

Perfil de risco: método multivariado de classificação sócio-econômica de microáreas urbanas – os setores censitários da região metropolitana do Rio de Janeiro

Risk profile: multivariate cluster analysis of urban small areas based on socioeconomic indicators – census tracts from the Greater Rio de Janeiro Metropolitan Area

Marília Sá Carvalho ¹
Oswaldo Gonçalves Cruz ²
Flávio Fonseca Nobre ³

¹ Departamento de Epidemiologia, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rua Leopoldo Bulhões 1480, Rio de Janeiro, RJ 21041-210, Brasil. marilia@malaria.procc.fiocruz.br

² Programa de Computação Científica, Fundação Oswaldo Cruz, Residência Oficial, Av. Brasil 4365, Rio de Janeiro, RJ, 21450-900, Brasil. oswaldo@malaria.procc.fiocruz.br

³ Programa de Engenharia Biomédica, Coordenação de Programas de Pós-Graduação em Engenharia, Universidade Federal do Rio de Janeiro, Centro de Tecnologia, Bloco H, Sala 327, C. P. 68510, Cidade Universitária, Rio de Janeiro, RJ 21945-970, Brasil. flavio@serv.peb.ufrj.br

Abstract A method based on small area data analysis was developed to build a health risk classification for the Greater Rio de Janeiro Metropolitan Area. The approach uses 1991 census data and studies data pertaining to sanitation, ownership and type of housing, size and occupancy of the household, demography, schooling, and income. Principal component analysis applied over each dimension allowed for the choice of 15 variables, which summarized most of the observed variances. Additional analysis with these variables suggested that just six variables are sufficient for the construction of a classification using k-means method of multivariate cluster analysis. Five classes were obtained: (A) high income; (B) lower income; (C) poor; (D) low schooling and income; (E) low-level access to sanitation. The existing inequality in each of the geopolitical established areas was clearly identified. The proposed method allowed for the construction of compound indices to evaluate quality of life, based on widespread and easily obtained data (the census). Moreover, the method contributed to the detection of socioeconomic inequality, identifying, not only the larger poor regions but also the small excluded areas.

Key words Demographic Census; Small-Area Analysis; Multivariate Analysis; Socioeconomic Analysis

Resumo Visando desenvolver métodos de estudo de microáreas urbanas, que permitam a integração de variáveis em um perfil sócio-econômico, foi analisado o censo demográfico de 1991, por setor censitário. Estudaram-se aspectos de: saneamento, ocupação do terreno, escolaridade, domicílio, demografia e renda. A análise de componentes principais para cada tema permitiu selecionar 15 indicadores com maior poder de explicação da variância observada, com base nos quais, em nova análise, foram escolhidos os sete indicadores finais. A classificação multivariada por partição gerou cinco tipos diferentes de setores censitários, assim descritos: (A) alta renda, (B) renda mais baixa, (C) pobre, com predomínio de casas alugadas, (D) baixa escolaridade e renda, (E) precariedade no saneamento básico. A análise segundo a divisão geopolítica tradicional permitiu identificar as desigualdades dentro de cada área. O método proposto permitiu a criação de índices compostos de avaliação de qualidade de vida, valendo-se de informações de fácil acesso e ampla cobertura. Além disso, contribuiu para detectar a desigualdade sócio-econômica, identificando não só as grandes regiões de pobreza, mas as microáreas de exclusão em meio à relativa riqueza.

Palavras-chave Censo Demográfico; Análise de Pequenas Áreas; Análise Multivariada; Perfil Sócio-econômico

Introdução

O desenvolvimento de métodos que permitam a integração e análise de grande número de variáveis, tendo por objetivo a construção de um índice único, capaz de sintetizar os diferentes aspectos descritos univariadamente, é o propósito geral deste trabalho. Métodos de redução da dimensionalidade e de classificação multivariada foram aplicados a um conjunto de indicadores criados a partir de todas as variáveis do questionário básico do censo demográfico, visando à construção de um índice classificatório que pode ser tomado enquanto um perfil sócio-econômico, usualmente relacionado a risco diferenciado de adoecer ou morrer por diversos agravos. A identificação destes grupos populacionais é uma das questões-chave da prevenção em saúde, permitindo o direcionamento de medidas, sejam elas de cunho ambiental, educacional ou assistencial (Castellanos, 1990; Cortinovis et al., 1993; Verhasselt & Mansourian, 1991).

Um aspecto importante na avaliação de risco diz respeito à metodologia de mensuração utilizada, à qual diferentes aspectos podem e devem ser incorporados, tais como medidas de atributos individuais, usualmente denominados fatores de risco, indicadores de ocorrência passada dos agravos em questão ou detecção de aspectos potencialmente danosos relacionados ao meio (Hayes, 1992). Em um contexto de avaliação de risco de grupos populacionais, a ênfase está nas doenças da população e não do indivíduo, a pergunta que se deseja responder não é sobre as causas dos casos de doença, mas sobre as causas da incidência da doença (Rose, 1985). Neste caso, duas questões sobressaem-se: a caracterização dos grupos sociais e o tratamento de diversos indicadores simultaneamente. A combinação de vários indicadores, os quais são medidas específicas que normalmente refletem apenas aspectos limitados da realidade, pode ser obtida através da criação de índices, que pretendem descrever de forma global determinado problema (Goldberg et al., 1979a;b).

Por outro lado, a caracterização sócio-econômica dos grupos populacionais, além de considerar todas as possíveis informações qualitativas – históricas, sociológicas, geopolíticas – pode ser abordada com base em informações amplamente acessíveis, como os indicadores do censo demográfico. Em que pesem seus limites quanto à qualidade, pertinência e oportunidade da informação, somente a ampla utilização das bases de dados secundários disponíveis permitirá aperfeiçoá-las.

As técnicas utilizadas na caracterização de perfis sócio-econômico foram a análise de componentes principais e método de classificação multivariado por partição, cujos conceitos básicos serão apresentados a seguir.

A análise de componentes principais é uma técnica multivariada utilizada para examinar, simultaneamente, a relação entre grande número de variáveis. Através da combinação linear das variáveis é possível calcular um conjunto de eixos não correlacionados (ortogonais), denominados autovetores, cada um dos quais capaz de representar parte da variância total do sistema. O primeiro componente é aquele que maximiza a variância total, o segundo maximiza o resíduo de variância entre as direções não correlacionadas ao primeiro e assim sucessivamente (Lebart et al., 1986). O peso relativo de cada variável na composição de cada eixo é medido através da sua correlação com este eixo. Desta forma, a análise de componentes principais pode ser utilizada como um instrumento de seleção de variáveis, na medida em que aquelas com maior peso na construção dos primeiros componentes, são as que possivelmente melhor representam o conjunto estudado.

A classificação multivariada, também denominada análise de *cluster*, é um conjunto de procedimentos que visa agrupar e discriminar grupos de indivíduos, regiões ou qualquer objeto. Estes agrupamentos ou *cluster* são constituídos definindo-se critérios baseados em distâncias. Distância é uma medida matemática de similaridade, que pode ser geográfica, temporal ou baseada em qualquer característica do objeto. Quando são utilizadas diversas variáveis é possível a construção de agrupamentos onde o critério de similaridade está distribuído entre diversas características. Na construção dos conglomerados, o objetivo é que os objetos próximos – segundo a métrica selecionada – fiquem no mesmo grupo, enquanto as maiores distâncias separem grupos. Entre as medidas de similaridade, as mais frequentes na literatura são: coeficientes de correlação, medidas de distância e coeficientes de associação.

Definida a métrica utilizada, diversos métodos de agrupamento podem ser empregados, hierárquicos e não hierárquicos. Os métodos hierárquicos podem ser ascendentes, nos quais o algoritmo de cálculo começa com tantos grupos quantos elementos e termina ao se reagrupar todo o conjunto, ou descendentes, em que o conjunto inicial de objetos vai sendo subdividido em aglomerados cada vez menores. O procedimento geral dos métodos não hierárquicos é a partição iterativa em k grupos, utilizando

critérios que diminuam a variância intragrupos e maximizem a variância intergrupos. O método não hierárquico *k-means* apresenta bom desempenho quando o número de grupos é pequeno em comparação com o número de elementos, tendo sido por isso selecionado para esta aplicação. Nos dois casos o objetivo é buscar que elementos semelhantes fiquem no mesmo grupo, e elementos diferentes fiquem em grupos distintos (Lebart et al., 1986).

Um aspecto importante nos trabalhos voltados para a caracterização de regiões é a extensão da base territorial de cada unidade geográfica, agregando grupos sociais distintos (favelas e áreas nobres), o que torna cada indicador selecionado média entre grupos populacionais diferentes. Evidentemente o acesso aos equipamentos urbanos, bem como os agravos à saúde não estão distribuídos homogeneamente na população. Entretanto, frequentemente as informações necessárias não estão disponíveis no nível de agregação desejado. Raras são as informações microlocalizadas, e ainda mais raras as análises.

O censo demográfico, com todo um conjunto de informações gerais sobre a população, maior detalhamento sobre o chefe da família e características do domicílio, pode ser trabalhado em nível de microárea, uma vez que a unidade de agregação é o setor censitário, em média com novecentos habitantes. Neste trabalho, os dados do censo de 1991 foram tratados, criando e identificando variáveis com maior capacidade de expressar as diferenças intersetores, e utilizando-as em um processo de classificação multivariado que gerou uma tipolo-

gia sócio-econômica dos setores censitários. Esta classificação e as variáveis selecionadas foram analisadas quanto à sua inserção na divisão administrativa usual da região, em Áreas Programáticas (APs) do Município do Rio de Janeiro e demais municípios da Região Metropolitana, utilizando o *boxplot*, técnica gráfica de análise exploratória de dados.

Metodologia

Dados

Foram utilizados os dados do censo de 1991, por setor censitário, disponibilizados pela Fundação IBGE em formato digital. A partir destas variáveis, foram criados diversos indicadores – proporções e médias de ocorrência por setor censitário – agrupados em diversos blocos por tema analisado. Foi analisada a Região Metropolitana do Rio de Janeiro, (Figura 1) subdividida em três grupos para fins de comparação interna: Baixada Fluminense, Rio de Janeiro (capital) e demais municípios (Tabela 1).

Foi utilizada a divisão geopolítica em vigor no ano do censo (1991), segundo a qual a Região Metropolitana do Rio de Janeiro é composta por dez municípios. O Município do Rio de Janeiro (capital) foi analisado subdividido nas 24 Regiões Administrativas então existentes. Somente foram utilizados os dados do questionário global, uma vez que as informações relativas à amostra não permitem análise por microáreas. Todas as variáveis são a contagem do evento descrito no setor censitário.

Tabela 1

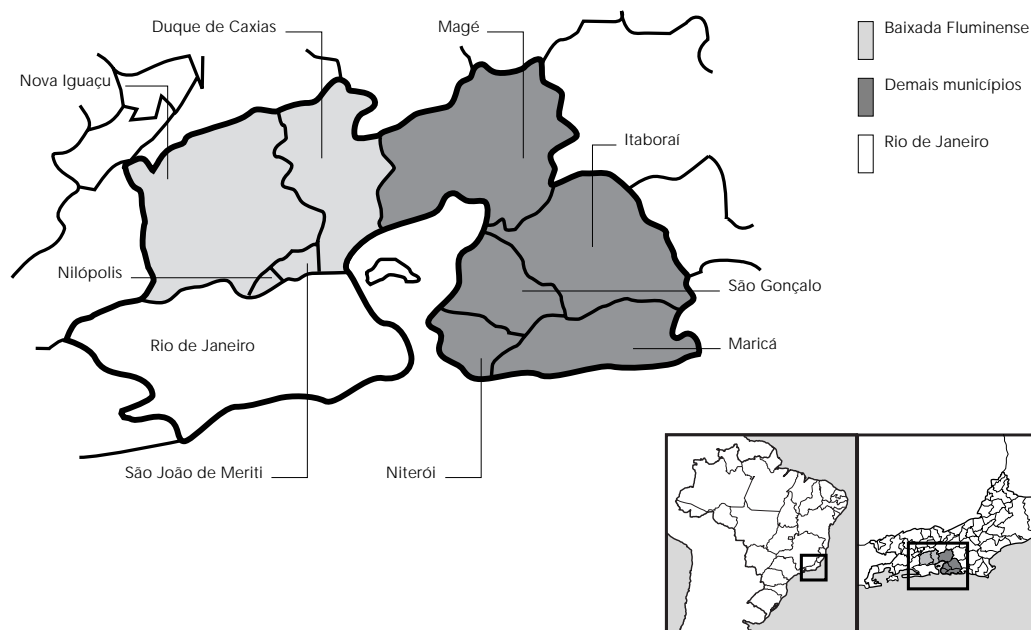
População e total de setores censitários¹ por município da Região Metropolitana do Rio de Janeiro, censo demográfico de 1991.

Sub-região	Municípios	Setores censitários	População
Rio (capital)	Rio de Janeiro	6.258	5.480.768
Baixada	Duque de Caxias	599	667.821
Baixada	Nilópolis	148	158.092
Baixada	Nova Iguaçu	2.320	1.297.704
Baixada	São João de Meriti	353	425.772
Outros	Itaguaí	190	162.742
Outros	Magé	265	191.734
Outros	Maricá	67	46.545
Outros	Niterói	480	436.155
Outros	São Gonçalo	676	779.832
Região Metropolitana		11.356	10.389.441

¹ somente os setores censitários normais e subnormais, excluídos os coletivos.

Figura 1

Rio de Janeiro, divisão geo-política de 1991, mapa de região metropolitana.



O questionário do censo divide-se em duas partes. A primeira dedicada ao domicílio e ao chefe da família, e a segunda dedicada a cada morador (FIBGE, 1994). Algumas definições das variáveis coletadas são necessárias visando esclarecer conceitos utilizados e melhorar a compreensão. Os setores censitários são classificados segundo a situação em urbanos e rurais. No Município do Rio de Janeiro todos os setores são considerados urbanos.

Quanto ao tipo, os setores podem ser:

Coletivos – Setores exclusivamente de habitações coletivas, como asilos, creches, quartéis, presídios e delegacias, que foram excluídos da análise, pois são em número reduzido e distorcem os indicadores, em função de sua feição especial.

Normais – Setores onde há estrutura urbana tradicional, com ruas, endereços, ainda que a propriedade do terreno eventualmente não seja bem definida.

Subnormais – A definição destes setores é pouco clara: “... conjunto constituído por unidades habitacionais ocupando ou tendo ocupado até período recente terrenos de propriedade alheia, dispostos, em geral, de forma desordenada e densa, e carentes, em sua maioria, de serviços públicos essenciais” (FIBGE, 1994).

Assim, em setores ditos ‘normais’, eventualmente podem ser encontrados domicílios precários, e, em setores considerados pelo censo ‘subnormais’, o tipo de moradia e situação de urbanização pode ser ‘normal’. Para fins deste trabalho, consideraram-se os setores ‘subnormais’ como sendo favelas, fazendo-se a análise conjunta dos dois tipos.

Quanto às características do domicílio, a primeira subdivisão separa-os em particulares e coletivos e, entre os particulares, em permanentes e improvisados. A pesquisa das demais características dos domicílios limitou-se aos domicílios particulares permanentes, e pelo mesmo motivo os indicadores criados têm como denominador os domicílios particulares permanentes.

Indicadores

Na Tabela 2 são apresentados todos os indicadores criados, acrescentando-se, quando necessário, as explicações pertinentes. As proporções foram calculadas tomando-se como base o total de domicílios particulares permanentes. A utilização eventual de outros denominadores será devidamente explicada. Na coluna de observações do quadro de indicadores é referida

a utilização de indicadores calculados pela própria FIBGE, e disponibilizados junto com os dados brutos. Algumas médias e medianas de variáveis contínuas foram também calculadas.

Seleção dos indicadores

Cada bloco de indicadores, exceto o demográfico devido à importância atribuída às quatro variáveis no perfil nosológico da população, foi submetido separadamente à análise de componentes principais. As duas ou três variáveis que apresentaram a maior contribuição na definição dos dois primeiros componentes foram selecionadas, sobrando desta primeira etapa da análise 11 variáveis, às quais se acrescentaram aquelas do bloco demográfico, totalizando 15. Estas foram submetidas em conjunto à nova análise de componentes principais, selecionando-se ao final os sete indicadores que melhor representam o conjunto das variáveis analisadas. A variável Renda Média Nominal do Chefe da Família foi reescalada para média 0 e desvio padrão 1.

A partir dos indicadores selecionados, o conjunto de setores censitários da Região Metropolitana do Rio de Janeiro foi submetido à análise classificatória multivariada pelo método *k-means* (Hartigan, 1975), resultando cinco diferentes perfis de classificação dos setores censitários.

Resultados

Os 15 indicadores selecionados inicialmente foram assinalados na Tabela 2 em negrito. A análise conjunta destes permitiu selecionar os sete indicadores utilizados na construção dos agrupamentos: *proporção de domicílios alugados* (CALUGA); *proporção de casas* (CASAS); *proporção de domicílios com água de rede pública* (REDE); *proporção de domicílios sem qualquer tipo de esgoto* (SEMESG); *proporção de população acima de cinco anos alfabetizada* (ALFABET); *proporção de chefes de família com escolaridade igual ou maior que o segundo grau* (2GRAU); *renda média nominal do chefe da família* (RNDCH).

Nenhum indicador demográfico foi utilizado na fase final de análise, pois com a inclusão de qualquer um deles o algoritmo empregado para a criação dos grupos não convergia. Isto é ocasionado pela forte relação entre estes indicadores e os indicadores de renda e escolaridade.

Classificação multivariada

A análise de *cluster* utilizando as sete variáveis selecionadas gerou cinco tipos diferentes de classes, descritas na Tabela 3, onde o valor de cada variável é a média das proporções da variável dentro de cada grupo.

O grupo de tipo "A" apresenta o menor número de setores, apenas 190, e a maior renda (4,8 desvios-padrão acima da média). A proporção de residências de tipo casa é pequena, em média apenas 13% dos domicílios. Os indicadores de escolaridade, especialmente a proporção de chefes de família com o segundo grau ou mais, diferenciam substancialmente este grupo dos demais.

A classe de setores de tipo "B" diferencia-se da anterior essencialmente pela renda, que é 1/3 da renda do grupo "A", embora 1,8 desvios-padrão acima da média geral. Escolaridade do chefe da família acima do segundo grau é um pouco menos freqüente (75%) e a proporção de residências alugadas aumenta de 18% para 26%.

O agrupamento de tipo "C" pode ser caracterizado como um grupo mediano. A renda aproxima-se da média (0,2), apenas 47% dos chefes de família têm escolaridade maior ou igual ao segundo grau. Nesta categoria concentra-se a maior parte das casas alugadas (32%). Os grupos "A", "B" e "C" somados atingem apenas 3.700 setores, ou seja, aproximadamente 33% do total.

Na caracterização do agrupamento "D", a baixa escolaridade do chefe da família (apenas 16% com segundo grau) e a maciça predominância de casas são fundamentais. A renda está ligeiramente abaixo da média global. Neste grupo, localizam-se mais de 50% dos setores da Região Metropolitana do Rio de Janeiro.

A categoria "E" apresenta a mais elevada proporção de analfabetos (33%), 5% dos domicílios particulares permanentes jogam esgotos na própria área da casa e apenas 16% está ligado à rede pública de água. Nos dois últimos tipos diminui novamente a proporção de residências alugadas.

O perfil dos municípios da Região Metropolitana (Figura 2) apresenta três subgrupos nitidamente diferentes:

- Itaguaí, Magé, Maricá, os municípios mais afastados e rurais da Região Metropolitana: os setores censitários concentram-se nas classes "E" e "D", com poucos setores do tipo "C", nenhum setor de tipo "B" ou "A", caracterizando os municípios como os de pior situação sócio-econômica segundo os indicadores escolhidos;
- Caxias, Nilópolis, Nova Iguaçu, São Gonçalo e São João de Meriti, com história semelhan-

Tabela 2

Descrição dos principais indicadores por grupo temático, Censo Demográfico do IBGE, 1991*.

2A) Primeiro Bloco Temático – Estrutura e Inserção do Domicílio.

Assunto	Indicador (proporção de)	Observações
Residência	domicílios improvisados casas (construção unifamiliar) apartamentos casas de cômodos domicílios coletivos residências (casas+ap ^{tos}) isoladas residências em conjuntos populares residências subnormais	Estes indicadores têm como denominador o total de domicílios, particulares ou coletivos, permanentes ou improvisados.
Propriedade	domicílios próprios em terreno próprio construção própria e terreno não próprio domicílios alugados domicílios cedidos	O denominador destes indicadores é o total de domicílios particulares permanentes.

* assinalados em negrito os indicadores selecionados após a primeira fase da análise.

2B) Segundo Bloco Temático – Saneamento.

Assunto	Indicador (proporção de)	Observações
Água	domicílios com ligação interna domicílios com água de rede pública domicílios c/ ligação interna e rede pública domicílios c/ ligação interna e s/ rede pública domicílios s/ ligação interna e água da rede pública domicílios s/ ligação interna e s/ rede pública	Denominador: total de domicílios particulares permanentes. Indicadores combinando duas variáveis: fonte da água e canalização interna.
Esgoto	domicílios c/ ligação à rede urbana de esgotos domicílios com fossas domicílios com outro tipo de esgotamento domicílios sem qualquer tipo de esgotos	Denominador: total de domicílios particulares permanentes.
Destino do lixo	domicílios c/ coleta regular de lixo domicílios onde o lixo é enterrado domicílios onde o lixo é queimado domicílios onde o lixo não recebe tratamento	Denominador: total de domicílios particulares permanentes.

2C) Terceiro Bloco Temático – Escolaridade.

Assunto	Indicador (proporção de)	Observações
Alfabetização	mulheres acima de cinco anos alfabetizadas homens acima de cinco anos alfabetizados população acima de cinco anos alfabetizada chefes de família mulheres alfabetizadas chefes de família homens alfabetizados	Denominador: população acima de cinco anos.
Escolaridade	chefes de família c/ escolaridade < 1 ano chefes de família c/ 1 – 3 anos de escolaridade chefes de família c/ 4 – 7 anos de escolaridade chefes de família c/ 8 – 10 anos de escolaridade chefes de família c/ 11 – 14 anos de escolaridade chefes de família c/ 15 e + anos de escolaridade chefes de família c/ segundo grau ou mais	Denominador: chefes de domicílios particulares permanentes. Utilizou-se também outros recortes para escolaridade.

2D) Quarto Bloco Temático – Renda.

Assunto	Indicador (proporção de)	Observações
Renda do chefe	renda média nominal dos chefes de família chefes c/ renda até meio salário mínimo chefes c/ renda de 1/2 a 1 salário mínimo chefes c/ renda > 1 a 2 salários mínimos chefes c/ renda > 2 a 3 salários mínimos chefes c/ renda > 3 a 5 salários mínimos chefes c/ renda > 10 a 15 salários mínimos chefes c/ renda > 15 a 20 salários mínimos chefes c/ renda > 20 salários mínimos	Denominador: chefes em domicílios particulares permanentes. O indicador renda média nominal do chefe é calculado por setor censitário pelo IBGE.
Empregados	média de empregados por domicílio	Total de empregados do setor por total de domicílios particulares permanentes.
Cômodos	média de cômodos por domicílios	Indicador calculado pelo IBGE.

2E) Quinto Bloco Temático – Demografia.

Assunto	Indicador	Observações
Estrutura etária	mediana etária de mulheres mediana etária de homens	A mediana é uma medida menos sensível à ocorrência de valores extremos sendo preferível à média.
Chefe da família	proporção de chefes de família do sexo masculino	Em relação ao total de domicílios particulares permanentes.
Densidade	média de habitantes por domicílio	Indicador calculado pelo IBGE.

Tabela 3

Média de cada indicador segundo grupos de classificação.

Grupo	Freq.	CALUGA	CASAS	REDE	SEMESG	ALFABET	2GRAU	RNDCH
"A"	190	0,18329	0,13464	0,97737	0,00299	0,96397	0,83297	4,83466
"B"	1.116	0,25938	0,13867	0,98289	0,00227	0,97012	0,74714	1,84796
"C"	2.394	0,32246	0,34883	0,98886	0,00257	0,95148	0,47100	0,23293
"D"	6.240	0,22374	0,94178	0,95250	0,01865	0,86476	0,15866	-0,43645
"E"	1.416	0,12028	0,98036	0,16639	0,05290	0,77013	0,06890	-0,57562

te e usualmente descritos como cidades-dormitório, apresentam maior proporção de setores do tipo "D", pequena proporção de setores do tipo "C" e parcela variável de setores do tipo "E". Em função da menor parcela de setores tipo "E", pode-se subdividir este grupo, ficando Nilópolis e São João de Meriti como os mais homogêneos, onde mais de 90% dos setores estão classificados no grupo "D";

• Niterói, capital do antigo Estado do Rio de Janeiro, com perfil inteiramente diferente dos

demaís, muito mais parecido com algumas regiões do Município do Rio de Janeiro.

O Figura 3 apresenta o perfil das APs do Município do Rio de Janeiro. A proporção de setores no grupo "E" é consideravelmente menor em relação a figura anterior. É interessante observar que o tipo de setor "A" aparece em duas APs: na 2, onde se localizam os bairros da zona sul da cidade, e na 4, composta pela Barra da Tijuca e Jacarepaguá, bairros de ocupação recente para onde vem se deslocando a classe

Figura 2

Distribuição dos setores censitários segundo classificação multivariada, Municípios da Região Metropolitana do Rio de Janeiro, 1991.

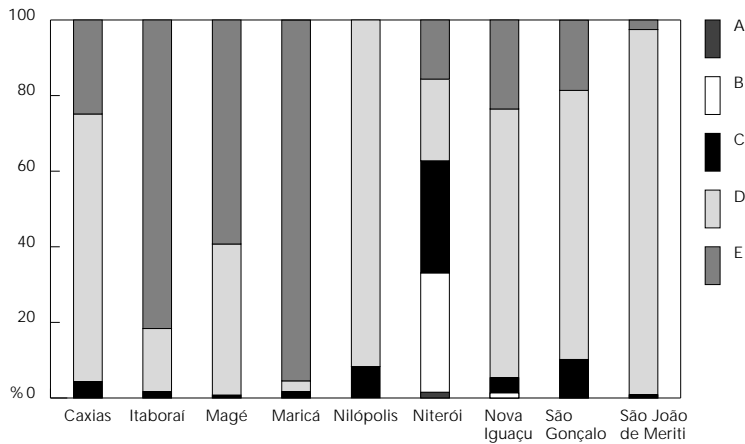
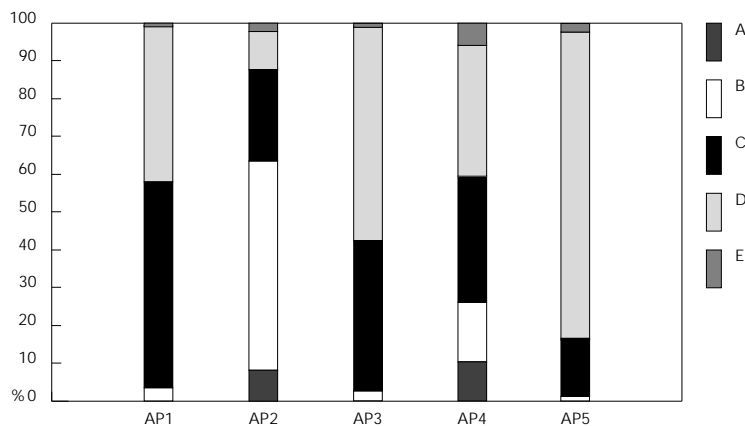


Figura 3

Distribuição dos setores censitários por grupo de classificação. Áreas programáticas do Município do Rio de Janeiro, 1991.



média e a população mais rica. Nestas APs também está presente o grupo "B", caracterizando-as como regiões efetivamente mais ricas. As APs 1 e 3, a primeira englobando as regiões do centro da cidade e a outra com os subúrbios tradicionais, apresentam perfil semelhante, com a maior parte dos setores censitários concentrados nos grupos "D" e "C". A AP 5 – Zona Oeste – é a área de expansão recente para as populações de baixa renda, ainda relativamente desprovida dos serviços urbanos.

Distribuição de indicadores selecionados

As figuras apresentadas apenas indicam a proporção de setores de cada tipo segundo região. Para estudar a distribuição dos indicadores, optou-se por utilizar *boxplots*, tipo de figura que sumariza medidas robustas de tendência central e dispersão, onde o retângulo representa o intervalo interquartil, a mediana apresenta-se ao meio na linha escura em destaque, a linha vertical pontilhada acima e abaixo dos quartis representa 1,5 vezes o intervalo entre a mediana e o respectivo quartil e os traços acima e abaixo destes valores representam *outliers*.

Em relação à proporção de residências com acesso à água da rede pública, o perfil dos municípios da Baixada Fluminense (Duque de Caxias, Nilópolis, Nova Iguaçu e São João de Meriti) e demais municípios da Região Metropolitana é substancialmente diferente das APs do Rio de Janeiro (Figura 4). Nestas, a mediana da proporção de residências com acesso à água da rede pública entre os setores censitários é acima de 95%, fazendo com que na figura fique ressaltada somente a presença de *outliers*. O acesso à água na Baixada e nos outros municípios da Região Metropolitana vai de 0 a 100%, sendo a mediana mais baixa (90%) nestes. É interessante observar que na AP 2 é pequeno o número de setores com valores nos extremos da distribuição, mais frequentes nas APs 4 e 5. Ou seja, localizar espacialmente estes setores e direcionar medidas concretas pode ser factível, uma vez que são poucos, e a medida teria provavelmente impacto na saúde. Evidentemente o nível de investimento necessário no restante da Região Metropolitana é muito maior.

A distribuição do indicador proporção de domicílios sem qualquer tipo de esgoto é inversa à primeira, e ainda mais concentrada. Neste caso, a detecção e localização dos valores extremos permitirá concentrar medidas focais de saneamento. Por exemplo, apenas na AP 4 e na Baixada Fluminense existem setores censitários onde 100% das residências lançam esgotos no peridomicílio, sem qualquer tratamento ou remoção (Figura 5).

Da mesma forma, a análise dos valores aberrantes da Figura 6, proporção de população acima de cinco anos alfabetizada, permite localizar áreas problemáticas. Note-se que na AP 2 a maior parte da população é alfabetizada, enquanto nas demais áreas esta distribuição não é tão concentrada.

A escolaridade do chefe da família apresenta grandes diferenças entre as APs (Figura 7) quanto à mediana, destacando-se a AP 2 como a mais alta proporção de escolaridade igual ou

maior do que o segundo grau. Os setores censitários com valores extremos em cada região indicam a desigualdade da distribuição do indicador.

As duas últimas figuras apresentam a estrutura residencial quanto à propriedade e à forma de ocupação do espaço, por meio dos indicadores proporção de domicílios alugados (Figura 8) e proporção de casas (Figura 9). O primeiro apresenta a distribuição mais suave de todas, indicando que há variedade grande neste indicador dentro de cada setor censitário e, em cada sub-região, a parcela de residências alugadas não está concentrada exclusivamente em algumas áreas. No caso das APs 1, 3 e 4, em metade dos setores, cerca de 80% dos domicílios são casas, mas em diversos setores pequenas parcelas de apartamentos são encontrados. Cabe lembrar que classificados como apartamento estão os sobrados ou pequenos prédios de duas ou três moradias. No caso da AP 2, mais de 75% dos setores têm menos de 10% de casas, enquanto na Baixada, em mais de 75% dos setores, a proporção de casas é acima de 85%.

Comentários finais

A análise realizada, utilizando as informações do censo demográfico por setor censitário, apresentou resultados interessantes em escalas diferentes. Em um nível mais macro, permitiu diferenciar facilmente cinco perfis sócio-econômicos na população da Região Metropolitana do Rio de Janeiro e apontou a grande diferença na composição dos municípios desta e APs da capital. Em escala maior, analisando por microárea, foi possível detectar os *outliers*, não apenas no sentido estatístico – localizados nos extremos da distribuição, mas no sentido também de marginal, à margem, excluído. Ou seja, embora no Município do Rio de Janeiro mais de 95% da população tenham acesso à rede de água, os setores censitários onde predominam casas sem este recurso são, evidentemente, as áreas mais carentes. Direcionar recursos para estas áreas terá, certamente, impacto no perfil de saúde da população.

Outras variáveis poderiam ser escolhidas, visando descrever a área segundo diferentes enfoques. A opção apresentada foi orientada pela perspectiva de analisar simultaneamente o conjunto, e, por isso, selecionaram-se indicadores menos correlacionados. Em alguns casos, quando duas variáveis apresentavam posicionamento semelhante na análise de componentes principais, optou-se por aquela que permitisse destacar situações excepcionais. É o

Figura 4

Boxplot da proporção de domicílios ligados à rede de água por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.

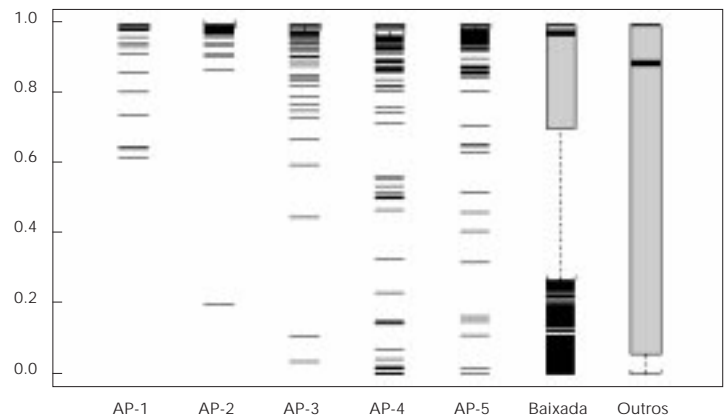


Figura 5

Boxplot da proporção de domicílios sem qualquer tipo de esgoto por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.

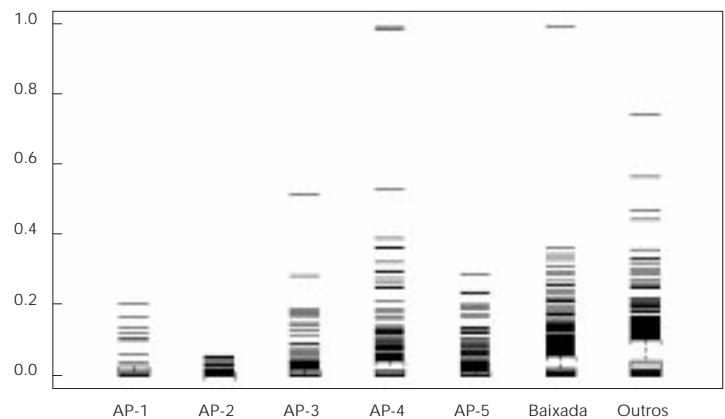


Figura 6

Boxplot da proporção de população alfabetizada por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.

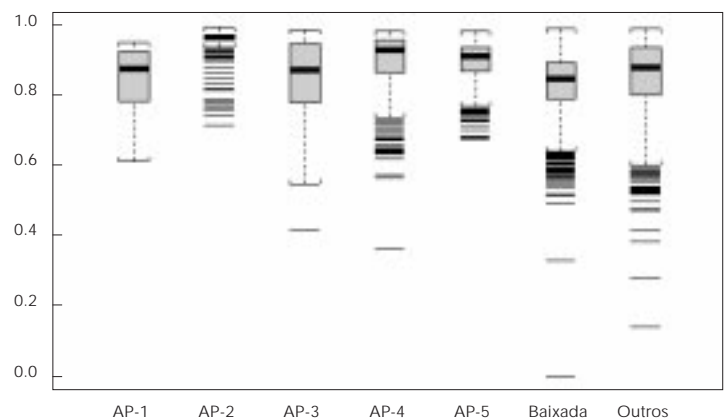


Figura 7

Boxplot da proporção de domicílios cujo chefe completou o 2º grau por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.

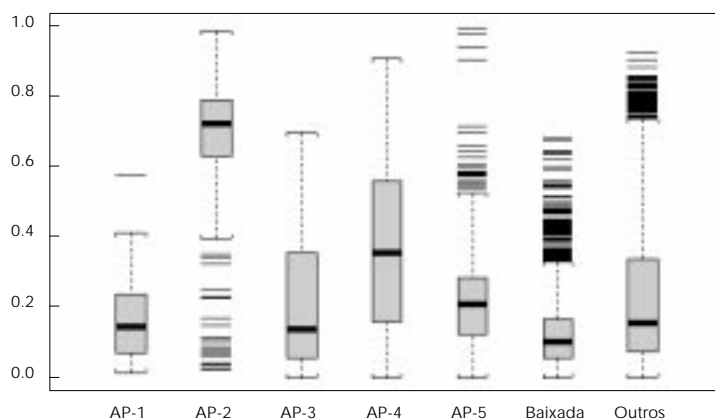


Figura 8

Boxplot da proporção de domicílios alugados por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.

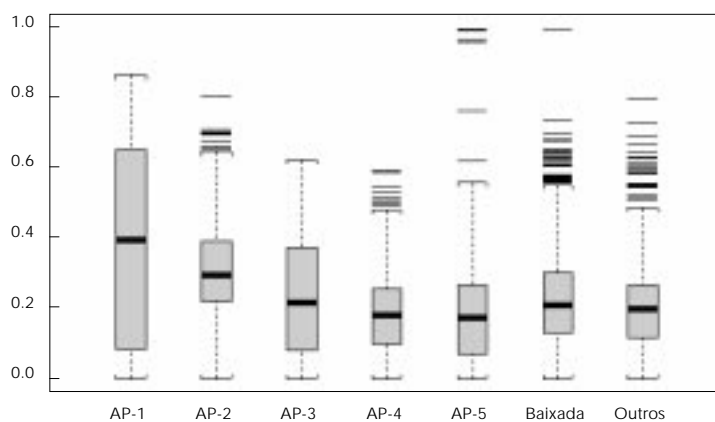
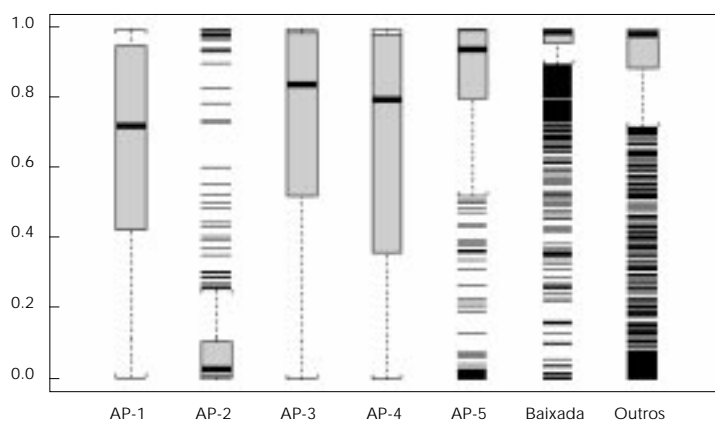


Figura 9

Boxplot da proporção de domicílios unifamiliares (casas) por Área Programática e Sub-regiões da Região Metropolitana do Rio de Janeiro, 1991.



caso da variável proporção de domicílios sem qualquer tipo de esgoto, que caracteriza uma realidade pouco freqüente, porém de extrema gravidade. Outros indicadores disponíveis, como, por exemplo, a proporção de domicílios que lançam esgotos em valas negras, muito mais freqüentes, também poderiam ter sido utilizados, porém a variável escolhida detecta a situação 'marginal'.

O padrão espacial dos grupamentos sócio-econômicos existentes na região estudada apresenta características de mosaico, muito comum no Rio de Janeiro. Os setores censitários mais ricos situam-se ao lado de setores excluídos. Ou seja, mesmo as áreas mais ricas, de mais alta renda, de moradores, por exemplo, da AP 2, apresentam setores censitários distribuídos por todas as classes, inclusive a "E". São mais homogêneas as regiões mais pobres, como os municípios de Itaguaí e Magé. Entretanto, seja morando no bairro rico, seja morando no subúrbio de baixa renda, grupos populacionais caracterizados no extremo inferior da escala aqui criada apresentam condições de vida igualmente precárias.

A classificação dos setores censitários foi feita para toda a Região Metropolitana do Rio de Janeiro, de forma a estabelecer comparação direta entre municípios e APs da capital. A mesma técnica, aplicada para regiões menores, permitiria maior discriminação interna (Carvalho et al., 1996). O mesmo tipo de técnica, aplicado em outras escalas geográficas - bairro e municípios -, permitiu também a criação de perfis sócio-econômicos claramente definidos, corroborando a utilização deste tipo de técnica para estas variáveis. É interessante observar que foram poucas as alterações na seleção de variáveis, ainda que eventualmente tenha sido incluído ou retirado algum indicador (Cruz, 1996; D'Orsi, 1996).

Evidentemente os indicadores do censo apresentam diversos problemas, seja quanto à inexistência de informações relevantes, ou quanto à qualidade do dado. Entretanto, o método proposto é relativamente simples, parte de informações de fácil acesso e ampla cobertura, e permite uma primeira aproximação na criação de índices compostos de avaliação de qualidade de vida. Além disso, contribui para apontar a desigualdade na distribuição dos recursos, seja mapeando grandes regiões de pobreza, seja identificando microáreas de exclusão em meio à relativa riqueza.

Referências

- CARVALHO, M. S.; CRUZ, O. G. & NOBRE, F. F., 1996. Spatial partition using multivariate cluster analysis and contiguity algorithm: application to Rio de Janeiro, Brazil. *Statistics in Medicine*, 15:1.885-1.894.
- CASTELLANOS, P. L., 1990. Sobre el concepto de salud-enfermedad. Descripción y explicación de la situación de salud. *Boletín Epidemiológico*, 10:1-7.
- CORTINOVIS, I.; VELLA, V. & NDIKU, J., 1993. Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Social Science and Medicine*, 36:1.087-1.097.
- CRUZ, O. G., 1996. *Homicídios no Estado do Rio de Janeiro: Análise da Distribuição Espacial e sua Evolução*. Dissertação de Mestrado, São Paulo: Faculdade de Saúde Pública, Universidade de São Paulo.
- D'ÓRSI, E., 1996. *Perfil de Nascimentos e Condições Sócio-Econômicas no Município do Rio de Janeiro: Uma Análise Espacial*. Dissertação de Mestrado, Rio de Janeiro: Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz.
- FIBGE (Fundação Instituto Brasileiro de Geografia e Estatística), 1994. *Censo Demográfico de 1991*. Rio de Janeiro: FIBGE.
- GOLDBERG, M.; DAB, W.; CHAPERON, J.; FULVIER, R. & GRÉMY, F., 1979a. Indicateurs de santé et "sanométrie": les aspects conceptuels des recherches récentes sur la mesure de l'état de santé – première partie. *Revue d'Épidémiologie et de Santé Publique*, 27:51-68.
- GOLDBERG, M.; DAB, W.; CHAPERON, J.; FULVIER, R. & GRÉMY, F., 1979b. Indicateurs de santé et "sanométrie": les aspects conceptuels des recherches récentes sur la mesure de l'état de santé – seconde partie. *Revue d'Épidémiologie et de Santé Publique*, 27:133-152.
- HARTIGAN, J. A., 1975. *Clustering Algorithms*. New York: John Wiley & Sons Inc.
- HAYES, M. V., 1992. On the epistemology of risk: language, logic and social science. *Social Science and Medicine*, 35:401-407.
- LEBART, L.; MORINEAU, A. & FÉNELON, J. P., 1986. *Traitement de Données Statistiques – Méthodes et Programmes*. 2^{ème} éd., Paris: Dunod.
- ROSE, G., 1985. Individuos enfermos y poblaciones enfermas. *Boletín Epidemiológico*, 6:1-8.
- VERHASSELT, Y. & MANSOURIAN, B., 1991. Método para la clasificación de los países de acuerdo con sus indicadores de salud. *Boletín de la Oficina Sanitaria Panamericana*, 110:319-323.