados que pertencem ao intervalo $(0,1)$, a exemplo de taxas ou proporções. Entre algumas aplicações do modelo de regressão beta estão: Simas, Barreto-Souza e Rocha (2010), Silva e Souza (2014), Júnior e Souza (2015), Pereira, Souza e Cribari-Neto (2014), Souza et al. (2016), entre outras. Ferrari e Cribari-Neto (2004) propuseram o modelo de regressão beta como uma alternativa de suprir as limitações apresentadas pelos modelos tradicionais (Modelos de regressão normal linear). Entretanto, para algumas situações, a variável de interesse pode assumir, inclusive, os valores 0 (zero) ou 1 (um), tornando o modelo de regressão beta inadequado, uma vez que essa distribuição modela variáveis distribuídas de forma contínua no intervalo $(0,1)$.

Desse modo, para se trabalhar com dados restritos aos intervalos $(0, 1]$, $[0, 1)$ e $[0, 1]$, Ospina e Ferrari (2010) apresentaram distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli degenerada em zero e/ou um, e essas distribuições são denominadas de beta inflacionadas. Uma vez que os dados de eficiência da presente dissertação são inflacionados em um, apresentaremos a distribuição beta inflacionada em zero ou um.

A função de distribuição acumulada do modelo beta inflacionado é escrita da seguinte forma

$$BI_c(y; \alpha, \mu, \phi) = \alpha \mathbb{I}_{[c,1]}(y) + (1 - \alpha)F(y; \mu, \phi),$$

em que $0 < \alpha < 1$, $0 < \mu < 1$, $\phi > 0$ e $\mathbb{I}_{[c,1]}(y)$ é uma função indicadora que assume valor 1 se $y \in A$ e 0 caso contrário. Aqui, $F(\cdot; \mu, \phi)$ é a função de distribuição acumulada beta $B(\mu, \phi)$ e $\alpha \in (0, 1)$ é o parâmetro de mistura da distribuição que é dado por $\alpha = Pr(y = c)$. A função BI_c tem um ponto de massa em $y = c$, em que $c = 0$ ou $c = 1$, não sendo, dessa forma, absolutamente contínua.

A função densidade de probabilidade de y é dada por

$$bi_c(y; \alpha, \mu, \phi) = \begin{cases} \alpha, & y = c, \\ (1 - \alpha)f(y; \mu, \phi), & y \in (0, 1), \end{cases} \quad (3.1)$$

em que $f(y; \mu, \phi)$ é a função densidade da $B(\mu, \phi)$. Para esta distribuição, $E(y) = \alpha c + (1 - \alpha)\mu$ e $Var(y) = (1 - \alpha)\frac{V(\mu)}{\phi + 1} + \alpha(1 - \alpha)(c - \mu)^2$, em que $V(\mu) = \mu(1 - \mu)$. Dessa forma, no modelo beta inflacionado a resposta é estimada através de $\hat{E}(y) = \hat{\alpha}c + (1 - \hat{\alpha})\hat{\mu}$, em que μ é a média condicional de y . Isto é, a média dado que y pertence ao intervalo $(0, 1)$. Se $c = 0$, a distribuição 3.1 é denominada distribuição beta inflacionada no ponto zero e representamos por $y \sim BEZI(\alpha, \mu, \phi)$. Se $c = 1$, denominamos a distribuição 3.1 de beta inflacionada no ponto um e representamos por $y \sim BEOI(\alpha, \mu, \phi)$.

Ospina e Ferrari (2012) propuseram modelos de regressão beta inflacionados com dispersão variável, que são extensões naturais do modelo de regressão beta proposto por Ferrari

e Cribari-Neto (2004), assumindo que a distribuição da variável dependente é beta inflacionada. Uma parametrização apresentada pelos autores permite modelar de forma direta a média da distribuição beta, a precisão e a massa de probabilidade em zero ou um, usando preditores lineares e funções de ligação adequadas, estendendo assim, o modelo de regressão beta inflacionado em um (HOOF, 2007) e o modelo de regressão beta inflacionado em zero (COOK; KIESCHNICK; MCCULLOUGH, 2006).

Sejam $y_1, \dots, y_n$ variáveis aleatórias independentes com distribuição beta inflacionada no ponto c ($c = 0$ ou $c = 1$). O modelo de regressão beta inflacionado em c é definido supondo que a média condicional de y_t , a massa de probabilidade em c e o parâmetro de precisão satisfazem as seguintes relações funcionais:

$$h(\alpha_t) = \sum_{i=1}^M z_{ti} \gamma_i = \zeta_t, \quad (3.2)$$

$$g(\mu_t) = \sum_{i=1}^m x_{ti} \beta_i = \eta_t, \quad (3.3)$$

$$b(\phi_t) = \sum_{i=1}^q s_{ti} \lambda_i = k_t, \quad (3.4)$$

$t = 1, \dots, n$, em que $\gamma = (\gamma_1, \dots, \gamma_M)^\top$, $\beta = (\beta_1, \dots, \beta_m)^\top$ e $\lambda = (\lambda_1, \dots, \lambda_q)^\top$ são vetores de parâmetros de regressão desconhecidos, tais que $\gamma \in \mathbb{R}^M$, $\beta \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}^q$, $z_{t1}, \dots, z_{tM}$, $x_{t1}, \dots, x_{tm}$ e $s_{t1}, \dots, s_{tq}$ são observações de variáveis conhecidas ($m + M + q < n$) que podem coincidir total ou parcialmente. As funções de ligação $h : (0, 1) \rightarrow \mathbb{R}$, $g : (0, 1) \rightarrow \mathbb{R}$ e $b : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Entre as funções de ligação mais utilizadas para α e μ estão a função logit: $g(\mu_i) = \log((\mu_i)/(1 - \mu_i))$, a função probit: $g(\mu_i) = \Phi^{-1}(\mu_i)$, em que $\Phi(\cdot)$ é a função de distribuição normal padrão, a complementar log-log: $g(\mu_i) = \log\{-\log(1 - \mu_i)\}$ e função log-log: $g(\mu_i) = -\log\{-\log(\mu_i)\}$, e a função logarítmica: $h(\phi) = \log(\phi)$ ou raiz quadrada: $h(\phi) = \sqrt{\phi}$ para ϕ . Uma discussão mais detalhada pode ser encontrada em Andrade (2007).

No modelo de regressão beta inflacionado temos que os parâmetros α_t , μ_t e ϕ_t são definidos como funções de γ , β e λ , respectivamente, através das relações apresentadas em 3.2, 3.3 e 3.4, ou seja, $\alpha_t = h^{-1}(\zeta_t)$, $\mu_t = g^{-1}(\eta_t)$ e $\phi_t = b^{-1}(k_t)$. Desse modo, a estimação dos parâmetros é realizada maximizando a função de log-verossimilhança em relação ao vetor $\theta = (\gamma^\top, \beta^\top, \lambda^\top)$, podendo ser expressa por

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \lambda),$$

em que

$$\ell_1(\gamma) = \sum_{t=1}^n \ell_t(\alpha_t) \quad \text{e} \quad \ell_2(\beta, \lambda) = \sum_{t: y_t \in (0,1)} \ell_t(\mu_t, \phi_t),$$

com

$$\begin{aligned} \ell_t(\alpha_t) &= \mathbb{I}_{\{c\}}(y_t) \log \alpha_t + (1 - \mathbb{I}_{\{c\}}(y_t)) \log(1 - \alpha_t), \\ \ell_t(\mu_t, \phi_t) &= \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) - \log \Gamma((1 - \mu_t) \phi_t) + (\mu_t \phi_t - 1) \log y_t \\ &\quad + \{(1 - \mu_t) \phi_t - 1\} \log(1 - y_t). \end{aligned}$$

Temos que para $t = 1, \dots, n$, a variável aleatória $\mathbb{I}_{\{c\}}(y_t)$ é do tipo Bernoulli, em que $Pr(\mathbb{I}_{\{c\}}(y_t) = 1) = \alpha_t$ é associada às covariáveis por meio de um preditor linear ζ_t e de uma função de ligação h . Desse modo, $\ell_1(\gamma)$ é a função de log-verossimilhança de um modelo linear generalizado com resposta binária e ℓ_2 representa a função de log-verossimilhança de um modelo de regressão beta com dispersão variável, onde a variável dependente é restrita ao intervalo $(0,1)$. Dessa maneira, uma vez que a função de verossimilhança pode ser fatorada em dois termos, os parâmetros são separáveis e a inferência por máxima verossimilhança de $(\beta^\top, \lambda^\top)^\top$ pode ser feita de forma independente do vetor de parâmetros γ . Maiores detalhes sobre aspectos inferenciais e expressões matriciais do vetor escore e matriz de informação de Fisher podem ser encontrados em Pereira e Cribari-Neto (2014) e Ospina e Ferrari (2012).

Devido a ortogonalidade do vetor de parâmetros γ e do vetor $(\beta^\top, \lambda^\top)^\top$, o estimador de máxima verossimilhança de γ é assintoticamente independente dos estimadores de máxima verossimilhança de β e λ . De forma similar ao modelo de regressão beta, os estimadores não possuem forma fechada e devem ser obtidos numericamente através da maximização da função de log-verossimilhança com o auxílio de algum algoritmo de otimização, a exemplo dos algoritmos de Newton ou quasi-Newton.

A matriz de Informação de Fisher é obtida através dos valores esperados das derivadas de segunda ordem da função de log-verossimilhança. Segue abaixo a expressão da matriz de informação de Fisher para o modelo beta inflacionado com dispersão variável

$$K(\theta) = \begin{pmatrix} K_\gamma(\gamma) & 0 \\ 0 & K_\vartheta(\vartheta) \end{pmatrix}. \quad (3.5)$$

em que $K_\gamma(\gamma) = K_{\gamma\gamma}$ é a matriz de Informação de Fisher de γ dada por $K_{\gamma\gamma} = Z^\top Q Z$, em que $Q = GPG = \text{diag}\{q_1, \dots, q_n\}$ é uma matriz diagonal e $q_t = p_t(d\alpha_t/d\zeta)^2$. Ademais, tem-se que

$$K_\vartheta(\vartheta) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}. \quad (3.6)$$

é a matriz de informação de Fisher de $\vartheta = (\beta^\top, \phi)$, com $K_{\beta\beta} = \phi^2 X^\top \Delta T W T X$, $K_{\beta\phi} = K_{\phi\beta}^\top = X^\top \Delta T c$ e $K_{\phi\phi} = \text{tr}(\Delta D)$. Em que, $\Delta = \text{diag} \delta_1, \dots, \delta_n$ e $W = \text{diag} w_1, \dots, w_n$. Para $t = 1, \dots, n$, $\delta_t = 1 - \alpha_t$ e $w_t = \psi'(\mu_t \phi) + \psi'((1 - \mu_t) \phi)$. Para mais detalhes ver Pereira e Cribari-Neto (2014).

3.2 ADEQUABILIDADE DO MODELO

Em algumas situações o parâmetro de dispersão do modelo pode variar ao longo das observações. Neste sentido, utilizou-se o teste da razão de verossimilhança (JÚNIOR; SOUZA, 2015; SOUZA et al., 2016) para verificar a hipótese nula de dispersão constante no modelo de regressão beta inflacionado. A hipótese nula a ser testada consiste em $\mathcal{H}_0 : \phi_1, \dots, \phi_n = \phi$, que é equivalente a testar $H_0 = \lambda_1 = \lambda_2, \dots, \lambda_q = 0$. A estatística do teste da razão de verossimilhança pode ser escrita como

$$RV = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\},$$

em que $\theta = (\gamma^\top, \beta^\top, \lambda^\top)^\top$, $\ell(\hat{\theta})$ é o valor maximizado do logaritmo da função de verossimilhança do modelo levando em conta a estrutura de regressão para ϕ e $\ell(\tilde{\theta})$ é o valor restrito maximizado do logaritmo da função de verossimilhança obtido pela imposição da hipótese nula. Uma vez que, sob condições usuais de regularidade e sob a hipótese nula, a estatística RV converge em distribuição para X_{q-1}^2 (uma qui-quadrado com $q-1$ graus de liberdade), o teste pode ser utilizado considerando valores críticos aproximados obtidos dessa distribuição.

3.3 ANÁLISE DE DIAGNÓSTICO

Após a estimação de um modelo de regressão é necessário verificar possíveis afastamentos das suposições iniciais feitas para o modelo, uma vez que trata-se de uma representação da realidade. Desse modo, torna-se importante a detecção de observações extremas no conjunto de dados e avaliação do impacto dessas observações nos resultados inferenciais. Além disso, também é importante avaliar a adequabilidade da distribuição de probabilidade imposta aos dados. Neste sentido, a análise de diagnóstico que será composta pela análise de resíduos, representa uma etapa indispensável no processo de estimação, tendo como pontos principais a avaliação da estabilidade e robustez do modelo, bem como a avaliação de adequabilidade da distribuição aos dados.

A análise de resíduos pode ser baseada nos resíduos ordinários, em versões padronizadas, em resíduos construídos a partir dos componentes da função desvio (MCCULLAGH; NELDER, 1989) ou em resíduos generalizados (COX; SNELL, 1989). Alguns trabalhos na literatura têm abordado diferentes aspectos dos resíduos, tais como definição, comportamento distribucional e correção (OSPINA; FERRARI, 2012). Neste contexto, destacam-se os trabalhos de Pregibon (1981), McCullagh (1989), Copas (1988), Davison e Gigli (1989), Souza e Paula (2002), Ferrari e Cribari-Neto (2004), Espinheira, Ferrari e Cribari-Neto (2008a), entre outros. Técnicas gráficas

utilizando resíduos são frequentemente adotadas para a análise de diagnóstico (SILVA, 2015). O uso de envelopes simulados, por exemplo, conforme proposto por Atkinson (1985) inicialmente para o modelo de regressão linear, permite avaliar a adequação do modelo ajustado.

3.3.1 Resíduos Quantis Aleatorizados

O resíduo quantil aleatorizado proposto por Dunn e Smyth (1996), é amplamente utilizado em análise de dados e será o resíduo considerado no presente trabalho. Ospina e Ferrari (2012) apresentam uma definição do resíduo quantil para o modelo de regressão beta inflacionado em zero ou um, definido por

$$r_t^q = \Phi^{-1}(u_t), \quad t = 1, \dots, n,$$

em que u_t é uma variável aleatória uniforme no intervalo $[a_t, b_t]$, sendo $a_t = \lim_{y \uparrow y_t} BI_c(y; \hat{\alpha}, \hat{\mu}, \hat{\phi})$, $b_t = BI_c(y_t; \hat{\alpha}, \hat{\mu}, \hat{\phi})$, $BI_c(y_t; \alpha, \mu, \phi)$ a função de distribuição beta inflacionada e Φ^{-1} a função quantil normal padrão. Sob estimativas consistentes dos parâmetros do modelo espera-se que o resíduo quantil siga distribuição normal (DUNN; SMYTH, 1996).

3.3.2 Medidas de Influência

No ajuste de modelos de regressão é importante verificar a presença de observações que podem causar efeito desproporcional no ajuste, podendo interferir, inclusive, em resultados inferenciais. A distância de Cook é uma medida que visa medir o impacto de uma observação particular nas estimativas dos parâmetros da regressão, a partir de sua exclusão do conjunto de dados. Por essa razão pode ser considerada uma medida de influência global. Apesar de ter sido desenvolvida originalmente para modelos normais lineares, aproximações para a distância de Cook são utilizadas em diversas classes de modelos, conforme apresentado em Pregibon (1981), Cook e Weisberg (1982), Cordeiro e Paula (1992).

3.3.2.1 Distância de Cook

Uma medida de influência de cada observação nas estimativas dos parâmetros é o "*likelihood displacement*" ou deslocamento pela verossimilhança (COOK; WEISBERG, 1982). No contexto de exclusão de observações uma medida de influência bastante geral motivada pela estatística da razão de verossimilhança, pode ser expressa da seguinte forma

$$LD_t = \frac{2}{d} \{ \ell(\hat{\theta}) - \ell(\hat{\theta}_{(t)}) \},$$

em que d é a dimensão de θ e $\hat{\theta}_{(t)}$ é o estimador de máxima verossimilhança de θ obtido com a remoção da t -ésima observação. Entretanto, esta definição não considera que θ apresenta dois

tipos de parâmetros distintos: o parâmetro da componente discreta e o parâmetro da componente contínua. Desse modo, é mais adequado considerar a influência da t -ésima observação sobre as estimativas de γ e (β, λ) separadamente. Neste sentido, Ospina e Ferrari (2012) propõe as seguintes estatísticas:

$$LD_t^D = \frac{2}{M} \{ \ell_1(\hat{\gamma}) - \ell_1(\hat{\gamma}_{(t)}) \}, \quad t = 1, \dots, n$$

$$LD_t^C = \frac{2}{q+m} \{ \ell_2(\hat{\beta}, \hat{\lambda}) - \ell_2(\hat{\beta}_t, \hat{\lambda}_t) \}, \quad t = y_t \in (0, 1).$$

Uma aproximação para as medidas LD_t^D e LD_t^C são dadas por

$$LD_t^D \approx C_{tt}^D = \frac{\hat{h}_{1tt}^*}{M(1 - \hat{h}_{1tt}^*)} (r_{pt}^D)^2, \quad t = 1, \dots, n.$$

$$LD_t^C \approx C_{tt}^C = \frac{\hat{h}_{2tt}^*}{(q+m)(1 - \hat{h}_{2tt}^*)} (r_{pt}^C)^2, \quad t = y_t \in (0, 1).$$

em que, $r_{pt}^{(D)} = (\mathbb{I}_{\{c\}}(y_t) - \hat{\alpha}_t) / \sqrt{\hat{\alpha}_t(1 - \hat{\alpha}_t)(1 - \hat{h}_{1tt}^*)}$ e $r_{pt}^{(C)} = (y_t^* - \hat{\mu}_t^*) / \sqrt{\hat{w}_t(1 - \hat{\alpha}_t)(1 - \hat{h}_{2tt}^*)}$ para $t = 1, \dots, n$ e $t : y_t \in (0, 1)$, respectivamente, representam os resíduos padronizados do componente discreto e contínuo do modelo de regressão beta inflacionado. Aqui, $t = 1, \dots, n$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)$ e $(y_1^*, \dots, y_n^*)$. Com

$$y_t^* = \begin{cases} \log(\frac{y_t}{1-y_t}), & y_t \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases} \quad (3.7)$$

e

$$\mu_t^* = \mathbf{E}(y_t^* | \mathbb{I}_c(y_t) = 0) = \psi(\mu_t \phi) - \psi((1 - \mu_t) \phi).$$

As quantidades $\hat{h}_{1tt}^*$ e $\hat{h}_{2tt}^*$ são os t -ésimos elementos da diagonal principal das matrizes de projeção

$$\hat{H}_1^* = \hat{Q}^{1/2} Z (Z^\top \hat{Q} Z)^{-1} Z^\top \hat{Q}^{1/2},$$

e

$$\hat{H}_2^* = (\Delta W)^{1/2} \Phi T X (X^\top \Delta T \Phi W \Phi T X)^{-1} X^\top T \Phi (\Delta W)^{1/2},$$

respectivamente, em que $Q = GPG = \text{diag}(q_1, \dots, q_n)$, Z é uma matriz $n \times M$ cuja t -ésima linha é $Z_t^\top = (z_{t1}, \dots, z_{tM})$, X é uma matriz $n \times m$ cuja t -ésima linha é $x_t^\top = (x_{t1}, \dots, x_{tm})$, $T = \text{diag}\{d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n\}$, $W = \text{diag}\{w_1, \dots, w_n\}$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$ e $\Delta = \text{diag}\{\delta_1, \dots, \delta_n\}$.

As aproximações consideradas acima são baseadas em esquema iterativo de avaliação dos estimadores de máxima verossimilhança de γ , β e λ e expansão em série de Taylor de $\ell_1(\hat{\gamma})$ e $\ell_2(\hat{\beta}_{(t)}, \hat{\lambda}_{(t)})$ para $\hat{\gamma}$ e $(\hat{\beta}, \hat{\lambda})$, respectivamente. Gráficos C_{it}^D e C_{it}^C versus ordem das observações podem ser úteis para destacar observações influentes (OSPINA; FERRARI, 2012).

4 METODOLOGIA DE PESQUISA

4.1 TIPOLOGIA DA PESQUISA

O estudo realizado é descritivo apresentando uma abordagem quantitativa e inferencial. Na presente dissertação foi abordado a classe de modelos de regressão beta inflacionados, com enfoque na aplicação a dados reais. Esse modelo caracteriza-se por permitir a modelagem de variáveis do tipo taxas ou proporções, dentro de um intervalo limitado de valores, que podem assumir, inclusive, o valor 0 (zero) ou 1 (um).

4.2 POPULAÇÃO E AMOSTRA

A amostra de dados secundários foi composta de 3.409 municípios brasileiros. Para a região nordeste e o estado da Paraíba foram considerados, respectivamente, 1.255 municípios e 154 municípios.

4.3 OBTENÇÃO DOS DADOS

A base de dados utilizada faz parte do trabalho realizado por Sousa (2011), que buscou computar índices de eficiência não-paramétricos para a AB nos municípios brasileiros. Para reduzir a influência de *outliers* e possíveis erros de medida no conjunto de observações, foram utilizados métodos bootstrap e o de reamostragem jackknife. Os escores foram obtidos através da técnica DEA (*Data Envelopment Analysis*) (CHARNES; COOPER; RHODES, 1978). A metodologia DEA fundamenta-se nas definições de eficiência técnica proposta por Farrell (1957). Segundo o autor dado um conjunto de insumos qual seria a quantidade de produtos a serem produzidos ou, alternativamente, dado um conjunto de produtos qual a menor quantidade de insumos para produzi-los. Existem outras metodologias utilizadas na avaliação da eficiência em saúde, a exemplo de fronteira estocástica. Entretanto, através de uma revisão de literatura pode-se verificar a ampla utilidade da técnica DEA na mensuração de eficiência em saúde. Em relação à atenção básica em saúde, Amado e Dyson (2008) revisaram os estudos sobre eficiência nesse nível de atenção, em particular os trabalhos que usaram a metodologia DEA (ZAVRAS et al., 2002; BRYCE; ENGBERG; WHOLEY, 2000; ROSENMAN; FRIESNER, 2004).

Neste estudo, os escores de eficiência obtidos representam a eficiência técnica de cada município no que tange ao gerenciamento dos recursos destinados à AB. A denominação de município "plenamente eficiente", refere-se aos municípios que apresentaram valor de eficiência igual a 1 (um) na amostra observada. A base de dados disponibilizada também contém informações de variáveis socioeconômicas e relacionadas à saúde, que foram coletadas em fontes

de domínio público como o IBGE (Instituto Brasileiro de Geografia e Estatística), DATASUS (Departamento de Informática do Sistema Único de Saúde), FIRJAN (Federação das Indústrias do estado do Rio de Janeiro), SIAB (Sistema de informação de atenção básica) e CNES (Cadastro Nacional de Estabelecimentos de saúde). Considerou-se um corte seccional para o ano de 2007.

O procedimento computacional foi desenvolvido utilizando o pacote GAMLSS (Modelos Aditivos Generalizados para Localização, Escala e Forma) (STASINOPOULOS; RIGBY, 2007) do *software* estatístico R (R Core Team, 2013). A análise espacial e a construção dos mapas foram realizados com o auxílio dos pacotes `spdep` e `maptools`, respectivamente.

5 ANÁLISE DE DADOS

O presente capítulo tem por objetivo apresentar uma análise descritiva das variáveis consideradas no estudo para os dados do Brasil, região nordeste e para o estado da Paraíba, através de medidas descritivas como média, mediana, mínimo e máximo das variáveis. A Tabela 1 apresenta a descrição das variáveis consideradas.

Tabela 1 – Descrição das variáveis utilizadas no ajuste dos modelos de regressão beta inflacionados.

Variável	Definição
EFIC	Escore de eficiência para a atenção básica - ano 2007.
NO	Dummy: igual a 1 se o município pertence a região norte, 0 caso contrário.
NE	Dummy: igual a 1 se o município pertence a região nordeste, 0 caso contrário.
SE	Dummy: igual a 1 se o município pertence a região sudeste, 0 caso contrário.
CO	Dummy: igual a 1 se o município pertence a região centro-oeste, 0 caso contrário.
PCI	Dummy: igual a 1 se o município participa de consórcio intermunicipal, 0 caso contrário.
PODER	Dummy: igual a 1 se o conselho municipal tem poder de decisão, 0 caso contrário.
REGIAO	Dummy: igual a 1 se o município pertence a uma região metropolitana, 0 caso contrário.
HOSP	Dummy: igual a 1 se houver hospital no município, 0 caso contrário.
CAP	Dummy: igual a 1 se o município for a capital do estado, 0 caso contrário.
POP	População do município.
IFDMR	Índice Firjan de desenvolvimento municipal - componente renda.
IFDME	Índice Firjan de desenvolvimento municipal - componente educação.
IFDMS	Índice Firjan de desenvolvimento municipal - componente saúde.
MED-ENF	Relação entre o número de médicos e enfermeiros da atenção básica municipal.
MEDIA	Média de médicos por estabelecimento da atenção básica.
CONSUL	Número de consultas por estabelecimento.
SANEA	Média dos percentuais de famílias atendidas pelas equipes de saúde da família que possuem água em rede, esgoto e coleta de lixo.
IDOS	Proporção da população municipal acima de 60 anos.
INFANT	Proporção da população municipal abaixo de 5 anos.
GASTO	Gasto per capita municipal em atenção básica.

5.1 ANÁLISE DESCRITIVA PARA OS DADOS DO BRASIL

A amostra foi composta por 3.409 municípios, distribuídos nas 5 regiões geográficas. Com base na análise descritiva foi possível observar particularidades nos estados e regiões com relação às variáveis. Por exemplo, o escore de eficiência médio foi 0,72, com valor mínimo de 0,159 no município de Rio Preto-MG. Do total da amostra 15,26% (520) dos municípios configuram-se como plenamente eficientes no gerenciamento de recursos destinados à AB. Na região nordeste, destacam-se os estados de Pernambuco com 28,82% dos municípios, Ceará e Bahia com 23,87% e 13,93%, respectivamente. Nos estados de Minas Gerais e São Paulo os percentuais de municípios plenamente eficientes foram, respectivamente, 10,35% e 28,77%.

O maior índice Firjan de desenvolvimento municipal componente na renda foi observado no município de Itabira em Minas Gerais com valor igual a 0,997, enquanto o município de São

Félix do Tocantins-TO apresentou menor índice na renda (0,041). A média do gasto per capita nos municípios foi de R\$ 141,80 reais e mediana de R\$ 121 reais, ou seja, 50% dos municípios observados gastam abaixo de R\$ 121,00 reais por pessoa na AB. O maior gasto per capita em AB foi realizado no município de Madre de Deus-BA com valor de R\$ 1.080,00 reais. Esse município apresentou alto escore de eficiência 0,987, no entanto não foi plenamente eficiente na aplicação dos recursos destinados á AB. De outro modo, vários municípios apresentaram valor igual a zero quando se refere ao gasto per capita em AB. Uma análise por região revelou que o centro-oeste apresentou o maior número de municípios com essa característica (20,58%), seguida da região norte (11,31%), nordeste (10,24%), sudeste (7,96%) e região sul (5,69%). Com relação aos estados, verifica-se que Minas Gerais e Goiás foram os que apresentaram o maior número de municípios sem gasto per capita em AB, sendo de 8,68% e 42,30%, respectivamente. Ademais, ainda destacam-se os estados do Rio Grande do Norte com 21,66%, Bahia com 6,8% e Paraíba com 14,93%, sem investimento em AB.

Na saúde os maiores índices Firjan de desenvolvimento municipal foram observados nos municípios de Flórida-PR, Alto Alegre-RS, Rondinha-RS e Saldanha Marinho-RS, com valores iguais a 1 (um), enquanto a cidade de São João do Piauí-PI apresentou o menor índice (0,361). A variável relação entre o número de médicos e enfermeiros na AB é muito importante, uma vez que a eficácia dos serviços está relacionada com as categorias profissionais dispostas nos estabelecimentos de saúde. O valor médio para essa variável foi de 1,21, indicando maior número de médicos em relação aos profissionais de enfermagem. Essa relação foi superior a 1 (um) principalmente nas regiões sudeste (76,73%), nordeste (47,89%) e sul (35%). Quanto aos estados, a maior ocorrência desse tipo de situação foi verificado na Bahia (43,65%), Minas Gerais (74,62%), São Paulo (83,01%), Paraná (60,96%), Santa Catarina (67,25%) e Rio Grande do Sul (67,50%). Quanto ao número de consultas realizadas por estabelecimento, o valor médio para essa variável foi de 5.547, 1, sendo o maior número de consultas observado no município de Piên no Paraná (41.750) e o menor número em Ananás no Tocantins com 44, 2.

A média de médicos por estabelecimento da atenção básica foi de 1,8, com destaque para a capital Vitória no Espírito Santo, apresentando maior média de médicos por estabelecimento da AB (12). Os municípios com média de médicos acima de 5, apresentam-se nas regiões sul (3,47%) e sudeste (7,33%). Com relação a existência de hospitais nos municípios, nota-se que a região norte se destaca com número reduzido de hospitais. No estado da Amazônia foi observado percentual de 19,5% e em Roraima apenas 33,3% dos municípios observados. No entanto, vale destacar que para os estados do Amapá e Roraima as amostras de municípios foram, relativamente, pequenas. Por outro lado, verifica-se que nos demais estados o número de municípios com hospitais é ligeiramente superior, com destaque para os estados da Bahia (63,46%), Minas Gerais (53,75%) e Paraná (73,97%). Quanto ao aspecto educação, o maior índice Firjan de desenvolvimento municipal na educação foi observado no município de Santa Rita d'Oeste no estado de São Paulo, com valor de 0,956, ao passo que na cidade de Wenceslau

Guimarães-BA, foi registrado o menor índice (0,297).

Os estados com maior número de municípios que participam de consórcio intermunicipal foram Minas Gerais (88,14%), Paraná (76,95%) e Rio Grande do Sul (52,5%). Quanto ao poder de decisão dos conselhos municipais, destacam-se os estados de Minas Gerais, Bahia, Rio Grande do Sul, Paraná, Santa Catarina e São Paulo, com percentuais de 33,88%, 43,65%, 45%, 52,41% e 53,30%, respectivamente. Apenas 8,95% dos municípios pertencem a região metropolitana.

Sabe-se que os aspectos ambientais, sanitários e de higienização são importantes na prevenção de danos para a saúde. Sobretudo, serviços como coleta de lixo, tratamento de água e manutenção de rede de esgoto, podem fazer um diferencial na qualidade de vida do indivíduo. Uma das variáveis analisadas diz respeito ao percentual médio de famílias que são assistidas por esses serviços nos municípios. Conforme a análise verificou-se que o menor percentual médio foi apresentado pelo município de Boqueirão do Leão (0,003), localizado no Rio Grande do Sul. Esse município foi considerado plenamente eficiente, um fato curioso que nos induz a investigar como a eficiência se relaciona com esses fatores.

Em contra partida, alguns municípios se destacam, positivamente, com relação a esse aspecto. Em Minas Gerais 33,38% dos municípios apresentaram um percentual médio acima de 80% de famílias com esses serviços. Já em São Paulo, esse percentual foi de 61,32%, no Rio de Janeiro sendo de 34,42% e no Rio Grande do Sul, com 8,93%. Apenas 0,12% (4) dos municípios observados apresentaram uma cobertura completa desses serviços, ou seja, a média percentual de famílias que são cobertas por esses serviços foi de 100%, sendo eles: Espírito Santo Pinhal, Ituverava, Junqueirópolis e São José dos Campos, todos localizados no estado de São Paulo. Os municípios de Espírito Santo do Pinhal e São José dos Campos foram considerados plenamente eficientes, isto é, escore de eficiência igual a 1 (um), enquanto os municípios de Ituverava e Junqueirópolis apresentaram eficiência de 0,63 e 0,51, respectivamente.

Com relação a estrutura etária da população, verificamos que proporção de idosos nos municípios é pequena, com valor mínimo de 0,011 no município de Maragogipe-BA e máximo em São José do Rio Preto-SP (0,230). Ao direcionar a análise por região e considerando uma proporção de idosos acima do valor mediano, isto é, considerando a mediana da variável IDOS que foi igual a 0,11, nota-se que as regiões sul e sudeste concentram o maior número de municípios com proporção de idosos acima de 11%, sendo os percentuais iguais a 71,8% e 62,49%, respectivamente. Por outro lado, a proporção da população infantil com menos de 5 anos foi mais expressiva nas regiões norte e nordeste do país. No nordeste, o percentual de municípios que apresentam uma proporção acima de 0,09 (valor mediano), ou seja, de 9%, foi de 84% e na região norte sendo de 92,22%. O centro-oeste também apresentou percentual elevado 47,05%.

No sul e sudeste, que podem ser consideradas regiões mais desenvolvidas em termos de emprego, renda, educação e saúde, esse percentual é ligeiramente inferior quando comparado

aos das regiões supracitadas, levando em consideração os tamanhos amostrais. Na região sul, esse percentual cai para 9,86% e no sudeste para 28,79%. Essa análise reforça mais uma vez o que pode ser encontrado na literatura, através de pirâmides etárias, que em regiões mais desenvolvidas há um maior número da população idosa, enquanto nas localidades com menor grau de desenvolvimento apresentam, em sua maioria, altas taxas de natalidade.

Tabela 2 – Descrição das variáveis contínuas e frequência das variáveis dummy para os dados do Brasil.

Variável	Mínimo	Média	$Q_{\{1/4\}}$	Mediana	$Q_{\{3/4\}}$	Máximo	DP	Frequência (%)
EFIC	0,1593	0,7223	0,577	0,708	0,891	1,000	0,193	-
POP	889	35.138,1	6.313,8	12.972,5	26.562,3	6.178.762	150.815,2	-
SANEA	0,003	0,507	0,334	0,477	0,662	1,000	0,232	-
IFDMS	0,361	0,731	0,638	0,738	0,832	1,000	0,123	-
IFDMR	0,041	0,406	0,301	0,379	0,474	0,997	0,152	-
IFDME	0,297	0,637	0,554	0,644	0,721	0,956	0,118	-
ME-ENF	0,075	1,207	0,982	1,075	1,350	5,200	0,493	-
INFANT	0,042	0,092	0,077	0,090	0,107	0,183	0,021	-
IDOS	0,011	0,109	0,088	0,108	0,129	0,230	0,031	-
GASTO	0,000	1,418	56,95	1,210	199,25	1.080,0	1,174	-
MEDIA	0,125	1,800	1,000	1,400	2,167	12,000	1,329	-
CONSULT	44,20	5.547,1	2.700,6	4.406,3	6.775,9	41.750,0	4.371,8	-
PODER	-	-	-	-	-	-	-	1492 (43,77)
PCI	-	-	-	-	-	-	-	1340 (39,31)
REGIAO	-	-	-	-	-	-	-	21 (0,61)
HOSP	-	-	-	-	-	-	-	1924 (56,45)
CAP	-	-	-	-	-	-	-	21

É possível observar a partir do histograma e *box-plot* da variável Eficiência, apresentados na 1, que a distribuição da variável é assimétrica à direita, com uma massa de pontos no valor igual a um. Além disso, a partir do *box-plot*, nota-se que a mediana está mais próximo do primeiro quartil e que não foi observado observações discrepantes, ou seja, pontos que excedem os limites do *box-plot*.

Para verificar a relação existente entre a variável resposta e as variáveis explicativas, calculamos as correlações entre as variáveis contínuas (Tabela 3). Verifica-se que nenhuma das covariáveis apresentam correlação superior a 0,80 entre si. Entretanto, as variáveis INFANT e IDOS mostram-se altamente correlacionadas negativamente ($-0,7206$), bem como os índices IFDME e IFDMS com valor igual a 0,6814.

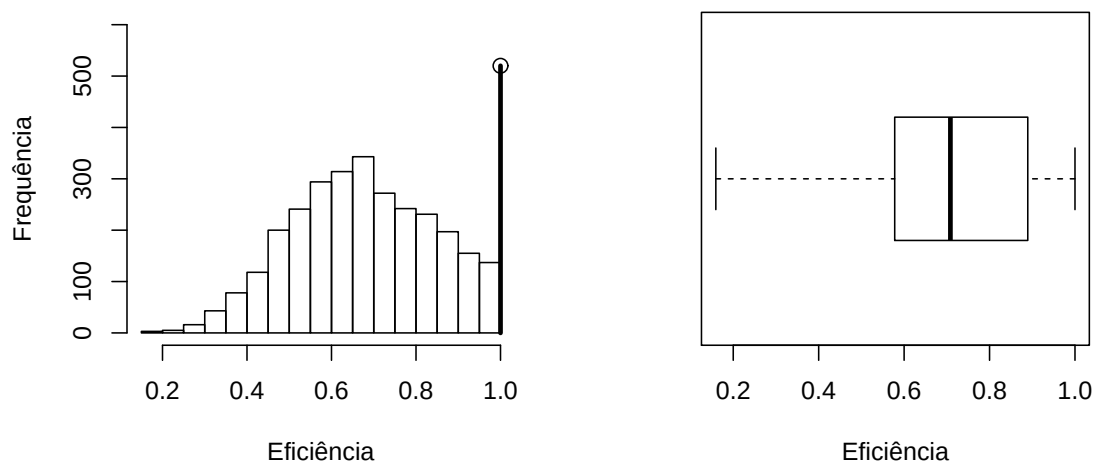
5.2 ANÁLISE DESCRITIVA PARA OS DADOS DA REGIÃO NORDESTE

Nesta seção, apresentaremos a análise descritiva dos dados para a região nordeste, cuja a amostra de municípios foi de 1.255. Na Tabela 4 são apresentadas algumas estatísticas descritivas

Tabela 3 – Matriz de correlação para as variáveis contínuas

	EFIC	POP	IFDMR	IFDME	IFDMS	GASTO	MED-ENF	MEDIA	CONSUL	SANEA	IDOS	INFANT
EFIC	1	0.1287	0.0670	-0.0225	-0.0968	-0.3117	-0.0694	-0.0382	0.0153	-0.0227	-0.1210	0.1566
POP	-	1	0.3435	0.0683	0.0663	-0.0789	0.0444	0.0105	0.0307	0.0402	-0.0104	-0.0125
IFDMR	-	-	1	0.4515	0.4563	0.0444	0.1719	0.2037	0.1902	0.2474	0.1301	-0.2848
IFDME	-	-	-	1	0.6814	0.1835	0.2548	0.3103	0.2762	0.4070	0.2755	-0.4843
IFDMS	-	-	-	-	1	0.2429	0.2704	0.2521	0.2244	0.3092	0.3136	-0.5085
GASTO	-	-	-	-	-	1	0.1093	0.1338	0.0666	0.0969	0.1389	-0.2036
MED-ENF	-	-	-	-	-	-	1	0.1374	0.0881	0.1835	0.1356	-0.1966
MEDIA	-	-	-	-	-	-	-	1	0.5946	0.3420	0.1054	-0.2918
CONSUL	-	-	-	-	-	-	-	-	1	0.3068	0.0121	-0.2073
SANEA	-	-	-	-	-	-	-	-	-	1	0.1701	-0.4460
IDOS	-	-	-	-	-	-	-	-	-	-	1	-0.7206
INFANT	-	-	-	-	-	-	-	-	-	-	-	1

Figura 1 – Histograma de frequências e *box-plot* da variável Eficiência.



das variáveis contínuas e frequência das variáveis dummy. Com base nos resultados verificamos que a menor eficiência foi apresentada pelo município de Arataca-BA, com escore de 0,263.

Na região nordeste a população média foi de 31.527,59 indivíduos, com população máxima registrada em Salvador no estado da Bahia (Tabela 4). O município com menor porte populacional foi Quixaba-PB, com 1.004 habitantes. O máximo e mínimo para a variável proporção de pessoas com mais de 60 anos de idade foram observados nas cidades de Cícero Dantas (0,011) e Maragogipe (0,175), respectivamente, ambas localizadas no estado da Bahia. Por outro lado, na cidade de Andorinha-BA foi encontrado a menor proporção de crianças, com 0,064 e máximo no município de Bom Jesus das Selvas-MA (0,174).

O gasto médio per capita em AB na região nordeste foi de R\$ 114,00 reais, apresentando valor máximo na cidade Madre de Deus-BA (R\$ 1.080). O maior índice de desenvolvimento na renda foi observado na cidade de Natal-RN (0,908) e o menor em Cacimba da Areia localizado no estado da Paraíba (0,046).

No tocante aos índices Firjan de desenvolvimento na saúde e educação, os maiores foram observados nos municípios de Santarém-PB (0,905) e Riacho de Santana-RN (0,79), respectivamente. Os menores índices na saúde e educação foram, respectivamente, para os municípios de Santo Amaro do Maranhão-MA (0,398) e Wenceslau Guimarães-BA (0,297).

No que diz respeito a variável média de médicos por estabelecimento da AB, seu valor máximo foi observado no município de Guaramiranga-CE (7,7), enquanto a menor média foi para as cidades de Colinas no estado do Maranhão e Guarabira na Paraíba (0,2). Os valores máximo e mínimo para a variável número de consultas por estabelecimento foram observados nas cidades de Parnamirim no Rio Grande do Norte e no município de Arame no Maranhão.

Além disso, apenas 48,02% dos municípios da amostra possuem hospital. Observa-se que a variável MED-ENF apresentou mediana igual a 1, uma evidência de que em 50% dos municípios observados há um número maior de médicos em relação ao de enfermeiros.

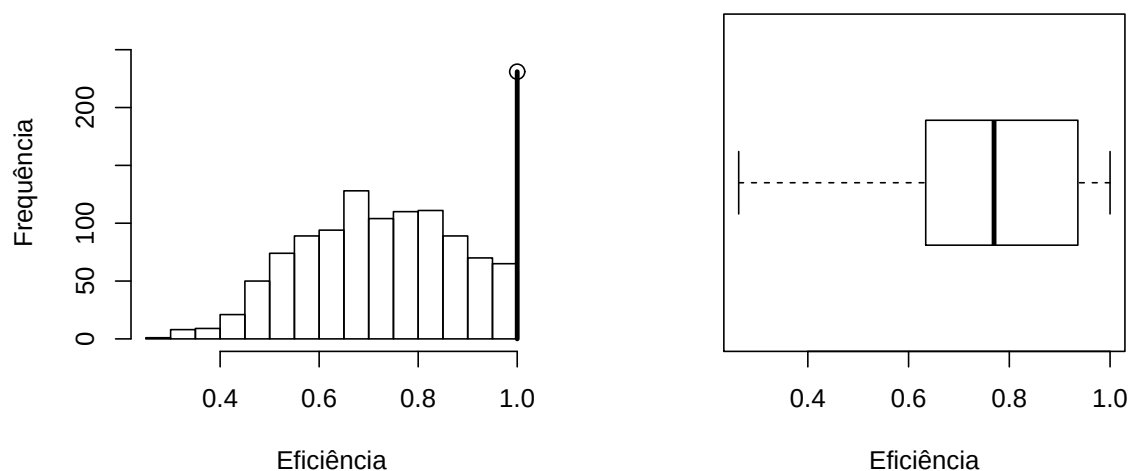
Com relação aos conselhos municipais, 40,7% dos municípios apresentam poder de decisão, 12,78% participam de consórcio intermunicipal e 5,95% pertencem a região metropolitana. Para a variável SANEA, que representa o percentual médio de famílias com acesso a serviços de saneamento básico, seu valor médio foi de 0,406, com valores máximo e mínimo de 0,988 e 0,004, respectivamente. O valor mínimo foi registrado pelo município Acopiara-CE (0,004) e o maior na cidade de Jaguarari-BA (0,988).

Tabela 4 – Descrição das variáveis contínuas e frequência das variáveis dummy para a região nordeste.

Variável	Mínimo	Média	$Q_{\{1/4\}}$	Mediana	$Q_{\{3/4\}}$	Máximo	DP	Frequência (%)
EFIC	0,263	0,768	0,634	0,770	0,935	1,000	0,177	-
POP	1.004	31.527,6	8.264,5	14.981	26.485,5	2.754.956	118.856,6	-
SANEA	0,004	0,406	0,282	0,384	0,521	0,988	0,178	-
IFDMS	0,398	0,647	0,589	0,645	0,704	0,905	0,086	-
IFDMR	0,046	0,326	0,258	0,303	0,369	0,908	0,117	-
IFDME	0,297	0,548	0,483	0,552	0,614	0,790	0,088	-
ME-ENF	0,075	1,074	0,944	1,000	1,173	4,033	0,291	-
INFANT	0,064	0,106	0,095	0,105	0,116	0,174	0,016	-
IDOS	0,011	0,102	0,084	0,101	0,118	0,175	0,023	-
GASTO	0,000	114,42	53,80	100,00	164,50	1.080,00	88,42	-
MEDIA	0,200	1,274	0,800	1,000	1,500	7,74	0,770	-
CONSULT	69,33	3.879,8	2.164,5	3.339	4.895,6	33.833	2.757,5	-
PODER	-	-	-	-	-	-	-	513 (40,7)
PCI	-	-	-	-	-	-	-	161 (12,8)
REGIAO	-	-	-	-	-	-	-	75 (5,9)
HOSP	-	-	-	-	-	-	-	605 (48,0)
CAP	-	-	-	-	-	-	-	7

Além disso, é possível observar que os desvios-padrão (DP) das variáveis observadas foram, em sua maioria, relativamente baixos, indicando menor dispersão em torno da média, exceto para as variáveis POP, GASTO e CONSULT. A Figura 2 apresenta o histograma de frequência e o *box-plot* da variável Eficiência. Neste gráfico, histograma de frequência, a barra vertical com o ponto acima representa a quantidade de uns na amostra (18,4%), ou seja, número de municípios que foram considerados plenamente eficientes. Ademais, com base no gráfico *box-plot*, nota-se que a mediana encontra-se mais próxima do primeiro quartil e que não foram observados pontos discrepantes.

Figura 2 – Histograma de frequência da variável Eficiência para os dados da região nordeste.



5.3 ANÁLISE DESCRITIVA PARA OS DADOS DA PARAÍBA

Nesta seção são apresentados os resultados da análise descritiva para os dados da Paraíba. A amostra foi composta por 154 municípios paraibanos. É possível verificar, Tabela (5), que o valor máximo e mínimo para a variável EFIC nos municípios foi de 1,000 e 0,351, respectivamente; e mediana igual a 0,654. Além disso, verificamos que a eficiência média dos municípios foi de 0,687, com desvio padrão (DP) igual a 0,177.

Quanto ao porte populacional dos municípios, nota-se que a população (POP) média observada foi de 20.118 habitantes. O maior e menor número de habitantes foram observados nos municípios de João Pessoa e Quixabá, com 683.278 e 1.004 habitantes, respectivamente. A proporção média de crianças e idosos nos municípios foram de 0,095 e 0,118, respectivamente. João Pessoa foi o município com menor proporção da população acima de 60 anos (5%).

A proporção média de famílias com serviços de saneamento básico (SANEA) foi de 0,418 com desvio padrão de 0,177. O menor percentual de famílias com esses serviços foi observado no município de Juazeirinho (0,023), enquanto o maior percentual no município Poço de José de Moura (0,926).

Com relação aos índices Firjan de desenvolvimento municipal, observa-se que o índice médio na saúde (IFDMS) foi de 0,681, na componente renda (IFDMR) a média foi de 0,307 e na educação sendo de 0,576. O município de Santarém apresentou o maior índice de desenvolvimento na saúde (0,905) e Curral Velho o menor (0,478). Para a capital do estado, o índice de desenvolvimento na saúde foi de 0,791. Considerando a componente educação, o município de Várzea foi o que apresentou maior valor (0,774), sobrepondo-se ao valor observado na capital do

estado (0,628). Por outro lado, São Bento foi o município com menor índice de desenvolvimento na educação (0,432).

A média de consultas por estabelecimento (CONSULT) foi de 3.489,76 consultas, com número mínimo de 327, observado no município de Riachão e máximo de 9.602 consultas, observado no município de Amparo. A variável relação entre o número de médicos em relação ao de enfermeiros apresentou valor médio acima de 1 (1,09), indicando a existência de um número relativamente maior de médicos em relação à profissionais da enfermagem nos serviços de AB. Nota-se que a mediana da variável média de médicos por estabelecimento da AB (MEDIA) foi igual a 1, indicando que 50% dos municípios observados possui mais de um profissional médico por estabelecimento da atenção básica. Salientando que o valor máximo observado para essa variável foi igual a 3. Apenas 41,56% dos municípios da amostra tem hospital.

Além disso, apenas 21,43% dos municípios observados têm poder de decisão através dos conselhos municipais e 22,73% participam de consórcio intermunicipal. Apenas 6,49% pertencem a uma região metropolitana. O gasto médio per capita em AB para a amostra de municípios da Paraíba foi de R\$143,08 reais, com valor máximo igual a R\$654,00 reais. Nota-se que os desvios-padrão (DP) das variáveis observadas foram, em sua maioria, relativamente baixos, indicando menor dispersão em torno da média.

Os municípios plenamente eficientes no gerenciamento de recursos destinados à AB foram: Aguiar, Algodão de Jandaíra, Amparo, Areia de Baraúnas, Belém do Brejo do Cruz, Guarabira, Riachão, Poço de José de Moura, Quixabá, Riacho de Santo Antônio, São Francisco, São José dos Cordeiros, Serra Grande e Zabelê. Observamos que a plena eficiência quanto ao gerenciamento dos recursos destinados à AB foi predominante nos municípios menores, com menos de cinco mil habitantes (IBGE, 2017), com exceção apenas das cidades de Aguiar, Belém do Brejo do Cruz e Guarabira, com populações de 5.562, 7.347 e 58.881 habitantes, respectivamente. Esse resultado corrobora com o estudo realizado por Queiroz et al. (2013), em cidades do Rio Grande do Norte, em que a plena eficiência no gasto público em saúde foi expressa por municípios de grande e pequeno porte populacional.

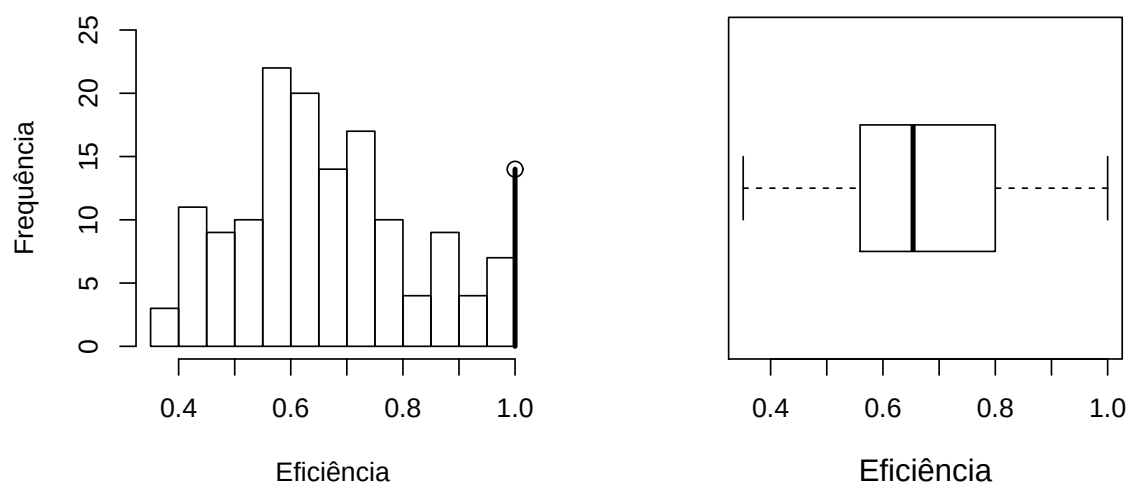
Por outro lado, Varela, Martins e Fávero (2012) que analisaram municípios paulistas, verificaram que o grupo de cidades ineficientes foi predominantemente formado por municípios de pequeno porte populacional. Desse modo, é importante salientar que não apenas o porte do município irá definir sua capacidade de ser ou não plenamente eficiente. O achado no presente estudo contribui para uma melhor compreensão da situação de assistência à saúde prestada à população, bem como o nível de gerenciamento dos recursos da AB no respectivo município. Uma vez que, os resultados podem diferir segundo a região analisada, conforme observado ao se comparar os resultados do presente trabalho com os apresentados por Varela, Martins e Fávero (2012).

Na Figura 3 apresentamos o histograma e o *box-plot* da variável EFIC. Observa-se que

Tabela 5 – Descrição das variáveis contínuas e frequência das variáveis dummy para os dados da Paraíba.

Variável	Mínimo	Média	$Q_{\{1/4\}}$	Mediana	$Q_{\{3/4\}}$	Máximo	DP	Frequência (%)
EFIC	0,351	0,687	0,560	0,654	0,798	1,000	0,177	-
POP	1.004	20.118	4.624,8	8.375,5	15.805,5	683.278	63.815,6	-
SANEA	0,023	0,418	0,298	0,388	0,531	0,926	0,177	-
IFDMS	0,479	0,681	0,622	0,679	0,730	0,905	0,084	-
IFDMR	0,046	0,307	0,259	0,292	0,333	0,863	0,105	-
IFDME	0,432	0,576	0,533	0,570	0,615	0,774	0,059	-
CONSULT	327	3.489	2.303,6	3.250	4.476,8	9.602	1.739,3	-
ME-ENF	0,550	1,090	1,000	1,000	1,142	3,000	0,259	-
INFANT	0,068	0,095	0,087	0,095	0,104	0,130	0,012	-
IDOS	0,050	0,118	0,105	0,119	0,135	0,171	0,023	-
MEDIA	0,200	1,135	0,922	1,000	1,321	3,000	0,479	-
GASTO	0,000	143,08	64,85	130,50	204,00	654,00	112,99	-
PODER	-	-	-	-	-	-	-	33 (21,43)
PCI	-	-	-	-	-	-	-	35 (22,73)
CAP	-	-	-	-	-	-	-	1
REGIAO	-	-	-	-	-	-	-	10 (6,49)
HOSP	-	-	-	-	-	-	-	64 (41,56)

a distribuição dos dados em (0,1] é assimétrica, com um registro de maior frequência entre os valores 0,5 a 0,7. A barra vertical com um ponto acima no histograma refere-se a quantidade de uns na amostra, que correspondeu a 9,09%, ou seja, 9% dos municípios analisados são classificados como plenamente eficiente. No gráfico *box-plot* nota-se que a mediana da variável encontra-se mais próxima do primeiro quartil e que não foram observados pontos discrepantes.

Figura 3 – Histograma de Frequência e *box-plot* da variável Eficiência para os dados da Paraíba.

6 AJUSTES DOS MODELOS DE REGRESSÃO BETA INFLACIONADOS

6.1 ESPECIFICAÇÃO DO MODELO PARA OS DADOS DO BRASIL

Nesta seção apresentamos os resultados do ajuste para explicar os escores de eficiência na AB para os dados do Brasil. Inicialmente, ao ajustar um modelo de regressão beta inflacionado, é indispensável verificar a suposição de dispersão constante para os dados. Segundo Smithson e Verkuilen (2006), modelos de regressão com dispersão variável precisam de uma estrutura em que seja possível modelar o parâmetro de precisão, de modo a melhorar os resultados inferenciais. Para tanto, foi utilizado o teste da razão de verossimilhança (JÚNIOR; SOUZA, 2015; SOUZA et al., 2016; SILVA; SOUZA, 2014) com o objetivo de verificar a hipótese nula de precisão fixa, ou seja, $\mathcal{H}_0 = \phi_1 \dots, \phi_n = \phi$.

Para a realização do teste consideramos na estrutura de regressão do parâmetro de precisão as variáveis apresentadas na Tabela 6. Como resultado obteve-se o p -valor = 0,0006, indicando que devemos rejeitar a hipótese nula de precisão fixa ao nível de significância de 5%. Desse modo, há indícios de que o parâmetro de precisão varia ao longo das observações, sendo necessário uma estrutura de regressão para modelar a precisão dos dados.

Na Tabela 6 são apresentadas as estimativas dos parâmetros, erros-padrão e p -valores das variáveis significativas para explicar o escore de eficiência. Para este modelo foram consideradas as funções de ligação logit, para a modelar a média e o parâmetro α , e a função de ligação log para o parâmetro de precisão. O nível de significância considerado na seleção das variáveis foi de 5%.

Conforme as estimativas dos parâmetros, Tabela 6, nota-se que a variável população do município (POP) influencia positivamente a eficiência média na AB, isto é, as cidades com maior população tendem a apresentar maiores escores de eficiência na AB. Por outro lado, verifica-se que a variável gasto per capita em atenção básica (GASTO) exerce efeito negativo na eficiência média, isto é, municípios com maiores gastos per capita tendem a apresentar menor eficiência média. Essa constatação pode ser explicada pelo tipo de cuidado que deve ser prioritariamente empregado na AB. Tais práticas devem perpassar as tecnologias leves do cuidado, com o objetivo de dar resolutividade às demandas e necessidades de saúde dos sujeitos em seus territórios (MERHY; FRANCO, 2003). Coelho e Jorge (2009) apresentam o acolhimento, o vínculo e a atenção integral como tecnologias leves no trabalho em saúde, as quais Mehry e Franco (2003) afirmam estarem centradas no campo das relações, evidenciando o diálogo e envolvimento entre profissionais e usuários.

Observa-se também que o índice Firjan de desenvolvimento municipal na saúde (IFDMS) exerce efeito positivo na eficiência média do município. Isto é, municípios com

melhores níveis de desenvolvimento na saúde tendem a apresentar maiores escores de eficiência na AB. A variável SANEA, que representa o percentual médio de famílias com acesso aos serviços essenciais de saneamento básico, também influencia de forma positiva a eficiência média do município. A população que convive com serviços de saneamento básico adequados, apresentam menores riscos de serem acometidas por algum tipo de doença (GONZALEZ; TOZONI-REIS; DINIZ, 2007).

Segundo Gonzalez, Tozoni-Reis e Diniz (2007), os transtornos oriundos do acúmulo de lixo em locais inadequados além de contribuir para o aparecimento de doenças, faz com que a comunidade encontre saídas próprias para a resolução deste problema, uma delas sendo a promoção de queimadas do lixo, que traz como consequência, a produção de fumaças tóxicas e o risco de incêndios, ambos prejudiciais à saúde.

Esse achado corrobora e evidencia o conceito ampliado de saúde, formulado durante a VIII Conferência Nacional de Saúde, em que a saúde é entendida como:

resultante das condições de alimentação, habitação, educação, renda, meio ambiente, trabalho, transporte, emprego, lazer, liberdade, acesso e posse da terra e acesso aos serviços de saúde. Sendo assim, é principalmente resultado das formas de organização social, de produção, as quais podem gerar grandes desigualdades nos níveis de vida (BRASIL, 1986).

Em relação às variáveis NO, NE, CO, é possível concluir que os municípios localizados nas regiões norte, nordeste ou centro-oeste, apresentam maior eficiência média na aplicação dos recursos destinados à AB, em relação aos municípios da região sul. Conforme Sousa (2011), esse resultado pode ser explicado pelo fato de que os bons escores na região nordeste, são reflexos de relações mais altas entre produtos e insumos e não necessariamente melhores prestações de serviços.

É provável que a qualidade dos serviços de AB seja melhor nas cidades do sul. Entretanto, se nessas cidades a produção desses serviços estiver relacionada a custos proporcionalmente maiores, essas, não poderão ser consideradas como exemplos de melhores práticas, isto é, com melhores desempenhos (SOUSA, 2011). Além disso, essa conclusão coincide com o apresentado por Dias (2010), que também observou menores escores de eficiência nas regiões sul e sudeste, justificando também pelas melhores relações de insumo e produto.

Para a estrutura de regressão do parâmetro α , que modela a probabilidade do município ser plenamente eficiente, destacam-se como significativas as variáveis PCI, GASTO, POP, IDOS, CONSUL, CAP, NE e SE. Neste sentido, os resultados sugerem que os municípios com maior população tendem a apresentar maior chance de serem plenamente eficientes. Do mesmo modo, os municípios localizados nas regiões nordeste ou sudeste, também apresentam maior chance de serem plenos no gerenciamento desses recursos. Com relação a variável número de consultas por estabelecimento, seu coeficiente positivo indica que estabelecimentos maiores tendem a apresentar maior chance de serem plenamente eficiente na aplicação dos recursos direcionados à

AB. Conforme Zavras et al. (2002), através de estudos anteriores, constatamos que a concentração da oferta de serviços da AB em alguns estabelecimentos de saúde eleva a eficiência.

Em contrapartida, municípios que participam de consórcio intermunicipal, apresentam maiores gastos per capita em AB, possuem maior proporção da população acima de 60 anos e são capitais, apresentam menores chances de serem plenamente eficientes. Algumas justificativas para tais achados são, por exemplo, a faixa etária da população acima de 60 anos, tende a apresentar maior incidência de doenças crônicas, necessitando de atendimento contínuo e, por sua vez, gerando maiores despesas. Isso faz com que sejam efetuados maiores gastos em saúde, ocasionando menores escores de eficiência (SOUSA, 2011).

O fato das capitais apresentarem menor chance de serem plenamente eficientes pode ser reflexo de alguns fatores, a saber: dificuldade de aplicação dos recursos, a diversidade de acometimentos em saúde ou a demanda de um contingente populacional que, geralmente, é maior do que nos outros municípios. Corroborando com o apresentado em Sousa (2011), a participação em consórcio intermunicipal não representou melhor desempenho na gestão dos recursos destinados à AB.

Com relação ao modelo para a precisão (ϕ), apenas as variáveis MED-ENF, NE e SE, foram significativas ao nível de significância de 5% (p-valor < 0,05). Esse resultado indica que os municípios com maior número de médicos em relação ao de enfermeiros na AB, tendem a apresentar escores de eficiência menos dispersos, assim como os municípios localizados nas regiões nordeste e sudeste.

Tabela 6 – Modelo de regressão beta inflacionado com dispersão variável para os dados do Brasil.

μ			
Variável	Estimativa	Erro-padrão	p -valor
Constante	0,2645	0,1258	0,0355
POP	$2,050 \times 10^{-06}$	$3,097 \times 10^{-07}$	$< 0,001$
IFDMS	0,5470	0,1482	0,0002
GASTO	$-1,484 \times 10^{-03}$	$1,252 \times 10^{-04}$	$< 0,001$
NO	0,2538	0,0646	$< 0,001$
NE	0,3776	0,0413	$< 0,001$
CO	0,1616	0,0649	0,0128
SANEA	0,1917	0,0677	0,0047
α			
Constante	-0,7092	0,2148	$< 0,001$
GASTO	-0,0054	$5,919 \times 10^{-04}$	$< 0,001$
POP	$3,374 \times 10^{-06}$	$6,567 \times 10^{-07}$	$< 0,001$
IDOS	-7,6010	1,7710	$< 0,001$
PCI	-0,4685	0,1243	$< 0,001$
CONSUL	$4,499 \times 10^{-05}$	$1,084 \times 10^{-05}$	$< 0,001$
CAP	-1,9760	0,8841	0,0255
NE	0,3045	0,1249	0,0149
SE	0,3539	0,1440	0,0141
ϕ			
Constante	1,6647	0,0701	$< 0,001$
SE	0,1914	0,0626	0,0023
NE	0,1424	0,0590	0,0159
ME-ENF	0,1179	0,0472	0,0125

Para avaliar a correta especificação do modelo ajustado utilizamos o teste RESET (PEREIRA; CRIBARI-NETO, 2014). O teste considera a hipótese nula de correta especificação, versus hipótese alternativa de má especificação. Essa má especificação estando relacionada, por exemplo, a variável regressora omitida ou função de ligação incorretamente especificada. O teste consiste em adicionar como variável de teste o preditor linear ao quadrado (PEREIRA; CRIBARI-NETO, 2014). Caso a variável de teste seja significativa, dizemos que o modelo não encontra-se bem especificado, do contrário, o modelo proposto apresenta forma funcional correta e não possui variáveis omitidas (RAMSEY, 1969). Os autores Ramsey e Gilbert (1972) sugeriram usar como variáveis de teste a segunda, terceira e quarta potência do valor ajustado. Para maiores detalhes (THURSBY; SCHMIDT, 1977; SHUKUR; EDGERTON, 2002; MANTALOS; SHUKUR, 2007; CRIBARI-NETO; LIMA, 2007). Na presente dissertação, o teste RESET foi realizado considerando a segunda potência do preditor estimado ($\hat{\eta}^2$), uma vez que Pereira e Cribari-Neto (2014) observaram, através de estudos de simulação, que os melhores resultados em termos de tamanho e poder se dão quando utilizada a segunda potência do preditor linear.

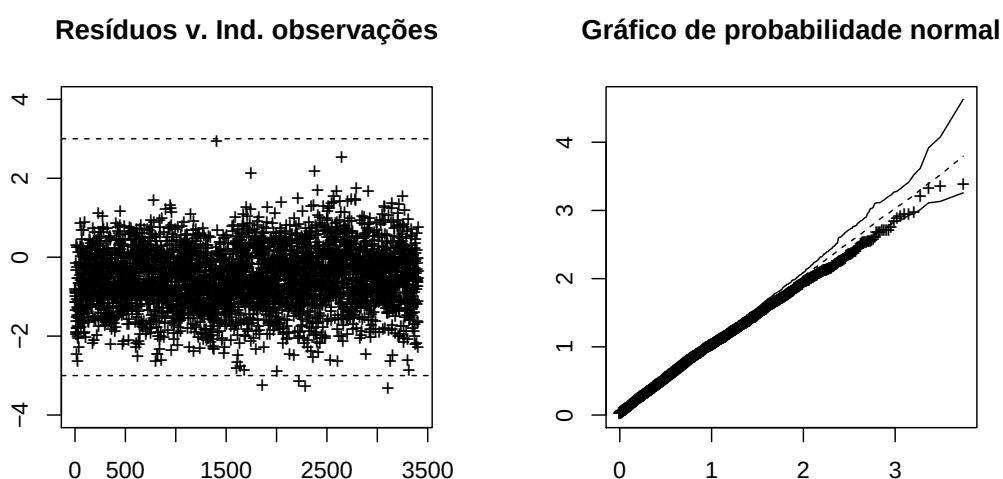
Uma medida para avaliar a adequabilidade do modelo proposto é denominada de

coeficiente de determinação (pseudo- R^2). Ela é bastante usual e pode ser definida pela razão entre o logaritmo da função de verossimilhança maximizada do modelo nulo (sem estrutura de regressão) em relação ao logaritmo da função de verossimilhança maximizada do modelo completo (modelo com a estrutura de regressão) (MCFADDEN, 1974).

Na aplicação do teste RESET, obtivemos p -valor igual a 0,7958 e, desse modo, não temos evidência suficiente para rejeitar a hipótese nula de que o modelo está bem especificado ao nível de significância de 5%. O pseudo- R^2 obtido da regressão beta inflacionada foi de 0,80, evidenciando a qualidade no ajuste do modelo. Salientando que o pseudo- R^2 do modelo foi alto e bastante expressivo.

A análise de diagnóstico representa uma etapa importante para avaliar possíveis desvios nas suposições impostas ao modelo e identificação de observações discrepantes no conjunto de dados. Desse modo, foram construídos os gráficos de resíduos versus ordem das observações e o gráfico de probabilidade normal com envelope simulados. Verifica-se, Figura 4, para o gráfico de resíduos, que os mesmos encontram-se dispostos de forma aleatória em torno do zero e dentro dos limites $(-3, 3)$, em geral. Quanto ao gráfico de probabilidade normal, observa-se que os pontos estão contidos dentro das bandas de confiança, de modo geral, não havendo indícios fortes de afastamento da suposição de que o modelo de regressão beta inflacionado é adequado para os dados.

Figura 4 – Gráfico dos resíduos quantis aleatorizados e normal de probabilidade para o modelo de regressão obtido com dados do Brasil.



Na Figura 5 são apresentados os gráficos das componentes discreta (r_{pt}^D) e contínua do modelo (r_{pt}^C). Esses gráficos são úteis para verificar a presença de observações que podem se destacar como possíveis pontos aberrantes. Nota-se que a observação 2222, que refere-se ao município de Carapebus localizado no Rio de Janeiro, destaca-se como possível ponto

aberrante na amostra. Esse município foi considerado plenamente eficiente no gerenciamento de recursos destinados à atenção básica e também apresentou elevados índices de desenvolvimento municipal na educação (0,72) e saúde (0,88). Na parte contínua do modelo, várias observações apresentam-se mais dispersas e com valor de resíduo acima de 4.

A Figura 6 apresenta os gráficos da distância de Cook versus ordem das observações. Esses gráficos permitem a identificação de pontos influentes na amostra que podem causar, em algumas situações, variações nos parâmetros do modelo e mudanças inferenciais. Conforme a Figura 6 verifica-se que as observações 1171 e 3106, configuram-se como pontos potencialmente influentes na parte discreta do modelo, enquanto a observação 1407 destaca-se na parte contínua. Essas observações referem-se aos municípios de Aracajú-SE, Porto Alegre-RS e Madre de Deus-BA, respectivamente. Particularmente para o município de Madre de Deus, verificamos que além de apresentar o maior gasto per capita em AB, esse município também apresentou elevado escore de eficiência (0,99) e alto índice de desenvolvimento municipal na saúde (0,78).

Figura 5 – Gráficos dos resíduos versus as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) do modelo para o Brasil.

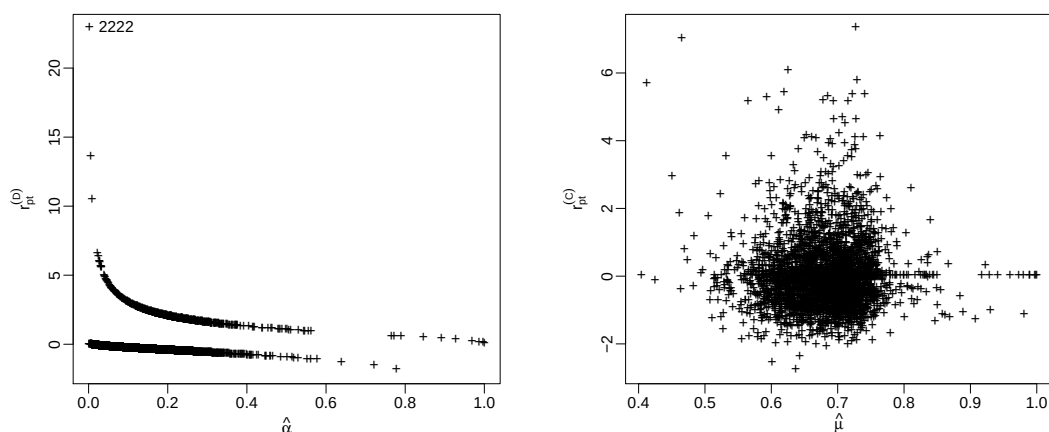
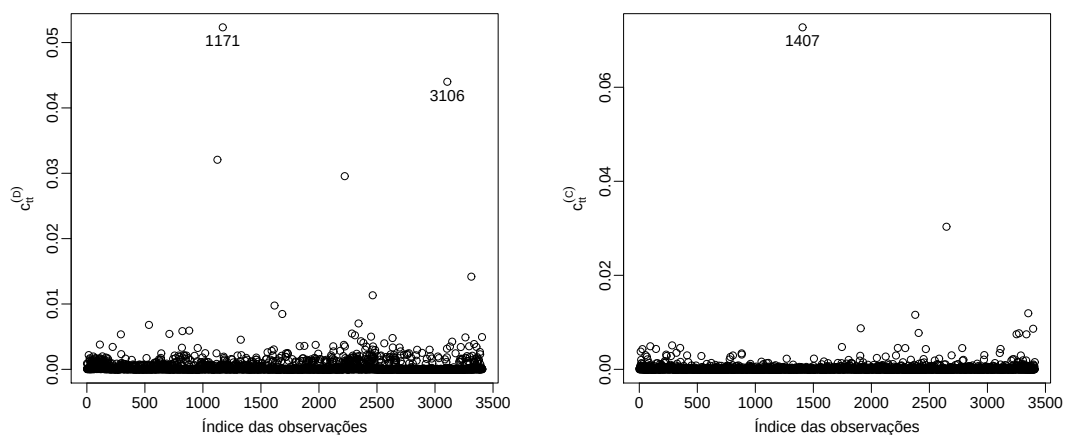


Figura 6 – Distância de Cook versus ordem das observações para a parte discreta e contínua do modelo de regressão obtido a partir dos dados do Brasil.



Para avaliar a influência das observações 1171, 3106 e 1407, nas estimativas dos parâmetros, ajustou-se novamente o modelo descrito anteriormente sem essas observações e foram calculadas as variações percentuais em relação às estimativas dos parâmetros, os resultados encontram-se na Tabela 7. De modo geral, observa-se que as variações com a exclusão das observações foram relativamente baixas, não excedendo 20%. No entanto, quando excluído o município de Aracaju-SE (observação 1171), verifica-se uma variação percentual de 32,29% na estimativa da variável CAP. Esse município foi considerado plenamente eficiente na aplicação de recursos destinados à AB. Quanto a retirada dos municípios de Porto Alegre-RS (observação 3106) e Madre de Deus-BA (observação 1407), não se observam altas variações.

Tabela 7 – Variação percentual das estimativas dos parâmetros sem as observações 1171, 3106 e 1407, que correspondem aos municípios de Aracaju-SE, Porto Alegre-RS e Madre de Deus-BA.

μ			
Variável	% (1171)	% (3106)	% (1407)
Constante	-0,02	-0,57	6,05
POP	-0,01	7,49	-1,93
IFDMS	0,00	-0,71	0,61
GASTO	-0,01	-0,55	8,66
NO	0,00	0,045	-2,64
NE	0,01	-0,00	-1,65
CO	-0,00	1,39	-0,22
SANEA	0,03	-0,49	4,20
α			
Constante	0,26	1,24	0,00
GASTO	0,93	-0,09	0,00
POP	5,40	6,76	-0,00
IDOS	-0,19	-0,03	0,00
PCI	-1,24	-1,06	-0,00
CONSUL	-0,64	-0,26	0,01
CAP	32,29	-10,18	0,00
NE	0,14	0,47	0,01
SE	-1,51	-1,87	-0,01
ϕ			
Constante	0,00	0,05	-0,40
SE	-0,00	-1,33	-0,87
NE	0,00	-2,31	16,43
MED-ENF	0,00	1,59	6,22

6.2 ESPECIFICAÇÃO DO MODELO PARA OS DADOS DA REGIÃO NORDESTE

A presente seção trás os resultados do ajuste referente aos dados da região nordeste. Novamente, após o ajuste do modelo de regressão foi realizado o teste da razão de verossimilhança afim de verificar a hipótese de precisão fixa. Vale ressaltar que o teste foi realizado considerando as variáveis apresentadas na Tabela 8. Como resultado obtivemos p -valor igual a 0,0026 (p -valor $< 0,05$). Dessa forma, rejeitamos ao nível de significância de 5% a hipótese de precisão fixa. Assim, é necessário considerar uma estrutura de regressão para o parâmetro de precisão do modelo. No ajuste do modelo utilizamos a função logit para a média e o parâmetro α , e a função log para o parâmetro de precisão. Foram utilizadas outras funções de ligação, contudo, a partir dessas foram obtidos os melhores resultados.

As estimativas dos parâmetros, erros-padrão e p -valores do modelo ajustado são apresentados na Tabela 8. É possível verificar que apenas às variáveis POP, IFDME e GASTO foram

significativas ao nível de 5% para explicar a eficiência média na AB nessa região. De acordo com as estimativas dos coeficientes percebe-se que as variáveis POP e IFDME exercem efeito positivo na eficiência média da AB. Conforme estudo realizado por Costa (2008), os resultados apontam um impacto positivo e significativo da educação sobre a saúde. Pessoas com nível educacional mais elevado tendem a ser mais informadas e buscam fazer uso dessa informação para alcançar um melhor nível de saúde, através de hábitos saudáveis, prevenção de doenças, etc. Varela (2008) chama a atenção para o fato de que populações mais escolarizadas podem ter um papel mais representativo no controle social, buscando assim, melhores prestações de serviços.

Por outro lado, a variável gasto per capita em AB, influencia negativamente a eficiência do município. O coeficiente negativo atrelado a essa variável é um indicativo de que maiores gastos, não necessariamente, estão relacionados a maior eficiência na AB. Segundo Faria, Jannuzzi e Silva (2008) que avaliaram a eficiência dos gastos municipais em saúde e educação no estado do Rio de Janeiro, concluíram que a eficiência não está relacionada à disponibilidade maior ou menor de recursos, pois, é possível que determinado município tenha um gasto relativamente pequeno, porém invista esses recursos de forma eficiente, ao passo que municípios com maiores recursos aplique-os de forma inadequada. Além disso, Dias (2010) verificou que a eficiência, em média, está associada a menor gasto per capita.

Quanto a variável população do município, seu coeficiente expressa o efeito positivo sobre a eficiência média na AB. Municípios com maior porte populacional tendem a apresentar maiores escores de eficiência na atenção básica, conforme também foi apontado por Varela (2008). Entretanto, vale ressaltar que apesar dessa variável apresentar significância estatística (p -valor $< 0,05$) o valor de sua estimativa é bem próximo de zero ($1,082 \times 10^{-06}$), indicando que a variável exerce pouca influência na variável resposta.

Para a estrutura de regressão do parâmetro α , que modela a probabilidade do município ser plenamente eficiente, apenas as variáveis GASTO, IFDME, POP, IDOS e MED-ENF foram significativas ao nível de significância de 5% (p -valor $< 0,05$). Em relação ao parâmetro de precisão, destacam-se como significativas, ao nível de significância de 5% (p -valor $< 0,05$), as variáveis MEDIA, MED-ENF e GASTO. Através dos coeficientes estimados para as variáveis MEDIA e MED-ENF é possível concluir que os municípios com maior média de médicos por estabelecimento da AB e maior quociente entre o número de médicos em relação ao de enfermeiros, tendem a apresentar respostas menos dispersas, ou seja, escores de eficiência menos dispersos.

Tabela 8 – Modelo de regressão beta inflacionado com dispersão variável ajustado para a região nordeste.

μ			
Variável	Estimativa	Erro-padrão	p -valor
Constante	0,0129	0,1385	0,9258
POP	$1,082 \times 10^{-06}$	$5,267 \times 10^{-07}$	0,0402
IFDME	1,9970	0,2502	< 0,001
GASTO	-0,0014	0,0003	< 0,001
α			
Constante	-1,3120	0,6354	0,0391
GASTO	$-4,837 \times 10^{-03}$	0,0010	< 0,001
IFDME	2,9780	0,8769	< 0,001
POP	$1,862 \times 10^{-06}$	$8,560 \times 10^{-07}$	0,0298
IDOS	-7,5920	3,2660	0,0203
MED-ENF	-0,6002	0,2853	0,0356
ϕ			
Constante	1,5489	0,1902	< 0,001
MEDIA	0,1254	0,0571	0,0283
MED-ENF	0,3296	0,1541	0,0326
GASTO	-0,0008	0,0004	0,0284

Para avaliar a correta especificação do modelo, o teste RESET foi realizado considerando a segunda potência do preditor estimado. Após a aplicação do teste o p -valor obtido foi de 0,4557 e, desse modo, a hipótese nula de que o modelo encontra-se bem especificado não é rejeitada ao nível de significância de 5%. O $pseudo-R^2$ do modelo foi de 0,67, indicando uma boa qualidade no ajuste.

Com objetivo de verificar possíveis afastamentos das suposições feitas para o modelo, a Figura 7 apresenta os gráficos dos resíduos quantis aleatorizados versus ordem das observações e o gráfico normal de probabilidade com envelopes simulados. Nota-se que os resíduos estão dispostos de forma aleatória em torno do zero e encontram-se dentro dos limites $(-3, 3)$. Além disso, a partir do gráfico de probabilidade normal, nota-se que os pontos situam-se, em geral, dentro das bandas de confiança dos envelopes simulados, indicando assim, que não há indícios claros de afastamento da suposição de que o modelo de regressão beta inflacionado é adequado para os dados.

Com o objetivo de identificar possíveis *outliers*, observações atípicas no conjunto de dados, foram construídos os gráficos de resíduos do componente discreto e contínuo do modelo (Figura 8). A observação 540 é destacada como possível ponto aberrante (na componente discreta do modelo). Na parte contínua, a observação 69 destaca-se como possivelmente aberrante.

Figura 7 – Gráfico dos resíduos quantis aleatorizados e normal de probabilidade para o modelo de regressão obtido a partir dos dados da região nordeste.

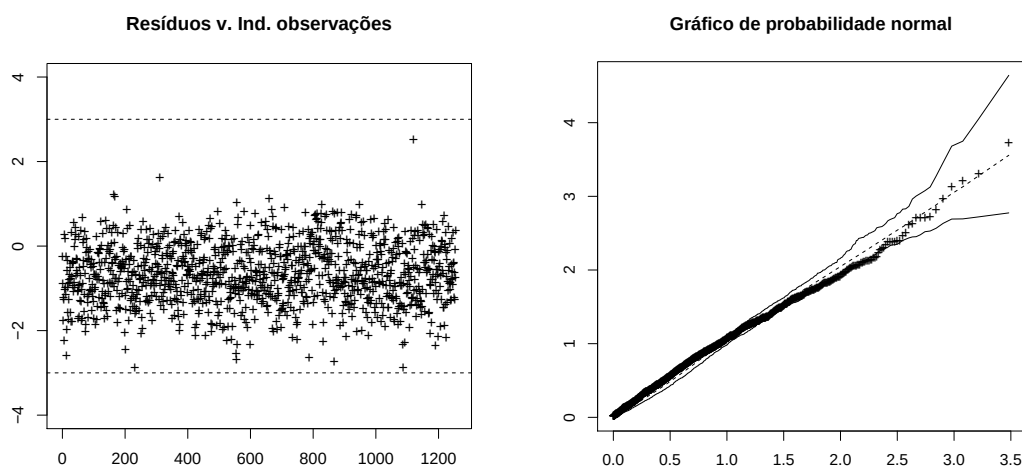
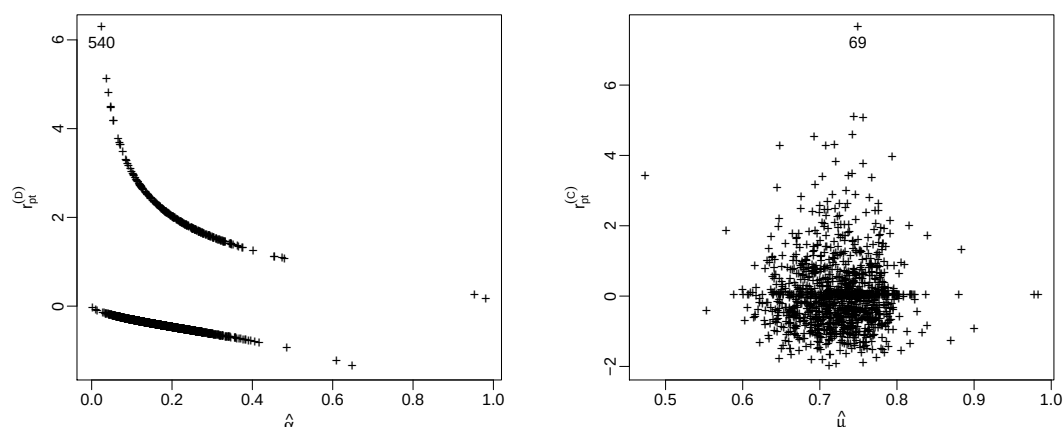
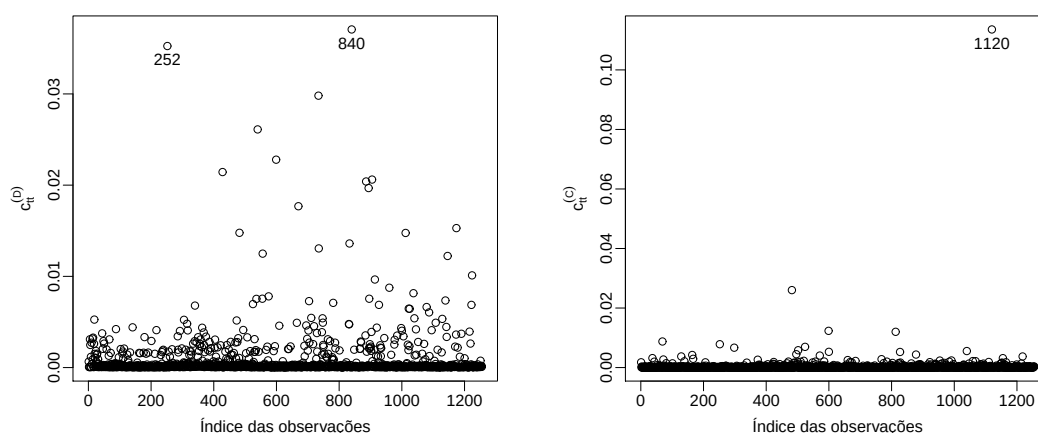


Figura 8 – Gráficos dos resíduos para as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) do modelo de regressão obtido a partir dos dados da região nordeste.



Adicionalmente, com o objetivo de complementar a análise dos resíduos apresentada anteriormente, construímos os gráficos da distância de Cook versus ordem das observações que são apresentados na Figura 9. É possível verificar que as observações 252 e 840 configuram-se como pontos influentes, destacando-se das demais. Essas observações referem-se as cidades de Teresina e Maceió, capitais dos estados do Piauí e Alagoas, respectivamente. Essas cidades apresentaram altos escores de eficiência no gerenciamento de recursos destinados à AB (0,79 e 1). Na parte contínua do modelo, destaca-se a observação 1120, que remete-se ao município de Madre de Deus localizado no estado da Bahia. Esse município apresentou alto escore de eficiência (0,99) e maior valor de gasto per capita em AB observado na amostra (R\$ 1.080,00).

Figura 9 – Distância de Cook versus ordem das observações para o ajuste da região nordeste.



Para avaliar o impacto das observações nas estimativas dos parâmetros, foram calculadas as variações percentuais das estimativas com a exclusão dos pontos potencialmente influentes, os resultados encontram-se na Tabela 9. Nota-se que a exclusão do município de Teresina-PI (observação 252), causa uma variação percentual acima de 24% na estimativa da variável POP na estrutura de regressão da média (μ) e do parâmetro α . No entanto, apesar do município apresentar uma população expressiva, acima de 800 mil habitantes, o mesmo não foi o mais populoso da amostra. De modo semelhante, a exclusão da capital Maceió (observação 840), provoca uma variação percentual na estimativa da variável POP de $-19,21\%$, mesmo também não sendo o município mais populoso do conjunto de dados. Por outro lado, a retirada do município de Madre de Deus-BA (observação 1120), causa alta variação percentual na estimativa da variável gasto per capita em AB. Esse município se destaca por apresentar o maior valor de gasto per capita observado na amostra (R\$ 1.080,00) e também alto escore de eficiência (0,99).

Tabela 9 – Variação percentual das estimativas dos parâmetros para o modelo com dados da região nordeste sem as observações 252, 840 e 1120, que representam os municípios de Teresina-PI, Maceió-AL e Madre de Deus-BA, respectivamente.

μ			
Variável	% (252)	% (840)	% (1120)
Constante	-53,06	0,00	3,54
POP	24,24	-0,03	-4,11
IFDME	-0,01	-0,02	-3,04
GASTO	-0,69	0,00	8,97
α			
Constante	1,43	-1,31	0,06
GASTO	0,56	0,23	-0,05
IFDME	0,75	1,67	0,01
POP	24,85	-19,21	0,01
IDOS	0,52	0,89	-0,01
MED-ENF	-0,14	4,64	0,00
ϕ			
Constante	0,00	0,00	-6,33
MEDIA	-0,33	0,00	-4,03
MED-ENF	0,48	0,00	-2,47
GASTO	-0,25	0,00	-123,09

6.3 ESPECIFICAÇÃO DO MODELO CONSIDERANDO OS MUNICÍPIOS PARA O ESTADO DA PARAÍBA

Na presente seção são descritos os resultados do ajuste do modelo de regressão beta inflacionado para os dados da Paraíba. Novamente, durante o ajuste do modelo, o interesse inicial foi verificar se a hipótese de precisão fixa está sendo violada, ou seja, se há ou não estrutura de regressão para o parâmetro de precisão (CRIBARI-NETO; PEREIRA, 2013). O teste da razão de verossimilhança foi realizado considerando para a estrutura de regressão do parâmetro de precisão as variáveis apresentadas na Tabela 10. Ao aplicarmos o teste da razão de verossimilhança obtivemos p -valor = 0,0013, logo, ao nível de significância de 5%, rejeitamos a hipótese nula de dispersão constante. Assim, além de modelarmos a média da resposta e a probabilidade do município ser plenamente eficiente, é necessário considerar uma estrutura de regressão para o parâmetro de precisão do modelo.

Para a amostra de dados da Paraíba, ao longo da análise, avaliamos a relação de dependência espacial entre os municípios, tendo em vista que municípios vizinhos tendem a ter características semelhantes, em geral. Com o objetivo de considerar tal informação na construção do modelo, realizamos o teste de Moran (CLIFF; ORD, 1973) para verificação de correlação espacial entre os municípios. O p -valor do teste foi de 0,4562, indicando que não rejeitamos a hipótese de independência entre os municípios ao nível de significância de 5%. Ou seja, não

existe evidência estatística de dependência entre os municípios paraibanos quanto a eficiência no gerenciamento de recursos destinados à AB. Além disso, pelo teste de permutação da estatística de Moran (CLIFF; ORD, 1981), também foi apontada a independência entre os municípios (p -valor = 0,45). Uma possível justificativa para isso pode estar no fato da AB ter territorialização. Isto é, todos os municípios têm sua área de atuação e população definida perante os serviços direcionados à AB. Desse modo, com a territorialização, os serviços oferecidos variam de acordo com a situação de saúde de cada população, existindo de certa forma, autonomia entre essas unidades.

A Tabela 10 apresenta as estimativas dos parâmetros, erros-padrão e p -valores das variáveis significativas ao nível de significância de 5%. Para o respectivo modelo, foram consideradas as funções de ligação logit, cloglog e log para modelar a média, o parâmetro α e a precisão do modelo, respectivamente. Com base na Tabela 10 verifica-se que as variáveis SANEIA, CONSUL, MEDIA e a interação entre as variáveis número de consultas por estabelecimento e média de médicos por estabelecimento da atenção AB, foram significativas ao nível de significância de 5%.

Através dos coeficientes estimados do modelo, Tabela 10, é possível concluir que a variável SANEIA exerce efeito positivo sobre a eficiência na AB. Municípios com maior cobertura dos serviços de saneamento básico tendem a apresentar maiores escores de eficiência. Deficiências na cobertura dos serviços de abastecimento de água e coleta de esgoto, foram associados aos índices de mortalidade infantil e mortalidade por doenças infecciosas e parasitárias em todas as faixas de idade (TEIXEIRA; GUILHERMINO, 2006). Conforme Cavinatto (1992), impedir a disseminação de doenças veiculadas por detritos na forma de esgotos e lixo é uma das principais funções do saneamento básico. Os profissionais que operam nesta área também tornam-se responsáveis pelo fornecimento e qualidade das águas que abastecem as populações.

Em contrapartida, as variáveis CONSUL e MEDIA exercem efeito negativo sobre a eficiência média. Municipalidades cuja a média de médicos por estabelecimento da AB é mais elevada, tendem a ser menos eficiente. Isso justifica-se, talvez, pelo fato do alto custo com mão de obra desses profissionais, isto é, o gasto com o salário dos médicos torna-se elevado. Além disso, em algumas situações, a concentração dos profissionais médicos em localidades específicas pode afetar o acesso da população mais carente, que precisa de maior cuidado e acompanhamento (SOUSA, 2011).

Considerando a estrutura de regressão para o parâmetro que modela a probabilidade do município ser plenamente eficiente (α), apenas a variável IFDME foi significativa ao nível de 5%. O nível de escolaridade do indivíduo é tido como um elemento essencial a ser considerado tanto na análise dos determinantes da saúde, bem como na abordagem da população para o avanço de práticas de promoção, prevenção e recuperação da saúde. O baixo grau de escolaridade pode afetar negativamente no desenvolvimento de conceitos de autocuidado em saúde, na conservação ambiental e na percepção de agir do indivíduo enquanto cidadão nos contextos sanitários coletivos

(BRASIL, 2004).

Para o modelo da precisão, à medida que a proporção da população acima de 60 anos aumenta, a precisão das respostas diminui, fazendo com que os escores de eficiência sejam mais dispersos. O envelhecimento populacional, traz consigo a maior propensão de doenças crônicas (DUNCAN et al., 2012) que, em geral, levam a tratamentos contínuos e dispendiosos. Neste sentido, as unidades municipais apresentam diferentes desempenhos quanto ao enfrentamento dos problemas de saúde dessa população em específico, fazendo com que os escores de eficiência, para esses municípios, sejam mais dispersos. De outro modo, as estimativas dos coeficientes para as variáveis PCI e HOSP, sugerem que os municípios que participam de consórcio intermunicipal e possuem hospital, apresentam escores de eficiência menos dispersos.

Tabela 10 – Modelo de regressão beta inflacionado com dispersão variável ajustado para o estado da Paraíba.

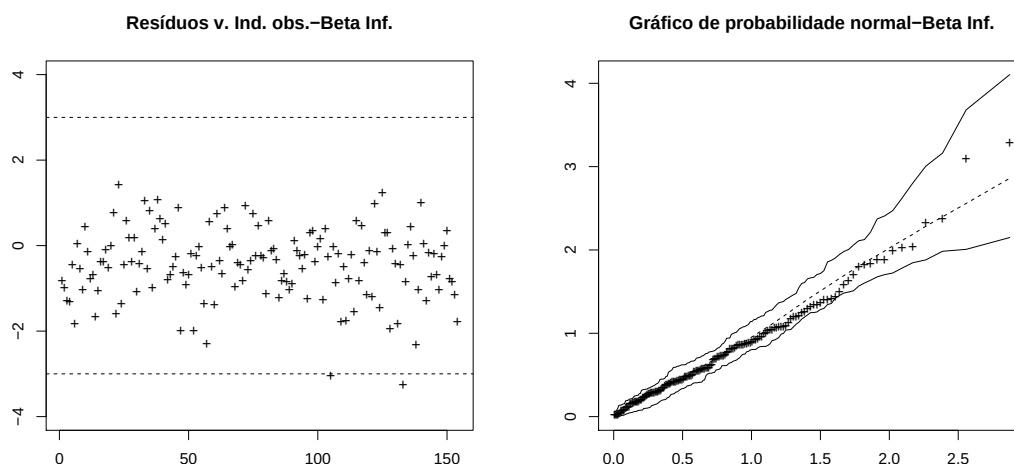
μ			
Variável	Estimativa	Erro-padrão	p -valor
Constante	1,2450	0,2183	< 0,001
SANEA	0,8410	0,3314	0,0122
CONSUL	-0,0003	$5,763 \times 10^{-05}$	< 0,001
MEDIA	-0,6061	0,1380	< 0,001
CONSUL:MEDIA	0,0002	$6,784 \times 10^{-05}$	0,0070
α			
Constante	-7,520	2,474	0,0028
IFDME	8,762	4,049	0,0321
ϕ			
Constante	3,2055	0,4922	< 0,001
IDOS	-13,2098	4,2652	0,0024
PCI	0,5968	0,2796	0,0345
HOSP	0,5189	0,2085	0,0140

Para avaliar a correta especificação do modelo ajustado, aplicamos o teste RESET considerando a segunda potência do preditor linear como variável de teste. O p -valor do teste foi igual a 0,1872, e desse modo, não rejeitamos a hipótese nula de que o modelo de regressão beta inflacionado encontra-se bem especificado, ao nível de significância de 5%. A medida de adequabilidade do modelo, pseudo- R^2 , foi de 0,74, indicando uma boa qualidade no ajuste. Novamente, destacamos o valor expressivo do pseudo- R^2 obtido a partir do modelo, uma vez que torna-se mais complicado a obtenção de medidas altas quando se trabalha com dados reais.

Para verificar possíveis afastamentos das suposições do modelo, a Figura 10 apresenta os gráficos dos resíduos quantis aleatorizados versus os índices das observações e o gráfico de probabilidade normal com envelopes simulados. Nota-se, Figura 10, que o modelo de regressão beta inflacionado parece estar bem ajustado, uma vez que os resíduos encontram-se dentro do intervalo $(-3,3)$. Além disso, pode-se verificar que os resíduos permanecem dentro das

bandas de confiança dos envelopes simulados, não havendo, portanto, indícios de afastamento da suposição de que o modelo de regressão beta inflacionado fornece uma boa representação para os dados observados.

Figura 10 – Gráficos de resíduos quantis aleatorizados versus ordem das observações e de probabilidade normal com envelopes simulados para o modelo de regressão obtido a partir dos dados da Paraíba.



A Figura 11 destaca as observações 6, 22, 109 e 133 como possíveis *outliers* na componente discreta do modelo (r_{pt}^D). Esses pontos referem-se, respectivamente, aos municípios de Algodão de Jandaíra, Belém do Brejo do Cruz, Quixabá e Santarém, que foram considerados plenamente eficientes, com exceção apenas do município de Santarém, com escore de eficiência igual a 0,90. Por outro lado a observação 75, que refere-se o município de Malta, apresenta comportamento atípico em relação às demais observações no gráfico da componente contínua do modelo. Esse município também apresentou alto escore de eficiência (0,99). Na Figura 12, através dos gráficos da distância de Cook versus ordem das observações, nota-se que os pontos 6, 154 e 75 configuram-se como influentes. Esses pontos representam os municípios de Algodão de Jandaíra, Zabelê e Malta, respectivamente.

Figura 11 – Gráficos dos resíduos versus as componentes discreta (r_{pt}^D) e contínua (r_{pt}^C) para o modelo de regressão obtido a partir dos dados da Paraíba.

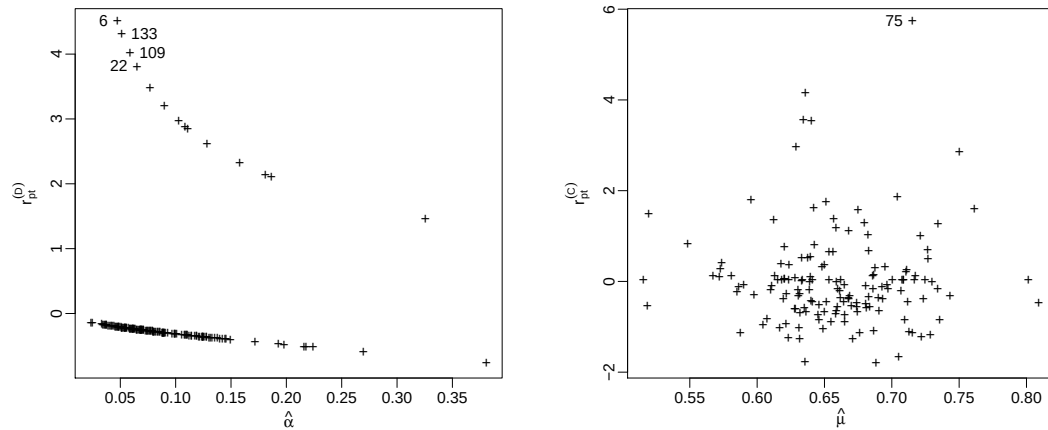
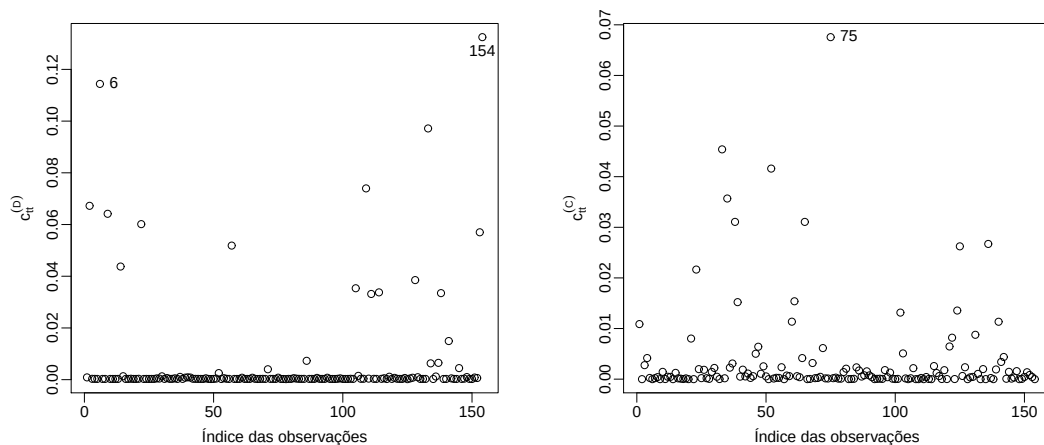


Figura 12 – Distância de Cook versus ordem das observações para o ajuste da Paraíba.



A Tabela 11 apresenta as variações percentuais das estimativas dos parâmetros com a retirada das observações influentes. É possível verificar que a exclusão do município de Algodão de Jandaíra, acarreta em uma variação percentual de 18,25% na estimativa da variável IFDME. O valor do IFDME para esse município foi de 0,51, não representando nem o valor máximo (0,77) e mínimo (0,43) da amostra. Entretanto, esse município além de ser considerado plenamente eficiente também apresenta alto valor de gasto per capita em AB (R\$ 449,00). Com relação ao município de Zabelê (observação 154), verificamos uma variação expressiva na estimativa da variável IFDME (−26,08), isso pode ser justificado pelo fato desse município apresentar o segundo maior valor da amostra para o índice Firjan de desenvolvimento na educação (0,75). De outro modo, a variação percentual associada à retirada do município de Malta (observação 75), foi mais expressiva na estimativa da variável HOSP (54,08%).

Tabela 11 – Variação percentual das estimativas dos parâmetros para o modelo com dados da Paraíba sem as observações 6, 154 e 75, que referem-se aos municípios de Algodão de Jandaíra, Zabelê e Malta.

μ			
Variável	% (6)	% (154)	% (75)
Constante	0,02	−0,02	−10,94
SANEA	0,00	0,00	0,02
CONSUL	−0,00	0,00	−10,44
MEDIA	0,00	0,00	−10,83
CONSUL:MEDIA	0,03	0,03	−8,65
α			
Constante	13,96	−17,46	0,25
IFDME	18,25	−26,08	0,53
ϕ			
Constante	0,00	0,00	−4,56
IDOS	0,01	0,00	−9,80
PCI	0,00	0,00	−7,75
HOSP	−0,01	0,01	54,08

6.4 MAPAS DAS EFICIÊNCIAS ESTIMADAS ATRAVÉS DOS MODELOS DE REGRESSÃO BETA INFLACIONADOS

Nas Figuras 13-15 são apresentados os mapas dos escores de eficiência estimados através dos modelos de regressão. Aqui, o objetivo foi de visualizar de forma mais completa como os escores estão distribuídos. Por exemplo, na Figura 13 é possível observar que no estado da Paraíba predominaram escores de eficiência entre os valores de 0,6 a 0,8. Os municípios que apresentaram escores de eficiência acima de 0,8, estão localizados nas mesorregiões do agreste e sertão paraibano. Para a região Nordeste e o Brasil os escores de eficiência estimados também predominaram entre os valores de 0,6 a 0,8 (Figuras 14 e 15). Além disso, para o Brasil é possível observar que alguns municípios apresentaram eficiência abaixo de 0,5.

Figura 13 – Mapa das eficiências estimadas para os dados da Paraíba.

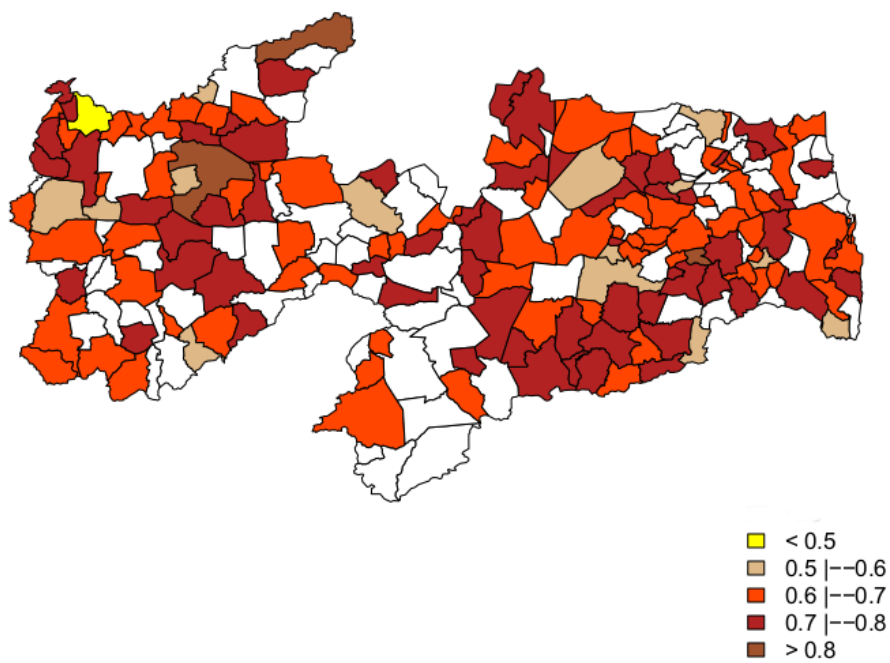


Figura 14 – Mapa das eficiências estimadas para os dados da região nordeste.

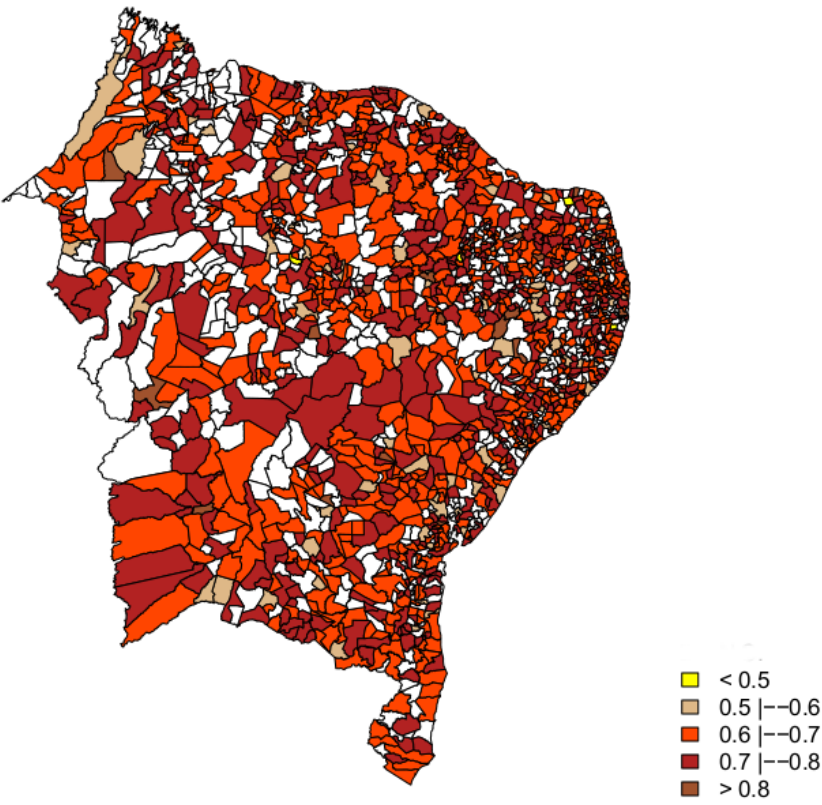
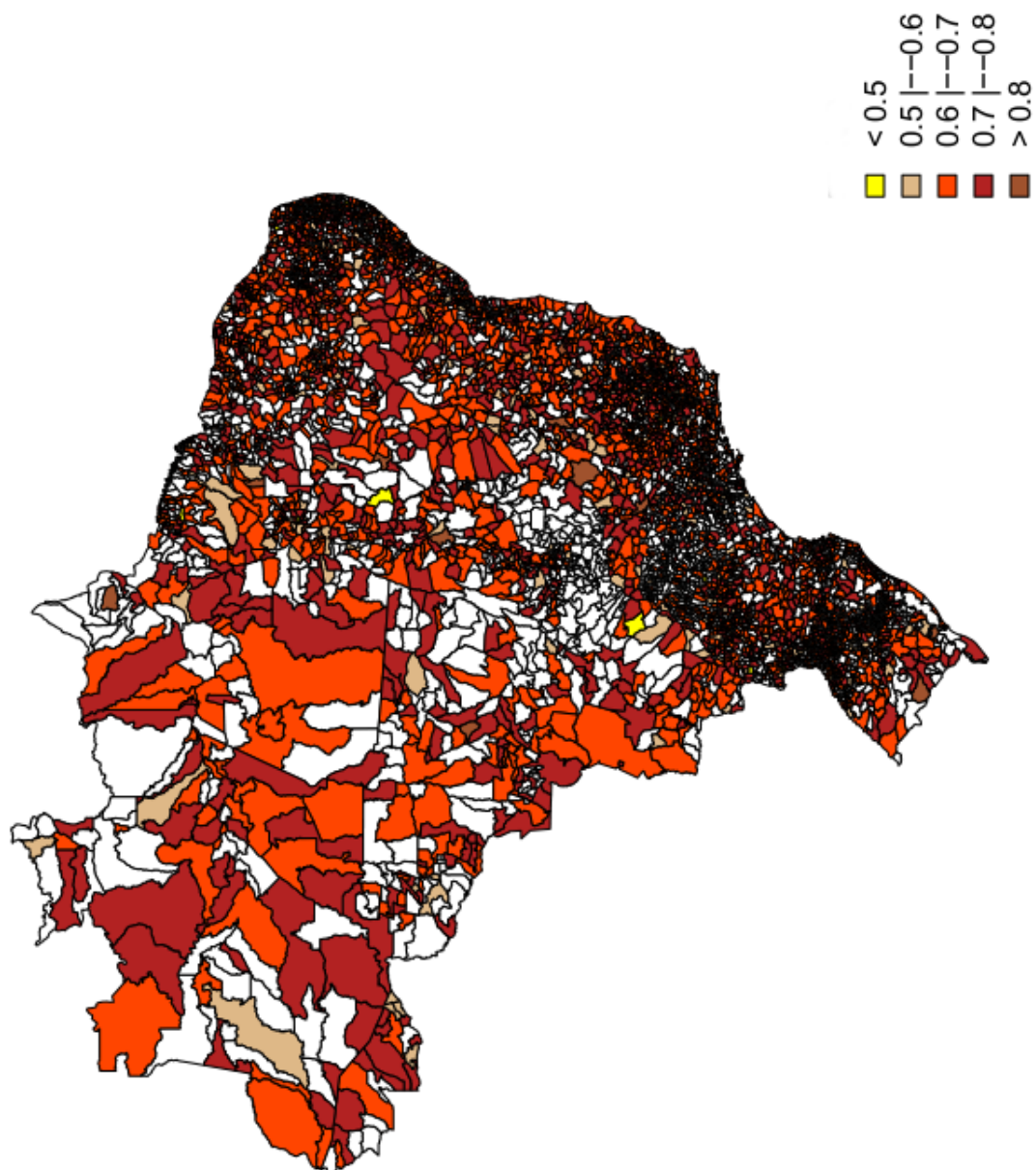


Figura 15 – Mapa das eficiências estimadas para os dados do Brasil.



6.5 ANÁLISE COMPARATIVA DOS RESULTADOS

Comparando-se os resultados obtidos através dos modelos de regressão beta inflacionados, foi possível verificar que algumas variáveis mostram-se mais pertinentes para explicar a eficiência média na AB. Dando enfoque às relacionadas ao contexto da saúde destacam-se as variáveis número de consultas por estabelecimento e média de médicos por estabelecimento da atenção básica. Essas, sendo significativas para o modelo com dados da Paraíba. Por outro lado, o índice Firjan de desenvolvimento municipal na saúde foi considerado um fator importante para explicar a eficiência média na AB, quando consideramos os dados do Brasil.

Para os modelos do Brasil e região nordeste, verificamos que a variável gasto per capita em AB foi significativa para explicar a eficiência média, entretanto, seu coeficiente estando sempre atrelado a valores bem próximos de zero, e com sinal negativo, evidenciando, desse modo, um pequeno impacto na variável resposta.

Com relação a estrutura etária da população, quando analisamos a amostra de dados do Brasil e região nordeste, a variável população do município exerce efeito positivo sobre a eficiência média. Além disso, a proporção da população de pessoas acima de 60 anos também foi relevante para os modelos com dados do Brasil e região nordeste, quando modelamos a probabilidade do município ser plenamente eficiente.

A variável que expressa o percentual de famílias com acesso aos serviços de saneamento básico foi significativa nos modelos com dados da Paraíba e do Brasil, em ambos os modelos seu coeficiente indica a influência positiva na eficiência média da AB. Por outro lado, para o ajuste com dados da região nordeste, a respectiva variável não foi significativa ao nível de 5%.

Em resumo, destaca-se a relevância dos níveis de desenvolvimento na educação e saúde, como uma direção para o alcance de melhor desempenho na gestão dos recursos municipais relacionados à saúde, uma vez que, esses indicadores foram considerados significativos nos modelos ajustados. Além disso, algumas variáveis foram destacadas exercendo efeito negativo sobre a eficiência média da atenção básica. Neste quesito, torna-se relevante uma investigação mais detalhada afim de verificar o comportamento entre as variáveis.

7 CONSIDERAÇÕES FINAIS

Nesta dissertação avaliamos os fatores que influenciam a eficiência técnica na aplicação de recursos destinados à atenção básica no Brasil, na região nordeste e no estado da Paraíba, através dos modelos de regressão beta inflacionados. A amostra foi composta por 3.409 municípios brasileiros, distribuídos nas 5 regiões. A amostra de dados para o nordeste contemplou 1.255 municípios e na Paraíba 154 cidades.

Com relação a análise descritiva pode-se destacar alguns resultados, entre eles de que a eficiência média observada foi de 0,72, com maior número de municípios plenamente eficientes nos estados de São Paulo e Minas Gerais. Apenas 0,12% dos municípios observados possui cobertura completa quanto aos serviços de saneamento básico, sendo Espírito Santo do Pinhal, Ituverava, Junqueirópolis e São José dos Campos, todos localizados no estado de São Paulo. Verificamos que as regiões norte e nordeste apresentam, em geral, maior proporção de crianças com menos de 5 anos de idade, quando comparadas às demais regiões. Entretanto, a proporção de idosos se destaca no sul e sudeste. Os melhores índices Firjan de desenvolvimento na renda e educação foram observados em Itabira-MG e Santa Rita d Oeste-SP, enquanto na saúde os melhores índices foram observados em Flórida no Paraná, Alto Alegre, Rondinha e Saldanha Marinho no Rio Grande do Sul. Do total de municípios da amostra (3.409), apenas 15,6% são considerados plenamente eficientes.

O uso da classe de modelos de regressão beta inflacionados acrescentou resultados relevantes ao estudo. Com base no modelo de regressão para os dados do Brasil, é possível concluir que as variáveis índice Firjan de desenvolvimento municipal na saúde e o percentual de famílias com serviços de saneamento básico, exercem efeito positivo na eficiência média do município. Além disso, os resultados também indicam que os municípios localizados nas regiões norte, nordeste ou centro-oeste, tendem a apresentar maior escore de eficiência médio, quando comparado aos municípios da região sul. No entanto, ressaltamos que o resultado apontado pode estar associado as maiores relações de insumos e produtos nas regiões norte e nordeste do país.

Além disso, para o modelo considerando os municípios do Brasil, o porte populacional do município também foi apontado como uma fator que exerce efeito positivo sobre a eficiência média da atenção básica. De outro modo, a variável gasto per capita exerce efeito negativo na eficiência média da atenção básica. Além disso, os resultados sugerem ainda que, a população municipal e o número de consultas por estabelecimento, são fatores que elevam a chance do município ser plenamente eficiente.

Para o modelo de regressão obtido a partir dos dados da região nordeste, verificamos que a população municipal e o índice Firjan de desenvolvimento na educação exercem efeito positivo sobre a eficiência média da atenção básica. Neste sentido, é possível verificar os determinantes

sociais que estão atrelados a bons escores de eficiência nessa região. Ademais, a população municipal e o índice Firjan de desenvolvimento na educação também são fatores que impulsionam a chance do município ser considerado plenamente eficiente.

O modelo de regressão para os dados da Paraíba destacou variáveis voltadas para o âmbito da saúde para explicar a eficiência na AB. Desse modo, observamos que o número de consultas por estabelecimento e a média de médicos por unidade de atenção básica exercem efeito negativo sobre a eficiência média do município. Relação contrária à apresentada pela variável que representa o percentual de famílias com acesso aos serviços essenciais de saneamento básico, que influencia positivamente a eficiência média da atenção básica. Para os municípios da Paraíba, o nível de desenvolvimento educacional mostra-se um fator relevante para os municípios que buscam a plena eficiência na aplicação dos recursos destinados à atenção básica.

Dado que no Brasil os recursos destinados aos estados e municípios para o provimento de seus serviços são limitados, inclusive, aqueles direcionados para aspectos essenciais como é o caso da saúde, ressalta-se a importância da aplicação dos recursos de forma eficiente, na tentativa de despender menos recursos e evidenciar um modelo de gestão eficaz. Contudo, os resultados encontrados podem proporcionar um melhor direcionamento de recursos destinados à atenção básica no âmbito municipal, permitindo maior eficácia nos serviços oferecidos à população. Ademais, também impulsiona os gestores a buscar administrar os recursos de forma eficiente, minimizando os custos e buscando ofertar um maior número de serviços de qualidade a população.

7.1 TRABALHOS FUTUROS

Dando continuidade, uma vez que a plena eficiência mostrou-se estritamente relacionada com o nível de desenvolvimento do município na educação, como visto nos modelos com dados do nordeste e Paraíba, torna-se necessário investigações mais aprofundadas a respeito do assunto, que possam ampliar os resultados encontrados no presente estudo. Além disso, outras variáveis também mostraram-se relevantes para explicar a eficiência média na atenção básica. Salientamos a importância de ampliar o leque de fatores que possam estar relacionados com a eficiência, uma vez que a base de dados estava restrita à determinadas variáveis.

- Estender os resultados dessa dissertação considerando as micros e mesorregiões do estado, visto que tendem a ser mais homogêneas na forma de estratificação;
- Utilizar modelos de regressão beta ou beta inflacionados para avaliar a eficiência técnica na atenção secundária e terciária na cidade de João Pessoa, com o objetivo de buscar melhorias nesses outros níveis de atenção.

- Avaliar a eficiência técnica na atenção secundária e terciária no estado da Paraíba, através de modelos de regressão beta inflacionados.

REFERÊNCIAS

- ALA-HARJA, M.; HELGASON, S. Em direção às melhores práticas de avaliação. *Revista do Serviço Público*, v. 51, n. 4, p. 5–60, 2000.
- AMADO, C. A.; DYSON, R. G. On comparing the performance of primary care providers. *European Journal Operational Research*, v. 3, p. 915–932, 2008.
- ANDRADE, A. C. G. d. *Efeitos da especificação incorreta da função de ligação no modelo de regressão beta*. Dissertação (Mestrado) — Universidade de São Paulo, 2007.
- ANDRADE, C. A. S. *Avaliação da eficiência técnica do SUS no Estado da Paraíba pelo método da Análise Envoltória de Dados*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2009.
- ATKINSON, A. C. *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. [S.l.]: New York: Oxford University Press, 1985.
- ATUN, R. What are the advantages and disadvantages of restructuring a health care system to be more focused on primary care services? *Health Evidence Network*, 2004.
- BASTOS, E. V. P.; GUIMARÃES, J. C. F.; SEVERO, E. A. Modelo de Regressão Linear para Análise de Investimentos em uma Empresa do Ramo Petrífero. *Revista Produção e Desenvolvimento*, v. 1, n. 1, p. 77–88, 2015.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society B*, v. 26, p. 211–252, 1964.
- BRASIL. *VIII Conferência Nacional de Saúde*. Brasília, 1986.
- BRASIL. *Saúde Brasil 2004 - uma análise da situação de saúde*. Brasília, 2004.
- BRYCE, C. L.; ENGBERG, J. B.; WHOLEY, D. R. Comparing the agreement among alternative models in evaluating hmo efficiency. *Health Service Research*, v. 35, p. 509–528, 2000.
- CAVINATTO, V. M. *Saneamento básico: fonte de saúde e bem estar*. [S.l.]: Ed. Moderna, 1992.
- CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, v. 2, p. 429–444, 1978.
- CLIFF, A. D.; ORD, J. K. *Spatial autocorrelation*. [S.l.]: London: Pion Press, 1973.
- CLIFF, A. D.; ORD, J. K. Spatial processes: models and application. *London: Pion*, v. 12, n. 4, p. 314, 1981.
- COELHO-BARROS, E. A. et al. Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos. *Revista Colombiana de Estadística*, v. 31, n. 1, p. 111–129, 2008.
- COELHO, M. O.; JORGE, M. S. B. Tecnologia das relações como dispositivo do atendimento humanizado na atenção básica à saúde na perspectiva do acesso, do acolhimento e do vínculo. *Ciênc. saúde coletiva*, v. 14, p. 1523–1531, 2009.

- COOK, D. O.; KIESCHNICK, R.; MCCULLOUGH, B. D. On the heterogeneity of corporate capital structures and its implications. 2006. Disponível em: <<http://ssrn.com/abstract=671061>>.
- COOK, R. D.; WEISBERG, S. *Residuals and influence in Regressions*. [S.l.]: London: Chapman and Hall, 1982.
- COPAS, J. B. Binary regression models for contaminated data. *Journal of the Royal Statistical Society B*, v. 50, p. 225–265, 1988.
- CORDEIRO, G. M.; PAULA, G. A. Estimation, large sample parametric tests and diagnostics for non-exponential family nonlinear models. *Communications in Statistics, Simulation and Computation*, v. 21, p. 149–172, 1992.
- COSTA, F. L.; CASTANHAR, J. C. Avaliação de programas públicos: desafios conceituais e metodológicos. *Revista de Administração Pública*, v. 37, n. 5, p. 969–992, 2003.
- COSTA, F. L. d.; FILHO, H. S. R. Avaliação da Eficiência técnica do sistema hospitalar do SUS na Paraíba. In: *Anais do congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural*. Parnaíba, PI, Brasil: [s.n.], 2013.
- COSTA, R. R. F. *O efeito da educação sobre o estado de saúde individual no Brasil*. Dissertação (Mestrado) — Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, 2008.
- COX, D. R.; SNELL, E. J. *Analysis of Binary Data*. [S.l.]: London: Chapman and Hall, 1989.
- CRIBARI-NETO, F.; LIMA, L. B. *A misspecification test for beta regressions*. [S.l.], 2007.
- CRIBARI-NETO, F.; PEREIRA, T. L. Avaliação da eficiência de administrações municipais no estado de São Paulo: uma nova abordagem via modelos de regressão beta. *Revista Brasileira de Biometria*, v. 31, p. 270–294, 2013.
- DAVISON, A. C.; GIGLI, A. Deviance residuals and normal scores plots. *Biometrika*, v. 76, p. 211–221, 1989.
- DIAS, R. H. *Eficiência da atenção primária à saúde nos municípios brasileiros*. Dissertação (Mestrado) — Universidade de Brasília, 2010.
- DUNCAN, B. et al. Doenças crônicas não transmissíveis no Brasil: prioridade para enfrentamento e investigação. *Revista Saúde Pública*, v. 46, n. 1, p. 126–134, 2012.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, v. 5, p. 1–10, 1996.
- ENGSTROM, S.; FOLDEVI, M.; BORQUIST, L. S. Is general practice effective? a systematic literature review. *Scandinavian Journal of Primary Health Care*, v. 19, p. 131–144, 2001.
- ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. *Journal of Applied Statistics*, v. 35, p. 407–419, 2008a.
- FADEL, C. B. et al. Administração pública: o pacto pela saúde como uma nova estratégia de racionalização das ações e serviços em saúde no Brasil. *Revista de Administração Pública*, v. 32, n. 2, p. 445–456, 2009.

FARIA, F. P.; JANNUZZI, P. M.; SILVA, S. J. Eficiência dos gastos municipais em saúde e educação: uma investigação através da análise envoltória no estado do Rio de Janeiro. *Revista de Administração Pública*, v. 42, n. 1, p. 155–177, 2008.

FARRELL, M. J. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (general)*, v. 120, p. 253–290, 1957.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v. 31, p. 799–815, 2004.

Folha de São Paulo. *Ranking de Eficiência dos Municípios- Folha mostra quais cidades entregam mais saúde, educação e saneamento com menos recursos*. [S.l.], 2017. Disponível em: <<http://temas.folha.uol.com.br/remf/ranking-de-eficiencia-dos-municipios-folha/ranking-inedito-revela-que-so-24-das-cidades-sao-eficientes.shtml>>.

FONSECA, P. C.; FERREIRA, M. A. M. Investigação dos níveis de eficiência na utilização de recursos no setor de saúde: uma análise das microrregiões de minas gerais. *Saúde e Sociedade*, v. 18, n. 2, p. 199–213, 2009.

FORREST, C.; STARFIELD, B. The effect of first-contact care with primary care clinicians on ambulatory health care expenditures. *Journal of Family Practice*, v. 43, p. 40–48, 1996.

GONZALEZ, L.; TOZONI-REIS, M.; DINIZ, R. Educação ambiental na comunidade: uma proposta de pesquisa-ação. *Rev. eletrônica Mestr. Educ. Ambient.*, v. 18, p. 379–398, 2007.

HASTIE, T. J.; TIBSHIRANI, R. J. *Generalized Additive Models*. [S.l.]: London: Chapman & Hal, 1990.

HOOFF, A. Second stage dea: comparison of approaches for modelling the dea score. *European Journal of Operational Research*, v. 181, p. 425–435, 2007.

IBGE. *Cidades. Estimativas da População 2017. Instituto Brasileiro de Geografia e Estatística*. 2017. Disponível em: <<https://cidades.ibge.gov.br/brasil/pb/panorama>>.

JÚNIOR, P. M. A.; SOUZA, T. C. Estimativas de votos de Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do bolsa família. *Ciência e Natura*, v. 37, n. 1, p. 12–22, 2015.

KRAVITZ, R.; DUAN, B. J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank quarterly*, v. 82, p. 61–87, 2004.

MANTALOS, P.; SHUKUR, G. The robustness of the reset test to non-normal error terms. *Computational Economics*, v. 30, p. 393–408, 2007.

MARINHO, A. Avaliação da eficiência técnica nos serviços de saúde nos municípios do estado do Rio de Janeiro. *Revista Brasileira de Economia*, v. 57, n. 3, p. 415–432, 2003.

MCCULLAGH, P. *Tensor Methods in Statistics*. [S.l.]: London: Chapman and Hall, 1989.

MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. [S.l.]: Chapman and Hall, 1989.

MCFADDEN, D. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, v. 1, p. 105–142, 1974.

- MENDES, A.; MARQUES, R. M. O financiamento da atenção básica e da estratégia saúde da família no sistema único de saúde. *Saúde em Debate*, v. 38, n. 103, p. 900–916, 2014.
- MERHY, E. E.; FRANCO, T. B. Por uma Composição Técnica do Trabalho Centrada nas Tecnologias Leves e no Campo Relacional. *Saúde em Debate*, v. 27, n. 65, p. 316–323, 2003.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. [S.l.]: John Wiley, 2012.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical A*, v. 135, p. 370–384, 1972.
- OSPINA, R.; FERRARI, S. L. P. Inflated beta distributions. *Statistical Papers*, v. 51, p. 111–126, 2010.
- OSPINA, R.; FERRARI, S. L. P. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, v. 56, p. 1609–1623, 2012.
- PAOLINO, P. Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, v. 9, n. 1, p. 325–246, 2001.
- PAULA, G. A. *Modelos de Regressão com Apoio Computacional*. [S.l.]: Universidade de São Paulo, 2004.
- PEÑA, C. R. Um Modelo de Avaliação da Eficiência da Administração Pública através do Método Análise Envolvória de Dados (dea). *Revista de Administração Contemporânea*, v. 12, n. 1, p. 83–106, 2008.
- PEREIRA, T. L.; CRIBARI-NETO, F. Detecting model misspecification in inflated beta regressions. *Communications in Statistics: Simulation and Computation*, v. 43, p. 631–656, 2014.
- PEREIRA, T. L.; SOUZA, T. C.; CRIBARI-NETO, F. Uma avaliação da eficiência do gasto público nas regiões do Brasil. *Ciência e Natura*, v. 36, p. 23–36, 2014.
- PREGIBON, D. Logistic regression diagnostics. *Annals of Statistics*, v. 9, p. 705–724, 1981.
- QUEIROZ, M. F. M. et al. Eficiência no gasto com saúde: uma análise nos municípios do Rio Grande do Norte. *Revista Econômica do Nordeste*, v. 44, n. 3, p. 761–776, 2013.
- R Core Team. *R: A language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.
- RAMSEY, J. B. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, v. 31, p. 350–371, 1969.
- RAMSEY, J. B.; GILBERT, R. A monte carlo study of some small sample properties of tests for specification error. *Journal of the American Statistical Association*, v. 67, p. 180–186, 1972.
- RIBEIRO, V. C. *Análise de Demanda por Transporte de Passageiros Via Modelos de Regressão Georeferenciados*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2012.
- RODRIGUES, L. B. B. et al. A atenção primária á saúde na coordenação das redes de atenção: uma revisão integrativa. *Ciência saúde coletiva*, v. 19, n. 2, 2014.

- ROSENMAN, R.; FRIESNER, D. Scope and scale inefficiencies in physician practices. *Health Economics*, v. 13, p. 1091–1116, 2004.
- SANTOS, E. G. F. A. *Uma avaliação comparativa da eficiência dos gastos públicos com saúde nos municípios brasileiro*. Tese (Doutorado) — Universidade de São Paulo, 2008.
- SHEA, S. et al. Predisposing factors for severe, uncontrolled hypertension in an inner-city minority population. *New England Journal of Medicine*, v. 327, p. 776–781, 1992.
- SHI, L. The relationship between primary care and life chances. *Journal of Health Care for the Poor and Underserved*, v. 3, p. 21–35, 1992.
- SHI, L. et al. The relationship between primary care, income inequality, and mortality in the United States. *Journal of the American Board of Family Practice*, v. 16, p. 412–422, 2003.
- SHI, L.; STARFIELD, B. Primary care, income inequality, and self-rated health in the U.S. *International Journal of Health Services*, v. 30, p. 541–555, 2000.
- SHUKUR, G.; EDGERTON, D. L. The small sample properties of the reset test as applied to systems of equations. *Journal of Statistical Computation and Simulation*, v. 72, p. 909–924, 2002.
- SILVA, A. O. *Regressão Simplex Não Linear: Inferência e Diagnóstico*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2015.
- SILVA, C. R.; SOUZA, T. C. Modelagem da taxa de analfabetismo no estado da Paraíba via modelo de regressão beta. *Revista Brasileira de Biometria*, v. 32, p. 345–359, 2014.
- SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, v. 54, p. 348–366, 2010.
- SLATER, B.; MARCHIONI, D. M. L.; VOCI, S. M. Aplicação de regressão linear para correção de dados dietéticos. *Rev. Saúde Pública*, v. 41, n. 2, p. 190–196, 2007.
- SMITHSON, M.; VERKUILEN, J. A better lemon-squeezer? maximum like-lihood regression with beta-distributed dependent variables. *Psychological Methods*, v. 11, p. 54–71, 2006.
- SOUSA, M. C. S. *Mensurando a eficiência da atenção primária à saúde em presença de outliers e fatores exógenos: uma análise em dois estágios*. Tese (Doutorado) — (Trabalho apresentado à banca examinadora como parte dos requisitos do Concurso de Professor Titular da Universidade Federal da Paraíba, na área de Economia do Setor Público), Brasília, 2011.
- SOUZA, F. A. M. D.; PAULA, G. A. Deviance residuals for an angular response. *Australian and New Zealand Journal of Statistics*, v. 44, p. 345–356, 2002.
- SOUZA, S. A. et al. Modelagem da proporção de obesos nos Estados Unidos utilizando o modelo de regressão beta. *Ciência e Natura*, v. 38, n. 3, p. 1146–1156, 2016.
- STARFIELD, B. *Primary care: balancing health needs, services, and technology*. [S.l.]: Oxford University Press, 1998.
- STASINOPOULOS, D. M.; RIGBY, R. A. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, v. 23, 2007.

TEIXEIRA, J. C.; GUILHERMINO, R. L. Análise da associação entre saneamento e saúde nos estados brasileiros, empregando dados secundários do banco de dados Indicadores e Dados Básicos para a Saúde 2003 - idb 2003. *Engenharia Sanitária e Ambiental*, v. 11, n. 3, p. 277–281, 2006.

THURSBY, J. G.; SCHMIDT, P. Some properties of tests for specification error in a linear regression model. *Journal of the American Statistical Association*, v. 72, p. 635–641, 1977.

VARELA, P. S. *Financiamento e Controladoria dos Municípios Paulistas no Setor Saúde: Uma Avaliação de Eficiência*. Tese (Doutorado) — Universidade de São Paulo, Departamento de Contabilidade e Atuária, 2008.

VARELA, P. S.; MARTINS, G. A.; FÁVERO, L. P. Desempenho dos municípios paulistas: uma avaliação de eficiência na atenção básica à saúde. *Revista de Administração*, v. 47, n. 4, p. 624–637, 2012.

VIACAVA, F. et al. Avaliação de desempenho de sistemas de saúde: um modelo de análise. *Ciência e saúde coletiva*, v. 17, n. 4, p. 921–934, 2012.

VILLALBI, J. R. et al. An evaluation of the impact of primary care reform on health. *Atencion Primaria*, v. 24, p. 468–474, 1999.

WEINBERGER, M.; ODDONE, E. Z.; HENDERSON, W. G. Does increased access to primary care reduce hospital readmissions? *New England Journal of Medicine*, v. 334, p. 1441–1447, 1996.

WOODWARD, C. A. et al. What is important to continuity in home care? perspectives of key stakeholders. *Social Science and Medicine*, v. 58, n. 1, p. 177–92, 2004.

ZAVRAS, A. I. et al. Using DEA to evaluate efficiency and formulate policy within a greek national primary health care network. *Journal of Medical Systems*, v. 26, p. 285–292, 2002.

Apêndice A — Script utilizado no software R

```
#Limpar memória
rm(list = ls(all=TRUE))

#=====Banco de Dados
dados = read.table("dados_dissertação_brasil.txt", header= T, dec = ",")
attach(dados)
names(dados)
#===== Pacotes úteis
library(fBasics)
library(lmtest)
library(gamlss)
library(nortest)
#=====Descritivas
par(mfrow=c(1,2))
n <- nrow(dados)
p = sum(y[y==1])
hist(y[y>0&y<1], freq = TRUE, main = "",
xlab = "Eficiência",ylab = "Frequência",ylim=c(0,600))
segments(1,0,1,p, lwd = 3)
points(1,p,lwd=1, cex = 1.3)
boxplot(y, xlab="Eficiência",cex.lab = 1.0,horizontal=T)
#===== Ajuste modelo
ajusteBrasil <- gamlss(y~(1),nu.formula=~(1),sigma.formula=~(1),family=BEOI,trace=F)
ajusteBrasil=gamlss(y~v1+v2+v3+v7+v8+v9+v10+v11+v12+v14+v16+v17+v18+v19+v20+v21+v24,
nu.formula=~(1),
sigma.formula=~(1),family=BEOI,trace=F)
tmu = stepGAIC(ajusteBrasil)
tsigma = stepGAIC(tmu, what="sigma", scope=~(v1+v2+v3+v7+v8+v9+v10+v11+v12+v14+v16+
v17+v18+v19+v20+v21+v24))
tsigma
tnu = stepGAIC(tsigma, what="nu", scope=~(v1+v2+v3+v7+v8+v9+v10+v11+v12+v14+v16+
v17+v18+v19+v20+v21+v24))
tnu
ajustemodelo = gamlss(y~ v3 + v7 + v8 + v9 + v16 + v17 + v19 + v21,
sigma.formula=~(v18 + v17 + v9 + v8 + v1),nu.formula=~v8+v3 + v24 + v1 + v11 + v12 +
v21 + v17 + v18,family = BEOI,trace=F)
summary(ajustemodelo)
modelofinal <- gamlss(formula = y ~(v3+v7+v8+v16+v17+v19+v21),
```

```

sigma.formula = ~(v18+v17+v9), nu.formula = ~(v8+v3 + v24 + v1 +
  v11 + v12 +v17+v18),
family=BEOI,trace=F)
summary(modelofinal)
=====
modelo para precisão dados : Brasil
=====
modeloprecisao= gamlss(y~v3+v7+v8+v16+v17+v19+v21,sigma.formula = ~(1),
nu.formula = ~(v8+v3 + v24 + v1 + v11 + v12 +v17+v18),
family = BEOI,trace = FALSE )
summary(modeloprecisao)
library(lmtest)
lrtest(modeloprecisao,modelofinal)
#### Teste de especificação - RESET
lrtest(modelofinal,~.+I(predict(modelofinal,type="link")^2))
=====
coeficiente de determinação
=====
ajustenovo= gamlss(y~ 1 ,sigma.formula = ~1 , nu.formula = ~1 ,
family = BEOI,trace = FALSE)
ltheta<-logLik(modelofinal)
lthetah0<-logLik(ajustenovo)
pR2<-1-(lthetah0/ltheta)
pR2

## Análise de diagnóstico
=====
Construindo o Gráfico de Envelope BEOI
=====
ajusterbeoif = modelofinal
ybeoi=ajusterbeoif$y #substituir "ajusterbeoif" pelo ajuste final do nosso modelo.
muhatbeoi=ajusterbeoif$mu.fv
nuhatbeoi=(ajusterbeoif$nu.fv)
phihatbeoi=(ajusterbeoif$sigma.fv)
Xbeoi=ajusterbeoif$mu.x
Xbeoi<-as.matrix(Xbeoi)
Xbeoin=ajusterbeoif$nu.x
Xbeoin<-as.matrix(Xbeoin)
Xbeois=ajusterbeoif$sigma.x
Xbeois<-as.matrix(Xbeois)
n <- nrow(Xbeoi)

```



```

j<-seq(n)
yhatbeoi<-meanBEOI(ajusterbeoif)
rqbeoi<-residuals(ajusterbeoif, what = "z-scores")
sim = 50
f<- matrix(0,n,sim)
f1<- numeric(n)
f2<- numeric(n)
f3<- numeric(n)
for(i in 1:sim) {
  respbeoi<- rBEOI(n, muhatbeoi, phihatbeoi, nuhatbeoi)
  fitbeoi <- gamlss(respbeoi~Xbeoi-1,nu.formula=~Xbeoi-1,
  sigma.formula=~Xbeois-1,family=BEOI)

  rqbeoibeta<-residuals(fitbeoi, what = "z-scores")
  rqbeoibeta<-sort(abs(rqbeoibeta))

  f[,i]<-rqbeoibeta
}

for(i in 1:n) {
  f1[i] <- min(f[i,])
  f2[i] <- mean(f[i,])
  f3[i] <- max(f[i,])
}

t<-1:n
qq <- qnorm((t+n-1/8)/(2*n+0.5))
# RESÍDUO QUANTIL ALEATORIZADO
par(mfrow=c(1,2))
plot(j, rqbeoi,main = "Resíduos v. Ind. observações",xlab="",ylab="",
pch = "+",ylim=c(-4,4))
abline(h = c(-3,3), lty = 2)
plot(qq, sort(abs(rqbeoi)),pch = "+",ylim=range(abs(rqbeoi),f1,f3),
main="Gráfico de probabilidade normal",xlab="",ylab="")
lines(qq,f1,lty=1)
lines(qq,f3,lty=1)
lines(qq,f2,lty=2)
#dev.off()

#### Calculo dos resíduos para as componentes discreta e contínua do modelo
res1 <- residuals(modelofinal, what=c("z-score"), type=c("weighted"))

```

```

res2 <- residuals(modelofinal, what="nu", type=c("weighted"))
res3 <- residuals(modelofinal, what="mu", type=c("weighted"))
pred1 <- fitted(modelofinal, what="nu", type=c("responce"))
pred2 <- fitted(modelofinal, what="mu", type=c("responce"))
pred3 <- fitted(modelofinal, what="sigma", type=c("response"))

### Resíduos versus predictor: componente discreto e contínuo
par(mfrow=c(1,2))
par(mgp = c(1.8, 0.5, 0))
plot(pred1, res2, pch="+", xlab= expression(paste(italic(hat(alpha)))),
      ylab=expression(paste(ylab=r[pt]^(D)))) # Parte Discreta
identify(pred1, res2, n=2)
plot(pred2, res3, pch="+", xlab= expression(paste(italic(hat(mu)))),
      ylab=expression(paste(ylab=r[pt]^(C)))) # Parte contínua
identify(pred2, res3, n=2)
=====
Distância de Cook
=====
##Parte discreta
M <- ncol(Xbeoin)
pt <- 1/(pred1*(1-pred1))
eta1 <- predict(modelofinal, what = "nu")
d <- exp(eta1)/((1 + exp(eta1))^2)
qt <- pt*(d^2)
Q <- diag(qt)
H1 <- Q^(1/2)%*%Xbeoin%*%solve(t(Xbeoin)%*%Q%*%Xbeoin)%*%t(Xbeoin)%*%Q^(1/2)
h1tt <- diag(H1)
cttD <- (h1tt/(M*(1-h1tt)))*(res2^2)
par(mgp = c(1.8, 0.5, 0))
plot(cttD, xlab="Índice das observações", ylab=expression(paste(ylab=c[tt]^(D))))
identify(cttD,,n=1)
## Parte contínua
m <- ncol(Xbeoi)
q <- ncol(Xbeois)
Phi <- diag(pred3)
wt <- trigamma(pred2*pred3)+trigamma((1-pred2)*pred3)
W <- diag(wt)
deltat <- 1 - pred1
Delta <- diag(deltat)
eta2 <- predict(modelofinal, what = "mu")
t <- exp(eta2)/((1 + exp(eta2))^2)

```

```

T <- diag(t)
H2 <- (Delta**W)^(1/2)**Phi**T**Xbeoi**solve(t(Xbeoi)
**Delta**T**Phi**W**Phi**T**Xbeoi)**t(Xbeoi)**T**Phi**(Delta**W)^(1/2)
h2tt <- diag(H2)
cttC <- (h2tt/((q + m)*(1-h2tt)))*(res3^2)
par(mgp = c(1.8, 0.5, 0))
plot(cttC,xlab="Índice das observações",ylab=expression(paste(ylab=c[tt]^(C))))
identify(cttC,,n=2)

### Construção dos mapas das eficiências estimadas

#carregar pacote
library(maptools)

#Carregar shape
mapa <- readShapePoly("F:/MAPAS 27112017/shapes_brasil/municipios/muniBR.shp")

#Lendo a lista pra não perder
listaBR<-read.csv2("F:/MAPAS 27112017/Mapas_efic_estimadas/listaBRfinal_est.csv")
listaBR<-listaBR[,-1] #removendo a primeira coluna que o R salva

#Ordenando
listaBR <- listaBR[order(listaBR[,1]),]

# Selecionar coluna da variável para a geração do mapa
var = listaBR[,5]

# Criar intervalo de classes
classes = c(-Inf, 0.00001, 0.4999999999, 0.5999999999, 0.6999999999,
0.7999999999, Inf)

# Criar legenda
legenda = c('N.S.', '< 0.5', '0.5 |--0.6', '0.6 |--0.7', '0.7 |--0.8', '> 0.8')

legenda = c('N.S.', '< 0.5', '0.5--0.6', '0.6|--0.7', '0.7|--0.8', '> 0.8')

# Definir as cores para as classes (máximo sete)
cores = c("white", "#e9afafff", "#ff8080ff", "#c83737ff", "#892626ff", "#280b0bff")

## Geração do PLOT para o Brasil
pdf("F:/MAPAS 27112017/Mapas_efic_estimadas/plot_brasil_est.pdf", width=8, height=8)

```

```
# Desenhar mapa coroplático com as classes e cores determinadas anteriormente
plot(mapa, border="black", axes=F, las=1, xlab="", ylab="",
col=cores[findInterval(var, classes, all.inside=TRUE)], forcefill=FALSE)
dev.off()
```

```
# Desenhar a legenda no mapa no canto inferior direito
legend(x="bottomright", legend=legenda, fill=cores, bty="n")
```

O procedimento de estimação, adequabilidade e análise de diagnóstico do modelo ajustado para o Brasil, pode ser estendido para os ajustes da região nordeste e o estado da Paraíba.