

# A HETEROGENEIDADE DA VULNERABILIDADE SOCIAL DAS JUVENTUDES: UMA PERSPECTIVA EMPÍRICA ATRAVÉS DO MÉTODO *GRADE OF MEMBERSHIP*

*Julimar Santos Pinto*<sup>1</sup>  
*e André Junqueira Caetano*<sup>2</sup>

## RESUMO

Considera-se que classificações de indivíduos definidas estritamente através de pertencimentos discretos, tendem a não permitir a análise empírica da heterogeneidade presente nas populações. Este artigo analisa a heterogeneidade da vulnerabilidade social das juventudes sob uma perspectiva empírica. Para alcançar esse objetivo, utiliza-se o método Grade of Membership distinguindo-o dos demais métodos de pesquisa quantitativa que não permitem incorporar a heterogeneidade ao nível dos indivíduos no processo de análise.

**Palavras-chave:** Juventude. Vulnerabilidade social. Heterogeneidade. *Grade of Membership*.

<sup>1</sup> Mestre em Ciências Sociais pela Pontifícia Universidade Católica de Minas Gerais, Brasil. julimar.jsp@gmail.com

<sup>2</sup> Doutor em Demografia, pesquisador do CEDEPLAR/Universidade Federal de Minas Gerais e professor Adjunto III da Pontifícia Universidade Católica de Minas Gerais PUC-Minas, Brasil. acaetano@pucminas.br

# THE HETEROGENEITY OF THE YOUTH SOCIAL VULNERABILITY: AN EMPIRICAL PERSPECTIVE THROUGH THE GRADE OF MEMBERSHIP METHOD

## ABSTRACT

Classifications of individuals strictly defined through discrete memberships are considered to tend not to allow the empirical analysis of heterogeneity present in populations. This article analyzes the heterogeneity of youth social vulnerability under an empirical perspective. In order to do that, we use the Grade of Membership method, which is distinct from the other quantitative research methods that do not allow incorporating the heterogeneity at the individual level in the analysis process

**Keywords:** Youth. Social vulnerability. Heterogeneity. Grade of Membership.

## INTRODUÇÃO

Houve, a partir da década de 1990, um aumento da produção acadêmica sobre o tema da juventude. No entanto, nota-se que a despeito do atual consenso na literatura sobre a heterogeneidade das juventudes, ainda permanece pouco explorado, teórica e empiricamente, as definições sobre quais “juventudes” são essas.

Por outro lado, reconhece-se que a complexidade de análise dos fenômenos nas ciências sociais, na demografia e em outras áreas do conhecimento, tem exigido respostas metodológicas que vão além das tradicionais análises descritivas. Assim, métodos matemáticos e estatísticos de agrupamento, redução e mineração de dados têm atraído, crescentemente, a atenção dos pesquisadores dessas áreas do conhecimento. Nesse contexto, o *Grade of Membership* (GoM) é um dos métodos de análise disponíveis.

Uma das vantagens do método GoM, em contraposição aos demais métodos de agrupamento, redução e mineração de dados, é o fato de que não apenas permite a inclusão analítica da heterogeneidade entre os indivíduos como também oferece uma medida que parametriza essa heterogeneidade ao nível de análise individual.

Ressalta-se, ainda, o desenvolvimento de uma nova rotina computacional criada com o objetivo de superar algumas limitações dos atuais *softwares* que realizam a modelagem matemática de GoM. Essa rotina pode ser executada no ambiente computacional *R* e, através deste artigo, divulga-se à comunidade científica.

Destaca-se que o objetivo deste artigo é apresentar metodologia que permita compreender empiricamente o aspecto teórico da heterogeneidade das juventudes, aspecto recorrente na literatura sobre o tema. Dessa forma, devido ao limite de espaço, o

aporte mais substantivo da literatura sobre juventudes e vulnerabilidade social foi desenvolvido considerando apenas a necessidade de situar o leitor no contexto mais geral desse tema.

## VULNERABILIDADE SOCIAL DAS JUVENTUDES

Desde meados dos anos 1990 tem havido, tanto nas Ciências Sociais como nas áreas de Educação, Psicologia e Serviço Social, um aumento da preocupação e, logo, da produção acadêmica sobre o tema da juventude. Esse crescente interesse justifica-se pelo aumento da exposição dos jovens a situações de violência, desemprego, evasão escolar, comportamentos sexuais de risco, uso e abuso de álcool e outras drogas, dentre outros fatores (DAYRELL; MOREIRA; STENGEL, 2011; PAIS, 1990).

Seguindo a compreensão sociológica, entende-se o conceito de juventude como uma

[...] categoria socialmente construída, passível de abrigar não apenas similaridades, mas também diferenças sociais entre os jovens. O uso indiscriminado do termo juventude pode encobrir aparentes unidades de realidade (PAIS, 1993), pois tal rótulo possui especificidades em termos de classe, gênero, raça/cor etc. (BOURDIEU, 1983). Os marcos etários que delimitam as fases do ciclo de vida ou as categorias de idade são móveis e variam ao sabor de novas concepções sociais acerca do humano e das relações intergeracionais (ARIÈS, 1978; DEBERT, 1999; PEIXOTO, 2000; LINS DE BARROS, 1998 apud HEILBORN; CABRAL, 2006, p. 226).

A literatura também evidencia que “[...] não existe somente um tipo de juventude, mas grupos juvenis que constituem um conjunto heterogêneo, com diferentes parcelas de oportunidades, dificuldades, facilidades e poder nas sociedades” (ESTEVES; ABRAMOVAY, 2007, p. 23). Assim, percebe-se a necessidade de que as juventudes sejam qualificadas a partir de distintas formas, ao mesmo tempo em que sejam compreendidas conceitualmente “[...] como uma categoria complexa e heterogênea, na busca de evitar simplificações e esquematismos” (VELHO, 2006, p. 192).

No entanto, nota-se que a despeito do atual consenso na literatura sobre a heterogeneidade das juventudes, ainda permanece pouco explorado, teórica e empiricamente, as definições sobre quais “juventudes” são essas. Ao se analisar a heterogeneidade das juventudes argumenta-se que é factível entrever diversos padrões nos quais um dado fenômeno seja matizado de acordo com suas distribuições sociais desiguais entre os indivíduos. Ou seja, possibilitando o desvelamento de quais são e como vive cada uma dessas diversas juventudes (PINTO, 2012).

Ainda que apenas o uso do recorte etário possa ser, com alguma razão, compreendido como insuficiente para prover uma análise sociológica das juventudes, a sua utilidade para o estabelecimento do objeto de estudo é indubitável. De outro modo, sem o balizamento etário não há como identificar juventude alguma. Por este motivo utilizou-se a mesma definição do Instituto de Pesquisa Econômica Aplicada - IPEA (2009) e dos recentes trabalhos empíricos realizados no Brasil nos quais a juventude é tratada sob o prisma da transição para a vida adulta (CAMARANO, 2006). Assim, o período etário de 15 a 29 anos configura, neste trabalho, a população de interesse. Ressalte-se que as delimitações de idade não devem ser utilizadas para homogeneizar os indivíduos que tenham o mesmo enquadramento etário (LEÓN, 2009).

Se considerarmos que a juventude assume importância por si mesma devido ao fato de que esta fase da vida se configura como parte do processo de constituição de sujeitos, então devemos levar em conta também que as experimentações desiguais de um dado fenômeno influirão no modo de reprodução social. Isto é, cada indivíduo tomado isoladamente, mas em relação com os demais indivíduos, apresenta um conjunto de aspectos concretos e observáveis que podem demonstrar maior ou menor probabilidade de se experimentar determinadas condições de vida, em que haja maior ou menor escassez de recursos necessários para se aproveitar de oportunidades e gerar novos recursos. Assim, se comprehende que, devido à heterogeneidade das juventudes, estas também sejam desigualmente vulneráveis socialmente.

Compreende-se o conceito de vulnerabilidade social como as condições de vida que um indivíduo pode experimentar, em um dado momento no tempo e contexto social, em que haja escassez de recursos necessários para que esse indivíduo possa se aproveitar de oportunidades e gerar novos recursos com a finalidade de controlar ou suportar forças que atuem sobre o seu bem-estar, condicionando-o a riscos e constrangimentos (KAZTMAN, 2000). Segundo Cunha et. al. (2006), a vulnerabilidade social possui características multifacetadas, não implicando, necessariamente, em uma dicotomização entre indivíduos vulneráveis e não vulneráveis. Por conseguinte, é possível entendê-la a partir da matização de diversos padrões de vulnerabilidade social que abordem as diferentes dimensões sociais desse fenômeno. Destaca-se que os indivíduos que compartilham uma condição semelhante quanto à escassez de determinados recursos tendem a possuir oportunidades sociais semelhantes entre si e desiguais em relação aos demais indivíduos.

Neste aspecto, sem perder o caráter de generalidade, é possível especificar determinados padrões em que a população jovem se distribua quanto às suas próprias especificidades de vulnerabilidade social. Dessa forma, nesse artigo também se explora o desenvolvimento de indicadores sociais, considerando a heterogeneidade dessa população. O ideal é que esses indicadores sejam confiáveis ao mesmo tempo em que

proporcione a comparação entre e dentro de determinadas áreas geográficas que circunscrevam as populações de interesse.

## O MODELO GRADE OF MEMBERSHIP

Os conjuntos nebulosos (*fuzzy sets*) foram propostos originalmente por Zadeh (1965) e se distinguem dos conjuntos discretos (*crisp sets*) porque cada indivíduo, ou elemento, pode ser matizado em um *continuum* de pertencimento entre 0 (zero) e 1 (um) em relação a diversos conjuntos ao mesmo tempo. Portanto, nos conjuntos nebulosos cada indivíduo é definido como membro parcial de todos os conjuntos. Já os conjuntos discretos admitem apenas dois estados de pertencimento para esses indivíduos, isto é, se pertencem a um dado conjunto não podem pertencer a outro ( $1 = \in | 0 = \notin$ ). De fato, os estados de pertencimento definidos pelos conjuntos discretos são exatamente os dois extremos do *continuum* admitido pelos conjuntos nebulosos (BAPTISTA, 2003).

O modelo matemático baseado na teoria dos conjuntos nebulosos que descreve a estimação dos parâmetros do método GoM e a função de máxima verossimilhança que define o processo iterativo de convergência desses parâmetros foram escritos originalmente por Woodbury e Clive (1974) e, posteriormente, por Manton, Woodbury e Tolley (1994). Nesse contexto, o método GoM foi proposto com a finalidade de identificar padrões latentes a partir de um conjunto de dados categóricos, ou seja, variáveis qualitativas ordinais ou nominais.

A principal diferença entre o método GoM em relação aos demais métodos de agrupamento, redução e mineração de dados, é que esses últimos se baseiam na teoria dos conjuntos discretos (*crisp sets*). Portanto, esses métodos fornecem resultados apenas sobre o pertencimento pleno de um indivíduo  $i$  a um dado conjunto  $k$ , ou seja, o indivíduo pertence ou não pertence a um dado conjunto. O GoM, por outro lado, possibilita a estimação dos graus de pertencimento  $g$  do indivíduo  $i$  a distintos conjuntos  $k$ . Dessa forma, ao se utilizar o método GoM assume-se que os indivíduos não se organizam segundo a lógica binária de pertencimento a conjuntos discretos (MANTON; WOOBURY; TOLLEY, 1994).

O modelo matemático de GoM estima iterativamente dois conjuntos de parâmetros. O parâmetro lambda (parâmetro locacional) descreve padrões de probabilidade de resposta. O segundo parâmetro, gama (parâmetro individual) informa a distância de cada indivíduo em relação a cada um dos padrões de probabilidade de resposta (MANTON; WOOBURY; TOLLEY, 1994).

O conjunto dos graus de pertencimento de cada indivíduo, em relação a cada padrão de probabilidade de resposta, é definido como partições nebulosas por meio das

quais se representa e se parametriza a heterogeneidade ao nível dos indivíduos (WOODBURY; CLIVE; GARSON JUNIOR, 1978; ZADEH, 1965). Em contrapartida, torna-se evidente que ao se considerar apenas os estados de pertencimento admitidos pelos conjuntos discretos então é “[...] necessário pressupor que o estado [ou conjunto] ao qual uma pessoa pertence é homogêneo” (BAPTISTA, 2003, p. 21). Portanto, observa-se que o método GoM pode ser satisfatoriamente aplicado a problemas de pesquisa que pressupõem que um dado fenômeno seja experimentado de modo heterogêneo, ao longo das diversas dimensões que esse fenômeno possa se apresentar para os indivíduos (PINTO, 2012).

As partições nebulosas são representadas pelos graus de pertencimento  $g$  de um cada indivíduo  $i$  em relação a diversos conjuntos  $k$  ( $g_{ik}$ ). Os gamas assumem valores no intervalo entre 0 (zero) e 1 (um) e a soma dos graus de pertencimento  $g$  de cada indivíduo  $i$  considerando cada conjunto  $k$ , de  $k_1$  a  $k_n=K$ , deve totalizar 1 (um) (MANTON; WOOBURY; TOLLEY, 1994).

O parâmetro gama ( $g_{ik}$ ), que parametriza os graus de pertencimento, define-se enquanto uma medida de distância em relação a cada um dos padrões de probabilidade de resposta  $k$ . Ou seja,  $g_{ik}$  descreve um atributo individual que indica o quanto das características referidas para o dado conjunto de padrões de probabilidade de resposta  $k$  pode ser esperado como realização para cada indivíduo  $i$  (MANTON; WOOBURY; TOLLEY, 1994).

Considera-se que para cada indivíduo  $i$  existem respostas para todas as  $J$  diferentes variáveis. Para cada variável  $j$  existem  $L_j$  possibilidades, ou categorias, de resposta, sendo que cada indivíduo  $i$  responde a uma, e somente a uma, dentre todas as possibilidades de resposta para cada variável  $j$ . Assim, para cada variável  $j$ , o conjunto das possibilidades de resposta  $L_j$  pode ser representado por variáveis binárias aleatórias  $Y_{ijl}$ . Ou seja, para cada possível resposta, de  $l_{ij}$  a  $l_{ij}=L_j$ , considerando uma mesma variável  $j$ , há uma variável binária aleatória dada por  $Y_{ijl}$  (MANTON; WOOBURY; TOLLEY, 1994). Abaixo, são estabelecidos os pressupostos do método GoM:

- a) As variáveis aleatórias  $Y_{ijl}$  são independentes para diferentes indivíduos  $i$ ;
- b) Os diferentes valores de  $g_{ik}$  são representados pelos componentes do vetor aleatório  $\xi_i = (\xi_{k1}, \xi_{k2} \dots \xi_{kn=K})$  com função de distribuição  $H(x) = P(\xi_i \leq x)$ ;
- c) Se  $g_{ik}$  é conhecido, então as respostas de cada indivíduo  $i$ , para cada variável binária aleatória  $Y_{ijl}$  são independentes para diferentes variáveis  $j$ ;

- d) A probabilidade da  $l$ -ésima resposta para a  $j$ -ésima variável pelo  $i$ -ésimo indivíduo é dada por  $\lambda_{kjl}$ ;

Considera-se que o parâmetro  $\lambda_{kjl}$  (lambda), também chamado de parâmetro locacional, diferentemente de  $g_{ik}$ , possui todas as propriedades matemáticas dos conjuntos discretos. O parâmetro  $\lambda_{kjl}$  descreve a probabilidade  $\lambda$  de ocorrência de cada resposta  $l$  para toda variável  $j$  para todo  $k$ . Porém, as probabilidades definidas por  $\lambda_{kjl}$  são válidas apenas quando se considera o indivíduo  $i$  com grau de pertencimento  $g_{ik}$  igual a 1 (um) para o referido conjunto de padrões de probabilidade de resposta  $k$  (MANTON; WOOBURY; TOLLEY, 1994; CAETANO; MACHADO, 2009). Assim, o último pressuposto do modelo GoM deriva dos pressupostos anteriores:

- e) A probabilidade de ocorrência da  $l$ -ésima resposta para a  $j$ -ésima variável pelo  $i$ -ésimo indivíduo, condicional aos distintos valores de  $g_{ik}$  é dada por:

$$P(Y_{ijl} = 1) = \sum_{k=1}^K g_{ik} \lambda_{kjl}$$

Considerando cada conjunto das probabilidades dadas por  $\lambda_{kjl} \forall l_j$ , mantendo-se  $k$  fixo, obtém-se um padrão de probabilidades de resposta para todo  $k$ . Cada padrão de probabilidades de respostas é denominado perfil extremo. Ao se utilizar variáveis que operacionalizem mais do que uma dimensão conceitual para compor o modelo GoM, em relação a quaisquer fenômenos, então cada perfil extremo pode ser chamado de perfil extremo multidimensional (PINTO, 2012).

Dado que os perfis extremos multidimensionais variam de  $k_1$  a  $k_n=K$  e possuem todas as propriedades matemáticas e estatísticas dos conjuntos discretos, então cada perfil extremo  $k$  possui uma relação condicional aos valores de  $g_{ik} \forall i$ . Logo, cada um dos indivíduos  $i$  possui uma medida de distância  $g_{ik}$  em relação a cada um dos padrões de probabilidade de resposta. Ou seja, para todo indivíduo  $i$  existe uma medida de distância em relação a cada um dos vários padrões de probabilidades de que determinadas características possam ser observadas nesse mesmo indivíduo  $i$ . Assim, o modelo de probabilidade de GoM, em que  $g_{ik} \forall i$  é supostamente conhecido, é dado por:

$$E(Y_{ijl}) = \sum_{k=1}^K g_{ik} \lambda_{kjl}$$

No processo de estimação do GoM o modelo de probabilidade é parte de uma função de máxima verossimilhança. Assumindo a independência dos indivíduos e condicionais aos valores de  $\lambda_{kjl}$  e  $g_{ik}$  a função de máxima verossimilhança é definida como:

$$\ln(Y) = \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^{L_J} \left( \sum_{k=1}^K g_{ik} \lambda_{kjl} \right)^{Y_{ijl}}$$

Esta função é estimada iterativamente, ou seja, com ajustamentos sucessivos e com alternância entre o parâmetro a ser ajustado e o parâmetro a ser mantido constante, até que se obtenha a convergência dos parâmetros  $\lambda_{kjl}$  e  $g_{ik}$ , permitindo a maximização da função máxima verossimilhança<sup>3</sup> (MANTON; WOOBURY; TOLLEY, 1994; MACHADO, 1997).

## FONTE DE DADOS E MATERIAIS DE ANÁLISE

A partir dos microdados do Censo Demográfico 2010 foram construídas 18 variáveis, perfazendo um total de 71 (setenta e uma) categorias de resposta, que se referem às dimensões de renda, educação, trabalho, tipos de arranjos familiares e condições de vida no domicílio. Todas essas variáveis foram observadas em nível de mensuração qualitativo ordinal ou nominal e as observações em branco foram categorizadas com código específico para não informação. Após o tratamento dos dados, utilizou-se o método *Grade of Membership* (GoM) em que foram considerados a amostra total de 92.765 jovens com idade entre 15 e 29 anos residentes em domicílios particulares permanentes na Região Metropolitana de Belo Horizonte – MG (RMBH).

As variáveis utilizadas para compor o modelo GoM são as seguintes<sup>4</sup>:

*Grupo Etário*: "De 15 a 17 anos"; "De 18 a 19 anos"; "De 20 a 22 anos"; "De 23 a 24 anos"; "De 25 a 29 anos".

*Sexo*: "Masculino"; "Feminino".

*Cor ou Raça*: "Pretos/Pardos/Indígenas"; "Brancos/Amarelos"; "Ignorado".

*Nível de Instrução Formal*: "Sem instrução e fundamental incompleto"; "Fundamental completo e médio incompleto"; "Médio completo e superior incompleto"; "Superior completo"; "Não determinado".

<sup>3</sup> A respeito da estimação iterativa dos parâmetros (lambda e gama) ver: Woodbury e Clive (1974), Manton, Woodbury e Tolley (1994), Baptista (2003), Cerqueira (2004) e Caetano e Machado (2009).

<sup>4</sup> Em *ítálico* apresenta-se o nome da variável e, entre “aspas”, as categorias que compõem a variável em questão.

*Defasagem Escolar Idade/Grau:* "Não estudantes"; "Estudantes defasados"; "Estudantes regulares"; "Não determinados".

*Condição Associada de Frequência Escolar e Trabalho:* "Não trabalha, não procura trabalho e não frequenta escola"; "Procura trabalho e não frequenta escola"; "Trabalha e não frequenta escola"; "Trabalha e frequenta escola"; "Procura trabalho e frequenta escola"; "Não trabalha, não procura trabalho e frequenta escola".

*Formalidade do Trabalho:* "Trabalhador informal"; "Trabalhador formal"; "Não trabalha".

*Renda do Trabalho:* "Não trabalha"; "Não remunerado/Sem rendimento/Sem informação"; "Menos de 1 SM"; "De 1 a menos de 2 SM"; "De 2 SM e mais".

*Renda Mensal Domiciliar Per Capita:* "Menos de 1 SM"; "De 1 a menos de 2 SM"; "De 2 SM e mais".

*Renda Mensal do Indivíduo em Relação à Renda Mensal do Domicílio:* "Não trabalha e/ou Não possui renda e/ou Domicílio não tem rendimento"; "De 50,0% ou mais"; "De 25 a 49,9%"; "Até 24,9%".

*Densidade de Moradores por Dormitório:* "Acima de 3"; "Acima de 2 até 3"; "Acima de 1 até 2"; "Até 1".

*Posição no Domicílio:* "Pessoa responsável"; "Cônjuge"; "Genro/nora"; "Filho(a)/enteado(a)"; "Outros".

*Estado Conjugal:* "Sim"; "Não, mas viveu"; "Não, nunca viveu".

*Número de Filhos:* "Nenhum Filho"; "1 filho"; "2 filhos"; "3 filhos e mais"; "Não identificável".

*Idade Com que Teve o Primeiro Filho:* "Menos de 20 anos"; "Com 20 anos ou mais"; "Não aplicável ou não identificável".

*Nível de Instrução Formal da Mãe:* "Sem instrução e fundamental incompleto"; "Fundamental completo e médio incompleto"; "Médio completo e superior incompleto"; "Superior completo"; "Educação da mãe não identificável ou não determinada".

*Migrante na Década:* "Migrante de retorno"; "Migrante por ocasião"; "Não migrante".

*Esgotamento Sanitário do Domicílio*: “Inexistência de sanitário ou buraco para dejeções”; “Outras formas de esgotamento sanitário”; “Rede geral de esgoto ou pluvial/Fossa séptica”.

Para realizar a análise de *Grade of Membership* utiliza-se em geral o programa computacional GOM3.EXE (versão 3.4)<sup>5</sup>. Porém, a capacidade de alocação de memória e processamento deste programa implica em severas limitações à análise dos dados, de modo que o total de variáveis a serem analisadas é de no máximo  $J = 28$  e no máximo  $I = 4.000$  observações, aproximadamente. Lage (2005) realizou algumas modificações no código fonte de GOM3.EXE e recompilou esse *software* com o nome gom3b, assim adaptando-o para a plataforma *LINUX*, suportando o total de  $J = 100$  variáveis e  $I = 1.024.000$  observações. Porém, a versão gom3b possui alguns problemas na definição das matrizes de valores iniciais para o conjunto dos parâmetros  $\lambda_{kjl}$ .

Em face da crescente demanda pelo GoM no Brasil, sobretudo no campo da demografia, e dado o seu potencial analítico no campo das Ciências Sociais, os autores se empenharam na construção de alternativas para as limitações do GOM3.EXE. Diante das dificuldades para uma nova atualização do código fonte de GOM3.EXE, por ser um código antigo e de difícil compreensão, optou-se por construir uma nova rotina computacional. É importante frisar que os algoritmos construídos por Woodbury e Clive (1974) e por Charpentier e Singer (1992), desenvolvidos computacionalmente através da linguagem *C*, foram integralmente incorporados a essa nova rotina computacional baseada no ambiente e na linguagem de programação *R*.

Destarte, construiu-se uma rotina de programação que empregou conjuntamente as linguagens *R* e *C/C++*. Essa rotina foi desenvolvida como um R-Script e foi denominado GoMRcpp.R<sup>6</sup>. Por meio do R-Script GoMRcpp.R é possível utilizar um amplo conjunto de  $J$  variáveis e uma considerável quantidade de  $I$  observações. Também foram corrigidos os problemas encontrados na versão gom3b para o *LINUX*. Vale destacar que, de modo geral, os algoritmos iterativos podem consumir muito tempo de análise se a quantidade de parâmetros a ser estimada for muito grande, isto é, quanto maiores forem os valores de  $I, K, J$  e  $\sum L_j = L_J$ .

## RESULTADOS

<sup>5</sup> O *software* GOM3.EXE foi desenvolvido em ambiente *MS-DOS*. Este software, que possui código fonte aberto e livre, foi escrito originalmente por Woodbury e Clive em 1975, e, posteriormente, modificado por Peter Charpentier e Burton Singer, do Departamento de Epidemiologia e Saúde Pública da Escola de Medicina da Universidade de Yale, EUA, em 1992. Disponível em: <<http://lib.stat.cmu.edu/DOS/general/.index.html>>.

<sup>6</sup> O R-Script GoMRcpp.R está disponível mediante solicitação por e-mail ao primeiro autor.

Após estimação dos modelos GoM, considerando distintos modelos com  $K$  perfis extremos multidimensionais de vulnerabilidade social das juventudes, aplicou-se o Critério de Informação de Akaike (AIC)<sup>7</sup> para se definir o modelo mais ajustado. Portanto, considera-se que um modelo GoM, assim ajustado, possua alto grau de robustez quanto à estimação dos seus dois conjuntos de parâmetros ( $\lambda_{kjl}$  e  $g_{ik}$ ). Diante dos resultados (Tabela 1), optou-se pelo modelo com  $K = 3$  perfis.

**Tabela 1** - Valores de AIC para modelos GoM com  $K$  perfis – RMBH, 2010

$K$	$P$	$\ln(L)$	AIC
2	185672	-1581596,88	3534537,75
3	<b>278508</b>	<b>-1263696,50</b>	<b>3084409,00</b>
4	371344	-1308490,13	3359668,25
5	464180	-1223070,88	3374501,75
6	557016	-1089990,75	3294013,50

Fonte: Elaboração dos autores.

A partir da aplicação do AIC aos modelos GoM e com intuito de delinejar perfis compostos por categorias de resposta relevantes – isto é, a partir de estimativas significativas de  $\lambda_{kjl}$  – descreve-se os perfis multidimensionais em função das categorias  $l$  das variáveis  $j$  em cada perfil extremo  $k$  que cumprem um dos seguintes critérios: a) Razão entre o total marginal estimado ( $\lambda_{kjl}$ ) e observado no total da população (%) igual ou maior que 1,2; b) Razão entre o total marginal estimado ( $\lambda_{kjl}$ ) e observado no total da população (%) igual ou maior que 1,2 e no mínimo 80% do perfil sendo definido pelo total marginal estimado; c) Razão entre o total marginal estimado ( $\lambda_{kjl}$ ) e observado no total da população (%) igual ou maior que 1,5 e no mínimo 20% do perfil sendo definido pelo total marginal estimado.

De fato, os perfis extremos multidimensionais de vulnerabilidade social das juventudes que foram encontrados se diferenciam quanto aos seus padrões de vulnerabilidade social e perpassam as dimensões de renda, educação, trabalho, tipos de arranjos familiares e condições de vida no domicílio. Portanto, apresenta-se abaixo a

<sup>7</sup> O AIC é uma medida baseada em teoria da informação estatística. Comparando os modelos GoM com o uso dessa medida, é preferido sempre o modelo com  $K$  perfis extremos que apresentar o menor valor para o AIC. Dessa forma, o AIC indica um melhor ajuste e uma maior parcimônia quanto ao número de perfis estimados, ao mesmo tempo em que afasta a propensão de “superajustamento” no modelo encontrado. Para maiores detalhes sobre o uso do Critério de Informação de Akaike como critério de ajuste entre modelos GoM ver: Guedes et al. (2010).

descrição, em termos das probabilidades de ( $\lambda_{kjl}$ ) relevantes<sup>8</sup>, dos 3 (três) perfis extremos gerados pelo método GoM:

*Perfil 1:* Este perfil é composto em sua maioria por jovens do grupo etário de 15 a 19 anos. É indistinto quanto ao sexo e etnia. Cerca de 30% possuem nível de instrução formal equivalente à “Sem instrução e fundamental incompleto” e 40% possuem “Fundamental completo e médio incompleto”. Um terço desses jovens possui defasagem escolar por idade/grau. Aproximadamente metade desses jovens está na condição de “Não trabalha, não procura trabalho e não frequenta escola” ou “Procura trabalho e não frequenta escola”. Nenhum jovem deste perfil exerce trabalho remunerado. A renda mensal domiciliar *per capita* para aproximadamente 73% dos jovens neste perfil é de menos de 1 (um) salário mínimo. Possuem exclusivamente posição de “filhos/enteados” ou “outros” no domicílio. A educação da mãe para aproximadamente 60% dos jovens deste perfil é de “Sem instrução e fundamental incompleto”.

*Perfil 2:* Este perfil é composto em sua maioria por indivíduos do grupo etário de 23 a 29 anos. É indistinto quanto ao sexo e etnia. Esse perfil é exclusivamente formado por jovens que abandonaram a escola. Todos os indivíduos desse perfil não frequentam escola, mas trabalham. Cerca de 20% desses jovens ocupam postos de trabalho informais. Para aproximadamente 74% dos jovens deste perfil a renda do trabalho é de mais de 1 (um) salário mínimo. Aproximadamente 75% desses jovens possuem renda de “50% ou mais” em relação à renda do domicílio em que vivem. A densidade morador/dormitório é de até 1 (um) para cerca de 20% desses jovens. Possuem exclusivamente posição de “Pessoa responsável”, “Cônjugue” ou “Genro/nora” no domicílio. Aproximadamente 92% desses jovens vivem em companhia de cônjuge ou companheiro (a). Cerca de 60% desses jovens possuem pelo menos 1 (um) filho, sendo que para 26% a idade ao primeiro filho foi com menos de 20 (vinte) anos.

*Perfil 3:* Este perfil é composto em sua maioria por jovens do grupo etário de 18 a 24 anos. É indistinto quanto ao sexo e etnia. O nível de instrução formal dos jovens desse perfil é bastante variável. Cerca de 20,0% são estudantes e nenhum deles se encontrava em defasagem escolar por idade/grau. Todos os jovens deste perfil trabalham e cerca de 30% desses jovens ocupam postos de trabalho informais. Para aproximadamente 45% dos jovens deste perfil a renda do trabalho é de menos de 1 (um) salário mínimo. Contudo, para aproximadamente 66% desses jovens a renda mensal domiciliar *per capita* é superior a 1 (um) salário mínimo. Todos os jovens desse perfil possuem renda de “Até 24,9%” ou “De 25 a 49,9%” em relação à renda do domicílio em que vivem. Possuem exclusivamente posição de “filhos/enteados” ou “outros” no

---

<sup>8</sup> A tabela com as informações sobre as frequências absolutas observadas (N), frequências relativas observadas (%), estimativas de  $\lambda_{kjl}$  e razões entre  $\lambda_{kjl}$  e frequências relativas observadas (%) está disponível mediante solicitação por e-mail ao primeiro autor.

domicílio. Aproximadamente 93% desses jovens nunca viveram em companhia de cônjuge ou companheiro (a). A educação da mãe para aproximadamente 60% dos jovens deste perfil é de “Sem instrução e fundamental incompleto”.

Por definição do método empregado cada indivíduo é necessariamente membro parcial de cada um dos perfis extremos multidimensionais descritos acima, e é necessário, para que áreas geográficas sejam analisadas, empregar uma medida que a sintetize as informações do conjunto dos indivíduos residentes em cada área. Porém, esse procedimento não deve levar à perca do foco analítico sobre o próprio indivíduo. Assim, sugere-se a realização do cálculo da taxa de prevalência dos perfis extremos multidimensionais, definida como:

$$P_{g_{ik_a}} = \left( \frac{\sum g_{ik_a}}{Pop_a} \right) * 1000$$

Onde:

$g_{ik_a}$  – Grau de pertencimento de cada indivíduo  $i$ , que seja residente na área  $a$ , em relação ao dado perfil  $k$ .

$Pop_a$  – População de indivíduos  $i$  residentes na área  $a$ .

$P_{g_{ik_a}}$  – Taxa de prevalência do perfil  $k$ , na área  $a$ , considerando a soma dos graus de pertencimentos de todos os indivíduos  $i$  residentes na área  $a$  em relação ao perfil  $k$ .

Desse modo,  $P_{g_{ik_a}}$  descreve uma taxa mínima da população da área  $a$  que possui grau de pertencimento  $g_{ik}$  diferente de 0 (zero). Logo,  $P_{g_{ik_a}}$  varia no intervalo de 0 (zero) a 1.000 (mil) para todas as áreas sob análise. Portanto  $P_{g_{ik_a}}$  será igual a 0 (zero) se  $\sum g_{ik_a} = 0$ , ou seja, se nenhum indivíduo possuir grau de pertencimento diferente de 0 (zero) para um determinado perfil extremo no dado território em questão. De outra forma,  $P_{g_{ik_a}}$  será igual a 1.000 (mil) se  $\sum g_{ik_a} = Pop_a$ , ou seja, se todos os indivíduos em um dado território possuírem grau de pertencimento pleno, igual a 1 (um), para um determinado perfil extremo.

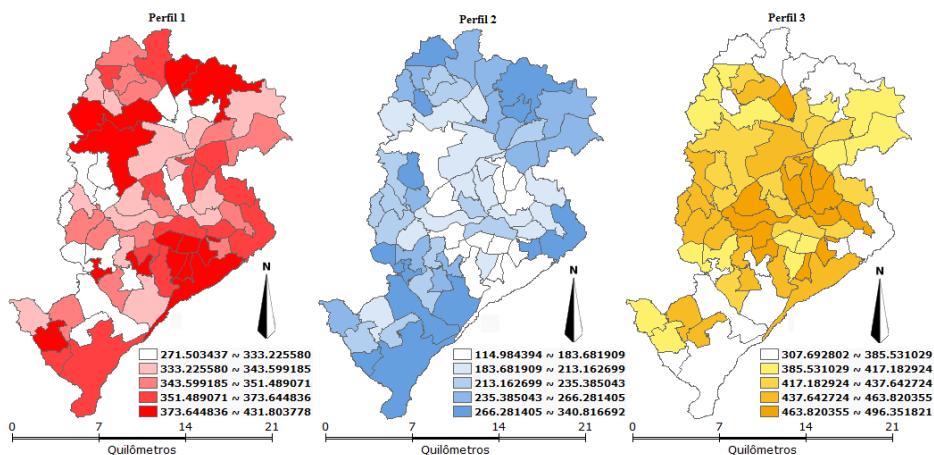
É possível verificar que  $P_{g_{ik_a}}$ , além de se constituir como uma medida ao nível de análise territorial e que, portanto, pode ser espacializada, também possui uma escala de fácil interpretação. Cada unidade a mais na escala de  $P_{g_{ik_a}}$  significa que, em um conjunto de mil indivíduos em um determinado território, haveria uma prevalência de pertencimento equivalente a um indivíduo a mais que viveria com pertencimento pleno ( $g_{ik} = 1$ ) em relação ao dado perfil em questão.

Contudo, deve-se observar que  $P_{gik_a}$  não é uma mensuração direta sobre o nível de análise individual, mas sim o  $g_{ik}$ , independente de qual seja a área  $a$ . Por outro lado, é possível que ambos,  $P_{gik_a}$  e  $g_{ik}$ , possuam algum grau de dependência espacial, mas somente  $P_{gik_a}$  pode se referir ao nível de análise territorial.

A dependência espacial pode ser entendida como o “[...] que Waldo Tobler chama de primeira lei da geografia: ‘todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes’.” (CÂMARA et al., 2004, p. 11). Analisando a dependência espacial é possível verificar a concentração territorial dos valores de  $P_{gik_a}$  e o quanto os territórios que possuem vizinhança entre si estão autocorrelacionados através desse atributo.

Para avaliar a distribuição e a dependência espacial de  $P_{gik_a}$  considerou-se apenas o município de Belo Horizonte, Minas Gerais, utilizando suas 67 áreas de ponderação como unidades territoriais de análise. Calculou-se o Índice Global de Moran para os valores de  $P_{gik_a}$  utilizando-se a matriz de proximidade normalizada na linha e segundo a contiguidade entre as áreas geográficas, para cada  $P_{gik_a}$ . Na Figura 1 podem ser observadas as distribuições dos  $P_{gik_a}$  por classes de quintis:

**Figura 1 – Distribuição Espacial de Classes de Quintis das Taxas de Prevalência dos Perfil Extremos de Vulnerabilidade Social das Juventudes ( $P_{gik_a}$ ) para as Áreas de Ponderação do Município de Belo Horizonte – MG, 2010\***



**Fonte:** Elaboração dos autores.

\*Moran Global – Autocorrelação do Perfil 1: 0,249348; p-valor: 0,01

\*Moran Global – Autocorrelação do Perfil 2: 0,305120; p-valor: 0,01

\*Moran Global – Autocorrelação do Perfil 3: 0,219893; p-valor: 0,02

Observa-se que, com 95,0% de confiança, é possível rejeitar a hipótese de inexistência de autocorrelação espacial entre as áreas. Ao mesmo tempo, para todos os perfis extremos, verifica-se que a autocorrelação espacial é positiva e sua magnitude é superior a 20%. Esses resultados evidenciam que os perfis de vulnerabilidade social das juventudes tendem a possuir forte estrutura de dependência espacial, isto é, as áreas mais próximas tendem a possuir valores mais semelhantes entre si.

Destaca-se que, apesar de  $P_{g_{ik_a}}$  indicar a concentração espacial de prevalências de pertencimento dos indivíduos aos perfis extremos multidimensionais de vulnerabilidade social das juventudes, verifica-se que em unidades territoriais com os menores valores nas classes de quintis, há um contingente populacional que possui grau de pertencimento ( $g_{ik}$ ) diferente de 0 (zero), para cada perfil, na ordem mínima de aproximadamente 11,4%, ou 114 indivíduos em 1000, considerando a prevalência do Perfil 2 na população jovem do território em questão.

## CONCLUSÃO

A heterogeneidade das juventudes é frequentemente considerada em sua perspectiva teórica, mas raramente é incorporada metodológica e empiricamente. Esse é o ponto de partida deste artigo, incorporar a condição inerente de heterogeneidade da juventude, ou seja, juventudes, e a multidimensionalidade da vulnerabilidade social. Para atingir este objetivo, empregamos o método *Grade of Membership* (GoM), ferramenta de análise empírica que permite incluir esses dois aspectos no que tange as formas latentes e os graus com que os indivíduos, espacialmente localizados em um dado território, experimentam os fenômenos sociais. A descrição dos três perfis delineados permite identificar um grupo de jovens em maior vulnerabilidade (Perfil 1), um segundo grupo de jovens que estabeleceram família e estão inseridos no mercado de trabalho (Perfil 2) e, por fim, um grupo de jovens que configuram uma mescla de situações, desde estudantes sem distorção idade/grau a trabalhadores ocupados no setor informal (Perfil 3). Os mapas temáticos permitem visualizar as áreas do município segundo a prevalência de cada perfil.

O GoM difere dos métodos de classificação frequentemente utilizados nas Ciências Sociais por não categorizar os elementos de um conjunto qualquer – jovens espacialmente referenciados no território metropolitano, neste caso – em termos discretos. Em outras palavras, um dado indivíduo pode pertencer simultaneamente a

todos os perfis delineados a partir das informações do conjunto da população alvo do território em questão.

De outra forma, a maioria das ferramentas de análise multivariada são, por construção, métodos de classificação discreta via maximização da homogeneidade intragrupo e da heterogeneidade intergrupos. Como consequência, as áreas que compõem o território em análise frequentemente assumem o lugar dos indivíduos. Assim, considerando uma análise espacial, cada área adquire uma coloração que diz respeito à média dos indivíduos que nelas residem, portanto, homogeneizando a heterogeneidade dos indivíduos. De modo geral, os indicadores sociais atualmente empregados no processo de análise das realidades sociais são ainda mais homogeneizadores, pois o índice de cada área é, de fato, uma média sintetizada de todas as variáveis incluídas no modelo de construção do indicador em questão.

O método GoM, leva em conta o nível de análise individual (grau de pertencimento) em conjunto com a latência das diversas formas com que um dado fenômeno social se apresenta (perfis extremos multidimensionais). É importante ter em mente que os parâmetros estimados pelo método permitem que o GoM seja, em uma série de situações, utilizado como meio para outras etapas analíticas, conforme pôde ser notado.

## REFERÊNCIAS

BAPTISTA, Dulce Benigna Dias Alvarenga. *Idosos no município de São Paulo: expectativa de vida ativa e perfis multidimensionais de incapacidade a partir da SABE*. 2003. Dissertação (Mestrado em Demografia) – Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em: <[http://www.cedeplar.ufmg.br/demografia/dissertacoes/2003/Dulce\\_Benigna.pdf](http://www.cedeplar.ufmg.br/demografia/dissertacoes/2003/Dulce_Benigna.pdf)>. Acesso em: 4 mar. 2010.

CAETANO, André Junqueira; MACHADO, Carla Jorge. Consistência e identificabilidade no modelo Grade of Membership: uma nota metodológica. *Revista Brasileira de Estudos Populacionais*, Rio de Janeiro, v. 26, n. 1, p. 145-149, jan./jun., 2009. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S010230982009000100012](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010230982009000100012)>. Acesso em: 7 dez. 2009.

CÂMARA, Gilberto et al. Análise espacial e geoprocessamento. In: DRUCK, Suzana et al. (Ed). *Análise espacial de dados geográficos*. Brasília: EMBRAPA, 2004. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>>. Acesso em: 7 maio 2012.

CAMARANO, Ana Amélia (Org.). *Transição para a vida adulta ou vida adulta em transição?* Rio de Janeiro: IPEA, 2006.

CERQUEIRA, Cezar Augusto. *Tipologia e características dos estabelecimentos escolares brasileiros*. Tese de Doutorado, Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional – CEDEPLAR, 2004. Disponível em: <[http://www.cedeplar.ufmg.br/demografia/teses/2004/Cezar\\_Augusto\\_Cerqueira.pdf](http://www.cedeplar.ufmg.br/demografia/teses/2004/Cezar_Augusto_Cerqueira.pdf)>. Acesso em: 07 dez. 2009.

CUNHA, José Marcos Pinto et al. A vulnerabilidade social no contexto metropolitano: o caso de Campinas. In: CUNHA, José Marcos Pinto (Org.). *Novas metrópoles paulistas: população, vulnerabilidade e segregação*. Campinas: Nipo/Unicamp, 2006. p.143-168.

DAYRELL, Juarez; MOREIRA, Maria Ignez Costa; STENGEL, Márcia (Org.). *Juventudes contemporâneas: um mosaico de possibilidades*. Belo Horizonte: Ed. PUC Minas, 2011.

ESTEVES, Luiz Carlos Gil; ABRAMOVAY, Mirian. Juventude, juventudes: pelos outros e por elas mesmas. In: ABRAMOVAY, Mirian; ANDRADE, Eliane Ribeiro; ESTEVES, Luiz Carlos Gil (Org.). *Juventudes: outros olhares sobre a diversidade*. Brasília: Ministério da Educação, Secretaria de Educação Continuada, Alfabetização e Diversidade; UNESCO, 2007. p. 21-56.

GUEDES, Gilvan Ramalho et al. Identificabilidade e estabilidade dos parâmetros no método Grade of Membership (GoM): considerações metodológicas e práticas. *Revista Brasileira de Estudos de População*, Rio de Janeiro, v. 27, n. 1, p. 21-33, 2010. Disponível em: <<http://www.scielo.br/pdf/rbepop/v27n1/03.pdf>>. Acesso em: 16 fev. 2011.

HEILBORN, Maria Luiza; CABRAL, Cristiane S. Parentalidade juvenil: transição condensada para a vida adulta. In: CAMARANO, Ana Amélia (Org.). *Transição para a vida adulta ou vida adulta em transição?* Rio de Janeiro: IPEA, 2006. Cap. 8, p. 225-256.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEA. PNAD 2008: primeiras análises. Comunicado da Presidência n. 36, dez. 2009. Disponível em: <[http://www.ipea.gov.br/sites/000/2/comunicado\\_presidencia/09\\_12\\_03\\_ComunicaPresi\\_36\\_PNAD2008\\_juventDesigual.pdf](http://www.ipea.gov.br/sites/000/2/comunicado_presidencia/09_12_03_ComunicaPresi_36_PNAD2008_juventDesigual.pdf)>. Acesso em: 20 mar. 2010.

KAZTMAN, Ruben. *Notas sobre la medición de la vulnerabilidad social*. México: BID-BIRF-CEPAL, 2000. Disponível em: <[www.eclac.cl/deype/mecovi/docs/TALLER5/24.pdf](http://www.eclac.cl/deype/mecovi/docs/TALLER5/24.pdf)>. Acesso em: 1 mar. 2010.

LAGE, Rafael Kelles Vieira. *Com 3b*. Belo Horizonte: [s. n], 2005. Cd-Rom.

LEÓN, Oscar Dávila. Uma revisão das categorias de adolescência e juventude. In: GUIMARÃES, Maria Tereza Canezin; SOUZA, Sônia M. Gomes (Org.). *Juventude e contemporaneidade: desafios e perspectivas*. Goiânia: Ed. UFG, 2009. p. 47-76.

MACHADO, Carla Jorge. *Perfis de morbi-mortalidade infantil no estado de São Paulo, 1994*: uma aplicação de grade of membership à análise de Causas Múltiplas de Morte. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional – CEDEPLAR, 1997. Disponível em: <[http://www.cedeplar.ufmg.br/demografia/dissertacoes/1997/Carla\\_Machado.pdf](http://www.cedeplar.ufmg.br/demografia/dissertacoes/1997/Carla_Machado.pdf)>. Acesso em: 04 mar. 2010.

MANTON, Kenneth G.; WOOMBURY, Max A.; TOLLEY, H. Dennis. *Statistical applications using fuzzy sets*. New York: John Wiley & Sons, 1994.

PAIS, José Machado. A construção sociológica da juventude: alguns contributos. *Análise Social*, Lisboa, v. 25, n. 105/106, p. 139-165, 1990.

PINTO, Julimar Santos. *Perfis multidimensionais de vulnerabilidade social das juventudes e indicadores sintéticos*. 2012. Dissertação (Mestrado em Ciências Sociais) – Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte.

VELHO, Gilberto. Juventudes, projetos e trajetórias na sociedade contemporânea. In: ALMEIDA, Maria Isabel Mendes de; EUGENIO, Fernanda (Org.). *Culturas jovens: novos mapas do afeto*. Rio de Janeiro: Jorge Zahar, 2006.

WOODBURY, Max A.; CLIVE, Jonathan. Clinical pure types as a fuzzy partition. *Journal of Cybernetics*, Washington, v. 4, n. 3, p. 111-121, 1974.

WOODBURY, Max A.; CLIVE, Jonathan; GARSON JUNIOR, Arthur. Mathematical Typology: a grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*, San Diego, v. 11, n. 3, p. 277-298, 1978.

ZADEH, Lotfi. A. Fuzzy sets. *Information and Control*, v. 8, p. 338-353, 1965.