

AMANDA DA SILVA LIRA

**TRANSITIVIDADE DE AGRUPAMENTOS ESPACIAIS PECUÁRIOS
BOVINOS NO NORDESTE DO BRASIL**

RECIFE-PE

2019



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA
APLICADA**

**TRANSITIVIDADE DE AGRUPAMENTOS ESPACIAIS PECUÁRIOS
BOVINOS NO NORDESTE DO BRASIL**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Doutora.

Área de Concentração: Biometria e Estatística Aplicada

Orientador: Prof. Dr. Kleber Régis Santoro

RECIFE-PE

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

L768t Lira, Amanda da Silva
Transitividade de agrupamentos espaciais pecuários bovinos no
Nordeste do Brasil / Amanda da Silva Lira. – 2019.
85 f.: il.

Orientador: Kleber Régis Santoro.

Tese (Doutorado) – Universidade Federal Rural de Pernambuco,
Programa de Pós-Graduação em Biometria e Estatística Aplicada,
Recife, BR-PE, 2019.

Inclui referências e apêndice(s).

1. Análise de regressão
2. Teoria bayesiana de decisão estatística
3. Análise por conglomerados
4. Pecuária – Brasil, Nordeste
5. Bovino I. Santoro, Kleber Régis, orient. II. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA
APLICADA

TRANSITIVIDADE DE AGRUPAMENTOS ESPACIAIS PECUÁRIOS BOVINOS NO
NORDESTE DO BRASIL

AMANDA DA SILVA LIRA

Tese julgada adequada para obtenção do título de Doutora em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em __/__/20__ pela Banca Examinadora.

Orientador:

Prof. Dr. Kleber Régis Santoro
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Adauto José Ferreira de Souza
Universidade Federal Rural de Pernambuco

Prof. Dr. Humber Agreli de Andrade
Universidade Federal Rural de Pernambuco

Prof^a. Dra. Daniela Moreira de Carvalho
Universidade Federal Rural de Pernambuco
Unidade Acadêmica de Garanhuns

Prof^a. Dra. Maria do Carmo Soares de Lima
Universidade Federal de Pernambuco

DEDICATÓRIA

Em especial a minha mãe M^a das Graças e
família.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, o grande responsável por escrever as nossas histórias no livro da existência. Fé é crer no que não se vê, é perseverar diante do impossível.

Aos meus familiares pela compreensão da minha ausência durante esse tempo de estudos no doutorado;

Ao meu esposo Luydson pelo amor, paciência, apoio e incentivo nos momentos mais difíceis;

A Universidade Federal Rural de Pernambuco (UFRPE) e a Universidade Federal de Sergipe (UFS) por terem contribuído para o meu crescimento como profissional e humano;

A todos os professores, pela dedicação a árdua tarefa de lecionar e incentivar a pesquisa, pois a influência de bons professores, ninguém consegue apagar, em especial ao meu orientador Kleber Régis com suas brilhantes pontuações, incentivos e ideia para o desenvolvimento deste trabalho;

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro através dos projetos 223635/2014 (DINTER), sem este apoio não teríamos conseguido realizar o doutorado a distância;

Aos colegas do Departamento de Estatística e Ciências Atuariais (DECAT) pelo apoio nesta jornada do doutorado, por terem tido paciência nas ausências das atividades que o departamento necessitava em especial ao prof. Kleber Fernandes que contribuiu com a empreitada de coordenar o DINTER, sem você não teríamos conseguido;

Aos amigos da pós-graduação em especial aos amigos Edyniesky, Elielma, Patrícia e Moises os “lobinhos” do Grupo de Pesquisa em Epidemiologia e Estatística Computacional idealizado pelo prof. Kleber Régis.

Aos amigos mais próximos, por entenderem minha ausência devido o tempo de estudos que o doutorado exigia em especial a Vanessa e Viviane Demarques que cuidaram das minhas plantinhas no meu afastamento, a Luanna por cuidar sempre dos meus pontos de tensões com ventosas milagrosas, pela torcida do pessoal do pilates e do meu clube de ciclismo. Também não posso deixar de agradecer as amigas Patrícia e Vanessa (Maria) pelo apoio e incentivo nas horas difíceis. A todos os amigos que fiz na cidade de Garanhuns. E por fim tenho muito a agradecer e a muitas pessoas. Então a todos que me auxiliaram até onde já cheguei,

Muito Obrigada!

“Deus Sabe...

Quando você está cansado e desencorajado por esforços que não deram frutos, Deus sabe o quanto você tentou quando você chorou por longo tempo, com o coração cheio de angústia, ele contou suas lágrimas.

Se você sente que sua vida está perdida e que muito tempo também se perdeu, Ele está confortando você....

Quando você está solitário e seus amigos estão muito ocupados para um simples telefonema, Ele acompanha você...

Quando você sente que já tentou de tudo e não sabe por onde recomeçar, Ele tem a solução...

Quando nada mais faz sentido e você se sente frustrado e deprimido, Ele tenta lhe mostrar respostas...

Se, de repente, tudo lhe parece mais brilhante e você percebe uma luz de esperança, nesse momento, Ele soprou nos seus ouvidos.

Quando as coisas vão bem e você tem muito para agradecer, Ele está festejando com você...

Quando algo lhe traz muita alegria e você se sente refortalecido, Ele está sorrindo para você...

Quando você tem um propósito a cumprir e um sonho para seguir, Ele abre seus olhos e o chama pelo nome...

Lembre-se que onde você estiver, seja na tristeza ou na felicidade, mesmo que ninguém mais saiba, com certeza, Deus sabe...”

Autor Desconhecido

RESUMO

Neste trabalho, foi analisada a dinâmica espacial do rebanho bovino na região Nordeste do Brasil, para o período de 1974 a 2015. Investigou-se a existência da autocorrelação espacial com a utilização do Índice global e local de Moran I. Foram calculadas as probabilidades das categorias de mudança e permanência dos agrupamentos formados, sendo tais parâmetros estimados através dos modelos de regressão Beta, em que realizou-se predições destes comportamentos. Também foram empregadas simulações para analisar a estabilidade de *clusters* espacial da probabilidade de mudança ou permanência do rebanho bovino com abordagem da regressão Beta Bayesiana. Além de analisar matrizes de estacionariedade dos dados observados e dos modelos de regressão Beta e Beta Bayesiana, também foi verificado o tempo de sobrevivência de um dos municípios de maior ocorrência de transições de *clusters* para o período estudado e assim efetivado comparações entre os modelos encontrados. Os resultados demonstraram que existem indícios de aglomeração do rebanho bovino para os municípios da região Nordeste do Brasil, sendo que foi predominante a existência de porções com baixo efetivo que são circundadas por outras porções também ostentando baixo efetivo do rebanho bovino. Também foi possível constatar que ocorreu uma probabilidade baixa nas mudanças de categorias dos tipos de *clusters*, e que ocorreu um decréscimo do rebanho bovino com o passar dos anos na região Nordeste do Brasil. Ademais, também foi possível observar que existe uma tendência de estacionariedade dos tipos de *clusters*, permanecendo com concentrações do efetivo do rebanho bovino na média da região do Nordeste do Brasil.

Palavras-chaves: *clusters*, dinâmica espacial, evolução da população, índice global e local de Moran, modelos de regressão Beta e Beta Bayesiana, rebanho bovino.

ABSTRACT

In this work, the spatial dynamics of the cattle herd in the Northeast region of Brazil was analyzed for the period from 1974 to 2015. The existence of spatial autocorrelation was investigated using the Moran I global and local index. categories of change and permanence of the formed groupings, being such parameters estimated through Beta regression models, in which predictions of these behaviors were made. Simulations were also carried out to analyze the stability of spatial clusters of the probability of change or permanence of the bovine herd with a Bayesian Beta regression approach. In addition to analyzing stationarity matrices of the observed data and Beta and Beta Bayesian regression models, the survival time of one of the municipalities with the highest occurrence of cluster transitions for the studied period was also verified, thus comparing the models found. The results showed that there is evidence of agglomeration of the bovine herd to the municipalities of the Northeast region of Brazil, being more prevalent the existence of low effective portions that are surrounded by other portions also showing low effective herds of cattle. It was also possible to verify that there was a low probability of changes in categories of clusters, and that there was a decrease in the bovine herd over the years in the Northeast region of Brazil. In addition, it was also possible to observe that there is a tendency of stationarity of the clusters types, remaining with concentrations of the bovine herd in the average of the region of Northeast of Brazil.

Keywords: clusters, spatial dynamics, population evolution, Moran global and local index, Beta and Beta Bayesian regression models, bovine herd.

LISTA DE FIGURAS

Capítulo 3

Figura 1	Região Nordeste do Brasil e seus respectivos Estados.	34
Figura 2	Índice Global de Moran I (1974 a 2015), significativo para todos os anos ($p < 0,001$).	37
Figura 3	Distribuição descritiva do efetivo do rebanho bovino em cabeças na região Nordeste do Brasil de 1974 a 2015.	38
Figura 4	Porcentagem de municípios da região Nordeste do Brasil por categoria de <i>cluster</i> a que pertence, segundo o tamanho do rebanho bovino de 1974 a 2015.	39
Figura 5	<i>Clusters</i> segundo o Índice Local de Moran I para o tamanho do rebanho bovino no Nordeste do Brasil a cada 10 anos (1975 a 2015).	40
Figura 6	Taxa de crescimento populacional da diferença de um ano para outro do rebanho bovino no Nordeste do Brasil (1974 a 2015).	43
Figura 7	Probabilidade de mudança e permanência das categorias de <i>clusters</i> para ligação logit.	45
Figura 8	Probabilidade predita das mudanças e permanências das categorias de <i>clusters</i> espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015), conforme a taxa de crescimento do rebanho (modelo com ligação logística).	45
Figura 9	Probabilidade predita da taxa de crescimento das categorias significativas de <i>clusters</i> espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015), conforme a taxa de crescimento do rebanho (modelo com ligação logística).	47
Figura 10	Projeção do modelo de permanência e mudança das categorias de <i>clusters</i> espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015), conforme a taxa de movimentação do rebanho.	49
Figura 11	Diagrama da transição para a matriz a partir dos valores observados (A) e o estado estacionário (B) de probabilidade para a classificação dos tipos de <i>clusters</i> (estágio atingido com 308 anos).	50
Figura 12	Probabilidade de permanência de categoria de <i>cluster</i> para o município de Serra do Ramalho (BA), sendo o primeiro modelo sem covariáveis (A) e o segundo com a covariável taxa de crescimento (B) para o rebanho bovino.	52

Capítulo 4

Figura 1	Representação gráfica das cadeias geradas (Modelo NS-HH com Burn: 10%, Jump: 10 nsim: 400 000).	58
Figura 2	Representação gráfica das cadeias geradas do Modelo NS-NS para um Burn: 10%, Jump: 20 e um nsim: de 400 000 mil.	62
Figura 3	Diagrama da transição e estacionaridade das matrizes para dados gerais, beta e beta bayesiana segundo a classificação dos tipos de <i>clusters</i> .	65
Figura 4	Número de transições entre os tipos de <i>clusters</i> ao longo dos anos (1974 a 2015).	67
Figura 5	Número de transições dos <i>clusters</i> em relação aos Estados da região	67

	Nordeste	
Figura 6	Probabilidade de permanência de categoria de <i>cluster</i> para o município de Serra do Ramalho (BA) com a covariável taxa de crescimento para dados observados.	68
Figura 7	Probabilidade de permanência de categoria de <i>cluster</i> para o município de Serra do Ramalho (BA), sendo o primeiro modelo com covariáveis para modelos Betas (A) e o segundo com a covariável para modelo Beta Bayesiano (B) para o rebanho bovino.	69

APÊNDICE

Figura 1	Distribuição da média e do desvio-padrão do efetivo do rebanho bovino em cabeças na região Nordeste e em seus municípios, de 1974 a 2015.	80
Figura 2	<i>Clusters</i> segundo o Índice local de Moran I para o tamanho do rebanho bovino no Nordeste do Brasil para os anos de 1974 a 2015.	81
Figura 3	Gráfico da probabilidade normal com envelope simulado (resíduo Pearson) para os modelos da permanência e mudança respectivamente das categorias de <i>clusters</i> .	82
Figura 4	Gráfico da probabilidade normal com envelope simulado (resíduo Pearson) para os modelos das mudanças respectivamente das categorias de <i>clusters</i>	83
Figura 5	Gráfico de resíduo Pearson versus os índices das observações para os modelos da permanência e mudança das categorias de <i>clusters</i> .	84
Figura 6	Gráfico dos resíduos de Pearson versus os índices das observações para os modelos da dos tipos de categorias de <i>clusters</i> .	85

LISTA DE TABELAS**Capítulo 3**

Tabela 1	Classificação dos tipos de <i>Clusters</i> do Índice de Moran I local e mudanças possíveis.	35
Tabela 2	Matriz geral de observação segundo a classificação dos tipos de <i>clusters</i> a partir da contagem frequentista.	35
Tabela 3	Frequências observadas para mudança de categoria do município dada o <i>cluster</i> a que pertence (1974-2015).	46

Capítulo 4

Tabela 1	Esquema de simulações realizadas	58
Tabela 2	Critérios de convergência para modelos Regressão beta bayesiano dos modelos de permanência e mudança das categorias de <i>clusters</i> espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015).	61
Tabela 3	Estimativas a posteriori por ponto e por intervalo de credibilidade dos parâmetros e erro das estimativas	63

SUMÁRIO

LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
Capítulo 1	14
1. INTRODUÇÃO	14
Capítulo 2	17
2. Revisão bibliográfica das técnicas estatísticas	17
2.1 Análise Exploratória de Dados Espaciais	17
2.2 Regressão Beta	19
2.2.1 Critério de Seleção de Modelos	22
2.3 Cadeias de Markov	22
2.4 Regressão Beta Bayesiana	23
2.4.1 Diagnóstico de Convergência	26
2.5 Análise de Sobrevivência	27
Capítulo 3	31
Alterações na agregação espacial da pecuária bovina devido à taxa de crescimento populacional (Proposta para o 1º Artigo)	31
3.1 Introdução	32
3.2 Material	33
3.2.1 A área em estudo	33
3.2.2 Métodos e técnicas	34
3.3 Resultados e discussões	37
3.4 Conclusões	53
Capítulo 4	54

Estabilidade de <i>Clusters</i> espacial pecuário bovino no Nordeste do Brasil (Proposta para o 2º Artigo)	54
4.1 Introdução	55
4.2 Material e Métodos	55
4.3 Resultados e Discussões	60
4.4 Conclusões	71
Capítulo 5	72
5.1 Conclusões gerais	72
5.2 Trabalhos Futuros	73
REFERÊNCIAS	74
APÊNDICE	80

Capítulo 1

1. INTRODUÇÃO

A expansão pecuária tem sido um tema que, gradualmente, vai se incorporando à agenda de pesquisa, por vezes procurando a relação entre desenvolvimento da atividade agrícola e da pecuária (BRANDÃO *et al.* 2005), devido ao processo de expansão de fronteiras na região Amazônica que tem provocado elevados índices de desflorestamento, em que um dos principais fatores apontados tem sido a ocupação pela pecuária bovina (FERRAZ 2001, MUCHAGATA e BROWN 2003).

Sob esse aspecto, a bovinocultura de corte brasileira passou por profundas modificações nos últimos dez anos, período em que foram observadas alterações significativas em questão de produção juntamente com a produtividade, ampliando as fronteiras agrícolas nas regiões Centro Oeste, Norte e Nordeste do país. Nestas regiões destacam os estados do Maranhão, Tocantins, Piauí e Bahia, por permitir acentuado crescimento do efetivo bovino e apresentando também elevado potencial de produção de grãos. Esse conjunto de estados foi denominado MATOPIBA, a qual, apesar de suas deficiências de infraestrutura, tem os preços de terras atrativos, clima correspondente ao do Cerrado e o relevo favorável (MAPA, 2017). Tal crescimento veio acompanhado de considerável aumento nos indicadores tecnológicos de produção e eficiência (ALITI, 2003).

A propensão do Nordeste brasileiro para a exploração pecuária é bastante conhecida. É uma região pioneira na criação de bovinos no Brasil (CAIRES *et al.*, 2009). Atualmente é composta basicamente por animais zebuínos, puros ou mestiços, criados a pasto. Isso se deve, principalmente, à grande adaptação desses animais às condições climáticas, uma vez que, ocorrem secas periódicas, irregularidades nas precipitações pluviais, limitações edáficas e outros limites de natureza ambiental que não têm permitido estabelecimentos de uma agricultura intensiva, mas têm encorajado a exploração pecuária nessa região. Na sua grande maioria, é composta por agricultores familiares, com baixa tecnificação e com produção oscilante, devido aos períodos chuvoso e seco do ano (GALVÃO JUNIOR *et al.*, 2015).

Diversos estudos apontam para o crescimento da produtividade agropecuária ao longo do tempo (GASQUES *et al.*, 2012; FORNAZIER e VIEIRA FILHO, 2013; VIEIRA FILHO, GASQUES e SOUSA, 2012). Frente ao cenário de crescimento, é preciso compreender os

padrões regionais, no intuito de assessorar o planejamento de políticas públicas de desenvolvimento regional e orientação a empreendedores na região Nordeste do Brasil.

Conhecer a natureza da estrutura, das interações e transformações de um fenômeno no espaço é importante para se sugerir padrões e recorrências que possam auxiliar tanto na elaboração de ações públicas por meio de políticas regionalizadas, como na maximização locativa dos recursos privados (ALMEIDA, 2004).

Estudos que abordam a análise espacial e que utilizam informações de indicadores produtivos e econômicos contribuem na estruturação e análise de riscos socioambientais e na identificação de áreas mais vulneráveis para a ocorrência de perdas de quantitativo do rebanho bovino de grande relevância, com vistas à tomada de decisões e implementação de medidas de diferentes magnitudes diante dos fatores predisponentes a sua ocorrência (ALMEIDA, 2004).

A dinâmica da distribuição do rebanho ao longo do tempo e espaço é de essencial importância para a compreensão do fenômeno populacional em seus vários aspectos, tais como implantação, expansão, retração, concentração, apropriação de espaços produtivo-exploratórios, implicações econômicas sociais e formação de *clusters* produtivos.

Com isto, em relação ao efetivo bovino na região Nordeste do Brasil como não há estudos baseados em análise espacial, surgem indagações do tipo: quais seriam os parâmetros que definiriam tal variável e como ela se comporta na modificação espacial nesta região? existe crescimento do rebanho bovino no Nordeste do país?

Logo, a contribuição deste trabalho é permitir um serviço de investigação em que socioeconomistas e especialistas em pecuária de corte poderão analisar e prever a direção da atividade econômica.

Assim, o principal objetivo é analisar as mudanças na agregação espacial do rebanho bovino no Nordeste do Brasil devido relativamente as taxas de crescimento desta população, no período de 1974 a 2015. As taxas de crescimento do rebanho, tanto em crescimento positivo (expansão) como negativo (retração), levariam ao deslocamento da distribuição do rebanho e, conseqüentemente, afetaria todos os atores envolvidos na cadeia produtiva.

Norteados por tais questionamentos, nesta tese foram utilizadas técnicas de estatística espacial que nos fornece indicador global e local de autocorrelação, sendo a variável em estudo o efetivo do rebanho bovino de corte. A escolha desta variável é justificada por sua criação ser a mais difundida mundialmente devido à utilidade que apresenta ao homem – força de trabalho, meio de transporte e principalmente fornecimento de carne, leite e couro (EMBRAPA, 2015).

Além das técnicas espaciais empregaram-se também modelos de regressão beta e beta bayesiano. A distribuição beta é muito versátil e flexível para modelar proporções, pois sua densidade pode assumir diferentes formas dependendo dos valores dos dois parâmetros que indexam a distribuição. Enquanto que a utilização da abordagem bayesiana se faz quando há necessidade do pesquisador de incorporar alguma informação sobre os parâmetros de interesse, ou seja, os parâmetros serão variáveis aleatórias e não mais constantes desconhecidas (REITMAN, 2007).

A presente tese de doutorado está dividida em cinco capítulos independentes. Em que no capítulo 2, apresentamos uma revisão bibliográfica sobre as técnicas estatísticas utilizadas para o desenvolvimento das análises. No capítulo 3, propomos o primeiro artigo da tese que disserta sobre autocorrelação espacial da taxa de crescimento populacional do rebanho bovino na região Nordeste do Brasil, de 1974 a 2015. No capítulo 4, sugerimos o segundo artigo, que trata de simulações da regressão Beta Bayesiana a taxa de crescimento populacional respeitando a mesma região brasileira e o mesmo período supracitados, além de analisar a matrizes de transição dos dados observados, da regressão Beta e Beta Bayesiana para averiguar a estacionariedade destas matrizes e, por fim, realizamos também um estudo do tempo de sobrevivência do município que mais apresentou transições de *clusters*. Finalmente no Capítulo 5 traz as conclusões gerais de todas as análises realizadas na tese. As referências bibliográficas utilizadas neste estudo encontram-se no fim deste trabalho. Alguns resultados que não fizeram parte dos artigos encontram-se ao final inserido em Apêndice.

Capítulo 2

Revisão bibliográfica das técnicas estatísticas

Nesse capítulo são apresentados conceitos básicos de cadeias de markov, análise exploratória de dados espaciais, modelos de regressão beta, beta bayesiano, e análise de sobrevivência para taxa de cura dos *clusters*, de forma a compor um embasamento teórico para a compreensão do trabalho.

2.1. Análise Exploratória de Dados Espaciais

Anselin (1995) e Almeida (2004) definem a Análise Exploratória de Dados Espaciais (AEDE), como sendo, um conjunto de técnicas para descrever e visualizar distribuições espaciais, identificar localidades atípicas (*outliers* espaciais), descobrir padrões de associação espacial (*clusters*) e sugerir diferentes regimes espaciais, ou seja, objetiva identificar padrões de associação espacial para as variáveis abordadas. Desta forma, os efeitos espaciais dizem respeito às comparações causadas pela autocorrelação espacial e pela estrutura espacial (PAULA *et al.*, 2013).

Conhecer a natureza da estrutura e das interações de um fenômeno no espaço é importante para se sugerir padrões e recorrências que possam auxiliar tanto na elaboração de ações públicas por meio de políticas regionalizadas, como na maximização alocativa dos recursos privados. Nesse sentido, a AEDE é uma ferramenta necessária, pois, de acordo com Perobelli *et al.* (2007), este método descreve a distribuição espacial do fenômeno e seus padrões de associação global e local (*clusters*). Entretanto, com a utilização da AEDE, torna-se necessário testar a hipótese de que os dados são distribuídos aleatoriamente, isto é, que os valores de um atributo em uma região não dependem dos valores do mesmo nas regiões vizinhas (ALMEIDA, 2004).

Em outras palavras, verifica se as áreas vizinhas mostram uma semelhança maior comparando a variável de interesse com um padrão aleatório. Portanto, o indicador é uma das maneiras de analisar se as regiões possuem algum padrão espacial de agrupamento (ALBUQUERQUE, 2009).

Para analisar se existe a autocorrelação espacial, inicialmente utiliza-se o coeficiente de correlação espacial global de Moran I,

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i w_{ij} (y_i - \bar{y})^2},$$

em que n é o número de unidades espaciais, y_i é a variável de interesse, w_{ij} é o peso espacial para o par de unidades espaciais i e j , medindo o grau de interação entre elas. A indicação de autocorrelação espacial positiva indica que há similaridade entre os valores dos atributos estudados e a localização do mesmo, ou seja, regiões com valores altos da variável a ser abordada são rodeadas por regiões com valores altos; e regiões com valores baixos consequentemente são rodeadas por regiões que apresentam valores baixos. Por outro lado, a indicação de autocorrelação espacial negativa indica dissimilaridade entre o valor do produto e sua localização (PAULA *et al.*, 2013).

Para possibilitar o cálculo do coeficiente Moran I, se faz necessário escolher uma matriz de pesos, que define o grau de proximidade entre os municípios. Utiliza-se para estimar a variabilidade espacial de dados de área, a matriz de proximidade espacial, também chamada matriz de vizinhança. Dado um conjunto de n áreas $\{A_1, \dots, A_n\}$, construímos a matriz $W^{(1)}$ ($n \times n$), onde cada um dos elementos w_{ij} representa uma medida de proximidade entre A_i e A_j (ALMEIDA, 2004).

Os padrões globais detectam a autocorrelação espacial para todo o espaço analisado. O problema do *Moran I* global é que ele pode esconder padrões locais ou ser influenciado por eles. Para superar esses problemas estatísticos torna-se imprescindível verificar a formação de *clusters* e/ou agrupamentos. Para tanto, Anselin (1995) propôs uma decomposição em categorias do indicador de *Moran I* local, dado por

$$I_i = \frac{(y_i - \bar{y}) \sum_j w_{ij} (y_j - \bar{y})}{\sum_i w_{ij} (y_i - \bar{y})^2 / n} = z_i \sum_j w_{ij} z_j,$$

em que z_i e z_j são variáveis padronizadas e a somatória sobre j é tal que somente os valores dos vizinhos $j \in J_i$ são incluídos. O conjunto J_i abrange os vizinhos da observação i .

De acordo com Almeida (2004), esse indicador, provê uma indicação do grau de agrupamento dos valores similares em torno de uma observação, identificando *clusters* espaciais, estatisticamente significantes. Tais *clusters* são divididos em quatro tipos de associação espacial, sendo: HH: Alto-Alto, LL: Baixo-Baixo, HL: Alto-Baixo; LH: Baixo-Alto. Os agrupamentos HH e LL indicam a ocorrência de autocorrelação espacial positiva,

enquanto que os agrupamentos HL e LH denotam a presença de autocorrelação espacial negativa (ESRI, 2009).

Subjacente ao cálculo do índice, Câmara *et al.* (2004), refere duas hipóteses a serem consideradas

- *Hipótese nula* H_0 : distribuição espacial aleatória dos dados obtidos para a variável em análise, ou seja, *independência espacial*;
- *Hipótese alternativa* H_1 : existência de *dependência espacial*.

Neste contexto, é aconselhável estimar a validade estatística da autocorrelação espacial através da significância do índice, sendo frequente associar a estatística do teste a uma distribuição normal. Se o valor de *p-value*, para a probabilidade, for de 0,05, significa que temos 5% de probabilidade de cometer um erro associado à rejeição nula da hipótese.

2.2 Regressão Beta

A análise de regressão beta, introduzida por Ferrari e Cribari-Neto (2004), é útil para modelar variáveis contínuas que assumem valores no intervalo (0,1), como ocorre, por exemplo, com taxas e proporções. A regressão Beta é desenvolvida assumindo que a variável resposta segue uma distribuição beta, sendo esta uma distribuição muito flexível para modelar proporções, uma vez que a sua função densidade pode tomar formas bastante distintas, dependendo dos valores dos seus parâmetros. Encontra-se mais detalhes em Cribari-Neto e Souza (2012), Simas *et al.* (2010), Espinheira *et al.* (2008b) e Espinheira *et al.* (2008a), e em Simithson e Verkuilen (2006).

Nos casos em que a variável resposta assume valores no intervalo [a,b] (com $a < b$ conhecidos e $a, b \in \mathbb{R}$), em Ferrari e Cribari-Neto (2004) é sugerida uma transformação da variável resposta de $Y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon_i$ para $\frac{y_i - a}{b - a}$, para que possa usar a regressão beta. Mas se a variável resposta assume os valores extremos 0 e 1, uma transformação útil é

$$\frac{y_i(n-1)+0.5}{n},$$

em que n é a dimensão da amostra, também referida em Ferrari e Cribari-Neto (2004). Esta sugestão foi útil na formulação de alguns modelos de ajustamento realizados neste trabalho.

A regressão beta baseia-se numa parametrização alternativa da função densidade beta, em termos da média das variáveis e do parâmetro de precisão. Usualmente, para $Y \sim \text{Beta}(p, q)$, a função densidade beta é expressa por

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad 0 < y < 1,$$

em que $p, q > 0$, são os parâmetros que indexam a distribuição e $\Gamma(\cdot)$ é a função Gama.

Ferrari e Cribari-Neto (2004) objetivando a modelação recorrendo a modelos lineares generalizados, propuseram uma reparametrização da função de densidade descrita anteriormente, definindo $\mu = \frac{p}{p+q}$ e $\phi = p + q$, e obtendo

$$f(y; \mu, \phi) = \frac{\phi}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

em que $0 < \mu < 1$ e $\phi > 0$. Assim $Y \sim \text{Beta}(\mu, \phi)$ e $\mathbb{E}[Y] = \mu$ e $\mathbb{V}[Y] = \frac{\mu(1-\mu)}{1+\phi}$.

O parâmetro ϕ é conhecido como parâmetro de precisão, uma vez que, para μ fixo, ϕ é maior quanto menor for a variância de Y e ϕ^{-1} é designado como parâmetro de dispersão.

Considerando Y_1, \dots, Y_n variáveis independentes, tais que $Y_i \sim \text{Beta}(\mu, \phi)$, $i = 1, \dots, n$, a regressão beta é definida como

$$\begin{cases} h(\mu_i) = \sum_{j=1}^n x_j^T \beta_j = \eta_i \\ g(\phi_i) = z_i^T \gamma \end{cases} \quad (1)$$

em que $\beta = (\beta_1, \dots, \beta_p)^T$ é um vetor de dimensão $(k \times 1)$ de parâmetros de regressão desconhecidos ($k < n$), $x_i = (x_{i1}, \dots, x_{ip})^T$ representa o vetor de covariáveis e η_i é o predictor linear. Além disso, β e γ são, respectivamente, os vetores de parâmetros de média e precisão, e x_i e z_i são, respectivamente, as variáveis explicativas de média e precisão para a i -ésima observação. Por fim, $h(\cdot)$ é a função de ligação, estritamente monótona e diferenciável de ordem dois.

Alguns exemplos de função de ligação usuais na regressão beta são as funções

- logit: $h(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$,
- probit: $h(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ é função de distribuição Normal e
- log-log: $h(\mu) = \log(\log(\mu))$.

Neste trabalho, ajustou-se as funções de ligação logit, probit e complemento log-log. A função preferível foi a função logit, uma vez que, obteve o menor AIC, como veremos nos próximos capítulos, em que

$$\mu_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}, \quad i = 1, \dots, n.$$

Para maiores detalhes sobre estas funções de ligações, ver McCullagh e Nelder (1989).

No modelo de regressão beta, pode-se considerar o método de máxima verossimilhança, em que os parâmetros são estimados através da maximização da função de verossimilhança. O método da máxima verossimilhança (MV) proposto por Fisher (1922) consiste em determinar os valores dos parâmetros que maximizam a função de verossimilhança dada por

$$L(\beta) = \prod_{t=1}^n f(y_t|\beta),$$

em que β é um vetor de parâmetros a ser estimado, y_t ($t = 1, 2, \dots, n$) são valores observados da variável resposta Y e $f(y_t|\theta)$ é a função densidade (probabilidade) proposta para a variável aleatória Y .

O logaritmo da função de verossimilhança para a distribuição beta, a partir de uma amostra de n observações independentes é

$$\ell(\beta, \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \phi_t),$$

em que $\ell_t(\mu_t, \phi_t) = \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) - \log \Gamma[(1 - \mu_t)\phi_t] + (\mu_t \phi_t - 1) \log y_t + [(1 - \mu_t)\phi_t - 1] \log(1 - y_t)$.

É possível utilizar testes assintóticos para fazer inferência dos parâmetros desconhecidos. Considere a hipótese nula $H_0: \beta = \beta^{(0)}$ versus a hipótese alternativa $H_1: \beta \neq \beta^{(0)}$, em que, $\beta = (\beta_1^{(0)}, \dots, \beta_m^{(0)})^T$, para $m < k$. A estatística da razão de verossimilhança é:

$$\bar{\omega}_1 = 2\{l(\hat{\beta}, \hat{\phi}) - l(\tilde{\beta}, \tilde{\phi})\},$$

em que $l(\hat{\beta}, \hat{\phi})$ é o logaritmo natural da função de verossimilhança e $l(\tilde{\beta}, \tilde{\phi})$ é o estimador de máxima verossimilhança restrito de (β, ϕ) obtido pela imposição da hipótese nula, sob condições gerais e de regularidade. Assim, o teste pode ser avaliado usando os valores críticos aproximados da distribuição qui-quadrado χ_m^2 , em que m é a dimensão do espaço paramétrico sob a hipótese nula.

Uma das etapas importantes de qualquer ajuste no modelo de regressão é a análise de diagnóstico e resíduo, com o objetivo de detectar pontos mal ajustados ou aberrantes, bem como a existência de observações extremas (*outliers*) que podem causar inconsistência no resultado final do ajuste, e verificar possíveis afastamentos das suposições sobre o modelo. As análises de resíduos podem basear-se nos resíduos ordinários ou em suas padronizações. Espinheira *et al.* (2008b) propuseram dois novos tipos de resíduos para regressão beta, são elas: os resíduos ponderados e os resíduos padronizados ponderados. Vale salientar que esses dois tipos de resíduos possuem um desempenho superior ao resíduo proposto por Ferrari e Cribari-Neto (2004), especialmente no sentido de identificar observações influentes para as estimativas das médias [ver Rocha e Simas (2011)].

2.3.1. Critério de Seleção de Modelos

Quando se ajusta um modelo a uma variável resposta e se pretende encontrar o melhor modelo, pode-se utilizar o Critério de Informação de Akaike (AIC), que foi desenvolvido por Hirotugu Akaike (1974), sendo uma medida de equilíbrio entre a qualidade do ajuste e o número de parâmetros incluídos no modelo. É uma ferramenta para selecionar um modelo dentre um conjunto de modelos. O AIC classifica os modelos e aquele que possuir o menor AIC deve ser considerado o melhor modelo dentre os estudos (SILVA, 2014).

A estatística correspondente para o modelo em H_0 é,

$$AIC = -2l(\beta; \phi, y) + 2p$$

em que $l(\beta; \phi, y)$ corresponde à função de log verossimilhança considerada, em que o viés é dado assintoticamente por p , em que p é o número de parâmetros a serem estimados no modelo (AKAIKE, 1974).

Outro critério utilizado é o Critério Bayesiano de Schwarz (BIC), proposto por Schwarz (1978), que tem como pressuposto a existência de um “modelo verdadeiro” que descreve a relação entre a variável dependente e as diversas variáveis explanatórias entre os diversos modelos sob seleção. Assim, o critério é definido como a estatística que maximiza a probabilidade de se identificar o verdadeiro modelo dentre os avaliados. O valor do critério BIC para um determinado modelo é dado por

$$BIC = -2 \log f(x_n | \theta) + p \log n,$$

em que $f(x_n | \theta)$ é o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra (SCHWARZ, 1978).

2.3. Cadeias de Markov

Uma Cadeia de Markov é um sistema matemático que experimenta transições de um estado para outro de acordo com um determinado conjunto de regras probabilísticas. Cadeias de Markov são processos estocásticos, mas diferem em que eles devem carecer de qualquer "memória". Ou seja, a probabilidade do próximo estado do sistema depende apenas do estado atual do sistema e não de quaisquer estados anteriores (ROSS, 1996). Isso é chamado de propriedade Markov:

$$P(X_n = i_n | X_{n-1} = i_{n-1}) = P(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}).$$

Para ter um modelo de cadeia de Markov funcional, é essencial definir uma matriz de transição P^t . Uma matriz de transição contém as informações sobre a probabilidade de transição entre os diferentes estados do sistema. Para que uma matriz de transição seja válida, cada linha deve ser um vetor de probabilidade e a soma de todos os seus termos deve ser 1 (um) (ROSS, 1996).

Matrizes de transição têm a propriedade de que o produto de matrizes subsequentes pode descrever as probabilidades de transição ao longo de um intervalo de tempo (ROBERT e CASELLA, 1996). Portanto, é possível modelar a probabilidade de estar em um determinado estado após k etapas, calculando

$$\text{Probabilidades Após } k \text{ passos} = \text{Estado Inicial} * P_t^k$$

A absorção de cadeias de Markov ocorre quando há pelo menos um estado que, uma vez atingido, a probabilidade de permanecer nele é 1 (você não pode abandoná-lo) (ROBERT e CASELLA, 1996).

Cadeias de Markov podem ser descritas por uma sequência de grafos dirigidos, em que as arestas do gráfico n são rotuladas por as probabilidades de ir de um estado no tempo n para outros estados no tempo $n+1$ (NORRIS, 1998) (Figura 1).

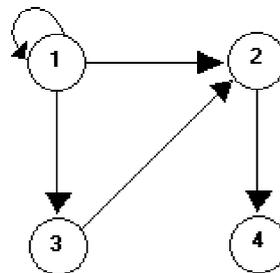


Figura 1. Grafo de uma cadeia de Markov simples de quatro estados.

Neste trabalho serão empregadas as cadeias de Markov para utilização das simulações de modelos bayesianos e para analisar a estacionariedade de matrizes formadas pelos modelos propostos através das regressões Beta e Beta Bayesiana dos tipos de *clusters* permanecerem ou modificarem seus estados.

2.4. Regressão Beta Bayesiana

Cepeda-Cuervo (2001) propuseram a utilização do método Bayesiano juntamente com o algoritmo Markov Chain Monte Carlo (MCMC), empregando o contexto do modelo de

regressão dupla generalizada, para ajustar o modelo de regressão beta. Neste trabalho, almejou-se empregar uma abordagem Bayesiana para estimar os parâmetros do modelo de regressão beta conjunta. Para tanto, necessitou-se especificar uma distribuição prévia para os parâmetros.

Assim, se $L(\Theta|\text{dados})$ denota a função de verossimilhança e $p(\Theta)$ a distribuição prévia conjunta, em que, $\Theta = (\beta, \gamma)$, então a distribuição posterior é dada por $\pi(\Theta|\text{dados}) \propto L(\Theta)p(\Theta)$. No entanto, dado que assumindo distribuições prévias normais para os parâmetros, a distribuição posterior $\pi(\Theta|\text{dados})$ é analiticamente intratável, Cepeda-Cuervo (2001) propôs obter amostras Θ usando um algoritmo iterativo alternativo, por amostragem β e γ .

A partir das distribuições condicionais posteriores $\pi(\beta|\gamma, \text{dados})$ e $\pi(\gamma|\beta, \text{dados})$, para o qual é necessário construir kernels de transição normais q_1 e q_2 , dado que também estas distribuições condicionais são analiticamente intratáveis.

Para construir as funções de transição do kernel, precisamos definir variáveis de observação para aproximar $h(\mu_i)$ e $g(\phi_i)$ em torno dos valores atuais de μ e ϕ , respectivamente. Elas são definidas como aproximações de Taylor de primeira ordem das funções reais $h(T_1)$ e $h(T_2)$, onde t_1 e t_2 são variáveis aleatórias apropriadas tais que $E(T_1) = \mu$ e $E(T_2) = \phi$. Assim, dado que para $E(T_1) = \mu$ se $t_1 = Y$, o modelo médio for dado por (1), a variável observacional é definida por

$$\tilde{y}_i = x_i' \beta^{(c)} + \frac{y_i - \mu_i^{(c)}}{((\mu_i^{(c)})(1 - \mu_i^{(c)}))}, i = 1, 2, \dots, n,$$

em que, $\mu^{(c)}$ e $\beta^{(c)}$ são os valores atuais de μ e β .

Assim, assumindo que \tilde{y}_i , $i = 1, \dots, n$, tem uma distribuição normal e assume a distribuição prévia normal condicional $\beta|\gamma \sim N(b, B)$, a função de transição do kernel q_1 é dada pela distribuição posterior obtida da combinação da distribuição anterior com o modelo de observação $\tilde{y}_i \sim N(x_i' \beta, \tilde{\sigma}_i^2)$, em que $\tilde{\sigma}_i^2 = \text{Var}(\tilde{y}_i)$. Isto é, por

$$q_1(\beta|\beta(C), \gamma(C)) = N(b^*, B^*), \quad (2)$$

em que

$$b^* = B^* (B^{-1}b + X' \Sigma^{-1} \tilde{Y}) \text{ e } B^* = (B^{-1} + X' \Sigma^{-1} X)^{-1},$$

em que Σ é uma matriz diagonal com entradas diagonais $\tilde{\sigma}_i^2$, $i = 1, 2, \dots, n$, [ver CEPEDA-CUERVO, 2001; CEPEDA e GAMERMAN, 2005; CEPEDA-CUERVO e GARRIDO, 2015]. Assim, os valores de β da amostra de distribuição posterior de $\pi(\beta, \gamma)$ são propostos do kernel de transição definido na equação (2).

Como a distribuição condicional completa $\pi(\gamma|\beta)$ também é analiticamente intratável e não é fácil gerar amostras a partir dela, precisamos construir uma transição do kernel q_2 para propor os valores de γ partir da distribuição posterior de Θ . Dado que $E(T_i) = \mu_i$ para $T_i = \frac{(p_i+q_i)^2}{p_i} Y_i$, se o modelo de dispersão é dado por (1), a variável observacional (2) é obtida a partir da aproximação de Taylor de primeira ordem em torno do valor atual de μ_i , dado pelos valores atuais dos parâmetros de regressão de dispersão $\gamma^{(c)}$, por sua vez dado por

$$\tilde{y}_i = z_i' \gamma^{(c)} + \frac{y_i}{\mu_i} - 1, i = 1, 2, \dots, n. \quad (3)$$

Assim, assumindo que a variável observacional (3) tem uma distribuição normal e desde que a distribuição prévia condicional seja dada por $\gamma|\beta \sim N(g, G)$, o kernel q_2 de transição normal é dado pela distribuição posterior obtida da combinação da distribuição anterior com o modelo observacional $\tilde{y}_i \sim N(z_i' \gamma, \tilde{\sigma}_i^2)$, em que $\tilde{\sigma}_i^2 = \text{Var}(\tilde{y}_i)$. Ou seja,

$$q_2(\gamma|\gamma^{(c)}, \beta^{(c)}) = N(g^*, G^*), \quad (4)$$

em que

$$g^* = G^* (G^{-1} g + Z' \Psi^{-1} \tilde{Y}) \text{ e } G^* = (G^{-1} + Z' \Psi^{-1} Z)^{-1}$$

e Ψ é uma matriz diagonal com entradas $\tilde{\sigma}_i^2$, para $i = 1, 2, \dots, n$. Exemplos de γ da distribuição posterior $\pi(\beta, \gamma)$ são obtidos do kernel q_2 da função de transição.

Com os kernels de transição fornecidos por (2) e (4), os componentes β e γ de $(\beta, \gamma)'$ são atualizados os seguintes passos

1. Inicia o contador de interação de cadeias em $j = 1$ e defina valores iniciais (β_0, γ_0) para $(\beta, \gamma)'$.
2. Mova o vetor β para um novo valor ψ gerado a partir da densidade proposta $q_1(\beta^{(j-1)}, \cdot)$.
3. Calcule a probabilidade de aceitação do movimento, $\alpha(\beta^{(j-1)}, \Psi)$. Se o movimento for aceito, então $\beta^{(j)} = \psi$. Se não for aceito, então $\beta^{(j)} = \beta^{(j-1)}$.
4. Mova o vetor γ para um novo valor ψ , gerado a partir da densidade proposta $q_2(\gamma^{(j-1)}, \cdot)$.
5. Calcule a probabilidade de aceitação do movimento, $\alpha(\gamma^{(j-1)}, \Psi)$. Se o movimento for aceito, então $\gamma^{(j)} = \psi$. Se não for aceito, então $\gamma^{(j)} = \gamma^{(j-1)}$.
6. Finalmente, mude o contador de j para $j + 1$ e volte para o passo 2 até que a convergência seja atingida.

Como uma ferramenta de avaliação para o ajuste do modelo, alguns tipos de resíduos são incluídos. Uma deles é residual de Pearson, definida por

$$r_{P_i} = \frac{Y_i - \hat{\mu}_i}{\sqrt{\text{Var}(Y_i)}}, \quad (5)$$

em que $\widehat{\text{Var}}(Y_i) = \hat{\mu}_i(1 - \hat{\mu}_i)/(1 + \hat{\vartheta}_i)$, $\hat{\mu}_i = g^{-1}(x_i'\hat{\beta})$ e $\hat{\vartheta}_i = h^{-1}(z_i'\hat{\gamma})$. Outro é um resíduo ponderado padronizado, definido como

$$r_i^w = \frac{y_i^* - \tilde{\mu}_i^*}{\sqrt{v_i(1 - h_{ii})}}$$

em que

$$y_i^* = \log\left(\frac{y_i}{1-y_i}\right),$$

$$\tilde{\mu}_i^* = \psi(\mu_i\phi) - \psi[(1-\mu_i)\phi],$$

$$v_i = \psi'(\mu_i\phi) + \psi'[(1-\mu_i)\phi]$$

e h_{ii} é o i -ésimo elemento diagonal da matriz $H = \tilde{W}^{-1/2}X(X'\tilde{W}X)^{-1}X'\tilde{W}^{1/2}$, em que W é uma matriz diagonal com elementos $w_i = \vartheta_i\{1/[g'(\mu_i)]^2\}$. Uma das propriedades interessantes do residual (5) é o fato de a sua distribuição poder ser bem aproximada por uma distribuição normal padrão quando comparado com os resíduos padronizados de Pearson. Além disso, dado que incorporam as observações de extrema, podem mais claramente identificar observações influentes para os parâmetros relacionados ao preditor linear (ESPINHEIRA *et al.* 2008a).

2.4.1. Diagnóstico de Convergência

Como os algoritmos MCMC são processos iterativos, são necessários métodos para a avaliação de suas convergências. Estes métodos podem ser classificados como formais e informais. Propostos por Gelfand e Smith (1990), os métodos informais são baseados em análises gráficas para a verificação da convergência. Tais métodos consistem em observar a trajetória da cadeia ao longo das iterações; assim, se após um período inicial a trajetória da cadeia apresentar certa estacionariedade, é possível concluir que a convergência foi atingida. Porém, a utilização apenas destes métodos informais não é recomendada, sendo necessária a utilização de alguns métodos formais. Dentre os métodos formais estão os propostos por Heidelberger & Welch (1983), Geweke *et al.* (1992) e Raftery & Lewis (1992).

O diagnóstico de Heidelberger e Welch (1983) avalia um estado inicial transiente em sequências simuladas de eventos discretos. O diagnóstico proposto é baseado na estatística de Cramer-Von-Mises para testar a hipótese de que os valores simulados formam um processo estacionário. Se a hipótese de estacionariedade não for rejeitada para a cadeia de interesse, o teste half-width é aplicado, consistindo em um intervalo de 95% de confiança para a média. Metade do tamanho desse intervalo é comparada com a média estimada. Se a razão entre a

metade do tamanho e a média estimada for inferior a certo *p-valor*, o parâmetro em teste não é rejeitado. Uma rejeição nesse teste pode ser um indício que o tamanho da cadeia não é grande o suficiente para estimar a média com precisão. Existem muitos métodos para verificar convergência. Para uma descrição e uma lista de referencia veja Gamerman (1997).

Já o de Geweke *et al.* (1992), baseado em métodos de séries temporais, testa se diferentes partições da cadeia se têm médias iguais, a partir de um teste Z. Por *default*, utilizou-se as frações 0.1 e 0.5 da cadeia. O critério proposto por Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o algoritmo apresente convergência a uma distribuição estacionária. O método fornece alternativas do período de aquecimento ou *burn in* (B), que se refere ao número de iterações que devem ser descartadas, do número de iterações que devem ser computadas (N) e o valor K, que é a distância mínima de uma iteração a outra (*thin*) para se obter uma sub-amostra aproximadamente independente. Esse critério consiste em apresentar resultados para todos os parâmetros estimados, portanto, devem-se tomar os maiores valores de B, N e K, pois a convergência só se dará quando as cadeias de todas as variáveis convergirem. Esse critério considera também o fator de dependência, que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência. De acordo com Raftery e Lewis (1992), se esse fator for maior que 5 (cinco), pode-se dizer que a convergência não foi obtida, necessitando reparametrizar o modelo em estudo (BORGES, 2008).

2.5. Análise de Sobrevivência

Nós últimos tempos, tem ocorrido um grande interesse em modelos para dados de sobrevivência com a presença de uma fração de sobreviventes de longa duração (curados ou imunes ao evento). Isto porque, usualmente, os modelos de análise de sobrevivência consideram que todos os indivíduos no estudo são suscetíveis ao evento de interesse, o que pode não ser adequado quando uma fração deles for não suscetível (curados ou imunes). Neste contexto, Corbière e Joly (2007) apresentaram um modelo com fração de cura que assume que a população estudada é uma mistura de indivíduos suscetíveis (não curados), que podem apresentar o evento de interesse, e indivíduos não suscetíveis (curados), os quais nunca apresentarão o evento de interesse. Neste contexto uma metodologia usada em análise de sobrevivência que considera uma proporção de curados são os modelos de longa duração, também chamados de modelos com fração de cura.

Os modelos de sobrevivência de longa duração têm grande importância em análise de dados de sobrevivência e confiabilidade e surgem em várias áreas tais como: biomedicina, finanças, confiabilidade industrial, entre outras. Por isso, diferentes métodos para ajustar tais modelos têm surgido na literatura. Diversos artigos tem abordado o problema de dados de longa duração como, por exemplo, Yakovlev e Tsodikov, 1996; Chen *et al.*, 1999; Tsodikov *et al.*, 2003. Em dados financeiros, o evento de interesse pode ser o desligamento do cliente de um banco devido a várias causas (HOGGART e GRIFFIN, 2001). Em dados biomédicos o evento de interesse pode ser a morte de um paciente submetido a certo tratamento devido a diferentes causas competitivas ou a recorrência do tumor pela presença de um número desconhecido de células cancerígenas.

Para este fim, foi utilizada a análise de sobrevivência aplicada a taxa de crescimento populacional do rebanho bovino dos municípios da região Nordeste do Brasil juntamente com a transições de *clusters* nestes mesmos municípios nos anos de 1974 a 2015.

Com isto a análise de sobrevivência, também chamada de análise de sobrevida, será utilizada quando o tempo (a variável aleatória contínua $T \geq 0$) for o objeto de interesse, seja este interpretado como o tempo até a ocorrência de um evento ou o risco de ocorrência de um evento por unidade de tempo (KLEINBAUM, 1995).

Os parâmetros mais importantes são a probabilidade de sobrevivência no curso de cada um dos intervalos considerados e a probabilidade de sobrevivência acumulada (taxa de sobrevida), isto é, a probabilidade de sobreviver do tempo zero até o tempo final considerado. Esta última equivale à probabilidade de sobreviver em todos os intervalos anteriores ao momento considerado e, usualmente, é denominado $S(t)$. Geralmente há uma variável de interesse, também chamada de variável dependente ou resposta. Às vezes, a variável dependente de interesse é o tempo decorrido até o aparecimento de algum evento. Há, ainda, uma ou mais variáveis, denominadas independentes, preditoras ou covariáveis, cujo relacionamento com a variável dependente, em geral, é imprescindível, pois os modelos estatísticos expressam a variável dependente como uma função matemática conhecida das variáveis independentes (LOUZADA-NETO *et al.*, 2002).

Entretanto, pode acontecer que na população considerada ocorra um grande número de observações censuradas durante o estudo ou ao fim dele, indicando, assim, que nessa população exista uma fração de indivíduos que não estão sujeitos ao evento de interesse. Esses indivíduos podem ser imunes ou curados do evento em questão, considerando que não são acometidos por ele durante o tempo de observação definido. Esta fração de indivíduos imunes ou que se curam é chamada de fração de cura. Nos modelos tradicionais de sobrevivência não

é possível estimar qual é a proporção de cura da população. Dessa forma, é necessária a utilização de modelos estatísticos que são capazes de incorporar na sua modelagem estatística a fração de cura. Assim surgiram vários modelos como uma maneira de superar as limitações dos modelos tradicionais, já que nessa nova classe de modelos admite-se que uma parte da população é imune ao evento de interesse (CALSAVARA, 2011).

Uma das técnicas para modelar a proporção de cura é ajustar um modelo paramétrico, considerando uma mistura de duas distribuições: uma que representa o tempo de sobrevivência para os indivíduos (tipo de *cluster*) suscetíveis e a outra uma distribuição que permite calcular o tempo de sobrevivência dos imunes. Este modelo assume que certa proporção da população são curados (ou não suscetível ao evento de interesse) e o seu complemento $(1-p)$ não são curados (aqueles suscetíveis ao evento de interesse), sendo $0 \leq p \leq 1$.

A função de risco descreve como a probabilidade instantânea de falha (taxa de falha) e se modifica com o passar do tempo, e também é conhecida como taxa de falha instantânea

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t / T \geq t)}{\Delta t}.$$

Ela pode ser utilizada para caracterizar classes especiais de distribuições de tempo de sobrevivência de acordo com o seu comportamento como função do tempo (constante, crescente, decrescente ou mesmo não-monótona). Torna-se, portanto necessária alguma metodologia para selecionar o modelo mais apropriado antes de se proceder ao ajuste. Informações estruturais acerca do fenômeno, vinculadas ao conhecimento do pesquisador sobre o mesmo, podem servir de indicações para a determinação da forma da função de risco. Quando a função de risco for monótona decrescente ou crescente em t , temos a indicação de uma distribuição Weibull (LOUZADA-NETO *et al.*, 2002).

A distribuição de Weibull foi proposta originalmente por Wallodi Weibull em 1951 e, devido a sua simplicidade, tem sido uma das mais utilizadas na modelagem em Análise de Sobrevivência. Sua densidade pode ser escrita na forma

$$f(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\},$$

em que $\gamma > 0$ e $\alpha > 0$, são os parâmetros de forma e escala, respectivamente, e α pode ser interpretado como sendo aproximadamente o i -ésimo percentil da distribuição da variável aleatória T . A partir da densidade se deduz a expressão da função de risco como sendo

$$h(t) = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{\gamma-1}$$

e a função de sobrevivência como

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}.$$

Os procedimentos de estimação de Máxima Verossimilhança para os parâmetros são para a situação em que temos disponível uma amostra aleatória n t_1, \dots, t_n de tempos de sobrevivência e os valores da variável indicadora, δ_i , é igual a 1 se t_i é exatamente observado e igual a zero se t_i é censurado à direita. Baseado nas informações (t_i, δ_i) , que o esquema de censura é não informativo e que os t_i são provenientes da mesma distribuição de probabilidade indexada pelo parâmetro θ , a Função de Máxima Verossimilhança é genericamente escrita na forma

$$L(\theta) = \prod_{i=1}^n h(t_i/\theta) \delta_i S(t_i/\theta)$$

Os modelos de sobrevivência com fração de cura (também chamados de modelos de longa duração) foram inicialmente empregados em estudos clínicos, porém com o avanço computacional várias áreas têm utilizado a aplicações de forma adaptada da sobrevivência, assim como na educação, na criminologia, na engenharia, na biologia, nas finanças, na demografia e etc (LOUZADA-NETO *et al.*, 2002).

Neste trabalho, foi verificada a probabilidade de permanência de categoria de *clusters* através da análise de sobrevivência com fração de cura, uma vez que, podem-se encontrar municípios suscetíveis e não suscetíveis a se modificaram para outro tipo de *clusters*. Foi realizada tal análise para um dos municípios que apresentou a maior quantidade de mudanças de *clusters* nos anos de 1974 a 2015 (tempo), para dois modelos, o primeiro modelo foi considerado sem covariáveis para os anos em estudo e o segundo modelo foi utilizado a covariável taxa de crescimento populacional do rebanho bovino. Em que, se considerou a não ocorrência de mudanças como 0 (zero) e a ocorrência de mudanças foi considerada como 1 (um), que para a sobrevivência seriam a censura e a observação, respectivamente. Realizou-se o ajuste da curva de sobrevivência estimada por Kaplan-Meier e pela distribuição Weibull, estimados através da máxima verossimilhança, devido ao comportamento da variável taxa de crescimento populacional do rebanho bovino e com intuito de comparações entre os modelos.

Capítulo 3

Alterações na agregação espacial da pecuária bovina devido à taxa de crescimento populacional (Proposta para o 1º Artigo)

Resumo

Neste trabalho foi analisada a dinâmica espacial do rebanho bovino na região Nordeste do Brasil, para o período de 1974 a 2015. A escolha do rebanho bovino foi justificada por sua criação ser responsável por uma das mais difundidas mundialmente devido à utilidade que apresenta ao homem – força de trabalho, meio de transporte e principalmente fornecimento de carne, leite e couro. Investigou-se a existência da autocorrelação espacial com a utilização do Índice Global de Moran I, em seguida, identificadas as regiões dos agrupamentos locais dos municípios da região Nordeste do Brasil segundo o efetivo do rebanho bovino com a utilização do Índice Local de Moran I; foram calculadas as probabilidades de mudança e permanência dos agrupamentos e estimados os parâmetros através dos modelos de regressão beta e, com isto, realizou-se predições destes comportamentos. Os resultados demonstraram que existe indícios de aglomeração do rebanho bovino para os municípios da região Nordeste do Brasil e que o comportamento da aglomeração espacial foi mais predominante a existência de porções com baixo rebanho bovino que são circundados por outras porções também ostentando baixo efetivo do rebanho bovino. Também foi possível constatar que ocorreu uma probabilidade baixa nas mudanças de categorias dos tipos de *clusters*, e que ocorreu um decréscimo do rebanho bovino com o passar dos anos na região Nordeste do Brasil.

Palavras-chaves: agrupamento, dinâmica, rebanho bovino, regressão.

Abstract

In this work the spatial dynamics of the bovine herd in the Northeast region of Brazil were analyzed for the period from 1974 to 2015. The choice of the cattle herd was justified by its creation being responsible for one of the most widespread in the world due to its usefulness to man - labor force, means of transport and mainly meat, milk and leather supply. We investigated the existence of spatial autocorrelation using the Moran I Global Index, then identified the regions of the local groups of the municipalities of the Northeast region of

Brazil according to the cattle herd using the Moran I Local Index; the probabilities of change and permanence of the clusters were calculated and the parameters were estimated through the beta regression models and, with this, predictions of these behaviors were made. The results showed that there is evidence of agglomeration of the bovine herd to the municipalities of the Northeast region of Brazil and that the behavior of spatial agglomeration was more predominant the existence of portions with low bovine herd that are surrounded by other portions also boasting low bovine herd . It was also possible to verify that there was a low probability of changes in categories of clusters, and that there was a decrease in the bovine herd over the years in the Northeast region of Brazil.

Keywords: grouping, dynamics, cattle herd, regression.

3.1. Introdução

O agronegócio brasileiro gera 33,5 milhões de empregos. A pecuária de corte é responsável por sete milhões de empregos diretos que geram três indiretos. Isso demonstra sua importância na economia brasileira (TEIXEIRA *et. al*, 2010). Em 2015, o faturamento com as exportações atingiu US\$ 5,9 bilhões, com volume exportado de 1,39 milhão de toneladas de carne bovina. As perspectivas para aumento das exportações é impulsionada pela reabertura de importantes mercados consumidores, como China, Irã, Egito, Japão, Arábia Saudita e Estados Unidos, que apontam para um caminho ainda mais promissor para a carne bovina com selo verde e amarelo (BATISTA FILHO, 2016).

Segundo o Departamento de Agricultura dos Estados Unidos (USDA), o rebanho bovino global está estimado em 1,03 bilhões de cabeças, o maior efetivo está na Índia, com 329,7 milhões (BATISTA FILHO, 2016). Sendo que o rebanho bovino brasileiro chegou a 212,3 milhões em 2014, colocando-o como segundo no ranking mundial (IBGE, 2016).

A bovinocultura de corte brasileira passou por profundas modificações nos últimos dez anos, onde foram observadas alterações significativas em questão de produção juntamente com a produtividade, ampliando as fronteiras agrícolas nas regiões Centro Oeste, Norte e Nordeste do país, nas quais se destacam os estados do Maranhão, Tocantins, Piauí e Bahia (MATOPIBA), permitindo acentuado crescimento do efetivo bovino e apresentando também elevado potencial de produção de grãos. Apesar de suas deficiências de infraestrutura, tem os preços de terras atrativos, clima corresponde ao do Cerrado e o relevo favorável (MAPA, 2017), e tal crescimento veio, acompanhado de considerável aumento nos indicadores tecnológicos de produção e eficiência (ALITI, 2003).

Percebe-se que a produção de carne bovina é uma das atividades agropecuárias de destaque na economia brasileira, sendo que a região Nordeste do Brasil possui o quarto maior rebanho bovino do país, concentrado principalmente nos estados da Bahia, Maranhão, Ceará, Pernambuco e Piauí, em sistema de criação predominantemente extensivo, mas com produtividade baixas em relação ao país, tanto em carne como em leite (EMBRAPA, 2015).

Conhecer a natureza da estrutura, das interações e transformações de um fenômeno no espaço é importante para se sugerir padrões e recorrências que possam auxiliar tanto na elaboração de ações públicas por meio de políticas regionalizadas, como na maximização locativa dos recursos privados (ALMEIDA, 2004).

Estudos que abordam a análise espacial e que utilizam informações de indicadores produtivos e econômicos contribuem na estruturação e análise de riscos socioambientais e na identificação de áreas mais vulneráveis para a ocorrência de perdas de quantitativo do rebanho bovino de grande relevância, com vistas à tomada de decisões e implementação de medidas de diferentes magnitudes diante dos fatores predisponentes a sua ocorrência (ALMEIDA, 2004).

A dinâmica da distribuição do rebanho ao longo do tempo e espaço é de essencial importância para a compreensão do fenômeno populacional em seus vários aspectos, tais como implantação, expansão, retração, concentração, apropriação de espaços produtivo-exploratórios, implicações econômicas sociais e formação de *clusters* produtivos. Sendo, assim, nos leva a hipótese que se as taxas de crescimento do rebanho, tanto em crescimento positivo (expansão) como negativo (retração), levariam ao deslocamento da distribuição do rebanho e, conseqüentemente, afetaria todos os atores envolvidos na cadeia produtiva.

Assim, o principal objetivo deste trabalho é analisar as mudanças na agregação espacial do rebanho bovino no Nordeste do Brasil devido relativamente as taxas de crescimento desta população, no período de 1974 a 2015.

3.2. Material

3.2.1 A área em estudo

O estudo abrange a região do Nordeste do Brasil (1,554 milhões km), equivalente a 18% do território nacional, na qual se inserem nove estados (Bahia/BA, Sergipe/SE, Alagoas/AL, Pernambuco/PE, Paraíba/PB, Rio Grande do Norte/RN, Ceará/CE, Piauí/PI, Maranhão/MA), em 1.794 municípios e habitados por 56.560.081 milhões de pessoas (IBGE, 2015) (Figura 1).

Os dados foram obtidos junto ao Sistema IBGE de Recuperação Automática (SIDRA) do Instituto Brasileiro de Geografia e Estatística (IBGE), representando o efetivo do rebanho bovino de corte por município para os anos de 1974 a 2015, totalizando 73.554 observações.



Figura 1. Região Nordeste do Brasil e seus respectivos Estados.
Fonte: IBGE, 2016.

3.2.2 Métodos e técnicas

Para a avaliação da autocorrelação espacial global e local do efetivo do rebanho bovino foram utilizados os índices global e local de Moran I. Uma vez que, a autocorrelação espacial mede a relação entre observações com proximidade espacial, considerando que observações próximas espacialmente possuam valores parecidos. Os indicadores globais de autocorrelação espacial (Moran I) fornecem uma única medida para o conjunto de todos os municípios, caracterizando toda a região de estudo. E os padrões de distribuição dos indicadores também foram examinados em menor escala por meio do Moran I local (LISA), produzindo um valor específico para cada município, permitindo a visualização de agrupamentos de municípios com valores similares para os indicadores selecionados (ANSELIN, 1995). No caso local se tem a tipologia dos padrões da correlação da forma que se segue, *Alto-Alto* (HH): agrupamento de valores altos e próximos; *Baixo-Baixo* (LL): agrupamento de valores baixos e

próximos; *Alto-Baixo* (HL): *outlier* de valores altos que não se agrupam, pois se encontram em meio a valores baixos; *Baixo-Alto* (LH): *outlier* de valores baixos que não se agrupam, em meio a valores altos. *Não Significativo* (NS): não se enquadram nos agrupamentos anteriores, conforme Anselin (1995).

Após análise do Índice de Moran I local foi realizada a contagem das frequências dos tipos de *clusters* a que pertenciam os municípios, que se formaram de um ano para outro (Tabela 1).

Tabela 1. Classificação dos tipos de *Clusters* gerados pelo índice local de Moran I e mudanças possíveis, avaliada com evolução temporal de ano.

Origem\Destino	NS	HH	HL	LH	LL
NS	Permanece	Mudança	Mudança	Mudança	Mudança
HH		Permanece	Mudança	Mudança	Mudança
HL			Permanece	Mudança	Mudança
LH				Permanece	Mudança
LL					Permanece

Considerou-se que, por exemplo, se em um determinado ano um município pertencia a um tipo de *cluster* HH e no ano seguinte ele passou a ser HL, foi classificado que houve mudança de categoria. Caso o *status* do município devido ao tipo de *cluster* fosse NS e no ano seguinte continuasse NS ele seria classificado como permanência, uma vez que, continuou no mesmo tipo de *cluster*. Com isto, realizou-se o cálculo das probabilidades dos municípios em relação aos tipos de *clusters* (Y) mudarem ou permanecerem nas categorias designadas pelo teste de Moran I local para os anos em análise (1974-2015) (Tabela 2).

Tabela 2. Matriz geral de número de ocorrências observado segundo a classificação dos tipos de *clusters*.

Origem/Destino	NS	HH	HL	LH	LL	Total
NS	41137	430	269	257	1149	43242
HH	416	7177	1	2	0	7596
HL	278	1	2327	0	5	2611
LH	255	45	0	1954	0	2254
LL	853	0	6	0	16992	17851
Total	42939	7653	2603	2213	18146	73554

Foi construída a variável crescimento populacional do rebanho bovino (X), que é a proporção de diferenças de mudança de um ano em relação a outro (1974 a 2015). Ou seja, a taxa de crescimento corresponde a proporção entre a diferença do tamanho do rebanho no ano

seguinte Z_{i+1} (próximo ano) menos o ano atual Z_i (ano atual) dividido pelo tamanho do rebanho no ano Z_{i+1} (próximo ano)

$$\text{Taxa Crescimento Populacional } (X_i) = \frac{Z_{i+1} - Z_i}{Z_{i+1}}$$

E através da modelagem das taxas de crescimento populacional do rebanho bovino e das probabilidades das mudanças ou permanências dos tipos de *clusters*, aplicou-se a regressão Beta (FERRARI e CRIBARI-NETO, 2004) para realizar as estimativas destas probabilidades utilizando ao modelo as ligações, logit, probit e complemento log-log. A escolha do melhor modelo empregou o critério de AIC (AKAIKE, 1974). Além disso, foi verificado o ajuste dos modelos através da análise dos resíduos (Apêndice). Sendo o modelo proposto da seguinte forma

$$\text{Prob. (tipos de clusters)} = \beta_0 + \beta_1 * \text{Taxa de Crescimento Populacional} + e,$$

em que a probabilidade dos tipos de *clusters* é a variável de interesse (por exemplo, HH-HH padrão alto-alto), a taxa de crescimento populacional é a matriz da variável explicativa e β é o vetor de coeficiente de associação e e é o erro aleatório do modelo.

Também foi realizada a análise das matrizes de transição e estacionárias dos coeficientes estimados da regressão Beta através da Cadeia de Markov (ROBERT e CASELLA, 1996).

E, por fim, foi verificada a probabilidade de permanência de categoria de *clusters* através da análise de sobrevivência com fração de cura, uma vez que, podem-se encontrar municípios suscetíveis e não suscetíveis a se modificarem para outro tipo de *clusters*. Foi realizada tal análise para um dos municípios que apresentou a maior quantidade de mudanças de *clusters* nos anos de 1974 a 2015 (tempo), para dois modelos, o primeiro modelo foi considerado sem covariáveis para os anos em estudo e o segundo modelo foi utilizado a covariável taxa de crescimento populacional do rebanho bovino. Em que, se considerou a não ocorrência de mudanças como 0 (zero) e a ocorrência de mudanças foi considerada como 1 (um), que para a sobrevivência seriam a censura e a observação, respectivamente. Realizou-se o ajuste da curva de sobrevivência estimada por Kaplan-Meier e pela distribuição Weibull, estimados através da máxima verossimilhança, uma vez que, são indicados devido ao comportamento da variável taxa de crescimento populacional do rebanho bovino e com intuito de realizar comparações entre os modelos.

Na análise estatística dos dados, foram empregados o software ArcMap (versão 10.4.1) para o estudo espacial, além do software estatístico R-project (versão 3.4.1) (R CORE TEAM, 2017). O pacote utilizado para modelagem de regressão Beta no R foi o *betareg*

(CRIBARI-NETO e ZELEIS, 2010). Foi utilizado o pacote *markovchain* para analisar a probabilidade de transição das matrizes e o comando *steadyStates*, para encontrar o estado estacionário. Em que os estados estacionários são identificados a partir dos quais os autovetores correspondem aos autovalores de identidade e depois os normalizam para somar a unidade. Quando valores negativos são encontrados na matriz, a extração de autovalores é realizada na submatriz de classes recorrentes (ROSS, 2010). E para análise de sobrevivência foi utilizado os pacotes *survival* e o *flexsurvreg*, devido a flexibilização da utilização da fração de cura que foi considerada pela mudança de um *cluster* passar para outro tipo de *cluster* com incidir dos anos.

3.3. Resultados e discussões

Através do índice global de Moran I foi possível verificar que ocorreram agrupamentos significativos na região ao longo do tempo com correlação espacial positiva (Figura 2), os padrões destes *clusters* foram verificados de baixos valores, indicando assim, que os municípios da região Nordeste do Brasil ocorrem baixo efetivo do rebanho bovino e que são circundados por outros que também ostentam baixos rebanhos.

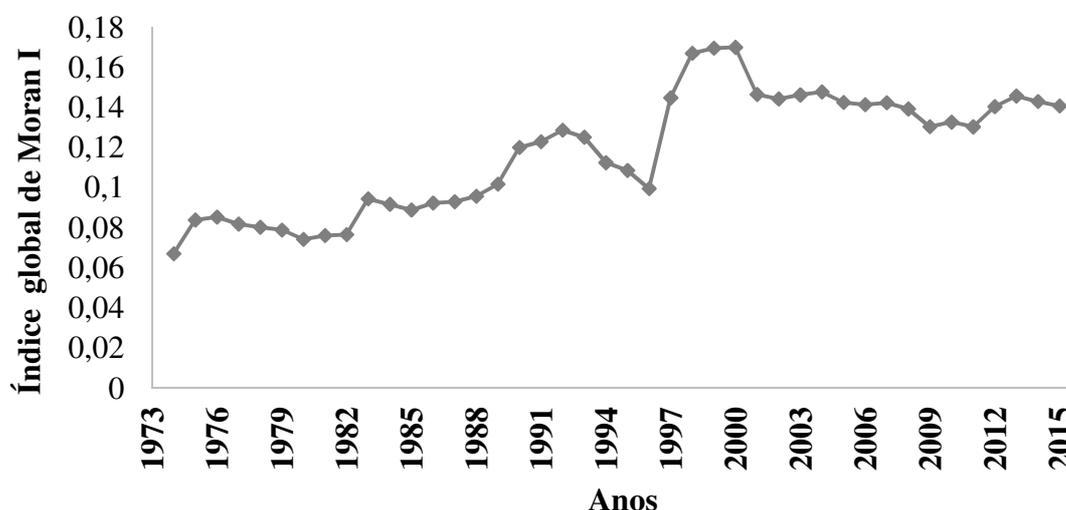


Figura 2. Índice Global de Moran I (1974 a 2015), significativo para todos os anos ($p < 0,001$).

Com o intuito de analisar com mais precisão a dinâmica do rebanho bovino na região Nordeste do Brasil, foi realizado um estudo do índice de Moran I local conjuntamente com a distribuição descritiva destes dados.

Foi possível verificar que a região Nordeste obteve um crescimento populacional do rebanho bovino de forma modesta (0,3%) (Figura 3), ao se comparar com os resultados apontados de que a região possui baixos agrupamentos do rebanho poderiam estar relacionados a propensão que a região tem para ocorrência de estiagem, que afeta diretamente o setor, uma vez que, pode promover reduções no consumo alimentar, na reprodução, na gestação, na lactação e, conseqüentemente, na eficiência produtiva dos rebanhos (AVENDAÑO-REYES *et al.*, 2006; GARCÍA-ISPIERTO *et al.*, 2006).

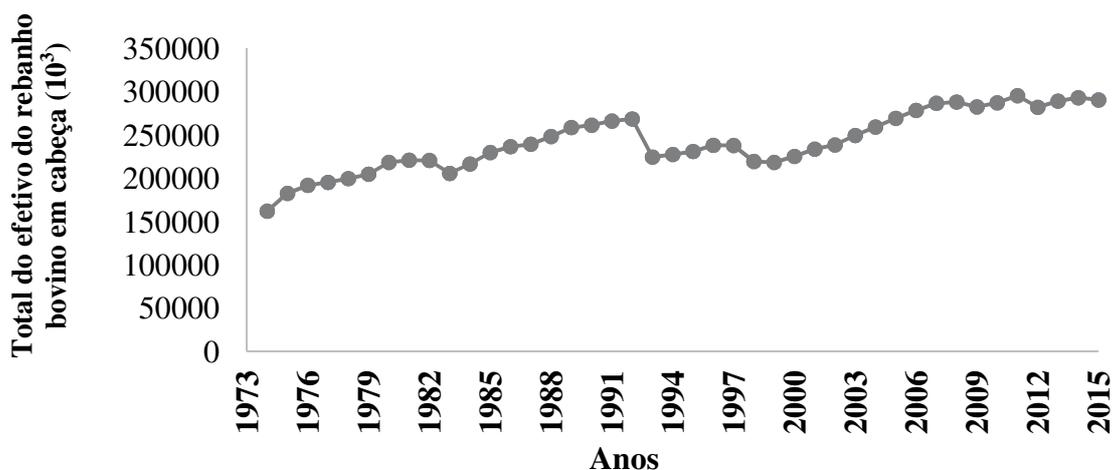


Figura 3. Distribuição do efetivo do rebanho bovino em cabeças na região Nordeste e em seus municípios, de 1974 a 2015.

Para a análise de agrupamentos locais, observa-se um comportamento de retração de municípios classificados como NS e um aumento dos agrupados em LL (Figura 4). Porém pode-se averiguar também pouca variação de *clusters* do padrão HH, em torno de 10%, ou seja, mostrando haver municípios com vizinhos de valores superiores à média do conjunto. Este comportamento, juntamente ao crescimento populacional da região (Figura 3), indica que os municípios começaram a se definir próximo ao ano de 2001 como criadores ou não de bovinos. Uma vez que, neste período entr 1974 a 2000 ocorreram intensas transformações necessitadas com a importância de se desenvolver políticas agrícolas para região Nordeste devido as estiagens prolongadas ocorridas.

Assim como descreve Mendes *et al.* (2009), o Brasil reduziu significativamente, desde a década de 1980, os investimentos em infraestrutura, o que possivelmente comprometeu a produtividade total dos fatores e a competitividade da agropecuária.

A partir da década de 1990, a adoção de mecanismos modernos de política agrícola fez com que, mesmo com a redução dos gastos públicos na agricultura, a iniciativa privada se

sentisse incentivada a conceder crédito ao setor, o qual pode em parte explicar o aumento na taxa média de crescimento da produtividade total dos fatores agrícola, principalmente a partir de 2001.

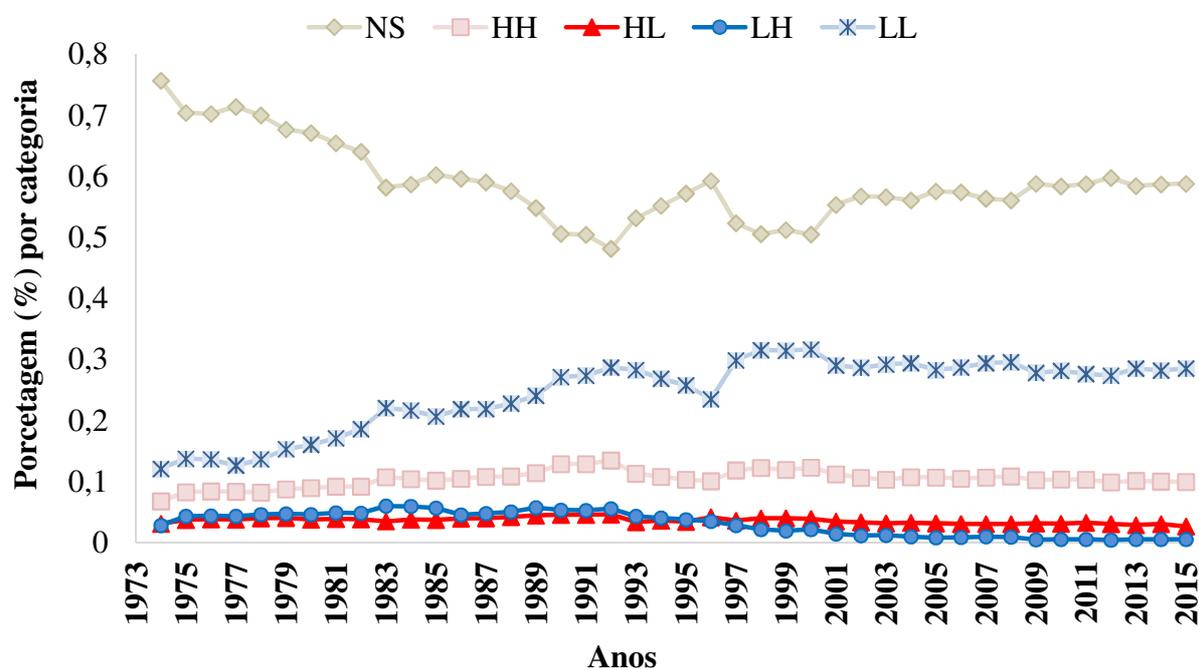


Figura 4. Porcentagem de municípios da região Nordeste do Brasil por categoria de *cluster* a que pertence, segundo o tamanho do rebanho bovino de 1974 a 2015.

Nessa mesma perspectiva, Rocha *et al.* (2013) destacam que a agricultura brasileira apresentou importantes transformações em sua estrutura, nesse processo. O governo vem sustentando esse setor por meio de política direcionada ao desenvolvimento tecnológico, e os resultados mostraram que, nas regiões menos desenvolvidas, os investimentos em ciência, tecnologia e inovação precisam ser complementados com investimentos em infraestruturas. Conforme Gomes *et al.* (2009), para a economia nordestina, a agropecuária possui uma representação expressiva; da forma como se deu o processo de mudanças tecnológicas no Brasil, torna-se necessário analisar as características regionais da estrutura produtiva agrícola. Nesse sentido, destacam ainda que a agropecuária nordestina apresentou ganhos na produtividade total de fatores no período de 1996 a 2006.

Ainda, na Figura 4, tais resultados corroboram com os autores, uma vez que observamos o fenômeno se comporta de forma inconstante nas décadas iniciais 1973 até início da década 2000, ou seja, existiu uma movimentação entre os grupos. Aquelas porções que estavam com criação de bovinos na média (NS) ao longo do tempo se tornaram de baixa criação (LL), pois ocorreu um aumento de *clusters* de baixa concentração (Figura 4).

Assim, através dessa análise pode-se dizer que talvez a criação de bovinos em determinadas regiões não tenha dado certo. E aqueles municípios que pertenciam a alta criação (HH) se mantiveram constantes ao longo dos anos. Também pode-se verificar que ocorreu uma retração a partir do ano 1997 da criação de *clusters* do tipo LH, ou seja, os municípios que eram de baixa criação e estavam cercados de municípios com alta criação, diminuíram ao longo do tempo e já os *clusters* do tipo HL se mantiveram constantes ao longo dos anos.

Foi realizada também a análise dos mapas de autocorrelação espacial local de Moran I (Figura 5), a cada 10 anos, no período de 1975, 1985, 1995, 2005 e 2015, em que se podem observar as concentrações espaciais (*clusters*) do rebanho bovino dos municípios da região Nordeste do Brasil.

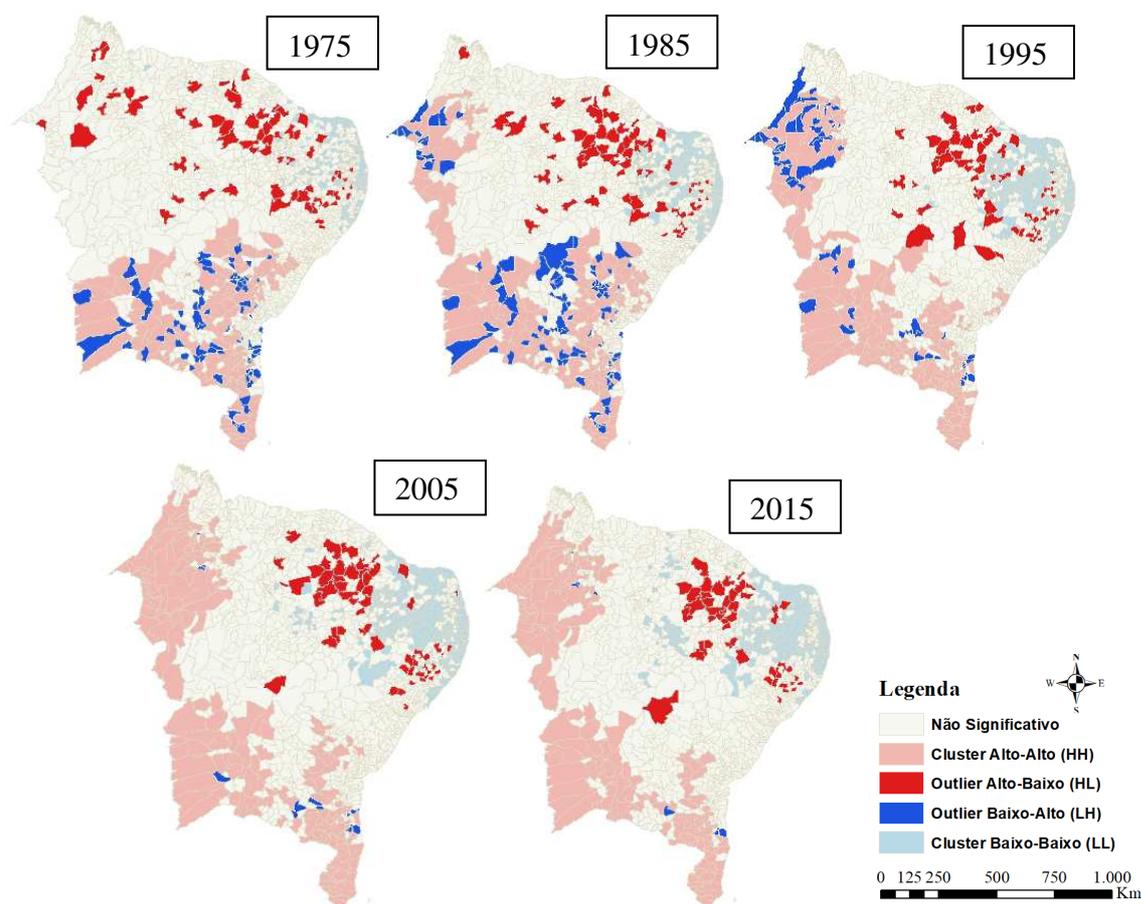


Figura 5. *Clusters* segundo o Índice Local de Moran I para o tamanho do rebanho bovino no Nordeste do Brasil a cada 10 anos (1975 a 2015).

No geral, observou-se que os agrupamentos que apresentaram valores baixos e vizinhos semelhantes (LL) concentram-se na porção nordeste e leste da região, ou seja, mais ao litoral

e ao centro nordestino, que correspondem aos estados do Rio Grande do Norte, Paraíba, Pernambuco e Alagoas. Também foi verificado que os *clusters* que apresentaram valores altos e vizinhos semelhantes (HH) e que se mantiveram constantes ao longo das décadas analisadas, se concentraram na porção sul da região, que corresponde ao estado da Bahia. Além disso, a partir da década de 1985, nota-se que ocorreu um aumento da concentração na porção do noroeste e oeste que correspondem a municípios dos estados do Maranhão e Piauí.

Esta região é considerada como MATOPIBA, acrônimo formado pelas primeiras letras dos estados do Maranhão, Tocantins, Piauí e Bahia. Dalla Nora *et al.* (2014) descrevem que no ano de 1995 ocorreu um cumprimento da legislação ambiental e acordos do setor privado que promoveram a diminuição do desmatamento no bioma amazônico, contribuindo para o desenvolvimento desta região agrícola no Cerrado, que atualmente representa uma das fronteiras agrícolas mais ativas no país, principalmente à custa de áreas de vegetação pristina (SPERA *et al.*, 2016). Logo, se percebe que de fato tal região houve uma evolução do rebanho bovino, devendo esta associada a tal desenvolvimento de políticas públicas nesta área.

Além, do que, observa-se, um crescimento e estabilidade de *clusters* do tipo HH na região interiorana que se inseri no estado da Bahia. De acordo com Cruz (2011) a partir de 1995, intensificou-se na Bahia uma política traduzida em esforços do Governo do Estado buscando acelerar o seu crescimento, que só no fim da década de 90 e início dos anos 2000, ocorreu o avanço significativo no setor pecuário. Pois o segmento pecuário baiano engajado nas diretrizes de plano de governo perseguiu a meta de desenvolvimento com geração de emprego e renda, o que alcançou desempenho traduzido pelos índices de crescimento do rebanho, aumento da oferta de carne, leite e derivados, aliados a diversificação de produtos e mais a atração de investimentos em diversos setores da pecuária. Assume ainda, uma posição de liderança regional na modernização de sua cadeia de produção de carne bovina, e detém o “Certificado de Zona Livre da Febre Aftosa”. Com isto, garante o seu acesso ao mercado internacional. Mas foi apenas a partir dos anos 80, que a agropecuária baiana melhorou seu perfil competitivo. Em 2008 a Bahia era responsável por 43% da produção de carne bovina do Nordeste Brasileiro e responde por 25% da produção nacional, com taxas de abate de 44,2% abaixo 1% da média nacional. Estas vantagens acabam por lhe colocar à frente dos demais estados do Norte e Nordeste, em termos de acessibilidade aos grandes mercados internos e externos (CRUZ, 2011)

O estudo acima mencionado corrobora com Neiva *et al.* (2003) que observou que na região Nordeste ocorreu um ligeiro aumento do efetivo bovino, tendo aumentado de 24,0 para 25,5 milhões de cabeças no período de 1994 a 2003. O aumento de 1,5 milhão de cabeças se

deu basicamente em dois Estados: Maranhão e Bahia. Nota-se que o crescimento se deu nos dois Estados com grandes áreas territoriais fora da região semiárida. O Maranhão que tem expandido a pecuária de corte em sua faixa amazônica e a Bahia na faixa litorânea e próximo ao centro-oeste. De maneira geral, os Estados que apresentam grande percentual do território no semiárido mantiveram seus rebanhos. É importante destacar que em relação à aptidão do rebanho no Nordeste brasileiro, alguns Estados apresentam o efetivo do rebanho bovino leiteiro superior ao de corte. Dois exemplos típicos são os Estados do Ceará e Pernambuco; logo, há necessidade de se avaliar a possibilidade do aproveitamento dos machos leiteiros para a pecuária de corte.

Em relação a porção ao interior no sentido norte (referente a municípios dos estados do Piauí e Ceará) apresenta exceções, valores altos com vizinhos que possuem valores altos do efetivo do rebanho (HL), porém, não se encontram agrupados, pois sua vizinhança apresenta valores baixos de rebanho bovino. A porção sul no ano de 1975 apresenta exceção LH, valores baixos com vizinhos que possuem alta concentração do rebanho bovino, tais pontos são referentes a municípios localizados no estado da Bahia. Já no ano de 1985 e 1995 apresentou este comportamento na porção noroeste (referente a municípios do estado do Maranhão), sendo que nos anos de 2005 e 2015 só apresentaram alguns municípios do tipo LH na porção sul (municípios do estado da Bahia). Os municípios considerados não significativos, não se enquadram nos agrupamentos, pois apresentam o efetivo do rebanho bovino próximo a média, assim como os valores dos municípios vizinhos (Figura 5).

Em resumo, conclui-se que na região Nordeste houve poucos grupos do tipo LH e HL ao longo do tempo; percebe-se que ocorreu uma movimentação de grupos do tipo HH na direção do noroeste da região (referente a municípios do estado do Maranhão), e que se permaneceram mais constantes ao longo dos anos na direção sul (referente a municípios do estado da Bahia). Já os *clusters* do tipo LL, ocorreram uma expansão para direção do interior do nordeste e leste (referente a municípios dos estados do Rio Grande do Norte, Paraíba, Pernambuco e Alagoas), se caracterizando por baixa criação de bovinos ao longo do tempo. Já nas áreas dos estados do Rio Grande do Norte, Paraíba, Pernambuco e Alagoas, podemos justificar a maior concentração dos *clusters* de baixa (LL) ao fato de essas regiões estarem localizadas no litoral; desta forma, possuem outras variedades de insumos e produções como, por exemplo, a cana de açúcar que é cultivada em grande parte em toda estas regiões.

Na Figura 6, é possível analisar o comportamento da variável explicativa (X) taxa de crescimento em relação a probabilidade de mudança de um determinado tipo de *clusters* se modificar para outro tipo de *clusters* ao longo dos anos em estudo. Observou-se que o efetivo

do rebanho bovino teve valor máximo no ano de 1974 para 1975, atingindo um crescimento do efetivo do rebanho de 12,63%. O mínimo foi observado entre os anos de 1992 e 1993, em que representou uma taxa de retração em torno de -16,37% do efetivo do rebanho bovino na região Nordeste do Brasil. A média da taxa de crescimento ao longo destes anos estudados está entre 1,53% com desvio padrão de 4,45%, demonstrando uma alta variabilidade desta taxa de mudança de *clusters*.

Isso pode ser justificado pelo fato de que, entre 1993 a 1996 houve um longo período de estiagem que atingiu duramente todos os estados do Nordeste e dizimou parte considerável de seus rebanhos, além de que também ocorreu alto endividamento do setor devido a questão do preço médio e do crédito, pela a falta de política para o desempenho desta cadeia (CRUZ, 2011). Assim como entre 1982 para 1983, de 1998 para 1999 e de 2012 para 2013, foram períodos de estiagem prolongada no Nordeste brasileiro.

Neste contexto, foi aplicado o modelo de regressão beta, sendo necessário realizar uma transformação da variável explicativa (X) taxa de crescimento da população bovina ao longo dos anos (Figura 6).

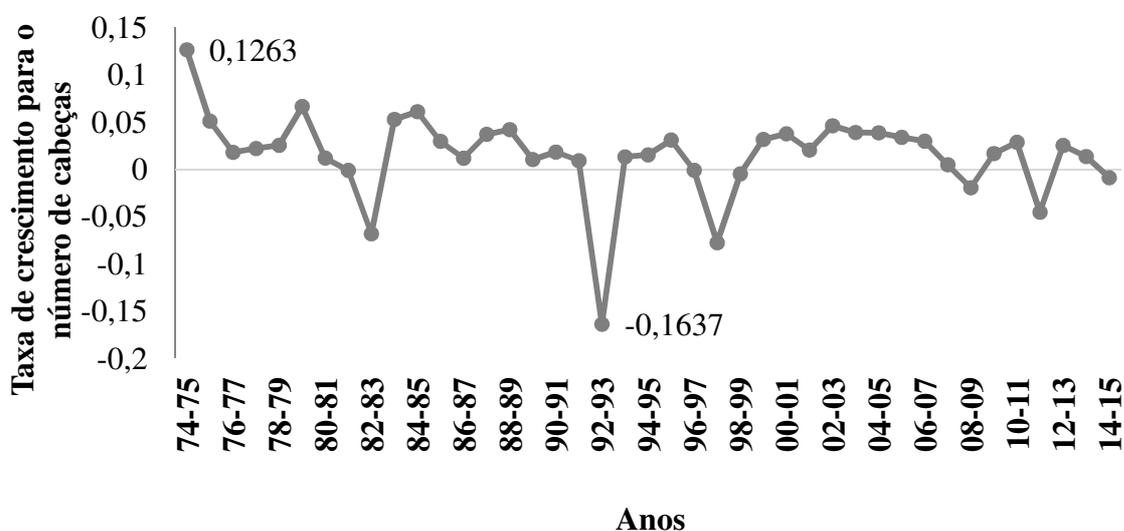


Figura 6. Taxa de crescimento populacional da diferença de um ano para outro do rebanho bovino no Nordeste do Brasil (1974 a 2015).

Para tanto, foi somado 0,17 aos dados para que tais estivessem entre 0 e 1, uma vez que, o modelo beta faz tal pressuposição. Após as análises foi retirado o acréscimo, e as covariáveis utilizadas no modelo foram as probabilidades de mudanças e ou permanências dos tipos de *clusters* espacial (Figura 6).

A Tabela 3 apresenta os coeficientes estimados para os modelos Betas baseados nas frequências observadas para mudança de categoria do município dada o *cluster* a que pertence. Para os modelos de regressão foram aplicadas as ligações Probit, Logit e Complemento log-log. Todos os modelos apresentaram boas estimativas para os tipos de *clusters*; porém, através do AIC, foi preferível a utilização dos modelos com ligação Logit. Das covariáveis selecionadas para os modelos, as do tipo NS-HL, NS-LH, HH-HH, HL-HL, LH-HH, LH-LH, LL-HL e LL-LL, não foram significativas, pois apresentaram p-valor acima de 5%. É possível verificar que as covariáveis do tipo de *clusters* NS-HH; NS-LL; NS-NS; HH-NS; HL-NS; LH-NS; LL-NS e o geral com todas as mudanças, obtiveram uma influência negativa em relação a taxa de crescimento de alteração de *cluster* (X). Isto é, ocorreu uma diminuição (retração) do rebanho bovino na região. Em contrapartida, a covariável do tipo de *cluster* geral com todas as permanências obteve uma influência positiva (expansão) na taxa de crescimento de *clusters*.

Segundo Teixeira e Hespanhol (2014), desde o início do processo de colonização do território brasileiro a atividade pecuária desempenhou papel importante na estrutura produtiva. Inicialmente, foi primordial no abastecimento dos núcleos urbanos e, posteriormente, expandiu-se em direção ao sertão nordestino, onde o rebanho bovino passou a ser criado solto em pastagens naturais.

Dentre os fatores que levaram a introdução e expansão do rebanho bovino no Nordeste, pode se destacar o relevo sem barreiras, facilitando os deslocamentos do gado; abundância de pastagens naturais; depósitos de sal-gema, importantes para a alimentação do gado; disponibilidade de água do rio São Francisco; exigência de reduzidos investimentos para a composição e custeio dos rebanhos e; mercado consumidor garantido tanto para o couro como para a carne, representado pelos engenhos (ADAS, 1983).

Na Figura 7, tem-se os gráficos dos valores observados versus os preditos (ajustados) do modelo de regressão Beta com ligação logit. Para o modelo da variável permanência dos *clusters*, observa-se que no geral, ocorreu um bom acompanhamento do ajuste aos dados observados, excetuando em alguns anos que não capturaram bem o comportamento, como nos anos iniciais da análise (1974 e 1975), de 1985 para 1986 e de 1995 para 1997. Já para o modelo da variável mudança de *clusters* foi mais sensível para as pequenas alterações no ajuste do modelo, a cada dois anos o modelo não capturou bem as variações; porém, no geral, o modelo acompanhou os dados observados aos ajustados.

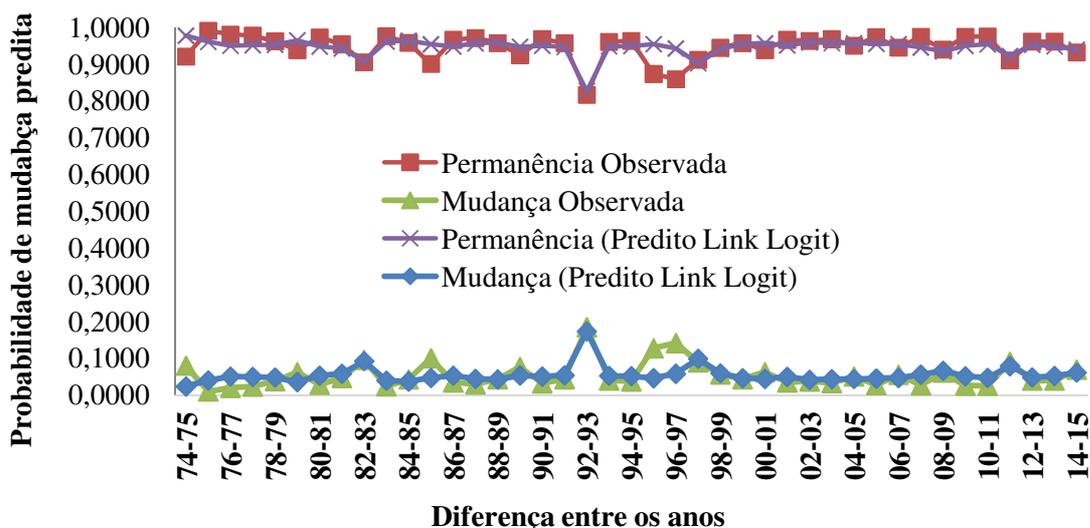


Figura 7. Probabilidade de mudança e permanência das categorias de *clusters* para ligação logit.

Após tais análises e visto que os modelos são adequados (análise dos resíduos no apêndice) e assim podem ser utilizados como estimadores desejou-se realizar projeções (Figura 8) para verificar como se comportariam as informações chegando aos seus limites nas taxas e nas probabilidades. Usamos a previsão das probabilidades da taxa de crescimento dos tipos de *clusters* espacial das mudanças e permanências do rebanho bovino da região Nordeste do Brasil, de 1974 a 2015 (Figura 8).

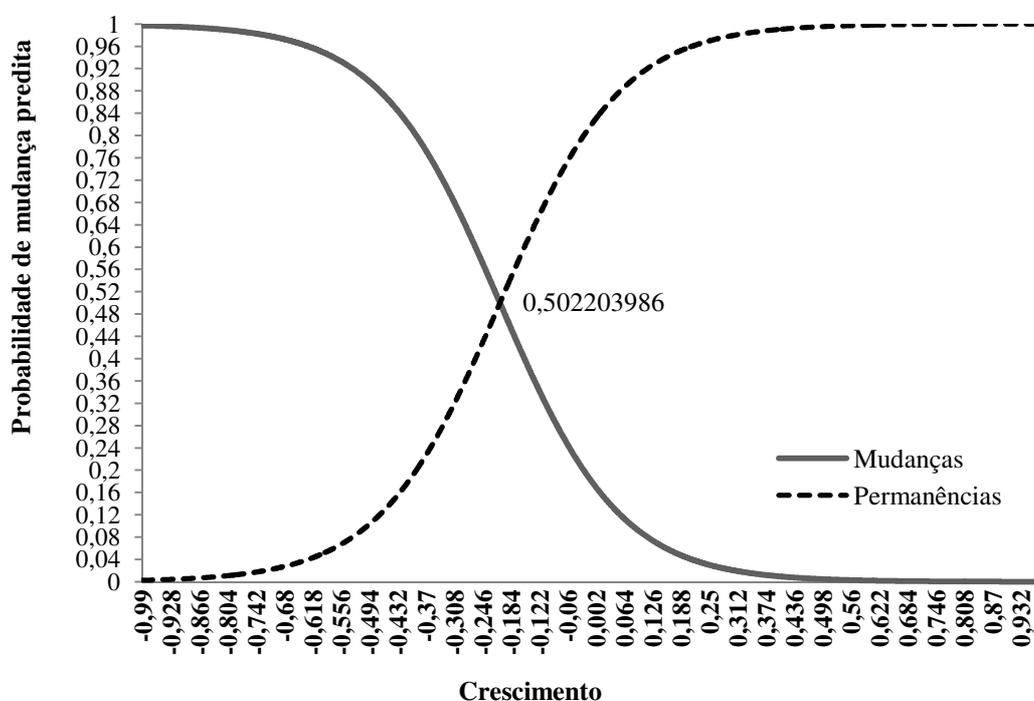


Figura 8. Probabilidade prevista das mudanças e permanências das categorias de *clusters* espaciais do rebanho bovino da região Nordeste do Brasil de 1974 a 2015, conforme a taxa de crescimento do rebanho (modelo com ligação logística).

Tabela 3. Frequências observadas para mudança de categoria do município dada o tipo de *cluster* a que pertence (1974-2015).

Tipo de <i>Cluster</i> Espacial (Y)	PROBIT				LOGIT				CLOGLOG			
	B ₀	B ₁	φ	AIC	B ₀	B ₁	φ	AIC	B ₀	B ₁	φ	AIC
PERMANÊNCIA	0,9806***	3,6253***	76,36***	-185,8	1,567***	7,564***	77,76***	-186,6	0,6283***	2,5441***	74,82***	-185,0
MUDANÇA	-0,9806***	-3,6253***	76,36***	-185,8	-1,567***	-7,564***	77,76***	-186,6	-1,6485***	-7,2625***	77,98***	-186,7
NS-HH	-2,1817***	-1,9551**	361,17***	-346,0	-4,1781***	-5,5408***	363,09***	-395,4	-4,1847***	-5,5199***	363,13***	-395,4
NS-HL	-2,6001***	-0,4614	890,40***	-395,4	-5,3516***	-1,4391	891,00***	-395,4	-5,3537***	-1,4375	891,00***	-395,4
NS-LH	-2,5312***	-0,9421	185,28***	-382,9	-5,1118***	-3,0785***	185,79***	-382,9	-5,1140***	-3,078	185,79***	-382,9
NS-LL	-1,6546***	-2,8977***	121,69***	-265,4	-2,9429***	-7,0078***	121,50***	-265,5	-2,9680***	-6,9102***	121,40***	-265,5
NS-NS	-0,1900*	1,8997***	73,59***	-112,1	-0,3035*	3,0358***	73,55***	-112,1	-0,5822***	2,1374***	73,78***	-112,2
HH-HH	-1,2333***	-0,3484	629,70***	-241,9	-2,1049***	-0,6697	629,20***	-241,9	-2,1634***	-0,6326	629,00***	-241,9
HH-NS	-1,9636***	-3,3664***	532,10***	-361,5	-3,5887***	-9,4476***	547,30***	-362,6	-3,5984***	-9,4102***	547,90***	-362,7
HL-HL	-1,8939***	0,20425	1310,8***	-315,5	-3,5058***	0,4710	1310,6***	-315,5	-3,5205***	0,4632	1310,6***	-315,5
HL-NS	-2,1750***	-2,9180***	434,4***	-380,1	-4,1427***	-8,487***	437,9***	-380,5	-4,1491***	-8,4619***	438***	-380,5
LH-HH	-1,8901***	-0,8935	62,14***	-279,5	-3,4829***	-2,2652	62,16***	-279,5	-3,4971	-2,2456	62,16***	-279,5
LH-LH	-2,0776***	0,7887	53,53***	-213,1	-3,9650***	1,995	53,58***	-213,1	-3,9758***	1,9789	53,58***	-213,1
LH-NS	-2,1127***	-3,5047***	218,05***	-394,8	-3,9318***	-10,452***	222,30***	-395,3	-3,9384***	-10,426***	222,33***	-395,4
LL-HL	-2,4969***	-0,0953	271,88***	-341,2	-5,0680***	-0,269	271,87***	-341,2	-5,0711***	-0,2681	271,87***	-341,2
LL-LL	-0,5423***	-1,0896	55,26***	-114,7	-0,8805***	-1,8183	55,17***	-114,7	-1,0620***	-1,5534	55,08***	-114,6
LL-NS	-1,9156***	-2,0364*	114,37***	-281,6	-3,4938***	-5,4966*	115,17***	-281,8	-3,5043***	-5,4712*	115,21***	-281,8

* P <0,05, ** P <0,01 e *** P <0,001.

Percebe-se que a curva da permanência de *clusters* indica um crescimento da taxa e já a de mudança aponta uma retração da taxa. Com 99,93% de probabilidade tem-se 77,3% de oportunidades do *cluster* permanecerem no mesmo local, ou seja, é provável de não modificar o efetivo do rebanho bovino quando ele se encontra estabelecido na região. Porém, com probabilidade de 0,1436% tem-se 65,8% de oportunidades do *cluster* mudarem da mesma região. Ou seja, uma parcela muito pequena pode tender a modificar seu tipo de *clusters* quanto ao efetivo do rebanho bovino na região Nordeste do Brasil.

Também foram analisadas as previsões para os modelos dos tipos de *clusters* (Figura 9). Observou-se que todos os modelos obtiveram decréscimo do efetivo do rebanho bovino, destacando-se o modelo para variável resposta NS-HH. Ou seja, *clusters* do tipo não significativos (NS) que foram para *clusters* do tipo Alto-Alto (HH), em que se obteve inicialmente 1,43% de taxa de decréscimo com 1% de probabilidade de ocorrência, chegaram a uma taxa de decréscimo de 9% com de probabilidade de 0,92% (Figura 9 A).

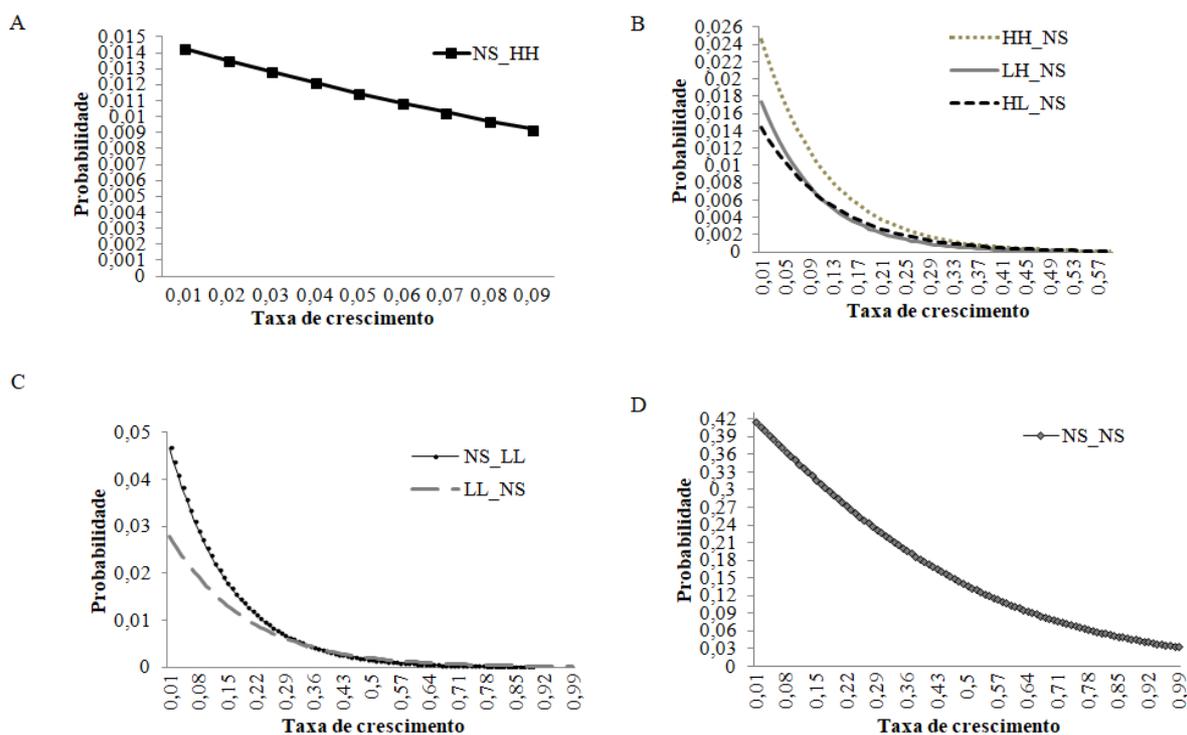


Figura 9. Probabilidade predita da taxa de crescimento das categorias significativas de *clusters* espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015), conforme a taxa de crescimento do rebanho (modelo com ligação logística).

Os *clusters* do tipo HH-NS, ou seja, *clusters* do tipo Alto-Alto (HH) que foram para *clusters* do tipo não significativos (NS), obtiveram inicialmente 1% de taxa de decréscimo com 2,45% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 59% com probabilidade de o evento ocorrer de 0,01 %. Os *clusters* do tipo LH-NS, ou seja, *clusters* do

tipo Baixo-Alto (LH) que foram para *clusters* do tipo não significativos (NS), obtiveram inicialmente 1% de taxa de decréscimo com 1,74% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 50% com probabilidade de o evento ocorrer de 0,01 %. E *clusters* do tipo HL-NS, ou seja, *clusters* do tipo Alto-Baixo (HL) que foram para *clusters* do tipo não significativos (NS), obtiveram inicialmente 1% de taxa de decréscimo com 1,44% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 59% com chances de probabilidade de o evento ocorrer de 0,01 % (Figura 9 B). Nos *clusters* do tipo NS-LL, ou seja, *clusters* do tipo não significativos (NS) que foram para *clusters* do tipo Baixo-Baixo (LL), obtiveram inicialmente 1% de taxa de decréscimo com 4,68% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 89% com probabilidade de o evento ocorrer de 0,01 %. Já *clusters* do tipo LL-NS, ou seja, *clusters* do tipo Baixo-Baixo (LL) que foram para *clusters* do tipo não significativos (NS), obtiveram inicialmente 1% de taxa de decréscimo com 2,80% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 98% com probabilidade de o evento ocorrer de 0,01 % (Figura 9 C). Já para os *clusters* do tipo NS-NS, ou seja, os *clusters* não significativos (NS) que permaneceram na mesma média do efetivo de bovino na região em estudo, obtiveram inicialmente 2% de taxa de decréscimo com 40,99% de probabilidade de ocorrência, chegando a uma taxa de decréscimo de 99% com probabilidade de o evento ocorrer de 3,53 % (Figura 9 D).

Com o propósito de verificar qual comportamento futuro dos tipos de *clusters*, foi realizada uma projeção para os modelos de permanência e mudança para 10 anos (Figura 10).

Observou-se que para o modelo da permanência ocorre uma tendência da taxa de crescimento que pode variar entre 90,09% e 99,11% em torno de 94,60% ao nível de 95% de significância, ou seja, uma convergência de os *clusters* permanecerem nas regiões aonde já se encontrava.

Já para a taxa de mudança, foi visto que pode variar de -1,52% a 12,30% sendo uma tendência em torno de 5,39%, também ao nível de 95% de significância, logo, espera-se que os *clusters* possam mudar de categoria ao longo dos anos, porém essa taxa tem possibilidades de decrescer. Com tais análises pode-se demonstrar que houve uma tendência a estabilização e especialização das cadeias produtivas de bovinos nos municípios da região Nordeste do Brasil.

Também foi avaliado a estacionariedade da matriz observada geral da frequência de mudança entre os tipos de *clusters* dada a sua origem e destino ao passar dos anos. Verificou-se que todos os tipos de *clusters* estabelecem probabilidades mesmo que baixas de mudarem para o tipo NS, sendo que a probabilidade dos *clusters* permanecerem no mesmo tipo fica em

torno dos 95%, ou seja, o rebanho bovino na região Nordeste possuem grandes chances de não mudarem da região que se encontram produções do rebanho bovino (Figura 11 A).

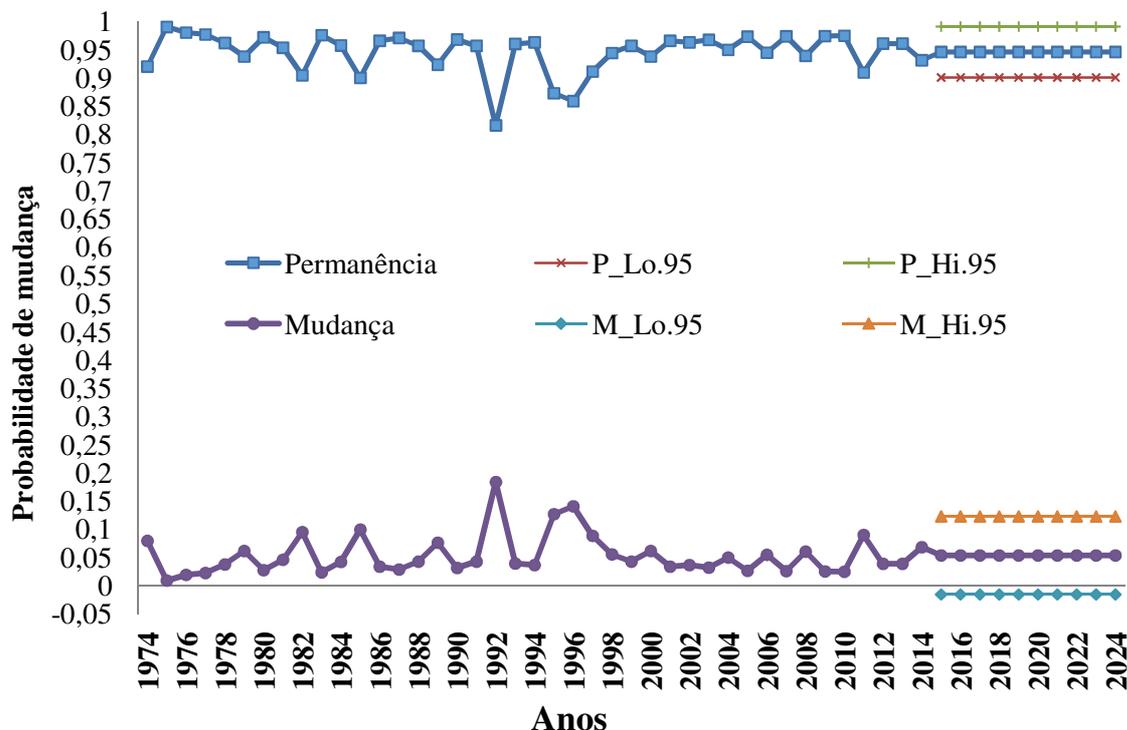


Figura 10. Projeção do modelo de permanência e mudança das categorias de *clusters* espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015), conforme a probabilidade de mudança.

Outro resultado relevante é que o *cluster* do tipo LL, ou seja, de baixa produção, não se modifica para *cluster* do tipo LH, nem HL, que são *clusters* de características de baixa produções ou alta, sendo que as regiões vizinhas que ocorram alta ou baixa produção respectivamente. Assim como HH, referente a altas produções, não migram para HL nem LH, que possam ter regiões vizinhas que ocorram alta ou baixa produção, respectivamente.

Através da estacionariedade da matriz foi verificado que para que a região do Nordeste se obtenha estacionariedade são necessários 308 anos a partir de 2015. Sendo assim foi verificado que no ano de 2323 as probabilidades do estado estacionários dos tipos de *clusters* por ano devido a movimentação do rebanho bovino na região Nordeste do Brasil tendem a possuir maior concentração de *clusters* do tipo NS, ou seja, a tendência de que nesta região se concentre o rebanho bovino na média, seguido de regiões com baixa produções e com vizinhos também de baixa produções. Além de que poucas regiões tendem a ter altas produções com vizinhos de alta concentração do rebanho bovino (Figura 11 B).

E, por fim, com intuito de avaliar a probabilidade de permanência de categoria de *cluster*, foi realizada a partir do município que mais apresentou modificações de *clusters*. Tal

análise foi efetivada para o municípios de Serra do Ramalho (BA) que foi o município a obter a maior quantidade de mudanças sendo esta de 16 vezes, em que a moda para mudanças foi de 4 vezes.

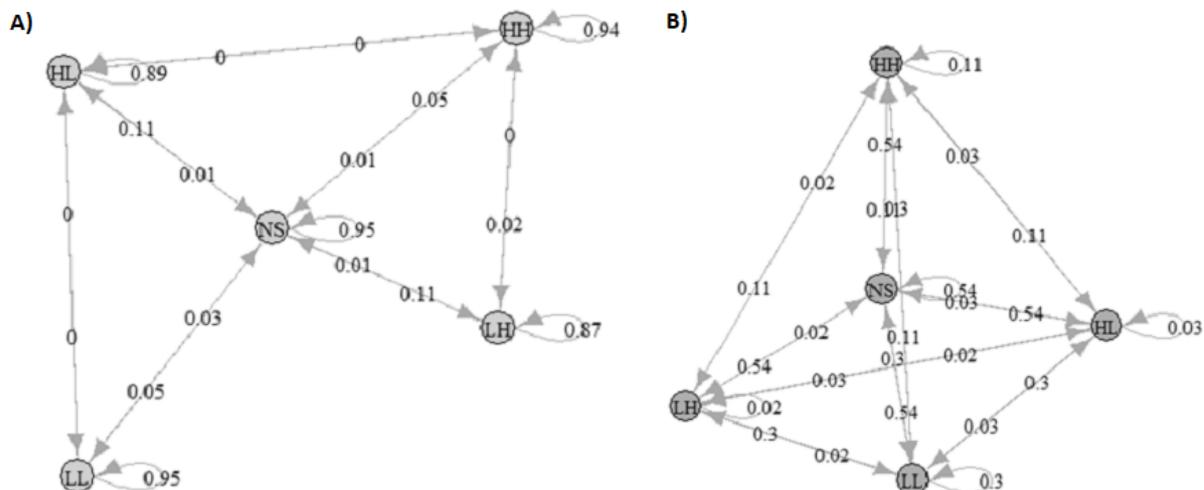


Figura 11. Diagrama da transição para a matriz a partir dos valores observados (A) e o estado estacionário (B) de probabilidade para a classificação dos tipos de *clusters* (estágio atingido com 308 anos)

Com intuito de analisar a probabilidade de permanência de categoria de *cluster* para o município de Serra do Ramalho (BA), foi avaliada a estimativa da função de sobrevivência por Kaplan-Meier com ajuste do modelo Weibull (Figura 12), sendo o primeiro modelo sem covariáveis (Figura 12 A) e o segundo com a covariável taxa de crescimento (Figura 12 B) para o rebanho bovino. Observa-se que ocorreu transição entre as categorias HH, HL, LH e NS. Para o município de Serra do Ramalho (BA) o *cluster* que teve mais ocorrências foi o do tipo NS, ou seja, o município esteve na maioria das vezes com a produção do rebanho bovino na média, porém conseguiu atingir produções altas (HH), com oscilações em produções baixas ao redor de altas (LH) concentrações, tendendo nos anos finais da análise para produções altas ao redor de municípios com baixa produção do rebanho bovino (HL).

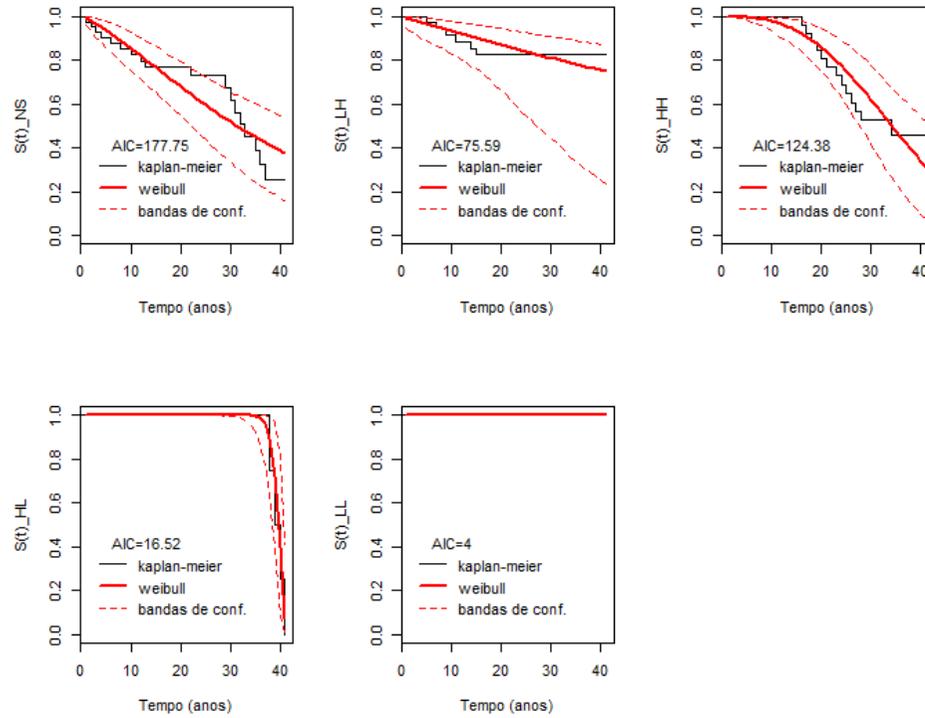
Nota-se ainda, o modelo para o *cluster* dos tipos NS e HH ajustou-se melhor para os modelos com covariável, e já os *clusters* dos tipos LH, HL e LL o melhor ajuste foi para modelo sem covariável. Ou seja, a taxa de crescimento pode influenciar o efetivo do rebanho bovino para se manter com produções do rebanho bovino na média e em regiões que possuam altas produções (Figura 12).

De acordo com Santos *et al.* (2011), a adaptação dos sistemas de produção existentes na região Nordeste depende ainda de informações mais detalhadas sobre: a necessidade de adequação do manejo de pastagens cultivadas aos cenários climáticos futuros; os efeitos das

mudanças climáticas sobre as pastagens nativas; os efeitos das interações entre as mudanças climáticas e outros fatores de estresse como, por exemplo, a salinidade do solo sobre o desenvolvimento das plantas forrageiras que são utilizadas de alimento para bovinocultura é de suma importância para tal crescimento. Estes e outros estudos necessários para garantir a sustentabilidade da atividade pecuária na região devem ser priorizados por instituições de pesquisa e universidades que atuam no Semiárido brasileiro.

Com isso, este trabalho buscou contribuir com uma pesquisa de prospecção em que socioeconomistas e especialistas em pecuária de corte poderão analisar e prever a direção da atividade econômica do efetivo bovino na região Nordeste do Brasil e, desta forma, auxiliar na tomada de decisões e estratégias deste seguimento nesta área.

A)



B)

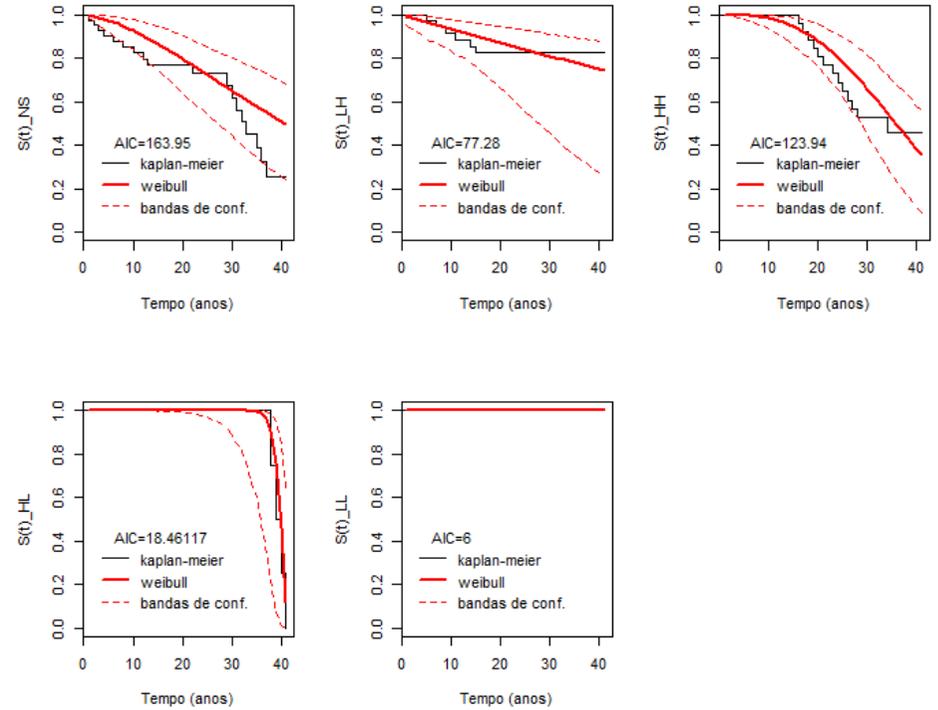


Figura 12. Probabilidade de permanência de categoria de *cluster* para o município de Serra do Ramalho (BA), sendo o primeiro modelo sem covariáveis (A) e o segundo com a covariável taxa de crescimento (B) para o rebanho bovino.

3.4. Conclusões

Os resultados revelaram que existem indícios de aglomeração do rebanho bovino para os municípios da região Nordeste, sendo mais predominante a existência de agrupamentos com baixo rebanho efetivo de bovino circundado também por baixos efetivos.

Entretanto, observou-se constância na proporção de rebanhos com alta intensidade ao longo do tempo e também se verifica mudança de agrupamento de baixo efetivo, que foram aqueles que estavam na média e foram para a baixa produção.

Foi possível verificar espacialmente que, na porção noroeste, oeste e sudoeste, que se localiza na região que compreende a MATOPIBA, ocorreu uma evolução dos *clusters* do tipo HH, ou seja, uma maior criação de rebanho bovino. Também teve avanço de *clusters* do tipo LL na porção do litoral noroeste, com um estreitamento dos *clusters* e locais com valores próximos a média na porção central ao norte e sudoeste, ou seja, as porções que estão acima da média do efetivo do rebanho bovino. Ou seja, de fato ocorreu uma movimentação do rebanho bovino nesta região, sendo que ao final dos anos observou-se que ocorreram concentrações maiores no interior e as menores concentrações se localizaram ao litoral nordeste e leste.

Constatou-se também que ocorreu uma probabilidade baixa nas mudanças de categorias dos tipos de *clusters*, e que ocorreu uma retração do rebanho bovino com o passar dos anos na região Nordeste do Brasil.

E por fim, de acordo com a matriz de transições observou-se uma tendência de que no ano de 2323, na região Nordeste do Brasil a concentração do rebanho bovino será na média, seguido de regiões com baixas produções e com vizinhos também de baixas produções. Além de que poucas regiões tenderão a ter altas produções com vizinhos de alta concentração do rebanho bovino.

Capítulo 4

Estabilidade de *clusters* espacial pecuário bovino no Nordeste do Brasil (Proposta para o 2º Artigo)

Resumo

Este capítulo centra-se na simulação através da regressão beta bayesiana, aplicada em análise espacial da probabilidade de mudança ou permanência de *clusters* do rebanho bovino na região Nordeste do Brasil, para o período de 1974 a 2015. Investigou-se a existência da autocorrelação espacial para identificação das regiões que possuíam agrupamentos locais com a utilização do Índice Local de Moran I. Foram calculadas as probabilidades de mudança ou permanência dos agrupamentos para serem utilizadas simulações destes parâmetros através do modelo de regressão beta bayesiano. O diagnóstico da convergência dos modelos foi baseado nos critérios de BIC, Geweke, Rafters, Heidel-berger e Welch. Também foi estudado a estacionariedade das matrizes para os modelos beta, beta bayesiano e dados observados, e modelos do tempo de sobrevivência para os *clusters* do município de Serra do Ramalho (BA).

Palavras-chaves: Bayes, rebanho, regressão beta, simulação.

Abstract

This chapter focuses on the simulation through Bayesian beta regression, applied in spatial analysis of the probability of change or permanence of clusters of the bovine herd in the Northeast region of Brazil, for the period from 1974 to 2015. The existence of spatial autocorrelation for the identification of the regions that had local groupings using the Moran I Local Index. The probabilities of change or permanence of the groupings were calculated to be used simulations of these parameters through the Bayesian beta regression model. The diagnosis of the convergence of the models was based on the criteria of BIC, Geweke, Rafters, Heidel-berger and Welch. We also studied the stationarity of the matrices for the beta, bayesian beta and observed models, and models of survival time for the clusters of the city of Serra do Ramalho (BA).

Keywords: Bayes, beta regression, herd, simulation.

4.1. Introdução

A expansão pecuária tem sido um tema que, gradualmente, vai se incorporando à agenda de pesquisa, por vezes procurando a relação entre desenvolvimento da atividade agrícola e da pecuária (BRANDÃO *et al.* 2005), devido ao processo de expansão de fronteiras na região Amazônica que tem provocado elevados índices de desflorestamento. Um dos principais fatores apontados tem sido a ocupação pelo gado (FERRAZ 2001, MUCHAGATA e BROWN, 2003).

Um estudo recente realizado pela Embrapa Gestão Territorial de São Paulo observou que o crescimento do rebanho bovino nacional tem se deslocado em direção oposta à da região amazônica por meio de dados obtidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A pesquisa registra que, a partir de 1974, a atividade pecuária se deslocava pelo País em direção a fronteira agrícola do norte, região que envolve a MATOPIBA (Maranhão, Tocantins, Piauí e Bahia). Esse movimento se manteve até 2006 quando tanto a pecuária de corte como a leiteira iniciou movimento contrário em direção ao sul do país. Uma hipótese que justificou atividade deixar de crescer em direção ao norte poderia ser a intensificação das ações de fiscalização na fronteira agrícola nesta região, desestimulando a abertura de novas áreas por lá para a bovinocultura (FREITAS, 2016).

Diversos estudos apontam para o crescimento da produtividade agropecuária ao longo do tempo (GASQUES *et al.*, 2012; FORNAZIER e VIEIRA FILHO, 2013; VIEIRA FILHO, GASQUES e SOUSA, 2012). Frente ao cenário de crescimento, é preciso compreender os padrões regionais, no intuito de assessorar o planejamento de políticas públicas de desenvolvimento regional, inclusive em regiões tradicionalmente desfavorecidas, como na região Nordeste do Brasil.

A movimentação da distribuição do rebanho ao longo do tempo e espaço é de essencial importância para a compreensão do fenômeno populacional em seus vários aspectos, tais como implantação, expansão, retração, concentração, apropriação de espaços produtivo-exploratórios, implicações econômicas sociais e formação de *clusters* produtivos.

Logo, este trabalho tem por objetivo analisar o comportamento da estabilidade de *clusters* espaciais da probabilidade de mudança ou permanência do rebanho bovino nos municípios da região Nordeste do Brasil no período de 1974 a 2015.

4.2. Material e Métodos

O estudo abrange a região do Nordeste do Brasil (1,554 milhões km²), equivalente a 18% do território nacional, na qual se inserem nove estados (Bahia/BA, Sergipe/SE, Alagoas/AL, Pernambuco/PE, Paraíba/PB, Rio Grande do Norte/RN, Ceará/CE, Piauí/PI, Maranhão/MA), em 1.794 municípios e habitados por 56.560.081 milhões de pessoas até 2015 (IBGE, 2015).

Os dados foram obtidos junto ao Sistema IBGE de Recuperação Automática (SIDRA) do Instituto Brasileiro de Geografia e Estatística (IBGE), representando o efetivo do rebanho bovino por município (1.794) para os anos de 1974 a 2015, totalizando 73.554 observações.

A avaliação da autocorrelação espacial local do efetivo do rebanho bovino foi utilizada o índice local de Moran I, uma vez que, a autocorrelação espacial mede a relação entre observações com proximidade espacial, considerando que observações próximas espacialmente possuam valores parecidos. Os padrões de distribuição dos indicadores foram examinados em menor escala por meio do Moran I local (LISA), produzindo um valor específico para cada município, permitindo a visualização de agrupamentos de municípios com valores similares para os indicadores selecionados (ANSELIN, 1995). No caso local se tem a tipologia dos padrões da correlação da forma que se segue, *Alto-Alto* (HH): agrupamento de valores altos e próximos; *Baixo-Baixo* (LL): agrupamento de valores baixos e próximos; *Alto-Baixo* (HL): *outlier* de valores altos que não se agrupam, pois se encontram em meio a valores baixos; *Baixo-Alto* (LH): *outlier* de valores baixos que não se agrupam, em meio a valores altos. *Não Significativo* (NS): não se enquadram nos agrupamentos anteriores, conforme Anselin (1995).

Após análise do Índice de Moran I local foi realizada a contagem das frequências dos tipos de *clusters* a que pertenciam os municípios, que se formaram de um ano para outro. Considerou-se que, por exemplo, se em um determinado ano um município pertencia a um tipo de *cluster* HH e no ano seguinte ele passou a ser HL foi classificado que houve mudança de categoria. Caso o *status* do município devido ao tipo de *cluster* fosse NS e no ano seguinte continuasse NS ele seria classificado como permanência, uma vez que, continuou no mesmo tipo de *cluster*. Com isto, realizou-se o cálculo das probabilidades dos municípios em relação aos tipos de *clusters* (Y) mudarem ou permanecerem nas categorias designadas pelo teste de Moran I local para os anos em análise (1974-2015).

Foi construída a variável crescimento populacional do rebanho bovino (X), que é a proporção de mudança de um ano em relação a outro (1974 a 2015). Ou seja, a taxa de crescimento corresponde a proporção entre a diferença do tamanho do rebanho no ano seguinte Z_{i+1} (próximo ano) menos o ano atual Z_i (ano atual) dividido pelo tamanho do rebanho no ano Z_{i+1} (próximo ano):

$$\text{Taxa Crescimento Populacional } (X_i) = \frac{Z_{i+1} - Z_i}{Z_{i+1}}$$

E através da modelagem da proporção de crescimento populacional do rebanho bovino e das probabilidades das mudanças ou permanências dos tipos de *clusters*, aplicou-se a regressão Beta Bayesiana para realizar as estimativas destes parâmetros. Foi utilizado o pacote Bayesianbetareg (CEPEDA-CUERVO *et al.*, 2016) desenvolvido na linguagem de programação estatística R-project (R CORE TEAM, 2017). Tal pacote contém todos os algoritmos necessários e opções personalizáveis para simulações - para a estimação do modelo de regressão beta bayesiano, onde ambos os parâmetros, média e precisão, são modelados. Com os parâmetros estimados, desvios-padrão e intervalos de credibilidade, apresentam a análise de resíduos, o critério de comparação (BIC) e o diagnóstico para este modelo.

Os modelos foram definidos como as probabilidades de mudança ou permanências dos 27 tipos de *clusters* que podem ocorrer, sendo o modelo proposto da seguinte forma

$$\text{Prob. (tipos de clusters)} = \beta_0 + \beta_1 * \text{Taxa de Crescimento Populacional} + e,$$

em que a probabilidade dos tipos de *clusters* é a variável de interesse (por exemplo, HH-HH padrão alto-alto), a taxa de crescimento populacional é a matriz da variável explicativa e β é o vetor de coeficiente de associação e e é o erro aleatório do modelo. Em que as prioris foram não informativas (JEFFREYS, 1961), obtendo, assim, as respostas das posteriores como os valores dos parâmetros estimados dos modelos.

Para escolher o melhor modelo utilizou-se o Critério de Informação Bayesiano (BIC), além de avaliar o diagnóstico da convergência para os parâmetros por meio dos testes de Geweke (1992), Raftery e Lewis (1992) e Heidelberger e Welch (1983), que são calculados através de algoritmos Metropolis-Hastings e amostrador de Gibbs que são processos iterativos, e as cadeias resultantes desses métodos necessitam ter sua convergência constatada. Para tal, utilizamos o pacote CODA (*Convergence Diagnosis and Output Analysis*), que corresponde a um conjunto de funções implementadas no software R-project (R CORE TEAM, 2017). Além disso, o gráfico com as trajetórias das cadeias, bem como os gráficos de autocorrelações foram obtidos. Porém, aqui se omitiram os gráficos referentes às cadeias das probabilidades de mudança ou permanência devido ao elevado número de figuras que tal tarefa demandaria. Entretanto, tais indicações de convergência também foram verificadas (RAMPASO, 2014).

Na Figura 1, tem-se um dos exemplos da simulação para o modelo do *cluster* do tipo NS-HH, na qual se observa que, apesar de ter uma boa convergência, não teve uma boa estacionariedade para seus parâmetros, principalmente para o de precisão (γ). Vários modelos obtiveram comportamento semelhante, não obtendo estacionariedade apesar do esforço intensivo de simulações.

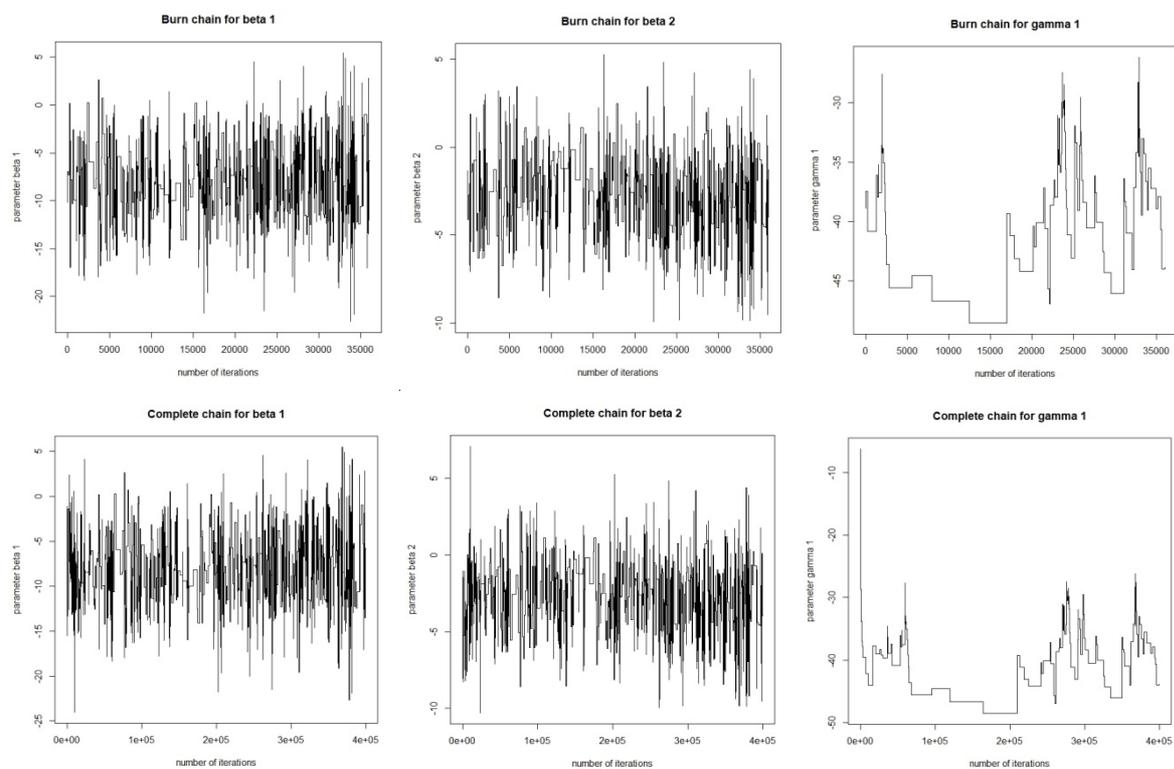


Figura 1. Representação gráfica das cadeias geradas do Modelo NS-HH com *Burn*: 10%, *Jump*: 10, *nsim*: 400.000 mil.

Aplicando os modelos definidos na metodologia, a estimação foi realizada por Monte Carlo via Cadeia de Markov (MCMC) através do software R-project (R CORE TEAM, 2017), com quatro combinações de tamanho de iterações iniciais, sendo de 10, 20, 30 e 40% descartadas para aquecimento (*burn*) da cadeia e variando em quatro tamanhos (*nsim*) de iterações subsequentes que foram de 100.000, 200.000, 300.000 e 400.000 mil. O salto (*jump*) utilizado foi combinado de 10, 20, 30 a 40. Em um primeiro momento, um estudo para verificar a indicação de convergência dos parâmetros foi conduzido.

Tabela 1. Esquema de simulações realizadas

<i>Burn in</i>	10%				20%				30%				40%			
<i>Jump</i>	10				20				30				40			
<i>Nsim</i> (10^5)	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

No método bayesiano proposto por Cepeda-Cuervo (2001) para ajustar os modelos de regressão beta, amostras da distribuição posterior são obtidas a partir das distribuições condicionais posteriores de dois blocos de parâmetros, um para os parâmetros médios e outro para o parâmetro de dispersão, assumindo distribuições prévias normais para os parâmetros de regressão de média e dispersão (CEPEDA e GAMERMAN, 2005). Essas distribuições posteriores condicionais são desconhecidas e analiticamente intratáveis. Assim, dois núcleos de transições normais são construídos para aplicar um algoritmo Metropolis Hastings para obter as amostras posteriores.

Com intuito de comparações entre as técnicas clássicas e bayesianas, modelos de regressão Beta com ligação logit também foram realizadas estimativas dos parâmetros para os tipos de *clusters* e a taxa do crescimento populacional do rebanho bovino, através do pacote *betareg* no software R-project (versão 3.4.1) (R CORE TEAM, 2017).

Também foi realizada a análise das matrizes de transição e estacionárias dos coeficientes estimados da regressão Beta e Beta Bayesiana através das Cadeias de Markov. Uma Cadeia de Markov pode ser descrita por uma sequência de grafos dirigidos, em que as arestas do gráfico n são rotulados pelas probabilidades de ir de um estado no tempo n para outros estados no tempo $n+1$, $\Pr(X_{n+1} = x | X_n = x_n)$. A mesma informação é representada pela matriz de transição de momento n para o tempo $n+1$. No entanto, as cadeias de Markov são assumidas frequentemente como sendo tempo-homogêneas. Nesse caso, o gráfico e a matriz são independentes de n e, portanto, não são apresentados como sequências. Os grafos são geralmente representados graficamente sendo desenhado um círculo para cada vértice, e para cada aresta é desenhado um arco conectando suas extremidades. Se o grafo for direcionado, seu sentido é indicado na aresta por uma seta (ROBERT e CASELLA, 1996).

Foi utilizado o pacote *markovchain* para analisar a probabilidade de transição das matrizes e o comando *steadyStates*, para encontrar o estado estacionário no software R-project (versão 3.4.1) (R CORE TEAM, 2017). Em que os estados estacionários são identificados a partir dos quais os autovetores correspondem aos autovalores de identidade e depois os normalizam para somar a unidade. Quando valores negativos são encontrados na matriz, a extração de autovalores é realizada na submatriz de classes recorrentes (ROSS, 2010).

E, por fim, com intuito de realizar a análise de sobrevivência para verificar a transição de um tipo de *cluster* se modificar para outro tipo, foi gerado através das estimativas das probabilidades de transição ajustadas do modelo de regressão Beta e Beta Bayesiano, o número de transições para um devido município escolhido por ter possuído o maior número

de mudanças como exemplificação, foram multiplicados o número total de anos pela probabilidade entre transição em meio a tipos de *clusters* de cada um dos modelos gerados.

E, por fim, foi verificada a probabilidade de permanência de categoria de *clusters* através da análise de sobrevivência com fração de cura, uma vez que, podem-se encontrar municípios suscetíveis e não suscetíveis a se modificarem para outro tipo de *clusters*. Foi realizada tal análise para um dos municípios que apresentou a maior quantidade de mudanças de *clusters* nos anos de 1974 a 2015 (tempo), para três modelos utilizando os valores ajustados pelas matrizes de transição para modelo Beta e o Beta Bayesiano, e os dados observados, em que a covariável foi à taxa do crescimento populacional do rebanho bovino. Se considerou a não ocorrência de mudanças como 0 (zero) e a ocorrência de mudanças foi considerada como 1 (um), que para a sobrevivência seriam a censura e a observação, respectivamente. Realizou-se o ajuste da curva de sobrevivência estimada por Kaplan-Meier e pela distribuição Weibull, estimados através da máxima verossimilhança, uma vez que, são indicados devido ao comportamento da variável taxa de crescimento populacional do rebanho bovino e com intuito de realizar comparações entre os modelos. Os pacotes *survival* e o *flexsurvreg* foram utilizados para tais análises no software R-project (versão 3.4.1) (R CORE TEAM, 2017).

4.3. Resultados e Discussões

Serão expostos os resultados a posteriori para os modelos de regressão Beta Bayesiano os quais foram verificados que atingiram as melhores convergências para os modelos, escolhidas através do BIC. Devido o alto número de simulações realizadas se faz extenso incluir todos os resultados obtidos.

Para averiguar a tendência das estimativas dos parâmetros dos modelos propostos para regressão Beta Bayesiana, foi testada convergência das cadeias que se encontram na Tabela 2. As cadeias com resultados a posteriores oferecem indícios de bom comportamento, um pequeno período transitório e pequenos desvios-padrão.

Os resultados (Tabela 2) indicam a convergência das cadeias para os parâmetros de regressão média, em um nível de 95%. Para os parâmetros de regressão de precisão, os testes indicam convergência, exceto para as estimativas dos parâmetros para os *clusters* do tipo HH-HH, HL-HL, LH-LH, NS-HH, NS-HL e NS-LL. As cadeias foram limpas usando uma queima (*burn*) em processo excluindo os primeiros 20% da amostra e um salto (*jump*) de 20 passos para coletar as amostras finais posteriores em um tamanho de 400.000 mil simulações.

De todas as simulações realizadas, os melhores resultados foram os que se encontram expostos na Tabela 2.

Tabela 2. Critérios de convergência para modelos regressão Beta Bayesiano dos modelos de permanência e mudança das categorias de *clusters* espaciais do rebanho bovino da região Nordeste do Brasil (1974 a 2015).

Parâmetros		Geweke	Raftery	Heidelberger e Welch		BIC
		R	Estacionária	Half-width		
Mudança	β_0	não	não	sim	sim	9,88
	β_1	sim	não	sim	sim	
	γ	não	não	sim	sim	
Permanência	β_0	não	sim	sim	sim	129,46
	β_1	não	sim	sim	sim	
	γ	não	sim	sim	sim	
NS-NS	β_0	não	sim	sim	sim	16,10
	β_1	não	sim	sim	sim	
	γ	não	sim	sim	sim	
HH-HH	β_0	não	sim	sim	sim	10,47
	β_1	não	sim	sim	sim	
	γ	não	não	sim	sim	
HL-HL	β_0	não	não	sim	sim	13,13
	β_1	não	não	sim	sim	
	γ	não	não	sim	sim	
LH-LH	β_0	não	não	sim	sim	15,55
	β_1	não	não	sim	sim	
	γ	não	não	sim	sim	
LL-LL	β_0	não	sim	sim	sim	14,53
	β_1	sim	sim	sim	sim	
	γ	sim	sim	sim	sim	
NS-HH	β_0	não	não	sim	sim	18,93
	β_1	sim	não	sim	sim	
	γ	não	não	sim	sim	
NS-HL	β_0	não	não	sim	sim	29,27
	β_1	não	não	sim	sim	
	γ	não	não	sim	sim	
NS-LL	β_0	não	não	sim	sim	12,33
	β_1	não	não	sim	sim	
	γ	sim	não	sim	sim	

Também pode-se observar a convergência das estimativas graficamente, assim como se segue para o modelo do *cluster* do tipo NS-NS, na Figura 2, que mostram um período transitório muito pequeno, produzindo uma indicação fortemente intuitiva da convergência, concordando com o resultado teórico da convergência. Os demais modelos apresentados

seguiram resultados similares, porém outros modelos que não foram apresentados ocorreram erros na simulação. Uma das justificativas para ocorrência nos erros das simulações, podem estar relacionadas às *prioris* que foram não informativas devido a restrição, em que não se pode indicar nos comandos do pacote implementado no R utilizado para tais análises (CEPEDA-CUERVO *et al.*, 2016).

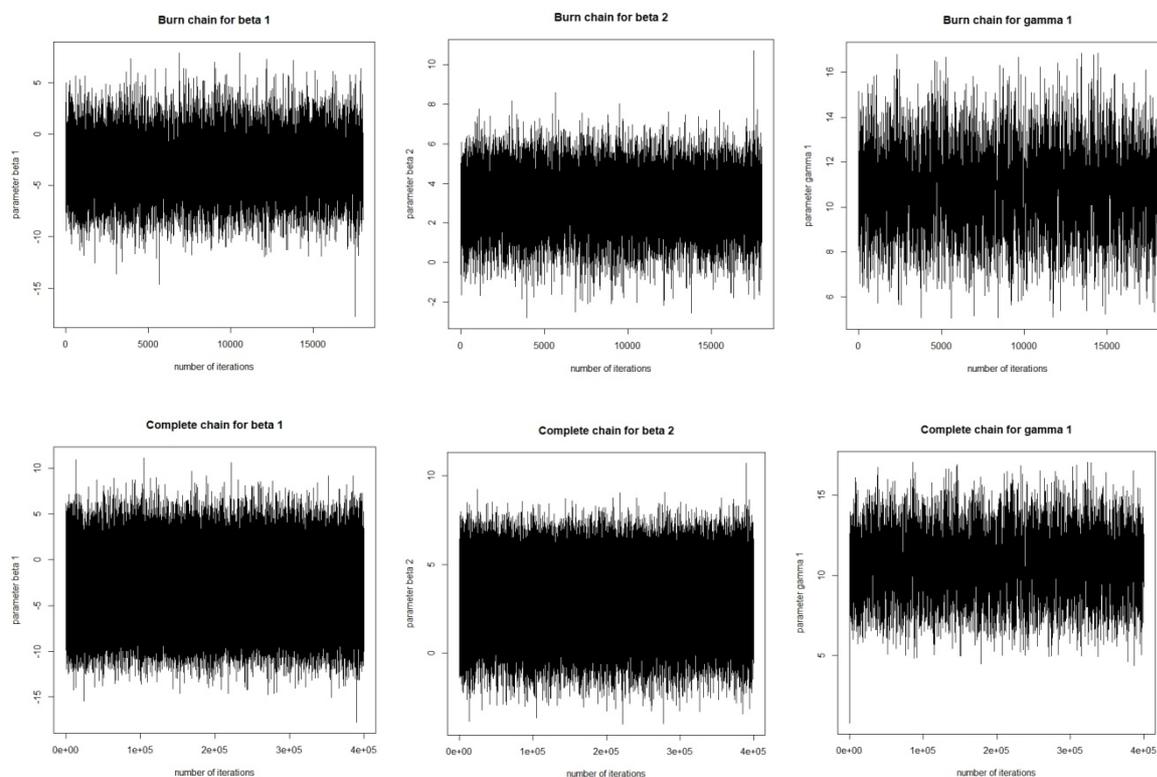


Figura 2. Representação gráfica das cadeias geradas do Modelo NS-NS para um *Burn*: 10%, *Jump*: 20 e um *nsim*: de 400 000 mil.

Os modelos de regressão Beta Bayesiano (Tabela 3), em que a média foi uma função dos tipos de *clusters* da mudança ou permanência das categorias e o parâmetro de dispersão foram modelados como uma função de probabilidade das modificações foi ajustada aos dados do efetivo do rebanho bovino na região Nordeste do Brasil. O modelo assume uma função com ligação logit para a média e uma função logarítmica para a precisão, incluindo os interceptos (Tabela 3). Nos resultados, os intervalos de credibilidade mostram que as estimativas têm uma probabilidade de 95% de serem diferentes de zero.

Foi avaliado que através das estimativas dos parâmetros (Tabela 3) para os modelos de regressão Beta Bayesiano, ocorreu uma retração do rebanho bovino na região do Nordeste do Brasil, porém para aqueles municípios com existência de *clusters* de permanência na mesma região indicou que houve um crescimento positivo (expansão).

Tabela 3. Estimativas a posteriori por ponto e por intervalo de credibilidade dos parâmetros e erro das estimativas.

Parâmetro		Estimativa	Erro	Intervalo de Credibilidade (95%)	
				Inferior	Superior
Mudança	β_0	-5,286	3,824	-12,746	2,076
	β_1	-3,06	2,13	-7,217	1,214
	γ	-28,734	5,998	-40,704	-17,846
Permanência	β_0	15,0657	3,6073	8,0932	22,12
	β_1	6,4433	1,8705	2,7819	10,1
	γ	14,3603	0,6501	13,1113	15,65
NS-NS	β_0	-3,0581	2,7332	-8,3096	2,43
	β_1	3,0454	1,4028	0,2153	5,701
	γ	10,9736	1,7451	7,6286	14,437
HH-HH	β_0	-6,036	3,866	-13,968	1,377
	β_1	-1,892	2,182	-6,152	2,491
	γ	-23,631	5,298	-34,618	-14,014
HL-HL	β_0	-7,185	3,748	-14,673	0,07
	β_1	-2,391	2,126	-6,537	1,812
	γ	-37,664	5,306	-474,764	-26,417
LH-LH	β_0	-6,986	4,033	-14,779	1,104
	β_1	-2,516	2,218	-6,866	2,131
	γ	-37,527	5,159	-46,384	-26,356
LL-LL	β_0	-4,129	3,74	-11,412	3,265
	β_1	-1,973	2,141	-6,16	2,235
	γ	-4,007	3,571	-10,253	3,493
NS-HH	β_0	-7,932	3,886	-15,65	-0,341
	β_1	-2,571	2,244	-7,029	1,996
	γ	-43,545	5,452	-50,232	-32,257
NS-HL	β_0	-8,082	4,068	-16,215	0,323
	β_1	-2,666	2,374	-7,32	2,122
	γ	-46,802	2,482	-47,86	-38,601
NS-LL	β_0	-6,992	3,866	-14,898	0,322
	β_1	-2,922	2,194	-6,965	1,48
	γ	-40,51	4,658	-47,411	-30,174

Também foi avaliada a estacionariedade da matriz observada geral (Figura 3A) da frequência de mudanças entre os tipos de *clusters*, dada a sua origem e destino ao passar dos anos. Verificou-se que todos os tipos de *clusters* estabeleceram probabilidades, mesmo que baixas, de modificarem para a média (NS), sendo que a probabilidade dos *clusters* permanecerem no mesmo tipo fica em torno dos 95%; ou seja, o rebanho bovino na região Nordeste possui grandes chances de não mudar do município que se encontra (Figura 3A).

Outro resultado relevante para estacionariedade da matriz geral é que o *cluster* do tipo LL, ou seja, de baixa produção do rebanho bovino, não se modifica para *cluster* do tipo LH,

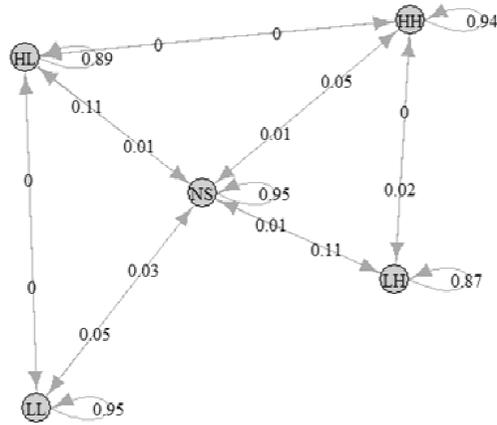
nem HL, que são *clusters* de características de baixa produções ou alta, respectivamente, sendo que as regiões vizinhas ocorram altas ou baixas produções, respectivamente. Assim como *clusters* de altas produções (HH), não migram para o tipo HL nem LH, que possam ter regiões vizinhas que ocorram alta ou baixa produção respectivamente.

Foi verificado que para os municípios da região do Nordeste obterem estacionariedade são necessários 308 anos a partir de 2015. Sendo assim, até o ano de 2323 as probabilidades do estado estacionário tenderão possuir maior concentração de *clusters* do tipo NS, ou seja, há disposição de que nesta região se concentrem o rebanho bovino na média, seguido de regiões com baixas (LL) produções e com vizinhos também de baixas produções. Além disso poucas regiões tendem a ter altas produções com vizinhos que também possuam alta concentração do rebanho bovino (Figura 3 B).

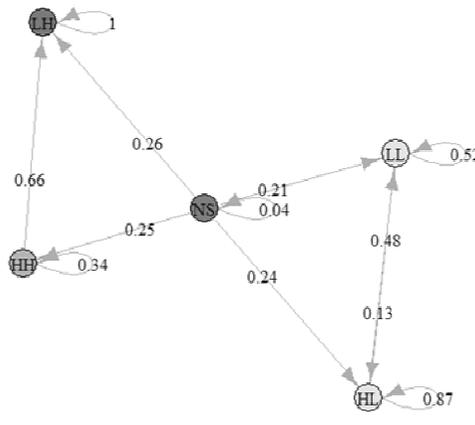
A partir das estimativas dos parâmetros da análise de regressão Beta para os tipos de *clusters* as estimativas dos parâmetros que foram significativos foram obtidas a estacionariedade da matriz para regressão Beta (Figura 3C). Verificou-se, através do diagrama de transição, que o *cluster* do tipo NS tem probabilidade de migrar para os demais tipos. Desta forma, nota-se que todos os *clusters* são absorventes, ou seja, se eles atingirem seus estados não modificarão para outro tipo de *clusters* no futuro. Isto é, aqueles municípios que possuem produções do rebanho bovino na média possuem probabilidades altas de modificarem para localizações de baixas produções ao redor de municípios com baixas produções (LL), como também para municípios que possuem altas produções ao redor de municípios com baixas produções (HL). O *cluster* do tipo LH, não tem probabilidade de migrar para outro tipo de *cluster*. Ou seja, municípios com baixas produções em que os vizinhos da região sejam de altas produções do rebanho bovino não ocorrem, segundo a regressão Beta.

Outro ponto importante foi analisado a partir da matriz de transição da regressão Beta (Figura 3D): foi visto que todos os *clusters* podem atingir absorvência, ou seja, uma vez que os municípios que tenham produções de rebanho bovino não vão deixar de produzir na devida localização, e que os municípios estão mais propensos à produções do rebanho bovino ao redor de municípios que tem altas produções do rebanho bovino, sendo que com baixa produtividade (HL). Também foi visto que podem ocorrer municípios com baixa concentração do rebanho bovino, porém ao redor destes municípios podem ocorrer alta produtividade do rebanho (LH) com probabilidades relevantes (Figura 3C).

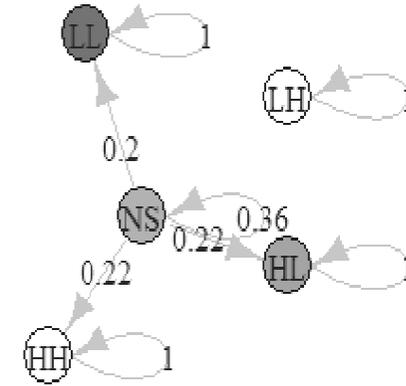
Também foi avaliado a estacionariedade para matriz dos modelos da regressão Beta (Figura 3D), em que foi verificado que são necessários 1700 anos a partir de 2015 para que se



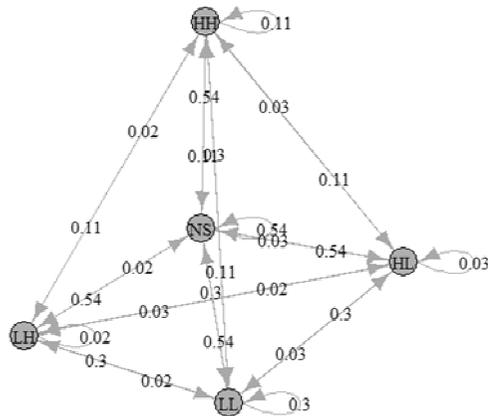
A) Matriz transição geral



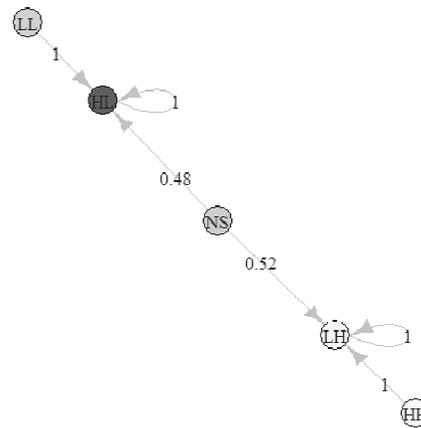
C) Matriz de transição beta



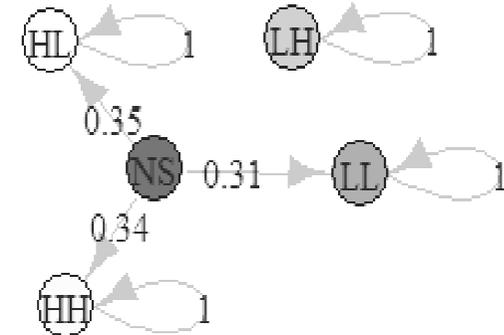
E) Matriz de transição beta bayesiana



B) Matriz estacionária geral



D) Matriz de estabilidade beta



F) Matriz de estabilidade beta bayesiana

Figura 3. Diagrama da transição e estacionaridade das matrizes para dados gerais, beta e beta bayesiana segundo a classificação dos tipos de *clusters*.

tenha estabilização do rebanho bovino na região Nordeste do país. Sendo assim, foi verificado que até o ano de 3715 segundo tais modelos, a movimentação do rebanho bovino na região Nordeste do Brasil tendem a possuir mais *clusters* que estiverem na média da produção do rebanho bovino, eles possuem alta probabilidade de se movimentarem para municípios com alta produções ao redor de municípios com baixas produções como também podem ocorrer movimentações em municípios com baixas produções de rebanho bovino ao redor de municípios com altas concentrações de rebanho bovino.

Na verificação da estacionariedade da matriz segundo os modelos de regressão Beta Bayesiana (Figura 3F) para os tipos de *clusters*, verifica-se que todos os *clusters* são absorventes, ou seja, se eles atingirem seus estados não modificarão para outro tipo de *clusters* no futuro. Só o *cluster* do tipo NS tem probabilidade de migrar para os demais tipos, exceto para o tipo LH, que possui a característica de se localizar ao redor de municípios com alta criação do rebanho bovino e tais municípios possuem baixa produtividade. Pode-se interpretar que não se estabelece a criação de rebanhos bovinos em municípios com essa característica (Figura 3E).

Para matriz dos modelos da regressão Beta Bayesiana, foi visto que são necessários 1000 anos para atingir o estado estacionário, a partir do ano de 2015. Assim, foi verificado que até o ano de 3015, segundo tais modelos, a movimentação do rebanho bovino na região Nordeste do Brasil tenha estabilidade, com uma tendência que o *cluster* do tipo NS se movimente com maiores probabilidades para os tipos HL (35%), HH (34%) e LL (31%). Ou seja, segundo a regressão Beta Bayesiana ocorrem probabilidades do rebanho bovino saírem da produção média (NS) podendo se movimentar para municípios com alta ou baixa produtividade do rebanho bovino (Figura 3F).

Logo, ambos os modelos para regressão Beta e Beta Bayesiano demoram mais anos para se tornarem estacionárias do que a matriz para dados observados. Sendo que o modelo para estimativas Beta Bayesiana encontrou-se estados estacionários com menor tempo. Ambos observaram que os municípios da região Nordeste tendem a possuir produção do rebanho bovino na média podendo haver baixas transições para outros tipos de *clusters*.

Ao longo dos anos de 1974 a 2015 ocorreram 3.965 transições entre os tipos de *clusters* para os municípios, ocorrendo um pico no período de 1992-1993 com 330 mudanças.

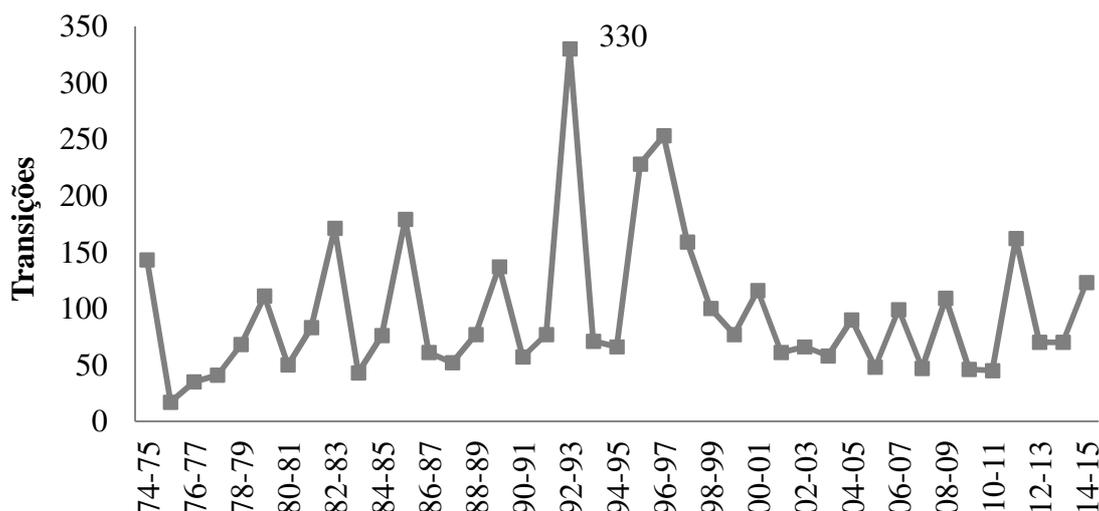


Figura 4. Número de transições entre os tipos de *clusters* ao longo dos anos (1974 a 2015).

O estado nordestino que maior ocorreu número de transições foi na Bahia e a menor ocorrência foi em Sergipe, isto pode ser explicado devido que, o estado da Bahia é o maior em território na região Nordeste, enquanto que Sergipe é o menor. O município que mais apresentou modificações foi o de Serra do Ramalho (BA) com 16 mudanças, sendo a moda para mudanças de 4 vezes (Figura 5).

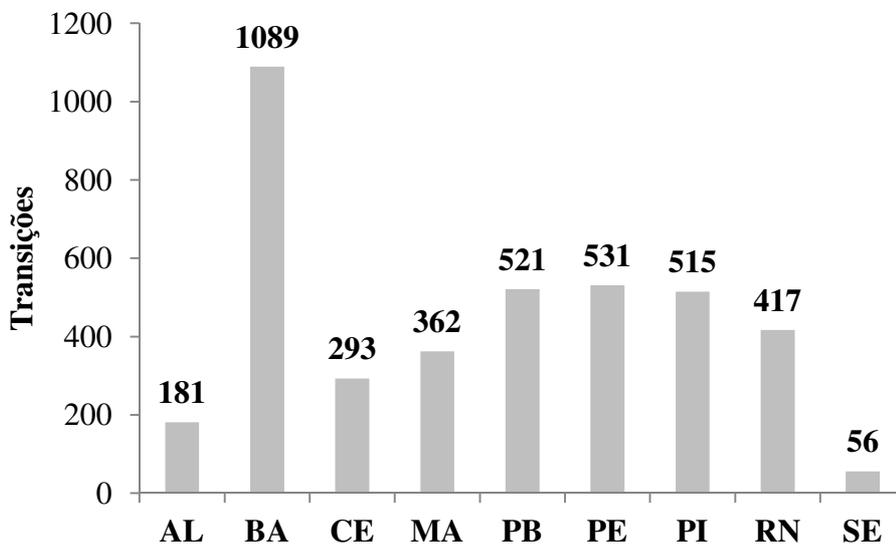


Figura 5. Número de transições dos *clusters* em relação aos estados da região Nordeste.

E, por fim, com intuito de avaliar a sobrevivência dos *clusters*, foi realizada uma aplicação ao município de Serra do Ramalho (BA), que apresentou o maior número de modificações, ressaltando que tal análise pode ser verificada para qualquer outro município. Temos também por objetivo comparar as estimativas dos modelos para regressão Beta e Beta

Bayesiana para assim poder verificar qual dos modelos ajustados se aproxima mais aos dados observados.

Na Figura 6, apresentamos a probabilidade de permanência de categoria de *cluster* para o município de Serra do Ramalho com a covariável taxa de crescimento para dados observados. Nota-se que neste município ocorreram transições entre as categorias NS, LH, HL e HH. O *cluster* que teve mais ocorrência foi o NS, ou seja, o município esteve na maioria das vezes com a produção do rebanho bovino na média, porém conseguiu atingir produções altas (HH), com oscilações em produções baixas ao redor de altas (LH) concentrações, tendendo nos anos finais produções altas ao redor de municípios com baixa (HL) produção do rebanho bovino.

Em comparação aos modelos Beta e Beta Bayesiano para a probabilidade da sobrevivência de permanência de categoria de *cluster* para o município de Serra do Ramalho (BA), observou-se que para os tipos de *clusters* NS e HH o melhor ajuste foi para o modelo Beta Bayesiano (Figura 7A), em que obtiveram menor AIC. Já os *clusters* LH e HL o melhor ajuste foi através do modelo Beta (Figura 7 B). E para LL os ajustes foram idênticos, uma que, neste município não houve produções de baixa concentração do rebanho bovino.

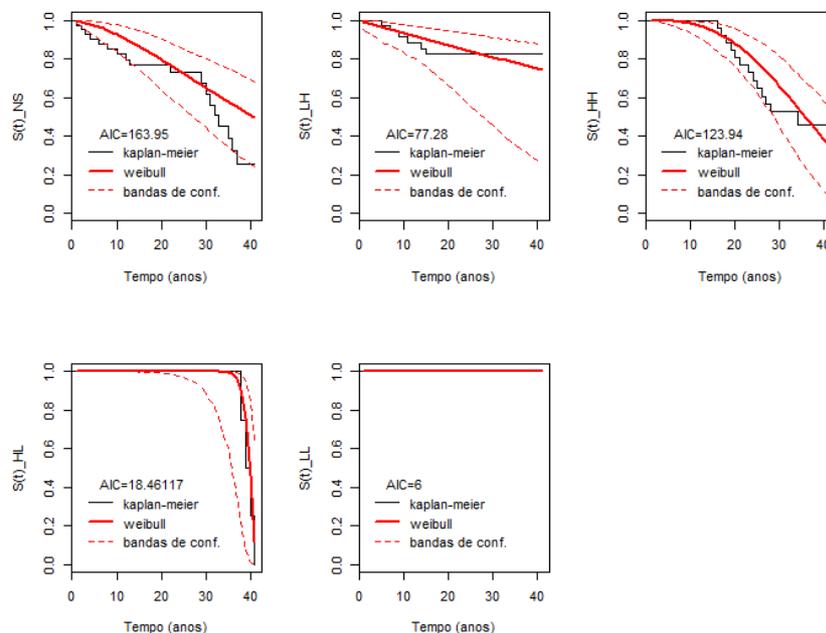
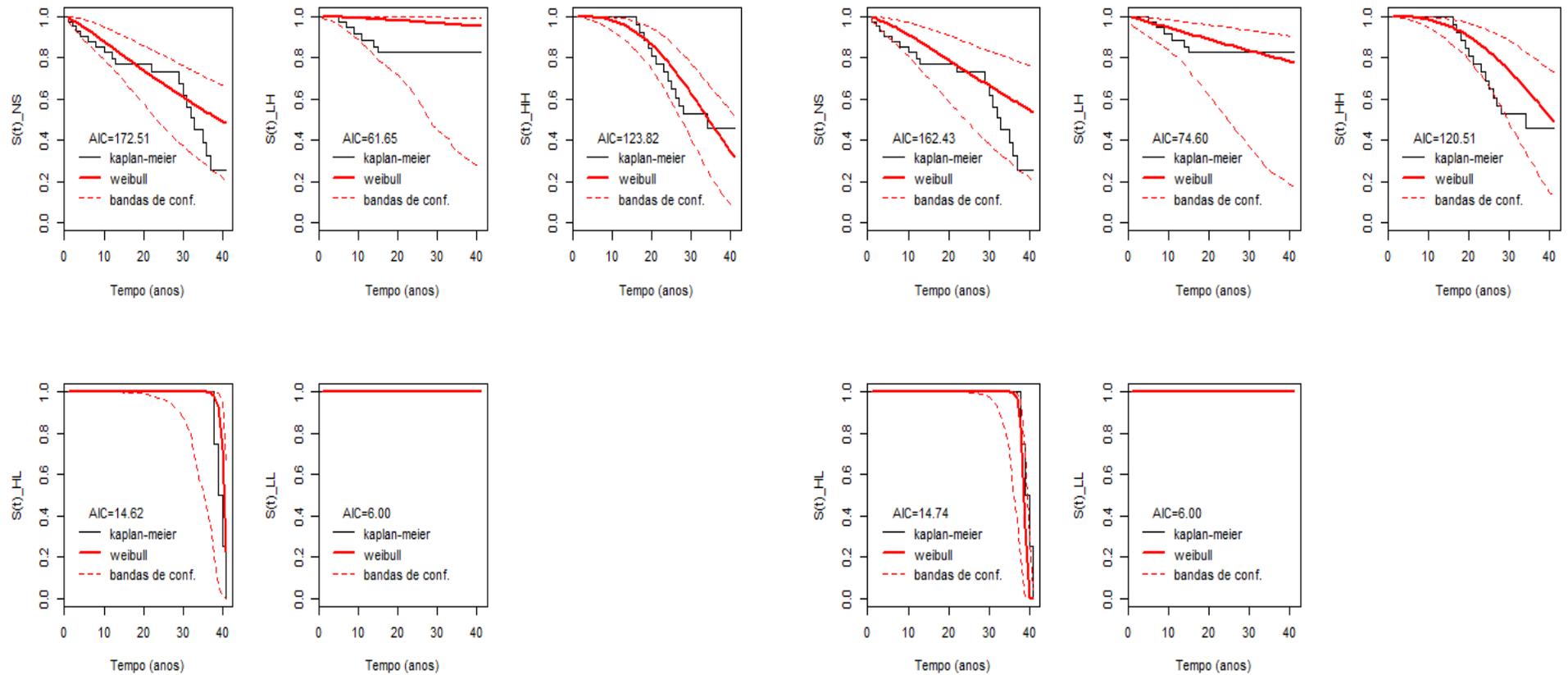


Figura 6. Probabilidade de permanência de categoria de *cluster* para o município de Serra do Ramalho (BA) com a covariável taxa de crescimento para dados observados.



A)

B.)

Figura 7. Probabilidade de permanência de categoria de *cluster* para o município de Serra do Ramalho (BA), sendo o primeiro modelo com covariável para modelos Betas (A) e o segundo com a covariável para modelo Beta Bayesiano (B) para o rebanho bovino.

A inferência bayesiana, ao contrário da inferência clássica, leva em conta, também, o conceito de probabilidade subjetiva, que mede o grau de incerteza que se tem sobre a ocorrência de um determinado evento do espaço amostral. Portanto, a análise bayesiana descreve toda quantidade desconhecida por meio de probabilidades (SOUSA, 2005).

Assim como na inferência frequentista, a inferência bayesiana trabalha na presença de observações y , cujo valor é inicialmente incerto e descrito por uma densidade $f(y/\theta)$. A quantidade θ serve como indexador da família de distribuições das observações, representando uma característica de interesse (GAMEMAN, 1997).

A diferença formal entre a inferência bayesiana e a inferência frequentista é que, para a inferência bayesiana, o parâmetro θ é uma variável aleatória, possuindo, então, uma distribuição de probabilidade, enquanto que para a inferência frequentista, os parâmetros são valores fixos ou constantes. Além disso, a inferência bayesiana usa toda a informação disponível a priori, enquanto a frequentista ignora esta informação.

A inferência bayesiana, criada antes mesmo da análise clássica utilizada atualmente, ficou esquecida durante anos pelo fato de que sua aplicação, a princípio, exigiria o cálculo analítico de integrais muito complicadas.

Com a utilização de métodos de simulação via Cadeias de Markov, nos casos em que as integrais não podem ser resolvidas analiticamente, passou a ser possível encontrar uma solução aproximada para o problema. Entre estes métodos, destacam-se os algoritmos do Amostrador de Gibbs e do Metropolis-Hastings.

Com isso, a partir de 1990, a inferência bayesiana tem sido bastante usada para resolver problemas em diversas áreas e os resultados coerentes tem atraído cada vez mais a atenção dos pesquisadores. No entanto, a facilidade de implementação dos algoritmos não deve substituir e, sim, complementar o pensamento crítico sobre o problema, por parte do pesquisador (SOUSA, 2005). Além disso, é importante usar um dos critérios disponíveis para monitorar a convergência da cadeia gerada, evitando, assim, que se utilize esforço computacional além do necessário, ou que se pare o processo antes da convergência, o que, certamente, conduziria a conclusões incorretas. Alguns dos critérios de convergência mais comumente utilizados estão apresentados.

Foi apresentada uma metodologia através da regressão Beta Bayesiana como proposta de estimar parâmetros dos *clusters* espaciais da probabilidade de mudança ou permanência do rebanho bovino nos municípios da região Nordeste do Brasil, em comparação com modelo de regressão Beta, e tal proposta mostrou-se uma alternativa viável, uma vez que, os modelos

Bayesianos obtiveram melhores ajustes, se mostrando assim com maior flexibilidade e se aproximando mais com os dados observados, em algum dos casos estudados.

Utilizou-se o pacote *Bayesianbetareg* do software R e foi visto que o mesmo não possui muitas aplicações para tal área de estudo, provavelmente devido ao fato de ser recém-publicado e também poder estar relacionado o imperativo do aporte computacional que se necessita a depender do tamanho da simulação para que os parâmetros tenham convergência. Outra dificuldade na utilização deste pacote foi a impossibilidade de estabelecer as prioris, e as mesmas tiveram de ser utilizadas de modo não informativas devido a implementação do pacote no R pelos autores Cepeda-Cuervo *et al.* (2016). Se fosse possível indicar prioris, talvez os resultados dos modelos pudessem ser melhores.

4.4. Conclusões

Os resultados dos modelos Betas Bayesianos sugerem que ocorrência de um decréscimo do rebanho bovino com o passar dos anos na região Nordeste do Brasil no que diz respeito ao comportamento das mudanças de *clusters*. Já para a permanência de *clusters* verificou-se que houve estimativa positiva, ou seja, aqueles municípios com criação de rebanho bovino continuaram com o passar dos anos com a produção do rebanho, sendo assim *clusters* estáveis.

Em relação a estabilidade das matrizes, o modelo de regressão Beta Bayesiana obteve estacionariedade com menor tempo e menos parâmetros estimados em relação a regressão Beta. Ambos demoram mais anos para se tornarem estacionários em relação à matriz de dados observada.

E, por fim, foi avaliada a sobrevivência dos *clusters* para o município de Serra do Ramalho localizado no estado da Bahia. Essa escolha se deu pelo fato deste município ser o maior em ocorrência de transições de categorias de *clusters*, e serviu de exemplo para tais análises, podendo ser realizado o mesmo estudo para qualquer outro município. Com isto também foi visto que o modelo de regressão Beta Bayesiana se revelou mais flexível que o modelo de regressão Beta.

Capítulo 5

Neste capítulo são apresentadas as conclusões do presente trabalho e as recomendações para a continuidade dos trabalhos nesta área de estudo.

5.1. Conclusões gerais

A presente tese teve o objetivo de realizar a investigação da movimentação de *clusters* por municípios da região Nordeste do Brasil, que foi utilizado através do estudo base que foi a análise de autocorrelação espacial, baseado no Índice local de Moran I, no período de 1974 a 2015, em que, criaram-se variáveis a partir de tais análises para averiguar a dinâmica do rebanho bovino na região Nordeste do país em 41 anos. Com relação às dificuldades para desenvolvimento do trabalho, o levantamento de dados foi uma etapa difícil do estudo. A falta e a qualidade de determinados dados sobre a localização do estudo foram um dos obstáculos encontrados. Quanto à revisão bibliográfica, foi realizada uma atualizada revisão bibliográfica abordando diversos conceitos relacionados a este caso, em que, é considerada uma análise inovadora nessa área de pesquisa.

Sendo assim, pode-se concluir que na região Nordeste houve poucos grupos com produções altas e baixas e vice-versa ao longo do tempo; percebeu-se que ocorreu uma movimentação de grupos do tipo de altas produções na direção do noroeste e sul, que seriam os estados referentes a região da MATOPIBA, citada na introdução como uma das regiões mais importantes atualmente, pois é considerada a grande fronteira agrícola nacional. Já os *clusters* do tipo de baixas concentrações, ocorreram uma expansão para direção do interior do nordeste e leste (referente a municípios dos estados do Rio Grande do Norte, Paraíba, Pernambuco e Alagoas), podemos ser justificada ao fato de essas regiões estarem localizadas no litoral; desta forma, possuem outras variedades de insumos e produções como, por exemplo, a cana de açúcar que é cultivada em grande parte em todas estas regiões diagnosticadas com baixas concentrações.

Foi possível diagnosticar que a produção do rebanho bovino na região Nordeste está retraindo com o passar dos anos, assim como verificado pelos modelos da regressão Beta, Beta Bayesiana e de sobrevivência.

Apesar de problemas para simulação dos parâmetros dos modelos Beta Bayesiano o mesmo obteve estacionariedade com menor tempo e com menos parâmetros estimados se comparado aos resultados das matrizes estáveis para os modelos Betas.

Ambos demoram mais anos para se tornarem estacionários em relação à matriz geral dos dados observados. Logo, pode-se concluir que o modelo Beta Bayesiano foi mais eficaz que o Beta. Mesmo com problemas de convergência na simulação para os modelos Bayesianos, foi possível obter resultados que se aproximavam mais aos dados observados, mostrando assim o quanto flexível e importante tal técnica.

5.2 Trabalhos Futuros

Como possíveis trabalhos futuros, pode-se apontar:

Faz-se necessário que os órgãos de desenvolvimento de banco de dados possam incluir aos seus sistemas o posicionamento global das variáveis que disponibilizam, pois desta forma pode-se realizar uma análise mais precisa utilizando a localização, neste caso o efetivo do rebanho bovino. Assim, como disponibilidade de dados climáticos consistentes, como também características do tipo de terra e de solo, além de dados como preço do efetivo por localização de todas as regiões do território nacional. Ou seja, com variáveis consistentes poderia se aprofundar o entendimento dos modelos propostos e tentar relacionar com outras variáveis quais os aspectos que de fato influenciam a movimentação do efetivo do rebanho bovino na região analisada, uma vez que, foi visto que houve sim uma dinâmica.

Com relação a programação do pacote para analisar a regressão beta bayesiana desenvolvido no programa R Project, um trabalho futuro necessário é que se possa incluir uma opção nos comandos das *prioris* dos dados, que as mesmas possam ser indicadas pelo pesquisador, uma vez que, esta foi uma de nossas dificuldade para obter convergência nos parâmetros dos modelos estudados. Neste sentido, faz-se necessário mantê-lo sempre atualizado com novas versões, dado que a análise bayesiana tem expandido nos últimos anos sua utilização necessitando as implementações desta técnica nos softwares.

Outro trabalho futuro bastante relevante diz respeito a modelar tais variáveis aqui utilizadas com outras distribuições e realizar comparações entre elas, como por exemplo, as novas classes de distribuições generalizadas, como a Beta G, que podem obter melhores resultados.

REFERÊNCIAS

- ADAS, Melhem. Panorama Geográfico do Brasil. São Paulo: Moderna, 1983.
- AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723, 1974.
- ALBUQUERQUE, P. H. M. Curso de Econometria Espacial Aplicada. Fortaleza:IPECE, 2009.
- ALITI, G. F. Fluxo da Informação em uma cadeia organizada da carne bovina no Rio Grande do Sul. Programa de Pós-Graduação em Agronegócio. UFRGS. Dissertação de Mestrado. 2003.
- ALMEIDA, E. Curso de Econometria Espacial Aplicada. ESALQ-USP: Piracicaba, 2004.
- ANSELIN, L. Local Indicators of Spatial Association – LISA. Geographical Analysis, 27, n.2, p. 93-115, 1995.
- BATISTA FILHO M. B. Evolução do Efetivo de Bovinos e da Produção do Gado de Corte no Brasil, Estado de Goiás e Município de Jataí (GO). Universidade Federal de Goiás Regional Jataí Curso de Graduação em Zootecnia. Relatório De Projeto Orientado, 2016.
- BORGES, L. C. Análise bayesiana do modelo fatorial dinâmico para um vetor de séries temporais usando distribuições elípticas. Tese Instituto de Matemática e Estatística da Universidade de São Paulo, 2008.
- BRANDÃO, Antonio S. P.; REZENDE, Gervásio Castro; MARQUES, Roberta W. da Costa. Crescimento agrícola no período 1999-2004: explosão da área plantada com soja e meio ambiente no Brasil. Texto para Discussão, Rio de Janeiro, n. 1062, 2005.
- CALSAVARA, V. Modelos de Sobrevivência com Fração de Cura usando um Termo de Fragilidade e Tempo de Vida Weibull Modificada Generalizada. 2011. 72f. Dissertação (Mestrado) – Departamento de Estatística, Universidade Federal de São Carlos. São Carlos- SP, Brasil.
- CÂMARA G, MONTEIRO AMV, DRUCK S, CARVALHO MS. Análise espacial e geoprocessamento. In: DRUCK, S.; CARVALHO, M.S.; CÂMARA, G; MONTEIRO, A. M. V. Análise espacial de dados geográficos. 2002. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/análise/>. Acessado em: 20 de novembro de 2011.

CEPEDA-CUERVO E., GAMERMAN D. Bayesian Methodology for Modelling Parameters in the two Parameter Exponential Family. *Rev. Estad.* 57: 93-105, 2005.

CEPEDA-CUERVO, E. (2001), Modelagem da variabilidade em modelos lineares generalizados, PhD thesis, Instituto de Matemática. Universidade Federal do Rio de Janeiro. URL: <http://www.bdigital.unal.edu.co/9394/1/modelagemvariabilidademodelos.pdf>.

CEPEDA-CUERVO, E. GARRIDO, L.. Bayesian beta regression models with joint mean and dispersion modeling. *Monte Carlo Methods and Applications*, 21(1), pp. 49-58, from doi:10.1515/mcma-2014-0007, 2015.

CEPEDA-CUERVO, E. JAIMES, D. MARÍN, M. ROJAS, J. "Bayesian Beta Regression with Bayesian-betareg R-package", *Computational Statistics*, vol. 31, no. 1, pp. 165-187, 2016.

CHEN, M.-H., IBRAHIM, J. G. SINHA, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447), 909–919.

CORBIÈRE, F. JOLY, P. A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine*, v. 85, n. 2, p. 173-180, 2007.

CRIBARI-NETO, F., SOUZA, T. C., Testing inference in variable dispersion beta regressions. *Journal of Statistical Computation and Simulation* 82, 1827–1843, 2012.

CRIBARI-NETO, F., ZELEIS, A., Beta regression in r. *Journal of Statistical Software* 34 (2), 2010.

DALLA-NORA, E. L. *et al.* Why have land use change models for the Amazon failed to capture the amount of deforestation over the last decade? *Land Use Policy*, v. 39, p. 403-411, 2014. ISSN 0264-8377. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0264837714000246>.

EMBRAPA Gado de Leite. Introdução a importância Econômica do Gado no Brasil. 2015. Acessado em: <http://www.cnp.gl.embrapa.br/sistemaproducao/31-introdu%C3%A7%C3%A3o-e-import%C3%A2ncia-econ%C3%B4mica>.

ESPINHEIRA, P. L., FERRARI, S., CRIBARI-NETO, F., On beta regression residuals. *Journal of Applied Statistics* 35, 407—419, 2008b.

ESPINHEIRA, P., FERRARI, S., CRIBARI-NETO, F., Influence diagnostics in beta regression. *Computational Statistics & Data Analysis* 52 (9), 4417–4431, 2008a.

ESRI, 2009. ArcGIS Desktop 9.3 Help [online]. Disponível em: <http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?tocVisible=0&ID=1&TopicName>

=How%20Spatial%20Autocorrelation:%20Moran's%20I%20(Spatial%20Statistics)%20works&pid=1. Acessado em 12 de dezembro de 2016.

FERRARI, S., CRIBARI-NETO, F., Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 (7), 799–815, 2004.

FERRAZ, Cláudio. Explaining agriculture expansion and deforestation: evidences from the Brazilian Amazon - 1980/ 98. Texto para discussão, Rio de Janeiro, n. 828, 2001.

FORNAZIER, A.; VIEIRA FILHO, J. E. R. Heterogeneidade estrutural na produção agropecuária: uma comparação da produtividade total dos fatores no Brasil e nos Estados Unidos. Brasília: Ipea, 2013. (Texto para Discussão, n. 1819).

FREITAS, T. Pecuária brasileira expande-se em direção ao sul. Conexão Tocantins Portal de Notícias. Disponível em: <http://coneaoto.com.br/2016/08/23/pecuaria-brasileira-expande-se-em-direcao-ao-sul>. Acesso em: 23 de agosto de 2016.

GAMERMAN, D. (1997), Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* 7(1), 57–68.

GASQUES, J. G. BASTOS E. T. VALDES, C. BACCHI M. R. P. Total factor productivity in Brazilian agriculture. In: FUGLIE, K. O.; WANG, S. L.; BALL, V. E. (Eds.). *Productivity growth in agriculture: an international perspective*. Oxfordshire: CAB International, 2012. p. 145-162.

GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, Taylor & Francis Group, v. 85, n. 410, p. 398–409, 1990.

GEWEKE, J. BERNARDO, J. M. BERGER, J. O. DAWID A. P. SMITCHE, A. F. M. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. [S.l.]: Federal Reserve Bank of Minneapolis, Research Department, 1991.

GOMES, A. P.; ALCANTARA FILHO, J. L.; SCALCO, P. R. Efi ciência, tecnologia e produtividade total dos fatores: uma análise das mudanças recentes na agropecuária do Nordeste. In: ENCONTRO REGIONAL DE ECONOMIA DO NORDESTE, 14., 2009, Fortaleza. Anais... Fortaleza: ANPEC/BNB, 2009.

HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. *Operations Research, INFORMS*, v. 31, n. 6, p. 1109–1144, 1983.

HOGGART, C. e GRIFFIN, J. E. (2001). A Bayesian partition model for customer attrition. In E. I. George, editor, *Bayesian Methods with Applications to Science, Policy, and Official Statistics (Selected Papers from ISBA 2000)*, pages 61–70, Creta, Greece. International Society for Bayesian Analysis, Proceedings of the the Sixth World Meeting of the International Society for Bayesian Analysis.

IBGE - INSTITUÍDO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICAS. Sistema IBGE de recuperação automática SIDRA. 2006. Disponível em: <<http://www.sidra.ibge.gov.br/bda/agric/default.asp?z=t&o=11&i=P>>. Acesso em: 7 de dezembro de 2015.

IBGE, Mapa da região Nordeste. Elaborado a partir de base cartográfica do IBGE, 2016. Disponível em: <http://www.baixarmapas.com.br/mapa-da-regiao-nordeste/>

IBGE. Indicadores IBGE: Estatística da Produção Pecuária. On-line. Disponível em: http://ibge.gov.br/home/estatistica/indicadores/agropecuaria/producaoagropecuaria/abat-e-leite-couro-ovos_201504_publ_completa.pdf. Acesso em: 11 de janeiro de 2016.

JACKSON, C. (2016). flexsurv: A Platform for Parametric Survival Modeling in R. *Journal of Statistical Software*, 70(8), 1-33. doi:10.18637/jss.v070.i08

JACKSON, C. H. AND SHARPLES, L. D. AND THOMPSON, S. G. (2010) Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. *International Journal of Biostatistics* 6(1):Article 34.

JEFFREYS H. *The Theory of Probability* (3rd ed.). Oxford. p. 432, 1961.

KLEINBAUM, D. G., 1995. *Survival Analysis: A Self-Learning Text*. New York: Springer.

LOUZADA-NETO, F, MAZUCHELI J., ACHCAR J.A. Análise de Sobrevivência e Confiabilidade. III Jornada Regional de Estatística e II Semana da Estatística, Maringá, 2002.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. PROJEÇÕES DO AGRONEGÓCIO, 8ª edição. Ano 2017. Disponível em: <<http://www.agricultura.gov.br/assuntos/politica-agricola/todas-publicacoes-de-politica-agricola/projecoes-do-agronegocio/projecoes-2017-finalizado.pdf/view>>.

MENDES, S. M.; TEIXEIRA, E. C.; SALVATO, M. A. Investimentos em infraestrutura e produtividade total dos fatores na agricultura brasileira: 1985-2004. *Revista Brasileira de Economia*, Rio de Janeiro, v. 63, n.2, p. 91-102, 2009.

MUCHAGATA, Márcia; BROWN, Katrina. Cows, colonists and trees: rethinking cattle and environmental degradation in the Brazilian Amazon. *Agricultural Systems*, New York, n. 76, 2003.

NEIVA, J. N. M. SANTANA, G. Z. M. NEIVA, A. C. G. R. PERSPECTIVAS PARA A PRODUÇÃO DE BOVINOS DE CORTE NO SEMIÁRIDO NORDESTINO. Associação Brasileira de Criadores. 2003. Disponível em: <http://www.abccriadores.com.br/images/upload/perspectivas%20para%20a%20producao%20de%20bovinos%20de%20corte%20no%20semi-.pdf>.

NORRIS, J.R., Markov Chains, Cambridge University Press, 1998.

PAULA, J.S; DALBERTO, C.R; BOHN, L. Determinantes da pobreza rural no Brasil: um enfoque espacial. 6ª Conferência Internacional sobre Estatísticas Agropecuárias. Rio de Janeiro, 2013.

PEROBELLI, Fernando Salgueiro et al. Análise espacial da produtividade do setor agrícola brasileiro: 1991-2003. Nova Economia, v. 17, p. 65-95, 2007.

R Development Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RAFTERY, A. E.; LEWIS, S. Comment: One long run with diagnostics: implementation strategies for markov chain monte carlo. Statistical Science, Hayward, v. 7, n. 4, p. 493-497, 1992.

RAMPASO, R. C. Análise Bayesiana de dados espaciais explorando diferentes estruturas de variância. Dissertação Programa de Pós-graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista UNESP, Presidente Prudente, 2014.

REITMAN, D. P. Uso de métodos clássicos e bayesianos em modelos de regressão beta. Dissertação de Mestrado. p.160. Universidade Federal de São Carlos, 2007.

ROBERT, C. CASELLA, G., Monte Carlo Statistical Methods, Springer; 2ª edição, 2005.

ROCHA, A., SIMAS, A., Influence diagnostics in a general class of beta regression models. Test 20, 95– 119, 2011.

ROCHA, L. A.; KHAN, A. S.; LIMA, P. V. P. S. Construindo competências tecnológicas na agricultura brasileira: articulando a política nacional de inovação com os investimentos em infraestrutura. In: ENCONTRO REGIONAL DE ECONOMIA, 18., 2013. Anais...Fortaleza, 2013.

ROSS, S.H. Stochastic Processes, Wiley Series in Probability and Mathematical Statistics, 1996.

ROSS, Sheldon. Um Primeiro Curso em Probabilidade (8ª Edição), Prentice Hall, 2010.

SANTOS, P. M. VOLTOLINI, T. V. CAVALCANTE, A C R. PEZZOPANE, J. R. M. MOURA, M. S. B. SILVA, T. G. F. BETTIOL, G. M. CRUZ, P. G. Mudanças Climáticas Globais e a Pecuária: Cenários Futuros para o Semiárido Brasileiro. Revista Brasileira de Geografia Física. v06 (2011) p.1176-1196. Disponível em: www.ufpe.br/rbgfe.

SCHWARZ, G. Estimating the dimensional of a model. *Annals of Statistics*, Hayward, v.6, n.2, p.461-464, Mar. 1978.

SILVA, A. F. A. V. Modelação do Risco de Crédito numa Carteira de Crédito ao Consumo. Dissertação. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa. 2014.

SILVA, THIERES. G. F. MOURA, M. S. B. SÁ, I. S. ZOLNIER, S. TURCO, S. H. N., JUSTINO, F. CARMO, J. F. A. SOUZA, L. S. B. Impactos das Mudanças Climáticas na Produção Leiteira do Estado de Pernambuco: Análise para os Cenários B2 e A2 do IPCC. *Revista Brasileira de Meteorologia*, v.24, n.4, 489-501, 2009.

SIMAS, A., BARRETO-SOUZA, W., ROCHA, A., Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis* 54, 348–366, 2010.

SMITHSON, M., VERKUILEN, J., A better lemonsqueezer? maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11, 54–71, 2006.

SPERA, S. A. GALFORD, G. L. COE, M. T. MACEDO, M. N. MOSTARDA, J. F. Land-Use Change Affects Water Recycling in Brazil's Last Agricultural Frontier. *Global Change Biology*, p. n/a-n/a, 2016. ISSN 1365-2486. Disponível em: <http://dx.doi.org/10.1111/gcb.13298> .

TEIXEIRA, J. C. e HESPANHOL A. N. A Trajetória Da Pecuária Bovina Brasileira. *Caderno Prudentino de Geografia, Presidente Prudente*, n.36, v.1, p.26-38, jan./jul. 2014.

TEIXEIRA, Jodenir Calixto; FREITAS, Wagner Suzano; AMORIM, Clarismundo José. A produção agropecuária e as novas tendências econômicas do Estado de Mato Grosso do Sul frente ao processo de modernização da agricultura. In: *Anais do XX ENGA – Encontro Nacional de Geografia Agrária*, Francisco Beltrão: Unioeste, 2010.

TSODIKOV, A., IBRAHIM, J. YAKOVLEV, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98(464), 1063–1078.

VIEIRA FILHO, J. E. R.; GASQUES, J. G.; SOUSA, A. G. Can Brazil feed the world? Not yet, but it has the potential! In: *THE FUTURES OF AGRICULTURE*. Brief n. 33: Global Forum on Agricultural Research (GFAR), Rome, 2012.

YAKOVLEV, A. TSODIKOV, A. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

Apêndice

Encontra-se na Figura 1, o comportamento da média e do desvio padrão para os dados do efetivo do rebanho bovino em cabeças (10^3) por municípios da região Nordeste do Brasil de 1974 a 2015. Já na Figura 2 encontram-se a evolução dos *clusters* formados de acordo com o Índice local de Moran I segundo o rebanho bovino na região Nordeste do Brasil.

Com interesse de realizar uma análise de resíduo e diagnóstico dos modelos de regressão Beta, existem algumas ferramentas gráficas que podem auxiliar na detecção de possíveis afastamentos das suposições feitas para o modelo de regressão. Entre elas, nas Figuras 3 e 4 é apresentado o gráfico da probabilidade normal com envelopes simulados. Foi utilizado o resíduo de Pearson para os modelos selecionados, pode-se ver nas Figuras 3 e 4, que os resíduos em geral, permanecem dentro dos limites de confiança dos envelopes simulados, então se designa que o modelo selecionado aparenta fornecer uma boa representação dos dados.

Ainda neste contexto, temos nas Figuras 5 e 6 os gráficos dos resíduos de Pearson versus os índices das observações. Pode-se ver que os modelos de regressão Beta selecionados aparentam estar bem ajustados, dado que apenas 2 resíduos se encontram fora do intervalo de $(-3,3)$ no tipo de *cluster* HL-NS, ou seja, do rebanho bovino que estava com *cluster* do tipo Alto-Baixo-HL e passou para Não Significativo-NS, que significa que a região Nordeste passou a estar na média do efetivo do rebanho bovino.

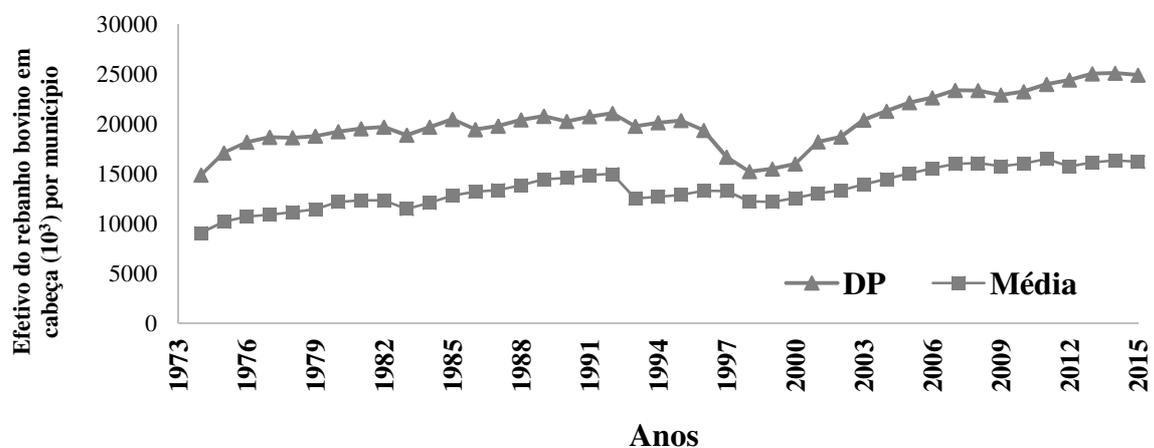


Figura 1. Distribuição da média e do desvio-padrão do efetivo do rebanho bovino em cabeças na região Nordeste e em seus municípios, de 1974 a 2015.

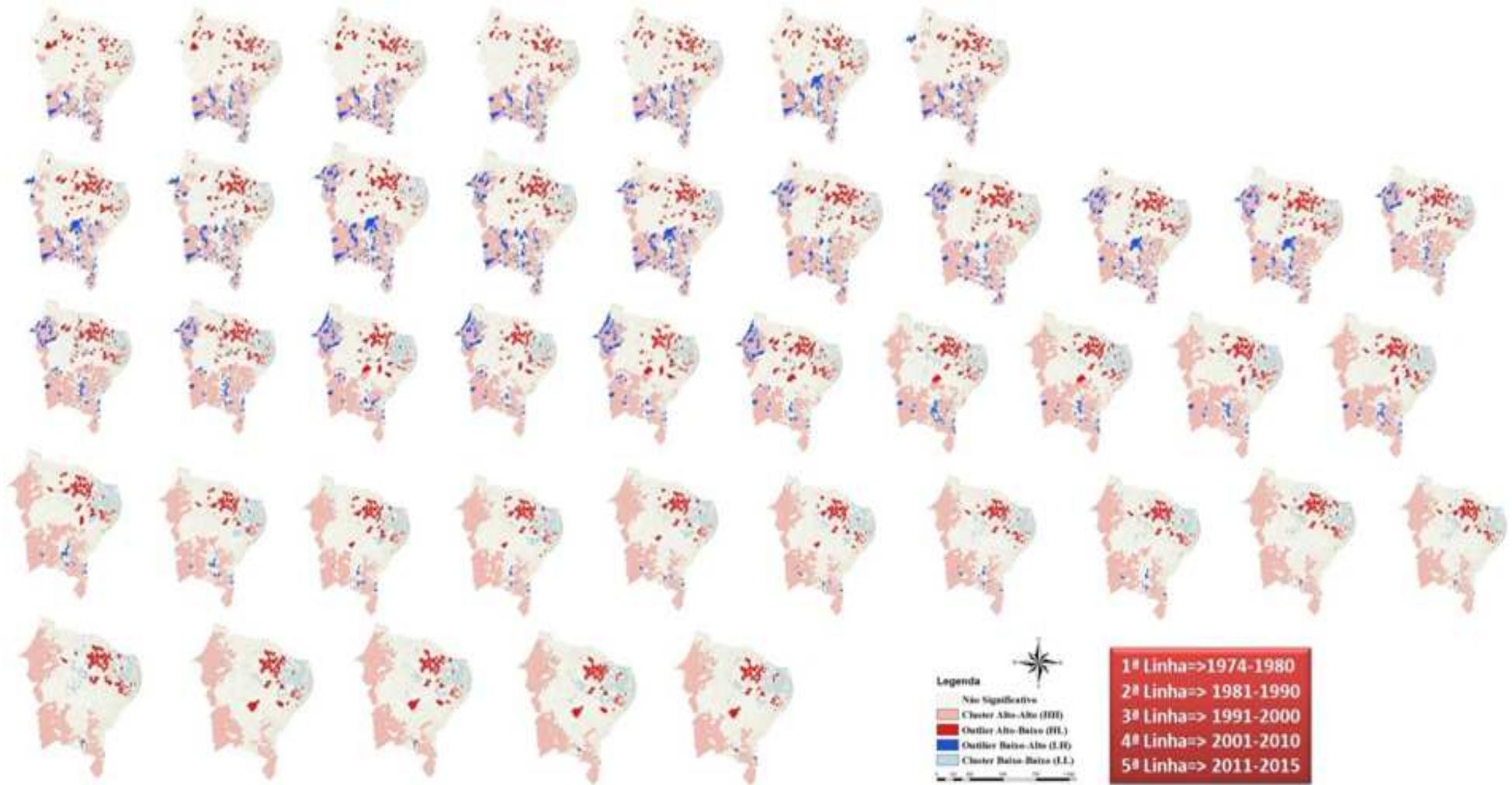


Figura 2. Clusters segundo o Índice local de Moran I para o tamanho do rebanho bovino no Nordeste do Brasil para os anos de 1974 a 2015.

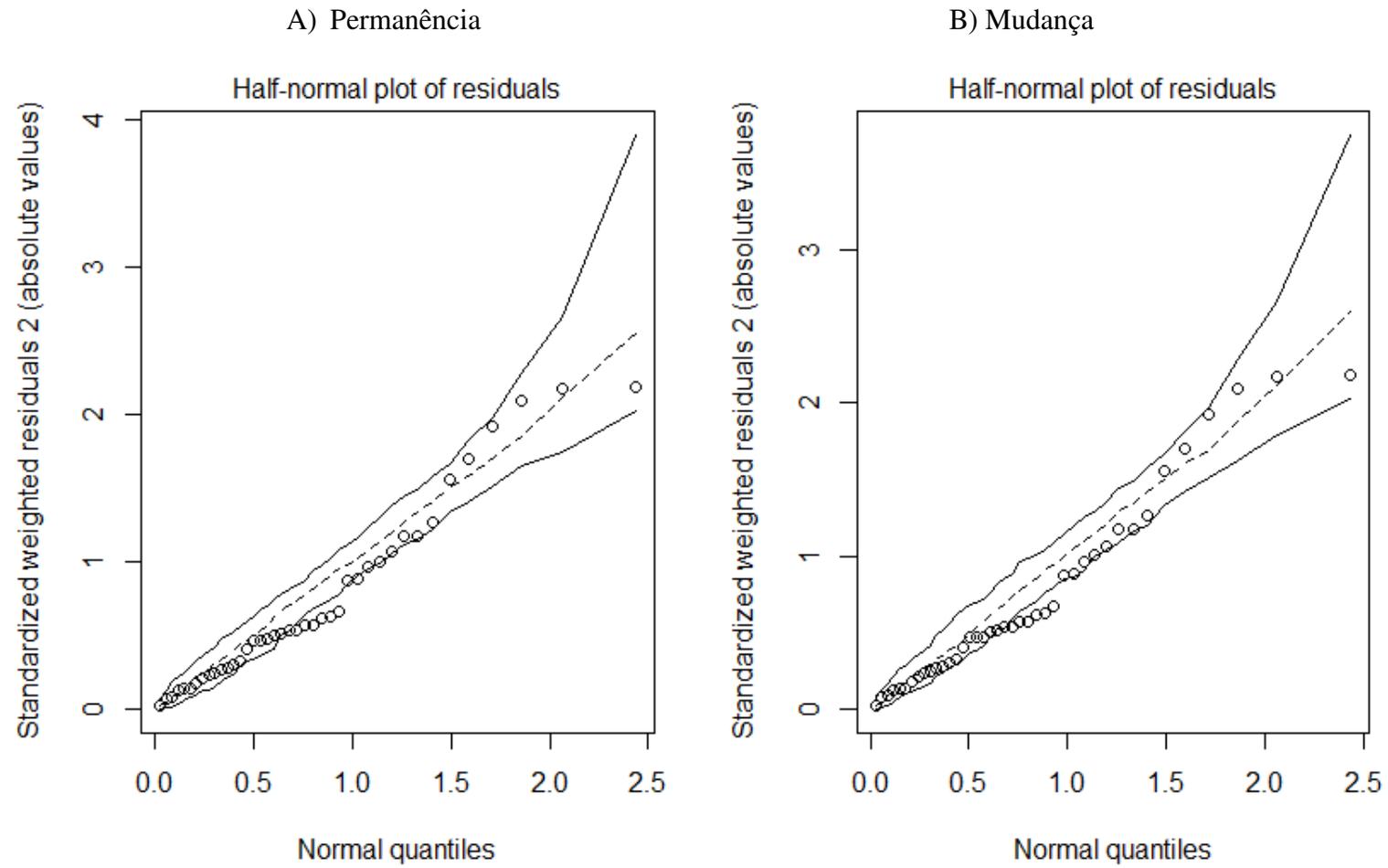


Figura 3. Gráfico da probabilidade normal com envelope simulado (resíduo Pearson) para os modelos da permanência e mudança respectivamente das categorias de *clusters*.

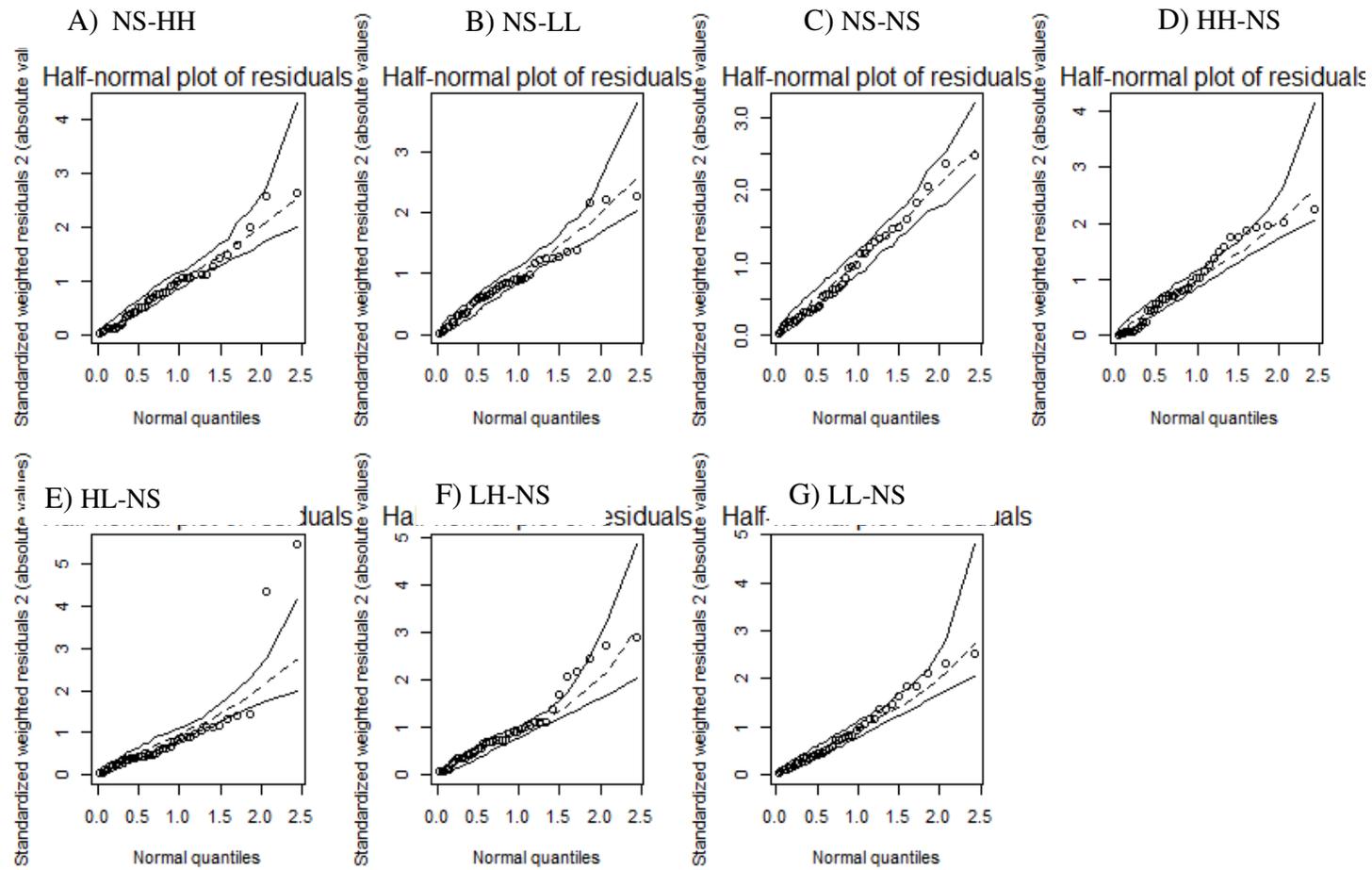


Figura 4. Gráfico da probabilidade normal com envelope simulado (resíduo Pearson) para os modelos das mudanças respectivamente das categorias de *clusters*.

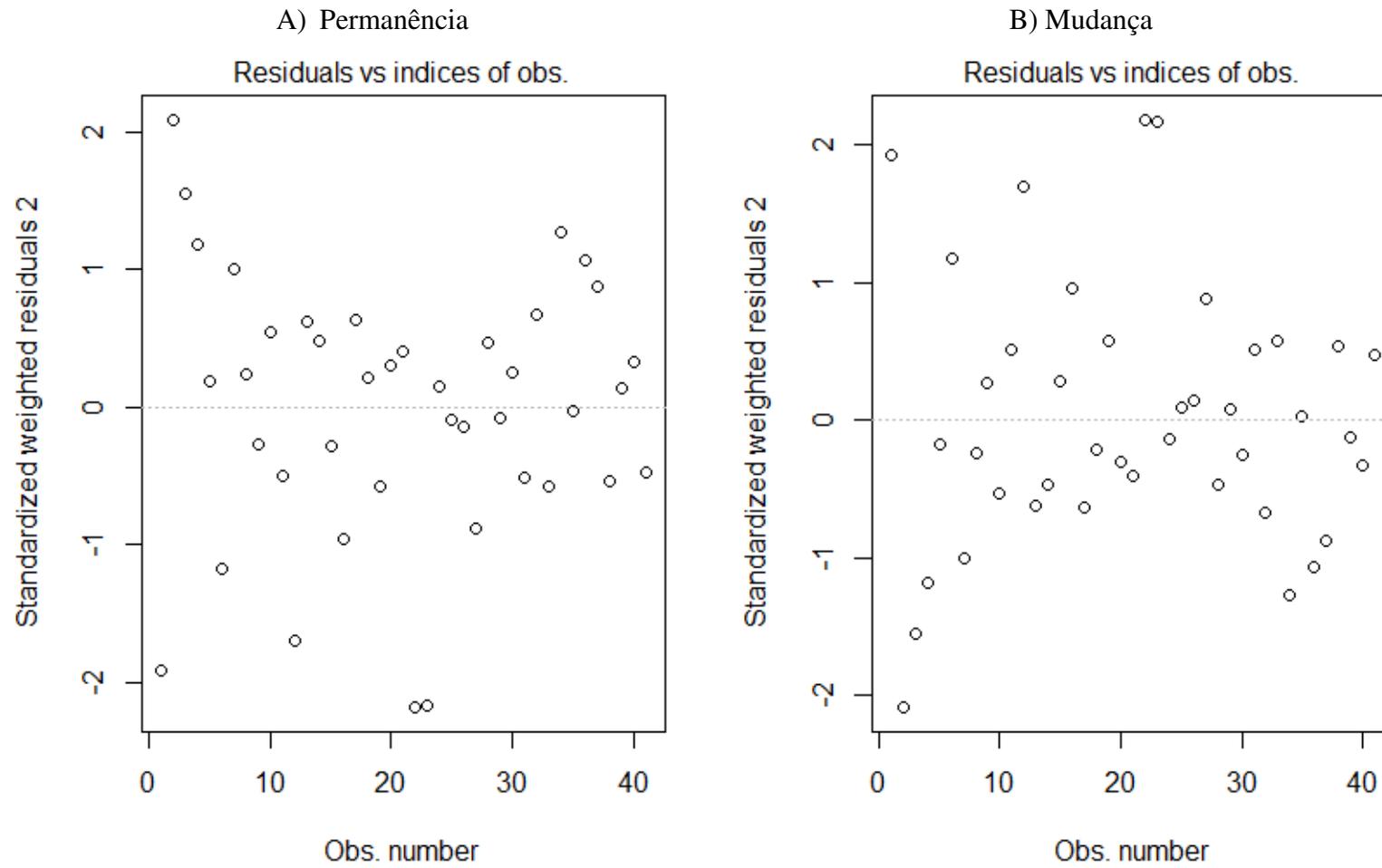


Figura 5. Gráfico de resíduo Pearson versus os índices das observações para os modelos da permanência e mudança das categorias de *clusters*.

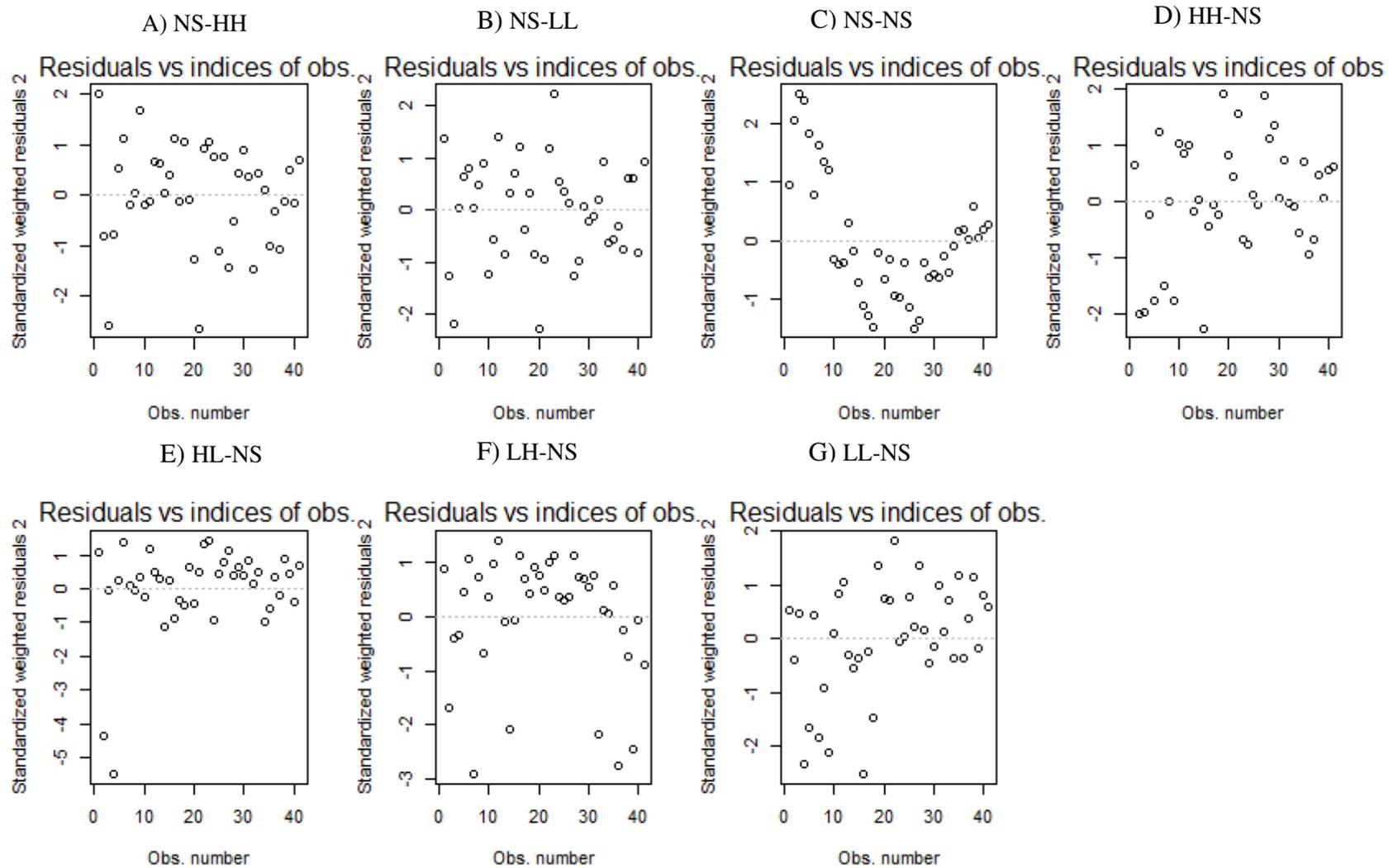


Figura 6. Gráfico dos resíduos de Pearson versus os índices das observações para os modelos dos tipos de categorias de *cluster*.