

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

Mapas da Pobreza:
Avaliação da metodologia e de sua aplicação para
retratar a pobreza no Rio Grande do Norte

MARIA BEATRIZ ASSUNÇÃO MENDES DA CUNHA

Rio de Janeiro
2007

Livros Grátis

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

Milhares de livros grátis para download.

MARIA BEATRIZ ASSUNÇÃO MENDES DA CUNHA

Mapas da Pobreza:
Avaliação da metodologia e de sua aplicação para
retratar a pobreza no Rio Grande do Norte

Dissertação apresentada como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Estudos Populacionais e Pesquisas Sociais da Escola Nacional de Ciências Estatísticas

Área de concentração: Estatística Social
Orientador: Prof. Dr. Pedro Luis do Nascimento Silva

Rio de Janeiro
2007

MARIA BEATRIZ ASSUNÇÃO MENDES DA CUNHA

Mapas da Pobreza:
Avaliação da metodologia e de sua aplicação para
retratar a pobreza no Rio Grande do Norte

Dissertação apresentada como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Estudos Populacionais e Pesquisas Sociais da Escola Nacional de Ciências Estatísticas

Área de concentração: Estatística Social
Orientador: Prof. Dr. Pedro Luis do Nascimento Silva

Rio de Janeiro, 09 de outubro de 2006.

Banca Examinadora:

Prof. Dr. Pedro Luis do Nascimento Silva
ENCE – Rio de Janeiro

Prof. Dr. Suzana Marta Cavenaghi
ENCE – Rio de Janeiro

Prof. Dr. Fernando Antônio da Silva Moura
UFRJ – Rio de Janeiro

Dedicatória

Dedico este trabalho especialmente aos meus pais José Mendes e Graça, pois eles foram meus primeiros incentivadores, me proporcionando educação para poder iniciar minha formação acadêmica e profissional.

Dedico também ao meu amado marido Herculano, por confiar integralmente na minha capacidade, e me oferecer imenso incentivo, apoio e carinho.

Agradecimentos

Minha caminhada em busca desta dissertação foi um pouco longa, teve muitas alegrias e algumas frustrações, por isso tenho que agradecer a muitas pessoas, mas tentarei ser breve. Desejo muito que todos vocês apreciem esta dissertação tanto quanto eu.

Primeiramente gostaria de agradecer ao meu orientador, a quem devo a mais sincera e efusiva admiração. Agradeço principalmente pela sua compreensão silenciosa nos momentos difíceis pelos quais passei, permitindo que meu tempo interno fluísse, respeitosamente. Pedro, a quem chamo carinhosamente de meu “pai intelectual”, pela sua sabedoria temporariamente compartilhada comigo, pela sua paciência incansável e pelos importantes ensinamentos, meu mais sincero agradecimento!

Sou imensamente grata também ao meu marido Herculano, que por muitas vezes tive que abandoná-lo mesmo em fins de semana de muito calor, pois precisava me dedicar integralmente à dissertação. A sua ajuda, paciência e otimismo foram fundamentais para o sucesso da minha trajetória acadêmica. Herculano, me desculpe se fiz te esperar tanto.

Aos meus pais, que desde pequenina me incentivaram e mostraram a importância dos estudos e das escolhas que fazemos na vida.

A todos os professores do mestrado, em especial a minha grande amiga e conselheira Denise Silva, seus conhecimentos compartilhados ainda na graduação e na fase de elaboração do trabalho final da Graduação, que serão guardados para sempre, foram todos reforçados com os ensinamentos adquiridos no mestrado.

Maurício Vasconcelos, muito obrigada! Sua imensa ajuda na compreensão e tratamento das bases de dados foi imprescindível.

A todos os meus amigos, que mesmo sem perceber me davam força para continuar batalhando. Um agradecimento em particular à minha grande amiga, quase irmã, Alessandra da Rocha Santos, pelo seu afeto e amizade sempre presentes nos principais momentos da minha vida acadêmica. Alê essa dissertação também é sua, pois você viveu todos os momentos dela!

Também gostaria de agradecer a outra grande amiga, a excelente fisioterapeuta Vanessa Moura, pois ela me ofereceu ajuda no momento mais tenso da elaboração da dissertação.

Aos meus colegas do grupo do Mapa da Pobreza do IBGE, especialmente Débora Souza e Viviane Quintaes, que desde o início da definição do tema desta dissertação se propuseram a me ajudar, e sempre tinham algo a me oferecer nos momentos em que estava com dificuldades.

Às minhas sobrinhas Nathália e Heloísa, que me contagiavam com sua constante alegria.

Sou muito a grata à Sala de Uso Público onde comecei a trabalhar há um ano, justamente no momento mais crítico do mestrado – a preparação da dissertação. A todos os usuários da Sala, meu mais sincero “obrigada”, por me proporcionar um maior conhecimento e aprofundamento das técnicas utilizadas na Estatística. É claro que devo muito à Suzana Cavenaghi, pois ela me deu condições de trabalho que possibilitaram a elaboração da minha dissertação.

Enfim, muito obrigada a todos que de alguma forma me ajudaram.

Resumo

A possibilidade de produzir indicadores para pequenas áreas está cada vez mais presente nas discussões acadêmicas. O Brasil é um país muito populoso e com grande extensão territorial e essas características dificultam a produção de indicadores em níveis menores de agregação, como municípios ou microrregiões. Numa tentativa de contornar esse problema, os pesquisadores Elbers, Lanjouw e Lanjouw (2002) do Banco Mundial construíram uma metodologia que combina pesquisas domiciliares com amostras pequenas mas de ampla abrangência temática, com os censos ou grandes amostras onde o detalhamento geográfico é fino, para estimar bem estar ou pobreza para pequenas áreas. Este trabalho visa então avaliar a metodologia criada por Elbers, Lanjouw e Lanjouw, bem como sua aplicação para retratar a pobreza no estado do Rio Grande do Norte. Foram consideradas duas variáveis comumente usadas para retratar a pobreza: renda domiciliar per capita e despesa domiciliar per capita. A avaliação foi feita a partir de comparações dos resultados obtidos através da metodologia proposta com resultados provenientes de outros métodos, tais como a estimação da proporção de pobres diretamente a partir dos dados do Censo Demográfico ou a utilização de modelos multiníveis em lugar dos modelos considerados por Elbers, Lanjouw e Lanjouw (2002). Apesar do grande esforço computacional exigido para sua implementação, a metodologia permite a identificação das microrregiões mais pobres do Rio Grande do Norte, uma vez que a classificação das microrregiões devido à pobreza foi praticamente a mesma nos diversos métodos usados. No entanto, os resultados da intensidade da pobreza nas microrregiões são bastante diferentes para cada metodologia. Os modelos multiníveis se mostraram mais satisfatórios, pois seus resultados se aproximaram mais dos resultados obtidos diretamente com os dados do censo.

Abstract

The possibility to produce information for small areas has been increasing in the academic discussions. Brazil is a very populous country, with vast territorial extension, and these characteristics make the production of information for small levels of aggregation difficult, as cities or micro regions. In an attempt to solve this problem, Elbers, Lanjouw and Lanjouw (2002) developed a methodology for welfare or poverty estimation for small areas that combines household samples with small samples but wide thematic scope with censuses or surveys having larger samples but for which fine geographic detail is feasible. This work evaluates the poverty mapping methodology created by Elbers, Lanjouw and Lanjouw, as it applies this approach to describe the poverty in the state of the Rio Grande do Norte. For the evaluation two variables commonly used to assess poverty have been used: per capita household income and per capita household expenditure. The evaluation is based on comparisons of the estimates obtained using the poverty mapping methodology with results obtained directly from the Demographic Census data, or from other modelling methods such as multilevel modelling. Despite the large computational effort required, the poverty mapping methodology allows the identification of the poorest regions in the state, because the classification of the micro regions into poverty was practically the same for the various methods used. However, the results for the intensity of the poverty in the micro regions are different according to which methodology was used. Multilevel models appeared as the more satisfactory option, because the results obtained for this method are closer to the direct census figures.

Índice

Dedicatória	ii
Agradecimentos	iii
Resumo	v
Abstract	vi
Lista de tabelas	ix
Lista de figuras	x
Lista de Gráficos	xi
1. Introdução	13
1.1. <i>Motivação</i>	13
1.2. <i>Contextualização Histórica</i>	16
1.3. <i>Estrutura da dissertação</i>	20
2. Pobreza	22
2.1. <i>Preliminares</i>	22
2.2. <i>Pobreza absoluta</i>	23
2.3. <i>Pobreza subjetiva</i>	26
2.4. <i>Mensuração de pobreza – Metodologia</i>	29
2.4.1. <i>Indicadores de pobreza</i>	29
2.5. <i>Literatura sobre indicadores de pobreza no IBGE</i>	32
3. Mapas de pobreza para pequenas áreas	36
3.1. <i>Introdução</i>	36
3.2. <i>O modelo de consumo/renda</i>	37
3.3. <i>Preparação dos dados</i>	40
3.4. <i>Ajuste do modelo</i>	43
3.5. <i>Método para seleção de modelos</i>	46

3.6. <i>Imputação da renda/despesa no Censo</i>	49
3.7. <i>O estimador de bem-estar</i>	51
4. Aplicação	52
4.1. <i>Delimitação da área de estudo</i>	52
4.2. <i>Fontes de dados</i>	55
4.2.1. POF	55
4.2.2. Censo 2000	60
4.3. <i>Seleção e ajuste de modelos</i>	64
4.3.1. Variáveis respostas	65
4.3.2. Variáveis preditoras	72
4.3.3. A renda no censo	89
4.3.4. Ajuste de modelos	91
4.4. <i>Avaliação de qualidade do ajuste</i>	96
4.4.1. Validação do modelo para a área rural	96
4.4.2. Avaliação dos modelos	98
4.5. <i>“Mapa” de pobreza – resultados</i>	100
5. Avaliação da Aplicação e da Metodologia	107
5.1. <i>Comparação entre as rendas</i>	107
5.2. <i>Comparação entre as despesa e renda na POF</i>	111
5.3. <i>Mapa de pobreza utilizando modelos multiníveis</i>	117
6. Conclusões e trabalhos futuros	125
7. Referências Bibliográficas	127
8. Anexos	135

Lista de tabelas

Tabela 4.1: População total do Rio Grande do Norte e das Microrregiões	54
Tabela 4.2: Número de Estratos, setores e domicílios pesquisados com entrevista completa no Rio Grande do Norte, por microrregião	58
Tabela 4.3: Número de Estratos, setores e domicílios pesquisados com entrevista completa no Rio Grande do Norte, por microrregião e tipo de área.....	59
Tabela 4.4: Frações de amostragem do Rio Grande do Norte e suas microrregiões .	62
Tabela 4.5: Número de setores e domicílios pesquisados com entrevista completa no Rio Grande do Norte, por microrregião.....	63
Tabela 4.6: Número de setores e domicílios pesquisados com entrevista completa na amostra do Censo 2000 do Rio Grande do Norte, por microrregião	64
Tabela 4.7: Medidas resumo das variáveis respostas	66
Tabela 4.8: Medidas resumo das variáveis respostas transformadas.....	68
Tabela 4.9: Evolução da distribuição das variáveis preditoras, cujo teste de igualdade entre Censo e POF foi rejeitado.....	77
Tabela 4.10: Medidas resumo da renda domiciliar per capita	90
Tabela 4.11: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da renda per capita	92
Tabela 4.12: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da despesa	92
Tabela 4.13: Estimativas dos betas e dos p-valores dos modelos de efeitos principais	94
Tabela 4.14: Teste do efeito das variáveis dos modelos com interações	95
Tabela 4.15: R^2 dos modelos de LNREND e LNDESP por tipo de área	97
Tabela 4.16: R^2 dos modelos para o ln da renda e da despesa	100

Tabela 4.17: Proporção de pessoas em domicílios pobres baseada nas variáveis originais das pesquisas - Rio Grande do Norte e Microrregiões	101
Tabela 4.18: Esquema dos 5 números e indicador de pobreza.....	102
baseados na renda imputada - Rio Grande do Norte e Microrregiões	102
Tabela 4.19: Esquema dos 5 números e indicador de pobreza baseados na despesa imputada - Rio Grande do Norte e Microrregiões	103
Tabela 5.1: Valores das linhas de pobreza hipotéticas baseadas na regressão de LNDESP contra LNREND	116
Tabela 5.2: Indicadores de pobreza baseados na renda e despesas imputadas a partir de modelos multiníveis - Rio Grande do Norte e Microrregiões	118
Tabela 5.3: Ranking da pobreza nas microrregiões do Rio Grande do Norte baseada na renda, segundo a metodologia.....	123
Tabela 5.4: Ranking da pobreza nas microrregiões do Rio Grande do Norte baseada na despesa, segundo a metodologia	124

Lista de figuras

Figura 2.1: Linha de pobreza subjetiva (z^*).....	27
Figura 3.1: Processo de seleção das variáveis preditoras para o ajuste do modelo....	43
Figura 4.1: Mapa do Rio Grande do Norte e microrregiões	55
Figura 4.2: Mapa da pobreza estimado a partir da renda imputada no Censo.	105
Figura 4.3: Mapa da pobreza estimado a partir da despesa imputada no Censo.	105
Figura 5.1: Mapa da pobreza estima a partir da renda imputada no Censo através de modelo multinível – Rio Grande do Norte	119
Figura 5.2: Mapa da pobreza estima a partir da despesa imputada no Censo através de modelo multinível – Rio Grande do Norte	120

Figura 5.3: Mapas de pobreza do Rio Grande do Norte baseados na renda e na despesa imputadas a partir das metodologias adotadas.121

Lista de Gráficos

Gráfico 1.1: Percentual de pessoas residentes de 10 anos ou mais de idade, sem instrução, por Unidade da Federação – Censo Demográfico 2000 19

Gráfico 4.1: Histograma da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados..... 67

Gráfico 4.2: Histograma da despesa total domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados 67

Gráfico 4.3: Histograma do logaritmo da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados 68

Gráfico 4.4: Gráfico Normal do logaritmo da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados..... 69

Gráfico 4.5: Histograma do logaritmo da Despesa total per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados 70

Gráfico 4.6: Gráfico Normal do logaritmo da despesa total per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados..... 70

Gráfico 4.7: Densidade do logaritmo da renda domiciliar per capita - Rio Grande do Norte..... 71

Gráfico 4.8: Densidade do logaritmo da despesa domiciliar per capita - Rio Grande do Norte..... 72

Gráfico 4.9: Resíduos do modelo de regressão para a renda ajustado com os dados da Pesquisa de Orçamentos Familiares – Rio Grande do Norte..... 98

Gráfico 4.10: Resíduos do modelo de regressão para a despesa ajustado com os dados da Pesquisa de Orçamentos Familiares – Rio Grande do Norte	99
Gráfico 5.2: Dispersão da LRDPC contra LREND_EST.....	111
Gráfico 5.3: Dispersão da DESPPC contra RDPC	112
Gráfico 5.4: Dispersão da LNDESP contra LNREND	114
Gráfico 5.5: Relação entre as duas variáveis, de acordo com a equação de regressão	116

1. Introdução

1.1. *Motivação*

O Brasil é um país de muitas facetas. Um território tão grande que abriga diversas culturas, comportamentos e religiões. Foi devido à necessidade de conhecer especificidades das muitas partes da população que pesquisadores estão buscando formas de retratar territórios por vezes não retratados pelas pesquisas mais abrangentes existentes no País.

A pesquisa domiciliar com maior detalhamento geográfico do Brasil é o Censo Demográfico, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Porém este só ocorre a cada 10 anos, e aborda um conjunto limitado de temas. A Pesquisa Nacional por Amostra de Domicílios (PNAD), também do IBGE, é mais abrangente no que se refere ao conjunto de temas pesquisados. Além disso, a PNAD é realizada todo ano, exceto quando é feito o Censo Demográfico. No entanto, a amostra da PNAD foi desenhada para representar as Unidades da Federação, e isso dificulta a realização de estudos voltados para retratar áreas menores, como por exemplo, municípios ou microrregiões.

O IBGE também produziu outras pesquisas importantes, que se propunham a mensurar as estruturas dos gastos e dos rendimentos das famílias. O ENDEF (Estudo Nacional de Despesa Familiar) ocorreu em 1974/75 e tinha âmbito territorial nacional, à exceção da área rural da Região Norte. Dez anos depois da realização do ENDEF, o IBGE passou a desenvolver a Pesquisa de Orçamentos Familiares (POF), que “tem

por objetivo fornecer informações sobre a composição orçamentária doméstica, a partir da investigação dos hábitos de consumo, da alocação de gastos e da distribuição dos rendimentos, segundo as características dos domicílios e das pessoas, bem como sobre a percepção das condições de vida da população brasileira”¹. Trata-se de uma pesquisa riquíssima, mas que não tem periodicidade fixa e foi realizada nos períodos 1987/88, 1995/96 e 2002/2003. As POFs de 1987/88 e 1995/96 foram desenvolvidas para atualizar a estrutura de consumo do índice de preços ao consumidor, abrangendo o perímetro urbano das regiões metropolitanas de Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre, além do Distrito Federal e o município de Goiânia. Só em 2002/03 é que a POF sofreu uma ampliação e sua abrangência passou a ser nacional, cobrindo todos os Estados, inclusive as áreas rurais da Região Norte.

O fato de no Brasil não existir uma pesquisa que combine ampla abrangência temática com detalhamento geográfico fino das estimativas, é que tornou necessário criar mecanismos para suprir a ausência destas particularidades nas pesquisas brasileiras. Elbers, Lanjouw e Lanjouw (2002) desenvolveram uma metodologia que permite estimar indicadores de bem-estar ou de pobreza para pequenas áreas tomando por base informações detalhadas sobre padrão de vida, disponível apenas em pesquisas domiciliares com pequenas amostras, combinadas com informações estruturais dos domicílios disponíveis nos censos ou grandes amostras. Combinando as vantagens de cada um destes tipos de fonte de dados, os estimadores obtidos a partir dessa nova metodologia podem ser usados em níveis geográficos mais desagregados que os permitidos pelas pesquisas com amostras pequenas, se usadas isoladamente. O capítulo 3 apresenta em detalhe essa nova metodologia.

¹ POF 2002/2003: primeiros resultados, p. 17.

Com o conhecimento de novas técnicas, é possível realizar estudos sobre um tema de grande relevo no cenário Brasileiro – a pobreza. No passado, o tema mais abordado nas discussões era o desenvolvimento econômico. A pobreza era vista apenas como uma conseqüência dos problemas gerados pela má administração da riqueza nacional. Hoje, a questão da pobreza aparece em primeiro plano, principalmente por requerer urgência na solução do problema.

No panorama político, a discussão da pobreza se mostra mais forte à medida que o governo federal volta seus olhares para uma parcela da população que não encontra recursos nem para comer. Esta é a pior forma de pobreza, conhecida pelos pesquisadores sociais como indigência. É difícil imaginar que a pobreza no Brasil tenha aumentado, enquanto houve crescimento econômico.

A metodologia proposta por Elbers, Lanjouw e Lanjouw (2002) permite pensar num mapeamento da pobreza no Brasil, a partir do cálculo de indicadores de pobreza para pequenas áreas, tais como municípios ou microrregiões, habitualmente não retratadas por pesquisas já desenvolvidas.

Já existem muitos estudos sobre a pobreza e a desigualdade social no Brasil como um todo, mas poucos estudos tratam as unidades da federação separadamente. As características culturais e alimentares diferem nas diversas regiões do país. Por outro lado, o esforço requerido para realizar tal trabalho para todos os estados do Brasil seria enorme. Então optamos por escolher uma única Unidade da Federação, e realizar o estudo detalhadamente para esta. Como a produção acadêmica a respeito do Estado do Rio Grande do Norte é pequena, e levando em conta a naturalidade da autora desta dissertação, resolveu-se focar o estudo sobre este estado que é muito conhecido por suas belas praias, mas pouco por suas características como educação, cultura, desigualdade social e pobreza.

1.2. Contextualização Histórica

O Programa das Nações unidas para o Desenvolvimento (PNUD) considera que a pobreza não deve ser abordada sem introduzir considerações culturais², tornando importante uma rápida menção da história e cultura potiguares. A cultura do Rio Grande do Norte começou a se estabelecer desde sua colonização, que aconteceu de forma atípica e turbulenta. Poucos sabem, mas esta área do nordeste brasileiro já foi disputada por Holandeses, por Franceses e, evidentemente por Portugueses. Os povos que os Portugueses encontraram no litoral da capitania do Rio Grande eram os potiguares³:

“de porte mediano, acima de 1,65cm, reforçados e bem feitos no físico. Olhos pequenos e amendoados como os da raça mongólica, escuros e encovados, de orelhas grandes, cabelos lisos e cortados redondos, arrancavam os pêlos da barba até as pestanas e sobrancelhas. Eram baços, claros, pintavam seus corpos com desenhos coloridos. Furavam o beijo, principalmente o inferior, assim como orelhas e o nariz”.⁴

A primeira grande contribuição para a cultura Norte-rio-grandense foi a agricultura. Ao chegarem ali os portugueses encontraram uma tribo indígena que dava seus primeiros passos na revolução agrícola, cultivando mandioca, milho, batata-doce, abóbora, algodão, tabaco e algumas árvores frutíferas. A rede servia para o

² Relatório do Desenvolvimento Humano (RDH): 2004, p. 13.

³ O interior da então capitania era habitado por índios Tapuias.

⁴ SUASSUNA & MARIZ: 1997, p. 51.

descanso e a canoa para locomoção e pesca, artefatos que ainda estão presentes na vida do habitante atual.

Durante todo o século XVI a Capitania do Rio Grande foi explorada por franceses, que traficavam o pau-brasil para a Europa. Os franceses se estabeleceram ali de forma pacífica e até firmaram amizade com os potiguares, e isso se deve ao tipo de interesse econômico que eles tinham com a terra. Seu único objetivo era tirar proveito da exploração e do comércio, ao contrário dos portugueses que queriam se fixar na terra, mudar costumes, hábitos e crenças, impondo uma nova lei e ordem aos indígenas. Esse foi o principal motivo da hostilidade dos potiguares para com os portugueses, o que tornou o domínio do território uma tarefa muito difícil, só se concretizando no final do século XVI, quando os franceses já haviam abandonado a terra devastada pela exploração.

Os holandeses também tiveram uma passagem bastante polêmica por terras potiguares. Após sair de Salvador, uma esquadra holandesa atracou na costa do Rio Grande do Norte devido a uma epidemia de escorbuto em sua tripulação, se atrasando em relação ao restante da frota. Dois meses depois, eles retornam à Holanda levando consigo vários índios potiguares, que voltariam falando holandês e conhecendo a bíblia, e que serviriam de peça chave quando do domínio holandês no Rio Grande do Norte.

Esta breve contextualização histórica serve para destacar que o estado do Rio Grande do Norte, mesmo pequeno, concentra diversas culturas, que podem ser explicadas pela influência das diferentes formas de colonização ali experimentadas. Além disso se há diversidade de culturas, há também diversidade econômica, pois desde a época da colonização já havia a figura/presença do pobre e do rico, aqui tipificados por índios e europeus.

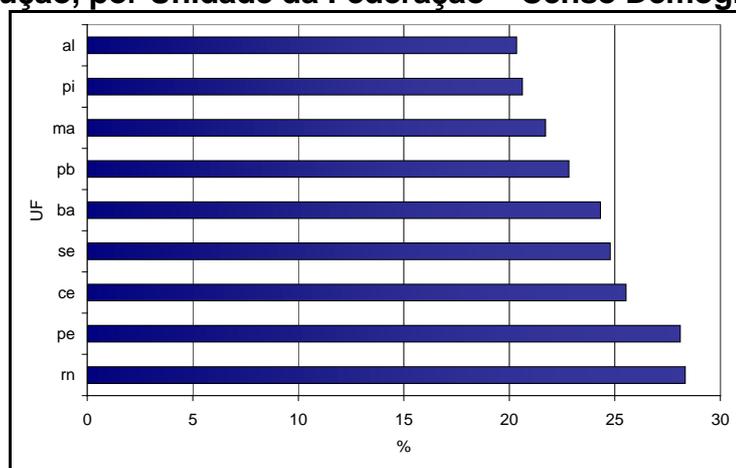
Hoje em dia, quando se fala em Rio Grande do Norte, lembra-se logo do seu lindo litoral, com belas paisagens e calor quase que constante. A idéia que se faz desse estado é resultado do forte crescimento do turismo ali observado, que pode até ter contribuído para o seu crescimento econômico, mas não influenciou as camadas mais pobres da população, que ainda sofrem com a seca, e conseqüentemente com a falta de água e comida.

O Rio Grande do Norte é o estado da Região Nordeste que tem limites a norte e a leste com o Oceano Atlântico, a oeste com o Ceará e ao sul com a Paraíba. Diz-se que este estado está localizado na “esquina do Brasil”. Isso porque o Rio Grande do Norte é o estado com maior projeção para o Atlântico, já que se situa em uma região onde o litoral brasileiro faz um ângulo agudo. Essa curvatura faz com que o estado possua o ponto brasileiro mais próximo da Europa e da África, fazendo do Rio Grande do Norte um ponto estratégico durante a II guerra mundial, época do governo de Getúlio Vargas. Naquela época os americanos instalaram ali uma base aérea muito importante para o sucesso do desembarque na Normandia. Por esse motivo tal base foi apelidada de Trampolim da Vitória.

De acordo com o Censo Demográfico 2000, a população Norte-rio-grandense era de 2.777.509 habitantes, distribuída nos seus 166 municípios. As cidades do Rio Grande do Norte ainda mantêm ares de cidades pequenas. Como exemplo disso, em 2000, as casas representavam quase totalidade dos domicílios particulares permanentes (cerca de 97%). Mesmo em Natal, capital do estado, tinha-se cerca de 91% de casas, contra apenas 8% de apartamentos. Por isso, ainda é possível ver crianças brincando nas ruas e não em condomínios fechados, muito comuns em grandes cidades como Rio de Janeiro e São Paulo.

Comparativamente a outros estados nordestinos, o Rio Grande do Norte apresenta boas condições de educação, sendo o estado com maior porcentagem de pessoas com mais de 10 anos de idade com ensino fundamental completo – 28%. Por outro lado, cerca de 13% de pessoas com mais de 10 anos de idade no Rio Grande do Norte não tinham instrução.

Gráfico 1.1: Percentual de pessoas residentes de 10 anos ou mais de idade, sem instrução, por Unidade da Federação – Censo Demográfico 2000



Por outro lado, na questão de infra-estrutura o Estado ainda tem muito a melhorar. Apenas 78% dos domicílios são abastecidos de água a partir de rede geral de distribuição. Mesmo na área urbana o abastecimento de água ainda não atingiu totalidade dos domicílios. E na área rural apenas 36% dos domicílios têm o luxo do acesso a água pela rede geral. De forma que ainda é muito comum presenciar a cena clássica de famílias carregando latas d'água. Se as condições do abastecimento de água no Rio Grande do Norte não são boas, as de esgotamento sanitário são piores ainda. Apenas 40% dos domicílios têm acesso a um escoamento sanitário adequado.

Muitas outras informações poderiam ser relevantes para um melhor conhecimento de um estado que tem múltiplas facetas, como conviver com condições precárias de infra-estrutura, mas manter uma educação superior à dos seus vizinhos.

O presente tópico foi uma tentativa de apresentar um pouco de história e atualidade do Rio Grande do Norte aos que não o conhecem, mas de forma breve para não fugir do tema central da dissertação.

1.3. Estrutura da dissertação

Nessa dissertação, procurou-se fazer uma aplicação da metodologia elaborada por Elbers, Lanjouw e Lanjouw (2002) para construir de um mapa de pobreza para o estado do Rio Grande do Norte. Além disso, faz-se uma avaliação da aplicação da metodologia, bem como dos resultados obtidos a partir dela.

No capítulo dois são abordadas algumas formas de análise da pobreza, e são expostos alguns indicadores comumente utilizados para realizar tais análises. Além disso, apresenta-se um relato da literatura recente do IBGE sobre os indicadores de pobreza.

No terceiro capítulo, é apresentada a metodologia proposta por Elbers, Lanjouw e Lanjouw (2002), explicitando todas as etapas necessárias para mensurar pobreza em pequenas áreas.

No quarto capítulo encontra-se a aplicação da metodologia para o Rio Grande do Norte. Também são apresentados a delimitação da área que será trabalhada, as fontes de dados utilizadas e o ajuste dos modelos. É neste capítulo que será mostrada uma avaliação da qualidade do ajuste do modelo adotado, e serão expostos os resultados do Mapa de Pobreza.

A avaliação da aplicação e da metodologia é apresentada no capítulo cinco. Neste capítulo serão discutidas as dificuldades de implementação da metodologia proposta para o caso de se trabalhar com dados da POF e do CENSO. Além disso é

realizada uma comparação dos resultados obtidos a partir da metodologia proposta com os outros de diferentes métodos.

Por fim, no sexto capítulo serão apresentadas as conclusões e as propostas para trabalhos futuros.

2. Pobreza

2.1. Preliminares

Não existe um consenso sobre o significado do termo pobreza, mas segundo o Programa das Nações Unidas para o Desenvolvimento (PNUD), “pobreza humana é a negação das escolhas e oportunidades básicas para o desenvolvimento humano refletida em vida curta, falta de educação elementar, falta de meios materiais, exclusão e falta de liberdade e dignidade.”⁵

A elaboração de medidas de pobreza geralmente parte do pressuposto que existem padrões de consumo predeterminados ou bem definidos, e isso permite a produção de linhas de pobreza, que vem a ser o nível mínimo de renda que seria necessária para garantir uma certa condição de subsistência para os indivíduos. Existem muitas formas de obter linhas de pobreza, e a maioria delas são realizadas de forma arbitrária ou por aproximações. Na maioria das sociedades, a noção da pobreza vai além da obtenção do mínimo necessário para a sobrevivência do indivíduo.

A forma de determinar a pobreza depende da disponibilidade de dados e da finalidade do estudo. A mais utilizada é sem dúvida a proporção de pobres calculada a partir da renda dos indivíduos. No entanto a renda de uma pessoa é uma medida muito imperfeita do seu bem-estar. Duas pessoas que têm rendas idênticas podem não ter o mesmo nível de satisfação. Por exemplo, uma dessas pessoas pode ter despesas com medicamentos que chegam a 1/3 da sua renda, enquanto a outra é uma pessoa saudável e não tem gastos com medicamentos. Ambos poderiam ser

⁵ Glossário da pobreza e desenvolvimento humano, RDH 1997, PNUD

tratados como pobres, ocasionando um erro de interpretação muito comum em pesquisas sociais que levam em conta apenas as rendas auferidas pelos indivíduos.

Além disso, a análise da distribuição de renda no Brasil é sempre baseada em dados advindos de questionários, e as respostas dadas pelos entrevistados podem não refletir a realidade, gerando erros nas conclusões. Estudos mostram que o grau de sub-declaração tende a crescer com o valor das rendas, principalmente quando o tipo de renda auferida não é de ordem monetária.⁶

Pensando no leque de opções para o estudo da pobreza, o Ministério de Desenvolvimento Social (MDS) e a Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) estabeleceram acordo de cooperação técnica para definir um conceito oficial de pobreza, que servirá de base para próximas ações de programas sociais do governo federal.⁷

Nas seções seguintes são apresentadas algumas formas de obtenção de indicadores de pobreza.

2.2. Pobreza absoluta

Muitas vezes o conceito de pobreza se confunde com o de desigualdade, à medida que a definição de “pobreza” depende de uma separação da população entre pobres e não pobres. Por isso, muitas vezes é usado o conceito de pobreza absoluta, onde o grau de pobreza não depende do nível de renda dos ricos. Para tanto, introduz-se o conceito de *linha de pobreza*. São consideradas pobres as pessoas (ou famílias) que tiverem renda inferior a esta linha. Como não há critério claro para

⁶ Hoffmann (1998, p. 214)

⁷ www.desigualdade.inf.br

estabelecer a linha de pobreza, a escolha desse valor é feita de forma muito arbitrária, explicando assim a existência de tantas linhas de pobreza no Brasil.

Um procedimento muito usado para a determinação de linhas de pobreza consiste em obter uma cesta básica de alimentos que atenda às necessidades nutricionais mínimas das famílias. Este procedimento leva ao problema de falta de informações sobre o consumo das pessoas. Segundo Rocha (2000a)⁸ “havendo disponibilidade de informações sobre a estrutura de consumo das famílias, esta é a fonte mais adequada para o estabelecimento de linhas de pobreza”. Por sorte a POF foi realizada há pouco tempo, e com seus dados é possível realizar diversos estudos a respeito das despesas das famílias, e conseqüentemente, da pobreza das famílias e pessoas. O consumo aqui considerado trata-se da quantidade de itens obtidos pela família, no entanto a POF não disponibiliza a informação de membros não residentes dividindo os mesmos itens adquiridos, como por exemplo num jantar com convidados. Na falta de informações sobre o consumo, a despesa pode ser utilizada como uma aproximação.

Segundo Rocha (2000b)⁹, “vale destacar que o resultado em termos de atendimento médio das necessidades nutricionais, resultante do confronto entre consumo agregado versus necessidades agregadas da população, é totalmente desprovido de sentido para fins de desenho e avaliação de políticas sociais, já que negligencia o aspecto distributivo do consumo alimentar.”

Rocha (2000) também afirma que vem sendo observado que, com a urbanização e a melhoria dos meios de comunicação, as escolhas alimentares das famílias têm se tornado cada vez piores quanto ao aporte nutricional e o preço dos alimentos. Desta forma o valor dos produtos normalmente consumidos aumentou mais

⁸ Rocha (2000a, p.109)

⁹ Rocha (2000b, p. 15)

que o índice de preços, devido a mudanças no padrão de consumo, mesmo para as famílias mais pobres. A população tem buscado alimentos mais caros e pouco eficientes no tocante à satisfação das necessidades nutricionais.

Como houve modificações na escolha dos alimentos, e valor da cesta aumentou muito, por isso faz sentido estabelecer a linha de pobreza a partir de estruturas de consumo, cujo valor estaria associado ao mínimo necessário para que uma pessoa funcione na sociedade.

No Brasil utiliza-se apenas o parâmetro das necessidades calóricas em vez de considerar toda a gama de necessidades nutricionais, que incluem necessidades de proteínas, vitaminas e minerais. Porém, estudos já realizados com o ENDEF (Estudo Nacional de Despesa Familiar) e mesmo a POF mostram que é suficiente escolher uma dieta que garanta os requisitos calóricos para que esta atenda também às demais necessidades nutricionais.¹⁰

Depois de obter as necessidades calóricas médias, deve-se obter a cesta alimentar observada de menor custo que permita o atendimento destas necessidades. É preciso derivar o aporte calórico per capita da cesta alimentar consumida em cada família. O passo seguinte é ordenar as famílias em função do seu consumo calórico per capita, na busca de identificar o intervalo de despesa alimentar mais baixo para o qual a necessidade calórica é atendida. Por exemplo, se a análise foi feita por décimos de distribuição, e as pessoas no 4º décimo consomem apenas o mínimo necessário, então o custo da cesta alimentar do 4º décimo será interpretado como a linha de pobreza em questão. É claro que é necessário fazer ajustamentos na quantidade de alimentos para fornecer a quantidade de calorias predeterminada.

¹⁰ Rocha (2000, p.111)

A adoção de cestas alimentares, que correspondam à ingestão recomendada de calorias, normalmente leva a um cálculo de proporção de indigência muito elevada. Por esta razão muitos pesquisadores adotam linhas de pobreza ainda mais arbitrárias, que não serão detalhadas nesta dissertação, por estar se afastando do objetivo proposto.

É sabido que as necessidades básicas do cidadão vão além da alimentação. O consumo de bens e serviços não alimentares é tão importante quanto o consumo de alimentos. Não existem normas para definir o consumo adequado de itens como vestuário, saúde, transporte, etc, ao contrário do que ocorre com o consumo alimentar. Ainda segundo Rocha (2000)¹¹, no Brasil o consumo alimentar não ultrapassa os 50% do total de consumo. Para entender esta relação do consumo, o coeficiente de Engel¹² foi consagrado como elemento central na determinação da linha de pobreza, existem muitos procedimentos para estabelecer a despesa não alimentar.

2.3. Pobreza subjetiva

Esta abordagem considera que definições de pobreza são inerentes a julgamentos subjetivos que as pessoas fazem sobre o padrão de vida mínimo aceitável numa determinada sociedade.

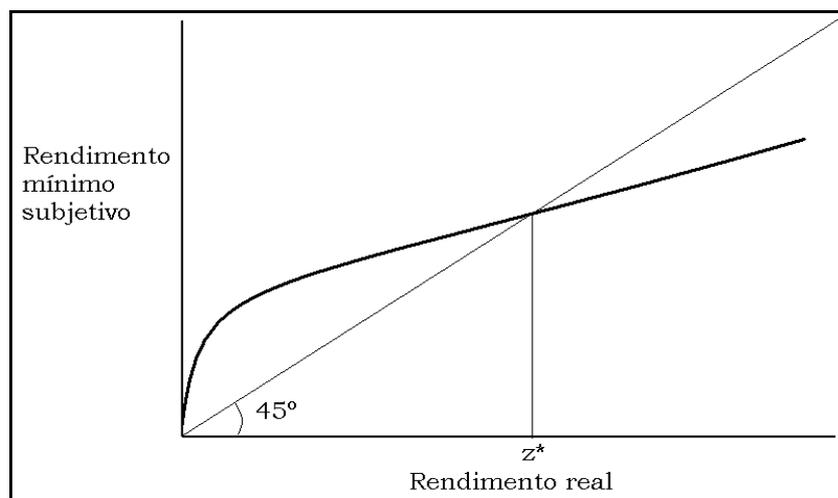
Pobreza subjetiva se deriva da opinião sobre qual deveria ser o padrão do nível de pobreza para todos, e se baseia em respostas à pergunta sobre rendimento mínimo (PRM), como definido por Kapteyn et al, 1988: “Qual é o nível de renda que você considera como absolutamente mínimo, tal que com menos que isso você não

¹¹ Rocha (2000, p.117)

¹² Coeficiente de Engel é a relação entre despesas alimentares e despesa total.

poderia chegar ao final do mês?”. Seria possível definir como pobre qualquer pessoa cujo rendimento fosse inferior ao montante que dera como resposta a essa pergunta. Entretanto, isso pode levar a medidas de pobreza inconsistentes, pois pessoas com a mesma renda, ou qualquer outra medida, seriam tratadas de forma diferente. É claro que pode acontecer diversidade nas respostas à PRM de pessoas com mesmo nível de bem-estar, mas elas têm que ser consideradas igualmente pobres para que se tenha uma medida consistente. Segundo Pradhan e Ravallion (2000), trabalhos anteriores mostram que o valor esperado da resposta a PRM tende a crescer com a renda, e essa relação está simbolicamente representada na figura 2.1.

Figura 2.1: Linha de pobreza subjetiva (z^*)



Fonte: Ravallion (1992, p.34)

A linha escura representa uma função de regressão meramente ilustrativa da renda com a resposta à PRM. O ponto z^* na figura é um óbvio candidato à linha de pobreza, uma vez que pessoas com renda real acima de z^* tendem a achar que seus rendimentos são adequados enquanto que as pessoas com renda real abaixo de z^* acham o contrário disso.

A POF 2002/03 traz o Questionário de Condições de Vida, que possibilita a realização de análise da pobreza através de linhas de pobreza subjetivas. As perguntas 1, 2 e 3 do questionário, dispostas a seguir, servem como exemplo da forma como são abordadas as questões sobre pobreza subjetiva.

- Pergunta 1:
 - Na sua opinião a renda total de sua família permite que você(s) leve(m) a vida até o fim do mês com:
 1. Muita dificuldade
 2. dificuldade
 3. alguma dificuldade
 4. alguma facilidade
 5. facilidade muita facilidade

- Pergunta 2:
 - Levando em conta a situação atual de sua família, qual seria a renda mensal mínima necessária para chegar até o final do mês?

- Pergunta 3:
 - Levando em conta a situação atual de sua família, qual seria o valor mínimo mensal de recursos para cobrir os gastos com alimentação?

Os entrevistados têm diferentes conceitos sobre renda, que nem sempre correspondem ao conceito usado pelo pesquisador. Alguns entrevistados sentem que sua renda é menor do que realmente é. Outros, por outro lado, tendem a superestimar sua renda. Talvez por causa disso, Pradhan e Ravallion (2000, p.463) não conheçam

nenhuma tentativa de aplicação do método de avaliação de pobreza subjetiva em países onde há esse tipo de problema, geralmente países em desenvolvimento.

2.4. Mensuração de pobreza – Metodologia

O estado de “pobreza” existe quando uma ou mais pessoas de uma dada sociedade não alcançam o nível de bem-estar material considerado como mínimo (linha de pobreza) para os padrões da sociedade. Partindo desta definição do termo “pobreza”, vários autores construíram indicadores para medir vários aspectos do problema.

2.4.1. Indicadores de pobreza

A Proporção de Pobres é um indicador muito utilizado, pois se trata de um indicador simples e, por isso, fácil de compreender. Dada uma população com N pessoas, das quais P são pobres, então a proporção de pobres é representada por H , ou seja:

$$H = \frac{P}{N} \quad (2.1)$$

A linha de pobreza é representada por z . Os pobres podem ser identificados através da sua renda per capita (x_i), por exemplo. Então as pessoas são ordenadas segundo suas rendas, e serão consideradas pobres aquelas que tiverem rendas inferiores z .

A proporção de pobres capta apenas a extensão ou a incidência da pobreza, não sendo capaz de informar nada sobre a intensidade da pobreza. Para isso introduz-se outro indicador: Insuficiência de Renda (S). A insuficiência de renda,

também conhecida como hiato de pobreza, é o montante que falta para que a renda de determinado indivíduo pobre alcance a linha de pobreza, e pode ser usada para indicar o quão pobres são os pobres. Assim, a insuficiência total de renda para todos os pobres é dada por:

$$S = \sum_{i=1}^P (z - x_i). \quad (2.2)$$

Este indicador é muito usado por pesquisadores em simulações para estimar quanta renda seria necessário transferir aos pobres para acabar com a pobreza, sendo muito útil no cálculo de custos de programas sociais.

O valor máximo da insuficiência de renda ocorre quando todos os pobres têm renda nula, e nesse caso, $S = p * z$. A razão entre a insuficiência de renda dos pobres e seu valor máximo possível dá origem a outro indicador, que é chamado de razão de insuficiência de renda (I).

$$I = \frac{S}{P * z} = \frac{1}{P} \sum_{i=1}^P \frac{(z - x_i)}{z}. \quad (2.3)$$

Se a média da renda dos pobres é dada por $M = \frac{1}{P} \sum_{i=1}^P x_i$, então verifica-se que

$$I = 1 - \frac{M}{z}. \text{ Isso mostra que, enquanto } H \text{ é insensível à intensidade da pobreza de}$$

cada pessoa, I é insensível ao número de pobres. Desta forma H e I são indicadores complementares.

Foram também propostas medidas de pobreza que levam em consideração tanto a incidência quanto a intensidade da pobreza. É o caso do índice de Sen (Sen, 1976), que é definido da seguinte forma:

$$SEN = \frac{2}{(P+1)Nz} \sum_{i=1}^P (z - x_i) \gamma(X; z), \quad (2.4)$$

onde $\gamma(X; z)$ é uma função que indica a posição de cada indivíduo abaixo da linha de pobreza na escala de renda.

Como em muitos casos, a pobreza está altamente relacionada com a desigualdade de renda, o índice de SEN também pode se relacionar a uma medida de desigualdade da seguinte forma:

$$SEN = H[I + (1 - I)G_p], \quad (2.5)$$

onde G_p é o índice de Gini para a distribuição de renda dos pobres.

Uma importante família de indicadores foi introduzida por Foster, Greer e Thorbecke (1984). Os indicadores são calculados a partir de uma fórmula geral dada por:

$$FGT(\alpha) = \frac{1}{Nz^\alpha} \sum_{i=1}^P (z - x_i)^\alpha \text{ com } \alpha \geq 0 \quad (2.6)$$

Os indicadores são facilmente derivados. Quando $\alpha = 0$, $FGT(0) = H$ é igual a proporção de pobres, e quando $\alpha = 1$, obtém-se o hiato de pobreza, isto é $FGT(1) = H * I$.

Quando $\alpha = 2$, obtém-se um indicador que é chamado de índice sintético de pobreza, e que merece destaque, pois é uma medida de “severidade” da pobreza. O indicador é dado pela expressão:

$$FGT(2) = \frac{1}{Nz^2} \sum_{i=1}^P (z - x_i)^2. \quad (2.7)$$

Esta medida está relacionada ao coeficiente de variação da renda dos pobres.

$$FGT = H \left[I^2 + (1 - I)^2 CV^2 \right] \quad (2.8)$$

onde CV é o coeficiente de variação da renda dos pobres.

Nesta dissertação apenas o indicador de proporção de pobres (equação 2.1) foi considerado.

2.5. Literatura sobre indicadores de pobreza no IBGE

O IBGE realizou desde 2000 uma série de estudos, com o objetivo de apoiar a seleção de metodologia para análise da desigualdade e da pobreza, apresentando os principais indicadores utilizados, assim como experimentando a aplicação desses indicadores com dados da PNAD do ano de 1999. Estes estudos são parte de um projeto do IBGE em parceria com o PNUD.

Ao todo foram produzidos seis relatórios, abordando os temas de desigualdade e pobreza. No primeiro relatório (Abramovitz, 2000) encontra-se uma breve introdução dos conceitos de desigualdade de renda, uma pequena análise da dimensão da desigualdade no Brasil, nas últimas quatro décadas, comparando com outros países. O texto também apresenta uma breve introdução sobre a Curva de Lorenz, e sobre técnicas de decomposição para os índices de Theil.

Na parte sobre pobreza do mesmo relatório (Da costa, 2001a), faz-se uma breve introdução das definições e conceitos de pobreza, aborda questões associadas à mensuração da pobreza, como medida de bem-estar individual, linhas de pobreza e insuficiência de renda. Além disso, o texto mostra a situação da incidência da pobreza no mundo, e em particular na América Latina e no Brasil. O texto traz também 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 se aproximado por uma amostra aleatória simples, visando tornar mais simples os cálculos dos índices selecionados.

O segundo relatório (Medici, 2001a) especifica os temas e variáveis relevantes para a análise e mensuração da desigualdade de renda, bem como apresenta um

resumo da metodologia definida para a estimação de medidas de pobreza. Numa segunda parte (Da Costa, 2001b) há uma breve introdução aos motivos pelos quais se deve estimar a variância dos estimadores estudados, os fatores que influenciam sua precisão, e as dificuldades de estimação. Também aborda aspectos metodológicos da PNAD, revisando seus objetivos, população-alvo, abrangência geográfica e desenho amostral.

O terceiro produto da série de relatórios (Medici, 2001b) tem como objetivo descrever a geração de bases de dados selecionados das pesquisas domiciliares do IBGE referidas ao tema de desigualdade de renda para a década de 90.

O quarto relatório (Medici, 2001c) apresenta a metodologia de decomposição de medidas de desigualdade e pobreza, assim como os indicadores selecionados para uma aplicação baseada na PNAD para os anos de 1997 a 1999. Na parte sobre a desigualdade, apresenta-se o método de decomposição para a Classe dos Índices de Entropia Generalizada. O estudo está interessado em desagregar as medidas de desigualdade baseadas na renda para alguns subgrupos da população, especificando o índice que melhor se adapta aos objetivos do trabalho. A parte sobre pobreza deste relatório traz uma revisão da metodologia usada na mensuração de pobreza, apresenta-se a metodologia para as decomposições das medidas, abordando condições e propriedades que uma medida deve satisfazer para que ela possa ser decomposta, o que justifica o uso dos indicadores FGT. Esse texto traz também uma aplicação da metodologia apresentada, a partir de características dos chefes de família – como sexo, cor/raça, idade, nível de escolaridade, tipo de família, região do domicílio, atividade e ocupação – com base nos dados da PNAD para os anos de 1997 a 1999.

O quinto relatório (Medici, 2001d) traz apenas uma exemplificação do método de decomposição apresentado no relatório anterior sobre a desigualdade de renda. A aplicação da parte de pobreza foi desenvolvida ainda no quarto relatório.

O sexto e último relatório (Medici, 2001e) apresenta um resumo da metodologia definida para a estimação de medidas de desigualdade e de pobreza e a teoria de escala de equivalência. Na parte referente à desigualdade, há uma breve revisão das medidas de desigualdade, novamente são apresentadas as técnicas de decomposição para a classe dos índices da entropia generalizada, bem como uma descrição dos procedimentos de estimação para calcular de forma mais precisa os índices em questão. Neste relatório, os autores acharam que era necessário fazer adaptações nos procedimentos de estimação, devido à complexidade do plano amostral da PNAD e ao fato das unidades na amostra de novas construções terem sido incluídas. Por fim apresenta-se uma aplicação da metodologia para os dados do Brasil de 1992 a 1999. Já na seção sobre pobreza, o relatório exhibe o conceito de escalas de equivalência, que é uma ferramenta que visa tratar a renda para que ela seja um indicador melhor do bem-estar econômico, traz uma revisão do conceito das principais medidas de pobreza, bem como os principais axiomas que essas medidas devem satisfazer. O texto também apresenta as técnicas de decomposição para a família dos índices FGT e expõe as aplicações que foram feitas para o Brasil para os anos de 1992 a 1999, para exemplificar a metodologia.

Está claro que o IBGE tem a preocupação de realizar estimações de desigualdade e pobreza para o Brasil, e até mesmo para áreas “menores”, como os estados. Mas as pesquisas realizadas pelo instituto ainda não comportam a análise da pobreza e da desigualdade em áreas muito pequenas, como municípios ou microrregiões. Isso ocorre porque, apesar de existirem pesquisas elaboradas para

fornecer dados municipais, essas pesquisas não incorporam perguntas que possibilitam a realização de tais análises. Por outro lado, as pesquisas com maior cobertura temática, não são desenhadas para representar essas pequenas áreas.

Faz-se necessária, então, a utilização de uma metodologia que combine a cobertura geográfica com a abrangência temática das diversas pesquisas produzidas pelo IBGE. Esse tipo de análise, ainda inédito no Brasil, pode ser implementado a partir da metodologia proposta por Elbers, Lanjouw e Lanjouw (2002), cujo detalhamento se encontra no próximo capítulo.

3. Mapas de pobreza para pequenas áreas

3.1. Introdução

Nos últimos anos cresceu a demanda por informações em pequenas áreas, o problema enfrentado é que as pesquisas domiciliares que incluem medidas razoáveis de renda ou consumo podem ser usadas para calcular medidas apenas para grandes áreas, mas para níveis menores de agregação, suas amostras raramente são representativas ou têm tamanho suficiente para fornecer estimativas estatisticamente confiáveis. Ao mesmo tempo, censos (ou grandes amostras), que têm tamanho suficiente para permitir tal desagregação, não trazem informações sobre renda ou consumo, ou abordam essas variáveis de maneira precária.

Elbers, Lanjouw e Lanjouw (2002) introduziram uma nova metodologia que permite construir indicadores de bem-estar para pequenas áreas, combinando informações detalhadas sobre padrões de vida obtidas em pesquisas domiciliares geralmente realizadas com pequenas amostras, com informações do Censo ou de grandes amostras, cuja cobertura é abrangente mas cujo detalhamento temático é pequeno ou até inexistente. Segundo esses autores, combinando o poder dos dois tipos de pesquisa, os estimadores calculados podem ser usados em níveis mais desagregados do que os comumente considerados.

Com medidas de bem-estar precisas para grupos do tamanho de cidades, vilas e até vizinhanças, os pesquisadores são capazes de testar hipóteses num nível de desagregação apropriado, onde suposições sobre uma estrutura básica estável são mais "aceitáveis". Medidas locais de pobreza e desigualdade também podem ser úteis

para analisar a focalização das políticas sociais, e podem ajudar os governantes a entender os prós e contras envolvidos na descentralização das suas decisões. Outro ponto importante ressaltado por Elbers, Lanjouw e Lanjouw (2002) é que, enquanto é benéfico conhecer as necessidades e prioridades locais, se a desigualdade local for grande e as decisões são tomadas pela elite, os projetos podem não beneficiar os mais pobres. Eles afirmam ainda que medidas de desigualdade locais, junto com dados de escolhas de projeto, possibilitam lançar uma luz acerca dos custos potenciais de descentralização.

A metodologia proposta por Elbers, Lanjouw e Lanjouw (2002)¹³ consiste, basicamente, das seguintes etapas:

- Formulação do modelo de consumo/renda;
- Preparação dos dados;
- Ajuste do modelo de consumo/renda com base na amostra pequena;
- Predição/imputação de valores de consumo/renda para o Censo ou grande amostra;
- Estimação das medidas de pobreza para as pequenas áreas de interesse.

Nas subseções seguintes há um detalhamento técnico da metodologia.

3.2. O modelo de consumo/renda

Nesta seção do trabalho são apresentados os modelos de consumo (ou renda). A primeira tarefa é desenvolver um modelo empírico para y_{ch} , que é o gasto (ou

¹³ A metodologia proposta do Elbers, Lanjouw e Lanjouw, neste trabalho será denominada de metodologia ELL, apenas com intuito de facilitar a referência.

renda) per capita do domicílio h pertencente ao conglomerado c . Seguindo Elbers, Lanjouw e Lanjouw (2002), propõe-se aproximar a distribuição de $\ln y_{ch}$ por:

$$\ln y_{ch} = E\left[\ln y_{ch} \mid x_{ch}^T\right] + u_{ch} = x_{ch}^T \beta + u_{ch} \quad (3.1)$$

onde x_{ch} é um vetor de variáveis preditoras observadas para o domicílio h do conglomerado c , β é um vetor de parâmetros desconhecidos de dimensão compatível com x_{ch} , e o vetor de erros $u = [u_{ch}]$ tem distribuição $N(0, \Sigma)$.

Como esta dissertação faz uma abordagem da pobreza no Rio Grande do Norte, o modelo poderia também ser feito diretamente para um indicador de pobreza, como proporção de pobres. No entanto, optou-se por estimar um modelo para o logaritmo natural da renda e outro para o logaritmo natural da despesa, desta forma é possível calcular outros indicadores de pobreza baseados nas variáveis modeladas, e não somente a proporção de pobres.

Um ponto que deve ser ressaltado na metodologia ELL (2002) é que ela não leva em conta o problema de causalidade inversa, pois entre as variáveis preditoras no modelo (3.1) podem ser incluídas variáveis que resultariam da renda ou consumo, mas são no modelo utilizadas como variáveis preditoras. O principal objetivo é dispor de um modelo que permita a melhor previsão possível do consumo (ou renda) a partir de variáveis preditoras que possam ser facilmente observadas mesmo em pesquisas pouco detalhadas, como ocorre com os censos, sem precisar fornecer interpretações econômicas para os parâmetros estimados desse modelo.

Os autores não indicam nenhum método de seleção de variáveis preditoras para inclusão no modelo (3.1). Na seção 3.4 desta dissertação é descrito um método para seleção de variáveis preditoras que permite levar em conta o poder de

explicação de cada variável sobre a variável resposta (consumo ou renda domiciliar per capita).

O modelo especificado em (3.1) pode ser usado diretamente para descrever a distribuição do consumo ou renda domiciliar per capita, mas está incompleto porque não incorpora efeitos de grupo, que poderiam contribuir para melhorar sua capacidade preditiva. Por este motivo, o passo seguinte será propor um modelo para os resíduos do modelo (3.1). Um modelo que permite incorporar tais efeitos de grupo pode ser formulado a partir da seguinte decomposição dos resíduos:

$$u_{ch} = \eta_c + \varepsilon_{ch} \tag{3.2}$$

onde η_c e ε_{ch} satisfazem

$$E(\eta_c) = 0 \forall c, \quad E(\varepsilon_{ch}) = 0 \forall c, h, \quad \eta_c \perp \eta_k \forall c \neq k, \quad \varepsilon_{ch} \perp \varepsilon_{kl} \forall (c,h) \neq (k,l), \quad \text{e}$$

$$\varepsilon_{ch} \perp \eta_k \forall c,h,k, \quad \text{e } \eta_c \sim N(0; \sigma_\eta^2) \text{ e } \varepsilon_{ch} \sim N(0; \sigma_\varepsilon^2).$$

Seja σ_u^2 a variância dos resíduos do modelo ao nível do domicílio h e conglomerado c . Com base no modelo (3.2), esta variância pode ser decomposta em duas parcelas:

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

Utilizando esta decomposição, é possível tentar incorporar efeitos de grupo e variáveis preditoras do nível do conglomerado. Nesta dissertação, os conglomerados nos dados da pesquisa correspondem a setores do censo. Assim as variáveis agregadas podem ser calculadas para todos os domicílios do setor e juntadas com a amostra menor. É importante que se calculem as variáveis agregadas através de dados de censos, pois medidas calculadas com mais domicílios fornecem indicadores com muito menos ruído que as calculadas a partir de amostras pequenas. Outras fontes de informação também podem ser usadas.

3.3. *Preparação dos dados*

Este tópico da dissertação aborda a forma como se devem identificar quais variáveis podem ser usadas na etapa de modelagem, discutida anteriormente.

Seja W um indicador de pobreza ou desigualdade, baseado na distribuição dos valores de uma variável de interesse y observada ao nível de domicílio. Usando uma amostra pequena, porém rica, estima-se a distribuição condicional de y dado um vetor de covariáveis x . Restringindo o conjunto de variáveis explicativas (x) para aquelas que também podem ser observadas nos domicílios do censo, a distribuição estimada na amostra pequena pode ser usada para gerar a distribuição de y , condicionada às características x da sub-população observada, para qualquer sub-população do censo. Essas variáveis, x , que serão usadas como variáveis de ligação entre as duas pesquisas, devem ser idênticas em termos de definição e distribuição, não bastando apenas assegurar que os questionários (perguntas) sejam iguais nas duas fontes. Em outras palavras, deve-se preparar um banco de dados de ambas as pesquisas, identificando as variáveis comuns às duas, e assegurar que as variáveis x sejam compatíveis em termos de definição e distribuição.

O primeiro passo desta etapa de trabalho seria escolher um rol de variáveis x “candidatas” a fazer parte do modelo. Isso se faz comparando os questionários e arquivos de documentação das duas pesquisas. Nesse procedimento, quaisquer variáveis que existam nas duas pesquisas, e que tenham definições semelhantes, farão parte do primeiro grupo de variáveis candidatas.

O passo seguinte é o de comparação entre as variáveis comuns às duas bases de dados que foram eleitas no primeiro passo, através do cálculo de médias,

variâncias, proporções, além do emprego de testes estatísticos para avaliar se as variáveis seguem a mesma distribuição. O teste usado para saber se as variáveis são iguais em distribuição foi o teste de igualdade de distribuições.

Considere que p_1 representa o vetor de proporções das categorias da variável na pesquisa 1 e p_2 é o vetor de proporções das categorias da variável na pesquisa 2, então a hipótese de igualdade de duas distribuições discretas¹⁴ pode ser expressa como $H_0 : p_1 = p_2$, excluindo-se uma categoria da variável testada, pois a soma das proporções é 1 em ambos os casos.

Sejam $\hat{p}_l = (\hat{p}_{l1}, \dots, \hat{p}_{lK-1})'$ os vetores de proporções estimadas ($l=1,2$), baseados em amostras independentes, onde $\hat{p}_{lk} = \frac{\hat{N}_{lk}}{\hat{N}_{l\bullet}}$ é um estimador consistente da proporção p_{lk} , e \hat{N}_{lk} e $\hat{N}_{l\bullet}$ são estimadores ponderados das frequências das categorias da variável de modo que $\sum_{k=1}^K \hat{N}_{lk} = \hat{N}_{l\bullet}$.

Sejam $\hat{V}_p(\hat{p}_1)$ e $\hat{V}_p(\hat{p}_2)$ estimadores consistentes das matrizes de variância dos vetores \hat{p}_1 e \hat{p}_2 , respectivamente. Uma vez que as amostras das duas pesquisas são independentes, a estatística de Wald para realizar o teste de igualdade de distribuições (ver Pessoa e Silva, 1998)¹⁵ é dada por:

$$X_W^2(H) = (\hat{p}_1 - \hat{p}_2)' [\hat{V}_p(\hat{p}_1) + \hat{V}_p(\hat{p}_2)]^{-1} (\hat{p}_1 - \hat{p}_2). \quad (3.3)$$

A estatística de Wald $X_W^2(H)$ tem distribuição assintótica *qui-quadrado* com ($K-1$) graus de liberdade. Esta estatística funciona adequadamente quando o número de unidades primárias na amostra é grande. Se isso não for verdade, podem ocorrer

¹⁴ As variáveis contínuas, como idade, devem ser discretizadas.

¹⁵ PESSOA, D. G. C. e SILVA, P. L. do N. (1998), p. 123.

problemas de instabilidade, e neste caso se recomenda, alternativamente, uma estatística F ajustada que tem distribuição de referência t.

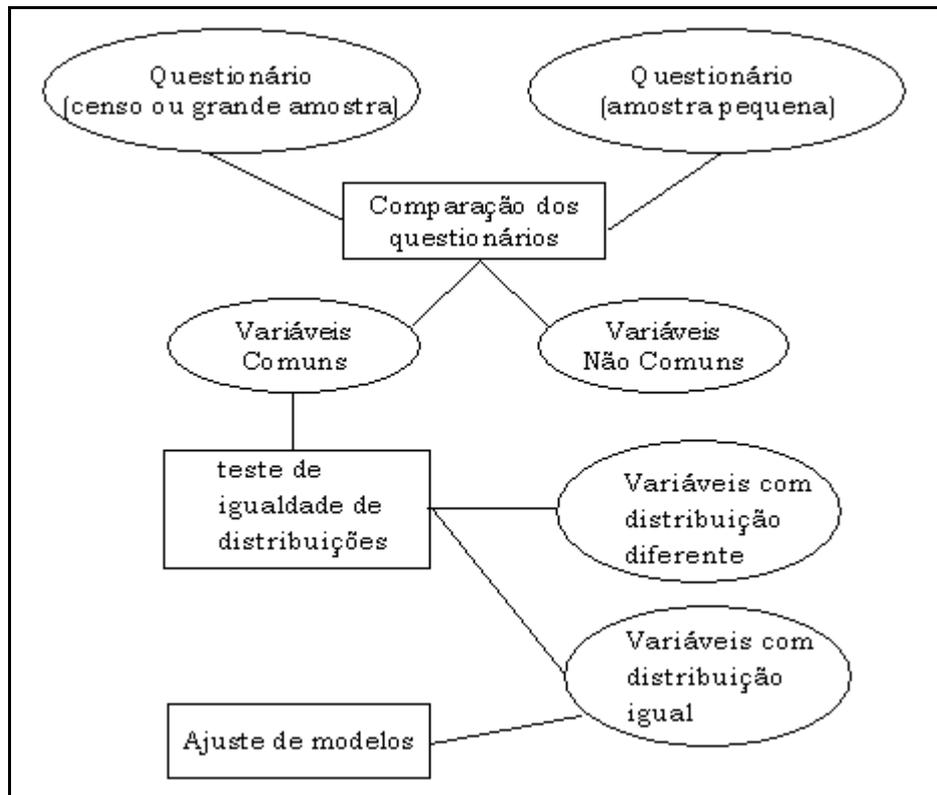
Está claro que, para a realização do teste, se faz necessário o cálculo de estimativas da matriz de covariâncias dos estimadores das distribuições das variáveis estudadas. No entanto, como os estimadores são quantidades não lineares, o método de estimação da matriz de covariâncias consiste em combinar o método de Linearização de Taylor com o método do Conglomerado Primário.

A técnica de Linearização de Taylor é utilizada para obter aproximações assintóticas para a variância do estimador do vetor de parâmetros, quando o vetor de parâmetros é função não linear das observações amostrais. Já o método do Conglomerado Primário (*Ultimate Cluster*) é usado para estimação de variâncias para estimadores de totais e médias em planos amostrais de múltiplos estágios, e considera apenas a variação entre informações disponíveis a nível das unidades primárias de amostragem (UPAs), ou seja, a nível dos conglomerados primários.

Esta dissertação não esgota a discussão sobre estes dois métodos, e um tratamento mais aprofundado pode ser encontrado no livro Särndal, Swensson e Wretman (1992).

Em suma, a escolha das variáveis que constarão na segunda etapa da elaboração no mapa da pobreza é realizada de acordo com uma série de procedimentos, tal qual está descrito na figura 3.1 a seguir.

Figura 3.1: Processo de seleção das variáveis preditoras para o ajuste do modelo



3.4. Ajuste do modelo

Uma primeira estimativa para o parâmetro β no modelo (3.1) pode ser obtida pelo método de Mínimos Quadrados Ordinários (MQO) ou Mínimos Quadrados Ponderados (MQP). Em notação matricial, o modelo pode ser escrito da seguinte forma:

$$\ln y = X\beta + u$$

onde $\ln y$ é o vetor dos logaritmos naturais da renda ou da despesa domiciliar per capita, β é o vetor dos coeficientes de regressão e u é o vetor de resíduos.

Os estimadores de β por MQO são calculados através da seguinte equação:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (3.4)$$

A matriz de variância-covariância de $\hat{\beta}$ é estimada por:

$$\hat{V}(\hat{\beta}) = \sigma_u^2 (X'X)^{-1} \quad (3.5)$$

onde σ_u^2 é a variância homocedástica dos resíduos da regressão (u), dada por

$$\sigma_u^2 = \frac{\sum_c \sum_h \hat{u}_{hc}^2}{n - k}, \text{ sendo que } n \text{ é o número de observações e } k \text{ é o número de}$$

parâmetros do modelo.

Suponha que tenhamos observações de u_{ch} , onde $u_{ch} = \eta_c + \varepsilon_{ch}$. O problema está em estimar $\sigma_\eta^2 = \text{var}(\eta)$. Então

$$E(u_{c.}^2) = \sigma_\eta^2 + \sigma_\varepsilon^2$$

Sendo que as estimativas para σ_η^2 e σ_ε^2 podem ser obtidas por:

$$\hat{\sigma}_\eta^2 = \max \left(\frac{\sum_c w_c (u_{c.} - u_{..})^2}{\sum_j w_j (1 - w_j)} - \frac{\sum_c w_c (1 - w_c) \hat{\sigma}_\varepsilon^2}{\sum_j w_j (1 - w_j)}; 0 \right) \quad (3.6)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n_c(n_c - 1)} \sum_h (\varepsilon_{ch} - \varepsilon_{c.})^2$$

E o estimador da variância de $\hat{\sigma}_\eta^2$ pode ser calculado usando simulações, que segundo Elbers, Lanjouw e Lanjouw (2002)¹⁶ é uma forma mais direta de estimar a $\text{var}(\hat{\sigma}_\eta^2)$. Evidentemente, o primeiro passo é estimar $\hat{\sigma}_\eta^2$, através da fórmula (3.6) acima. Em seguida estimar a variância do modelo de heterocedasticidade $\sigma_{\varepsilon, ch}^2$, exposto anteriormente. Isso fornece $\hat{\sigma}_{\varepsilon, ch}^2$. Usando as componentes de variância estimadas, e assumindo que η_c e ε_{ch} são independentes e seguem uma distribuição normal com média zero, são gerados novos valores para u_{ch} , a partir de

$$u_{ch} = \eta_c + \varepsilon_{ch}.$$

¹⁶ Elbers, Lanjouw e Lanjouw (2002, p. 54)

Em seguida novas estimativas para σ_{η}^2 são calculadas a partir da equação (3.6) para $\hat{\sigma}_{\eta}^2$ dada acima. As operações são repeditas várias vezes, guardando os valores simulados de σ_{η}^2 . O conjunto de valores simulados para σ_{η}^2 pode ser usado para estimar a variância amostral de $\hat{\sigma}_{\eta}^2$, diretamente.

No artigo, Elbers, Lanjouw e Lanjouw (2002) afirmam que, na prática, $\hat{\sigma}_{\eta}^2$ muitas vezes é tão pequena que a equação pode fornecer muitas estimativas nulas para a variância de η , ou seja, o modelo (3.2) não teria efeitos de grupo.

A seguir, é necessário simular os resíduos η e ε . Uma distribuição apropriada pode ser determinada a partir do resíduo de conglomerado $\hat{\eta}_c$ e resíduo padronizado de domicílio

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

onde H_c é o número total de domicílios no conglomerado c . Estimar diretamente dos resíduos padronizados evita fazer qualquer suposição a respeito da distribuição dos resíduos. Alternativamente, os percentis de uma distribuição empírica de resíduos padronizados podem ser comparados com os percentis correspondentes de uma distribuição Normal Padrão, t , ou outra distribuição.

3.5. Método para seleção de modelos

Durante o ajuste de modelos de regressão, é comum o pesquisador ficar em dúvida de qual conjunto de variáveis deverá compor seu modelo final. Muitas vezes é quase impossível realizar um estudo detalhado de todos os modelos de regressão possíveis. Entretanto, existem métodos de seleção de variáveis cujo propósito é auxiliar o pesquisador a decidir qual modelo deverá ser ajustado.

Neste trabalho, a variável resposta sempre será denotada por y e as variáveis explicativas, por x . Em geral, espera-se que o número de variáveis explicativas (J) seja menor que o número de observações (n). O problema consiste na busca de um conjunto de variáveis (x_1, x_2, \dots, x_j) que minimize a soma dos quadrados dos erros de previsão, dada por:

$$S = \sum_{i=1}^n \left(y_i - \sum_{j=1}^J \hat{\beta}_j x_{ij} \right)^2$$

onde y_i é o valor da variável resposta para a i -ésima observação, $i = 1, 2, \dots, n$, x_{ij} é o valor da j -ésima variável explicativa para a i -ésima observação e $\hat{\beta}_j$ é o efeito da j -ésima variável explicativa sobre a média da variável resposta.

Na prática, o valor de J não é pré-determinado, e queremos encontrar um subconjunto parcimonioso de variáveis explicativas que explique bem o comportamento da variável resposta.

O procedimento seleção passo a passo para frente¹⁷ é um bom método para este trabalho. De acordo com o procedimento de seleção passo a passo para frente, as variáveis são adicionadas ao modelo uma a uma, a partir de um algoritmo de

¹⁷ Este método de seleção também é conhecido como *forward selection*. Para maiores detalhes, consultar Miller (1990).

entrada de variáveis descrito em (Neter, Wasserman e Kutner, 1990, p.453), de forma que a primeira variável selecionada, é a variável x_j que minimiza

$$S = \sum_{i=1}^n (y_i - \hat{\beta}_j x_{ij})^2$$

onde $\hat{\beta}_j$ minimiza S_j para a variável x_j , e é dado por

$$\hat{\beta}_j = \frac{\sum_{i=1}^n x_{ij} y_i}{\sum_{i=1}^n x_{ij}^2}$$

daí tem-se que

$$S = \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n x_{ij} y_i \right)^2 / \sum_{i=1}^n x_{ij}^2$$

Portanto, a variável selecionada é aquela que maximiza a parcela

$$\left(\sum_{i=1}^n x_{ij} y_i \right)^2 / \sum_{i=1}^n x_{ij}^2, \quad (3.7)$$

entre todas as variáveis preditoras.

Suponha que a primeira variável selecionada seja denotada por x_{j1} . A partir daí são ajustados modelos com duas variáveis, x_{j1} mais uma variável do rol de variáveis explicativas ainda não incluídas no modelo. Os resíduos $y - x_j \hat{\beta}_j$ são ortogonais a x_{j1} , então na medida em que as variáveis vão sendo adicionadas ao modelo, a soma de quadrados dos resíduos vai diminuindo.

Agora, para cada variável x_j , além da que já foi selecionada, pode-se formar $x_{j1} = x_j - \hat{\beta}_{j1} x_j$, onde $\hat{\beta}_{j1}$ é o coeficiente da regressão de x_j sobre x_{j1} . A próxima variável selecionada é aquela que maximiza a função (3.7), quando y é substituído por $y - x_{j1} \hat{\beta}_{j1}$ e x_j por x_{j1} . A soma de quadrados e a soma de produtos cruzados podem ser diretamente calculadas a partir das somas de quadrados e de produtos anteriores,

sem precisar calcular as componentes ortogonais de cada observação. Se a média fosse subtraída de cada variável, então a nova variável a ser selecionada é aquela que possuir maior correlação parcial, em valor absoluto, com y depois que o modelo com a variável x_{j-1} foi ajustado.

Assim, as variáveis x_1, x_2, \dots, x_J são progressivamente adicionadas à equação de predição, e cada variável é selecionada porque minimiza a soma de quadrados dos resíduos ao ser adicionada ao modelo que já contém outras variáveis previamente selecionadas. Uma vez que a variável é selecionada, não pode mais ser retirada do modelo.

Como neste trabalho se busca um modelo com um bom poder de previsão da variável resposta (renda ou despesa), optou-se por determinar a seleção de variáveis explicativas a partir da estatística F_j^* , definida adiante. Primeiramente é ajustado um modelo de regressão linear simples para cada uma das $(J-1)$ potenciais variáveis explicativas (x). Para cada modelo, a estatística F é calculada para testar a sua significância.

$$F_j^* = \frac{MQR(X_j)}{MQE(X_j)} \quad (3.8)$$

Onde $MQR(x_j)$, é a média dos quadrados dos resíduos da regressão da variável x_j , e mede a redução na variação total de y , associada ao uso da variável x_j . Assim, a variável x que apresentar o maior F_j^* é a primeira candidata a compor o modelo. Segundo Miller (1990, p.50), se este valor F_j^* excede o valor 4, a variável x_j é adicionada ao modelo. Caso contrário, o processo encerra sem nenhuma variável suficientemente boa para compor o modelo. Miller define o valor limite da estatística F_j^* como F de entrada (F_e), e sugere o valor $F_e = 4$.

Imagine que X_7 seja a variável a ser adicionada no primeiro passo. Agora devem ser ajustados todos os modelos com duas variáveis, sendo que X_7 estará presente em todos esses modelos. Isto é, são ajustados vários modelos com 2 variáveis, X_7 e mais uma. Para cada modelo é obtida a estatística F:

$$F_j^* = \frac{SQR_j - SQR_{j+1}}{SQR_{j+1}/(n - j - 2)} \quad (3.9)$$

onde SQR_j é a soma de quadrados dos resíduos do modelo com j variáveis e onde SQR_{j+1} denota a soma de quadrados dos resíduos do modelo com $(j+1)$ variáveis. E esta é a estatística para testar se $\beta_{j+1} = 0$, quando x_j já está no modelo. A variável X com maior F de entrada (F_e) é a candidata a ser adicionada ao modelo. Mais uma vez é realizado um teste, e se $F_j^* > 4$ a variável é adicionada.

O procedimento é repetido enquanto estatística F^* apresentar valor maior que 4. Suponha que esteja sendo testada a inclusão de 5 variáveis, e o maior F^* obtido não superou 4, então o procedimento deve ser encerrado, e o modelo final é aquele com 4 variáveis.

É importante lembrar que uma vez que a variável é incluída, ela não deve ser retirada do modelo.

3.6. Imputação da renda/despesa no Censo

Nesta etapa são descritos os procedimentos usados na imputação das variáveis de renda ou despesa no Censo. Para imputar o valor da renda ou da despesa *per capita* de cada domicílio no Censo, os parâmetros estimados a partir dos dados da amostra pequena são utilizados para calcular

$$\hat{y}_{ch}^r = \exp[x'_{ch}\hat{\beta} + u_{ch}^r] \quad (3.10)$$

Onde $\hat{\beta}$ é o vetor de parâmetros do modelo de consumo ou renda (3.1) estimado com base na amostra pequena usando os métodos descritos na seção 3.4, x_{ch} é um vetor de valores das variáveis preditoras para o domicílio h do conglomerado c observado no Censo, e o resíduo \hat{u}_{ch}^r é obtido por simulação, mediante uma das duas abordagens descritas a seguir. É importante ressaltar que não há ligação entre os dados individuais do domicílio na pesquisa de amostra pequena e na pesquisa com a amostra grande.

Abordagem paramétrica para simulação dos resíduos

Sob esta abordagem, os resíduos u_{ch} do modelo (3.1) são considerados como provenientes da distribuição paramétrica $N(0;\Sigma)$. Então para obter valores simulados dos resíduos u_{ch} basta gerar vetores de resíduos u com a distribuição $N(0;\hat{\Sigma})$, onde $\hat{\Sigma}$ é o estimador de Σ descrito na seção 3.4. Tantas réplicas do vetor de resíduos u são geradas quantas sejam necessárias, sempre de forma independente.

Abordagem semi-paramétrica

Considerando o modelo proposto em (3.2) para decomposição dos resíduos u_{ch} , onde $u_{ch} = \eta_c + \varepsilon_{ch}$, nesta abordagem a idéia é obter \hat{u}_{ch}^r mediante adição de dois resíduos “empíricos”, $\hat{\eta}_c^r$ e e_{ch}^{*r} obtidos como descrito a seguir. Estes dois resíduos empíricos representam os erros dos níveis de conglomerado e domicílio.

Primeiro o efeito do nível de conglomerado ou geográfico $\hat{\eta}_c^r$ é selecionado aleatoriamente com reposição do conjunto de todos os resíduos empíricos $\hat{\eta}_c$. Em seguida, e_{ch}^{*r} é obtido também por amostragem com reposição do conjunto de todos os resíduos empíricos e_{ch}^* que correspondem a domicílios pertencentes ao conglomerado cujo efeito do nível geográfico $\hat{\eta}_c$ foi selecionado na primeira etapa. A soma destes dois resíduos gera o valor do resíduo a ser utilizado na equação (3.10) para obter os valores de renda e despesas de cada réplica simulada no Censo.

3.7. O estimador de bem-estar

O processo de imputação descrito em 3.5 é repetido R vezes, gerando a cada vez uma réplica “completa” dos valores de renda ou despesa per capita para todos os domicílios existentes no Censo. Com cada conjunto de valores simulado, obtém-se um valor para o indicador ou medida 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, e é dado por:

$$\tilde{W}_a = \frac{1}{R} \sum_{r=1}^R \hat{W}_a^r .$$

onde \hat{W}_a^r é o valor do indicador para a área a obtido com os valores simulados de renda ou despesa domiciliar per capita para a réplica r.

4. Aplicação

O objetivo deste capítulo é apresentar a construção dos mapas de pobreza para o estado do Rio Grande do Norte, empregando a metodologia ELL (2002). Com os mapas da pobreza construídos, será feita uma avaliação da aplicação da metodologia, através da comparação destes resultados com os obtidos de forma direta no Censo 2000, com base na variável de rendimento, investigada no questionário da amostra. Outro enfoque adotado para avaliar a metodologia consiste em comparar os resultados do indicador de pobreza estimado para pequenas áreas com o resultado do mesmo indicador obtido de forma direta para a grande área (no caso, o estado do Rio Grande do Norte).

Estas duas abordagens permitem uma avaliação da aplicabilidade e adequação da metodologia. Nas seções a seguir serão discutidos: a delimitação da área de estudo; os métodos adotados para estimação; as fontes de dados utilizadas neste trabalho; a seleção e ajuste de modelos; a avaliação da qualidade do ajuste; e finalmente serão apresentados os resultados do mapa da pobreza.

4.1. Delimitação da área de estudo

No momento da definição do tema desta dissertação surgiu a idéia de realizar um estudo sobre pobreza para o Brasil, mas com a abordagem de pequenas áreas. Então deveria ser realizado um mapeamento da pobreza para cada uma das 27 Unidades da Federação. No entanto, o esforço computacional requerido para realizar tal trabalho para todos os estados do Brasil seria muito grande. Então optamos por escolher uma única Unidade da Federação, e realizar o estudo detalhadamente para

esta. Constatou-se que a produção acadêmica a respeito do Estado do Rio Grande do Norte é pequena, e levando em conta a naturalidade da autora deste, resolveu-se focar o estudo sobre este estado, que é muito conhecido por suas belas praias, mas pouco por suas características como educação, cultura, desigualdade social e pobreza.

Segundo o último Censo Demográfico, realizado em 2000, o Rio Grande do Norte tem população de cerca de 2.8 milhões de habitantes, residentes em 166 municípios. O grande número de municípios associado à pequena população, nos levou a pensar que alguns municípios teriam população muito pequena. De fato, 21 dos 166 municípios têm população inferior a 3 mil habitantes, sendo que em 3 destes municípios a população não chega a 2 mil habitantes.

O estudo poderia tentar estimar o indicador de pobreza para cada um dos 166 municípios, mas em virtude da pequena população de alguns deles, suspeitou-se que a qualidade das informações estimadas para essas áreas poderia ser baixa. Então, optou-se por realizar o estudo para uma agregação de municípios, que permitisse a avaliação de pequenas áreas. A agregação escolhida foi microrregião.

Os 166 municípios do Rio Grande do Norte estão divididos em 19 microrregiões. A tabela abaixo mostra a população estimada do estado e das suas das microrregiões.

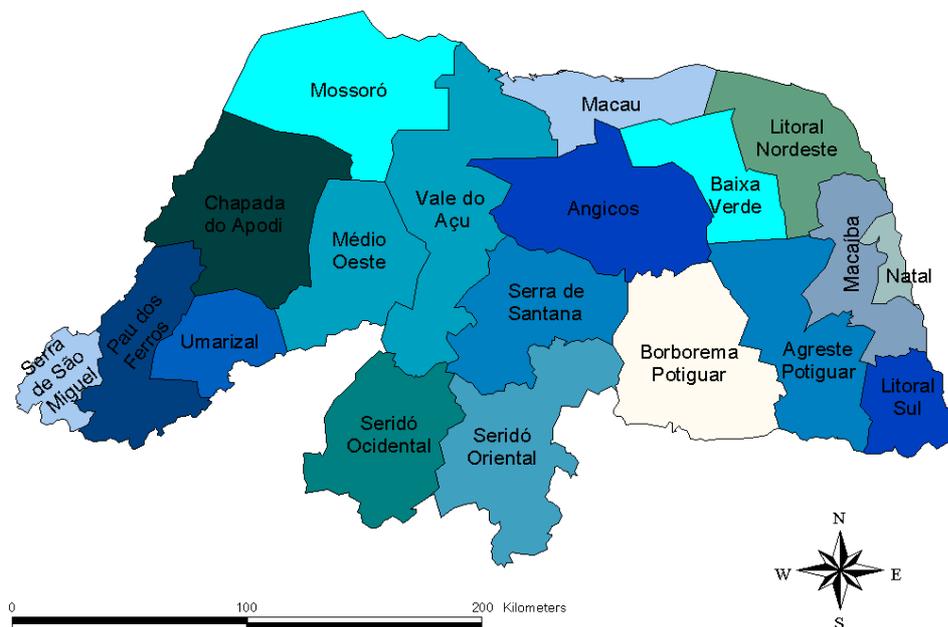
Tabela 4.1: População total do Rio Grande do Norte e das Microrregiões

MICRORREGIÃO	População estimada	%
Rio Grande do Norte	2.777.509	100,00
Médio Oeste	37.085	1,34
Macau	44.901	1,62
Angicos	50.932	1,83
Baixa Verde	56.912	2,05
Serra de São Miguel	58.579	2,11
Serra de Santana	60.682	2,18
Umarizal	64.043	2,31
Chapada do Apodi	70.290	2,53
Litoral Nordeste	75.188	2,71
Seridó Ocidental	89.562	3,22
Pau dos Ferros	112.349	4,04
Litoral Sul	113.576	4,09
Seridó Oriental	114.535	4,12
Vale do Açu	124.753	4,49
Borborema Potiguar	125.767	4,53
Agreste Potiguar	206.106	7,42
Macaíba	240.694	8,67
Mossoró	274.976	9,90
Natal	856.579	30,84

Fonte: Censo Demográfico 2000

A figura 4.1 a seguir representa o estado do Rio Grande do Norte e suas microrregiões.

Figura 4.1: Mapa do Rio Grande do Norte e microrregiões



4.2. Fontes de dados

4.2.1. POF

A Pesquisa de Orçamentos Familiares (POF) realizada em 2002/2003 é a quarta pesquisa desenvolvida pelo IBGE sobre orçamentos familiares¹⁸. A primeira foi o ENDEF (Estudo Nacional de Despesa Familiar), que ocorreu em 1974/75, e tinha âmbito territorial nacional, à exceção da área rural da Região Norte. Mais de 10 anos depois da realização do ENDEF é que foi implementada a primeira POF, que foi realizada em 1987/88, sendo a segunda realizada em 1995/96. Essas duas últimas

¹⁸ Pesquisa de orçamentos familiares 2002-2003: Primeiros resultados, p. 9.

foram realizadas para atualizar as estruturas de consumo dos índices de preços ao consumidor do IBGE.

A investigação mais recente, POF 2002/2003, além de servir como base para o cálculo dos índices de preços, como as pesquisas anteriores, tem como principal objetivo medir as estruturas de consumo, dos gastos e dos rendimentos das unidades de consumo, permitindo esboçar um perfil das condições de vida dos brasileiros.

A POF 2002/2003 cobriu todo o território brasileiro, uma vez que o desenho da sua amostra permite a estimação de resultados para o Brasil e Grandes Regiões, além da situação urbana e rural. Apesar do título da pesquisa abordar o tema família, a unidade básica da pesquisa é a unidade de consumo (UC), que compreende um conjunto de moradores compartilhando despesas com moradia ou a mesma fonte de alimentação. A pesquisa foi realizada durante o período de julho de 2002 a junho de 2003, para que fosse possível captar as alterações a que estão sujeitas as despesas, quantidade de bens e rendimentos durante o ano.

Devido à grande diversidade de itens de despesas, com diferentes valores e frequência de aquisição, foram definidos quatro períodos de referência: sete dias, 30 dias, 90 dias e 12 meses. Vale ressaltar que, como a coleta de informações ocorreu durante 12 meses, o período de referência para cada domicílio selecionado é diferente. Além disso, 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 POF 2002/2003 foi de 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

censitários¹⁹. As UPAs foram selecionadas através de amostragem sistemática com probabilidade proporcional ao número de domicílios no setor. As unidades secundárias de amostragem são os domicílios particulares permanentes, e estes foram selecionados 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: Norte, Nordeste, Sudeste, Centro-Oeste e Sul.

A variável utilizada para a estratificação estatística foi a média dos anos de estudo do responsável pelo domicílio, obtida a partir dos dados da Contagem Populacional de 1996, por ser a fonte censitária mais recente na época do planejamento da pesquisa.

Foram selecionados 3.984 setores censitários para a amostra. 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 resultou em 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 4.2 apresenta os números de estratos, setores e domicílios efetivamente pesquisados no Rio Grande do Norte.

¹⁹ Os setores censitários correspondem à base geográfica do Censo Demográfico de 2000.

Tabela 4.2: Número de Estratos, setores e domicílios pesquisados com entrevista completa no Rio Grande do Norte, por microrregião

Microrregião	Estratos	Setores	Domicílios
Médio Oeste	1	1	12
Macau	3	3	32
Angicos	2	2	24
Baixa Verde	2	3	29
Serra de São Miguel	1	1	12
Serra de Santana	3	3	38
Umarizal	2	2	28
Chapada do Apodi	3	3	41
Litoral Nordeste	1	2	31
Seridó Ocidental	4	4	46
Pau dos Ferros	4	5	60
Litoral Sul	3	5	70
Seridó Oriental	4	6	84
Vale do Açu	4	5	62
Borborema Potiguar	3	3	28
Agreste Potiguar	4	7	82
Macaíba	7	10	144
Mossoró	5	14	165
Natal	9	48	560
Total do estado	65	127	1548

Fonte: POF 2002/2003

A diferença entre a quantidade de domicílios ou setores selecionados e efetivamente entrevistados ocorre porque foi feita a opção de selecionar um número maior de domicílios do que o estipulado durante o dimensionamento da amostra, para compensar a perda de domicílios por entrevista não realizada na coleta de dados.

Observe que em algumas microrregiões do Rio Grande do Norte o número de setores entrevistados é bem pequeno. A situação é mais grave nas regiões do Médio Oeste e Serra de São Miguel, onde apenas 1 setor foi selecionado e pesquisado.

A circunstância se torna ainda mais crítica ao analisar a amostra sob o olhar do tipo de área. 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. Observe na tabela 4.3 que 9 das 19 microrregiões do estado não tiveram setor selecionado na área rural.

Essa peculiaridade no aspecto da amostragem da POF 2002/2003 ressalta a importância de se conhecer uma metodologia que possibilite a estimação para pequenas áreas, uma vez que a amostra da pesquisa, quando usada isoladamente, não permitiria obter estimadores diretos para áreas rurais das microrregiões do estado, por exemplo.

Tabela 4.3: Número de Estratos, setores e domicílios pesquisados 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	-
Total do estado	51	14	112	15	1287	261

Fonte: POF 2002/2003

Outro ponto importante a ser ressaltado na pesquisa é a expansão da amostra. Como a pesquisa foi realizada através de amostra, cada domicílio representa um conjunto de domicílios, por isso a cada domicílio da amostra é associado um peso amostral ou fator de expansão, que permite a obtenção de estimativas para o universo da pesquisa. Num primeiro momento, os pesos amostrais da POF foram calculados a partir do método de amostragem utilizado. Em seguida foram feitos ajustes de

calibração, para obter estimativas para o total de pessoas em algumas categorias, iguais às respectivas projeções populacionais para 15 de janeiro de 2003 para cada Unidade da Federação. A lista de variáveis de calibração pode ser encontrada no relatório da publicação da pesquisa, bem como informações mais detalhadas sobre o método de calibração.

4.2.2. Censo 2000

A palavra Censo, segundo o Dicionário de Termos Estatísticos²⁰, significa “uma enumeração da população ou grupo num determinado período com respeito a características bem definidas”. Neste sentido, a palavra denota a averiguação de algumas características para a totalidade da população.

No Brasil, os censos são realizados em diversas esferas temáticas, formando um conjunto de censos, denominado recenseamento. Os recenseamentos são realizados desde 1872, com periodicidade normalmente de 10 anos.

Este trabalho usou o Censo Demográfico realizado no ano 2000 (Censo 2000) como uma das principais fontes de dados. Encontra-se abaixo uma explicação geral sobre esta pesquisa.

Os Censos no Brasil são legalmente amparados, através das leis 5534 e 8184, que garantem que as informações prestadas são obrigatórias e confidenciais. Para o CENSO 2000, todas as pessoas moradoras em domicílios particulares no Território Nacional deveriam ser recenseadas. A data de referência adotada pelo IBGE para a pesquisa foi 1º de agosto de 2000, e o período de coleta foi o de 1º de agosto a 30 de novembro de 2000.

²⁰ Kendal e Buckland (1957), p. 37.

O Censo foi realizado através de dois questionários: básico e da amostra. O primeiro foi aplicado a todos os domicílios particulares do país que não tivessem sido selecionados para a amostra, e abordava poucas perguntas. Já o segundo foi aplicado somente para uma amostra de domicílios, que além das perguntas contidas no questionário básico, abrangia questões mais detalhadas sobre as características dos domicílios e das pessoas residentes, referentes aos temas: religião, cor ou raça, deficiência, migração, escolaridade, fecundidade, nupcialidade, trabalho e rendimento.

O desenho amostral adotado compreendeu a seleção sistemática com eqüiprobabilidade, dentro de cada setor censitário, 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 maior que 15.000 habitantes e 20% para os demais municípios. As frações amostrais calculadas para o Estado do Rio Grande do Norte e suas microrregiões encontram-se na tabela abaixo.

Tabela 4.4: Frações de amostragem do Rio Grande do Norte e suas microrregiões

Microrregião	Fração de amostragem
Médio Oeste	21,2
Macau	14,4
Angicos	19,8
Baixa Verde	15,7
Serra de São Miguel	17,4
Serra de Santana	21,5
Umarizal	20,8
Chapada do Apodi	13,7
Litoral Nordeste	16,8
Seridó Ocidental	14,0
Pau dos Ferros	18,4
Litoral Sul	16,5
Seridó Oriental	15,3
Vale do Açu	15,4
Borborema Potiguar	18,4
Agreste Potiguar	17,5
Macaíba	10,4
Mossoró	10,9
Natal	10,0
Total do estado	13,7

Fonte: Censo Demográfico 2000

Da mesma forma que na POF, no Censo também foram calculados pesos para cada domicílio e morador pesquisado, a fim de se obter expansão dos dados coletados pelo questionário da amostra. Inicialmente, os pesos foram calculados como o inverso das frações de amostragem. Em seguida, foi aplicado o método de calibração²¹ em relação a um conjunto de variáveis auxiliares, cujos totais populacionais já são conhecidos. A calibração visa ajustar os pesos iniciais a fim de obter os totais conhecidos, ao aplicar os pesos calibrados às variáveis auxiliares.

A amostra do censo, por definição, é muito mais robusta que a da POF, uma vez que foram pesquisados 92,6 mil domicílios em 2623 setores no Rio Grande do Norte. Conseqüentemente, as microrregiões apresentam amostras muito maiores. A microrregião que apresenta menor número de setores pesquisados é o Médio Oeste,

²¹ Para maiores detalhes, ver Censo Demográfico 2000: Documentação dos microdados da amostra, pág 11.

com 38 setores, em seguida aparece Macau, com 55 setores. Conforme esperado, a microrregião que contém a capital é a que apresenta maior amostra: 662 setores e cerca de 21,7 mil domicílios pesquisados. Observe que a desagregação da amostra em áreas urbanas e rurais continua revelando uma maior robustez para a amostra do censo, comparativamente à da POF.

Tabela 4.5: Número de setores e domicílios pesquisados com entrevista completa no Rio Grande do Norte, por microrregião

Microrregião	Setores	Domicílios
Médio Oeste	38	1.875
Macau	55	1.646
Angicos	57	2.443
Baixa Verde	73	2.000
Serra de São Miguel	63	2.414
Serra de Santana	63	3.049
Umarizal	77	3.412
Chapada do Apodi	72	2.402
Litoral Nordeste	86	2.724
Seridó Ocidental	88	3.190
Pau dos Ferros	121	5.092
Litoral Sul	102	4.157
Seridó Oriental	113	4.456
Vale do Açu	129	4.683
Borborema Potiguar	129	5.580
Agreste Potiguar	212	8.553
Macaíba	224	5.901
Mossoró	259	7.348
Natal	662	21.748
Total do estado	2.623	92.673

Fonte: Censo Demográfico 2000

Tabela 4.6: Número de setores e domicílios pesquisados com entrevista completa na amostra do Censo 2000 do Rio Grande do Norte, por microrregião

Microrregião	Setores		Domicílios	
	Urbana	Rural	Urbana	Rural
Médio Oeste	16	22	1.059	816
Macau	28	27	1.062	584
Angicos	29	28	1.717	726
Baixa Verde	27	46	1.237	763
Serra de São Miguel	25	38	1.229	1.185
Serra de Santana	24	39	1.510	1.539
Umarizal	35	42	2.202	1.210
Chapada do Apodi	34	38	1.239	1.163
Litoral Nordeste	18	68	1.000	1.724
Seridó Ocidental	55	33	2.416	774
Pau dos Ferros	58	63	3.119	1.973
Litoral Sul	44	58	2.436	1.721
Seridó Oriental	71	42	3.476	980
Vale do Açu	64	65	2.919	1.764
Borborema Potiguar	66	63	3.448	2.132
Agreste Potiguar	85	127	4.586	3.967
Macaíba	75	149	2.465	3.436
Mossoró	219	40	6.560	788
Natal	636	26	21.194	554
Total do estado	1.609	1.014	64.874	27.799

Fonte: Censo Demográfico 2000

4.3. Seleção e ajuste de modelos

O ajuste de modelos de renda e despesa é uma etapa muito importante na construção dos mapas da pobreza. A partir dos modelos ajustados com os dados da POF 2002/2003 para o Rio Grande do Norte, serão imputados valores para as variáveis respostas no Censo 2000.

Diferentemente das análises mais comuns, estes modelos não precisam ter coerência estatística ou econômica, mas sim um grande poder preditivo, pois a finalidade dos modelos é que eles sejam capazes de explicar a renda *per capita* e a despesa total *per capita*. Para viabilizar a etapa de imputação da renda e da despesa no censo, as variáveis preditoras devem ser iguais em termos de distribuição, ao

comparar as proporções das duas fontes de dados. Nas subseções seguintes, encontram-se as análises das variáveis de interesse (resposta) e das possíveis variáveis preditoras (explicativas) consideradas.

4.3.1. *Variáveis respostas*

Existem vários tipos de dados para serem utilizados na análise da pobreza. Em alguns estudos foram utilizadas as renda individuais das pessoas economicamente ativas (PEA). Pode ser adotada a renda total de todas as fontes ou apenas a do trabalho principal. Esta última abordagem é perfeita para o caso de estar analisando o mercado de trabalho.

Se o objetivo for analisar o nível de vida das pessoas, a variável mais adequada não é a renda das pessoas economicamente ativas, nem a renda individual, como se pode imaginar. Pode ser, por exemplo, que uma pessoa ganhe relativamente pouco, mas pertença a um domicílio com renda alta. Então, a alternativa mais adequada para analisar a pobreza é classificando as pessoas de acordo com seu rendimento domiciliar *per capita*. A construção de indicadores de pobreza baseados na renda domiciliar *per capita* é muito comum e fácil de compreender.

Também é possível classificar as pessoas ou domicílios de acordo com suas despesas. Da mesma forma que a renda, a variável de despesa mais adequada para a análise da pobreza é a despesa domiciliar *per capita*. Na impossibilidade de obtenção de dados sobre consumo, a despesa *per capita* pode ser usada como uma aproximação do consumo das pessoas, que também é uma abordagem muito freqüente nas discussões sobre pobreza.

Este trabalho não objetiva responder qual variável é melhor para avaliar a pobreza, nem qual é a melhor forma de captar rendimentos ou despesas em pesquisas amostrais, muito menos quer corrigir tais variáveis. Mesmo assim, achou-se necessário uma análise das variáveis para melhor entendimento do método que está sendo aplicado.

Neste trabalho foram ajustados dois modelos de regressão: um para a renda domiciliar per capita e outro para a despesa domiciliar per capita. As variáveis que foram utilizadas como resposta nos modelos a seguir foram analisadas quanto à adequação de suas distribuições ao modelo normal. A tabela 4.7 abaixo apresenta algumas medidas resumo que fornecem informações sobre a distribuição observada na amostra da POF 2002/2003 para as duas variáveis respostas consideradas.

Tabela 4.7: Medidas resumo das variáveis respostas

	Renda per capita	Despesa total per capita
Média	R\$ 367,22	R\$ 255,75
Mínimo	R\$ 3,32	R\$ 9,39
1º quartil	R\$ 97,83	R\$ 87,46
Mediana	R\$ 176,92	R\$ 150,53
3º quartil	R\$ 338,28	R\$ 273,34
Máximo	R\$ 14.015,26	R\$ 12.635,75
Desvio padrão	R\$ 16.408,88	R\$ 9.782,46

Fonte: Pof 2002/2003

Essas medidas mostram uma forte assimetria à direita nas distribuições. As duas variáveis apresentam médias muito superiores às medianas, caracterizando grande concentração dos valores nas camadas mais baixas de renda ou despesa e por outro lado, poucas observações com renda ou despesa elevadas. Além disso, os desvios padrões são muito grandes. A hipótese da assimetria verificada a partir das medidas resumo é confirmada pelos próximos gráficos.

Gráfico 4.1: Histograma da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados

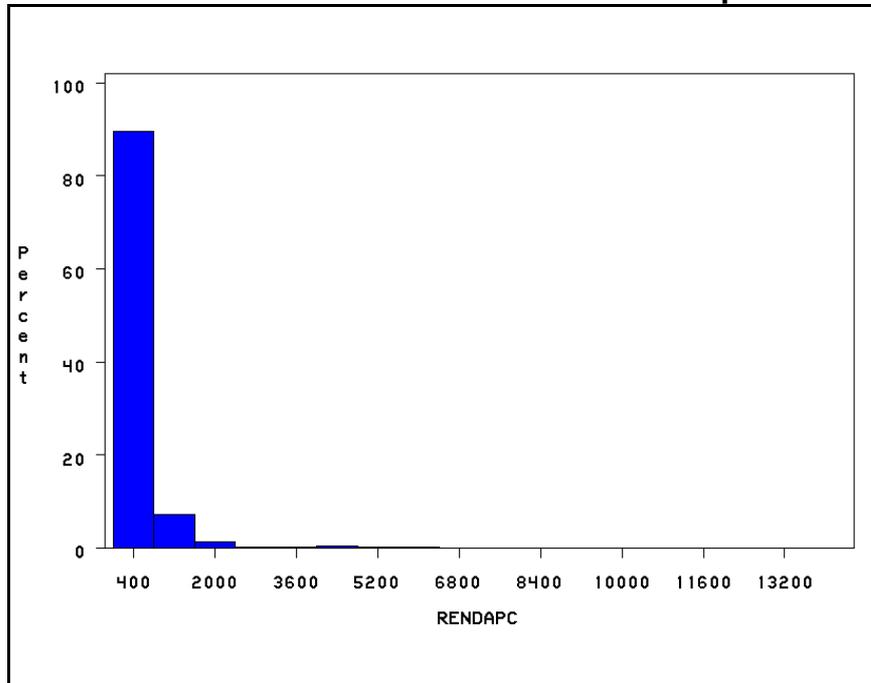
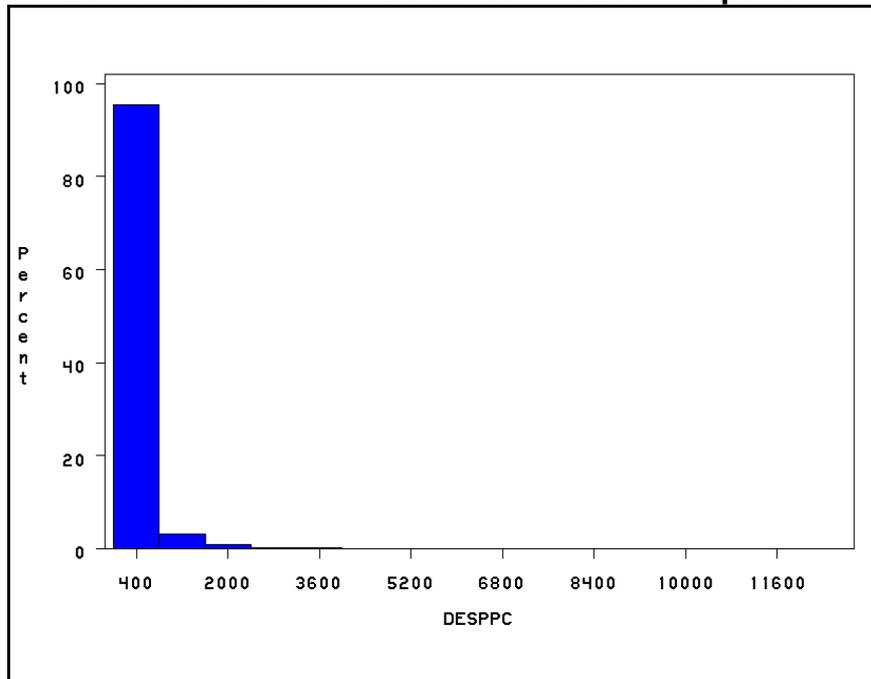


Gráfico 4.2: Histograma da despesa total domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados



Além da assimetria acentuada, os gráficos confirmam a idéia de que as variáveis de renda e despesa, quando trabalhadas na escala original, não seguem a distribuição normal. Para que a adequação ao modelo normal seja feita, as variáveis passaram por transformações logarítmicas. A tabela 4.8 mostra as medidas resumo das variáveis transformada. Observa-se que a média e a mediana assumem valores bem mais próximos para ambas as variáveis.

Tabela 4.8: Medidas resumo das variáveis respostas transformadas

	Ln da Renda per capita	Ln da Despesa total per capita
Média	5,26	5,07
Mínimo	1,20	0,00
1º quartil	4,58	3,73
Mediana	5,18	4,02
3º quartil	5,82	4,47
Máximo	9,55	6,64
Desvio padrão	22,07	19,41

Fonte: Pof 2002/2003

Gráfico 4.3: Histograma do logaritmo da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados

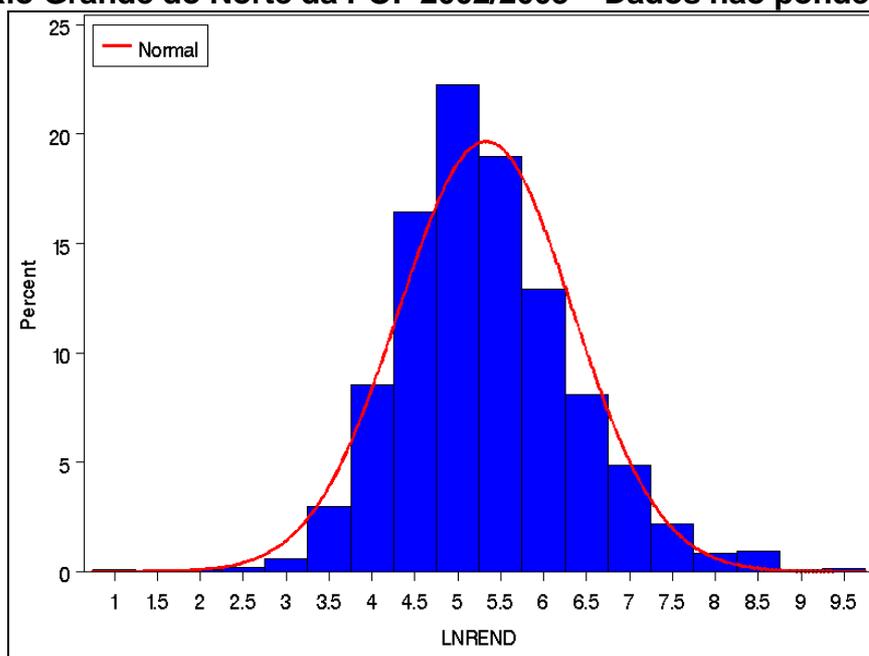
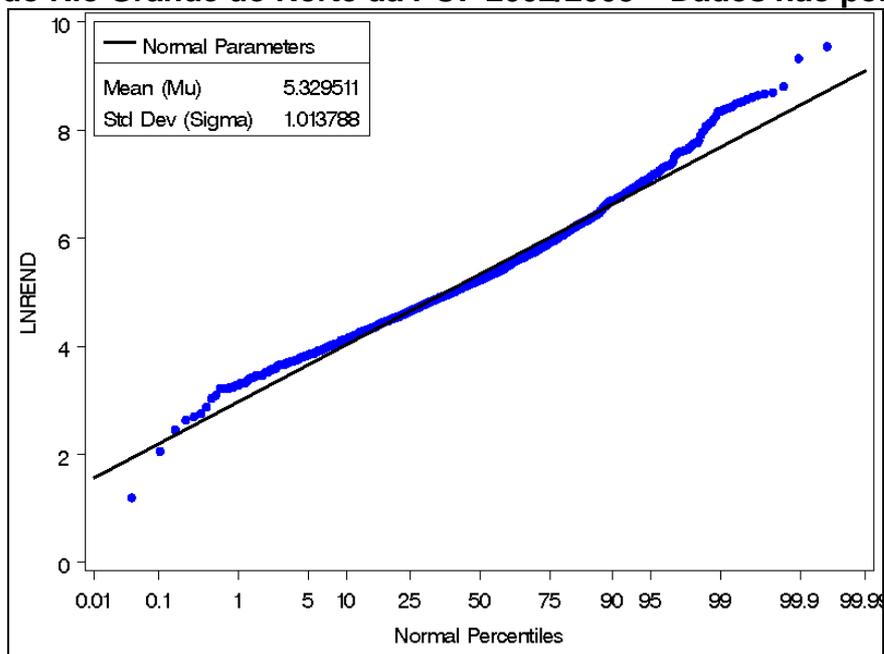


Gráfico 4.4: Gráfico Normal do logaritmo da renda domiciliar per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados



A transformação logarítmica na renda domiciliar per capita parece suficiente para garantir a adequação dos dados a uma distribuição normal. O histograma mostra que existem poucas observações nas caudas, inferior e superior, e forte concentração no meio da distribuição. Esse comportamento dos dados é o que se espera de uma distribuição normal. Observe no gráfico normal acima que o desenho formado pelas observações está muito próximo de uma reta, com exceção de um pequeno desvio na cauda superior, confirmando a adequação da aproximação Normal.

A análise dos gráficos do logaritmo da despesa total domiciliar *per capita*, feita de forma análoga à do logaritmo da renda, também revela que a variável está bem adequada à distribuição normal. Apesar do histograma apresentar uma leve assimetria à direita, o gráfico normal está muito próximo de uma reta.

Gráfico 4.5: Histograma do logaritmo da Despesa total per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados

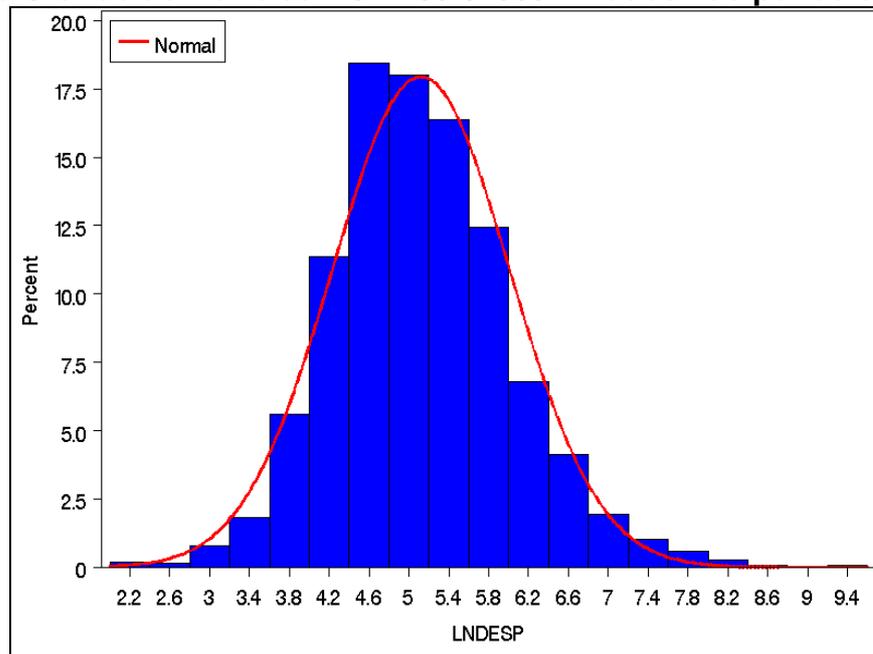
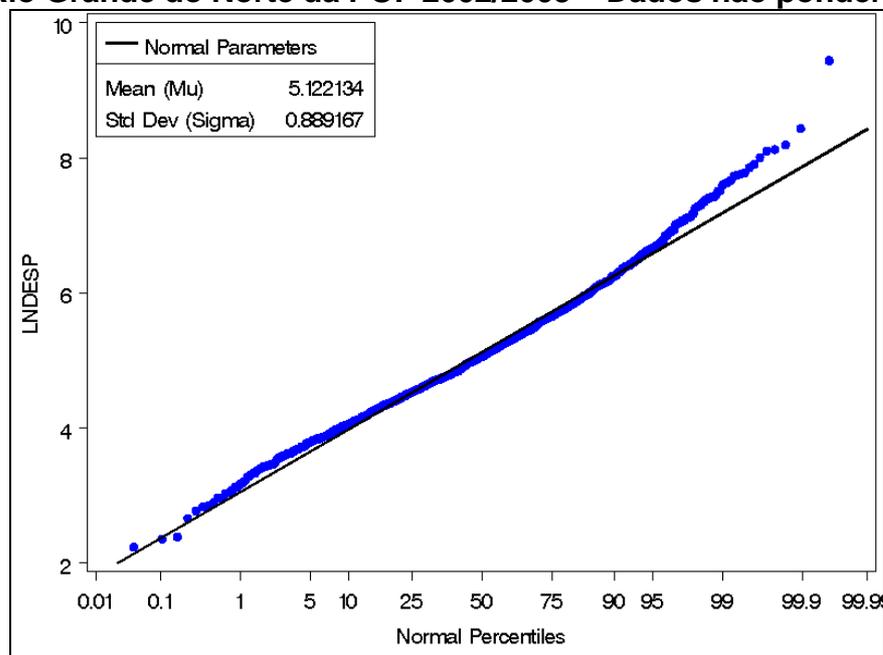


Gráfico 4.6: Gráfico Normal do logaritmo da despesa total per capita na amostra do Rio Grande do Norte da POF 2002/2003 – Dados não ponderados



Mesmo que os histogramas e os gráficos das normais sem considerar o plano amostral tenham se mostrado compatíveis com uma distribuição normal, achou-se

importante o cálculo da densidade dos dados considerando os pesos amostrais da POF.

Observe que o gráfico da densidade do ln da renda domiciliar *per capita* demonstra que os dados estão quase perfeitamente ajustados a uma distribuição normal. Fato que não ocorre para a variável despesa. Note que a cauda inferior não está próxima a zero e é bem menor que a cauda superior. Mesmo assim, a curva está muito parecida com a curva da distribuição normal.

Gráfico 4.7: Densidade do logaritmo da renda domiciliar per capita - Rio Grande do Norte

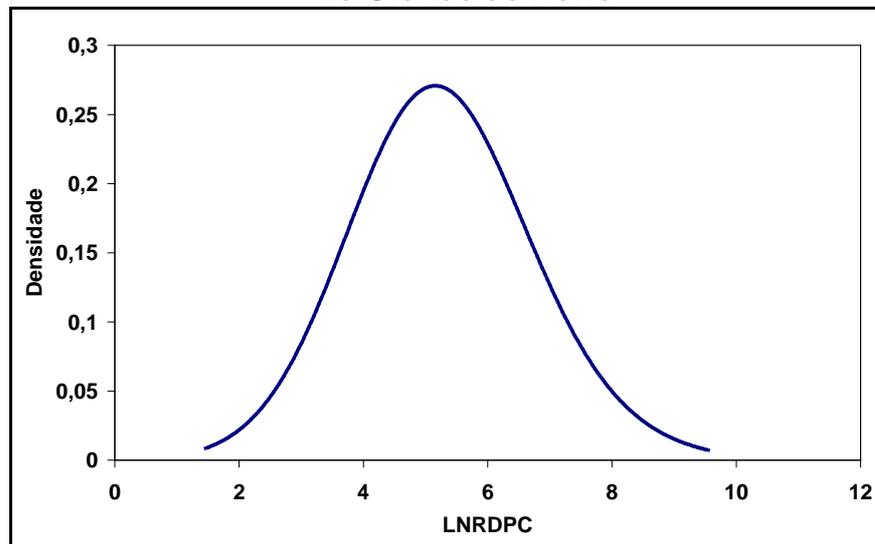
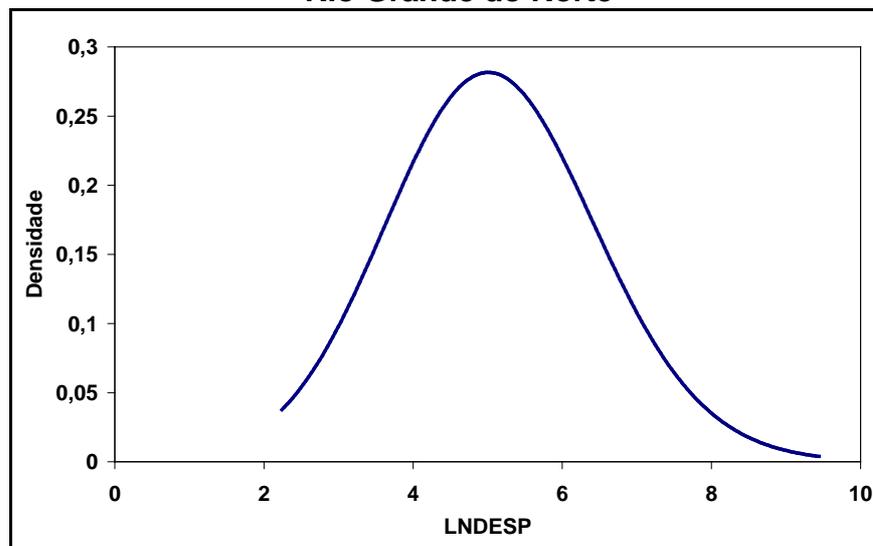


Gráfico 4.8: Densidade do logaritmo da despesa domiciliar per capita - Rio Grande do Norte



O procedimento para cálculo da densidade considerando os pesos amostrais foi realizado através de programação em SAS. Enfim, as transformações feitas nas variáveis permitem que sejam realizados testes estatísticos e que modelos sejam ajustados.

Enfim, as transformações feitas nas variáveis permitem que sejam realizados testes estatísticos e que modelos sejam ajustados nessa escala.

4.3.2. Variáveis preditoras

A aplicação de uma nova metodologia requer a realização de testes para verificar a sua adequação e viabilidade. Muitas vezes o pesquisador fica na expectativa de obter uma resposta positiva nos testes, e pode inconscientemente influenciar nos resultados. Para evitar que esse problema ocorra e, principalmente, para reduzir esforços repetitivos e às vezes desnecessários, achou-se apropriado

identificar quais foram as principais dificuldades encontradas na utilização da metodologia do mapa da pobreza.

O Brasil possui uma grande produção de pesquisas domiciliares e cadastrais, no entanto, cada pesquisa tem objetivos bem definidos fazendo com que as diversas fontes de dados nem sempre sejam exatamente compatíveis entre si. O pesquisador que precisa combinar ou comparar indicadores de duas ou mais fontes, deve estar disposto a lidar com a diferença de definições das variáveis, e o mais importante, deve saber como realizar análises corretas.

Na seção de preparação dos dados (3.3.1), foi exposto que para a elaboração do mapa da pobreza é necessário eleger um rol de variáveis preditoras compatíveis nas duas pesquisas utilizadas. Nesta etapa, já podem ser identificadas algumas dificuldades.

As variáveis são inicialmente selecionadas através da comparação dos questionários das pesquisas, identificando as perguntas comuns. Nesta primeira fase foi possível identificar 31 perguntas comuns, ou seja, que estavam presentes tanto na POF quanto no Censo. Nem sempre as perguntas apresentam as mesmas categorias de respostas, e é comum realizar agregações, para que as variáveis utilizadas tenham conceitos semelhantes. A lista de perguntas comuns às duas pesquisas é apresentada no quadro 4.1.

Quadro 4.1: Lista de perguntas comuns na POF 2002/2003 e Censo 2000

Nome abreviado	Descrição
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	Iluminação Elétrica
PAVIMEN	Pavimentação
AR	Ar condicionado
AUTO	Automóvel
COMPUT	Computador
GELFREE	Geladeira ou freezer
MICROO	Microondas
MQLAVAR	Máquina de lavar
RADIO	Rádio
TV	Televisão
VIDEO	Vídeo cassete
SEX	Sexo da pessoa de referência da UC
COR4	Cor da pessoa de referência da UC
FREQ	Frequência à escola
ESCOLA	Nível de escolaridade da pessoa de referência do domicílio
ANOEST	Anos de estudo da pessoa de referência do domicílio
CONJUGE	Existência de cônjuge
IDOSOS	Número de idosos no domicílio
CRIAN	Número de crianças no domicílio
ADOLES	Número de adolescentes no domicílio
TIPOFAM	Tipo de família
POS	Posição na ocupação da pessoa de referência do domicílio
MORA	Quantidade de moradores no domicílio

Durante a etapa de comparação dos questionários, observou-se que as pesquisas adotam alguns conceitos bastante similares para algumas variáveis, essa singularidade facilita o trabalho do pesquisador. No entanto a comparação dos questionários não é suficiente para considerar a compatibilidade das variáveis. No tópico 3.1 do capítulo que trata da metodologia do mapa da pobreza, foi detalhado o procedimento de teste de hipóteses para comparação das proporções estimadas a

partir das duas bases de dados. Na busca de um modelo com grande poder de explicação, procurou-se um grande rol de variáveis compatíveis na POF e no CENSO, mas na etapa de teste de homogeneidade das proporções, cujo resultado se encontra na anexo 8.1, apenas 20 variáveis podem ser consideradas iguais:

Quadro 4.2: Lista de perguntas comuns consideradas iguais na POF 2002/2003 e Censo 2000

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

Estas variáveis podem seguramente ser utilizadas na etapa seguinte do trabalho, que é a construção de um modelo na POF que posteriormente será utilizado para imputar a variável resposta (renda ou despesa) no Censo. Os demais indicadores não devem ser descartados ainda. Quando o teste de hipótese indica que as proporções são diferentes nas duas pesquisas, ele pode estar apontando para uma evolução natural dos dados. Para avaliar esta hipótese, as PNADs dos anos 2001, 2002 e 2003 foram utilizadas. Além disso, para assegurar a similaridade entre as

pesquisas, foram analisadas algumas instruções dadas às equipes de campo, conceitos adotados pelas pesquisas, e formatos das perguntas. Na tabela 4.9 encontra-se a evolução da distribuição das variáveis preditoras, cujo teste de igualdade entre Censo e POF foi rejeitado.

Tabela 4.9: Evolução da distribuição das variáveis preditoras, cujo teste de igualdade entre Censo e POF foi rejeitado

	CENSO	PNAD2001	PNAD2002	PNAD2003	POF
Total de cômodos servindo de dormitório					
1 dormitório	26.0	29.4	27.3	29.7	27.1
2 dormitórios	46.0	46.3	46.3	44.6	49.6
3 dormitórios	22.3	20.1	21.7	20.4	19.8
4 dormitórios	4.6	3.7	4.2	4.3	2.7
5 dormitórios ou mais	1.1	0.5	0.5	1.1	0.9
Total de cômodos					
Menos que 3 cômodos	12.4	11.9	9.7	10.6	9.0
Quatro cômodos	16.5	17.7	16.4	14.0	14.4
Cinco cômodos	24.3	27.2	24.1	28.6	26.8
Seis cômodos	20.0	19.5	22.0	21.1	22.6
Sete cômodos ou mais	26.8	23.9	27.8	25.6	27.2
Abastecimento de água					
Rede geral	78.6	82.5	83.7	85.3	83.7
Poço ou nascente	8.1	6.5	7.4	4.9	2.6
Outra	13.3	11.0	8.9	9.8	13.7
Existência de banheiro no domicílio					
Tem	90.6	92.7	94.5	94.9	94.0
Não tem	9.4	7.3	5.5	5.1	6.0
Quantidade de banheiros no domicílio*					
Não tem banheiro	9.4		5.6	5.2	6.0
1 banheiro	75.6		79.7	79.5	77.1
2 banheiros	10.3		10.2	10.9	12.5
3 banheiros	3.0		2.9	3.1	3.0
4 banheiros ou mais	1.7		1.7	1.3	1.4
Posse de geladeira ou freezer					
Tem	72.3	73.3	76.5	78.3	79.1
Não tem	27.8	26.7	23.5	21.7	20.9
Posse de rádio					
Tem	81.7	80.1	79.0	80.1	61.5
Não tem	18.3	19.9	21.0	19.9	38.5
Frequenta creche ou escola					
Sim, rede privada	0.7	0.5	0.9	0.7	1.2
Sim, rede pública	3.5	4.2	4.8	5.6	7.1
Não, mas já freqüentou	76.9	77.6	78.7	79.8	69.0
Nunca freqüentou	18.9	17.6	15.7	13.9	22.7
Número de crianças no domicílio					
Nenhuma criança	60.5	62.8	64.4	64.4	64.8
1 criança	25.4	25.5	23.3	25.0	23.2
2 crianças	10.2	8.7	9.5	8.8	9.3
3 crianças	3.1	2.1	2.3	1.1	2.1
4 crianças ou mais	0.9	0.9	0.6	0.8	0.7

Fonte: Censo, PNAD 2001, PNAD 2002, PNAD 2003 e POF 2003/2003

* Informação não disponível na PNAD 2001

Em alguns casos a PNAD serviu para a confirmar a existência de evolução, em outros o resultado se mostrou inconclusivo, com valores muito diferentes nas

pesquisas. Neste caso, surge a suspeita de que as pesquisas captam coisas diferentes. Observe, por exemplo, que a proporção de domicílios que possuem apenas um cômodo servindo de dormitório oscila entre 26% e 29% a cada ano de realização de uma pesquisa. E não é possível afirmar que houve aumento nem redução deste indicador. Ao passo que se o pesquisador conhecesse apenas as informações do Censo e da POF, ele afirmaria que houve aumento de 2% de domicílios com essa característica. Essa falta de clareza nos movimentos dos dados também é seguida pelo total de cômodos. Os dados das PNADs não explicam as diferenças entre Censo e POF nessas duas variáveis através da evolução.

Para explicar essa discordância nas distribuições das duas variáveis que poderiam descrever o tamanho do domicílio, achou-se necessário recorrer aos conceitos adotados em cada pesquisa.

Foram disponibilizadas, em diversos quadros a seguir, as definições de cada variável constantes nas notas metodológicas, bem como as instruções dadas aos agentes de pesquisa ou recenseadores, incluídas nos manuais de entrevista.

Quadro 4.3: Definição, instrução e pergunta do total de cômodos e total de cômodos servindo de dormitório

	POF	CENSO	PNAD
Total de Cômodos			
Definição e instrução	Você deverá considerar como cômodo: todo compartimento do domicílio limitado por paredes, inclusive banheiro e cozinha; e os existentes na parte externa do prédio que sejam parte integrante do domicílio. Não considere como cômodo os: corredores, alpendres, áreas de serviço, terraços e varandas abertas, e garagem, depósito e outros compartimentos utilizados para fins não residenciais.	Cômodos são todos os compartimentos integrantes do domicílio separados por paredes, inclusive banheiros e cozinha, e os existentes na parte externa do prédio, desde que constituam parte integrante do domicílio. Não são considerados os corredores, alpendres, varandas abertas e outros compartimentos utilizados para fins não-residenciais como garagens, depósitos etc.	Considerou-se como cômodo todo compartimento, coberto por um teto e limitado por paredes, que fosse parte integrante do domicílio particular permanente, com exceção de corredor, alpendre, varanda aberta, garagem, depósito e outros compartimentos utilizados para fins não-residenciais.
Pergunta	Cômodos: Total	Quantos cômodos existem neste domicílio?	Quantos cômodos tem este domicílio?
Total de Cômodos servindo de dormitório			
Definição e Instrução	Registre, com dois dígitos, o número de cômodos que estiverem servindo de dormitório aos moradores do domicílio, em caráter permanente. Você também deve incluir neste item: os cômodos que estejam servindo de dormitório, permanentemente, por falta de acomodações adequadas a este fim; e os cômodos situados na parte externa do prédio que sejam usados, permanentemente, como dormitório, por moradores do domicílio. Não considere como servindo de dormitório os cômodos utilizados, somente, como: quarto de vestir, de hóspede, de costura, escritório, etc.	O número de dormitórios corresponde ao total de cômodos integrantes do domicílio que estavam servindo, em caráter permanente, de dormitório para os moradores, nele incluídos aqueles que assim são utilizados em função de não haver acomodação adequada para esta finalidade.	Considerou-se como dormitório o cômodo que estivesse, em caráter permanente, sendo utilizado para esta finalidade por morador do domicílio particular permanente.
Pergunta	Cômodos: Servindo de dormitório	Quantos cômodos servem de dormitórios para os moradores deste domicílio?	Quantos Cômodos estão servindo permanentemente de dormitório para os moradores deste domicílio?

Observe que as definições tanto de cômodo quanto de cômodos servindo de dormitório são muito parecidas nas três pesquisas, e somente o formato da pergunta na POF é diferente das demais. Uma vez que houve consonância entre as definições, não é possível explicar as diferenças nas distribuições. Além disso, a PNAD não foi suficiente para assegurar a hipótese da evolução dos indicadores dessas duas variáveis. Portanto, elas não devem constar na etapa seguinte do trabalho.

No tocante ao abastecimento de água, a POF apresenta um maior percentual de domicílios ligados à rede geral de abastecimento de água que o Censo. Os dados das PNADs revelam que pode sim ter ocorrido aumento no percentual de domicílios ligados à rede geral entre os anos de 2000 e 2003, e conseqüentemente, uma redução no percentual de domicílios com poço ou nascente.

Mesmo que os dados apresentados tenham mostrado uma certa lógica, achou-se útil disponibilizar os conceitos referentes ao abastecimento de água adotados pelas pesquisas.

Observe no quadro 4.4 que a grande diferença no conceito de abastecimento de água ocorre no tipo de canalização. Na POF, apenas uma pergunta é feita sobre o abastecimento de água, e o tipo de canalização está incluído nas categorias de resposta, fato que não ocorre no Censo nem na PNAD, onde são feitas duas perguntas: uma para o tipo de abastecimento e outra para a canalização. Neste trabalho não foram analisados nem o tipo nem a existência de canalização, então, para haver comparabilidade, na POF as categorias foram agregadas em rede geral, poço ou nascente e outro tipo de abastecimento, independente da canalização. Na PNAD e no Censo o tipo de canalização foi ignorado, pois só interessava a forma de abastecimento de água.

A variável de abastecimento de água é um excelente exemplo da importância de examinar em detalhe os conceitos adotados nas pesquisas. Mesmo que as perguntas tenham formatos diferentes, talvez a singularidade dos conceitos adotados tenha sido relevante para que fosse possível identificar o movimento de crescimento de uma determinada forma de abastecimento de água.

Quadro 4.4: Definição, instrução e pergunta do abastecimento de água

	POF	CENSO	PNAD
Abastecimento de água			
Definição e Instrução	Assinale com um X a quadrícula correspondente à forma de abastecimento de água do domicílio. Rede Geral com Canalização Interna: quando o domicílio for servido de água canalizada proveniente de rede geral de abastecimento, com distribuição interna para um ou mais cômodos. Poço ou Nascente com Canalização Interna (na propriedade): quando o domicílio for servido de água canalizada ligada a poço ou nascente, com distribuição interna para um ou mais cômodos. Outra Forma com Canalização Interna: quando o domicílio tiver distribuição interna de água, mas o reservatório ou caixa d'água for abastecido por carro-pipa, coleta de chuva, etc.	Classificação quanto à origem da água que abastece o domicílio. Rede geral: quando o domicílio, o terreno, ou a propriedade onde ele está localizado é servido de água ligada à rede geral de abastecimento. Poço ou nascente: quando o domicílio é servido por poço ou nascente localizados no terreno ou na propriedade onde está construído. Outra: quando o domicílio é servido de água de reservatório (ou caixa), abastecido com água das chuvas, por carro-pipa ou, ainda, por poço ou nascente localizados fora do terreno ou da propriedade onde está construído. Branco: para domicílio particular improvisado e domicílio coletivo.	A proveniência da água utilizada nos domicílios foi classificada em: Rede geral: Quando o domicílio fosse servido por água proveniente de uma rede geral de distribuição, com canalização interna ou, pelo menos, para o terreno ou propriedade em que se situava; Poço ou nascente: Quando o domicílio fosse servido por água, com canalização interna, proveniente de poço ou nascente ou, sem canalização interna, proveniente de poço ou nascente localizado no terreno ou na propriedade em que se situava; ou Outra proveniência: Quando o domicílio fosse servido por água proveniente de reservatório abastecido por carro-pipa, coleta de chuva ou outra procedência que não se enquadrasse nas anteriormente descritas.
	Rede Geral sem Canalização Interna: quando o domicílio for servido de água proveniente de uma rede geral canalizada para a propriedade, sem haver distribuição interna. Poço ou Nascente sem Canalização Interna (na propriedade): quando o domicílio for servido de água proveniente de poço ou nascente próprios, sem distribuição interna. Outra Forma sem Canalização Interna: quando a água utilizada no domicílio for apanhada em fonte pública, poço, bica, etc., localizados fora da propriedade, e não houver distribuição interna.	Classificação quanto à existência de canalização para a distribuição de água no domicílio. Canalizada em pelo menos um cômodo: quando o domicílio é servido de água canalizada com distribuição interna para um ou mais cômodos. Canalizada só na propriedade ou terreno: quando a água utilizada no domicílio chega à propriedade, sem haver distribuição interna. Não canalizada: quando a água utilizada no domicílio não chega canalizada à propriedade ou terreno e não tem distribuição interna. Branco: para domicílio particular improvisado e domicílio coletivo.	Investigou-se a existência de água canalizada nos domicílios particulares permanentes e a sua proveniência. Quanto à existência de água canalizada, os domicílios foram classificados em: Com canalização interna: Para o domicílio que tivesse água canalizada para, pelo menos, um cômodo; ou Sem canalização interna: Para o domicílio que não tivesse água canalizada para nenhum cômodo.
Pergunta	Abastecimento de água Com canalização interna: 1 - rede geral; 2 - poço ou nascente (na propriedade); 3 - outra forma. Sem canalização interna: 4 - rede geral; 5 - poço ou nascente (na propriedade); 6 - outra forma.	A forma de abastecimento de água utilizada neste domicílio é: 1 - rede geral; 2 - poço ou nascente (na propriedade); 3 - outra forma. A água utilizada neste domicílio chega: 1 - canalizada em pelo menos um cômodo; 2 - canalizada só na propriedade ou terreno; 3 - não canalizada.	Este domicílio tem água canalizada para, pelo menos, um cômodo? 1 - sim; 3 - não. A água utilizada neste domicílio é proveniente de: 2 - rede geral de distribuição; 4 - poço ou nascente; 6 - outra proveniência. A água utilizada neste domicílio é canalizada de rede geral de distribuição para a propriedade? 1 - sim; 3 - não. A água utilizada neste domicílio é de poço ou nascente localizado na propriedade? 2 - sim; 4 - não.

No tocante ao inventário de bens, observa-se na tabela 4.9 que o indicador de posse de geladeira ou freezer pode ser aproveitado no modelo, pois se nota claramente que o número de domicílios que possuem esse tipo de bem aumentou no período compreendido entre o Censo e a POF (2000 e 2003).

O mesmo não pode ser afirmado para a posse de rádio. Observe que o percentual de domicílios que não possuem o bem é mais do que o dobro na POF, comparado às demais fontes. Pode ser notado no quadro 4.5, que existe algumas

diferenças na forma da pergunta. Justamente a POF, que apresentou valor tão discrepante para a posse de rádio, não apresenta a pergunta de posse do bem, passando diretamente para a pergunta da quantidade do bem, de forma que é necessário considerar que o domicílio que registrou quantidade zero de determinado bem, não possui o tal bem. A lógica da construção do indicador pode parecer muito simples e até mesmo intuitiva, mas a forma diferente de construir o indicador na POF pode ser a responsável pela diferença dos resultados.

A disposição das perguntas nos questionários é outro fator que pode ser determinante para explicar tal diferença. Os questionários da POF e do Censo apresentam um bloco específico para captar o inventário de bens, e nos manuais de entrevista possuem instruções gerais para o bloco como um todo. O Censo apresenta ainda instruções para cada bem. Já na PNAD, as perguntas sobre inventário de bens são feitas dentro do conceito de domicílio, e as instruções aos entrevistadores são dadas exclusivamente para cada bem, não havendo uma instrução geral sobre o inventário de bens. Além disso, a PNAD pergunta separadamente a existência de geladeira e de freezer, e foi necessário juntar as duas variáveis para poder comparar com as outras fontes.

Quadro 4.5: Definição, instrução e pergunta do inventário de bens

	POF	CENSO	PNAD
Inventário de bens			
	Destina-se a pesquisar, dentre os bens duráveis relacionados, aqueles que a Unidade de Consumo dispõe no domicílio principal, que estejam em uso ou aguardando reparo por ocasião da pesquisa. Registre também neste Quadro, os bens: de propriedade do locador, existentes no domicílio principal, no caso de o mesmo ser alugado; e existentes no domicílio principal e que tenham sido emprestados à Unidade de Consumo.	Deverão ser considerados apenas os aparelhos que estiverem em condições de uso, independente de serem próprios, cedidos ou alugados.	
Posse de geladeira ou freezer			
Definição e Instrução	Registre, na linha correspondente, a quantidade de cada tipo de bem pesquisado de que dispõe a Unidade de Consumo.	Assinale conforme o caso: 1 - SIM - quando houver geladeira (de uma ou mais portas) ou freezer, no domicílio, mesmo que seja a gás ou querosene; 2 - NÃO	Geladeira: Nos domicílios particulares permanentes, pesquisou-se a existência de geladeira de duas portas (ou seja, o aparelho que acopla dois compartimentos independentes, sendo um de refrigeração e o outro de congelamento de alimentos) e para os que não tivessem este tipo de aparelho, investigou-se a existência de geladeira de uma porta. Freezer: Nos domicílios particulares permanentes, pesquisou-se a existência de freezer
Pergunta	Inventário de bens duráveis do domicílio principal: Quantidade	Neste domicílio existe: geladeira ou freezer?	Este domicílio tem geladeira? Este domicílio tem freezer?
Rádio			
Definição e Instrução	Registre, na linha correspondente, a quantidade de cada tipo de bem pesquisado de que dispõe a Unidade de Consumo.	Assinale conforme o caso: 1 - SIM - quando no domicílio houver qualquer tipo de aparelho de rádio (inclusive a pilha ou acoplado a outros aparelhos); 2 - NÃO	Nos domicílios particulares permanentes, pesquisou-se a existência de rádio, mesmo que fizesse parte de conjunto que acoplasse outros aparelhos, tais como: rádio-gravador, rádio toca-fitas etc.
Pergunta	Inventário de bens duráveis do domicílio principal: Quantidade	Neste domicílio existe: rádio?	Este domicílio tem rádio?

Outro indicador que apresenta grande diferença entre as pesquisas é a quantidade de banheiros no domicílio. Esta diferença ocorre principalmente no percentual de domicílios que não possuem banheiro ou que possuem apenas um. E a pesquisa em que os dados se apresentaram diferentes foi o Censo, pois POF e PNAD mostram distribuições muito parecidas para esta variável. O Censo apresentou um percentual de domicílios sem banheiros (9,4%) mais elevado que a POF (6%) e a PNAD (cerca de 5,5%). Conseqüentemente, o percentual de domicílios que possuem apenas um banheiro é menor no Censo do que nas demais pesquisas. Além disso, a análise da evolução não mostrou uma tendência clara. Nas demais categorias da variável não houve discrepância, o que pode significar que não há divergência na

distribuição, mas sim dos conceitos adotados principalmente na existência de banheiro.

Em nenhuma das pesquisas utilizadas é perguntado se existe banheiro no domicílio, partindo-se direto para a pergunta sobre a quantidade de banheiros no domicílio. Tanto na POF quanto na PNAD o conceito de banheiro engloba o de sanitário, que segundo o Censo vem a ser o “cômodo ou o local limitado por paredes de qualquer material, coberto ou não por um teto, que disponha de aparelho sanitário ou buraco para dejeções”²². Já no CENSO, é pesquisada adicionalmente a existência do sanitário, mas não é questionada a quantidade. Além disso, a não existência de banheiro na POF foi definida a partir do tipo de escoamento sanitário, ou seja, quando o domicílio não possui escoamento, considerou-se que também não possui banheiro. Para que esta diferença de conceito fique mais evidente, foram disponibilizadas abaixo as orientações dadas aos recenseadores e agentes de pesquisa sobre a definição da quantidade de banheiros.

²² Manual do Recenseador, Censo Demográfico, Rio de Janeiro: 2000.

Quadro 4.6: Definição, instrução e pergunta da quantidade de banheiros e existência de sanitários

	POF	CENSO	PNAD
Quantidade de banheiros no domicílio			
Definição e Instrução	Registre, com dois dígitos, o total de banheiros localizados dentro da propriedade, que estiverem servindo como tal aos moradores do domicílio, em caráter permanente.	Considere, como banheiro, o cômodo que dispõe de chuveiro ou banheira e aparelho sanitário. Registre o número total de banheiros de uso exclusivo dos moradores do domicílio. Se existirem 9 (nove) banheiros ou mais, registre 9 (nove). Se não existir banheiro, registre 0 (zero).	Pesquisou-se, também, o número de banheiros ou sanitários de uso exclusivo dos moradores em domicílios particulares permanentes. Considerou-se como banheiro o cômodo destinado a banho e que também dispusesse de vaso sanitário ou buraco para dejeções. Considerou-se como sanitário o cômodo ou o local limitado por paredes de qualquer material, coberto, ou não, por um teto e que dispusesse de vaso sanitário ou buraco para dejeções.
Pergunta	Cômodos: Total de banheiros	Quantos banheiros existem neste domicílio?	Quantos banheiros ou sanitários tem este domicílio?
Existência de sanitário no domicílio			
Definição e Instrução		Considere, como sanitário, o cômodo ou o local limitado por paredes de qualquer material, coberto ou não por um teto, que disponha de aparelho sanitário ou buraco para dejeções. Assinale conforme o caso: 1 - SIM - quando no domicílio, no terreno ou propriedade em que se localiza, existir sanitário para uso de seus moradores, comum ou não a mais de um domicílio; 2 - NÃO - quando no domicílio, no terreno ou propriedade em que se localiza, não existir sanitário para uso dos seus moradores.	
Pergunta		Neste domicílio, terreno ou propriedade existe sanitário utilizado pelos moradores?	

É evidente que não dá para considerar a igualdade desses indicadores através da POF e do Censo, quando a existência de banheiros compõe uma categoria isolada. O conceito de existência de banheiro difere nas pesquisas. Entretanto, quando os modelos foram ajustados, observou-se que essa variável tinha um poder de predição muito alto tanto para a renda quanto para a despesa. Como a diferença ocorria apenas nas duas primeiras categorias, optou-se por agregar essas categorias e realizar o teste novamente. Então a variável agora tem as categorias 0 ou um banheiro, 2 banheiros, 3 banheiros, e 4 banheiros ou mais, e o resultado do teste foi satisfatório, ou seja, as proporções na POF e no censo podem ser consideradas iguais. Conseqüentemente a variável de quantidade de banheiros, com essa nova agregação, pode compor o modelo.

O teste para o número de crianças no domicílio foi rejeitado porque a POF apresentou um percentual de domicílios sem crianças superior ao do Censo. As PNADs mostram que pode estar ocorrendo um aumento de domicílios, onde não há crianças.

Por fim, as duas últimas variáveis que resultaram em diferença entre os indicadores da POF e Censo, se referem à educação da pessoa responsável pela família, no caso do Censo e da PNAD, e da pessoa de referência da unidade de consumo, na POF.

A variável de frequência à creche ou escola não parece mostrar indícios de evolução no período entre a POF e o Censo. Aliás, a PNAD em nada contribuiu para a avaliação desta variável, pois, enquanto a PNAD parece mostrar um crescimento da proporção de responsáveis que não freqüentam escolas mas já freqüentaram, a POF apresenta valor inferior. O contrário ocorre com a proporção dos que nunca freqüentaram escola: PNAD mostrando queda e POF, aumento.

Já no que se refere ao grau de instrução, os dados da PNAD serviram para mostrar que pode ter havido um aumento de escolarização dos responsáveis pela família. Observe que a proporção dos responsáveis com ensino fundamental sofreu uma leve queda, e a proporção dos que concluíram o ensino médio apresentou um pequeno aumento. Houve também um aumento na POF na proporção dos tiveram instrução ao nível de creche, pré-escolar e classe de alfabetização, mas não foi acompanhado pelos indicadores da PNAD. Mesmo assim, pode considerar que os dados de grau de instrução apresentaram evolução, e pode ser incluído na etapa seguinte do processo de mapeamento da pobreza.

As definições e instruções das variáveis de frequência à escola e grau de instrução foram suprimidas desta dissertação, por se tratar de um texto muito extenso.

Mas quaisquer informações sobre conceitos e instruções nas três pesquisas abordadas aqui são facilmente encontradas na *homepage* do IBGE.

A PNAD não pergunta a posse de vídeo nem de automóvel, por isso não foi possível analisar a evolução dessas variáveis. Como o teste de hipóteses revelou diferença entre as distribuições dessas variáveis entre Censo e POF, optou-se por não incluí-las na etapa de modelagem. Entretanto, apenas para diminuir a curiosidade da autora, todas as variáveis foram testadas no modelo e constatou-se que algumas variáveis que não poderiam compor o modelo, por não terem as mesmas distribuições e não apresentarem quadro de evolução, apresentaram grande poder de predição tanto para o logaritmo da renda *per capita* quanto para o logaritmo da despesa *per capita*. De certa forma é lamentável que essas variáveis não possam ser incluídas do modelo final, pois elas acrescentariam muito no poder de predição de ambas as variáveis respostas.

A lista de variáveis que devem ser testadas para inclusão no modelo pode ser encontrada no quadro 4.7, onde também estão apresentados tanto os códigos como os nomes completos das variáveis e de suas respectivas categorias.

Quadro 4.7: Descrição das variáveis que poderão ser utilizadas no ajuste de modelos (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

Quadro 4.7: Descrição das variáveis que poderão ser utilizadas no ajuste de modelos (continuação)

Variáveis	Categorias
IDOSOS 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
ADOLES 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
CRIAN 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
TIPOFAM 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
COR4 Cor do chefe da UC	1 Branca 2 Não branca
SEX Sexo do chefe da UC	1 Masculino 2 Feminino
ANOEST 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
POS 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

4.3.3. A renda no censo

Depois de analisar as variáveis de renda e despesa da POF, que são as variáveis respostas do modelo a ser ajustado a seguir, achou-se necessário realizar uma rápida avaliação da renda no CENSO. Como não será ajustado nenhum modelo baseado na renda do censo, não é necessário verificar a adequação de sua

distribuição ao modelo normal, como feito na seção 4.3.1 para as variáveis de renda e despesa da POF. Esta parte da dissertação destina-se apenas a fazer uma breve análise da renda, com a finalidade de comparar resultados obtidos posteriormente, além de fornecer uma base para conclusões futuras.

Observe na tabela 4.8 que em 2000 o Rio Grande do Norte apresentava uma renda domiciliar per capita de R\$ 231,00. O valor da média é bem superior ao da mediana (R\$ 99,06), indicando que a distribuição da renda domiciliar per capita no Rio Grande do Norte apresenta forte assimetria à direita, com muitas observações na parte inferior da distribuição. Isso também pode ser verificado pela grande diferença entre os valores do primeiro e terceiro quartis (R\$ 49,53 e R\$199,44, respectivamente).

Tabela 4.10: Medidas resumo da renda domiciliar per capita

Renda per capita	
Média	231.72
Mínimo	0.00
1º quartil	49.53
Mediana	99.06
3º quartil	199.44
Máximo	50,082.90
Desvio padrão	3,527.75

Fonte: Censo 2000

Nessa distribuição, a média ultrapassou também o valor do terceiro quartil. Isso ocorre devido ao fato de se ter na distribuição alguns valores extremamente elevados. Esses altos valores poderiam puxar a média para cima, fazendo com que ela não represente um bom indicador para avaliação de pobreza.

4.3.4. Ajuste de modelos

O método usado para a seleção do modelo foi o de seleção passo a passo para frente descrito na seção 3.5. Primeiramente são ajustados modelos com apenas uma variável explicativa, e a primeira variável a ser adicionada no modelo é aquela que apresenta maior estatística F, cuja expressão é dada pela fórmula 3.8 do capítulo 3. Sendo que este F tem que ser maior que um nível predeterminado (4).

A variável que mais contribui para a predição do logaritmo da renda *per capita* e do logaritmo da despesa *per capita* é a posse de computador. Esta é a primeira variável adicionada nos dois modelos.

Em seguida foram ajustados todos os modelos com duas variáveis, sendo que uma delas é a posse de computador. As variáveis são novamente avaliadas quanto ao poder de predição, sendo que a partir de agora a estatística de teste é a definida pela expressão (3.9). Nesta etapa, a variável com o segundo maior poder de explicação para o modelo da renda é diferente da do modelo para a despesa. No primeiro modelo a segunda variável a ser adicionada foi a classe de anos de estudo (ANOEST), medida através das categorias: sem instrução, 1 a 3 anos, 4 a 7 anos, 8 a 10 anos, 11 a 14 anos e 14 anos ou mais. Já no modelo do *ln* da despesa, a segunda variável a entrar no modelo foi a quantidade de moradores no domicílio (MORA, com as categorias: 1, 2, 3, 4 e 5 moradores ou mais).

O procedimento foi repetido até que a estatística obtida pela expressão 3.9 fosse inferior a 4. Nas tabelas 4.11 e 4.12 são apresentados os valores da estatística de teste F^* correspondentes ao ajuste de cada modelo. Nota-se que o modelo para o *ln* da renda contém 19 variáveis, e para o *ln* da despesa, 14.

Tabela 4.11: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da renda per capita

Nº	Variável adicionada	F-parcial
1	COMPUT	341,7
2	ANOEST	327,4
3	MORA	279,3
4	QTDBANH	171,3
5	GELFREE	195,2
6	MQLAVAR	68,7
7	IDOSOS	52,3
8	POS	38,5
9	ADOLES	34,5
10	CRIAN	41,6
11	AR	31,7
12	MICROO	20,5
13	ABASTEC	10,7
14	CONDDOM	9,6
15	TIPOFAM	10,8
16	TIPOESC	7,2
17	COR4	5,7
18	PAVIMEN	5,2
19	SEX	4,8

Fonte: POF 2002/2003

Tabela 4.12: Variável adicionada e F-parciais de cada modelo com efeitos principais ajustado para o logaritmo da despesa

Nº	Variável adicionada	F-parcial
1	COMPUT	331,7
2	MORA	397,3
3	QTDBANH	287,9
4	GELFREE	213,3
5	ANOEST	111,2
6	MQLAVAR	41,6
7	ADOLES	28,6
8	CRIAN	41,6
9	MICROO	26,9
10	AR	19,4
11	TIPOESC	15,3
12	CONDDOM	16,3
13	IDOSOS	15,2
14	ABASTEC	14,0

Fonte: POF 2002/2003

Observe que todas as 14 variáveis que compõem o modelo do ln da despesa estão contidas no modelo do ln da renda. Então optou-se por trabalhar com o mesmo

conjunto de variáveis nos dois modelos, pois a inclusão de uma nova variável não diminui o poder de predição do modelo, além disso facilita a elaboração dos próximos procedimentos.

A tabela 4.13 apresenta as estimativas dos betas e dos p-valores dos dois modelos de efeitos principais.

Tabela 4.13: Estimativas dos betas e dos p-valores dos modelos de efeitos principais

Parâmetro	β	LNDESP		LNREND	
		Betas	P-valor	Betas	P-valor
Intercept	β_0	4,759	0,000	4,485	0,000
COMPUT 1	β_1	0,455	0,000	0,372	0,000
ANOEST 1	β_2	-0,365	0,015	-0,273	0,171
ANOEST 2	β_3	-0,249	0,071	-0,230	0,231
ANOEST 3	β_4	-0,194	0,145	-0,186	0,311
ANOEST 4	β_5	0,025	0,855	0,016	0,932
ANOEST 5	β_6	0,047	0,744	0,299	0,143
ANOEST 6	β_7	0,197	0,253	0,673	0,007
MORA 1	β_8	1,177	0,000	1,030	0,000
MORA 2	β_9	0,582	0,000	0,412	0,000
MORA 3	β_{10}	0,369	0,000	0,255	0,000
MORA 4	β_{11}	0,156	0,002	0,077	0,112
QTDBANH 1	β_{12}	-0,917	0,000	-0,836	0,001
QTDBANH 2	β_{13}	-0,650	0,000	-0,553	0,000
QTDBANH 3	β_{14}	-0,331	0,035	-0,230	0,058
QTDBANH 4	β_{15}	-0,268	0,120	-0,283	0,057
GELFREE 1	β_{16}	0,371	0,000	0,474	0,000
MQLAVAR 1	β_{17}	0,200	0,000	0,296	0,000
IDOSOS 1	β_{18}	-0,680	0,023	-0,477	0,064
IDOSOS 2	β_{19}	-0,699	0,017	-0,263	0,321
IDOSOS 3	β_{20}	-0,597	0,051	-0,222	0,392
POS 1	β_{21}	-0,057	0,650	-0,119	0,305
POS 2	β_{22}	0,044	0,417	-0,023	0,721
POS 3	β_{23}	0,095	0,425	0,410	0,002
POS 4	β_{24}	0,035	0,488	-0,085	0,193
POS 5	β_{25}	-0,082	0,746	0,360	0,064
ADOLES 1	β_{26}	0,393	0,000	0,573	0,000
ADOLES 2	β_{27}	0,309	0,002	0,429	0,003
ADOLES 3	β_{28}	0,140	0,186	0,227	0,087
ADOLES 4	β_{29}	0,000	0,999	0,257	0,060
CRIAN 1	β_{30}	0,669	0,002	0,371	0,052
CRIAN 2	β_{31}	0,527	0,009	0,227	0,233
CRIAN 3	β_{32}	0,489	0,025	0,132	0,495
CRIAN 4	β_{33}	0,291	0,230	-0,108	0,594
AR 1	β_{34}	0,360	0,000	0,449	0,000
MICROO 1	β_{35}	0,363	0,000	0,394	0,000
ABASTEC 1	β_{36}	0,105	0,091	0,161	0,042
ABASTEC 2	β_{37}	-0,093	0,276	0,345	0,001
CONDDOM 1	β_{38}	0,160	0,003	0,148	0,010
CONDDOM 2	β_{39}	0,029	0,654	0,021	0,738
TIPOFAM 2	β_{40}	0,090	0,284	0,163	0,156
TIPOFAM 3	β_{41}	0,121	0,197	0,096	0,362
TIPOFAM 4	β_{42}	0,030	0,821	-0,087	0,586
TIPOFAM 5	β_{43}	0,114	0,221	0,175	0,057
TIPOESC 1	β_{44}	-0,137	0,248	-0,138	0,451
TIPOESC 2	β_{45}	-0,254	0,021	-0,214	0,237
TIPOESC 3	β_{46}	-0,415	0,022	-0,459	0,011
COR4 1	β_{47}	0,027	0,408	0,083	0,020
PAVIMEN 1	β_{48}	0,038	0,367	0,085	0,029
SEX 1	β_{49}	0,000	0,995	0,149	0,044

Fonte: POF 2002/2003

O passo seguinte é testar algumas interações. Foram testadas 3 interações: COMPUT*ANOEST, MORA*QTDBANH E SEX*ANOEST. As três interações foram significativas para o modelo da LNREND, embora a interação SEX*ANOEST não seja significativa no modelo para LNDESP. Mesmo assim todas as interações serão mantidas em ambos os modelos, pelo mesmo motivo dos efeitos principais.

Tabela 4.14: Teste do efeito das variáveis dos modelos com interações

Variável	LNREND			LNDESP		
	GL*	Valor F	P-valor	GL*	Valor F	P-valor
Modelo	75	781,530	<.0001	75	834,500	<.0001
Intercepto	1	2487,500	<.0001	1	1735,260	<.0001
COMPUT	1	36,680	<.0001	1	56,810	<.0001
ANOEST	6	8,140	<.0001	6	4,740	0,000
MORA	3	6,510	0,000	3	14,050	<.0001
QTDBANH	4	8,710	<.0001	4	17,040	<.0001
GELFREE	1	66,560	<.0001	1	47,220	<.0001
MQLAVAR	1	28,510	<.0001	1	16,080	0,000
IDOSOS	3	5,350	0,002	3	2,440	0,068
POS	5	5,680	<.0001	5	0,460	0,806
ADOLESC	4	8,810	<.0001	4	10,180	<.0001
CRIAN	4	8,000	<.0001	4	7,040	<.0001
AR	1	29,910	<.0001	1	23,880	<.0001
MICROO	1	22,480	<.0001	1	26,060	<.0001
ABASTEC	2	7,240	0,001	2	2,950	0,057
CONDDOM	2	4,600	0,012	2	7,600	0,001
TIPOFAM	4	2,360	0,057	4	0,830	0,512
TIPOESC	3	2,900	0,038	3	5,010	0,003
COR4	1	4,920	0,029	1	0,790	0,377
PAVIMEN	1	4,820	0,030	1	1,060	0,306
SEX	1	0,360	0,549	1	0,460	0,499
COMPUT*ANOEST	5	5,350	0,000	5	3,450	0,006
ANOEST*SEX	6	2,500	0,026	6	0,700	0,652
MORA*QTDBANH	15	3,110	0,000	15	2,020	0,019

Fonte: POF 2002/2003

*GL = Graus de liberdade

4.4. Avaliação de qualidade do ajuste

4.4.1. Validação do modelo para a área rural

Uma vez definido o modelo que deverá ser usado na etapa de imputação da renda no censo, é necessária a avaliação da qualidade do ajuste deste modelo. A primeira avaliação da qualidade do modelo é feita a partir de uma análise da validação do modelo escolhido para as duas áreas geográficas: Urbana e Rural. O teste de validação se faz imprescindível devido a uma característica no desenho da amostra da POF 2002/2003. Sabe-se que os estratos de seleção das áreas rurais não foram definidos para cada Unidade da Federação, no entanto, alguns domicílios da amostra se localizam em estratos rurais de seleção. Em outras palavras, dos 1548 domicílios que compõem a amostra da POF 2002/2003, 261 se localizam em áreas rurais. Estes domicílios foram selecionados para compor o estrato rural da região Nordeste, e a utilização destes domicílios na análise do Rio Grande do Norte pode estar errada. Mas a simples retirada dos 261 domicílios da amostra pode resultar numa grande perda de informação, visto que eles representam cerca de 17% de todos os domicílios pesquisados no estado. Então se optou por realizar um teste de validação do modelo ajustado para a área rural e urbana. De modo que se o R^2 do modelo da área rural fosse muito diferente do R^2 para a área urbana, os dados referentes à área rural seriam excluídos.

A estatística de teste é dada pela equação:

$$R_j^2 = \frac{\sum_{i \in j} e_i^2}{\sum_{i \in j} z_i^2},$$

sendo que i representa o número da observação ($i=1, 2, \dots, 1548$), j representa o tipo de área (urbana ou rural), $e_i = \hat{y}_i - y_i$ e $z_i = y_i - \bar{y}$, onde y_i é o valor observado da variável resposta do domicílio i , \hat{y}_i é o valor ajustado da variável resposta do domicílio i , a partir do modelo ajustado, e \bar{y} é a média dos valores observados. Então as estatísticas R^2 para cada modelo ajustado são apresentadas na tabela a seguir.

Tabela 4.15: R^2 dos modelos de LNREND e LNDESP por tipo de área

	LNREND	LNDESP
AREA URBANA	0.390	0.405
AREA RURAL	0.395	0.389

Fonte: POF 2002/2003

Pode-se observar que os R^2 dos modelos para as áreas rurais não são muito diferentes que os das áreas urbanas, então não é necessária a exclusão dos dados das áreas rurais.

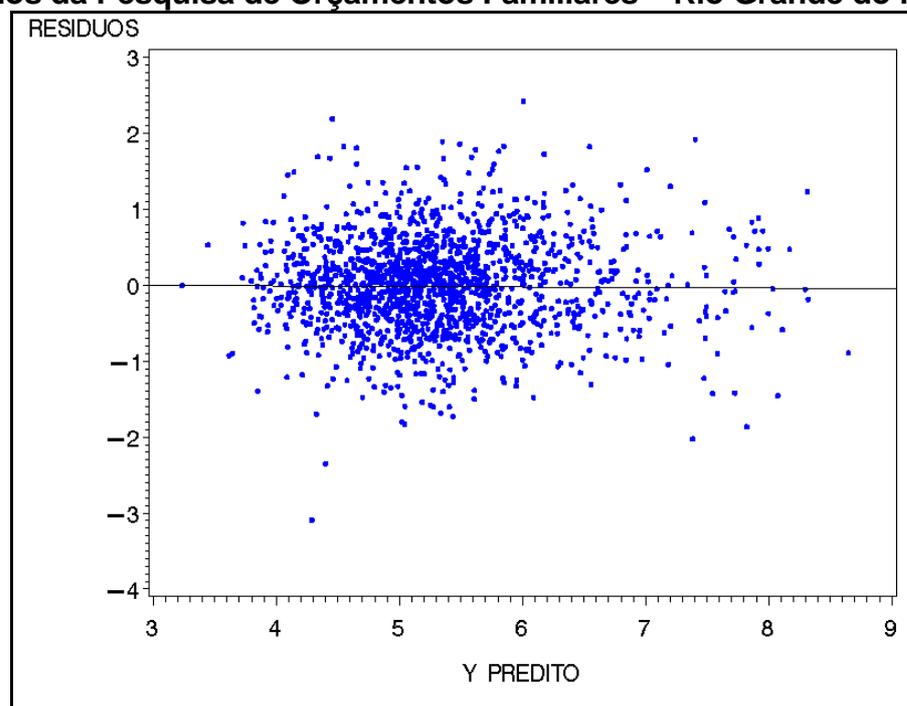
4.4.2. Avaliação dos modelos

A avaliação do ajuste dos modelos desenvolvidos nesta dissertação foi feita através de duas maneiras: avaliação dos gráficos dos resíduos contra os y preditos, e análise do poder de predição – R^2 .

Os gráficos dos resíduos são importantes para constatar possíveis variações na variância dos modelos. Se os pontos dos gráficos apresentam alguma forma definida, então há indícios que a hipótese de homocedasticidade foi violada, ou seja, o modelo não tem variância constante.

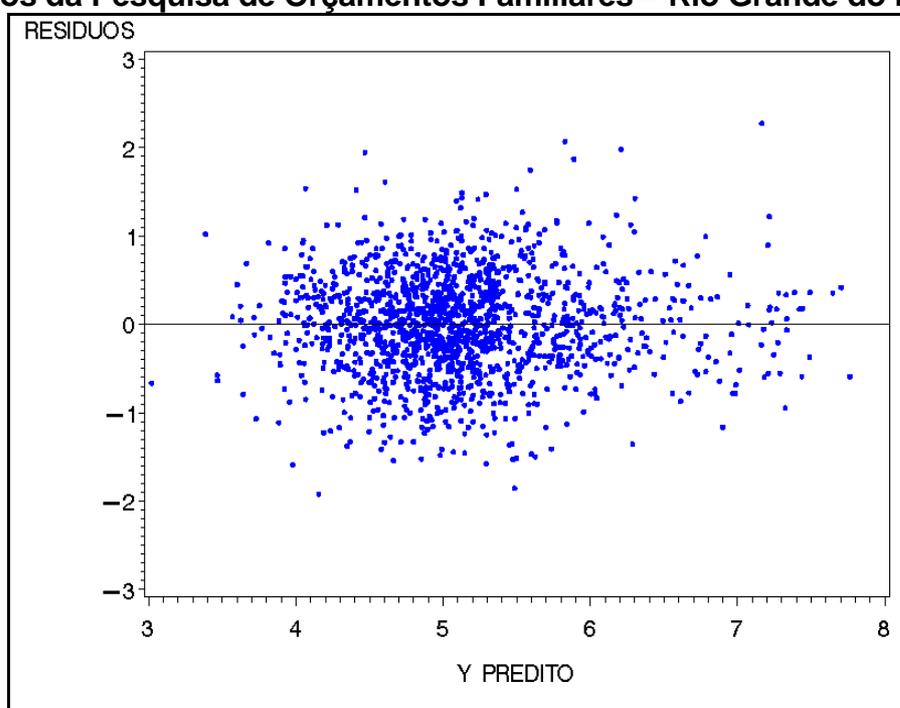
O gráfico 4.9 mostra os resíduos do modelo para o \ln da renda domiciliar *per capita*. Observa-se que os pontos estão aleatoriamente distribuídos em torno de zero, indicando que o modelo é homocedástico.

Gráfico 4.9: Resíduos do modelo de regressão para a renda ajustado com os dados da Pesquisa de Orçamentos Familiares – Rio Grande do Norte



O gráfico 4.10 mostra os resíduos do modelo do \ln da despesa *per capita*, onde se vê também que não apresentou nenhuma forma padrão, com pontos também aleatoriamente distribuídos em torno de zero. Existem alguns poucos pontos mais afastados da nuvem de pontos, que podem ser informações discrepantes e deveriam ser analisados com mais cuidado. Mas não acredito que estas informações possam afetar da qualidade dos resultados obtidos.

Gráfico 4.10: Resíduos do modelo de regressão para a despesa ajustado com os dados da Pesquisa de Orçamentos Familiares – Rio Grande do Norte



Com relação ao poder de explicação dos modelos, observa-se que tanto o modelo para o \ln da renda quanto o modelo para o \ln da despesa apresentam altos poderes de explicação. O primeiro modelo apresenta R^2 de 0,6415, e o segundo de 0,6155, ou seja, em ambos os modelos mais de 60% da variação total do \ln da renda *per capita* ou do \ln da despesa *per capita* é explicada pelo modelo.

Tabela 4.16: R² dos modelos para o ln da renda e da despesa

Modelo	R ²
Renda	0,6415
Despesa	0,6155

4.5. “Mapa” de pobreza – resultados

Para que seja feita uma mensuração da pobreza, é necessário primeiramente que se defina uma linha de pobreza, que será utilizada como ponto de corte da distribuição de renda ou despesa domiciliar *per capita*. Conforme mencionada no capítulo 2, a determinação da linha de pobreza geralmente remete a métodos arbitrários, que na maioria das vezes são escolhas pessoais dos pesquisadores. Barros, Henriques e Mendonça (2000)²³, por exemplo, definem a linha de pobreza como múltiplo da linha de indigência, que por sua vez se refere ao custo de uma cesta alimentar, regionalmente definida, que contemple as necessidades de consumo calórico mínimo de um indivíduo.

Já Hoffmann (2000)²⁴, utilizou em seu artigo para o livro *Desigualdade e Pobreza no Brasil*, do IPEA, a linha de pobreza de meio salário mínimo. Segundo Rocha (2000)²⁵, utilizar múltiplos do salário mínimo como linha de pobreza é uma prática relativamente comum entre os pesquisadores do Brasil.

Numa tentativa de não fugir ao que é comumente feito pelos pesquisadores da pobreza, neste trabalho a linha de pobreza é de R\$ 100,00, que representa meio salário mínimo em janeiro de 2003, período de referência da POF, e também é o período para o qual os valores monetários desta dissertação estão deflacionados.

²³ Barros, Henriques e Mendonça (2000), pág 23.

²⁴ Hoffmann (2000), pág 97.

²⁵ Rocha (IPEA: 2000), pág 109

Entretanto, são realizadas duas diferentes abordagens para determinar a população pobre. A primeira toma como base a renda domiciliar *per capita*, e a outra leva em conta as despesas domiciliares *per capita*. A linha de pobreza é a mesma para os dois métodos, de forma que serão considerados pobres, aqueles que tiverem renda *per capita* ou despesa *per capita* inferiores à R\$100,00²⁶.

O indicador utilizado na análise é o percentual de pobres e foi inicialmente calculado 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, e os resultados se encontram na tabela 4.16 a seguir.

Tabela 4.17: Proporção de pessoas em domicílios pobres baseada nas variáveis originais das pesquisas - Rio Grande do Norte e Microrregiões

RN e Microrregiões	POF		Censo
	Renda	Despesa	
RN	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: POF 2002/2003 e Censo 2000.

²⁶ O salário mínimo na época da POF era de R\$200,00.

O indicador também é obtido a partir da média das réplicas da variável em questão imputadas no Censo 2000 conforme descrito na seção 3.7. A renda domiciliar *per capita* e despesa domiciliar *per capita* foram imputadas com base na construção de 100 réplicas de cada variável. As réplicas foram calculadas com base numa distribuição normal. Para cada domicílio foram calculadas as médias destas 100 réplicas e este resultado é usado como estimativa para que seja feita a análise da pobreza nas 19 microrregiões do Rio Grande do Norte.

Nas tabelas 4.17 e 4.18, encontram-se os resultados do indicador da proporção de pessoas em domicílios pobres (média) estimado por microrregião, baseados na renda e despesa imputadas (Renda_imp e Desp_imp, respectivamente), bem como o esquema dos 5 números baseado nas réplicas e os percentis 3 e 97, que são análogos ao intervalo de confiança de 94%.

Tabela 4.18: Esquema dos 5 números e indicador de pobreza baseados na renda imputada - Rio Grande do Norte e Microrregiões

RN e Microrregiões	Mínimo	P03	Q1	Q2	Q3	P97	Máximo	Média
RN	35,5%	35,6%	35,7%	35,8%	35,9%	36,1%	36,2%	35,8%
Mossoró	28,9%	29,0%	29,6%	30,0%	30,4%	31,0%	31,4%	30,0%
Chapada do Apodi	40,6%	41,0%	42,2%	42,9%	43,6%	44,7%	45,6%	42,9%
Médio Oeste	45,2%	45,9%	47,5%	48,2%	48,9%	50,4%	51,4%	48,2%
Vale do Açu	37,7%	38,0%	38,7%	39,2%	39,6%	40,5%	40,7%	39,2%
Serra de São Miguel	46,0%	46,3%	47,7%	48,4%	49,0%	50,4%	50,9%	48,3%
Pau dos Ferros	40,3%	40,8%	41,7%	42,2%	42,7%	43,5%	43,6%	42,2%
Umarizal	40,3%	41,3%	42,0%	42,6%	43,1%	43,9%	44,2%	42,6%
Macau	32,1%	32,6%	34,0%	34,8%	35,6%	37,4%	37,9%	34,9%
Angicos	42,9%	43,6%	44,9%	45,5%	46,5%	47,5%	47,6%	45,6%
Serra de Santana	48,9%	49,8%	50,4%	51,0%	51,6%	52,3%	52,4%	51,0%
Seridó Ocidental	29,0%	29,1%	30,1%	30,8%	31,3%	32,0%	32,9%	30,7%
Seridó Oriental	32,1%	32,2%	33,3%	33,8%	34,2%	35,2%	35,4%	33,8%
Baixa Verde	46,6%	47,2%	49,0%	49,8%	50,5%	52,1%	53,2%	49,7%
Borborema Potiguar	48,2%	48,4%	49,3%	49,8%	50,1%	50,9%	51,4%	49,7%
Agreste Potiguar	48,6%	48,9%	49,5%	49,9%	50,2%	50,7%	50,9%	49,9%
Litoral Nordeste	49,0%	49,1%	50,7%	51,3%	51,9%	53,0%	53,6%	51,3%
Macaíba	37,0%	38,4%	39,4%	39,8%	40,2%	40,8%	41,1%	39,8%
Natal	22,0%	22,2%	22,6%	22,7%	22,9%	23,3%	23,4%	22,7%
Litoral Sul	44,6%	45,3%	46,3%	46,9%	47,2%	48,4%	49,1%	46,8%

Tabela 4.19: Esquema dos 5 números e indicador de pobreza baseados na despesa imputada - Rio Grande do Norte e Microrregiões

RN e Microrregiões	Mínimo	P03	Q1	Q2	Q3	P97	Máximo	Média
RN	42,9%	43,0%	43,2%	43,3%	43,5%	43,7%	43,9%	43,3%
Mossoró	35,3%	36,1%	37,0%	37,3%	37,6%	38,5%	38,5%	37,3%
Chapada do Apodi	49,6%	49,9%	51,3%	52,0%	52,6%	54,1%	54,6%	52,0%
Médio Oeste	55,1%	56,0%	57,7%	58,4%	59,2%	60,3%	60,8%	58,4%
Vale do Açu	46,1%	46,5%	47,4%	47,9%	48,4%	49,4%	49,6%	47,9%
Serra de São Miguel	53,7%	54,7%	56,0%	56,8%	57,6%	58,7%	59,5%	56,8%
Pau dos Ferros	48,1%	48,5%	49,4%	49,8%	50,3%	51,2%	51,5%	49,8%
Umarizal	49,7%	50,0%	51,1%	51,7%	52,1%	53,4%	53,8%	51,7%
Macau	42,0%	42,8%	44,4%	45,2%	46,2%	48,3%	49,0%	45,3%
Angicos	50,1%	51,4%	52,5%	53,1%	53,8%	55,1%	56,1%	53,2%
Serra de Santana	55,4%	56,0%	56,8%	57,4%	57,9%	59,1%	59,5%	57,4%
Seridó Ocidental	35,8%	36,1%	37,2%	37,9%	38,3%	39,5%	40,5%	37,8%
Seridó Oriental	39,3%	39,8%	40,6%	41,1%	41,8%	42,8%	43,3%	41,2%
Baixa Verde	55,0%	55,7%	57,3%	58,1%	58,9%	60,0%	60,5%	58,0%
Borborema Potiguar	55,3%	56,4%	57,2%	57,8%	58,2%	58,8%	59,2%	57,7%
Agreste Potiguar	56,2%	56,6%	57,2%	57,5%	57,8%	58,6%	59,3%	57,5%
Litoral Nordeste	57,4%	58,7%	59,6%	60,3%	60,8%	61,8%	62,0%	60,3%
Macaíba	47,7%	47,8%	48,7%	49,1%	49,6%	50,3%	50,4%	49,1%
Natal	27,9%	28,0%	28,4%	28,6%	28,9%	29,3%	29,4%	28,6%
Litoral Sul	54,7%	55,0%	56,0%	56,5%	57,1%	58,2%	58,5%	56,5%

Em ambas as tabelas acima, a coluna da média representa a proporção média de pessoas em domicílios pobres estimada a partir das 100 réplicas da renda e da despesa imputadas. Esta média é a variável que será usada para fazer a análise da pobreza nas microrregiões do estado do Rio Grande do Norte. Para facilitar a compreensão da análise, esta variável será chamada apenas de proporção de pobres ou percentual de pobres. Observe que os percentuais de pobres são maiores quando é utilizada a abordagem da despesa. Isso era esperado pelo fato de que normalmente as pessoas têm estruturas de gastos inferiores aos seus rendimentos e porque utilizamos a mesma linha de pobreza para as duas variáveis. No Estado como um todo a pobreza atinge cerca de 35,8% da população, quando a analisamos a partir da renda, e 43,3%, quando a analisamos a partir da despesa. Observe também que as proporções são muito diferentes para as microrregiões. Em ambas as abordagens a microrregião da capital (Natal) é a que apresenta a menor proporção de pobres, 22,7% para a renda e 28,6% para a despesa. Por outro lado, a microrregião que

possui a maior proporção de pobres tanto na abordagem da renda quanto na despesa é o Litoral Nordeste (51,3% e 60,3, respectivamente).

Já era de se esperar que a capital atingisse valores menores para proporção de pobres. Isso provavelmente acontece em todos os estados do Brasil, pois geralmente é na capital onde se concentra o maior número de empresas e escolas, fazendo com que a população tenha mais acesso a bons empregos e educação, e conseqüentemente atingindo níveis menores de pobreza.

Como já foi dito, a proporção de pobres aqui abordada é uma média de várias réplicas de renda ou despesa. A análise da média sozinha não fornece informação sobre como as réplicas se comportam quanto à distribuição dos dados. Nas tabelas 4.17 e 4.18 acima se encontra o esquema de 5 números das réplicas da proporção de pobres (mínimo, 1º quartil, 2º quartil, 3º quartil e máximo). Este conjunto de estatísticas ilustra a variabilidade dos indicadores de pobreza estimados para as microrregiões. Além dos 5 números a tabela apresenta os percentis 3 e 97, que são análogos ao intervalo de confiança de 94%. O conjunto destas 5 estatísticas possibilita uma avaliação da imprecisão das réplicas da renda imputada. Note que a imputação da renda através da metodologia ELL pode fornecer estimativas da proporção de pobres no Rio Grande do Norte variando de 35,5% (mínima proporção de pobres das réplicas) a 36,2% (máxima proporção de pobres), quando está usando a abordagem da renda. Já para a pobreza baseado na despesa, a proporção de pobres pode variar de 42,9% a 43,9%.

Numa tentativa de evidenciar mais as diferenças nos indicadores das microrregiões, estão sendo disponibilizados dois cartogramas a seguir. Eles foram construídos no software arcview. As classes de proporção de pobres dos mapas foram divididas para que tenham aproximadamente a mesma quantidade de

microrregiões. Note que quanto mais escura for a área do mapa, mais pobre será a microrregião.

Figura 4.2: Mapa da pobreza estimado a partir da renda imputada no Censo.

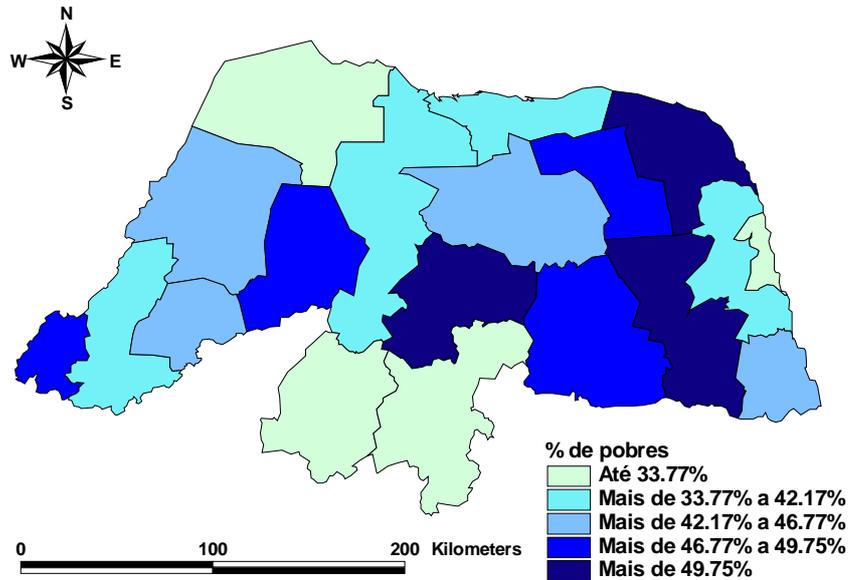
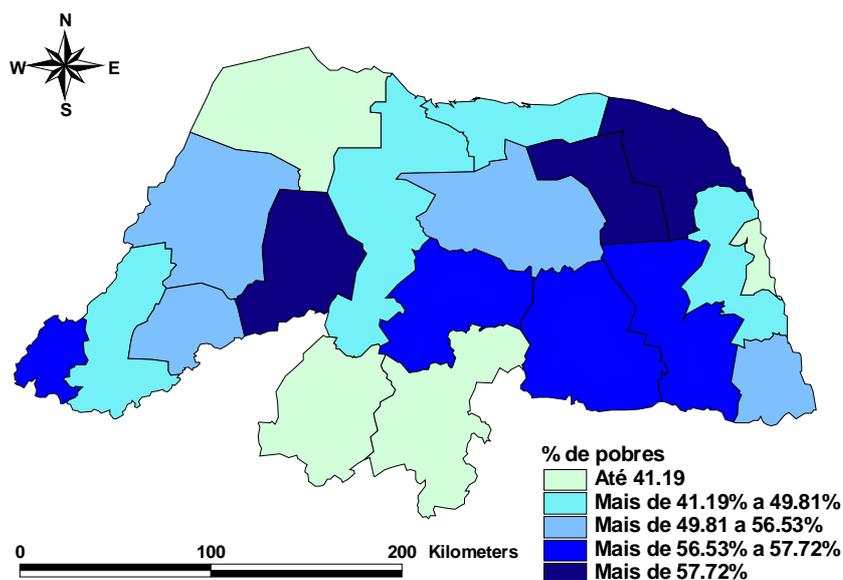


Figura 4.3: Mapa da pobreza estimado a partir da despesa imputada no Censo.



Comparando os resultados das duas metodologias apresentadas nessa seção, pode-se verificar que ambas apresentam resultados parecidos. Note que no Rio Grande do Norte a proporção de pobres baseada na renda estimada diretamente pela POF é 33%, enquanto que se baseando na despesa, este resultado é de 38,9%. Ao utilizar a metodologia ELL os resultados passam para 35,8% e 43,3%, respectivamente. Os resultados são semelhantes, no entanto, os erros padrões das estimativas da POF são muito elevados, 20,2 e 20,9, respectivamente, o que não ocorre com método ELL, uma vez que as estimativas dos 5 números apresentam valores parecidos, conforme mostram as tabelas 4.18 e 4.19.

5. Avaliação da Aplicação e da Metodologia

5.1. Comparação entre as rendas

Neste trabalho, os mapas da pobreza foram realizados a partir de duas variáveis imputadas, despesa domiciliar *per capita* e renda domiciliar *per capita*. A renda foi escolhida pois, além de ser um critério muito utilizado para medir pobreza, está presente nas duas pesquisas aqui utilizadas (Censo e POF), o que possibilita a comparação de resultados obtidos através da metodologia abordada nesta dissertação e através da variável original do censo.

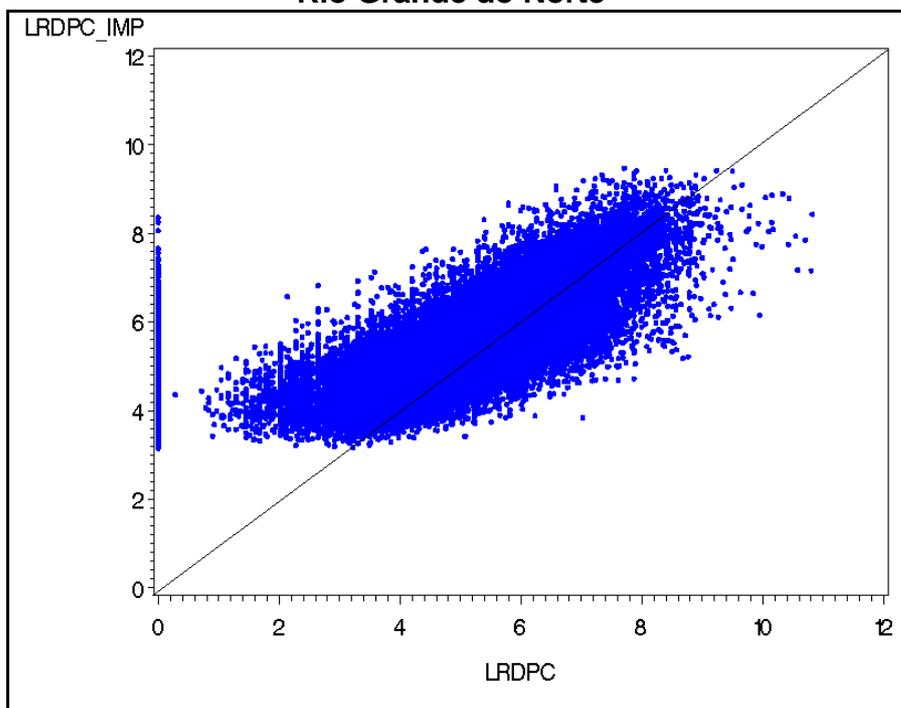
Uma primeira forma de comparação dos resultados pode ser feita a partir da análise gráfica da dispersão do logaritmo natural da renda domiciliar per capita (LRDPC)²⁷ contra o logaritmo natural da renda domiciliar per capita imputada (LRDPC_IMP). O gráfico 5.1 sugere não linearidade na relação entre essas variáveis, pois a nuvem de pontos não está formando uma reta, mas sim uma curva.

A variável original do censo (LRDPC) possui vários casos de rendas nulas, o que não ocorre na renda imputada (LRDPC_IMP), em virtude de não haver renda nula na POF. Esta conclusão leva a outra discussão. Porque não existe renda nula na POF? Seria impossível um domicílio não possuir renda naquele mês da pesquisa? Na verdade, na etapa de crítica da POF, se fossem detectadas rendas muito baixas, consideradas *outliers*, elas seriam substituídas por uma renda imputada. Além disso, a renda na POF é composta por uma componente monetária e outra não monetária. A renda não monetária é tudo “que é produzido, pescado, caçado, coletado ou recebido

²⁷ Como existem muitos casos domicílios com renda nula no censo, foi somado 1 à RDPC para que o cálculo do logaritmo natural pudesse ser realizado.

em bens (troca, doação, retirada do negócio, produção própria e salário em bens) utilizados ou consumidos durante o período de referência da pesquisa.”²⁸

**Gráfico 5.1: Dispersão da LRDPC contra LRDPC_IMP
– Rio Grande do Norte**



Vale ressaltar também que a variável original do censo apresenta valores geralmente menores que a variável imputada. Observe que a quantidade de pontos acima da reta que corta o gráfico ao meio, formando um ângulo de 45° com os eixos, é maior que a quantidade de pontos abaixo da reta. Além disso, a variável original apresenta alguns valores abaixo de 2, enquanto que a variável imputada não apresenta nenhum. Como o enfoque deste trabalho é a pobreza, observa-se que cerca de metade dos pontos da distribuição da renda original têm valores menores que 4,6 (que é o logaritmo natural de 100, valor estipulado para a linha de pobreza), enquanto que na variável imputada, o número de pontos abaixo de 4,6 é bem menor.

²⁸ Pesquisa de orçamentos familiares 2002-2003: Primeiros resultados, p. 23.

Então o gráfico já mostra sinais de que o percentual de pobres estimado a partir da variável original do Censo é superior ao percentual de pobres estimado pela imputação da variável a partir da metodologia ELL (2002). Ou seja, a metodologia subestima a pobreza, quando são utilizadas a POF e o Censo.

Talvez essa diferença nos resultados indique que as duas pesquisas captam a renda de modo desigual. Como já foi dito nesta dissertação²⁹, a forma da pergunta influencia na captação da resposta, e a POF aborda os rendimentos das famílias de maneira muito mais completa que o Censo. Vale lembrar que a POF foi uma pesquisa desenhada principalmente para obter informações sobre a composição orçamentária das famílias brasileiras, por isso a renda é pesquisada com muito mais detalhe.

Ainda pensando na relação entre as variáveis imputada e original, foi realizada uma regressão entre as duas variáveis, na expectativa de conhecer o poder de predição da renda imputada. Como se esperava uma relação linear entre as duas variáveis, foi feita uma regressão linear simples da seguinte forma:

$$LRDPC = \alpha + \beta * LRDPC_IMP + \varepsilon \quad (5.1)$$

Vale salientar que a renda imputada é obtida a partir de uma função de várias variáveis da POF combinadas com variáveis do censo, a equação acima na verdade significa a relação entre a renda do censo e essa função de variáveis das duas pesquisas. A equação do modelo é:

$$LRDPC = -2,048 + 1,2358 * LRDPC_IMP . \quad (5.2)$$

O principal resultado da regressão é que a renda imputada explica apenas 34% da variação da renda original do censo. Ou seja, a metodologia ELL não foi capaz de captar a renda como no Censo. Vale salientar que o poder de predição do modelo para a renda estimado com os dados da POF foi de 64%.

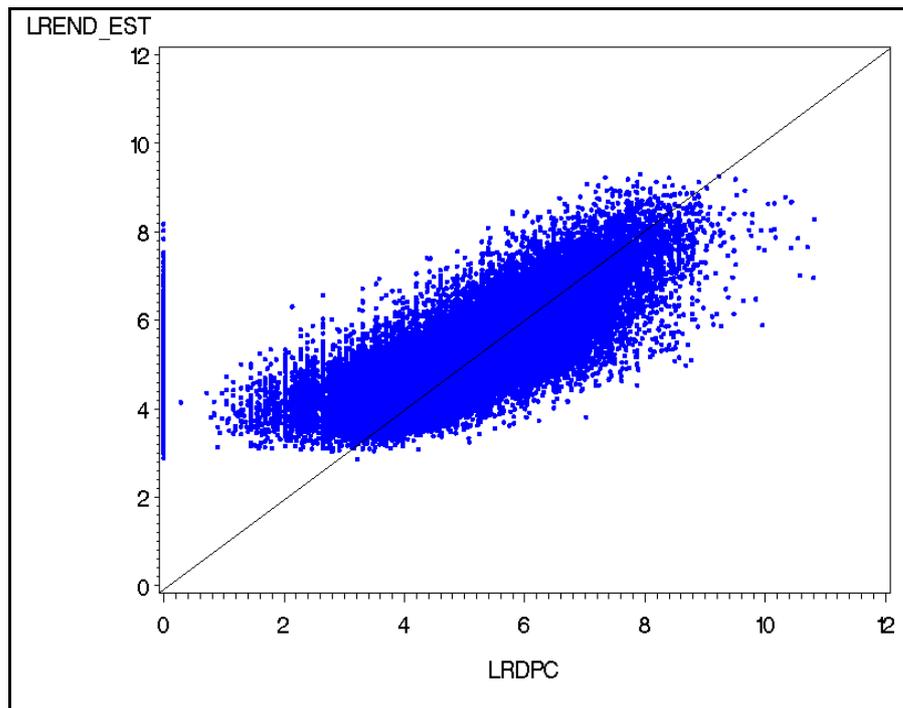
²⁹ Ver capítulo 4.

Essa constatação leva mais uma vez para a discussão sobre qual é a melhor forma de captar a renda. Talvez o fato de existir apenas uma pergunta sobre renda no Censo torne mais fácil o entrevistado errar ou omitir sua renda. É possível que na POF este problema seja amenizado pelo fato de existirem muitas perguntas sobre a renda, e também por ter a parcela não monetária da renda. Obter resposta a esta questão é uma tarefa muito difícil, que envolve fatores psicológicos dos entrevistados, forma e ordem das perguntas, tamanho do questionário, dentre outros. Esta dissertação se propõe apenas a avaliar a aplicabilidade da metodologia ELL, fazendo comparações com outros resultados possíveis, de forma que a avaliação da melhor forma de captar renda fica para outro momento.

Outro ponto importante na avaliação da aplicação da metodologia é o fato de que ela exige um enorme esforço computacional no cálculo da decomposição dos resíduos para a imputação no censo. A mesma regressão que foi estimada entre a renda imputada e a renda original foi feita para a renda estimada sem incorporar os resíduos, que aqui é chamada de LREND_EST. O fato inusitado é que o poder de explicação da LREND_EST é quase o mesmo do obtido pela LRDPC_IMP, ou seja, 34%. Então nessa metodologia talvez não fosse necessário fazer a decomposição dos resíduos nos 3 níveis estabelecidos nessa dissertação: microrregião, município e domicílio. Pois, aparentemente, os resultados obtidos pela LREND_EST são os mesmos que os da LRDPC_IMP, sendo que seriam economizados tempo e esforço se o estudo da pobreza fosse realizado a partir da LREND_EST.

O gráfico 5.2 confirma que as relações das variáveis LREND_EST e LRDPC_IMP com a renda original do censo são muito parecidas. Observe que os formatos das nuvens de pontos são praticamente os mesmos.

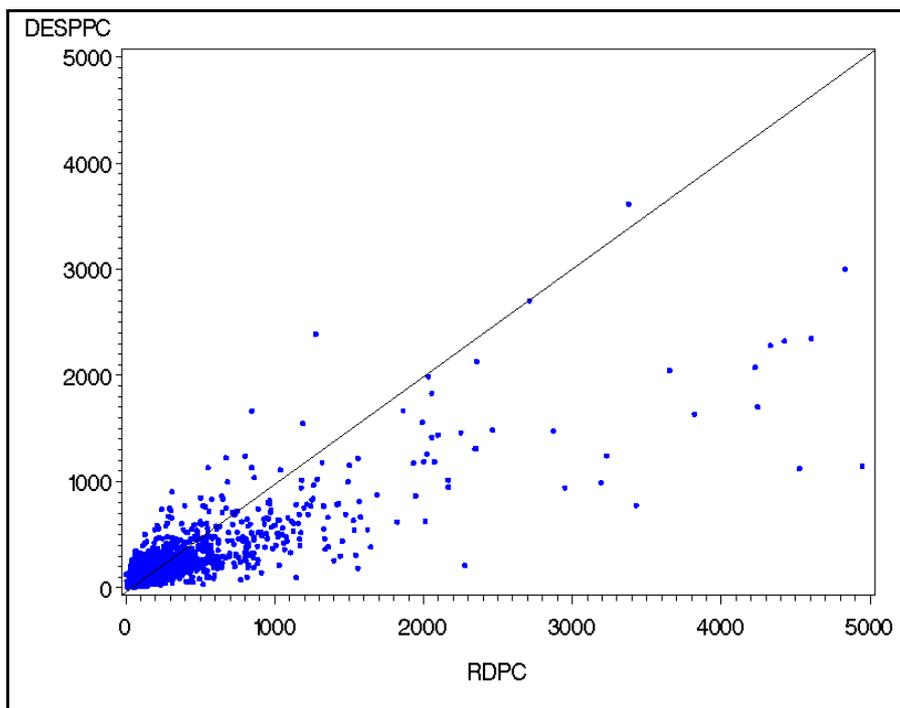
Gráfico 5.2: Dispersão da LRDPD contra LREND_EST



5.2. Comparação entre as despesa e renda na POF

Espera-se que quanto maior for a renda maior deveriam ser as despesas, além disso o total das despesas não deve ultrapassar o valor da renda. No entanto essa norma nem sempre é observada, caso contrário não existiriam. O gráfico abaixo mostra exatamente esse comportamento da relação entre despesas e rendas na POF. As variáveis analisadas são a renda domiciliar per capita (RDPC) e a despesa domiciliar per capita (DESPPC).

Gráfico 5.3: Dispersão da DESPPC contra RDPC



Observe a maioria dos pontos está concentrada junto à origem do gráfico, indicando que existem muitas rendas e despesas pequenas e poucas rendas e despesas altas. Percebe-se também que a nuvem de pontos forma um cone com abertura para os valores mais altos de RDPC e DESPPC. Isso indica que a variância dos resíduos da regressão entre as duas variáveis não é constante. Em linguagem matemática:

$$DESPPC = \alpha + \beta * RDPC + \varepsilon$$

onde, $E(\varepsilon) = 0$ e

$$V(\varepsilon) = \sigma^2 * RDPC^\gamma, \text{ com } \gamma > 0. \tag{5.3}$$

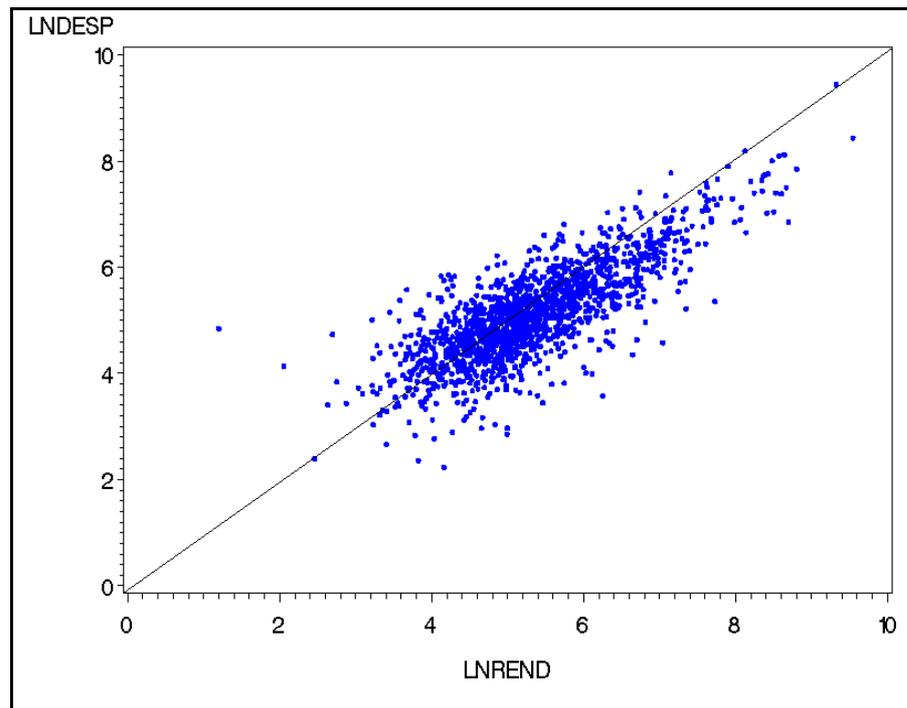
Para verificar a relação com mais qualidade, um modelo de regressão linear simples foi ajustado, e sua equação é a seguinte:

$$DESPPC = 70,9003 + 0,5078 * RDPC$$

Observe que a relação entre as variáveis é realmente positiva e crescente, uma vez que coeficiente angular é positivo, ou seja, quanto mais renda tem a pessoa maior é seu nível de despesa. A renda per capita possui um alto poder de explicação para a despesa per capita, cerca de 69%. No entanto, de acordo com o gráfico 5.3 se percebe que quando a renda é alta, as pessoas tendem a poupar, pois as despesas são menores que as rendas. Veja que a maioria dos pontos está abaixo da linha de 45°. Esse comportamento não é estranho, pois se espera que apenas as pessoas com rendas altas tenham renda capaz de gerar poupança. Ao passo que quando a renda é baixa, algumas pessoas não conseguem guardar dinheiro e podem até gastar mais do que têm, por isso existem alguns pontos acima da reta de 45° próximos a origem. Estes dois comportamentos caracterizam a existência de poupanças (por exemplo) e dívidas, respectivamente, correspondentes ao período de referência da pesquisa.

A mesma análise é feita com as variáveis transformadas a partir do logaritmo natural. O gráfico a seguir representa a relação entre as variáveis RDPC e DESPPC, em escala de logaritmo natural, LNREND e LNDESP, respectivamente.

Gráfico 5.4: Dispersão da LNDESP contra LNREND



Depois da transformação observa-se um maior espalhamento dos dados, sem que eles estejam concentrados próximos à origem do gráfico, como ocorre no gráfico das variáveis originais. Além disso, nota-se que o cone agora está com a abertura voltada para a origem, indicando que a transformação aplicada é muito forte, fazendo o γ da equação (5.3) passe a ser negativo ($\gamma < 0$).

$$R^2=64\%$$

$$LNDESP = 1,3867 + 0,7009 * LNREND$$

Um outro cuidado que o pesquisador deve tomar, se refere ao período de referência da pergunta. A POF pesquisa os mais variados tipos de despesas, elas podem ser feitas uma vez a cada ano, a cada mês, a cada semana, ou diariamente. Mas não é perguntada a frequência desta despesa, a pesquisa questiona apenas se houve despesa naquele período definido. Na despesa com alimentação, por exemplo, alguns domicílios não apresentam essa despesa. Por outro lado, outros domicílios apresentam despesa com alimentação naquela semana, mas não necessariamente fazem tal despesa toda semana. De forma que não dá para separar o grupo que tem despesas semanais do que fez a despesa naquele período de referência, mas não fazem a despesa semanalmente. A falta de informação sobre periodicidade da despesa pode gerar erro de análise, pois o pesquisador pode querer imputar uma despesa nos domicílios que não fizeram tal despesa.

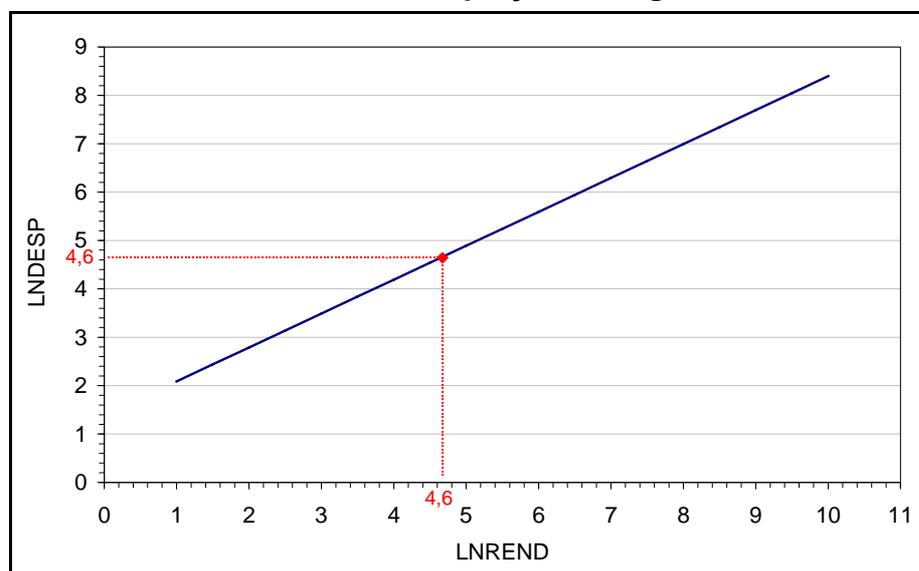
Um ponto importante a ser ressaltado é que neste trabalho foi utilizada a mesma linha de pobreza para a renda e para a despesa. Conforme visto anteriormente, as estruturas de gastos normalmente são inferiores às rendas, por isso seria coerente usar uma linha de pobreza um pouco menor para a despesa. A fim de exemplificar, na tabela abaixo se encontram os valores da linha de pobreza para a abordagem da despesa que são equivalentes a algumas linhas de pobreza hipotéticas. De acordo com a equação de regressão, se substituir a LNREND na equação por 4,6 (que é o valor da linha de pobreza de R\$100,00 em escala logarítmica), resulta em uma LNDESP de 4,61. Então até é sim possível utilizar a mesma linha de pobreza para as duas abordagens, pois por coincidência a equação de regressão fornece o mesmo valor para as linhas de pobreza. No entanto se fosse utilizada outra linha de pobreza baseada na renda (como 1 por exemplo) resultaria em um valor diferente para a linha de pobreza baseada na despesa (2,088), conforme

mostra a tabela 5.1. Além disso o gráfico, logo após a tabela, mostra a relação entre as duas variáveis, de acordo com a equação de regressão, com destaque para o valor usado como linha de pobreza.

Tabela 5.1: Valores das linhas de pobreza hipotéticas baseadas na regressão de LNDESP contra LNREND

LNREND	LNDESP
1,0	2,088
1,5	2,438
2,0	2,789
2,5	3,139
3,0	3,489
3,5	3,840
4,0	4,190
4,5	4,541
4,6	4,614
5,0	4,891
5,5	5,242
6,0	5,592
6,5	5,943
7,0	6,293
7,5	6,643
8,0	6,994
8,5	7,344
9,0	7,695
9,5	8,045
10,0	8,396

Gráfico 5.5: Relação entre as duas variáveis, de acordo com a equação de regressão



5.3. Mapa de pobreza utilizando modelos multiníveis

Como já foi dito anteriormente, a implementação da metodologia ELL exige intenso esforço computacional. Em virtude disso nesta dissertação tentou-se realizar as mesmas análises através de uma metodologia com emprego mais fácil e rápido. Os modelos multiníveis possibilitam a imputação de variáveis no censo e análise do mapa de pobreza no Rio Grande do Norte mantendo a mesma hierarquia de áreas que a metodologia criada. Além disso, sua utilização é amplamente difundida pelo mundo, facilitando acesso a livros e textos. Sua metodologia já está introduzida em alguns softwares estatísticos, e isso é primordial para facilitar o processamento dos dados. O modelo multinível é escrito da seguinte forma:

$$y = X\beta + Z\gamma + \varepsilon$$

onde γ é o vetor desconhecido de parâmetros fixos de efeitos aleatórios; Z é a matriz de efeitos aleatórios; ε é o vetor de erros aleatórios.

Nesta dissertação, os modelos multiníveis foram desenvolvidos com três hierarquias: domicílio, município e microrregião. A estimativa da proporção de pobres foi calculada para o estado do Rio Grande do Norte e suas microrregiões, assim como na metodologia criada por Elbers, Lanjouw e Lanjouw (2002). Os resultados se encontram na tabela a seguir.

Tabela 5.2: Indicadores de pobreza baseados na renda e despesas imputadas a partir de modelos multiníveis - Rio Grande do Norte e Microrregiões

RN e Microrregiões	Modelo	
	Renda	Despesa
RN	45.3%	51.6%
Natal	31.7%	36.8%
Mossoró	39.3%	45.4%
Seridó Ocidental	39.9%	46.7%
Macau	44.2%	55.7%
Seridó Oriental	44.9%	49.8%
Macaíba	49.9%	57.2%
Vale do Açu	50.3%	57.2%
Pau dos Ferros	52.9%	58.9%
Umarizal	53.3%	61.0%
Angicos	55.4%	63.1%
Chapada do Apodi	55.6%	60.7%
Médio Oeste	57.4%	67.9%
Agreste Potiguar	57.7%	64.7%
Serra de São Miguel	57.9%	65.8%
Borborema Potiguar	57.9%	63.4%
Litoral Sul	58.0%	64.6%
Serra de Santana	58.0%	65.8%
Litoral Nordeste	60.0%	68.2%
Baixa Verde	60.6%	66.4%

Observe que os resultados encontrados através dos modelos multiníveis atingem patamares superiores aos da metodologia ELL. Através do uso de modelos multiníveis, o Rio Grande do Norte apresentou proporção de pobres estimada em 45,3% e 51,6%, para a renda e despesa, respectivamente. Os resultados obtidos pela metodologia ELL são de 35,8% e 43,3%, respectivamente. No entanto, a proporção de pobres estimada a partir do modelo multinível para renda é muito mais próxima do valor da proporção de pobres estimada a partir da variável de renda original do Censo. Daí surge a dúvida se é realmente válido aplicar uma metodologia para cálculos de indicadores tão complicada e de difícil implementação para estimar indicadores de pobreza em pequenas áreas, se existem outros métodos mais simples, principalmente no tocante à estimação das variâncias, e cujos resultados se aproximaram mais

daqueles obtidos com indicadores derivados das medidas de renda disponíveis no censo. Infelizmente análise comparativa similar não é possível para a medida de pobreza baseada na variável despesa, em virtude de não existir essa variável no Censo.

As figuras a seguir representam os mapas de pobreza do Rio Grande do Norte estimados a partir dos modelos multiníveis para a renda e a despesa, respectivamente.

Figura 5.1: Mapa da pobreza estima a partir da renda imputada no Censo através de modelo multinível – Rio Grande do Norte

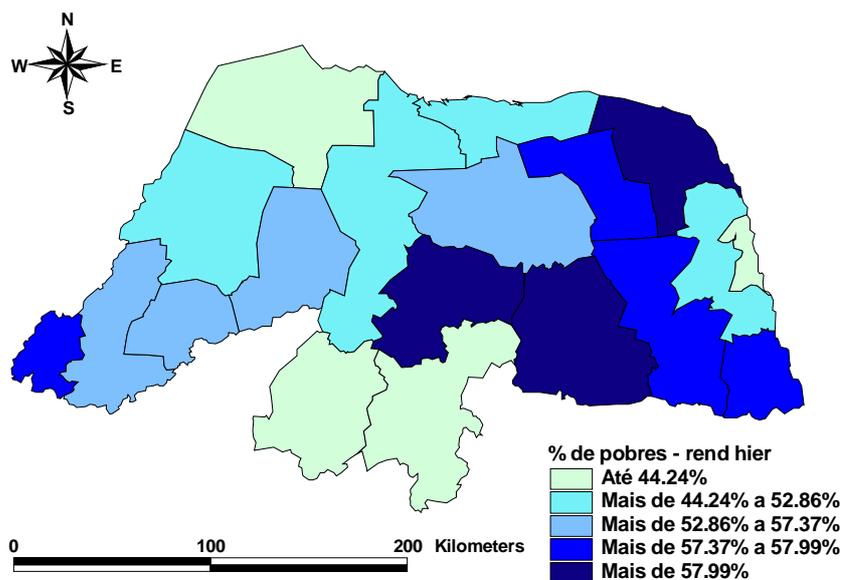
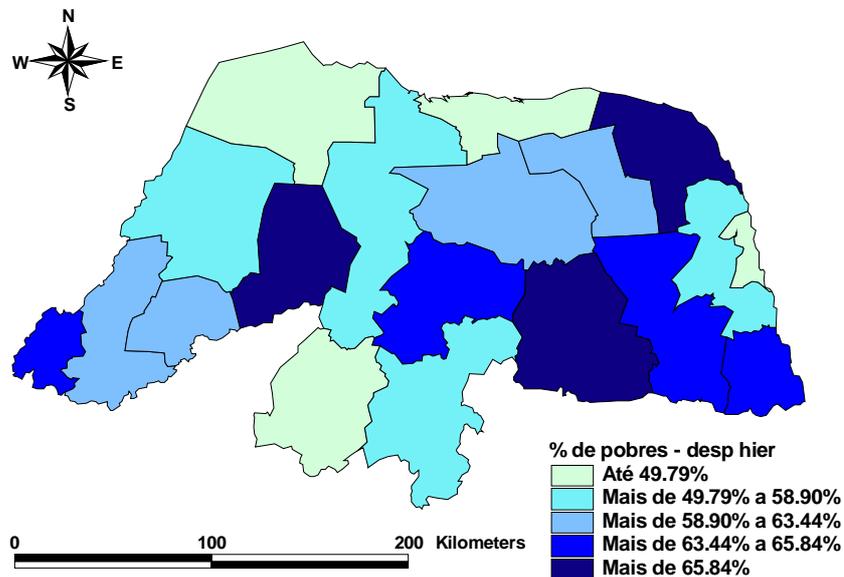
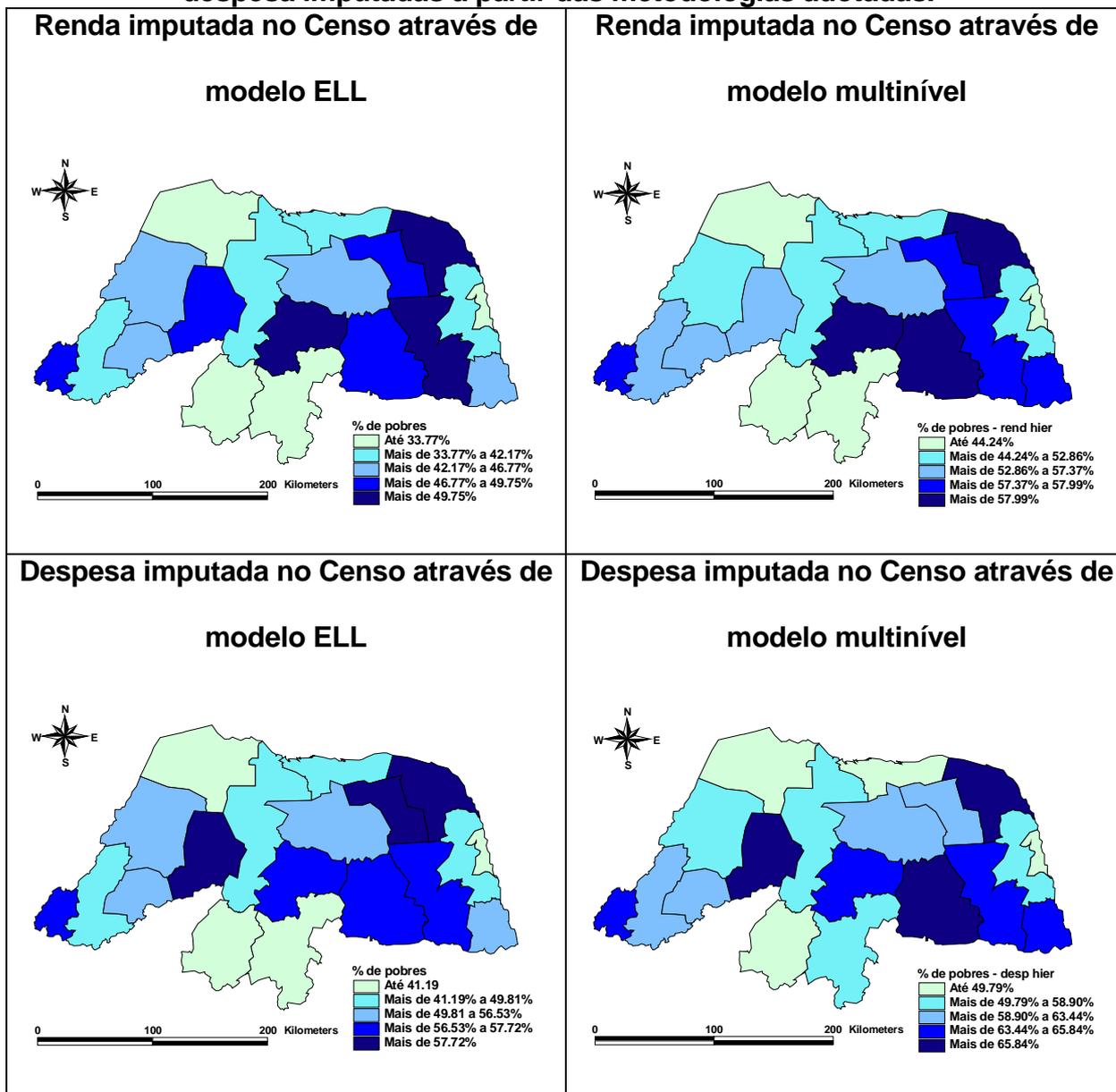


Figura 5.2: Mapa da pobreza estima a partir da despesa imputada no Censo através de modelo multinível – Rio Grande do Norte



Para comparar as duas metodologias com mais rapidez, os mapas obtidos por ambas metodologias foram postos lado a lado. Observe no quadro abaixo que os mapas são muito parecidos, o que reforça a conclusão de que é possível realizar uma análise semelhante à proposta pelos pesquisadores Elbers, Lanjouw e Lanjouw, mas com muito menor esforço e mais qualidade estatística.

Figura 5.3: Mapas de pobreza do Rio Grande do Norte baseados na renda e na despesa imputadas a partir das metodologias adotadas.



Para comparar todos os métodos de estimação do mapa de pobreza (ELL, POF sozinha, Censo sozinho e multinível), observe as duas tabelas a seguir, que apresentam ordenações das microrregiões do Rio Grande do Norte segundo os indicadores de pobreza derivados para cada uma das metodologias consideradas. De acordo com a tabela 5.3, as três primeiras posições, que representam as microrregiões menos pobres são praticamente ocupadas pelas mesmas microrregiões

em todas as metodologias, com exceção da que usa apenas a POF como fonte de dados. Já para as microrregiões mais pobres, não há tanta concordância entre as metodologias. No entanto as 8 microrregiões mais pobres são as mesmas para as metodologias ELL e Multinível, mudando apenas suas posições. Além disso, ao comparar a ordem das microrregiões por cada metodologia, observa-se uma concordância de 7 posições quando é comparada a metodologia multinível com a metodologia ELL.

A concordância entre as metodologias ELL e multinível é ainda maior quando o indicador de pobreza está baseado na despesa. De acordo com a tabela 5.4 há concordância em 10 posições. Além disso as 5 microrregiões menos pobres, de acordo tanto com as duas metodologias são as mesmas.

Tabela 5.3: Ranking da pobreza nas microrregiões do Rio Grande do Norte baseada na renda, segundo a metodologia

	Censo	POF	ELL	Multinível
Menos pobre	Natal	Natal	Natal	Natal
	Seridó Ocidental	Litoral Nordeste	Mossoró	Mossoró
	Mossoró	Macau	Seridó Ocidental	Seridó Ocidental
	Seridó Oriental	Seridó Ocidental	Seridó Oriental	Macau
	Macau	Mossoró	Macau	Seridó Oriental
	Macaíba	Macaíba	Vale do Açu	Macaíba
	Vale do Açu	Pau dos Ferros	Macaíba	Vale do Açu
	Angicos	Chapada do Apodi	Pau dos Ferros	Pau dos Ferros
	Chapada do Apodi	Vale do Açu	Umarizal	Umarizal
	Pau dos Ferros	Seridó Oriental	Chapada do Apodi	Angicos
	Umarizal	Litoral Sul	Angicos	Chapada do Apodi
	Borborema Potiguar	Borborema Potiguar	Litoral Sul	Médio Oeste
	Litoral Sul	Umarizal	Médio Oeste	Agreste Potiguar
	Serra de Santana	Angicos	Serra de São Miguel	Serra de São Miguel
	Baixa Verde	Agreste Potiguar	Borborema Potiguar	Borborema Potiguar
	Agreste Potiguar	Baixa Verde	Baixa Verde	Litoral Sul
	Médio Oeste	Serra de São Miguel	Agreste Potiguar	Serra de Santana
	Serra de São Miguel	Médio Oeste	Serra de Santana	Litoral Nordeste
Mais Pobre	Litoral Nordeste	Serra de Santana	Litoral Nordeste	Baixa Verde

Tabela 5.4: Ranking da pobreza nas microrregiões do Rio Grande do Norte baseada na despesa, segundo a metodologia

	POF	ELL	Multinível
Menos pobre	Natal	Natal	Natal
	Angicos	Mossoró	Mossoró
	Macau	Seridó Ocidental	Seridó Ocidental
	Seridó Ocidental	Seridó Oriental	Seridó Oriental
	Macaíba	Macau	Macau
	Litoral Nordeste	Vale do Açu	Macaíba
	Mossoró	Macaíba	Vale do Açu
	Litoral Sul	Pau dos Ferros	Pau dos Ferros
	Pau dos Ferros	Umarizal	Chapada do Apodi
	Serra de São Miguel	Chapada do Apodi	Umarizal
	Borborema Potiguar	Angicos	Angicos
	Seridó Oriental	Litoral Sul	Borborema Potiguar
	Vale do Açu	Serra de São Miguel	Litoral Sul
	Baixa Verde	Serra de Santana	Agreste Potiguar
	Umarizal	Agreste Potiguar	Serra de Santana
	Médio Oeste	Borborema Potiguar	Serra de São Miguel
	Agreste Potiguar	Baixa Verde	Baixa Verde
	Chapada do Apodi	Médio Oeste	Médio Oeste
Mais Pobre	Serra de Santana	Litoral Nordeste	Litoral Nordeste

6. Conclusões e trabalhos futuros

Esta dissertação se propôs a avaliar a metodologia proposta por Elbers, Lanjouw e Lanjouw (2002), e discutir até que ponto a sua aplicabilidade é viável. Para tal análise foi escolhida uma Unidade da Federação, em virtude do esforço computacional exigido para aplicação da metodologia. As áreas de análise são as microrregiões do Rio Grande do Norte.

Os resultados discutidos neste trabalho mostram que a metodologia ELL até permite elaborar mapas de pobreza para pequenas áreas, combinando as vantagens das duas fontes de dados utilizadas (POF 2002/2003 e Censo 2000), uma vez que possibilitaram a identificação das Microrregiões mais pobres no Rio Grande do Norte. Esta informação não seria possível ser obtida caso o estudo fosse baseado apenas na POF, ou no Censo para o caso do indicador de pobreza baseado na despesa.

Entretanto o enorme esforço computacional, diversas vezes comentado, dificulta que o estudo seja ampliado para todas as Unidades da Federação. E por isso estudou-se uma outra forma de captar as informações, com o auxílio de modelos multiníveis. Os resultados obtidos a partir deste modelo são mais apropriados, se comparados com o percentual de pobres obtido a partir da variável de renda do Censo 2000. O fato de haver maior acesso a documentos abordando a temática da modelagem multinível também é mais um motivo para confiar mais nesta metodologia, pois não há dúvidas quanto a aceitação desta pela academia. De fato, pareceu à autora que a tentativa dos três pesquisadores era ajustar um modelo hierárquico, o que foi feito de forma ingênua, sem benefício dos modernos métodos e sistemas computacionais disponíveis para este fim.

Além disso, a possibilidade de ampliação do estudo para todas as Unidades da Federação, e conseqüentemente, execução do mapa da pobreza para o Brasil como um todo, torna-se muito mais viável a partir da aplicação de modelos multiníveis, em virtude de demandar menor esforço e tempo na execução da programação. Essa ampliação do estudo para todos os outros estados do Brasil, não pôde ser executada neste momento, mas fica aqui a proposta para um trabalho futuro.

Observou-se que as duas metodologias fornecem estimativas do percentual de pobres com patamares diferentes, mas fornecem *rankings* de pobreza similares para as microrregiões do Rio Grande do Norte. Ou seja, tanto para a metodologia ELL quanto para multiníveis, obtém-se praticamente a mesma ordenação das microrregiões segundo o percentual de pobres. Este resultado é importante quando há interesse em identificar as áreas com maior incidência de pobreza.

Todavia, quando há necessidade de analisar a intensidade da pobreza, há uma grande diferença entre as metodologias. Este trabalho não responde com exatidão qual é a melhor metodologia para estimação de mapas de pobreza, e as conclusões tiradas por esta avaliação são fundamentadas em hipóteses tiradas pela autora. Dessa forma, responder qual método é mais eficaz configura-se em mais uma proposta para estudo futuro.

7. Referências Bibliográficas

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, 2000. Projeto BRA/97/013.

AITKIN, M.; LONGFORD, N. ***Statistical modeling in school effectiveness studies.*** Journal of the Royal Statistical Society. Series A, v.149, p.1-43, 1986.

ARAÚJO, F. das C. de S. O. ; SILVA, F. V. da; MACÊDO, M. das V. de A. & SILVA, M. E. da. **Potiguares.** In: *História do RN n@ WEB [On-line]*. Disponível em: <URL: www.seol.com.br/rnnaweb/>

ASSUNÇÃO, R. M., CAVENAGHI, S. M., et al. ***Empirical Bayes estimation of demographic schedules for small areas.*** Demography. 42: 537-558 p. 2005.

BARROS, R. P. de, CARVALHO, M. de. **Desafios Para A Política Social Brasileira,** Rio de Janeiro: IPEA, 2003. Texto para discussão nº 985.

BARROS, R. P. de, HENRIQUES, R. e MENDONÇA, R., **A estabilidade inaceitável: desigualdade e pobreza no Brasil,** in: Henriques, R. (Org.), **Desigualdade e pobreza no Brasil,** IPEA, Rio de Janeiro, 2000, 21-47.

CAVENAGHI, S. M., POTTER, J. E., *et al.* ***Fertility and development: evidence from Brazil.*** Demography. 39: 739-761 p. 2002.

Censo Demográfico, 2000. Documentação dos microdados da amostra. Rio de Janeiro:IBGE, novembro de 2002. 166 p.

DA COSTA, A. W. N. **Relatório da metodologia de decomposição das medidas de pobreza e produção de indicadores, tabelas e gráficos sobre a pobreza relativa.** Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001a. Projeto BRA/97/013.

DA COSTA, A. W. N. **Relatório técnico com avaliação da metodologia e testes estatísticos utilizados na medição de pobreza, definição de variáveis classificatórias e tratamento de renda.** Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001b. Projeto BRA/97/013.

DEATON, A. e ZAIDI, S. ***Guidelines for constructing consumption aggregates for Welfare Analysis.*** Washington DC: World Bank, Maio, 2002. Working Paper nº 135. ISBN 0-8213-4990-2.

ELBERS, C., LANJOUW, J.O., LANJOUW, P. (2001) ***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

ELBERS, C.; LANJOUW, J. O.; *et al.* ***Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data***. August 25, 2001.

ELBERS, C.; LANJOUW, J. O.; LANJOUW, P. ***Micro-level estimation of welfare***. Washington DC: World Bank, October, 2002. 64p. Policy Research Working Paper nº 2911

ESRI. **Manual de referência do usuário do sistema ArcView 3.2**, 2001.

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

FOSTER, J.; GREER, J. e THORBECKE, E. ***A Class of Decomposable Poverty Measures*** Econometrica 52(3): 761-765, May 1984.

GALVÃO, F. de A; MACÊDO, J. E. de; LIMA, M. D. dos A. & MEDEIROS, N. dos S. **Os franceses na costa potiguar**. In: História do RN n@ WEB [On-line]. Disponível em: <URL: www.seol.com.br/rnnaweb/>

Glossário da Pobreza e desenvolvimento humano, RDH 1997 PNUD

HISTÓRIA DO RIO GRANDE DO NORTE NA WEB

<http://www.serido.net/frame.php?&url=http://www.seol.com.br/ceres>

HOFFMANN, R. **Distribuição de renda: Medidas de Desigualdade e Pobreza**. São Paulo: Editora da Universidade de São Paulo, 1998. 275p. ISBN: 85-314-0473-8.

HOFFMANN, R., **Mensuração da desigualdade e da pobreza no Brasil**, in: Henriques, R. (Org.), **Desigualdade e pobreza no Brasil**, IPEA, Rio de Janeiro, 2000, 81-107.

JOSUÉ DE CASTRO

www.josuedecastro.com.br

KAPTEYN, A., KOOREMAN, P., AND WILLEMSE, R., ***Some Methodological Issues in the Implementation of Subjective Poverty Definitions***, in: *The Journal of Human Resources*, vol 23 (1988), 222-242.

KENDAL, M. G. e BUCKLAND, W. R., ***A dictionary of statistical terms***. Edinburgh: *The International Statistical Institute*, 1957, p. 37.

LONGFORD, N.T. ***Random coefficient models***. Oxford: Clarendon Press, 1993. 270p.

MAIA, G. (1998). **A invasão holandesa no Rio Grande (resumo)**. In: História do RN n@ WEB [On-line]. Disponível em: <URL: ww.seol.com.br/rnnaweb/>

MEDEIROS, M. **A Trajetória do *Welfare State* no Brasil: Papel Redistributivo das Políticas Sociais dos Anos 1930 aos Anos 1990**, Brasília: IPEA, 2001. Texto para discussão nº 852.

MEDICI, E. V. **Relatório com a especificação de temas variáveis e indicadores selecionadas**. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001a. Projeto BRA/97/013.

MEDICI, E. V. **Relatório com arquivos gerados das variáveis selecionadas**. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001b. Projeto BRA/97/013.

MEDICI, E. V. **Arquivo com indicadores, tabelas e gráficos sobre desigualdade social**. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001c. Projeto BRA/97/013.

MEDICI, E. V. **Relatório metodológico com avaliação da metodologia e testes estatísticos utilizados**. Projeto 5 – Sistema de indicadores sócio-demográficos. Rio de Janeiro: IBGE, 2001d. Projeto BRA/97/013.

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, 2001e. Projeto BRA/97/013.

MILLER, Alan J. ***Subset selection in regression***, In: Monographs on statistics and applied probability N° 40, Londres: Chapman and Hall, 1990. 229 p. ISBN 0-412-35380-6

MOURA, F. A. S. e HOLT, D. **Small area estimation using multilevel models**. Survey Methodology, 1999, Vol.25, N° 1, p. 73-80.

Pesquisa de orçamentos familiares 2002-2003 - primeiros resultados: Brasil e grandes regiões, Rio de Janeiro: IBGE, Coordenação de Índices de Preços, 2004. 276p. ISBN 85-240-3721-0

Pesquisa Nacional por Amostra de Domicílios, Rio de Janeiro: IBGE, 2003. v. 24, 120p. ISSN 0101-6822

PESSOA, D. G. C.; SILVA, P. L. do N. ***Análise de dados amostrais complexos***. São Paulo: Associação Brasileira de Estatística. 1998

PESSOA, D. G. C.; et al. **Comparando distribuições etárias em pesquisas por amostragem: PNAD 95 e PPV 96/97**

PRADHAN, M. e RAVALLION, M. ***Measuring poverty using qualitative perceptions of consumption adequacy***. The Review of Economics and Statistics, August 2000, 82(3): 462–471.

RAVALLION, M. ***Poverty comparisons: A guide to concepts and methods.***

Washington DC: World Bank, Fevereiro, 1992. 138p. Working Paper nº 88. ISBN 0-8213-2036-X.

Relatório de Revisão Bibliográfica Referente à Seleção de Metodologia Estatística para Análise da Pobreza, Rio de Janeiro: IBGE, 2000.

Relatório do Desenvolvimento Humano 2004, Nova York: PNUD, 2004.
ISBN 972-8730-18-7

RIO DE JANEIRO: TRABALHO E SOCIEDADE. **Conhecimento e desigualdade**. Rio de Janeiro: IETS, 2002. Edição Especial.

ROCHA, S., **Estimação de linhas de indigência e de pobreza: opções metodológicas no Brasil**, in: Henriques, R. (Org.), **Desigualdade e pobreza no Brasil**, IPEA, Rio de Janeiro, 2000, 109 -127.

ROCHA, S. **As Pesquisas Nacionais sobre Consumo no Brasil: Objetivos e Aplicações**, in: **Consumo alimentar: as grandes bases de informação: simpósio**. São Paulo: Instituto Danone, 2000. 80p. ISBN 85-87923-01-3

SÄRNDAL, C. E., SWENSSON, B., and WRETMAN, J.H. ***model-assisted Survey Sampling.*** New York: Springer, 1991.

SAS Institute Inc. 2004. **SAS OnlineDoc® 9.1.3**. Cary, NC: SAS Institute Inc.

SCHWARTZMAN, S. **As causas da pobreza no Brasil**. 1ª ed. Rio de Janeiro: Editora FGV, 2004, 208 p. ISBN 85-225-0462-8.

SEN, A. **Poverty: an ordinal approach to measurement**. *Econometrica*, Vol.44, Nº 2, p.219-231, 1976.

SUASSUNA, L. E. B.; MARIZ, M. da S. **História do Rio Grande do Norte Colonial (1597/1822)** Natal; 1997

URANI, A. **Desenvolvimento e Distribuição de Renda** In: CASTRO, A. C. (org). **Desenvolvimento em debate: Painéis do desenvolvimento brasileiro**. Rio de Janeiro: BNDES, 2002. p.155-191.

<http://www.rn.gov.br/principal/riograndedonorte.asp>

<http://www.rn.gov.br>

www.desigualdade.inf.br

8. Anexos

Anexo 8.1: Comparação das distribuições das variáveis preditoras entre a POF e o CENSO

(continua)

		TOTCOMDO				
		1	2	3	4	5 ou mais
CENSO		26.0	46.0	22.3	4.6	1.1
POF		27.1	49.6	19.8	2.7	0.9
	CHIW 21.59		P-valor: 0		Decisão: REJEITO IGUALDADE	
		TIPODOM				
		Casa	Apartamento	Cômodo		
CENSO		96.4	2.9	0.7		
POF		97.8	1.8	0.5		
	CHIW 2.95		P-valor 0.23		Decisão: NÃO REJEITO IGUALDADE	
		TOTCOM				
		3 ou menos	4	5	6	7 ou mais
CENSO		12.4	16.5	24.3	20.0	26.8
POF		9.0	14.4	26.8	22.6	27.2
	CHIW 19.84		P-valor: 0		Decisão: REJEITO IGUALDADE	
		ABASTEC				
		Rede Geral	Poço ou nascente	Outra forma		
CENSO		78.6	8.1	13.3		
POF		83.7	2.6	13.7		
	CHIW 18.69		P-valor 0		Decisão: REJEITO IGUALDADE	
		QTDBANH				
		0 ou 1	2	3	4 banheiros ou mais	
CENSO		85.0	10.3	3.0	1.7	
POF		83.2	12.5	3.0	1.4	
	CHIW 6.18		P-valor 0.1		Decisão: NÃO REJEITO IGUALDADE	
		TIPOESC				
		Rede Geral	Fossa Vala, Rio ou Outro	Não tem escoamento		
CENSO		17.2	71.8	1.6	9.4	
POF		17.4	75.2	1.4	6.0	
	CHIW 8.02		P-valor 0.05		Decisão: NÃO REJEITO IGUALDADE	
		EXISTBAN				
		Tem banheiro	Não tem banheiro			
CENSO		90.6	9.4			
POF		94.0	6.0			
	CHIW 7.83		P-valor 0.01		Decisão: REJEITO IGUALDADE	
		CONDDOM				
		Próprio	Alugado	Cedido ou outro		
CENSO		77.1	12.0	10.8		
POF		75.6	13.5	11.0		
	CHIW 1.71		P-valor 0.42		Decisão: NÃO REJEITO IGUALDADE	
		ILUMINA				
		Tem iluminação elétrica	Não tem iluminação elétrica			
CENSO		94.4	5.6			
POF		96.9	3.1			
	CHIW 6.04		P-valor 0.01		Decisão: NÃO REJEITO IGUALDADE	
		PAVIMEN				
		Sim	Não			
CENSO		55.7	44.3			
POF		59.8	40.2			
	CHIW 1.83		P-valor 0.18		Decisão: NÃO REJEITO IGUALDADE	
		AR				
		Tem	Não tem			
CENSO		4.5	95.5			
POF		4.1	95.9			
	CHIW 0.44		P-valor 0.51		Decisão: NÃO REJEITO IGUALDADE	

Anexo 8.1: Comparação das distribuições das variáveis preditoras entre a POF e o CENSO

(continuação)

AUTO					
		Tem	Não tem		
CENSO		20.1	79.9		
POF		16.4	83.6		
	CHIW 12.85		P-valor 0	Decisão: REJEITO IGUALDADE	
COMPUT					
		Tem	Não tem		
CENSO		5.4	94.6		
POF		6.3	93.7		
	CHIW 1.48		P-valor 0.22	Decisão: NÃO REJEITO IGUALDADE	
GELFREE					
		Tem	Não tem		
CENSO		72.3	27.8		
POF		79.1	20.9		
	CHIW 14.55		P-valor 0	Decisão: REJEITO IGUALDADE	
MICROO					
		Tem	Não tem		
CENSO		5.5	94.5		
POF		5.5	94.5		
	CHIW 0		P-valor 0.99	Decisão: NÃO REJEITO IGUALDADE	
MQLAVAR					
		Tem	Não tem		
CENSO		13.6	86.5		
POF		15.7	84.3		
	CHIW 4.67		P-valor 0.03	Decisão: NÃO REJEITO IGUALDADE	
RADIO					
		Tem	Não tem		
CENSO		81.7	18.3		
POF		61.5	38.5		
	CHIW 143.77		P-valor 0	Decisão: REJEITO IGUALDADE	
TV					
		Tem	Não tem		
CENSO		84.3	15.7		
POF		86.8	13.2		
	CHIW 2.95		P-valor 0.09	Decisão: NÃO REJEITO IGUALDADE	
VIDEO					
		Tem	Não tem		
CENSO		18.9	81.1		
POF		15.6	84.4		
	CHIW 13.74		P-valor 0	Decisão: REJEITO IGUALDADE	
SEX					
		Masculino	Feminino		
CENSO		76.5	23.5		
POF		77.2	22.8		
	CHIW 0.32		P-valor 0.57	Decisão: NÃO REJEITO IGUALDADE	
COR					
		Branca	Preta	Parda	Outras
CENSO		40.8	5.7	52.9	0.7
POF		38.3	3.1	58.0	0.6
	CHIW 33.98		P-valor 0	Decisão: REJEITO IGUALDADE	
COR2					
		Branca	Preta	Outras	
CENSO		40.8	5.7	53.5	
POF		38.3	3.1	58.6	
	CHIW 33.67		P-valor 0	Decisão: REJEITO IGUALDADE	

Anexo 8.1: Comparação das distribuições das variáveis preditoras entre a POF e o CENSO

(conclusão)

		COR4					
		Branca	Não Branca				
CENSO		40.8	59.3				
POF		38.3	61.7				
	CHIW 1.89		P-valor 0.17		Decisão: NÃO REJEITO IGUALDADE		
		FREQ					
		Sim, rede privada	Sim, rede pública	Não, mas já frequentou	Nunca frequentou		
CENSO		0.7	3.5	76.9	18.9		
POF		1.2	7.1	69.0	22.7		
	CHIW 23.32		P-valor 0		Decisão: REJEITO IGUALDADE		
		ESCOLA					
		Sem instrução	creche/alfabetização	Fundamental	Médio	Superior	
CENSO		23.7	1.0	54.0	15.1	6.3	
POF		23.1	3.3	50.3	17.9	5.5	
	CHIW 28.14		P-valor 0		Decisão: REJEITO IGUALDADE		
		ANOEST					
		Sem instrução	1 a 3 anos	4 a 7 anos	8 a 10 anos	11 a 14 anos	14 anos ou mais
CENSO		24.9	24.7	23.7	9.2	12.8	4.7
POF		24.6	23.1	23.3	10.8	14.2	3.9
	CHIW 7.76		P-valor 0.17		Decisão: NÃO REJEITO IGUALDADE		
		CONJUGE					
		Tem	Não tem				
CENSO		73.3	26.7				
POF		72.3	27.7				
	CHIW 0.75		P-valor 0.39		Decisão: NÃO REJEITO IGUALDADE		
		IDOSOS					
		Não tem	1 idoso	2 idosos	3 idosos ou mais		
CENSO		79.6	14.9	5.3	0.2		
POF		80.5	13.9	5.2	0.4		
	CHIW 1.85		P-valor 0.6		Decisão: NÃO REJEITO IGUALDADE		
		CRIAN					
		Não tem	1 criança	2 crianças	3 crianças	4 crianças ou mais	
CENSO		60.5	25.4	10.2	3.1	0.9	
POF		64.8	23.2	9.3	2.1	0.7	
	CHIW 14.88		P-valor 0		Decisão: REJEITO IGUALDADE		
		ADOLES					
		Não tem	1 adolescente	2 adolescentes	3 adolescentes	4 adolescentes ou mais	
CENSO		56.1	24.8	12.9	4.6	1.6	
POF		56.3	27.2	11.8	3.3	1.4	
	CHIW 7.95		P-valor 0.09		Decisão: NÃO REJEITO IGUALDADE		
		TIPOFAM					
		Unipessoal	Casal sem filho	Casal com filho	Homem chefe s/ conj	Mulher Chefe s/ conj	Outros
CENSO		6.8	11.1	62.3	1.8	14.3	3.8
POF		7.1	12.6	59.7	2.1	13.3	5.3
	CHIW 10.34		P-valor 0.07		Decisão: NÃO REJEITO IGUALDADE		
		POS					
		Emp. Doméstica	Empregado	Empregador	Conta-própria/Próprio consumo	Estagiário/voluntário	
CENSO		2.9	54.6	3.0	28.7	0.4	
POF		2.5	55.6	2.8	34.0	0.3	
	CHIW 50.1		P-valor 0		Decisão: REJEITO IGUALDADE		
		MORA					
		1	2	3	4	5 ou mais	
CENSO		6.8	14.9	20.7	22.6	35.1	
POF		7.1	16.8	21.7	23.8	30.6	
	CHIW 12.55		P-valor 0.01		Decisão: NÃO REJEITO IGUALDADE		

Anexo 8.2: Saída do modelo da POF para renda.

Estatísticas de análises	
R ²	0.6415
Root MSE	0.6209
Graus de liberdade (GL)	117

ANOVA do modelo para LNREND					
	GL	Soma de quadrados	Quadrados médios	Valor F	P-valor
Model	70	484060.5	6915.150	37.81	<.0001
Error	1479	270474.4	182.877		
Corrected Total	1549	754534.9			

Testes dos efeitos dos modelos			
Efeito	GL	Valor F	P-valor
Model	70	623.18	<.0001
Intercept	1	2558.56	<.0001
COMPUT	0	.	.
ANOEST	0	.	.
MORA	0	.	.
QTDBANH	0	.	.
GELFREE	1	68.40	<.0001
MQLAVAR	1	28.66	<.0001
IDOSOS	3	5.63	0.0012
POS	5	5.63	0.0001
ADOLES	4	8.79	<.0001
CRIAN	4	8.93	<.0001
AR	1	29.95	<.0001

Testes dos efeitos dos modelos			
Efeito	GL	Valor F	P-valor
MICROO	1	22.35	<.0001
ABASTEC	2	6.03	0.0032
CONDDOM	2	4.56	0.0124
TIPOFAM	4	2.31	0.0617
TIPOESC	3	2.35	0.0762
COR4	1	5.53	0.0204
PAVIMEN	1	4.59	0.0342
SEX	0	.	.
COMPUT_ANOEST	5	5.34	0.0002
SEX_ANOEST	6	2.52	0.0247
MORA_QTDBANH	11	4.01	<.0001

The denominator degrees of freedom for the F tests is 117.

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
Intercept	4.3382375	0.45985874	9.43	<.0001
COMPUT 1	0.7022240	0.30454731	2.31	0.0229
COMPUT 2	0.0000000	0.00000000	.	.
ANOEST 1	-0.4969403	0.37705149	-1.32	0.1901
ANOEST 2	-0.5158507	0.36841632	-1.40	0.1641
ANOEST 3	-0.6809819	0.37312883	-1.83	0.0705
ANOEST 4	-0.4050088	0.40744195	-0.99	0.3223
ANOEST 5	-0.0523660	0.37960244	-0.14	0.8905
ANOEST 6	0.7718761	0.43886672	1.76	0.0812
ANOEST 9	0.0000000	0.00000000	.	.

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
MORA 1	0.8135677	0.17298859	4.70	<.0001
MORA 2	0.2702972	0.50970253	0.53	0.5969
MORA 3	1.0496111	0.27705418	3.79	0.0002
MORA 4	0.4804801	0.24087434	1.99	0.0484
MORA 5	0.0000000	0.00000000	.	.
QTDBANH 1	-0.2568050	0.26398182	-0.97	0.3327
QTDBANH 2	0.0129164	0.25906055	0.05	0.9603
QTDBANH 3	-0.1022765	0.24425306	-0.42	0.6762
QTDBANH 4	0.0000000	0.00000000	.	.
GELFREE 1	0.4838855	0.05850774	8.27	<.0001
GELFREE 2	0.0000000	0.00000000	.	.
MQLAVAR 1	0.2988280	0.05582180	5.35	<.0001
MQLAVAR 2	0.0000000	0.00000000	.	.
IDOSOS 1	-0.5748106	0.25088923	-2.29	0.0237
IDOSOS 2	-0.3537677	0.25610771	-1.38	0.1698
IDOSOS 3	-0.3252309	0.25688001	-1.27	0.2080
IDOSOS 4	0.0000000	0.00000000	.	.
POS 1	-0.0777870	0.11423006	-0.68	0.4972
POS 2	-0.0108323	0.06381790	-0.17	0.8655
POS 3	0.4693692	0.12951561	3.62	0.0004
POS 4	-0.0725192	0.06398381	-1.13	0.2594
POS 5	0.4632495	0.20093896	2.31	0.0229
POS 9	0.0000000	0.00000000	.	.
ADOLE 1	0.5747888	0.14387206	4.00	0.0001
ADOLE 2	0.4143315	0.14227399	2.91	0.0043

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
ADOLE 3	0.2170918	0.13756402	1.58	0.1172
ADOLE 4	0.2334163	0.14076149	1.66	0.0999
ADOLE 5	0.0000000	0.00000000	.	.
CRIAN 1	0.3712595	0.19190548	1.93	0.0555
CRIAN 2	0.2261019	0.19203497	1.18	0.2414
CRIAN 3	0.1271818	0.19381964	0.66	0.5130
CRIAN 4	-0.1238541	0.20450722	-0.61	0.5459
CRIAN 5	0.0000000	0.00000000	.	.
AR 1	0.4962720	0.09068665	5.47	<.0001
AR 2	0.0000000	0.00000000	.	.
MICROO 1	0.3886922	0.08221608	4.73	<.0001
MICROO 2	0.0000000	0.00000000	.	.
ABASTEC 1	0.1700339	0.07835672	2.17	0.0320
ABASTEC 2	0.3481234	0.10508743	3.31	0.0012
ABASTEC 3	0.0000000	0.00000000	.	.
CONDDOM 1	0.1447600	0.05685004	2.55	0.0122
CONDDOM 2	0.0285095	0.06505161	0.44	0.6620
CONDDOM 3	0.0000000	0.00000000	.	.
TIPOFAM 1	0.0000000	0.00000000	.	.
TIPOFAM 2	0.1987966	0.11679439	1.70	0.0914
TIPOFAM 3	0.1259415	0.10465685	1.20	0.2313
TIPOFAM 4	-0.0333499	0.15382210	-0.22	0.8287
TIPOFAM 5	0.2292019	0.09768844	2.35	0.0206
TIPOFAM 6	0.0000000	0.00000000	.	.
TIPOESC 1	0.0827773	0.09014710	0.92	0.3604

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
TIPOESC 2	0.0231189	0.08288722	0.28	0.7808
TIPOESC 3	-0.2552758	0.13491154	-1.89	0.0609
TIPOESC 4	0.0000000	0.00000000	.	.
COR4 1	0.0800697	0.03404513	2.35	0.0204
COR4 2	0.0000000	0.00000000	.	.
PAVIMEN 1	0.0865256	0.04036825	2.14	0.0342
PAVIMEN 2	0.0000000	0.00000000	.	.
SEX 1	-0.3723847	0.35616325	-1.05	0.2979
SEX 2	0.0000000	0.00000000	.	.
COMPUT_ANOEST 12	0.2384410	0.33535702	0.71	0.4785
COMPUT_ANOEST 13	0.1714122	0.35434366	0.48	0.6295
COMPUT_ANOEST 14	-0.3874128	0.44445313	-0.87	0.3852
COMPUT_ANOEST 15	-0.4436988	0.32782353	-1.35	0.1785
COMPUT_ANOEST 16	-0.6471512	0.37141411	-1.74	0.0841
COMPUT_ANOEST 19	0.0000000	0.00000000	.	.
COMPUT_ANOEST 21	0.0000000	0.00000000	.	.
COMPUT_ANOEST 22	0.0000000	0.00000000	.	.
COMPUT_ANOEST 23	0.0000000	0.00000000	.	.
COMPUT_ANOEST 24	0.0000000	0.00000000	.	.
COMPUT_ANOEST 25	0.0000000	0.00000000	.	.
COMPUT_ANOEST 26	0.0000000	0.00000000	.	.
COMPUT_ANOEST 29	0.0000000	0.00000000	.	.
SEX_ANOEST 11	0.4046865	0.34977658	1.16	0.2496
SEX_ANOEST 12	0.4735791	0.37098372	1.28	0.2043
SEX_ANOEST 13	0.7337342	0.36761123	2.00	0.0483

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
SEX_ANOEST 14	0.6458103	0.40951142	1.58	0.1175
SEX_ANOEST 15	0.5939371	0.37443265	1.59	0.1154
SEX_ANOEST 16	0.1526708	0.46030614	0.33	0.7407
SEX_ANOEST 19	0.0000000	0.00000000	.	.
SEX_ANOEST 21	0.0000000	0.00000000	.	.
SEX_ANOEST 22	0.0000000	0.00000000	.	.
SEX_ANOEST 23	0.0000000	0.00000000	.	.
SEX_ANOEST 24	0.0000000	0.00000000	.	.
SEX_ANOEST 25	0.0000000	0.00000000	.	.
SEX_ANOEST 26	0.0000000	0.00000000	.	.
SEX_ANOEST 29	0.0000000	0.00000000	.	.
MORA_QTDBANH 11	0.1981419	0.17405428	1.14	0.2573
MORA_QTDBANH 12	0.4616376	0.22760856	2.03	0.0448
MORA_QTDBANH 13	0.0000000	0.00000000	.	.
MORA_QTDBANH 21	0.0833155	0.50550110	0.16	0.8694
MORA_QTDBANH 22	0.3168180	0.52882599	0.60	0.5503
MORA_QTDBANH 23	0.4911679	0.58466063	0.84	0.4026
MORA_QTDBANH 24	0.0000000	0.00000000	.	.
MORA_QTDBANH 31	-0.8315104	0.27786767	-2.99	0.0034
MORA_QTDBANH 32	-0.7259762	0.31256951	-2.32	0.0219
MORA_QTDBANH 33	-0.8532729	0.29060390	-2.94	0.0040
MORA_QTDBANH 34	0.0000000	0.00000000	.	.
MORA_QTDBANH 41	-0.4066780	0.24613664	-1.65	0.1012
MORA_QTDBANH 42	-0.4978009	0.26317569	-1.89	0.0610
MORA_QTDBANH 43	-0.3602349	0.31282692	-1.15	0.2519

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
MORA_QTDBANH 44	0.0000000	0.00000000	.	.
MORA_QTDBANH 51	0.0000000	0.00000000	.	.
MORA_QTDBANH 52	0.0000000	0.00000000	.	.
MORA_QTDBANH 53	0.0000000	0.00000000	.	.
MORA_QTDBANH 54	0.0000000	0.00000000	.	.

The denominator degrees of freedom for the t tests is 117.
Matrix $X'WX$ is singular and a generalized inverse was used to solve the normal equations.
Estimates are not unique.

Anexo 8.3: Saída do modelo da POF para despesa

Estatísticas de análises	
R-square	0.6155
Root MSE	0.5655
Denominator DF	117

ANOVA do modelo para LNDESP					
	GL	Soma de quadrados	Quadrados médios	Valor F	P-valor
Model	70	359198.8	5131.412	33.83	<.0001
Error	1479	224366.7	151.702		
Corrected Total	1549	583565.6			

Testes dos efeitos dos modelos			
Efeito	GL	Valor F	P-valor
Model	70	576.54	<.0001
Intercept	1	1739.99	<.0001
COMPUT	0	.	.
ANOEST	0	.	.
MORA	0	.	.
QTDBANH	0	.	.
GELFREE	1	51.25	<.0001
MQLAVAR	1	15.55	0.0001
IDOSOS	3	2.50	0.0627
POS	5	0.48	0.7932
ADOLES	4	10.40	<.0001
CRIAN	4	6.46	<.0001
AR	1	23.94	<.0001

Testes dos efeitos dos modelos			
Efeito	GL	Valor F	P-valor
MICROO	1	26.08	<.0001
ABASTEC	2	2.52	0.0851
CONDDOM	2	7.63	0.0008
TIPOFAM	4	0.88	0.4790
TIPOESC	3	2.75	0.0459
COR4	1	0.84	0.3613
PAVIMEN	1	0.81	0.3696
SEX	0	.	.
COMPUT_ANOEST	5	3.33	0.0075
SEX_ANOEST	6	0.74	0.6160
MORA_QTDBANH	11	2.37	0.0112

The denominator degrees of freedom for the F tests is 117.

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
Intercept	4.4453786	0.49644500	8.95	<.0001
COMPUT 1	0.6069343	0.26632090	2.28	0.0245
COMPUT 2	0.0000000	0.00000000	.	.
ANOEST 1	-0.5742327	0.28524406	-2.01	0.0464
ANOEST 2	-0.5699072	0.28859351	-1.97	0.0506
ANOEST 3	-0.4903471	0.29031606	-1.69	0.0939
ANOEST 4	-0.3341560	0.32498720	-1.03	0.3060
ANOEST 5	-0.1940027	0.29377624	-0.66	0.5103
ANOEST 6	0.1344409	0.34698794	0.39	0.6991
ANOEST 9	0.0000000	0.00000000	.	.

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
MORA 1	1.3194181	0.18005766	7.33	<.0001
MORA 2	0.6995591	0.45050099	1.55	0.1232
MORA 3	0.9792178	0.28992524	3.38	0.0010
MORA 4	0.7241376	0.29736052	2.44	0.0164
MORA 5	0.0000000	0.00000000	.	.
QTDBANH 1	-0.2806414	0.27822590	-1.01	0.3152
QTDBANH 2	-0.0376051	0.28688899	-0.13	0.8959
QTDBANH 3	-0.0551119	0.26748199	-0.21	0.8371
QTDBANH 4	0.0000000	0.00000000	.	.
GELFREE 1	0.3844167	0.05369638	7.16	<.0001
GELFREE 2	0.0000000	0.00000000	.	.
MQLAVAR 1	0.2081368	0.05277438	3.94	0.0001
MQLAVAR 2	0.0000000	0.00000000	.	.
IDOSOS 1	-0.7294059	0.29827398	-2.45	0.0160
IDOSOS 2	-0.7410500	0.29205558	-2.54	0.0125
IDOSOS 3	-0.6613500	0.30809800	-2.15	0.0339
IDOSOS 4	0.0000000	0.00000000	.	.
POS 1	-0.0370017	0.12940242	-0.29	0.7754
POS 2	0.0460149	0.05303018	0.87	0.3873
POS 3	0.1616476	0.11697115	1.38	0.1696
POS 4	0.0380252	0.04958082	0.77	0.4447
POS 5	-0.0516790	0.27316641	-0.19	0.8503
POS 9	0.0000000	0.00000000	.	.
ADOLES 1	0.3961335	0.09875984	4.01	0.0001
ADOLES 2	0.3016063	0.09800930	3.08	0.0026

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
ADOLE 3	0.1349546	0.10475775	1.29	0.2002
ADOLE 4	-0.0158630	0.12505723	-0.13	0.8993
ADOLE 5	0.0000000	0.00000000	.	.
CRIAN 1	0.6969611	0.21590009	3.23	0.0016
CRIAN 2	0.5588481	0.20835814	2.68	0.0084
CRIAN 3	0.5030411	0.22152618	2.27	0.0250
CRIAN 4	0.3075907	0.24922209	1.23	0.2196
CRIAN 5	0.0000000	0.00000000	.	.
AR 1	0.3751091	0.07665867	4.89	<.0001
AR 2	0.0000000	0.00000000	.	.
MICROO 1	0.3538952	0.06929674	5.11	<.0001
MICROO 2	0.0000000	0.00000000	.	.
ABASTEC 1	0.1144808	0.06522498	1.76	0.0818
ABASTEC 2	-0.0911486	0.08696118	-1.05	0.2967
ABASTEC 3	0.0000000	0.00000000	.	.
CONDDOM 1	0.1469037	0.05316139	2.76	0.0066
CONDDOM 2	0.0264188	0.06532098	0.40	0.6866
CONDDOM 3	0.0000000	0.00000000	.	.
TIPOFAM 1	0.0000000	0.00000000	.	.
TIPOFAM 2	0.1110384	0.08251086	1.35	0.1810
TIPOFAM 3	0.1405504	0.08998628	1.56	0.1210
TIPOFAM 4	0.0725430	0.12338168	0.59	0.5577
TIPOFAM 5	0.1446389	0.09429262	1.53	0.1277
TIPOFAM 6	0.0000000	0.00000000	.	.
TIPOESC 1	0.0741632	0.10880946	0.68	0.4968

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
TIPOESC 2	-0.0386048	0.09878258	-0.39	0.6967
TIPOESC 3	-0.2302683	0.18963304	-1.21	0.2271
TIPOESC 4	0.0000000	0.00000000	.	.
COR4 1	0.0287422	0.03136020	0.92	0.3613
COR4 2	0.0000000	0.00000000	.	.
PAVIMEN 1	0.0382252	0.04244189	0.90	0.3696
PAVIMEN 2	0.0000000	0.00000000	.	.
SEX 1	-0.3774271	0.29446902	-1.28	0.2025
SEX 2	0.0000000	0.00000000	.	.
COMPUT_ANOEST 12	0.1311391	0.28922955	0.45	0.6511
COMPUT_ANOEST 13	0.1233387	0.28716759	0.43	0.6683
COMPUT_ANOEST 14	-0.4089561	0.32675084	-1.25	0.2132
COMPUT_ANOEST 15	-0.1270576	0.27327563	-0.46	0.6428
COMPUT_ANOEST 16	-0.3579774	0.32690071	-1.10	0.2757
COMPUT_ANOEST 19	0.0000000	0.00000000	.	.
COMPUT_ANOEST 21	0.0000000	0.00000000	.	.
COMPUT_ANOEST 22	0.0000000	0.00000000	.	.
COMPUT_ANOEST 23	0.0000000	0.00000000	.	.
COMPUT_ANOEST 24	0.0000000	0.00000000	.	.
COMPUT_ANOEST 25	0.0000000	0.00000000	.	.
COMPUT_ANOEST 26	0.0000000	0.00000000	.	.
COMPUT_ANOEST 29	0.0000000	0.00000000	.	.
SEX_ANOEST 11	0.3225809	0.29368481	1.10	0.2743
SEX_ANOEST 12	0.4620921	0.30353905	1.52	0.1306
SEX_ANOEST 13	0.4121607	0.30667449	1.34	0.1816

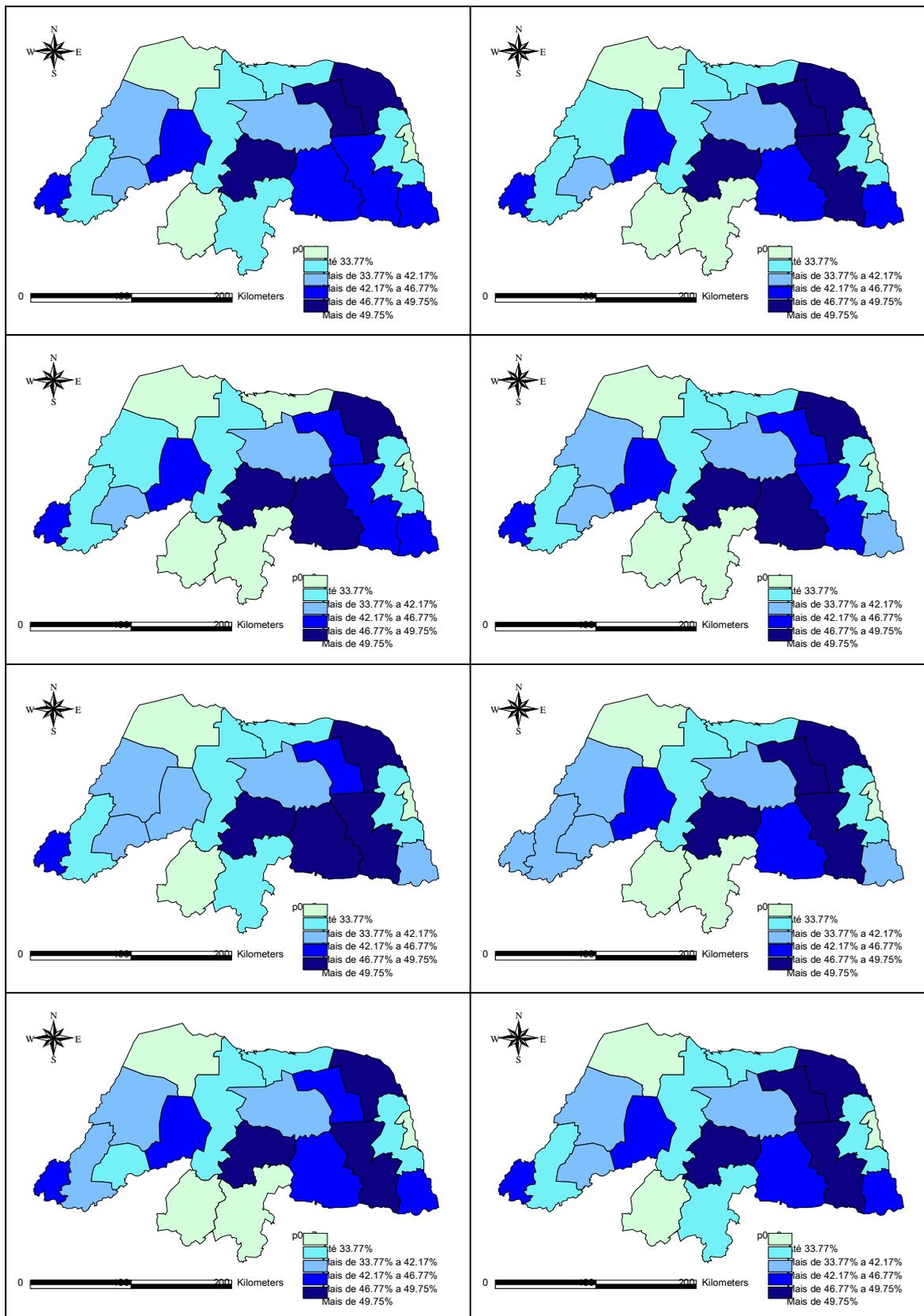
Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
SEX_ANOEST 14	0.5140474	0.34071579	1.51	0.1341
SEX_ANOEST 15	0.3580495	0.30000873	1.19	0.2351
SEX_ANOEST 16	0.1995067	0.35245439	0.57	0.5724
SEX_ANOEST 19	0.0000000	0.00000000	.	.
SEX_ANOEST 21	0.0000000	0.00000000	.	.
SEX_ANOEST 22	0.0000000	0.00000000	.	.
SEX_ANOEST 23	0.0000000	0.00000000	.	.
SEX_ANOEST 24	0.0000000	0.00000000	.	.
SEX_ANOEST 25	0.0000000	0.00000000	.	.
SEX_ANOEST 26	0.0000000	0.00000000	.	.
SEX_ANOEST 29	0.0000000	0.00000000	.	.
MORA_QTDBANH 11	-0.1665864	0.17713811	-0.94	0.3489
MORA_QTDBANH 12	0.0182268	0.24563590	0.07	0.9410
MORA_QTDBANH 13	0.0000000	0.00000000	.	.
MORA_QTDBANH 21	-0.1963711	0.44475474	-0.44	0.6596
MORA_QTDBANH 22	0.1674221	0.48520713	0.35	0.7307
MORA_QTDBANH 23	0.2879435	0.52839365	0.54	0.5868
MORA_QTDBANH 24	0.0000000	0.00000000	.	.
MORA_QTDBANH 31	-0.6590378	0.28676399	-2.30	0.0233
MORA_QTDBANH 32	-0.4582762	0.30862460	-1.48	0.1403
MORA_QTDBANH 33	-0.6699606	0.32121283	-2.09	0.0392
MORA_QTDBANH 34	0.0000000	0.00000000	.	.
MORA_QTDBANH 41	-0.5623631	0.29767102	-1.89	0.0613
MORA_QTDBANH 42	-0.7344848	0.32134619	-2.29	0.0241
MORA_QTDBANH 43	-0.4633427	0.33449279	-1.39	0.1686

Coefficientes estimados da regressão				
Parâmetro	Estimativa	Erro Padrão	Valor t	P-valor
MORA_QTDBANH 44	0.0000000	0.00000000	.	.
MORA_QTDBANH 51	0.0000000	0.00000000	.	.
MORA_QTDBANH 52	0.0000000	0.00000000	.	.
MORA_QTDBANH 53	0.0000000	0.00000000	.	.
MORA_QTDBANH 54	0.0000000	0.00000000	.	.

The denominator degrees of freedom for the t tests is 117. Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

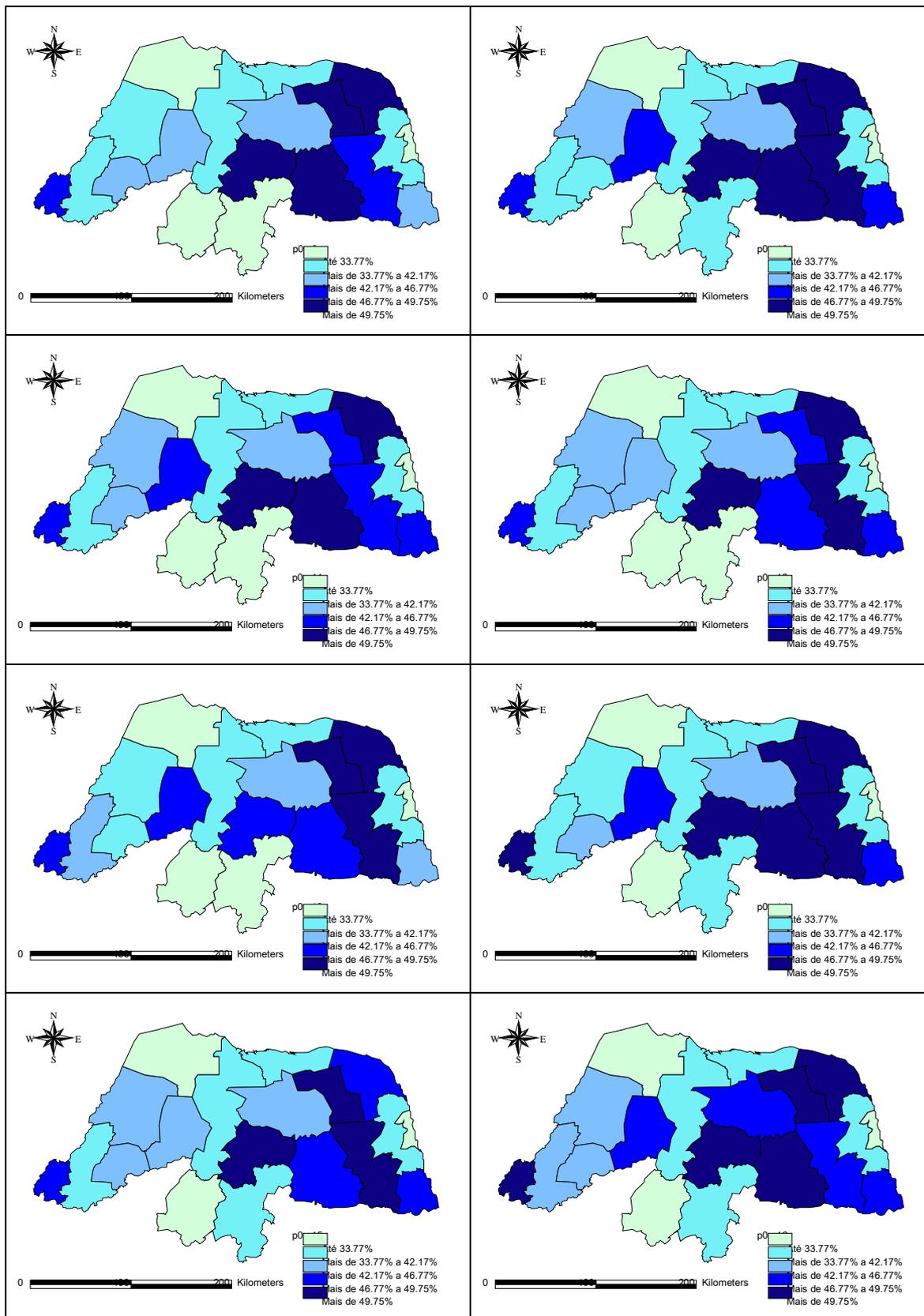
Anexo 8.4:

Mapas de 24 réplicas da renda imputada pela metodologia ELL. (continua)



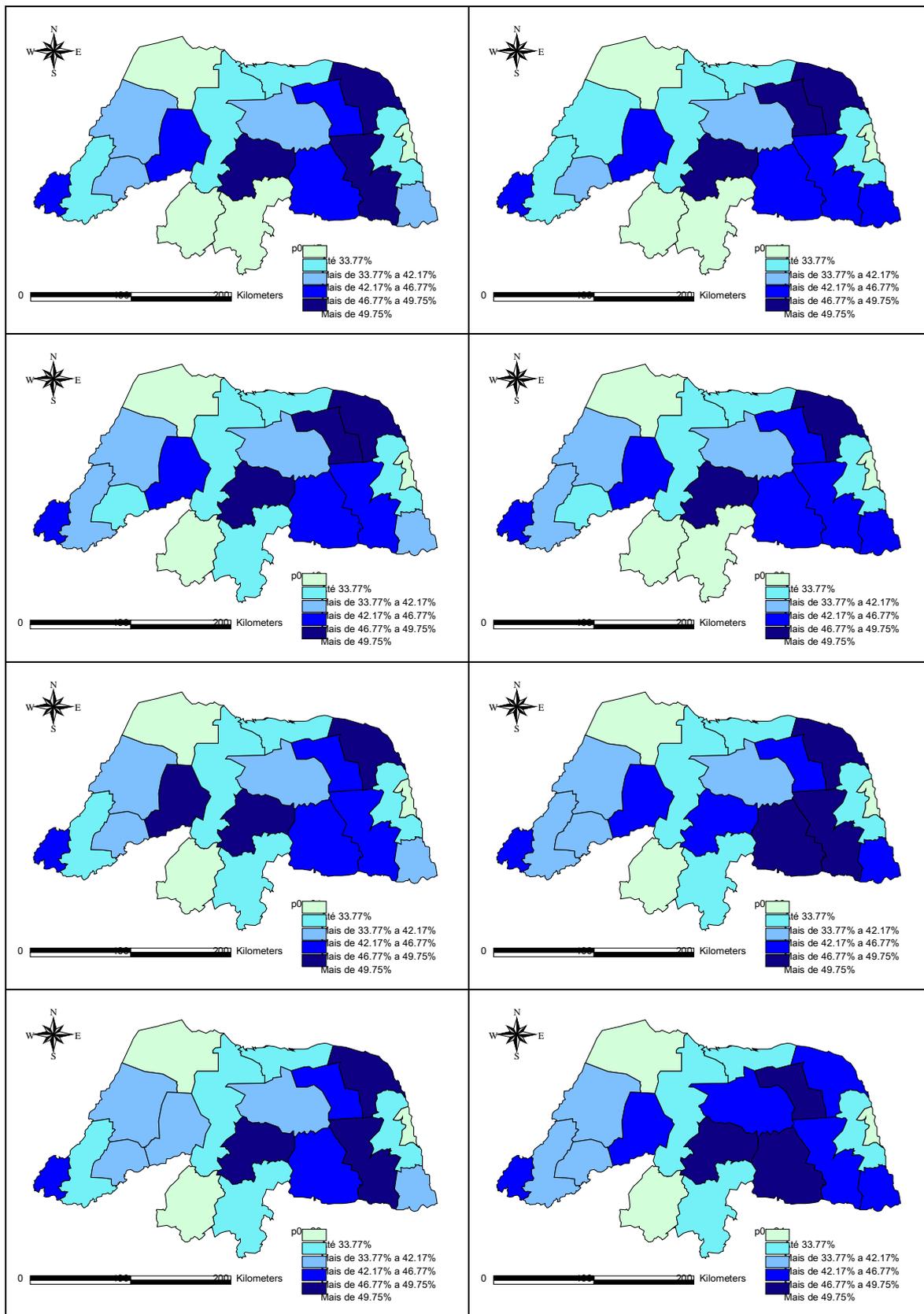
Anexo 8.4:

Mapas de 24 réplicas da renda imputada pela metodologia ELL. (continua)



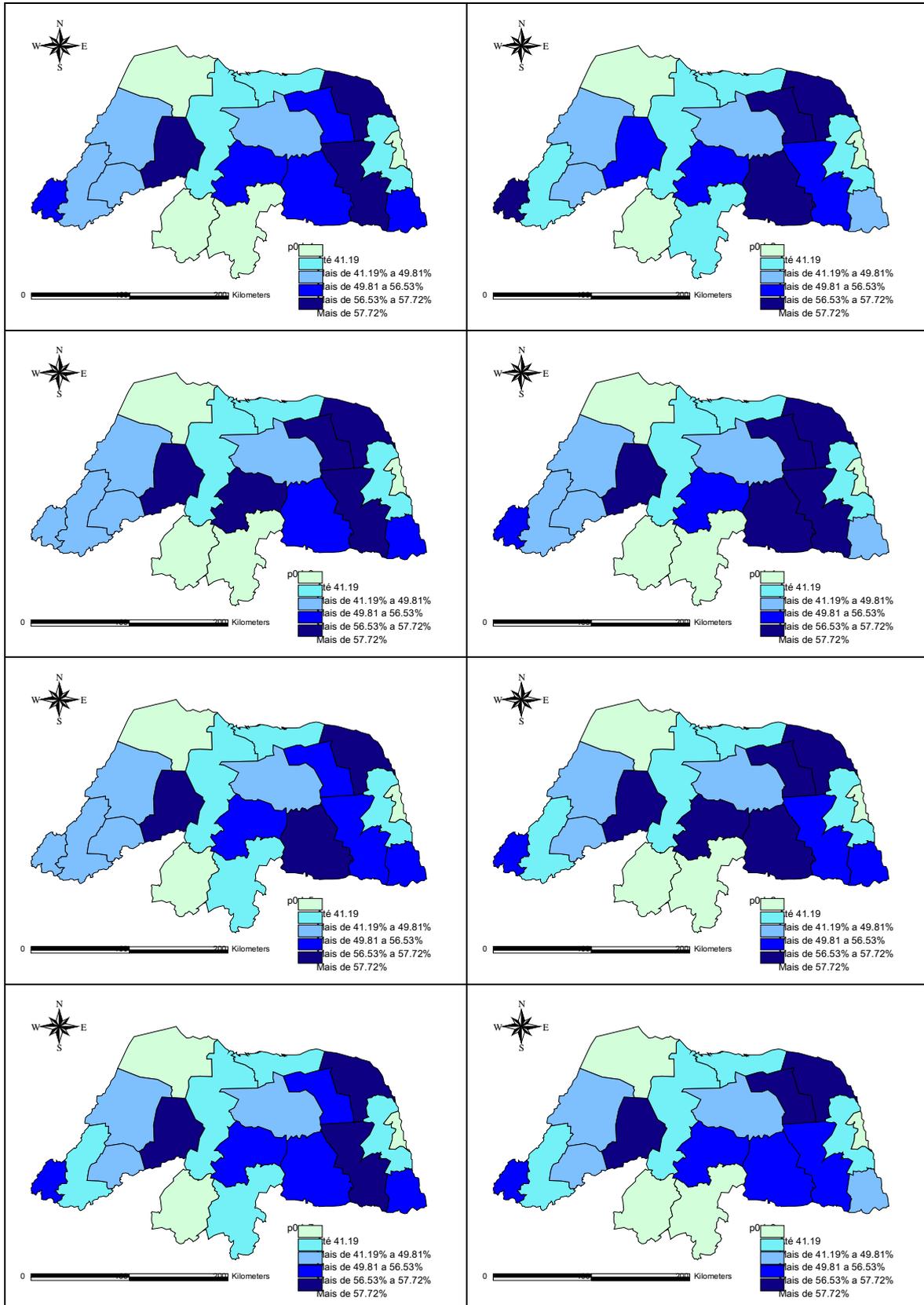
Anexo 8.4:

Mapas de 24 réplicas da renda imputada pela metodologia ELL. (Continuação)



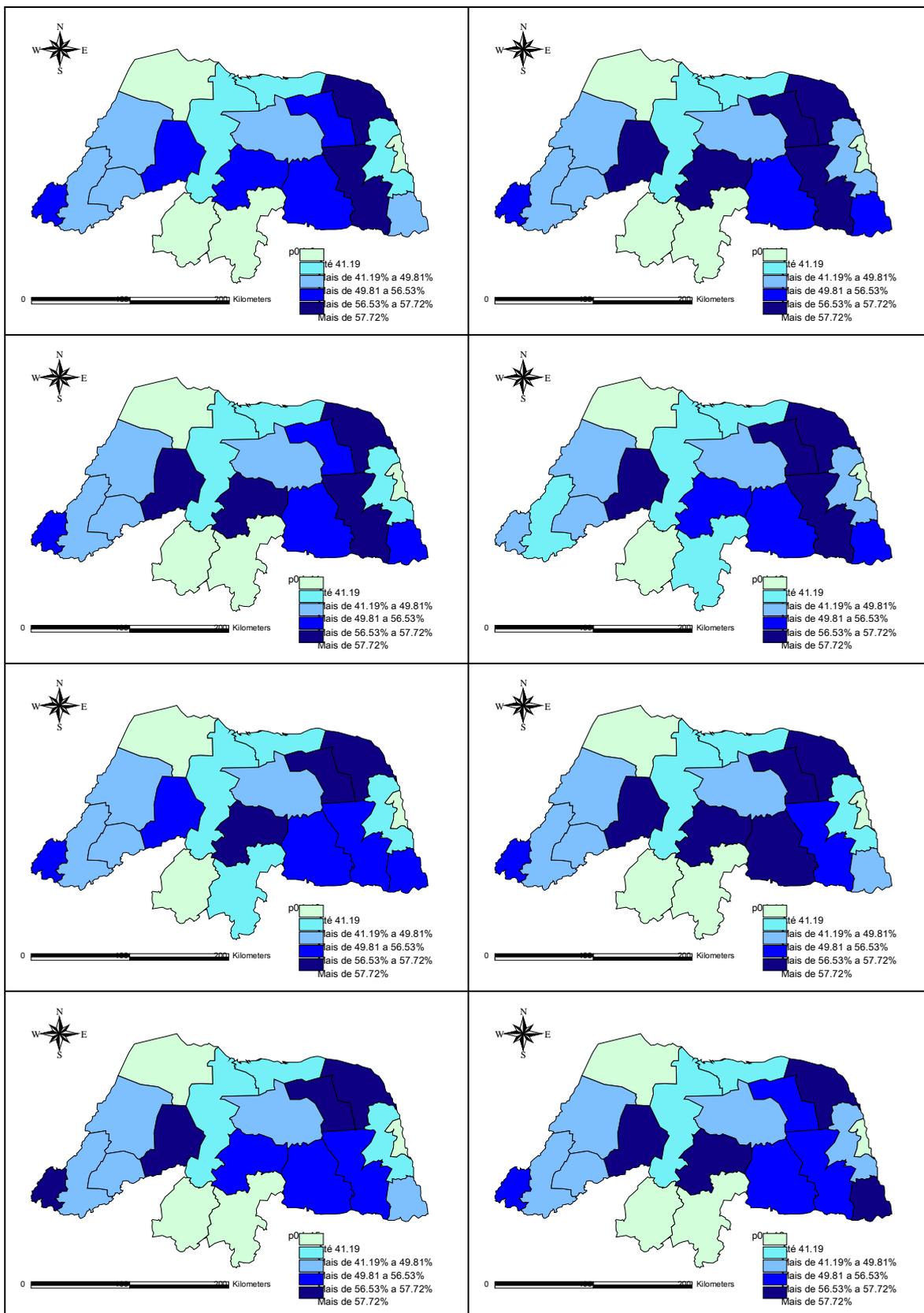
Anexo 8.5:

Mapas de 24 réplicas da despesa imputada pela metodologia ELL. (continua)



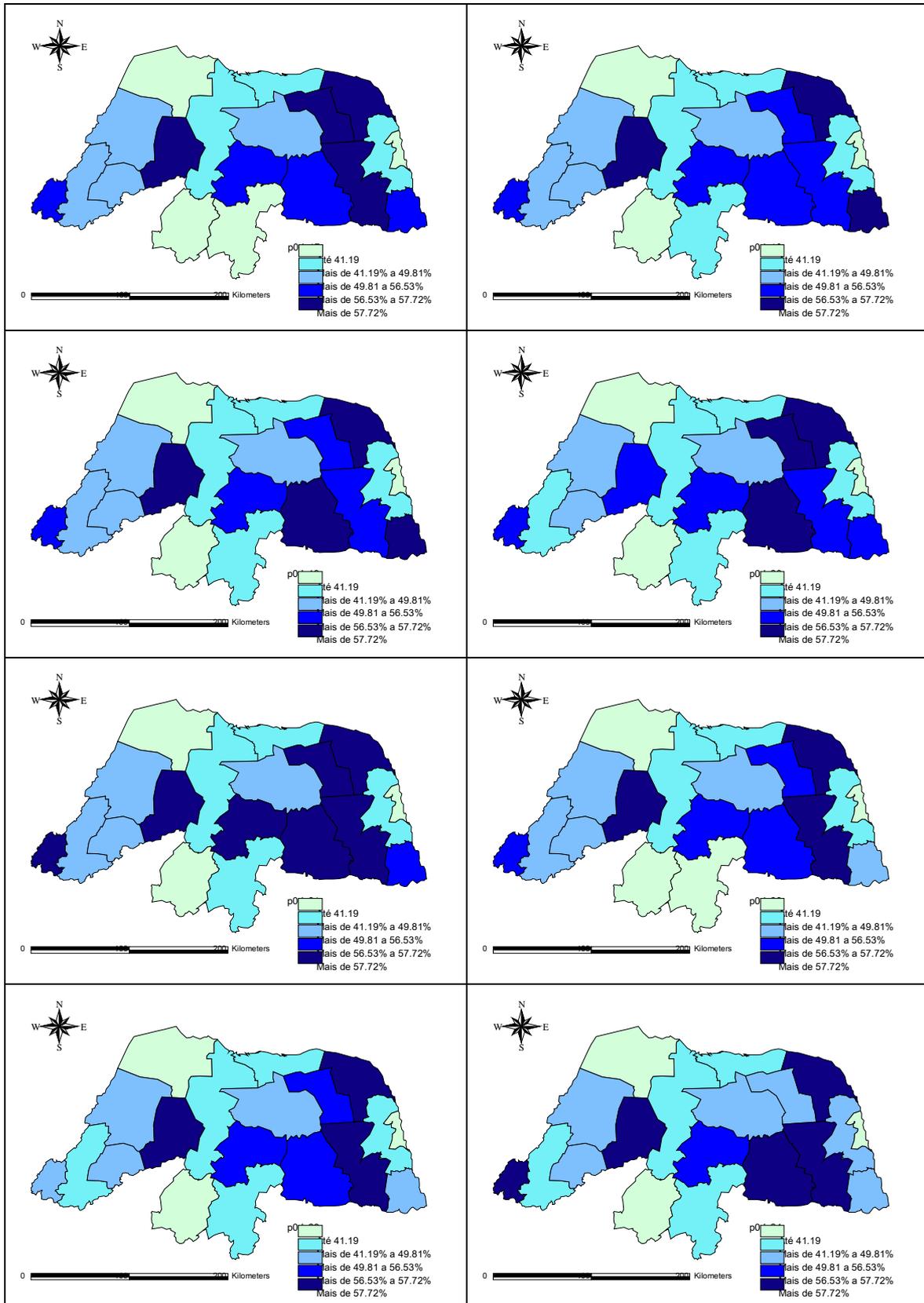
Anexo 8.5:

Mapas de 24 réplicas da despesa imputada pela metodologia ELL. (continua)



Anexo 8.5:

Mapas de 24 réplicas da despesa imputada pela metodologia ELL. (Continuação)



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](#)