UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

**DIFICULDADES ORÇAMENTÁRIAS BÁSICAS DAS FAMÍLIAS BRASILEIRAS:
UM CONVITE À REFLEXÃO A PARTIR DE REDES BAYESIANAS**

Cláudia Mendes Nogueira

Orientador: Prof Dr. Hiroo Takaoka

SÃO PAULO
2012

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Reinaldo Guerreiro
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Adalberto Américo Fischmann
Chefe do Departamento de Administração

Prof. Dr. Lindolfo Galvão de Albuquerque
Coordenador do Programa de Pós-Graduação em Administração

CLÁUDIA MENDES NOGUEIRA

**DIFICULDADES ORÇAMENTÁRIAS BÁSICAS DAS FAMÍLIAS BRASILEIRAS :
UM CONVITE À REFLEXÃO A PARTIR DE REDES BAYESIANAS**

Tese apresentada ao Departamento de Administração na Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Hiroo Takaoka

Versão Corrigida

(Versão original disponível na Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2012

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Nogueira, Cláudia Mendes

Dificuldades orçamentárias básicas das famílias brasileiras: um convite à reflexão a partir de redes bayesianas / Cláudia Mendes Nogueira - São Paulo, 2012.

263 p.

Dissertação (Doutorado) – Universidade de São Paulo, 2012.

Orientador: Hiroo Takaoka

1. Inteligência artificial 2. Redes bayesianas 3. Probabilidade 4. Rendas 5. Consumo 6. Condições de Vida 7. Família – Estrutura 8. Classes econômicas I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 006.3

Aos queridos
Clayton, Vitor, Ana Luisa e Beatriz

AGRADECIMENTOS

Ao meu orientador **Prof. Dr. Hiroo Takaoka**, por me acolher com paciência e sabedoria oriental. Ao **Prof. Dr. Ronaldo Zwicker**, em memória, por ter ajudado na escolha do tema e pela confiança que depositou em mim.

Aos professores da FEA que muito representaram nesta jornada, em especial : à **Prof. Dra. Maria Aparecida Gouvea**, pela didática e vocação para o ensino ; Ao **Prof. Dr. Abraham Yu**, que me apresentou temas desafiadores em processos decisórios; aos professores da banca de qualificação : **Prof. Dr. Nuno Fouto** e **Prof. Dr. Adolpho Walter P. Canton**. Aos meus colegas da FEA, especialmente **Fernando Alves de Moura, Jorge Pasquotto, Marcelo Araujo, Rosária Russo** pelo apoio nas horas difíceis e pelas realizações cooperadas.

A toda a equipe da **secretaria de pós graduação**, sempre atenta às nossas necessidades e em especial à Cida pelas orientações precisas.

À **Fundação Instituto de Administração (FIA)** por ter financiado o software utilizado, sem isso seria difícil a evolução. Um agradecimento especial à Licia, sinônimo de competência e assertividade.

Ao **Lionel Jouffe**, Phd e responsável pelo software BayesiaLab, por ter respondido perguntas técnicas de “bate-pronto” e liberado o uso do software para a finalização desta tese.

Ao **André Martins**, do IBGE, por ter esclarecido muitas dúvidas críticas sobre a POF.

Aos amigos da Oficina de Valor Consultoria : **Helena Gomes Marinotti**, que me ajudou a adentrar no mundo da POF com tanta paciência, dedicação e precisão. **Melaine Cristina de Oliveira** pela paciência com minhas teimosias na área, tentando esclarecer dúvidas apesar das minhas limitações técnicas ; **Pedro Paulo Cunha**, por todo suporte, apoio na minha ausência e pelas perguntas críticas durante a tese, as quais me fizeram repensar os caminhos escolhidos.

Ao **Sr. Cláudio Heitor Moreira de Alvarenga**, mestre do início da carreira que me fez conhecer os métodos quantitativos aplicados efetivamente na prática.

Aos meus pais **José e Ida**, por me ensinarem desde cedo a valorizar a educação como forma de crescimento e desenvolvimento pessoal.

À **Maria Lidia e Darzísio** pelo apoio incondicional em todas as fases da minha vida.

À **Alba e João** pela amizade, camaradagem e prontidão de sempre.

À **Patricia** pela simpatia e suporte nesta reta final.

Ao meu marido **Clayton**, companheiro de todas as horas, sem ele eu não conseguiria chegar onde estou, em todos os sentidos. Ao meu filho **Vítor** por ter me apresentado “bibliografia improvável”. Às minhas meninas : **Ana Luísa** pelo carinho e preciosos comentários e **Beatriz** pelo apoio e alegria em período tão difícil para mim.

“Essentially, all models are wrong, but some are useful. ”

George Box & Norman Draper

RESUMO

Este estudo visa compreender a adequação dos rendimentos às necessidades e condições de vida dos brasileiros. Observando os dados da Pesquisa de Orçamentos Familiares (POF) realizada pelo IBGE (Instituto Brasileiro de Geografia e Estatística) para o período: 2008 e 2009, o estudo identifica um modelo que se concentra na investigação sobre o fato de 75% dos domicílios brasileiros declararem dificuldades orçamentárias. Para desenvolver um modelo, foi utilizada a percepção declarada e subjetiva de adequação da renda, informada pelo chefe de família ou pessoa de referência no domicílio. O referencial teórico baseia-se no comportamento do consumidor e foca nos recursos econômicos. O método quantitativo foi desenvolvido com Inteligência Artificial, mais especificamente Redes Bayesianas. Redes Bayesianas são estruturas em forma de grafos onde as distribuições de probabilidade são representadas por nós ligados por arcos acíclicos, que podem representar ou não relações causais entre as variáveis. No final pretende-se contribuir para o conhecimento e melhoria no desenho de políticas públicas e para as empresas em geral, dando um panorama sobre o que afeta as dificuldades das famílias, proporcionando uma visão que vai além da tradicional divisão de classes econômicas.

Palavras Chave : Renda, Consumo, Condições de Vida, Estrutura Familiar, Classes Econômicas, Inteligência Artificial, Probabilidade, RedesBayesianas.

ABSTRACT

This study aims to understand the adequacy of Brazilians' income to their needs and living conditions. According to the data from the Household Budget Survey (POF) conducted by IBGE (Brazilian Institute of Geography and Statistics) for the years of 2008 - 2009, the study identifies a model which focuses on the investigations about the fact that 75% of Brazilian households reported budgetary difficulties. To develop a model, was used the perceived adequacy of income declared by the householder or reference person in the household. The theoretical framework was based on consumer behavior and focuses on economic resources. The quantitative method was developed by Artificial Intelligence, specifically Bayesian Networks. Bayesian Networks are structures in the form of graphs for which the probability distributions are represented by nodes connected by acyclic arcs, which may or may not represent causal relationships between variables. At the end we intend to contribute to knowledge and improvement in the design of public policies and business in general, giving a more detailed look at what affects the difficulties of families, providing a vision that goes beyond the traditional division of economic classes.

Keywords: Income, Consumption, Living Conditions, Family Structure, Economic Classes; Artificial Intelligence, Probabilities, Bayesian Networks

CONTEÚDO

CONTEÚDO	1
LISTA DE ABREVIATURAS E SIGLAS	4
LISTA DE TABELAS	5
LISTA DE GRÁFICOS.....	7
LISTA DE ILUSTRAÇÃO / FIGURAS	8
1 INTRODUÇÃO E FORMULAÇÃO DO PROBLEMA DE PESQUISA	11
1.1 Relevância do Tema.....	13
1.2 Objetivo do Trabalho e Problema da Pesquisa	15
1.3 Descrição dos capítulos.....	17
2 MOTIVAÇÃO TEÓRICA.....	19
2.1 Recursos do Consumidor e Ciclo de Vida das Famílias	19
2.2 Alocação do orçamento familiar	21
2.3 Classes Econômicas	22
2.4 A POF -Pesquisa de Orçamentos Familiares	25
3 ASPECTOS METODOLÓGICOS	31
3.1 Complexidade	32
3.2 Redes Bayesianas - Introdução	33
3.2.1 Redes Bayesianas - Nomenclatura e Controvérsias	35
3.2.2 Redes Bayesianas -Justificativa para utilização.....	39
3.2.3 Redes Bayesianas -Conexões Básicas	40
3.2.4 A aprendizagem com Redes Bayesianas.....	42
3.3 Teoria da Informação e Entropia	47
3.3.1 Métricas Utilizadas: Entropia Condicional, Relativa e Informação Mútua.	50
3.4 Uso de um software	52
3.4.1 Algoritmos de classificação disponíveis	53
3.4.2 Algoritmos para descoberta de associações entre variáveis :	56
3.4.3 Parâmetros para avaliar redes bayesianas para efeito de classificação:.....	57

3.4.4	Estrutura das Redes Bayesianas	61
3.5	Redes Bayesianas - Relações Causais.....	63
3.6	Tratamento de variáveis para redução da complexidade	66
3.7	Passos para a mineração de dados	68
3.8	Esquema metodológico para o desenvolvimento do trabalho.....	71
4	APRESENTAÇÃO DA MODELAGEM E ANÁLISE DE RESULTADOS.....	73
4.1	Objetivos da Modelagem	73
4.2	Considerações Iniciais	74
4.3	Preparação e seleção de variáveis:	77
4.3.1	Discretização de Variáveis e Fusão de Categorias.	77
4.3.2	Agrupamento e Eliminação de Variáveis	84
4.4	Primeira Visão da Floresta.....	85
4.4.1	Renda e informalidade no trabalho	101
4.4.2	Presença de cônjuge e gênero	103
4.4.3	Adimplência no pagamento de despesas.....	104
4.4.4	Idosos com aposentadoria, empregadores , previdência privada.....	105
4.4.5	Despesas Básicas, Diferenciadas e Necessidades Específicas de Saúde	106
4.4.6	Moradia e Alimentação	108
4.4.7	Classe Econômica, escolaridade e despesas diferenciadas.	108
4.4.8	Transporte público, acesso à educação, à saúde e ao lazer.	110
4.4.9	Condições físicas das moradias	111
4.4.10	Habitação e Comprometimento de Renda	111
4.4.11	Quantidade de Pessoas na Família, ciclo de vida familiar e Educação	113
4.4.12	Acesso aos serviços básicos.....	114
4.4.13	Renda de Transferências do Governo	115
4.4.14	Raça, cor da pele e região geográfica.	116
4.5	Visão Seleccionada da Floresta	119
4.5.1	Primeiro Markov Blanket	119

4.5.2	Segundo Markov Blanket	123
4.5.3	Variáveis incluídas além do Markov Blanket.....	128
4.5.4	Processo final de seleção de variáveis:	129
4.5.5	Agrupamento de variáveis para a seleção dos nós do(s) modelo(s) finais	130
4.5.6	Interferência na heurística dos modelos -arcos proibidos.....	135
4.5.7	Interferência na heurística dos modelos - presença e sentido dos arcos	140
4.6	Modelos Finais.....	140
4.7	Modelo Final Selecionado e proposição ajustes	148
4.8	Análise Geral do Modelo	157
4.9	Agrupamentos por dificuldade.....	179
4.10	Validação da Rede	191
4.11	Validação contra a variável renda média per capita:	191
4.12	Validação dos arcos do modelo	196
4.13	Validação pela verossimilhança:.....	201
5	REFLEXÕES A PARTIR DAS REDES: QUESTÕES PARA NOVOS ESTUDOS.	203
5.1	“Ranking” de probabilidade de estar em dificuldade e a alocação dos recursos de transferências do governo.....	203
5.2	Cotas raciais nas universidades e a estrutura observada na rede bayesiana.	206
6	LIMITAÇÕES DO ESTUDO E SUGESTÕES DE MELHORIA	213
7	CONCLUSÃO.....	215
	REFERÊNCIAS BIBLIOGRÁFICAS SELECIONADAS.....	217
	REFERÊNCIAS BIBLIOGRÁFICAS COMPLEMENTARES	233
	APÊNDICE 1A: VARIÁVEIS POR ORDEM ALFABÉTICA	237
	APÊNDICE 1B: VARIÁVEIS POR ORDEM NUMÉRICA	245
	APÊNDICE 2 : VARIÁVEIS SELECIONADAS E RELAÇÃO COM A DIFICULDADE	251

LISTA DE ABREVIATURAS E SIGLAS

ABA - Associação Brasileira de Anunciantes

ABIPEME - Associação Brasileira dos Institutos de Pesquisa de Mercado

ABEP - Associação das empresas de Pesquisa

ANB - Rede Bayesiana Aumentada

AUROC - Área sobre a curva ROC

DIF - Diferenciados ou diferenciadas

DKL - Divergência de Kullback-Leibler

IDH - Índice de Desenvolvimento Humano

LSE - Levantamento Sócio Econômico

MB - Markov Blanket

MB1- Markov Blanket 1

MB2 - Markov Blanket 2

MDL - Minimum Descriptive Length

PIB - Produto Interno Bruto

PNAD - Pesquisa Nacional de Amostra por Domicílios

POF - Pesquisa de Orçamentos Familiares

POF 6 - Identifica o questionário que trata de condições de vida na POF

ROC - Receiver Operating Characteristics

TO - Tabu Order

V - Variável

WOE -Weight of Evidence

LISTA DE TABELAS

Tabela 1 - Comparativo Entropia	47
Tabela 2 - Comparativos entre métodos de classificação	58
Tabela 3 - Tabela com exemplo de Verdadeiros Positivos e Verdadeiros Negativos.....	59
Tabela 4 - AUROC e a discriminação de modelos	61
Tabela 5 - Percepção de “dificuldade”	76
Tabela 6 - Percepção de Dificuldade 2003_2009.....	77
Tabela 7 - Fusão de Categorias - Análise Bivariada	79
Tabela 8 - Fusão de Categorias - resultado da categorização	79
Tabela 9 – Classificação das despesas diferenciadas - percentual sobre montante em Reais..	82
Tabela 10 - Categorização despesas com habitação.....	83
Tabela 11 - Origem das Rendas Familiares – percentual sobre o total da renda monetária .	83
Tabela 12 - Validação do Fator 3	98
Tabela 13 - Variável Área Geo	101
Tabela 14 - Primeiro Markov Blanket - Variáveis Seleccionadas - continuação	121
Tabela 15 - Variáveis Adicionais Seleccionadas no segundo Markov Blanket -	125
Tabela 16 - Comparação entre Markov Blanket (MB) 1 e 2.....	127
Tabela 17 - Mudanças entre o Primeiro e Segundo Markov Blanket	128
Tabela 18 -Variáveis agrupadas para suportar a seleção para os modelos finais.....	133
Tabela 19 -Variáveis Seleccionadas para os modelos finais.....	134
Tabela 20 - Variáveis Seleccionadas para o Modelo Final – Probabilidades Marginais	135
Tabela 21 - Causas e Sintomas, definidos a priori	137
Tabela 22 - Área sobre a curva ROC - modelos ANB e TO	144
Tabela 23 -Acerto Geral dos Modelos ANB e TO.....	147
Tabela 24 - Tabela com critério de escolha entre TO e ANB	148
Tabela 25 -Arcos Invertidos e Arcos Adicionados	152
Tabela 26 - Grupos de Dificuldade	179
Tabela 27 - Tipo de Alimento e Grupos de Dificuldade	181
Tabela 28- Quantidade de Alimento e Grupos de Dificuldade	181
Tabela 29 - Classes Econômicas (Critério Brasil -ABEP) e Grupos de Dificuldade.....	182
Tabela 30 -Atraso de despesas básicas e Grupos de Dificuldade	183
Tabela 31 - Necessitou de serviços de saúde ou medicamentos e Grupos de Dificuldade	183
Tabela 32 - Plano de Saúde e Grupo de Dificuldade	184
Tabela 33 - Condições da Moradia e Grupos de Dificuldade	184
Tabela 34 - Atraso em prestações e Grupos de Dificuldade	185
Tabela 35 -Situação na Rede de Ensino e Grupos de Dificuldade.....	185
Tabela 36 - Grupos de Unidades da Federação - Grupos de Dificuldade	186

Tabela 37 - Quantidade de Moradores -Grupos de Dificuldade	186
Tabela 38 - Transferências do Governo - Grupos de Dificuldade	187
Tabela 39 - Raça e Cor - Grupos de Dificuldade	187
Tabela 40 -Tipos de Família -Grupos de Dificuldade.....	188
Tabela 41 - Área geográfica - Grupo de Dificuldade.....	188
Tabela 42 - Acesso à Educação, Saúde e Lazer e Grupos de Dificuldade	189
Tabela 43- Contribuição para a previdência e Grupos de Dificuldade	189
Tabela 44 -Serviços de escoamento da agua de chuva -Grupos de Dificuldade.....	190
Tabela 45 Sexo/Gênero do responsável pela família -Grupos de Dificuldade	191
Tabela 46 -Resultados Gerais da validação dos arcos.....	198
Tabela 47 -Arcos Duvidosos	199
Tabela 48 - Validação do arco : previdência -> atraso de prestações	199
Tabela 49 -Validação do arco Unidades da Federação -> Grupo Fator 11	200
Tabela 50 –Arcos Questionáveis.....	201
Tabela 51 - Famílias beneficiadas - Bolsa família, grupos de UF e renda per capita – POF 2009	205
Tabela 52 - Famílias beneficiadas - Bolsa Família, por grupo de UF e probabilidade - POF 2009	205
Tabela 53 : Percentual de chefes de família que se declararam negros, pardos e indígenas..	207

LISTA DE GRÁFICOS

Gráfico 1 - Crises Internacionais, PIB e Consumo das Famílias	11
Gráfico 2 -Taxa de Dependência Demográfica no Brasil 1970 - 2090	12
Gráfico 3 -BRASIL : Despesas Familiares sobre Renda Média POF 2003 e 2009	21
Gráfico 4 -Critério Brasil - Comparativo entre Distribuição na ABEP e na POF	24
Gráfico 5 -Período de coleta da POF 2008-2009 e as variações no PIB.....	27
Gráfico 6 -Período de Coleta e Percepção Subjetiva sobre a Facilidade	28
Gráfico 7 : Distribuição das famílias, por referência à dificuldade	29
Gráfico 8 - Entropia -Moeda	49
Gráfico 9 - IDH x Dificuldade Percebida.....	80
Gráfico 10 - Contribuição das variáveis selecionadas pelo Markov Blanket	122
Gráfico 11 - Contribuição das variáveis selecionadas pelo Segundo Markov Blanket	125
Gráfico 12 - Dificuldade ; Classe Econômica; Quantidade de Pessoas na Família	126
Gráfico 13 -Erro na Estimativa da Probabilidade, modelo MB1	141
Gráfico 14 - Rede Naive Bayes Aumentada (ANB)	
Gráfico 15 -Rede Tabu Order (TO).....	142
Gráfico 16 -Comparativo entre as redes : ANB (Naive Bayes Aumentada) e TO (Tabu Order)	143
Gráfico 17 - Curva ROC : modelos ANB (Naive Bayes Aumentada) E TO (Tabu Order)..	144
Gráfico 18 -Tabu Order - Estimativa por percentil de probabilidade	145
Gráfico 19 - ANB - Estimativa por percentil de probabilidade	145
Gráfico 20 -Comparativo entre probabilidades estimadas pelos modelos	146
Gráfico 21 - Efeito do ajuste estrutural nas proporções conjuntas.....	151
Gráfico 22 - Ajustes no modelo selecionado -Arcos Adicionados e Removidos.	153
Gráfico 23 -Probabilidade de estar com “dificuldade” e percepção informada pelas famílias	157
Gráfico 24 : Informação mútua variáveis selecionadas para o modelo final	158
Gráfico 25 - Boxplot : probabilidade de estar em dificuldade por grupo	180
Gráfico 26 -Grupos de Dificuldade & Classes Econômicas	182
Gráfico 27 -Decis de renda per capita e percepção sobre a dificuldade	192
Gráfico 28 - Renda per capita média por décimos de probabilidade	193
Gráfico 29 - Múltiplos da Renda por percentil da probabilidade.....	194
Gráfico 30- Decis de renda per capita e grupos de dificuldade	195
Gráfico 31 - Decis de renda per capita e probabilidade de estar em dificuldade.....	196
Gráfico 32 - Redes de comparação : validação dos arcos	197
Gráfico 33 - Dificuldade percebida e variação entre raça e cor dentro das mesmas classes econômicas.	210
Gráfico 34 - Dificuldade percebida e variação entre raça e cor dentro das mesmas faixa de renda familiar per capita – em cada categoria estão 10% das famílias	211

LISTA DE FIGURAS E ILUSTRAÇÕES

Figura 1 - Agenciamento da complexidade.....	31
Figura 2 - Conexões em Redes Bayesianas e Condições de Markov.....	40
Figura 3 - D-Separações e Tipos de Conexões em Redes Bayesianas.....	41
Figura 4 - Aprendizado de Estrutura - Teoria da Informação -MDL.....	45
Figura 5 - Dados e Estruturas em Redes Bayesianas - Algoritmos Utilizados	46
Figura 6 -Diagrama de Venn para representação das entropias, entropias condicionais e	50
Figura 7 - Exemplo cálculo do DKL.....	51
Figura 8 - DKL - Exemplo - Clima.....	51
Figura 9 - Entropia Relativa e Força dos Arcos	52
Figura 10 -Naive Bayes.....	54
Figura 11 - Naive Bayes Aumentada	54
Figura 12 - Markov Blanket.....	55
Figura 13 - Exemplo interpretativo da curva ROC	60
Figura 14 - Exemplo Didático Rede Bayesiana : Condição da Moradia	61
Figura 15 - Exemplo Didático Rede Bayesiana -Distribuições Marginais -A Priori	62
Figura 16- Exemplo Didático Rede Bayesiana -Distribuições Marginais -A Posteriori.....	63
Figura 17 - Exemplo Didático Rede Bayesiana – Condição da Moradia.....	63
Figura 18 - Primitivas do Processo de Mineração de Dados.....	68
Figura 19 - Esquema para Mineração de Dados	71
Figura 20 - Máximo Spanning Tree - Todas as Variáveis	86
Figura 21 - Máximo Spanning Tree : Tipos de Renda _Previdência_Tempo Sem Trabalho	87
Figura 22 - Máximo Spanning Tree : Alimento, Vestuário, Moradia, Outras Despesas	88
Figura 23 - Máximo Spanning Tree - Escolaridade & Estrutura das Famílias	89
Figura 24 - Máximo Spanning Tree – Renda, Classe Economica , Moradia	90
Figura 25 - Máximo Spanning Tree - Renda, Adimplência, Moradia, Saúde.....	91
Figura 26 - Máximo Spanning Tree - Moradia e Acesso a Serviços Públicos.....	92
Figura 27 - Máximo Spanning Tree - Acesso a Serviços Públicos Básicos	93
Figura 28 - Força dos Arcos.....	94
Figura 29 - Algoritmo EQ aplicado sobre a rede Máximo Spanning Tree	95
Figura 30 - Agrupamento de Variáveis	97
Figura 31- Representação do Fator 3 com Variável Latente.....	97
Figura 32 - Priori e Posteriori - Fator 3 & Serviço de Agua	98
Figura 33 - G Teste e relação do mesmo com Informação Mútua e Entropia Cruzada	99
Figura 34 - Fator 3 -Urbano, Rural, Acesso à agua.....	100
Figura 35 - Primeiro Markov Blanket (MB1) -Variáveis Seleccionadas.....	120
Figura 36 - Segundo Markov Blanket (MB2) com Variáveis Seleccionadas	124

Figura 37 - Rede Naive Aumentada com as 33 variáveis selecionadas após MB1 e MB2. .	131
Figura 38 - Destaque dos 10 grupos identificados na Rede Naive Bayes Aumentada	133
Figura 39 - Área Geográfica e Sexo do Chefe da Família	149
Figura 40 - Area Rural e Sexo do Chefe da Família	150
Figura 41 - Análise da relação entre variáveis - identificação de problemas nas distribuições conjuntas.....	150
Figura 42 - Melhor representação das distribuições conjuntas após ajuste estrutural na rede	151
Figura 43 - Mudança do sentido do arco na probabilidade calculada: simulação de 1 caso.	154
Figura 44 - Variáveis para representar efeitos regionais e locais na Rede Bayesiana	155
Figura 45 - Mudança do sentido do arco para melhor representar Transferências do Governo	156
Figura 46 - Variáveis Instanciadas: Tipo e Quantidade de Alimento, reflexo nas demais ...	159
Figura 47 - Variável Instanciada na Rede : Plano de Saúde e reflexo nas demais.....	159
Figura 48 - AREA_GEO	160
Figura 49 - ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE	161
Figura 50 – Classe ABEP.....	162
Figura 51 – DIFICULDADE_REC.....	164
Figura 52 – FATOR_10_FAMILIAS	165
Figura 53 – FATOR_11_EDUC_SAUDE_LAZER-.....	166
Figura 54 FATOR_12_GOTEIRA_FUNDACAO_PAREDE_ETC.....	167
Figura 55 - FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	168
Figura 56 - FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	169
Figura 57 - PREVIDENCIA	170
Figura 58 - QTDE_MORADORES	171
Figura 59 - QUANTIDADE_DE_ALIMENTO_REC.....	172
Figura 60 - RACA_COR_PESSOA_REFERENCIA	173
Figura 61 - SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	174
Figura 62 - SEXO DO CHEFE DA FAMÍLIA.....	175
Figura 63 - STATUS_REDE_ENSINO.....	176
Figura 64 - TEM_PLANO_DE_SAUDE.....	177
Figura 65 - UF_GRUPO	178
Figura 66 : Raça e Cor e a Dificuldade - Modelo TO	208
Figura 67 : Raça e Cor, Renda Per Capita, Classe Econômica e Dificuldade.	209

LISTA DE QUADROS

Quadro 1 – Renda e informalidade no trabalho	102
Quadro 2 – Presença de cônjuge e gênero.....	104
Quadro 3 – Adimplência e pagamento de despesas	104
Quadro 4 – Aposentados, empregadores e previdência privada.	105
Quadro 5 – Despesas básicas, diferenciadas e necessidades de saúde (continua...)	106
Quadro 6 – Moradia e Alimentação.	108
Quadro 7 – Classe Econômica, escolaridade,despesas diferenciadas e veículos.	109
Quadro 8 – Transporte público, acesso à educação, saúde e lazer.....	110
Quadro 9 – Condições físicas da moradia.	111
Quadro 10 – Habitação e comprometimento de renda.	112
Quadro 11 – Quantidade de pessoas na família, ciclo de vida familiar e educação.....	114
Quadro 12 – Acesso aos serviços básicos	115
Quadro 13 – Transferências do Governo	116
Quadro 14 – Cor da pele, religião e região geográfica.....	118
Quadro 15 – Variáveis que reduzem a incerteza sobre a dificuldade.	132

1 INTRODUÇÃO E FORMULAÇÃO DO PROBLEMA DE PESQUISA

O Brasil nos últimos anos vivencia grandes mudanças em sua estrutura social e econômica, migrando para uma posição de destaque internacional (THE ECONOMIST, 2009). Os anos 90 foram marcados por reformas estruturais que impulsionaram a privatização e reorientaram a política social (LAMOUNIER & SOUZA, 2010). Após o Plano Real, o Brasil sofreu com diversas crises mundiais, como apresenta o Gráfico 1. Estas crises tiveram origem em lugares diferentes no mundo incluindo o México (Dezembro de 1994), a Ásia (Julho de 1997), a Rússia (Agosto de 1998) os Estados Unidos, com o colapso do fundo LTCM, (*Long-Term Capital Management*; Setembro de 1998). Na sequência veio a Bolha da Internet em 2001. Após 2003, favorecido pela inexistência de crises internacionais, o Brasil ganhou fôlego, ampliou o PIB (Produto Interno Bruto), acelerou medidas de impacto social chegando em 2008 em novas condições, apenas abaladas pela crise internacional de 2009, seguida de uma recuperação favorecida pela Economia Chinesa e por medidas internas para a manutenção e estímulo ao consumo.

Gráfico 1 - Crises Internacionais, PIB e Consumo das Famílias

Fonte: Inspirado em THE ECONOMIST (2009); Dados IPEA (2011)

Esta tese olha para 2008-2009, faz breve comparação sobre a percepção das famílias em relação a suas dificuldades orçamentárias em 2002-2003 e convida a olhar para 2030. O ano de

2030 é apenas uma referência simbólica do futuro próximo, período coincidente com previsões para a inversão na taxa de dependência populacional, representada pelo Gráfico 2. Esta informação é uma compilação de dados apurados em 2010 pelo Departamento de Economia e Assuntos Sociais das Nações Unidas, os quais já haviam sido apresentados nos estudo de Alves (2006).

Gráfico 2 -Taxa de Dependência Demográfica no Brasil 1970 - 2090

FONTE:Departamento de Assuntos Econômicos e Sociais das Nações Unidas - UNPP 2010

Como bem apresenta este pesquisador e demógrafo a alteração da estrutura etária populacional acontece em fases sequenciais: (1) reduzindo o peso relativo das crianças e aumentando, em primeiro lugar, o peso dos adultos; (2) período posterior, onde aumenta o peso dos idosos. Desta forma percebe-se pelo Gráfico 2, num primeiro momento, a diminuição na taxa de dependência demográfica. Com a permanente queda da fecundidade, a taxa de dependência foi se reduzindo e deve permanecer em seu valor mais baixo entre 2015 e 2025. Neste período, a taxa de dependência de crianças (0 a 14 anos) continuará caindo persistentemente, no entanto a elevação da dependência dos idosos (65 anos e mais) aumenta. A partir de 2025/2030 esta taxa volta a se elevar, por conta dos idosos e termina o momento de “Bônus Demográfico”. O estudo “Envelhecendo em um Brasil mais Velho” (BANCO MUNDIAL, 2011) traz dados e análises preocupantes das implicações do envelhecimento populacional para o crescimento econômico, redução da pobreza, finanças públicas e prestação de serviços.

1.1 Relevância do Tema

Atualmente os meios de comunicação no Brasil noticiam o nascimento de uma “nova classe média”. Ilustram as reportagens com pessoas sorridentes, felizes pelas realizações de consumo. Estas notícias tomam como base estudos conduzidas para compreender e definir a Classe Média, ora denominado “Classe C” ou a “Classe Baixa”, ora nomeados como advindos das Classes “D e E”. As notícias veiculadas nos principais meios de comunicação colaboram para simplificar a visão sobre os consumidores brasileiros, criando rótulos para aglomerados com milhões de pessoas, sem esclarecer que nem sempre apresentam as mesmas condições de vida, consumo, renda e estrutura familiar. Em abril de 2011 este debate tomou dimensões interessantes até dentro dos partidos políticos. Após artigo do ex-presidente Fernando Henrique Cardoso (2011), os jornais publicaram resultados sobre a preferência das “classes econômicas” pelos atuais partidos (FRANCO, 2011). O próprio ex-presidente acrescentou em seu texto original: “ Existe toda uma gama de classes médias [...] às quais se soma o que vem sendo chamado sem muita precisão de “classe C” ou de nova classe média”.

De acordo com Lamounier e Souza (2010, p. 157) a classe C deixou de ser “baixa” e começa a ser “média”, disputando espaço com os estratos situados acima dela: as “classes médias tradicionais”. Apesar do título do livro conter a expressão: “A Classe Média Brasileira”, os autores não apresentam apenas uma medida para “operacionalizá-la” em classes econômicas. Utilizam diversos critérios separadamente ou de forma combinada, ilustrados com resultados multifacetados da pesquisa. Analisam os dados por renda familiar, educação, padrão de consumo, ocupação, identidade de classe. Para os autores “tentar definir classe média em termos que possam ser aceitos por todos os pesquisadores é cortejar a frustração” (LAMOUNIER E SOUZA, 2010, p. 21). Este estudo sugere que as classes econômicas para serem compreendidas precisam ser analisadas de forma multidimensional.

Para Neri (2011, p. 83) “ os limites da Classe C seriam as fronteiras para o lado indiano e para o lado Belga da nossa Belíndia”. Ou seja, existe um universo de pessoas diferentes dentro da Classe Média. Seu trabalho revela um lado bastante promissor sobre os avanços em direção à diminuição de desigualdades e ascensão das classes mais baixas para a “classe média”.

A Secretaria de Assuntos Estratégicos (SAE) divulgou em 2012 novos critérios para identificação da classe média brasileira, utilizando o critério renda per capita familiar para defini-la. De acordo com Ricardo Paes de Barros em matéria publicada no jornal Valor Econômico, reportagem de Pedroso (2012), a Classe Média é composta por famílias com renda per capita mensal entre R\$291 a R\$ 1019, representando 54% da população brasileira. Acrescenta Paes de Barros que o estudo deverá ser divulgado em meados de 2012 e adianta que a “Nova Classe Média” será dividida em três grupos. Além destas subdivisões fala-se em outras classes, visando representar as demais populações aquém e além da classe média. Pode-se depreender da matéria, com alguns ajustes necessários ao entendimento, que o Brasil estaria assim configurado:

Extremamente pobres.....	Até R\$81;
Pobres, mas não extremamente pobres....	R\$81 a R\$162;
Vulneráveis.....	R\$162 a R\$291;
Baixa Classe Média	R\$291 a R\$ 441;
Classe Média.....	R\$441 a R\$ 641;
Classe Média Alta.....	R\$641 a R\$ 1.019;
Baixa Classe Alta.....	R\$1.019 a R\$ 2.480;
Classe Alta.....	Acima de R\$ 2.480.

Outra questão em pauta em 2010 e 2011 foi a definição da linha da pobreza. Muitos esforços são realizados para publicar um limiar, geralmente discutido por estudiosos para orientar as políticas públicas. Estudos neste sentido podem ser observados em Neri (2011) e recentemente pela Secretaria de Assuntos Estratégicos (SAE), conforme citado por PEDROSO (2012). No entanto, não se espera que esta linha corrobore com a autoidentificação das pessoas com uma determinada classe social (LAMOUNIER & SOUZA, 2010, p. 21).

Apesar da renda per capita ser um indicador bastante objetivo para mensurar necessidades e desigualdades, e talvez o mais preciso, a sua obtenção é bastante difícil e controversa. Por exemplo, relevante estudo realizado pelo IPEA, Instituto de Pesquisa Econômica Aplicada (BARROS et al., 2007) aponta as diferenças entre rendas calculadas pela POF (Pesquisa de Orçamentos Familiares) e PNAD (Pesquisa Nacional de Amostra de Domicílios). Neste estudo a renda total da POF era 26% maior que a calculada pela PNAD, sendo que 63% da diferença podia ser explicada por conta de outras rendas do trabalho, não mensuradas pela PNAD. Estas rendas são: auxílio-refeição, auxílio-transporte, décimo terceiro salário, adicio-

nal de férias, horas extras, dentre outras. Outros 14% foram atribuídos às rendas de ativos como aluguéis, juros, dividendos e outros 11% às rendas não monetárias.

Na esfera das pesquisas de mercado, fora do setor público, a imprecisão é maior. Além da difícil obtenção, ao se perguntar sobre a renda mensal de uma pessoa ela, por exemplo, não considera toda a renda do ano, incluindo férias, décimo terceiro e divide mentalmente pelo número de meses, para informar ao pesquisador. Se ela está desempregada, vai informar que não tem renda. Neste sentido os institutos de pesquisa de opinião, acabam criando critérios de aproximação para mensurar o poder econômico das famílias.

Assim, a simplificação para obter as classes econômicas, como o “Critério Brasil”, desenvolvido pela ABEP, Associação das Empresas de Pesquisa tem suas justificativas (ABEP, 2008). Ao mesmo tempo a simplificação representa também um problema quando não mais se questiona e toma-se a divisão de classes sem muita reflexão.

Neste sentido o presente estudo procura evidenciar uma teia de dados que permita refletir sobre o Brasil e toma como perspectiva as dificuldades básicas percebidas pelas famílias brasileiras num momento de mudanças socioeconômicas. Para isso utiliza dados da POF, Pesquisa de Orçamentos Familiares, coletados pelo IBGE entre 2008 e 2009.

1.2 Objetivo do Trabalho e Problema da Pesquisa

Este trabalho surgiu da necessidade de identificar alguns aspectos que se relacionam direta e indiretamente com a percepção subjetiva de adequação da renda das famílias brasileiras.

Duas perguntas estão orientadas para o problema da pesquisa:

- Como mensurar a “dificuldade” percebida pelas famílias brasileiras para adequar orçamento e renda ?
- Como as variáveis selecionadas, de forma combinada, permitem compreender esta dificuldade?

De forma complementar foram identificadas perguntas que norteiam os objetivos do trabalho:

- Como a dificuldade percebida difere por estrato de renda ?
- Além da renda, que variáveis estão relacionadas direta e indiretamente com esta percepção?

- De que forma as classes econômicas fariam parte do estudo? Seria importante incluí-las na análise? Que ganhos trariam ao entendimento e condução do estudo?
- O que mais revelam estas análises? Que hipóteses podem ser levantadas?

Para responder às perguntas, propõe-se o desenvolvimento de um modelo probabilístico, baseado na análise dos dados da POF 2008-2009, que permita representar a inter-relação entre as variáveis selecionadas, formando um contexto onde a dificuldade percebida pelas famílias é a variável central. Este modelo deve permitir simular diferentes situações a partir do conjunto de dados selecionados, visando estimular a compreensão, estabelecimento de hipóteses e aquisição de novos conhecimentos.

Neste sentido, o trabalho contribui na medida em que olha para a percepção subjetiva das famílias sobre a adequação de sua renda e busca explicitar uma “teia de dados”, incluindo as linhas imaginárias que dividem a sociedade em classes econômicas. Ao estabelecer uma “teia”, rede de dados, busca evidenciar a complexidade e a diferença entre as famílias colocando no centro da análise a percepção pessoal dos envolvidos.

1.3 Descrição dos capítulos

Modelo de apresentação inspirado em Rocha, 2007.

2 MOTIVAÇÃO TEÓRICA

Os elementos que motivaram o desenvolvimento do trabalho estão inseridos na teoria que trata do comportamento do consumidor e foca especificamente as famílias brasileiras. Neste sentido, seleciona três abordagens teóricas, utilizadas como pano de fundo e elemento inicial para reflexão:

- Recursos do consumidor e a alocação do orçamento familiar.
- Composição das famílias e seu ciclo de vida.
- Classes econômicas e o momento atual brasileiro.

Como elemento chave utiliza-se também de dados disponíveis na POF (Pesquisa de Orçamentos Familiares) e evidencia a importância do uso e mineração de informações públicas para a compreensão da realidade nacional.

2.1 Recursos do Consumidor e Ciclo de Vida das Famílias

Engel et al. (1995) ressaltam diferenças importantes entre os consumidores em diversos aspectos: conhecimentos, atitudes, motivações, personalidade, valores, estilo de vida. No entanto, inicia o estudo sobre as diferenças individuais entre os consumidores sob a óptica dos recursos que possuem, sendo estes: econômicos, tempo e cognitivos. Os recursos são utilizados nos processos de troca com o mercado. Assim, o mercado compete por recursos tangíveis (econômicos) e intangíveis (tempo destinado ao consumo e forma como consumidores processam informações durante a decisão de compra). O presente estudo seleciona o “recurso econômico” como ponto principal de análise. No entanto, aproxima-se dos aspectos relacionados ao tempo e cognição, a partir de itens consumidos pelas famílias como alimentação fora de casa, por exemplo, ou através do nível de educação, que tende a favorecer os aspectos cognitivos. Parafraseando Engel et al. (1995 p. 316) : as famílias podem ter altas rendas, mas são pobres em tempo. Ou seja, o consumo de produtos e serviços que poupam o tempo, depende de uma renda adequada. Neste sentido o tripé : tempo, cognição e renda, parece ter no último a sua maior sustentação quando se analisa a população em geral e não nichos de mercado. A renda é usualmente utilizada em estudos sobre condições de vida e pobreza.

Em geral os critérios de classificação econômica olham para a família. Conforme apresenta Neri (2010) a renda per capita familiar tem sido priorizada como indicador de poder econô-

mico de um indivíduo e de sua família. Neste critério estão embutidos dois conceitos: o valor da renda e a quantidade de pessoas na família.

Neste sentido o ciclo de vida familiar complementa o conceito de renda per capita na compreensão sobre os “recursos” dos consumidores. Na literatura relativa ao comportamento do consumidor, três modelos de ciclo de vida familiar podem ser referenciados, considerando diferentes épocas: Wells e Gubar (1966), Murphy e Staples (1979) Gilly e Enis (1982), os quais inspiraram estudos posteriores sobre o tema. O Estudo de Saraiva Jr (2005) para o Brasil, apresenta além dos anteriores o modelo proposto por Redondo-Bellón et al. (2001), que concilia os modelos tradicionais com os aspectos culturais do ambiente espanhol, incorporando mudanças na estrutura familiar contemporânea. Este estudo propôs uma divisão para as famílias brasileiras, apresentando adaptações e identificando 11 tipos de famílias com base nos microdados da POF 2003. Os elementos para a formação destes tipos familiares são: idade da pessoa de referência no domicílio, estado civil, número de moradores, presença de casais, presença de filhos, outros moradores do domicílio, família monoparental e idade do filho mais novo. Nestes tipos de divisão procuram-se referências para marcos como casamento, entrada e saída de crianças na família, divórcio separação e envelhecimento. Limeira (2008, p. 184) apresenta as mudanças na estrutura familiar brasileira com base em dados do IBGE, indicando modelos menos clássicos de família como as unipessoais e as mulheres sem cônjuges. Alguns estudos combinam ciclo de vida das famílias com renda, visando estabelecer segmentos mais homogêneos em termos mercadológicos, conforme apresenta Engel et al. (1995, p. 753-755).

O ciclo de vida também já foi estudado pelos economistas através de microdados, como apresenta o texto de Blundell et al. (1994), cujo objetivo foi avaliar como padrões de consumo se alteram ao longo do ciclo de vida das famílias, neste caso sem a preocupação de criar agrupamentos de famílias, mas identificando variáveis relacionadas: demográficas e relativas ao mercado de trabalho. O presente trabalho beneficia-se de estudos econométricos que se apoiaram na teoria do Ciclo de Vida, conforme apresenta Fouto (2008 p. 17).

2.2 Alocação do orçamento familiar

O consumo das famílias no Brasil representa 60% do PIB (IPEA, 2011). O estudo de sua composição em tipos de despesa pode revelar maior ou menor diversificação no padrão de consumo das famílias. Com o desenvolvimento da economia, a participação do consumo de itens mais básicos, como alimentação, habitação e vestuário tende a diminuir dando espaço a novas formas de consumo. Engel et al. (1995, p. 298) apresenta 2 Gráficos relativos à economia americana no início e final do século passado. Estes Gráficos são similares aos que aqui estão apresentados para representar a diversificação do consumo brasileiro.

Gráfico 3 -BRASIL : Despesas Familiares sobre Renda Média POF 2003 e 2009

Fonte: baseados em Engel et al.; dados : POF 2003 e 2009.

Segundo os autores, enquanto numa economia desenvolvida as despesas com alimentação, moradia e vestuário representam menos que 50% da renda média familiar, em países “menos-desenvolvidos” estas despesas ultrapassam os 50%. Se tomar isso como uma verdade, abstrair eventuais diferenças no computo das rendas e despesas, e comparar o Gráfico 4 do

Brasil de 2009 com o apresentado por Engel et al. (p.296) pode-se dizer que em 2009 o Brasil apresentou alocação de renda familiar em despesas de forma semelhante à americana do final do século XX.

No entanto, o mais importante é observar que nos anos de 2003 e 2009 as “outras despesas” passaram de 33% para 42% da renda média familiar. Estes 42% representam categorias de despesa que dinamizam o mercado e fazem do entendimento do consumidor algo mais desafiador, conforme explica Engel et al. (p.295).

Com a evolução econômica o consumidor aumenta suas exigências, se antes era a alimentação que deveria caber em seu bolso, agora é o celular a fonte das “dificuldades” orçamentárias. Ao migrar pela pirâmide das necessidades o consumidor descobre novas despesas que limitam o seu orçamento. Para o consumidor urbano de grandes centros, por exemplo, o transporte tem maior peso orçamentário e transforma-se em item básico, juntamente com a alimentação, a moradia e o vestuário.

Além disso, observado o Gráfico 3, percebe-se que existe uma pequena diferença entre consumo médio e renda média no Brasil, quantificado como renda livre (1%). Neste sentido, as empresas brigam por fatias diferentes num bolo de renda altamente comprometido. Neste contexto, entra o crédito ampliando os prazos, viabilizando parte do consumo, mas incluindo novas fontes de despesa orçamentária: juros e parcelas. Observando-se a POF 2003, pode-se perceber que parte dos que declaravam dificuldades orçamentárias haviam atrasado parcelas de carnês, contas de energia, água, gás, além de alugueis. Estes dados ainda persistem na POF 2009, no entanto não é possível uma comparação direta das duas pesquisas, pois na POF 2003, atrasos estavam classificados de forma diferente de 2009. Em 2009, felizmente, as variáveis de atraso apresentam melhores agrupamentos e serão utilizadas neste trabalho.

2.3 Classes Econômicas

As classes econômicas têm sido amplamente mencionadas em estudos sobre o consumo, sendo muitas vezes a base da segmentação em muitas empresas brasileiras. Independente da compreensão do seu significado, é comum a associação do poder econômico com as letras que vão de A a E, sem que haja uma correspondência entre os diferentes critérios utilizados para essa atribuição.

O presente estudo utiliza e testa a classificação econômica recomendada pela Associação das Empresas de Pesquisa (ABEP, 2008). Esta entidade é responsável pela manutenção do Critério Brasil o qual é aberto e pode ser consultado diretamente no “site” da entidade, referenciado na bibliografia. Este critério contém uma fórmula de cálculo baseada na posse de bens de consumo, na escolaridade do chefe de família, no número de banheiros da residência e utilização de serviços de empregados domésticos mensalistas. Sofre atualização periódica e avaliação anual, sendo uma de suas características principais a simplicidade de aplicação.

A pontuação calculada com os dados da POF, resultante da aplicação do método, pode ser considerada como uma “aproximação” com o Critério Brasil (ABEP-2008), pois utiliza os itens de consumo com base nas informações colhidas no inventário domiciliar de bens, escolaridade do chefe da família, informação sobre analfabetismo, número de banheiros e serviços de empregados domésticos. Além das diferenças em termos amostrais, podem ocorrer diferentes interpretações sobre o significado dos itens que compõem a fórmula. Um exemplo: o entrevistado respondeu que não tem rádio, mas tem um equipamento de som, então foi computado pelo dado da POF que o domicílio possui um ou mais rádios. É importante este esclarecimento, pois um item a mais ou a menos e a família muda de classe econômica, dado que pode aumentar ou diminuir sua pontuação, conforme cálculo proposto no Critério Brasil. Quanto à amostra, a ABEP utiliza anualmente o Levantamento Sócio Econômico (LSE) do IBOPE Media, unidade de negócios do Grupo IBOPE responsável por pesquisas de audiência, consumo e mídia na América Latina. Este estudo se propõe a mapear as características sociais, demográficas e econômicas das famílias em regiões metropolitanas do país e Distrito Federal : Grande São Paulo, Grande Rio de Janeiro, Grande Porto Alegre, Grande Curitiba, Grande Belo Horizonte, Distrito Federal, Grande Salvador, Grande Recife, Grande Fortaleza e Grande Florianópolis. O estudo levanta as características físicas de cada domicílio pesquisado, os dados demográficos dos moradores, a posse de bens, a utilização de serviços e a renda familiar, o que permite a definição da classe econômica segundo o sistema de pontos do Critério Brasil. No Gráfico a seguir são apresentadas as diferenças de estimativa entre a distribuição informada pela ABEP para o ano de 2008 e as obtidas com o processamento dos dados da POF 2008-2009, considerando as áreas pesquisadas pela ABEP e o total do Brasil. A curva do Brasil, calculada com os dados da POF, possui inclinação mais para a direita, pois inclui população interiorana e rural, que não são pesquisadas pelo Levantamento Sócio Econômico (LSE) utilizado pela ABEP. Dentro das mesmas áreas pesquisadas as distribui-

ções são mais parecidas, apesar da distribuição da POF estar também mais inclinada à direita, também nestas localidades.

Gráfico 4 -Critério Brasil - Comparativo entre Distribuição na ABEP e na POF

Fonte : POF 2009 e ABEP -Critério Brasil 2008

O Critério ABA_ABIPEME, que antecedeu o Critério Brasil, recebeu muitas críticas de Mattar (1995), que ressaltou a inadequação do método para classificar famílias posicionadas nos extremos da distribuição de renda, em especial as mais pobres. Mais recentemente, Francisco (2010) critica também o Critério Brasil, apresenta revisão bibliográfica sobre o uso das classes econômicas no Brasil, conta a história do critério e sugere adotar o consumo de energia elétrica como uma boa aproximação para renda familiar. O autor utiliza métodos quantitativos baseados em regressão GWR (Geographically Weighted Regression).

Kamakura e Mazzon (2011) apresentam extensa revisão da literatura sobre as classes econômicas em caráter mundial e lembram que estes construtos, utilizados na prática do Marketing há mais de sete décadas foram pouco investigados pela comunidade acadêmica visando propor novas formas de classificação. Os autores propõem um novo critério, alternativo ao Critério Brasil, que também utiliza variáveis relativas à posse de bens, escolaridade e outras informações, toma como base para o desenvolvimento do modelo os dados da POF 2002_2003 e POF 2008_2009. Os autores informam que a nova estratificação proposta per-

mite melhor representar as recentes mudanças socioeconômicas no Brasil. Segundo os autores o consumo em muitas categorias de produto é mais uniforme do que se imagina entre as diferentes classes, o que não era possível de ser visto com o Critério Brasil. Os autores utilizam um modelo baseado em métodos de classes latentes. Dado a recência do artigo não foi possível testar a classificação proposta no presente trabalho.

Vale lembrar que inicialmente foi colocada uma questão que norteia um dos objetivos a serem atingidos, ou seja, “de que forma as classes econômicas fariam parte do modelo desenvolvido?”. Neste sentido, adotou-se o Critério Brasil como aproximação para a Classe Econômica, apesar das críticas observadas. Esta escolha foi feita por alguns motivos : (1) o modelo não utiliza diretamente a variável renda, mas informa que tem alta correlação com renda familiar, acima de 0.79 e que este índice tem aumentado nas últimas revisões (ABEP, 2008); (2) é um modelo adotado pelas empresas de pesquisa de mercado e em consequência por muitas empresas no mercado; (3) é simples e possível de aplicação na POF 2009 e anteriores.

2.4 A POF -Pesquisa de Orçamentos Familiares

Kotler e Keller (2006) ressaltam a importância de utilizar as fontes de dados governamentais e como estas podem auxiliar na segmentação de mercado, destacando as variáveis relativas à demografia. Além da segmentação estas informações podem ser utilizadas para a compreensão do macro ambiente, apontando tendências e forças que revelam ameaças e oportunidades para as organizações. O ambiente demográfico, segundo os autores, pode ser estudado a partir de diversas variáveis, como: crescimento populacional, composição etária, grupos étnicos, grau de instrução, religiões declaradas, estruturas familiares, movimentação geográfica da população e identificação de diferentes gerações.

Conforme apresentado por Silveira et al. (2006) os principais objetivos da POF são: (1) estabelecer a estrutura de gastos com consumo para os cálculos dos índices de custo de vida, determinando e atualizando as estruturas de ponderação dos dois principais índices de preços do IBGE, o INPC, Índice Nacional de Preços ao Consumidor, e o IPCA, Índice Nacional de Preços ao Consumidor Amplo e (2) investigar a parcela do consumo das famílias nas Contas Nacionais. Assim, a pesquisa investiga o orçamento das famílias e mensura com detalhes os

seus rendimentos, possibilitando novos estudos e análises. O Instituto Brasileiro de Geografia e Estatística (IBGE) apresenta, além dos resultados analíticos desta pesquisa, os “microdados”, termo utilizado pelo IBGE para designar bases de dados detalhadas, “questionário a questionário”, tratados e disponibilizados com o intuito de servir à comunidade e pesquisadores em diversas áreas. Este fato viabilizou este trabalho, considerando que para sua realização uma grande amostra com representatividade nacional seria necessária.

A POF tem objetivos específicos e uma base de dados muito peculiar, seu papel dentro do sistema nacional de estatística complementa e supera outras pesquisas como a PNAD, Pesquisa Nacional de Amostras por Domicílio, quando se fala por exemplo de compreender a renda dos brasileiros e de suas famílias.

Existe uma tendência de aproximação cada vez maior entre a POF e PNAD, no que se refere à captação das principais variáveis socioeconômicas (IBGE 2010). As diferenças e similaridades entre os dados socioeconômicos da POF e PNAD podem ser observadas em Cunha et al. (2006).

A POF tem previsão de ser realizada a cada 5 anos, mas a última ocorreu após intervalo de 6 anos, dado questões orçamentárias durante sua realização. A última pesquisa, foi realizada entre 2008 e 2009 e foram entrevistados 55.970 domicílios urbanos e rurais.

O IBGE (2010) sinaliza os diferentes usos analíticos que podem ser feitos a partir da POF:

- Mensurar as estruturas de consumo, dos gastos, dos rendimentos e parte da variação patrimonial das famílias.
- Traçar perfil das condições de vida da população brasileira a partir da análise de seus orçamentos domésticos.
- Estudar a composição dos gastos das famílias segundo as classes de rendimentos, disparidades regionais, áreas urbana e rural.
- Verificar a extensão do endividamento familiar.
- Dimensionar o mercado consumidor para grupos de produtos e serviços.
- Compreender as condições de vida das famílias, a partir da investigação qualidade percebida pelas famílias de forma subjetiva.

A última POF teve período de coleta de 12 meses: Mai 2008 a Maio de 2009, coincidiu parcialmente com a grande crise financeira internacional. No entanto, nesse período, o lado do consumo das famílias foi menos afetado que o resto do PIB. Mesmo assim, foi prudente considerar este fato antes da análise dos dados da POF 2009. O IBGE (2010) coletou a da-

ta/semana em que a família foi pesquisada e assim foi possível avaliar se houve alguma alteração na percepção subjetiva que as famílias fizeram neste período. A POF 2008-2009 (IBGE,2010) aplica, no último dia de entrevista, em cada domicílio, um questionário destinado a avaliar aspectos ligados à qualidade de vida das famílias : condições de moradia, facilidade para chegar ao final do mês com os rendimentos familiares, avaliação do alimento consumido, etc. As respostas dadas a este questionário são opiniões individuais dos entrevistados. São percepções subjetivas, que podem funcionar como um contraponto para outros estudos, mas no presente trabalho têm aspecto central. Assim, é importante avaliar se a crise provocou alguma alteração na percepção ao longo da realização da pesquisa.

Observa-se visualmente pelo Gráfico 5 que a pesquisa ocorreu em meio a uma turbulência econômica. Entretanto no Gráfico 6 não se percebe influência da crise na resposta à pergunta que trata da facilidade para chegar ao final do mês com os rendimentos familiares. Utilizou-se o critério informado pelos relatórios da POF (IBGE 2010), ou seja, foram somadas as respostas: Alguma Facilidade; Muita Facilidade; Facilidade à pergunta sobre adequação renda e orçamento. Este número é aproximadamente 25% e não sofreu grandes alterações durante a coleta. Por este critério observa-se que 75% em média reportaram “dificuldade” e 25% “facilidade”. Nesta tese, todas as vezes que aparece o termo “dificuldade” das famílias ele estará relacionado a esta proporção informada, o que será esclarecido mais adiante.

Gráfico 5 -Período de coleta da POF 2008-2009 e as variações no PIB

Fonte : IPEA 2010 - adaptado pela autora representar o período da POF 2008 2009

Fonte : Elaborado pela autora com base nos dados da POF 2008 2009 -amostra ponderada

Para compreender a POF, suas potencialidades, riquezas e algumas limitações sugere-se consultar os dois volumes organizados por Silveira et al. (2006 e 2007) contendo uma coletânea de artigos sobre o gasto e o consumo das famílias brasileiras, referentes a POF 2002_2003.

Observando os relatórios do IBGE publicados em 2010 sobre condições de vida das famílias brasileiras, percebe-se um fato, que isolado, parece contraditório com as notícias mais otimistas sobre a ascensão de consumo das classes baixas e médias: 75% declaram viver com dificuldade.

Distribuição das famílias, por referência à dificuldade para chegar ao fim do mês com o rendimento monetário familiar - Brasil - período 2008-2009

Gráfico 7 : Distribuição das famílias, por referência à dificuldade

Fonte: IBGE - PESQUISA DE CONDIÇÕES DE VIDA POF 2009 -2010

3 ASPECTOS METODOLÓGICOS

Do ponto de vista metodológico o presente trabalho pode ser dividido em dois grandes blocos. A Figura 1 ilustra dois aglomerados. À esquerda estão os temas referentes ao pano de fundo teórico, que motivaram este estudo e que foram apresentados nos itens anteriores. À direita estão as Redes Bayesianas utilizadas para agenciar a complexidade que emerge do tema. O termo “agenciar” surgiu da leitura de Vassão (2010). O agenciamento é compreendido como a criação de um espaço específico para que se possa atuar a partir dele, fazendo alterações oportunas ou necessárias para compreender o contexto observado e a relação entre os objetos.

Figura 1 - Agenciamento da complexidade

FONTE : PREPARADO PELA AUTORA

Esta tese se propõe a desenvolver um modelo quantitativo que seja um espaço de significação, reflexão e mediação para compreender uma realidade complexa, focada na “dificuldade” percebida pelas famílias brasileiras, através de uma teia de dados que contextualiza esta percepção subjetiva.

Este trabalho pode ser enquadrado como uma pesquisa exploratória que emprega métodos quantitativos, baseados em técnica de mineração de dados. Para Zikmund (1997, p. 51) pesquisas exploratórias visam encontrar a dimensão de um problema, coletando informação e abrindo espaço para que, no futuro, outras pesquisas mais específicas possam ser conduzidas. A mineração de dados é uma técnica adequada à pesquisa exploratória, uma vez que permite explorar dados em busca de evidências esperadas e inesperadas.

Este capítulo discute brevemente o conceito de complexidade, inerente ao tema estudado e apresenta uma introdução sobre Redes Bayesianas

3.1 Complexidade

Para Morin (2000, p. 213) o pensamento complexo é essencialmente o pensamento que trata com a incerteza, sendo capaz de conceber a organização, de reunir o *complexus* (“aquilo que é tecido em conjunto”) de contextualizar, de globalizar, e ao mesmo tempo reconhecer o singular, o individual, o concreto.

O complexo e o complicado fazem parte do tipo de modelagem apresentada nesta tese. Complicado pela quantidade de peças - variáveis selecionadas - e complexo pelo fato de que o todo não deve representar apenas a soma das partes. Do todo emerge algo diferente como um novo conhecimento.

Espera-se não ser reducionista ao tentar operacionalizar um modo para abstrair sobre a realidade observada. De acordo com Vassão (2010) alguns creem que todas as coisas são irreduzíveis, complexas e que não se deve tentar simplificar. No entanto ele pergunta: se o complexo é de difícil compreensão como reduzi-lo e projetá-lo? Este autor assume que a simplificação é inevitável e que a própria simplificação está repleta de complexidade.

Deve-se ter cuidado ao citar Morin (2000) que sugere em sua obra uma leitura menos tecnicista para a compreensão de fenômenos complexos. O presente trabalho, assim como em Vassão (2008 p. 121), não consegue “rechaçar” (nas palavras deste autor) a validade da ciência instrumental e tecnicista para compreender realidades complexas. Usa o instrumental como objeto de reflexão e inspira-se nas ideias de Morin ao aceitar que o instrumental pode dizer alguma verdade a respeito do mundo, mas não “toda a verdade”.

Becker (2008, p.23) fala sobre a utilidade de métodos que ajudam a pensar os problemas sociais como a modelagem matemática. O autor apresenta também “truques” para realização da pesquisa social. No entanto, o autor ressalta que a palavra “truque” dá a conotação de que as coisas tornam-se “mais fáceis”, mas na verdade estes “truques” tornam as coisas mais difíceis, pois interferem nas “confortáveis rotinas de pensamento”: sugerem maneiras diferentes de pensar, apontam novas perspectivas, criam novos problemas a pesquisar, apresentam novas possibilidades de comparar, inventam categorias ainda não vislumbradas.

A ciência informacional é aqui utilizada como elemento de elucidação, como objeto ou lente que permite refletir e aprender sobre a realidade. Nas palavras de Levy (1993, p. 157) :

[...] as tecnologias intelectuais amplificam, exteriorizam e modificam numerosas funções cognitivas humanas: memória (bancos de dados, hiperdocumentos, arquivos digitais de todos os tipos), imaginação (simulações), percepção (sensores digitais, telepresença, realidades virtuais), raciocínios (inteligência artificial, modelização de fenômenos complexos) .

O presente trabalho utiliza novas tecnologias analíticas menos para “projetar respostas” e mais para “projetar perguntas”. Quando Vassão (2010 p. 13) fala da possibilidade de fazer “design” do complexo, sugere que toda e qualquer resposta seja antes de tudo uma pergunta disfarçada. No caso do presente estudo ao procurar pela resposta “operacionalizável” que permite calcular a probabilidade de uma família viver com “dificuldade” orçamentária, na verdade lança-se a primeira pergunta : o que a leva a esta percepção ?

A expectativa com esta tese é que as Redes Bayesianas possam representar um caminho para lidar com o pensamento complexo. Neste sentido espera-se que sejam agenciadoras e facilitadoras do pensamento e não uma ferramenta utilizada de forma tecnicista ou reducionista. Assim, se deve considerar de antemão as palavras de Becker (2008, p. 64) para quem os fenômenos sociais estão conectados de tantas maneiras e sob tantas condições que (talvez) nunca se consiga obter um projeto adequado. No entanto, segundo Becker (2008, p. 66):

[...] é um bom truque pensar sobre algum conjunto de atividades sociais como dotado desse caráter orgânico, procurando todas as conexões que contribuem para o resultado em que estamos interessados, vendo como eles se afetam mutuamente, cada um criando condições para que os outros operem.

Passa-se agora para uma breve revisão sobre Redes Bayesianas, uma vez que este método será utilizado para agenciar a complexidade fornecendo uma “máquina abstrata” (VASSÃO, 2010, p.69), repleta de conexões cujas inter-relações serão objeto de análise.

3.2 Redes Bayesianas - Introdução

Com base no levantamento bibliográfico sobre as Redes Bayesianas pode-se dizer que elas permitem a representação de um conhecimento e visam diminuir a incerteza sobre o objeto de estudo, a partir de uma modelo baseada em probabilidades.

As Redes Bayesianas são suportadas por grafos e têm o potencial de ajudar a lidar com a complexidade. Segundo Vassão (2010) a complexidade é ao mesmo tempo concreta e abstrata e para projetá-la são necessários diagramas, como os grafos, que são figuras simples, corriqueiras e prosaicas. Entretanto, salienta o autor, podem render percepções poderosas sobre o modo como as coisas estão configuradas. Além disso, os grafos e diagramas formalizam matematicamente questões relacionadas à causalidade, impossíveis de serem operacionalizadas apenas com equações (PEARL, 2009, p. 416-417).

Quando do início desta tese o tema Redes Bayesianas soava como desconhecido e novo. No entanto digitando o tema “Bayesian Networks” no Google surgiram mais de 3 milhões de documentos, sendo que a mesma expressão em língua portuguesa apresentou mais de 39.000 resultados. A maioria dos documentos é relevante e trata de questões teóricas, com resumos bem feitos e inteligíveis. Dado o interesse crescente dos estudiosos, espera-se que as aplicações aumentem nos próximos anos inclusive pela quantidade de dados disponíveis.

Atualmente no banco de teses da Universidade de São Paulo, o tema “Redes Bayesianas” aparece vinculado a 16 teses e dissertações em diferentes escolas e áreas do conhecimento (USP, 2012), mostrando também a aplicação da técnica em diversas áreas no Brasil, mesmo que os números sejam ainda tímidos. O interesse pelas Redes Bayesianas é relativamente novo, uma vez que o seu desenvolvimento está associado com o aumento da capacidade de processamento dos computadores e ao aumento da quantidade de dados disponíveis e passíveis de análise para geração de conhecimento. Estes fatos combinados com a teoria probabilística permitiram o uso prático das Redes Bayesianas em processos analíticos (PEARL, 1988; JENSEN e NIELSEN, 2007).

Com tantas publicações internacionais e nacionais e tendo em cada uma delas uma introdução, opta-se pela brevidade selecionando apenas os pontos que podem ser relevantes para situar e compreender a origem, a extensão do termo Redes Bayesianas e algumas controvérsias a este respeito. Além disso, serão apresentados conceitos importantes para compreender o funcionamento das Redes Bayesianas e sua aplicação nesta tese, além de uma breve cronologia que pode ser útil para os futuros interessados situarem-se neste universo de informações disponíveis.

3.2.1 Redes Bayesianas - Nomenclatura e Controvérsias

O termo Redes Bayesianas foi cunhado por Judea Pearl aproximadamente em 1985, sendo apresentado na Sétima Conferência da Sociedade de Ciências Cognitivas da Universidade da Califórnia (PEARL, 1985, p.329-334). No entanto, o termo tornou-se mais conhecido a partir de 1988 quando da publicação de seu livro sobre o assunto (PEARL, 1988).

O nome Bayesiano homenageia Thomas Bayes (1701 - 1761), cujos estudos deram origem ao teorema que é a base conceitual da regra para atualizar probabilidades à luz de novas evidências. Bayes foi pouco conhecido e deixou um manuscrito que ao ser encontrado entre outros documentos por Richard Price (1763) foi encaminhado à Royal Society, da qual Bayes era membro. O documento intitula-se : “Ensaio sobre a solução de um problema da doutrina das chances” . Para os mais curiosos, uma cópia do original acompanhada da carta de Price pode ser encontrada no site da UCLA (University of California, Los Angeles) , conforme referenciado neste texto. Outros materiais sobre a teoria de Bayes auxiliam na compreensão sobre a doutrina das chances e estão na Enciclopédia de Filosofia de Stanford (STANFORD, 2012).

Vale apresentar nesta introdução as considerações de Silander (2009, p4) a respeito do sobrenome “bayesiano” que foi adicionado às redes quando estudiosos da estatística e da inteligência artificial (IA) ainda caminhavam distantes uns dos outros. Este fato trouxe muita confusão, pois nem sempre uma Rede Bayesiana é baseada em Estatística Bayesiana. Elas podem ser desenvolvidas seguindo as convenções da estatística bayesiana, mas muitas vezes o emprego do nome é apenas uma coincidência, o que provoca uma grande confusão entre pesquisadores. Assim, é importante compreender que enunciados diferentes podem ser utilizados para a construção das redes : os bayesianos, os frequentistas ou nenhum deles como sugere a visão pragmática de D’Ambrosio (1999, p35, tradução nossa).

Matematicamente as redes bayesianas são representações de distribuições de probabilidade na forma fatorizada e assim é ignorante sobre a guerra filosófica existente entre bayesianos e frequentistas. No entanto, como um objeto de engenharia para lidar com conhecimento, é utilizada para expressar conhecimento a priori, julgamentos ou crenças subjetivas.

Para tentar esclarecer sobre a perspectiva bayesiana Heckerman (1996, p.279) apresenta um exemplo que ilustra a diferença básica entre ambas: suponha que um alfinete com uma ponta arredondada seja lançado várias vezes sobre uma superfície lisa. Em algumas das vezes o alfinete ficará com sua aste para cima (supondo 0.55 dos casos) e e outras com a aste para

baixo (supondo 0.45 dos casos). Um estatístico com perspectiva bayesiana reconhece os valores possíveis desta fração como uma variável denominada “ Θ ”, cujo verdadeiro valor é uma incerteza. No entanto ele não deixa de lado seu conhecimento prévio e para expressar esta incerteza utiliza-se de uma distribuição de probabilidade a priori, $p(\Theta | \varepsilon)$, sendo que ε representa seu conhecimento prévio. Este conhecimento é atualizado, modificado à medida que se observa a posição dos alfinetes, $p(\Theta | \text{"aste para cima"} | \varepsilon)$. Esta nova informação culmina numa distribuição posteriori. Para tentar deixar mais claro, utiliza-se o raciocínio de Pereira (2006) que sugere uma sequência de perguntas e respostas para operacionalizar o conceito de informação. A estas perguntas foi acrescido o item (b).

a) A Informação é sobre o que?

Sobre o valor de Θ que é descrita pela distribuição atual de Θ (priori)

b) De que tipo é esta informação?

Do tipo experimental, ou seja, relativa à realização de experimentos representados por variáveis aleatórias, passíveis de observação e que têm o potencial de esclarecer sobre Θ .

c) Onde está esta informação ?

Como a informação é algo que tem o poder de modificar o conhecimento sobre Θ ela está em $L(\Theta/x)$. Ou seja, no cálculo da probabilidade de Θ , dado que existe uma nova informação.

d) Quanto de informação é usada ?

A informação utilizada é mensurada pela distância entre o conhecimento do parâmetro que se quer compreender e a variação percebida, quando se tem em mãos uma variável aleatória x , o que pode ser representado por

$$D\{P(\Theta|X), P(\Theta)\}$$

e) Como a informação é extraída ?

Pela operação de Bayes, a qual permite atualizar probabilidades à luz de novas evidências $\{x_1, x_2, x_3, x_4\}$

$$P(\Theta|x_1) = \frac{P(x_1|\Theta) * P(\Theta)}{P(x_1)}$$

Para iniciar modelagem com Redes Bayesianas esta sequência de perguntas pode também ser utilizada, no entanto enquanto a abordagem bayesiana irá utilizar como ponto de partida uma distribuição a priori, mesmo que não informativa (uma uniforme, por exemplo) a frequentista usará outro raciocínio. Ou seja : tomando como base o exemplo do alfinete, para um frequentista a proporção 0.55 encontrada no experimento é assumida como “o valor da

probabilidade”, pois considera que os alfinetes foram atirados várias vezes em um experimento aleatório que permitiu este cálculo. Assim, para os frequentistas a probabilidade de um evento (“aste para cima”) é calculada pela divisão do número de eventos de interesse pelo total de eventos observados (ex: 0,55). Os enunciados probabilísticos frequentistas são feitos exclusivamente com base na frequência de ocorrência de um evento: “um único resultado em uma seqüência infinita de um metaprocessos estocástico imaginário” (STERN 2006 p.573).

Stern (2006) apresenta a definição de três enunciados probabilísticos : (1) frequentistas são feitos exclusivamente com base na frequência de ocorrência do evento em estudo (2) epistêmicos são feitos com base no grau de crença ou possibilidades de um evento dentre os possíveis resultados gerados por uma variável aleatória (3) bayesianos são enunciados probabilísticos epistêmicos gerados pela aplicação recursiva da fórmula de Bayes.

É importante salientar que a abordagem bayesiana para estatística tem uma longa história, no entanto antes dos anos 90 ela teve um papel marginal, considerando as ciências humanas e sociais. Segundo Lynch (2010) isso se deve provavelmente às seguintes razões : (1) oposição ao uso de “distribuições a priori” e ao uso de probabilidades subjetivas, (2) falta de capacidade dos computadores para realizar análises Bayesianas completas e realistas. (3) Uma terceira razão que pode justificar o baixo uso da abordagem bayesiana especificamente pelos cientistas sociais: muitos não são preparados em seus cursos para lidar com a teoria da probabilidade e com os cálculos ligados ao raciocínio bayesiano, necessitando de muito esforço para avançar no uso da técnica.

Para conhecer a visão bayesiana da estatística sugere-se consultar Pereira (2006) e Stern (2006) onde se pode aprofundar sobre importantes diferenças entre os conceitos, com explicações sobre o tratamento de hipóteses em cada um dos enunciados. Para iniciar de forma mais básica consultar Wonnacott e Wonnacott (1990 p. 564-636) e Lynch (2010), tendo este último um enfoque aplicado às ciências sociais.

No entanto, as abordagens apresentadas são governadas pelas mesmas leis da probabilidade (DARWICHE,2009), apesar de conceitualmente muito diferentes em seus enunciados. Assim, nem sempre as ferramentas de IA (Inteligência Artificial) estão alinhadas com o reconhecimento de uma distribuição a priori e nem com probabilidades subjetivas, pertencentes à filosofia bayesiana. Utilizam-se muitas vezes da probabilidade sob a óptica “frequentista”, conhecida também como “probabilidade física” Heckerman (1996, p.279).

Silander (2009) sugere que o termo “redes de dependência” poderia ser um nome melhor, mas parece tarde demais para isso. Hoje em dia, segundo o autor, as Redes Bayesianas são vistas como parte de uma grande família de modelos Gráficos e não necessariamente da Estatística Bayesiana. Apesar destas advertências sobre a confusão de nomenclaturas tanto Silander (2009) quanto Darwiche (2009) são comprometidos com “a filosofia” bayesiana em seus trabalhos.

A seguinte cronologia permite evidenciar a evolução inicial das Redes Bayesianas, conforme apresentado por Jensen e Nielsen (2007 p. 45), dentro das ciências da computação :

- a. O trabalho de Minsky (1961) sobre os passos em direção à inteligência artificial. Ele foi um dos fundadores da robótica. Estudou a interação entre a neurociência, psicologia e ciências da computação na área de inteligência artificial.
- b. Primeira metade dos anos 80 quando Pearl (1985) e Spiegelhalter e Knill-Jones (1984) trabalharam com a temática em sistemas especialistas e diagnósticos médicos.
- c. Estudos na área das ciências da computação demonstrando que inferência em Redes Bayesianas é um problema computacional “*NP-Hard*” por Cooper (1987). A partir daí as Redes Bayesianas foram muito estudadas visando desvendar seu desafio computacional o que aperfeiçoou o desenvolvimento de softwares e algoritmos nesta área.
- d. Quando as primeiras aplicações reais de Redes Bayesianas foram apresentadas por Heckerman et al. (1992) .
- e. Com os trabalhos de D’Ambrosio (1999) que apresentou métodos para viabilizar a inferência em Redes Bayesianas. Seu trabalho sobre a representação e inferência visou encontrar formalismos eficientes para capturar os aspectos estruturais e algoritmos de inferência.

Nos últimos 25 anos, desde as primeiras aplicações em diagnósticos médicos até o momento, quando o poder computacional e as ferramentas estão cada vez melhores, as Redes Bayesianas demonstram ser uma forma exequível de representação do conhecimento e compreensão de realidades complexas. Diversos trabalhos realizados com Redes Bayesianas foram citados por Pedreti (2007, p.42) relacionados aos processos decisórios em finanças. Na área de Marketing aplicações são encontradas em Marketing Direto, onde a rede Bayesiana é utilizada em modelos preditivos de resposta apresentando resultados bastante favoráveis como demonstra Cui et al. (2006). Alguns trabalhos práticos em planejamento e pesquisa de Marketing foram apresentados por Conrady e Jouffe (2011). Um interessante trabalho envolvendo modelo de escore para crédito foi desenvolvido por Karcher (2009), no entanto este explora

mais a capacidade preditiva de Redes Bayesianas e menos seu poder de transparência e interpretabilidade como em Cui *et al.* (2004).

Neste sentido o presente trabalho contribui para reflexões sobre o uso da técnica, sua aplicação e limitações visando apresentar suas capacidades : preditiva, de interpretabilidade e transparência do modelo.

3.2.2 Redes Bayesianas -Justificativa para utilização.

As Redes Bayesianas apresentam visualmente variáveis em nós cujos vértices revelam a dependência entre as variáveis mensuradas pelas relações de probabilidade. No presente trabalho as Redes Bayesianas são tratadas como parte dos modelos gráficos que auxiliam na organização do conhecimento e agenciamento da complexidade. Utilizando as palavras de Jordan (1999) elas são o casamento entre a teoria da probabilidade e a teoria dos grafos, onde a teoria da probabilidade provê “a cola” que permite reunir e combinar as partes, representadas através de uma interface intuitiva e atraente.

As Redes Bayesianas visam redução de incertezas a partir da observação de evidências, as quais podem mudar o conhecimento sobre determinada situação . Neste sentido elas têm o potencial de ajudar no raciocínio a partir da observação de variáveis relacionadas em um diagrama em forma de grafos com capacidade de inferência bidirecional Pearl & Russel (2003 p.2).

As Redes Bayesianas têm o potencial de ativar o raciocínio, pois como é possível instanciar os estados das variáveis do modelo enquanto o observa, pode-se “descobrir” informações de forma interativa, como num modelo de simulação. Pearl e Russel (2003) salientam que o processo de raciocínio pode ser operado nas Redes Bayesianas pela propagação da informação em qualquer direção. Ou seja, dado que a informação se propaga é possível raciocinar “da causa para o efeito” e do “efeito para a causa”, o que não é possível nas redes neurais ou sistemas baseados em regras.

Outro aspecto valorizado para o uso de Redes Bayesianas é a possibilidade de representar a realidade, mesmo que de forma simplificada, através de dados. Para Pearl e Russel (2003) uma das questões que emerge quanto ao uso de Redes Bayesianas como potencial modelo de cognição humana é quanto ao seu poder expressivo para lidar com os tipos de raciocínio so-

bre os indivíduos, relações, propriedades e questões universais que permeiam o pensamento humano. No entanto sua aplicação em ciências sociais ainda é pequena (LYNCH, 2010).

3.2.3 Redes Bayesianas -Conexões Básicas

Uma rede Bayesiana satisfaz a condição de Markov local, ou seja: cada variável (nó da rede) é condicionalmente independente de seus não descendentes dado o “nó pai”. Observa-se na figura a seguir como se opera a condição de Martkov em redes bayesianas.

Figura 2 - Conexões em Redes Bayesianas e Condições de Markov

Fonte: preparado pela autora com base em notas de aula Métodos de Inferência Aplicados à Administração - FEA-USP Prof Ronaldo Zwicker 2009 com base em Neapolitan 2007.

Segundo Jensen e Nielsen (2007) a d-separação, conceito ligado à condição de Markov, é um elemento muito importante na interpretação de Redes Bayesianas. Algumas representações estão ilustradas na Figura 3, as quais permitem avaliar como o conhecimento sobre uma variável pode mudar o conhecimento sobre outra variável.

Figura 3 - D-Separações e Tipos de Conexões em Redes Bayesianas

Fonte: preparado pela autora com base em notas de aula Métodos de Inferência Aplicados à Administração - FEA-USP Prof Ronaldo Zwicker 2009.

São três as conexões básicas de uma Rede Bayesianas, conforme exemplos extraídos de Jensen e Nielsen (2007, p. 27-30):

- 1) **Nós Convergentes:** os nós são independentes, ou seja, um não influencia no outro. (A) velas de ignição sujas; (B) nível de gasolina; (C) carro ligando ou não. A é independente de B, mas conhecer B pode mudar a crença em A, dependendo do conhecimento de C. Ou seja, se sei que a gasolina está em nível adequado e que o carro não liga então minha crença de que as velas de ignição estão sujas, aumenta.
- 2) **Nós Lineares :** A evidência é transmitida de forma serial, ao menos que a variável na conexão seja conhecida. Exemplo: (A) chuva ; (B) nível da água ; (C) inundação. Se o nível da água não foi observado, sabendo que houve inundação aumenta a crença de que o nível da água está marcado como alto, o que pode revelar que choveu. No entanto, se já se sabe o nível da água, saber que houve inundação não muda o conhecimento sobre “chuva”, pois já se sabe pelo medidor que choveu.
- 3) **Nós Divergentes :** a evidência pode ser transmitida através da conexão divergente, ao menos que o nó pai já seja conhecido. Exemplo: (A) Sexo, (B) Comprimen-

to do cabelo e (C) Estatura. Se nada se sabe sobre o sexo da pessoa, o comprimento do cabelo pode ser um bom indicativo do sexo, mas se já se sabe o sexo o comprimento do cabelo não é mais informativo.

Ao conhecer as conexões é possível interpretar qualitativamente as relações estabelecidas numa rede bayesiana, acatando ou criticando os resultados apresentados por redes obtidas de forma automática.

3.2.4 A aprendizagem com Redes Bayesianas

No início as Redes Bayesianas foram desenvolvidas supondo conhecimento de especialistas como base para a definição da estrutura da rede, ou seja, dado que uma estrutura ou topologia fosse informada por um especialista, a Rede Bayesiana seria capaz de estimar as probabilidades condicionais (RAMONI & SEBASTIANI, 1995, p.401). No entanto, este “conhecimento” ou estrutura de relacionamento das variáveis pode ser atualmente descoberto com a utilização de algoritmos de aprendizagem de máquina. Conforme Heckerman (1999, p. 301-354) e Ling & Zhang (2003 p. 709-721) o aprendizado por máquina nas Redes Bayesianas refere-se à construção de uma rede a partir de distribuições conjuntas, dispostas em estrutura gráfica (qualitativa) e um conjunto de parâmetros numéricos (quantitativos). Tanto a estrutura da rede quanto as distribuições conjuntas e parâmetros podem ser obtidos por algoritmos matemáticos.

Vale salientar que o especialista pode interferir numa rede estabelecida de forma automática incorporando posteriormente suas heurísticas, as quais podem não ter sido capitadas pelos algoritmos que podem ser testados através de métricas estabelecidas para avaliar a rede, gerando um aprendizado híbrido.

Um dos métodos mais comuns de aprendizado de estrutura é o método de pontuação. Para visão introdutória sobre este método consultar Veiga (2002) e Cui et al. (2006). Na busca da melhor estrutura pelo método da pontuação, procura-se uma dentre todas as redes segundo uma métrica. O espaço de busca da melhor rede apresenta uma enorme quantidade de redes alternativas. Quanto maior o número de variáveis maior é este espaço de busca, daí a necessidade de uma métrica que ajude a buscar a rede mais adequada. A métrica MDL - *Minimum Descriptive Length* (RISSANEM, 1978; SUZUKI 1996; LAM; BACCHUS, 1994) é passível de decomposição, permitindo construir redes a partir de uma rede inicial, decidindo em cada

passo pela inclusão ou exclusão (automática) de dependências (arcos) que conferem maior contribuição informacional para a rede final, segundo a métrica adotada.

Para aprimorar a percepção intuitiva sobre o funcionamento de um método de busca de estrutura de rede sugere-se consultar e testar o site B-Course, referenciado na bibliografia. Dado seu conceito e formato didático permite compreender o funcionamento de um espaço de busca de uma estrutura de rede.

De acordo com Carvalho (2011) existem dois grandes blocos de algoritmos utilizados para o aprendizado de redes bayesianas e cada um possui base teórica e métricas próprias. Os dois blocos são baseados nas seguintes teorias : (1) Teoria da Informação e (2) Funções “Bayesianas”. A autora apresenta a cronologia em que os diferentes algoritmos foram desenvolvidos:

Teoria da Informação :

- LL (Log-likelihood) (1912-22)
- AIC (Akaike Information Criterion) (1974)
- MDL (Minimum Description Length) (1978)
- BIC (Bayesian Information Criterion) (1978)
- MIT (Mutual Information Tests) (2006)
- NML (Normalized Minimum Likelihood) (2008)

Funções Bayesianas :

- BDeu ("u" para designer “uniform joint distribution”) (1991)
- K2 (1992)
- BD (Bayesian Dirichlet) (1995)
- BDe ("e" para designar “likelihood-equivalence”) (1995)

No presente trabalho os principais algoritmos estão baseados na Teoria da Informação, a qual será tratada no próximo tópico. A aprendizagem automática de estrutura utiliza a métrica MDL - Minimum Descriptive Length, uma vez que Redes Bayesianas complexas demandam muitos cálculos probabilísticos, uma grande capacidade computacional de processamento e podem apresentar resultados muito difíceis de interpretar. Foi Rissanen (1978) quem propôs este princípio de busca para modelagem, sendo que Lam e Bacchus (1994) que o adotou para avaliação de Redes Bayesianas. Esta métrica tem suas raízes na teoria da informação, sendo, de certa forma, equivalente à métrica BIC -Bayesian Information Criteria ou AIC -Akaike Information Criteria, para compreender esta métrica consultar Karcher (2009).

O método MDL, conforme Jensen e Nielsen (2007) considera uma abordagem baseada nos princípios de Ockham's Razor, traduzido como a Navalha de Ockham ou Occam, que valoriza o princípio da parcimônia (*Lex Parsimoniae*).

De acordo com Hiroshi (1997) este princípio lógico foi atribuído a um franciscano inglês que viveu no século XIV. Para Willian de Ockham, todo o estudo deve assumir apenas as premissas necessárias à explicação do fenômeno. Das afirmativas abaixo, segundo este autor, as duas primeiras aparecem nas obras de Willian de Ockham e a terceira foi escrita por seus seguidores e resumem bem o significado deste princípio: “Pluralitas non est ponenda sine neccesitate”: Entidades não devem ser multiplicadas desnecessariamente (tradução nossa); “Frustra fit per plura quod potest fieri per pauciora”: várias coisas que podem ser realizadas através de menos (tradução nossa); “Entia non sunt multiplicanda praeter necessitatem”: As coisas não devem ser multiplicadas além da necessidade (tradução nossa).

As ideias de Willian de Ockham estão relacionadas à simplicidade e não é tarefa corriqueira lidar com ela, pressupõe método e certo grau de perseverança para consegui-la. Assim neste trabalho utilizou-se três perguntas baseadas em Barker (2011) para esclarecer sobre a simplicidade almejada, sendo a simplicidade compreendida aqui como parcimônia: (1) Como a simplicidade é compreendida neste trabalho? (2) Qual é o papel dos princípios da simplicidade para lidar com o assunto? (3) Qual a justificativa racional para o uso dos princípios de simplicidade?

A Simplicidade remete à complexidade, conforme apresentado no item 3.1. No presente trabalho simplicidade é um dos objetivos da “máquina abstrata” que está sendo projetada para agenciar o entendimento de um tema complexo. O termo máquina abstrata foi utilizado por Vassão (2010, p.69). O papel dos princípios da simplicidade é tornar um assunto inteligível compreensível para a mente humana que tem dificuldades para lidar com muitas entidades (variáveis) ao mesmo tempo.

Operacionalmente a simplicidade visa tornar a “máquina abstrata” exequível, dentro dos seus parâmetros limitadores de memória, processamento, heurísticas computacionais. Neste sentido vem a resposta para a terceira pergunta, que trata da justificativa racional: Jensen e Nielsen (2007, p. 240) afirmam que no caso das Redes Bayesianas este princípio tem justificativa por si só, pois a complexidade de uma rede pode ser facilmente observada pelo número de ligações e parâmetros independentes.

A Figura 4 apresenta um resumo sobre o MDL e a relação com a verossimilhança (LL).

Figura 4 - Aprendizado de Estrutura - Teoria da Informação -MDL

FONTE : baseado em WITTEN et al., 2005

A verossimilhança (ou LL - Log Likelihood) permite avaliar a aderência do modelo aos dados tende a favorecer uma estrutura completa, onde os parâmetros estão todos interligados; o MDL funciona como uma fórmula que penaliza o número de nós filhos por variável. Enfim, o objetivo da métrica MDL, Minimum Descriptive Length, é obter um balanceamento entre a acurácia do modelo e a sua complexidade, sendo largamente empregada como metodologia para avaliar Redes Bayesianas, servindo-se também como um mecanismo para controlar “overfitting” (CUI et al., 2006). Em algoritmos de IA, Inteligência Artificial, o “overfitting” ocorre quando um modelo começa a memorizar os dados de treinamento ao invés de aprender e generalizar a partir do modelo.

O aprendizado da estrutura de Redes Bayesianas pode ser baseado tanto em variáveis discretas (que assumem valores de um conjunto finito de possibilidades e estados discretos ou qualitativos), ou contínuas (que podem assumir infinitos valores), ou ainda mistas (ambas). Matsuura (2003) explica que a diferença entre as Redes Discretas, Contínuas ou Mistas está na forma de representar as probabilidades condicionais. Nas redes discretas as probabilidades condicionais são representadas pelas tabelas de contingência, nas contínuas elas são expressas por funções de densidade de distribuição e nas mistas há uma composição destas duas opções. A complexidade das redes contínuas origina-se nas estimativas das probabilidades condicionais que devem ser representadas considerando tanto a função distribuição de probabilidade quanto seus parâmetros, demandando conhecimentos específicos sobre estas funções, a priori.

O aprendizado de redes discretas também possui complexidade, no entanto está mais relacionada ao processamento computacional, à necessidade de critérios de discretização de variáveis e à quantidade de arcos e nós. A complexidade de uma rede discreta é dimensionada

pelo número de variáveis a serem tratadas e pelo o número de atributos ou classes destas variáveis. O presente trabalho utiliza Redes Bayesianas discretas.

Por trás dos softwares escondem-se questões que não são escopo do presente trabalho como por exemplo a escolha de algoritmos de aprendizagem (MATSUURA,2003, p.25), detalhes técnicos sobre os algoritmos de aprendizagem de estrutura, de tratamento de “missing” cases, dentre outros. É sabido, por exemplo, que para tratar casos “missing” os softwares contam geralmente com algoritmo EM, Expectation Maximization. Este algoritmo opera durante a aprendizagem da Rede Bayesiana, calculando os valores durante o processo, considerando as relações que são encontradas entre as variáveis neste momento, o que pode variar a cada processamento. Mais detalhes sobre este tipo de algoritmo e aplicação em Redes Bayesianas estão presentes no trabalho de (LUNA; 2004).

O trabalho desenvolvido possui estrutura desconhecida (a ser aprendida) e dados incompletos (mas em pequena quantidade), neste sentido, pode ser classificado da seguinte forma:

DADOS	ESTRUTURA DA REDE	
	CONHECIDA	DESCONHECIDA
COMPLETOS	A	B
INCOMPLETOS	C	D

DADOS	ESTRUTURA DA REDE	
	CONHECIDA	DESCONHECIDA
COMPLETOS	A	B
INCOMPLETOS	C	MDL - "Minimum Descriptive Length" (ESTRUTURA DA REDE) + EM, Expectation Maximization (DADOS INCOMPLETOS)

Figura 5 - Dados e Estruturas em Redes Bayesianas - Algoritmos Utilizados

Como o software selecionado utiliza a métrica **MDL(Minimum Descriptive Length)**, então toma como base a Teoria da Informação, conforme Carvalho (2011), tema que será tratado de forma breve no próximo tópico.

3.3 Teoria da Informação e Entropia

Este tópico foi preparado a partir de uma seleção de indicadores selecionados por Conrady e Jouffe (2011). A esta seleção foram incorporadas novas informações originadas no levantamento bibliográfico visando maior compreensão dos conceitos.

Neste trabalho o uso da palavra Entropia está relacionado à Teoria da Informação e a compreensão do seu significado tem origem nas ideias de Shannon (1948).

Pineda (2006) apresenta as semelhanças entre entropia utilizada na teoria cinética dos gases desenvolvida por Boltzmann, 1877, e a teoria da informação iniciada com Shannon, 1948, o que pode ajudar a compreender, através de analogia, o significado de entropia

ENTROPIA		
AUTOR	BOLTZMANN	SHANNON
OBJETO	GASES	INFORMAÇÃO
CONCEITO	DESORDEM MOLECULAR	QUANTIDADE DE INFORMAÇÃO
MEDIDA	DISTRIBUIÇÃO DE VELOCIDADES DAS MOLÉCULAS	PROBABILIDADE DE OCORRÊNCIA DOS SÍMBOLOS
VARIAÇÃO	AUMENTA À MEDIDA QUE O MOVIMENTO DAS MOLÉCULAS SE TORNA ALEATÓRIO	AUMENTA À MEDIDA QUE A OCORRÊNCIA DOS SÍMBOLOS SE TORNA ALEATÓRIA (EQUIPROVÁVEIS)

Tabela 1 - Comparativo Entropia

FONTE : BASEADO PINEDA , 2006 p. 88

Antes de esclarecer sobre entropia como apresentado por Shannon, vale salientar que não se apresenta nesta tese definição do termo informação, considerando que o mesmo está referenciado em diversos trabalhos como em Setzer (2001), apenas para citar um autor com muitas discussões sobre o termo. Foca-se aqui na visão de Shannon (1948) que trata a informação independente do seu significado, reduzindo-a a um nível básico, cuidando da sua codificação e garantindo que possa ser quantificada - ideia esta mais próxima do conceito de “dado” como entidade matemática, conforme apresentado por Setzer (2001).

Shannon trabalhou e desenvolveu o conceito de entropia nos seus experimentos e realizações com mensagens de texto, tendo como objeto de trabalho as palavras e códigos. Este conceito foi aos poucos estendido para outras áreas da Teoria da Informação, mantendo-se inclusive a formula de calculo para medir a entropia da informação.

Pineda (2006, p.31) resume as ideias de Shannon sobre entropia da seguinte forma:

Shannon propõe que a quantidade de informação deve ser entendida como a entropia mecânica da estatística, e observa que... à medida que a ocorrência de um grupo de símbolos se torna mais provável que a dos outros ... a entropia decresce. A entropia máxima só é atingida quando a ocorrência de todos os símbolos é equiprovável (ou seja, não existe tendência de concentração de probabilidades em algum grupo de símbolos). Quando existe certeza sobre qual o símbolo vai ser transmitido, a entropia é zero.

De acordo com Conrady e Jouffe (2011) a entropia mede a incerteza inerente à distribuição de uma variável aleatória “x” que pode assumir diversos estados, sendo representada pela seguinte fórmula:

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

Esta mesma fórmula já aparecia em Shannon (1948), assim como o Gráfico 8, onde o autor já utilizava a medida de entropia em bits. Observa-se o uso do logaritmo na base dois, expressando os bits.

Para ilustrar e compreender a entropia Conrady e Jouffe (2011) tomam como exemplo o lançamento de uma moeda, não viciada, onde dois estados são equiprováveis $X=\{\text{Cara, Coroa}\}$ e a entropia igual a 1 bit. Como apresentado nas ideias de Shannon nesta situação a ocorrência de cara ou coroa são equiprováveis, com entropia máxima.

$$\begin{aligned} H(X_{\text{Não_Viciada}}) &= -p(\text{Cara}) \log_2 p(\text{Cara}) - p(\text{Coroa}) \log_2 p(\text{Coroa}) \\ &= -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 0.5 + 0.5 = 1 \text{ bit} \end{aligned}$$

Caso a moeda fosse viciada e os dois estados não tivessem a mesma probabilidade de ocorrência $X=\{\text{Cara, Coroa}\}$, a entropia não seria mais a máxima e sim :

$$H(X_{\text{Viciada}}) = -0.7 \log_2 0.7 - 0.3 \log_2 0.3 = 0.88$$

A entropia binária pode ser visualizada num Gráfico 8 onde todas as possibilidades (incertezas) estariam representadas para a possibilidade de se obter “cara”.

Gráfico 8 - Entropia -Moeda

FONTE : CONRADY E JOUFFE (2011)

Através do Gráfico 8 pode-se deduzir que a entropia nunca é negativa, no mínimo é zero, ela cresce com a incerteza da variável e o ponto de maior valor representa a incerteza máxima. Para uma variável binária é fácil saber que neste ponto a probabilidade é de 0.50 (possibilidades equiprováveis). De acordo com Camorim (2010) é possível provar que o mesmo ocorre para qualquer que seja o número de estados, sendo que o valor máximo assumido pela entropia, no caso geral, é :

$$H_{\max}(X) = \log_2 |X| ; \text{ onde } |X| \text{ é o número de elementos de } X.$$

A compreensão do conceito de entropia é a base para a compreensão das métricas utilizadas para as redes bayesianas neste trabalho, além disso, este conceito é fundamental nos dias de hoje onde a informação cada vez mais passa a ser vista como um fenômeno físico e em grandes proporções. Para maior compreensão sobre os diversos aspectos da entropia na teoria da informação sugere-se consultar Gray (2009, p. 39).

3.3.1 Métricas Utilizadas: Entropia Condicional, Relativa e Informação Mútua.

A **entropia condicional** está relacionada ao ganho de informação que uma variável pode trazer sobre o conhecimento de outra. Ou seja, duas variáveis X e Y podem estar relacionadas, e ao conhecer sobre uma pode-se também ganhar informação sobre a outra. Em outras palavras à medida que se tem o valor de Y a incerteza sobre X pode diminuir. Se Y não confere nenhuma informação sobre X então $H(X|Y) = H(X)$.

A fórmula da entropia condicional de X , dado Y é dada por :

$$H(X|Y) = H(X,Y) - H(Y)$$

De acordo com Conrady e Jouffe (2011) talvez este seja o conceito mais importante nos cálculos feitos pelo software que está sendo utilizado neste trabalho (Bayesia, 2012).

Já Informação mútua, representada por $I(X;Y)$ mede o quanto (em média) o conhecimento ou observação de uma variável (X) aleatória reduz a incerteza sobre outra (Y). Ou seja, quanto de entropia de X é reduzido dado uma informação sobre Y .

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

A informação mútua é simétrica, ou seja conhecendo-se uma reduz-se a incerteza sobre a outra, e vice-versa. Na sequência apresenta-se um diagrama com objetivo de esclarecer sobre as diversas entropias.

Figura 6 -Diagrama de Venn para representação das entropias, entropias condicionais e informação mútua

FONTE : BASEADO EM CAMORIM, R.D. L ;2010; P11

A entropia relativa conhecida também como Divergência de Kullback-Leibler (DKL) ou entropia cruzada foi introduzida por Solomon Kullback e Richard Leibler em 1951. Esta métrica permite calcular a similaridade entre distribuições de probabilidade (GRAY,2009). Em redes bayesianas é utilizada também para medir o ganho de informação quando se move

da distribuição priori para a posteriori. Se um novo dado ($Y = y$) é conhecido, ele pode ser utilizado para atualizar a distribuição de probabilidade de X . Vale observar que esta medida, apesar de ser referenciada muitas vezes como distância, não tem este efeito pois não é simétrica.

Para ficar mais claro o cálculo do DKL volta-se ao exemplo da moeda (Cara ou Coroa). Primeiramente assume-se que uma moeda não é viciada ($p=0,50$ e $q=0,50$). Depois, jogando a moeda várias vezes observa-se que a probabilidade é diferente ($p=0.75$ e $q = 0.25$). Usando a medida de DKL pode-se calcular distorção entre o que fora assumido e o que realmente aconteceu. Neste caso a medida está sendo utilizada para comparar as distribuições a priori e posteriori.

$$\begin{aligned}
 D_{KL} &= (p(X) \parallel q(X)) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)} \\
 &= p(\text{cara}) \log_2 \frac{p(\text{cara})}{q(\text{cara})} + p(\text{coroa}) \log_2 \frac{p(\text{coroa})}{q(\text{coroa})} = 0.75 \log_2 \frac{0.75}{0.5} + 0.25 \log_2 \frac{0.25}{0.5} \\
 &= 0.188722 \text{ bits}
 \end{aligned}$$

Figura 7 - Exemplo cálculo do DKL

FONTE : CONRADY & JOUFFE (2011)

Outro exemplo dado pelos mesmos autores Conrady e Jouffe (2011): deseja-se conhecer a previsão de chuva em duas cidades do estado de Victória, na Australia. Usando a previsão de chuva medida no aeroporto Geelong e no aeroporto Tullamarine em Melbourne, os quais estão aproximadamente a 80 km de distância. Utilizou-se todo o calendário de 2010. Dado a proximidade das duas cidades poderia se esperar climas parecidos. O tempo em Geelong não é reportado no jornal e os viajantes precisam usar o clima de Melbourne como estimativa. O observatório do clima diz que em 40.3% dos dias chove em Melbourne, enquanto em Geelong este índice é de 47.4% no ano. Percebe-se pela Figura 8 que a métrica DKL não é simétrica.

$$\begin{aligned}
 D_{KL} &= (p_{\text{Geelong}}(X) \parallel p_{\text{Melbourne}}(X)) = \sum_{x \in X} p_{\text{Geelong}}(x) \log_2 \frac{p_{\text{Geelong}}(x)}{p_{\text{Melbourne}}(x)} \\
 &= p_{\text{Geelong}}(x = \text{No Rain}) \log_2 \frac{p_{\text{Geelong}}(x = \text{No Rain})}{p_{\text{Melbourne}}(x = \text{No Rain})} + p_{\text{Geelong}}(x = \text{Rain}) \log_2 \frac{p_{\text{Geelong}}(x = \text{Rain})}{p_{\text{Melbourne}}(x = \text{Rain})} \\
 &= 0.526 \log_2 \frac{0.526}{0.597} + 0.474 \log_2 \frac{0.474}{0.403} = 0.0148958 \text{ bits} \\
 \\
 D_{KL} &= (p_{\text{Melbourne}}(X) \parallel p_{\text{Geelong}}(X)) = \sum_{x \in X} p_{\text{Melbourne}}(x) \log_2 \frac{p_{\text{Melbourne}}(x)}{p_{\text{Geelong}}(x)} \\
 &= p_{\text{Melbourne}}(x = \text{Rain}) \log_2 \frac{p_{\text{Melbourne}}(x = \text{Rain})}{p_{\text{Geelong}}(x = \text{Rain})} + p_{\text{Melbourne}}(x = \text{No Rain}) \log_2 \frac{p_{\text{Melbourne}}(x = \text{No Rain})}{p_{\text{Geelong}}(x = \text{No Rain})} \\
 &= 0.403 \log_2 \frac{0.403}{0.474} + 0.597 \log_2 \frac{0.597}{0.526} = 0.0147077 \text{ bits}
 \end{aligned}$$

Figura 8 - DKL - Exemplo -Clima

FONTE : CONRADY & JOUFFE (2011)

As forças dos arcos das redes bayesianas desenvolvidas neste trabalho usam prioritariamente a entropia relativa, a qual descreve a força e o sentido dos arcos, conforme ilustrado na Figura 9.

Figura 9 - Entropia Relativa e Força dos Arcos

3.4 Uso de um software

Para a modelagem será utilizado um software. Segundo Vassão (2010) todos os programas de computador são “meta-entidades”. Um exemplo bem simples dado pelo autor é um software denominado MetaFont, que é um programa de computador que desenha fontes tipográficas de maneira semiautomática, ao contrário do uso tradicional do “type design”, em que o “designer” cria letra por letra. O que muda quando se usa o MetaFont é que o “designer” indicará os ajustes necessários sendo que o autor do programa de computador cuidou de estabelecer as bases da tipografia, dentro do que entendia ser possível e mais universal possível. Este fato de um lado, facilita o trabalho do “designer”, que por sua vez deve aprender a lidar com as potencialidades e limitações do MetaFont. O mesmo acontece com as Redes Bayesianas ou outras metodologias de IA quando se escolhe um meta-programa. O “designer” da informação estará submetido às escolhas e possibilidades do software. Para um generalista a quantidade de cortes nos caminhos analíticos, a capacidade de gerar os grafos automaticamente e relatórios esclarecedores superam as desvantagens de não ter este ou aquele teste específico. Assim, os programas de computador são, segundo Vassão (2010, p. 19-21) objetos abstratos que ajudam a construir, norteiam outros objetos abstratos, dentro de um conceito de *MetaDesign*.

O software utilizado foi o BayesiaLab (BAYESIA,2012), escolhido por permitir trabalhar com casos perdidos (“missing cases”), possuir interface amigável e uma lista de relatórios

facilitadores para a interpretação dos resultados e principalmente por fornecer : (a) Algoritmos para descoberta da associação entre variáveis com meta-heurísticas que revelam a estruturas possíveis para os dados (b) Algoritmos para classificação; (c) Métricas que permitem avaliar a adequação das estruturas obtidas ;(d) Possibilidade de interferir na heurística dos modelos incluindo arcos proibidos (e) Capacidade de validação das estruturas e arcos que ligam os nós, identificando-os como prováveis ou não.

3.4.1 Algoritmos de classificação disponíveis

Este tópico e o seguinte foram desenvolvidos com base na experimentação do software e no sistema de ajuda do mesmo e visa esclarecer sobre os algoritmos disponíveis.

O primeiro conjunto de algoritmos permitem a aprendizagem de Redes Bayesianas dedicadas à caracterização de uma variável resposta ou destino, ou seja possuem como função maior implementar um modelo de classificação. Neste trabalho estes algoritmos são importantes pois pretende-se chegar a um “ranking” que classifique as famílias brasileiras de acordo com a probabilidade de estarem em dificuldade.

Naive Bayes: Valorizada pelos que procuram praticidade e simplicidade e criticada pelos que necessitam considerar adequadamente a dependência entre as variáveis. Possui uma arquitetura pré-definida na qual o nó de destino é o pai de todos os outros nós. Apesar desta hipótese de independência das variáveis ser geralmente falsa, na maioria dos casos, o baixo número de probabilidades para estimar torna esta estrutura robusta e com um tempo de aprendizado muito curto. Apesar da acuracidade na classificação , este tipo de rede não é informativo sobre a relação entre as variáveis, pressupondo independência dos nós filhos. Na sequência apresenta-se um exemplo da estrutura de uma rede Naive Bayes.

Figura 10 -Naive Bayes

Fonte: Extraída do Bayesia a partir dos dados da POF 2009.

Naive Bayes Aumentada: permite relaxar a forte restrição de independência entre as variáveis conforme acontece no exemplo anterior. Esta arquitetura enriquece uma Naive Bayes com relações entre os nós filhos. A precisão desse algoritmo é melhor que do “naive bayes”, além disso, a relação entre as variáveis passa a ser conhecida e as Tabelas de distribuição conjunta apresentam boa aderência com as Tabelas observadas através do processamento do banco de dados.

Figura 11 - Naive Bayes Aumentada

Fonte: Extraída do Bayesia a partir dos dados da POF 2009.

Árvore Naive Bayes Aumentada: Esta arquitetura é também naive, em que uma árvore de abrangência máxima é aprendida. A precisão da previsão desse algoritmo é melhor do que aqueles obtidos pela arquitetura “naive”, mas não tão boa quanto a obtida com Naive Bayes Aumentada, no entanto, esse algoritmo é muito mais rápido do que o anterior e pode ser uma saída para redes muito complexas.

Filhos e Pais : Esta estrutura é, em certa medida uma arquitetura “naive” aumentada em que os filhos não são fixos, a priori, mas procurados de acordo com a dependência marginal dos nós em relação ao alvo. Este algoritmo tem a vantagem de destacar os nós que não estão relacionados diretamente com o alvo, ou variável resposta. A duração da aprendizagem é comparável ao de uma arquitetura naive aumentada.

Markov Blanket: algoritmo que busca os nós pertencentes ao Markov Blanket do nó de destino, alvo ou resposta. A busca por essa estrutura é totalmente focada no nó “alvo”, fazendo com que seja possível obter mais rapidamente o subconjunto de nós que são mais importantes para compreender a variável resposta. É um algoritmo de seleção muito poderoso em processos de classificação e redução de escopo de variáveis. Em uma rede bayesiana, qualquer variável é influenciada diretamente somente pelas variáveis presentes no seu Markov Blanket.

Figura 12 - Markov Blanket

Fonte: Preparada pelo autor baseado em BAI (2005, p. 5)

Além do Markov Blanket existem outras variações como o Markov Blanket Aumentado. Quando se procura por uma tarefa de predição pura (como por exemplo uma função do tipo “escoragem” ou “ranking”), o algoritmo Markov Blanket Aumentado é geralmente mais preciso que o anterior.

Neste estudo os algoritmos de classificação serão necessários, uma vez que se espera obter um eixo de probabilidade que permita classificar as famílias de acordo com sua propensão a estar em “**dificuldade**”, que passa a ser o nó de destino, alvo ou variável resposta.

3.4.2 Algoritmos para descoberta de associações entre variáveis :

O número de Redes Bayesianas a partir de um determinado número de variáveis é tão grande que torna impossível realizar uma busca exaustiva das melhores redes. Os algoritmos de aprendizagem possuem heurísticas que permitem reduzir este espaço de busca. O software utilizado suporta cinco algoritmos de aprendizagem para descoberta da estrutura da rede e estimativa das Tabelas de probabilidade condicional. As heurísticas de cada método são diferentes, assim os resultados de cada um podem ser também diferentes. No entanto, como todos os métodos de aprendizagem utilizam a mesma métrica comparativa, MDL, *Minimum Descriptive Length*, as redes podem ser comparadas entre si. Neste trabalho não será dada tanta importância à comparação do MDL, deixando esta tarefa para a máquina.

São apresentados, na sequência do mais rápido para o mais lento, os diferentes métodos aplicados neste trabalho. O conceito de cada método utilizado não será detalhado, uma vez que não existem muitas referências acadêmicas a trabalhos similares executados com estes procedimentos.

Máximo Spanning Tree : Este algoritmo de aprendizagem é mais rápido dentre os não supervisionados. Ele se baseia apenas em duas passagens. A primeira consiste em calcular o valor das relações entre todas as variáveis, o segundo consiste então na construção da rede. Mesmo que a rede resultante não seja a ideal, este método pode ser usado para (1) uma primeira imputação de valores perdidos (“missing cases”) (2) como rede inicial antes de outras redes. Além disso, este método pode ser útil como base de processos de agrupamento de variáveis, através da métrica de Kullback-Leibler , conforme apresentam Conrady e Jouffe (2011).

Tabu : o aprendizado estrutural através deste algoritmo no espaço das Redes Bayesianas é particularmente útil para a refinar uma rede já construída por especialistas ou para atualizar uma rede cuja aprendizagem ocorreu em um conjunto de dados diferentes. Nesta rede é possível definir número máximo de pais e filhos permitidos.

EQ: Método de pesquisa que procura classes equivalentes de Redes Bayesianas. Este método é muito eficiente porque se propõe a evitar uma grande quantidade de “mínimos locais” e reduz fortemente o tamanho do espaço de busca. O EQ pode ser calculado a partir de uma rede informada, sendo possível escolher a estrutura da rede para iniciar a aprendizagem. Uma referência a este tipo de método é Munteau, e Bendou (2001). Este trabalho propõe

uma teoria e uma estrutura algorítmica para a análise eficiente de algoritmos de aprendizagem que exploram o espaço de classes de equivalência das estruturas de rede Bayesiana. Aplicações práticas com este método podem ser observadas em Conrady e Jouffe (2011).

SopLEQ : um método de pesquisa de estrutura baseado numa caracterização global dos dados e na exploração das propriedades de equivalência de Redes Bayesianas. Pouca referência foi encontrada para este método (JOUFFE , 2001).

Tabu Order : método de aprendizagem utilizando Tabu, busca no espaço da ordem dos nós da rede Bayesiana para encontrar a melhor rede para uma ordem de nós fixos. Consiste em escolher os pais de um nó entre os nós que aparecem antes da ordem considerada. Este é o método de busca mais completo, mas também o mais demorado. Além das opções de padrão, é possível definir o tamanho da lista para o “Tabu Order”. Este algoritmo foi o escolhido para o modelo final e será apresentado durante o tópico que apresenta a modelagem (4.6). Mais referência sobre este algoritmo combinado com Markov Blanket encontra-se em Bai (2005).

3.4.3 Parâmetros para avaliar redes bayesianas para efeito de classificação:

Os resultados observados por Karcher (2009) e Cui *et al.* (2006) demonstram que as Redes Bayesianas têm capacidade de classificação superior ou similar a outras técnicas de classificação. Apesar dos dois trabalhos reforçarem a superioridade das Redes Bayesianas não é conveniente considerá-las como muito melhores quando o objetivo é a mera classificação, sem observar valores das probabilidades calculadas ou a topologia da rede. No entanto, quando o objetivo é também conhecer, através do modelo, a relação e interação entre as variáveis, de forma visual, justifica-se o uso das redes bayesianas conforme CUI *et al.* (2006).

O presente trabalho representa uma oportunidade de testar um número maior de variáveis dispostas numa Rede Bayesiana, neste sentido a “interpretabilidade” visual das redes poderá ser testada em condições mais complexas.

COMPARATIVO DE MÉTODOS ALTERNATIVOS - MODELO COM VARIÁVEL RESPOSTA BINÁRIA ⁽¹⁾				
CRITÉRIO	REDES BAYESIANAS	REDES NEURAIIS	CART ⁽²⁾	REGRESSÃO LOGÍTICA
ACURÁCIA	ALTA	ALTA	MÉDIA	ALTA
TRANSPARÊNCIA	ALTA	BAIXA	BAIXA	MÉDIA
INTERPRETABILIDADE	ALTA	BAIXA	BAIXA	MÉDIA

(1) RESPOSTA A CAMPANHA DE MARKETING DIRETO
(2) CLASSIFICATION AND REGRESSION TREE

Tabela 2 - Comparativos entre métodos de classificação

FONTE : BASEADO EM CUI, WONG, LUI , 2006 p. 597-61

Parâmetros relacionados à transparência e interpretação podem ser considerados como medidas subjetivas, não sendo possível estabelecer neste trabalho uma métrica, uma vez que o potencial de alavancar a interpretação e a transparência dos modelos finais não será testado com outras pessoas. As conclusões sobre estes parâmetros serão apenas baseadas na percepção do próprio pesquisador, podendo ser um tópico para futuros estudos a possibilidade de verificar percepções sobre a transparência e interpretabilidade das técnicas entre especialistas e não especialistas em análise de dados

Já a acurácia pode ser medida e para isso foi selecionada a seguinte métrica : curva ROC - “Receiver Operating Characteristic” dado que o modelo proposto trata-se de um “ranking” de probabilidades que se propõe a classificar uma população através de uma variável resposta binária. As mesmas métricas podem ser observadas em Hoogenberg e Brinke (2004), Lecumberri e Duarte (2003) ; Sicsu (2010) e Annibal, (2008).

A análise ROC (Receiver Operating Characteristics), tem origem na teoria estatística desenvolvida entre 1950 e 1960 (BRAGA, 2000), permite avaliar os impactos do erro nos modelos, considerando sua imperfeição. É uma forma de avaliar o potencial de um modelo discriminar (classificar corretamente) uma determinada população.

A curva ROC é caracterizada pela representação gráfica de todas as possibilidades de Verdadeiros Positivos (VP), nas ordenadas, comparados a todas as possibilidades de Falsos Positivos (FP), nas abscissas, resultantes da variação do valor de corte ao longo de um eixo de decisão. O eixo do VP é também denominado “sensitividade” e o eixo do VN “especificidade” do modelo. O ponto de corte é denominado “limiar” ou “*threshold*”, utilizando termo em inglês, aqui apresentado para facilitar correspondência com a literatura e parâmetros fornecidos pelos softwares.

A Tabela 3 representa os verdadeiros positivos e os verdadeiros negativos para um modelo ilustrativo no qual se compara a classificação de dois grupos o A e o B. Foi classificado como grupo A todos os que apresentaram probabilidade menor que 0,50 (ponto de corte ou limiar) e como grupo B os que apresentaram probabilidade maior do que 0,50. Com este corte, hipotético, observou-se um acerto geral de 83%, no exemplo dado.

EXEMPLO : CLASSIFICAÇÃO MODELO				
DADOS MERAMENTE ILUSTRATIVOS	CLASSIFICAÇÃO PELO MODELO		CLASSIFICAÇÃO PELO MODELO	
	GRUPO A'	GRUPO B'	GRUPO A	GRUPO B
GRUPO A	8,600	1,800	VP	FN
GRUPO B	1,800	8,600	FP	VN
ACERTO = (VP+VN)/TOTAL		83% ——— 17.200	VP = Verdadeiro Positivo FN = Falso negativo FP = Falso Positivo VN = Verdadeiro Negativo	
		20,800		
GRUPO A' = PONTO DE CORTE : probabilidade < 0,50				
GRUPO B' = PONTO DE CORTE : probabilidade >= 0,50				

Tabela 3 - Tabela com exemplo de Verdadeiros Positivos e Verdadeiros Negativos

Fonte : baseado em PAULA (2004: 113) e HOOGENBERG & BRINKE (2004 :243)

A Figura 12, a seguir, foi desenvolvida com base em simulações didáticas feitas no site <http://www.anaesthetist.com/mnm/stats/roc>, representando o efeito de diferentes limiares (pontos de corte) no computo de Verdadeiros Positivos e Falsos Positivos ao longo da curva ROC.

Os exemplos abaixo mostram pontos de corte extremos, apenas para que se possa observar a interferência destes na proporção de verdadeiros positivos e verdadeiros negativos, dado um modelo de classificação.

Figura 13 - Exemplo interpretativo da curva ROC

FONTE : Ilustração preparada com base em simulações calculadas
em: <http://www.anaesthetist.com/mnm/stats/roc/Findex.htm>

Apenas para esclarecer, os formatos em curva normal na Figura 13 não representam nenhum pressuposto de normalidade para o teste, as formas foram usadas apenas para efeito didático. Quanto maior a AUROC (“Area Under ROC Curve”) maior o poder de discriminação de um modelo. Num modelo, a proporção de acertos dependerá do ponto de corte utilizado para a classificação e também da área sobre a curva ROC (AUROC).

Hosmer e Lemeshow (2000, p. 162) sugerem uma interpretação para a capacidade de discriminação de um modelo logístico, considerando o valor obtido pela AUROC :

Área abaixo da curva Roc e a Discriminação de Modelos	
AUROC	Discriminação
AUROC = 0,50	Sugere que o modelo não discrimina
$0,50 < \text{AUROC} < 0,70$	Pouca Discriminação
$0,70 \leq \text{AUROC} < 0,80$	Discriminação Aceitável / Boa Discriminação
$0,80 \leq \text{AUROC} < 0,90$	Excelente Discriminação
$\text{AUROC} \geq 0,90$	Não usual - fora do comum

Tabela 4 - AUROC e a discriminação de modelos – tabela adaptada

FONTE : Baseado em HOSMER & LEMENSHOW, 2000, p. 162

Esta Tabela foi utilizada no presente trabalho para avaliar as classificações feitas através das Redes Bayesianas e visa modelos com ROC superior a 0,70, pelo menos.

3.4.4 Estrutura das Redes Bayesianas

Os nós (representados por círculos) nas Redes Bayesianas representam as variáveis de interesse. As ligações (arcos ou flechas) representam dependências ou em algumas situações, relações (causais) entre as variáveis. As dependências são quantificadas por probabilidades condicionais dado os antecessores da variável na rede, conforme Figura 14. A estrutura da rede pode ser determinada pelo próprio especialista ou pesquisador, manualmente, ou a partir de bases de dados e algoritmos que permitem o aprendizado automático.

Um exemplo didático de Redes Bayesianas é apresentado na Figura 14, o qual baseou-se em bancos de dados para a descoberta de estrutura.

Figura 14 - Exemplo Didático Rede Bayesiana : Condição da Moradia

Fonte: Elaborado a partir de dados da POF 2003

Na Figura 14, o nó (M) Condição da Moradia, tem como pai os nós (C) Casa Escura, (G) Goteira e (U) Umidade e um ancestral (E) Problema na Estrutura da Casa. Goteira (G) além de interferir na Umidade (U) interfere diretamente na percepção de Condição da Moradia (M). Para cada nó existe uma Tabela de contingência conforme Figura 15.

Figura 15 - Exemplo Didático Rede Bayesiana -Distribuições Marginais -A Priori

Fonte: Elaborado pelo autor a partir de dados da POF 2003

Quando cada um dos nós é instanciado, os demais nós apresentam automaticamente novos valores, dependendo das probabilidades condicionais entre eles. O processo de instanciar um nó pode ocorrer em qualquer parte da rede de forma bidirecional. A Figura 15 apresenta uma nova configuração dos resultados de probabilidade *a posteriori*, dado que “Umidade” e “Goteira” passaram a ser conhecidos, foram instanciados. Percebe-se agora que a probabilidade da condição de moradia que *a priori* era de 27 % de ser “ruim”, passa para 99%, representando *a posteriori*. Mudanças também foram observadas nas variáveis: Casa Escura e Problema na Estrutura da Casa, uma vez que ao instanciar uma variável, toda a rede é afetada, a partir das dependências informadas. Caso o modelo esteja bem definido e aderente ao mundo real, produz respostas coerentes em relação ao pensamento racional.

Pela Figura 17 observa-se que três nós apontam para a condição de moradia “Casa Escura”, “Umidade” e “Goteira” o que significa que no nó Condição da Moradia existe uma Tabela de contingência com as 4 variáveis.

Figura 16- Exemplo Didático Rede Bayesiana -Distribuições Marginais -A Posteriori

Fonte: Elaborado pelo autor a partir de dados da POF 2003

UMIDADE (U)	GOTEIRA (G)	CASA_ESC...	RUIM	BOA
Nao	Nao	Nao	0.000	100.000
		Sim	0.000	100.000
	Sim	Nao	38.342	61.658
		Sim	51.185	48.815
Sim	Nao	Nao	29.960	70.040
		Sim	45.582	54.418
	Sim	Nao	99.963	0.037
		Sim	99.987	0.013

Figura 17 - Exemplo Didático Rede Bayesiana – Condição da Moradia

Fonte: Elaborado pelo autor a partir de dados da POF 2003

Uma das grandes vantagens das redes bayesianas está na capacidade de inferir as probabilidades conjuntas a partir dos dados, o que seria extremamente complexo para um ser humano.

3.5 Redes Bayesianas - Relações Causais

Causalidade é compreendida como a relação entre uma causa e seu efeito. Parece um conceito óbvio, pois o ser humano lida com isso no raciocínio do dia-a-dia. No entanto, este conceito não é tão simples assim, sendo motivo de reflexão filosófica desde a antiguidade. Não é objetivo analisar profundamente a filosofia da causalidade, no entanto faz-se necessário esclarecer causalidade nesta tese uma vez que Redes Bayesianas são também conhecidas como

modelos causais, quando podem ser compreendidas desta forma. Para Pearl (2009) uma das perspectivas mais interessantes tem sido a possibilidade da utilização de Redes Bayesianas para descobrir estruturas causais em dados brutos. Para ele o aspecto mais importante de uma rede bayesiana é que as mesmas são a representação direta do mundo, não do processo de raciocínio. O autor afirma que a palavra causalidade tem sido evitada em muitos estudos científicos.

Para Aristóteles questões do tipo “por quê?” poderiam ser respondidas de quatro diferentes maneiras, conforme apresenta Chauí (2000) : com base na causa material , causa formal, causa eficiente e causa final. Um exemplo clássico: uma casa tem sua causa material nos tijolos e demais materiais, sua causa formal no projeto, sua causa eficiente no trabalho do pedreiro e sua causa final na sua função, que é morar. Este pensamento predominou no pensamento ocidental, até o renascimento e o iluminismo. Nesta época a explicação causal estava baseava na ideia da linearidade que vê o efeito presente, na causa que o precedeu, que por sua vez relaciona-se a outra causa anterior e assim por diante. A determinação depende do passado, numa única linha causal explicada pelas condições iniciais do universo, que no final é uma causa sem causa. Segundo Barata (1997) o pensamento causal sofreu modificações importantes com Galileu Galilei, no século XVI, no momento em que a ciência começa a se distanciar da filosofia e da religião. Galileu passa a utilizar a dedução matemática no lugar da dedução lógica, observando fatos, quantificando os elementos, e verificando através de experimentos imaginários, onde as leis matemáticas expressariam relações de causa e efeito. A causa dos fenômenos naturais começa a ser atribuída ao movimento atômico.

No final do iluminismo, apresenta-se na Wikipédia (2011) : a ideia de causalidade linear perde espaço para explicar a maioria dos fenômenos sejam eles naturais, sociais, biológicos, entre outros, necessitando de novos modelos de causalidade. Surgem as reflexões sobre a causalidade complexa, em que nem todo efeito está totalmente contido na causa anterior e vê a determinação ocorrendo não no passado, mas no presente ou na simultaneidade dos processos ou eventos.

Conforme apresenta Pessoa Jr (2004) para o iluminista Hume – 1739 - as conexões causais resultam de hábitos que estabelecemos, não havendo relação de causalidade de fato. Kant também aceitou esta conclusão de Hume e colocou a causalidade como uma “categoria de entendimento” e não como uma coisa em si. Segundo o autor, no final do século XIX a causalidade foi banida até das explicações científicas, com a ascensão do positivismo. Surge a Estatística e o conceito de correlação ganha o espaço antes reservado para causalidade. A

partir de Karl Pearson - 1911- a correlação entre duas variáveis passa a ser a medida científica da relação entre duas coisas, oferecendo instrumento para analisar relações, não necessariamente causas e efeitos. Ronald Fisher, -1936 - introduziu a noção de “experimento randomizado” , estabelecendo método de controle para testar eficácia de um procedimento. Por exemplo, para testar se um remédio tem eficácia ou não deve-se observar um grupo de controle. Este fato permitiria identificar causas através de um procedimento que não só observa, mas também intervém.

Bender (1996 p.299, tradução nossa) alerta : “causalidade é a dádiva da nossa existência e a maldição de nossa matemática. Nós confiamos na causalidade em nossas interações com a realidade, mas achamos difícil captura-la em nossa abstração”.

A causalidade em Redes Bayesianas está ligada à inteligibilidade que esta rede pode trazer para a compreensão de fenômenos, inclusive sociais. Segundo Pearl (2009) a análise causal visa inferir não apenas as probabilidades dentro de condições estáticas, mas também a dinâmica de crenças que podem mudar estas condições. Judea Pearl é otimista e acredita que a ideia de intervir nas variáveis, e numa rede de variáveis, é uma novidade conceitual importante na teoria da causalidade contemporânea.

Em Redes Bayesianas nem sempre é possível assegurar relações causais, apesar do grande salto que isso pode representar para os estudos analíticos. O entusiasmo de Pearl (2009) e seu conhecimento sobre o raciocínio e a filosofia da causalidade são convincentes e não se pode deixar de tentar exercitar estas possibilidades através das redes bayesianas em estudos futuros. O autor criou um método denominado “cirúrgico” (PEARL, 2010, p.428) para análise da causalidade em grafos acíclicos o qual não foi testado no presente trabalho . O autor convida utilizar as Redes Bayesianas no entendimento dos grandes problemas da humanidade (PEARL, 2010, p.428, tradução nossa) :

“... o problema desafiador está ainda por vir : continuamos não tendo um entendimento sobre a pobreza, o câncer, a intolerância, apenas acumulando dados e insights ... poder-se-á eventualmente lidar com tal entendimento ... Os dados estão em todos os lugares e agora o ábaco está a disposição, também. Espero que esta combinação amplifique cada um desses componentes”.

Combinada com estudos da causalidade as redes bayesianas poderão promover grande salto analítico para a pesquisa científica. Infelizmente esta tese não consegue ser conclusiva neste sentido, e espera-se ver muitos outros trabalhos nesta linha. As redes aqui apresentadas limi-

tam-se a representar conexões entre variáveis e conceitos, sendo que algumas podem sofrer interpretações causais. Segundo Becker (2008, p. 66) conexão é uma palavra vaga pois é utilizada como sinônimo de “influência”, “causalidade” ou “dependência”, sendo indicativa de variações de processos, uns sobre os outros de modo complexo.

Uma rede bayesiana pode ser construída a partir de pesquisas prévias e dados consolidados como apresenta Blodgett e Anderson (2000) ou mesmo a partir do conhecimento prévio de um especialista, conforme Cooper (2000). Quando existem bases de dados disponíveis pesquisadores podem utilizar as Redes Bayesianas para “compreender” a distribuição de probabilidade que melhor descreve a relação causal entre as variáveis. Vale salientar que mecanismos híbridos baseados em dados explícitos, presentes em bases de dados, e conhecimento tácitos, do especialista, podem ser utilizados para compor uma Rede Bayesiana. As Redes Bayesianas são representadas em grafos e estes, além da capacidade de comunicar a ideia final do modelo, são também uma forma importante no projeto e análise de algoritmos para aprendizado de máquinas. Um modelo Gráfico confere modularidade e diminui a complexidade, dado que combina partes e assegura a consistência do sistema (JORDAN, 1999, p.1).

Neste trabalho será priorizada a obtenção do conhecimento através do tratamento dos dados e aprendizado de máquina, considerando que o volume disponibilizado pela POF encoraja tal empreitada. Além disso, considera-se um avanço poder raciocinar sobre fenômenos sociais com ajuda da máquina. Para tentar se aproximar de conceitos causais algumas variáveis serão apontadas oportunamente como “proxy de causa” e outras como “proxy de consequência”, permitindo ao pesquisador interferir no processo de aprendizado de máquina. Este procedimento será apresentada no item 4.5.6, no entanto não garante que relações causais possam ser comprovadas.

3.6 Tratamento de variáveis para redução da complexidade

De acordo com Vassão (2010) as escalas de complexidade podem ser entendidas sob a perspectiva da Cibernética, citando Ashby (1952), para quem a complexidade seria função do número de elementos de um sistema, ou seja, quanto maior o número de entidades, maior sua complexidade. Esta noção de complexidade é diferente da apresentada no item 3.1, mas em Redes Bayesianas é também desta forma que a complexidade é abordada, ou seja, pelo número de elementos da rede. Neste sentido a abstração e a criação de ontologia, compreendida

aqui como seleção, nomeação dos objetos e especificação das categorias para representar um determinado contexto, são elementos básicos na redução da complexidade.

Na revisão da literatura foram identificados procedimentos que auxiliam nesta redução, particularmente de forma aplicada às redes Bayesianas. Tais procedimentos são: discretização de variáveis contínuas, fusão de categorias (KARCHER, 2009; WITTEN et al., 2005) e agrupamento de variáveis (CONRADY ; JOUFFE , 2011). Tais procedimentos serão utilizados no capítulo 4.

Matsuura (2003 p.25) apresenta como a discretização de variáveis pode ser útil em Redes Bayesianas e levanta as vantagens deste procedimento:

- Facilidade de explicar características discretas, facilitando o processo de compreensão.
- Melhoria no poder preditivo dos modelos de classificação.
- Diminuição da necessidade computacional, pois redes bayesianas possuem grande complexidade no processamento, dado o número de probabilidades conjuntas calculadas e a quantidades de redes possíveis a partir de um grupo de variáveis.

No presente trabalho a discretização de variáveis é necessária por dois motivos (1) é um requisito do software selecionado que as variáveis estejam discretizadas e de preferência que o número de categorias seja o menor possível. (2) existem evidências empíricas que mostram que variáveis categorizadas apresentam mais vantagens que desvantagens em modelos de classificação binária logística (SICSU, 2010, p. 63), sendo que o mesmo foi observado em modelos de classificação usando redes bayesianas (KARCHER, 2009 p.55).

Espera-se também que a capacidade de interpretação do modelo seja favorecida com este procedimento.

Vale salientar que utilizando tais procedimentos redutores de complexidade, opta-se de forma pragmática pelo que deve e que não deve entrar para a estrutura da rede bayesiana. Ou seja, como apresenta Vassão (2010 p. 36) :

“Em computação, as Ontologias procuram apenas oferecer o esquema de classificação das entidades contempladas pelo sistema em questão, e nada além. Desse modo, elas assumem um papel de organizar nossa cognição de acordo com contextos específicos (...) manipular ontologias é manipular a realidade, pois alteram-se as vias pelas quais (...) a cognição e a criação podem ocorrer.”

Neste sentido, ao buscar pela redução da complexidade através de processos de abstração e criação de variáveis, circunscreve-se os modelos à realidade que está sendo observada, a qual depende de quem a observa e escolhe as ontologias.

3.7 Passos para a mineração de dados

Este trabalho visa descobrir padrões e estruturas nos dados selecionados. Para isso utiliza técnica de mineração de dados. Mais especificamente métodos relacionados ao “aprendizado de máquina”. As técnicas de aprendizado de máquina assimilaram testes estatísticos para avaliação de modelos e ampliaram a capacidade de generalizações. No mundo contemporâneo atributos se multiplicam em bases de dados, muitas vezes rarefeitas. Neste campo a mineração de dados ganha destaque.

Segundo Chen (2001) é importante compreender os elementos necessários para definir tarefas em mineração de dados. O autor aponta sete elementos, denominados “primitivas” que caracterizam esta atividade. A Figura 18 apresenta estes elementos de forma adaptada, os quais serão utilizados como direcionadores para o método de desenvolvimento no presente trabalho.

Figura 18 - Primitivas do Processo de Mineração de Dados

Fonte : Adaptado de CHEN, Z. **Data Mining and Uncertain Reasoning -An Integrated Approach**, p. 14.

Definição do Problema : em mineração de dados é extremamente importante que o problema esteja bem definido, pois as demais atividades dependem disso. Ao contrário do que pode

pensar o senso comum, mineração de dados não é uma panaceia automática de descoberta de conhecimento. Assim como nos modelos estatísticos tradicionais, depende de clara especificação do problema.

Dados Relevantes: a seleção de dados relevantes é o elemento central na descoberta de padrões e conhecimento. Atributos selecionados ou não pelo pesquisador acabam por direcionar os resultados. Um exemplo é a classificação de cervejas em grupos homogêneos, considerando suas características. Se a variável “teor de álcool” não é selecionada para compor um modelo, jamais serão reveladas as “cervejas sem álcool”. Este exemplo consta de um antigo manual do software SPSS do qual não se tem mais a referência, pois a leitura se deu em 1984 aproximadamente.

Tipos de “Conhecimento”: esta atividade especifica os desafios a serem atingidos com a mineração de dados, por exemplo: classificação, previsão, detecção de padrões, extração de regras e raciocínio sob incerteza.

Conhecimento Prévio (*Background*): é o conhecimento no qual o processo de mineração de dados está embasado, está associado com o domínio do problema a ser minerado e com a seleção de variáveis que deverão compor a modelagem. Assim, é importante o levantamento bibliográfico prévio visando ampliar o conhecimento sobre determinado assunto.

Algoritmos de Mineração de Dados: por conta da diversidade de algoritmos existentes é importante selecionar o(s) mais adequado(s) para lidar com o problema proposto, tipo de dado, conhecimento prévio existente, objetivo da mineração, resultados esperados e limitações dos métodos.

Modelos Desenvolvidos e Padrões Extraídos: este é o objetivo do processo de mineração de dados, sendo igualmente importante sua apresentação, visualização e caracterização para que leigos possam compreendê-lo.

Usuário / Pesquisador: os processos de mineração de dados podem auxiliar o raciocínio. Apesar de ser um processo automatizado ele é fortemente dependente da pessoa que está no comando do processo uma vez que é ela quem define o problema, as variáveis, as métricas e detém o conhecimento prévio sobre o assunto. Desta forma não há como dissociar o fator humano do processo de criação de um modelo. A subjetividade está presente mesmo que o processo seja abordado de forma objetiva. Pode-se mitigar a influência qualitativa da estruturação do processo, porém não há como eliminá-la. Para Keeney (2004) todas as análises quantitativas são fundamentadas em uma estruturação qualitativa.

Métricas de Interesse: as métricas de interesse servem como guia para mensurar o conhecimento descoberto. As métricas podem ser objetivas (estatísticas, indicadores) ou subjetivas (simplicidade, capacidade de revelar fatos novos, capacidade de replicação).

Para Vassão (2010) modelos surgem de ato criativo não podendo ser entendidos apenas como racionais. Com base nos estudos de (BAZERMAN, 2004) foram selecionados dois aspectos subjetivos para avaliação de modelos: capacidade de corroborar com heurísticas “positivas”, capacidade de mitigar heurísticas “negativas”. As heurísticas positivas representam a experiência prática e subjetiva que o especialista tem com a situação a ser modelada. Antes de partir para uso de processos que envolvam modelos matemáticos os especialistas certamente utilizam suas heurísticas, pois elas estão disponíveis e dependem exclusivamente de seus conhecimentos e experiências prévias. Um modelo pode, então, ser a representação do conhecimento e heurísticas dos especialistas de forma otimizada, quantificada e automatizada. Nesse sentido, os modelos tendem a corroborar com realidades conhecidas, sendo capazes de explicitar percepções de especialistas e outras obtidas por levantamento bibliográfico e pesquisa sobre determinado tema. Especialistas esperam que o comportamento das variáveis do modelo corroborem, de certa forma, com o que conhecem. Se um modelo foge muito da heurística conhecida os especialistas recomendam que as amostras de desenvolvimento sejam avaliadas, que as variáveis do estudo sejam checadas quanto ao seu conteúdo e o seu processamento. As heurísticas têm o potencial de ajudar no processo de construção e validação de modelos. As heurísticas negativas advêm da própria experiência do especialista ou grupo de pessoas envolvidos com o problema e estão baseadas na visão limitada sobre a realidade. Ou seja, muitas vezes conhecimentos prévios são imagens parciais de uma realidade. O uso de técnicas de análise de dados amplia a capacidade de mitigar ou eliminar heurísticas negativas, mensurando a influência de algumas variáveis de forma objetiva, mitigando vieses cognitivos. Esse fato confere aos modelos matemáticos a capacidade de aportar novas informações, combinar variáveis e rever modelos mentais criados.

Segundo Bazerman (2004) as pessoas são muito melhores para selecionar informação do que para integrar dados, principalmente quando envolve números. Ao integrar informações em busca de probabilidades conjuntas, por exemplo, envolvendo duas ou mais variáveis a mente é muito ineficiente.

3.8 Esquema metodológico para o desenvolvimento do trabalho

Respeitadas “as primitivas” para a mineração dos dados foi proposto um esquema metodológico para nortear o desenvolvimento do trabalho. Durante a execução do projeto de análise dos dados o esquema metodológico foi ajustado chegando à configuração final conforme apresentado a seguir.

O esquema proposto visou nortear o planejamento das etapas a serem executadas durante o processo de mineração e foi desenvolvido após pré-testes realizados com o software escolhido durante a revisão bibliográfica, denominado BayesiaLab. O esquema, apresentado na Figura 19 está dividido em três grandes etapas : (A) entendimento do objetivo e dos dados disponíveis, (B) criação e escolha do modelo, (C) análises finais e conclusões.

Figura 19 _Esquema para Mineração de Dados

FONTE : PREPARADO PELA AUTORA

4 APRESENTAÇÃO DA MODELAGEM E ANÁLISE DE RESULTADOS

Neste capítulo estão apresentados os detalhes da amostra e do processo de modelagem, bem como as análises realizadas e conclusões principais em cada etapa da modelagem.

4.1 Objetivos da Modelagem

Uma Rede Bayesiana diz respeito a um conjunto de informações que devem contribuir para mudar o conhecimento a respeito do estado de natureza de um determinado parâmetro. Neste trabalho, o parâmetro é a “facilidade” ou “dificuldade” de uma família levar a vida até o final do mês com sua renda” e será denominado simplesmente de “dificuldade”. Tanto o parâmetro quanto a informação sobre ele serão obtidos a partir das variáveis observadas na POF-2009. Assim, o primeiro passo é definir tal parâmetro “operacionalmente”.

Como já apresentado anteriormente será considerada “dificuldade” a agregação das seguintes classes de resposta: “alguma dificuldade”, “dificuldade” ou “muita dificuldade”, totalizando 75% da amostra. O percentual obtido é o mesmo utilizado nas publicações do IBGE (2010) tornando mais fácil o intercâmbio analítico entre os relatórios gerados. Vale salientar que poderia ter-se decidido por um modelo multinomial, no qual a variável resposta seria cada um dos “graus” de dificuldade ou os mesmos agrupados de outra forma que não a apresentada. No entanto, o objetivo foi conseguir um eixo de probabilidade que resuma o efeito das variáveis sobre a “dificuldade”. Espera-se ao final do modelo : probabilidades médias aderentes com cada uma das classes de “dificuldade” apresentadas na POF, da menor à maior.

No presente trabalho a probabilidade de uma família estar em “dificuldade” é estimada pela observação de quantas famílias foram consideradas com “dificuldade” em relação ao total de famílias observadas. Todas as famílias observadas receberam valor 0 ou 1, de acordo com as classes de resposta informadas, ou seja a “dificuldade” 0 = Não e 1 = Sim . Dessa forma, o modelo final de probabilidade apresenta um resultado entre 0 e 1 para cada uma dessas famílias, indicando a incerteza associada à percepção de “dificuldade”. Para mudar a percepção sobre esta probabilidade inicial são necessárias variáveis que ao serem instanciadas ou conhecidas mudam o valor desta probabilidade, ou seja, mudam o conhecimento sobre a situação das famílias estudadas. Assim, para a realização da análise, considerando cada variável

selecionada, deve-se medir a “distância” entre a probabilidade observada Θ (0.75) e a probabilidade atualizada, quando as variáveis são instanciadas ou conhecidas. Por exemplo: ter “a pele com cor branca” é uma informação que pode mudar o conhecimento sobre Θ ?

Pretende-se assim compreender o quanto as variáveis trazem de informação para a compreensão da “**dificuldade**” (Θ), e como o modelo estrutural (a rede) permite mudar o conhecimento sobre ela.

Ao final pretende-se obter:

- Uma rede de dados, passível de interação, que apresente os relacionamentos entre as variáveis entre si e com a variável “**dificuldade**”.
- Um eixo de probabilidade que resuma o efeito das variáveis para cada caso na amostra
- A análise do erro de classificação por percentil de probabilidade
- Uma classificação em grupos de “**dificuldade**”, com base no modelo de probabilidade e no erro observado.

4.2 Considerações Iniciais

De acordo com o IBGE (2010) a POF 2008_2009 pesquisou uma amostra com 56.091 unidades de consumo, destas 55.412 responderam também ao questionário denominado POF6, Questionário de Condições de Vida, focado apenas em percepções subjetivas. Esta foi a segunda vez que o IBGE realizou este tipo de avaliação. A anterior ocorreu juntamente com a POF 2002-2003.

De acordo com a POF (2010) é importante esclarecer sobre a adoção do termo “família” e “unidade de consumo” na pesquisa.

“A unidade de consumo compreende um único morador ou conjunto de moradores que compartilham da mesma fonte de alimentação, isto é, utilizam um mesmo estoque de alimentos e/ou realizam um conjunto de despesas alimentares comuns. Nos casos onde não existia estoque de alimentos nem despesas alimentares comuns, a identificação ocorreu através das despesas com moradia (...) família refere-se às pessoas ligadas por laços de parentesco, dependência doméstica ou normas de convivência, sem referência explícita ao consumo ou despesas”

No entanto, na maior parte das situações, a unidade de consumo da POF coincide com a “família” e os termos são utilizados de forma intercambiada no presente trabalho e nos relatórios da POF. Para se ter uma ideia de grandeza, de todas as unidades de consumo analisadas 5% tinham mais do que uma família na mesma unidade de consumo. Assim, o número de famílias na unidade de consumo foi computado como variável para o modelo, pois espera-se que a percepção de “dificuldade” orçamentária seja maior nas unidades de consumo onde residam mais do que uma família.

Os questionários foram aplicados ao longo dos 12 meses de realização da pesquisa, sendo que o de “Condições de Vida” foi coletado no último dia do levantamento em cada unidade de consumo. Ou seja, as respostas foram obtidas após o preenchimento dos demais formulários que investigaram os dados objetivos sobre a moradia, pessoas e o orçamento doméstico. Este fato provavelmente ajudou o entrevistado a responder com um pouco mais de objetividade à pergunta principal do questionário da POF6: “A renda total da sua família permite que você(s) leve(m) a vida até o final do mês com”...

() Muita Dificuldade , () Dificuldade, () Alguma Dificuldade, () Alguma Facilidade, () Muita Facilidade, () Facilidade.”

O questionário tem ao todo 10 perguntas inicia-se pela percepção sobre renda e orçamento, passando por alimentação, condições de moradia, percepção de violência e finalizando com aspectos da adimplência no pagamento de despesas básicas e prestações.

As respostas são individuais e um entrevistado foi indicado pela família como respondente mais apropriado. Em média, 36% dos respondentes não eram a “pessoa de referência” da unidade de consumo, no entanto este fato não impactou em diferença significativa na percepção de adequação orçamentária, conforme Tabela 5. Segundo a POF (2010) foi considerada como pessoa de referência :

“ aquela pessoa responsável por uma das seguintes despesas: aluguel, prestação do imóvel ou outras despesas de habitação (condomínio, imposto predial, serviços, taxa, etc.). No caso em que nenhum morador satisfizesse a pelo menos uma das condições acima, a pessoa de referência foi aquela assim considerada pelos moradores da unidade de consumo. Se mais de uma pessoa foi identificada pelos moradores, estabeleceu-se a idade mais alta como critério de escolha.”

Assim, a percepção apresentada pela pessoa selecionada para a resposta foi considerada como uma percepção familiar (ou da unidade de consumo), pois ela responde em nome dos demais membros. Na Tabela 5 pode-se observar a quantidade de famílias pesquisadas, a distribuição de respostas à pergunta indicada por “Responsável” e “Não Responsável” bem como o tamanho total da amostra utilizada.

Para o desenvolvimento do estudo foram utilizados 55.411 casos, ou seja, todos os que responderam à primeira pergunta do questionário de condições de vida (POF 6) que de forma ponderada representam mais de 56 milhões de unidades de consumo, ou 98% do total dos domicílios pesquisados.

A RENDA TOTAL DA SUA FAMÍLIA PERMITE QUE VOCÊ(S) LEVE(M) A VIDA ATÉ O FINAL DO MÊS COM :							
RESPOSTAS	RESPONSÁVEL		Total	RESPONSÁVEL		Total	TAMANHO DA AMOSTRA
	NÃO	SIM		NÃO	SIM		
MUITA_DIFICULDADE	3,503,955	6,673,380	10,177,335	17%	18%	18%	10,526
DIFICULDADE	4,537,176	7,619,191	12,156,367	22%	21%	21%	12,250
ALGUMA_DIFICULDADE	7,777,231	12,639,931	20,417,162	37%	34%	35%	19,507
ALGUMA_FACILIDADE	2,864,392	5,261,685	8,126,077	14%	14%	14%	7,607
FACILIDADE	1,885,622	3,536,634	5,422,256	9%	10%	9%	4,976
MUITA_FACILIDADE	188,858	401,839	590,697	1%	1%	1%	545
NAO_RESPONDEU A PERGUNTA	-	1,502	1,502	0%	0%	0%	1
RESPONDERAM POF6	20,757,234	36,132,660	56,889,894	100%	98%	98%	55,412
NÃO RESPONDEU POF 6		925,209	925,209	0%	2%	2%	679
TOTAL DE DOMICÍLIO	20,757,234	37,057,869	57,815,103	100%	100%	100%	56,091

Tabela 5 - Percepção de “dificuldade”

FONTE : POF 2008-2009

Vale salientar que a percepção de “dificuldade” das famílias brasileiras melhorou entre os períodos de 2003 e 2009 conforme apresenta a Tabela 6. Apesar dos avanços anunciados na mídia relativos ao consumo das famílias, 75% dessas em 2008-2009 percebem “dificuldade”. Neste sentido o modelo pretende investigar o que leva a esta percepção.

A RENDA TOTAL DA SUA FAMÍLIA PERMITE QUE VOCÊ(S) LEVE(M) A VIDA ATÉ O FINAL DO MÊS COM :

RESPOSTAS	OBSERVAÇÕES		
	(A) 2008_2009	(B) 2002_2003	(A) / (B)
MUITA_DIFICULDADE	18%	27%	0.66 ↓
DIFICULDADE	21%	24%	0.89 ↓
ALGUMA_DIFICULDADE	36%	35%	1.04 →
ALGUMA_FACILIDADE	14%	9%	1.57 ↑
FACILIDADE	10%	5%	2.00 ↑
MUITA_FACILIDADE	1%	1%	1.43 ↑
TOTAL	100%	100%	1.00
ALGUMA DIFICULDADE	75%	86%	0.88 ↓
ALGUMA FACILIDADE	25%	15%	1.71 ↑
TOTAL	100%	100%	1.00

Tabela 6 - Percepção de Dificuldade 2003_2009

Fonte: POF 2002-2003 e POF 2008-2009

4.3 Preparação e seleção de variáveis:

De forma descritiva os dados foram analisados visando compreendê-los. Assim como em outras técnicas analíticas a preparação dos dados é elemento fundamental no desenvolvimento de Redes Bayesianas. Esta etapa é importante não só para ampliar o conhecimento sobre cada uma das variáveis selecionadas como também para reduzir a complexidade da rede, a qual aumenta quanto maior o número de variáveis, atributos ou categorias.

Neste sentido foram utilizados três procedimentos para a preparação e seleção de variáveis: a discretização das variáveis e fusão de categorias e o agrupamento de variáveis.

4.3.1 Discretização de Variáveis e Fusão de Categorias.

As variáveis contínuas foram discretizadas e as qualitativas passaram por processos de agrupamento ou fusões de categorias.

Os processos de discretização e fusão de categorias foram realizados inicialmente usando procedimentos de pré-categorização disponíveis no software: (1) Árvore de decisão: através de indução supervisionada a variável é categorizada de forma que maximize a relação com

uma variável resposta (no caso a “dificuldade”); (2) Frequências iguais : neste caso cada categoria deve apresentar proporção similar, no caso deste estudo foi selecionado inicialmente 5% dos casos em cada categoria, ou seja as variáveis foram divididas em 20 faixas (3) Manual ou semi manual: após os dois procedimentos anteriores indicadores estatísticos foram utilizados para validar e realizar a fusão de categorias, quando necessário. Foram utilizados os seguintes índices informativos : análise *do Risco Relativo* (AGRESTI, 2002 p.43) e o WOE - *Weights of Evidence* (HAND-HENLEY, 1997). Estes mesmos indicadores foram utilizados em Karcher (2009 p.56-58) e Sicsu (2010, p. 72-73). No presente trabalho tomou-se a liberdade de substituir o termo Risco Relativo por Análise Bivariada, pois o nome Risco Relativo tem uma carga de significado mais próxima da Epidemiologia e das análises de Risco de Crédito.

Os indicadores calculados auxiliam não só no entendimento das variáveis em relação à variável “dificuldade” como também na determinação de um critério para a fusão de categorias.

Na sequência é apresentado o cálculo com os indicadores selecionados para uma das variáveis. Visando facilitar a compreensão, os cálculos podem ser feitos a partir da própria observação da Tabela 7, sem apresentação de fórmulas matemáticas. Na Tabela 7, apresenta-se uma variável qualitativa nominal denominada “Unidades da Federação”. O objetivo neste caso é o de diminuir o número de categorias distintas e para isso foi necessário considerar a relação dos atributos desta variável com a variável em estudo: “dificuldade”. Na Tabela é apresentado o cálculo da Bivariada e da Evidência (WOE).

Com base na análise das evidências (WOE), foi estabelecida a última coluna à direita da Tabela 7, nela são apresentados indicadores cuja interpretação visual e dos indicadores , norteou a fusão de categorias, conforme Tabela 8.

Observa-se pela análise Bivariada e WOE na Tabela 7 que o Maranhão possui a pior situação percebida pelas famílias. Santa Catarina possui o melhor indicador. No entanto, incorporar Estado por Estado na Rede Bayesiana aumenta em muito sua complexidade. Assim, considerando a Bivariada e o WOE, as Unidades da Federação foram reunidas em 6 grupos de acordo com o apresentado na Tabela 8. É importante informar que os softwares, em geral, permitem automaticamente criar categorias ou reagrupá-las, no entanto, optou-se por este processo manual visando ganhar maior atenção para cada atributo em particular e tomar algumas decisões que automaticamente não seriam possíveis, como por exemplo: criar um grupo apenas com São Paulo, Minas e Rio de Janeiro. Considerando as métricas automáticas, estes Estados seriam agrupados com Acre, Roraima, Goiás e Amazonas, o que não pareceu

interessante para a modelagem. De acordo com Sicsu (2010, p. 75) a fusão de atributos muitas vezes é realizada de forma “lógica” e em função da experiência de quem analisa o problema.

UNIDADES DA FEDERAÇÃO	%	PERC_DIFICULDADE		%	BIVARIADA	WOE =	IMPACTO NA PERCEPÇÃO
		NÃO	SIM				
MARANHAO	86%	2%	3%	3%	1.97	0.68	Bem Acima da Media
ALAGOAS	84%	1%	2%	2%	1.71	0.53	Bem Acima da Media
SERGIPE	84%	1%	1%	1%	1.70	0.53	Bem Acima da Media
PIAUI	83%	1%	2%	2%	1.66	0.51	Bem Acima da Media
BAHIA	82%	5%	8%	7%	1.47	0.38	Acima da Media
PARAIBA	82%	1%	2%	2%	1.46	0.38	Acima da Media
CEARA	81%	3%	4%	4%	1.44	0.37	Acima da Media
PARA	81%	2%	3%	3%	1.43	0.36	Acima da Media
RIO_GRANDE_DO_NORTE	81%	1%	2%	2%	1.42	0.35	Acima da Media
PERNAMBUCO	81%	4%	5%	5%	1.40	0.34	Acima da Media
TOCANTINS	81%	0%	1%	1%	1.39	0.33	Acima da Media
ACRE	76%	0%	0%	0%	1.06	0.06	Media
RORAIMA	76%	0%	0%	0%	1.03	0.03	Media
GOIAS	75%	3%	3%	3%	1.01	0.01	Media
AMAZONAS	74%	2%	1%	1%	0.94	0.06	Media
MINAS_GERAIS	74%	11%	10%	11%	0.94	0.06	Media
RIO_DE_JANEIRO	74%	10%	9%	9%	0.93	0.08	Media
SAO_PAULO	74%	24%	22%	22%	0.92	0.08	Media
MATO_GROSSO_DO_SUL	73%	1%	1%	1%	0.89	0.11	Abaixoda Media
ESPIRITO_SANTO	72%	2%	2%	2%	0.86	0.15	Abaixoda Media
RONDONIA	71%	1%	1%	1%	0.81	0.21	Abaixoda Media
PARANA	71%	7%	5%	6%	0.81	0.21	Abaixoda Media
AMAPA	70%	0%	0%	0%	0.78	0.25	Abaixoda Media
DISTRITO_FEDERAL	69%	2%	1%	1%	0.73	0.31	Bem Abaixo da Media
MATO_GROSSO	68%	2%	2%	2%	0.70	0.35	Bem Abaixo da Media
RIO_GRANDE_DO_SUL	67%	8%	6%	6%	0.66	0.41	Bem Abaixo da Media
SANTA_CATARINA	65%	5%	3%	3%	0.60	0.51	Bem Abaixo da Media
TOTAL	75%	100%	100%	100%	1.00		Media

Tabela 7 - Fusão de Categorias - Análise Bivariada

FONTE : PREPARADA PELA AUTORA COM BASE NOS DADOS DA POF 2009.

FUSÃO DE CATEGORIAS UNIDADES DA FEDERAÇÃO	%	PERC_DIFICULDADE		%	BIVARIADA	WOE =	IMPACTO NA PERCEPÇÃO
		NÃO	SIM				
MA_PI_AL_SE	84%	4%	8%	7%	1.79	0.58	Bem Acima da Media
BA_PE_CE_PA_PB_RN_TO	81%	17%	25%	23%	1.44	0.36	Acima da Media
GO_AM_AC_RR	75%	5%	5%	5%	1.00	0.00	Media
SP_MG_RJ	74%	44%	41%	42%	0.93	0.07	Media
PR_ES_MS_RO_AP	71%	11%	9%	10%	0.83	0.19	Abaixoda Media
RS_SC_MT_DF	67%	17%	11%	13%	0.66	0.42	Bem Abaixo da Media
TOTAL	75%	100%	100%	100%	1.00		Media

Tabela 8 - Fusão de Categorias - resultado da categorização

FONTE : PREPARADO PELA AUTORA COM BASE NOS DADOS DA POF 2009.

A estatística Bivariada e o WOE, neste trabalho, ajudaram a objetivar a lógica e conhecimento prévio. Por exemplo: a interpretação da Tabela 7, construída a partir de variável que re-

sume percepções subjetivas das famílias, corrobora com alguns indicadores apresentados por NERI (2011, p. 282-303) e com o IDH 2000 (PNUD, 2012). O Gráfico 9 apresenta a relação entre o percentual de famílias com dificuldades e o IDH - Índice de Desenvolvimento Humano, para cada uma das Unidades da Federação, calculado para o ano de 2000. O IDH visa oferecer um contraponto ao Produto Interno Bruto (PIB) per capita, que considera apenas dimensão econômica do desenvolvimento e pretende ser uma medida que sintetiza o desenvolvimento humano passível de comparação entre países e regiões. Leva em conta também o PIB (utilizado para mensurar a renda) e mais dois componentes: a longevidade (expectativa de vida ao nascer) e a educação (índice de analfabetismo e taxa de matrícula em todos os níveis de ensino).

Gráfico 9 - IDH x Dificuldade Percebida

FONTE : DESENVOLVIDO PELA AUTORA -POF (2009) E PNUD (2000)

Outras variáveis, contínuas, tiveram tratamento similar ao apresentado para as Unidades da Federação. Por exemplo: para as variáveis com origem nos valores alocados pelos domicílios no orçamento familiar foram feitos três tratamentos: (1) agrupamento por tipos de despesas (2) qualificação de tais despesas como básicas ou diferenciadas (3) categorização de cada subgrupo identificando faixas percentuais apontando o impacto da despesa no orçamento familiar.

Primeiramente foram formados os grandes grupos de despesa: Alimentação, Habitação, Vestuário, Transporte, Moradia e Outros conforme item 2.2. Na sequência parte destas despesas

foram marcadas como diferenciadas, conforme Tabela 9. Despesas diferenciadas, neste trabalho, são aquelas que impactam mais no orçamento das famílias que têm menor “dificuldade” orçamentária. No geral, são despesas “não básicas”, como por exemplo : “empregados mensalistas”. Despesas como “alimentação em casa”, por exemplo, impactam mais no orçamento dos que vivem com maior dificuldade, não sendo consideradas diferenciadas.

A Alimentação (em casa) originou-se do agrupamento de muitas outras categorias, pois a POF possui esta riqueza. Na sequência da Tabela 9 está o item “alimentação fora de casa”. Este último surpreendeu uma vez que se esperava impacto maior dessa categoria para os que não têm “dificuldade”. A explicação para este fato vem do próprio crescimento da alimentação fora de casa conforme apresentado pelo IBGE (2010): a participação alimentação fora do domicílio nos gastos com alimentação urbana subiu de aproximadamente 26% para 33%, e a rural subiu de 13% para 17,5%. O fato é que alimentação fora de casa aumentou seu impacto inclusive para os que têm maior “dificuldade” orçamentária não sendo então, pelo critério adotado, classificada como despesa diferenciada. Para identificar itens diferenciadores dentro do grupo Alimentação seria necessário adentrar em menores níveis de despesa. No entanto, a POF não tem como objetivo analisar família a família em relação aos itens consumidos, ou seja, nem tudo é pesquisado para todos. Por exemplo: se o entrevistador foi numa casa em determinada semana e esta não comprou pizza no período da coleta dos dados, então este item não está associado à cesta de compras dessa família, mesmo que seja consumidora habitual do produto. Assim, não é possível analisar cestas de compras família por família na POF, mas apenas em agregados, preferencialmente por classes de rendimento. Num primeiro momento tentou-se trabalhar com os dados em categorias menores, no entanto foi necessário diminuir o detalhamento para viabilizar o estudo. Para mais detalhes sobre as ponderações da POF ver a publicação completa da POF (2010).

A Tabela 10 apresenta um exemplo de como foi realizado o agrupamento de categorias de despesa. Tomando-se como base o % de gasto em Reais sobre o montante de despesa realizado pelas famílias com ou sem “dificuldade”. Foram estabelecidos agrupamentos com base na Bivariada e no WOE, e identificadas as despesas “diferenciadas”. Por exemplo : o item “aluguel monetário” está contido dentro do grupo Habitação (não diferenciada) e “mensalistas” estão contidas dentro do grupo “Habitação Diferenciada”.

O processo de categorização apresentado na Tabela 9 foi realizado para todos os demais grupos de despesa: Vestuário, Moradia, Transporte, Habitação e Outras Despesas. Resultados

desta classificação podem ser consultados no Apêndice 2 onde os grupos de despesas “diferenciadas” aparecem acompanhadas com o sufixo “_DIF”.

DESCRIPTIVO VARIÁVEIS	DIFICULDADE % COL			RR (A) / (B)	WOE	CARACTERIZAÇÃO
	(A) SIM	(B) NÃO	% TOTAL			
Alimentação em casa	13.31%	8.06%	11.1%	1.65	0.50	ALIMENTACAO
Alimentação fora	5.11%	4.79%	5.0%	1.07	0.06	ALIMENTACAO
Telefone Fixo	1.13%	0.91%	1.0%	1.24	0.21	HABITACAO
Telefone Celular	1.01%	0.98%	1.0%	1.03	0.03	HABITACAO
Prestação do Imóvel	0.64%	0.64%	0.6%	1.00	0.00	HABITACAO
Mobiliários simples para o lar	0.53%	0.47%	0.51%	1.12	0.12	HABITACAO
ML Lenha e Carvão	0.10%	0.03%	0.1%	3.73	1.32	HABITACAO
Gás Doméstico	1.01%	0.47%	0.8%	2.14	0.76	HABITACAO
Energia Elétrica	2.63%	1.84%	2.3%	1.43	0.36	HABITACAO
Eletrodomésticos	2.23%	1.89%	2.1%	1.18	0.16	HABITACAO
Consertos de artigos lar	0.27%	0.19%	0.2%	1.39	0.33	HABITACAO
Artigo de Limpeza	0.68%	0.44%	0.6%	1.55	0.44	HABITACAO
Aluguel Monetário	1.98%	1.58%	1.8%	1.25	0.23	HABITACAO
Aluguel Não Monetário	11.71%	9.79%	10.9%	1.20	0.18	HABITACAO
Água e Esgoto	1.08%	0.60%	0.9%	1.79	0.58	HABITACAO
Reforma Imóvel	1.74%	1.92%	1.8%	0.90	0.10	HABITACAO_DIFERENCIADO
Pacote TV Internet fone	0.52%	0.66%	0.6%	0.78	0.24	HABITACAO_DIFERENCIADO
Outras Taxas Imóveis	0.06%	0.06%	0.1%	0.93	0.08	HABITACAO_DIFERENCIADO
Outras Taxas	0.30%	0.50%	0.4%	0.60	0.51	HABITACAO_DIFERENCIADO
Móveis para o lar diferenciadas	1.19%	1.37%	1.3%	0.87	0.14	HABITACAO_DIFERENCIADO
Mensalistas sem INSS	0.37%	0.55%	0.4%	0.67	0.40	HABITACAO_DIFERENCIADO
Mensalistas com INSS	0.41%	1.12%	0.7%	0.37	1.00	HABITACAO_DIFERENCIADO
Manutenção do lar (ML)	2.25%	2.31%	2.3%	0.97	0.03	HABITACAO_DIFERENCIADO
Diarista sem INSS	0.33%	0.65%	0.5%	0.51	0.67	HABITACAO_DIFERENCIADO
Diarista com INSS	0.04%	0.07%	0.1%	0.57	0.55	HABITACAO_DIFERENCIADO
Condomínio	0.57%	1.24%	0.8%	0.46	0.78	HABITACAO_DIFERENCIADO
Aquisição Imóvel	1.52%	2.49%	1.9%	0.61	0.49	HABITACAO_DIFERENCIADO
Aquisição Imóvel	1.44%	2.84%	2.0%	0.51	0.68	HABITACAO_DIFERENCIADO

Tabela 9 – Classificação das despesas diferenciadas - percentual sobre montante em Reais

FONTE : PREPARADO PELA AUTORA COM BASE NA POF 2009

Após este procedimento os subgrupos de despesa identificados foram tratados individualmente, conforme Tabela 10. Pode-se perceber que famílias que têm maior impacto das despesas de habitação em seus orçamentos, por exemplo, são as que apresentam também maior proporção de “dificuldade”.

Exemplificando a leitura desta Tabela observa-se : dos que gastam mais de 30% do orçamento com habitação, 83% vivem com “dificuldade”, sendo a média igual a 75%.

EVIDÊNCIAS	%	DIFICULDADE		Total	BIVARIADA	EVIDÊNCIA WOE	DIFICULDADE
PERC_HABITACAO	DIFICULDADE	NÃO	SIM		SIM / NÃO		RELAÇÃO COM MEDIA
<=0.05	65%	14%	8%	10%	0.62	-	0.48
<=0.10	70%	29%	23%	24%	0.77	-	0.26
<=0.15	76%	23%	24%	24%	1.02		0.02
<=0.20	78%	14%	17%	16%	1.17		0.16
<=0.30	81%	12%	17%	16%	1.38		0.32
>0.30	83%	7%	12%	11%	1.56		0.45
	75%	100%	100%	100%	1.00		

Tabela 10 - Categorização despesas com habitação

Fonte : Preparada pela autora com base na POF 2009

Algumas variáveis foram convertidas em *dummies* ou binárias, cujos atributos 0 (zero) ou 1 (um) foram convertidos em categorias “não” e “sim”. Um exemplo foram as variáveis relacionadas à renda, cada tipo de renda transformou-se numa variável binária, indicando assim todas as origens de renda disponíveis nas famílias.

ORIGEM DAS RENDAS \$ REAIS	DIFICULDADE (A)	FACILIDADE (B)	TOTAL	(A) / (B)	WOE
Emprego	47.31%	43.72%	45.70%	1.08	0.08
Empregador	4.07%	9.32%	6.42%	0.44	0.83
Conta Própria	15.16%	12.76%	14.09%	1.19	0.17
Aposentadoria	16.40%	15.90%	16.18%	1.03	0.03
Prev Privada	0.42%	1.11%	0.73%	0.38	0.98
Benefícios 13o e férias	4.77%	4.98%	4.86%	0.96	0.04
Pensão, Mesada, Doação	1.89%	1.47%	1.70%	1.28	0.25
Transferências Governamentais	2.31%	0.57%	1.53%	4.02	1.39
Seguros - Inclusive Obrigatório	0.08%	0.07%	0.07%	1.11	0.11
Renda de Aluguel	1.41%	2.61%	1.95%	0.54	0.62
Outras Rendas	1.64%	1.90%	1.76%	0.86	0.15
Variações Patrimoniais	4.55%	5.59%	5.01%	0.81	0.21
RENDIA TOTAL	100%	100%	100%		

Tabela 11 - Origem das Rendas Familiares – percentual sobre o total da renda monetária

FONTE : Preparado pela autora a partir dos dados da POF (2009)

Todas as variáveis originadas de despesas, renda, tipos de família, relação com o trabalho, escolaridade dentre outras foram classificadas, discretizadas, tiveram redução de categorias, quando necessário, seguindo a metodologia aqui apresentada.

Metodologias específicas para discretização de variáveis visando especificamente redes Bayesianas podem ser obtidas em Matsuura (2003). Tais métodos não foram adotados neste trabalho por conveniência e tempo, pois não estão disponíveis no software utilizado e os procedimentos adotados incluindo a Bivariada e o WOE foram validados no trabalho de outros

autores como Karcher (2009 p.56-58) e considerados suficientes neste trabalho. No entanto, este pode ser um tema de futuros estudos visando aprimorar o processo de simplificação, necessário à criação de Redes Bayesianas discretas.

4.3.2 Agrupamento e Eliminação de Variáveis

Em análise estatística é comum a utilização da análise fatorial para o tratamento e redução de variáveis que apresentam correlação entre si, pois a medida que o número de variáveis aumenta, aumenta também a complexidade e a necessidade de conhecer melhor as inter-relações entre elas (HAIR et al., 2005, p. 89). Através da criação de fatores comuns preserva-se grande parte da variabilidade original dos dados e diminui-se a complexidade do conjunto. Esta técnica estatística deve ser utilizada a partir de variáveis quantitativas em escala razão e pode ter caráter confirmatório ou exploratório. Assim como na análise fatorial, o presente estudo necessita reduzir o conjunto de dados a uma menor dimensão, no entanto faz uso de técnica que possui similaridade quanto ao objetivo exploratório da análise fatorial, denominado aqui como Agrupamento de Variáveis, tendo a peculiaridade de tratar variáveis categóricas, utilizando as redes bayesianas como técnica analítica.

O objetivo da aplicação da técnica de agrupamento de variáveis nesta tese é (1) reduzir o escopo de variáveis na análise (2) conhecer melhor o conjunto de dados e a inter-relação entre as variáveis. (3) identificar redundâncias ou variáveis das quais se pode abrir mão inicialmente.

No entanto, o primeiro passo, antes de proceder com o agrupamento de variáveis, foi preparar uma visão genérica da rede permitindo examinar as principais relações entre os nós. Para isso foi utilizado inicialmente o algoritmo *Máximo Spanning Tree*, e na sequência o EQ, conforme será apresentado no item 4.4. Estes métodos permitiram uma visão geral do relacionamento das 98 variáveis selecionadas, bem como a mensuração da relação entre elas, favorecendo o surgimento de grupos conceituais de variáveis. Os agrupamentos “conceituais” foram obtidos através do cálculo da entropia relativa entre as variáveis, conforme será apresentado mais adiante. Foram formados macro grupos, facilitando a compreensão do conjunto de dados. Este procedimento será apresentado durante o desenvolvimento deste trabalho.

4.4 Primeira Visão da Floresta.

A primeira rede “descoberta automaticamente” possui a configuração apresentada na Figura 20 e será visualmente nas próximas páginas, dado que possui grande quantidade de variáveis não foi possível sua completa visualização no papel.

Esta representação permite uma visão “sobre a floresta”. Nas páginas seguintes à Figura 20 serão apresentadas “partes” da rede, de modo que se possa visualizar a relação inicial encontrada entre as variáveis e assim validar mentalmente se existe sentido ou lógica nos relacionamentos encontrados. Vale salientar que neste momento não existe preocupação em apontar ou restringir relações entre os arcos. Espera-se simplesmente conhecer o conjunto dos dados e as relações primárias identificadas entre as variáveis. Para isso houve um relaxamento sobre a hipótese de independência entre as variáveis.

O Apêndice 1 (A) apresenta a lista de variáveis selecionadas por ordem alfabética. O Apêndice 1(B) apresenta a lista de variáveis em ordem numérica. Nestes Apêndices estão apresentadas as variáveis inicialmente construídas, a indicação dos agrupamentos posteriormente realizados, marcados com um “X”, e o significado de cada uma, conforme exemplo abaixo.

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES X		SIGNIFICADO
V	71	AREA_GEO	AREA_GEO		Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água.
V	16	NECESS_MEDICAMENTO	FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	X	Indica se a família necessitou de medicamentos nos últimos meses, mas não foi atendida neste sentido.
V	16	NECESS_SERVICOS_SAUDE	FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	X	Indica se a família necessitou de serviços de saúde nos últimos meses, mas não foi atendida neste sentido.

O Apêndice 2 apresenta apenas o conteúdo das variáveis selecionadas para a modelagem final.

Figura 20 - Máximo Spanning Tree - Todas as Variáveis

FONTE : DADOS DA POF 2009

Figura 21 - Máximo Spanning Tree : Tipos de Renda _Previdência _Tempo Sem Trabalho

FONTE : DADOS DA POF 2009

Figura 22 - Máximo Spanning Tree : Alimento, Vestuário, Moradia, Outras Despesas
 FONTE : DADOS DA POF 2009

Figura 23 - Máximo Spanning Tree - Escolaridade & Estrutura das Famílias

FONTE : DADOS DA POF 2009

Figura 24 - Máximo Spanning Tree – Renda, Classe Economica , Moradia
FONTE : DADOS DA POF 2009

Figura 25 - Máximo Spanning Tree - Renda, Adimplência, Moradia, Saúde
FONTE : DADOS DA POF 2009

Figura 26 - Máximo Spanning Tree - Moradia e Acesso a Serviços Públicos
FONTE : DADOS DA POF 2009

Figura 27 - Máximo Spanning Tree - Acesso a Serviços Públicos Básicos

FONTE : DADOS DA POF 2009

Na sequência se apresenta como os arcos foram configurados, utilizando os conceitos de entropia, conforme item 3.3 e disponíveis no software BayesiaLab. Como dito anteriormente as forças dos arcos nas redes bayesianas são determinados pela informação mútua e pela entropia relativa, as quais estão exemplificadas na Figura 28. Observa-se que a entropia relativa entre o “Tipo de Alimento” e “Condições de Moradia” é de 6,4% (azul), o que determinou o sentido do arco, que no sentido contrário é 5,18% (vermelho). Ou seja, as flechas apontam geralmente no sentido da maior entropia relativa. Assim, não se pode esperar que automaticamente seja possível estabelecer relações de causa e efeito, mas apenas que existe uma relação entre estas variáveis e que ao se conhecer uma, algo se pode saber sobre a outra. Na Figura 28, acima das métricas de entropia relativa, apresenta-se a informação mútua (em preto) que mede, de forma simétrica, o quanto (em média) o conhecimento ou observação de uma variável (tipo de alimento) reduz a incerteza sobre outra (condições de moradia).

Figura 28 - Força dos Arcos

FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009

Por questão de espaço e limitações do papel não é possível apresentar estas métricas para todos os arcos, até porque eles irão se alterar conforme evolui análise. Por exemplo: não se pode esperar que “condições de moradia” apresente dependência apenas da variável “Tipo de Alimento”.

Após a avaliação inicial do conjunto de variáveis, utilizou-se o algoritmo EQ para aprendizagem complementar de estrutura. Mantendo a estrutura anterior (Figura 20), obteve-se novas relações entre as variáveis. Este procedimento foi utilizado para aperfeiçoar o processo de agrupamento de variáveis. A Figura 29 não tem objetivo de leitura, apenas é uma ilustração que mostra como as relações entre as variáveis são muito mais complexas (com mais arcos) do que havia sido proposto pelo algoritmo *Maximo Spanning Tree*, cujo objetivo foi fornecer a visão inicial sobre a “floresta”.

Figura 29 - Algoritmo EQ aplicado sobre a rede Máximo Spanning Tree

FONTE : FONTE : DADOS DA POF 2009

Para cada um dos arcos existentes foram estabelecidas medidas de entropia relativa baseadas na métrica de divergência de Kullback-Leibler. Com base nesta métrica foram formados os agrupamentos de variáveis conforme pode ser observado pela Figura 30. Ao todo foram identificados 29 grupos de variáveis (do Fator_0 a Fator_28). Cada um destes agrupamentos são variáveis categóricas contendo, na presente análise, entre 2 a 7 categorias. Vale esclarecer que algumas variáveis não sofreram nenhum agrupamento. No entanto, as que foram

agrupadas podem ser interpretadas observando apenas o nome das variáveis e o seu grupo respectivo, denominado Fator.

Fatores	Nodes
[Fator_0]	CONDICOES_DE_MORADIA
	TIPO_ALIMENTO_RAZAO_REC
	QUANTIDADE_DE_ALIMENTO_REC
	CASA_ESCURA
[Fator_1]	VIOLENCIA_OU_VANDALISMO
	POLUICAO_OU_PROBLEMAS_AMBIENTAIS
	RUA_OU_VIZINHOS_BARULHENTOS
[Fator_2]	UF_GRUPO
	UF
	RACA_COR_PESSOA_REFERENCIA
	REGIAO
[Fator_3]	TIPO_AREA
	URBANO
	PROVENIENCIA_AGUA
	SERVICO_DE_AGUA
[Fator_4]	X_RENDA_CONTA_PROPRIA
	RENDA_PREDOMINANTE
	X_RENDA_EMPREGADO
	X_RENDA_EMPREGADOR
[Fator_5]	COMODOS_DORMITORIOS
	POUCO_ESPACO
	QTDE_BANHEIROS
	QTDE_COMODOS
[Fator_6]	SERVICO_DE_FORNECIMENTO_DE_ENERGIA_ELETRICA
	SERVICO_DE_COLETA_DE_LIXO
	SERVICO_DE_ILUMINACAO_DE_RUA
[Fator_7]	ATRASO_DE_ALUGUEL_OU_PRESTACAO_DA_RESIDENCIA
	CONDICAO_OCUPACAO_REC
	PERC_ALUGUEL_ESTIMADO
	PERC_HABITACAO
[Fator_8]	FAMILIAS_E_IDOSOS
	X_RENDA_BENEFIC_ASSALARIADO
	X_RENDA_APOSENTADORIA
[Fator_9]	PERC_TARNSPORTE_DIF
	ESCOLARIDADE_PESS_REFERENCIA
	PONTOS_ABEP
	CLASSE_ABEP
[Fator_10]	QTDE_FILHOS_REC
	ESCOLARIDADE_MENORES
	NR_DEPENDENTES_REC
	QTDE_MORADORES
	NINHOS_POR_FAIXA_ETARIA
[Fator_11]	SERVICO_DE_EDUCACAO
	SERVICO_DE_SAUDE
	SERVICO_DE_LAZER_E_ESPORTE
[Fator_12]	MADEIRAS_DAS_JANELAS_PORTAS_OU ASSOALHOS_DETERIO
	FUNDACAO_PAREDES_OU_CHAO_UMIDOS
	TELHADO_COM_GOTEIRA
[Fator_13]	ESCOADOURO_SANITARIO
	EXISTE_AGUA_CANALIZADA
	SERVICO_DE ESGOTAMENTO_SANITARIO
[Fator_14]	STATUS_REDE_ENSINO
	PESSOAS_ESTUDANDO_sum
	CONJ_OU_RESP_ESTUDANDO_sum
[Fator_15]	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA
	INUNDACOES
[Fator_16]	PERC_OUTRAS_DIF
	PERC_DESP_INDIV
	PERC_OUTRAS
[Fator_17]	TITULAR_DA_CONTA_CORRENTE
	TEM_CARTAO_DE_CREDITO

Figura 30 Agrupamento de Variáveis – continua (...)

FONTE : DADOS POF 2009.

Fatores	Nodes
[Fator_18]	PERC_RENDA_NAO_MONET
	COMPROM_RENDA
	RENDA_TOT_ANO
	RENDA_MONET_PER_CAPTA
[Fator_19]	PERC_HABITACAO_DIF
	PERC_A_CREDITO
[Fator_20]	SEXO
	TEM_CONJUGE_sum
	STATUS_CONJUGE
[Fator_21]	RENDA_MONET_NECESS
	PERC_VAL ALIM_MIN
[Fator_22]	MATERIAL_PREDOMINA_PAREDES_EXTERNAS
	MATERIAL_PREDOMINA_PISO
[Fator_23]	EXISTENCIA_PAVIMENTACAO_RUA
	SERVICO_DE_LIMPEZA_E_MANUTENCAO_DE_RUA
[Fator_24]	TIPO_DOMICILIO
	MATERIAL_PREDOMINA_COBERTURA
[Fator_25]	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIR
	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS
[Fator_26]	PREVIDENCIA
	TEMPO_SEM_TRAB_PRINC_REC
	RENDA_DAS_PESSOAS_UC
[Fator_27]	NECESS_SERVICOS_SAUDE_max
	NECESS_MEDICAMENTO
[Fator_28]	TIPO_CONTRATO_ALUGUEL
	TEMPO_ALUGUEL

Figura 30 - Agrupamento de Variáveis - continuação

FONTE : DADOS POF 2009

Os grupos de variáveis apresentam-se como pequenas redes bayesianas do tipo “naive”, onde o fator aponta para um conjunto de variáveis, representando “o potencial” de resumir o conjunto dos dados, como se fosse uma variável latente.

Figura 31- Representação do Fator 3 com Variável Latente

FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009.

Algumas métricas são utilizadas para a validação de cada um dos grupos de variáveis. Tomando como exemplo o Fator 3 estas métricas são apresentadas na Tabela 12.

Análise Fator 3

NÓ PAI	NÓ FILHO	Kullback-Leibler-D (entropia relativa)	Peso Relativo	Contribuição Global	GKL-test (G*)	Graus de Liberdade	p-valor
[Factor_3]	URBANO	0.627	1.000	0.390	48,155	2.000	0.00
	TIPO_AREA	0.627	1.000	0.390	48,146	6.000	0.00
	PROVENIENCIA_AGUA	0.213	0.340	0.133	16,392	4.000	0.00
	SERVICO_DE_AGUA	0.139	0.222	0.087	10,664	2.000	0.00

Tabela 12 - Validação do Fator 3

FONTE : DADOS POGF 2009

A métrica Kullback-Leibler refere-se à entropia cruzada entre os nós filhos e o pai (variável latente). Esta métrica é calculada para cada nó individualmente em relação ao fator. Como apresentado no item 3.3.1, em redes bayesianas esta métrica é utilizada também para “medir o ganho de informação” quando se move da distribuição priori para a posteriori.

PRIORI PARA O FATOR 3

CONHECIMENTO DE QUE O SERVIÇO DE AGUA NÃO EXISTE OU É RUIM

POSTERIORI FATOR 3 : DEPOIS DE SABER SOBRE O SERVIÇO DE AGUA

Figura 32 - Priori e Posteriori - Fator 3 & Serviço de Agua

FONTE : DADOS POF 2003

A entropia cruzada entre o FATOR 3 e a variável PROVENIÊNCIA_AGUA, a partir da fórmula de Kullback-Leibler é igual a 0.213, representando um "Peso Relativo" de 0.34 no fator. A métrica de "Contribuição Global" está subestimada para esta variável e para SERVIÇO_DE_AGUA, pois na verdade as duas primeiras variáveis (URBANO e TIPO_ÁREA) referem-se à mesma informação, como será apresentado a seguir. O teste G_{KL}

assemelha-se em funcionalidade ao teste Qui_quadrado, recebe também o nome de log-verossimilhança ou teste de razão de verossimilhança (MCDONALD, 2008, p.42).

Assim como no Qui_quadrado a interpretação do resultado é influenciada pelo tamanho da amostra (WONNACOTT & WONNACOTT, p.552).

Neste trabalho a amostra possui mais de 55.000 casos. Neste caso, o maior valor do teste está em informar que se tem em mãos uma quantidade suficiente de dados para discernir sobre as prováveis hipóteses (WONNACOTT & WONNACOTT, p.552).

A fórmula utilizada para o G_{KL}^* , deriva-se do teste G, apresenta a seguinte configuração e pode ser calculada utilizando a métrica de Kullback-Leibler, onde N é o número de casos na amostras, $I(X,Y)$ a informação mútua e $DKL(P \parallel Q)$ é a divergência de Kullback-Leibler.

$$G = 2 \cdot N \cdot \ln(2) \cdot I(X, Y)$$

$$G^* = 2 \cdot N \cdot \ln(2) \cdot D_{KL}(P \parallel Q)$$

Figura 33 - Fórmula do G Teste e relação do mesmo com Informação Mútua e Entropia Cruzada

FONTE : Material interno “Bayesia” cedido para uso acadêmico por JOUFFE, Lionel, abril/2012

O teste G passou a ser mais utilizado em pesquisas após a publicação do livro de Sokal & Rohlf (1981) com aplicações em biologia. A relação entre o G_{KL} , a entropia e a informação mútua também pode ser consultada na WIKIPEDIA (2011), como material complementar, dado que existem poucas referências a este teste.

Voltando ao FATOR 3 pode-se perceber que URBANO e TIPO_ÁREA representam a mesma informação, pois “urbano” está contido na variável TIPO_ÁREA, que assume também outras categorias para representar a localidade (como região metropolitana, capital, interior). Não faz sentido manter ambas no modelo. Entretanto, existem informações novas: a proveniência e o serviço de água, associados à região onde a família mora : “urbana” ou “rural”.

O FATOR 3, de forma automática, dividiu o Brasil em “Urbano”, “Rural com água” e “Rural com dificuldade de acesso à água”, conforme apresenta a Figura 34. Este fato pode ser ilustrado com notícias que mostram que enquanto a renda no NE aumentou 42% entre 2001 e

2009, o número da rede de saneamento aumentou apenas 6,9%. Imagens surpreendentes de casas com bens de consumo, antenas parabólicas, mas miséria em termos de saneamento e acesso á água foi retratada em reportagem da Folha de São Paulo (GUIBU, 2012).

Figura 34 -Fator 3 -Urbano, Rural, Acesso à água

Fonte : Preparada pela autora com base na POF 2009

Como existe o “pré-requisito” de diminuição de complexidade para facilitar o processamento da Rede Bayesiana, percebeu-se que o software opta por perder informação para diminuir complexidade, a decisão de continuar a análise com os agrupamentos torna-se decisão do pesquisador.

Na sequência ilustra-se como neste trabalho mesclaram-se as descobertas automáticas de redução de dados com a necessidade de preservar algumas informações o que implicou em processamentos adicionais.

Optou-se por incorporar o novo conhecimento (rural com e sem água), mas preservar as informações que existiam na variável “Tipo_Area”. Assim criou-se manualmente um novo agrupamento, uma nova variável derivada, denominada Área_Geo. Esta variável foi numerada como 71, constando na lista das evidências estudadas, conforme apresentado na Tabela 13.

71	EVIDÊNCIA	Total	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	AREA_GEO		DIFIC	NÃO	SIM	Total			
	CAPITAL	13,452,333	70%	29%	22%	24%	0.76	-	0.28
	INTERIOR	27,175,057	75%	48%	48%	48%	0.99	-	0.01
	REGIÕES METROPOLITANAS	7,333,112	79%	11%	14%	13%	1.28	-	0.24
	RURAL_COM_AGUA	4,940,343	78%	8%	9%	9%	1.20	-	0.18
	RURAL_SEM_AGUA	3,989,049	84%	5%	8%	7%	1.68	-	0.52
	Total	56,889,894	75%	100%	100%	100%	1.00	-	-

Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água

Tabela 13 - Variável Área Geo

Fonte: POF 2009

Os grupos de variáveis calculados utilizam uma relação de rede do tipo Naive e por isso cuidados devem ser tomados neste sentido, pois existe o pressuposto de que as variáveis que compõem o fator são independentes. Assim, neste trabalho, os grupos de variáveis foram utilizados mais como suporte para a compreensão dos dados e reflexão sobre o motivo do agrupamento, para depois sugerir algo diferente do proposto no procedimento automático. Optou-se em muitos casos por seguir com as variáveis originais, pois se pretendia analisa-las contra a percepção subjetiva de “dificuldade”. Para melhor visualização o dos quadros apresentados na sequência sugere-se consultar também o Apêndice 2. O apêndice 1 apresenta o dicionário de variáveis em ordem alfabética (1A) ou numérica 1(B).

4.4.1 Renda e informalidade no trabalho

As variáveis deste bloco informam as relações entre renda e contribuição com a previdência pública sob a perspectiva da “dificuldade”.

Prevalece entre as famílias que vivem com “dificuldade” a não contribuição com a previdência pública por nenhum membro da família (43%) provavelmente reflexo de duas situações observadas: a informalidade das relações destas pessoas com o trabalho e o não exercício do trabalho principal pelo chefe de família já há algum tempo. De acordo com a publicação da

POF (2009) a definição de trabalho principal leva em consideração um conjunto de critérios que combinam (a) o trabalho atual (b) o último trabalho (c) o trabalho em que a pessoa esteve no maior número de horas no último mês (d) aquele que proporcionava maior rendimento mensal. A característica evidenciada como “ Não tem trabalho principal” refere-se aos aposentados.

Nota-se, pelos quadros abaixo que , 35% das família recebem rendas advindas de trabalhadores que atuam por conta própria (V5) e 69% de trabalhadores que são empregados (V6) . Uma família pode ter mais do que uma origem de renda. Estas duas evidências não diferenciam muito os grupos de “dificuldade”/ “facilidade”. O que os diferencia, na verdade, são a informalidade, representada pela variável que trata da previdência pública (V1) e rendas muito baixas, inferiores ao salário mínimo (V2) e o tempo sem o trabalho principal (V3). Aposentados e famílias cujos trabalhadores são formais, onde todos contribuem com a previdência pública, reportam maior facilidade.

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO			WOE	
V1 PREVIDENCIA PUBLICA				Total	SIM / NÃO		DE DIFICULDADE
APOSENTADO	73%	15%	13%	13%	0.89	-	0.12
NINGUEM CONTRIBUI	80%	32%	43%	40%	1.34	-	0.29
PARTE CONTRIBUI	74%	25%	23%	24%	0.93	-	0.07
TODOS CONTRIBUEM	69%	28%	21%	23%	0.74	-	0.30
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica se alguém da família contribui para a previdência social ou se está aposentado							
V2 RENDA DAS PESSOAS UC				Total	SIM / NÃO		DE DIFICULDADE
PREDOMINA APOSENTADORIA	73%	15%	13%	13%	0.89	-	0.12
PREDOMINA MAIS QUE SALARIO MINIMO	68%	66%	47%	51%	0.71	-	0.34
PREDOMINA SALARIO MENOS QUE SALÁRIO MÍNIMO	86%	20%	40%	35%	2.03	-	0.71
PREDOMINA SALARIO MINIMO	88%	1%	3%	3%	2.52	-	0.93
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica o tipo de rendimento (salários mínimos) que predomina na família							
V3 TEMPO SEM TRAB. PRINC. REC				Total	SIM / NÃO		DE DIFICULDADE
A_SEMPRE COM TRABALHO PRINCIPA	71%	57%	47%	49%	0.82	-	0.20
B_NAO_TEM TRABALHO_PRINCIPAL (aposentado)	73%	15%	13%	13%	0.89	-	0.12
C_DE 1 A 3 MESES	79%	11%	13%	13%	1.27	-	0.24
D_DE 4 A 6 MESES	81%	11%	15%	14%	1.40	-	0.34
E_MAIS DE 6 MESES	83%	7%	12%	11%	1.66	-	0.50
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica se o chefe da família sempre tem trabalho principal , quanto tempo no ano ficou sem rendimento deste trabalho ou se está aposentado							
V4 X_RENDA_BENEFIC_ASSALARIADO				Total	SIM / NÃO		DE DIFICULDADE
NAO	82%	26%	38%	35%	1.47	-	0.39
SIM	72%	74%	62%	65%	0.84	-	0.18
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica se no domicílio há recebimento de benefícios típicos dos assalariados EX : Recebe décimo terceiro, bonus de férias, vale alimentação.							
V5 X_RENDA_CONTA_PROPRIA				Total	SIM / NÃO		DE DIFICULDADE
NAO	74%	68%	64%	65%	0.94	-	0.07
SIM	77%	32%	36%	35%	1.13	-	0.13
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica se no domicílio tem renda de trabalhador por conta própria							
V6 X_RENDA_EMPREGADO				Total	SIM / NÃO		DE DIFICULDADE
NAO	74%	33%	31%	31%	0.94	-	0.06
SIM	76%	67%	69%	69%	1.03	-	0.03
TOTAL	75%	100%	100%	100%	1.00	-	0.00
Indica se no domicílio tem renda de empregado							

Quadro 1 – Renda e informalidade no trabalho

4.4.2 Presença de cônjuge e gênero

Observando as variáveis individualmente percebe-se que : o fato de não ter cônjuge parece não diferenciar os dois grupos analisados (facilidade/”dificuldade”) e o fato do responsável pelo domicílio ser mulher também não aparenta diferenciar os grupos por percepção de “dificuldade”. Isso contraria, de certa forma, conclusões apresentadas em estudos prévios que indicam ações inclusivas por gênero (SILVEIRA, 2007, p.201). Neste sentido o comportamento individual desta variável precisa ser melhor investigado no presente trabalho. Por exemplo : percebe-se que reuniu-se a este grupo de variáveis a informação que indica que a família recebe pensão, mesada ou doação, e para estas a percepção de “dificuldade” é bem acima da média. Levanta-se a hipótese que a questão das mulheres esteja localizada provavelmente em certa fase de suas vidas quando possuem filhos pequenos, não têm acesso a creches, não podem trabalhar e melhorar a condição de renda da família. Assim, parte vive de pensão, mesada ou doação.

Observa-se que a escolaridade do cônjuge diferencia bastante os dois grupos, constatou-se que esta variável está relacionada com a própria escolaridade do responsável pelo domicílio. A informação da escolaridade do responsável pelo domicílio não se encontra apresentada no quadro abaixo, pois faz parte do computo dos pontos para cálculo da classe econômica . A V7, status do cônjuge, reflete questões vinculadas à escolaridade. De acordo com Menezes Filho (2012) entre 2005 e 2009 a escolaridade média dos 20% mais pobres dobrou, passando de dois anos para quatro anos em média. Os 20% mais ricos passaram de oito para dez anos. Ou seja, o “avanço” levou os mais pobres apenas ao “antigo primário” e os mais ricos ao antigo “colegial incompleto”, o que mostra o “atraso” em relação ao “capital humano” nacional. Entende-se nesta tese o “capital humano” vinculado à educação.

V7	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	STATUS_CONJUGE	DIFICULDADE	NÃO	SIM	Total		WOE	
	A_ANALFABETO	89%	3%	8%	7%	2.55	0.94	Muito Acima da Media
	B_0 A 3	81%	6%	9%	8%	1.46	0.38	Acima da Media
	C_4 A 7	82%	13%	19%	18%	1.52	0.42	Bem Acima da Media
	D_8 A 10	78%	9%	10%	10%	1.16	0.14	Media
	E_11 A 14	69%	24%	17%	19%	0.73	0.31	Bem Abaixo da Media
	F_15 OU MAIS	45%	11%	3%	5%	0.27	1.30	Muito Abaixo da Media
	G_NAO DECLARADO	76%	0%	0%	0%	1.03	0.03	Media
	NAO_TEM_CONJUGE	75%	34%	33%	33%	0.97	0.03	Media
		75%	100%	100%	100%	1.00	0.00	Media
Informa a escolaridade do conjuge ou se não tem conjuge								
V8	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	SEXO	DIFICULDADE	NÃO	SIM	Total		WOE	
	FEMININO	77.01%	29%	32%	31%	1.11	0.10	Media
	MASCULINO	74.32%	71%	68%	69%	0.96	0.04	Media
		75.15%	100%	100%	100%	1.00	-	Media
Indica o sexo do chefe da familia								
V9	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	TEM_CONJUGE_sum	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO	75%	34%	33%	33%	0.97	0.03	Media
	SIM	75%	66%	67%	67%	1.02	0.02	Media
		75%	100%	100%	100%	1.00	-	Media
Indica se o chefe de familia tem conjuge								
V10	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	X_RENDA_PENSAO_MESADA_DOACAO	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO	74%	93%	89%	90%	0.96	0.04	Media
	SIM	82%	7%	11%	10%	1.52	0.42	Bem Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Indica se no domicilio existe renda de pensão, mesada ou doação								

Quadro 2 – Presença de cônjuge e gênero

4.4.3 Adimplência no pagamento de despesas

É importante observar que 36% das famílias declararam atrasar despesas básicas com água, eletricidade e gás (V11), e declaram viver com “dificuldade”, acima da média. Para algumas famílias essa pergunta “não se aplica”, pois provavelmente não têm acesso ao serviço.

Estar na posição de “devedor” ou “atrasado com prestações” (V12) aparenta ser pior em termos de dificuldade do que não ter acesso aos serviços ou ao crédito (item não se aplica).

V11	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO	67%	80%	53%	60%	0.67	0.41	Bem Abaixo da Media
	NAO_SE_APLICA	83%	3%	5%	4%	1.63	0.49	Bem Acima da Media
	SIM	88%	17%	42%	36%	2.47	0.90	Muito Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Declara já ter atrasado contas de água ou eletricidade ou gás.								
V12	EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO	69%	82%	61%	66%	0.74	0.30	Bem Abaixo da Media
	NAO_SE_APLICA	81%	4%	6%	5%	1.45	0.37	Acima da Media
	SIM	88%	14%	33%	28%	2.40	0.88	Muito Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Declara já ter atrasado outras contas relativas a bens ou serviços adquiridos								

Quadro 3 – Adimplência e pagamento de despesas

4.4.4 Idosos com aposentadoria, empregadores , previdência privada

Apesar destas variáveis não apresentarem relação direta entre si, considerando a entropia, estão apresentadas em conjunto por representarem um certo “privilégio conquistado” : os idosos com aposentadoria, os empregadores e os que recebem renda de previdência privada podem se considerar privilegiados, pois declaram viver com mais facilidade. Infelizmente são poucos os empregadores e muito menos ainda os que contam com previdência privada. Aposentados são muitos e tendem a crescer, no entanto não se pode esperar a mesma situação de relativo “bem estar” no futuro próximo, considerando o nível de informalidade no Brasil e o consequente déficit previdenciário.

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO SIM			WOE	
V13 FATOR 8 IDOSO_COM_APOSENTADORIA				Total	SIM / NÃO		DE DIFICULDADE
NAO	76%	73%	78%	77%	1.07	0.06	Media
SIM	71%	27%	22%	23%	0.82	0.20	Abaixo da Media
	75%	100%	100%	100%	1.00	-	Media
Indica se no domicilio tem idoso recebendo aposentadoria							
EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO SIM			WOE	
V24 X_RENDA_EMPREGADOR				Total	SIM / NÃO		DE DIFICULDADE
NAO	76%	92%	97%	96%	1.06	0.06	Media
SIM	53%	8%	3%	4%	0.37	0.99	Muito Abaixo da Media
	75%	100%	100%	100%	1.00	-	Media
Indica se no domicilio tem renda de empregador							
EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO SIM			WOE	
V68 X_RENDA_PREV_PRIVADA				Total	SIM / NÃO		DE DIFICULDADE
NAO	75%	99%	100%	99%	1.01	0.01	Media
SIM	49%	1%	0%	1%	0.31	1.16	Muito Abaixo da Media
	75%	100%	100%	100%	1.00	-	Media
Indica se no domicilio tem renda de previdência privada							

Quadro 4 – Aposentados, empregadores e previdência privada.

4.4.5 Despesas Básicas, Diferenciadas e Necessidades Específicas de Saúde

A forma como o orçamento doméstico está comprometido difere-se bastante para as famílias que percebem maior “dificuldade”. Utilizam seus orçamentos com despesas básicas como alimentação e vestuário, não diversificando muito a cesta de produtos domésticos (V14) ou de alimentos (V15), além disso, apresentam poucos gastos individuais (V18), ou seja, o orçamento é concentrado nas despesas compartilhadas entre os familiares.

Boa parte (36%) das famílias com “dificuldade” tiveram necessidades não atendidas em relação a medicamentos ou serviços de saúde (V16), nos últimos meses antes da pesquisa.

O levantamento das despesas na POF, relativo às despesas pessoais e coletivas, não é realizado para todas as famílias da mesma forma, o que foi considerado neste estudo como “sinal de alerta” sobre o uso destes dados para o fim a que se propõe esta tese: analisar família por família.

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO			WOE	
V14 CADERNETA_COM_DIFERENCIADOS				Total	SIM / NÃO		DE DIFICULDADE
A_NAO_COMPROU_NENHUMA_CATEGORIA_DIFERENCIADA	81%	26%	36%	33%	1.38	0.32	Acima da Media
B_COMPROU_1_CATEGORIA_DIFERENCIADA	76%	26%	28%	28%	1.07	0.07	Media
C_COMPROU_2_CATEGORIAS_DIFERENCIADAS	72%	23%	20%	20%	0.86	0.16	Abaixoda Media
D_SEM_INFORMACAO_CADERNETA	71%	8%	7%	7%	0.82	0.20	Abaixoda Media
E_COMPROU_3_OU MAIS_CATEGORIAS_DIFERENCIADAS	64%	17%	10%	12%	0.60	0.52	Bem Abaixo da Media
	75%	100%	100%	100%	1.00	-	Media
Informa presença de itens diferenciados na caderneta de despesa							
EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO			WOE	
V15 DIVERSIDADE_DESPESAS_ALIMENTACAO				Total	SIM / NÃO		DE DIFICULDADE
<=4 TIPOS	76%	34%	35%	35%	1.02	0.02	Media
<=7	77%	26%	28%	27%	1.08	0.08	Media
<=10	76%	22%	23%	23%	1.03	0.03	Media
>10 TIPOS	71%	18%	14%	15%	0.79	0.23	Abaixoda Media
	75%	100%	100%	100%	1.00	0.00	Media
Apresenta a diversidade observada em itens de alimentação							

Quadro 5 – Despesas básicas, diferenciadas e necessidades de saúde (continua...)

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V16 FATOR_27_NECESITOU_SERV_SAUDE_OU_MEDICAMENTO		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
NAO	68%	89%	64%	70%	0.72	-	Bem Abaixo da Media
SIM	91%	11%	36%	30%	3.28	1.19	Muito Acima da Media
	75%	100%	100%	100%	1.00	0.00	Media

Indica se a família necessitou de medicamentos nos últimos meses mas não foi atendida neste sentido

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V17 IMPACTO_ALIMENTACAO		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
Z_SEM_DESPESA_ALIMENTACAO_0_0	77%	3%	4%	4%	1.10	0.09	Media
B_MUITO_ABAIXO_0_0.13	67%	42%	28%	31%	0.66	0.42	Bem Abaixo da Media
C_ABAIXO_0.13_0.20	72%	22%	19%	20%	0.86	0.15	Abaixoda Media
D_MEDIA_0.20_0.25	78%	9%	10%	10%	1.15	0.14	Media
E_ALTO_0.25_0.33	80%	12%	16%	15%	1.35	0.30	Acima da Media
F_MUITO_ALTO_0.33_0.43	83%	7%	11%	10%	1.66	0.51	Bem Acima da Media
G_BEM_MAIS_ALTO_0.43_OU MAIS	88%	5%	12%	10%	2.44	0.89	Muito Acima da Media
	75%	100%	100%	100%	1.00	-	Media

Faixa Percentual que representa impacto de gastos com alimentação

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V18 PERC_DESP_INDIV		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
ATE_27_PERC	80%	15%	19%	18%	1.30	0.26	Acima da Media
28_40_PERC	78%	17%	20%	19%	1.19	0.17	Media
40_52_PERC	75%	21%	21%	21%	1.01	0.01	Media
52_64_PERC	72%	24%	20%	21%	0.86	0.15	Abaixoda Media
MAIS_DE_64_PERC	71%	25%	20%	21%	0.82	0.20	Abaixoda Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com despesas individuais

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V19 IMPACTO_OUTRAS		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
E_SEM_OUTRAS_DEPESAS_0_0	82%	1%	1%	1%	1.46	0.38	Acima da Media
B_MUITO_ALTO_0_0.05	67%	32%	21%	24%	0.67	0.41	Bem Abaixo da Media
C_ACIMA_0.05_0.08	71%	17%	14%	15%	0.82	0.19	Abaixoda Media
D_MEDIA_0.08_0.14	76%	29%	31%	30%	1.07	0.07	Media
E_ABAIXO_0.14_0.20	80%	12%	16%	15%	1.33	0.29	Acima da Media
F_BEM_ABAIXO_0.20_ou mais	84%	10%	17%	15%	1.76	0.57	Bem Acima da Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com outras despesas

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V20 IMPACTO_OUTRAS_DIF		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
E_SEM_DESPESA_OUTRAS_DIF_0_0	89%	4%	12%	10%	2.73	1.00	Muito Acima da Media
A_BEM_ALTO_0.27_OU MAIS	53%	18%	7%	10%	0.38	0.97	Muito Abaixo da Media
B_MUITO_ALTO_0.17_0.27	64%	21%	13%	15%	0.60	0.51	Bem Abaixo da Media
C_ACIMA_0.14_0.17	72%	6%	5%	5%	0.83	0.18	Abaixoda Media
D_MEDIA_0.04_0.14	76%	34%	36%	35%	1.05	0.05	Media
E_ABAIXO_0.03_0.04	79%	4%	5%	5%	1.24	0.22	Acima da Media
F_BEM_ABAIXO_0.01_0.03	83%	7%	11%	10%	1.60	0.47	Bem Acima da Media
G_MUITO_ABAIXO_0_0.01	87%	5%	12%	10%	2.27	0.82	Muito Acima da Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com outras despesas diferenciadas

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V21 IMPACTO_VESTUARIO		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
Z_SEM_DESPESA_VESTUARIO_0_0	80%	10%	13%	13%	1.31	0.27	Acima da Media
B_MUITO_ABAIXO_0_0	69%	3%	2%	2%	0.74	0.30	Bem Abaixo da Media
C_ABAIXO_0_0.02	72%	17%	14%	15%	0.85	0.16	Abaixoda Media
D_MEDIA_0.02_0.12	74%	62%	59%	60%	0.97	0.04	Media
E_ALTO_0.12_OU MAIS	79%	8%	11%	10%	1.27	0.24	Acima da Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com vestuário

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V22 FATOR_19_CREDITO_E_OU_HABITACAO_DIFERENC		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
NÃO_TEVE_ACESSO	83%	20%	31%	29%	1.61	0.48	Bem Acima da Media
TEVE_ACESSO	72%	80%	69%	71%	0.85	0.16	Abaixoda Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com crédito e habitação

EVIDÊNCIA	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE					
V23 IMPACTO_HABITACAO_DIF		NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
E_SEM_DESPESA_HABIT_DIF_0_0	85%	14%	28%	24%	1.93	0.66	Bem Acima da Media
B_MUITO_ALTO_0.06_OU MAIS	66%	49%	31%	35%	0.63	0.46	Bem Abaixo da Media
C_ACIMA_0.04_0.06	71%	12%	10%	10%	0.83	0.19	Abaixoda Media
D_MEDIA_0.02_0.04	76%	14%	15%	15%	1.07	0.07	Media
E_ABAIXO_0.01_0.02	81%	4%	5%	5%	1.43	0.36	Acima da Media
F_BEM_ABAIXO_0_0.01	82%	8%	12%	11%	1.52	0.42	Bem Acima da Media
	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com habitação diferenciada

Quadro 5 – Despesas básicas, diferenciadas e necessidades de saúde (continuação)

4.4.6 Moradia e Alimentação

Das famílias que vivem com “dificuldade” 33% declaram que a alimentação às vezes não é suficiente, 56% declaram que a alimentação nem sempre é do tipo que quer, pois a renda não permite. Esta situação de “dificuldade” se agrava na medida em que as moradias são também ruins, refletindo real pobreza e “dificuldade” de levar a vida com o orçamento doméstico. Considerando o WOE pode-se observar que estas variáveis terão grande importância para a Rede Bayesiana. As estatísticas abaixo ilustram estas afirmações.

V31	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	CONDICOES_DE_MORADIA	DIFICULDADE	NÃO	SIM	Total			
	BOAS	65%	74%	44%	52%	0.60	-	0.51
	RUINS	96%	1%	12%	9%	7.92	2.07	Muito Acima da Media
	SATISFATORIAS	84%	25%	44%	39%	1.76	0.56	Bem Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa declarou sua percepção geral sobre a sua moradia								
V33	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	QUANTIDADE_DE_ALIMENTO_REC	DIFICULDADE	NÃO	SIM	Total			
	AS_VEZES_NAO_E_SUFICIENTE	94%	7%	33%	26%	4.83	1.58	Muito Acima da Media
	E_SEMPRE_SUFICIENTE	65%	92%	56%	65%	0.61	0.50	Bem Abaixo da Media
	NORMALMENTE_NAO_E_SUFICIENTE	96%	2%	12%	9%	7.77	2.05	Muito Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa declarou sua percepção se a quantidade de alimentos está adequada às necessidades familiares.								
V34	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	TIPO_ALIMENTO_RAZAO_REC	DIFICULDADE	NÃO	SIM	Total			
	ALIMENTOS NEM SEMPRE DISPONIVEIS OU OUTROS	58%	10%	5%	6%	0.45	0.80	Muito Abaixo da Media
	NEM SEMPRE DO TIPO QUE QUER POIS A RENDA NAO	90%	18%	56%	47%	3.14	1.14	Muito Acima da Media
	RARAMENTE DO TIPO QUE QUER POIS A RENDA NAO	96%	2%	16%	12%	7.71	2.04	Muito Acima da Media
	SEMPRE DO TIPO QUE QUER	51%	70%	24%	35%	0.34	1.08	Muito Abaixo da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa declarou sua percepção se os tipos de alimentos estão adequados e a razão de não se alimentar como quer.								

Quadro 6 – Moradia e Alimentação.

4.4.7 Classe Econômica, escolaridade e despesas diferenciadas.

A Classe Econômica (ABEP, 2008), foi calculada a partir do Critério Brasil na forma de categorias ou “pontos”. Utilizou-se da fórmula vigente para o ano de 2008, e calculou-se a relação com a percepção de “dificuldade”.

Associada à Classe Econômica (V35) está a escolaridade do chefe da família e o fator “Espaço em Casa” (V38), pois estas variáveis fazem parte do modelo adotado pela ABEP. No caso da variável “Espaço_Casa” a influência vem do número de banheiros, utilizado na fórmula do critério Brasil e na composição desta variável.

Além disso, aparecem reunidas a este grupo de variáveis a posse de cartões de crédito e conta corrente (V37), itens estes associados às Classes Econômicas mais altas pelo Critério Bra-

sil . Vale salientar que estas informações relativas ao crédito e conta corrente estão disponíveis para as pessoas que responderam ao questionário de orçamento. A informação de conta corrente apenas foi coletada quando a família possuía cheque especial. Esta forma de captação do dado pareceu estranha e pode ser otimizada nas próximas pesquisas pelo IBGE, mudando-se apenas a ordem das perguntas ou suas dependências no questionário. As despesas com “transporte diferenciado” (V39) ou o tipo do veículo que possui (V69), também apareceram associadas aos grupos com maior ou menor facilidade orçamentária.

V35	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	CLASSE_ABEP	DIFICULDADE	NÃO	SIM	Total		WOE	
	A1+A2	32%	7%	1%	2%	0.16	-	1.84
	B1	46%	13%	4%	6%	0.28	-	1.26
	B2	59%	22%	11%	14%	0.47	-	0.75
	C1	72%	24%	20%	21%	0.83	-	0.18
	C2	81%	18%	25%	23%	1.42	-	0.35
	D	87%	15%	32%	28%	2.20	-	0.79
	E	91%	2%	8%	6%	3.40	-	1.22
Classe ABEP - agregado da pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família e número de banheiros.								
V36	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	ESCOLARIDADE_PESS_REFERENCIA	DIFICULDADE	NÃO	SIM	Total		WOE	
	A_ANALFABETO	87%	7%	16%	14%	2.14	-	0.76
	B_0 A 3	82%	10%	15%	14%	1.46	-	0.38
	C_4 A 7	81%	19%	26%	25%	1.37	-	0.32
	D_8 A 10	79%	10%	13%	12%	1.25	-	0.23
	E_11 A 14	69%	32%	24%	26%	0.75	-	0.29
	F_15 OU MAIS	44%	20%	5%	9%	0.25	-	1.37
	G_NAO_DECLARADO	80%	1%	1%	1%	1.35	-	0.30
		75%	100%	100%	100%	1.00	-	Media
Escolaridade da pessoa de referência ou responsável pela família								
V37	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	Fator_17_CARTAO_CONTA_CORRENTE	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO	82%	46%	70%	64%	1.52	-	0.42
	SIM	63%	54%	30%	36%	0.56	-	0.58
		75%	100%	100%	100%	1.00	-	Media
Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Cartão de Crédito ou para as quais foi colhida informação de Conta Corrente apenas quando possui cheque especial								
V38	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	FATOR_5_ESPACO_CASA	DIFICULDADE	NÃO	SIM	Total		WOE	
	ATE_4_COMODOS_1_BANHEIRO_1_COMO_DORM	86%	7%	15%	13%	2.01	-	0.70
	PRED_4_5_COMODOS_1_BANHEIRO_2_DORM	80%	28%	37%	34%	1.31	-	0.27
	PRED_2_A_3_DORM_6 MAIS_COMODOS_1_BANHEIRO	77%	29%	32%	32%	1.12	-	0.11
	PRED_2_A_3_DORM_7 MAIS_COMODOS_2_BANHEIROS	62%	21%	12%	14%	0.54	-	0.61
	PRED_9_OU MAIS_COMODOS_3_OU MAIS_BAHEIROS E DORM	48%	15%	4%	7%	0.31	-	1.18
		75%	100%	100%	100%	1.00	-	Media
Nr de cômodos que estão sendo utilizados como dormitório no domicílio_Quantidade de banheiros no domicílio_Quantidade de cômodos no domicílio_A pessoa declarou ter pouco espaço na casa								
V39	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	IMPACTO_TRANSPORTE_DIF	DIFICULDADE	NÃO	SIM	Total		WOE	
	E_SEM_DESPESA_TRANSP_DIF_0_0	83%	30%	51%	46%	1.67	-	0.51
	A_BEM_ALTO_0.31_OU MAIS	59%	8%	4%	5%	0.47	-	0.76
	B_MUITO_ALTO_0.04_0.31	66%	41%	26%	30%	0.63	-	0.46
	C_ACIMA_0.03_0.04	70%	6%	5%	5%	0.75	-	0.28
	D_MEDIA_0_0.03	76%	14%	15%	15%	1.07	-	0.06
		75%	100%	100%	100%	1.00	-	Media
Faixa Percentual que representa impacto de gastos com transporte diferenciado								
V69	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	TIPO_VEICULO	DIFICULDADE	NÃO	SIM	Total		WOE	
	BICICLETA	86%	13%	26%	23%	1.99	-	0.69
	NAO POSSUI VEICULO	81%	25%	35%	33%	1.40	-	0.33
	MOTOCICLETA	79%	7%	9%	8%	1.23	-	0.21
	SEM INFORMACAO	72%	5%	5%	5%	0.85	-	0.16
	UM AUTOMOVEL	64%	37%	22%	26%	0.59	-	0.53
	DOIS OU MAIS AUTOMOVEIS	47%	12%	4%	6%	0.29	-	1.22
	Total	75%	100%	100%	100%	1.00	-	Media
Índice se veículo prioritário é carro, moto, bicicleta, nenhum								

Quadro 7 – Classe Econômica, escolaridade, despesas diferenciadas e veículos.

4.4.8 Transporte público, acesso à educação, à saúde e ao lazer.

A presença das variáveis relacionadas aos serviços de transporte (V43), associadas ao acesso à educação, saúde e lazer (V41) mostra que muitos brasileiros não têm às mesmas oportunidades devido à dificuldade de locomoção ou mesmo pelo fato da despesa de transporte pesar muito no bolso dos que vivem com dificuldade.

É interessante também observar que no processo de agrupamento de variáveis, as percepções sobre o acesso à educação, à saúde e ao lazer foram reunidas (V41), refletindo que 61% das famílias têm problemas nestas áreas.

A questão do transporte coletivo agrava esta situação, indicando que os que passam por maiores dificuldades vivem nos rincões do Brasil ou em grandes centros, periferias, sendo obrigados a gastar grande parte de seus orçamentos com transporte ou viverem longe das possibilidades de educação, lazer e saúde.

V 41	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	DIFICULDADE	NÃO	SIM	Total			
	TEM BOM ACESSO	68%	50%	35%	39%	0.69	-	0.37
	NÃO TEM BOM ACESSO	80%	50%	65%	61%	1.31	0.27	Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
Acesso a serviços de educação_Acesso a serviços de lazer e esporte_Acesso a serviços de saúde								
V42	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	PERC_TRANSPORTE	DIFICULDADE	NÃO	SIM	Total			
	<=0	77%	29%	32%	31%	1.10	0.09	Media
	<=0.05	69%	40%	30%	33%	0.74	0.30	Bem Abaixo da Media
	<=0.10	77%	16%	18%	17%	1.08	0.08	Media
	<=0.15	80%	7%	9%	9%	1.30	0.26	Acima da Media
	<=0.25	82%	5%	8%	7%	1.56	0.44	Bem Acima da Media
	>0.25	82%	2%	4%	3%	1.54	0.43	Bem Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
Faixa Percentual que representa impacto de gastos com transporte								
V 43	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	SERVICO_DE_TRANSPORTE_COLETIVO	DIFICULDADE	NÃO	SIM	Total			
	_BOM+NAO_RES	71%	58%	46%	49%	0.79	0.23	Abaixo da Media
	_NAO_TEM+RUIM	80%	42%	54%	51%	1.29	0.25	Acima da Media
		75%	100%	100%	100%	1.00	-	Media
A pessoa apresentou sua percepção sobre transporte coletivo no local onde mora								

Quadro 8 – Transporte público, acesso à educação, saúde e lazer.

4.4.9 Condições físicas das moradias

Nota-se que 22% entre os que têm “dificuldade” afirmam viver em casa escuras (V44) , bem como têm a presença de goteiras, problemas com a fundação das casas, paredes, chão úmidos e madeiras deterioradas (V45) . Metade dos domicílios brasileiros têm pelo menos um destes problemas em suas casas, indicando dificuldades.

V44	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
	CASA_ESCURA		NÃO	SIM			WOE	
	NAO	72%	90%	78%	81%	0.87	0.14	Abaixo da Media
	SIM	87%	10%	22%	19%	2.23	0.80	Muito Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa declarou que percebe sua casa escura								

V45	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS		NÃO	SIM				
	NAO	65%	70%	43%	50%	0.62	0.48	Bem Abaixo da Media
	SIM	85%	30%	57%	50%	1.88	0.63	Bem Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa declarou ter paredes e chão umidos_A pessoa declarou ter janelas, portas ou assoalhos deteriorados_A pessoa declarou ter telhado com goteira								

Quadro 9 – Condições físicas da moradia.

4.4.10 Habitação e Comprometimento de Renda

Apesar das condições de moradia não serem boas para todas as famílias brasileiras, a maioria declara ter casa própria (V46) . O valor da moradia própria é computado na POF como aluguel estimado e este valor impacta tanto nas despesas quanto na renda, identificada como “não monetária” (V50). Quando este valor é proporcionalmente alto em relação às despesas totais ou à renda significa que as famílias têm praticamente só a casa, e nem sempre em boas condições.

O comprometimento de renda das famílias que vivem com “dificuldade” relaciona-se também com as despesas básicas de habitação (V49) relativas a gás, energia, água, etc, as quais tomam grande parte de seus orçamentos. Estas são despesas classificadas como não diferenciadas em habitação.

Pelo menos metade das famílias têm mais do que 100% dos seus rendimentos monetários comprometidos (V45 b) e a partir deste nível de comprometimento as famílias começam a declarar “dificuldade”. Caso fossem computadas despesas de aluguel como provisões para melhorias habitacionais a proporção de famílias com comprometimento acima de 100% subi-

ria para 61% (V45 a), indicando que poucos são os que conseguiriam fazer economias para investir na reposição ou melhoria das próprias moradias.

EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
		DIFICULDADE	NÃO			WOE	
V45a	COMPROM_RENDA						
	sem informação	88%	0%	0%	2.46	0.90	Muito Acima da Media
	<=0.5	68%	7%	5%	0.71	0.34	Bem Abaixo da Media
	<=0.6	69%	6%	4%	0.73	0.31	Bem Abaixo da Media
	<=0.7	70%	7%	6%	0.77	0.26	Abaixada Media
	<=0.8	69%	9%	7%	0.74	0.30	Bem Abaixo da Media
	<=1	72%	18%	15%	0.84	0.17	Abaixada Media
	>1	78%	53%	63%	1.19	0.18	Media
		75%	100%	100%	1.00	0.00	Media
Comprometimento de renda. Considerando somatório de todas as despesas pela renda total							
V46	CONDICAO_OCUPACAO_REC						
	ALUGADO	75%	16%	17%	1.01	0.01	Media
	CEDIDO_OUTRA_FORMA	83%	6%	9%	1.58	0.45	Bem Acima da Media
	PROPRIO_OU_CEDIDO_POR_EMPREGADOR	74%	78%	74%	0.95	0.05	Media
		75%	100%	100%	1.00	0.00	Media
Apresenta a condição de ocupação do domicílio : alugado, cedido ou próprio							
V48	FATOR_28_MORA_DE_ALUGUEL						
	NAO	75%	84%	83%	1.00	0.00	Media
	SIM	75%	16%	17%	1.01	0.01	Media
		75%	100%	100%	1.00	0.00	Media
Há quanto tempo a pessoa fez o contrato de aluguel? Apresenta o tipo de contrato de aluguel, sendo formal ou informal							
V48a	IMPACTO_ALUGUEL_ESTIM						
	Z_SEM_DESPESA_ALUGUEL_ESTIM_0_0	75%	16%	17%	1.02	0.02	Media
	B_MUITO_ABAIXO_0_0.04	68%	4%	3%	0.70	0.35	Bem Abaixo da Media
	C_ABAIXO_0.04_0.11	72%	29%	24%	0.83	0.19	Abaixada Media
	D_MEDIA_0.11_0.33	76%	43%	46%	1.07	0.07	Media
	E_ALTO_0.33_OU MAIS	81%	8%	11%	1.37	0.31	Acima da Media
	SEM INFORMACAO	100%	0%	0%			
		75%	100%	100%	1.00	-	Media
Faixa Percentual que representa impacto de gastos com habitação (não diferenciada)							
V49	PERC_HABITACAO						
	<=0.05	65%	14%	8%	1.0%	0.62	Bem Abaixo da Media
	<=0.1	70%	29%	23%	24%	0.77	0.26
	<=0.15	76%	23%	24%	24%	1.02	0.02
	<=0.2	78%	14%	17%	16%	1.17	0.16
	<=0.3	81%	12%	17%	16%	1.38	0.32
	>0.3	83%	7%	12%	11%	1.56	0.45
		75%	100%	100%	100%	1.00	0.00
Faixa Percentual que representa impacto de gastos com habitação (não diferenciada)							
V50	PERC_REnda_NAO_MONET						
	sem inforção	100%	0%	0%	0%		
	<=0.0	69%	10%	7%	8%	0.73	0.32
	<=0.1	66%	36%	23%	26%	0.66	0.42
	<=0.2	75%	29%	30%	30%	1.01	0.01
	<=0.3	80%	14%	18%	17%	1.32	0.28
	<=0.4	85%	5%	10%	9%	1.84	0.61
	> 0.4	86%	6%	11%	10%	2.00	0.69
		75%	100%	100%	100%	1.00	0.00
Faixa Percentual da renda que não é monetária.							

Quadro 10 – Habitação e comprometimento de renda.

4.4.11 Quantidade de Pessoas na Família, ciclo de vida familiar e Educação

Quanto mais pessoas na família (V54), maior a percepção de “dificuldade”. No entanto essa “dificuldade” está centrada em famílias que possuem filhos dependentes de renda (V53).

Provavelmente o “bônus demográfico” ajude a diminuir a percepção média de “dificuldade”. Nos próximos anos a taxa de dependência demográfica tende a diminuir alcançando patamares historicamente mais baixos por volta de 2015/2020, no entanto para alcançar o futuro. Após 2025/2030, será necessária uma nova agenda de desenvolvimento social, conforme estudo do Banco Mundial (2011), já citado.

Aproximadamente, metade das famílias vivem com “dificuldade” e podem ser encontradas nas escolas pública (V56), com contato praticamente diário, sem que necessariamente este fato seja aproveitado de fato para alavancar o desenvolvimento do “capital humano” nacional.

Os solitários e as famílias com duas pessoas e sem filhos (V53) são os que menos reportam “dificuldade”.

Observa-se também poucos cônjuges estudando (V51), mesmo que os adultos brasileiros necessitem de capacitação e tenham pela frente um longo caminho de trabalho.

V51	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	CONJ OU RESP ESTUDANDO_sum	DIFICULDADE	NÃO	SIM	Total		WOE	
	PELO MENOS UM ESTUDANDO	70%	11%	8%	9%	0.77	- 0.24	Abaixada Media
	NENHUM ESTUDANDO	76%	89%	92%	91%	1.03	0.03	Media
		75%	100%	100%	100%	1.00	0.00	Media
Indicador se o conjuges ou o responsável está estudando								
V52	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	ESCOLARIDADE_MENORES	DIFICULDADE	NÃO	SIM	Total		WOE	
	NAO TEM MENORES DE 18 ANOS	68%	58%	41%	45%	0.71	- 0.34	Bem Abaixo da Media
	TEM PESSOA DE 7 A 17 ANOS E TODAS ESTAO NA ESCOLA	80%	22%	28%	27%	1.28	0.25	Acima da Media
	TEM PESSOA DE 7 A 17 ANOS E FORA DA ESCOLA	85%	1%	2%	2%	1.89	0.64	Bem Acima da Media
	TEM_MENOR_6_ANOS_COM_CRECHE_ESCOLA	80%	9%	11%	11%	1.31	0.27	Acima da Media
	TEM_MENOR_6_ANOS_SEM_CRECHE_ESCOLA	83%	10%	17%	15%	1.64	0.50	Bem Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Situação da escolaridade dos menores de 18 anos								
V53	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	FATOR_10_FAMILIAS	DIFICULDADE	NÃO	SIM	Total		WOE	
	ATE DOIS FILHOS ADULTOS OU ADOLESCENTES	75%	26%	26%	26%	0.98	- 0.02	Media
	DUAS PESSOAS SEM FILHOS	67%	24%	16%	18%	0.66	0.42	Bem Abaixo da Media
	FAMILIAS COM CRIANCAS ATE 6 ANOS	82%	19%	28%	26%	1.49	0.40	Acima da Media
	MAIS DE DOIS FILHOS PREDOM ADOLESCENTES	81%	14%	20%	18%	1.39	0.33	Acima da Media
	SOLITARIOS	65%	16%	10%	12%	0.62	0.47	Bem Abaixo da Media
		75%	100%	100%	100%	1.00	0.00	Media
Tipos de família considerando a idade do filho mais novo_Número de dependentes de renda na família_Quantidade de filhos na família								
V54	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	QTDE_MORADORES	DIFICULDADE	NÃO	SIM	Total		WOE	
	1	65%	17%	10%	12%	0.63	- 0.47	Bem Abaixo da Media
	2	69%	28%	21%	22%	0.75	0.29	Abaixada Media
	3	74%	25%	24%	25%	0.97	0.04	Media
	4	79%	18%	22%	21%	1.24	0.21	Acima da Media
	5	81%	8%	12%	11%	1.43	0.36	Acima da Media
	6 ou mais pessoas	88%	4%	10%	9%	2.44	0.89	Muito Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Quantidade de moradores na família								
V55	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	QUANTIDADE_FAMILIAS (na mesma UC)	DIFICULDADE	NÃO	SIM	Total		WOE	
	1	75%	97%	94%	95%	0.97	0.03	Media
	2 ou mais	84%	3%	6%	5%	1.72	0.54	Bem Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Indica se existe mais de uma família morando no mesmo domicílio								
V56	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	STATUS_REDE_ENSINO	DIFICULDADE	NÃO	SIM	Total		WOE	
	NINGUEM NO DOM CRECHE OU ESCOLA	71%	51%	40%	43%	0.79	- 0.23	Abaixada Media
	PARCIALMENTE NA REDE PARTICULAR	75%	5%	5%	5%	0.98	0.02	Media
	TODOS NA REDE PARTICULAR	60%	19%	9%	12%	0.49	0.71	Muito Abaixo da Media
	TODOS NA REDE PUBLICA	84%	25%	46%	41%	1.80	0.59	Bem Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
Situação da rede de ensino utilizada pela família : escolas e creches, particulares e públicas								

Quadro 11 – Quantidade de pessoas na família, ciclo de vida familiar e educação.

4.4.12 Acesso aos serviços básicos

O acesso aos serviços básicos de caráter público está identificado dentro deste macro grupo de variáveis, refletindo que muitos no Brasil ainda sofrem com este tipo de carência. Estas carências estão presentes mais nas áreas rurais e nas periferias das regiões metropolitanas.

	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
			DIFICULDADE	NÃO				
V57	FATOR_13_ESGOTO_ESCOADOURO				Total	SIM / NÃO	WOE	DE DIFICULDADE
	RUIIM_NAO_TEM	81%	30%	43%	40%	1.44	0.36	Acima da Media
	SIM_BOM	71%	70%	57%	60%	0.81	0.21	Abaixoda Media
		75%	100%	100%	100%	1.00	0.00	Media
Presença de escoadouro sanitário_Presença de agua canalizada_Tipo de serviço de esgotamento sanitário disponível no domicilio								
V58	FATOR_23_LIMPEZA_MANUTENCAO_PAVIMENTACAO				Total	SIM / NÃO	WOE	DE DIFICULDADE
	NAO	83%	17%	29%	26%	1.67	0.51	Bem Acima da Media
	SIM	72%	83%	71%	74%	0.86	0.15	Abaixoda Media
		75%	100%	100%	100%	1.00	0.00	Media
Acesso a serviços de limpeza e manutenção de ruas_Indica sobre a pavimentação da rua onde mora								
V59	FATOR_3_ACESSO_A_AGUA_RURAL				Total	SIM / NÃO	WOE	DE DIFICULDADE
	RURAL_COM_ACESSO_A_AGUA	78%	8%	9%	9%	1.20	0.18	Acima da Media
	RURAL_COM_DIFICULDADE_DE_ACESSO_A_AGUA	84%	5%	8%	7%	1.68	0.52	Bem Acima da Media
	URBANO	74%	88%	83%	84%	0.95	0.05	Media
		75%	100%	100%	100%	1.00	-	Media
Proveniência da agua utilizada no domicilio_Tipo de serviço de agua disponível para o domicilio_Indica se urbano ou Rural								
V60	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO				Total	SIM / NÃO	WOE	DE DIFICULDADE
	NAO	80%	35%	47%	44%	1.34	0.29	Acima da Media
	SIM	71%	65%	53%	56%	0.82	0.20	Abaixoda Media
		75%	100%	100%	100%	1.00	0.00	Media
Acesso a serviços de coleta de lixo_Acesso a serviços de fornecimento de energia elétrica_Acesso a serviços de fornecimento de iluminação de rua								
V61	INUNDACOES				Total	SIM / NÃO	WOE	DE DIFICULDADE
	NAO	74%	93%	89%	90%	0.96	0.04	Media
	SIM	83%	7%	11%	10%	1.62	0.48	Bem Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa apresenta inundações no local onde mora								
V62	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA				Total	SIM / NÃO	WOE	DE DIFICULDADE
	BOM	71%	71%	59%	62%	0.83	0.19	Abaixoda Media
	RUIIM_NAO_TEM_NR	81%	29%	41%	38%	1.41	0.35	Acima da Media
		75%	100%	100%	100%	1.00	0.00	Media
A pessoa apresentou sua percepção sobre escoamento de agua no local onde mora								
V63	TIPO_AREA				Total	SIM / NÃO	WOE	DE DIFICULDADE
	CAPITAL	70%	29%	22%	24%	0.76	0.28	Abaixoda Media
	RESTO DA RM	79%	11%	14%	13%	1.28	0.24	Acima da Media
	RESTO DA UF	75%	48%	48%	48%	0.99	0.01	Media
	RURAL	81%	12%	17%	16%	1.38	0.33	Acima da Media
		75%	100%	100%	100%	1.00	-	Media
Tipo da área onde mora								

Quadro 12 – Acesso aos serviços básicos

4.4.13 Renda de Transferências do Governo

Como se pode naturalmente esperar, dos 22% de famílias que recebem a transferências do governo, 88% declaram viver com dificuldade (V64). Neste sentido os programas de transferência são paliativos, não resolvem o problema de todos que estão com dificuldade, pois foca-se na miséria. Observando o Bolsa Família, foi informado em reportagem de Máximo (2011) que mais de 2,2 milhões de famílias tiveram suas transferências canceladas por aumento da renda per capita desde que o programa foi criado em 2003 (com este nome). Tal resultado deve ser analisado com cautela, como afirma Tiago Falcão, secretário nacional de

renda de cidadania do Ministério do Desenvolvimento Social, pois não avalia que existem famílias entrando e saindo do bolsa família. Segundo ele os dados mostram que a quantidade de famílias beneficiada é muito menor que o contingente que vive em condições precárias.

A condição duradoura de emprego não é realidade para as camadas mais vulneráveis da sociedade (V3, apresentada no item 4.4.1).

V64	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO DE DIFICULDADE
	X_RENDA_TRANSF_GOV	DIFICULDADE	NÃO	SIM	Total	SIM / NÃO	WOE	
	NAO	72%	89%	74%	78%	0.83	-	0.19
	SIM	88%	11%	26%	22%	2.42	0.88	Muito Acima da Media
		75%	100%	100%	100%	1.00	-	Media

Indica se no domicílio de renda de transferências do governo

Quadro 13 – Transferências do Governo

4.4.14 Raça, cor da pele e região geográfica.

A “Raça ou Cor” na POF é uma característica declarada pelas pessoas de acordo com as seguintes opções: branca, preta, amarela, parda ou indígena. As categorias coletadas para esta variável foram reduzidas após o processo de fusão apresentado em 4.3.1, e ficou representada apenas por duas categorias (V66) : (1) brancos, amarelos e os que não sabem sua cor e (2) negros/pretos, pardos e indígenas, ambos com aproximadamente 50% da população. Neste trabalho utiliza-se a terminologia negra/negro ou preta/preto de forma intercambiada, o IBGE utiliza apenas a terminologia “preta/preto” para identificar a cor da pele.

A cor da pele esteve sempre ligada às Unidades da Federação neste estudo, considerando a entropia entre estas variáveis. Nos Estados do Norte e Nordeste o número de pessoas declaradas como pardas, pretas/negras e indígenas é bem maior do que no Sul. Neste sentido, quanto da “dificuldade” percebida pelas famílias tem origem na região onde moram ou na raça ou cor que têm na pele? Assim, decidiu-se manter estas variáveis em separado, e não criar um “Grupo” que as resuma, por mais que a análise da entropia tenha sugerido este agrupamento. Estas variáveis têm o potencial de representar duas situações relacionadas, mas que podem ser estudadas também em separado, através de simulações possíveis nas Redes Bayesianas: (1) a cor da pele e a herança de dificuldade originada desde a época da colonização (índios), da escravidão (negros/pretos e índios) e agravada com vinda dos imigrantes europeus brancos, mais preparados para o trabalho livre. (2) o desenvolvimento ou (falta) de desenvolvimento regional, associado às políticas e economias locais.

Vale lembrar, que não existe necessariamente nenhuma fronteira regional Levy (1993, p. 143), o que existe são gradientes e descontinuidades, o recorte é fruto de um ou mais critérios que separam o que está dentro do que está fora da região. A escolha dos critérios de divisão são convenções, fruto da história e de circunstâncias específicas. De acordo com Pozenato (2011), mesmo que as fronteiras não possam ser nitidamente definidas, elas funcionam como um traço simbólico de separação e de exclusão, pois estando “limitada” dentro de fronteiras, a região é um espaço periférico com relação ao centro, sendo este o polo que hierarquiza o acesso aos recursos socioeconômicos. Ao redor do centro gravita a periferia, o interior, a província, cada um com seu significado.

Toma-se como exemplo o Maranhão: que de acordo com Furtado (2000, p. 68) foi criado com objetivos políticos no século XVII como uma colônia de povoamento. Esta Unidade da Federação parece não ter encontrado um caminho de desenvolvimento adequado: seus solos, segundo o autor, não apresentavam a mesma fecundidade que o massapê nordestino para a produção do açúcar. Depois, ficaram isolados pela ocupação holandesa em Pernambuco, permaneceram abandonados pelo governo português nos séculos XVII. No século XVIII tentaram sobreviver como exportadores de produtos florestais, baseados na escravização de indígenas, tiveram como destino abraçar a cultura de subsistência e adentraram ao século XXI ainda com muita precariedade.

Ou seja, há muito os problemas persistem no Maranhão: o baixo desenvolvimento, condições precárias da população e o trabalho escravo, recentemente noticiado na mídia em geral. Pode-se observar no Gráfico 9, já apresentado, que o Maranhão tem o menor IDH brasileiro, ou seja as adversidades econômicas e descaso político vêm de muito tempo no local. Assim, a “dificuldade” de hoje tem causas históricas e complexas não passíveis de isolamento perfeito para cálculo do seu efeito sobre a percepção subjetiva de “dificuldade”. No entanto, no limite, esta perspectiva geográfica revela que em alguns lugares as pessoas nascem e crescem excluídas (NASCIMENTO, 2000). Infelizmente não é possível no modelo, por limitações na complexidade da técnica, manter cada Unidade da Federação em aberto, conforme apresentado anteriormente na Tabela 7.

Assim, a variável relativa às Unidades da Federação agrupadas pela “dificuldade” percebida por seus moradores, guardam aspectos regionais determinados pela geografia, história, economia e política e serão mantidas até o final da modelagem separadas da raça ou cor, assim será possível observar o papel de cada uma no modelo.

Fernandes (2008) relata os legados, as pendências e as carências dos descendentes negros libertos e descartados com a vinda dos imigrantes europeus.

Os negros/pretos, pardos e indígenas são hoje a maioria nas classes mais baixas da população e compartilham esta posição com pessoas brancas ou que não sabem sua cor. Estas provavelmente desconhecem sua história, como questionado no capítulo 5.

A POF traz uma grande variedade de religiões, neste sentido foi necessário agrupá-las com base na dificuldade percebida pelas famílias. O agrupamento obtido pareceu muito ligado às classes econômicas. Por exemplo: os espíritas kardecistas, religiões orientais, judeus e muçulmanos são religiões ligadas às classes mais altas, os católicos são a maioria da população, especialmente em zonas rurais, mais tradicionais onde os pentecostais ainda não chegaram. Estes últimos concentram-se nas populações mais pobres de centros urbanos. Neste sentido, a variável religião acabou sendo deixada de lado dada sua relação com as classes econômicas. Além disso, a POF não coleta se a pessoa ou família é praticante de religião, o que no nosso entendimento é diferente de apenas declarar uma religião. Assim achou-se por bem não mais considerar esta variável no estudo uma vez que demandaria investigações adicionais.

	EVIDÊNCIA	%	PERC_DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA PERCEPÇÃO
			DIFICULDADE	NÃO				
V65	GRUPO DE RELIGIOES				Total	SIM / NÃO	WOE	DE DIFICULDADE
	SEM INFORMACAO	83%	6%	9%	8%	1.59	0.46	Bem Acima da Media
	CATOLICA+OUTRAS	75%	86%	83%	84%	0.97	0.03	Media
	ESPIRITAS+INDIGENA+JUDAICAS+MUCULMANO+ORIENTAI	62%	4%	2%	2%	0.54	0.62	Bem Abaixo da Media
	NEOPENTECOSTAL_EVANG	79%	5%	6%	6%	1.22	0.20	Acima da Media
		75%	100%	100%	100%	1.00	-	Media
	Grupo de religião ao qual pertence o chefe da família							
V66	RACA_COR_PESSOA_REFERENCIA				Total	SIM / NÃO	WOE	IMPACTO NA PERCEPÇÃO
	_AMARELA+BRANCA+NAO SABE	69%	62%	46%	50%	0.74	0.31	Bem Abaixo da Media
	_INDIGENA+PARDA+PRETA	81%	38%	54%	50%	1.44	0.36	Acima da Media
		75%	100%	100%	100%	1.00	-	Media
	Raça ou cor do chefe da família							
V67	REGIAO				Total	SIM / NÃO	WOE	IMPACTO NA PERCEPÇÃO
	00 'CENTRO OESTE'	72%	8%	7%	8%	0.86	0.15	Abaixoda Media
	00 'NORDESTE'	82%	19%	29%	26%	1.52	0.42	Bem Acima da Media
	00 'NORTE'	78%	6%	7%	7%	1.15	0.14	Media
	00 'SUDESTE'	74%	47%	43%	44%	0.92	0.08	Media
	00 'SUL'	68%	20%	14%	15%	0.70	0.36	Bem Abaixo da Media
		75%	100%	100%	100%	1.00	0.00	Media
	Região Geográfica							
V70	UF				Total	SIM / NÃO	WOE	IMPACTO NA PERCEPÇÃO
	MA_PI_AL_SE	84%	4%	8%	7%	1.79	0.58	Bem Acima da Media
	BA_PE_CE_PA_PB_RN_TO	81%	17%	25%	23%	1.44	0.36	Acima da Media
	GO_AM_AC_RR	75%	5%	5%	5%	1.00	0.00	Media
	SP_MG_RJ	74%	44%	41%	42%	0.93	0.07	Media
	PR_ES_MS_RO_AP	71%	11%	9%	10%	0.83	0.19	Abaixoda Media
	RS_SC_MT_DF	67%	17%	11%	13%	0.66	0.42	Bem Abaixo da Media
		75%	100%	100%	100%	1.00	0.00	Media
	Agrupamentos das Unidades da Federação considerando percepção média de dificuldade em cada uma delas							
V71	AREA_GEO				Total	SIM / NÃO	WOE	IMPACTO NA PERCEPÇÃO
	CAPITAL	70%	29%	22%	24%	0.76	0.28	Abaixoda Media
	INTERIOR	75%	48%	48%	48%	0.99	0.01	Media
	REG_METR	79%	11%	14%	13%	1.28	0.24	Acima da Media
	RURAL_COM_AGUA	78%	8%	9%	9%	1.20	0.18	Acima da Media
	RURAL_SEM_AGUA	84%	5%	8%	7%	1.68	0.52	Bem Acima da Media
	Total	75%	100%	100%	100%	1.00	-	Media
	Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água							

Quadro 14 – Cor da pele, religião e região geográfica

4.5 Visão Seleccionada da Floresta

Para compreender o problema em meio a tantas variáveis e aumentar a capacidade preditiva do modelo foi necessário utilizar o procedimento denominado Markov Blanket, conforme 3.4.1. Optou-se, então por : (1) diminuir o escopo de variáveis, a princípio ; (2) posteriormente incluir algumas outras que estejam fora do Markov Blanket mas que são consideradas como importantes para a compreensão do problema.

O Markov Blanket foi realizado em duas etapas. O Primeiro Markov Blanket visa avaliar quais variáveis estão diretamente relacionadas à “dificuldade”, segundo esta heurística computacional, e para o Segundo Markov Blanket refina-se esta análise excluindo algumas variáveis manualmente e processando novamente o Markov Blanket. Ao final desta etapa tem-se um conjunto de variáveis mínimo para a análise.

O Segundo Markov Blanket terá a função de identificar variáveis que possam substituir o valor da renda, por exemplo, e outras que forem consideradas como de maior dificuldade de obtenção em pesquisas primárias. Considera-se que as variáveis com o valor da renda e orçamento são as mais críticas para obtenção. O motivo é tentar chegar num conjunto de variáveis que possa ser coletado em pesquisas futuras e que não dependa exclusivamente da Pesquisa de Orçamentos Familiares.

Ao final desta etapa para mensurar eventual perda para o modelo, devido à exclusão de variáveis do primeiro Markov Blanket, utiliza-se a curva ROC para efeitos comparativos. Se a perda de discriminação for alta, não se pode prescindir das variáveis contendo valor da renda e orçamento. Entretanto, se o novo conjunto de dados apresentar discriminação boa ou equivalente, torna-se possível seguir com o modelo sem estas variáveis.

4.5.1 Primeiro Markov Blanket

Dando sequência ao estudo, as variáveis foram submetidas ao Markov Blanket. O objetivo desta primeira rodada é compreender as variáveis que diretamente influenciam o nó “dificuldade”.

Assim pretende-se observar o subconjunto selecionado e avaliar se alguma variável poderia ser excluída da análise visando simplificação. A Figura 35 e a Tabela 14 apresentam as variáveis selecionadas no Primeiro Markov Blanket (MB1).

Figura 35 -Primeiro Markov Blanket (MB1) -Variáveis Selecionadas

FONTE : DADOS DA POF 2009

Fator_17_CARTAO_OU_TCH_ESPECIAL	ACESSO_ FINANCEIRO	Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Cartão de Crédito e Cheque Especial
ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	ADIMPLÊNCIA	Declara já ter atrasado contas de água ou eletricidade ou gás.
ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	ADIMPLÊNCIA	Declara já ter atrasado outras contas relativas a bens ou serviços adquiridos
QUANTIDADE_DE_ALIMENTO_REC	ALIMENTAÇÃO	CONTINUA CONTINUAÇÃO A pessoa declarou sua percepção sobre a adequação da quantidade de alimentos e às necessidades familiares.
		TABELA 14 CONTINUA ...

TIPO_ALIMENTO_RAZAO_REC	ALIMENTAÇÃO	CONTINUAÇÃO A pessoa declarou sua percepção sobre adequação dos tipos de alimentos e a razão de não se alimentar como quer.
CLASSE_ABEP	CLASSE ECONOMICA	Classe ABEP - agregado da pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família e número de banheiros e empregados domésticos mensalistas.
FATOR_8_IDOSO_COM_APOSENTADORIA	FAMILIA	Indica se no domicilio tem idoso recebendo aposentadoria
CONDICOES_DE_MORADIA	MORADIA	A pessoa declarou sua percepção geral sobre a sua moradia
CASA_ESCURA	MORADIA	A pessoa declarou que percebe sua casa escura
FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	MORADIA	A pessoa declarou ter problemas na moradia
RENDA_MONET_PER_CAPITA	RENDA	Faixa de renda monetária per capita
RENDA_MONET_NECESS	RENDA	Quanto de renda necessita para viver até o final do mês em relação à renda que possui.
X_RENDA_TRANSF_GOVERNO		Indica se o domicílio recebe renda de transferências do governo
FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	SAUDE	Indica se a família necessitou de medicamentos nos últimos meses, mas não teve esta necessidade atendida.
TEM_PLANO_DE_SAUDE	SAUDE	Indica se algum morador possui plano de saúde
FATOR_11_EDUC_SAUDE_LAZER	SERVIÇOS PÚBLICOS	Percepção sobre o acesso aos serviços de educação, saúde e lazer.
INUNDACOES	SERVIÇOS PÚBLICOS	A pessoa declara ter inundações no local onde mora
SERVICO_DE_TRANSPORTE_COLETIVO	SERVIÇOS PÚBLICOS	Percepção sobre transporte coletivo no local onde mora
X_RENDA_EMPREGADO	RENDA	Indica se se no domicílio tem renda de empregado.

Tabela 14 - Primeiro Markov Blanket - Variáveis Seleccionadas - continuação

Pela mensuração da entropia cruzada, divergência de Kullback-Leibler, foi possível observar o percentual de contribuição de cada variável para a compreensão, ou mudança no conhecimento, sobre a “dificuldade”.

Gráfico 10 - Contribuição percentual das variáveis seleccionadas pelo Markov Blanket

FONTE : DADOS DA POF 2009

Através do Markov Blanket foi possível reduzir consideravelmente a quantidade de variáveis no modelo e em consequência a sua complexidade, como foi observado na Figura 34. No entanto, esta visão representa uma grande simplificação, pois as variáveis não são na realidade independentes como apresenta o modelo. Este procedimento é utilizado para compreender as relações mais fortes com a variável em estudo : “dificuldade” selecionando-as para dar continuidade à análise.

Pelo Gráfico 10, além da forte relação com renda e classes econômicas, observa-se que a “dificuldade” está relacionada a questões básicas como: alimentação, moradia, acesso à saúde e capacidade de pagamento de despesas essenciais como (água, luz e energia elétrica).

Este modelo (MB1) apresentou boa capacidade de discriminação: a AUROC (área sobre a curva ROC), possui valor igual a 0,84, indicando excelente discriminação. No entanto, como a configuração da rede é do tipo “naive”, o modelo, apesar da boa capacidade de discriminação, ainda precisa ser melhor trabalhado para representar a relação entre as variáveis seleccionadas. Assim, o objetivo desta fase foi apenas selecionar variáveis importantes, assegurar alguma acurácia para a classificação e dar subsídio para a compreensão das próximas fases.

4.5.2 Segundo Markov Blanket

Neste passo optou-se por excluir do universo de possibilidades as variáveis relacionadas ao valor da renda e todas as ligadas ao orçamento familiar, considerando a dificuldade de obtê-las em outras pesquisas e a inadequação verificada nas variáveis relacionadas ao orçamento familiar para analisar família por família. Vale lembrar que nenhuma variável do orçamento familiar foi selecionada pelo Primeiro Markov Blanket.

Com a exclusão das variáveis contendo informações que dependem do valor da renda : “Renda_Monetária_Per_Capita (V25) ” e “Renda_Monetária_Necessária (V26)” pode-se observar que a Classe Econômica, calculada com os dados da POF, têm o potencial para substituir as variáveis que usam valor da renda, retirada do modelo.

De acordo com a ABEP os pontos obtidos através do Critério Brasil têm alta correlação com a renda familiar, conforme demonstram em seus estudos esta correlação é de 0,79 (ABEP, 2008).

No presente trabalho os índices de correlação entre os “pontos calculados a partir dos dados da POF 2009” e a “Renda Monetária Familiar Mensal” também apurada na POF foram de (1) 0,56 em áreas fora da região de pesquisa da ABEP e (2) 0,60 para a mesma região de pesquisa da ABEP.

A diferença entre estes índices de correlação e o informado pela ABEP é esperada pois (1) as amostras utilizadas são diferentes, (2) os pontos foram calculados com base nos itens informados na POF e (3) a forma de apurar a renda na POF obedece metodologia distinta, englobando diferentes rendas.

Com a retirada das duas variáveis, que usam valor da renda apurada pela POF, espera-se também que, além de classe econômica, outras sejam adicionadas no novo processamento utilizando o Markov Blanket. Observa-se que após as exclusões, 12 novas variáveis entraram para o modelo. Na Figura 35 as variáveis mais ao centro, foram todas as que passaram a ser consideradas no Segundo Markov Blanket (MB2).

Figura 36 - Segundo Markov Blaket (MB2) com Variáveis Seleccionadas
após exclusão do valor da renda e outras de difícil obtenção em pesquisa primária

FONTE : DADOS DA POF 2009

O modelo torna-se mais complexo. Na sequência a Tabela 15 apresenta um breve descritivo das variáveis seleccionadas (MB2), além das que já estavam no modelo (MB1) e foram apresentadas na Tabela 15 abaixo.

FATOR_10_FAMILIAS	FAMILIA	Tipos de família considerando a idade do filho mais novo; número de dependentes de renda na família; quantidade de filhos na família.
QTDE_MORADORES	FAMILIA	Quantidade de moradores na família, limitada na última categoria a 5 ou mais pessoas.
STATUS_REDE_ENSINO	ESCOLARIDADE	Situação da rede de ensino utilizada pela família: escolas e creches, particulares e pública.
SERVICO_DE_ESCOAMENTO_DA_AGUA_D E_CHUVA	SERVIÇOS PÚBLICOS	A pessoa apresentou sua percepção sobre escoamento de agua no local onde mora
RACA_COR_PESSOA_REFERENCIA	RAÇA_COR	Raça ou cor declarada pela pessoa de referência no domicílio.
STATUS_CONJUGE	ESCOLARIDADE	Informa a escolaridade do cônjuge ou se o responsável não tem cônjuge.
PREVIDENCIA	TRABALHO_RENDA	Indica se alguém da família contribui para a previdência social ou se está aposentado
TEMPO_SEM_TRAB_PRINC_REC	TRABALHO_RENDA	Indica se o chefe da família sempre tem trabalho principal, quanto tempo no ano ficou sem rendimento deste trabalho ou se está aposentado.

CONTINUA

		CONTINUAÇÃO
X_RENDA_EMPREGADOR	REND A	Indica se no domicílio tem renda de empregador.
FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	SERVIÇOS PÚBLICOS	Acesso a serviços de coleta de lixo
X_RENDA_BENEFIC_ASSALARIADO	TRABALHO_RENDA	Indica se no domicílio há recebimento de benefícios típicos dos assalariados EX : Recebe décimo terceiro, bônus de férias, vale alimentação, transporte.
X_RENDA_PENSAO_MESADA_DOACAO	REND A	Indica se no domicílio existe renda de pensão, mesada ou doação

Tabela 15 - Variáveis Adicionais Seleccionadas no segundo Markov Blanket - após exclusão do valor da renda e outras
FONTE : DADOS DA POF 2009

Novamente, utilizando o procedimento da entropia cruzada (Kullback-Leibler) foi possível observar o percentual de contribuição de cada variável, da mesma forma que fora apresentado no primeiro Markov Blanket (MB1).

Gráfico 11 - Contribuição percentual das variáveis selecionadas pelo Segundo Markov Blanket

FONTE : DADOS DA POF 2009

Na Figura 35 (MB 2) foram destacadas duas flechas apontando para “dificuldade” a partir das variáveis : Classe_Abep e Quantidade de Moradores. A Classe_Abep pode ajudar a representar o valor da Renda Familiar Monetária, e em composição com Quantidade de Moradores pode ser uma *proxy* para a renda monetária familiar per capita.

Este fato leva à seguinte reflexão: ao aplicar o critério ABEP em pesquisas de mercado vale a pena captar também o número de pessoas nas famílias, pois este fato muda a forma como as unidades de consumo lidam com seu orçamento e renda. A “dificuldade” percebida pelas famílias tende a aumentar quanto pior a classe econômica. Entretanto, dentro das mesmas classes econômicas, a percepção de “dificuldade” varia de acordo com a quantidade de moradores. Por exemplo, analisando o Gráfico 12 percebe-se que: as unidades de consumo da classe C1 com 2 moradores, por exemplo, apresentam proporção de famílias com “dificuldade” na ordem de 64% , enquanto que as famílias com 5 moradores, nesta mesma classe, reportam que 80% estão com alguma “dificuldade”.

Gráfico 12 -% Com dificuldade; Classe Econômica; Quantidade de Pessoas na Família

FONTE : DADOS DA POF 2009

Para avaliar se houve perda da acurácia entre os modelos do Primeiro e Segundo Markov Blanket, calculou-se a Curva Roc, conforme item 3.4.3.

AUROC - AREA UNDER ROC CURVE

RESULTADOS DO TESTE	AUROC	Erro Padrão (a)	Sig Assintótica (b)	Assint.95% Intervalo de Confiança	
				Limite Inferior	Limite Superior
DIFICULDADE_MB1	.836	.002	.000	.832	.839
DIFICULDADE_MB2	.821	.002	.000	.817	.825

a. Pressuposto não paramétrico

b. H_0 : AUROC = 0.5

Tabela 16 - Comparação entre Markov Blanket (MB) 1 e 2

FONTE : DADOS POF 2009

Percebe-se pela AUROC, que houve alguma perda na capacidade preditiva, mas ambos os modelos apresentam discriminação excelente.. A AUROC nos dois casos é maior que 0,80 e a significância do teste permite dizer que existem evidências para rejeitar a hipótese de que os modelos não discriminam os dois grupos em estudo (dificuldade e facilidade).

Mesmo com a ligeira perda de discriminação no Segundo Markov Blanket (MB2) e aumento no número de variáveis, foi dada continuidade na modelagem com base nesta última seleção.

Na sequência apresenta-se a tabela 17 com as principais modificações ocorridas entre o Primeiro e o Segundo Markov Blanket, ressaltando que a variável “renda de empregado” que aparecia no Primeiro Markov Blanket foi eliminada. Ou seja, todas as demais permaneceram.

A Tabela 17 mostra a comparação das duas redes estudadas (MB1) e (MB2) a partir dos seus arcos comuns, invertidos e incluídos.

COMPARAÇÃO ENTRE AS REDES MB1 E MB2		
ARCOS COMUNS		
DIFICULDADE_REC	SERVICO DE TRANSPORTE COLETIVO	
DIFICULDADE_REC	CASA ESCURA	
DIFICULDADE_REC	INUNDACOES	
DIFICULDADE_REC	CONDICOES DE MORADIA	
DIFICULDADE_REC	ATRASO DE DESPESAS COM AGUA ELETRICIDADE E GAS	
DIFICULDADE_REC	ATRASO DE PRESTACOES DE BENS OU SERVICOS ADQUIRIDOS	
DIFICULDADE_REC	TEM PLANO DE SAUDE	
DIFICULDADE_REC	X RENDA TRANSF GOVERNO	
DIFICULDADE_REC	TIPO ALIMENTO RAZAO_REC	
DIFICULDADE_REC	QUANTIDADE DE ALIMENTO_REC	
DIFICULDADE_REC	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIR	
DIFICULDADE_REC	FATOR_11_EDUCAO_SAUDE_LAZER	
FATOR_8_IDOSO_COM_APOSENTADORIA	Fator_17_CARTAO_CONTA_CORRENTE	
DIFICULDADE_REC	Fator_17_CARTAO_CONTA_CORRENTE	
DIFICULDADE_REC	FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	
ARCOS INVERTIDOS		
CLASSE ABEP	DIFICULDADE_REC	
ARCOS INCLUÍDOS		
DIFICULDADE_REC	SERVICO DE ESCOAMENTO DA AGUA DE CHUVA	
DIFICULDADE_REC	X RENDA EMPREGADOR	
DIFICULDADE_REC	X RENDA BENEFIC ASSALARIADO	
DIFICULDADE_REC	X RENDA PENSAO MESADA DOACAO	
DIFICULDADE_REC	RACA COR PESSOA REFERENCIA	
QTDE_MORADORES	DIFICULDADE_REC	
DIFICULDADE_REC	STATUS CONJUGE	
DIFICULDADE_REC	STATUS REDE ENSINO	
DIFICULDADE_REC	TEMPO SEM TRAB PRINC_REC	
DIFICULDADE_REC	PREVIDENCIA	
DIFICULDADE_REC	FATOR_10_FAMILIAS	
DIFICULDADE_REC	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	
ARCOS REMOVIDOS		
X_RENDA_EMPREGADO	CLASSE ABEP	
DIFICULDADE_REC	RENDA_MONET_PER_CAPTA	
DIFICULDADE_REC	RENDA_MONET_NECESS	
ESTRUTURAS EM V COMUNS		
FATOR_8_IDOSO_COM_APOSENTADOR	Fator_17_CARTAO_CONTA_CORRENTE	DIFICULDADE_REC
ESTRUTURAS EM V ADICIONADAS		
CLASSE ABEP	DIFICULDADE_REC	QTDE_MORADORES
ESTRUTURAS EM V REMOVIDAS		
X_RENDA_EMPREGADO	CLASSE ABEP	DIFICULDADE_REC

OBS : estrutura em V consideram a relação entre 3 variáveis, d separadas.

Tabela 17 - Mudanças entre o Primeiro e Segundo Markov Blanket

FONTE : DADOS DA POF 2009

4.5.3 Variáveis incluídas além do Markov Blanket

Para dar continuidade à análise foram adicionadas, por conveniência analítica, algumas variáveis além das que foram obtidas pelo Segundo Markov Blanket :

Sexo da pessoa responsável pelo domicílio (SEXO) : dado a importância desta variável torna-se relevante sua avaliação no modelo final . O motivo dela não ter sido selecionada pelo Markov Blanket foi a relação com o indicador de recebimento de Pensão Mesada ou Doação.

Outras Rendas -(X_OUTRAS_RENDAS) : indicador da presença de outras rendas como aluguel, aplicações financeiras o que pode ser útil para discriminar a “dificuldade”/ “facilidade” entre as classes mais altas.

Grupos das Unidades de Federação (UF_GRUPO) esta variável não entrou no Markov Blanket por conta da forte relação com Raça ou Cor, no entanto será mantida no modelo final, visando compreender melhor a “dificuldade” nas diferentes localidades, sob esta variável escondem-se diferentes histórias, políticas sociais, econômicas, conforme já apresentado.

Áreas Geográficas (AREA_GEO) : Apresenta se o domicílio está em que tipo de área urbana ou rural, se na área rural verifica facilidade de acesso à água, esta variável complementa a variável relativa às Unidades da Federação, pois detalha o quão próximas as famílias estão do centro do poder e economia dentro de suas regiões.

4.5.4 Processo final de seleção de variáveis:

O tratamento realizado nos itens anteriores, principalmente através do Markov Blanket, tem como objetivo reduzir o escopo de variáveis de forma semi-controlada, usando a máquina e a compreensão analítica do pesquisador. Conforme alertam Jensen e Nielsen (2007, p. 240) ao se propor aprender uma estrutura através de experimentos com dados é importante estar atento para o princípio denominado “*Ockham’s razor*”, traduzido como a Navalha de Ockham, conhecido também como princípio da parcimônia (*Lex Parsimoniae*), já abordado anteriormente.

Então nesta fase, partindo das variáveis pré-selecionadas desenvolveu-se mais um “processo de navalha”, além da própria métrica MDL, conforme item 3.2.4. Ao final restaram as variáveis identificadas como focais para a compreensão da “dificuldade” e alguns de seus relacionamentos. A própria navalha de Ockham também deve ser utilizada com parcimônia: colocou-se como premissa que o modelo não perca significativamente sua capacidade discriminatória (considerando a estatística KS já observada no item anterior) e que o conjunto de dados faça sentido para a compreensão da realidade. Muitas vezes os métodos automáticos excluem variáveis que no entendimento do pesquisador devem permanecer no modelo. Ainda mais no caso das redes bayesianas onde é possível simular o efeito da variável de forma propagada

pela rede. Assim, mesmo que seja para mostrar que uma variável sozinha não é tão relevante para o modelo, ela pode ser importante do ponto de vista da compreensão dos fatos e na inter-relação com outros nós na rede. Tomando como exemplo a variável sexo do chefe da família : comprovadamente a mulher em outros estudos (SILVEIRA et al., 2007, p. 201) apresenta renda inferior. No entanto, em termos da percepção subjetiva sobre dificuldade, a diferença entre gêneros não parece tão significativa observando apenas esta variável no conjunto de dados aqui estudados. De qualquer forma é preciso seguir em frente com esta análise, pois é possível identificar na rede o que levaria à maior desigualdade entre gêneros.

Os próximos passos incluem as seguintes etapas : (4.5.5) agrupamento de variáveis para a seleção dos nós para o(s) modelo(s) finais (4.5.6) inclusão de arcos proibidos na heurística dos modelos visando facilitar a interpretação final.

4.5.5 Agrupamento de variáveis para a seleção dos nós do(s) modelo(s) finais

Com base nas 31 variáveis selecionadas anteriormente foi desenvolvido uma nova rede, que busca representar a dependência entre variáveis. Através dessa rede é possível observar o quanto complexa pode ser a relação entre as variáveis. Além disso, muitas outras configurações seriam possíveis para este conjunto de dados, considerando a quantidade de nós, parâmetros e diferentes heurísticas disponíveis para chegar numa estrutura plausível e representativa de algum conhecimento “aprendido” através dos dados.

O algoritmo utilizado inicialmente nesta fase foi o Naive Bayes Aumentada, que considera os nós filhos da “dificuldade” não independentes entre si . A relação entre as variáveis passa a ser melhor representada, impactando positivamente no cálculo das distribuições de probabilidade entre os nós.

Figura 37 -Rede Naive Aumentada com as 33 variáveis selecionadas após MB1 e MB2.

FONTES : DADOS POF 2009

Na representação da Figura 37 existe uma linha unindo todos os nós filhos à “dificuldade”. Apenas para efeito visual estas linhas estão expostas de forma mais delgada, evitando maior confusão com as demais setas.

Na sequência, as variáveis desta rede são apresentadas de acordo com a Informação Mútua, que mede o quanto o conhecimento ou observação de cada uma delas reduz a incerteza sobre a “dificuldade”. O Gráfico apresenta a posição relativa das variáveis, dado que todas as Informações Mútuas foram somadas e para cada uma foi calculado o % de contribuição em relação ao total.

Assim como já havia sido observado no cálculo da entropia cruzada, ou Kullback-Leibler, nos modelos MB1 e MB2, as variáveis que mais contribuem com Informação Mútua são as relativas às necessidades mais básicas como alimentação, moradia e acesso à saúde.

Quadro 15 – Variáveis que reduzem a incerteza sobre a dificuldade.

Estas informações têm relação umas com as outras, como pode ser observado pela Figura 37.

Visando simplificar a análise, conforme também fora feito no item 4.3.2, a rede foi submetida a um novo processo de agrupamento gerando 10 grupos distintos.

Figura 38 -Destaque dos 10 grupos identificados na Rede Naive Bayes Aumentada

Na Tabela 18 estão representadas as variáveis e o quanto cada uma delas contribui mutuamente com o seu respectivo grupo e a decisão de mantê-las ou não na análise.

SIGNIFICÂNCIA DO NÓ : GANHO DE INFORMAÇÃO EM CADA NOVO GRUPO DE VARIÁVEIS - 33 VARIÁVEIS ; 10 GRUPOS

NÓ OU VARIÁVEL	Informação		Variável Selecionada	MOTIVO DA SELEÇÃO	GRUPO
	Mútua	Relativa			
QUANTIDADE_DE_ALIMENTO_REC	0.40	1.00	SIM	TODAS FORAM MANTIDAS POIS REPRESENTAM SITUAÇÕES COMPLEMENTARES IMPORTANTES PARA O MODELO FINAL	0
TIPO_ALIMENTO_RAZAO_REC	0.33	0.82	SIM		
FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	0.13	0.33	SIM		
UF_GRUPO	0.07	0.18	SIM		
RACA_COR_PESSOA_REFERENCIA	0.07	0.18	SIM		
QTDE_MORADORES	0.52	1.00	SIM	TODAS FORAM MANTIDAS POIS REPRESENTAM SITUAÇÕES COMPLEMENTARES IMPORTANTES PARA O MODELO FINAL	1
FATOR_10_FAMILIAS	0.52	0.99	SIM		
STATUS_REDE_ENSINO	0.12	0.22	SIM		
X_RENDA_TRANSF_GOV	0.02	0.04	SIM		
X_OUTRAS_RENDAS	0.07	1.00	NÃO	EXCLUÍDAS - BAIXA REPRESENTATIVIDADE	2
X_RENDA_EMPREGADOR	0.01	0.10	NÃO		
FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	0.52	1.00	SIM	APENAS TRANSPORTE COLETIVO FOI EXCLUÍDO, APESAR DO SEU POTENCIAL PARA COMPREENDER ACESSO À EDUCAÇÃO E SAÚDE. DECISÃO PARA SIMPLIFICAÇÃO	3
AREA_GEO	0.18	0.35	SIM		
SERVICO_DE_TRANSPORTE_COLETIVO	0.09	0.18	NÃO		
FATOR_11_EDUC_SAÚDE_LAZER	0.03	0.05	SIM		
SEXO	0.49	1.00	SIM	MANTIDA APENAS A VARIÁVEL SEXO, PARA MELHOR INVESTIGAR NA REDE	4
STATUS_CONJUGE	0.46	0.94	NÃO		
X_RENDA_PENSAO_MESADA_DOACAO	0.04	0.08	NÃO	A VARIÁVEL MAIS RELEVANTE PARA REPRESENTAR AS CONDIÇÕES DA MORADIA FOI MANTIDA	5
FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DE	0.67	1.00	SIM		
CONDICOES_DE_MORADIA	0.09	0.14	NÃO		
CASA_ESCURA	0.07	0.10	NÃO		
PREVIDENCIA	1.12	1.00	SIM	A VARIÁVEL PREVIDÊNCIA REPRESENTA TANTO A FORMALIZAÇÃO DO TRABALHO QUANTO A PRESENÇA DE APOSENTADOS E ACESSO A RENDA DE EMPREGADO	6
TEMPO_SEM_TRAB_PRINC_REC	0.61	0.54	NÃO		
X_RENDA_EMPREGADO	0.43	0.38	NÃO		
X_RENDA_BENEFIC_ASSALARIADO	0.31	0.27	NÃO		
FATOR_8_IDOSO_COM_APOSENTADORIA	0.12	0.11	NÃO	APENAS A VARIÁVEL CARTÃO OU CHEQUE ESPECIAL ELIMINADA, DADO A SUA FORMA DE CAPTAÇÃO.	7
CLASSE_ABEP	0.42	1.00	SIM		
Fator_17_CARTAO_OU_TCH_ESPECIAL	0.31	0.74	NÃO		
TEM_PLANO_DE_SAUDE	0.27	0.65	SIM	SERVIÇO DE ESCOAMENTO PRIORIZADA PARA REPRESENTAR O FATOR	8
SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	0.69	1.00	SIM		
INUNDACOES	0.06	0.09	NÃO		
ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	0.47	1.00	SIM	AMBAS MANTIDAS	9
ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	0.35	0.75	SIM		

Tabela 18 -Variáveis agrupadas para suportar a seleção para os modelos finais.

A partir da análise dos grupos foram selecionadas as principais variáveis que constam no modelo final. A seleção não levou em conta somente os agrupamentos automáticos, eles facilitaram o processo. Mais uma vez, o procedimento automático de agrupamento de variáveis não se demonstrou satisfatório, do ponto de vista subjetivo, para substituir totalmente o conjunto de variáveis por apenas uma variável latente (representada pelo grupo). Ao analisar os grupos a conclusão é que muita informação estaria perdida em termos do detalhamento almejado. Assim, como critério geral procurou-se manter, pelo menos, uma variável de cada grupo: algumas variáveis foram selecionadas por conveniência analítica (exemplo da variável sexo), alguns grupos tiveram todas as suas variáveis mantidas para a análise final. Assim, o conjunto de variáveis selecionadas está representado na Tabela 19 e 20 e a partir deste conjunto serão desenvolvidas as representações finais do modelo.

Nó ou Variável - Seleção Final
QUANTIDADE_DE_ALIMENTO_REC
TIPO_ALIMENTO_RAZAO_REC
FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO
UF_GRUPO
RACA_COR_PESSOA_REFERENCIA
QTDE_MORADORES
FATOR_10_FAMILIAS
STATUS_REDE_ENSINO
X_RENDA_TRANSF_GOVERNO
FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO
AREA_GEO
FATOR_11_EDUCAO_SAUDE_LAZER
SEXO
FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA
PREVIDENCIA
CLASSE_ABEP
TEM_PLANO_DE_SAUDE
SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA
ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS
ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS

Tabela 19 -Variáveis Selecionadas para os modelos finais

Tabela 20 - Variáveis Seleccionadas para o Modelo Final – Probabilidades Marginais
FONTE : Dados POF 2009

4.5.6 Interferência na heurística dos modelos -arcos proibidos

Jensen e Nielsen (2007, p. 60 ; p. 247) ressaltam que a detecção de arcos e suas direções é uma tarefa difícil em Redes Bayesianas. Segundo o autor uma forma de aperfeiçoar o espaço de busca de estrutura é estabelecer algumas restrições aos arcos que podem ser incluídas como parte do algoritmo.

Em Redes Bayesianas é possível incluir crenças ou conhecimento do pesquisador para uso durante o aprendizado de uma estrutura. Ou seja, é possível interferência na heurística com-

putacional. Este procedimento diminui o espaço de busca para as redes e previne contra “arcos indesejáveis”, na concepção do pesquisador. Desta forma o modelo passa a considerar também em suas relações gráficas e métricas um conhecimento externo, além dos dados.

Jensen e Nielsen (2007, p.60) tomam como exemplo a dificuldade que se tem para estabelecer relações causais, que geralmente não são óbvias e nem bem compreendidas. O que se tem nas Redes Bayesianas geralmente não são relações causais, mas variáveis que ao serem conhecidas podem mudar o conhecimento sobre outras variáveis. No entanto, em alguns casos, o conhecimento causal pode estar na cabeça do pesquisador e pode facilitar o processo de busca incluindo esta informação como parte da heurística.

Neste trabalho foram tomadas algumas providências neste sentido. As variáveis foram classificadas em dois grupos e para estes grupos foi estabelecida uma regra que “proíbe” a troca das setas de forma livre durante o processo de aprendizado da estrutura.

Os grupos foram denominados como proxys de “causa” e “sintoma”. A regra informa que sintoma não deve apontar para causa. Exemplo : “dificuldade orçamentária” apontando para “quantidade de pessoas na família”, não é um arco permitido. Conforme comenta Becker (2008, p. 93) os “expedientes visuais” ordenam as variáveis num diagrama conectado por setas, o que dá a entender que existe uma sequência temporal, mas o tempo é apenas uma metáfora visual para poder representar um problema sob a óptica da pesquisa em curso.

Apesar do peso da palavra “causa” e do fato dela ser evitada por grande parte dos pesquisadores como anuncia Pearl (2009 p. 401) decidiu-se por correr o risco e , quem sabe, progredir a partir de uma maior compreensão da técnica proposta aplicada na busca de relações causais em problemas sociais. Não é escopo e nem possibilidade neste trabalho fazer um levantamento exaustivo de causas. Vale esclarecer que as variáveis identificadas para investigação foram denominadas como “*proxy*” de “causa” e “sintoma” . A Tabela 21 apresenta um quadro com as variáveis classificadas nos dois grupos e na sequência uma justificativa baseada “na lógica do pesquisador” para cada uma delas. Espera-se que a técnica seja mesmo promissora, na prática, para o entendimento de causas e que novos trabalhos possam apontar direções mais fortes neste sentido, pois no momento existe um otimismo que aqui é apresentado para encorajar outros pesquisadores na adoção desta linha de pesquisa. Becker (2008, p. 91), quando aborda a questão das causas, lembra as ideias de Thomas Kuhn, datada no início dos anos 70, que ressaltou que técnicas paradigmáticas nunca fazem o que prometem, deixando “anomalias terríveis na esteira do seu uso”, apresentam “falhas quanto a lógica que a embasa”, sendo assim “vulneráveis a ataques” que mostram que “fazem menos do que ale-

gam”. No entanto lembra que sem elas não podemos fazer coisa alguma. Neste sentido, nesta tese opta-se por incorrer em riscos.

NÓS DA REDE	CLASSES A PRIORI
UF_GRUPO	CAUSA
RACA_COR_PESSOA_REFERENCIA	CAUSA
QTDE_MORADORES	CAUSA
FATOR_10_FAMILIAS	CAUSA
FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	CAUSA
AREA_GEO	CAUSA
FATOR_11_EDUCAO_SAUDE_LAZER	CAUSA
SEXO	CAUSA
CLASSE_ABEP	CAUSA
SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	CAUSA
QUANTIDADE_DE_ALIMENTO_REC	SINTOMA
TIPO_ALIMENTO_RAZAO_REC	SINTOMA
FATOR_27_NECESITOU_SERV_SAUDE_OU_MEDICAMENTO	SINTOMA
STATUS_REDE_ENSINO	SINTOMA
X_RENDA_TRANSF_GOVERNO	SINTOMA
FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	SINTOMA
PREVIDENCIA	SINTOMA
TEM_PLANO_DE_SAUDE	SINTOMA
ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	SINTOMA
ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	SINTOMA

Tabela 21 - Causas e Sintomas, definidos a priori

FONTE : ABEP 2009

A Justificativa para esta classificação a priori é apresentada na sequência :

UF_GRUPO (CAUSA) : as Unidades da Federação foram utilizadas como construtos que representam o resultado agregado de políticas públicas, cujos aspectos históricos diferenciam as regiões e a economia. As Unidades da Federação foram agrupadas com base na percepção de dificuldade dos seus moradores, ou seja, pela percepção média dos que ali moram.

RACA_COR_PESSOA_REFERENCIA (CAUSA): a cor da pele, denominada no IBGE também como raça, foi identificada a priori como Causa, assim pode-se investigar o efeito desta variável na dificuldade simulando sua composição com as demais variáveis.

QTDE_MORADORES (CAUSA): a quantidade de moradores no domicílio notadamente parece ser uma causa de dificuldade, dado que interfere diretamente na renda per capita familiar.

FATOR_10_FAMILIAS (CAUSA) : A composição da família, especialmente com presença de crianças e jovens provoca maior restrição orçamentária, pois a renda além de ser dividida também com os dependentes inibi o trabalho feminino, o que reduz a renda familiar.

FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO (CAUSA): Este agrupamento de variáveis foi selecionado para oferecer mais um nível de informação, tendo o potencial de dividir as Unidades da Federação e Áreas Geográficas de acordo com o acesso aos serviços de caráter básico. Por exemplo: mora-se em São Paulo, na Região Metropolitana, mas o acesso a estes serviços é ruim ou praticamente não existe. Isso diz muito sobre a dificuldade em que se encontra uma família.

AREA_GEO (CAUSA) : A área geográfica ilustra a proximidade com a economia e com o poder. Quanto mais longe do poder, maior tende a ser a dificuldade. As famílias podem estar em zonas rurais ou em regiões periféricas das grandes metrópoles, o que as coloca em condição de exclusão ou até de “esquecimento”.

FATOR_11_EDUC_SAUDE_LAZER (CAUSA) : Este agrupamento de variáveis visa representar a qualidade do acesso aos serviços de educação, saúde e lazer, e assim como no caso dos serviços básicos confere mais detalhes sobre a situação interna das Unidades da Federação e Regiões Geográficas.

SEXO (CAUSA) : Esta variável será testada como causa e objetiva-se avaliar o real impacto no modelo a partir da composição com as demais variáveis.

CLASSE_ABEP (CAUSA) : Considerada como causa pois neste modelo está sendo usada para representar a renda familiar, seu significado mais direto envolve posse de bens, escolaridade do chefe da família , tamanho da casa / moradia, presença de serviços domésticos.

SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA (CAUSA) : Este agrupamento de variáveis visa esclarecer um pouco melhor a relação com as Unidades da Federação, ilustrando acesso a serviços públicos básicos.

QUANTIDADE_DE_ALIMENTO_REC (SINTOMA): Quando as coisas vão mal, a quantidade de alimento é restrita e a família necessita reduzir este item, sinal de que ela está na base da pirâmide de necessidades, o que tem grande efeito sobre a percepção de dificuldade.

TIPO_ALIMENTO_RAZAO_REC (SINTOMA) : Quando as coisas vão mal o tipo de alimentação também é restrita, no entanto este item pode gerar algumas dúvidas, pois os tipos de alimentação desejados podem ter caráter subjetivo.

FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO (SINTOMA) : As pessoas menos favorecidas estão geralmente longe dos aparatos da saúde e dos medicamentos e acabam sem acesso a estes serviços e a declaração dessa necessidade indica dificuldade.

STATUS_REDE_ENSINO (SINTOMA) : Quando as coisas vão bem as famílias matriculam seus filhos na escola particular, o contrário as leva para a escola pública. No entanto, manter filhos nas escolas particulares é tarefa que pode trazer “dificuldade”, neste sentido a causa da dificuldade estaria na estrutura familiar (com presença de filhos).

X_RENDA_TRANSF_GOVERNO (SINTOMA) : O recebimento de rendas de transferência tem foco nos mais necessitados, sendo uma consequência da dificuldade .

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA (SINTOMA) Quando as coisas vão bem as famílias mantêm suas casas em boas condições, quando não, todo o conjunto tende a se deteriorar. Este também pode ser um indicador de pobreza arraigada há muito tempo na família, a ponto da manutenção da moradia não ser prioridade.

PREVIDENCIA (SINTOMA) : As relações formais ou não com o trabalho foram consideradas como sintoma, especialmente quando se observa sua relação com a classe econômica. As pessoas menos favorecidas, têm menor instrução, menor poder de barganha e menor “capital humano”. Assim esta variável é consequência de algo maior, recebendo o status inicial de sintoma e não de causa. No entanto, no modelo, pode assumir também papel de causa em relação às outras variáveis, como posse de plano de saúde.

TEM_PLANO_DE_SAUDE (SINTOMA) : Em todo o processo de modelagem o plano de saúde esteve ligado à classe econômica e à renda, não houve sustentação para ligar esta variável à outra que trata da necessidade de serviços de saúde. Assim, é tratado como provável sintoma da renda e da formalização do trabalho. No entanto, no modelo esta variável pode assumir status de causa. O acesso a plano de saúde também pode indicar acesso à saúde como um todo, pois a questão da saúde no Brasil é grave e os planos mantêm suas carteiras entre os que têm dupla cobertura (pública e privada). Como bem apresenta Medici (2011, p. 23), os “mais ricos” têm mais informação, conhecimento e influência e assim conseguem vantagens no sistema de saúde público, passando menos dificuldade.

ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS (SINTOMA) : Quando o aperto orçamentário é grande as pessoas atrasam contas básicas, o que pode desencadear em outros processos de inadimplência, pagamento de juros e descontrole financeiro.

ATRASSO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS (SINTOMA):

Quando as coisas vão mal as pessoas tendem a atrasar prestações de serviços ou bens adquiridos, o que pode desencadear em outros processos de inadimplência, pagamento de juros e descontrole financeiro.

4.5.7 Interferência na heurística dos modelos - presença e sentido dos arcos

Em redes bayesianas, assim como é possível inibir alguns arcos durante o processo de aprendizado da estrutura é também possível adicionar, "deletar" ou inverter o sentido do arco após o "aprendizado". De acordo com Jensen e Nielsen (2007, p. 245) estas interferências dependem do julgamento de quem analisa a rede e somente podem ser aplicadas quando não ferem a regra básica que mantém a rede acíclica, característica básica de uma Rede Bayesiana.

Neste trabalho estes procedimentos são utilizados quando necessário e documentados para que se tenha controle do aprendizado descoberto automaticamente e das interferências do pesquisador.

4.6 Modelos Finais

A partir das variáveis selecionadas foram testadas as diversas formas de descoberta de estrutura. Ao final desta fase foram eleitos dois modelos: um que tem origem no conjunto que prioriza a classificação, conforme item 3.4.1 e outro que pertence à categoria que busca a relação entre as variáveis, sugerindo uma estrutura, conforme 3.4.2.

O primeiro critério de seleção destes modelos baseou-se nos resultados obtidos pela curva ROC e na inspeção visual da rede. As áreas sobre as curvas ROC (AUROC) foram similares entre os modelos, indicando que as diferentes heurísticas possuem boa capacidade de classificação entre "dificuldade" e "não dificuldade", confirmando a capacidade de classificação das Redes Bayesianas. Outro critério também observado foi o erro da estimativa de probabilidade em relação ao valor esperado, pois um modelo pode classificar bem um conjunto de dados, dado um ponto de corte, e apresentar valor de probabilidade que não reflete adequadamente a situação específica de cada uma das observações. Para a avaliação visual do erro as probabilidades calculadas foram dispostas em percentis, "ranking" de 1 em 1 %, onde a

probabilidade média do percentil estimado foi comparada com a proporção média de famílias com dificuldade (valor esperado). A diferença observada foi denominada “erro das estimativas por percentil”. Este procedimento permite observar vieses na estimativa da probabilidade. Vale esclarecer que modelos exclusivamente utilizando o Markov Blanket ou mesmo Naive Bayes, apesar de conferirem boa classificação, dado um limiar ou ponto de corte, apresentam problemas quanto à estimativa da probabilidade. Como exemplo, o Gráfico 13 apresenta um comparativo entre a proporção da “dificuldade” e a probabilidade calculada pelo modelo MB1 em cada percentil.

Gráfico 13 -Erro na Estimativa da Probabilidade, modelo MB1

FONTE : DADOS POF 2009

Os modelos selecionados na etapa final devem minimizar “o erro” na estimativa da probabilidade.

Os melhores modelos considerando a curva ROC e a análise “do erro” (como no Gráfico 13), foram também as melhores redes em termos de apresentação visual.

Na sequência são apresentados os dois modelos selecionados sendo que um deles foi o escolhido como espaço de reflexão sobre a dificuldade orçamentária das famílias brasileiras.

Os Gráficos 14 e 15 permitem observar comparativamente as redes selecionadas. A que representa o conjunto de modelos focados em classificação é Naive Bayes Aumentada e a que busca a melhor relação entre as variáveis foi a rede baseada na heurística denominada “Tabu

Order”. Durante o processo de modelagem focou-se na possibilidade de reunir os dois atributos à modelagem: capacidade de classificação e capacidade de representação visual do problema.

Gráfico 14 - Rede Naive Bayes Aumentada (ANB)

Gráfico 15 - Rede Tabu Order (TO)

Deve-se esclarecer que os grafos apresentados foram obtidos por processos automáticos e que após estes processos houve interferência do pesquisador quanto ao sentido de algumas

setas. A ideia inicial foi a de mexer o menos possível nas redes originalmente “descobertas”. Assim em ambas as redes foram feitas as seguintes adequações/inversões de alguns arcos.

[RAÇA → CLASSE]; [SEXO → CLASSE], pois não faz sentido a raça e o sexo antecederem classe. Além disso, apenas para a rede “Tabu Order” (TO) foi promovida a seguinte adequação: [QTDE_MORADORES → DIFICULDADE], pois não faz sentido a dificuldade anteceder a quantidade de moradores.

Não foi possível modificar a seta :

[ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_GAS] por conta da regra que mantém a rede acíclica, esta variável será tratada nos próximos passos.

Estas últimas modificações não foram implementadas na Naive Bayes Aumentada (ANB), dado sua configuração : todas as variáveis têm como o nó pai a variável “dificuldade”.

O Gráfico 16 apresenta as principais diferenças entre as duas estruturas quanto à disposição das setas

Gráfico 16 -Comparativo entre as redes : ANB (Naive Bayes Aumentada) e TO (Tabu Order)

Tomando-se como base a Rede ANB e comparando-a com a TO verifica-se : em vermelho - os arcos que foram excluídos ; em azul - os arcos que foram adicionados ; em preto - os arcos que foram mantidos. As representações circulares em torno de alguns nós representam

as estruturas em V (d-separações e conexões conforme apresentado em 3.2.3) que foram adicionadas de um modelo para o outro. Percebe-se que o modelo “Tabu Order” incluiu mais estruturas em V do que o ANB.

A principal diferença entre as estruturas das redes é que na Rede Naive Aumentada existem arcos apontando da variável dificuldade para todas as variáveis, pois este é um pressuposto da técnica, o que pode não ser muito realista. No entanto, se pensar que a dificuldade emerge de forma complexa de todos estes fatores esta pode ser uma boa representação visual ou uma abstração. No entanto no Gráfico 15 percebe-se uma rede “mais provável”, onde algumas setas de forma prioritária apontam como “causa” da dificuldade.

Na sequência são apresentadas as curvas ROC de forma comparativa para ambos os modelos.

Gráfico 17 - Curva ROC : modelos ANB (Naive Bayes Aumentada) E TO (Tabu Order)

AUROC - Área Sobre a Curva ROC					
Resultado dos Testes	Área	Erro Padrão ^a	Sig. Assintótica ^b	95% Intervalo de Confiança	
				Limite Inferior	Limite Superior
TABU_ORDER_PROB (TO)	,841	,002	,000	,837	,845
NAIVE BAYES AUMENTADA (ANB)	,843	,002	,000	,839	,846

a. Pressuposto não paramétrico

b. Hipótese Nula : Área = 0.5

Tabela 22 - Área sobre a curva ROC - modelos ANB e TO

Percebe-se que os dois modelos possuem AUROC em torno de 0,84, podendo ser considerados modelos excelentes do ponto de vista da capacidade de classificação, de acordo com o apresentado no item 3.4.3.

Avaliando o “ranking” de probabilidade por percentil, assim como fora feito no Gráfico 13, compara-se a dificuldade esperada em cada percentil com as probabilidades auferidas em

cada um dos modelos analisados. Pode-se perceber que os modelos ANB e TO encontram-se mais aderentes ao valor esperado do que o que fora apresentado no Gráfico 13. No entanto, apesar de ter diminuído o erro entre o esperado e probabilidade média, ambos continuam subestimar a dificuldade nos menores percentis. Mais adiante a informação sobre o este “erro” será novamente tratada. Neste momento esta análise é mais um subsídio para a escolha do modelo final que será utilizado para concluir a reflexão sobre a “dificuldade”.

Gráfico 18 -Tabu Order - Estimativa por percentil de probabilidade

Gráfico 19 - ANB - Estimativa por percentil de probabilidade

Ainda avaliando a classificação dos modelos foi preparado o Gráfico 20 que mostra a relação entre a probabilidade calculada pelos diferentes modelos apresentados e a variável original, estabelecida a partir da pergunta : “A renda total da sua família permite que você(s) leve(m) a vida até o final do mês com : Muita Dificuldade, Dificuldade, Alguma Dificuldade, Facilidade; Muita Facilidade”.

Gráfico 20 -Comparativo entre probabilidades estimadas pelos modelos

FONTE : DADOS POF 2009

Percebe-se que os modelos MB1 e MB2 conferem maior diferença entre a primeira e a última classe de probabilidade, no entanto os erros em média são maiores conforme apresenta o Gráfico 13. Os modelos MB1 e MB2 foram apresentados no Gráfico 20 apenas para que se compreenda o efeito da premissa de independência dos parâmetros sobre o computo final da probabilidade.

Observando os diferentes aspectos apresentados até o momento pode-se dizer que os modelos ANB e TO são similares quanto à capacidade de classificação. De forma “seca”, usando o valor da probabilidade igual a 0,75, como suposto corte ou limiar para classificar os que declaram viver em “dificuldade”, observa-se que os dois acertam em 77% dos casos, conforme Tabela 23.

MODELO TABOO ORDER (TO)		"DIFICULDADE" DECLARADA		Total
		NÃO	SIM	
"DIFICULDADE" MODELADA				
QUANTIDADES	NÃO	11,277,554	10,244,473	21,522,027
	SIM	2,861,476	32,506,389	35,367,865
	TOTAL	14,139,030	42,750,862	56,889,892
CONFIABILIDADE	NÃO	52%	48%	100%
	SIM	8%	92%	100%
	TOTAL	25%	75%	100%
PRECISÃO	NÃO	80%	24%	38%
	SIM	20%	76%	62%
	TOTAL	100%	100%	100%

MODELO NAIVE BAYES AUMENTADO (ANB)		"DIFICULDADE" DECLARADA		Total
		NÃO	SIM	
"DIFICULDADE" MODELADA				
QUANTIDADES	NÃO	11,416,519	10,525,357	21,941,876
	SIM	2,722,511	32,225,506	34,948,017
	TOTAL	14,139,030	42,750,863	56,889,893
CONFIABILIDADE	NÃO	52%	48%	100%
	SIM	8%	92%	100%
	TOTAL	25%	75%	100%
PRECISÃO	NÃO	81%	25%	39%
	SIM	19%	75%	61%
	TOTAL	100%	100%	100%

* SUPONDO O LIMAR DE PROBABILIDADE = A 0.75
 ACERTO TOTAL = (VP + VN) / TOTAL = 77%
 VP = SIM E SIM ; VN = NÃO ENÃO

Tabela 23 -Acerto Geral dos Modelos ANB e TO

FONTE : DADOS PNAD 2009

A precisão entre os que não declararam “dificuldade” está é em torno de 80% e para os que declararam “dificuldade”, 76%. A Confiabilidade das estimativas quando o modelo classifica dificuldade como “não” é de 52% e de 92% quando classifica como “sim”. Estes indicadores são os mesmos para os dois modelos em foco.

A Tabela 24 apresenta o resumo dos indicadores selecionados para tomar a decisão entre um ou outro modelo para ser usado como espaço de reflexão sobre a dificuldade das famílias brasileiras.

CRITÉRIO PARA ESCOLHA DE UM DOS MODELOS	TO	ANB
AUROC - ÁREA SOBRE A CURVA ROC	0.84	0.84
ACERTO GERAL	77%	77%
SUBESTIMA DIFICULDADE NOS MENORES PERCENTIS	MENOS	MAIS
O SENTIDO DAS FLECHAS APARENTA SER MAIS PROVÁVEIS (VISUALMENTE)	MAIS	MENOS
MODELO ELEITO COMO ESPAÇO MÍNIMO PARA A REFLEXÃO	SIM	NÃO
TO : TABU ORDER ; ANB : NAIVE BAYES AUMENTADA		

Tabela 24 - Tabela com critério de escolha entre TO e ANB

FONTE : DADOS POF 2009

É importante salientar que a expectativa inicial, neste trabalho, era poder contar com muitas outras variáveis no modelo final, no entanto, a luta contra os princípios de Ockham's Razor foram inglórias. A parcimônia é algo sempre lembrado na teoria, no entanto neste trabalho este aspecto foi aprendido a duras penas e o modelo final deve então ser classificado como um espaço mínimo para a reflexão

4.7 Modelo Final Selecionado e proposição ajustes

A heurística computacional selecionada, “Tabu Order” (TO), foi a mais adequada às redes bayesianas estudadas, dado o objetivo de descobrir a relação entre variáveis e não só agir buscando a melhor classificação entre “dificuldade” e “não dificuldade”. O processo de modelagem com o algoritmo TO não age de forma aleatória dentro do espaço de busca de uma estrutura. Ele considera uma variável de memória (lista interna) que armazena as decisões de escolha já realizadas durante o aprendizado para poder dar os próximos passos e descobrir o melhor nó de ligação, passo a passo. Este fato é interessante, pois ele não busca apenas encontrar um nó de ligação entre outros nós e assim sucessivamente, o “Tabu Order” possui um processo de busca mais completo. É necessário informar ao sistema o parâmetro “tamanho da lista”, que são os passos anteriores que devem ser memorizados para tomar a decisão do próximo nó de ligação. Na rede selecionada foi fornecido parâmetro igual a 10 para a lista de memória. Foram testados 15 e 20, no entanto os resultados foram os mesmos. O “Tabu Order” implementado no software utilizado (Bayesia) considera os arcos proibidos, complementando assim a heurística do modelo com conhecimento externo do pesquisador.

Para a exploração do modelo selecionado foi utilizado um “play ground” de variáveis. O termo “play ground” é inspirado no projeto de software on line denominado B-Course (2012). Realmente esta metáfora (“play ground”) expressa o sentimento que dá ao mexer (brincar) com todas as variáveis de forma interativa. Percebe-se que esta é uma preocupação dos softwares que tratam com redes bayesianas, pois a simulação e interação homem-máquina fazem parte do processo de escolha da rede, da aceitação dela como um modelo possível, do processo de aprendizado e exploração dos dados e também da descoberta de fatos escondidos ou não esperados.

Por exemplo: ao observar a rede no Gráfico16 percebe-se que existe um arco apontando da AREA_GEO para SEXO, num primeiro momento isso provoca dúvida ou curiosidade, então entra-se no “playground”, seleciona-se as variáveis em questão, investiga-se de forma visual e rapidamente o cérebro capta o sentido do que vê.

Figura 39 - Área Geográfica e Sexo do Chefe da Família

FONTE : DADOS POF -2009

Ao selecionar, conforme Figura 39, as barras “RURAL_COM_AGUA” e depois “RURAL_SEM_AGUA” percebe-se que mulher “responsável pela família” é um fenômeno urbano. Se em média 69% dos homens são chefes de família, quando a área é rural este percentual sobe para 80%, pois as relações familiares são mais tradicionais e o ambiente e trabalho são menos adequados à mulher que vive sozinha, ou seja, o ambiente parece condicionar o sexo do chefe da família. Na amostra existem 6900 casos de “RURAL_COM_AGUA” e 6400 casos de “RURAL_SEM_AGUA”, que depois de ponderados com o fator de expansão fornecido pela POF permitem chegar à distribuição apresentadas nas Figuras 39.

Figura 40 - Area Rural e Sexo do Chefe da Família

FONTE : DADOS POF -2009

O objetivo deste exemplo é mostrar que a rede passou por um processo de validação lógica, e quantitativa. A validação quantitativa visa verificar se as principais probabilidades conjuntas obtidas na rede estão próximas das observadas no banco de dados. Por exemplo: no “play ground” de variáveis ao se observar que a percepção sobre o acesso à educação, saúde e lazer é muito parecida em todas as classes econômicas, desconfia-se desta informação. Ela é improvável, pois não se espera que a Classe A1 tenha a mesma percepção que a Classe D sobre o assunto, conforme pode ser observado na Figura 41. Percebe-se que as relações entre as variáveis podem não ter sido construídas de forma adequada pelos processos automáticos.

Figura 41 - Análise da relação entre variáveis - identificação de problemas nas distribuições conjuntas

FONTE : DADOS POF -2009

Para checar, processa-se uma “query” diretamente na base de dados para constatar se a estimativa acima está mesmo diferente do esperado. A solução é tentar propor uma nova configuração para a rede, não detectada pela heurística computacional. Assim torna-se necessário propor nova estrutura para o modelo “Tabu Order” (TO), intervindo no modelo manualmente. Antes de apresentar a nova estrutura proposta apresenta-se como ficou a nova estimativa para o Fator 11 . Agora parece mais adequado, como mostra a Figura 42.

Figura 42 -Melhor representação das distribuições conjuntas após ajuste estrutural na rede

FONTE : DADOS POF -2009

Este tipo de intervenção é necessária, pois as probabilidades estimadas de forma conjunta (e seus erros) se propagam pela rede, dado a ligação dos arcos. O processo deve ser criterioso.

Gráfico 21 - Efeito do ajuste estrutural nas proporções conjuntas

FONTE : DADOS POF -2009

No Gráfico_21 é apresentado o efeito do ajuste na estrutura da rede em relação à estimativa de perceber a educação, saúde e lazer como ruins. A maior diferença foi para as classes A1 e B1.

No entanto, antes de aplicar este critério “manual” foi importante encontrar a heurística computacional que minimizasse este trabalho de ajustes finais, pois os ajustes manuais são trabalhosos e nem sempre possíveis, pois a rede deve manter-se acíclica e cuidar desta tarefa pode

ser muito trabalhoso. Fora isso, os algoritmos cuidam para que uma consistência interna entre mínimos e máximos locais sejam tratados e o ser humano não tem como fazer isso.

Outras variáveis passaram pelo mesmo processo de validação e novos ajustes foram propostos ao modelo previamente apresentado. As alterações na estrutura da rede estão apresentadas na Tabela 25 e representadas no Gráfico 22 sobre a rede, indicadas com arcos coloridos. Os azuis são os arcos adicionados e os laranja/âmbar são os arcos invertidos.

Na sequência serão apresentadas as justificativas para as modificações sugeridas nos arcos :

GRUPOS		ARCOS INVERTIDOS	
1	FATOR_10_FAMILIAS	QTDE_MORADORES	
2	TEM_PLANO_DE_SAUDE	DIFICULDADE_REC	
	AREA_GEO	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	
3	AREA_GEO	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	
	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	
4	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	
	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	
5	DIFICULDADE_REC	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	
6	STATUS_REDE_ENSINO	X_RENDA_TRANSF_GOVERNO	
		ARCOS ADICIONADOS	
3	AREA_GEO	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	
	UF_GRUPO	FATOR_11_EDUC_SAUDE_LAZER	
7	CLASSE_ABEP	FATOR_11_EDUC_SAUDE_LAZER	

Tabela 25 -Arcos Invertidos e Arcos Adicionados

FONTE : DADOS POF -2009

Gráfico 22 - Ajustes no modelo selecionado -Arcos Adicionados e Removidos.

FONTE : DADOS POF 2009

O primeiro arco invertido foi o $FATOR_{10_FAMILIAS} \rightarrow QTDE_MORADORES$. Com esta nova configuração a variável que apresenta a composição familiar passa a influenciar a quantidade de moradores e não o inverso. Considerando as D-Separações e Tipos de Conexão em Redes Bayesianas, conforme apresentado no item 3.2.3, Figura 2, percebe-se que a junção sugerida para $[FATOR_{10_FAMILIAS} \rightarrow QTDE_MORADORES \rightarrow DIFICULDADE]$, apresenta nós com configuração linear e na rede anterior apresentava-se um nó divergente com a seguinte representação: $[FATOR_{10_FAMILIAS} \leftarrow QTDE_MORADORES \rightarrow DIFICULDADE]$. Esta configuração não parece satisfatória. Mesmo que se conheça quantidade de pessoas na família, saber que no domicílio existem crianças e adolescentes (dependentes de renda) muda a crença sobre a quantidade de pessoas moradoras e sobre a dificuldade familiar. Para deixar o modelo mais simples não estabeleceu-se uma ligação direta entre o $FATOR_{10_FAMILIAS}$ e a dificuldade.

O próximo arco invertido foi o $[TEM_PLANO_SAÚDE \rightarrow DIFICULDADE_REC]$. Para que se compreenda o efeito prático desta mudança é apresentada uma simulação com uma unidade de consumo.

Se considerar que ter plano de saúde é consequência da dificuldade, então a probabilidade desta família estar em dificuldade seria 69,63%. Mas se considerar que o plano de saúde influencia na percepção de dificuldade, então o valor cai para 58,70%. Esta modificação pareceu não apenas mais lógica, mas capaz de acomodar melhor a influência da variável “previdência”, que mostra a relação da família com o trabalho formal ou com o fato de ter um aposentado como responsável (não possui trabalho principal).

Figura 43 -Efeito da Mudança do sentido do arco na probabilidade calculada: simulação de 1 caso

FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009

As setas em posição divergente não representam adequadamente a situação e por isso foi modificada para uma configuração linear [PREVIDÊNCIA→ TEM PLANO DE SAUDE → DIFICULDADE].

No conjunto de variáveis do GRUPO 3, Tabela 25, as alterações foram realizadas para estabelecer um “conjunto” que pudesse facilitar a simulação de questões relacionadas à região. O esquema apresentado na Figura 44 visa representar que tanto a Unidade da Federação como as Áreas Geográficas onde as famílias residem influenciam nas diferentes percepções sobre os serviços de caráter público. Além disso, este conjunto tem reflexo na percepção da Educação, da Saúde e do Lazer. Para a compreensão dos efeitos destas variáveis é apresentado um exemplo conforme Figura 44. No exemplo simulado observa-se que 60% das pessoas que moram no Estado de São Paulo acham que a Educação, a Saúde e o Lazer são ruins. Mas ao saber que mora na região metropolitana e que os serviços de caráter público são ruins, a percepção sobre a Educação, Saúde e Lazer deve ser pior que a média. É o que mostra a Figura 44. Desta forma tem-se a possibilidade de “quebrar” um pouco melhor as diferentes

localidades, indo além das Unidades da Federação ou Áreas Geográficas, incorporando serviços públicos para caracterizar os lugares onde as famílias habitam.

Figura 44 -Composição de variáveis para representar efeitos regionais e locais na Rede Bayesiana

FONTE: PREPARADA PELA AUTORA COM BASE NA POF 2009.

O GRUPO 4 de alterações, conforme apresentado na Tabela 25, inverte os arcos da variável [FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA] pois intende-se que este conjunto reflete a dificuldade que perdura na família há algum tempo, ao ponto das deficiências na moradia representarem tal situação. Neste sentido a inversão desta seta aponta para:

- [ATRASSO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS] -que são despesas básicas.
- [FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO]. -que mostra a incapacidade financeira para lidar com necessidades essenciais.

Quanto ao GRUPO 5 : [ATRASSO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS], verifica-se que esta variável já havia sido considerada anteriormente como proxy de “causa” e neste

momento, com os ajustes, foi possível inverter o arco, sem ferir a propriedade acíclica da rede, o que não era possível anteriormente.

No GRUPO 6, referente a [STATUS_REDE_ENSINO] e [RENDA_TRANSF_GOVERNO] foi feita a inversão no sentido da seta pois apesar do governo atrelar o benefício do Bolsa Família às matrículas em escolas públicas, este fato é uma redundância do sistema, como constatou SOUZA (2011). Segundo o autor a maioria das crianças e adolescentes já frequenta a escola, mesmo os mais vulneráveis (95%). Assim o reflexo do programa nas matrículas é marginal. Além disso, em termos de permanência na escola, de acordo com cálculos feitos pelo autor : se a criança recebesse a transferência mensal do Bolsa Escola nos valores do ano 2000 por oito anos consecutivos, aos 15 anos de idade, o beneficiário teria apenas 0,2 anos de escolaridade a mais.

A simulação feita e apresentada na Figura 45 mostra que das famílias que receberam transferências do governo, 69% tinham filhos matriculados na rede pública, pois as famílias com crianças e adolescentes são as mais vulneráveis e presentes na escola pública.

Neste sentido a rede proposta visa representar o que já fora sugerido por Souza (2011), corroborando com a ideia de que os programas de transferência podem ser eficazes para focalizar os mais pobres, mas não estimula a “acumulação de capital humano das novas gerações”, usando as palavras do autor.

Figura 45 - Efeito da Mudança do sentido do arco para melhor representar Transferências do Governo

FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009

A alteração no GRUPO 7 , referente a [CLASSE_ABEP] e [EDUC_SAUDE_LAZER], já foi apresentada na Figura 40 e 41, justificando a necessidade do procedimento de ajustes para calibrar as estimativas do valor da probabilidade.

4.8 Análise Geral do Modelo

Após os ajustes propostos no item anterior, o modelo final apresentou os mesmos indicadores para a curva ROC com AUROC igual a 0,841, sendo o intervalo para esta estimativa entre 0,834 até 0,845. Ou seja, o resultado é praticamente o mesmo obtido para os modelos MB1 e MB2, cujas AUROCs foram de 0,83 e 0,82 respectivamente.

Nesse sentido os ajustes no modelo selecionado não interferiram na capacidade de classificação, mas permitiram incorporar relações mais adequadas entre as variáveis.

O Gráfico 23 apresenta a média da probabilidade calculada por categoria da variável referente à pergunta “A renda total da sua família permite que você(s) leve(m) a vida até o final do mês com : Muita Dificuldade, Dificuldade, Alguma Dificuldade, Facilidade; Muita Facilidade ?”.

Observa-se também que os ajustes propostos não alteraram esta estimativa, comparando o Gráfico 23 com o Gráfico 20, já apresentado anteriormente.

Gráfico 23 -Probabilidade de estar com “dificuldade” e percepção informada pelas famílias

Fonte : Preparada pela autora com base na POF 2009

Também não houve mudança em relação ao “teste seco” medindo a assertividade do modelo. Supondo o mesmo limiar de probabilidade (0,75) para determinar a precisão temos : entre os que não declararam “dificuldade” a precisão está é em torno de 80% e para os que declara-

ram “dificuldade”, 76%. A Confiabilidade das estimativas quando o modelo classifica dificuldade como “não” é de 52% e de 92% quando classifica como “sim”. Estes indicadores são os mesmos do modelo anterior. Apenas para esclarecer o “teste seco” adota a medida 0.75 (que é a proporção geral dos que vivem com dificuldade) apenas como um parâmetro para a avaliação geral e não necessariamente deve ser usado para classificar as famílias.

Para detectar a dificuldade das famílias o modelo desenvolvido utilizou dois blocos classificados a priori como *proxy* de “causa” e “sintoma”, conforme apresentado anteriormente na Tabela 21. Esta classificação está representada no Gráfico 24 e percebe-se que os sintomas selecionados conferem muito mais informação para revelar se uma família está ou não em dificuldade.

A interpretação do Gráfico 24 é a seguinte: ordenando todas as variáveis de acordo com o grau de informação mútua que conferem para a compreensão da “dificuldade” e instanciando-as na rede pela ordem do valor da informação, o valor esperado de cada variável vai sendo revelado. Observa-se um efeito dominó, quando rede “é instanciada” da informação “mais forte” para a “menos forte”, em termos de informação mútua. Tanto a variável “dificuldade” quanto as demais vão sendo reveladas, pois estão todas interligadas.

Gráfico 24 : Informação mútua variáveis selecionadas para o modelo final

No exemplo apresentado na Figura 45, foram instanciadas as variáveis tipo de alimento e quantidade de alimento. A partir desta informação, grandes são as chances da pessoa pertencendo a uma família com dificuldade.

cer à classe D, já ter atrasado despesas, ter necessidades na área da saúde, sendo que plano de saúde é uma possibilidade remota para esta família.

As informações adicionais vão ajustando a percepção de “dificuldade”. No exemplo, sabe-se que a pessoa tem plano de saúde (variável instanciada na Figura 47), então a quantidade de alimento deve ser sempre suficiente e a classe econômica deve ser B2 e ou C1 e a “dificuldade” esperada é menor que no exemplo anterior.

O que se vê é que as variáveis mais relevantes têm também o poder de desvendar, ou revelar as demais, assim as variáveis que aparecem com o menor valor para a informação mútua são “quase” desnecessárias para mensurar a “dificuldade”. No entanto defende-se neste trabalho que são importantes para contextualizar o problema e ter ideia do panorama geral e relação entre as variáveis.

Figura 46 Variáveis Instanciadas na Rede : Tipo e Quantidade de Alimento, reflexo nas demais
FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009

Figura 47 - Variável Instanciada na Rede : Plano de Saúde e reflexo nas demais
FONTE : PREPARADA PELA AUTORA COM BASE NA POF 2009

Esta é uma característica importante das redes bayesianas, confirmada neste trabalho, Conforme já apresentado anteriormente : as Redes Bayesianas têm o potencial de ativar o raciocínio, pois como é possível instanciar as variáveis do modelo enquanto o observa, pode-se “descobrir” informações de forma interativa. Pearl e Russel (2003) salientam que o processo de raciocínio pode ser operado nas Redes Bayesianas pela propagação da informação em qualquer direção. Ou seja, dado que a informação se propaga é possível raciocinar “da causa para o sintoma” e do “sintoma para a causa”, o que não é possível nas redes neurais ou sistemas baseados em regras, por exemplo.

Se o Gráfico 24 mostra a importância das variáveis para compreender a “dificuldade”, os arcos e suas origens, apresentados na sequência, é que permitem compreensão do funcionamento da rede. Assim, serão apresentadas as justificativas lógicas, sob a óptica do pesquisador, que conferem significado para os arcos em evidência. Serão tomadas as variáveis, uma a uma, como ponto de partida e serão justificadas as relações que estabelecem com as demais variáveis no modelo, tomando como baliza a óptica do pesquisador, ou seja, aquilo que faz sentido subjetivamente para quem observa o modelo.

Inicia-se pela variável AREA_GEO e seus relacionamentos.

Figura 48 - AREA_GEO e seus relacionamentos.

AREA_GEO : entende-se que as áreas geográficas, representadas pela região metropolitana, capital, interior e rural com água e sem água, guardam maior ou menor aproximação com o poder político e econômico.

AREA_GEO->ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS: as áreas rurais muitas vezes não têm acesso a estes serviços e assim não podem atrasar estas despesas.

AREA_GEO->CLASSE_ABEP : quanto mais “próximo do poder”, maior a classe econômica.

AREA_GEO->FATOR_11_EDUC_SAÚDE_LAZER: quanto mais próximo do poder melhores são estes serviços.

AREA_GEO->FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA: nas áreas rurais as condições de moradia tendem a ser piores.

AREA_GEO->FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO, quanto mais próximo do poder melhores são estes serviços.

AREA_GEO->SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA: quanto mais próximo do poder melhores são estes serviços

AREA_GEO->SEXO : na área rural predominam os homens chefes de família, dado as condições do local e às relações mais tradicionais em família

Figura 49 - ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE e seus relacionamentos.

ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS->ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS:
O descontrole no pagamento de despesas básicas parece evidenciar o atraso de prestações de bens e serviços.

Figura 50 – Classe ABEP e seus relacionamentos.

CLASSE_ABEP->ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS : as classes mais baixas tendem a ter dificuldade com o pagamento de despesas básicas .

CLASSE_ABEP->DIFICULDADE_REC : as classes, aqui representando renda, influenciam diretamente na dificuldade, pois a falta de dinheiro é a causa mais óbvia para a dificuldade. Por outro lado as classes são obtidas considerando a posse de bens e educação, assim quanto menos se tem maior o sentimento de viver com dificuldade.

CLASSE_ABEP->FATOR_11_EDUC_SAUDE_LAZER : o acesso à educação, saúde e lazer de qualidade custam caro e o sistema público deixa muito a desejar, assim as famílias mais abastadas têm maior acesso que as demais a estes serviços.

CLASSE_ABEP->FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA : quando não se tem recursos financeiros é muito difícil manter a casa em boa condição.

CLASSE_ABEP->FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO : a situação da saúde pública no Brasil é ruim, neste sentido, quando os menos favorecidos precisam de atendimento e aquisição de medicamentos não conseguem obtê-los facilmente.

CLASSE_ABEP->PREVIDENCIA : as famílias de classes mais altas já investiram muito mais em capital humano, principalmente através da educação, assim, conseguem empregos melhores e formais, o que implica na capacidade destas famílias contribuírem com a previdência pública.

CLASSE_ABEP->QTDE_MORADORES : as famílias de classes mais baixas são maiores, tendo a presença de maior número de filhos, em algumas situações dividem a moradia com outra(s) família.

CLASSE_ABEP->QUANTIDADE_DE_ALIMENTO_REC : quanto maior a renda maior a quantidade e diversidade de alimentos no domicílio. Se a situação é restritiva as famílias acabam passando necessidades de alimentação, declarando que a quantidade de alimento é insuficiente.

CLASSE_ABEP->STATUS_REDE_ENSINO : famílias de classes altas matriculam os filhos em escolas particulares e as de classe baixa em escolas públicas.

CLASSE_ABEP->TEM_PLANO_DE_SAUDE : se a família tem boas condições econômicas e financeiras pagará um plano de saúde ou terá um emprego com benefícios que incluem o plano de saúde.

CLASSE_ABEP->TIPO_ALIMENTO_RAZAO_REC : o tipo de alimento é escolhido de acordo com a renda disponível, os mais pobres nem sempre comem o alimento do tipo que querem, pois o orçamento é restrito.

CLASSE_ABEP->X_RENDA_TRANSF_GOVERNO : se a família pertence às classes mais vulneráveis, em algumas situações, poderá contar com a ajuda das transferências governamentais.

Figura 51 – DIFICULDADE_REC e seus relacionamentos

DIFICULDADE_REC->ATRASSO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS: a dificuldade provoca o atraso de despesas.

DIFICULDADE_REC->ATRASSO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS: a dificuldade provoca o atraso nas prestações de bens ou serviços.

DIFICULDADE_REC->QUANTIDADE_DE_ALIMENTO_REC: a dificuldade provoca a redução na quantidade de alimentos consumidos.

DIFICULDADE_REC->TIPO_ALIMENTO_RAZAO_REC: a dificuldade restringe a liberdade de escolha de alimentos.

Figura 52 – FATOR_10_FAMILIAS e seus relacionamentos

FATOR_10_FAMILIAS->QTDE_MORADORES : A formação da família e seu ciclo de vida determinam o número de moradores de um domicílio.

FATOR_10_FAMILIAS->STATUS_REDE_ENSINO : O ciclo de vida familiar determina se a família tem ou não filhos matriculados na rede de ensino.

Figura 53 – FATOR_11_EDUC_SAUDE_LAZER- e seus relacionamentos

FATOR_11_EDUC_SAUDE_LAZER->DIFICULDADE_REC : A percepção negativa sobre o acesso à educação, saúde e lazer tendem a provocar sentimento de dificuldade, pois sabe-se que para obtê-los seriam necessários muitos recursos.

Figura 54 FATOR_12_GOTEIRA_FUNDACAO_PAREDE_ETC e seus relacionamentos

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA->DIFICULDADE_REC a as condições da moradia podem representar o “estoque de pobreza” de uma família. Quanto pior, há mais tempo ela está com dificuldade.

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA-

>ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS : a pobreza pode ser observada pela condição da moradia e estando em más condições a família está em estado de privação, o que implica no atraso de despesas básicas.

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA-

>ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS : a pobreza pode ser observada pela condição da moradia . Nestas condições a família pode vir a atrasar despesas de bens e produtos adquiridos.

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA-

>FATOR_27_NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO : tanto a pobreza quanto as condições da moradia (umidade por exemplo) podem levar a família a ter problemas com a saúde.

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA-

>QUANTIDADE_DE_ALIMENTO_REC : a pobreza pode ser observada pela condição e qualidade da moradia

Figura 56 - FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO e seus relacionamentos

FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO->FATOR_11_EDUC_SAÚDE_LAZER : o fator 6 é o representante da infraestrutura, que chega sempre antes das escolas, dos hospitais e das atividades de lazer.

FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO->SERVIÇO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA : : geralmente a energia elétrica chega antes dos serviços de escoamento de água.

Figura 57 - PREVIDENCIA e seus relacionamentos

PREVIDENCIA->ATRASSO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS : quando pelo menos parte da família tem trabalho formal, existe maior chance de acesso ao crédito.

PREVIDENCIA->TEM_PLANO_DE_SAUDE : trabalhos formais tendem a oferecer benefícios de plano de saúde.

Figura 58 - QTDE_MORADORES e seus relacionamentos

QTDE_MORADORES->DIFICULDADE_REC : quanto mais moradores num domicílio menor a renda per capita e maior a dificuldade

QTDE_MORADORES->PREVIDENCIA : famílias com poucas pessoas tentem a ser habitadas por aposentados (casais ou solitários).

QTDE_MORADORES->X_RENDA_TRANSF_GOV : quanto mais moradores num domicílio maiores as chances de buscar auxílio do governo.

Figura 59 - QUANTIDADE_DE_ALIMENTO_REC e seus relacionamentos

QUANTIDADE_DE_ALIMENTO_REC->TIPO_ALIMENTO_RAZAO_REC : famílias que restringem a quantidade de alimentos tendem a restringir também o tipo de alimento.

Figura 60 - RACA_COR_PESSOA_REFERENCIA e seus relacionamentos

RACA_COR_PESSOA_REFERENCIA->CLASSE_ABEP: considerando a situação atual e a longa história da composição de raças e cor no Brasil, pode-se dizer que a raça e a cor tendem a influenciar na ascensão econômica, uma vez que as raças e diferentes cores têm um legado histórico, associa-se para negros/pretos, pardos e índios as piores condições de vida.

Figura 65 - UF_GRUPO e seus relacionamentos

UF_GRUPO->AREA_GEO : as Unidades da Federação possuem configurações diferentes, com mais, menos ou nenhuma região metropolitana, população mais interiorana, ou com mais pessoas na zona rural

UF_GRUPO->CLASSE_ABEP : os desenvolvimentos regionais acabam interferindo na composição de classes de cada unidade da federação.

UF_GRUPO->FATOR_11_EDUC_SAÚDE_LAZER : estes serviços possuem grandes diferenças regionais, resultado da adoção de políticas públicas.

UF_GRUPO->FATOR_27_NECESSITOU_SERV_SAÚDE_OU_MEDICAMENTO : estes serviços possuem grandes diferenças regionais, inclusive em relação ao aparelhamento da saúde (redes de hospitais, por exemplo).

F_GRUPO->FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO : estes serviços também possuem diferenças regionais

UF_GRUPO->RACA_COR_PESSOA_REFERENCIA : as unidades da federação possuem histórias e heranças diferentes relacionadas à época da escravidão e imigrações para o Brasil, o que acabou determinando a cor dos habitantes nos diferentes estados.

UF_GRUPO->SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA : estes serviços também possuem diferenças regionais.

4.9 Agrupamentos por dificuldade

Tanto o tipo da pesquisa utilizada (POF) quanto as variáveis selecionadas levam à identificação de dificuldades básicas da população brasileira. Neste sentido o modelo está direcionado para necessidades básicas desde a escolha das variáveis. No entanto, com tantas simulações possíveis na rede tornou-se necessário quantificar o grupo de pessoas que poderiam estar com dificuldades básicas. Pelo modelo sabe-se o que influencia a dificuldade, mas quantos são os influenciados?

Neste sentido confrontando o modelo de probabilidade, produto da rede bayesiana, com a “verdadeira” resposta dada pelas famílias, chegou-se em quatro grupos distintos e homogêneos quanto a probabilidade de estar em dificuldade.

GRUPOS DE FAMÍLIAS	ESTIMATIVAS BASEADAS NA POF 2009	
	Total	%
OUTRAS DIFICULDADES	8,369,024	15%
VIVEM BEM	8,922,587	16%
DIFICULDADE BASICA	34,381,838	60%
DIFICULDADE REMEDIADA	5,216,444	9%
TOTAL ANALISADO	56,889,893	100%

Tabela 26 - Grupos de Dificuldade

Para a obtenção dos grupos as famílias foram primeiro separadas pelo agrupamento original : “dificuldade” e “não dificuldade” . Na sequência, utilizando o procedimento de agrupamento por K-médias (ou K-Means) , considerando a variável que consolida probabilidade calculada para dificuldade, foram criados mais dois subgrupos, conforme apresentado no Gráfico 25. Assim, evitou-se adotar um limiar arbitrário para “corte” da probabilidade, como o 0,75 anteriormente praticado, garantindo maior homogeneidade para as médias dentro dos grupos. Observa-se pelo Gráfico 25 que o grupo Dificuldades Básicas é o mais homogêneo, apesar de maior. Não existe a expectativa de encontrar grupos com tamanhos similares, mas apenas

grupos que possam permitir confrontar a probabilidade estimada à resposta dada pelo representante da família.

Gráfico 25 - Boxplot : probabilidade de estar em dificuldade por grupo

FONTE :DADOS POF 2009

Em linhas gerais os grupos foram assim definidos :

“DIFICULDADE BÁSICA” : possui a mais alta probabilidade de estar em dificuldade e afirmou estar com “dificuldade”.

“VIVEM BEM” : ao contrário, possui a mais baixa probabilidade de dificuldade e declarou não estar nesta condição.

“DIFICULDADE REMEDIADA” : possui alta probabilidade de estar com dificuldade, mas não se declara dessa forma, pode estar conformado com sua dificuldade ou vem de situação tão mais difícil que a atual pode ser considerada satisfatória.

“OUTRAS DIFICULDADES” : que foi assim nomeado pelo fato do modelo tratar essencialmente de dificuldades básicas e as deste grupo podem ser outras . Ou seja, algumas famílias declararam estar com “dificuldade” e o modelo não as classificou assim, provavelmente outras dificuldades menos básicas incomodam estas famílias, as quais não estão apresentadas no modelo.

No corpo deste texto são apresentadas “apenas” as variáveis utilizadas no modelo selecionado :

Percebe-se que o grupo Outras Dificuldades assemelha-se aos que Vivem Bem, no entanto parecem estar num patamar econômico abaixo deste, mas acima dos que se encontram classificados no grupo “Dificuldades Básicas”. Este último, por sua vez, declara que a alimentação é um problema: 65% informam que nem sempre a alimentação é do tipo que quer (Tabela 27) e 40% informam que as vezes ou normalmente não é suficiente (Tabela 28).

TIPO_ALIMENTO_RAZAO_REC	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
ALIMENTOS NEM SEMPRE DISPONÍVEIS OU OUTROS MOTIVOS	10%	11%	3%	9%	6%	3,373,197
NEM SEMPRE DO TIPO QUE QUER POIS A RENDA NÃO PERMITE	19%	4%	65%	41%	47%	26,502,239
RARAMENTE DO TIPO QUE QUER POIS A RENDA NÃO PERMITE	1%	0%	19%	5%	12%	7,007,591
SEMPRE DO TIPO QUE QUER	70%	85%	12%	44%	35%	20,006,866
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 27 - Tipo de Alimento e Grupos de Dificuldade

FONTE : DADOS POF 2009

Já os que estão no grupo Dificuldade Remediada, parecem viver muito modestamente, apesar de parte declarar que a alimentação “nem sempre é do tipo que quer”, a maioria a considera suficiente.

QUANTIDADE_DE_ALIMENTO_REC	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
AS VEZES NÃO É SUFICIENTE	4%	1%	40%	16%	26%	14,973,665
E SEMPRE SUFICIENTE	95%	98%	46%	81%	65%	36,697,579
NORMALMENTE NÃO É SUFICIENTE	1%	1%	14%	3%	9%	5,218,651
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 28- Quantidade de Alimento e Grupos de Dificuldade

FONTE : DADOS POF 2009

Observando os grupos através das Classes Econômicas (Tabela 29), mais uma vez aparecem as semelhanças e diferenças entre os grupos. Destacam-se os que Vivem Bem nas classes mais abastadas. O grupo Outras Dificuldades pode ser chamados de representante da classe média, “B2_C1”, que não têm necessidades tão básicas mas almejam por outras conquistas, não retratadas neste estudo. As Classes “D e E” predominam entre os que estão com “Difi-

culdades Básicas” e as famílias com “Dificuldade Remediada” pertencem mais às Classes C1 e C2 e percebem até alguma facilidade.

CLASSE_ABEP	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
A1+A2	3%	10%	1%	1%	2%	1,373,217
B1	10%	18%	2%	4%	6%	3,391,303
B2	22%	27%	8%	15%	14%	7,691,457
C1	27%	22%	18%	27%	21%	11,734,803
C2	20%	13%	26%	26%	23%	13,131,290
D	16%	10%	36%	23%	28%	15,895,402
E	2%	2%	9%	3%	6%	3,672,420
TOTAL %	100%	100%	100%	100%	100%	56,889,892

Tabela 29 - Classes Econômicas (Critério Brasil -ABEP) e Grupos de Dificuldade

FONTE : DADOS POF

Observando os mesmos dados, mas tendo os percentuais calculados dentro de cada Classe Econômica, tem-se a distribuição apresentada no Gráfico 26.

Gráfico 26 -Grupos de Dificuldade & Classes Econômicas

Percebe-se que o modelo capta a informação de classes e classifica as famílias, na maioria das vezes, dentro de uma categoria esperada do ponto de vista lógico.

Considerando que o grupo “Dificuldades Básicas” predomina abaixo da classe C2 é de se esperar que a inadimplência de contas básicas como energia, água e gás sejam realidade em

suas vidas. A dificuldade do dia a dia leva parte destas famílias e mesmo as remediadas a não cumprirem com os compromissos mais básicos.

ATRASO_DE_DESPESAS_COM_ AGUA_ELETRICIDADE_E_GAS	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO	82%	90%	46%	64%	60%	34,212,606
NAO_SE_APLICA	2%	3%	5%	3%	4%	2,388,752
SIM	15%	8%	48%	33%	36%	20,288,534
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,025	8,922,586	34,381,837	5,216,444	56,889,892	

Tabela 30 -Atraso de despesas básicas e Grupos de Dificuldade

FONTE : DADOS POF 2009

A necessidade de saúde é premente para muitos dos classificados no grupo de Dificuldades Básicas (Tabela 31). Enquanto isso nos dois grupos “mais elitizados” muitos podem contar com Plano de Saúde (Tabela 32).

FATOR_27_NECESSITOU_SERV_ SAUDE_OU_MEDICAMENTO	GRUPO_DIFICULDADE				Total %	Total
	DIFICULDADES	VIVEM BEM	BÁSICAS	REMEDIADA		
NAO	91%	96%	57%	78%	70%	39,956,185
SIM	9%	4%	43%	22%	30%	16,933,708
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,838	5,216,444	56,889,893	

Tabela 31 - Necessitou de serviços de saúde ou medicamentos e Grupos de Dificuldade

FONTE : DADOS POF 2009

Dentre os que Vivem Bem 61% declaram ter plano de saúde. No entanto mais da metade dos que têm Outras Dificuldades não o tem, podendo esta ser uma de suas necessidades, uma vez que já superaram necessidades mais básicas, como a alimentação.

TEM_PLANO_DE_SAUDE	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO	53%	39%	84%	78%	72%	41,050,238
SIM	47%	61%	16%	22%	28%	15,839,655
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,587	34,381,837	5,216,444	56,889,893	

Tabela 32 - Plano de Saúde e Grupo de Dificuldade

FONTE : DADOS POF 2009

As casas deterioradas fazem parte do universo dos que possuem dificuldades básicas, pois mais de 63% declararam ter problemas na moradia. Altos índices também são observados para os remediados. Esta variável, como dito anteriormente, visa representar no modelo o “estoque” de pobreza de longa data.

FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO	71%	81%	37%	51%	50%	28,432,892
SIM	29%	19%	63%	49%	50%	28,457,003
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 33 - Condições da Moradia e Grupos de Dificuldade

FONTE : DADOS POF 2009

Acompanhando a mesma tendência quanto ao atraso de serviços básicos os grupos menos favorecidos têm bastante dificuldade quanto a manter outras prestações em dia . Atrasar é o primeiro passo para inadimplir e atrasos acabam acarretando despesas com juros altos, que sufocam orçamentos muito restritos. Muitas vezes por conta da inadimplência, os mais necessitados ficam fora do mercado de crédito e acabam provavelmente em busca de fontes não oficiais para financiamento que cobram juros exorbitantes ampliando ainda mais suas dificuldades. O comparativo entre os grupos mostra que crédito é uma necessidade não atendida a contento para as classes menos favorecidas.

ATRASO_DE_PRESTACOES_DE_BENS_OU _SERVICOS_ADQUIRIDOS	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO	83%	91%	56%	68%	66%	37,747,866
NAO_SE_APLICA	4%	4%	6%	5%	5%	3,119,339
SIM	13%	6%	38%	27%	28%	16,022,688
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,587	34,381,837	5,216,444	56,889,893	

Tabela 34 - Atraso em prestações e Grupos de Dificuldade

FONTE : DADOS POF 2009

É na rede pública de ensino que se encontram os mais necessitados, o que significa também que dentre este grupo estão mais famílias com crianças e adolescentes. Percebe-se que os que têm Outras Dificuldades e Vivem Bem têm mais da metade da família fora da escola, ou seja, predomina população adulta e não dependente de renda.

STATUS_REDE_ENSINO	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NINGUEM NO DOM CRECHE OU ESCOLA	56%	56%	37%	42%	43%	24,430,771
PARCIALMENTE NA REDE PARTICULAR	5%	4%	5%	6%	5%	2,718,532
TODOS NA REDE PARTICULAR	18%	23%	7%	11%	12%	6,650,205
TODOS NA REDE PUBLICA	22%	16%	52%	41%	41%	23,090,384
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,025	8,922,587	34,381,837	5,216,443	56,889,892	

Tabela 35 -Situação na Rede de Ensino e Grupos de Dificuldade

FONTE : DADOS POF 2009

Como esperado, a maioria dos que passam por dificuldades estão nos estados do Norte e Nordeste, mas em quantidade é também surpreendente a população que passa por dificuldade básica em São Paulo, Minas e Rio de Janeiro, normalmente associados ao desenvolvimento. Nestes estados também se concentram metrópoles e grandes periferias, onde estão os marginalizados.

UF_GRUPO	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
BA_PE_CE_PA_PB_RN_TO	14%	13%	28%	24%	23%	13,131,388
GO_AM_AC_RR	5%	5%	5%	6%	5%	3,001,234
MA_PI_AL_SE	3%	3%	9%	7%	7%	3,949,034
PR_ES_MS_RO_AP	12%	12%	9%	11%	10%	5,678,205
RS_SC_MT_DF	19%	19%	9%	13%	13%	7,214,167
SP_MG_RJ	46%	48%	40%	38%	42%	23,915,867
TOTAL %	100%	100%	100%	100%	100%	56,889,894
TOTAL	8,369,025	8,922,586	34,381,838	5,216,445	56,889,894	

Tabela 36 - Grupos de Unidades da Federação - Grupos de Dificuldade

FONTE : DADOS POF 2009

A quantidade de moradores também corrobora com a ideia de que nos grupos menos favorecidos concentram-se famílias mais numerosas, dado a presença de crianças e adolescentes, dependentes de renda. Nos domicílios menos favorecidos também se encontram mais de uma família vivendo conjuntamente, tornando-a ainda mais numerosa.

QTDE_MORADORES	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
<=1	19%	21%	8%	9%	12%	6,771,090
<=2	29%	31%	19%	22%	22%	12,742,720
<=3	26%	24%	24%	27%	25%	14,061,790
<=4	19%	16%	23%	22%	21%	12,175,144
<=5	6%	7%	13%	11%	11%	6,151,829
>5	2%	2%	12%	8%	9%	4,987,321
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,025	8,922,587	34,381,837	5,216,443	56,889,892	

Tabela 37 - Quantidade de Moradores -Grupos de Dificuldade

FONTE : DADOS POF 2009

Nos grupos com as dificuldades básicas são mais presentes as transferências governamentais. No modelo observou-se que as transferências do governo recebem influência direta da quantidade de moradores nos domicílios, da classe econômica, da presença em escola pública e da variável que indica que a família necessitou de serviços de saúde nos últimos meses e não foi atendida. Este fato evidencia o grau de precariedade destas famílias.

X_RENDA_TRANSF_ GOVERNO	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO	93%	94%	70%	81%	78%	44,329,427
SIM	7%	6%	30%	19%	22%	12,560,465
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,025	8,922,586	34,381,837	5,216,444	56,889,892	

Tabela 38 - Transferências do Governo - Grupos de Dificuldade

FONTE : DADOS POF 2009

Como os recursos são escassos o governo estabelece regras de priorização baseadas em renda per capita e outros requisitos. Os recursos de transferências do governo estão alocados entre os mais pobres, conforme também já se comprovou em outros de estudos apresentados por Souza (2011). No entanto, visando contribuir com a melhor localização da pobreza a ser assistida, será proposto um exercício de reflexão sobre as transferências do governo no item 5.

Dando a continuidade à análise das demais variáveis constata-se que os negros/pretos, pardos e índios estão dentre os que têm maior dificuldade orçamentária. Neste sentido, dado que o assunto é controverso será proposto também um exercício para reflexão no item 5, tendo como objeto de análise as cotas raciais e a dificuldade percebida pelas famílias. Percebe-se que a situação é mesmo muito grave: enquanto os negros/pretos, pardos e índios representam 28% dos classificados como “Vivem Bem”, este número sobe para 59% entre os que têm Dificuldades Básica.

RACA_COR_PESSOA_REFERENCIA	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADE S BÁSICAS	DIFICULDADE REMEDIADA		
BRANCOS, AMARELOS E NÃO SABE	67%	72%	41%	46%	50%	28,383,382
NEGROS, PARDOS E INDIOS	33%	28%	59%	54%	50%	28,506,510
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,025	8,922,586	34,381,837	5,216,444	56,889,892	

Tabela 39 - Raça e Cor - Grupos de Dificuldade

FONTE : DADOS POF 2009

Enquanto os solitários estão dentro dos grupos mais abastados, as famílias com crianças adolescentes são a maioria nos outros dois. Muitas vezes a dificuldade das famílias vem do fato das mulheres não poderem trabalhar ou assumirem empregos menos formais. Em São Paulo,

por exemplo, muitas creches fecham uma vez por mês, na última sexta feira, para cuidar de assuntos internos. Assim, mesmo mulheres que têm filhos em creches precisam conseguir trabalhos onde possam faltar pelo menos uma vez por mês, caso não tenham outra opção para cuidar dos filhos. O problema da mulher começa provavelmente na condução da família e não apenas no fato dela ser mulher, o que justificaria ações por gênero. A variável “sexo” ficou em último lugar neste estudo quanto à métrica da informação mútua. Ao saber sobre filhos pequenos e adolescentes no domicílio, muitas a famílias estão em dificuldade, não só a mulher, mas todos que nela habitam.

FATOR_10_FAMILIAS	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
ATE DOIS FILHOS ADULTOS OU ADOLESCENTES	27%	26%	25%	27%	26%	14,690,057
DUAS PESSOAS SEM FILHOS	25%	28%	13%	18%	18%	10,164,717
FAMILIAS COM CRIANCAS ATE 6 ANOS	16%	14%	31%	28%	26%	14,830,903
MAIS DE DOIS FILHOS PREDOM ADOLESCENTES	13%	12%	22%	18%	18%	10,471,032
SOLITARIOS	19%	21%	8%	9%	12%	6,733,186
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 40 -Tipos de Família -Grupos de Dificuldade

FONTE : DADOS POF 2009

O interior dos estados abriga a maior parte das famílias, mas as regiões metropolitanas e áreas rurais também abrigam os que passam por dificuldade. Muitos dos que vivem bem estão mais próximos das relações com o poder político e econômico, nas capitais.

AREA_GEO	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
CAPITAL	28%	33%	20%	22%	24%	13,452,332
INTERIOR	50%	48%	47%	49%	48%	27,175,057
REG METROPOLITANA	10%	9%	15%	14%	13%	7,333,112
RURAL_COM_AGUA	8%	7%	9%	9%	9%	4,940,342
RURAL_SEM_AGUA	4%	4%	9%	6%	7%	3,989,049
TOTAL %	100%	100%	100%	100%	100%	56,889,894
TOTAL	8,369,025	8,922,586	34,381,838	5,216,445	56,889,894	

Tabela 41 - Área geográfica - Grupo de Dificuldade

FONTE : DADOS POF 2009

A percepção sobre o acesso que têm à Educação, Lazer e Saúde corrobora com a realidade vivida. Os mais necessitados têm consciência sobre a carência destes benefícios, como pode ser visto na Tabela 42.

Pergunta-se : Onde está o acesso à educação, à saúde e ao lazer de qualidade, que possa abarcar as famílias que vivem com dificuldade, que têm filhos pequenos e adolescentes, que

se encontram “não por acaso” dentro das escolas públicas? Sabendo que mais da metade das famílias do grupo menos favorecido se encontra diariamente nas escolas públicas, dividindo as mesmas dificuldades, pergunta-se: qual seria o papel deste espaço ? O professor é adequadamente valorizado na escola pública ? Deveria ter uma enfermeira, uma nutricionista por lá? Deveria passar um dentista, um médico? As crianças poderiam estar estudando e praticando esporte neste local? As mães poderiam trabalhar mais tranquilas liberando o bolsa família ? Os adolescentes teriam melhor instrução sobre controle da natalidade e seus benefícios? A diferença entre raças e cores diminuiria? A renda das famílias aumentaria? E a violência? Qual o tamanho da renda “não monetária” necessária para saltar qualitativamente nesta área?

FATOR_11_EDUCAO_SAUDE_LAZER	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
BOM_NAO_RESPONDEU	54%	59%	30%	36%	39%	21,936,079
RUIM_NAO_TEM	46%	41%	70%	64%	61%	34,953,814
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 42 - Acesso à Educação, Saúde e Lazer e Grupos de Dificuldade

FONTE : DADOS POF 2009

O acesso aos itens nomeados anteriormente estão, de certa forma, ligados ao acesso ao trabalho formal. Dentre os que têm dificuldade muitos não contribuem com a previdência e isso remete a 2030, quando espera-se uma inversão na curva da dependência demográfica. Estudo do Banco Mundial (2011) mostra o país envelhecendo e os efeitos dessa mudança no crescimento, saúde, educação e previdência.

PREVIDENCIA	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
NAO POSSUI TRABALHO PRINCIPAL	17%	17%	12%	11%	13%	7,580,673
NINGUEM CONTRIBUI	28%	28%	46%	38%	40%	22,803,810
PARTE CONTRIBUI	23%	23%	23%	28%	24%	13,488,372
TODOS CONTRIBUEM	32%	32%	18%	22%	23%	13,017,037
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,586	34,381,837	5,216,445	56,889,893	

Tabela 43- Contribuição para a previdência e Grupos de Dificuldade

FONTE : DADOS POF 2009

Na sequência apresenta-se o acesso aos serviços de utilidade, deixando claras o quão básicas são as necessidades das camadas mais pobres da população. Os serviços de escoamento de água são ainda ruins para muitos.

SERVICO_DE_ESCOAMENTO_ DA_AGUA_DE_CHUVA	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
BOM	75%	78%	55%	58%	62%	35,116,149
RUIM_NAO_TEM	25%	22%	45%	42%	38%	21,773,744
TOTAL %	100%	100%	100%	100%	100%	56,889,892
TOTAL	8,369,024	8,922,586	34,381,838	5,216,444	56,889,892	

Tabela 44 -Serviços de escoamento da agua de chuva -Grupos de Dificuldade

FONTE : DADOS POF 2009

O sexo/gênero do responsável pela família. Percebe-se uma pequena elevação no percentual de mulheres responsáveis pela família dentre os que têm dificuldades básicas. Verificou-se através de simulações na rede bayesiana que “mulher responsável pela família” é uma característica que aumenta conforme as classes econômicas vão diminuindo de poder aquisitivo. Além disso, as mulheres chefes de família das classes mais baixas tendem a não contar com a presença do cônjuge. Dentro da Classe D, 36% dos responsáveis são mulheres e destas 82% não têm cônjuge. Comparando com a classe B observa-se 25% de responsáveis do sexo feminino, sendo que 65% não têm cônjuge.

No entanto o motivo delas estarem em último lugar neste estudo é que suas características já foram esgotadas nos itens anteriores, ou seja, para “o modelo” depois de tudo conhecer, saber que a mulher é a responsável pelo domicílio nada mais agrega. Como dito, algumas das causas da dificuldade da mulher chefe de família não é só o fato de ser mulher, mas de não ter como cuidar dos filhos, não ter creche em número suficiente, necessitar de trabalhos menos formais ou de meio período.

SEXO / GENERO	GRUPO_DIFICULDADE				Total %	Total
	OUTRAS DIFICULDADES	VIVEM BEM	DIFICULDADES BÁSICAS	DIFICULDADE REMEDIADA		
FEMININO	28%	28%	32%	29%	31%	17,538,686
MASCULINO	72%	72%	68%	71%	69%	39,351,207
TOTAL %	100%	100%	100%	100%	100%	56,889,893
TOTAL	8,369,025	8,922,587	34,381,837	5,216,444	56,889,893	

Tabela 45 Sexo/Gênero do responsável pela família -Grupos de Dificuldade

FONTE : DADOS POF 2009

4.10 Validação da Rede

Para a validação da rede, além da curva ROC, já apresentada, que visa identificar a capacidade classificatória da rede, foram realizados mais três avaliações: (1) validação da probabilidade calculada contra a variável renda média per capita; (2) validação dos arcos através do procedimento de K-Amostras e (3) verificação pela verossimilhança..

4.11 Validação contra a variável renda média per capita:

Uma vez que a variável relativa à renda per capita foi propositadamente excluída da análise após o primeiro modelo (MB1 -Markov Blanket 1) foi possível utilizá-la para comparar com a probabilidade calculados a partir do modelo selecionado. Esta é uma avaliação descritiva e qualitativa.

Antes desta análise apresenta-se o Gráfico 27 que mostra os décimos da renda média per capita e a percepção de dificuldade informada pelas famílias brasileiras. A renda per capita exprime a quantidade de renda monetária dividida pelo número de pessoas no domicílio. Os “decis” ou décimos representam as famílias agrupadas de 10 em 10% por ordem de renda familiar per capita.

Como esperado, as menores rendas têm a maior proporção de pessoas com dificuldade.

Gráfico 27 -Decis de renda per capita e percepção sobre a dificuldade

FONTES : DADOS POF 2009

Para avaliar o modelo sob a perspectiva da renda foram feitas avaliações por décimos da probabilidade: famílias divididas de 10 em 10% por ordem de probabilidade de estarem em dificuldade.

Observa-se pelo Gráfico 28 que nos menores decis, onde se encontram as famílias mais ricas a renda média é bem superior que nos últimos decis, corroborando assim com os resultados esperados para o modelo.

Outra validação considerou a renda informada pelas famílias como necessária para levar a vida com tranquilidade e a dividiu pela renda média per capita familiar, apurada pela POF. Obteve-se assim o múltiplo de renda que a família precisa em relação à renda coletada na POF, por exemplo : 2,5 vezes, 1,5 vezes, etc.

Gráfico 28 - Renda per capita média por décimos de probabilidade

FONTE : DADOS POF 2009

Pelo Gráfico 29, a seguir, pode-se verificar que os 10% com maiores probabilidades de dificuldade necessitariam de 2,5 vezes a sua renda familiar atual para viver com mais tranquilidade. Neste sentido é preciso mais que transferências do governo para atender suas necessidades.

Gráfico 29 - Múltiplos da Renda por percentil da probabilidade.

FONTE : DADOS POF 2009.

Em relação às necessidades estas famílias, do último decil, declaram que destinariam 45% da renda necessária para a alimentação. Este ponto corrobora com o fato da alimentação (obtida por outras variáveis) contribuir tanto para compreender a dificuldade.

Gráfico 30 - Múltiplos da Renda por percentil da probabilidade.

FONTE : DADOS POF 2009.

Considerando os grupos propostos no item 4.9 e voltando a observar as famílias por decimos de renda per capita, verifica-se que nas menores rendas concentra-se o grupo “Dificuldades Básicas”. Fazendo cálculos mentais e observando os valores das rendas médias per capita, percebe-se que muitos cuidam mesmo do que é essencial. Abaixo de R\$ 585 per capi-

ta a maioria são famílias classificadas como “Dificuldades Básicas”. Percebe-se que as linhas de corte apuradas pelo governo para estabelecer o público que deve ser assistido é muito mais operacional, pois não dá para atender a todos. As carências da população são muito maiores do que os recursos disponíveis

Gráfico 30- Decis de renda per capita e grupos de dificuldade

FONTE : DADOS POF 2009

Neste sentido, cada vez mais, torna-se necessário avaliar a realidade atual e o destino das famílias de rendas intermediárias.

De forma resumida o valor médio da probabilidade de estar em “dificuldade” é apresentado no Gráfico 31, a seguir. No primeiro “decil” a probabilidade é 0.93 e no último “decil” igual a 0.39, quanto menor o decil de probabilidade apurado pelo modelo, maior a dificuldade percebida pelas famílias.

Gráfico 31 - Decis de renda per capita e probabilidade de estar em dificuldade

FONTE : DADOS POF 2009

4.12 Validação dos arcos do modelo

Para a validação da estrutura da rede o procedimento utilizado foi o método k-Fold este tipo de validação é referenciado na literatura por Witten et al. (2005, p. 153). O método consiste em dividir o banco de dados em 10 amostras de igual tamanho, mas antes deste procedimento o banco de dados é aleatorizado para evitar viés quanto à ordenação dos dados. São utilizadas 10 amostras para aprendizado de 10 diferentes redes. A quantidade de 10 amostras é considerada arbitrária, mas o mais comum, segundo os autores.

Assim, são geradas 10 redes que podem ser diferentes umas das outras. Estas novas estruturas são comparadas à rede referência. A rede utilizada como referência é a rede que está sendo validada. Ela pode ser encontrada ou não entre as novas redes geradas.

Ao final são realizadas comparações arco a arco, visando compreender o que mudou e o que poderia ser aperfeiçoado na rede.

No caso deste trabalho obteve-se a seguinte configuração inicial :

Ao rodar o procedimento K-fold obteve-se cinco redes distintas, sendo que em duas delas houve maior número de coincidências : para a rede do tipo A foram encontradas 3 estruturas idênticas, para a rede D foram encontradas 3 estruturas idênticas. Em nenhuma delas foi en-

contrada a estrutura referência. Este fato já é esperado, pois houve ajuste manual nos arcos, conforme 4.7.

REDES PARA COMPARAÇÃO	ESTRUTURAS	CONTÉM A ESTRUTURA REFERÊNCIA ?
A	3	Não
B	1	Não
C	2	Não
D	3	Não
E	1	Não
TOTAL	10	Não

Gráfico 32 - Redes de comparação : validação dos arcos

As redes apesar de diferentes, guardam alguma semelhança com a rede testada como referencial. Assim é importante conhecer quais foram os arcos que sofreram mudanças. A avaliação destes arcos pode ser utilizada para modificar a rede desenvolvida ou para refletir sobre as alterações propostas.

No presente trabalho os resultados foram bons, sendo que 80% dos arcos foram validados. O “acerto” geral foi computado da seguinte forma : somatório de todos os corretos, divididos pelo total de arcos : $48 / 60$, ou 80%. Os corretos estão identificados na Tabela 46 e podem variar em 3 tipos (Correto 1,2,3).

A rede utilizada como referência e apresentada com detalhes no item 4.7 apresenta 60 arcos. As redes alternativas geradas pelo procedimento de validação sugeriram mais 7 arcos. De forma geral os resultados estão resumidos na tabela a seguir.

ANÁLISE DE CONFIABILIDADE - PROCEDIMENTO JACKKNIFE 10 AMOSTRAS - TABOO ORDER			
AVALIAÇÃO	NOVOS	REFERENCIA	Total Geral
CORRETO 1		25	25
CORRETO 2		11	11
CORRETO 3		12	12
DUVIDOSO		2	2
INSUFICIENTE	6		6
INVERTIDO		8	8
POTENCIAL	1		1
QUESTIONÁVEL		2	2
Total Geral	7	60	67

Tabela 46 -Resultados Gerais da validação dos arcos

Para a compreensão Tabela 46 foi criado um “dicionário” explicando cada um dos status de validação para os 60 arcos que já estavam na rede.

- CORRETO 1 :Arcos que aparecem iguais em 100% das amostras
- CORRETO 2 : Arcos que aparecem iguais em 90% das amostras
- CORRETO 3 : Arcos que podem ser considerados corretos, pois independem do sentido das setas e foram encontrados na maioria das amostras (90%). Mudanças no sentido destes arcos não interferem no computo geral da probabilidade.
- DUVIDOSO : Arcos que aparecem em 50% ou menos na amostras
- QUESTIONÁVEIS : Arcos que não aparecem em nenhuma das 10 redes.
- INVERTIDOS :Arcos que aparecem invertidos na maioria das amostras

Quanto aos 7 novos arcos propostos :

- INSUFICIENTES : apareceram nas outras redes mas num número insignificante de vezes.
- POTENCIAL : arco que apareceu em pelo menos 4 redes, dentre as 10 e pode ser avaliado como potencial arco.

Na sequência são evidenciados todos os arcos que apresentaram discrepância e foram classificados como duvidosos, questionáveis e potencial :

ARCOS DUVIDOSOS	FREQUENCIA DO ARCO	ARCOS INVERTIDOS	ARCOS SEM DIREÇÃO ESPECÍFICA	TOTAL	TOTAL
PREVIDENCIA -> ATRASSO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	50%	0%	0%	50%	50%
UF_GRUPO -> FATOR_11_EDUCAO_SAUDE_LAZER	0%	0%	10%	10%	10%

Tabela 47 -Arcos Duvidosos

O arco relativo à variável previdência pública aparece em metade das amostras testadas, no entanto este arco será mantido considerando a análise adicional que se faz da relação entre as duas variáveis através de uma Tabela cruzada. A aprovação deste arco deve-se à lógica da análise: aposentados e as famílias em que todos contribuem com a previdência pública apresentam índices de atraso menor que os que residem em famílias onde ninguém contribui ou parte contribui. Os casos de “não se aplica”, podem indicar que famílias não têm acesso ao crédito, por isso não podem atrasar prestações. Se retirar o “não se aplica” do cômputo das famílias em que ninguém contribui para a previdência pública o índice de atraso sobe para 42%.

VALIDAÇÃO DO ARCO : PREVIDENCIA -> ATRASSO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS		ATRASSOU ?		
CONTRIBUI COM A PREVIDÊNCIA ?	NAO	NAO_SE_APLICA	SIM	Total
NAO POSSUI TRABALHO / APOSENTADO	5,662,754	197,161	1,720,758	7,580,673
NINGUEM CONTRIBUI	12,280,509	1,604,116	8,919,184	22,803,809
PARTE CONTRIBUI	7,637,746	280,998	5,569,629	13,488,373
TODOS CONTRIBUEM	8,631,597	306,477	4,078,964	13,017,038
Total #	34,212,606	2,388,752	20,288,535	56,889,893
NAO POSSUI TRABALHO PRINCIPAL	75%	3%	23%	100%
NINGUEM CONTRIBUI	54%	7%	39%	100%
PARTE CONTRIBUI	57%	2%	41%	100%
TODOS CONTRIBUEM	66%	2%	31%	100%
Total %	60%	4%	36%	100%

Tabela 48 - Validação do arco : previdência -> atraso de prestações

Assim, com base nestas considerações e dado que os arcos apareceram em pelo menos 50% das redes alternativas, então entende-se que este arco deva ser mantido.

O arco relativo às Unidades da Federação e acesso à Educação Saúde e Lazer, apareceram apenas em 1 das redes, provavelmente pelo fato da relação encontrada entre estas variáveis passar na maioria dos casos primeiramente pela Energia Elétrica, condicionando a Educação,

Saúde e Lazer a este serviço. No entanto, isso não é totalmente verdade e pode ser produto da heurística do “Tabu order”. Na rede proposta (referência) os arcos saem em direção à Educação Saúde e Lazer, tanto da energia elétrica quanto diretamente das Unidades da Federação. Na sequência a relação entre os dois grupos de variáveis :

UF_GRUPO -> FATOR_11_EDUCAO_SAUDE _LAZER	FATOR_11_EDUCAO_SAUDE_LAZER		Total
	BOM_NAO_RESPONDEU	RUIM_NAO_TEM	
BA_PE_CE_PA_PB	4032404	9098984	13131388
GO_AM_AC_RR	1154668	1846566	3001234
MA_PI_AL_SE	1131581	2817453	3949034
PR_ES_MS_RO_AP	2566796	3111408	5678204
RS_SC_MT_DF	3516233	3697933	7214166
SP_MG_RJ	9534397	14381469	23915866
total	21936079	34953813	56889892
BA_PE_CE_PA_PB	31%	69%	100%
GO_AM_AC_RR	38%	62%	100%
MA_PI_AL_SE	29%	71%	100%
PR_ES_MS_RO_AP	45%	55%	100%
RS_SC_MT_DF	49%	51%	100%
SP_MG_RJ	40%	60%	100%
total	39%	61%	100%

Tabela 49 -Validação do arco Unidades da Federação -> Grupo Fator 11

Pela Tabela 49 pode-se perceber a relação entre estas variáveis e de forma consciente não se pode permitir que o arco saia direto e apenas do nó relativo aos demais serviços públicos. A decisão é mantê-lo também desta forma e posteriormente procurar uma maneira de melhor acomodar a informação, pois a relação entre Unidades da Federação e as demais informações relativas à prestação de serviços públicos está confusa na rede, como pode ser observado por este processo de validação. As alterações neste arco podem ser objeto de futuros estudos.

Os arcos invertidos já foram tratados anteriormente e a justificativa já foi apresentada no item 4.7.

Os arcos apresentados abaixo, não apareceram em nenhuma das 10 redes alternativas e neste sentido precisarão ser melhor compreendidos em próximos trabalhos, nota-se que um deles diz respeito novamente ao Fator 11 , Relativo à Educação, Saúde e Lazer. Talvez seja necessário voltar às variáveis originais, antes do agrupamento em um fator. Este trabalho também pode ser conduzido em futuros trabalhos.

ARCOS QUESTIONÁVEIS	FREQUENCIA DO ARCO	ARCOS INVERTIDOS	ARCOS SEM DIREÇÃO ESPECÍFICA	TOTAL	TOTAL	COMENTÁRIO	ORIGEM
AREA_GEO -> FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MA DEIRA_DETERIORADA	0%	0%	0%	0%	0%	QUESTIONÁVEL	REFERENCIA
CLASSE_ABEP -> FATOR_11_EDUCAO_SAUDE_LAZER	0%	0%	0%	0%	0%	QUESTIONÁVEL	REFERENCIA

Tabela 50 –Arcos Questionáveis

Nesta validação o algoritmo utilizado para a validação e desenvolvimento foi o mesmo: o “Tabu Order”. Para trabalhos futuros sugere-se que a validação dos arcos seja realizada também com base em diferentes tipos de algoritmo. Desta forma os arcos mais robustos poderiam ser identificados, independente da técnica de aprendizado de estrutura da rede.

4.13 Validação pela verossimilhança:

Como saída do modelo tem-se uma medida baseada na verossimilhança, denominada no software como “Contingence Fit”. Ela é utilizada para representar proporcionalmente a aderência do modelo aos dados. Ou seja, é uma medida que identifica o quanto o modelo se ajusta ao conjunto de dados informados. Quanto mais próxima de 100% mais aderente o modelo. No modelo proposto obteve-se um índice de 30%, o que foi considerado baixo, no entanto não se encontrou nenhuma medida comparativa na literatura. Para compreender melhor a métrica foram selecionadas poucas variáveis, criou-se um modelo alternativo, bem restrito o qual apresentou baixa AUROC (através da análise da curva ROC) , no entanto obteve-se uma medida bem melhor para o “Contingence Fit” , em torno de 65%. Isso informa que modelos com muitos dados tendem para uma métrica inferior relativa à verossimilhança, pois muitos outros modelos seriam possíveis. Neste sentido, o resultado deste teste não foi conclusivo. Decidiu-se então seguir com a modelagem observando a AUROC para checar o poder de classificação e a análise dos arcos e rede para avaliar a estrutura de forma qualitativa.

5 REFLEXÕES A PARTIR DAS REDES: QUESTÕES PARA NOVOS ESTUDOS.

A partir da rede selecionada é possível raciocinar sobre diferentes aspectos da dificuldade dos brasileiros. O presente capítulo visa exemplificar como operar analiticamente através do modelo, considerando duas abordagens possíveis: uma que utiliza o “ranking” da probabilidade de estar em dificuldade, como um vetor de análise, e a outra que explora a estrutura da rede bayesiana, como elemento de reflexão sobre a inter-relação dos nós.

Para isso foram selecionadas duas situações dentre outras possíveis de observação na rede proposta, lembrando que o presente trabalho utiliza novas tecnologias analíticas menos para “projetar respostas” e mais para “projetar perguntas”. Assim, foram analisadas as seguintes situações:

- 1) A capacidade de identificar famílias-alvo para as transferências governamentais, a partir da comparação de duas ordenações : “ranking” de probabilidade de estar em dificuldade” e outro por “renda familiar per capita”.
- 2) A questão de cotas raciais a partir da estrutura da rede bayesiana, representada pela ordenação dos nós e isolamento de alguma variáveis consideradas para operacionalizar a determinação das cotas: “cor da pele” e “poder econômico das famílias”, representado ora pela renda familiar per capita ora pela classe econômica

5.1 “Ranking” de probabilidade de estar em dificuldade e a alocação dos recursos de transferências do governo.

Depreende-se da leitura de Sousa (2011, p. 166) que os programas de transferências diretas são mais eficientes e eficazes quanto mais se associa (1) à capacidade de identificar os mais pobres com (2) a capacidade de estabelecer medidas complementares de maior impacto sobre o desenvolvimento do capital humano no médio e longo prazo. Quanto ao segundo ponto, observa o autor que os programas de transferência de renda apresentaram resultados bastante insatisfatórios até o momento de sua análise. Quanto ao primeiro, não se questiona a necessidade de atendimento de populações muito pobres e a necessidade de diminuir a desigualdade entre os brasileiros, identificando-os corretamente.

O presente exercício de reflexão visa contribuir com a primeira parte do problema: identificar os mais pobres. Vale salientar que para esta reflexão, do conjunto de benefícios de transferências governamentais, apenas foi selecionado o Bolsa Família. Para isso toma como base duas dimensões analíticas: a ordenação por renda per capita e a ordenação pela probabilidade de estar em dificuldade. Os dois eixos foram utilizados, pois enquanto a renda per capita é uma medida mais objetiva, a probabilidade de estar em dificuldade, desenvolvida nesta tese, é uma medida mais subjetiva, dado que se baseia, parcialmente, na percepção das famílias sobre suas condições de vida.

Observa-se pela Tabela 51 que na mesma faixa de renda per capita familiar, os programas beneficiam mais algumas Unidades da Federação que outras. Por exemplo, das famílias do grupo “BA_PE_CE_PA_PB_RN_TO”, no primeiro décimo de renda per capita, 62% recebem o benefício. Enquanto isso famílias de outros Estados, na mesma faixa de renda, não recebem o benefício na mesma proporção. O agrupamento utilizado para as Unidades da Federação foi o mesmo do critério apresentado na Tabela 8, item 4.3.1. Além disso, em algumas das localidades os beneficiados avançam além dos primeiros décimos da distribuição da renda. Observando a Tabela 51 (que ordena por “decil” de renda per capita média) e a Tabela 51 (que ordena por “decil” da probabilidade), pode-se observar esta desigualdade distributiva.

Sendo os Estados do Norte e Nordeste mais pobres é natural que recebam mais recursos, como demonstra as linhas identificadas como “TOTAL” das Tabelas 51 e 52. No entanto, famílias de outros Estados, dentro das mesmas faixas de renda (Tabela 51) e probabilidade de estarem em dificuldade (Tabela 52), não recebem o benefício na mesma proporção.

Vale salientar que determinar a renda monetária per capita é difícil, principalmente em locais onde há maior informalidade nas relações de trabalho. A POF realiza esta atividade com qualidade. Entretanto, nem sempre se consegue tal rigor na coleta da renda quando da implementação das políticas sociais. Acrescenta-se a isso a volatilidade da renda e seu poder aquisitivo em diferentes locais do Brasil.

% DE FAMÍLIAS BENEFICIADAS COM BOLSA FAMÍLIA POR GRUPOS DE UNIDADES DA FEDERAÇÃO & DECIL DE RENDAS PER CAPTA

RENDA MONETÁRIA PER_CAPITA	BA_PE_CE_PA_PB_RN_TO	GO_AM_AC_RR	MA_PI_AL_SE	PR_ES_MS_RO_AP	SP_MG_RJ	RS_SC_MT_DF	TOTAL
<=113	61%	41%	60%	24%	29%	24%	48%
<=191	46%	30%	46%	26%	27%	21%	36%
<=267	33%	16%	28%	16%	15%	12%	22%
<=363	20%	13%	21%	10%	10%	8%	13%
<=455	10%	7%	7%	5%	4%	4%	6%
<=585	7%	6%	7%	4%	2%	3%	4%
<=784	6%	3%	5%	2%	2%	1%	2%
<=1091	2%	1%	5%	1%	1%	1%	1%
<=1881	2%	1%	2%	1%	0%	1%	1%
>1881	1%	0%	0%	0%	0%	0%	0%
TOTAL	28%	14%	31%	8%	6%	5%	13%

Tabela 51 - Famílias beneficiadas - Bolsa família, grupos de UF e renda per capita – POF 2009

FONTE : DADOS POF 2009

Na sequência apresenta-se a Tabela 52 onde a probabilidade das famílias estarem em dificuldade (representada em percentual) foi agrupada em décimos, caracterizando um “ranking”, onde em cada um estão 10% das famílias. Esta dimensão foi utilizada no lugar da renda per capita familiar, visando avaliar se a conclusão seria a mesma quanto à capacidade de identificar públicos para o Bolsa Família.

% DE FAMÍLIAS BENEFICIADAS COM BOLSA FAMÍLIA POR GRUPOS DE UNIDADES DA FEDERAÇÃO & DECIL DE PROBABILIDADE DE ESTAR COM DIFICULDADE (QUANTO MAIOR O DECIL MAIOR A DIFICULDADE ESPERADA)

DÉCIMOS DA PROBABILIDADE	BA_PE_CE_PA_PB_RN_TO	GO_AM_AC_RR	MA_PI_AL_SE	PR_ES_MS_RO_AP	SP_MG_RJ	RS_SC_MT_DF	TOTAL
1	2%	1%	0%	0%	0%	0%	0%
2	6%	3%	6%	1%	2%	1%	2%
3	11%	8%	12%	6%	3%	3%	6%
4	13%	10%	15%	4%	3%	3%	6%
5	21%	11%	24%	7%	5%	5%	10%
6	21%	11%	24%	8%	7%	5%	11%
7	31%	14%	33%	12%	9%	9%	18%
8	33%	20%	37%	16%	13%	13%	22%
9	45%	25%	46%	19%	15%	20%	31%
10	50%	36%	52%	26%	21%	26%	38%
TOTAL	28%	14%	31%	8%	6%	5%	13%

Tabela 52 - Famílias beneficiadas - Bolsa Família, por grupo de UF e probabilidade - POF 2009

FONTE : DADOS POF 2009

Modelos como o proposto neste trabalho que lidam com “proxys” de renda e de dificuldade podem ser elementos facilitadores ou complementares na identificação de públicos necessários. O exercício realizado na Tabela 51 foi repetido apenas para as famílias que tinham todos os filhos matriculados na escola pública, um dos requisitos para acesso ao Bolsa Família. Ao fazer a análise para este segmento da população que representava em 2009 cerca de 40% das famílias, obteve-se a mesma conclusão: parece que existem fortes evidências de

problemas na identificação e transferência de renda para os públicos elegíveis ao Bolsa Família.

Esta tese mostrou que quando os mais pobres têm aumento de suas rendas, focalizam estes recursos para a alimentação ou despesas muito básicas, o que é esperado. Para Souza (2011 p. 174 – 180) o Bolsa Família tem diminuído desigualdade de renda no curto prazo mas não apresenta nenhum ou quase nenhum impacto de longo prazo na formação do capital humano, educação, saúde, trabalho infantil, fecundidade e oferta de trabalho para os adultos.

O Bolsa Família é visto neste trabalho como “necessário” para diminuição da desigualdade que afeta pobres e os extremamente pobres. Neste sentido, sugere-se para os próximos trabalhos o aprimoramento das “proxys” propostas para que se faça cautelosa distribuição do Bolsa Família, uma vez que a própria Secretaria de Assuntos Estratégicos (SAE), conforme apresentado na página 14 desta tese, caracterizou os extremamente pobres e pobres como aqueles que recebem até R\$ 162 de renda familiar per capita mensal e os Vulneráveis até R\$ 291. Utilizando a Tabela 50 percebe-se que a população dos “Extremamente Pobres” e “Pobres” está até os dois primeiros décimos e os Vulneráveis até o terceiro décimo de renda familiar per capita. A Tabela 51 confirma o avanço distributivo para além dos dois primeiros décimos, principalmente em estados no Norte e Nordeste, além da disparidade entre os Estados. Então, levanta-se questionamento sobre a lógica distributiva vigente que não parece tão eficiente como apresentam estudos baseados em médias gerais por Estado (SOUZA 2011, p. 171).

Os recursos são escassos e sua origem, em grande parte, vem do PIS (Programa de Integração Social), da COFINS (Contribuição para o Financiamento da Seguridade Social) e da CSLL (Contribuição sobre o lucro líquido) o que pode gerar impactos duvidosos sobre a economia e emprego, conforme apresentado pelo autor, neste sentido seu uso deve ser racionalizado.

5.2 Cotas raciais nas universidades e a estrutura observada na rede bayesiana.

Enquanto nas classes mais altas existem poucos negros/pretos, pardos e índios, nas classes mais baixas esta proporção é bem maior, como pode ser observado na Tabela 53.

Existe na sociedade descontentamento com tamanha desigualdade e no poder público a necessidade de resposta através de políticas sociais. Assim, surgem as ações inclusivas e afir-

mativas visando mitigar o problema e reparar “questões históricas”. Estas propostas incluem a adoção de cotas raciais.

Ações afirmativas são segundo Gomes (2001) “políticas e mecanismos de inclusão concebidas por entidades públicas, privadas e por órgãos dotados de competência jurisdicional, com vistas à concretização de um objetivo constitucional universalmente reconhecido – o da efetiva igualdade de oportunidades a que todos os seres humanos têm direito”.

CLASSE ABEP	TOTAL DE FAMÍLIAS	%NEGROS, PARDOS E ÍNDIOS	EVIDENCIA
A1+A2	1,373,217	17%	Muito Abaixo da Media
B1	3,391,303	20%	Muito Abaixo da Media
B2	7,691,457	30%	Muito Abaixo da Media
C1	11,734,804	41%	Bem Abaixo da Media
C2	13,131,289	55%	Acima da Media
D	15,895,402	65%	Bem Acima da Media
E	3,672,420	79%	Muito Acima da Media
Total	56,889,893	50%	Media

Tabela 53 : Percentual de chefes de família que se declararam negros/pretos, pardos e indígenas.

FONTE : Dados da POF 2009

Neste trabalho pode-se observar que a cor da pele (denominado de forma simplista como raça ou cor) está muito associada à geografia. Em estados do Norte e Nordeste existem muito mais negros/pretos, pardos e indígenas do que nos estados do Sul e Sudeste. No entanto a análise do DNA da população brasileira nas diferentes regiões, feita por Pena et al. (2011), mostra como o Brasil é miscigenado, independente da cor que a população estampa na pele. Assim, muitos desconhecem a sua história.

As atuais ações inclusivas olham de forma simplista apenas para duas variáveis visando mobilizar “a tecnologia social” em prol da solução de um problema complexo. Neste sentido este exercício de reflexão toma apenas estas variáveis simplificadoras e avalia o impacto sobre a percepção de dificuldade das famílias brasileiras, bem como sobre a questão das cotas raciais.

O modelo evidencia, a partir da sequência de arcos destacados na Figura 66, que o fato do responsável pela família ser negro, pardo ou indígena interfere na classificação socioeconômica da família e que esta classificação, por sua vez, interfere na percepção de dificuldade.

Temos então uma configuração linear destes nós, conforme fora apresentado no item 3.2.3. Ou seja, é certo que a dificuldade depende da cor ou raça, mas se conheço a classe econômica de uma família então a dificuldade independe da raça e cor. Explicando ainda de outra forma temos: as setas destacadas da Figura 66 indicam que ao se saber a classe econômica não é necessário conhecer a raça ou cor para detectar se uma família vive em dificuldade.

Da observação desta sequência de arcos e nós, surge a pergunta: duas pessoas, uma branca e outra negra ou parda, dentro da mesma classe econômica têm percepções muito diferentes sobre a dificuldade orçamentária enfrentada pelas suas famílias ?

Figura 66 : Relação entre Raça e Cor e a Dificuldade - Modelo TO

Para responder esta pergunta de forma mais cuidadosa, isolou-se apenas as variáveis raça/cor, classe econômica e renda monetária per capita e dificuldade. Utilizando métrica Kullback-Leibler, pode-se checar a força dos arcos entre estas variáveis comprovando o baixo poder informativo da raça/cor neste contexto. A dificuldade foi representada de duas formas: reclassificada em apenas duas categorias conforme já visto nesta tese (DIFICULDADE_REC= SIM ou NÃO) e também na forma original (DIFICULDADE =

Muita Dificuldade, Dificuldade, Alguma Dificuldade, Alguma Facilidade, Facilidade, Muita Facilidade).

Observa-se na Figura 67, através da métrica de Kullback-Leibler que não é mesmo possível sustentar a seta que parte diretamente da raça/cor para a “Dificuldade”.

Figura 67 : Relação entre Raça e Cor, Renda Per Capita, Classe Econômica e Dificuldade.

Cada uma das formas de representar a “dificuldade” foram contrastadas com as variáveis “Classe de Renda Per Capita Familiar” e “Classe Econômica”, gerando 4 possibilidades analíticas identificadas como (a), (b), (c) e (d).

As 4 versões testadas comprovam que quando já se conhece a classe econômica ou a renda per capita familiar a informação sobre raça/cor pouco informa sobre se uma família está em dificuldade ou não.

Neste sentido, questiona-se a necessidade de criar ações que usem a informação da cor da pele para oferecer acesso às políticas públicas inclusivas quando a classe econômica ou a renda familiar per capita são suficientes para identificar a população carente e excluída.

Além disso, se considerar, por exemplo, que 50% da população da Classe C é branca, amarela ou não sabe sua cor, haveria injustiça na implementação de cotas raciais, pois estas famílias declaram também estar em dificuldade e também estão, na sua maioria, nas escolas públicas brasileiras, necessitando das mesmas oportunidades.

Para tornar a análise mais clara propõe-se observar a proporção de famílias com dificuldade, por “Classe Econômica”, comparando-as pela cor da pele. Dado o que fora apresentado na Figura 67 espera-se pouco descolamento entre as curvas, conforme Gráfico 33.

Gráfico 33 - Dificuldade percebida e variação entre raça e cor dentro das mesmas classes econômicas.

O mesmo pode ser observado por Classe de Renda per Capita Familiar, onde se observa uma diferença ainda menor.

Gráfico 34 - Dificuldade percebida e variação entre raça e cor dentro das mesmas faixas de renda familiar per capita – em cada categoria estão 10% das famílias

Como se pode observar pelos gráficos 33 e 34, a diferença entre as curvas é mínima, dado a cor da pele, o que já estava representado através da força dos arcos na Figura 67. Neste sentido, levanta-se a questão : as ações inclusivas envolvendo raça ou cor são realmente justas ?

Este exercício mostra que ao se saber a classe econômica ou a renda monetária per capita, a cor da pele nada mais informa sobre o fato de uma pessoa estar em dificuldade. Estando em dificuldade, ela estará “privada dos meios e instrumentos de competição” e consequentemente da “igualdade de oportunidades”. Nas palavras de Gomes (2001, p. 61)

“... a discriminação entendida sob essa ótica como uma privação de meios ou de instrumentos de competição resulta igualmente em privação de oportunidades. Consequentemente, reduzem-se as perspectivas. Para a teoria da justiça compensatória, a melhor forma de correção e de reparação desse estado de coisas consistiria em aumentar (via ações afirmativas) as chances dessas vítimas históricas de obterem os empregos e as posições de prestígio que elas naturalmente obteriam caso não houvesse discriminação”

Qual a história dos brancos que vivem na classe C, D e E ? As suas histórias não importam? Por que estes brancos vivem com tanta dificuldade? Ao estabelecer medidas que favoreçam uma cor em detrimento da outra não seria uma tremenda simplificação e injustiça, ainda mais sabendo que passam pelas mesmas dificuldades orçamentárias e desconhecem a sua história?

Observando os dados e estudando-os com cautela pergunta-se ainda : Se existirem cotas no Brasil, não deveriam ser por classe econômica, por renda familiar per capita ou por probabilidade de estar em dificuldade ?

Como dito anteriormente, o objetivo é levantar perguntas importantes a partir de um exercício numérico e lógico que confere novas perspectivas analíticas para o problema das cotas raciais.

6 LIMITAÇÕES DO ESTUDO E SUGESTÕES DE MELHORIA

O modelo foi desenvolvido em apenas uma amostra, decidiu-se manter uma grande quantidade de casos, dado que muitas tabelas conjuntas são estabelecidas nas redes bayesianas. Assim, priorizou-se manter o maior número de casos possível. Para lidar com esta limitação utilizou-se o método de partição de “K” amostras para a validação do modelo final

A POF 2003 pode ser utilizada para validar o modelo. Este outro período permitirá também avaliar o comportamento do modelo em outro contexto (de dificuldades ainda maiores). Esta validação não coube no cronograma, dado que a criação, processamento das variáveis e principalmente a identificação do modelo na POF 2009 foram atividades trabalhosas. Além disso, as variáveis da POF 2003 não são totalmente compatíveis com a POF 2009, o que requer cuidados adicionais com o tratamento das variáveis.

Redes Bayesianas é um tema novo para administradores e áreas afins e com mais conhecimento sobre o assunto os modelos podem ficar ainda melhores. O presente trabalho pode conter erros por desconhecimento de alguns aspectos da técnica. O assunto tem bibliografia dispersa e tecnicamente complexa. Espera-se com este trabalho ajudar aos administradores e áreas afins na compreensão das redes bayesianas visando raciocinar a partir dos dados, cada vez mais disponíveis. No entanto, considera-se este passo apenas um início, deve-se lembrar do que recomenda Pereira (2006, p. 4) :

Devemos sempre questionar métodos que estão sendo aceitos universalmente para não permitir que sejam utilizados equivocadamente. Esse questionamento deve ser sobre a lógica do método e não sobre seu uso imediato. Isto porque, ao entender completamente um método, o profissional saberá quando e como usá-lo.

Esta tese pretende abrir caminho para a busca de técnicas alternativas visando apoiar o raciocínio e a tomada de decisão em estudos sociais, organizacionais e mercadológicos e em especial para as redes bayesianas.

A escolha das variáveis e as validações foram realizadas com base nas heurísticas pessoais da pesquisadora, neste sentido a subjetividade pode ser considerada como limitadora do presente estudo.

7 CONCLUSÃO

Para mensurar a dificuldade percebida pelas famílias brasileiras para adequar orçamento e renda utilizou-se inicialmente uma variável colhida a partir da resposta sobre a percepção de dificuldade informada à POF (Pesquisa de Orçamentos Familiares 2008_2009). Com base nesta variável e em outras selecionadas desenvolveu-se um modelo baseado em Redes Bayesianas. A técnica permitiu calcular a probabilidade de cada família estar com dificuldade e ao mesmo tempo evidenciou a relação entre as variáveis através de grafos acíclicos, o que possibilitou uma melhor percepção sobre os problemas básicos enfrentados pelas famílias no Brasil.

Foi adotado um processo de mineração de dados e sendo a rede bayesiana um “playground” de variáveis algumas reflexões não planejadas puderam ser vislumbradas como no exemplo apresentado envolvendo a questão das cotas raciais. A estrutura da rede fez pensar que a relação entre raça e dificuldade passa pela classe econômica das famílias, o que oferece novas perspectivas analíticas sobre este fato.

Optou-se durante a condução deste trabalho por não incluir variáveis relacionadas ao valor da renda no modelo, dado a dificuldade de obtê-las em pesquisas de forma geral. No entanto, o modelo de probabilidade obtido, mesmo sem a variável renda e suas derivadas, mas contando com a “proxy” do Critério Brasil, demonstrou ser uma alternativa para identificar a percepção de dificuldade das famílias, podendo ser adotado como um critério complementar em avaliações sócio econômicas.

Mesmo tendo sido adotado um processo de aprendizado de estrutura através de dados, os algoritmos permitiram inclusão de algumas crenças no modelo, as quais foram declaradas na forma de “arcos proibidos” entre variáveis ou mesmo modificando a direção das setas dos grafos. O processo foi, ao mesmo tempo, tecnológico e artesanal. Os modelos selecionados permitiram vivenciar que diferentes relações podem ocorrer entre as variáveis, considerando os processos computacionais e as heurísticas utilizadas. Este fato permite perceber, na prática analítica, que um modelo é apenas uma das representações possíveis de uma realidade complexa, carrega consigo potencialidades e limitações da simplificação adotada pelo pesquisador e pela tecnologia. Existem muitos outros modelos possíveis e este é apenas um deles.

Os modelos desenvolvidos, nas diferentes fases da modelagem foram utilizados para “agenciar” a compreensão da complexidade, criando um espaço para a reflexão, assim espera-se

que a “maquina abstrata” desenvolvida seja útil, que permita observar e agir sobre as dificuldades reais e mais prementes dos brasileiros. Ao conseguir estabelecer um modelo estrutural sobre a dificuldade das famílias brasileiras e associa-lo a um modelo de probabilidades tem-se um avanço em termos de modelagem aplicada à compreensão de problemas sociais.

O modelo representa a dificuldade orçamentária das famílias brasileiras de forma multidimensional e identifica que seu combate está nas questões estruturais, não se trata apenas de carência de renda. Os nós do problema puderam ser identificados em macro áreas que representam a miséria (questão da alimentação e necessidade de serviços de saúde), pobreza (precariedade das casas e atraso nas despesas), infraestrutura (impactando na educação, saúde e lazer), desenvolvimento político e regional (dependendo de onde se vive ou nasce a dificuldade pode ser maior), desigualdade (representada pelas classes econômica e acesso aos melhores empregos), capital humano (representado pelas famílias e pela rede de ensino que tem nas mãos o brasileiro do futuro).

O desafio é pensar um Brasil diferente, que não esteja olhando para as classes econômicas atuais apenas sob a óptica do consumo. O Brasil retratado neste trabalho tem mais de 60% da população vivendo e cuidando de dificuldades básicas. Como o Brasil pretende chegar em 2030? Provavelmente não é apenas pela seta do assistencialismo ou do consumo.

REFERÊNCIAS BIBLIOGRÁFICAS SELECIONADAS

ABEP, 2008 B. **Adoção do CCEB 2008, Critério de Classificação Econômica Brasil.**

Disponível em: <<http://www.abep.org/novo/CMS/Utils/FileGenerate.ashx?id=13>>. Acesso em: 29 Mai. 2011.

ABEP, 2007, **Breve histórico dos critérios sócio econômicos no Brasil**,

<http://www.viverbem.fmb.unesp.br/docs/classificacaobrasil.pdf>

AGRESTI, Alan. **Categorical Data Analysis**. 2.ed. New York: John Wiley and Sons, 2002. 710p.

ALVES, José Eustáquio Diniz. **O Bônus demográfico e o crescimento econômico no Brasil**, 2006. Disponível em: <http://www.abep.nepo.unicamp.br/docs/PopPobreza/Alves.pdf>

Acesso em: 22 Mai. 2011

ANAESTHETIST.COM. Disponível em :

<<http://www.anaesthetist.com/mnm/stats/roc/Findex.htm>>. Acesso em 29 Mai. 2011

ANNIBAL C. A. **Validação do poder discriminante das classificações de operações de crédito das instituições financeira**. 2008. Dissertação de Mestrado -Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2008.

ASHBY, W. Ross. **Design for a brain: The origin of adaptive behavior**. 1st ed.

London: Chapman & Hall, 1952. Disponível em:

<<http://archive.org/stream/designforbrainor00ashb#page/40/mode/2up>>. Acesso em: 05 Abr. 2011.

BAI, Xue. **Tabu Search Enhanced Markov Blanket Classifier for High Dimensional Data Sets**. Center for Automated Learning Discovery. School of Computer Science. Carnegie Mellon University. Pittsburgh, 2005

BANCO MUNDIAL, **Envelhecendo em um Brasil mais Velho, Implicações do envelhecimento populacional para o crescimento econômico, a redução da pobreza, as finanças públicas e a prestação de serviços**, relatório, 2011.62p

BARATA, R. B. **Causalidade e Epidemiologia. História da Saúde. Manguinhos IV, Congresso**, 1, 32-49, mar a jun 1997. disponível em : <http://www.scielo.br/pdf/hcsm/v4n1/v4n1a02.pdf>, acesso : 01/11/2010.

BARKER, Alan. Simplicity. **The Stanford Encyclopedia of Philosophy**. Summer 2011. Disponível em: <http://plato.stanford.edu/archives/sum2011/entries/simplicity/>. Acesso em: 04 Dez. 2011.

BARROS, Ricardo Paes (org). INSTITUTO DE PESQUISA ECONÔMICA APLICADA IPEA. Texto Para Discussão 1263, **A Desigualdade de Renda no Brasil Encontra-se Subestimada? Uma Análise Comparativa com Base na PNAD, na POF e nas Contas Nacionais**. Rio de Janeiro, março de 2007. Disponível em http://www.ipea.gov.br/sites/000/2/publicacoes/tds/td_1263.pdf. Acesso em: 11 Mai. 2011

BAYESIA . Disponível em: <<http://www.bayesia.com>>. Acesso em: 04 Mai. 2012.

BAZERMAN, M. H. **Processo Decisório: para cursos de Administração e Economia e MBAs**. Rio de Janeiro: Elsevier, 2004.

BECKER, Howard S., **Segredos e Truques da Pesquisa**, Zahar, 2008, 295 p; trad. Maria Luiza X. de Borges;

BENDER, E. A., **Mathematical methods in Artificial Intelligence**. Palo Alto, CA : IEEE Computer Socitey Press, 1996

B-COURSE. Disponível em: <<http://b-course.cs.helsinki.fi/obc/>>. Acesso em: 12 Jun. 2012

BLODGETT, J.; ANDERSON, R. A **Bayesian network model of the consumer complaint process**. Journal of Service Research, v. 2, n. 4, p.321-339, May, 2000

BLUNDELL, Richard; BROWNING, Martin; MEGHIR, Costas. **Consumer demand and the life-cycle allocation of household expenditures. The Review of Economic Studies**, University of Oxford, Institute of Economics and Statistics, Oxford, UK : v.61,n.1,p.57-80, Jan 1994.

BRAGA, A. C. S. **Curvas ROC : Aspectos Funcionais e Aplicações. 2000.** Tese de Doutorado -Engenharia de Produção e Sistemas, Métodos Numéricos e Estatísticos, Universidade do Minho, Portugal, Dez. 2000.

BUSSAB, Wilton O ; FERREIRA, M. **Crítério Brasil de Estratificação Socioeconômica: Aspectos Demográficos.** In: CLADEA, 34, 1999, San Juan, Porto Rico.

CAMORIM, Ricardo Dória Loyola. **Codificação de canal com taxa variável.** 2010. Tese de Doutorado - Escola Politécnica, Departamento de Eletrônica e de Computação, Universidade Federal do Rio de Janeiro, 2010.

CARDOSO, Fernando Henrique. **O papel da Oposição.** Revista Interesse Nacional, Abril, 2011. Disponível em: <<http://interessenacional.uol.com.br>>. Acesso em: 22 Abr. 2011.

CARVALHO, Alexandra M. **Scoring functions for learning Bayesian networks.** Disponível em: <www.lx.it.pt/~asmc/pub/talks/09-TA/ta_pres.pdf>. Acesso em: 20 Mai. 2011

CHAUÍ, Marilena. **Convite à Filosofia.** Ática, São Paulo, 2000, 567p.

CHEN, Z. **Data Mining and Uncertain Reasoning: An Integrated Approach.** New York: John Wiley & Sons, 2001. 370p.

CONRADY , Stefan; JOUFFE, Lionel. **Introduction to Bayesian Networks, Practical and Technical Perspectives,** white paper, February 15, 2011. Disponível em: <<http://www.conradyscience.com>>. Acesso em: 22 Abr. 2011

CONRADY, Stefan; JOUFFE, Lionel. **Probabilistic Latent Factor Induction and Statistical Factor Analysis - A Comparison of Methods**. Abril 011. Disponível em: <<http://www.conradyscience.com>>. Acesso em: 15 Mai.2011.

COOPER G. F. **Probabilistic Inference Using Belief Networks is NP-Hard**. Artificial Intelligence, v. 42, p. 393-405, 1987.

COOPER, L. G. **Strategic Marketing Planning For Radical New Products**. Journal of Marketing, v. 64, n. 1, p.1-16, 2000.

CUI, Geng; WONG, Man Leung; LUI, Hong-kwong. **Machine Learning For Direct Marketing Response Models: Bayesian Network With Evolutionary Programming**. Management Science, v. 52, n. 4, Abril, p. 597-612, 2006

CUNHA, Maria B. A. M., SILVA, Pedro L. N., VASCONCELLOS M. T. L.; **Análise comparativa de conceitos em pesquisas domiciliares POF, PNAD e Censo Demográfico**, Trabalho apresentado no XV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu- MG -Brasil, Setembro de 2006.

D'AMBROSIO, Bruce. **Inference in Bayesian Networks**. AI Magazine, v. 20, n. 2, p. 21-36, 1999.

DARWICHE, Adnan. Bayesian Networks. **Communications of the ACM**, v.53, p. 80-90, dez, 2010. Disponível em: <<http://www.ics.uci.edu/~dechter/courses/ics-275b/spring-11/bn-cacm.pdf>> Acesso em: 22 Abr. 2011

DARWICHE, Adnan. **Modeling and Reasoning With Bayesian Networks**. New York: Cambridge University Press, 2009 (versão eletrônica) D'AMBROSIO, Bruce. Página Pessoal. Disponível em: <<http://web.engr.oregonstate.edu/~dambrosi/bayesian/frame.html>>. Acesso em: 23 Mai. 2011

ENGEL ,J. ;Blackwell R.D.; MINIARDI Paul W . **Consumer Behavior : The Dryden Press** 1995, 951p.

FERNANDES, Florestan. **A integração do negro na sociedade de classes**. São Paulo: Editora Globo, v.2, 2008, 624p.

FOUTO, Numo M. Martins Dias. **Mercado de Consumo Brasileiro: Evolução e Determinantes do Volume de Vendas, Uma Análise do Período Pós Real**. 2011. Tese de Doutorado - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2011.

FRANCISCO, E. R. **Indicadores de Renda baseados em Consumo de Energia Elétrica: Abordagens Domiciliar e Regional na Perspectiva da Estatística Espacial**. Tese (doutorado) - Escola de Administração de Empresas de São Paulo, 2010

FRANCO, B. M. **PT Supera PSDB em Briga Pela Nova Classe Média**. Folha de São Paulo, A4, 22 Abr. 2011.

FURTADO, Celso. **Formação Econômica do Brasil**. 27. ed. São Paulo: Companhia Editora Nacional: Publifolha, 2000. 276p.

GILLY, Mary C.; ENIS, Ben. M. **Recycling the family life cycle: a proposal for redefinition** em : Advances in Consumer Research Volume 09, eds. Andrew Mitchell, Ann Abor : Association for Consumer Research, Pages: 271-276, 1982

GOMES, Joaquim B. Barbosa. Ação Afirmativa & Princípio Constitucional da Igualdade. O Direito como Instrumento de Transformação Social. A Experiência dos EUA. Renovar, Rio de Janeiro/São Paulo, 2001, p. 41.

GRAY, Robert M. **Entropy and Information Theory**; Stanford University, Information Systems Laboratory; Electrical Engineering Department; Stanford University 16/07/2009. Disponível em: <<http://ee.stanford.edu/~gray/it.pdf>>. Acesso em: 16 Mai.2012

GUIBU, Fábio, CARVALHO, Daniel. **Sertanejo tem máquina de lavar, só que falta água**, Folha de São Paulo, Cotidiano, C6, 29 Abr. 2012.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L. & BLACK, W. C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

HAND, D. J.; HENLEY; W. E. **Statistical classification methods in consumer Credit Scoring: a review**. Journal of the Royal Statistical Society: Series A (Statistics in Society), v.160, n.3, p.523-541, 1997

HECKERMAN, D. , **A Tutorial Learning With Bayesian Network**, Microsoft Research, Março de 1995, Revisado em Novembro de 1996, Technical Report, Advanced Technology Division, Microsoft Corporation. Disponível em: <<http://research.microsoft.com/en-us/um/people/heckerman/tutorial.pdf>>. Acesso em: 22 Abr. 2011

HECKERMAN, D. **A tutorial on learning with Bayesian networks: learning in graphical models**. Cambridge, MA: MIT Press, 1999.

HECKERMAN, D. **Bayesian Networks for Knowledge Discovery in Advances. In: Knowledge Discovery and Data Mining**. American Association for Artificial Intelligence, Cambridge: MIT Press, p.273-305, 1996

HECKERMAN, D.; HORWITZ, E.; NATHWANI, B. **Toward normative expert systems: Part I, The Pathfinder Project. Methods of Information in Medicine**, v. 31, p. 90-105, 1992.

HECKERMAN, D; GEIGER, D; CHICKERING, D M. **Learning Bayesian networks: The combination of knowledge and statistical data**. Redmond: Microsoft Research, 1994.

HIROSHI, Sugihara. **What is Occam's Razor?** Atualização em 1997, original por: Phil Gibbs, 1996, Physics FAQ. Disponível em: <<http://math.ucr.edu/home/baez/physics/General/Ockham.html>>

HOOGENBERG, Marc; BRINKE, Sander auf dem. **Decision making in an uncertain environment: An application of ROC analysis for credit scoring in mobile telephone mar-**

ket. Journal of Database Marketing & Customer Strategy Management, v. 11, n. 3, p. 241-254, Apr, 2004

HOSMER, David. W.; LEMESHOW, Stanley Jr. **Applied logistic regression**. New York: John Wiley & Sons, 2nd ed., 2000. 375p.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Pesquisa de Orçamentos Familiares, POF 2002-2003 E POF 2008 -2009**. Disponível em: <<http://www.ibge.gov.br>>. Acesso em 22 Abr. 2011

IBGE, Instituto Brasileiro de Geografia e Estatística, **Pesquisa de Orçamentos Familiares, POF 2008 -2009 , Despesa, Rendimento e Condições de Vida** , Janeiro de 2010. Disponível em:

<http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/pof/2008_2009/POFpublicacao.pdf> . Acesso em: 07 Mai. 2011

IBGE, Instituto Brasileiro de Geografia e Estatística, **Pesquisa de Orçamentos Familiares, POF 2008 -2009 , Despesa, Rendimento e Condições de Vida** , Junho de 2010. Disponível em:<http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=1648&id_pagina=1>. Acesso em: 07 Mai. 2011

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA - **Carta de Conjuntura** 2011, número12, Séries Históricas, Disponível em:

<http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=8005:carta-de-conjuntura-no-12&catid=107:dimac&Itemid=3>. Acesso em: 15 Mai. 2011.

JENSEN, F. V.; NIELSEN T. D. **Bayesian Networks and Decision Graphs . Information Science & Statistics Series**. 2nd ed. Heidelberg: Springer, fev. 2007

JORDAN, Michael I. **Learning In graphical models**. Cambridge, MA: MIT Press, 1999.

JOUFFE, L.; MUNTEAU, P. **New Search Strategies for Learning Bayesian Networks**. in:International Symposium on applied stochastic models, data analysis, Compiègne. Pro-

ceedings of Tenth International Symposium on Applied Stochastic Models, Data Analysis. Compiègne, jun., 2001

KAMAKURA, Wagner A. ; MAZZON, José A. **Socioeconomic Status and Consumption in an Emerging Economy**. Forthcoming at the International Journal of Research in Marketing, In press, Dez 2011.

KARCHER, Cristiane. **Redes Bayesianas Aplicadas à Análise do Risco de Crédito**. 2009. Dissertação de Mestrado - Escola Politécnica, Universidade de São Paulo, 2009.

KEENEY, R.L. **Making Better Decisions Makers**. Decisions Analysis, v. 1, n. 4, Dez, p. 193-204, 2004

KOTLER, Philip; KELLER Kevin L. **Administração de Marketing** : Prentice Hall, 2006.

LAM, W.; BACCHUS, F. **Learning Bayesian belief networks: An approach based on the MDL principle**. Computatinal Intelligence, v. 10, n. 4, 1994.

LAMOUNIER, B.; SOUZA, A. **A Classe Média Brasileira, Ambições, Valores e Projetos de Sociedade**. Rio de Janeiro: Elsevier, 2010. 174p.

LECUMBERRI, L.F.L; DUARTE, A. M. **Uma Metodologia para o gerenciamento de modelos de escoragem em operações de crédito de varejo no Brasil**. Revista de Economia Aplicada, v. 7, n. 4, p. 795-818, 2003.

LEVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro: Editora 34, 1993.

LIMEIRA, Tânia M. V. **Comportamento do Consumidor Brasileiro** : Saraiva, 2009, 377.p

LING, C. X.; ZHANG, H. **The representational power of discrete Bayesian networks**. The Journal of Machine Learning Research, v. 3, p. 709-721, 2003.

LUNA, José Eduardo Ochoa. **Algoritmos EM para Aprendizagem de Redes Bayesianas a Partir de Dados Incompletos**. 2004. Dissertação de Mestrado - Departamento de Computação e Estatística do Centro de Ciências Exatas e Tecnologia da Universidade de Mato Grosso do Sul, Mato Grosso do Sul, 2004.

LYNCH, Scott M. **Introduction to Applied Bayesian Statistics and Estimation for Social Scientists**. Springer, 2010, 357p.

MATSUURA, Jackson Paul. **Discretização para aprendizagem bayesiana : aplicação no auxílio à validação de dados em proteção ao voo**. Dissertação de Mestrado, Instituto Tecnológico de Aeronáutica, São José dos Campos, SP -Brasil; 2003.

MATTAR, Fauze N. **Estudo sobre Estratificação Social para Utilização em Marketing e Pesquisas de Marketing : Proposta de Novo Modelo para Estratificação Sócio-Econômica**. 1995. Tese (Livre-Docência em Administração) : FEA-USP, São Paulo.

MÁXIMO, Luciano. **Desistência do Bolsa Família por iniciativa própria chega a 40%**. Folha de São Paulo, Reportagem, Brasil, 17 Out. 2011.

MCDONALD, John H, **Handbook of Biological Statistics** . University of Delaware, Baltimore: Maryland: Sparky House Publishing. Disponível em: <<http://udel.edu/~mcdonald/statintro.html>> . Acesso em: 19 Mai. 2011

MEDICI, André. **Propostas para melhorara a cobertura, a eficiência e a qualidade no setor saúde**. In: BACHA, E. L.; SCHWARTZMAN, S (orgs). A Nova Agenda Social. Livros Técnicos e Científicos Editora, 2011, 365p.

MENEZES, Naercio Filho. **Educação, Bolsa Família e Desigualdade**. Valor Econômico, Opinião. 20 Jan. 2012

MINSKI, M. **Steps towards artificial intelligence**. IN: Proceedings of the Institute of Radio Engineers, v. 49, p. 8-30, 1961.

MORIN, Edgar; LE MOIGNE, Jean-Louis. **A Inteligência da Complexidade**. São Paulo: Petrópolis, 2000.

MURPHY, P.E.; STAPLES, W.A. **A modernized family life cycle**. Journal of Consumer Research, v. 6, p. 12-22, 1979.

MUNTEAU, P., BENDOU, M. **The EQ Framework for Learning Equivalence Classes of Bayesian Networks**. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING (IEEE ICDM), 1, 2001. Proceedings of First IEEE International Conference on Data Mining, San José, nov. 2001.

NASCIMENTO, Elimar Pinheiro. **Dos excluídos necessários aos excluídos desnecessários**. In: BURSZTYN, Marcel (org). No meio da rua. Nômades, excluídos e viradores. Rio de Janeiro: Garamond, 2000, p. 56-87..

NEAPOLITAN, R. E.; JIANG, X. **Probabilistic Methods for Financial and Marketing Informatics**. 1st ed. San Francisco: Morgan Kaufmann, 2007.

NERI, M. C. **A Nova Classe Média**. Rio de Janeiro: Fundação Getúlio Vargas, Instituto Brasileiro de Economia, Centro de Políticas Sociais, 2008.

NERI, Marcelo Cortes; **O Lado Brilhante dos Pobres**, Rio de Janeiro, FGV, CPS, 2010, 149p. Disponível em <<http://www.fgv.br/cps/nem>>. Acesso em : 01 Mai. 2011.

PAULA, G.A. **Modelos de Regressão Com Apoio Computacional**. Versão Preliminar, Instituto de Matemática e Estatística, USP, Jun 2004.

PEARL, J. **Bayesian networks: A model of self-activated memory for evidential reasoning**. In: CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY, 7, 1985. Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, Agosto, 1985, p. 329–334.

PEARL, J. **Causality, Reasoning and Inference**. 2 ed. New York: Cambridge University Press, 2009.

PEARL, J. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. San Mateo, CA: Morgan Kaufmann Publishers, 1988.

PEARL, J. **The Foundation of Causal Inference**. Sociological Methodology, v. 40, p. 75–149. Disponível em: <http://ftp.cs.ucla.edu/pub/stat_ser/r355-corrected-reprint.pdf>. Acesso em: 01 Mar. 2010

PEARL, J.; RUSSEL, S. Bayesian Networks. In: ARBIB, M. A. (ed.). **The Handbook of Brain Theory and Neural Networks**. 2nd ed. Cambridge: MIT Press, 2003. Disponível em: <<http://www.cs.berkeley.edu/~russell/papers/hbttnn-bn.pdf>>. Acesso em: 22 Abr. 2011

PEARL, Judea. **Probabilistic Reasoning in Intelligent Systems: networks of plausible inference**. 2nd ed. California, San Mateo: Morgan Kaufmann Publishers, 1988. 552p.

PEDRETI CHAGAS, Ricardo. **Aplicação de Redes Bayesianas na Previsão de Crescimento de Fluxos de Caixa**. 2007. Tese de Mestrado, Fundação Getúlio Vargas, 2007.

PEDROSO, RODRIGO. **Nova classe média ganha até R\$ 1.019 per capita**, Valor Econômico, em 30.05.2012.

PENA, Sergio D.J; DI PIETRO Giuliano; FUCHSHUBER-MORAES Mateus; GENRO Julia Pasqualini ; HUTZ Mara H; KEHDY, Fernanda de Souza; KOHLRAUSCH Fabiana; MAGNO Luiz Alexandre Viana ; MONTENEGRO Raquel Carvalho ; MORAES Manoel Odorico ; MORAES Maria Elisabete Amaral ; MORAES Milene Raiol ; OJOPI Élidea B.; PERINI Jamila A. ; RACCIOPI Clarice ; SANTOS Ândrea Kely Campos Ribeiro-dos; SANTOS Fabrício Rios- ; SILVA Marco A. Romano, SORTICA Vinicius A.; KURTZ Guilherme Suarez-3. **The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected** ; PLoS ONE 6(2) : e 17063. doi : 10.1371/journal.pone.0017063;
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0017063>, acesso em 07/04/2012.

PEREIRA, Carlos Alberto Bragança. **Estatística Aplicada : Uma Visão Genuinamente Bayesiana**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 15. Anais da Decima Quinta Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, Instituto de Biociências da Universidade Estadual Paulista (UNESP), 2006.

PESSOA JR.; Osvaldo. **Modelos causais em história da ciência**. In: MARTINS, R. A.; MARTINS, L.A.C.P.; SILVA, C.C.; FERREIRA, J.M.H. (eds.). *Filosofia e História da Ciência no Cone Sul: 3º Encontro*. Campinas: AFIHC, 2004. p. 341-346. Disponível em: <<http://www.fflch.usp.br/df/opessoa/AFHIC-04.pdf>> Acesso em: 28 Ago. 2011.

PINEDA, José Octávio de Carvalho. **A entropia segundo Claude Shannon: O Desenvolvimento do Conceito Fundamental da Teoria da Informação**. 2006. Dissertação de Mestrado, Pontifícia Universidade Católica (PUC), São Paulo, 2006

PNUD. **Programa das Nações Unidas para o Desenvolvimento**. Disponível em: <<http://www.pnud.org.br/idh/>>. Acesso em: 9 Mai. 2012.

POZENATO, José Clemente; **Algumas considerações sobre região e regionalidade. A região: uma rede de relações, Letras e Cultura Regional**. Disponível em: <https://www.ucs.br/ucs/tplPadrao/tplPOSLetras/posgraduacao/strictosensu/letras/professores/jose_pozenato/imprimir/artigo.pdf>. Acesso em: 30 Jan. 2012

PRICE, Richard; **Transcrição da carta de Price para a Royal Society apresentando o manuscrito original de Bayes**. Disponível em: <<http://www.stat.ucla.edu/history/essay.pdf>>. Acesso em: 04 Mai. 2012.

RAMONI, M.; SEBASTIANI, P. **Learning Bayesian networks from incomplete databases**. IN: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL, 1995. Proceedings of Conference On Uncertainty In Artificial, Morgan Kaufman, 1995. p.401-408.

REDONDO-BELLÓN, Ignacio Redondo; ROYO-VELA, Marcelo; ALDAS-MANZANO, Joaquim. **A family life cycle model adapted to the Spanish environment**. European Journal of Marketing, v. 35, p. 612-38, 2001

RISSANEN, J. **Modeling by shortest data description**. Automatica, v. 14, p. 465-471, 1978.

ROCHA, Thelma. **Marketing de Relacionamento e Competitividade no Mercado Empresarial : Um Estudo de Caso em Uma Empresa Multinacional Agroquímica**. 2007. Tese (Doutorado em Administração) -Faculdade de Economia e Administração -FEA- Universidade de São Paulo, São Paulo, 2007.

SARAIWA JR., Francisco Ison. **Em Busca de Um Modelo Brasileiro de Ciclo de Vida Familiar para Segmentação de Mercado**. 2005. Dissertação de Mestrado, Fundação Getúlio Vargas, São Paulo, 2005.

SHANNON, Claude E. **A Mathematical Theory of Communication**, Vol XXVII, n,3, Jul, 1948, p. 379-422.

SICSU , Abraham Laredo, **Credit Scoring : Desenvolvimento, Implantação e Acompanhamento**. Blucher, 2010.

SILANDER, Tomi. **The most probable bayesian network and beyond**, Phd Thesis, Series os Publications A, Report A-2009, Helsinki, Abril, 2009.

SETZER, Valdemar W. **Dado, Informação, Conhecimento e Competência, Departamento de Ciência da Computação**, Universidade de São Paulo Disponível em: <<http://www.ime.usp.br/~vwsetzer/dado-info.html>>, Acesso em: 15 Mai. 2011.

SILVEIRA, Fernando G. et al.. (Org). **Gasto e Consumo das Famílias Brasileiras Contemporâneas**. Brasília: IPEA, 2006, v.1

SILVEIRA, Fernando G. et al.. (Org). **Gasto e Consumo das Famílias Brasileiras** Contemporâneas. Brasília: IPEA, 2007, v.2

SOKAL, R. R. , ROHLF, F. J. **Biometry: the principles and practice of statistics in biological research**, New York: Freeman, 1981.

SOUZA, André. **Políticas de distribuição de renda no Brasil**. In: BACHA, E. L.; SCHWARTZMAN, S. **A Nova Agenda Social**. Livros Técnicos e Científicos Editora, 2001, p. 165-186.

SPIEGELHALTER, D. J.; KNILL-JONES, R. P. **Statistical and Knowledge-based approaches to clinical decision-support systems**. Journal of the Royal Statistical Society, Series A, v. 147, n.1, p. 35-77, 1984.

SPIRITES, P.; CLARK, G.; RICHARD, S. **Causation, Prediction, and Search**, 2nd ed. Cambridge: MIT Press, 2001.

STANFORD. **Enciclopédia de Filosofia**. Disponível em: <<http://plato.stanford.edu/entries/bayes-theorem/>>. Acesso em: 04 Mai. 2012.

STERN, Julio Michael. **Construtivismo cognitivo e estatística bayesiana**. Scienti e Studia, São Paulo, v. 4, n. 4, p. 563-588, 2006. Disponível em: <http://www.scientiaestudia.org.br/revista/pdf/04_04_02.pdf>. Acesso em: 10 Mai. 2011.

SUZUKI, J. **Learning Bayesian belief networks based on the MDL principle: An efficient algorithm using the branch and bound technique**. Paper In : IECE TRANS.INF &SYST., VOL E82-D, NO. 2. Feb, 1999. P. 356 – 367. Disponível em : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.7807&rep=rep1&type=pdf>, acesso 04/06/2012

THE ECONOMIST, **Brazil Takes Off**, Special report on Latin America's big success history , Nov. 2009, 14 p.

UNPP, United Nations Population Division, Department of Economic and Social Affairs **World Population Prospects: The 2010 Revision**, Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, 2010 Disponível em: <<http://esa.un.org/unpd/wpp/index.htm>>. Acesso em: 23 Mai. 2011

USP - UNIVERSIDADE DE SÃO PAULO Biblioteca Digital de Teses e Dissertações da USP. Disponível em: <<http://www.teses.usp.br>>. Acesso em: 04 Mai. 2012.

VASSÃO, Caio Adorno, **Metadesign : Ferramentas , estratégia e ética para a complexidade** . Blucher, 2010, 134p.

VASSÃO, Caio Adorno, **Arquitetura Livre : Complexidade, Metadesign, Ciência Nômade**, 2008. Tese Doutorado, FAU -Universidade de São Paulo, 2008.

VEIGA, Samuel C. A. **Grau de Informação de Evidências em Redes Bayesianas**, Dissertação de Mestrado, Universidade de Brasília, CIC -Departamento de Ciência da Computação - Inteligência Artificial, Brasília, 2002. Disponível em: <<http://www.samuelveiga.pro.br/arq/Redes%20Bayesianas%20%20Uma%20visao%20geral.pdf>>. Acesso em: 01 Jun. 2011.

WELLS, W.D.; GUBAR, G. **Life cycle concept in marketing research**. Journal of Marketing Research, v. 3, p. 355-63, 1966.

WIKIPEDIA. **G-Test**. Disponível em <<http://en.wikipedia.org/wiki/G-test>>. Acesso em: 18 Mai. 2011.

WIKIPEDIA., **Causalidade**. <http://pt.wikipedia.org/wiki/Causalidade> Acesso em: 20 Dez. 2011.

WITTEN, Ian H.; FRANK, Eibe; HALL Mark A. **Data Mining: Practical Machine Learning Tool and Techniques**. Elsevier, 2011. 629p.

WONNACOTT, Thomas H. WONNACOTT, Ronald. **Introductory Statistics for Business and Economics**. 4th ed. Canada: John Wiley & Sons, 1990. 815p.

ZIKMUND, William G. **Exploring marketing research**. 6th ed. Orlando: The Dryden Press, 1997

REFERÊNCIAS BIBLIOGRÁFICAS COMPLEMENTARES

ALBA, Henrique de; MENDOZA, Manuel. **Bayesian forecasting methods for short time series**. Foresight: The International Journal of Applied Forecasting, v.8, p.41-44, 2007. Disponível em: <<http://allman.rhon.itam.mx/~mendoza/Foresight.pdf>>. Acesso em: 01 Nov. 2011

BARBER, D. **Bayesian Reasoning and Machine Learning**. Disponível em: <<http://www.cs.ucl.ac.uk/staff/d.barber/brml>>. Acesso em: 22 Abr. 2011

BARROS, Ricardo Paes. **Bolsa Família Parte Agora Para a Inserção Produtiva, Rumos da Economia**, entrevista dada a Chico Santos. p. F7. Valor Econômico, 30 Abr. 2011

BAUMGARTEN, Maira; TEIXEIRA, Alex N.; LIMA, Gilson. **Sociedade e conhecimento: novas tecnologias e desafios para a produção de conhecimento**. Sociedade e Estado, Brasília, v. 22, n. 2, p. 401-433, maio/ago. 2007. Disponível em: <<http://www.scielo.br/pdf/se/v22n2/06.pdf>>. Acesso em : 10/04/2011>. Acesso em: 10 Abr. 2011

BELLO, Laurentino; GONZÁLEZ, Ana M. **The construct “lifestyle” in Market segmentation**. European Journal of Marketing, v. 36, n. 1/2, p. 51-85, 2002

BOUNTINE, W. **Theory Refinement on Bayesian Networks**. In: Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publisher, San Mateo, CA, 52-60, 1991.

BOX, G. E. P.; DRAPER, N. R. **Empirical Model-Building and Response Surfaces**. New York: John Wiley & Sons, 1987. 663 p.

CASTILLO, E.; GUTIÉRRES, J. M.; HADI, A. S. **Expert systems and probabilistic network models**. Monographs in Computer Science, Springer-Verlag, New York, 1997.

CLEMEN, Robert T. **Making hard decisions**. 2nd.ed. Belmont: Duxbury Press, 1996.

GAAG, Linda C. **Bayesian Belief Networks: Odds and Ends**. The Computer Journal, v. 39, n. 2, p. 97-113, 1996

HUGIN. Disponível em: < <http://www.hugin.com/>>. Acesso em: Mai. 2012.

HULSWIT, Menno. **A Short History of Causation**. University of Nijmegen. Disponível em: <<http://www.library.utoronto.ca/see/SEED/Vol4-3/Hulswit.htm>>, Acesso em: 28 Ago. 2011.

IFRAH, Georges. **História Universal dos Algarismos**. Rio de Janeiro: Nova Fronteira, Tomo 2, Segunda Ed, 2000, 1046p.

JOUFFE, L. **Nouvelle classe de méthodes d'apprentissage de réseaux bayésiens**. Journées francophones d'Extraction et de Gestion des Connaissances (EGC), Montpellier, France, janvier 2002.

KOLLER, Daphne.; NIR Friedman. **Probabilistic Graphical Models: Principles and Techniques**. Cambridge: MIT Press, 2009.

McCARTHY, J. HAYES Patrick J., **Some Philosophical Problems From the Standpoint of Artificial Intelligence**, 1969, Machine Intelligence , 4, Computer Science Department, Stanford University, Stanford, CA 94305, Disponível em: < <http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html>>. Acesso em: 22 Abr. 2011

MITCHELL, A. A. (Ed.). **Advances in Consumer Research**, v.9, Association for Consumer Research, p. 271-6, 1982

MOWEN, John C.; MINOR, Michael S. **Comportamento do Consumidor**. São Paulo; Prentice Hall, 2003.

NASSAR, Silvia Modesto. **Material Didático**. Disponível em:
http://www.inf.ufsc.br/~silvia/disciplinas/sep/material_didatico.
 Acesso em: 27 Mai. 2010

NETICA. Disponível em: <<http://www.norsys.com/>>. Acesso em: Mai. 2012

PUTLER, Daniel S.; KALYANAM, Kirthi; HODGES, James S. **A Bayesian Approach for Estimanting Target Market Potencial With Limited Geodemographic Information**. Journal of Marketing Research, v. 33, May,1996.

PUTLER, Daniel S; LI Tieshan ; LIU, Yong; **The effectiveness of household life-cycle models in predicting household expenditure behavior: an empirical examination**, Canadian Journal of Administrative Sciences, 24 (4): 284-299.

SCHANINGER, C.M.; DANKO, W.D. **A conceptual and empirical comparison of alternative household life cycle models**. Journal of Consumer Research, v. 19, March, p. 580-94, 1993.

SPIR, P.; CLARK, G.; RICHARD, S. **Causation, Prediction, and Search**, 2nd ed. Cambridge: MIT Press, 2001.

SZAFIR-GOLDSTEIN, Cláudia. **O conceito de valor percebido e as estratégias de marketing: estudo de caso em empresa fornecedora de sistemas de gestão empresarial**. Dissertação (Mestrado em Administração) -Programa de Pós-Graduação em Administração, Faculdade de Economia, Administração e Contabilidade. Universidade de São Paulo, São Paulo, 2000.

US CENSUS BUREAU. Table 685. **Average Annual Expenditures of all Consumer Units by Race and Age of Householder**. Disponível em:
 <www.census.gov/compendia/statab/2011/tables/11s0685.xls> . Acesso em: 17 Mai. 2011

US CENSUS BUREAU. **The 2011 Statistical Abstract The National Data Book**. Disponível em: <<http://www.census.gov/compendia/statab>>. Acesso em: 17 Mai. 2011.

WILSON, Robert A.; KEIL Frank C. **The MIT Encyclopedia Of The Cognitive Sciences.**
First MIT Press Paperbak Edition, 2001. 831p.

APÊNDICE 1A: VARIÁVEIS POR ORDEM ALFABÉTICA

* O NÚMERO DA VARIÁVEL É O MESMO QUANDO AGRUPADAS EM FATORES

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
V	71	AREA_GEO	AREA_GEO	Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água
V	77	ATRASO_DE_ALUGUEL_OU_PRESTACAO_DA_RESIDENCIA	FATOR_7_MORA_DE_ALUGUEL	X Informa se a família atrasou aluguel ou prestação da casa
V	11	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	Declara já ter atrasado contas de água, eletricidade ou gás.
V	12	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	Declara já ter atrasado outras contas relativas a bens ou serviços adquiridos.
V	14	CADERNETA_COM_DIFERENCIADOS	CADERNETA_COM_DIFERENCIADOS	Informa presença de itens diferenciados na caderneta de despesa.
V	44	CASA_ESCURA	CASA_ESCURA	A pessoa declarou que percebe sua casa escura.
V	35	CLASSE_ABEP	CLASSE_ABEP	Classe ABEP - agregado da pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família, número de banheiros e presença de domésticos.
V	38	COMODOS_DORMITORIOS	FATOR_5_ESPACO_CASA	X Número de cômodos que estão sendo utilizados como dormitório no domicílio.
V	46	COMPROM_RENDA	COMPROM_RENDA	Comprometimento de renda. Considerando somatório de todas as despesas pela renda total.
V	47	CONDICAO_OCUPACAO_REC	CONDICAO_OCUPACAO_REC	Apresenta a condição de ocupação do domicílio: alugado, cedido ou próprio.
V	31	CONDICOES_DE_MORADIA	CONDICOES_DE_MORADIA	A pessoa declarou sua percepção geral sobre a sua moradia.
V	51	CONJ_OU_RESP_ESTUDANDO	CONJ_OU_RESP_ESTUDANDO	Indicador se o cônjuge ou o responsável estão estudando.
V	32	DIFICULDADE2	DIFICULDADE_REC	Percepção sobre a adequação da renda agrupada em duas categorias, sendo que uma delas agrupa: Muita Dificuldade, Dificulda-

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
				de e Alguma Dificuldade
V	15	DIVERSIDADE_DESPESAS_ ALIMENTACAO	DIVERSIDADE_ DESPESAS_ ALIMENTACAO	Apresenta a diversidade observada em itens de alimentação.
V	57	ESCOADOURO_SANITARIO	FATOR_13_ESGOT O_ESCOADOURO_ AGUA_CANALIZA DA	X Presença de escoadouro sanitário.
V	7	ESCOLARIDADE_CONJUGE	ESCOLARIDADE_ CONJUGE	Informa a escolaridade do cônjuge ou se não tem cônjuge.
V	52	ESCOLARIDADE_MENORES	ESCOLARIDADE_ MENORES	Situação da escolaridade dos menores de 18 anos.
V	36	ESCOLARIDADE_PESS_REF ERENCIA	ESCOLARIDADE_ PESS_ REFERENCIA	Escolaridade da pessoa de referência ou responsável pela família.
V	57	EXISTE_AGUA_ CANALIZADA	FATOR_13_ESGOT O_ESCOADOURO_ AGUA_CANALIZA DA	X Presença de água canalizada.
V	58	EXISTENCIA_PAVIMENTAC AO_RUA	FATOR_23_LIMPE ZA_MANUTEN CAO_PAVIMEN TACAO	X Indica sobre a pavimentação da rua onde mora.
V	13	FAMILIAS_E_IDOSOS	FATOR_8_IDOSO_ COM_APOSENTA DORIA	X Indica se no domicílio tem idoso recebendo aposentadoria.
V	45	FUNDACAO_PAREDES_OU_ CHAO_UMIDOS	FATOR_12_GOTEI RA_FUNDACAO_P AREDE_CHAO_U MIDOS_MADEIRA _DETERIORADA	X A pessoa declarou ter paredes e chão úmidos.
V	65	GRUPO_DE_RELIGIOES	GRUPO_DE_ RELIGIOES	Grupo de religião ao qual pertence o chefe da família.
V	61	INUNDACOES	INUNDACOES	A pessoa apresenta inundações no local onde mora.
V	45	MADEIRAS_DAS_JANELAS_ PORTAS_OU ASSOALHOS_ DETERIORADOS	FATOR_12_GOTEI RA_FUNDACAO_ PAREDE_CHAO_ UMIDOS_MADEIR A_DETERIORADA	X A pessoa declarou ter janelas, portas ou assoalhos deteriorados.
V	80	MATERIAL_PREDOMINA_C OBERTURA	FATOR_24_ TIPO_DOMICILIO_ COBERTURA	X Tipo do material para cobertura da moradia
V	79	MATERIAL_PREDOMINA_P AREDES_EXTERNAS	FATOR_22_MATE RIAL PAREDES_PISO	X Tipo do material para revestimento das paredes externas
V	79	MATERIAL_PREDOMINA_PI SO	FATOR_22_MATE RIAL PAREDES_PISO	X Tipo do material predominante para a cobertura ou não do piso.

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
V	16	NECESS_MEDICAMENTO	FATOR_27_NECES SITOU_SERV_ SAUDE_OU_ MEDICAMENTO	X Indica se a família necessitou de medicamen- tos nos últimos meses, mas não foi atendida neste sentido.
V	16	NECESS_SERVICOS_SAUDE	FATOR_27_NECES SITOU_SERV_ SAUDE_OU_ MEDICAMENTO	X Indica se a família necessitou de serviços de saúde nos últimos meses, mas não foi atendi- da neste sentido.
V	53	NINHOS_POR_FAIXA_ETAR IA	FATOR_10_ FAMILIAS	X Tipos de família considerando a idade do filho mais novo.
V	53	NR_DEPENDENTES_REC	FATOR_10_ FAMILIAS	X Número de dependentes de renda na família.
V	22	PERC_A_CREDITO	FATOR_19_CREDI TO_E_OU_HABITA CAO_DIFERENC	Faixa Percentual que representa impacto de gastos com crédito e habitação.
V	17	PERC_ALIMENTACAO	PERC_ALIMENTA CAO	Faixa Percentual que representa impacto de gastos com alimentação.
V	48	PERC_ALUGUEL_ ESTIMADO	PERC_ALUGUEL_ ESTIMADO	Faixa Percentual de aluguel estimado sobre total de despesas.
V	18	PERC_DESP_INDIV	PERC_DESP_ INDIV	Faixa Percentual que representa impacto de gastos com despesas individuais.
V	49	PERC_HABITACAO	PERC_ HABITACAO	Faixa Percentual que representa impacto de gastos com habitação “não diferenciados”.
V	23	PERC_HABITACAO_DIF	PERC_ HABITACAO_DIF	Faixa Percentual que representa impacto de gastos com habitação “diferenciados”.
V	19	PERC_OUTRAS	PERC_OUTRAS	Faixa Percentual que representa impacto de gastos com outras despesas.
V	20	PERC_OUTRAS_DIF	PERC_OUTRAS _DIF	Faixa Percentual que representa impacto de gastos com outras despesas diferenciadas.
V	50	PERC_RENDA_NAO_MONET	PERC_RENDA_ NAO_MONET	Faixa Percentual da renda que não é monetá- ria.
V	39	PERC_TRANSPORTE_DIF	PERC_TRANSPOR TE_ DIFERENCIADO	Faixa Percentual que representa impacto de gastos com transporte diferenciado.
V	42	PERC_TRANSPORTE	PERC_TRANSPOR TE	Faixa Percentual que representa impacto de gastos com transporte.
V	25	PERC_VAL ALIM_MIN	PERC_VAL ALIM_ MIN	Percentual da renda mínima que designaria para a alimentação dentro da renda que de- clarou como necessária para viver.
V	21	PERC_VESTUARIO	PERC_VESTUARIO	Faixa Percentual que representa impac- to de gastos com vestuário.
V	78	PESSOAS_ESTUDANDO_sum	FATOR_14_ FAMILIAS	X Presença de pessoas estudando no domicílio.

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
			PESSOAS_ ESTUDANDO_ REDE_ENSINO	
V	72	POLUICAO_OU_PROBLEMA S_AMBIENTAIS	FATOR_1_VIOLÊN CIA_VADALISMO_ POLUIÇÃO	X No local onde mora há problemas ambien- tais.
V	40	PONTOS_ABEP	PONTOS_ABEP	Pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, esco- laridade do chefe da família e número de banheiros.
V	38	POUCO_ESPACO	FATOR_5_ESPACO _CASA	X A pessoa declarou ter pouco espaço na casa.
V	1	PREVIDENCIA	PREVIDENCIA	Indica se alguém da família contribui para a previdência social ou se está aposentado.
V	59	PROVENIENCIA_AGUA	FATOR_3_ACESSO _A_AGUA_RURAL	X Proveniência da água utilizada no domicílio.
V	38	QTDE_BANHEIROS	FATOR_5_ESPACO _CASA	X Quantidade de banheiros no domicílio.
V	38	QTDE_COMODOS	FATOR_5_ESPACO _CASA	X Quantidade de cômodos no domicílio.
V	53	QTDE_FILHOS_REC	FATOR_10_FAMILI AS	X Quantidade de filhos na família.
V	54	QTDE_MORADORES	QTDE_MORADOR ES	Quantidade de moradores na família.
V	33	QUANTIDADE_DE_ALIMEN TO_REC	QUANTIDADE_DE _ALIMENTO_REC	A pessoa declarou sua percepção se a quanti- dade de alimentos está adequada às necessi- dades familiares.
V	55	QUANTIDADE_FAMILIAS	QUANTIDADE_FA MILIAS	Indica se existe mais de uma família moran- do no mesmo domicílio.
V	66	RACA_COR_PESSOA_ REFERENCIA	RACA_COR_ PESSOA_ REFERENCIA	Raça ou cor do chefe da família.
V	67	REGIAO	REGIAO	Região Geográfica representada pela UF.
V	2	RENDAS_PESSOAS_UC	RENDAS_PESSOAS_UC	Indica o tipo de rendimento que predomina na família (mais ou menos que salário míni- mo)
V	26	RENDAS_MONET_NCESS	RENDAS_MONET_ NECESS	Quanto da renda declarou precisar para viver até o final do mês.
V	27	RENDAS_MONET_PER_CAPI TA	RENDAS_MONET_P ER_CAPITA	Faixa de renda monetária per capita.
V	75	RENDAS_PREDOMINANTE	FATOR_4_TIPOS_ DE_RENDA	X Identifica o tipo de renda predominante no domicílio.
V	28	RENDAS_TOT_ANO	RENDAS_TOT_ANO	Faixa de renda total recebida no ano.

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	<u>ORIGINAIS OU AGRUPADAS EM FATORES (X)</u>	SIGNIFICADO
V	72	RUA_OU_VIZINHOS_BARULHENTOS	FATOR_1_VIOLENCIA_VADALISMO_POLUIÇÃO	X No local onde moram os vizinhos são barulhentos.
V	59	SERVICO_DE_AGUA	FATOR_3_ACESSO_A_AGUA_RURAL	X Tipo de serviço de água disponível para o domicílio.
V	60	SERVICO_DE_COLETA_DE_LIXO	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X Acesso a serviços de coleta de lixo.
V	41	SERVICO_DE_EDUCACAO	FATOR_11_EDUC_SAÚDE_LAZER	X Acesso a serviços de educação
V	62	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	A pessoa apresentou sua percepção sobre escoamento de água no local onde mora
V	57	SERVICO_DE ESGOTAMENTO_SANITARIO	FATOR_13_ESGOTO_ESCOAD OURO	X Tipo de serviço de esgotamento sanitário disponível no domicílio
V	60	SERVICO_DE_FORNECIMENTO_DE_ENERGIA_ELETRICA	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X Acesso a serviços de fornecimento de energia elétrica
V	60	SERVICO_DE_ILUMINACAO_DE_RUA	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X Acesso a serviços de fornecimento de iluminação de rua
V	41	SERVICO_DE_LAZER_E ESPORTE	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	X Acesso a serviços de lazer e esporte
V	58	SERVICO_DE_LIMPEZA_E_MANUTENCAO_DE_RUA	FATOR_23_LIMPEZA_MANUTENCAO_PAVIMENTACAO	X Acesso a serviços de limpeza e manutenção de ruas
V	41	SERVICO_DE_SAUDE	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	X Acesso a serviços de saúde
V	43	SERVICO_DE_TRANSPORTE_COLETIVO	SERVICO_DE_TRANSPORTE_COLETIVO	A pessoa apresentou sua percepção sobre transporte coletivo no local onde mora
V	8	SEXO	SEXO	Sexo da pessoa responsável pelo domicílio
V	56	STATUS_REDE_ENSINO	STATUS_REDE_ENSINO	Situação da rede de ensino utilizada pela família: escolas e creches, particulares e públicas
V	45	TELHADO_COM_GOTEIRA	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERI	X A pessoa declarou ter telhado com goteira

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
			ORADA	
V	37	TEM_CARTAO_DE_CREDITO	FATOR_17_CARTAO_OU_TCH_ESPECIAL	X Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Cartão de Crédito
V	9	TEM_CONJUGE_sum	TEM_CONJUGE_sum	Indica se o responsável possui cônjuge.
V	29	TEM_PLANO_DE_SAUDE	TEM_PLANO_DE_SAUDE	Indica se algum morador possui plano de saúde
V	48	TEMPO_ALUGUEL	FATOR_28_MORA_DE_ALUGUEL	X Há quanto tempo a pessoa fez o contrato de aluguel
V	3	TEMPO_SEM_TRAB_PRINC_REC	TEMPO_SEM_TRAB_PRINC_REC	Indica se o chefe da família sempre tem trabalho principal, quanto tempo no ano ficou sem rendimento deste trabalho ou se está aposentado
V	34	TIPO_ALIMENTO_RAZAO_REC	TIPO_ALIMENTO_RAZAO_REC	A pessoa declarou sua percepção se os tipos de alimentos estão adequados e a razão de não se alimentar como quer.
V	63	TIPO_AREA	TIPO_AREA	Tipo da área onde mora
V	47	TIPO_CONTRATO_ALUGUEL	FATOR_28_MORA_DE_ALUGUEL	X Apresenta o tipo de contrato de aluguel, sendo formal ou informal.
V	80	TIPO_DOMICILIO	FATOR_24_TIPO_DOMICILIO_COBERTURA	X Informa em que tipo de domicílio mora : casa, apto.
V	69	TIPO_VEICULO	TIPO_VEICULO	Indica se veículo prioritário é carro, moto, bicicleta, nenhum.
V	37	TITULAR_DA_CONTA_CORRENTE (APENAS QDO TEM CHEQUE ESPECIAL)	Fator_17_CARTAO_OU_TCH_ESPECIAL	X Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Conta Corrente _ apenas quando possui cheque especial
V	75	UF	FATOR_2_COR_RACA_LOCALIDADE	X Sigla das Unidades da Federação Brasileira
V	70	UF_GRUPO	UF_GRUPO	Agrupamentos das Unidades da Federação considerando percepção média de dificuldade em cada uma delas
V	59	URBANO	FATOR_3_ACESSO_A_AGUA_RURAL	X Indica se urbano ou Rural
V	72	VIOLENCIA_OU_VANDALISMO	FATOR_1_VIOLENCIA_VANDALISMO_POLUIÇÃO	X No local onde mora há presença de violência e vandalismo.
V	30	X_OUTRAS_RENDAS	X_OUTRAS_RENDAS	Indica presença de outras rendas no domicílio (ex aluguel, rendimento de aplicações, etc).
V	13	X_RENDA_APOSENTADORIA	FATOR_8_IDOSO_COM_APOSENTADORIA	X Indica se no domicílio tem idoso recebendo aposentadoria
V	4	X_RENDA_BENEFIC_ASSALARIADO	X_RENDA_BENEFIC_ASSALARIADO	Indica se no domicílio há recebimento de

V	*	<u>VARIÁVEIS INICIALMENTE CONSTRUÍDAS</u>	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
				benefícios típicos dos assalariados EX: Recebe décimo terceiro, bônus de férias, vale alimentação.
V	5	X_RENDA_CONTA_PROPRI A	X_RENDA_CONTA _PROPRIA	Indica se no domicílio tem renda de trabalha- dor por conta própria
V	6	X_RENDA_EMPREGADO	X_RENDA_EMPRE GADO	Indica se no domicílio tem renda de empre- gado
V	24	X_RENDA_EMPREGADOR	X_RENDA_EMPRE GADOR	Indica se no domicílio tem renda de empre- gador
V	10	X_RENDA_PENSAO_MESAD A_DOACAO	X_RENDA_PENSA O_MESADA_DOA CAO	Indica se no domicílio existe renda de pen- são, mesada ou doação.
V	68	X_RENDA_PREV_PRIVADA	X_RENDA_PREV_ PRIVADA	Indica se no domicílio tem renda de previ- dência privada
V	64	X_RENDA_TRANSF_GOVER NO	X_RENDA_TRANS F_GOVERNO	Indica se no domicílio de renda de transfe- rências do governo

APÊNDICE 1B: VARIÁVEIS POR ORDEM NUMÉRICA

* O NÚMERO DA VARIÁVEL É O MESMO QUANDO AGRUPADAS EM FATORES

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
V	1	PREVIDENCIA	PREVIDENCIA	Indica se alguém da família contribui para a previdência social ou se está aposentado
V	2	RENDA_DAS_PESSOAS_UC	RENDA_DAS_PESSOAS_UC	Indica o tipo de rendimento (salários mínimos) que predomina na família
V	3	TEMPO_SEM_TRAB_PRINC_REC	TEMPO_SEM_TRAB_PRINC_REC	Indica se o chefe da família sempre tem trabalho principal, quanto tempo no ano ficou sem rendimento deste trabalho ou se está aposentado
V	4	X_RENDA_BENEFIC_ASSALARIADO	X_RENDA_BENEFIC_AS SALARIADO	Indica se no domicílio há recebimento de benefícios típicos dos assalariados EX: Recebe décimo terceiro, bônus de férias, vale alimentação.
V	5	X_RENDA_CONTA_PROPRIA	X_RENDA_CONTA_PROPRIA	Indica se no domicílio tem renda de trabalhador por conta própria
V	6	X_RENDA_EMPREGADO	X_RENDA_EMPREGADO	Indica se no domicílio tem renda de empregado
V	7	ESCOLARIDADE_CONJUGE	ESCOLARIDADE_CONJUGE	Informa a escolaridade do cônjuge ou se não tem cônjuge
V	8	SEXO	SEXO	Sexo da pessoa responsável pelo domicílio
V	9	TEM_CONJUGE	TEM_CONJUGE	Indica se a pessoa possui cônjuge
V	10	X_RENDA_PENSAO_MESADA_DOACAO	X_RENDA_PENSAO_MESADA_DOACAO	Indica se no domicílio existe renda de pensão, mesada ou doação.
V	11	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	ATRASO_DE_DESPESAS_COM_AGUA_ELETRICIDADE_E_GAS	Declara já ter atrasado contas de água ou eletricidade ou gás.
V	12	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	ATRASO_DE_PRESTACOES_DE_BENS_OU_SERVICOS_ADQUIRIDOS	Declara já ter atrasado outras contas relativas a bens ou serviços adquiridos
V	13	FAMILIAS_E_IDOSOS	FATOR_8_IDOSO_COM_APOSENTADORIA	X Indica se no domicílio tem idoso recebendo aposentadoria
V	13	X_RENDA_APOSENTADORIA	FATOR_8_IDOSO_COM_APOSENTADORIA	X Indica se no domicílio tem idoso recebendo aposentadoria
V	14	CADERNETA_COM_DIFERENCIADOS	CADERNETA_COM_DIFERENCIADOS	Informa presença de itens diferenciados na caderneta de despesa
V	15	DIVERSIDADE_DE_SPESAS_ALIMENTACAO	DIVERSIDADE_DESPESAS_ALIMENTACAO	Apresenta a diversidade observada em itens de alimentação
V	16	NECESS	FATOR_27	X Indica se a família necessitou de medicamentos nos últi-

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
		MEDICAMENTO	NECESSITOU_SERV_SAUDE_OU_MEDICAMENTO	mos meses, mas não foi atendida neste sentido.
V	16	NECESS_SERVICO_S_SAUDE	FATOR_27_NECESITOU_SERV_SAUDE_OU_MEDICAMENTO	X Indica se a família necessitou de serviços de saúde nos últimos meses, mas não foi atendido neste sentido.
V	17	PERC_ALIMENTACAO	PERC_ALIMENTACAO	Faixa Percentual que representa impacto de gastos com alimentação
V	18	PERC_DESP_INDIV	PERC_DESP_INDIV	Faixa Percentual que representa impacto de gastos com despesas individuais
V	19	PERC_OUTRAS	PERC_OUTRAS	Faixa Percentual que representa impacto de gastos com outras despesas
V	20	PERC_OUTRAS_DIF	PERC_OUTRAS_DIF	Faixa Percentual que representa impacto de gastos com outras despesas diferenciadas
V	21	PERC_VESTUARIO	PERC_VESTUARIO	Faixa Percentual que representa impacto de gastos com vestuário
V	22	PERC_A_CREDITO	FATOR_19_CREDITO_E_OU_HABITACAO_DIFERENC	Faixa Percentual que representa impacto de gastos com crédito e habitação
V	23	PERC_HABITACAO_DIF	PERC_HABITACAO_DIF	Faixa Percentual que representa impacto de gastos com habitação diferenciada
V	24	X_RENDA_EMPREGADOR	X_RENDA_EMPREGADOR	Indica se no domicílio tem renda de empregador
V	25	PERC_VAL ALIM_MIN	PERC_VAL ALIM_MIN	Percentual da renda mínima que designaria para a alimentação dentro da renda que declarou como necessária para viver.
V	26	RENDA_MONET_NECES	RENDA_MONET_NECES	Quanto da renda declarou precisar para viver até o final do mês
V	27	RENDA_MONET_PER_CAPITA	RENDA_MONET_PER_CAPITA	Faixa de renda monetária per capita
V	28	RENDA_TOT_ANO	RENDA_TOT_ANO	Faixa de renda total recebida no ano
V	29	TEM_PLANO_DE_SAUDE	TEM_PLANO_DE_SAUDE	Indica se algum morador possui plano de saúde
V	30	X_OUTRAS_RENDAS	X_OUTRAS_RENDAS	Indica presença de outras rendas no domicílio (ex aluguel, rendimento de aplicações, etc)
V	31	CONDICOES_DE_MORADIA	CONDICOES_DE_MORADIA	A pessoa declarou sua percepção geral sobre a sua moradia
V	32	DIFICULDADE2	DIFICULDADE_REC	Percepção sobre a adequação da renda agrupada em duas categorias, sendo que uma delas agrupa: Muita Dificuldade, Dificuldade e Alguma Dificuldade
V	33	QUANTIDADE_DE_	QUANTIDADE_DE_ALI	A pessoa declarou sua percepção se a quantidade de ali-

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
		ALIMENTO_REC	MENTO_REC	mentos está adequada às necessidades familiares.
V	34	TIPO_ALIMENTO_RAZAO_REC	TIPO_ALIMENTO_RAZAO_REC	A pessoa declarou sua percepção se os tipos de alimentos estão adequados e a razão de não se alimentar como quer.
V	35	CLASSE_ABEP	CLASSE_ABEP	Classe ABEP - agregado da pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família, número de banheiros e presença de domésticos
V	36	ESCOLARIDADE_PESS_REFERENCIA	ESCOLARIDADE_PESS_REFERENCIA	Escolaridade da pessoa de referência ou responsável pela família
V	37	TEM_CARTAO_DE_CREDITO	FATOR_17_CARTAO_OU_TCH_ESPECIAL	X Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Cartão de Crédito
V	37	TITULAR_DA_CONTA_CORRENTE (APENAS QDO TEM CHEQUE ESPECIAL)	FATOR_17_CARTAO_OU_TCH_ESPECIAL	X Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Conta Corrente _ apenas quando possui cheque especial
V	38	COMODOS_DORMITÓRIOS	FATOR_5_ESPACO_CASA	X Número de cômodos que estão sendo utilizados como dormitório no domicílio
V	38	POUCO_ESPACO	FATOR_5_ESPACO_CASA	X A pessoa declarou ter pouco espaço na casa
V	38	QTDE_BANHEIROS	FATOR_5_ESPACO_CASA	X Quantidade de banheiros no domicílio
V	38	QTDE_COMODOS	FATOR_5_ESPACO_CASA	X Quantidade de cômodos no domicílio
V	39	PERC_TRANSPORTE_DIF	PERC_TRANSPORTE_DIFERENCIADO	Faixa Percentual que representa impacto de gastos com transporte diferenciado
V	40	PONTOS_ABEP	PONTOS_ABEP	Pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família e número de banheiros.
V	41	SERVICO_DE_EDUCACAO	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	X Acesso a serviços de educação
V	41	SERVICO_DE_LAZER_E_ESPORTE	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	X Acesso a serviços de lazer e esporte
V	41	SERVICO_DE_SAÚDE	FATOR_11_EDUCAÇÃO_SAÚDE_LAZER	X Acesso a serviços de saúde
V	42	PERC_TRANSPORTE	PERC_TRANSPORTE	Faixa Percentual que representa impacto de gastos com transporte
V	43	SERVICO_DE_TRANSPORTE_COLETIVO	SERVICO_DE_TRANSPORTE_COLETIVO	A pessoa apresentou sua percepção sobre transporte coletivo no local onde mora
V	44	CASA_ESCURA	CASA_ESCURA	A pessoa declarou que percebe sua casa escura
V	45	FUNDACAO_PAREDES_OU_CHAO_UMIDOS	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEI	X A pessoa declarou ter paredes e chão úmidos

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)	SIGNIFICADO
			RA_DETERIORADA	
V	45	MADEIRAS_DAS_JANELAS_PORTAS_OU ASSOALHOS_DETERIORADOS	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	X A pessoa declarou ter janelas, portas ou assoalhos deteriorados
V	45	TELHADO_COM_GOTEIRA	FATOR_12_GOTEIRA_FUNDACAO_PAREDE_CHAO_UMIDOS_MADEIRA_DETERIORADA	X A pessoa declarou ter telhado com goteira
V	46	COMPROM_RENDA	COMPROM_RENDA	Comprometimento de renda. Considerando somatório de todas as despesas pela renda total
V	47	CONDICAO_OCUPACAO_REC	CONDICAO_OCUPACAO_REC	Apresenta a condição de ocupação do domicílio : alugado, cedido ou próprio
V	47	TIPO_CONTRATO_ALUGUEL	FATOR_28_MORA_DE_ALUGUEL	X Apresenta o tipo de contrato de aluguel, sendo formal ou informal
V	48	PERC_ALUGUEL_ESTIMADO	PERC_ALUGUEL_ESTIMADO	Faixa Percentual de aluguel estimado sobre total de despesas
V	48	TEMPO_ALUGUEL	FATOR_28_MORA_DE_ALUGUEL	X Há quanto tempo a pessoa fez o contrato de aluguel
V	49	PERC_HABITACAO	PERC_HABITACAO	Faixa Percentual que representa impacto de gastos com habitação (não diferenciada)
V	50	PERC_RENDA_NAO_MONET	PERC_RENDA_NAO_MONET	Faixa Percentual da renda que não é monetária.
V	51	CONJ_OU_RESP_ESTUDANDO_sum	CONJ_OU_RESP_ESTUDANDO_sum	Indicador se o cônjuge ou o responsável estão estudando
V	52	ESCOLARIDADE_MENORES	ESCOLARIDADE_MENORES	Situação da escolaridade dos menores de 18 anos
V	53	NINHOS_POR_FAMILIA_ETARIA	FATOR_10_FAMILIAS	X Tipos de família considerando a idade do filho mais novo
V	53	NR_DEPENDENTES_REC	FATOR_10_FAMILIAS	X Número de dependentes de renda na família
V	53	QTDE_FILHOS_REC	FATOR_10_FAMILIAS	X Quantidade de filhos na família
V	54	QTDE_MORADORES	QTDE_MORADORES	Quantidade de moradores na família
V	55	QUANTIDADE_FAMILIAS	QUANTIDADE_FAMILIAS	Indica se existe mais de uma família morando no mesmo domicílio
V	56	STATUS_REDE_ENSINO	STATUS_REDE_ENSINO	Situação da rede de ensino utilizada pela família : escolas e creches, particulares e públicas
V	57	ESCOADOURO_SANITARIO	FATOR_13_FATOR_ESCOADOURO_AGUA_CANALIZADA	X Presença de escoadouro sanitário
V	57	EXISTE_AGUA_CANALIZADA	FATOR_13_ESGOTO_ESCOADOURO_AGUA_CANALIZADA	X Presença de água canalizada
V	57	SERVICO_DE ESGOTAMENTO_SANITARIO	FATOR_13_ESGOTO_ESCOADOURO_AGUA_CANALIZADA	X Tipo de serviço de esgotamento sanitário disponível no domicílio

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)		SIGNIFICADO
V	58	EXISTENCIA_PAVIMENTACAO_RUA	FATOR_23_LIMPEZA_MANUTENCAO_PAVIMENTACAO	X	Indica sobre a pavimentação da rua onde mora
V	58	SERVICO_DE_LIMPEZA_E_MANUTENCAO_DE_RUA	FATOR_23_LIMPEZA_MANUTENCAO_PAVIMENTACAO	X	Acesso a serviços de limpeza e manutenção de ruas
V	59	PROVENIENCIA_AGUA	FATOR_3_ACESSO_AGUA_RURAL	X	Proveniência da água utilizada no domicílio
V	59	SERVICO_DE_AGUA	FATOR_3_ACESSO_AGUA_RURAL	X	Tipo de serviço de água disponível para o domicílio
V	59	URBANO	FATOR_3_ACESSO_AGUA_RURAL	X	Indica se urbano ou Rural
V	60	SERVICO_DE_COLETA_DE_LIXO	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X	Acesso a serviços de coleta de lixo
V	60	SERVICO_DE_FORNECIMENTO_DE_ENERGIA_ELETRICA	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X	Acesso a serviços de fornecimento de energia elétrica
V	60	SERVICO_DE_ILUMINACAO_DE_RUA	FATOR_6_BOA_ILUM_ENERGIA_COLETA_LIXO	X	Acesso a serviços de fornecimento de iluminação de rua
V	61	INUNDACOES	INUNDACOES		A pessoa apresenta inundações no local onde mora
V	62	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA	SERVICO_DE_ESCOAMENTO_DA_AGUA_DE_CHUVA		A pessoa apresentou sua percepção sobre escoamento de água no local onde mora
V	63	TIPO_AREA	TIPO_AREA		Tipo da área onde mora
V	64	X_RENDA_TRANSF_GOV	X_RENDA_TRANSF_GOV		Indica se no domicílio de renda de transferências do governo
V	65	GRUPO_DE_RELIGIOES	GRUPO_DE_RELIGIOES		Grupo de religião ao qual pertence o chefe da família
V	66	RACA_COR_PESSOA_REFERENCIA	RACA_COR_PESSOA_REFERENCIA		Raça ou cor do chefe da família
V	67	REGIAO	REGIAO		Região Geográfica
V	68	X_RENDA_PREV_PRIVADA	X_RENDA_PREV_PRIVADA		Indica se no domicílio tem renda de previdência privada
V	69	TIPO_VEICULO	TIPO_VEICULO		Indica se veículo prioritário é carro, moto, bicicleta, nenhum
V	70	UF_GRUPO	UF_GRUPO		Agrupamentos das Unidades da Federação considerando percepção média de dificuldade em cada uma delas
V	71	AREA_GEO	AREA_GEO		Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água
V	72	POLUICAO_OU_PROBLEMAS_AMBIENTAIS	FATOR_1_VIOLENCIA_VADALISMO_POLUICAO	X	No local onde mora há problemas ambientais

V	*	VARIÁVEIS INICIALMENTE CONSTRUÍDAS	ORIGINAIS OU AGRUPADAS EM FATORES (X)		SIGNIFICADO
V	72	RUA_OU_VIZINHOS_ BARULHENTOS	FATOR_1_VIOLENCIA_ VADALISMO_ POLUIÇÃO	X	No local onde moram os vizinhos são barulhentos
V	72	VIOLENCIA_OU_ VANDALISMO	FATOR_1_VIOLENCIA_ VADALISMO_ POLUIÇÃO	X	No local onde mora há presença de violência e vandalismo.
V	75	REDA_PREDOMINANTE	FATOR_4_TIPOS_DE_ REDA	X	Identifica o tipo de renda predominante no domicílio
V	75	UF	FATOR_2_COR_RACA_ LOCALIDADE	X	Sigla das Unidades da Federação Brasileira
V	77	ATRASO_DE_ ALUGUEL_OU_PR ESTACAO_DA_RES IDENCIA	FATOR_7_MORA_DE _ALUGUEL	X	Informa se a família atrasou aluguel ou prestação da casa
V	78	PESSOAS_ESTUDANDO	FATOR_14_PESSOAS_ES TUDANDO_REDE_ENSI NO	X	Presença de pessoas estudando no domicílio
V	79	MATERIAL_ PREDOMINA_PAREDES_ EXTERNAS	FATOR_22_MATERIAL PAREDES_PISO	X	Tipo do material para revestimento das paredes externas
V	79	MATERIAL_PREDOMINA_PISO	FATOR_22_MATERIAL PAREDES_PISO	X	
V	80	MATERIAL_PREDOMINA_COBERTURA	FATOR_24_ TIPO_DOMICILIO_COBERTURA	X	Tipo do material para cobertura da moradia
V	80	TIPO_DOMICILIO	FATOR_24_ TIPO_DOMICILIO_COBERTURA	X	Informa em que tipo de domicílio mora: casa, apto

APÊNDICE 2 : VARIÁVEIS SELECIONADAS E RELAÇÃO COM A DIFICULDADE

VERIFICAR NAS PRÓXIMAS PÁGINAS

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
1	PREVIDENCIA PUBLICA									
	APOSENTADO	7,580,674	73%	15%	13%	13%	0.89	-	0.12	Abaixoda Media
	NINGUEM CONTRIBUI	22,803,810	80%	32%	43%	40%	1.34	-	0.29	Acima da Media
	PARTE CONTRIBUI	13,488,372	74%	25%	23%	24%	0.93	-	0.07	Media
	TODOS CONTRIBUEM	13,017,037	69%	28%	21%	23%	0.74	-	0.30	Bem Abaixo da Media
	TOTAL	56,889,893	75%	100%	100%	100%	1.00	-	0.00	Media
Indica se alguém da família contribui para a previdência social ou se está aposentado										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
2	RENDA DAS PESSOAS UC									
	PREDOMINA APOSENTADORIA	7,580,674	73%	15%	13%	13%	0.89	-	0.12	Abaixoda Media
	PREDOMINA MAIS QUE SALARIO MINIMO	29,180,988	68%	66%	47%	51%	0.71	-	0.34	Bem Abaixo da Media
	PREDOMINA SALARIO MENOS QUE SALÁRIO MÍNIMO	20,128,230	86%	20%	40%	35%	2.03	-	0.71	Muito Acima da Media
	PREDOMINA SALARIO MINIMO	1,555,648	88%	1%	3%	3%	2.52	-	0.93	Muito Acima da Media
	TOTAL	56,889,892	75%	100%	100%	100%	1.00	-	0.00	Media
Indica o tipo de rendimento (salários mínimos) que predomina na família										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
3	TEMPO SEM TRAB PRINC REC									
	A SEMPRE COM TRABALHO PRINCIPA	28,112,925	71%	57%	47%	49%	0.82	-	0.20	Abaixoda Media
	B_NAO_TEM_TRABALHO_PRINCIPAL (aposentado)	7,580,674	73%	15%	13%	13%	0.89	-	0.12	Abaixoda Media
	C DE 1 A 3 MESES	7,196,017	79%	11%	13%	13%	1.27	-	0.24	Acima da Media
	D DE 4 A 6 MESES	7,812,689	81%	11%	15%	14%	1.40	-	0.34	Acima da Media
	E MAIS DE 6 MESES	6,187,590	83%	7%	12%	11%	1.66	-	0.50	Bem Acima da Media
	TOTAL	56,889,892	75%	100%	100%	100%	1.00	-	0.00	Media
Indica se o chefe da família sempre tem trabalho principal, quanto tempo no ano ficou sem rendimento deste trabalho ou se está aposentado										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
4	X RENDA BENEFIC ASSALARIADO									
	NAO	19,694,855	82%	26%	38%	35%	1.47	-	0.39	Acima da Media
	SIM	37,195,037	72%	74%	62%	65%	0.84	-	0.18	Abaixoda Media
	TOTAL	56,889,892	75%	100%	100%	100%	1.00	-	0.00	Media
Indica se no domicílio há recebimento de benefícios típicos dos assalariados EX : Recebe décimo terceiro, bonus de férias, vale alimentação.										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
5	X RENDA CONTA PROPRIA									
	NAO	36,842,928	74%	68%	64%	65%	0.94	-	0.07	Media
	SIM	20,046,965	77%	32%	36%	35%	1.13	-	0.13	Media
	TOTAL	56,889,893	75%	100%	100%	100%	1.00	-	0.00	Media
Indica se no domicílio tem renda de trabalhador por conta própria										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
6	X RENDA EMPREGADO									
	NAO	17,666,221	74%	33%	31%	31%	0.94	-	0.06	Media
	SIM	39,223,672	76%	67%	69%	69%	1.03	-	0.03	Media
	TOTAL	56,889,893	75%	100%	100%	100%	1.00	-	0.00	Media
Indica se no domicílio tem renda de empregado										
EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA		EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM		SIM / NÃO			
7	STATUS CONJUGE									
	A ANALFABETO	3,903,864	89%	3%	8%	7%	2.55	-	0.94	Muito Acima da Media
	B. 0 A 3	4,699,300	81%	6%	9%	8%	1.46	-	0.38	Acima da Media
	C. 4 A 7	9,974,840	82%	13%	19%	18%	1.52	-	0.42	Bem Acima da Media
	D. 8 A 10	5,439,337	78%	9%	10%	10%	1.16	-	0.14	Media
	E. 11 A 14	10,860,849	69%	24%	17%	19%	0.73	-	0.31	Bem Abaixo da Media
	F. 15 OU MAIS	2,868,971	45%	11%	3%	5%	0.27	-	1.30	Muito Abaixo da Media
	G. NAO DECLARADO	276,674	76%	0%	0%	0%	1.03	-	0.03	Media
	NAO TEM CONJUGE	18,866,058	75%	34%	33%	33%	0.97	-	0.03	Media
	TOTAL	56,889,893	75%	100%	100%	100%	1.00	-	0.00	Media

CONTINUA ...

8	EVIDÊNCIA									
	SEXO	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	FEMININO	17,538,687	77.01%	29%	32%	31%	1.11	0.10	Media	
	MASCULINO	39,351,207	74.32%	71%	68%	69%	0.96	0.04	Media	
		56,889,894	75.15%	100%	100%	100%	1.00	-	Media	
	Indica o sexo do chefe da família									
9	EVIDÊNCIA									
	TEM CONJUGE sum	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	NAO	18,866,058	75%	34%	33%	33%	0.97	0.03	Media	
	SIM	38,023,836	75%	66%	67%	67%	1.02	0.02	Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
	Indica se o chefe de família tem conjuge									
10	EVIDÊNCIA									
	X RENDA PENSAO MESADA DOACAO	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	NAO	51,252,572	74%	93%	89%	90%	0.96	0.04	Media	
	SIM	5,637,322	82%	7%	11%	10%	1.52	0.42	Bem Acima da Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
	Indica se no domicílio existe renda de pensão, mesada ou doação									
11	EVIDÊNCIA									
	ATRASO DE DESPESAS COM AGUA ELETRICIDADE E G	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	NAO	34,212,607	67%	80%	53%	60%	0.67	0.41	Bem Abaixo da Media	
	NAO SE APLICA	2,388,752	83%	3%	5%	4%	1.63	0.49	Bem Acima da Media	
	SIM	20,288,535	88%	17%	42%	36%	2.47	0.90	Muito Acima da Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
Declara já ter atrasado contas de agua ou eletricidade ou gás.										
12	EVIDÊNCIA									
	ATRASO DE PRESTACOES DE BENS OU SERVICOS ADQU	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	NAO	37,747,866	69%	82%	61%	66%	0.74	0.30	Bem Abaixo da Media	
	NAO SE APLICA	3,119,340	81%	4%	6%	5%	1.45	0.37	Acima da Media	
	SIM	16,022,688	88%	14%	33%	28%	2.40	0.88	Muito Acima da Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
Declara já ter atrasado outras contas relativas a bens ou serviços adquiridos										
13	EVIDÊNCIA									
	FATOR 8 IDOSO COM APOSENTADORIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	NAO	43,590,091	76%	73%	78%	77%	1.07	0.06	Media	
	SIM	13,299,803	71%	27%	22%	23%	0.82	0.20	Abaixoda Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
	Indica se no domicilio tem idoso recebendo aposentadoria									
14	EVIDÊNCIA									
	CADERNETA COM DIFERENCIADOS	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	A NAO COMPROU NENHUMA CATEGORIA DIFERENCIADA	18,861,444	81%	26%	36%	33%	1.38	0.32	Acima da Media	
	B COMPROU 1 CATEGORIA DIFERENCIADA	15,681,392	76%	26%	28%	28%	1.07	0.07	Media	
	C COMPROU 2 CATEGORIAS DIFERENCIADAS	11,577,381	72%	23%	20%	20%	0.86	0.16	Abaixoda Media	
	D SEM INFORMACAO CADERNETA	3,986,165	71%	8%	7%	7%	0.82	0.20	Abaixoda Media	
	E COMPROU 3 OU MAIS CATEGORIAS DIFERENCIADAS	6,783,512	64%	17%	10%	12%	0.60	0.52	Bem Abaixo da Media	
		56,889,894	75%	100%	100%	100%	1.00	-	Media	
Informa presença de itens diferenciados na caderneta de despesa										
15	EVIDÊNCIA									
	DIVERSIDADE DESPESAS ALIMENTACAO	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
				NAO	SIM		SIM / NÃO	WOE	DE DIFICULDADE	
	<=4 TIPOS	19,795,407	76%	34%	35%	35%	1.02	0.12	Media	
	<=7	15,480,138	77%	26%	28%	27%	1.08	0.18	Media	
	<=10	13,091,551	76%	22%	23%	23%	1.03	0.13	Media	
	>10 TIPOS	8,522,797	71%	18%	14%	15%	0.79	0.23	Abaixoda Media	
		56,889,893	75%	100%	100%	100%	1.00	0.10	Media	
Apresenta a diversidade observada em itens de alimentação										

CONTINUA ...

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
FATOR 27 NECESSITOU SERV SAUDE OU MEDICAMENT				NAO	SIM				
16	NAO	39,956,186	68%	89%	64%	70%	0.72	-	0.33 Bem Abaixo da Media
	SIM	16,933,707	91%	11%	36%	30%	3.28	1.19	Muito Acima da Media
		56,889,893	75%	100%	100%	100%	1.00	-	0.00 Media
Indica se a família necessitou de medicamentos nos últimos meses mas não foi atendida neste sentido									

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
IMPACTO ALIMENTACAO				NÃO	SIM				
17	Z SEM DESPESA ALIMENTACAO 0 0	2.126.539	77%	3%	4%	4%	1.10	0.09	Media
	B MUITO ABAIXO 0 0.13	17.673.981	67%	42%	28%	31%	0.66	0.42	Bem Abaixo da Media
	C ABAIXO 0.13 0.20	11.392.685	72%	22%	19%	20%	0.86	0.15	Abaixoda Media
	D MEDIA 0.2 0.25	5.673.892	78%	9%	10%	10%	1.15	0.14	Media
	E ALTO 0.25 0.33	8.574.667	80%	12%	16%	15%	1.35	0.30	Acima da Media
	F MUITO ALTO 0.33 0.43	5.713.688	83%	7%	11%	10%	1.66	0.51	Bem Acima da Media
	G BEM MAIS ALTO 0.43 OU MAIS	5.734.442	88%	5%	12%	10%	2.44	0.89	Muito Acima da Media
		56.889.894	75%	100%	100%	100%	1.00	-	Media
Faixa Percentual que representa impacto de gastos com alimentação									

18	EVIDÊNCIA			Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA	
	PERC	DESP	INDIV			NÃO	SIM					Total
ATE 27 PERC				10,122,766	80%	15%	19%	18%	1.30	-	0.26	Acima da Media
28 40 PERC				10,879,468	78%	17%	20%	19%	1.19	-	0.17	Media
40 52 PERC				11,812,217	75%	21%	21%	21%	1.01	-	0.01	Media
52 64 PERC				12,022,532	72%	24%	20%	21%	0.86	-	0.15	Abaixoda Media
MAIS DE 64 PERC				12,052,910	71%	25%	20%	21%	0.82	-	0.20	Abaixoda Media
				56,889,893	75%	100%	100%	100%	1.00	-	0.00	Media
Faixa Percentual que representa impacto de gastos com despesas individuais												

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA	IMPACTO NA DE DIFICULDADE
IMPACTO OUTRAS				NÃO	SIM			WOE	
19	E SEM OUTRAS DESPESAS 0 0	668.176	82%	1%	1%	1%	1.46	0.38	Acima da Media
	B MUITO ALTO 0 0.05	13.527.381	67%	32%	21%	24%	0.67	0.41	Bem Abaixo da Media
	C ACIMA 0.05 0.08	8.533.050	71%	17%	14%	15%	0.82	0.19	Abaixoda Media
	D MEDIA 0.08 0.14	17.087.562	76%	29%	31%	30%	1.07	0.07	Media
	E ABAIXO 0.14 0.20	8.548.258	80%	12%	16%	15%	1.33	0.29	Acima da Media
	F BEM ABAIXO 0.20 ou mais	8.525.466	84%	10%	17%	15%	1.76	0.57	Bem Acima da Media
		56.889.893	75%	100%	100%	100%	1.00	0.00	Media
Faixa Percentual que representa impacto de gastos com outras despesas									

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
IMPACTO OUTRAS DIF				NÃO	SIM				
E SEM DESPESA OUTRAS DIF 0 0		5.677.466	89%	4%	12%	10%	2.73	1.00	Muito Acima da Media
A BEM ALTO 0.27 OU MAIS		5.595.018	53%	18%	7%	10%	0.38	0.97	Muito Abaixo da Media
B MUITO ALTO 0.17 0.27		8.471.541	64%	21%	13%	15%	0.60	0.51	Bem Abaixo da Media
C ACIMA 0.14 0.17		2.825.231	72%	6%	5%	5%	0.83	0.18	Abaixoda Media
D MEDIA 0.04 0.14		19.996.078	76%	34%	36%	35%	1.05	0.05	Media
E ABAIXO 0.03 0.04		2.863.224	79%	4%	5%	5%	1.24	0.22	Acima da Media
F BEM ABAIXO 0.01 0.03		5.716.449	83%	7%	11%	10%	1.60	0.47	Bem Acima da Media
G MUITO ABAIXO 0 0.01		5.744.885	87%	5%	12%	10%	2.27	0.82	Muito Acima da Media
		56.889.892	75%	100%	100%	100%	1.00	0.00	Media
Faixa Percentual que representa impacto de gastos com outras despesas diferenciadas									

CONTINUA ...

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM				
21	IMPACTO VESTUARIO								
	Z_SEM_DESPESA_VESTUARIO_0_0	7.146.584	80%	10%	13%	13%	1.31	0.27	Acima da Media
	B_MUITO_ABAIXO_0_0	1.396.512	69%	3%	2%	2%	0.74	0.30	Bem Abaixo da Media
	C_ABAIXO_0_0.02	8.527.312	72%	17%	14%	15%	0.85	0.16	Abaixoda Media
	D_MEDIA_0.02_0.12	34.123.143	74%	62%	59%	60%	0.97	0.04	Media
	E_ALTO_0.12_OU MAIS	5.696.340	79%	8%	11%	10%	1.27	0.24	Acima da Media
		56.889.891	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com vestuário

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM				
22	FATOR 19 CREDITO E OU HABITACAO DIFERENC								
	NÃO TEVE ACESSO	16,230,726	83%	20%	31%	29%	1.61	0.48	Bem Acima da Media
	TEVE_ACESSO	40,659,167	72%	80%	69%	71%	0.85	0.16	Abaixoda Media
		56.889.893	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com crédito e habitação

EVIDÊNCIA		Nr Famílias	% DIFÍCIL	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
IMPACTO HABITACAO DIF				NÃO	SIM				
23	E_SEM_DESPESA_HABIT_DIF_0_0	13,797,462	85%	14%	28%	24%	1.93	0.66	Bem Acima da Media
	B_MUITO_ALTO_0.06_OU MAIS	19,941,607	66%	49%	31%	35%	0.63	0.46	Bem Abaixo da Media
	C_ACIMA_0.04_0.06	5,715,777	71%	12%	10%	10%	0.83	0.19	Abaixoda Media
	D_MEDIA_0.02_0.04	8,546,995	76%	14%	15%	15%	1.07	0.07	Media
	E_ABAIXO_0.01_0.02	2,843,047	81%	4%	5%	5%	1.43	0.36	Acima da Media
	F_BEM_ABAIXO_0_0.01	6,045,007	82%	8%	12%	11%	1.52	0.42	Bem Acima da Media
		56,889,895	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com habitação diferenciada

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
				NÃO	SIM				
24	X RENDA EMPREGADOR								
	NAO	54,349,379	76%	92%	97%	96%	1.06	0.06	Media
	SIM	2,540,515	53%	8%	3%	4%	0.37	0.99	Muito Abaixo da Media
		56,889,894	75%	100%	100%	100%	1.00	-	Media

Indica se no domicílio tem renda de empregador

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
25	PERC VAL ALIM MIN			NÃO	SIM				
	sem resposta	491,878	47%	2%	1%	1%	0.29	-1.23	Muito Abaixo da Media
	<=0.25	21,251,434	66%	51%	33%	37%	0.64	-0.44	Bem Abaixo da Media
	<=0.5	27,907,152	80%	39%	52%	49%	1.34	0.29	Acima da Media
	>0.5	7,239,429	84%	8%	14%	13%	1.78	0.58	Bem Acima da Media
		56,889,893	75%	100%	100%	100%	1.00	0.00	Media

Percentual da renda mínima que designaria para a alimentação dentro da renda que declarou como necessária para viver.

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
26	RENDA MONET NECESS			NÃO	SIM				
Não resposta		72.170	88%	0%	0%	0%	2.46	0.90	Muito Acima da Media
<=0.53		5,682,921	48%	21%	6%	10%	0.31	1.17	Muito Abaixo da Media
<=0.75		5,683,064	60%	16%	8%	10%	0.51	0.68	Muito Abaixo da Media
<=0.92		5,679,233	67%	13%	9%	10%	0.67	0.40	Bem Abaixo da Media
<=1.06		5,685,203	71%	12%	9%	10%	0.79	0.24	Abaixoda Media
<=1.22		5,679,514	79%	9%	10%	10%	1.21	0.19	Acima da Media
<=1.40		5,681,152	82%	7%	11%	10%	1.47	0.38	Acima da Media
<=1.63		5,681,045	84%	6%	11%	10%	1.76	0.57	Bem Acima da Media
<=1.98		5,681,726	85%	6%	11%	10%	1.94	0.66	Bem Acima da Media
<=2.75		5,681,881	88%	5%	12%	10%	2.32	0.84	Muito Acima da Media
>2.75		5,681,986	88%	5%	12%	10%	2.35	0.86	Muito Acima da Media
		56.889.895	75%	100%	100%	100%	1.00	0.00	Media

Quanto da renda declarou precisar para viver até o final do mês

CONTINUA ...

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
27	RENDA MONET PER CAPITA			NAO	SIM				
	<=113	5,689,535	94%	2%	13%	10%	5.55	1.71	Muito Acima da Media
	<=191	5,689,760	92%	3%	12%	10%	3.81	1.34	Muito Acima da Media
	<=267	5,688,057	90%	4%	12%	10%	2.93	1.08	Muito Acima da Media
	<=364	5,688,336	86%	6%	11%	10%	2.03	0.71	Muito Acima da Media
	<=455	5,689,162	81%	8%	11%	10%	1.37	0.32	Acima da Media
	<=585	5,691,074	76%	10%	10%	10%	1.04	0.04	Media
	<=784	5,688,679	73%	11%	10%	10%	0.91 -	0.09	Media
	<=1091	5,687,047	66%	14%	9%	10%	0.64 -	0.45	Bem Abaixo da Media
	<=1881	5,688,930	57%	17%	8%	10%	0.43 -	0.84	Muito Abaixo da Media
	>1881	5,689,312	37%	25%	5%	10%	0.19 -	1.64	Muito Abaixo da Media
		56,889,892	75%	100%	100%	100%	1.00	0.00	Media

Faixa de renda monetária per capita

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
28	RENDA TOT ANO			NAO	SIM				
	<=6615	5,689,613	92%	3%	12%	10%	3.67	1.30	Muito Acima da Media
	<=9480	5,692,481	90%	4%	12%	10%	2.88	1.06	Muito Acima da Media
	<=12275	5,684,896	87%	5%	12%	10%	2.18	0.78	Muito Acima da Media
	<=15198	5,689,316	84%	7%	11%	10%	1.72	0.54	Bem Acima da Media
	<=18917	5,688,320	81%	8%	11%	10%	1.42	0.35	Acima da Media
	<=23981	5,691,039	78%	9%	10%	10%	1.16	0.15	Media
	<=30887	5,687,048	73%	11%	10%	10%	0.89 -	0.12	Abaixoda Media
	<=42333	5,690,859	67%	13%	9%	10%	0.67 -	0.41	Bem Abaixo da Media
	<=68583	5,687,608	61%	16%	8%	10%	0.51 -	0.67	Bem Abaixo da Media
	>68583	5,688,715	40%	24%	5%	10%	0.22 -	1.52	Muito Abaixo da Media
		56,889,895	75%	100%	100%	100%	1.00 -	0.00	Media

Faixa de renda total recebida no ano

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
29	TEM PLANO DE SAUDE			NAO	SIM				
	NAO	41,050,238	82%	53%	78%	72%	1.46	0.88	Acima da Media
	SIM	15,839,656	58%	47%	22%	28%	0.47	0.76	Muito Abaixo da Media
		56,889,894	75%	100%	100%	100%	1.00	-	Media

Indica se algum morador possui plano de saúde

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
30	X OUTRAS RENDAS			NAO	SIM				
	NAO	43,075,347	78%	68%	78%	76%	1.15	0.14	Media
	SIM	13,814,546	67%	32%	22%	24%	0.68	0.38	Bem Abaixo da Media
		56,889,893	75%	100%	100%	100%	1.00 -	0.00	Media

Indica presença de outras rendas no domicílio (ex aluguel, rendimento de aplicações, etc)

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
31	CONDICOES DE MORADIA			NAO	SIM				
	BOAS	29,375,434	65%	74%	44%	52%	0.60	0.51	Bem Abaixo da Media
	RUINS	5,290,225	96%	1%	12%	9%	7.92	2.07	Muito Acima da Media
	SATISFATORIAS	22,224,234	84%	25%	44%	39%	1.76	0.56	Bem Acima da Media
		56,889,893	75%	100%	100%	100%	1.00	0.00	Media

A pessoa declarou sua percepção geral sobre a sua moradia

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
33	QUANTIDADE DE ALIMENTO REC			NAO	SIM				
	AS VEZES_NAO_E_SUFICIENTE	14,973,665	94%	7%	33%	26%	4.83	1.58	Muito Acima da Media
	E SEMPRE_SUFICIENTE	36,697,577	65%	92%	56%	65%	0.61	0.50	Bem Abaixo da Media
	NORMALMENTE_NAO_E_SUFICIENTE	5,218,651	96%	2%	12%	9%	7.77	2.05	Muito Acima da Media
		56,889,893	75%	100%	100%	100%	1.00 -	0.00	Media

A pessoa declarou sua percepção se a quantidade de alimentos está adequada às necessidades familiares.

CONTINUA ...

34	EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
	TIPO ALIMENTO RAZAO_REC			DIFIC	NAO				
	ALIMENTOS NEM SEMPRE DISPONIVEIS OU OUTROS	3,373,197	58%	10%	5%	6%	0.45	-	0.80
	NEM SEMPRE DO TIPO QUE QUER POIS A RENDA NAO	26,502,239	90%	18%	56%	47%	3.14	-	1.14
	RARAMENTE DO TIPO QUE QUER POIS A RENDA NAO	7,007,591	96%	2%	16%	12%	7.71	-	2.04
	SEMPRE DO TIPO QUE QUER	20,006,866	51%	70%	24%	35%	0.34	-	1.08
		56,889,893	75%	100%	100%	100%	1.00	-	0.00
A pessoa declarou sua percepção se os tipos de alimentos estão adequados e a razão de não se alimentar como quer.									

A pessoa declarou sua percepção se os tipos de alimentos estão adequados e a razão de não se alimentar como quer.

35	EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
	CLASSE	ABEP			NÃO	SIM				
	A1+A2		1,373,217	32%	7%	1%	2%	0.16	-	1.84
	B1		3,391,303	46%	13%	4%	6%	0.28	-	1.26
	B2		7,691,457	59%	22%	11%	14%	0.47	-	0.75
	C1		11,734,804	72%	24%	20%	21%	0.83	-	0.18
	C2		13,131,290	81%	18%	25%	23%	1.42	-	0.35
	D		15,895,402	87%	15%	32%	28%	2.20	-	0.79
	E		3,672,420	91%	2%	8%	6%	3.40	-	1.22

Classe ABEP - agregado da pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado.

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
ESCOLARIDADE	PESS REFERENCIA			NAO	SIM				
36	A ANALFABETO	7,898,594	87%	7%	16%	14%	2.14	0.76	Muito Acima da Media
	B 0 A 3	7,887,678	82%	10%	15%	14%	1.46	0.38	Acima da Media
	C 4 A 7	14,043,503	81%	19%	26%	25%	1.37	0.32	Acima da Media
	D 8 A 10	6,850,574	79%	10%	13%	12%	1.25	0.23	Acima da Media
	E 11 A 14	14,738,754	69%	32%	24%	26%	0.75	-0.29	Bem Abaixo da Media
	F 15 OU MAIS	5,098,427	44%	20%	5%	9%	0.25	-1.37	Muito Abaixo da Media
	G NAO DECLARADO	372,364	80%	1%	1%	1%	1.35	0.30	Acima da Media
		56,889,894	75%	100%	100%	100%	1.00	-	Media

Escolaridade da pessoa de referência ou responsável pela família

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
37 Fator 17 CARTAO CONTA CORRENTE				NAO	SIM				
			DIFIC			Total			
NAO		36,227,686	82%	46%	70%	64%	1.52	0.42	Bem Acima da Media
SIM		20,662,207	63%	54%	30%	36%	0.56	0.58	Bem Abaixo da Media
		56,889,893	75%	100%	100%	100%	1.00	0.00	Media

Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Cartão de Crédito_Pessoas que responderam ao questionário de orçamento e despesa e para as quais foi colhida informação de Conta Corrente apenas quando possui cheque especial

38	EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
	FATOR 5 ESPACO CASA			DIFIC	NAO				
	ATE 4 COMODOS 1 BANHEIRO 1 COMO DORM	7,457,755	86%	7%	15%	13%	2.01	-0.70	Muito Acima da Media
	PRED 4 5 COMODOS 1 BANHEIRO 2 DORM	19,595,993	80%	28%	37%	34%	1.31	-0.27	Acima da Media
	PRED 2 A 3 DORM 6 MAIS COMODOS 1 BANHEIRO	17,958,356	77%	29%	32%	32%	1.12	-0.11	Media
	PRED 2 A 3 DORM 7 MAIS COMODOS 2 BANHEIROS	7,902,432	62%	21%	12%	14%	0.54	-0.61	Bem Abaixo da Media
	PRED 9 OU MAIS COMODOS 3 OU MAIS BANHEIROS E	3,975,358	48%	15%	4%	7%	0.31	-1.18	Muito Abaixo da Media
		56,889,894	75%	100%	100%	100%	1.00	-	Media

Nr de cômodos que estão sendo utilizados como dormitório_Quantidade de banheiros no domicílio_Quantidade de cômodos no domicílio_A pessoa declarou ter pouco espaço na casa

CONTINUA ...

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NÃO	SIM				
39 IMPACTO TRANSPORTE DIF								
E SEM DESPESA TRANSP DIF 0 0	25,888,613	83%	30%	51%	46%	1.67	0.51	Bem Acima da Media
A_BEM_ALTO_0.31_OU MAIS	2,834,932	59%	8%	4%	5%	0.47	0.76	Muito Abaixo da Media
B_MUITO_ALTO_0.04_0.31	17,022,995	66%	41%	26%	30%	0.63	0.46	Bem Abaixo da Media
C_ACIMA_0.03_0.04	2,842,806	70%	6%	5%	5%	0.75	0.28	Abaixoda Media
D_MEDIA_0_0.03	8,300,548	76%	14%	15%	15%	1.07	0.06	Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media

Faixa Percentual que representa impacto de gastos com transporte diferenciado

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NÃO	SIM				
40 PONTOS ABEP								
<=9	6,125,068	90%	4%	13%	11%	2.95	1.08	Muito Acima da Media
<=11	5,659,283	87%	5%	12%	10%	2.28	0.82	Muito Acima da Media
<=13	7,783,471	86%	8%	16%	14%	2.08	0.73	Muito Acima da Media
<=14	3,590,242	83%	4%	7%	6%	1.57	0.45	Bem Acima da Media
<=16	6,653,558	80%	9%	13%	12%	1.36	0.30	Acima da Media
<=18	5,702,717	79%	9%	11%	10%	1.23	0.21	Acima da Media
<=20	4,990,243	71%	10%	8%	9%	0.80	0.22	Abaixoda Media
<=23	5,679,558	68%	13%	9%	10%	0.69	0.37	Bem Abaixo da Media
<=27	5,102,435	58%	15%	7%	9%	0.45	0.80	Muito Abaixo da Media
>27	5,603,318	44%	22%	6%	10%	0.26	1.36	Muito Abaixo da Media
TOTAL	56,889,893	75%	100%	100%	100%	1.00	0.00	Media

Pontuação calculada a partir da fórmula do Critério Brasil, fornecido pela Associação das Empresas de Pesquisa de Mercado, a partir dos dados do inventário de bens, escolaridade do chefe da família e número de banheiros.

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NÃO	SIM				
41 FATOR 11 EDUCAÇÃO SAUDE LAZER								
TEM BOM ACESSO	21,936,079	68%	50%	35%	39%	0.69	0.37	Bem Abaixo da Media
NÃO TEM BOM ACESSO	34,953,813	80%	50%	65%	61%	1.31	0.27	Acima da Media
	56,889,892	75%	100%	100%	100%	1.00	0.00	Media

Acesso a serviços de educação_Acesso a serviços de lazer e esporte_Acesso a serviços de saúde

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NÃO	SIM				
42 PERC TRANSPORTE								
<=0	17,525,788	77%	29%	32%	31%	1.10	0.09	Media
<=0.05	18,539,487	69%	40%	30%	33%	0.74	0.30	Bem Abaixo da Media
<=0.10	9,880,551	77%	16%	18%	17%	1.08	0.08	Media
<=0.15	5,043,890	80%	7%	9%	9%	1.30	0.26	Acima da Media
<=0.25	4,042,113	82%	5%	8%	7%	1.56	0.44	Bem Acima da Media
>0.25	1,858,064	82%	2%	4%	3%	1.54	0.43	Bem Acima da Media
	56,889,893	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com transporte

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NÃO	SIM				
43 SERVICO DE TRANSPORTE COLETIVO								
_BOM+NAO_RES	27,885,671	71%	58%	46%	49%	0.79	0.23	Abaixoda Media
_NAO_TEM+RUIM	29,004,223	80%	42%	54%	51%	1.29	0.25	Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media

A pessoa apresentou sua percepção sobre transporte coletivo no local onde mora

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
44 CASA ESCURA								
NAO	46,127,794	72%	90%	78%	81%	0.87	0.14	Abaixoda Media
SIM	10,762,099	87%	10%	22%	19%	2.23	0.80	Muito Acima da Media
	56,889,893	75%	100%	100%	100%	1.00	0.00	Media

A pessoa declarou que percebe sua casa escura

CONTINUA ...

EVIDÊNCIA				Nr Famílias	%	DIFICULDADE		Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE	
FATOR 12 GOTEIRA FUNDACAO PAREDE CHAO UMIDOS						NAO	SIM					
45	NAO				28,432,891	65%	70%	43%	50%	0.62	0.48	Bem Abaixo da Media
	SIM				28,457,002	85%	30%	57%	50%	1.88	0.63	Bem Acima da Media
					56.889.893	75%	100%	100%	100%	1.00	0.00	Media

A pessoa declarou ter paredes e chão umidos_A pessoa declarou ter janelas, portas ou assoalhos deteriorados_A pessoa declarou ter telhado com

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
45	COMPROM RENDA			NÃO	SIM				
sem informação		72,170	88%	0%	0%	0%	2.46	<div><div></div></div> 0.90	Muito Acima da Media
<=0.5		3,258,249	68%	7%	5%	6%	0.71	<div><div></div></div> 0.34	Bem Abaixo da Media
<=0.6		2,556,676	69%	6%	4%	4%	0.73	<div><div></div></div> 0.31	Bem Abaixo da Media
<=0.7		3,402,598	70%	7%	6%	6%	0.77	<div><div></div></div> 0.26	Abaixoda Media
<=0.8		4,092,143	69%	9%	7%	7%	0.74	<div><div></div></div> 0.30	Bem Abaixo da Media
<=1		8,890,376	72%	18%	15%	16%	0.84	<div><div></div></div> 0.17	Abaixoda Media
>1		34,617,683	78%	53%	63%	61%	1.19	<div><div></div></div> 0.18	Media
		56,889,895	75%	100%	100%	100%	1.00	<div><div></div></div> 0.00	Media

Comprometimento de renda. Considerando somatório de todas as despesas pela renda total

46	EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
	CONDICAO OCUPACAO REC			DIFIC	NÃO				
	ALUGADO	9,372,960	75%	16%	17%	16%	1.01	0.01	Media
	CEDIDO OUTRA FORMA	4,701,737	83%	6%	9%	8%	1.58	0.45	Bem Acima da Media
	PRÓPRIO OU CEDIDO POR EMPREGADOR	42,815,195	74%	78%	74%	75%	0.95	0.05	Media
		56,889,892	75%	100%	100%	100%	1.00	0.00	Media

Apresenta a condição de ocupação do domicílio : alugado, cedido ou próprio

EVIDÊNCIA		Nr Famílias	%	DIFICULDADE		%	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
FATOR 28 MORA DE ALUGUEL				NAO	SIM				
NAO		47,516,933	75%	84%	83%	84%	1.00	0.00	Media
SIM		9,372,960	75%	16%	17%	16%	1.01	0.01	Media
		56.889.893	75%	100%	100%	100%	1.00	0.00	Media

Há quanto tempo a pessoa fez o contrato de aluguel_Apresenta o tipo de contrato de aluguel, sendo formal ou informal

48	EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
	IMPACTO ALUGUEL ESTIM			NAO	SIM				
	Z_SEM_DESPESA_ALUGUEL_ESTIM_0_0	9.473.784	75%	16%	17%	17%	1.02	0.02	Media
	B_MUITO_ABAIXO_0_0.04	1.861.124	68%	4%	3%	3%	0.70	0.35	Bem Abaixo da Media
	C_ABAIXO_0.04_0.11	14.256.033	72%	29%	24%	25%	0.83	0.19	Abaixoda Media
	D_MEDIA_0.11_0.33	25.641.519	76%	43%	46%	45%	1.07	0.07	Media
	E_ALTO_0.33_OU MAIS	5.656.107	81%	8%	11%	10%	1.37	0.31	Acima da Media
	SEM INFORMAÇÃO	1.327	100%	0%	0%	0%			
		56.889.894	75%	100%	100%	100%	1.00	-	Media

Faixa Percentual que representa impacto de gastos com habitação (não diferenciada)

EVIDÊNCIA		Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
49	PERC HABITACAO			NAO	SIM				
	<=0.05	5,481,015	65%	14%	8%	10%	0.62	-0.48	Bem Abaixo da Media
	<=0.1	13,881,582	70%	29%	23%	24%	0.77	-0.26	Abaixoda Media
	<=0.15	13,433,456	76%	23%	24%	24%	1.02	0.02	Media
	<=0.2	9,048,248	78%	14%	17%	16%	1.17	0.16	Media
	<=0.3	9,052,183	81%	12%	17%	16%	1.38	0.32	Acima da Media
	>0.3	5,993,409	83%	7%	12%	11%	1.56	0.45	Bem Acima da Media
		56,889,893	75%	100%	100%	100%	1.00	0.00	Media

Faixa Percentual que representa impacto de gastos com habitação (não diferenciada)

CONTINUA ...

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
50 PERC_RENDA_NAO_MONET								
sem inforção	1,843	100%	0%	0%	0%			
<=0.0	4,568,752	69%	10%	7%	8%	0.73	-	Bem Abaixo da Media
<=0.1	15,070,741	66%	36%	23%	26%	0.66	-	Bem Abaixo da Media
<=0.2	16,795,159	75%	29%	30%	30%	1.01		Media
<=0.3	9,753,596	80%	14%	18%	17%	1.32	-	Acima da Media
<=0.4	5,111,260	85%	5%	10%	9%	1.84	-	Bem Acima da Media
> 0.4	5,588,542	86%	6%	11%	10%	2.00	-	Muito Acima da Media
	56,889,893	75%	100%	100%	100%	1.00	-	Media

Faixa Percentual da renda que não é monetária.

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
51 CONJ OU RESP ESTUDANDO_sum								
PELO MENOS UM ESTUDANDO	4,941,883	70%	11%	8%	9%	0.77	-	Abaixo da Media
NENHUM ESTUDANDO	51,948,009	76%	89%	92%	91%	1.03	-	Media
	56,889,892	75%	100%	100%	100%	1.00	-	Media

Indicador se o conjuges ou o responsável estão estudando

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
52 ESCOLARIDADE_MENORES								
NAO TEM MENORES DE 18 ANOS	25,844,190	68%	58%	41%	45%	0.71	-	Bem Abaixo da Media
TEM PESSOA DE 7 A 17 ANOS E TODAS ESTAO NA ESCOLA	15,112,165	80%	22%	28%	27%	1.28	-	Acima da Media
TEM PESSOA DE 7 A 17 ANOS E FORA DA ESCOLA	1,102,635	85%	1%	2%	2%	1.89	-	Bem Acima da Media
TEM_MENOR_6_ANOS_COM_CRECHE_ESCOLA	6,119,783	80%	9%	11%	11%	1.31	-	Acima da Media
TEM_MENOR_6_ANOS_SEM_CRECHE_ESCOLA	8,711,121	83%	10%	17%	15%	1.64	-	Bem Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media

Situação da escolaridade dos menores de 18 anos

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
53 FATOR_10_FAMILIAS								
ATE DOIS FILHOS ADULTOS OU ADOLESCENTES	14,690,057	75%	26%	26%	26%	0.98	-	Media
DUAS PESSOAS SEM FILHOS	10,164,717	67%	24%	16%	18%	0.66	-	Bem Abaixo da Media
FAMILIAS COM CRIANCAS ATE 6 ANOS	14,830,904	82%	19%	28%	26%	1.49	-	Acima da Media
MAIS DE DOIS FILHOS PREDOM ADOLESCENTES	10,471,031	81%	14%	20%	18%	1.39	-	Acima da Media
SOLITARIOS	6,733,186	65%	16%	10%	12%	0.62	-	Bem Abaixo da Media
	56,889,895	75%	100%	100%	100%	1.00	-	Media

Tipos de família considerando a idade do filho mais novo_Número de dependentes de renda na família_Quantidade de filhos na família

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
54 QTDE_MORADORES								
1	6,771,089	65%	17%	10%	12%	0.63	-	Bem Abaixo da Media
2	12,742,720	69%	28%	21%	22%	0.75	-	Abaixo da Media
3	14,061,791	74%	25%	24%	25%	0.97	-	Media
4	12,175,145	79%	18%	22%	21%	1.24	-	Acima da Media
5	6,151,829	81%	8%	12%	11%	1.43	-	Acima da Media
6 ou mais pessoas	4,987,320	88%	4%	10%	9%	2.44	-	Muito Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media

Quantidade de moradores na família

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
55 QUANTIDADE_FAMILIAS (na mesma UC)								
1	53,873,262	75%	97%	94%	95%	0.97	-	Media
2 ou mais	3,016,632	84%	3%	6%	5%	1.72	-	Bem Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media

Indica se existem mais de uma família morando no mesmo domicílio

CONTINUA ...

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
56 STATUS REDE ENSINO		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
NINGUEM NO DOM CRECHE OU ESCOLA	24,430,771	71%	51%	40%	43%	0.79	-	0.23
PARCIALMENTE NA REDE PARTICULAR	2,718,532	75%	5%	5%	5%	0.98	-	0.02
TODOS NA REDE PARTICULAR	6,650,206	60%	19%	9%	12%	0.49	-	0.71
TODOS NA REDE PUBLICA	23,090,383	84%	25%	46%	41%	1.80	-	0.59
	56,889,892	75%	100%	100%	100%	1.00	-	0.00

Situação da rede de ensino utilizada pela família : escolas e creches, particulares e públicas

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
57 FATOR 13 ESGOTO ESCOADOURO		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
RUIIM_NAO_TEM	22,641,339	81%	30%	43%	40%	1.44	-	0.36
SIM_BOM	34,248,554	71%	70%	57%	60%	0.81	-	0.21
	56,889,893	75%	100%	100%	100%	1.00	-	0.00

Presença de esgoto sanitário_Presença de água canalizada_Tipo de serviço de esgotamento sanitário disponível no domicílio

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
58 FATOR 23 LIMPEZA MANUTENCAO PAVIMENTACAO		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
NAO	14,901,882	83%	17%	29%	26%	1.67	-	0.51
SIM	41,988,011	72%	83%	71%	74%	0.86	-	0.15
	56,889,893	75%	100%	100%	100%	1.00	-	0.00

Acesso a serviços de limpeza e manutenção de ruas_Indica sobre a pavimentação da rua onde mora

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
59 FATOR 3 ACESSO A AGUA RURAL		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
RURAL COM ACESSO A AGUA	4,940,343	78%	8%	9%	9%	1.20	-	0.18
RURAL COM DIFICULDADE DE ACESSO À AGUA	3,989,049	84%	5%	8%	7%	1.68	-	0.52
URBANO	47,960,502	74%	88%	83%	84%	0.95	-	0.05
	56,889,894	75%	100%	100%	100%	1.00	-	0.00

Proveniência da água utilizada no domicílio_Tipo de serviço de água disponível para o domicílio_Indica se urbano ou Rural

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
60 FATOR 6 BOA ILUM ENERGIA COLETA LIXO		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
NAO	25,044,194	80%	35%	47%	44%	1.34	-	0.29
SIM	31,845,699	71%	65%	53%	56%	0.82	-	0.20
	56,889,893	75%	100%	100%	100%	1.00	-	0.00

Acesso a serviços de coleta de lixo_Acesso a serviços de fornecimento de energia elétrica_Acesso a serviços de fornecimento de iluminação de rua

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
61 INUNDACOES		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
NAO	51,465,252	74%	93%	89%	90%	0.96	-	0.04
SIM	5,424,641	83%	7%	11%	10%	1.62	-	0.48
	56,889,893	75%	100%	100%	100%	1.00	-	0.00

A pessoa apresenta inundações no local onde mora

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
			NAO	SIM				
62 SERVICO DE ESCOAMENTO DA AGUA DE CHUVA		DIFIC			Total	SIM / NÃO	WOE	DE DIFICULDADE
BOM	35,116,149	71%	71%	59%	62%	0.83	-	0.19
RUIIM_NAO_TEM_NR	21,773,744	81%	29%	41%	38%	1.41	-	0.35
	56,889,893	75%	100%	100%	100%	1.00	-	0.00

A pessoa apresentou sua percepção sobre escoamento de água no local onde mora

CONTINUA ...

EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
63 TIPO AREA								
CAPITAL	13,452,333	70%	29%	22%	24%	0.76	0.28	Abaixoda Media
RESTO DA RM	7,333,112	79%	11%	14%	13%	1.28	0.24	Acima da Media
RESTO DA UF	27,175,057	75%	48%	48%	48%	0.99	0.01	Media
RURAL	8,929,392	81%	12%	17%	16%	1.38	0.33	Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media
Tipo da área onde mora								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
64 X RENDA TRANSF GOVERNO								
NAO	44,329,428	72%	89%	74%	78%	0.83	0.19	Abaixoda Media
SIM	12,560,466	88%	11%	26%	22%	2.42	0.88	Muito Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media
Indica se no domicilio de renda de transferências do								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
65 GRUPO DE RELIGIOES								
CATOLICA+OUTRAS	4,520,572	83%	6%	9%	8%	1.59	0.46	Bem Acima da Media
ESPIRITAS+INDIGENA+JUDAICAS+MUCULMANO+ORIENT	47,813,072	75%	86%	83%	84%	0.97	0.03	Media
NEOPENTECOSTAL_EVANG	1,369,508	62%	4%	2%	2%	0.54	0.62	Bem Abaixo da Media
	3,186,742	79%	5%	6%	6%	1.22	0.20	Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media
Grupo de religião ao qual pertence o chefe da família								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
66 RACA COR PESSOA REFERENCIA								
AMARELA+BRANCA+NAO SABE	28,383,384	69%	62%	46%	50%	0.74	0.31	Bem Abaixo da Media
INDIGENA+PARDA+PRETA	28,506,510	81%	38%	54%	50%	1.44	0.36	Acima da Media
	56,889,894	75%	100%	100%	100%	1.00	-	Media
Raça ou cor do chefe da família								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
67 REGIAO								
00 'CENTRO OESTE'	4,318,785	72%	8%	7%	8%	0.86	0.15	Abaixoda Media
00 'NORDESTE '	14,922,452	82%	19%	29%	26%	1.52	0.42	Bem Acima da Media
00 'NORTE'	3,901,058	78%	6%	7%	7%	1.15	0.14	Media
00 'SUDESTE'	24,990,728	74%	47%	43%	44%	0.92	0.08	Media
00 'SUL'	8,756,869	68%	20%	14%	15%	0.70	0.36	Bem Abaixo da Media
	56,889,892	75%	100%	100%	100%	1.00	0.00	Media
Região Geográfica								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
68 X RENDA PREV PRIVADA								
NAO	56,492,313	75%	99%	100%	99%	1.01	0.01	Media
SIM	397,580	49%	1%	0%	1%	0.31	1.16	Muito Abaixo da Media
	56,889,893	75%	100%	100%	100%	1.00	0.00	Media
Indica se no domicilio tem renda de previdência privada								
EVIDÊNCIA	Nr Famílias	% DIFIC	DIFICULDADE		% Total	BIVARIADA SIM / NÃO	EVIDÊNCIA WOE	IMPACTO NA DE DIFICULDADE
			NÃO	SIM				
69 TIPO VEICULO								
BICICLETA	12,979,638	86%	13%	26%	23%	1.99	0.69	Bem Acima da Media
NAO POSSUI VEICULO	18,562,084	81%	25%	35%	33%	1.40	0.33	Acima da Media
MOTOCICLETA	4,680,456	79%	7%	9%	8%	1.23	0.21	Acima da Media
SEM INFORMACAO	2,756,284	72%	5%	5%	5%	0.85	0.16	Abaixoda Media
UM AUTOMOVEL	14,722,249	64%	37%	22%	26%	0.59	0.53	Bem Abaixo da Media
DOIS OU MAIS AUTOMOVEIS	3,189,183	47%	12%	4%	6%	0.29	0.22	Muito Abaixo da Media
Total	56,889,894	75%	100%	100%	100%	1.00	-	Media
Indica se veiculo prioritário é carro, moto, bicicleta ou nenhum veículo								

CONTINUA ...

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
UF		DIFIC	NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
MA_PL_AL_SE	3,949,033	84%	4%	8%	7%	1.79	0.58	Bem Acima da Media
BA_PE_CE_PA_PB_R_N_TO	13,131,388	81%	17%	25%	23%	1.44	0.36	Acima da Media
GO_AM_AC_RR	3,001,234	75%	5%	5%	5%	1.00 -	0.00	Media
SP_MG_RJ	23,915,867	74%	44%	41%	42%	0.93 -	0.07	Media
PR_ES_MS_RO_AP	5,678,204	71%	11%	9%	10%	0.83 -	0.19	Abaixoda Media
RS_SC_MT_DF	7,214,167	67%	17%	11%	13%	0.66 -	0.42	Bem Abaixo da Media
	56,889,893	75%	100%	100%	100%	1.00 -	0.00	Media

Agrupamentos das Unidades da Federação considerando percepção média de dificuldade em cada uma delas

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
AREA GEO		DIFIC	NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
CAPITAL	13,452,333	70%	29%	22%	24%	0.76 -	0.28	Abaixoda Media
INTERIOR	27,175,057	75%	48%	48%	48%	0.99 -	0.01	Media
REG METR	7,333,112	79%	11%	14%	13%	1.28	0.24	Acima da Media
RURAL_COM_AGUA	4,940,343	78%	8%	9%	9%	1.20	0.18	Acima da Media
RURAL_SEM_AGUA	3,989,049	84%	5%	8%	7%	1.68	0.52	Bem Acima da Media
Total	56,889,894	75%	100%	100%	100%	1.00	-	Media

Apresenta se o domicílio está em área urbana ou rural e se na área rural verifica acesso à água

EVIDÊNCIA	Nr Famílias	%	DIFICULDADE		%	BIVARIADA	EVIDÊNCIA	IMPACTO NA
UF		DIFIC	NÃO	SIM	Total	SIM / NÃO	WOE	DE DIFICULDADE
MARANHAO	1,658,086	86%	2%	3%	3%	1.97	0.68	Bem Acima da Media
ALAGOAS	864,492	84%	1%	2%	2%	1.71	0.53	Bem Acima da Media
SERGIPE	560,541	84%	1%	1%	1%	1.70	0.53	Bem Acima da Media
PIAUI	865,915	83%	1%	2%	2%	1.66	0.51	Bem Acima da Media
BAHIA	4,116,049	82%	5%	8%	7%	1.47	0.38	Acima da Media
PARAIBA	1,097,826	82%	1%	2%	2%	1.46	0.38	Acima da Media
CEARA	2,291,675	81%	3%	4%	4%	1.44	0.37	Acima da Media
PARA	1,795,559	81%	2%	3%	3%	1.43	0.36	Acima da Media
RIO GRANDE DO NORTE	874,030	81%	1%	2%	2%	1.42	0.35	Acima da Media
PERNAMBUCO	2,593,840	81%	4%	5%	5%	1.40	0.34	Acima da Media
TOCANTINS	362,411	81%	0%	1%	1%	1.39	0.33	Acima da Media
ACRE	182,255	76%	0%	0%	0%	1.06	0.06	Media
RORAIMA	107,192	76%	0%	0%	0%	1.03	0.03	Media
GOIAS	1,867,306	75%	3%	3%	3%	1.01	0.01	Media
AMAZONAS	844,481	74%	2%	1%	1%	0.94 -	0.06	Media
MINAS GERAIS	6,016,721	74%	11%	10%	11%	0.94 -	0.06	Media
RIO DE JANEIRO	5,166,511	74%	10%	9%	9%	0.93 -	0.08	Media
SAO PAULO	12,732,634	74%	24%	22%	22%	0.92 -	0.08	Media
MATO GROSSO DO SUL	738,511	73%	1%	1%	1%	0.89 -	0.11	Abaixoda Media
ESPIRITO SANTO	1,074,861	72%	2%	2%	2%	0.86 -	0.15	Abaixoda Media
RONDONIA	451,007	71%	1%	1%	1%	0.81 -	0.21	Abaixoda Media
PARANA	3,255,672	71%	7%	5%	6%	0.81 -	0.21	Abaixoda Media
AMAPA	158,153	70%	0%	0%	0%	0.78 -	0.25	Abaixoda Media
DISTRITO FEDERAL	745,316	69%	2%	1%	1%	0.73 -	0.31	Bem Abaixo da Media
MATO GROSSO	967,653	68%	2%	2%	2%	0.70 -	0.35	Bem Abaixo da Media
RIO GRANDE DO SUL	3,543,617	67%	8%	6%	6%	0.66 -	0.41	Bem Abaixo da Media
SANTA CATARINA	1,957,581	65%	5%	3%	3%	0.60 -	0.51	Bem Abaixo da Media
	56,889,895	75%	100%	100%	100%	1.00	-	Media