

**Instituto Brasileiro de Geografia e Estatística**  
**Escola Nacional de Ciências Estatísticas**

**UMA ABORDAGEM BAYESIANA  
PARA ESTIMAÇÃO DE POBREZA  
EM PEQUENOS DOMÍNIOS**

**Autor: Cleber Nascimento do Carmo**

RIO DE JANEIRO  
AGOSTO/2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**Cleber Nascimento do Carmo**

**UMA ABORDAGEM BAYESIANA PARA  
ESTIMAÇÃO DE POBREZA EM PEQUENOS  
DOMÍNIOS.**

Dissertação apresentada ao Programa de Pós-Graduação da Escola Nacional de Ciências Estatísticas, como parte dos requisitos necessários à obtenção do título de Mestre em Estudos Populacionais e Pesquisas Sociais.

Orientação: Suzana Marta Cavenaghi, Phd.

Co-Orientação: Fernando Antônio da Silva Moura, Phd.

RIO DE JANEIRO  
AGOSTO/2007

# UMA ABORDAGEM BAYESIANA PARA ESTIMAÇÃO DE POBREZA EM PEQUENOS DOMÍNIOS.

**Cleber Nascimento do Carmo**

Dissertação de Mestrado submetida ao Programa de Pós Graduação da Escola Nacional de Ciências Estatísticas, do Instituto Brasileiro de Geografia e Estatística, como parte dos requisitos necessários à obtenção do título de Mestre em Estudos Populacionais e Pesquisas Sociais, área de concentração em Estatística Social.

Banca Examinadora:

Prof<sup>o</sup> : \_\_\_\_\_  
Suzana M. Cavenaghi, Phd - Orientadora.  
Professora - ENCE/IBGE

Prof<sup>o</sup>: \_\_\_\_\_  
Fernando A. Moura, Phd – Co-orientador.  
Professor - IM/UFRJ

Prof<sup>o</sup>: \_\_\_\_\_  
Denise B. do Nascimento Silva, Phd.  
Professora - ENCE/IBGE

Prof<sup>o</sup>: \_\_\_\_\_  
Ismênia Blavatsky de Magalhães, Dr.  
Pesquisadora - IBGE

Prof<sup>o</sup>: \_\_\_\_\_  
Marcel de Toledo Vieira, Phd.  
Professor - UFJF

Rio de Janeiro, agosto de 2007.

### Ficha catalográfica

CARMO, CLEBER NASCIMENTO DO

Uma abordagem bayesiana para estimação de  
Pobreza em pequenos domínios [Rio de Janeiro] 2007.

XI, 129p. 29,7 cm (ENCE/IBGE, M.Sc.,  
Estudos Populacionais e Pesquisas Sociais, 2007)

Dissertação - Escola Nacional de Ciências  
Estatísticas, ENCE/IBGE

1. Estimação em pequenas áreas

I. ENCE/IBGE II. Título ( série )

# AGRADECIMENTOS

À professora Dra. Suzana Cavenaghi tenho a agradecer por ter aceitado a tarefa de me orientar, apesar da urgência dos prazos, além da paciência e atenção dispensados na realização desse trabalho.

Ao professor Dr. Fernando Moura, por aceitar fazer parte deste trabalho como co-orientador e, assim, contribuir para o amadurecimento de questões relativas ao tema.

Aos professores doutores Maurício Vasconcellos, Neide Patarra e José Eustáquio Alves, membros da banca do processo seletivo do curso de mestrado da ENCE/IBGE, por acreditarem no meu potencial.

Sou imensamente grato à Maria Beatriz Cunha e a Débora Souza pela importantíssima ajuda nos programas computacionais SAS e WinBUGS.

A Deus. À minha família, em especial à Regina, minha mãe, por ter adiado a realização de muitos de seus sonhos em benefício da realização dos meus.

Aos amigos do curso de mestrado com os quais convivi e que passaram pelo mesmo árduo processo.

À Escola Nacional de Ciências Estatísticas – ENCE/IBGE, em especial ao Programa de Mestrado em Estudos Populacionais e Pesquisas Sociais, pela formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo importantíssimo apoio financeiro durante o curso.

## RESUMO

Pesquisas por amostragem, geralmente, não fornecem estimativas em nível geográfico muito desagregado. Entretanto, há interesse em se obter estimativas para pequenas áreas, já que o tamanho amostral das pequenas áreas é pequeno. Em alguns casos, não há unidades amostrais nos domínios de interesse. O uso de estimadores diretos de expansão poderia, de maneira inadequada, apresentar grandes desvios padrão. Portanto, novas abordagens baseadas em modelos estatísticos são recomendadas para o problema de estimação em pequenas áreas. Nesta dissertação, apresentamos aplicações de abordagem estatística tanto clássica como bayesiana (baseada em métodos MCMC), para estimar proporções de pobres em pequenas áreas. Entre os diversos métodos existentes para este fim, aplicamos a metodologia proposta pelo Banco Mundial e modelos hierárquicos de regressão logística. As metodologias são aplicadas para dados socioeconômicos coletados no Estado do Rio Grande do Norte e são comparadas sob diferentes situações. Como principal resultado obtido, identificou-se que o uso de modelos hierárquicos é mais adequado para o problema de estimação da porcentagem de pobres nas microrregiões potiguares, pois permite a obtenção de estimadores com maiores taxas de acertos que os estimadores dos modelos de regressão propostos pelos pesquisadores do Banco Mundial.

Palavras-chave: Estimação em pequenas áreas; regressão logística hierárquica, MCMC.

# ABSTRACT

Sample survey designs, usually, do not provide accuracy at a low level. However, sometimes the aim is on small area estimation. In this case the sample size in each area is usually small. In some cases, there is no information in the domains of interest. The use of direct estimators such as design-based estimators would have unacceptable large standard errors. Therefore, new model-based approaches are recommended to address small area estimation problem. In this thesis, we present a classical and a full bayesian approach (based on MCMC methods) for estimating proportions of poor people in small areas. Among the several existent methods, we applied the World Bank method and hierarchical logistic regression models. The two methodologies are fitted to socioeconomic data collected in Rio Grande do Norte State and comparisons are made under different settings. As main result, it was found that hierarchical models are more adequate to potiguar microregions estimation of proportion of poor people problem because they produce estimators with bigger hit proportions than the regression models estimators proposed by the World Bank methodology group.

Keywords: Small area estimation; hierarchical logistic regression; MCMC.



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>01</b>
1.1	Objetivos	04
1.2	Organização dos capítulos posteriores	05
<b>2</b>	<b>CONCEITUAÇÃO E MENSURAÇÃO DA POBREZA</b>	<b>08</b>
2.1	A variedade do conceito de Pobreza	09
2.1.1	Pobreza Absoluta versus Relativa	09
2.2	As dimensões da Pobreza	12
2.2.1	Renda	12
2.2.2	Necessidades Básicas	13
2.2.3	Capacidades	14
2.3	A mensuração da Pobreza	15
2.3.1	Proporção de Pobres	15
2.3.2	Profundidade de Pobreza	16
2.3.3	Severidade de Pobreza	16
2.4	A mensuração da Pobreza no Brasil	17
2.5	Descrição das fontes de dados	20
2.5.1	A Pesquisa de Orçamentos Familiares – POF 2002/2003	20
2.5.2	O Censo Demográfico de 2000	24
2.6	Contextualização socioeconômica e demográfica do Rio Grande do Norte	26
2.7	As variáveis de interesse	30
2.7.1	O rendimento domiciliar per capita	31
2.7.2	O consumo domiciliar per capita	32
2.8	As variáveis explicativas	33
<b>3</b>	<b>ESTIMAÇÃO EM PEQUENAS ÁREAS</b>	<b>37</b>
3.1	Estimação da pobreza em pequenas áreas	42
3.1.2	A proposta do Banco Mundial	43
3.1.2.1	O modelo de renda domiciliar per capita (ou consumo domiciliar per capita)	44
3.1.2.2	O procedimento de implementação	46
3.2	A abordagem bayesiana	53
3.2.1	Modelos Lineares Generalizados	57
3.2.2	Modelos Hierárquicos propostos	58
3.2.3	Modelo Autoregressivo Condicional (CAR) proposto	60
<b>4</b>	<b>APLICAÇÃO DOS MÉTODOS ESTATÍSTICOS</b>	<b>65</b>
4.1	Modelagem sob a abordagem clássica	65
4.2	Modelagem sob a abordagem bayesiana	74
<b>5</b>	<b>CONCLUSÕES E RECOMENDAÇÕES FINAIS</b>	<b>91</b>
5.1	Conclusões	91
5.2	Recomendações para trabalhos futuros	93
	<b>Referências bibliográficas</b>	<b>96</b>
	<b>Anexos</b>	<b>102</b>

## Lista de Gráficos e Figuras

<b>Figura 2.1:</b> Distribuição da variável Renda per capita. Rio Grande do Norte, 2000	28
<b>Gráfico 2.1:</b> Distribuição da variável Anos de Estudo do Chefe do Domicílio por classes de anos, Rio Grande do Norte, 2000 -----	29
<b>Figura 3.1:</b> Esquema de seleção das variáveis explicativas para o ajuste do modelo -----	49
<b>Gráfico 4.1:</b> Resíduos dos modelos de regressão para as variáveis dependentes, logaritmo da renda e logaritmo do consumo, respectivamente, Rio Grande do Norte -----	68
<b>Figura 4.1:</b> Distribuição espacial da proporção de pobres, com base no rendimento domiciliar per capita, Método do Banco Mundial. Rio Grande do Norte, 2003 -----	72
<b>Figura 4.2:</b> Distribuição espacial da proporção de pobres, com base no consumo domiciliar per capita, Método do Banco Mundial. Rio Grande do Norte, 2003 -----	73
<b>Figura 4.3:</b> Distribuição espacial da proporção de pobres, com base no rendimento domiciliar per capita, Enfoque bayesiano. Rio Grande do Norte, 2003	88
<b>Figura 4.4:</b> Distribuição espacial da proporção de pobres, com base no consumo domiciliar per capita, Enfoque bayesiano. Rio Grande do Norte, 2003 -----	88
<b>Gráfico 4.2:</b> Distribuição das diferenças entre as taxas de acerto nos modelos hierárquicos e o modelo proposto pelo Banco Mundial, para o logaritmo da renda domiciliar per capita e logaritmo do consumo domiciliar per capita, respectivamente -----	90

## Lista de Tabelas e Quadros

<b>Tabela 2.1:</b> Número de setores censitários e domicílios pesquisados na POF com entrevista completa no Rio Grande do Norte, por microrregião -----	23
<b>Tabela 2.2:</b> Número de estratos, setores censitários e domicílios pesquisados na POF com entrevista completa no Rio Grande do Norte, por microrregião e tipo de área -----	24
<b>Tabela 2.3:</b> Número de setores censitários e domicílios pesquisados no Censo e na POF com entrevista completa no Rio Grande do Norte, por microrregião -----	25
<b>Tabela 2.4:</b> Distribuição dos municípios, segundo faixa de população. Rio Grande do Norte, 2000 -----	27
<b>Tabela 2.5:</b> Domicílios particulares permanentes, segundo algumas características domiciliares. Rio Grande do Norte, 2000 -----	30
<b>Quadro 2.1:</b> Lista de perguntas comuns nas duas pesquisas -----	34
<b>Quadro 2.2:</b> Lista de perguntas comuns consideradas compatíveis pelo teste de igualdade de proporções -----	36
<b>Tabela 3.1:</b> Resultados de R <sup>2</sup> ajustado das aplicações da metodologia do Banco Mundial em alguns países -----	44
<b>Tabela 4.1:</b> Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da renda domiciliar per capita -----	67
<b>Tabela 4.2:</b> Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo do consumo domiciliar per capita -----	67
<b>Tabela 4.3:</b> R <sup>2</sup> dos modelos para o logaritmo da renda e do consumo -----	69
<b>Tabela 4.4:</b> Distribuição dos 5 números da proporção de pobres, baseados no RDPC imputado, Rio Grande do Norte e Microrregiões, 2003 -----	70
<b>Tabela 4.5:</b> Distribuição dos 5 números da proporção de pobres, baseados no CDPC imputado, Rio Grande do Norte e Microrregiões, 2003 -----	71
<b>Tabela 4.6:</b> Proporção de domicílios pobres, com base nas variáveis originais, POF 2002/2003 e Censo 2000 – Rio Grande do Norte e Microrregiões -----	74
<b>Tabela 4.7:</b> Parâmetros de regressão estimados a partir de 15000 amostras geradas via amostrador de Gibbs, Rendimento domiciliar per capita -----	77
<b>Tabela 4.8:</b> Parâmetros de regressão estimados a partir de 15000 amostras geradas via amostrador de Gibbs, Consumo domiciliar per capita -----	78

<b>Tabela 4.9:</b> Distribuição das estimativas de proporção de pobres, baseadas no rendimento domiciliar per capita imputado na amostra do censo (Modelo Hierárquico). Microrregiões do Rio Grande do Norte, 2003 -----	79
<b>Tabela 4.10:</b> Distribuição das estimativas de proporção de pobres, baseadas no consumo domiciliar per capita imputado na amostra do censo (Modelo Hierárquico). Microrregiões do Rio Grande do Norte, 2003 -----	80
<b>Tabela 4.11:</b> DIC dos modelos hierárquicos -----	81
<b>Tabela 4.12:</b> Parâmetros de regressão estimados a partir de 15000 amostras, Rendimento domiciliar per capita. (modelo CAR) -----	83
<b>Tabela 4.13:</b> Parâmetros de regressão estimados a partir de 15000 amostras, Consumo domiciliar per capita (modelo CAR) -----	84
<b>Tabela 4.14:</b> Distribuição das estimativas de proporção de pobres, baseadas no rendimento domiciliar per capita imputado na amostra do censo (modelo CAR). Microrregiões do Rio Grande do Norte, 2003 -----	85
<b>Tabela 4.15:</b> Distribuição das estimativas de proporção de pobres, baseadas no consumo domiciliar per capita imputado na amostra do censo (modelo CAR). Microrregiões do Rio Grande do Norte, 2003 -----	86
<b>Tabela 4.16:</b> DIC dos modelos CAR -----	86
<b>Tabela 4.17:</b> DIC dos modelos sob o enfoque bayesiano -----	87
<b>Tabela 4.18:</b> Proporção de acerto nos modelos ajustados na POF -----	89

## Lista de Siglas

ENCE	Escola Nacional de Ciências Estatísticas.
IBGE	Instituto Brasileiro de Geografia e Estatística.
UFJF	Universidade Federal de Juiz de Fora.
UFRJ	Universidade Federal do Rio de Janeiro.
POF	Pesquisa de Orçamentos Familiares.
PNAD	Pesquisa Nacional por Amostras de Domicílios.
PNUD	Programa das Nações Unidas para o Desenvolvimento.
MCMC	Método de Monte Carlo via Cadeias de Markov.
RDPC	Rendimento domiciliar per capita.
CDPC	Consumo domiciliar per capita.
CAR	Condicional autoregressivo.
ONU	Organização das Nações Unidas.
ONG	Organização Não Governamental
IDH-M	Índice de Desenvolvimento Humano Municipal

# CAPÍTULO 1

## INTRODUÇÃO

Muitos pesquisadores têm se interessado pelo estudo da pobreza, ocasionando uma multiplicidade de opiniões e avaliações objetivas e subjetivas. O debate sobre pobreza está constantemente na pauta de instituições como o Banco Mundial, a ONU, governos e ONGs. A literatura brasileira sobre o tema contém inúmeros estudos documentando a incidência da pobreza com base em diversas variáveis socioeconômicas, tais como: insuficiência de renda, más condições de saúde, desemprego etc. (por exemplo, ROCHA, 2003; MONTEIRO, 2003; FERREIRA *et al.*, 2003).

A mensuração da pobreza, entretanto, depende do conjunto de dados disponível e do propósito do estudo. A maneira mais utilizada atualmente para calcular quantidade de indivíduos pobres é, certamente, com base numa medida de proporção. E, geralmente, essa proporção refere-se a uma linha de pobreza baseada na insuficiência de rendimentos. Neste contexto, ROCHA (2004) destaca que a insuficiência de rendimentos é uma boa *proxy* para investigar a incidência de pobreza e que o bem-estar familiar está intimamente ligado ao nível de sua renda familiar.

E, independente das áreas geográficas ou medidas de mensuração de pobreza consideradas nos estudos realizados, há interesse em estender suas análises e resultados para níveis administrativos locais. O governo brasileiro tem interesse nisso, por exemplo, pela atual legislação que define a distribuição de verbas federais a partir de critérios populacionais e, por outro lado, pela necessidade das autoridades locais em obter um quadro mais fidedigno e atualizado da sua área de jurisdição. Adicionalmente,

governos estaduais e federal também precisam de informações referentes a um nível geográfico mais desagregado para elaboração de políticas públicas locais.

Além disso, há também uma preocupação sobre as diferenças da pobreza ao longo das regiões geográficas. Assim, métodos estatísticos para dados observados ao longo de uma região geográfica assumem cada vez mais importância no planejamento de políticas públicas locais. Tais métodos são usados para capturar a possível associação espacial existente entre as localidades e para fazer previsões em áreas sem observações disponíveis. Em estudos de dados de área, assume-se que as observações não são independentes ao longo da região sob estudo (BAILEY & GATRELL, 1995).

Pode-se dizer que uma dificuldade para a obtenção de estimativas para pequenas áreas geográficas é que, em muitos casos, o tamanho amostral resultante de uma pesquisa de cobertura nacional é muito pequeno para que estimativas baseadas somente no plano amostral apresentem precisão adequada. Outro problema pode ser também a inexistência de unidades amostrais nas regiões de interesse.

De acordo com ASSUNÇÃO et. al. (1998), uma área geográfica é chamada de pequena área (ou pequeno domínio) quando a variabilidade de estimativas utilizando dados apenas desta área produz desvios padrão inaceitáveis pelo seu enorme tamanho. Os autores continuam afirmando que isso é devido à pequena quantidade de pessoas em risco na área de estudo, em relação à raridade dos eventos do fenômeno em questão. Assim, de posse de uma mesma base de dados, uma área pode ser considerada pequena em certos casos e não, em outros. Cabe ainda ressaltar que o termo “área” não é aplicado apenas às áreas geográficas, mas também a qualquer domínio de interesse. Outros termos utilizados para descrição de estimadores de pequenas áreas incluem: estimadores de pequenos domínios (SÄRNDAL, 1984); estimadores para pequenos

grupos (HOLT et. al., 1979), entre outros. Neste texto, os termos “pequenas áreas” e “pequenos domínios” são usados sem distinção.

O problema de estimação em pequenos domínios tem recebido muita atenção nos últimos anos devido à necessidade de produção de estimativas para pequenas áreas administrativas, ao avanço de técnicas estatísticas (SCHMIDT et. al.) e, também, devido à disponibilidade de computadores pessoais mais velozes, com maior capacidade de processamento e relativamente baratos.

Há, na literatura estatística, diversas propostas de métodos estatísticos para estimação em pequenos domínios. Especificamente para o caso da pobreza, destaca-se o método proposto por ELBERS et al. (2000), que vem sendo aplicado em diversos países, dentre eles o Brasil, para estimação de medidas de pobreza. Uma deficiência constatada nas aplicações desta proposta é que sua capacidade preditiva, representada pelo coeficiente de determinação ou de explicação não é elevada.

Recentemente, CUNHA (2006) aplicou este procedimento para estimar a proporção de pobres no Estado do Rio Grande do Norte, região nordeste do Brasil, no nível de microrregiões, encontrando coeficientes de determinação moderados, com valores de 61% e 64%, aproximadamente, para as variáveis consumo domiciliar per capita e renda domiciliar per capita, respectivamente. A autora procurou, também, alternativas para aumentar o poder de predição dos modelos. Especificamente, utilizando modelos hierárquicos sob o enfoque da Estatística Clássica com três níveis hierárquicos: domicílios, município e microrregião; encontrando estimativas mais elevadas para a proporção de pobres.

Nesta dissertação, propomos a aplicação de métodos estatísticos sob a abordagem bayesiana para a estimação da proporção de pobres nas microrregiões potiguares, a saber: modelos hierárquicos de regressão logística e modelos CAR. Para



tal, antes replicamos o método proposto pelos pesquisadores do Banco Mundial para fins comparativos dos resultados. A unidade geográfica utilizada foi microrregião. Apesar de não ser a unidade geográfica mais natural quando se estuda o problema de estimação em pequenas áreas, esta unidade é utilizada aqui para que se mantenha uma comparabilidade de resultados com o trabalho desenvolvido por CUNHA (2006). Em sua justificativa, a autora diz que não considerou o problema de estimação em nível municipal porque tinha indícios de que a qualidade das estimações seria baixa devido à pequena população de muitos desses municípios.

O primeiro método aplicado é o procedimento proposto pelos pesquisadores do Banco Mundial (ELBERS et. al. 2000, 2001 e 2002). E no sentido de tentarmos aumentar a capacidade preditiva dos modelos, aplicamos métodos estatísticos sob a abordagem bayesiana, na qual a estrutura hierárquica dos dados é considerada por meio de modelos hierárquicos de regressão logística e os efeitos espaciais analisados por meio de modelos CAR. Com a inclusão de efeitos espaciais, assumimos implicitamente que microrregiões geograficamente próximas têm probabilidades de incidência mais similares do que as mais distantes.

## **1.1 Objetivos**

A perspectiva geral deste estudo é explorar, à luz do referencial teórico, técnicas estatísticas, sob a abordagem bayesiana, para estimar a proporção de pobres em pequenas áreas para anos não censitários. Mais especificamente, em nível de microrregiões e para o ano de 2003, tendo como base os microdados da parte amostral do Censo Demográfico de 2000 e os microdados da Pesquisa de Orçamentos Familiares 2002/2003. Para o ajuste dos modelos estatísticos são utilizados dados do Estado do Rio

Grande do Norte. Ainda fazem parte do escopo deste trabalho os seguintes objetivos específicos:

- i) Traçar, com base nas informações contidas nos dados do Censo Demográfico de 2000 e na Pesquisa de Orçamentos Familiares 2002/2003, um perfil das condições socioeconômicas do Estado do Rio Grande do Norte;
- ii) Aplicar o procedimento proposto pelos pesquisadores do Banco Mundial para mapear a pobreza daquele Estado em nível de microrregião;
- iii) Estimar a proporção de pobres nas microrregiões potiguares para o ano de 2003, usando a abordagem bayesiana. Detalhadamente, aplicamos modelos de regressão logística hierárquicos e modelos autoregressivos condicionais (modelos CAR).

## **1.2 Organização dos capítulos posteriores**

A seguir, é feita uma descrição dos capítulos deste trabalho. Na primeira seção do segundo capítulo, analisamos o fenômeno da pobreza, expondo a variedade de conceitos existentes, destacando-se limitações existentes para cada um deles. Na segunda seção, focalizamos as diferentes dimensões da pobreza, ressaltando limitações de todas as diferentes abordagens. Na seqüência, as fontes de dados e os indicadores utilizados, neste trabalho, para mensuração da pobreza são apresentados. Por último, é feita uma breve descrição das condições socioeconômicas do Estado do Rio Grande do Norte com o propósito de mostrar o quão pobre esse Estado é, a partir de diversas variáveis, tais como: renda, escolaridade e condições de habitação.

No terceiro capítulo, é feita uma breve revisão da literatura existente sobre o problema de estimação em pequenas áreas, tanto sob a abordagem da Estatística clássica como da estatística bayesiana. Os estudos abordados tratam desde o estimador sintético proposto por GONZALES (1973) até o uso de modelos hierárquicos, com variáveis contínuas (MOURA e HOLT, 1999) e para o caso de variáveis categorizadas em que a variável de interesse é uma proporção (MOURA e MIGON, 2002). Dando seqüência, apresentamos, especificamente, os métodos utilizados para estimação dos indicadores de pobreza nas pequenas áreas em questão (microrregiões). Primeiramente, a metodologia proposta pelos pesquisadores do Banco Mundial, que utiliza métodos da estatística clássica, é apresentada. Posteriormente, é descrito o procedimento da abordagem bayesiana, fazendo também uma breve revisão dos métodos computacionais de Monte Carlo via Cadeias de Markov (MCMC) utilizados para obtenção de uma amostra da distribuição a posteriori dos parâmetros dos modelos. Quando o cálculo analítico das distribuições a posteriori não é possível, utilizamos esses métodos computacionais para retirar amostras dessas distribuições. Na seqüência, fazemos uma apresentação dos modelos lineares generalizados, uma vez que os modelos de regressão logística são membros dessa classe de modelos e, posteriormente, dos modelos hierárquicos. Mais especificamente, modelos hierárquicos de regressão logística com intercepto aleatório são utilizados como ferramenta estatística na estimação da proporção de indivíduos pobres. Por fim, descrevemos também os modelos autoregressivos condicionais (conhecidos também como Modelos CAR).

Já no quarto capítulo, tratamos da investigação empírica. O primeiro tópico é destinado à aplicação e apresentação dos resultados dos métodos propostos pelos pesquisadores do Banco Mundial e aplicados recentemente por CUNHA (2006) no Estado

do Rio Grande do Norte, região Nordeste do Brasil. Em seguida, prossegue-se a análise com a aplicação e os resultados dos métodos sob abordagem bayesiana.

Finalmente, o quinto capítulo é destinado às conclusões, considerações finais e sugestões de trabalhos futuros.

## **CAPÍTULO 2**

# **CONCEITUAÇÃO E MENSURAÇÃO DA POBREZA**

A redução e erradicação da pobreza têm sido um tema central em diversas instituições, como governos e ONGs. Os programas de combate à pobreza se multiplicam e com eles as análises dos níveis e da composição da pobreza. Entretanto, a teoria por trás do tema é bastante heterogênea. Acredita-se que entender a heterogeneidade da pobreza é requisito para a formulação e implantação de um conjunto mais personalizado de políticas públicas. Neste capítulo, apresentamos e discutimos essas diferenças e comentamos algumas de suas implicações na orientação de políticas públicas de combate à pobreza. Na seqüência, descrevemos as variáveis respostas usadas para mensurar pobreza, as bases de dados utilizadas e uma breve descrição das condições socioeconômicas do Estado do Rio Grande do Norte.

No caso do Brasil, especificamente, o problema da pobreza tornou-se tema particular de preocupação durante os últimos quinze anos, sob um contexto de uma frustração com a persistência desse fenômeno. Frequentemente, o assunto é tratado pela mídia com afirmativas sobre o agravamento da pobreza e referências a melhores condições de vida quando o país era essencialmente rural (ROCHA, 2004). Com o intuito de reduzir o quadro de pobreza do país, diversos programas sociais foram criados pelo governo brasileiro ao longo dos anos. No presente texto, é utilizado o critério do programa Bolsa-Família para classificação dos indivíduos como pobres, o qual é um programa social de auxílio financeiro que, em 2003, visava atender aquelas famílias com renda de até R\$100,00 *per capita* mensais.

Não há um conceito único para a pobreza na literatura especializada. Sendo assim, é comum a existência de diversos trabalhos apresentando resultados distintos acerca desse mesmo tema, dadas as diferenças na metodologia utilizada. Além disso, conceituar pobreza torna-se mais difícil por ser um conceito “em deslocamento” dentro de uma sociedade. A seguir, sintetizam-se alguns desses conceitos.

## **2.1 A variedade do conceito de Pobreza**

### **2.1.1 Pobreza Absoluta versus Relativa**

Entre os diversos conceitos de pobreza existentes, citam-se os conceitos de pobreza absoluta e pobreza relativa, presentes na literatura há séculos.

Mais recentemente, de acordo com SALLES & TUIRÁN (2000), o conceito de pobreza absoluta diz respeito à determinação de um nível mínimo de renda capaz de satisfazer um conjunto de necessidades básicas para a manutenção de um nível mínimo de condições de vida, sendo classificados como pobres aqueles com uma renda abaixo desse limite. Essa renda mínima é chamada de linha de pobreza, uma vez que ela atua como ferramenta discriminatória para pobres e não-pobres. Cabe ressaltar que, apesar da facilidade operacional desse termo, o cálculo do valor a ser utilizado como linha de pobreza não é nada fácil. Isso porque a escolha desse parâmetro passa por uma necessidade de inclusão de subjetividade ao problema, pois não há um consenso sobre o que é indispensável para atender as necessidades básicas de um indivíduo. Com o intuito de adoção de um critério mais concreto, foram desenvolvidas algumas soluções para a determinação de uma linha de pobreza: como requisitos alimentares, necessidades básicas e renda.

A mais básica de todas as necessidades do ser humano é, sem dúvida, a alimentação. Sem alimentar-se, a pessoa não consegue sequer sobreviver. Assim, o enfoque dos requisitos alimentares enfatiza, como o próprio nome sugere, o atendimento dos requisitos energéticos, partindo do princípio de que existe uma forte correlação entre o estado nutricional do indivíduo e fatores socioeconômicos, de acordo com LUSTOSA e FIGUEIREDO (1989). Os autores ainda afirmam que uma família só deixará de suprir os requerimentos energéticos do conjunto de seus entes ou membros familiares quando já não puder mais retirar de outros itens do orçamento doméstico, como a habitação, vestuário, educação, etc. Todavia, segundo ROCHA (1988) esse pressuposto não pode ser tomado como verdade absoluta, pois, de acordo com a autora, deve ser considerada também a existência de diferentes custos e preferências com alimentação entre pessoas de diferentes faixas de renda. A autora sugere que se trabalhe com a estrutura de consumo dos mais pobres.

No caso das necessidades básicas, a identificação da pobreza é dada de forma mais abrangente, envolvendo outros itens considerados indispensáveis além de alimentação. Porém, LUSTOSA e FIGUEIREDO (1989) relatam que uma dificuldade, neste caso, seria identificar quais itens deveriam constar na lista de necessidades básicas a serem analisadas.

Já a abordagem da renda, representada pelo salário mínimo, possui como premissa a idéia de que o salário mínimo serve aos propósitos da sua elaboração, segundo a qual este deveria se constituir da “remuneração mínima devida a todo trabalhador adulto, sem distinção de sexo, por dia normal de serviço, e capaz de satisfazer, em determinada época e região do país, às suas necessidades normais de alimentação, habitação, vestuário, higiene e transporte” (Decreto-lei 399, de 30 de abril de 1938, apud ROCHA, 1988). Contudo, a mesma autora no mesmo trabalho diz que

“mesmo nas suas origens, houve um descolamento do valor do salário mínimo do custo efetivo de atendimento das necessidades básicas, já que seus valores resultavam de acordos (...), representando uma solução politicamente possível, mas não o valor de uma cesta básica” (ROCHA, 1988, p.2).

De acordo com MOTTA (2000), um problema com relação à utilização da medida de pobreza absoluta refere-se ao fato de que esta não levaria em conta as particularidades de cada região em relação às diversas características como cultura, condições socioeconômicas e desigualdade de renda. A autora ainda relata que isso levaria a uma sobrestimação ou subestimação da quantidade de pobres em decorrência da utilização de um único critério para determinar o que deve entrar no cômputo da linha de pobreza em diferentes localidades. Para isso teríamos que considerar o problema da aplicação da política pública com base nesse estudo, além do problema de mensurar pobreza, já que estaríamos utilizando critérios inapropriados em termos da composição e da quantidade de pobres com relação ao padrão de vida da região sob estudo.

No que se refere à pobreza relativa, não é definido um pacote mínimo de necessidades a serem cumpridas, mas um conjunto médio ou mediano de tais necessidades. Pode-se dizer que ela define o indivíduo pobre como sendo o indivíduo pertencente às camadas inferiores da distribuição de renda, como por exemplo, os 20% mais pobres da população. Adicionalmente, pode-se dizer que a pobreza relativa é fruto de uma comparação de um indivíduo com outros, dentro de uma mesma região.

Para SEN (1983), há um consenso na utilização da pobreza relativa em países ricos, porque a proporção de população em condição de pobreza seria muito pequena, sendo a desigualdade um melhor indicador para a pobreza. O mesmo autor, entretanto, destaca que há uma confusão em torno dos conceitos de pobreza e desigualdade, já que uma sociedade pode apresentar distribuição de renda igualitária,



mas composta apenas por indivíduos pobres. O contrário também poderia acontecer. A noção de pobreza relativa deveria ser utilizada apenas para medir o padrão de distribuição de renda de uma sociedade, e não seu grau ou intensidade de pobreza (MOTTA, 2000; ROMÃO, 1993).

## **2.2 As dimensões da Pobreza**

Antes de abordar as diferentes maneiras possíveis de se mensurar a pobreza, é necessário definir um outro conceito geralmente confundido com a pobreza: a fome. Para MACHADO (2006), a fome é caracterizada pela ingestão insuficiente de alimentos necessários para a subsistência. Apesar de todos os indivíduos que passam fome sejam considerados como pobres, nem todos os pobres passam fome. A fome seria apenas um aspecto da pobreza, dentre outros, que serão discutidos a seguir.

### **2.2.1 Renda**

Esta abordagem mensura pobreza com base na carência de rendimentos, dada uma determinada linha de pobreza. As linhas de pobreza podem ser estabelecidas a partir de diversos critérios, que variam de salários mínimos até linhas formuladas com base em nutrientes necessários para a manutenção de uma determinada condição nutricional. SCHWARTZMAN (2006) afirma que não há uma linha de pobreza oficial no Brasil. O autor ainda comenta que esse cálculo varia entre regiões, estados e áreas urbana, rural e metropolitana, e depende de uma série de suposições e estimativas sobre custos e padrões de consumo das diferentes populações.

Cabe ressaltar que apesar da possibilidade de cálculo e sua utilização no cotidiano, não é fácil medir a pobreza através da renda. As inúmeras metodologias existentes de cálculo e a própria definição de uma linha de pobreza torna arbitrária a sua determinação e nos fornece uma imensa quantidade de estimativas, mesmo fixando-se a população e o período de tempo em estudo.

ROCHA (2003) destaca uma série de problemas com a variável renda na sua utilização como medida de pobreza, mas reconhece que ela é a melhor aproximação de bem-estar do indivíduo devido a forte correlação que há entre a renda e indicadores de bem-estar físico. Adicionalmente, MATTOS e WAQUIL (2006) afirmam que o amplo uso dessa abordagem refere-se à sua adequação ao postulado da microeconomia de maximização da utilidade pelos consumidores. Segundo os autores, o amplo uso dessa abordagem é devido à existência de uma certa renda mínima a todos os indivíduos e à adequação desse método como aproximação do bem-estar e da pobreza e sua possibilidade de cálculo através da base de dados disponível.

## **2.2.2 Necessidades Básicas**

A abordagem de mensuração de pobreza com base na privação de necessidades básicas passou a ser utilizada por organismos internacionais a partir da década de 70 e tem como principal vantagem a retirada da renda como elemento central para uma abordagem multidimensional da pobreza (ROCHA, 2004). Sob essa abordagem, é considerado pobre aquele indivíduo que não tem acesso a certas necessidades básicas, como habitação, esgotamento sanitário, água encanada e educação. Uma pergunta que se faz ao utilizar essa abordagem é: quais indicadores devem ser utilizados? A mesma autora relata que os itens alimentação correta, boas

condições sanitárias e educação sempre estão entre os indicadores selecionados. MACHADO (2006) complementa dizendo que outros itens, como acesso a bens públicos, coleta de lixo e transporte, seriam utilizados de acordo com a amplitude, objeto de estudo e conveniência do pesquisador.

### **2.2.3 Capacidades**

A maior parte dos estudos sobre pobreza dá ênfase essencialmente a apenas uma dimensão; não analisando, portanto, a importância de outros fatores. Sob o contexto da abordagem das capacitações, Amartya Sen, pioneiro e maior expoente desse enfoque, afirma que “a utilidade da riqueza está nas coisas que ela nos permite fazer – as liberdades substantivas que ela nos ajuda a obter...” (SEN, 2000, p.28). Segundo o autor, a pobreza pode ser definida como privação de capacidades de ser e fazer aquilo que valorizam. Porém, é destacado ainda que definir quais informações são – ou não são – consideradas relevantes depende da base informacional disponível. Nessa abordagem, são levados em consideração elementos do indivíduo desde estar devidamente nutrido, estar saudável, até características mais complexas, tais como ser feliz, ter respeito próprio, tomar parte na vida da sociedade, e assim por diante (SEN, 2001, p.79).

Mas de acordo com MACHADO (2006), o estabelecimento de uma linha de pobreza nessa abordagem também seria problemático, pois ou se identifica cada capacidade individualmente ou se estabelece um indicador agregado das capacidades. O autor exemplifica que o Índice de Desenvolvimento Humano (IDH) e o Índice de Pobreza Humana (IPH) partiram desse ponto, mas a agregação das variáveis seria um problema, e não deixaria de ser arbitrária. Adicionalmente, MATTOS e WAQUIL (2006) afirmam que

não existe, até então, nenhum método consolidado que operacionalize os conceitos complexos dos quais se valem esta abordagem.

## **2.3 A mensuração da Pobreza**

Depois da apresentação e descrição de diferentes formas de conceituação e de dimensionar a pobreza, pode-se proceder a uma quantificação do número de pobres e à construção de um índice de pobreza. De acordo com ANGOLA (2005), tendo por referência a linha de pobreza, podem-se calcular três indicadores que servem para medir a pobreza na população: (i) Proporção de Pobres, (ii) Profundidade da Pobreza e (iii) Severidade da Pobreza. A seguir, apresentamos uma breve descrição de cada um deles, com detalhes para o indicador de proporção de pobres – escolhido para mensurar a pobreza nesta dissertação.

### **2.3.1 Proporção de Pobres**

Pode-se ter uma noção da extensão da pobreza com o quociente do número de pobres pelo número de pessoas da localidade em questão. Este indicador representa a proporção da população que se encontra em condição de pobreza. Alguns autores também o designam de índice de incidência de pobreza (VIEIRA, 2005).

Os pobres são identificados pela sua renda, por exemplo, sendo considerados pobres aqueles indivíduos com rendas inferiores à linha de pobreza estabelecida. Uma limitação deste indicador é não oferecer nenhuma indicação da intensidade ou profundidade da pobreza. Mas apesar disso, a proporção de pobres é um indicador muito

utilizado, pois seu cálculo é muito simples e, por isso, fácil de compreender. O indicador de proporção de pobres nos dá uma idéia da extensão da pobreza.

A seguir, os demais indicadores são apresentados, de maneira sucinta.

### **2.3.2 Profundidade de Pobreza**

O índice de profundidade da pobreza é a distância média, em percentagem, entre a condição de pobreza do indivíduo e a linha de pobreza estabelecida. No caso, por exemplo, de uma linha de pobreza identificada pelo nível de consumo, o índice de profundidade é o mesmo que a diferença média entre os níveis de consumo dos pobres e a linha de pobreza, multiplicada pela incidência de pobreza. De acordo com ANGOLA (2005), com este indicador é possível observar mudanças de rendimento dos pobres, traduzido como afastamento ou aproximação da linha de pobreza, mesmo quando o indicador de proporção de pobres permanece inalterado.

### **2.3.3 Severidade de Pobreza**

O índice de profundidade da pobreza ao quadrado é a média do quadrado da profundidade da pobreza. Ele mede a severidade da pobreza, e mensura as desigualdades entre os pobres. VIEIRA (2005) exemplifica comentando que, por exemplo, se uma transferência é feita de uma pessoa apenas um pouco abaixo da linha de pobreza para uma pessoa muito abaixo da linha de pobreza, o índice de profundidade da pobreza ao quadrado irá reduzir-se porque os padrões de vida dos mais pobres entre os pobres terão melhorado. Em contraste, tal transferência não afetaria nem o índice de incidência, nem o índice de profundidade da pobreza. Para maiores detalhes, ver ANGOLA (2005).

É importante ressaltar que não faz parte do escopo deste trabalho discutir qual o melhor indicador capaz de mensurar a pobreza. Diferentes indicadores fornecem diferentes resultados para a investigação do fenômeno da pobreza e cada um deles apresenta uma limitação específica. A idéia aqui é, com base em um indicador qualquer, buscar modelos estatísticos com maior poder preditivo para estimar o contingente populacional de indivíduos pobres numa dada região. Como salientado anteriormente, não há uma linha de pobreza oficial no Brasil, o que significa que saber a quantidade de pessoas pobres ainda é importante para a formulação de políticas públicas. Nesse sentido, optou-se por utilizar nesse estudo o indicador mais comumente utilizado, o de proporção de pobres como uma primeira aproximação ao tema.

## **2.4 A mensuração da pobreza no Brasil**

Em relação aos dados sobre pobreza produzidos pelo IBGE, CUNHA (2006) descreve que uma série de estudos foi feita desde o ano 2000 com o objetivo de selecionar uma metodologia para a análise da desigualdade e da pobreza no Brasil, com a apresentação dos principais indicadores utilizados, assim como a implementação desses indicadores com dados da PNAD do ano de 1999. A autora ainda diz que estes estudos são parte de um projeto do Instituto em parceria com o Programa das Nações Unidas para o Desenvolvimento (PNUD). Brevemente, citaremos cada um dos seis estudos realizados.

No primeiro relatório (ABRAMOVITZ, 2000), é feita uma análise da dimensão da desigualdade e da pobreza no Brasil nas últimas quatro décadas, comparando a situação brasileira com a de outros países. O texto traz uma aplicação dos índices de

pobreza estudados para o município de São Paulo para o ano de 1999, usando dados da PNAD e considerando que o desenho amostral da PNAD pode ser aproximado por uma amostra aleatória simples, visando tornar mais simples os cálculos dos índices selecionados.

No segundo relatório (MEDICI, 2001a), encontramos uma explicação dos motivos pelos quais se deve estimar a variância dos estimadores estudados, os fatores que influenciam sua precisão e as dificuldades encontradas na estimação.

O terceiro relatório (MEDICI, 2001b), por sua vez, tem como objetivo descrever a geração de bases de dados selecionados das pesquisas domiciliares do IBGE referidas ao tema de desigualdade de rendimentos para a década de 90.

O quarto relatório da série (MEDICI, 2001c) traz uma revisão da metodologia usada para a mensuração da pobreza, apresenta a metodologia para a decomposição das medidas, abordando condições e propriedades que uma medida deve satisfazer para que ela possa ser decomposta. Esse texto traz também uma aplicação da metodologia apresentada, a partir de características dos chefes de família, com base nos dados da PNAD para os anos de 1997 a 1999.

No quinto relatório (MEDICI, 2001d), há uma ilustração do método de decomposição apresentado no relatório anterior sobre a desigualdade de renda. A aplicação da parte referente à pobreza foi desenvolvida ainda no quarto relatório.

O último relatório (MEDICI, 2001e) traz um resumo da metodologia definida para a estimação de medidas de pobreza e a teoria de escala de equivalência - uma ferramenta que visa tratar a renda para que ela seja um melhor indicador do bem-estar econômico.

Por estes relatórios, fica evidente a preocupação do governo brasileiro em obter estimativas de pobreza para o país; não apenas em nível nacional, mas também em

regiões geográficas menores. Mas de acordo com CUNHA (2006), as pesquisas realizadas até o momento ainda não possibilitam estender a análise da pobreza em áreas geográficas pequenas, como municípios ou microrregiões. A autora justifica que isso ocorre devido ao fato das pesquisas referentes a unidades geográficas menores, como municípios, não incorporarem variáveis necessárias para a realização de tais análises. Por outro lado, as pesquisas que possuem maior profundidade nos temas abordados não são elaboradas para representarem pequenas áreas geográficas.

Esta carência de informações justifica a utilização de uma metodologia que combine pesquisas de ampla cobertura geográfica com pesquisas de abrangência temática mais aprofundada. Esse tipo de análise, proposta por ELBERS et. al. (2000), foi aplicada recentemente no Brasil por CUNHA (2006). A autora aplica a metodologia para a estimação da proporção de pobres nas microrregiões potiguares, como resultado, elabora um mapeamento da pobreza naquele Estado. Entretanto, os coeficientes de determinação encontrados pela autora foram moderados, com valores um pouco acima dos 60%. É importante ressaltar que a autora buscou avanços nos resultados obtidos, como a aplicação de modelos hierárquicos. A presente dissertação também busca melhores resultados, como alternativa aos métodos propostos por ELBERS et. al. (2000). No entanto, diferentemente do que foi tentado no trabalho de CUNHA (2006), o enfoque bayesiano é utilizado nas análises. O presente estudo teve como proposta ajustar modelos hierárquicos de regressão logística e modelos autoregressivos condicionais na amostra da POF. E, em seguida, predizer os valores da variável resposta para a amostra do Censo Demográfico. Nos itens a seguir são descritos os dados utilizados no presente estudo e as variáveis utilizadas.



## **2.5 Descrição das fontes de dados**

Na delimitação da área de estudo, dois quesitos foram considerados: (i) o esforço computacional e (ii) o contexto social do local. O Estado do Rio Grande do Norte atende a esses dois critérios – é relativamente pobre (com grandes desigualdades socioeconômicas) e pequeno em extensão geográfica. Uma breve análise das condições socioeconômicas do Estado potiguar é feita, logo após a descrição das fontes de dados consideradas. Os dados utilizados foram extraídos dos microdados da parte amostral do Censo Demográfico de 2000 e da Pesquisa de Orçamentos Familiares de 2002/2003. É importante ressaltar que, além dos microdados, utilizamos um arquivo auxiliar com informação sobre identificação dos setores censitários para cada domicílio entrevistado nas pesquisas.

A população alvo é constituída dos habitantes do Estado do Rio Grande do Norte, para o ano de 2003. Na amostra do Censo de 2000, para aquele Estado, foram pesquisados aproximadamente 92,6 mil domicílios. Na POF, por sua vez, foram entrevistados 1548 domicílios. A seguir, são apresentados em maiores detalhes, as pesquisas POF 2002/2003 e o Censo Demográfico de 2000.

### **2.5.1 A Pesquisa de Orçamentos Familiares – POF 2002/2003**

A Pesquisa de Orçamentos Familiares teve como principal objetivo medir a estrutura de consumo, dos gastos e dos rendimentos das unidades de consumo, permitindo traçar um perfil das condições de vida dos brasileiros, além de servir como base para o cálculo dos índices de preços (IBGE, 2004). No caso, o termo unidade de

consumo é definido como sendo a unidade básica para investigação e análise dos orçamentos. Ela pode ser um único morador ou um conjunto de moradores que compartilham da mesma fonte de alimentação.

A POF cobriu todo o território brasileiro. Com base nessa pesquisa, é possível a estimação de resultados para o Brasil e Grandes Regiões, além de resultados para a situação urbana e rural. A pesquisa foi realizada durante os meses de julho de 2002 a junho de 2003, no intuito de captar possíveis alterações na estrutura de despesas (consumo), quantidade de bens e rendimentos domiciliares durante o ano, quando comparadas com as pesquisas anteriores.

No intuito de analisar toda a diversidade de itens de despesas, com diferentes valores e frequência de aquisição, foram definidos 4 períodos de referência para a pesquisa: 7 dias, 30 dias, 90 dias e 12 meses. Cabe ressaltar ainda que, durante o período de coleta, ocorreram mudanças nos preços, justificando a criação de uma data de referência para a apresentação dos resultados da POF. Essa data foi 15 de janeiro de 2003.

O plano amostral da pesquisa foi uma amostragem conglomerada em dois estágios, com estratificação geográfica e estatística (a partir dos estratos socioeconômicos) das unidades primárias de amostragem (UPA), que são os setores censitários. Essas UPAs foram selecionadas através de uma amostragem sistemática com probabilidade proporcional à quantidade de domicílios no setor. As unidades secundárias de amostragem foram os domicílios particulares permanentes, e estes foram sorteados por amostragem aleatória simples sem reposição, dentro dos setores selecionados.

A estratificação geográfica teve a finalidade de espalhar geograficamente a amostra. Para a área urbana de cada Unidade da Federação, foram definidos os estratos:

município da capital, restante da região metropolitana sem a capital e restante da área urbana. Os estratos de seleção das áreas rurais foram determinados para as cinco grandes regiões do país. No que se refere à estratificação estatística, a variável utilizada foi a média de anos de estudo do responsável pelo domicílio, obtida a partir dos dados da Contagem Populacional de 1996 (IBGE, 2004), por ser a fonte censitária mais recente na época do planejamento da pesquisa.

Para a amostra, foram selecionados 3.984 setores censitários no Brasil. Na área urbana foram selecionados 13 domicílios em cada setor, e na área rural foram selecionados 20 (nas regiões Norte e Nordeste) ou 25 (nas demais regiões). Isso totalizou 48.470 domicílios selecionados no Brasil como um todo. No Rio Grande do Norte, a amostra foi de 132 setores e 1919 domicílios. A Tabela 2.1 apresenta o número de setores censitários e de domicílios efetivamente pesquisados nesse Estado por microrregião. A quantidade de domicílios selecionada é maior devido a ter que compensar a perda de domicílios por entrevista não realizada durante a coleta de dados. Como pode ser observado, em algumas microrregiões o número de setores entrevistados é bem pequeno, no Médio Oeste e Serra de São Miguel, por exemplo, apenas 1 setor foi selecionado e pesquisado.

**Tabela 2.1: Número de setores censitários e domicílios pesquisados na POF com entrevista completa no Rio Grande do Norte, por microrregião.**

<b>Microrregião</b>	<b>Setores</b>	<b>Domicílios</b>
Médio Oeste	1	12
Macau	3	32
Angicos	2	24
Baixa Verde	3	29
Serra de São Miguel	1	12
Serra de Santana	3	38
Umarizal	2	28
Chapada do Apodi	3	41
Litoral Nordeste	2	31
Seridó Ocidental	4	46
Pau dos Ferros	5	60
Litoral Sul	5	70
Seridó Oriental	6	84
Vale do Açu	5	62
Borborema Potiguar	3	28
Agreste Potiguar	7	82
Macaíba	10	144
Mossoró	14	165
Natal	48	560
<b>Total do Estado</b>	<b>127</b>	<b>1548</b>

**Fonte: Microdados da POF 2002/2003**

A situação é ainda mais crítica quando discriminamos a amostra por tipo de área: urbana e rural. Como a estratificação nas áreas rurais não foi definida para cada Unidade da Federação, nem todas as microrregiões do Rio Grande do Norte tiveram setores selecionados para compor o estrato rural da região Nordeste. Na Tabela 2.2 podemos observar que 9 das 19 microrregiões do Estado não tiveram setores censitários selecionados na área rural.

**Tabela 2.2: Número de estratos, setores censitários e domicílios pesquisados na POF com entrevista completa no Rio Grande do Norte, por microrregião e tipo de área.**

Microrregião	Estratos		Setores		Domicílios	
	Urbana	Rural	Urbana	Rural	Urbana	Rural
Médio Oeste	1	-	1	-	12	-
Macau	3	-	3	-	32	-
Angicos	1	1	1	1	9	15
Baixa Verde	2	-	3	-	29	-
Serra de São Miguel	1	-	1	-	12	-
Serra de Santana	2	1	2	1	20	18
Umarizal	1	1	1	1	10	18
Chapada do Apodi	2	1	2	1	25	16
Litoral Nordeste	1	-	2	-	31	-
Seridó Ocidental	4	-	4	-	46	-
Pau dos Ferros	3	1	4	1	42	18
Litoral Sul	2	1	4	1	50	20
Seridó Oriental	3	1	5	1	65	19
Vale do Açu	3	1	4	1	46	16
Borborema Potiguar	3	-	3	-	28	-
Agreste Potiguar	2	2	5	2	52	30
Macaíba	3	4	5	5	53	91
Mossoró	5	-	14	-	165	-
Natal	9	-	48	-	560	-
<b>Total do Estado</b>	<b>51</b>	<b>14</b>	<b>112</b>	<b>15</b>	<b>1287</b>	<b>261</b>

Fonte: Microdados da POF 2002/2003

## 2.5.2 O Censo Demográfico de 2000

A data de referência adotada pelo IBGE para o censo foi 01 de agosto de 2000 e o período de coleta foi o dia 01 de agosto a 30 de novembro de 2000.

O Censo foi realizado com base em dois questionários: o básico e o da amostra; sendo o primeiro aplicado a todos os domicílios particulares do país que não tinham sido selecionados para a amostra, abordando um conjunto não muito extenso de perguntas. Por sua vez, o questionário da amostra foi aplicado somente para uma amostra de domicílios, que além das perguntas contidas no questionário básico, abordava questões mais detalhadas sobre características dos domicílios e das pessoas residentes, referentes aos temas: religião, cor/raça, deficiência, migração, escolaridade, fecundidade,

nupcialidade, trabalho e rendimento. O presente estudo faz uso dos dados da parte amostral do Censo 2000. O desenho amostral da pesquisa constituiu numa seleção sistemática com equi-probabilidade, dentro de cada setor, de uma amostra de domicílios e da investigação das pessoas que residiam neles. Foram definidas duas frações amostrais distintas: 10% para os municípios com população estimada maior que 15.000 habitantes e 20% para os demais municípios. A amostra do Censo é muito maior que a da POF, uma vez que foram pesquisados 92,6 mil domicílios em 2623 setores no Rio Grande do Norte, conforme pode ser observado na Tabela 2.3.

**Tabela 2.3: Número de setores censitários e domicílios pesquisados no Censo e na POF com entrevista completa no Rio Grande do Norte, por microrregião.**

Microrregião	Censo		POF	
	Setores	Domicílios	Setores	Domicílios
Médio Oeste	38	1.875	1	12
Macau	55	1.646	3	32
Angicos	57	2.443	2	24
Baixa Verde	73	2.000	3	29
Serra de São Miguel	63	2.414	1	12
Serra de Santana	63	3.049	3	38
Umarizal	77	3.412	2	28
Chapada do Apodi	72	2.402	3	41
Litoral Nordeste	86	2.724	2	31
Seridó Ocidental	88	3.190	4	46
Pau dos Ferros	121	5.092	5	60
Litoral Sul	102	4.157	5	70
Seridó Oriental	113	4.456	6	84
Vale do Açu	129	4.683	5	62
Borborema Potiguar	129	5.580	3	28
Agreste Potiguar	212	8.553	7	82
Macaíba	224	5.901	10	144
Mossoró	259	7.348	14	165
Natal	662	21.748	48	560
<b>Total do Estado</b>	<b>2.623</b>	<b>92.673</b>	<b>127</b>	<b>1.548</b>

Fonte: Microdados da POF 2002/2003 e do Censo 2000.

É importante ressaltar, mais uma vez, que as informações em nível de setor censitário não estão disponíveis no arquivo de microdados de acesso público, para isto foi necessário a obtenção de um arquivo auxiliar com identificação dos setores censitários, para cada domicílio. Além disso, ressaltamos ainda que, atualmente, o IBGE desenvolve um estudo de mapeamento da pobreza no Brasil, porém até o momento os resultados não foram divulgados.

## **2.6 Contextualização socioeconômica e demográfica do Rio Grande do Norte**

O Rio Grande do Norte é um Estado relativamente pequeno, tanto em termos de população quanto em termos de extensão geográfica, quando comparado a outros Estados brasileiros. De acordo com dados do Censo Demográfico de 2000, é a segunda Unidade da Federação da região Nordeste menos populosa, com menos de três milhões de habitantes (dos quais 2.036.673 vivem na área urbana, o que representa 73,35% de sua população total), superado apenas pelo Estado de Sergipe. O Estado concentra cerca de metade da sua população urbana em apenas três municípios com população superior a 100 mil habitantes: Natal (a capital potiguar), Mossoró e Parnamirim. Dos 166 municípios existentes no ano de 2000, 97 possuíam até 5.000 habitantes que somadas concentravam apenas 11,9% da população urbana do Estado. Maiores detalhes, na Tabela 2.4 a seguir.

**Tabela 2.4: Distribuição dos municípios, segundo faixa de população residente em área urbana. Rio Grande do Norte, 2000.**

Faixa de população (em 1.000 hab.)	Municípios		População	
	Total	%	Total	%
Até 2	34	20,5	46.463	2,3
De 2 até 5	63	38,0	194.794	9,6
De 5 até 10	44	26,4	312.497	15,3
De 10 até 100	22	13,3	462.382	22,7
Acima de 100	3	1,8	1.020.537	50,1
<b>Total</b>	<b>166</b>	<b>100,0</b>	<b>2.036.673</b>	<b>100,0</b>

Fonte: Microdados do CENSO 2000.

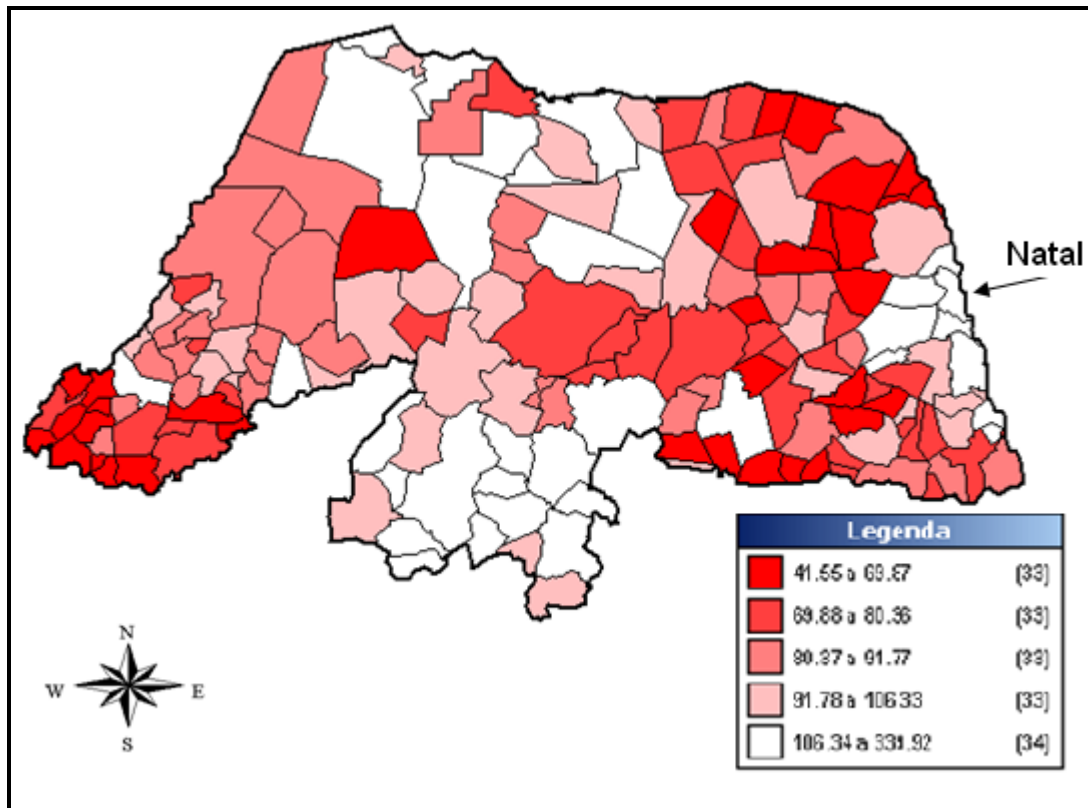
Em relação ao Índice de Desenvolvimento Humano Municipal (IDH-M), o Rio Grande do Norte (0,705) estava ligeiramente abaixo do índice nacional (0,766), de acordo com o Atlas de Desenvolvimento no Brasil (2003).

Sua renda per capita, com base no Censo 2000, (R\$176,21) estava bem abaixo da média brasileira (R\$297,00). O Estado também apresenta grandes disparidades entre seus municípios, podendo-se verificar que a renda per capita do município de Venha-Ver (R\$41,55) é cerca de sete vezes menor que o valor observado no município de Natal (R\$339,92), a capital do Estado. Com base na Figura 2.1, é possível verificar, de maneira geral, o quão pobre o estado é, a partir da variável renda per capita. Nota-se que de maneira geral, o Estado é muito pobre. Dos 166 municípios, 139 possuem renda per capita inferior a R\$116,14 e apenas 1 (Natal) possui renda per capita superior a R\$265,00.



**Figura 2.1: Distribuição da variável Renda per capita.**

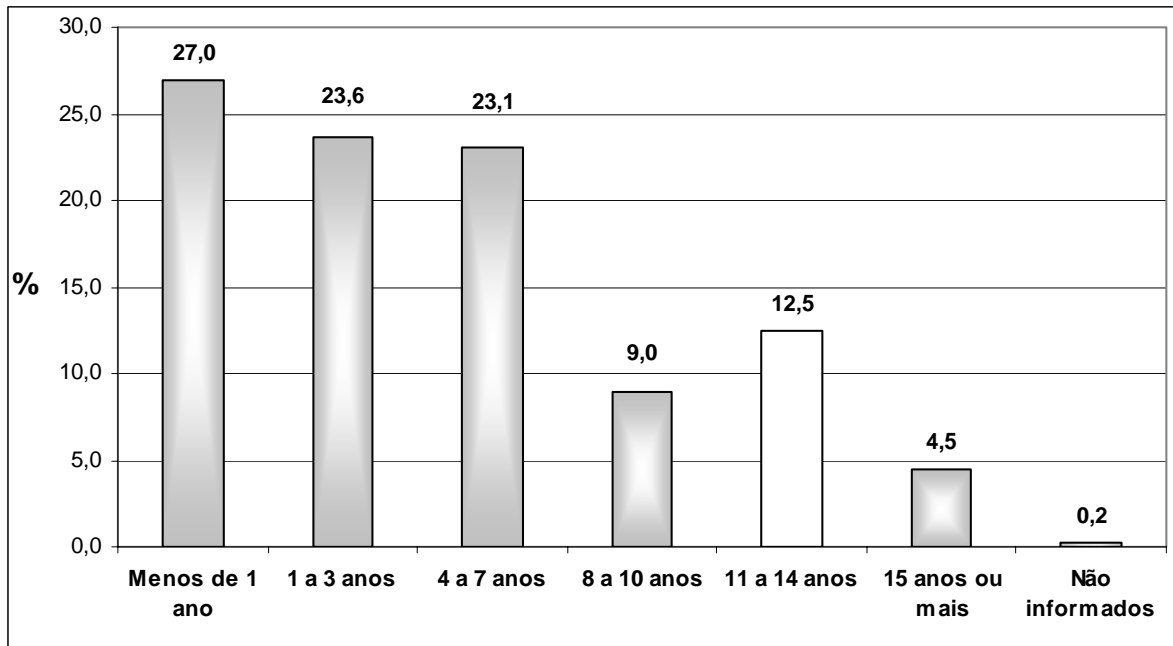
**Rio Grande do Norte, 2000.**



**Fonte: Microdados do censo 2000 apud Atlas do Desenvolvimento Humano no Brasil (2003)**

Em relação à educação, é detectada uma taxa de alfabetização da população de 10 anos ou mais de 76,3%, o que significa dizer que 23,7% da população desta faixa etária é analfabeta. No que diz respeito aos anos de estudo do chefe de domicílio, verificou-se um baixo nível de escolarização, uma vez que 50% destes se encontrarem nas faixas de menos de 1 ano e de 1 a 3 anos de estudo. É possível observar que metade dos chefes de domicílio do Estado não chegou a completar o primeiro ciclo, referente ao ensino fundamental. Maiores detalhes podem ser constatados no Gráfico 2.1.

**Gráfico 2.1: Distribuição da variável Anos de Estudo do Chefe do Domicílio por classes de anos, Rio Grande do Norte, 2000.**



Fonte: Microdados do censo 2000 apud Atlas do Desenvolvimento Humano no Brasil (2003)

Os dados da Tabela 2.5 ainda revelam a precariedade do sistema potiguar de esgotamento sanitário, no qual 47,5% dos domicílios ainda utilizam fossa rudimentar e apenas 16,5% estão ligados a rede geral de esgotos. Quanto à existência de banheiro ou sanitário, 10% dos domicílios não dispõem de instalações sanitárias. Pode-se dizer que esta parcela da população está mais exposta a proliferação de doenças devido às condições inadequadas de higiene.

Com relação ao destino do lixo, é possível verificar que 26,3% dos domicílios dão aos resíduos outros destinos que não a coleta organizada. Outro item que merece destaque é a rede geral de abastecimento de água que atende a somente 78% dos domicílios potiguares.

**Tabela 2.5: Domicílios particulares permanentes, segundo algumas características domiciliares. Rio Grande do Norte, 2000.**

<b>Algumas características domiciliares</b>	<b>Domicílios particulares permanentes</b>
<b>Abastecimento de água</b>	
Rede geral	78,2
Poço/Nascente	2,0
Outra forma	15,1
<b>Esgotamento sanitário</b>	
Rede geral de esgoto	16,5
Fossa séptica	24,3
Fossa rudimentar	47,5
Outra forma ou sem instalação	11,6
Sem declaração	0,0
<b>Destino do lixo</b>	
Coletado Serviço de Limpeza	68,2
Colocado em Caçamba	5,5
Outro	26,3

Fonte: Microdados do censo 2000 apud Atlas do Desenvolvimento Humano no Brasil (2003)

## 2.7 As variáveis de interesse

Uma das principais questões ao se estudar o tema pobreza está em definir um indicador mais apropriado para mensurar o bem-estar dos indivíduos. Ainda há muita controvérsia na literatura sobre qual é o melhor indicador para mensurar a pobreza. Pode-se construir indicadores tanto a partir da renda quanto do consumo. O presente estudo utilizará os dois indicadores: o rendimento domiciliar per capita e o consumo domiciliar per capita. O primeiro se justifica, segundo ROCHA (2003), como sendo a melhor aproximação para o bem-estar do indivíduo devido a forte correlação existente entre a renda e indicadores de bem-estar físico. O segundo, por sua vez, reflete não apenas o que uma família pode fazer com sua renda, mas também os diversos hábitos culturais e regionais, como destacam QUINTAES et. al. (2006).

## **2.7.1 O rendimento domiciliar per capita**

O rendimento domiciliar per capita é construído por meio das várias fontes de rendimento recebidas pelas pessoas pertencentes ao domicílio. De acordo com o questionário da POF, há 2 quesitos para a variável rendimento: o rendimento monetário e o rendimento não monetário. O primeiro é referente a todo e qualquer tipo de rendimento monetário recebido durante o período de referência de 12 meses anteriores à data de realização da coleta dos dados. No segundo quesito foi computado tudo aquilo que foi produzido, pescado, caçado ou recebido em bens utilizados ou consumidos durante o período de referência da POF. Maiores detalhes podem ser obtidos na documentação da pesquisa (IBGE, 2004). O rendimento domiciliar é formado pela soma de todos esses rendimentos.

O rendimento domiciliar per capita foi obtido por meio da divisão do total obtido com a soma dos rendimentos pelo número de pessoas residentes no domicílio, excluindo-se o rendimento de pensionistas, empregados domésticos e seus parentes. Com esse procedimento avaliamos a condição dos domicílios sem a contribuição dos rendimentos obtidos por essas pessoas. Como os valores de rendimento foram obtidos em diferentes momentos, foi necessário deflacioná-los para o ajuste em relação ao efeito inflacionário entre os períodos de coleta. Com base na POF, a data de referência dos valores deflacionados foi 15 de janeiro de 2003.

No Censo, de acordo com o questionário da amostra, há 7 quesitos para a variável rendimento: (i) trabalho principal, (ii) demais trabalhos, (iii) aposentadorias ou pensões, (iv) aluguel, (v) pensão alimentícia, mesada, doação, (vi) renda mínima, bolsa escola, seguro desemprego e (vii) outros rendimentos (tais como, abono de permanência

em serviço, juros de aplicações financeiras, etc.). O rendimento domiciliar é formado pela soma de todos esses rendimentos.

No Censo, o rendimento domiciliar per capita foi obtido por meio da divisão do total obtido com a soma dos rendimentos pelo número de pessoas residentes no domicílio, excluindo-se pensionistas, empregados domésticos e seus parentes. Os valores de rendimento foram deflacionados para o ajuste em relação ao efeito inflacionário entre os períodos de coleta. Com base na POF, a data de referência dos valores deflacionados foi 15 de janeiro de 2003.

### **2.7.2 O consumo domiciliar per capita**

Nesse item, vale dizer que estamos considerando que, em termos econômicos, consumo costuma ser definido como o valor (ou quantidade) das compras adquiridas, apesar de essa idéia ser restritiva quando os objetivos incluem avaliação nutricional. Para fins de padronização, utilizaremos o conceito de consumo para representar os gastos das unidades de consumo, mesmo sabendo que nem sempre eles são sinônimos.

A construção desta variável é baseada no estudo de QUINTAES et. al. (2006). Os autores defendem o uso deste indicador argumentando que os gastos de consumo refletem não apenas o que uma família pode fazer com seus rendimentos, mas também nos diz se tem acesso aos mercados de crédito. O consumo domiciliar é composto por duas partes: alimentar e não-alimentar. A primeira compreende toda a aquisição (monetária e não-monetária) com refeições. A segunda, por sua vez, é composta pelo total de todos os gastos com produtos e serviços feitos pela família. Cabe dizer que essa variável leva em consideração todo o período de pesquisa, ou seja, os 12 meses de

realização da POF. Assim como feito para a variável rendimento domiciliar per capita, todos os valores foram deflacionados para a data de referência da pesquisa, 15 de janeiro de 2003.

Para a construção da variável consumo domiciliar per capita, foram levados em consideração os seguintes itens: (i) Alimentação dentro e fora do domicílio; (ii) Habitação; (iii) Transporte; (iv) Saúde e educação; (v) Vestuário, cultura e lazer, serviços pessoais e higiene e cuidado pessoal; (vi) Fumo; (vii) Outras despesas; (viii) Outras despesas correntes; (ix) Aumento do ativo e (x) Diminuição do passivo. Maiores detalhes sobre a construção desta variável podem ser encontrados em QUINTAES et. al. (2006)

Cabe dizer que ambas as variáveis dependentes (consumo e renda) eram bastante assimétricas. Numa tentativa de contornar isso, adotamos a transformação logarítmica nos valores das duas variáveis, como feito na grande maioria dos estudos desta natureza.

## **2.8 Variáveis explicativas**

Para a execução das análises, é necessário identificar um conjunto de variáveis explicativas compatíveis nas duas pesquisas em estudo. Neste item, seguimos os mesmos passos do trabalho de CUNHA (2006). A idéia é, usando uma amostra pequena, porém rica em conteúdo, estimar a distribuição condicional da variável de interesse  $y$  observada no nível de domicílio, dado um vetor de covariáveis  $x$ . Quando restringimos o conjunto de variáveis explicativas para aquelas que também podem ser observadas nos domicílios da amostra do Censo Demográfico, a distribuição estimada na amostra pequena pode ser usada para gerar a distribuição de  $y$ , condicionadas às características de  $x$  da subpopulação observada, para qualquer subpopulação da amostra

do Censo. Essas covariáveis, que serão utilizadas como variáveis de ligação entre as duas pesquisas devem ser idênticas no que se refere a sua definição e distribuição, não sendo suficiente apenas assegurar que as perguntas sejam iguais em ambas as pesquisas.

**Quadro 2.1: Lista de perguntas comuns na POF e no Censo Demográfico**

<b>Nome abreviado</b>	<b>Descrição</b>
TOTCOMDO	Quantidade de dormitórios
TIPODOM	Tipo de domicílio
TOTCOM	Total de cômodos
ABASTEC	Forma de abastecimento de água
EXISTBAN	Existência de banheiro
QTDBANH	Quantidade de banheiros
TIPOESC	Tipo de esquadro
CONDDOM	Condição do domicílio
ILUMINA	Existência de Iluminação Elétrica
PAVIMEN	Existência de Pavimentação
AR	Posse de Ar condicionado
AUTO	Posse de Automóvel
COMPUT	Posse de Computador
GELFREE	Posse de Geladeira ou freezer
MICROO	Posse de Microondas
MQLAVAR	Posse de Máquina de lavar
RADIO	Posse de Rádio
TV	Posse de Televisão
VIDEO	Posse de Vídeo cassete
SEX	Sexo da pessoa de referência da Unid. de Cons. ou domicílio
COR4	Cor da pessoa de referência da Unid. de Cons. ou domicílio
FREQ	Frequência à escola
ESCOLA	Nível de escolaridade da pessoa de referência do domicílio ou Unid. de Consumo
ANOEST	Anos de estudo da pessoa de referência do domicílio ou Unid. de Consumo
CONJUGE	Existência de cônjuge
IDOSOS	Número de idosos no domicílio ou Unid. de Consumo
CRIAN	Número de crianças no domicílio ou Unid. de Consumo
ADOLES	Número de adolescentes no domicílio ou Unid. de Consumo
TIPOFAM	Tipo de família
POS	Posição na ocupação da pessoa de referência do domicílio ou Unid. de Consumo
MORA	Quantidade de moradores no domicílio ou Unid. de Consumo

Fonte: Adaptado de CUNHA (2006).

Num primeiro momento, as variáveis foram selecionadas através da comparação dos dois questionários das pesquisas, identificando perguntas comuns. Como resultado, foram identificadas 31 perguntas presentes tanto na POF quanto no Censo Demográfico. Cabe ressaltar que nem sempre as perguntas apresentavam as mesmas categorias de respostas, justificando-se a realização de agregações, para que as variáveis utilizadas medissem a mesma quantidade. A lista de perguntas comuns às duas pesquisas é apresentada no Quadro 2.1.

Dando prosseguimento à comparação dos questionários, foi verificada a compatibilidade das variáveis por meio de teste de hipóteses para verificar a igualdade das proporções. O procedimento de teste de hipóteses é descrito em detalhes no Capítulo 4. Após os testes de hipóteses, restaram apenas vinte variáveis que puderam ser consideradas compatíveis, conforme mostrado no Quadro 2.2.

Estas variáveis compuseram a etapa seguinte do trabalho, que é a construção dos modelos estatísticos e que, posteriormente, seriam utilizados para imputar a variável resposta no Censo Demográfico. De acordo com ELBERS et. al. (2002), somente essas variáveis podem compor os modelos. Por outro lado, CUNHA (2006) relata que não se deve descartar o uso de variáveis apenas pelo fato das proporções terem sido consideradas diferentes pelos testes de hipóteses nas duas pesquisas e que isto pode sugerir uma evolução natural dos dados. Para avaliar esta hipótese, a autora utilizou uma comparação com dados das PNAD dos anos 2001, 2002 e 2003. No presente estudo, optou-se por seguir a recomendação original de ELBERS et. al. (2002) e descartar, de imediato, as variáveis consideradas diferentes pelos testes de hipóteses.

A descrição completa da lista de variáveis que podem ser testadas nos modelos pode ser encontrada no Quadro A1, em anexo.



**Quadro 2.2: Lista de perguntas comuns consideradas compatíveis pelo teste de igualdade de proporções.**

<b>Nome abreviado</b>	<b>Descrição</b>
TIPODOM	Tipo de domicílio
QTDBANH	Quantidade de banheiros
TIPOESC	Tipo de escoadouro
CONDDOM	Condição do domicílio
ILUMINA	Existência de Iluminação Elétrica
PAVIMEN	Existência de Pavimentação
AR	Posse de Ar condicionado
COMPUT	Posse de Computador
MICROO	Posse de Microondas
MQLAVAR	Posse de Máquina de lavar
TV	Posse de Televisão
SEX	Sexo da pessoa de referência da Unid. de Cons. ou domicílio
COR4	Cor da pessoa de referência da Unid. de Cons. ou domicílio
ANOEST	Anos de estudo da pessoa de referência do domicílio ou Unid. de Consumo
CONJUGE	Existência de cônjuge
IDOSOS	Número de idosos no domicílio ou Unid. de Consumo
ADOLES	Número de adolescentes no domicílio ou Unid. de Consumo
TIPOFAM	Tipo de família
POS	Posição na ocupação da pessoa de referência do domicílio ou Unid. de Consumo
MORA	Quantidade de moradores no domicílio ou Unid. de Consumo

Fonte: Adaptado de CUNHA (2006).

## **CAPÍTULO 3**

# **ESTIMAÇÃO EM PEQUENAS ÁREAS**

Neste capítulo, é feita uma breve revisão das técnicas de estimação em pequenas áreas freqüentemente utilizadas, considerando-se principalmente aquelas que são baseadas em modelos estatísticos.

Estimação em pequenas áreas é uma grande área da estatística que se destina a estudar o problema de fornecer estimativas confiáveis de uma variável de interesse num conjunto de áreas geográficas pequenas (RAO, 2003).

Um procedimento usualmente utilizado em estudos de estimação de pequenas áreas é a combinação de dados amostrais com uma fonte auxiliar de informação, tais como dados de registros ou censos. Neste contexto, têm-se desenvolvido estimadores baseados em modelos estatísticos que tomam “emprestado” informação de áreas vizinhas com o intuito de obter estimadores mais precisos. Uma das primeiras aplicações dessa metodologia foi o estimador sintético proposto por GONZALES (1973), que posteriormente foi usado em GONZALES e HOZA (1978). Sucintamente, o procedimento de estimação sintética consiste em assumir uma disponibilidade de estimativas para grandes unidades da população. Estas grandes unidades da população podem ser grandes unidades geográficas (i.e., estados ou regiões) ou grupos demográficos (i.e., grupos etários). Pesos são, então, atribuídos à essas estimativas das grandes unidades da população para se obter as estimativas desejadas das pequenas áreas em questão. Geralmente, esses pesos são obtidos com base em dados censitários. Entretanto, de acordo com CASTAÑEDA (1999), tais métodos de estimação costumam ser tendenciosos por assumirem duas hipóteses. Primeiro, a premissa de homogeneidade de taxas ou proporções é difícil de ser satisfeita, ou seja, é pouco provável que a relação encontrada nas áreas maiores

seja verificada nos pequenos domínios. Além disso, a estrutura do fenômeno em estudo pode ter se alterado no intervalo entre as pesquisas.

Numa tentativa de superar essa deficiência, estimadores baseados em modelos de regressão foram propostos posteriormente, como pode ser visto, por exemplo, em HOLT et. al. (1979), CASTAÑEDA (1999) e MOURA e MIGON (2002). No método do modelo de regressão, a variável resposta é geralmente obtida através de uma amostra e os previsores são formados por um conjunto de variáveis auxiliares disponíveis para as unidades amostrais. Cabe ressaltar que se supõem conhecidas as médias populacionais destas variáveis auxiliares para cada pequena área em questão. Formalmente, o modelo pode ser escrito sob a seguinte forma:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + e_{ij},$$

no qual  $y_{ij}$  é o valor da variável resposta  $y$  para a  $j$ -ésima unidade amostral da  $i$ -ésima pequena área sob estudo,  $i = 1, \dots, m$  e  $j = 1, \dots, n_i$ . De maneira similar, é definida a matriz  $x_{kij}$ ,  $k = 1, \dots, p$ , de variáveis explicativas. Assume-se que os termos residuais,  $e_{ij}$ , seguem distribuição normal, com média zero e variância  $\sigma_\varepsilon^2$ . O vetor de parâmetros  $\beta = (\beta_0, \dots, \beta_p)^T$  é estimado pelo método dos mínimos quadrados, com base na informação amostral de todas as pequenas áreas.

Mesmo assim, estes métodos não permitem capturar toda a heterogeneidade presente nas pequenas áreas. CASTAÑEDA (1999) diz que os estimadores de regressão captam pouca variação entre as pequenas áreas e conseqüentemente as estimativas obtidas são geralmente concentradas em torno de um valor médio de previsão. Tal fato conduziu a outros métodos baseados em modelos mais gerais que levam em consideração a heterogeneidade entre as diversas pequenas áreas do estudo.

Diversos autores propuseram a utilização de modelos hierárquicos com o intuito de captar alguma variação entre as pequenas áreas que não podem ser atribuídas a diferenças entre os valores das variáveis auxiliares.

Como exemplo inicial, podemos citar o trabalho de FAY e HERRIOT (1979), que propuseram modelos lineares hierárquicos para a estimação da renda per capita em pequenas áreas e fizeram uma aplicação aos dados do Censo Demográfico Americano de 1970. O modelo de FAY e HERRIOT (1979) pode ser escrito como:

$$\tilde{\theta}_i = \theta_i + e_i$$

$$\theta_i = x_i^T \beta + u_i.$$

No qual  $u_i$  e  $e_i$  são assumidos independentes. Aqui,  $\tilde{\theta}_i$  é média amostral da característica de interesse na  $i$ -ésima pequena área;  $e_i$  é o erro amostral, assumidamente, com média zero e variância conhecida  $Var_D(e_i) = \sigma_{Di}^2$ ,  $x_i = (x_{i1}, \dots, x_{ip})^T$  é um vetor de variáveis auxiliares conhecido para todas as áreas  $i = 1, \dots, m$ ;  $\beta = (\beta_1, \dots, \beta_p)$  é um vetor de parâmetros desconhecidos. Este modelo incorpora a estrutura do desenho, leva em consideração um erro de efeito aleatório  $u_i$  e um erro amostral  $e_i$ , sendo os dois erros independentes.

BATTESE *et. al.* (1988) propuseram e aplicaram um modelo hierárquico de intercepto aleatório para a estimação da média de produção de soja em doze pequenas áreas no estado americano de Iowa, utilizando dados amostrais e como variável auxiliar dados fornecidos por satélite. Este estudo mostrou a superioridade dos estimadores baseados no modelo de intercepto aleatório em relação àqueles baseados em modelos de regressão simples. Este modelo pode ser escrito como:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + u_{0i} + e_{ij}.$$

A diferença, neste modelo, em relação aos modelos de regressão, é a incorporação de um efeito aleatório para cada pequena área  $i$  sob estudo,  $i = 1, \dots, m$ . Com base nesse modelo, pode-se obter o seguinte estimador para a média da  $i$ -ésima pequena área:

$$\hat{\mu}_{1i} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_{1i} + \dots + \hat{\beta}_p \bar{X}_{pi} + \hat{u}_{0i}.$$

Há vários métodos disponíveis para se estimar os parâmetros de regressão e o efeito aleatório  $u_{0i}$ ; veja PRASAD e RAO (1990) para maiores detalhes. Especificamente, os autores obtiveram o erro quadrático médio, uma eficiência relativa desse modelo em relação ao modelo de regressão.

MOURA & HOLT (1999) propuseram um modelo hierárquico mais geral para estimação em pequenas áreas para variáveis contínuas em que todos os coeficientes do modelo podem ser aleatórios. O modelo faz uso de variáveis no nível de unidades amostrais como também no nível de pequenas áreas. Os autores obtiveram o EBLUP (sigla em inglês para melhor preditor linear empírico não tendencioso) para a média da pequena área  $\mu_i$  e um estimador não tendencioso do erro quadrático médio. O modelo hierárquico proposto pelos autores pode ser escrito como:

$$Y_i = X_i \beta_i + e_i, \quad i = 1, \dots, m,$$

$$\beta_i = Z_i \gamma + u_i,$$

onde:  $m$  é o número de pequenas áreas na população em questão;  $X_i$  é uma matriz de dimensão  $p \times q$  de variáveis auxiliares para a  $i$ -ésima pequena área em estudo;  $\gamma$  é um vetor de dimensão  $q$  de coeficientes fixos e  $u_i = (u_{i1}, \dots, u_{ip})^T$  é o vetor de dimensão  $p$  de efeitos aleatórios para a  $i$ -ésima pequena área. Os termos  $u_i$  são supostamente independentes com  $E(u_i) = 0$  e  $V(u_i) = \Phi$ , no qual:

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1p} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{p1} & \Phi_{p2} & \cdots & \Phi_{pp} \end{bmatrix}$$

Há diversos estudos com aplicações de modelos hierárquicos ao problema de estimação em pequenas áreas sob a abordagem bayesiana. Como exemplo, podemos citar: DATTA e GHOSH (1991), que apresentaram uma abordagem bayesiana para modelos lineares generalizados; YOU e RAO (1997) que consideraram uma abordagem hierárquica bayesiana com variância não-constante no último nível da hierarquia. Os autores aplicaram o método do amostrador de Gibbs para retirar amostras da distribuição a posteriori para a média de cada pequena área.

Outros autores estudaram o problema de estimação de taxas e proporções para pequenos domínios utilizando a abordagem bayesiana. Como exemplo, podemos citar o trabalho de DEMPSTER e TOMBERLIN (1980), em que os autores propuseram um método bayesiano empírico para estimar a proporção de perdas censitárias para pequenas áreas baseada em modelos de regressão logística contendo parâmetros fixos e aleatórios. FARREL, MACGIBBON e TOMBERLIN (1997) obtiveram estimativas de proporções para pequenas áreas pelo método bayesiano empírico e forneceram critérios de escolha para distribuição da priori que depende da existência de valores discrepantes nas pequenas áreas. FARRELL (2000), por sua vez, aplica

modelos de regressão logística com efeitos aleatórios para estimar a proporção de participação na força de trabalho local.

Na mesma linha, pode-se citar também o estudo de MOURA e MIGON (2002) em que os autores apresentam um modelo de regressão logística com efeito aleatório para estimação da proporção de alunos com grau de proficiência escolar considerado satisfatório para 34 regiões do município do Rio de Janeiro para o ano de 1996.

Uma outra possibilidade de utilização conjunta de informações de todas as áreas para melhorar a estimação em pequenas áreas pode ser feita por meio de modelos hierárquicos espaciais. Nestes modelos são consideradas informações de áreas geograficamente próximas, onde várias definições e medidas de proximidade podem ser assumidas. Cita-se também o trabalho de MOURA e MIGON (2002), em que modelos de regressão logística hierárquicos de dois níveis com efeitos espaciais são utilizados para a estimação de proporções. Os autores mostram a superioridade dos estimadores de regressão logística com efeitos espaciais quando comparados com aqueles que não consideram efeitos espaciais e com os que não consideram níveis de hierarquia nem efeitos espaciais nos dados.

### **3.1 Estimação da pobreza em pequenas áreas**

Nesta seção, apresentamos as duas abordagens para a análise estatística dos dados, ou seja, para o mapeamento da pobreza. Utilizamos métodos tradicionais, da estatística clássica, e de estatística bayesiana. Em relação aos métodos estatísticos tradicionais ou clássicos, apresentamos a metodologia desenvolvida por pesquisadores do Banco Mundial (ELBERS et. al., 2000) para estimação de indicadores de pobreza aplicados em diversos países, tais como: Equador (ELBERS et. al., 2001), Brasil (CUNHA, 2006), Uganda (SCHIPPER e HOOGEVEEN, 2005) entre outros. Especificamente, aplicaremos a metodologia no Estado do Rio Grande

do Norte, assim como CUNHA (2006). Para a abordagem bayesiana, a proposta é utilizar modelos hierárquicos de regressão logística (MOURA e MIGON, 2002) e modelos que incorporam efeitos espaciais; no caso, aplicaremos modelos autoregressivos condicionais (SCHIMIDT et al., 2003).

### **3.1.2 A proposta do Banco Mundial**

O procedimento proposto por pesquisadores do Banco Mundial para a construção de indicadores de bem-estar para pequenas áreas visa combinar a vantagem da pesquisa amostral domiciliar em se ter informação detalhada para os domicílios com a cobertura abrangente de um censo ou pesquisa de amostra maior. A idéia básica do procedimento é combinar o poder dos dois tipos de pesquisa (realizadas em períodos próximos) para que os estimadores calculados possam ser usados, de maneira satisfatória, em níveis mais desagregados do que os comumente considerados. DAVIS (2003) recomenda que as pesquisas tenham ocorrido em intervalo de tempo próximo e que o maior intervalo de tempo possível entre as pesquisas variará de acordo com a situação econômica do país em questão.

Numa aplicação a dados brasileiros, citamos anteriormente o estudo de CUNHA (2006), que encontrou valores moderados de coeficientes de determinação, 0,61 e 0,64 para consumo e rendimento, respectivamente. E, para fins de ilustração, consideremos a Tabela 3.1, a seguir, com resultados de coeficientes de determinação para a aplicação da metodologia em alguns países, sob diferentes situações, como diferentes unidades geográficas (estados, distritos, área urbana e rural etc) e para indicadores de rendimento e de consumo. Como pode ser observado, uma característica desse procedimento é a obtenção de valores de coeficientes de determinação moderados.



**Tabela 3.1: Resultados de R<sup>2</sup> ajustado de aplicações da metodologia do Banco Mundial em alguns países.**

Países	Resultados de R <sup>2</sup>
Sri Lanka	Entre 0,28 e 0,63
Papua Nova Guiné	0,34
Madagascar	Entre 0,24 e 0,64
Equador	Entre 0,46 e 0,74

Fonte: WORLD BANK (2005) para resultados do Sri Lanka; GIBSON et al. (2004) para Papua Nova Guiné; MISTIAEN et al. (2001) para o Madagascar e HENTSCHEL et al. (2000) para o Equador.

A idéia básica do procedimento consiste em utilizar os dados da pesquisa amostral menor e ajustar um modelo estatístico para qualquer indicador de nível domiciliar ou do indivíduo como uma função de variáveis que estão disponíveis em ambas as pesquisas.

As estimativas resultantes dos parâmetros desse modelo são então usadas num procedimento de imputação para prever as observações das variáveis de interesse, para cada domicílio, no censo. Em outras palavras, para cada domicílio do Censo Demográfico, multiplicamos suas características (variáveis) pelas estimativas dos parâmetros da regressão para obtermos um valor imputado para o logaritmo da renda domiciliar *per capita* (e do consumo domiciliar *per capita*). Com base nesses valores preditos, medidas de pobreza dos domicílios são então calculadas e agregadas para pequenas áreas, como bairros, municípios ou microrregiões. Cabe ressaltar que ao aplicar esta metodologia, levamos em consideração o plano amostral.

### **3.1.2.1 O modelo de renda domiciliar per capita (ou consumo domiciliar per capita)**

De acordo com ELBERS *et al.* (2001, 2002), o modelo empírico para a variável consumo domiciliar *per capita* (ou renda domiciliar *per capita*) é definido como:

$$\ln y_{cd} = E(y_{cd} | x_{cd}) + u_{cd}, \quad (3.1)$$

no qual  $\ln y_{cd}$  é o logaritmo da renda *per capita* (ou consumo total *per capita*) do domicílio  $d$  no conglomerado  $c$ ,  $x_{cd}$  é o vetor de características observadas deste domicílio (incluindo variáveis no nível de conglomerados), e  $u_{cd}$  é o termo residual do modelo e segue distribuição  $N(0, \Sigma)$ . Segundo os autores, o termo  $u_{cd}$  é não-correlacionado com  $x_{cd}$ . O modelo descrito acima pode ser reescrito usando uma aproximação linear da esperança condicional  $E(y_{cd} | x_{cd})$  e decompondo os termos residuais  $u_{cd}$  em parcelas não-correlacionadas:

$$u_{cd} = \eta_c + \varepsilon_{cd}, \quad (3.2)$$

no qual  $\eta_c$  representa um termo residual no nível do conglomerado em questão (microrregião) comum a todos os domicílios dentro do conglomerado, e o termo  $\varepsilon_{cd}$  representa o resíduo referente ao domicílio. Cabe ressaltar que  $\eta_c$  e  $\varepsilon_{cd}$  satisfazem:

$$E(\eta_c) = 0, \forall c, \quad E(\varepsilon_{cd}) = 0, \forall c, d, \quad \eta_c \perp \eta_e \forall c \neq e, \quad \varepsilon_{cd} \perp \varepsilon_{el} \forall (c, d) \neq (e, l),$$

e  $\varepsilon_{cd} \perp \eta_e \forall c, h, e,$  e  $\eta_c \sim N(0, \sigma_\eta^2)$  e  $\varepsilon_{cd} \sim N(0, \sigma_\varepsilon^2)$ .

Seja  $\sigma_u^2$  a variância dos resíduos do modelo no nível do domicílio  $d$  e conglomerado  $c$ .

Com base no modelo (3.2), esta variância pode ser decomposta em duas parcelas:

$$V(u_{cd}) = V(\eta_c) + V(\varepsilon_{cd}) = \sigma_\eta^2 + \sigma_\varepsilon^2.$$

Utilizando esta decomposição, é possível tentar incorporar os efeitos de grupo e variáveis preditoras do nível de conglomerados. Com base nessas premissas, tem-se que a equação (3.1) pode ser simplificada por:

$$\ln y_{cd} = x_{cd}\beta + \eta_c + \varepsilon_{cd}. \quad (3.3)$$

Vale dizer que, neste modelo, os conglomerados em questão são as microrregiões potiguares.

### 3.1.2.2 O procedimento de implementação

Primeiramente, os dados da pesquisa de menor amostra são usados para se ajustar um modelo de predição da variável de interesse. Com isso, as estimativas dos parâmetros da regressão são aplicadas aos dados da pesquisa maior para o cálculo das medidas de pobreza de interesse. Sucintamente, o procedimento de implementação do método pode ser sintetizado nos seguintes passos:

#### i) Preparação dos dados:

No intuito de obtermos estimativas dos níveis de consumo e de renda dos domicílios no Censo, as variáveis explicativas selecionadas para a determinação do modelo de consumo/renda têm que existir e terem sido captadas da mesma maneira em ambas as pesquisas. Supondo que a amostra da pesquisa amostral domiciliar menor tenha sido sorteada de maneira aleatória e possui representatividade nacional, espera-se que a distribuição de cada variável explicativa seja a mesma da distribuição das variáveis do Censo. A seguir, é detalhado o procedimento adotado para o cumprimento desta etapa do método.

Inicialmente, um conjunto de possíveis variáveis explicativas foi selecionado. Isso foi feito comparando-se os questionários e a documentação das duas pesquisas. Todas as variáveis existentes em ambas as pesquisas e que tinham definições parecidas compuseram o primeiro grupo de variáveis candidatas. Posteriormente, foi feita uma comparação entre as variáveis comuns às duas pesquisas, por meio de procedimentos estatísticos, como estatísticas descritivas e testes estatísticos para avaliar se as variáveis seguiam a mesma distribuição. Em caso afirmativo, as variáveis eram utilizadas na etapa posterior, a estimação do modelo de regressão. Assim, foi utilizado o teste de igualdade de distribuições (CUNHA, 2006). A seguir, apresenta-se uma descrição formal do teste.

Seja  $p_1$  o vetor de proporções das categorias da variável na pesquisa  $A$  e  $p_2$  o vetor de proporções das categorias da variável na pesquisa  $B$ . Cabe ressaltar que todas as variáveis utilizadas são discretas, originalmente ou foram discretizadas. A hipótese de igualdade de duas distribuições discretas pode ser descrita como  $H_0 : p_1 = p_2$ , excluindo-se uma categoria da variável testada, uma vez que a soma das proporções é igual a 1 em ambos os casos.

Sejam  $\hat{p}_l = (\hat{p}_{l1}, \dots, \hat{p}_{lg-1})'$  os vetores de proporções estimadas ( $l=1,2$ ), considerando-se amostras independentes, no qual  $\hat{p}_{lg} = \frac{\hat{N}_{lg}}{\hat{N}_{l+}}$  é um estimador

consistente da proporção  $p_{lg}$  na população correspondente, e  $\hat{N}_{lg}$  e  $\hat{N}_{l+}$  são estimadores ponderados das frequências das categorias da variável de modo que

$$\sum_{g=1}^G \hat{N}_{lg} = \hat{N}_{l+}.$$

Sejam agora,  $\hat{V}_p(\hat{p}_1)$  e  $\hat{V}_p(\hat{p}_2)$  estimadores consistentes das matrizes da variância amostral dos vetores  $\hat{p}_1$  e  $\hat{p}_2$ , respectivamente. De acordo com PESSOA e SILVA (1998), supondo independência entre as amostras das duas pesquisas, a estatística de Wald baseada no plano amostral para efetuar o teste de igualdade entre as distribuições é dada por:

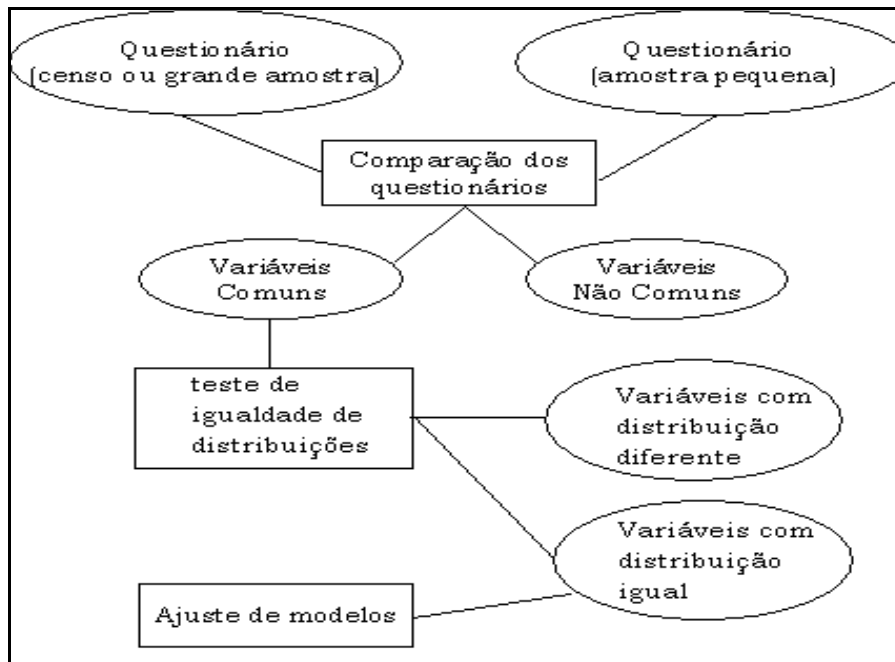
$$X_{Wald}^2(H) = (\hat{p}_1 - \hat{p}_2)' \left[ \hat{V}_p(\hat{p}_1) + \hat{V}_p(\hat{p}_2) \right]^{-1} (\hat{p}_1 - \hat{p}_2). \quad (3.4)$$

A estatística de Wald  $X_{Wald}^2(H)$  tem distribuição assintótica qui-quadrado com (G-1) graus de liberdade. Os mesmos autores dizem que esta estatística funciona adequadamente quando o número de unidades primárias na amostra é grande. Quando isto não ocorre, podem ocorrer problemas de instabilidade e usa-se, alternativamente, uma estatística F corrigida de Wald.

Para a realização do teste, é necessário o cálculo de estimativas da matriz de covariâncias dos estimadores das distribuições das variáveis estudadas. No entanto, de acordo com CUNHA (2006), como os estimadores são quantidades não lineares, o método de estimação da matriz de covariâncias consiste em combinar o método de linearização de Taylor com o método do conglomerado primário. Segundo a autora, a técnica de linearização de Taylor é utilizada na obtenção de aproximações assintóticas para a variância do estimador do vetor de proporções populacionais quando o vetor de parâmetros é função não linear das observações amostrais.

E o método do Conglomerado Primário é usado para a estimação de variâncias para estimadores de totais e médias em planos amostrais de múltiplos estágios, e considera apenas a variação entre informações disponíveis no nível das unidades primárias de amostragem, i.e., no nível dos conglomerados primários.

**Figura 3.1: Esquema de seleção das variáveis explicativas para o ajuste do modelo**



Fonte: Adaptado de CUNHA (2006).

**ii) Estimação dos parâmetros do modelo;**

Uma estimativa inicial para os  $\beta$ 's do modelo (3.1) pode ser obtida por mínimos quadrados ordinários ou mínimos quadrados ponderados. Matricialmente, o modelo pode ser escrito da seguinte forma:

$$\ln y = X\beta + u, \quad (3.5)$$

no qual  $\ln y$  representa o vetor de logaritmos da variável de interesse,  $\beta$  representa o vetor de coeficientes de regressão e  $u$  representa o vetor de termos residuais.

Os estimadores de  $\beta$  por mínimos quadrados ordinários, são obtidos através da seguinte equação:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Já o cálculo da matriz de variância-covariância de  $\hat{\beta}$  é estimado por:

$$\hat{V}(\hat{\beta}) = \sigma_u^2 (X'X)^{-1},$$

no qual  $\sigma_u^2$  representa a variância homocedástica dos resíduos da regressão ( $u$ ), dada por:

$$\sigma_u^2 = \frac{\sum_c \sum_d \hat{u}_{dc}^2}{n - g},$$

lembrando que  $n$  representa o número de observações e  $k$  o número de parâmetros do modelo.

Seguindo ELBERS *et. al.* (2001 e 2002), suponha que tenhamos observações de  $u_{cd}$ , no qual  $u_{cd} = \eta_c + \varepsilon_{cd}$ . Temos que estimar  $\sigma_\eta^2 = \text{Var}(\eta)$ . Logo:

$$E(u_c^2) = \sigma_\eta^2 + \sigma_\varepsilon^2 \tag{3.6}$$

Sendo que as estimativas para  $\sigma_\eta^2$  e  $\sigma_\varepsilon^2$  podem ser obtidas por meio de:

$$\hat{\sigma}_\eta^2 = \max \left( \frac{\sum_c w_c (u_c - u_{..})^2}{\sum_j w_j (1 - w_j)} - \frac{\sum_c w_c (1 - w_c) \hat{\sigma}_\varepsilon^2}{\sum_j w_j (1 - w_j)}; 0 \right) e$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n_c (n_c - 1)} \sum_d (\varepsilon_{cd} - \varepsilon_c)^2.$$

O estimador da variância de  $\sigma_\eta^2$  pode ser calculado usando simulações.

Em seguida, deve-se estimar a variância do modelo de heterocedasticidade  $\sigma_{\varepsilon, cd}^2$ .

Para maiores detalhes, ver ELBERS et. al. (2002). De posse das componentes de variância estimadas e assumindo que  $\eta_c$  e  $\varepsilon_{cd}$  são independentes e seguem uma distribuição normal com média zero, são gerados novos valores para  $u_{cd}$ , com base em  $u_{cd} = \eta_c + \varepsilon_{cd}$ . O procedimento é repetido várias vezes. De acordo com ELBERS et. al. (2002), o conjunto de valores simulados para  $\sigma_\eta^2$  pode ser usado para estimar a variância amostral de  $\hat{\sigma}_\eta^2$ , diretamente. No mesmo estudo, os autores ainda relatam que, na prática,  $\hat{\sigma}_\eta^2$  muitas vezes é tão pequeno que a equação pode fornecer estimativas nulas para  $\sigma_\eta^2$ , ou seja, o modelo (3.2) não teria efeitos de grupo. Resta agora simular os resíduos  $\eta$  e  $\varepsilon$ . Os autores destacam que uma distribuição adequada pode ser obtida a partir do resíduo de conglomerado  $\hat{\eta}_c$  e do resíduo padronizado de domicílio.

$$e_{cd}^* = \frac{e_{cd}}{\hat{\sigma}_{\varepsilon, cd}} - \left[ \frac{1}{H_c} \sum_h \frac{e_{cd}}{\hat{\sigma}_{\varepsilon, cd}} \right],$$



no qual,  $H_c$  é o número total de domicílios no conglomerado  $c$ . CUNHA (2006) diz que estimar diretamente dos resíduos padronizados evita fazer qualquer suposição a respeito da distribuição dos resíduos. De maneira alternativa, os percentis da distribuição empírica dos resíduos padronizados podem ser comparados com os percentis correspondentes de uma distribuição normal padrão.

### iii) Imputação dos valores das variáveis de interesse no Censo;

Para a imputação dos valores da renda domiciliar per capita e da despesa domiciliar per capita dos domicílios na amostra do Censo, as estimativas dos parâmetros, obtidas com base nos dados da amostra menor (no caso, POF 2002-2003) são utilizados na seguinte equação:

$$\hat{y}_{cd}^r = \exp[x'_{cd} \hat{\beta} + u_{cd}^r], \quad (3.7)$$

no qual  $\hat{\beta}$  representa o vetor de parâmetros do modelo estimado com base em 3.1;  $x_{cd}$  representa o vetor de valores das variáveis preditoras para o domicílio  $d$  do conglomerado  $c$  observado no Censo e,  $\hat{u}_{cd}^r$ , o termo residual é estimado por simulações, por abordagens paramétrica e semi-paramétrica. É importante ressaltar que não há ligação entre os dados individuais do domicílio da pesquisa de amostra menor (no caso, a POF) e na pesquisa de amostra maior (no caso, a parte amostral do censo).

Sob a abordagem paramétrica, os resíduos  $u_{cd}$ , do modelo 3.1 são vistos como originários de uma distribuição normal de média 0 e matriz de covariância  $\Sigma$ . Logo, para simulá-los basta gerar vetores de resíduos com a distribuição  $N(0, \hat{\Sigma})$ , no

qual  $\hat{\Sigma}$  representa o estimador de  $\Sigma$ . O procedimento de simulação é repetido inúmeras vezes (no caso, 100 vezes), sempre de maneira independente.

Na abordagem semi-paramétrica, a idéia é estimar  $\hat{u}_{cd}^r$  com base na soma de dois resíduos empíricos,  $\hat{\eta}_c^r$  e  $e_{cd}^{*r}$ .

Num primeiro momento, o efeito do nível de conglomerado  $\hat{\eta}_c^r$  é selecionado de maneira aleatória e com reposição do rol de todos os resíduos empíricos  $\hat{\eta}_c$ . Posteriormente, o termo  $e_{cd}^{*r}$  é obtido da mesma forma que o termo  $\hat{\eta}_c^r$ . A soma destes dois resíduos gera o valor do resíduo a ser utilizado na equação (3.6) para obter os valores das variáveis dependentes (renda e consumo domiciliar per capita) de cada réplica simulada no Censo.

O processo de imputação é repetido um número  $R$  de vezes (no caso, 100), gerando a cada vez uma réplica dos valores das variáveis dependentes para todos os domicílios existentes no Censo. De posse de cada conjunto de valores simulado, obtém-se um valor para o indicador de pobreza para cada uma das áreas de interesse, e também para o conjunto da população. O valor utilizado como estimador para o indicador de pobreza para cada área é a média das  $R$  réplicas.

### 3.2 A abordagem bayesiana

Enquanto os métodos da estatística clássica baseiam suas decisões na análise de dados amostrais, a estatística bayesiana incorpora, a estes dados amostrais, a subjetividade do pesquisador. A estatística bayesiana utiliza-se da função de verossimilhança e de uma informação adicional (subjetiva, uma vez que expressa a incerteza do pesquisador sobre o parâmetro, antes da observação dos dados). Esta incerteza é expressa por uma distribuição de probabilidade. Combinando-se a função de verossimilhança com a distribuição a priori por meio do Teorema de Bayes, obtém-

se a distribuição a posteriori (distribuição dos parâmetros do modelo condicionado aos valores observados na amostra).

Na abordagem bayesiana, o processo de inferência é baseado na distribuição do vetor de parâmetros  $\theta$  de interesse, após a observação dos dados. Esta distribuição é denominada distribuição a posteriori, e é obtida seguindo o teorema da bayes.

$$p(\theta | y) \propto f_n(Y | \theta)p(\theta),$$

onde  $f_n(Y | \theta)$  representa a função de verossimilhança, que descreve a informação proveniente dos dados, e  $p(\theta)$  representa a distribuição a priori, que descreve a informação que temos sobre  $\theta$  antes da observação dos dados.

Em muitas situações práticas, não é possível obter a distribuição a posteriori em forma fechada, ou seja, de forma analítica. Portanto, é preciso fazer uso de métodos de simulação estocástica para obter amostras de uma determinada densidade a posteriori, cuja geração direta é de difícil implementação. Pode-se dizer que um algoritmo de Monte Carlo via Cadeia de Markov para simular  $p(\cdot)$  é qualquer método que produza uma cadeia de Markov homogênea, ergódica e irredutível cuja distribuição estacionária seja  $p(\cdot)$ . A idéia de tais métodos é obter uma amostra da distribuição de interesse e calcular estimativas de características dessa distribuição, utilizando-se de técnicas de simulação iterativa, com base em Cadeias de Markov. Para maiores detalhes, ver GAMERMAN (1997).

A seguir, são apresentados os métodos MCMC mais utilizados, o algoritmo Amostrador de Gibbs e o Metropolis-Hastings.

#### **i) Amostrador de Gibbs**

É um algoritmo que gera uma seqüência

$$\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots\},$$

a partir de uma cadeia de Markov, cuja distribuição limite é  $p(\theta)$  e cujo núcleo de transição é formado pelo produto das distribuições condicionais completas.  $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots\}$  representam amostras dos parâmetros do modelo. Maiores detalhes sobre o Amostrador de Gibbs podem ser obtidos, por exemplo, em GAMERMAN (1997). Sucintamente, pode se dizer que a atualização do Amostrador de Gibbs é dada pelas seguintes etapas:

a) Iniciar as iterações a partir de um ponto arbitrário

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)}) \text{ e fazer } j = 1;$$

b) Fazer  $i = 1$ ;

c) Simular  $\theta_i^{(j)} \sim \pi(\cdot, \theta_i^{(j)})$ ;

d) Se  $i < p$ , fazer  $i = i + 1$  e retornar a etapa (c);

e) Fazer  $j = j + 1$  e retornar a etapa (b) até obter convergência.

### **i) Metropolis-Hastings**

Da mesma forma que o Amostrador de Gibbs, é um algoritmo que gera uma seqüência

$$\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots\},$$

a partir de uma cadeia de Markov, cuja distribuição limite é  $p(\theta)$ .

O algoritmo de Metropolis-Hastings será usado para amostrar das distribuições condicionais completas dos parâmetros que não possuem condicionais completas de forma fechada. Maiores detalhes em METROPOLIS et al. (1953) e HASTINGS (1970). Basicamente, o algoritmo consiste nos seguintes passos:

- 1) Na  $i$ -ésima iteração do algoritmo, sorteamos um valor proposto para  $\theta, \xi$  de uma distribuição proposta  $q(\xi | \theta^{(i-1)})$ ;
- 2) Aceitamos o valor proposto com probabilidade  $\alpha$ , onde

$$\alpha = \min \left\{ 1, \frac{\frac{p(\xi)}{q(\xi | \theta^{(i-1)})}}{\frac{p(\theta^{(i-1)})}{q(\theta^{(i-1)} | \xi)}} \right\},$$

caso contrário, permanecemos com o valor corrente de  $\theta$ .

De posse da simulação da cadeia, é necessário verificar se a convergência foi atingida. Em caso afirmativo, os valores simulados podem ser considerados como amostra da distribuição a posteriori das quantidades desconhecidas do modelo, sendo, portanto, utilizados para fazermos inferência. Vale ressaltar que os valores da amostra obtidos antes da convergência devem ser desconsiderados da análise.

A implementação computacional na estatística bayesiana merece especial destaque, uma vez que diversos autores mostraram que as técnicas bayesianas, para a maioria dos propósitos, apresentam vantagens ou melhorias quando comparadas com demais métodos (GOSH e RAO, 1994). De acordo com GAMERMAN (1997), um dos grandes entraves no desenvolvimento da Inferência Bayesiana tem sido a sua dificuldade de implementação em problemas reais. São exemplos dessas dificuldades: a especificação da distribuição à priori e a convergência da posteriori resultante. O autor afirma que a primeira dificuldade está sendo, aos poucos, eliminada pela disponibilidade de linguagens simbólicas nas quais várias especificações para a

distribuição a priori podem ser facilmente implementadas em programas computacionais. Em relação à segunda dificuldade, o comentário feito é de que ela está sendo minimizada devido à introdução de métodos iterativos como o Amostrador de Gibbs, possibilitando a análise de modelos complexos por meio de sucessivas decomposições em distribuições condicionais completas.

O programa computacional utilizado para a análise bayesiana é o WinBugs, versão em plataforma Windows do programa Bugs (*Bayesian Using Gibbs Sampling*), implementado por THOMAS *et. al.* (1992). O referido programa utiliza um conjunto de procedimentos que permite a especificação de diversos tipos de modelos e faz a estimação de seus parâmetros via MCMC. Além do Amostrador de Gibbs, o programa utiliza o algoritmo de Metropolis-Hastings. Nos modelos sob abordagem bayesiana, o plano amostral não foi considerado nas análises.

### 3.2.1 Modelos Lineares Generalizados

A classe dos modelos lineares generalizados é uma extensão dos modelos lineares. Tal classe de modelos relaxa a suposição de que  $Y$  segue distribuição normal e permite  $Y$  seguir qualquer distribuição pertencente à família exponencial com um parâmetro, na forma canônica, como no presente caso em que analisamos uma variável dicotômica (1 – pobre e 0 – não pobre). RESENDE e BIELE (2002) destacam que as generalizações ocorrem em duas direções: (i) permitem que a esperança  $\pi$ , de  $Y$ , seja uma função monotonicamente diferenciável do preditor linear  $\eta = \sum x_i \beta_i$  de forma que  $\pi = f(\eta) = f(\sum x_i \beta_i)$ ; (ii) ou, por inversão,  $g(\pi) = \eta$ , em que  $g$  é a função de ligação, a qual liga a média ao preditor linear. Maiores informações sobre estes modelos podem ser encontrados em McCULLAGH e NELDER (1989).

### 3.2.2 Modelos Hierárquicos propostos

De acordo com SNIJDERS e BOSKER (1999), uma característica comum nas pesquisas sociais é fazer inferências tanto para indivíduo quanto para o grupo no qual ele pertence. No mesmo contexto, DUNCAN *et. al.* (1998) destacam que a noção de que os indivíduos não são afetados por fatores sociais, culturais, econômicos ou legislativos e, que definem suas atitudes livremente, precisa ser refutada.

A utilização de estruturas hierárquicas em modelos lineares generalizados pode ser muito útil, sobretudo nas Ciências Sociais. Geralmente, os indivíduos e os grupos sociais são conceituados como um sistema hierárquico de indivíduos e grupos.

Para ilustrar, consideremos o modelo com dois níveis de hierarquia do atual estudo, para a variável resposta  $y_{ij}$ , variável binária, dada por:

$$y_{ij} = \begin{cases} 1, & \text{se a unidade } ij \text{ possui o atributo de interesse, ser pobre.} \\ 0, & \text{caso contrário} \end{cases}$$

onde o índice  $ij$  representa os dois níveis hierárquicos estabelecidos,  $j$  representa o índice da unidade amostral no último nível (nível mais desagregado; no caso, domicílios),  $i$  representa o índice correspondente às pequenas áreas (segundo nível; no caso, microrregiões). Então,  $y_{ij} = 10$ , por exemplo, indica a resposta da  $j$ -ésima unidade amostral no último nível (domicílios) dentro da  $i$ -ésima unidade amostral no primeiro nível correspondente às pequenas áreas (microrregiões).

Supõe-se que a variável resposta de interesse  $y_{ij}$  tenha a seguinte distribuição de probabilidades:

$$y_{ij} | \pi_{ij} \sim \text{Bernoulli}(\pi_{ij}),$$

Para relacionar a probabilidade de ser pobre  $\pi_{ij}$  e o vetor de variáveis explicativas, propõe-se o uso do seguinte modelo de regressão logística com efeitos aleatórios:

$$\text{Logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = X_{ij}^T \beta_i, \quad i = 1, \dots, m. \quad (3.9)$$

onde:  $\beta_i = Z_i \gamma + u_i$ ,

$\beta_i$  é um vetor de parâmetros, no qual todas as componentes do vetor podem assumir efeitos aleatórios.  $X_{ij}^T$  representa o vetor de variáveis explicativas associadas aos efeitos fixos,  $Z_i$  é um vetor de variáveis associadas com as pequenas áreas,  $\gamma$  é o vetor de parâmetros de efeitos fixos na regressão logística. A quantidade  $u_i$  tem média 0 e variância  $\Phi$ , dadas pela matriz descrita na página 39.

Modelos com mais níveis de hierarquia podem ser estendidos, naturalmente. E, para o caso particular em que apenas o intercepto varia com as pequenas áreas, tem-se que  $v_i = u_i + e_{ij}$ , no qual  $u_i$  é um efeito aleatório para a  $i$ -ésima pequena área, o termo  $e_{ij}$  representa o efeito para a  $j$ -ésima unidade amostral do segundo nível de hierarquia, dentro da  $i$ -ésima pequena área.

Cabe mencionar que a obtenção de amostras de populações, com estrutura hierárquica, pode ser feita em múltiplos estágios. Primeiramente, considerando-se amostras de unidades no nível maior (por exemplo, de microrregiões) e, posteriormente, amostrando-se subunidades destas unidades selecionadas (como, por exemplo, domicílios dentro das microrregiões). Pode-se dizer que, em tais amostras, as unidades individuais não são totalmente independentes (algumas



microrregiões podem atrair pessoas com alto poder aquisitivo; entretanto, outras microrregiões podem atrair pessoas mais pobres). Como resultado, a correlação intraclasse pode ser alta (domicílios numa mesma microrregião), no entanto a correlação extraclasse pode ser baixa (entre microrregiões). No que se refere ao significado do coeficiente de correlação intraclasse, GOLDSTEIN (1995) afirma que esta estatística mede o quanto da variação total é explicada pelo nível maior.

### 3.2.3 Modelo Autoregressivo Condicional (CAR) proposto

Outro aspecto importante dos dados do presente estudo, é que supõe-se a existência de uma autocorrelação espacial entre eles; sendo assim, as estimativas do modelo devem incorporar esta estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo.

Quando se trata do estudo de dependência espacial de dados de área, um elemento muito importante na análise é o conceito de matriz de proximidade espacial. A matriz de proximidade espacial ( $W$ ) estima a variabilidade espacial de dados de área. É uma matriz quadrada, não estocástica, cujos elementos  $w_{ij}$  refletem a intensidade da interdependência existente entre as regiões  $i$  e  $j$  (CRESSIE, 1993). Na matriz  $W$  ( $n \times n$ ) cada elemento  $w_{ij}$  representa uma medida de proximidade entre  $A_i$  e  $A_j$ , sendo  $A_i$  e  $A_j$  as zonas (pequenas áreas) que estão sendo analisadas.

O valor de  $W$  pode ser baseado em algumas medidas de adjacência, tais como:

- $w_{ij} = 1$ , se o centróide de  $A_i$  está a uma determinada distância de  $A_j$ , e  $w_{ij} = 0$ , caso contrário;
- $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , e  $w_{ij} = 0$ , caso contrário;

- $w_{ij} = 1_{ij} / 1_i$ , onde  $1_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $1_i$  é o perímetro de  $A_i$ .

Nesta dissertação, a proximidade espacial foi considerada em termos de vizinhança, ou seja, de microrregiões adjacentes. A matriz de vizinhança foi construída no programa computacional ArcView 3.0 e a quantidade de vizinhos de cada microrregião foi calculada pelo programa Geoda através do método de Hook. Esses dois resultados, ou seja, uma matriz compostas por 0 e 1 e outra com a quantidade de vizinhos de cada microrregião foram armazenados no programa computacional WinBUGS para a modelagem estatística.

A análise de regressão em dados espaciais incorpora, na modelagem, a dependência espacial entre os dados, melhorando o poder preditivo do modelo. SCHIMIDT et. al. (2003) apresentam dois tipos básicos de modelagem que permitem incorporar o efeito espacial: as de forma local e as de forma global.

- ***Modelos com efeitos espaciais locais***

Nesse caso, assume-se que a natureza do processo espacial é de tal forma que diversos padrões de associação espacial podem estar presentes. Em tais casos, devem-se aplicar modelos que variam no espaço, chamados de modelos com efeitos espaciais locais. Esta dissertação não considerou tal situação. Para maiores detalhes sobre esses modelos podem ser obtidos em CÂMARA et. al. (2002).

Partimos do princípio que o processo espacial subjacente aos dados era estacionário. Isso implica que os padrões de autocorrelação espacial existentes nos dados podem ser capturados num único parâmetro.

- **Modelos com efeitos espaciais globais**

De acordo com CÂMARA (2002), os modelos de forma global capturam a estrutura espacial através de um único parâmetro que é adicionado ao modelo de regressão tradicional. Os modelos mais simples são: modelo espacial auto-regressivo misto (*Spatial Auto-Regressive = SAR*) e modelo condicional autoregressivo ou modelo de erro espacial (*Conditional Auto-Regressive = CAR*).

No modelo SAR, a autocorrelação espacial ignorada é atribuída à variável dependente. Considera-se a dependência espacial através da adição, ao modelo de regressão, de um novo termo na forma de uma relação espacial para a variável dependente. Maiores detalhes sobre estes modelos podem ser encontrados em CRESSIE (1993) e em ANSELIN (2002).

Na segunda alternativa de análise, considera-se que os efeitos espaciais são um ruído, ou seja, um fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo residual  $\varepsilon$  e o modelo pode ser expresso por:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W + \xi,$$

no qual  $W$  é a matriz de proximidade espacial (ou matriz de vizinhança), onde vizinho é definido aqui como sendo uma microrregião adjacente;  $\lambda$  é o coeficiente autoregressivo e  $\xi$  é a componente do erro com variância constante e não correlacionada.

Neste trabalho, o modelo CAR foi preferido em relação ao modelo SAR por apresentar a propriedade de que sua especificação fornece diretamente as distribuições condicionais completas dos parâmetros do modelo, fator determinante

para o uso do amostrador de Gibbs em Monte Carlo via Cadeia de Markov. E, de acordo com SCHMIDT et. al. (2003), no contexto bayesiano geralmente o modelo CAR é usado como informação a priori de um parâmetro do modelo para o processo de interesse. A seguir descreveremos, em detalhes, um modelo genérico para exemplificar a modelagem bayesiana de um processo condicionalmente autoregressivo.

Assumindo que a área de interesse possa ser dividida em  $m$  sub-regiões, regulares ou não, e  $Y_i$  representa a quantidade de interesse que observamos em cada sub-região  $i$ ,  $i = 1, \dots, m$ . Um possível modelo para  $Y_i$  é:

$$Y_i = \mu + \sum_{j=1}^t \alpha_j X_{ij} + b_i, \quad i = 1, \dots, m; j = 1, \dots, t,$$

onde  $\mu$  representa um nível geral, comum a toda região sob estudo;  $X_i = (X_{i1}, \dots, X_{it})$  um vetor de possíveis covariáveis para a  $i$ -ésima área, que podem explicar o processo, e  $\alpha_j$  o efeito da  $j$ -ésima covariável na resposta  $Y_i$ , as componentes  $b = (b_1, \dots, b_m)$  são efeitos aleatórios que podem ser vistos como variáveis latentes que capturam efeitos desconhecidos ou não medidos pelas covariáveis. Supondo que essas covariáveis não medidas apresentam uma estrutura espacial (por exemplo, efeitos ambientais ou características sociais não medidos) então o modelo para  $b_i$  deve permitir a presença de tal estrutura.

Sob o enfoque bayesiano, a equação anterior representa o primeiro nível de hierarquia do modelo. No segundo nível devemos especificar a distribuição a priori do vetor paramétrico  $\theta = (\mu, \alpha_1, \alpha_2, \dots, \alpha_j, b)$ . Geralmente assume-se a priori que esses parâmetros são independentes e que  $\mu, \alpha_1, \alpha_2, \dots, \alpha_j, b$ , seguem uma distribuição a priori normal centrada em 0 (zero) com baixa precisão. Dessa forma, deixamos que os dados nos dêem maiores informações sobre tais parâmetros. Para

os efeitos aleatórios  $b_i$ 's, assumimos uma priori autoregressiva condicional intrínseca.

Maiores detalhes sobre tais modelos podem ser obtidos em SCHMIDT et. al. (2003).

Especificamente, analisamos um modelo logístico sob a seguinte forma:

$$\text{Logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = X_{ij}^T \beta_i + b_{j(i)}$$

Aqui o vetor  $X_j$  multiplicado pelo vetor de parâmetros da regressão  $\beta_i$  representa os efeitos fixos. As componentes  $b = (b_1, \dots, b_j)$  são efeitos aleatórios que podem ser vistos como variáveis latentes que capturam efeitos desconhecidos ou não medidos pelas covariáveis. Vale ressaltar que essa abordagem requer uma estrutura de vizinhança espacial. Consideramos duas microrregiões como vizinhas se elas fossem adjacentes.

# **CAPÍTULO 4**

## **APLICAÇÃO DOS MÉTODOS ESTATÍSTICOS**

Neste capítulo, são implementados os métodos estatísticos considerados no presente estudo, o ajuste dos modelos e, finalmente, apresentados os resultados dos modelos aplicados. Deve-se destacar que para a modelagem da Estatística Clássica foi utilizado o pacote estatístico SAS, com códigos fornecidos por CUNHA (2006), que tinham sido previamente adaptados do programa estatístico SPSS (fornecidos pelo Banco Mundial) para a linguagem SAS pela equipe da Coordenadoria de Metodologia e Qualidade do IBGE.

### **4.1 Modelagem sob a abordagem clássica**

CUNHA (2006) afirma que, diferente dos modelos estatísticos usuais, esses modelos propostos pelos pesquisadores do Banco Mundial não precisam ter coerência estatística ou econômica, mas sim um grande poder de predição. A grande finalidade da aplicação desses modelos é explicar, de maneira adequada, a variação dos indicadores de pobreza.

Após termos feito a identificação de variáveis comuns nas duas pesquisas, obtendo um total de 31 perguntas comuns, e realizado os testes de hipóteses para verificar quais variáveis tinham proporções estatisticamente iguais ( $p. 0,05$ ), procedemos ao ajuste dos modelos.

O ajuste dos modelos estatísticos é uma etapa de grande importância na construção do mapeamento dos indicadores de pobreza. A partir dos modelos ajustados, com base na POF, os valores para as variáveis dependentes serão imputados nos dados da amostra do Censo Demográfico de 2000, numa tentativa de ter informação para um

número maior de domicílios ou, ainda, atualizar a informação que fora pesquisada pelo censo.

Num primeiro momento do ajuste de modelos, adotou-se um método para a seleção de modelos - seleção passo a passo pra frente (do inglês, *forward selection*). Neste método de seleção de variáveis, a idéia é adicionar as variáveis (uma a uma) por importância ao modelo. Isso significa adicionar aquela variável que apresentar o maior valor da estatística F.

De início, são ajustados modelos com apenas uma variável explicativa, e a primeira variável a ser adicionada no modelo é aquela que apresenta maior estatística F. Sendo que este F tem que ser maior que um nível predeterminado, comumente esse valor adotado é igual a 4.

Este procedimento foi repetido até que a estatística F calculada fosse inferior a 4. Nas Tabelas 4.1 e 4.2 são apresentados os valores da estatística de teste correspondentes ao ajuste de cada modelo. Nota-se que o modelo para o logaritmo da renda domiciliar per capita contém 17 variáveis, e para o logaritmo do consumo domiciliar per capita, 12.

**Tabela 4.1: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da renda domiciliar per capita**

<b>n</b>	<b>Variável adicionada</b>	<b>F parcial</b>
1	Posse de computador	385,73
2	Posse de microondas	156,7
3	Número de moradores	65,94
4	Posse de máquina de lavar	114,54
5	Posse de aparelho de ar condicionado	41,59
6	Existência de pavimentação	34,15
7	Qtde de banheiros	32,52
8	Posse de TV	18,58
9	Existência de Iluminação elétrica no domicílio	3,08
10	Anos de estudo	16,14
11	Número de idosos	14,53
12	Condição do domicílio	11,22
13	Tipo de escoamento	7,58
14	Existência de cônjuge	7,49
15	Cor do chefe do domicílio	6,83
16	Posição na ocupação	5,95
17	Número de adolescentes	4,09

**Tabela 4.2: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo do consumo domiciliar per capita.**

<b>n</b>	<b>Variável adicionada</b>	<b>F parcial</b>
1	Posse de computador	325,12
2	Posse de microondas	135,26
3	Número de moradores	119,97
4	Posse de máquina de lavar	79,37
5	Posse de TV	51,07
6	Posse de aparelho de ar condicionado	38,84
7	Qtde de banheiros	26,95
8	Condição do domicílio	10,87
9	Anos de estudo	9,89
10	Existência de pavimentação	9,78
11	Tipo de domicílio	5,00
12	Número de adolescentes	4,77

Com base nas duas tabelas anteriores, percebe-se o quão importante é o fato de ter computador no domicílio. Esta variável é a que mais contribui nos dois modelos, o de rendimento domiciliar per capita e de consumo domiciliar per capita. Outros aparelhos

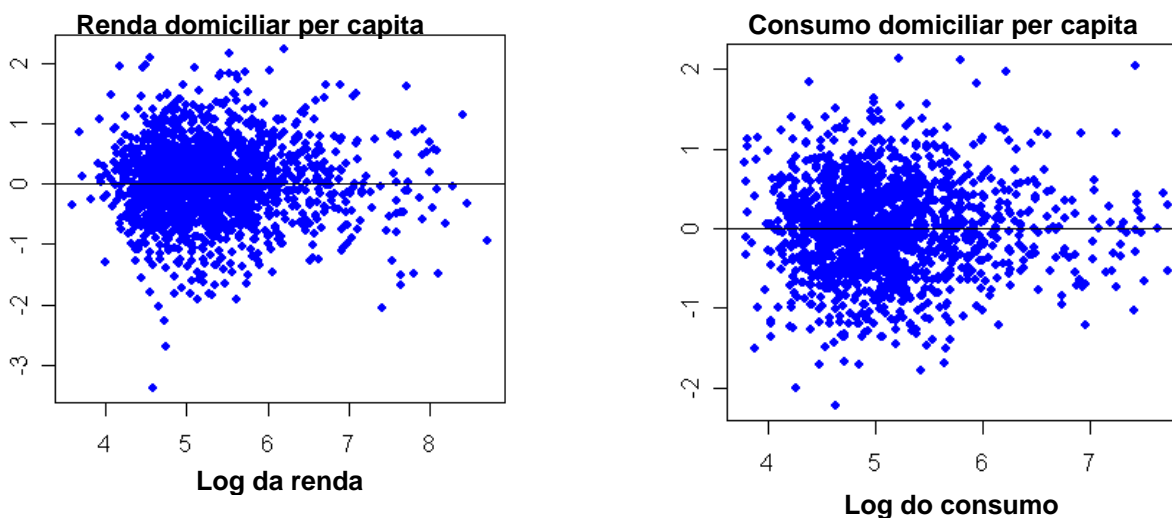


eletrodomésticos também se destacam, nesse mesmo contexto, como é o caso de máquina de lavar e forno microondas.

As estimativas dos parâmetros e dos p-valores dos dois modelos de efeitos principais são apresentados nas Tabelas A2 e A3, em anexo.

Feito o ajuste dos modelos estatísticos, usados para imputação das variáveis de interesse nos dados da amostra do Censo, é necessário mensurar a qualidade do ajuste realizado. Isso foi feito com base em dois quesitos: avaliação dos gráficos dos resíduos contra os valores preditos e análise do poder de predição ( $R^2$ ).

**Gráfico 4.1: Resíduos dos modelos de regressão para as variáveis dependentes, logaritmo da renda e logaritmo do consumo, respectivamente, Rio Grande do Norte.**



Com base nesses gráficos de resíduos versus valores preditos, nota-se que não há nenhum padrão definido nos dados, com pontos distribuídos de maneira aleatória em torno do valor zero. Nota-se a presença de alguns pontos mais distantes da nuvem de pontos, que podem ser considerados como observações discrepantes e que deveriam ser analisados com mais detalhes. Mas não acreditamos que essas informações possam afetar a qualidade dos resultados obtidos.

No que se refere à capacidade preditiva dos modelos, é possível observar, com base na Tabela 4.3, que os dois modelos (o de rendimento e o de consumo) apresentam valores moderados de explicação. O primeiro modelo (referente ao logaritmo do rendimento domiciliar per capita) apresenta R<sup>2</sup> de 0,5928, e o segundo de 0,6153, ou seja, em ambos os casos os modelos explicam em torno de 60% da variação total do logaritmo da renda domiciliar per capita e do logaritmo do consumo domiciliar per capita.

**Tabela 4.3: R<sup>2</sup> dos modelos para o logaritmo da renda e do consumo**

<b>Variável dependente</b>	<b>R<sup>2</sup></b>
Logaritmo do Rendimento Domiciliar per capita	0,5928
Logaritmo do Consumo Domiciliar per capita	0,6153

Como ponto de corte para a classificação dos indivíduos em pobres e não pobres, a distribuição dos valores de renda domiciliar per capita e consumo domiciliar per capita foi categorizada em valores acima da linha de pobreza e valores abaixo da linha de pobreza, no caso R\$100,00. A escolha dessa linha de pobreza é arbitrária, e a justificativa de escolha desse valor foi mencionada anteriormente.

A proporção populacional na *i*-ésima pequena área de interesse, ou seja, microrregião, ( $\theta_i$ ) para a característica de interesse *y*, pode ser escrita como:

$$\theta_i = \frac{\sum_j y_{ij}}{N_i}, \quad (4.1)$$

No qual *j* são as unidades correspondentes ao segundo nível (domicílio);  $y_{ij}$  representa a característica de interesse para toda a população;  $N_i$  representa o número total de unidades na *i*-ésima pequena área, tendo como premissa que  $N_i$  é conhecido. No

presente caso, como mencionado anteriormente, o  $N_i$  refere-se à amostra do Censo. Ou seja, o estimador para a proporção populacional é a média das réplicas do processo de imputação, descrito na seção 3.1.2.2, no passo (iii).

O indicador foi obtido a partir da média das 100 réplicas das variáveis respostas sob estudo imputadas na amostra do Censo 2000, conforme descrito no item 3.1.2.2. As réplicas foram calculadas com base numa distribuição normal. Em cada domicílio, foram calculadas as médias destas 100 réplicas e este resultado é usado como estimativa da proporção de pobres nas 19 microrregiões potiguares. As Tabelas 4.4 e 4.5 apresentam os resultados do indicador da proporção de domicílios pobres (média) estimado por microrregião, baseados na renda e consumo domiciliar per capita imputadas, respectivamente, bem como o esquema dos 5 números e os percentis 3 e 97.

**Tabela 4.4: Distribuição dos 5 números e média da proporção de pobres, baseados no Rendimento domiciliar per capita imputado, Rio Grande do Norte e Microrregiões, 2003.**

<b>Estado e Microrregiões</b>	<b>Mín</b>	<b>P03</b>	<b>Q1</b>	<b>Média</b>	<b>Q3</b>	<b>P97</b>	<b>Máx</b>
<b>Estado</b>	34,6	34,7	34,9	36,1	35,8	36,5	36,7
Mossoró	27,7	29,2	29,5	30,4	30,7	30,9	31,0
Chapada do Apodi	41,0	41,2	41,7	42,2	43,1	44,6	45,3
Médio Oeste	45,2	45,3	47,4	48,4	48,8	50,6	51,4
Vale do Açu	37,6	38,2	38,5	39,7	40,0	40,4	40,7
Serra de São Miguel	46,0	46,2	47,5	49,1	49,5	50,5	50,9
Pau dos Ferros	40,2	41,0	41,5	42,2	42,9	43,1	43,6
Umarizal	40,3	41,2	42,2	42,6	43,9	44,0	44,2
Macau	32,0	32,4	34,1	34,3	35,2	37,2	37,9
Angicos	42,4	43,2	44,7	45,6	46,8	47,0	47,6
Serra de Santana	47,9	48,2	49,9	51,3	51,7	52,0	52,4
Seridó Ocidental	29,2	29,4	29,9	30,6	31,1	32,1	32,9
Seridó Oriental	32,3	32,5	33,4	34,1	34,5	35,1	35,4
Baixa Verde	46,6	47,1	48,1	49,2	50,6	52,7	53,2
Borborema Potiguar	48,2	48,3	48,6	49,1	50,6	50,9	51,4
Agreste Potiguar	48,6	48,7	48,9	49,2	50,3	50,6	50,9
Litoral Nordeste	49,0	49,6	50,1	52,3	52,8	53,1	53,6
Macaíba	37,0	37,2	38,1	39,4	40,2	40,9	41,1
Natal	22,0	22,1	22,3	22,5	22,8	23,1	23,4
Litoral Sul	44,6	45,1	46,8	47,1	48,0	48,3	48,8

**Tabela 4.5: Distribuição dos 5 números e média da proporção de pobres, baseados no consumo domiciliar per capita imputado, Rio Grande do Norte e Microrregiões, 2003.**

<b>Estado e Microrregiões</b>	<b>Mín</b>	<b>P03</b>	<b>Q1</b>	<b>Média</b>	<b>Q3</b>	<b>P97</b>	<b>Máx</b>
<b>Estado</b>	41,8	42,1	43,0	44,7	46,1	49,6	50,9
Mossoró	35,1	35,3	35,5	35,8	36,0	36,9	37,3
Chapada do Apodi	49,7	49,9	51,0	51,3	51,6	52,2	52,9
Médio Oeste	53,9	54,3	54,9	57,0	57,3	57,8	58,8
Vale do Açu	45,3	45,5	46,1	48,3	48,8	49,9	50,2
Serra de São Miguel	51,0	51,6	52,0	53,8	54,2	55,0	55,2
Pau dos Ferros	44,9	45,0	45,5	46,8	47,0	47,2	47,3
Umarizal	49,7	50,0	51,3	52,1	53,1	54,0	54,1
Macau	42,0	42,3	43,6	45,5	46,7	48,3	49,2
Angicos	50,1	50,5	51,0	52,7	54,1	54,9	56,3
Serra de Santana	55,4	55,7	56,3	59,2	59,4	59,7	59,9
Seridó Ocidental	35,8	35,9	36,2	36,8	37,2	38,7	39,2
Seridó Oriental	39,3	39,7	39,9	41,3	41,6	42,9	43,1
Baixa Verde	55,0	55,5	56,4	58,1	58,7	60,1	60,7
Borborema Potiguar	53,9	54,2	55,2	55,9	56,2	57,8	58,1
Agreste Potiguar	54,8	55,0	55,3	56,2	56,6	57,1	57,8
Litoral Nordeste	57,4	57,8	58,4	60,4	60,7	61,5	61,8
Macaíba	47,7	48,0	48,6	49,2	49,8	50,4	51,2
Natal	27,9	28,0	28,1	28,3	28,5	29,5	30,1
Litoral Sul	54,7	55,1	55,4	57,3	58,0	58,2	58,6

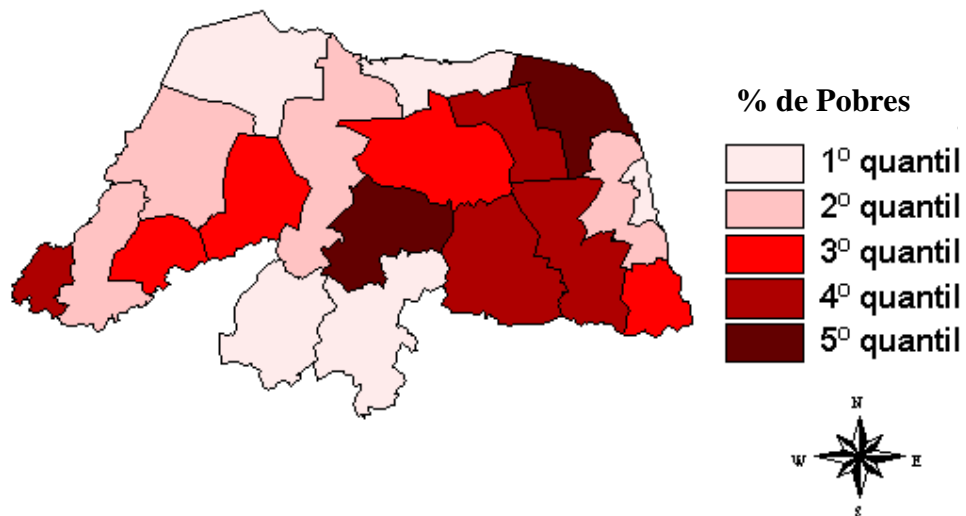
Como pode ser visto nas Tabelas 4.4 e 4.5, as proporções de pobres são maiores, quando mensuradas pelo enfoque do consumo domiciliar per capita. No Estado potiguar, como um todo, a proporção de pobres é de 36,1% da população, pelo enfoque do rendimento domiciliar, e 44,7%, sob o enfoque do consumo domiciliar. É nítido também que existe heterogeneidade entre as microrregiões potiguares sob os dois enfoques. A microrregião da capital Natal (composta pelos municípios de Extremoz, Natal e Parnamirim) apresenta, em ambos os enfoques, a menor proporção de pobres, 22,5% no enfoque do rendimento e 28,3% no enfoque do consumo. No outro extremo, a microrregião do Litoral Nordeste (composta por 7 municípios: Maxaranguape, Pedra Grande, Pureza, Rio do Fogo, São Miguel de Touros, Taipu e Touros) é a que apresenta

a maior proporção de pobres nos dois enfoques (52,3% e 60,4%, respectivamente, para rendimento e consumo).

A seguir, são apresentados os mapas temáticos com as estimativas dessas proporções nos dois enfoques de mensuração da pobreza (rendimento e consumo). Com base nesses mapas, a heterogeneidade da proporção de pobres ao longo das microrregiões potiguares fica ainda mais evidente. Quanto maior a proporção de pobres, mais escura é a cor da área correspondente a microrregião. Esses mapas corroboram o mapa produzido anteriormente no Capítulo 2, com base na variável renda per capita. O Estado é muito pobre, de maneira geral. Mas a microrregião da capital potiguar Natal se destaca nos dois enfoques (rendimento e de consumo) como sendo a microrregião de menor proporção de domicílios pobres.

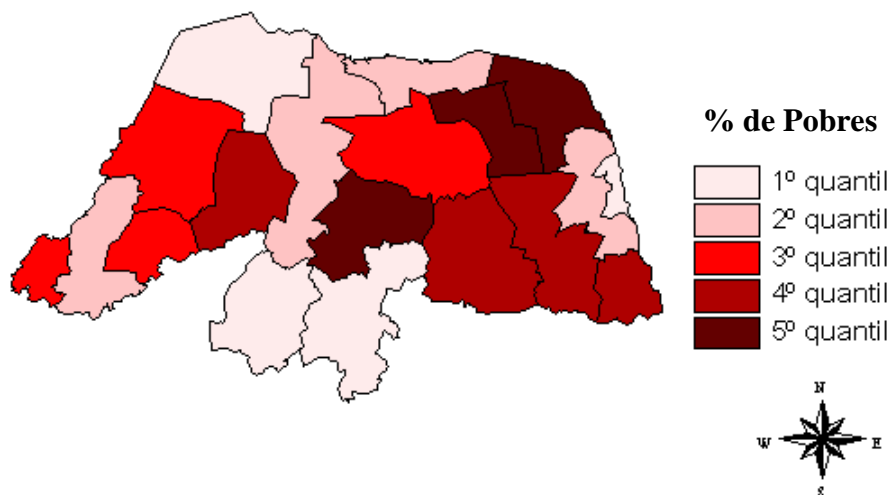
**Figura 4.1: Distribuição espacial da proporção de pobres, com base no rendimento domiciliar per capita, Método do Banco Mundial.**

**Microrregiões do Rio Grande do Norte, 2003.**



**Figura 4.2: Distribuição espacial da proporção de pobres, com base no consumo domiciliar per capita, Método do Banco Mundial.**

**Microrregiões do Rio Grande do Norte, 2003.**



Para fins comparativos, as proporções foram também calculadas a partir dos dados originais das duas pesquisas (Censo 2000 e POF 2002/2003), para cada microrregião e para o Estado do Rio Grande do Norte, como mostra a Tabela 4.6.

**Tabela 4.6: Proporção de domicílios pobres, com base nas variáveis originais POF 2002/2003 e Censo 2000 - Rio Grande do Norte e Microrregiões.**

Estado e Microrregiões	POF 2002/2003		Censo - renda
	Renda	Consumo	
Estado	33,0	38,9	50,6
Natal	17,1	23,3	29,7
Litoral Nordeste	18,5	33,5	73,1
Macau	21,3	28,3	53,1
Seridó Ocidental	24,4	30,7	40,8
Mossoró	24,5	35,5	43,9
Macaíba	29,4	31,4	57,3
Pau dos Ferros	31,0	44,2	64,0
Chapada do Apodi	32,8	64,9	63,6
Vale do Açu	38,6	52,1	59,2
Seridó Oriental	41,6	51,8	46,3
Litoral Sul	43,0	43,0	68,7
Borborema Potiguar	43,1	51,0	68,6
Umarizal	50,6	59,4	65,5
Angicos	56,7	23,7	63,6
Agreste Potiguar	57,5	62,6	69,9
Baixa Verde	61,2	55,9	69,5
Serra de São Miguel	62,4	50,4	72,9
Médio Oeste	71,5	59,9	70,1
Serra de Santana	79,1	79,0	69,5

**Fonte: Microdados Censo 2000 e POF 2002/2003.**

## 4.2 Modelagem sob a abordagem bayesiana

Na abordagem bayesiana, seguindo a idéia por trás do método do Banco Mundial, ajustamos um modelo com base nos dados da POF e, em seguida, imputamos as proporções de domicílios pobres na amostra do Censo. Os códigos na linguagem WinBugs para ajustes dos modelos estatísticos, sob a abordagem bayesiana, são apresentados em anexo.

Idealmente, pretendia-se fazer uma listagem dos domicílios que não foram pesquisados na POF e, em seguida, incluí-los na análise para uma estimação direta da proporção de pobres nas microrregiões potiguares para o ano de 2003. Porém não há

informação suficiente nos microdados (tanto da amostra do Censo, quanto da POF) para obter tais dados. Mesmo se essa informação estivesse disponível, encontraríamos um problema: estamos utilizando a parte amostral do censo. Poderia acontecer que um domicílio não pesquisado na POF também não fosse pesquisado na amostra do censo.

Então, com base nas estimativas dos parâmetros dos modelos ajustados na POF, imputamos valores de probabilidade do domicílio ser classificado como pobre, com base no rendimento e no consumo domiciliar per capita na amostra do censo. Isso foi feito multiplicando-se as estimativas dos parâmetros obtidos com modelos ajustados na POF pelas variáveis da amostra do Censo. Embora, não sejamos capazes de termos informação de todos os domicílios com a amostra do censo, obteremos estimativas mais robustas das proporções de pobres nas microrregiões potiguares.

Consideramos o modelo 4.2 com função de ligação logística para a variável resposta do estudo, para o qual foi utilizado o Amostrador de Gibbs.

$$y_{ij} | \pi_{ij} \approx i.i.d. \text{ Bernoulli}(\pi_{ij}),$$

$$\text{Logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = X_{ij}^T \beta_i \quad (4.2)$$

O modelo acima descrito é conhecido como modelo hierárquico de regressão logística, no qual apenas o intercepto leva em conta o efeito da pequena área  $v_{0i}$ . Este modelo leva em consideração dois níveis de hierarquia (domicílio e microrregião).

Pressupõe-se que a distribuição condicional de  $v_{0i}$  dado a precisão  $\tau_0$  (inverso da variância) tem média 0 e precisão  $\tau_0$ . No modelo, dado pela equação (4.2), assumimos distribuições a priori vagas para os parâmetros de regressão associados aos



efeitos fixos e aleatórios. Vale ressaltar que as prioris vagas para os hiperparâmetros são: os parâmetros de regressão fixos seguem uma distribuição Normal com média 0 e precisão  $10^{-3}$ , o hiperparâmetro  $\tau_0$  tem distribuição Gama com parâmetros 0,5 e  $5^{-4}$ .

Consideramos 20.000 amostras geradas via MCMC (Amostrador de Gibbs), descartamos as primeiras 5.000 amostras e, para obtermos a média à posteriori de  $\theta_i$ , consideramos as 15.000 últimas amostras obtidas da seqüência da cadeia de Markov.

Vale ressaltar que a convergência foi verificada para todos os parâmetros dos modelos a partir do comportamento das trajetórias das cadeias ao longo das iterações. As Figuras A1 e A2, em anexo, apresentam os gráficos de diagnóstico de convergência de Gelman e Rubin para os modelos de rendimento e de consumo domiciliar per capita.

Nas Tabelas 4.7 e 4.8 apresentamos o sumário da distribuição a posteriori dos parâmetros para os modelos que consideram as variáveis respostas rendimento domiciliar per capita e consumo domiciliar per capita, respectivamente.

**Tabela 4.7: Parâmetros de regressão estimados a partir de 15000 amostras geradas via amostrador de Gibbs, Rendimento domiciliar per capita.**

Node	média	Desvio	MC error	2.5%	mediana	97.5%	Iniciar	Finalizar
COMPUT	-2.141	1.319	0.0933	-5.252	-1.943	-0.1474	5.001	20.000
PAVIMEN	-0.5373	0.1522	0.1845	-0.836	-0.5361	-0.242	5.001	20.000
QTBANH1	15.86	9.654	0.6854	2.33	14.11	32.05	5.001	20.000
QTBANH2	14.22	9.662	0.6855	0.6124	12.42	30.5	5.001	20.000
TV	-0.6624	0.2245	0.1638	-1.091	-0.6634	-0.2219	5.001	20.000
IDOSOS1	-6.871	8.992	0.6383	-18.76	-9.489	6.556	5.001	20.000
IDOSOS2	-7.723	9.0	0.6386	-19.65	-10.31	5.674	5.001	20.000
IDOSOS3	-9.163	9.019	0.6388	-21.24	-11.8	4.361	5.001	20.000
IDOSOS4	-7.448	9.187	0.6376	-20.69	-10.03	7.129	5.001	20.000
TIPOESC1	-6.415	3.742	0.2633	-13.55	-6.059	1.793	5.001	20.000
TIPOESC2	-6.336	3.738	0.2633	-13.44	-5.979	1.892	5.001	20.000
TIPOESC3	-5.779	3.788	0.2646	-13.09	-5.443	2.358	5.001	20.000
TIPOESC4	-5.212	3.75	0.2635	-12.37	-4.845	2.984	5.001	20.000
MORA2	0.709	0.4802	0.1741	-0.2079	0.7013	1.677	5.001	20.000
ADOLEES1	-10.2	3.304	0.2322	-16.01	-10.42	-1.509	5.001	20.000
ADOLEES2	-10.09	3.303	0.2322	-15.93	-10.27	-1.415	5.001	20.000
ADOLEES3	-9.572	3.304	0.2319	-15.45	-9.756	-0.9028	5.001	20.000
ADOLEES4	-9.749	3.32	0.2318	-15.55	-9.922	-1.118	5.001	20.000
ADOLEES5	-8.077	3.378	0.2323	-13.99	-8.235	0.5312	5.001	20.000
MORA3	1.472	0.4506	0.1774	0.6245	1.465	2.393	5.001	20.000
TIPODOM2	-24.76	19.12	0.3308	-70.52	-20.91	-0.2963	5.001	20.000
MORA4	2.102	0.4594	0.186	1.228	2.093	3.031	5.001	20.000
MORA5	2.713	0.4596	0.3898	1.836	2.702	3.643	5.001	20.000
MQLAVAR	-2.296	0.4026	0.4976	-3.14	-2.276	-1.55	5.001	20.000
Intercepto	-5.401	14.99	0.8072	-24.27	-7.929	25.56	5.001	20.000

**Tabela 4.8: Parâmetros de regressão estimados a partir de 15000 amostras geradas via amostrador de Gibbs, Consumo domiciliar per capita.**

Node	média	desvio	MC error	2.5%	mediana	97.5%	Iniciar	Finalizar
COMPUT	-27.4	18.59	0.2302	-72.43	-23.59	-3.796	5.001	20.000
MICROO	-2.55	1.322	0.0983	-5.672	-2.36	-0.5315	5.001	20.000
CONDDOM1	-0.9085	7.497	0.6319	-10.72	-1.881	10.72	5.001	20.000
CONDDOM2	-0.2982	7.495	0.6314	-10.12	-1.191	11.36	5.001	20.000
CONDDOM3	-0.1202	7.496	0.5317	-9.964	-1.222	11.51	5.001	20.000
MORA1	-0.917	0.4398	0.6514	-1.815	-0.9053	-0.09135	5.001	20.000
MORA3	0.6378	0.25	0.5620	0.1535	0.6366	1.133	5.001	20.000
MORA4	1.021	0.2584	0.6996	0.5223	1.019	1.54	5.001	20.000
MORA5	1.908	0.2618	0.7665	1.405	1.905	2.428	5.001	20.000
ANOEST1	1.14	0.2124	0.3385	0.731	1.138	1.566	5.001	20.000
ANOEST2	0.7684	0.2092	0.3461	0.3609	0.7672	1.184	5.001	20.000
ANOEST3	0.2517	0.2112	0.3161	-0.1634	0.2505	0.6669	5.001	20.000
PAVIMEN	-0.3399	0.145	0.2211	-0.6252	-0.3401	-0.0559	5.001	20.000
TIPODOM1	-3.45	5.089	0.3603	-12.45	-3.931	6.523	5.001	20.000
TIPODOM2	-27.72	18.7	0.2737	-72.33	-24.07	-1.005	5.001	20.000
TIPODOM3	-1.971	5.146	0.3602	-11.17	-2.39	8.141	5.001	20.000
ADOLE1	-0.7867	0.2099	0.5769	-1.205	-0.7846	-0.3775	5.001	20.000
ADOLE2	-0.4209	0.2027	0.4522	-0.8209	-0.4193	-0.02631	5.001	20.000
MQLAVAR	-1.338	0.2689	0.2244	-1.881	-1.33	-0.8312	5.001	20.000
TV	-0.4788	0.2124	0.6015	-0.8911	-0.4793	-0.05624	5.001	20.000
AR	-2.184	1.355	0.1021	-5.395	-1.992	-0.03553	5.001	20.000
Intercepto	-10.68	4.286	0.229	-19.64	-10.96	-2.306	5.001	20.000

Para ambas as variáveis resposta, obtivemos estimativas bastante coerentes com a realidade. Pode ser observado que, nos dois casos, o fato do domicílio possuir bens de consumo, tais como: computador (COMPUT), aparelho de tv (TV) ou máquina de lavar (MQLAVAR), reduz a probabilidade de o domicílio ser pobre. Outro fato é que residir em apartamento (TIPODOM2) reduz também a probabilidade de o domicílio ser pobre.

A partir dos parâmetros estimados na POF, foi possível prever na amostra do censo os valores para as variáveis respostas. As Tabelas 4.9 e 4.10 apresentam as estimativas da proporção de domicílios pobres (média) estimado por microrregião, baseados no rendimento e consumo domiciliar per capita imputados, respectivamente.

**Tabela 4.9: Distribuição das estimativas de proporção de pobres, baseadas no rendimento domiciliar per capita imputado na amostra do censo (Mod. Hierárquico) Microrregiões do Rio Grande do Norte, 2003.**

<b>Microrregião</b>	<b>Média</b>	<b>Desvio</b>	<b>MC error</b>	<b>2.50%</b>	<b>mediana</b>	<b>97.50%</b>	<b>Iniciar</b>	<b>Finalizar</b>
Mossoró	0.3202	0.0265	0.2040	0.2997	0.3198	0.3332	5001	20000
Chapada do Apodi	0.4269	0.0614	0.3186	0.398	0.4269	0.4475	5001	20000
Médio Oeste	0.4955	0.0844	0.6167	0.4702	0.4959	0.5195	5001	20000
Vale do Açu	0.4066	0.0614	0.8146	0.3924	0.4052	0.4295	5001	20000
Serra de São Miguel	0.4261	0.0999	0.1210	0.3967	0.4285	0.437	5001	20000
Pau dos Ferros	0.4516	0.0629	0.7493	0.4308	0.4512	0.4765	5001	20000
Umarizal	0.4587	0.0927	0.1095	0.4352	0.4574	0.4637	5001	20000
Macau	0.422	0.0852	0.1030	0.4166	0.4191	0.4557	5001	20000
Angicos	0.4718	0.0941	0.1148	0.4638	0.4711	0.4999	5001	20000
Serra de Santana	0.4656	0.0953	0.9289	0.4379	0.4661	0.4755	5001	20000
Seridó Ocidental	0.4208	0.0564	0.5053	0.4137	0.4196	0.4398	5001	20000
Seridó Oriental	0.4839	0.0439	0.3950	0.4573	0.4843	0.5098	5001	20000
Baixa Verde	0.6016	0.0680	0.5873	0.5881	0.6018	0.6244	5001	20000
Borborema Potiguar	0.4977	0.0628	0.5282	0.4768	0.4962	0.5243	5001	20000
Agreste Potiguar	0.515	0.0415	0.3187	0.4941	0.5148	0.5278	5001	20000
Litoral Nordeste	0.5297	0.0696	0.5881	0.5049	0.5294	0.5376	5001	20000
Macaíba	0.4429	0.0358	0.2769	0.4238	0.443	0.4736	5001	20000
Natal	0.2917	0.0136	0.1067	0.2654	0.2915	0.3089	5001	20000
Litoral Sul	0.5084	0.0441	0.3545	0.4744	0.508	0.548	5001	20000

**Tabela 4.10: Distribuição das estimativas de proporção de pobres, baseadas no consumo domiciliar per capita imputado na amostra do censo (Mod. Hierárquico) Microrregiões do Rio Grande do Norte, 2003.**

<b>Microrregião</b>	<b>Média</b>	<b>Desvio</b>	<b>MC error</b>	<b>2.50%</b>	<b>mediana</b>	<b>97.50%</b>	<b>Iniciar</b>	<b>Finalizar</b>
Mossoró	0.4593	0.0356	0.3155	0.4292	0.4593	0.4787	5001	20000
Chapada do Apodi	0.5914	0.0591	0.5948	0.5786	0.5912	0.6276	5001	20000
Médio Oeste	0.5483	0.0759	0.8284	0.5206	0.5442	0.5784	5001	20000
Vale do Açu	0.5421	0.0494	0.3625	0.5063	0.5418	0.5692	5001	20000
Serra de São Miguel	0.564	0.0674	0.5046	0.5450	0.5630	0.6088	5001	20000
Pau dos Ferros	0.5758	0.0437	0.3841	0.5417	0.5751	0.6028	5001	20000
Umarizal	0.5612	0.0605	0.4451	0.5453	0.5606	0.5826	5001	20000
Macau	0.5747	0.0631	0.4669	0.5591	0.5751	0.5979	5001	20000
Angicos	0.5258	0.0619	0.5809	0.5037	0.5264	0.5463	5001	20000
Serra de Santana	0.677	0.0590	0.5833	0.6426	0.6768	0.6934	5001	20000
Seridó Ocidental	0.4598	0.0566	0.4217	0.4487	0.4603	0.4807	5001	20000
Seridó Oriental	0.4527	0.0422	0.3707	0.4104	0.4526	0.4661	5001	20000
Baixa Verde	0.6264	0.0561	0.4532	0.6043	0.6235	0.6420	5001	20000
Borborema Potiguar	0.652	0.0541	0.4582	0.6321	0.6595	0.6750	5001	20000
Agreste Potiguar	0.5944	0.0250	0.2017	0.5775	0.5935	0.6257	5001	20000
Litoral Nordeste	0.6538	0.0467	0.4172	0.6360	0.6526	0.6777	5001	20000
Macaíba	0.5895	0.0317	0.2610	0.5482	0.5895	0.6021	5001	20000
Natal	0.3099	0.0243	0.2661	0.2821	0.3099	0.3130	5001	20000
Litoral Sul	0.6185	0.0465	0.3719	0.5960	0.6187	0.6586	5001	20000

Assim como observado na abordagem do Banco Mundial, a microrregião de Natal (composta pelos municípios de Extremoz, Natal e Parnamirim) apresenta as menores proporções de pobres, tanto no enfoque do rendimento quanto do consumo domiciliar per capita.

As proporções, como um todo, também apresentam valores superiores aos obtidos com a metodologia do Banco Mundial.

A verificação do ajuste e discriminação de modelos é feita a partir do DIC (Critério de informação baseado no Deviance). O modelo de melhor ajuste é aquele que

apresenta o menor valor para o DIC. De acordo com SOUZA (2004), a quantidade DIC pode ser considerada como medida bayesiana de ajuste ou adequação, recebendo uma punição representada pelo termo de complexidade do modelo sob estudo. Maiores detalhes sobre o cálculo do DIC podem ser encontrados em SOUZA (2004) e THOMAS et al. (1992). Para os modelos de rendimento domiciliar per capita (RDPC) e consumo domiciliar per capita (CDPC), encontramos valores da estatística DIC, conforme mostra Tabela 4.11.

**Tabela 4.11: DIC dos modelos hierárquicos**

Variável Resposta	DIC
Logaritmo do rendimento domiciliar per capita	1250,4
Logaritmo do consumo domiciliar per capita	1392,7

Até aqui, os modelos bayesianos apresentados consideravam a hierarquia dos dados, como no caso, de acordo com a microrregião de residência. Agora se faz a consideração de que um par de microrregiões vizinhas (sendo que vizinhas são consideradas, aqui, como microrregiões adjacentes) possui proporções de pobres mais semelhantes quando comparadas com duas microrregiões escolhidas arbitrariamente. Para isso um modelo CAR é utilizado por apresentar a propriedade de que sua especificação fornece diretamente as distribuições condicionais completas dos parâmetros do modelo, fator determinante para o uso do amostrador de Gibbs em Monte Carlo via Cadeia de Markov. No contexto bayesiano, geralmente o modelo CAR é usado como informação a priori de um parâmetro do modelo para o processo de interesse (SCHIMIDT et al, 2003).

Da mesma forma, consideramos 20.000 amostras geradas via MCMC (Amostrador de Gibbs), descartamos as primeiras 5.000 amostras e, para obtermos a

média à posteriori de  $\theta_i$ , consideramos as 15.000 últimas amostras obtidas da seqüência da cadeia de Markov.

A convergência também foi verificada para todos os parâmetros dos modelos a partir do comportamento das trajetórias das cadeias ao longo das iterações, por gráficos de diagnóstico de convergência de Gelman e Rubin para os modelos de rendimento e de consumo domiciliar per capita.

A seguir, apresentamos nas Tabelas 4.12 e 4.13 os sumários da distribuição a posteriori dos parâmetros para os modelos que consideram as variáveis respostas rendimento domiciliar per capita e consumo domiciliar per capita, respectivamente.

Assim como no caso anterior, temos estimativas de parâmetros bastante coerentes. De maneira geral, os mesmos comentários feitos com base nos modelos hierárquicos podem ser reproduzidos aqui. Temos que a posse de bens de consumo se relaciona de maneira negativa com a probabilidade do domicílio ser classificado como pobre. Destaca-se, desta vez, o fato da presença de idosos no domicílio (IDOSOS1, IDOSOS2, IDOSOS3 e IDOSOS4) ser significativo e com coeficientes negativos sob o enfoque dos rendimentos, o que significa que a presença dessas pessoas no domicílio ajuda a diminuir a pobreza, possivelmente, contribuindo com as despesas do lar.

**Tabela 4.12: Parâmetros de regressão estimados a partir de 15000 amostras, Rendimento domiciliar per capita. (modelo CAR).**

node	Media	Desvio	MC error	2.5%	mediana	97.5%	Iniciar	Finalizar
COMPUT	-2.249	1.33	0.08876	-5.439	-2.049	-0.2234	5.001	20.000
PAVIMEN	-0.5389	0.1508	0.01936	-0.832	-0.5401	-0.2427	5.001	20.000
QTBANH1	8.975	4.487	0.3173	2.777	8.107	19.72	5.001	20.000
QTBANH2	7.284	4.504	0.3175	0.9798	6.408	18.09	5.001	20.000
TV	-0.6601	0.2263	0.5997	-1.107	-0.6597	-0.2134	5.001	20.000
IDOSOS1	-9.507	4.955	0.3508	-21.93	-8.413	-3.003	5.001	20.000
IDOSOS2	-10.35	4.962	0.3509	-22.74	-9.25	-3.801	5.001	20.000
IDOSOS3	-11.8	4.982	0.3503	-24.22	-10.63	-5.148	5.001	20.000
IDOSOS4	-10.02	5.294	0.3484	-22.79	-8.958	-2.259	5.001	20.000
TIPOESC1	-3.143	6.162	0.4366	-14.51	-3.228	5.465	5.001	20.000
TIPOESC2	-3.056	6.159	0.4366	-14.39	-3.15	5.551	5.001	20.000
TIPOESC3	-2.555	6.181	0.4362	-14.1	-2.497	6.13	5.001	20.000
TIPOESC4	-1.934	6.167	0.4368	-13.36	-2.004	6.66	5.001	20.000
MORA2	0.7397	0.4919	0.2024	-0.1976	0.7293	1.728	5.001	20.000
ADOLEES1	14.12	7.37	0.5226	3.714	11.77	33.46	5.001	20.000
ADOLEES2	14.23	7.37	0.5226	3.823	11.87	33.54	5.001	20.000
ADOLEES3	14.75	7.369	0.5224	4.346	12.38	34.08	5.001	20.000
ADOLEES4	14.58	7.375	0.5224	4.19	12.23	33.94	5.001	20.000
ADOLEES5	16.26	7.4	0.5224	5.828	13.95	35.57	5.001	20.000
MORA3	1.517	0.4705	0.2094	0.6293	1.505	2.48	5.001	20.000
TIPODOM2	-25.0	19.16	0.1356	-70.85	-21.19	-0.451	5.001	20.000
MORA4	2.144	0.4768	0.02136	1.257	2.131	3.109	5.001	20.000
MORA5	2.75	0.4826	0.02175	1.856	2.734	3.734	5.001	20.000
MQLAVAR	-2.316	0.4011	0.0325	-3.156	-2.299	-1.576	5.001	20.000
Intercepto	-12.63	4.732	0.3349	-20.64	-12.84	-3.433	5.001	20.000



**Tabela 4.13: Parâmetros de regressão estimados a partir de 15000 amostras, Consumo domiciliar per capita (modelo CAR).**

<b>Node</b>	<b>média</b>	<b>desvio</b>	<b>MC error</b>	<b>2.5%</b>	<b>mediana</b>	<b>97.5%</b>	<b>Iniciar</b>	<b>Finalizar</b>
COMPUT	-27.27	18.5	0.1342	-43.02	-23.38	-3.887	5.001	20.000
MICROO	-2.578	1.33	0.0945	-5.718	-2.379	-0.5633	5.001	20.000
CONDDOM1	-7.349	3.057	0.2155	-12.01	-8.038	-1.162	5.001	20.000
CONDDOM2	-6.733	3.06	0.2154	-11.42	-7.401	-0.5336	5.001	20.000
CONDDOM3	-6.539	3.058	0.2153	-11.23	-7.214	-0.3328	5.001	20.000
MORA1	-0.8939	0.4301	0.0636	-1.777	-0.876	-0.09229	5.001	20.000
MORA3	0.6409	0.2538	0.0557	0.1447	0.64	1.139	5.001	20.000
MORA4	1.032	0.2567	0.0614	0.5411	1.029	1.543	5.001	20.000
MORA5	1.951	0.2595	0.0659	1.455	1.948	2.467	5.001	20.000
ANOEST1	1.121	0.2139	0.0424	0.7064	1.12	1.541	5.001	20.000
ANOEST2	0.7662	0.2093	0.0398	0.356	0.7647	1.173	5.001	20.000
ANOEST3	0.257	0.2122	0.0378	-0.1524	0.2547	0.6766	5.001	20.000
PAVIMEN	-0.3258	0.1424	0.1985	-0.6054	-0.3251	-0.05002	5.001	20.000
TIPODOM1	-1.093	3.77	0.2665	-8.206	-0.5938	9.496	5.001	20.000
TIPODOM2	-26.0	18.81	0.2025	-70.53	-22.22	-0.5938	5.001	20.000
TIPODOM3	0.4143	3.85	0.2666	-6.953	0.8079	10.88	5.001	20.000
ADOLE1	-0.7536	0.2066	0.0439	-1.158	-0.7544	-0.3474	5.001	20.000
ADOLE2	-0.409	0.2039	0.0340	-0.8099	-0.4079	-0.01181	5.001	20.000
MQLAVAR	-1.355	0.272	0.2238	-1.894	-1.354	-0.8376	5.001	20.000
TV	-0.4833	0.2141	0.5018	-0.8948	-0.4896	-0.0552	5.001	20.000
AR	-2.237	1.357	0.1007	-5.41	-2.049	-0.1035	5.001	20.000
Intercepto	7.367	4.239	0.2999	-1.91	7.754	14.62	5.001	20.000

Dando seqüência, apresentamos a seguir nas Tabelas 4.14 e 4.15 as estimativas da proporção de domicílios pobres (média) estimadas por microrregião, baseadas no rendimento e consumo domiciliar per capita imputados na amostra do censo, respectivamente.

**Tabela 4.14: Distribuição das estimativas de proporção de pobres, baseadas no rendimento domiciliar per capita imputado na amostra do censo (modelo CAR) Microrregiões do Rio Grande do Norte, 2003.**

<b>Microrregião</b>	<b>Média</b>	<b>Desvio</b>	<b>MC error</b>	<b>2,50%</b>	<b>mediana</b>	<b>97,50%</b>	<b>Iniciar</b>	<b>Finalizar</b>
Mossoró	0.4327	0.0194	0,2672	0.3926	0.4334	0.4694	5001	20000
Chapada do Apodi	0.5672	0.0296	0,4006	0.5459	0.5681	0.5818	5001	20000
Médio Oeste	0.6067	0.0314	0,4486	0.5423	0.6077	0.6157	5001	20000
Vale do Açu	0.6256	0.0257	0,3718	0.5772	0.6265	0.6439	5001	20000
Serra de São Miguel	0.6929	0.0293	0,3828	0.6522	0.6941	0.7168	5001	20000
Pau dos Ferros	0.5565	0.0273	0,3614	0.5304	0.5579	0.5871	5001	20000
Umarizal	0.5244	0.0258	0,3636	0.4915	0.5256	0.5426	5001	20000
Macau	0.3845	0.0245	0,3364	0.3734	0.3852	0.4009	5001	20000
Angicos	0.5427	0.0258	0,3781	0.5136	0.5481	0.5552	5001	20000
Serra de Santana	0.5975	0.0280	0,3978	0.5829	0.5987	0.6396	5001	20000
Seridó Ocidental	0.3695	0.0234	0,3461	0.3514	0.3702	0.3914	5001	20000
Seridó Oriental	0.3644	0.0233	0,3307	0.3471	0.3649	0.3988	5001	20000
Baixa Verde	0.5637	0.0283	0,4074	0.5258	0.5646	0.5967	5001	20000
Borborema Potiguar	0.5511	0.0255	0,3841	0.5282	0.5523	0.5984	5001	20000
Agreste Potiguar	0.5472	0.0256	0,3822	0.5134	0.5485	0.5745	5001	20000
Litoral Nordeste	0.5817	0.0295	0,4395	0.5506	0.583	0.6068	5001	20000
Macaíba	0.5073	0.0241	0,3423	0.4872	0.5086	0.5325	5001	20000
Natal	0.2843	0.0159	0,2350	0.2514	0.2921	0.3143	5001	20000
Litoral Sul	0.5509	0.0262	0,3622	0.5161	0.5521	0.5792	5001	20000

**Tabela 4.15: Distribuição das estimativas de proporção de pobres, baseadas no consumo domiciliar per capita imputado na amostra do censo (modelo CAR) Microrregiões do Rio Grande do Norte, 2003.**

<b>Microrregião</b>	<b>Média</b>	<b>Desvio</b>	<b>MC error</b>	<b>2,50%</b>	<b>mediana</b>	<b>97,50%</b>	<b>Iniciar</b>	<b>Finalizar</b>
Mossoró	0.426	0.01601	0,1292	0.395	0.426	0.4571	5001	20000
Chapada do Apodi	0.5593	0.01862	0,1554	0.5231	0.5592	0.5958	5001	20000
Médio Oeste	0.6672	0.1751	0,1365	0.6332	0.667	0.7018	5001	20000
Vale do Açu	0.5554	0.01731	0,1439	0.5418	0.5552	0.5894	5001	20000
Serra de São Miguel	0.6737	0.01651	0,1293	0.6414	0.6738	0.7061	5001	20000
Pau dos Ferros	0.5512	0.01511	0,1334	0.5219	0.551	0.5814	5001	20000
Umarizal	0.5408	0.01757	0,1348	0.5165	0.5406	0.5756	5001	20000
Macau	0.5447	0.01933	0,1614	0.5268	0.5444	0.5821	5001	20000
Angicos	0.6325	0.01815	0,1509	0.5975	0.6322	0.6682	5001	20000
Serra de Santana	0.6705	0.01839	0,1736	0.644	0.6705	0.7066	5001	20000
Seridó Ocidental	0.5052	0.01677	0,1215	0.4727	0.505	0.5383	5001	20000
Seridó Oriental	0.5148	0.01598	0,1226	0.4937	0.5148	0.546	5001	20000
Baixa Verde	0.6087	0.01748	0,1472	0.5744	0.6084	0.6438	5001	20000
Borborema Potiguar	0.6242	0.01613	0,1447	0.593	0.624	0.6566	5001	20000
Agreste Potiguar	0.5895	0.0111	0,1111	0.5684	0.5894	0.6121	5001	20000
Litoral Nordeste	0.6964	0.01644	0,1406	0.6749	0.6964	0.7291	5001	20000
Macaíba	0.4894	0.01589	0,1317	0.4587	0.4893	0.5209	5001	20000
Natal	0.4518	0.01602	0,1445	0.4404	0.4516	0.4735	5001	20000
Litoral Sul	0.5797	0.01784	0,1559	0.5547	0.5797	0.615	5001	20000

De maneira geral, os valores são um pouco inferiores aos obtidos com o ajuste de modelos hierárquicos, com algumas exceções.

Os valores da estatística DIC para os modelos de rendimento domiciliar per capita (RDPC) e consumo domiciliar per capita (CDPC) são mostrados na Tabela 4.16.

**Tabela 4.16: DIC dos modelos CAR**

<b>Variável Resposta</b>	<b>DIC</b>
RDPC	1254,1
CDPC	1393,6

Para a seleção do modelo de melhor ajuste, utilizamos uma medida fornecida pelo WinBUGS para a verificação do ajuste e discriminação entre modelos competitivos que é o DIC (Critério de Informação “*Deviance*”), cujo cálculo é baseado nos valores de máxima verossimilhança dos diferentes modelos, ponderando pela diferença no número de parâmetros estimados (THOMAS et al, 1992). O modelo de melhor ajuste é aquele que apresenta o menor valor para o DIC.

**Tabela 4.17: DIC dos modelos sob o enfoque bayesiano**

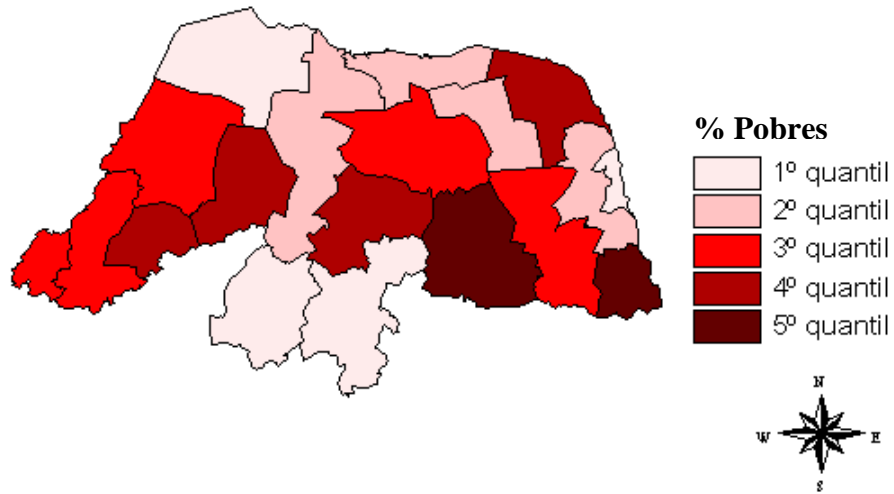
<b>Variável Resposta</b>	<b>Modelos Hierárquicos (DIC)</b>	<b>Modelos CAR (DIC)</b>
RDPC	1250,4	1254,1
CDPC	1392,7	1393,6

Uma simples observação dos resultados da Tabela 4.17 mostra que os modelos hierárquicos fornecem um melhor ajuste quando comparados com os modelos CAR, embora essa melhora seja pequena.

A seguir são feitos mapas temáticos dos resultados obtidos com os modelos hierárquicos. Quanto mais escura a área correspondente no gráfico, mais pobre é a microrregião.

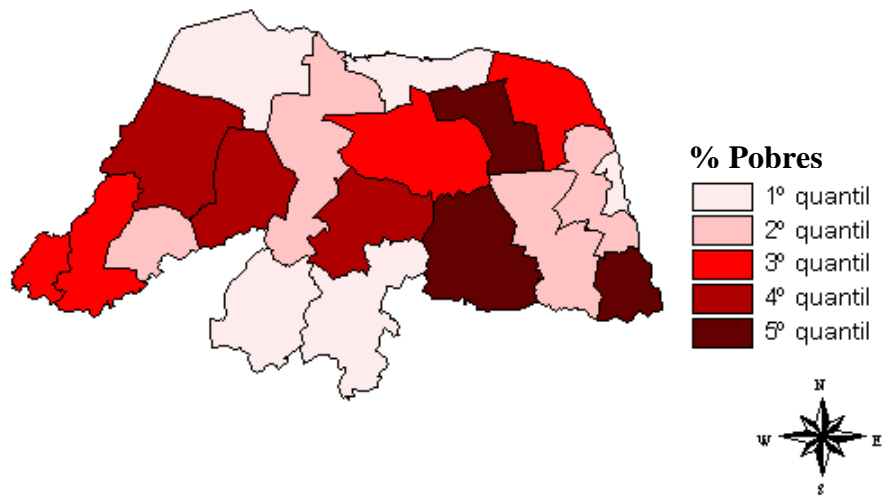
**Figura 4.3: Distribuição espacial da proporção de pobres, com base no rendimento domiciliar per capita, Modelos Hierárquicos.**

**Microrregiões do Rio Grande do Norte, 2003.**



**Figura 4.4: Distribuição espacial da proporção de pobres, com base no consumo domiciliar per capita, Modelos Hierárquicos.**

**Microrregiões do Rio Grande do Norte, 2003.**



Numa tentativa de comparação entre as duas metodologias utilizadas (Banco Mundial e a dos Modelos Hierárquicos), colocamos a proporção de acerto entre os valores observados e ajustados na POF em cada microrregião potiguar na Tabela 4.18. Nesta dissertação não adicionamos cálculos de intervalos de confiança dos modelos analisados.

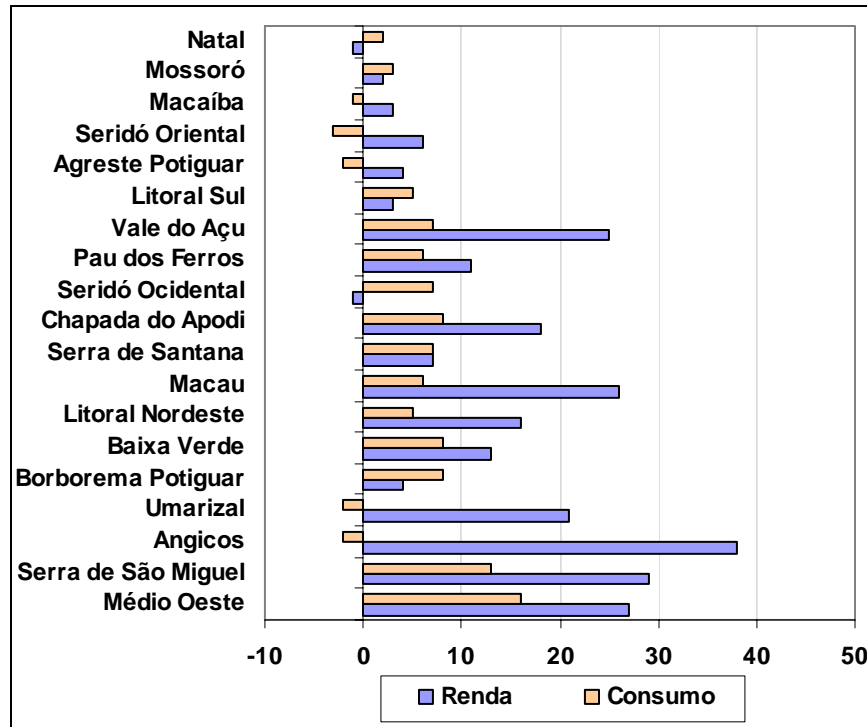
**Tabela 4.18: Proporção de acerto nos modelos ajustados na POF.**

Microrregião	Total de domicílios	Rendimento			Consumo		
		Banco Mundial	Modelo Hierárquico	Modelo CAR	Banco Mundial	Modelo Hierárquico	Modelo CAR
Mossoró	165	0,71	0,73	0,73	0,65	0,68	0,68
Chapada do Apodi	41	0,58	0,76	0,78	0,57	0,65	0,64
Médio Oeste	12	0,42	0,69	0,72	0,48	0,64	0,64
Vale do Açu	62	0,65	0,90	0,86	0,61	0,68	0,67
Serra de São Miguel	12	0,59	0,88	0,93	0,53	0,66	0,64
Pau dos Ferros	60	0,77	0,88	0,90	0,60	0,66	0,66
Umarizal	28	0,74	0,95	0,96	0,66	0,64	0,63
Macau	32	0,69	0,95	0,94	0,63	0,69	0,70
Angicos	24	0,57	0,95	0,93	0,68	0,66	0,64
Serra de Santana	38	0,88	0,95	0,93	0,59	0,66	0,65
Seridó Ocidental	46	0,68	0,67	0,68	0,69	0,76	0,74
Seridó Oriental	84	0,60	0,66	0,66	0,69	0,66	0,66
Baixa Verde	29	0,54	0,67	0,66	0,62	0,70	0,70
Borborema Potiguar	28	0,63	0,67	0,66	0,53	0,61	0,60
Agreste Potiguar	82	0,61	0,65	0,66	0,65	0,63	0,64
Litoral Nordeste	31	0,56	0,72	0,71	0,60	0,65	0,65
Macaíba	144	0,70	0,73	0,71	0,70	0,69	0,68
Natal	560	0,75	0,74	0,73	0,78	0,80	0,81
Litoral Sul	70	0,67	0,70	0,69	0,61	0,66	0,67

Com base na Tabela 4.18 é possível perceber que na abordagem bayesiana, no caso com modelos hierárquicos, as taxas de acerto encontradas são superiores às encontradas pelo modelo sugerido pelo Banco Mundial. Em particular, chama atenção o fato de que a superioridade do ajuste pelos modelos hierárquicos fica mais evidente quando há poucos domicílios entrevistados na microrregião. Essa mesma impressão fica ainda mais evidente quando se coloca num gráfico as diferenças entre as taxas de acerto

dos modelos hierárquicos e do modelo proposto pelo Banco Mundial. Maiores detalhes no Gráfico 4.2.

**Gráfico 4.2: Distribuição das diferenças entre as taxas de acerto nos modelos hierárquicos e modelo proposto pelo Banco Mundial, para o logaritmo da renda domiciliar per capita e logaritmo do consumo domiciliar per capita, respectivamente.**



As barras do Gráfico 4.2 foram construídas com base nas diferenças de proporção de acerto dos modelos hierárquicos e dos modelos de regressão do Banco Mundial. Uma barra com valor positivo, por exemplo, indica superioridade dos modelos hierárquicos. Os valores estão ordenados, segundo o total de domicílios pesquisados na POF. A microrregião Médio Oeste possui a menor quantidade de domicílios pesquisada e, no outro extremo, Natal é a microrregião com maior quantidade de domicílios pesquisada. Nos dois casos, renda e consumo, o modelo hierárquico foi superior em quase todas as

microrregiões, com destaque para as microrregiões em que o total de domicílios pesquisados é muito reduzido.



# **CAPÍTULO 5**

## **CONCLUSÕES E RECOMENDAÇÕES FINAIS**

Neste capítulo, apresentamos as principais conclusões deste trabalho de pesquisa, suas contribuições e algumas propostas de trabalhos futuros.

### **5.1 Conclusões**

Este trabalho apresentou um estudo comparativo dos resultados da aplicação de dois métodos estatísticos para o problema de estimação da proporção de pobres em pequenas áreas para o estudo da pobreza: um sob a abordagem da estatística clássica e o outro sob a abordagem bayesiana. Devido ao enorme esforço computacional e acesso aos dados necessários ao ajuste, apenas uma Unidade da Federação foi escolhida para o estudo. Assim como feito em CUNHA (2006), a área de estudo se restringiu ao estado do Rio Grande do Norte e as pequenas áreas ou pequenos domínios analisados foram as 19 microrregiões potiguares.

Dada a complexidade e multidimensionalidade do fenômeno pobreza, uma questão fundamental encontrada, logo de início, foi definir um indicador apropriado para captar o bem-estar das famílias. Com base na literatura pesquisada, foi possível constatar que ainda existe grande controvérsia sobre quais indicadores, baseados na renda ou consumo, são mais adequados para mensurar a pobreza. Numa tentativa de não fugir do que é comumente utilizado, adotamos os indicadores baseados no rendimento domiciliar per capita e no consumo domiciliar per capita. Além disso, outra questão de difícil resposta está em como classificar um indivíduo ou uma família como pobre, pois identificar os pobres implica em realizar comparações entre distintos níveis de bem-estar.

A resposta para esta questão depende do conceito de pobreza adotado, bem como da linha de pobreza considerada.

Nesta dissertação, os indivíduos foram classificados como pobres com base no critério do Programa Bolsa-Família, o qual é um programa social de auxílio financeiro que, em 2003 (época de sua criação), visava atender aquelas famílias com renda de até R\$100,00 per capita mensais.

Na abordagem da estatística clássica, o método proposto pelos pesquisadores do Banco Mundial foi aplicado. Este método tem sido aplicado em diversos países em desenvolvimento, no intuito de estimar medidas de pobreza em pequenas áreas. Uma característica observada nas aplicações dessa metodologia é que o coeficiente de determinação ( $R^2$ ) obtido é moderado, raramente superando o patamar de 70% de explicação. Como exemplo, foram citados os resultados obtidos em: Sri Lanka (entre 28% e 63%), Papua Nova Guiné (34%), Madagascar (entre 24% e 64%) e Equador (entre 46% e 74%). Recentemente, CUNHA (2006) aplicou tal método (sugerindo algumas melhorias sob a abordagem da estatística clássica) para estimação da proporção de pobres nas microrregiões do Estado do Rio Grande do Norte; entretanto, seus resultados também foram moderados (61% e 64%). Neste estudo, o trabalho de CUNHA (2006) foi replicado. Porém, seguindo rigorosamente o procedimento original, obtendo-se coeficientes de determinação um pouco menores (59% e 61%). No sentido de aumentar essa capacidade de explicação, métodos sob o enfoque bayesianos foram considerados.

Sob o enfoque bayesiano, foram aplicados modelos hierárquicos de regressão logística (com intercepto aleatório) e modelos condicionais autoregressivos (modelos CAR) na tentativa de incorporar efeitos espaciais. Vale ressaltar que os métodos bayesianos foram aplicados no mesmo contexto dos métodos propostos pelos pesquisadores do Banco Mundial, ou seja, ajustaram-se modelos na POF e, em seguida, valores das variáveis respostas foram imputados na amostra do Censo.

Os resultados da aplicação, para o caso do estado do Rio Grande do Norte, da teoria e metodologia utilizadas no presente trabalho, visando dar suporte a necessidade dos governantes de ter subsídios técnicos para formulação de políticas públicas no combate a pobreza demonstram que é possível melhorar a qualidade das estimativas dos modelos estatísticos propostos pelos pesquisadores do Banco Mundial. Um fato imediato que pôde ser verificado é que as estimativas da proporção de pobres encontradas pela abordagem bayesiana foram mais elevadas; tanto utilizando modelos hierárquicos quanto utilizando modelos CAR, se considerado que a matriz de vizinhança utilizada foi a de adjacência. Além disso, verificou-se, através do DIC, que os modelos hierárquicos obtiveram melhor desempenho em relação aos modelos CAR.

Corroborando o que foi encontrado nesse estudo, BATTESE *et. al.* (1988) mostraram empiricamente que o modelo com intercepto aleatório produz estimadores significativamente mais precisos que os estimadores de regressão. Podemos concluir nesse caso específico que, para variáveis binárias, as estimativas em pequenas áreas são melhoradas usando modelos hierárquicos. Vale ressaltar que essa melhoria foi constatada apenas com cálculos de proporções de acerto. Com base nesse critério, os modelos hierárquicos e/ou modelos CAR foram mais adequados ao problema de inferência no estudo de estimação de medidas de pobreza em pequenas áreas, pois forneceram um melhor ajuste aos dados, principalmente quando se tem pouquíssimas informações amostrais.

## **5.2 Recomendações para trabalhos futuros**

Este trabalho mostrou um possível caminho para melhorar a metodologia proposta pelo Banco Mundial. Outros aperfeiçoamentos e melhorias podem ser feitos para avançar em trabalhos futuros. Dentre estes podemos citar:

Primeiramente, adotar uma unidade geográfica menor para as pequenas áreas para ver se o modelo se ajusta a contento; como por exemplo, utilizar municípios. Este interesse pode ser atribuído, principalmente, pela legislação vigente que define a distribuição de verbas federais (Fundo de Participação dos Municípios, saúde e educação) a partir de critérios populacionais e por outro lado pela necessidade das autoridades locais de se obter um quadro mais preciso e atualizado da sua área de jurisdição. Além disso, os governos estaduais e federal também necessitam de informações em nível geográfico mais desagregado no intuito de identificar subregiões menos desenvolvidas, auxiliando na elaboração de planos de desenvolvimento regionais.

Estender os métodos aplicados para mais de dois níveis hierárquicos; como domicílio, setor censitário e município. Acredita-se que quanto mais informação for adicionada ao modelo, melhor será o ajuste.

Considerar a hipótese de plano amostral informativo sob a abordagem bayesiana também é um ponto interessante a se considerar em trabalhos futuros. Pois, quando ignoramos o mecanismo através do qual os dados foram obtidos, as unidades amostrais são analisadas como se fossem independentes e identicamente distribuídas, o que nem sempre é verdadeiro, pois o plano amostral pode ser do tipo informativo, ou seja, ser um desenho onde as probabilidades de seleção dos elementos da população estão correlacionadas com as variáveis respostas.

Propor a utilização de indicadores multidimensionais ou indicadores de intensidade de pobreza. Também seria interessante estender a análise para desigualdade de pobreza e não somente proporção de pobres, como feito no presente estudo. Foi verificado nesse estudo que as estruturas de gastos são, geralmente, inferiores às rendas, por isso seria importante utilizar diferentes linhas de pobreza para rendimento e consumo, sendo uma linha de pobreza um pouco menor para a despesa.

Finalmente, buscar outros métodos para avaliação dos métodos, além da comparação de porcentagem de acertos para os valores preditos para a proporção de pobres.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ABRAMOVITZ, L. Relatório da revisão bibliográfica referente à seleção de metodologia estatística para análise da pobreza. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2000.
- [2] ANGOLA. Estratégia de combate à pobreza: reinserção social, reabilitação e reconstrução e estabilização econômica. Ministério do Planeamento – Edição Revista, United Nations Development Programme, 2005.
- [3] ANSELIN, L. Under the hood: issues in the specification and interpretation of spatial regression models (forthcoming: Agriculture Economics), obtido em <http://agec221.agecon.uiuc.edu/users/anselin/papers/hood.pdf> Acesso em 05/12/2007.
- [4] ATLAS DO DESENVOLVIMENTO HUMANO NO BRASIL (*software*). PNUD, FJP, IPEA, 2003.
- [5] ASSUNÇÃO, R.; POTTER, J.; CAVENAGHI, S. Estimação de Curvas de Fecundidade para Pequenas Áreas com Modelos Bayesianos Espaciais.. In: XI Encontro Nacional de Estudos Populacionais (ABEP), Caxambu, MG. Associação Brasileira de Estudos Populacionais - ABEP, p. 2851-2862, 1998.
- [6] BAILEY, T. C. & GATRELL, A. C. Interactive Spatial Data Analysis. Essex: Longman, 1995.
- [7] BATTESE, G.E., HARTER, R.M. e FULLER, W.A. An Error Components Model for Prediction of County Crop Areas Using Survey and Satellite Data, Journal of American Statistical Association, 83, 28-36, 1988.
- [8] CÂMARA, G., CARVALHO, M.S., CRUZ, O.G., CORREA, V. *Análise Espacial de Áreas*, Em: *Análise Espacial de Dados Geográficos*, Eds Fuks, S.D., CARVALHO, M.S., CAMARA, G., MONTEIRO, A.M.V. DPI/INPE. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/> Acesso em 05/12/2007.
- [9] CASTAÑEDA, D. F. N. Estimação em Pequenas áreas para dados discretos via modelos hierárquicos. [Dissertação de Mestrado], IM-UFRJ, 1999.
- [10] CRESSIE, N.A.C. Statistics for Spatial Data. Wiley & Sons, 1993.
- [11] CUNHA, M.B.A.M. Mapas da Pobreza: Avaliação da metodologia e de sua aplicação para retratar a pobreza no Rio Grande do Norte, 2006 127f. Dissertação de mestrado, ENCE/IBGE, 2006.
- [12] DATTA, G.S. e GHOSH, M. Bayesian prediction in linear models: applications to small area estimation. Annals of Statistics, 13, 262-271, 1991.

- [13] DAVIS, B. Choosing a method for poverty mapping. Agriculture and Economic Development Analysis Division, Food and Agriculture Organization of the United Nations, Rome, 2003.
- [14] DEMPSTER, A.P. e TOMBERLIN, T.J. The analysis of census undercount from a postenumeration survey, Proceedings of the Conference on Census Undercount, Arlington, Va., 88–94, 1980.
- [15] DUNCAN, C., JONES, K.; MOON, G. Context, Composition and Heterogeneity: Using Multilevel Models in Health Research. Social Science and Medicine, 46: 97-117, 1998.
- [16] ELBERS, C., LANJOUW, J.O., LANJOUW, P. **Welfare in Villages and Towns: Micro-Level Estimation of Poverty and Inequality**, mimeo, Development Economics Research Group, the World Bank. TINBERGEN INSTITUTE WORKING PAPER, 2000.
- [17] ELBERS, C., e LANJOUW, P. Intersectoral Transfer, Growth, and Inequality in Rural Ecuador," World Development, 29, 3, 481-96, 2001.
- [18] ELBERS, C.; LANJOUW, J. O.; LANJOUW, P. **Micro-level estimation of welfare**. Washington DC: World Bank, October, 64p. Policy Research Working Paper nº 2911, 2002.
- [19] FARRELL, P.J., MACGIBBON, B. e TOMBERLIN, T.J. Protection against outliers in empirical Bayes estimation, Canadian Journal of Statistics, 22, 365–376, 1997.
- [20] FARRELL, P.J. Bayesian Inference for Small Area Proportions, The Indian Journal of Statistics, 62, Series B, Pt. 3, pp. 402–416, 2000.
- [21] FAY, R. E., HERRIOT, R. A. Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269-277, 1979.
- [22] FERREIRA, F. H. G.; LANJOUW, P.; NERI, M. **A new poverty profile for Brazil using PPV, PNAD and census data**. Revista Brasileira de Economia, v. 57,n.1, p.59-92, 2003.
- [23] GAMERMAN, D. Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference. Chapman & Hall, 1997.
- [24] GEMAN, S. and GEMAN, D. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741, 1984.
- [25] GIBSON, J., G. DATT, B. ALLEN, V. HWANG, R.M. BOURKE, D. PARAJULI, "Mapping Poverty in Rural Papua New Guinea," mimeo, 2004.
- [26] GOLDSTEIN, H. Multilevel Statistical Models. London: Arnold. (Disponível na página [www.mlwin.com](http://www.mlwin.com)), 1995.
- [27] GONZALES, M. E. "Use and Evaluation of Synthetic Estimates". Proceedings of the American Statistical Association, Social Statistics Section, pp. 33-36, 1973.

- [28] GONZALES, M. E. and HOZA, C. "Small Area Estimation with Application to Unemployment and Housing Estimates." *J. Amer. Statist. Assoc.*, vol. 73, pp. 7-15, 1978.
- [29] GOSH, M. and RAO, J.N.K. (1994). Small area estimation: an appraisal, *Statistical Science*, 9, 55–93.
- [30] HASTINGS, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57,97-109, 1970.
- [31] HENTSCHEL, J., LANJOUW, J.O. P., LANJOUW, and POGGI, J. "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador," *The World Bank Economic Review*, 14 (1): 147-65, 2000.
- [32] HOLT, D., SMITH, T.M.F. e TOMBERLIN, T.J. A Model-Based Approach to Estimation for Small Subgroups of a Population. *J.Amer.Stat.Assoc.* 74, 405-410, 1979.
- [33] IBGE. Pesquisa de orçamentos familiares 2002-2003 - Primeiros resultados: Brasil e grandes regiões, Rio de Janeiro: IBGE, 2004.
- [34] JESUS, C. D. Uma análise bayesiana para dados educacionais. [Dissertação de mestrado]. IM-UFRJ, 2004.
- [35] LUSTOSA, T Q O e FIGUEIREDO, J B B. Pobreza no Brasil: métodos de análise e resultados. Rio de Janeiro: IEI/UFRJ, (Texto para Discussão nº 205), 1989.
- [36] MACHADO, K.M. Uma comparação entre a renda e as abordagens multidimensionais da pobreza na Bahia, II Encontro de Economia baiana, Bahia, 2006.
- [37] MATTOS, E.J., WAQUIL, P.D., Pobreza rural no Rio Grande do Sul: comparando abordagens. 3º Encontro de Economia Gaúcha, 2006.
- [38] McCULLAGH, P. and NELDER, J. A. *Generalized Linear Models*. 2nd ed. Chapman and Hall, Londres, 1989.
- [39] MEDICI, E. V. Relatório com a especificação de temas variáveis e indicadores selecionadas. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2001a.
- [40] MEDICI, E. V. Relatório com arquivos gerados das variáveis selecionadas. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2001b.
- [41] MEDICI, E. V. Arquivo com indicadores, tabelas e gráficos sobre desigualdade social. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2001c.
- [42] MEDICI, E. V. Relatório metodológico com avaliação da metodologia e testes estatísticos utilizados. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2001d.



- [43] MEDICI, E. V. Relatório técnico com a metodologia utilizada, principais indicadores e medidas de desigualdade. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE. Projeto BRA/97/013, 2001e.
- [44] METROPOLIS, N., ROSENBLUTH, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1089-1091, 1953.
- [45] MIGON, H. S. & GAMERMAN, D. *Statistical Inference - an integrated approach*. Arnold, 1999.
- [46] MISTIAEN, J., B. OZLER, T. RAZAFIRMANANTENA, and J. RAZAFIRMANANTENA “Putting Welfare on the Map in Madagascar,” African Region Working Paper Series No.34, 2001.
- [47] MONTEIRO, C.A. A dimensão da pobreza, da desnutrição e da fome no Brasil. *Estudos Avançados*, São Paulo, v. 17, n. 48, p. 7-20, 2003.
- [48] MOTTA, G.O. Pobreza, desigualdade e heterogeneidade nas favelas do Rio de Janeiro, 2000 66f. Monografia, Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.
- [49] MOURA, F.A.S. and HOLT, D. Small Area Estimation Using Multilevel Models. *Survey Methodology*. Vol 25, 1, 73-80, 1999.
- [50] MOURA, F. A. S.; MIGON, H. S. Bayesian Spatial Models for small area estimation of proportions. *Statistical Modeling*, v. 2, n. 3, p. 183-201, 2002.
- [51] PESSOA, D. G. C.; SILVA, P. L. do N. **Análise de dados amostrais complexos**. São Paulo: Associação Brasileira de Estatística. 1998.
- [52] PRASAD, N.G.N. e RAO, J.N.K. The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, 85, 163-171, 1990.
- [53] QUINTAES, VCC; SOUZA, D.F.; MACHADO, D.C.; ASSIS, J.M.; DUARTE, C.B.; SANTOS, I.C.M. Análise do Bem-estar a partir de um indicador de consumo agregado com base nas informações da POF 2002-2003. XV Encontro Nacional de Estudos Populacionais - ABEP, 2006.
- [54] RAO, J. N. K. Small Area Estimation. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.
- [55] RESENDE, M. D. V. e BIELE, J. Estimção e predição em modelos lineares generalizados mistos com variáveis binomiais. *Rev. Mat. Estat.*, São Paulo, 20: 39-65, 2002.
- [56] ROCHA, S. Estabelecimento e comparação de linhas de pobreza para o Brasil. Rio de Janeiro: IPEA. (Texto para Discussão n° 153), 1988.

- [57] ROCHA, S. A Investigação da Renda nas Pesquisas Domiciliares. *Economia e Sociedade*, Campinas, v. 12 (2), p. 205-224, 2003.
- [58] ROCHA, S. Pobreza no Brasil: O que mudou nos últimos 30 anos. *Estudos e Pesquisas*, nº83. Seminário Especial, Mini-fórum em homenagem aos 40 anos do IPEA, Rio de Janeiro, 2004.
- [59] ROMÃO, M. C. Pobreza: Conceito e Mensuração. Rio de Janeiro: IPEA. (Caderno de Economia nº13), 1993.
- [60] SALLES, V.; TUIRÁN, R. ¿Cargan lãs mujeres com el peso de la pobreza? Puntos de vista de um debate. In: GARCIA, B. (Coord.). *Mujer, género e población em México*. México: SOMEDE/El Colégio de México, 2000.
- [61] SÄRNDAL, C.E. Design-Consistent versus Model-Dependent Estimation for Small Domains, *J.Amer.Stat.Assoc.* 79,624-631, 1984.
- [62] SCHIPPER, Y., HOOGEVEEN, J. G. Which inequalities matters? Growth Evidence Based on Small Area Welfare Estimates in Uganda, *World Bank Policy Research Working Paper 3592*, 2005.
- [63] SCHMIDT, A. M., NOBRE, A. A., FERREIRA, G. S. Alguns aspectos da modelagem de dados espacialmente referenciados (artigo) Departamento de Métodos Estatísticos do IM/UFRJ, 2003.
- [64] SCHWARTZMAN, S. Redução da desigualdade, da pobreza e os programas de transferência de renda. Instituto de Estudos do Trabalho e Sociedade, 2006.
- [65] SEN, A. Poor, relatively speaking. *Oxford Economic Papers*, vol. 35, nº 1, p. 153-169, mar. 1983.
- [66] SEN, A. *Desenvolvimento como liberdade*, São Paulo, Companhia das Letras, 2000.
- [67] SEN, A. *Desigualdade reexaminada*, Rio de Janeiro, Record, 2001.
- [68] SNIJDERS, T. & BOSKER, R. *An Introduction to basic and Advanced Multilevel Modeling*. London: Sage, 1999.
- [69] SOUZA, D.F., *Estimação de População em Nível Municipal via Modelos Hierárquicos e Espaciais*. [Dissertação de mestrado]. IM-UFRJ, 2004.
- [70] THOMAS, A., SPIEGELHALTER, D. J. and GILKS, W. R. BUGS: A program to perform Bayesian inference using Gibbs Sampling. Em *Bayesian Statistics*. 4, 837-842 (Editors Bernardo, J. M., Berger, J. O., Dawid, A.P. and Smith, A. F. M.), 1992.
- [71] VIEIRA, S. P. Crescimento econômico, desenvolvimento humano e pobreza: análise da situação em Moçambique. Documentos de trabalho nº 68, Centro de Estudos sobre África e do Desenvolvimento do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa, Lisboa, 2005.

- [72] WORLD BANK, "Poverty and Social Impact Assessment Case Study: Welfare Reform in Sri Lanka," forthcoming as a South Asia PREM Policy Note, 2005.
- [73] YOU, Y, RAO J.N.K. Hierarchical Bayes small area estimation using multi-level models. A draft paper. Department of Mathematics and Statistics, Carleton University. Canada, 1997.

## Anexos

**Quadro A1: Descrição das variáveis utilizadas no ajuste dos modelos estatísticos (continua)**

Variáveis	Categorias
ABASTEC Forma de abastecimento de água	1 Rede geral 2 Poço ou Nascente 3 Outra forma
TIPOESC Tipo de escoamento	1 Rede geral 2 Fossa 3 Vala, rio e outro 4 Não tem
QTDBANH Número de banheiros no domicílio	1 0 ou 1 banheiro 2 2 banheiros 3 3 banheiros 4 4 banheiros ou mais
CONDDOM Condição do domicílio	1 Próprio 2 Alugado 3 Cedido ou outro
TIPODOM Tipo de domicílio	1 Casa 2 Apartamento 3 Cômodo
PAVIMEN Existência de pavimentação da rua do domicílio	1 Tem 2 Não tem
ILUMINA Existência de iluminação elétrica no domicílio	0 Não tem 1 Tem
COMPUT Posse de computador	0 Não tem 1 Tem
GELFREE Posse de geladeira ou freezer	0 Não tem 1 Tem
MQLAVAR Posse de máquina de lavar	0 Não tem 1 Tem
AR Posse de ar condicionado	0 Não tem 1 Tem
MICROO Posse de microondas	0 Não tem 1 Tem
TV Posse de Televisão	0 Não tem 1 Tem
MORA Classes de total de moradores no domicílio	1 1 morador 2 2 moradores 3 3 moradores 4 4 moradores 5 5 moradores ou mais
CONJUGE Existência de cônjuge no domicílio	0 Não tem 1 Tem

Fonte: Adaptado de CUNHA (2006).

**Quadro A1: Descrição das variáveis utilizadas no ajuste dos modelos estatísticos  
(continuação)**

Variáveis	Categorias
<b>IDOSOS</b> Quantidade de idosos (mais de 64 anos de idade) no domicílio	1 0 idoso 2 1 idoso 3 2 idosos 4 3 idosos ou mais
<b>ADOLES</b> Quantidade de adolescentes (7 a 14 anos de idade) no domicílio	1 0 adolescente 2 1 adolescente 3 2 adolescentes 4 3 adolescentes 5 4 adolescentes ou mais
<b>CRIAN</b> Quantidade de crianças (até 6 anos de idade) no domicílio	1 0 criança 2 1 criança 3 2 crianças 4 3 crianças 5 4 crianças ou mais
<b>TIPOFAM</b> Tipos de família	1 Unipessoais 2 Casal sem filhos 3 Casal com filhos 4 Homem chefe com filhos 5 Mulher chefe com filhos 6 Outras
<b>COR4</b> Cor do chefe da UC	1 Branca 2 Não branca
<b>SEX</b> Sexo do chefe da UC	1 Masculino 2 Feminino
<b>ANOEST</b> Classe de anos de estudo do chefe da UC	1 Sem instrução ou menos de 1 ano 2 1 a 3 anos 3 4 a 7 anos 4 8 a 10 anos 5 11 a 14 anos 6 15 anos ou mais 9 Ignorado
<b>POS</b> Posição na ocupação do chefe da UC	1 Empregada(o) doméstica(o) 2 Empregado(a) 3 Empregador(a) 4 Conta-própria 5 Aprendiz, estagiário e voluntário sem rem. 6 Próprio consumo 9 Outros rendimentos ou desempregado

Fonte: Adaptado de CUNHA (2006).

**Tabela A1: Listagem das microrregiões potiguares e seus municípios.**

<b>Código</b>	<b>Microrregião</b>	<b>Municípios</b>
24001	<b>Mossoró</b>	Areia Branca, Baraúna, Grossos, Mossoró, Serra do Mel, Tibau
24002	<b>Chapada do Apodi</b>	Apodi, Caraúbas, Felipe Guerra, Governador Dix-Sept Rosado
24003	<b>Médio Oeste</b>	Augusto Severo, Janduís, Messias Targino, Paraú, Triunfo Potiguar, Upanema
24004	<b>Vale do Açu</b>	Açu, Alto do Rodrigues, Carnaubais, Ipanguaçu, Itajá, Jucurutu, Pendências, Porto do Mangue, São Rafael
24005	<b>Serra de São Miguel</b>	Água Nova, Coronel João Pessoa, Doutor Severiano, Encanto, Luís Gomes, Major Sales, Riacho de Santana, São Miguel, Venha-Ver
24006	<b>Pau dos Ferros</b>	Alexandria, Francisco Dantas, Itaú, José da Penha, Marcelino Vieira, Paraná, Pau dos Ferros, Pilões, Portalegre, Rafael Fernandes, Riacho da Cruz, Rodolfo Fernandes, São Francisco do Oeste, Severiano Melo, Taboleiro Grande, Tenente Ananias, Viçosa
24007	<b>Umarizal</b>	Almino Afonso, Antônio Martins, Frutuoso Gomes, João Dias, Lucrecia, Martins, Olho d'água do Borges, Patu, Rafael Godeiro, Serrinha dos Pintos, Umarizal
24008	<b>Macau</b>	Caiçara do Norte, Galinhos, Guamaré, Macau, São Bento do Norte
24009	<b>Angicos</b>	Afonso Bezerra, Angicos, Caiçara do Rio do Vento, Fernando Pedroza, Jardim de Angicos, Lajes, Pedra Preta, Pedro Avelino
24010	<b>Serra de Santana</b>	Bodó, Cerro Corá, Florânia, Lagoa Nova, Santana do Matos, São Vicente, Tenente Laurentino Cruz
24011	<b>Seridó Ocidental</b>	Caicó, Ipueira, Jardim de Piranhas, São Fernando, São João do Sabugi, Serra Negra do Norte, Timbaúba dos Batistas
24012	<b>Seridó Oriental</b>	Acari, Carnaubas dos Dantas, Cruzeta, Currais Novos, Equador, Jardim do Seridó, Ouro Branco, Parelhas, Santana do Seridó, São José do Seridó
24013	<b>Baixa Verde</b>	Bento Fernandes, Jandaíra, João Câmara, Parazinho, Poço Branco
24014	<b>Borborema Potiguar</b>	Barcelona, Campo Redondo, Coronel Ezequiel, Jaçaná, Japi, Lagoa de Velhos, Lajes Pintadas, Monte das Gameleiras, Ruy Barbosa, Santa Cruz, São Bento do Trairí, São José do Campestre, São Tomé, Serra de São Bento, Sítio Novo, Tangará,
24015	<b>Agreste Potiguar</b>	Bom Jesus, Brejinho, Ielmo Marinho, Januário Cicco, Lagoa D'anta, Lagoa de Pedras, Lagoa Salgada, Monte Alegre, Nova Cruz, Passa e Fica, Passagem, Presidente Juscelino, Riachuelo, Santa Maria, Santo Antônio, São Paulo do Potengi, São Pedro, Senador Elói de Souza, Serrinha, Varzea, Vera Cruz
24016	<b>Litoral Nordeste</b>	Maxaranguape, Pedra Grande, Pureza, Rio do Fogo, São Miguel de Touros, Taipu, Touros
24017	<b>Macaíba</b>	Ceará-Mirim, Macaíba, Nísia Floresta, São Gonçalo do Amarante, São José do Mipibu
24018	<b>Natal</b>	Extremoz, Natal, Parnamirim
24019	<b>Litoral Sul</b>	Arês, Baía Formosa, Canguaretama, Espírito Santo, Goianinha, Montanhas, Pedro Velho, Senador Georgino Avelino, Tibau do Sul, Vila Flor

Fonte: Microdados do Censo 2000.

**Tabela A2: Estimativas dos parâmetros e dos p-valores dos modelos de efeitos principais – Logaritmo do Rendimento Domiciliar per capita – Metodologia do Banco Mundial.**

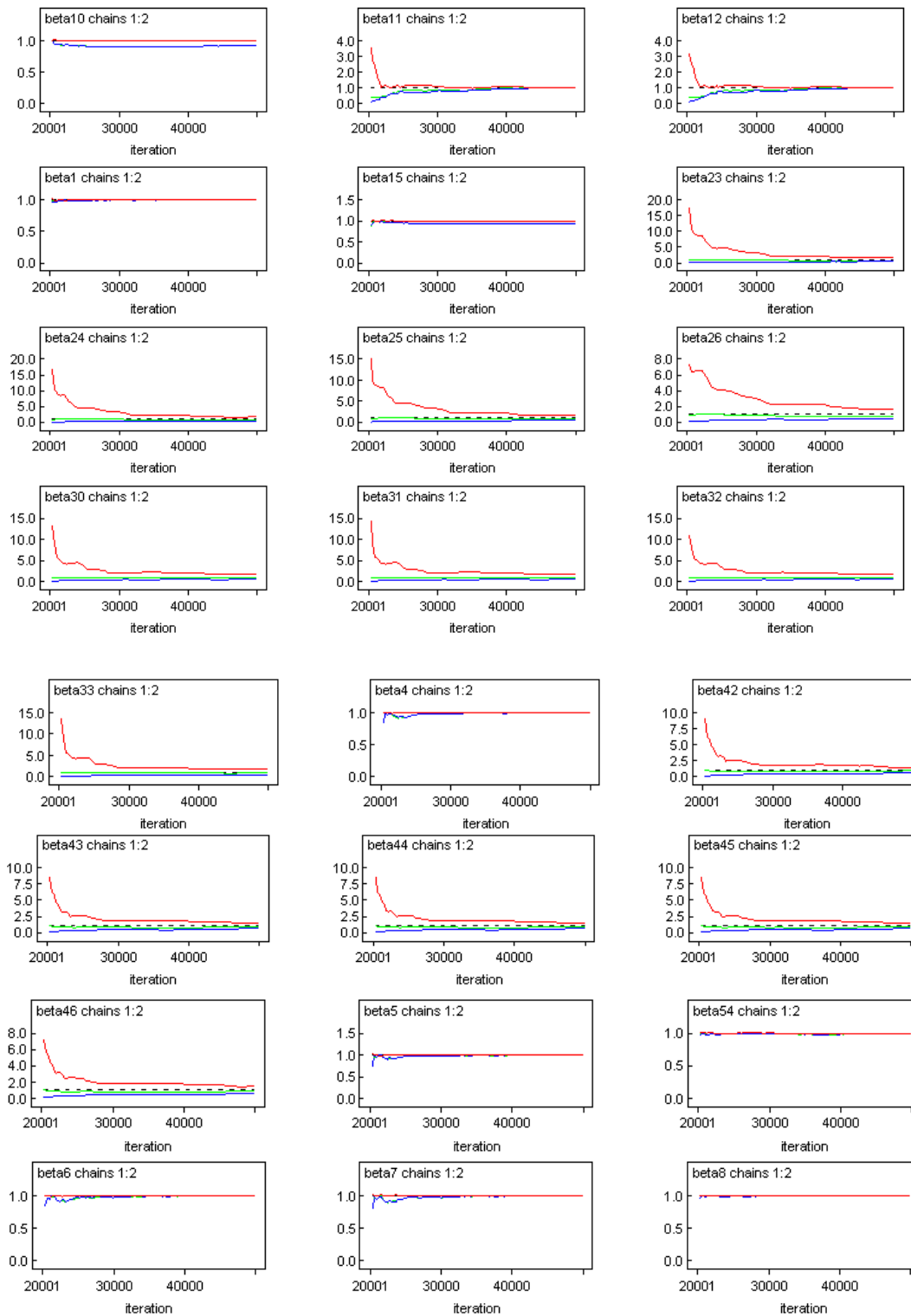
Parâmetros	$\beta$	RENDA	
		Estimativa	P-valor
Intercepto	$\beta_0$	4,5280916	<,0001
COMPUT 1	$\beta_1$	0,6308940	0,0439
MICROO 1	$\beta_2$	0,3688927	<,0001
MORA 1	$\beta_3$	0,7830180	<,0001
MORA 2	$\beta_4$	0,4050915	0,4978
MORA 3	$\beta_5$	1,0311705	0,0003
MORA 4	$\beta_6$	0,4608760	0,0594
MQLAVAR 1	$\beta_7$	0,3506460	<,0001
AR 1	$\beta_8$	0,4722942	<,0001
PAVIMEN 1	$\beta_9$	0,1655225	0,0002
QTDBANH 1	$\beta_{10}$	-0,4179692	0,1122
QTDBANH 2	$\beta_{11}$	-0,0070248	0,9783
QTDBANH 3	$\beta_{12}$	-0,2479228	0,317
TV 1	$\beta_{13}$	0,1416237	0,0351
ILUMINA 1	$\beta_{14}$	0,0575054	0,6555
ANOEST 1	$\beta_{15}$	-0,2264057	0,3222
ANOEST 2	$\beta_{16}$	-0,1931825	0,3858
ANOEST 3	$\beta_{17}$	-0,1216418	0,574
ANOEST 4	$\beta_{18}$	0,1106508	0,6111
ANOEST 5	$\beta_{19}$	0,4152942	0,0864
ANOEST 6	$\beta_{20}$	0,9215417	0,003
IDOSOS 1	$\beta_{21}$	-0,5128820	0,0345
IDOSOS 2	$\beta_{22}$	-0,3203834	0,1906
IDOSOS 3	$\beta_{23}$	-0,2167443	0,3914
CONDDOM 1	$\beta_{24}$	0,2402307	<,0001
CONDDOM 2	$\beta_{25}$	0,0475373	0,499
TIPOESC 1	$\beta_{26}$	0,3413894	0,0002
TIPOESC 2	$\beta_{27}$	0,2857897	0,0003
TIPOESC 3	$\beta_{28}$	0,0163607	0,9165
CONJUGE 1	$\beta_{29}$	0,1281552	0,0038
COR4 1	$\beta_{30}$	0,0915751	0,0172
POS 1	$\beta_{31}$	-0,2172441	0,0505
POS 2	$\beta_{32}$	-0,0564109	0,404
POS 3	$\beta_{33}$	0,4557777	0,0004
POS 4	$\beta_{34}$	-0,1077098	0,1353
POS 5	$\beta_{35}$	0,5226505	0,0097
ADOLE 1	$\beta_{36}$	0,4169790	0,0089
ADOLE 2	$\beta_{37}$	0,3107255	0,0445
ADOLE 3	$\beta_{38}$	0,1652052	0,2904
ADOLE 4	$\beta_{39}$	0,1447168	0,3827

**Tabela A3: Estimativas dos parâmetros e dos p-valores dos modelos de efeitos principais – Logaritmo do Consumo Domiciliar per capita– Metodologia do Banco Mundial.**

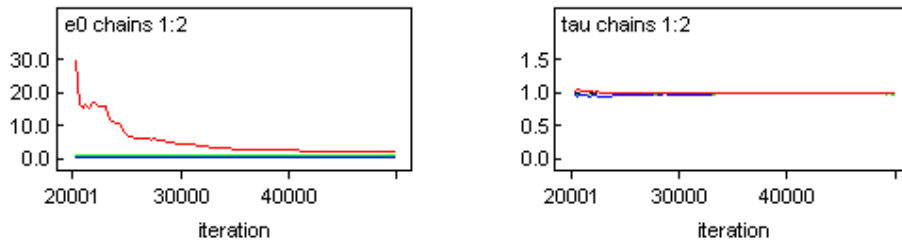
Parâmetros	$\beta$	CONSUMO	
		Estimativa	P-valor
Intercepto	$\beta_0$	4,0189064	<.0001
COMPUT 1	$\beta_1$	0,4213627	0.1134
MICROO 1	$\beta_2$	0,3479109	<.0001
MORA 1	$\beta_3$	1,2205792	<.0001
MORA 2	$\beta_4$	0,8059386	0.1116
MORA 3	$\beta_5$	1,0202548	0.0012
MORA 4	$\beta_6$	0,7934077	0.0148
MQLAVAR 1	$\beta_7$	0,2690898	<.0001
TV 1	$\beta_8$	0,2256942	<.0001
AR 1	$\beta_9$	0,3765046	<.0001
QTDBANH 1	$\beta_{10}$	-0,4450831	0.1509
QTDBANH 2	$\beta_{11}$	-0,0976270	0.7580
QTDBANH 3	$\beta_{12}$	-0,1885518	0.5265
CONDDOM 1	$\beta_{13}$	0,2514276	<.0001
CONDDOM 2	$\beta_{14}$	0,0858213	0.1895
ANOEST 1	$\beta_{15}$	-0,4684026	0.0047
ANOEST 2	$\beta_{16}$	-0,3730270	0.0192
ANOEST 3	$\beta_{17}$	-0,2810348	0.0613
ANOEST 4	$\beta_{18}$	-0,0113723	0.07
ANOEST 5	$\beta_{19}$	-0,0151126	0.09
ANOEST 6	$\beta_{20}$	0,2292067	0.02
PAVIMEN 1	$\beta_{21}$	0,1334368	0.09
TIPODOM 1	$\beta_{22}$	0,4515811	0.23
TIPODOM 2	$\beta_{23}$	0,4635569	0.08
ADOLE 1	$\beta_{24}$	0,3602187	0.44
ADOLE 2	$\beta_{25}$	0,31138	0.04
ADOLE 3	$\beta_{26}$	0,1892987	0.69
ADOLE 4	$\beta_{27}$	-0,0100937	0.08



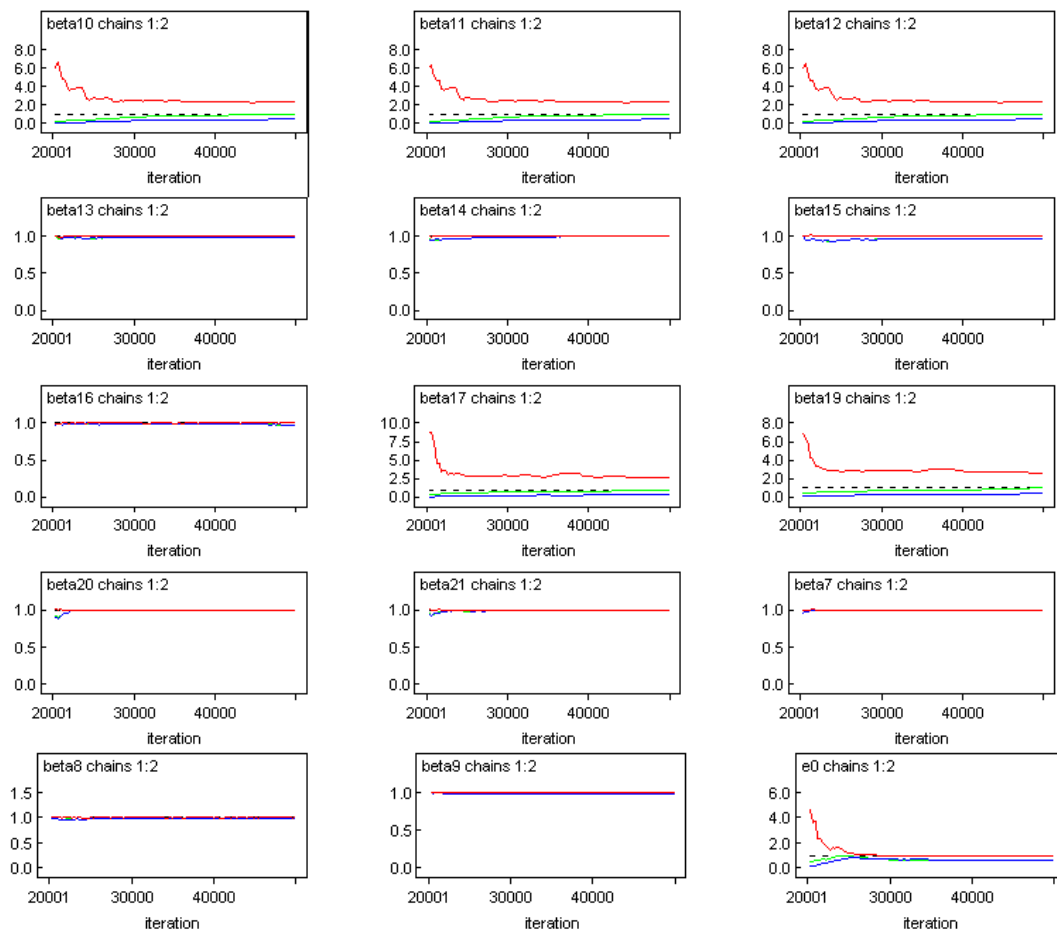
**Figura A1:** Gráficos de diagnóstico de convergência de Gelman e Rubin – Rendimento domiciliar per capita (continua)



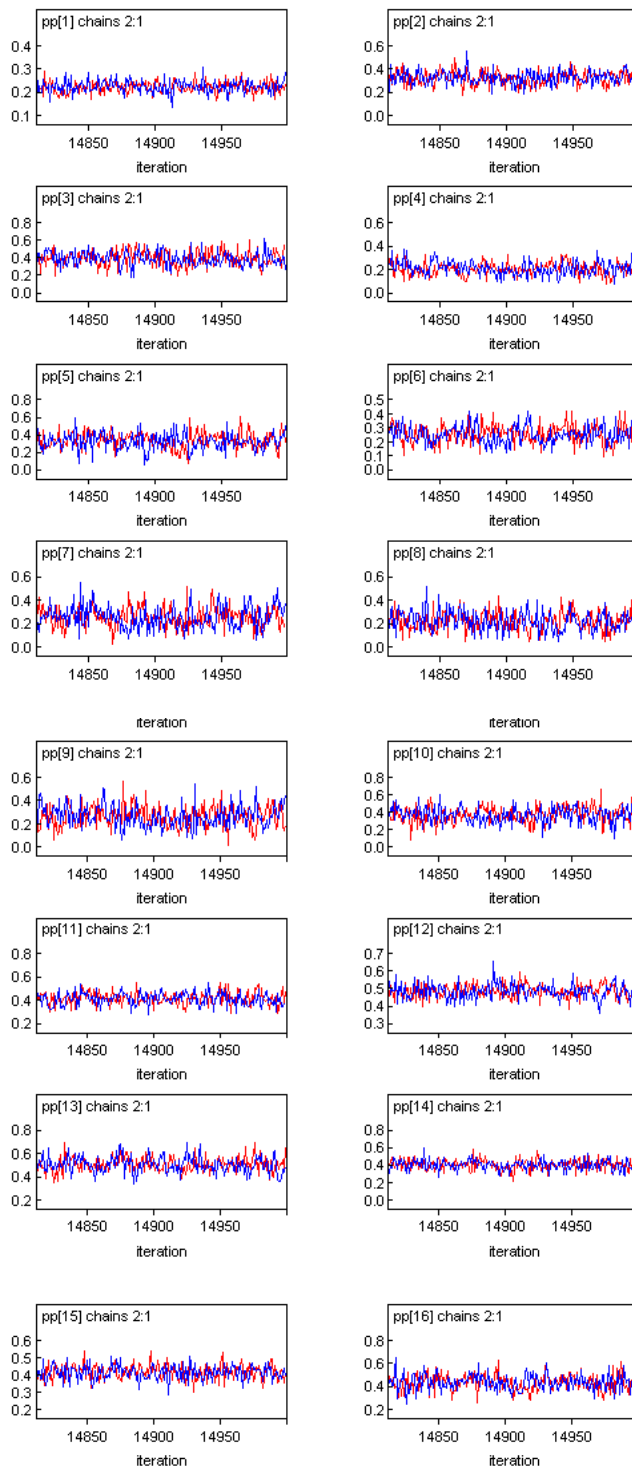
**Figura A1:** Gráficos de diagnóstico de convergência Gelman e Rubin – Rendimento domiciliar per capita (continuação)



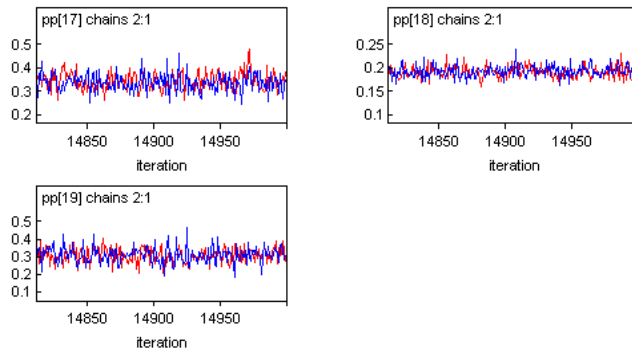
**Figura A2:** Gráficos de diagnóstico de convergência de Gelman e Rubin – Consumo domiciliar per capita (continuação)



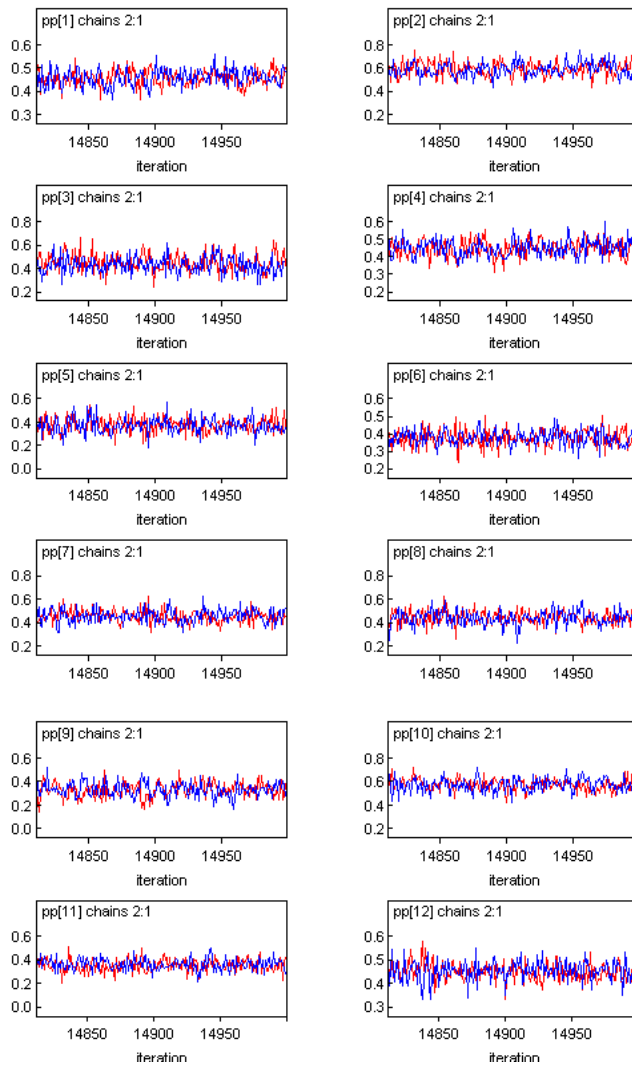
**Figura A3:** Gráficos das trajetórias das cadeias geradas para as proporções estimadas – Rendimento domiciliar per capita (continua).



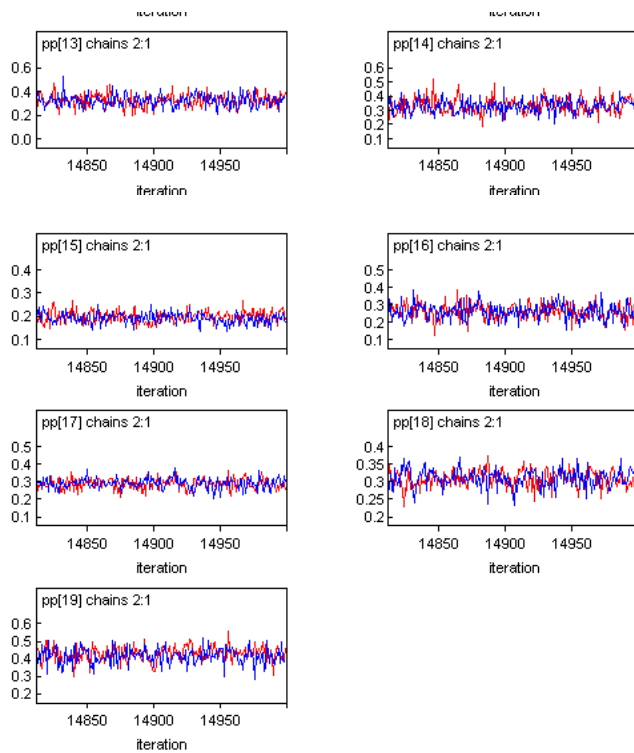
**Figura A3:** Gráficos das trajetórias das cadeias geradas para as proporções estimadas – Rendimento domiciliar per capita (continuação).



**Figura A4:** Gráficos das trajetórias das cadeias geradas para as proporções estimadas – Consumo domiciliar per capita (continua).



**Figura A4:** Gráficos das trajetórias das cadeias geradas para as proporções estimadas – Consumo domiciliar per capita (continuação).



## Códigos Utilizados – WinBUGS

### 1. Modelo Hierárquico de intercepto aleatório – Rendimento domiciliar per capita

```
model
{

for(i in 1:19) {
#erro do nível 1
erro1[i] ~ dnorm(0.0, tau)
  for (k in n[i] + 1:n[i+1]) {
    emicro[k]<-erro1[i]
  }
}

for(j in 1:1550) {

  P0_R[j] ~ dbern(p[j])

  logit(p[j]) <- (beta0 + emicro[j]) + (beta1*COMPUT[j] + beta4*MORA2[j]
+ beta5*MORA3[j] + beta6*MORA4[j] + beta7*MORA5[j] +
beta8*MQLAVAR[j] + beta10*PAVIMEN[j]+ beta11*QTDBANH1[j] +
beta12*QTDBANH2[j] + beta15*TV[j] + beta23*IDOSOS1[j] +
beta24*IDOSOS2[j] + beta25*IDOSOS3[j] + beta26*IDOSOS4[j] +
beta30*TIPOESC1[j] + beta31*TIPOESC2[j] + beta32*TIPOESC3[j] +
beta33*TIPOESC4[j] + beta42*ADOLEES1[j] + beta43*ADOLEES2[j] +
beta44*ADOLEES3[j] + beta45*ADOLEES4[j] + beta46*ADOLEES5[j] +
beta54*TIPODOM2[j])

#criando réplica dos dados
y[j]~dbern(p[j])

#criando réplica dos dados para a classificação
yr[j]~dbern(p[j])

a [j] <- equals (y[j], yr[j])

# Proporcao de acerto nas microrregioes da POF
}
for(i in 1:19) {
  papof[i] <- sum(a[(n[i]+1):(n[i+1]])) / (n[i+1] - n[i])
}

#Prioris para Efeitos Fixos
  beta0 ~ dnorm(0.0,0.001)
  beta1 ~ dnorm(0.0,0.001)
  beta4 ~ dnorm(0.0,0.001)
  beta5 ~ dnorm(0.0,0.001)
  beta6 ~ dnorm(0.0,0.001)
  beta7 ~ dnorm(0.0,0.001)
  beta8 ~ dnorm(0.0,0.001)
  beta10 ~ dnorm(0.0,0.001)
  beta11 ~ dnorm(0.0,0.001)
  beta12 ~ dnorm(0.0,0.001)
  beta15 ~ dnorm(0.0,0.001)
  beta23 ~ dnorm(0.0,0.001)
  beta24 ~ dnorm(0.0,0.001)
  beta25 ~ dnorm(0.0,0.001)
  beta26 ~ dnorm(0.0,0.001)
```

```

beta30 ~ dnorm(0.0,0.001)
beta31 ~ dnorm(0.0,0.001)
beta32 ~ dnorm(0.0,0.001)
beta33 ~ dnorm(0.0,0.001)
beta42 ~ dnorm(0.0,0.001)
beta43 ~ dnorm(0.0,0.001)
beta44 ~ dnorm(0.0,0.001)
beta45 ~ dnorm(0.0,0.001)
beta46 ~ dnorm(0.0,0.001)
beta54 ~ dnorm(0.0,0.001)

#Prioris para Efeitos Aleatórios

tau ~ dgamma(0.5, 0.0005)

# IMPUTAÇÃO DA RENDA NO CENSO

for(i in 1:19) {
  for(k in nc[i] + 1:nc[i+1]) {
    emicroc[k]<-erro1[i]
  }
}

for(j in 1:91805) {

logit(pc[j]) <- (beta0 + emicroc[j]) + (beta1*COMPUTC[j] +
beta4*MORA2C[j] + beta5*MORA3C[j] + beta6*MORA4C[j] +
beta7*MORA5C[j] + beta8*MQLAVARC[j] + beta10*PAVIMENC[j]+
beta11*QTDBANH1C[j] + beta12*QTDBANH2C[j] + beta15*TVC[j] +
beta23*IDOSOS1C[j] + beta24*IDOSOS2C[j] + beta25*IDOSOS3C[j] +
beta26*IDOSOS4C[j] + beta30*TIPOESC1C[j] + beta31*TIPOESC2C[j] +
beta32*TIPOESC3C[j] + beta33*TIPOESC4C[j] + beta42*ADOLESC1C[j] +
beta43*ADOLESC2C[j] + beta44*ADOLESC3C[j] + beta45*ADOLESC4C[j] +
beta46*ADOLESC5C[j] + beta54*TIPODOM2C[j])

#criando réplica dos dados da amostra do Censo para o ajuste
yc[j]~dbern(pc[j])

}

for(i in 1:19) {
# Proporção de pobres nas microrregiões do CENSO

pp[i]<-sum(yc[(nc[i]+1):(nc[i+1]]))/(nc[i+1] - nc[i])
}
}

```

## 2. Modelo Hierárquico de intercepto aleatório – Consumo domiciliar per capita

```
model
{
  for(i in 1:19) {
    #erro do nível 1
    erro1[i] ~ dnorm(0.0, tau)
    for(k in n[i] + 1:n[i+1]) {
      emicro[k] <- erro1[i]
    }
  }

  for(j in 1:1550) {

    P0_D[j] ~ dbern(p[j])

    logit(p[j]) <- (beta0+emicro[j]) + (beta1*COMPUT[j] + beta2*MICROO[j]
+ beta3*MORA1[j] + beta4*MORA3[j]+ beta5*MORA4[j] + beta6*MORA5[j] +
beta7*MLAVAR[j] + beta8*TV[j]+ beta9*AR[j] + beta10*CONDDOM1[j] +
beta11*CONDDOM2[j]+ beta12*CONDDOM3[j]+ beta13*ANOEST1[j]+
beta14*ANOEST2[j]+ beta15*ANOEST3[j] + beta16*PAVIMEN[j] +
beta17*TIPODOM1[j]+ beta18*TIPODOM2[j]+ beta19*TIPODOM3[j] +
beta20*ADOLESI1[j] + beta21*ADOLESI2[j])

    #criando réplica dos dados
    y[j]~dbern(p[j])

    #criando réplica dos dados para a classificação
    yr[j]~dbern(p[j])

    a[j] <- equals(y[j], yr[j])

  }

  # Proporcao de acerto nas microrregioes da POF

  for(i in 1:19) {
    papof[i] <- sum(a[(n[i]+1):(n[i+1])]) / (n[i+1] - n[i])
  }

  #Priors para Efeitos Fixos
  beta0 ~ dnorm(0.0,0.001)
  beta1 ~ dnorm(0.0,0.001)
  beta2 ~ dnorm(0.0,0.001)
  beta3 ~ dnorm(0.0,0.001)
  beta4 ~ dnorm(0.0,0.001)
  beta5 ~ dnorm(0.0,0.001)
  beta6 ~ dnorm(0.0,0.001)
  beta7 ~ dnorm(0.0,0.001)
  beta8 ~ dnorm(0.0,0.001)
  beta9 ~ dnorm(0.0,0.001)
  beta10 ~ dnorm(0.0,0.001)
  beta11 ~ dnorm(0.0,0.001)
  beta12 ~ dnorm(0.0,0.001)
  beta13 ~ dnorm(0.0,0.001)
  beta14 ~ dnorm(0.0,0.001)
  beta15 ~ dnorm(0.0,0.001)
  beta16 ~ dnorm(0.0,0.001)
  beta17 ~ dnorm(0.0,0.001)
}
```



```

beta18 ~ dnorm(0.0,0.001)
beta19 ~ dnorm(0.0,0.001)
beta20 ~ dnorm(0.0,0.001)
beta21 ~ dnorm(0.0,0.001)

```

```
#Prioris para Efeitos Aleatórios
```

```
tau ~ dgamma(0.5, 0.0005)
```

```
# IMPUTAÇÃO DO CONSUMO NO CENSO
```

```

for(i in 1:19) {
  for (k in nc[i] + 1:nc[i+1]) {
    emicroc[k]<-erro1[i]
  }
}

for(j in 1:91805) {

logit(pc[j]) <- (beta0+emicroc[j]) + (beta1*COMPUTC[j] +
beta2*MICROOC[j] + beta3*MORA1C[j] + beta4*MORA3C[j]+ beta5*MORA4C[j]
+ beta6*MORA5C[j] + beta7*MQLAVARC[j] + beta8*TVC[j]+ beta9*ARC[j] +
beta10*CONDDOM1C[j] + beta11*CONDDOM2C[j]+ beta12*CONDDOM3C[j]+
beta13*ANOEST1C[j]+ beta14*ANOEST2C[j]+ beta15*ANOEST3C[j] +
beta16*PAVIMENC[j] + beta17*TIPODOM1C[j]+ beta18*TIPODOM2C[j]+
beta19*TIPODOM3C[j] + beta20*ADOLESC1C[j] + beta21*ADOLESC2C[j])

#criando réplica dos dados da amostra do Censo para o ajuste
yc[j]~dbern(pc[j])
}

for(i in 1:19) {
# Proporção de pobres nas microrregiões do CENSO

pp[i]<-sum(yc[(nc[i]+1):(nc[i+1])]) / (nc[i+1] - nc[i])

}

}

```

### 3. Modelo Autoregressivo Condicional – Rendimento domiciliar per capita

```
model
{
  for(i in 1:19) {
    for(k in n[i] + 1:n[i+1]) {
      emicro[k] <- e0 + espac[i]
    }
  }

  for(j in 1:1550) {

    P0_R[j] ~ dbern(p[j])

    logit(p[j]) <- (emicro[j] + beta1*COMPUT[j] + beta4*MORA2[j] +
beta5*MORA3[j] + beta6*MORA4[j] + beta7*MORA5[j] +
beta8*MQLAVAR[j] + beta10*PAVIMEN[j] + beta11*QTDBANH1[j] +
beta12*QTDBANH2[j] + beta15*TV[j] + beta23*IDOSOS1[j] +
beta24*IDOSOS2[j] + beta25*IDOSOS3[j] + beta26*IDOSOS4[j] +
beta30*TIPOESC1[j] + beta31*TIPOESC2[j] + beta32*TIPOESC3[j] +
beta33*TIPOESC4[j] + beta42*ADOLES1[j] + beta43*ADOLES2[j] +
beta44*ADOLES3[j] + beta45*ADOLES4[j] + beta46*ADOLES5[j] +
beta54*TIPODOM2[j])

    #criando réplica dos dados para o ajuste
    y[j] ~ dbern(p[j])

    #criando réplica dos dados para a classificação
    yr[j] ~ dbern(p[j])

    a[j] <- equals(y[j], yr[j])

  }

  # Proporcao de acerto nas microrregioes da POF

  for(i in 1:19) {
    papof[i] <- sum(a[(n[i]+1):(n[i+1])]) / (n[i+1] - n[i])
  }

  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  espac[1:19] ~ car.normal(adj[], weights[], num[], tau)

  #Prioris para Efeitos Fixos
  beta1 ~ dnorm(0.0, 0.001)
  beta4 ~ dnorm(0.0, 0.001)
  beta5 ~ dnorm(0.0, 0.001)
  beta6 ~ dnorm(0.0, 0.001)
  beta7 ~ dnorm(0.0, 0.001)
  beta8 ~ dnorm(0.0, 0.001)
  beta10 ~ dnorm(0.0, 0.001)
  beta11 ~ dnorm(0.0, 0.001)
  beta12 ~ dnorm(0.0, 0.001)
  beta15 ~ dnorm(0.0, 0.001)
  beta23 ~ dnorm(0.0, 0.001)
```

```

beta24 ~ dnorm(0.0,0.001)
beta25 ~ dnorm(0.0,0.001)
beta26 ~ dnorm(0.0,0.001)
beta30 ~ dnorm(0.0,0.001)
beta31 ~ dnorm(0.0,0.001)
beta32 ~ dnorm(0.0,0.001)
beta33 ~ dnorm(0.0,0.001)
beta42 ~ dnorm(0.0,0.001)
beta43 ~ dnorm(0.0,0.001)
beta44 ~ dnorm(0.0,0.001)
beta45 ~ dnorm(0.0,0.001)
beta46 ~ dnorm(0.0,0.001)
beta54 ~ dnorm(0.0,0.001)

#Prioris para Efeitos Aleatórios

tau ~ dgamma(0.5, 0.0005)
e0~dflat()

# IMPUTAÇÃO DA RENDA NO CENSO

for(i in 1:19) {
  for (k in nc[i] + 1:nc[i+1]) {
    emicroc[k]<-e0
  }
}

for(j in 1:91805) {

logit(pc[j]) <- (emicroc[j] + beta1*COMPUTC[j] + beta4*MORA2C[j] +
beta5*MORA3C[j] + beta6*MORA4C[j] + beta7*MORA5C[j] +
beta8*MQLAVARC[j] + beta10*PAVIMENC[j]+ beta11*QTDBANH1C[j] +
beta12*QTDBANH2C[j] + beta15*TVC[j] + beta23*IDOSOS1C[j] +
beta24*IDOSOS2C[j] + beta25*IDOSOS3C[j] + beta26*IDOSOS4C[j] +
beta30*TIPOESC1C[j] + beta31*TIPOESC2C[j] + beta32*TIPOESC3C[j] +
beta33*TIPOESC4C[j] + beta42*ADOLES1C[j] + beta43*ADOLES2C[j] +
beta44*ADOLES3C[j] + beta45*ADOLES4C[j] + beta46*ADOLES5C[j] +
beta54*TIPODOM2C[j])

#criando réplica dos dados da amostra do Censo para o ajuste
yc[j]~dbern(pc[j])

}

for(i in 1:19) {
# Proporção de pobres nas microrregiões do CENSO

pp[i]<-sum(yc[(nc[i]+1):(nc[i+1])]) / (nc[i+1] - nc[i])

}
}

```

#### 4. Modelo Autoregressivo Condicional – Consumo domiciliar per capita

```
model
{
  for(i in 1:19) {
    #erro do nível 1
    #erro1[i] ~ dnorm (0.0, tau)
    for (k in n[i] + 1:n[i+1]) {
      emicro[k]<-e0+espac[i]
    }
  }

  for(j in 1:1550) {

    P0_D[j] ~ dbern(p[j])

    logit(p[j]) <- (emicro[j] + beta1*COMPUT[j] + beta2*MICROO[j] +
    beta3*MORA1[j] + beta4*MORA3[j]+ beta5*MORA4[j] + beta6*MORA5[j] +
    beta7*MQLAVAR[j] + beta8*TV[j]+ beta9*AR[j] + beta10*CONDDOM1[j] +
    beta11*CONDDOM2[j]+ beta12*CONDDOM3[j]+ beta13*ANOEST1[j]+
    beta14*ANOEST2[j]+ beta15*ANOEST3[j] + beta16*PAVIMEN[j] +
    beta17*TIPODOM1[j]+ beta18*TIPODOM2[j]+ beta19*TIPODOM3[j] +
    beta20*ADOLEES1[j] + beta21*ADOLEES2[j])

    #criando réplica dos dados
    y[j]~dbern(p[j])

    #criando réplica dos dados para a classificação
    yr[j]~dbern(p[j])

    a [j] <- equals (y[j],yr[j])

    # Proporcao de acerto nas microrregioes da POF
  }

  for(i in 1:19) {
    papof[i]<-sum(a[(n[i]+1):(n[i+1]])] / (n[i+1] - n[i])
  }

  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  espac[1:19] ~ car.normal(adj[], weights[], num[], tau)

  #Prioris para Efeitos Fixos
  beta1 ~ dnorm(0.0,0.001)
  beta2 ~ dnorm(0.0,0.001)
  beta3 ~ dnorm(0.0,0.001)
  beta4 ~ dnorm(0.0,0.001)
  beta5 ~ dnorm(0.0,0.001)
  beta6 ~ dnorm(0.0,0.001)
  beta7 ~ dnorm(0.0,0.001)
  beta8 ~ dnorm(0.0,0.001)
  beta9 ~ dnorm(0.0,0.001)
  beta10 ~ dnorm(0.0,0.001)
  beta11 ~ dnorm(0.0,0.001)
```

```

beta12 ~ dnorm(0.0,0.001)
beta13 ~ dnorm(0.0,0.001)
beta14 ~ dnorm(0.0,0.001)
beta15 ~ dnorm(0.0,0.001)
beta16 ~ dnorm(0.0,0.001)
beta17 ~ dnorm(0.0,0.001)
beta18 ~ dnorm(0.0,0.001)
beta19 ~ dnorm(0.0,0.001)
beta20 ~ dnorm(0.0,0.001)
beta21 ~ dnorm(0.0,0.001)

#Prioris para Efeitos Aleatórios

tau ~ dgamma(0.5, 0.0005)
e0~dflat()

# IMPUTAÇÃO DO CONSUMO NO CENSO

for(i in 1:19) {
  for(k in nc[i] + 1:nc[i+1]) {
    emicroc[k]<-e0
  }
}

for(j in 1:91805) {

logit(pc[j]) <- (emicroc[j] + beta1*COMPUTC[j] + beta2*MICROOC[j] +
beta3*MORA1C[j] + beta4*MORA3C[j]+ beta5*MORA4C[j] + beta6*MORA5C[j] +
beta7*MQLAVARC[j] + beta8*TVC[j]+ beta9*ARC[j] + beta10*CONDDOM1C[j] +
beta11*CONDDOM2C[j]+ beta12*CONDDOM3C[j]+ beta13*ANOEST1C[j]+
beta14*ANOEST2C[j]+ beta15*ANOEST3C[j] + beta16*PAVIMENC[j] +
beta17*TIPODOM1C[j]+ beta18*TIPODOM2C[j]+ beta19*TIPODOM3C[j] +
beta20*ADOLES1C[j] + beta21*ADOLES2C[j])

#criando réplica dos dados da amostra do Censo para o ajuste
yc[j]~dbern(pc[j])

}

for(i in 1:19) {
# Proporção de pobres nas microrregiões do CENSO

pp[i]<-sum(yc[(nc[i]+1):(nc[i+1]])) / (nc[i+1] - nc[i])

}

}

```

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)