

ANÁLISE DE COMPONENTES PRINCIPAIS PARA CONSTRUÇÃO DE INDICADORES SOCIAIS

Dalson Britto FIGUEIREDO FILHO¹
Ranulfo PARANHOS²
Enivaldo Carvalho da ROCHA¹
José Alexandre Da SILVA JR²
Romero G. MAIA³

- RESUMO: O que são, para que servem e como são construídos indicadores sociais? O principal objetivo desse trabalho é responder essas questões. O foco repousa sobre a compreensão intuitiva dos principais conceitos, propriedades e construção de indicadores sociais. Nosso público alvo são estudantes de graduação, pós-graduação e pesquisadores em geral com limitado conhecimento de cálculo e álgebra. Metodologicamente, demonstramos como a técnica de análise de componentes principais pode ser utilizada para construir indicadores sociais. Em particular, replicamos os dados do Atlas do Desenvolvimento Humano para construir o Índice de Qualidade de Moradia (IQM) por Estado. Os resultados sugerem que: (1) São Paulo, Distrito Federal e Santa Catarina apresentam as melhores condições de qualidade de moradia, enquanto Acre, Piauí e Maranhão apresentam as piores condições; (2) existe uma correlação negativa e estatisticamente significativa (-0,703; p-valor<0,000) entre o IQM e a taxa de mortalidade; e (3) observa-se uma correlação positiva e estatisticamente significativa (0,804; p-valor<0,000) entre o Índice de Qualidade da Moradia e a renda *per capita*.
- PALAVRAS-CHAVES: Indicadores sociais; métodos quantitativos; análise de componentes principais; Índice de Qualidade de Moradia; replicabilidade.

1 Introdução

O principal objetivo desse trabalho é apresentar uma introdução aos indicadores sociais. O foco repousa sobre a compreensão intuitiva dos principais conceitos, propriedades e construção de indicadores sociais. Nosso público alvo são estudantes de graduação, pós-graduação e pesquisadores em geral com limitado conhecimento de cálculo e álgebra. Metodologicamente, utilizamos a técnica de análise de componentes

¹ Universidade Federal de Pernambuco - UFPE, Departamento de Ciência Política, CEP: 50670-901, Recife, PE, Brasil. E-mail: dalsonbritto@yahoo.com.br / enivaldocrocha@gmail.com.

² Universidade Federal de Alagoas - UFAL, Instituto de Ciências Sociais, CEP: 57072-900, Alagoas, Brasil. E-mail: ranulfoparanhos@me.com / jasjunior2007@yahoo.com.br.

³ Instituto Brasileiro de Geografia e Estatística - IBGE, Analista de Informações Estatísticas, CEP: 57072-900, Recife, PE, Brasil. E-mail: romeromaia@gmail.com.

principais para demonstrar o passo a passo de construção de indicadores sociais. Em particular, replicamos os dados do Atlas do Desenvolvimento Humano para construir o Índice de Qualidade de Moradia (IQM) por unidade da federação. O índice foi construído a partir da redução, via análise de componentes principais, de cinco variáveis observadas:

1. Porcentagem de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo (LIXO);
2. Porcentagem de pessoas que vivem em domicílios com energia elétrica (ENERGIA);
3. Porcentagem de pessoas que vivem em domicílios com energia elétrica e geladeira (ENERGIA/GELADEIRA);
4. Porcentagem de pessoas que vivem em domicílios com água encanada (ÁGUA) e
5. Porcentagem de pessoas que vivem em domicílios com água encanada e banheiro (ÁGUA/BANHEIRO).

O índice é padronizado, ou seja, tem média zero e desvio padrão igual a um. Quanto maior, melhor é a qualidade da moradia.

Nossa motivação central é facilitar a compreensão do conceito, aplicação e construção de indicadores sociais não só entre a comunidade acadêmica, mas também entre os formuladores de políticas públicas. Para Henriot (1970), os indicadores sociais procuram solucionar uma questão antiga: como utilizar a informação disponível sobre a realidade social para otimizar o desenho e a implementação de políticas públicas? Dessa forma, uma primeira função dos indicadores sociais é descrever a realidade a partir de dados objetivos e comparáveis e/ou critérios considerados socialmente relevantes.

Estimar em que medida as ações governamentais produzem os seus efeitos esperados é um dos principais desafios enfrentados por estudiosos do assunto e formuladores de políticas públicas. Esse argumento ganha mais força ao se considerar a realidade de países em que as instituições não dispõem de tecnologias para a coleta, o processamento e a divulgação sistemática de informações. Esse impedimento gera uma série de efeitos negativos. Primeiro, a falta de informação limita a possibilidade de aprimoramento da legislação. Segundo, reduz a transparência das ações públicas, violando o princípio da publicidade. Terceiro, a falta de informação dificulta a realização de estudos em perspectiva comparada, inibindo a difusão de práticas institucionais eficientes. Por fim, restringe a capacidade do poder público identificar quais são as demandas mais latentes e solucionar os principais problemas sociais. Em conjunto, esses obstáculos comprometem tanto a eficiência quanto a efetividade das políticas públicas. É nesse sentido que a construção e a divulgação de indicadores sociais tem um papel fundamental no planejamento e execução das ações públicas.

O artigo está dividido da seguinte forma: a próxima seção define o que são indicadores sociais e discute a sua importância tanto na pesquisa acadêmica quanto na formulação e avaliação de políticas públicas. Depois disso, a seção 2.2 apresenta algumas das propriedades desejáveis dos indicadores sociais e fornece exemplos de repositórios onde diferentes indicadores podem ser acessados. A parte 2.3 ilustra o passo a passo que deve ser seguido para construir um indicador a partir da análise de componentes principais e apresenta a variação do Índice de Qualidade de Moradia. A conclusão sumariza nossas principais recomendações.

2 O que são indicadores sociais?

O termo *indicadores sociais* surgiu no início da década de 1960 no contexto da corrida espacial norte-americana (LAND, 1983). Em termos acadêmicos, é importante destacar o papel de *Ogburn* e seus colaboradores na Universidade de Chicago como entusiastas intelectuais e institucionais de diferentes esforços no sentido de produzir indicadores sociais que revelassem os padrões de mudança social da época. De acordo com Land, Michalos e Sirgy (2012), as ideias de Ogburn sobre a mensuração de fenômenos sociais influenciou vários de seus estudantes, entre eles Albert D. Biderman, Otis Dudley Duncan, Albert J. Reiss e Eleonor Bernert Sheldon que desempenharam papel fundamental no desenvolvimento desse campo de pesquisa nas décadas de 1960 e 1970.

O aparecimento e desenvolvimento dos indicadores sociais está intrinsecamente ligado à consolidação das atividades de planejamento do setor público ao longo do século XX. Embora se possa citar algumas contribuições importantes para a construção de um marco conceitual sobre os Indicadores Sociais nos anos 20 e 30, o desenvolvimento da área é recente, tendo ganhando corpo científico em meados dos anos 60 no bojo das tentativas de organização de sistemas mais abrangentes de acompanhamento das transformações sociais e aferição do impacto das políticas sociais nas sociedades desenvolvidas e subdesenvolvidas (JANNUZZI, 2002: 1).

Mas o que é um indicador social afinal? Land (1971) e Wilcox e Brooks (1971) definem indicadores sociais como componentes do sistema social que descrevem o funcionamento do próprio sistema (HENRIOT, 1970). Para Jannuzzi (2005), “no campo aplicado das políticas públicas, os indicadores sociais são medidas usadas para permitir a operacionalização de um conceito abstrato ou de uma demanda de interesse programático” (JANNUZZI, 2005: 138). Mais adiante, Jannuzzi (2005) afirma que “os indicadores apontam, indicam, aproximam, traduzem em termos operacionais as dimensões sociais de interesse definidas a partir de escolhas teóricas ou políticas realizadas anteriormente” (JANNUZZI, 2005: 138). Nesse artigo adotamos a definição proposta por Jannuzzi (2005).

E qual é a relação entre indicadores sociais e gestão governamental? Os indicadores sociais têm um papel fundamental no desenho, na implementação e na avaliação de políticas públicas. Os indicadores informam ao gestor a respeito da quantidade de alunos por escola/sala de aula, do número de detentos por presídio/cela, do número de homicídios por estado/cidade/bairro, do montante de pessoas desempregadas, etc. Ou seja, sempre que existir um interesse programático em uma área específica da atividade governamental, haverá a necessidade de utilizar indicadores sociais já que eles podem ser utilizados para abordar diferentes temas. Para Jannuzzi (2005), “os indicadores guardam, pois, relação direta com o objetivo programático original, na forma operacionalizada pelas ações e viabilizada pelos dados administrativos e pelas estatísticas públicas disponíveis” (JANNUZZI, 2005: 139).

Logisticamente, é impensável cogitar qualquer intervenção estatal sem conhecer detalhadamente a realidade social que se deseja transformar. Uma política de segurança pública dificilmente produzirá os seus resultados esperados se o gestor não souber qual é a modalidade criminal mais recorrente e onde ela ocorre com mais intensidade. Similarmente, uma política de transporte tende ao fracasso quando o gestor desconhece o tamanho da frota e localidades de maior fluxo. No entanto, a simples compilação de dados

não garante a eficiência, eficácia e efetividade das políticas. Tão importante quanto coletar e processar os dados de forma sistemática é assegurar as ferramentas mais adequadas à sua análise. Por exemplo, as técnicas multivariadas, quando corretamente utilizadas, podem fornecer estimativas eficientes e elevar o grau de profundidade analítica das investigações. No entanto, a maior parte dos relatórios técnicos são majoritariamente descritivos e/ou se limitam a utilizar técnicas básicas de estatística inferencial.

Por se tratar de um trabalho introdutório, iremos cobrir apenas conceitos básicos e aplicações relativamente simples da construção de indicadores sociais. Para os interessados em saber mais sobre esse tema sugerimos cobrir as referências bibliográficas. Em particular, sugerimos acompanhar a produção do professor Paulo de Martino Jannuzzi. Para uma compilação comentada da bibliografia disponível sobre indicadores sociais ver Wilcox *et al.* (1972). Para uma contextualização histórica do surgimento e desenvolvimento dos indicadores sociais ver Land (1983), Land, Michalos e Sirgy (2012) e Soligo (2012).

2.1 Propriedades desejáveis dos indicadores sociais

A literatura especializada aponta a confiabilidade e a validade como termos essenciais da linguagem básica da mensuração (ZELLER e CAMINES, 1982). A definição clássica de Nunnally (1967) postula que a confiabilidade diz respeito à repetibilidade da mensuração, ou seja, o grau em que medidas repetidas sobre as mesmas unidades produzem resultados similares. Uma forma intuitiva de entender o conceito de confiabilidade é imaginar uma balança. Se a cada vez que o mesmo indivíduo subir na balança ela apontar valores diferentes, conclui-se que o instrumento não é confiável. Isso quer dizer que quanto maior a confiabilidade da medida, menor a quantidade de erro aleatório no processo de mensuração. Para uma introdução à mensuração em Ciências Sociais ver Zeller e Carmines (1980). Para uma abordagem mais avançada ver Blalock (1984).

Por sua vez, a validade refere-se ao grau de correspondência entre o que se mediu e o que se queria medir (ZELLER e CARMINES, 1980; EVERITT e SKRONDAL, 2010). Nas palavras de Jannuzzi (2005), “validade é outro critério fundamental na escolha de indicadores, pois é desejável que se disponha de medidas tão próximas quanto possível do conceito abstrato ou da demanda política que lhes deram origem” (JANNUZZI, 2005: 139).

Depois de examinar os conceitos de confiabilidade e validade, o próximo passo é analisar outras características importantes dos indicadores sociais. Jannuzzi (2005) sugere doze propriedades desejáveis na construção dessas medidas. O Quadro 1 reproduz essas informações.

Quanto mais propriedades o indicador apresentar, tanto melhor é a medida. Comparativamente, o indicador 1 (*escore* 4) prescinde de várias características desejáveis como por exemplo a transparência metodológica na construção. Para Jannuzzi (2005), “a boa prática de pesquisa social recomenda que os procedimentos de construção dos indicadores sejam claros e transparentes, que as decisões metodológicas sejam justificadas, que as escolhas subjetivas – invariavelmente frequentes – sejam explicitadas de forma objetiva” (JANNUZZI, 2005: 141). Essa é a essência da replicabilidade científica, ou seja, o processo pelo qual novas análises podem ser realizadas a partir de um banco de dados já existente com o objetivo de aprimorar os resultados de pesquisa.

Quadro 1 - Propriedades desejáveis dos indicadores sociais

Propriedades	Indicador 1	Indicador 2	Indicador 3
Relevância para agenda política	+	+	+
Validade de representação do conceito	+	+	+
Confiabilidade da medida	+	+	+
Cobertura populacional	+	+	+
Sensibilidade às ações previstas		+	+
Especificidade ao programa		+	+
Transparência metodológica na construção		+	+
Comunicabilidade ao público		+	+
Factibilidade operacional para sua obtenção			+
Periodicidade de sua atualização			+
Desagregabilidade populacional e territorial			+
Comparabilidade da série histórica			+
Total	4	8	12

Fonte: Elaboração dos autores a partir de Jannuzzi (2005).

Com efeito, se não é possível entender como o indicador foi construído, fica impossível utilizá-lo tanto em pesquisas acadêmicas quanto na formulação e avaliação de políticas públicas. Por exemplo, quando um indicador não é periodicamente atualizado, a sua comparação no tempo (série histórica) fica comprometida. Por sua vez, o indicador 2 (*escore* 8) é superior ao indicador 1, no entanto, ainda carece de propriedades importantes. Por exemplo, ele não é passível de desagregação em outros níveis de análise territorial nem permite a comparação entre diferentes segmentos populacionais. A experiência prática de pesquisa sugere que dificilmente o pesquisador vai encontrar indicadores que apresentem todas as propriedades desejáveis elencadas por Jannuzzi (2005). Entretanto, é extremamente importante que o pesquisador considere essas propriedades na fase inicial do desenho de sua pesquisa e reporte, detalhadamente, como cada indicador foi construído. Para Jannuzzi (2005), “nem sempre o indicador de maior validade é o mais confiável, nem sempre o mais confiável é o mais sensível; nem sempre o mais sensível é o mais específico; enfim, nem sempre o indicador que reúne todas essas qualidades é passível de ser obtido na escala territorial e na periodicidade requerida” (JANNUZZI, 2005: 142).

2.2. Como classificar os indicadores sociais?

Existem diferentes critérios para classificar os indicadores sociais (CARLEY, 1985). Por exemplo, Land (1983) propõe três classes de indicadores: (1) *normative welfare indicators*, (2) *satisfaction indicators* e (3) *descriptive social indicators*. A perspectiva normativa assume que a própria definição de indicador deve ser orientada em termos de objetivos concretos de planejamento de políticas.

Por sua vez, os indicadores de satisfação têm sua origem no livro *The Human Meaning of Social Change*, de Campbell e Converse (1972). Parte-se do pressuposto de que a avaliação das atitudes, expectativas, sentimentos, aspirações e valores são importantes para compreender os processos de mudança social (LAND, 1983).

Por fim, a perspectiva descritiva enfatiza o papel dos indicadores como estimativas das condições sociais que podem variar no tempo e no espaço. Eles podem ser taxas, razões, índices, escalas, etc. Além disso, esses indicadores podem ser objetivos ou subjetivos, a depender dos propósitos do pesquisador, diferente dos indicadores de satisfação que são, necessariamente, subjetivos.

De acordo com Jannuzzi (2005), a classificação mais recorrente é a divisão dos indicadores por área temática. Por exemplo, o IPEADATA agrupa os dados a partir de três principais temas: (1) macroeconômico; (2) regional e (3) social. Dentro de cada tema, tem-se ainda uma divisão mais específica de modo que a partir de macroeconomia é possível acessar indicadores referentes a balanço de pagamentos, câmbio, comércio exterior, consumo e vendas, contas nacionais, etc. O tema regional existem indicadores relativos a agropecuária, eleições, emprego, moeda e crédito, estoque de capital, etc. Por fim, no tema social pode-se ter acesso a outros subtemas e seus respectivos indicadores como assistência social, demografia, desenvolvimento humano, educação, habitação, etc. O Quadro 2 ilustra diferentes indicadores sociais a partir do banco de dados do IPEA.

Quadro 2 - Temas, subtemas e indicadores do IPEADATA⁴

Tema	Subtema	Indicadores
Macroeconômico	Balanço de pagamentos	(1) Total erros e omissões (anual)
		(2) Total de captação de recursos externos – empréstimos em moeda – (mensal)
		(3) Total de captação de recursos externos – financiamentos (mensal)
Regional	População	(1) Número de pessoas de 0 a 4 anos (homens)
		(2) Número de pessoas de 10 a 14 (mulheres)
		(3) População residente (urbana)
Social	Habitação	(1) Número de domicílios com água encanada
		(2) Número de domicílios com iluminação elétrica
		(3) Número de domicílios com instalações sanitárias

Fonte: Elaboração dos autores

Jannuzzi (2005) afirma que outra classificação usual é dividir os indicadores sociais entre objetivos e subjetivos. Em suas palavras, “os indicadores objetivos referem-se a ocorrências concretas ou a entes empíricos da realidade social, construídos a partir de estatísticas públicas disponíveis (...) os indicadores subjetivos, por outro lado, correspondem a medidas construídas a partir da avaliação dos indivíduos ou especialistas com relação a diferentes aspectos da realidade” (JANNUZZI, 2005: 143). Outra possibilidade analítica de classificação diz respeito a distinção dos indicadores entre insumos, processo, resultado e impacto. Por exemplo, ao se considerar a base de dados disponibilizada pelo *Quality of Government Institute*, é possível identificar a presença de

⁴Os dados estão disponíveis em: <http://www.ipeadata.gov.br/>

indicadores objetivos e subjetivos na mesma base⁵. O Quadro 3 sumariza alguns exemplos de indicadores dessa base.

Quadro 3 – Indicadores por tipo a partir do *Quality of Government Institute*⁶

Tipo	Variável	Descrição
Objetivo	wdi_co2	<i>Carbon Dioxide Emissions</i> (toneladas per capita). Quantidade de dióxido de carbono emitido per capita.
Subjetivo	wvs_a008	<i>Feeling of Happiness</i> . Assume valores entre 1 e 4. Quanto maior, mais feliz ⁷ .

Fonte: Elaboração dos autores.

A quantidade de dióxido de carbono emitido é um indicador quantitativo objetivo. Por ser uma medida contínua, o pesquisador pode utilizar técnicas paramétricas para analisar a sua distribuição e como ele se relaciona com outras variáveis. Por exemplo, é possível calcular a média de emissão por continente, o desvio padrão, variância, etc. Além disso, ela pode ser utilizada como variável dependente ou independente em modelos explicativos. Outros exemplos de indicadores objetivos são a taxa de desemprego, taxa de analfabetismo, índice de Gini, *déficit* de vagas no sistema prisional, etc. A medida de felicidade, por outro lado, é um indicador subjetivo. O pesquisador deve analisar a distribuição de frequência de cada categoria e/ou analisar como esse indicador se relaciona com outras variáveis utilizando técnicas apropriadas ao seu nível de mensuração. Outros exemplos de indicadores subjetivos são o nível de satisfação com a administração do governo, percepção sobre a incidência da corrupção, opinião sobre a qualidade da saúde pública, etc.

A literatura também distingue entre medidas analíticas e medidas sintéticas. As medidas analíticas são geralmente utilizadas para medir dimensões específicas da realidade. Por exemplo, a taxa de homicídios por 100 mil habitantes mensura uma dimensão específica da violência. Uma localidade pode ter uma alta taxa de homicídios mas apresentar uma taxa baixa de roubos e furtos. Os indicadores sínteses, por sua vez, agrupam diferentes dimensões em uma mesma medida. O Índice de Desenvolvimento Humano (IDH) é o exemplo mais amplamente difundido de um índice composto. Ele é calculado a partir da média entre um indicador de educação, um indicador de saúde e um indicador de renda. É nesse sentido que o IDH sintetiza em uma única estimativa informações referentes a três diferentes áreas de interesse governamental. Outros exemplos de indicadores sintéticos são o Índice de Desenvolvimento Municipal (IDH-M) e o Índice de Condições de Vida Municipal (ICV), formulados pela Fundação João Pinheiro (MG), o Índice de Qualidade Municipal (IQM - verde) elaborado pela Fundação CIDE (RJ), o Índice Paulista de Vulnerabilidade Social desenvolvido pela Fundação SEADE/SP, o Índice de Qualidade Institucional Municipal elaborado pelo Ministério do Planejamento, entre tantos outros.

⁵Bancos de dados disponíveis em: <http://www.qog.pol.gu.se/data/datadownloads/>

⁶Livro de códigos disponível em:

http://www.qog.pol.gu.se/digitalAssets/1373/1373416_qog_basic_codebook_120608.pdf

⁷ (1) *very happy*; (2) *quite happy*; (3) *not very happy* e (4) *not at all happy*.

2.3 Como construir indicadores sociais?

O primeiro passo para construir um indicador é definir que tipo de indicador se deseja criar. Existem diferentes maneiras de transformar um conceito abstrato em um indicador empiricamente observável. Um dos procedimentos mais comuns é a elaboração de índices e escalas. Para Babbie (2005), “índices e escalas (especialmente escalas) são dispositivos de redução de dados, as várias respostas de um respondente podem ser resumidas num único escore, e mesmo assim os detalhes específicos daquelas respostas serem mantidos quase que na totalidade” (BABBIE, 2005: 214).

Segundo Babbie (2005), um índice é construído a partir da soma simples dos escores atribuídos a respostas específicas aos itens individuais. Já a escala é construída pela atribuição de escores a padrões de resposta entre os vários itens que formam a escala. Uma escala difere do índice por possuir uma estrutura de intensidade entre os itens individuais que a compõem. Por exemplo, suponha que o pesquisador deseja identificar qual é o melhor time de futebol do Brasil. Uma das opções é construir uma escala para medir a qualidade das equipes.

- () Campeão do Mundo
- () Campeão da Libertadores
- () Campeão do Brasileirão
- () Campeão da Copa do Brasil

Observe que existe uma estrutura de intensidade entre os itens individuais. Um time pode ser campeão da Copa do Brasil ou do Brasileirão mas nunca vencer a Libertadores e/ou a Copa dos Campeões⁸. No entanto, para disputar a Libertadores, o time deve, necessariamente, ou ter vencido a Libertadores do ano anterior ou ter vencido o Campeonato Brasileiro, ou ter ficado em segundo ou terceiro lugar, ou ter vencido a Copa do Brasil ou ter ganhado a Copa Sul Americana. Ser campeão do mundo é muito mais importante do que ser campeão da Copa do Brasil, por exemplo. Se o pesquisador atribuir o mesmo peso a cada item, a sua medida será necessariamente inválida, já que não consegue capturar a estrutura de intensidade presente na realidade.

Para Babbie (2005), “a medição de variáveis, frequentemente, é tarefa difícil. Normalmente, é impossível chegar a uma medida totalmente inequívoca e completamente aceitável de qualquer variável. Mesmo assim, os pesquisadores não desistem de tentar criar medidas cada vez melhores e mais úteis” (BABBIE, 2005: 213). A Figura 1 ilustra a lógica de criação de um indicador a partir da técnica de análise de componentes principais. O objetivo é demonstrar o padrão de correlação entre as variáveis originais e os componentes extraídos.

⁸É o exemplo do egrégio Sport Clube do Recife que venceu o Brasileirão de 1987 e a Copa do Brasil de 2008.

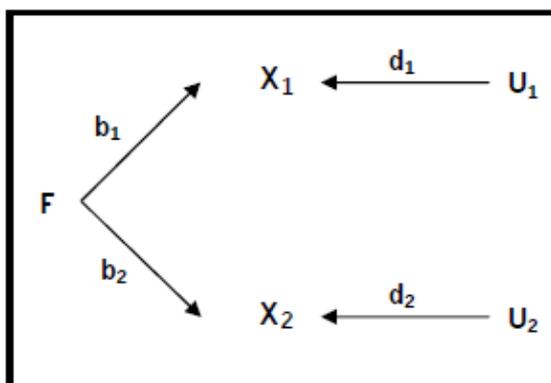


Figura 1 - Modelo das vias para duas variáveis, modelo de um fator comum
 Fonte: Asher, 1983.

Algebricamente,

$$F_j = \sum_{i=1}^i \omega_{ji} X_i$$

Onde F_j representa os fatores comuns, ω_{ji} são os coeficientes dos escores fatoriais e X_i são as variáveis originais observadas. As cargas fatoriais medem o grau de correlação entre as variáveis originais e os fatores/componentes. O fator é o resultado da combinação linear entre as variáveis observadas e pode ser utilizado para explicar/representar as variáveis originais. O autovalor (*eigenvalue*) representa o poder explicativo do fator/componente em relação à variância das variáveis originais. Apenas são extraídos os fatores/componentes com autovalor acima de um, já que eles explicam mais do que a variável original. No que concerne ao padrão de correlação entre as variáveis, a matriz de correlações deve exibir a maior parte dos coeficientes com valor acima de 0,30. O teste de Kaiser-Meyer-Olkin (KMO) varia entre 0 e 1. Quanto mais perto de 1, tanto melhor. Pallant (2007) sugere 0,6 como um limite razoável.

Fundamentalmente, o que o modelo de análise de componentes principais vai fazer é estimar em que medida a correlação entre as variáveis observadas podem ser agrupadas em um número menor de variáveis latentes (componentes).

O próximo passo é ilustrar a construção de um indicador a partir de dados reais. Utilizamos os dados do Atlas de Desenvolvimento Humano, tendo como referência o ano de 2000⁹. Existem diferentes indicadores por tema (educação, renda, população, etc.) e nível de agregação (municipal, estadual, regional, nacional). Optamos por utilizar os dados referentes à habitação para construir o Índice de Qualidade de Moradia (IQM). O índice foi construído a partir da redução, via análise de componentes principais, de cinco variáveis observadas: (1) % das pessoas que vivem em domicílios urbanos com serviço de coleta de lixo (LIXO); (2) % das pessoas que vivem em domicílios com energia elétrica

⁹Ver://www.pnud.org.br/IDH/Atlas2003.aspx?indiceAccordion=1&li=li_Atlas2003

(ENERGIA); (3) % das pessoas que vivem em domicílios com energia elétrica e geladeira (ENERGIA/GELADEIRA); (4) % das pessoas que vivem em domicílios com água encanada (ÁGUA) e (5) % das pessoas que vivem em domicílios com água encanada e banheiro (ÁGUA/BANHEIRO). O índice é padronizado, ou seja, tem média zero e desvio padrão igual a um. Quanto maior, melhor é a qualidade da moradia. A comparação será realizada a partir dos estados brasileiros. As Tabelas 1 e 2 resumem as estatísticas de interesse.

Tabela 1 - Matriz de correlação das variáveis observadas

	LIXO	ENERGIA	EN./GEL.	ÁGUA	ÁGUA/BANHEIRO
LIXO	1	0,848	0,789	0,897	0,889
ENERGIA		1	0,840	0,893	0,867
EN./GEL.			1	0,912	0,887
ÁGUA				1	0,992
ÁGUA/BANHEIRO					1

EN./GEL. - ENERGIA/GELADEIRA; N = 27 (Unidades da Federação)

Fonte: Elaboração dos autores a partir do PNUD (2000).

Tabela 2 - Comunalidades

LIXO	0,864
ENERGIA	0,873
ENERGIA/GELADEIRA	0,866
ÁGUA	0,975
ÁGUA/BANHEIRO	0,951

Fonte: Elaboração dos autores

Sempre que o pesquisador se deparar com muitas variáveis correlacionadas entre si, ele pode optar por alguma técnica de redução de dados¹⁰. Quanto maior o grau de correlação recíproca, tanto mais adequadas serão as técnicas de redução de dados.

De modo geral, existem três estágios que devem ser seguidos para empregar a técnica de análise de componentes principais ou análise fatorial para reduzir diferentes variáveis em um número menor de componentes/fatores. Para Garson (2009:1), "a análise de componentes principais é em geral preferida para fins de redução de dados (traduzindo o espaço das variáveis num espaço ótimo de fatores), enquanto a análise fatorial é em geral preferida quando o objetivo da pesquisa é detectar a estrutura dos dados ou a

¹⁰ Para uma introdução intuitiva à lógica da análise fatorial em português ver Figueiredo Filho e Silva Júnior (2010).

modelagem causal". De acordo com Hair *et al.* (2006), na maioria dos casos tanto a ACP, quanto a AF, chegam aos mesmos resultados quando o número de variáveis supera 30 ou se as comunalidades excedem 0,60 para a maior parte das variáveis. O primeiro estágio diz respeito à adequabilidade da base de dados. O pesquisador deve observar o nível de mensuração das variáveis, o tamanho da amostra, a razão entre o número de casos, a quantidade de variáveis e o padrão de correlação entre as variáveis. O segundo estágio consiste na escolha do método de extração (os mais comuns são: *principal component, principal factors, image factoring, maximum likelihood factoring, unweight least squares e generalized least squares*). O último estágio consiste em decidir o tipo de rotação dos componentes/fatores. Nesse trabalho, optamos pelo método de extração de componentes principais, e como iremos extrair apenas um componente não faz sentido rotacioná-lo.

As variáveis são altamente correlacionadas o que sinaliza que a redução de dados é uma técnica potencialmente útil para analisar essas informações. Além disso, as comunalidades sugerem uma forte associação entre o componente extraído e as variáveis originais.

As comunalidades representam a proporção da variância de cada variável observada que é explicada pelo componente extraído. Observe que quanto maior a correlação entre as variáveis maior é a comunalidade.

O teste KMO foi de 0,827, com um BTS de 223,915 e estatisticamente significativo ($p > 0,000$), reforçando a noção de que a base de dados é adequada. A Figura 2 ilustra a dispersão dos componentes do *Scree Test*.

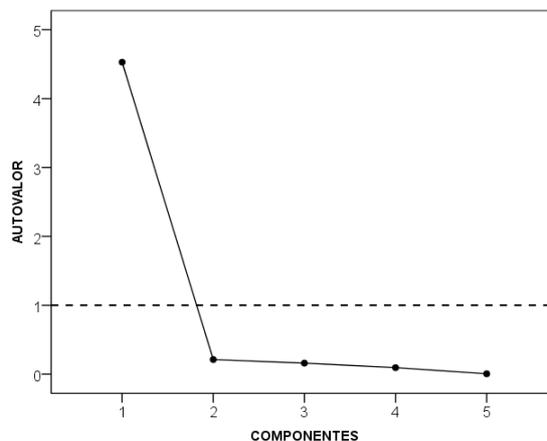


Figura 2 - Dispersão dos componentes do *Scree Test*

A linha pontilhada ilustra o critério de Kaiser, ou seja, deve-se apenas extrair componentes com autovalor maior do que um. Tanto Hair *et al.* (2006) quanto Schawb (2007) sugerem que a extração deve continuar até o pesquisador captar, pelo menos, 60% da variância. Em nosso exemplo, o primeiro componente extraído apresentou um autovalor de 4,53 e carregou 90,57% da variância das variáveis originais. A Tabela 3 sumariza essas informações.

Tabela 3 - Variância total explicada (variáveis observadas)

Componente	Autovalor inicial			Extração da soma de quadrados das cargas		
	Total	% da variância	% acumulada	Total	% da variância	% acumulada
1	4,53	90,57	90,57	4,53	90,57	90,57
2	0,21	4,24	94,82			
3	0,16	3,19	98,01			
4	0,09	1,89	99,90			
5	0,01	0,10	100,0			

Fonte: Elaboração dos autores

As regiões Norte ($\bar{X} = -0,82$; $dp = 0,52$; $n = 7$) e Nordeste ($\bar{X} = -0,58$; $dp = 0,75$; $n = 9$) apresentam piores condições de moradia quando comparadas com o Centro-Oeste ($\bar{X} = 0,78$; $dp = 0,44$; $n = 4$), Sudeste ($\bar{X} = 1,09$; $dp = 0,25$; $n = 4$) e Sul ($\bar{X} = 1,16$; $dp = 0,07$; $n = 3$). Observa-se ainda que o Sul é a região mais homogênea do Brasil. O próximo passo é analisar a distribuição do IQM por unidade da federação.

Para garantir resultados mais robustos optamos por utilizar diferentes métodos de extração. O coeficiente de correlação de Pearson entre o componente extraído via análise de componentes principais e os outros métodos de extração (*unweight least squares*, *generalized least squares* e *maximum likelihood*) foi igual ou maior a 0,980 (p -valor < 0,000), sugerindo que eles covariam no mesmo sentido. Estatisticamente, portanto, eles produzem os mesmos resultados.

Depois de extraído, o próximo passo é decidir o que fazer com o indicador¹¹. Aqui iremos explorar duas possibilidades. A primeira é apresentar um *ranking*, ou seja, dispor os casos analisados (unidades da federação) a partir do IQM (Índice de Qualidade de Moradia). A segunda é examinar em que medida o IQM se relaciona com outras variáveis. Em particular, se nosso índice for válido, devemos observar uma correlação positiva entre renda *per capita* e Índice de Qualidade de Moradia e uma correlação negativa entre o IQM e a taxa de mortalidade até cinco anos de idade. As Figuras 3 e 4 ilustram essas informações.

¹¹O componente pode ser utilizado em três principais perspectivas: (1) variável independente; (2) variável dependente e (3) índice que sumariza a dimensão de interesse.

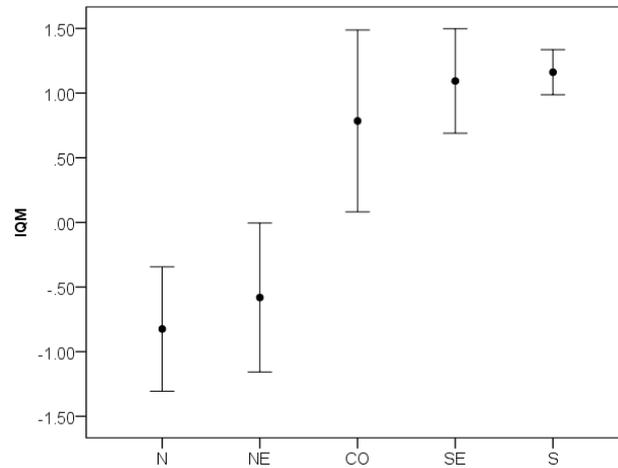


Figura 3 - Média do IQM por região (I.C 95%)
 Fonte: Elaboração dos autores

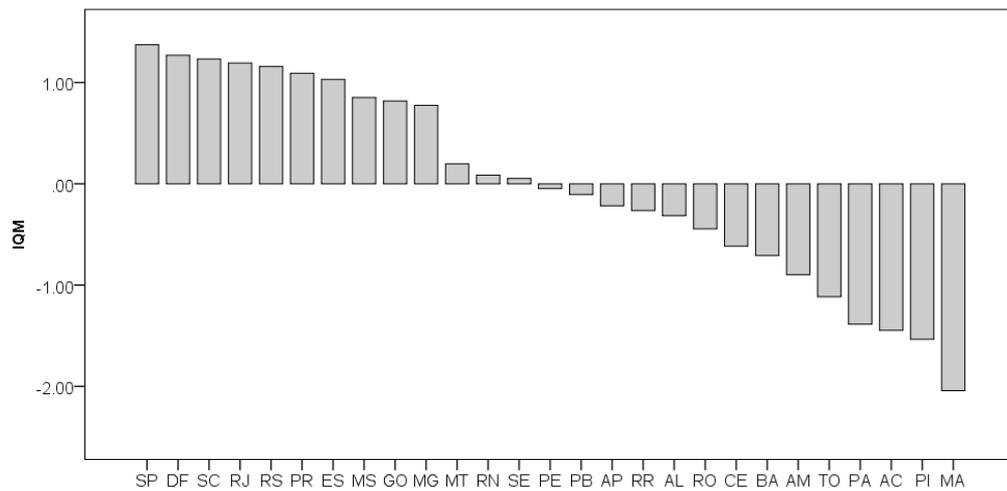


Figura 4 - *Ranking* do IQM por unidade da federação (decrecente)
 Fonte: Elaboração dos autores

Piauí e Maranhão apresentam, comparativamente, as piores condições de moradia do Brasil. O primeiro estado do Nordeste a aparecer no *ranking* é o Rio Grande do Norte na 12ª posição. O Amapá, por sua vez, é a primeira unidade da federação da região Norte a aparecer no ranking na 16ª posição. Por fim, dos dez últimos colocados no *ranking*, todos são estados do Norte ou Nordeste. Fica evidente a enorme disparidade de qualidade de moradia não só entre as regiões do Brasil mas também entre as unidades da federação.

Depois de analisar o *ranking*, o próximo passo é examinar como o IQM se relaciona com a renda *per capita* e a mortalidade. A Figura 5 (5a e 5b) ilustra a dispersão dessas informações.

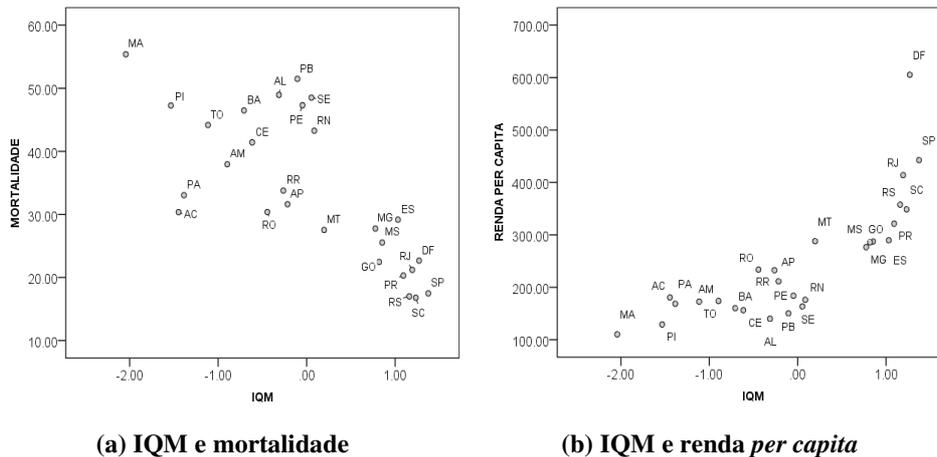


Figura 5 - IQM, mortalidade e renda per capita

Fonte: Elaboração dos autores

Existe uma correlação negativa (-0,703) e estatisticamente significativa (p -valor<0,000) entre o IQM e a taxa de mortalidade. Ou seja, quanto melhor a qualidade da moradia, menor é a taxa de mortalidade. O Maranhão apresenta a maior taxa de mortalidade (55,38) e as piores condições de moradia (-2,04). Contrariamente, observa-se uma correlação positiva (0,804) e estatisticamente significativa (p -valor<0,000) entre o Índice de Qualidade da Moradia e a renda *per capita*. Ou seja, quanto maior a qualidade da moradia, maior é a renda *per capita*.

Uma forma mais simples de construção de índices consiste em somar as variáveis de interesse e depois dividir pela quantidade de variáveis incluídas na análise. Uma eventual desvantagem dessa abordagem é não ponderar o peso de cada variável na construção do indicador final. Para superar esse problema, o pesquisador pode atribuir peso aos itens com o objetivo de ponderar o grau de importância de cada questão. Uma terceira alternativa, específica para dados categóricos, consiste em utilizar a técnica de análise de correspondência. Em nosso exemplo, o coeficiente de correlação entre o IQM e a média aritmética dos cinco indicadores é de 0,998 (p -valor<0,000).

Conclusões

O principal objetivo desse trabalho foi oferecer uma introdução aos indicadores sociais. Apresentamos a definição do conceito, as propriedades desejáveis dos indicadores e alguns critérios para classificá-los. Depois disso, o foco recaiu sobre o processo de construção do Indicador de Qualidade de Moradia (IQM).

Ressaltamos que os indicadores sociais tem um papel fundamental na formulação, implementação e avaliação de políticas públicas. São os indicadores que informam a desigualdade de renda de um país, o grau de violência de um estado e a taxa de desemprego do município. São os indicadores que permitem estimar a efetividade das ações governamentais e avaliar em que medida o dinheiro público está sendo eficientemente utilizado. No entanto, os indicadores apenas podem cumprir esses papéis quando os pesquisadores compreendem efetivamente o que eles são, quais são as suas características e como eles são construídos. Há mais de 30 anos, o professor Hubert Blalock afirmou que ainda que o desenvolvimento de teorias seja intrinsecamente importante, os problemas mais sérios e importantes que merecem nossa atenção imediata são aqueles de conceitualização e mensuração (BLALOCK, 1967). Similarmente, o físico Erwin Schroedinger afirmou que “há uma diferença entre uma fotografia tremida ou desfocada e um instantâneo de nuvens e bancos de nevoeiro”. O conhecimento científico simplesmente não pode avançar enquanto as nossas medidas não forem válidas e confiáveis. Esperamos com esse artigo ajudar estudantes de graduação, pós-graduação e pesquisadores em geral a não só interpretarem, mas também construírem os seus próprios indicadores.

Agradecimentos

Ao aporte financeiro da CAPES e do CNPq. Eventuais imprecisões são integralmente creditadas aos autores.

FIGEIREDO FILHO, D. B. F.; PARANHOS, R.; ROCHA, E. C.; SILVA JR, J. A.; MAIA, R. G. Principal components analysis for social indicator's construction . *Rev. Bras. Biom.*, São Paulo, v.31, n.1, p. 61-78, 2013.

- *ABSTRACT: What is, which it does and how we do construct social indicators? The principal aim of this paper is to answer these questions. The focus relies on the intuitive comprehension of the main concepts, characteristics and measurement of social indicators. Our targeting audience is both undergraduate, graduate students and researchers in general with limited knowledge in calculus and algebra. On methodological grounds, we show how principal component analysis technique can be used to construct social indicators. In particular, we replicate data from the Human Development Atlas to estimate Household Quality Index (HQI) per state. The results suggest that: (1) São Paulo, Distrito Federal e Santa Catarina show the higher levels of household quality while Acre, Piauí and Maranhão have the lower levels; (2) there is a negative and statistically significant correlation (-0,703; p-value<0,000) between the HQI and mortality levels and (3) there is a positive and statistically significant correlation (0,804; p-value <0,000) between the Household Quality Index and income per capita.*
- *KEYWORDS: Social indicators, quantitative methods, principal component analysis; Household Quality Index; replicability.*

Referências

ANDERSON, G. J. Causal models and social indicators: toward the development of social systems models. *Am. Sociol. Rev.*, Thousand Oaks, v.38, n.3., p. 285-301, 1973.

BABBIE, E. *Métodos de pesquisas de survey*. Belo Horizonte: UFMG, 2005. 276p.

- BARTHOLOMEW, D. J. The foundations of factor analysis, *Biometrika*, Oxford, v.71, p.221-232, 1984.
- BLALOCK, H. M. *Measurement in the social sciences: theories and strategies*, 1984. 464p.
- BLALOCK, H. M. *Toward a Theory of minority-group relations*. New York: Wiley, 1967. 227p.
- BOLLEN, K. A.; ARMINGER, G. Observational residuals in factor analysis and structural equation models. *Sociol. Methodol.*, Thousand Oaks, v.21, p.235-262, 1991.
- BONJEAN, C. M.; BROWNING, H. L. Toward comparative community research: a factor analysis of united states counties. *Sociol. Quart.*, Ames, v.10, n.2, p.157-176, 1969.
- CAMPBELL, A.; PHILIP, E. C. (Ed). *The human meaning of social change*. New York: Russell Sage Foundation, 1972. 548p.
- CARLEY, M. *Indicadores sociais: teoria e prática*. Rio de Janeiro: Zahar, 1985. 216p.
- COOPER, J. C. B. Factor analysis: an overview. *Am. Stat.*, Alexandria, v.37, n.2, p.141-147, 1983.
- COSTELLO, A. B.; OSBORNE, J. W. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Res. Eval.*, College Park, v.10, n.7, p.13-24, 2005.
- CRAMER, D.; HOWITT, D. (Ed). *The SAGE dictionary of statistics*. London: SAGE Publications, 2004. 199p.
- DANCEY, C.; REIDY, J. *Estatística sem matemática para psicologia: usando SPSS para windows*. Porto Alegre: Artmed, 2006. 608p.
- DECOSTER, J. *Overview of factor analysis*, 1998. Disponível em: <<http://www.stat-help.com/notes.html>>, Acesso em: 15 out. 2012.
- DUNN, M. J.; SCHNECK, R.; LAWSON, J. A test of the uni-dimensionality of various political scales through factor analysis: a research note. *Canad. J. Political Sci. / Rev. Can. Sci. Pol.*, Cambridge, v.6, n.4, p.664-669, 1973.
- EVERITT, B. S.; SKRONDAL, A. *The Cambridge dictionary of statistics*. Cambridge University Press, 2010. 449p.
- FIGUEIREDO FILHO, D. B.; SILVA Jr., J. A. da. Visão além do alcance: uma introdução à análise fatorial. *Opin. Publica*, Campinas, v.16, n.1, p. 160-185, 2010.
- GARSON, G. D. *Statnotes: topics in multivariate analysis*. 2009. Disponível em: <<http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>>. Acesso em: 22 out. 2012.
- GRUMM, J. G. A Factor analysis of legislative behavior. *Midwest J. Polit. Sci.*, Bloomington, v.7, n.4, p.336-356, 1963.
- GUIMARÃES, J. R. S.; JANNUZZI, P. M. IDH, indicadores sintéticos e suas aplicações em políticas públicas: uma análise crítica. *Rev. Bras. Estud. Urbanos Reg.*, Recife, v.7, n.1, p.73-90, 2005.
- HAIR Jr., J.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. *Multivariate data analysis*. 7th.ed. Prentice Hall, 2009. 928p.

- HARMAN, H. H. *Modern factor analysis*. 2.ed. Chicago: University of Chicago Press, 1967. 508p.
- HENRIOT, P. Political questions about social indicators. *Western Political Quarterly*, v.23, n.2, p.235-55, 1970.
- ISOGAWA, Y and OKAMOTO, M. linear prediction in the factor analysis model. *Biometrika*, Oxford, v.67, n.2, p.482-484, 1980.
- JANNUZZI, P. M. Indicadores para diagnóstico, monitoramento e avaliação de programas sociais no Brasil. *Rev. Serv. Público*, Brasília, v.56, n.2, p.137-160, 2005.
- JANNUZZI, P. M. Considerações sobre uso, abuso e mau uso de indicadores nas políticas públicas municipais. *Rev. Adm. Pública*, Rio de Janeiro, v.36, n.1, p.51-72, 2002.
- KIM, J.; MUELLER, C. W. *Factor analysis: statistical methods and practical issues*. Beverly Hills, CA: Sage, 1978a. 88p.
- KIM, J.; MUELLER, C. W. *Introduction to factor analysis - what it is and how to do it*. Beverly Hills, CA: Sage, 1978b. 79p.
- LAND, K. Social Indicators. *Ann. Rev. Sociol.*, Palo Alto, v. 9, p.1-26, 1983.
- LAND, K. On the definition on social indicators. *Am. Sociol.* New York, v.6, p.322-325, 1971.
- LAND, K.; MICHALOS, A. C.; SIRGY, M. J. (Ed.) *Handbook of social indicators and quality-of-life research*. Dordrecht, Netherlands: Springer Publishers, 2002.
- LAWLEY, D. N.; MAXWELL, A. E. Regression and factor analysis. *Biometrika*, Oxford, v.60, n.2, p.331-338, 1973.
- LIGNY, C. L.; NIEUWDORP, G. H. E; BREDERODE, W. K; HAMMERS, W. E; HOUWELINGEN, J. C. van. An application of factor analysis with missing data. *Technometrics*, Milwaukee, v.23, n.1, p.91-95, 1981.
- MACKELPRANG, A. J. Missing data in factor analysis and multiple regression. *Midwest J. Polit. Sci.*, v.14, n.3, p.493-505, 1970.
- OLSON, M. *The logic of collective action*. Cambridge: Harvard University Press, Cambridge, MA, 1965. 176p,
- RUMMEL, R. J. *Applied factor analysis*. Evanston: Northwestern University Press, 1970. 617p.
- SCHAWB, A. J. *Eletronic classroom*. 2007. Disponível em: <<http://www.utexas.edu/ssw/eclassroom/schwab.html>>. Acesso em: 22 set. 2012.
- SCHRÖDINGER, E. Die gegenwärtige situation in der quantenmechanik (A situação atual da mecânica quântica). *Naturwissenschaften*, 1935. Heidelberg.
- SHELDON, M. E. Investment and involvement as mechanism producing organizational commitment. *Admin. Sci. Quart.*, Thousand, v.16, p.143-150, 1971.
- SLATIN, G. T. A factor analytic comparison of ecological and individual correlations: some methodological implications. *Sociol. Quart.*, v.15, n.4, p.507-520, 1974.
- SOLIGO, V. Indicadores: conceito e complexidade do mensurar em estudos de fenômenos sociais. *Est. Aval. Educ.*, São Paulo, v. 23, n. 52, p. 12-25, 2012.

TABACHNICK, B.; FIDELL, L. *Using multivariate analysis*. Needham Heights: Allyn & Bacon, 2007. 966p.

VERMUNT, J. K.; MAGIDSON, J. Factor analysis with categorical indicators: a comparison between traditional and latent class approaches. In: VAN DER ARK, A.; CROON, M.; SITSTMA, K. *Advancements in categorical data analysis*. London: Erlbaum, 2004. P.1-41.

WILCOX, L. D.; BROOKS, R. M.; BEAL, G. M.; KLONGLAN, G. E. *Social indicators and societal monitoring*. San Francisco: Jossey-Bass, 1972. 64p.

WILCOX, L. D.; BROOKS, R. M. Toward the development of social indicators for policy planning. In: ANNUAL MEETING OF THE OHIO VALLEY SOCIOLOGICAL SOCIETY, 1971, Cleveland.

YALCIN, I.; AMEMIYA, Y. Nonlinear factor analysis as a statistical method. *Stat. Sci.*, Bethesda, v.16, n.3, p.275-294, 2001.

ZELLER, R. A.; CARMINES, E. G. *Measurement in the social sciences: the link between theory and data*. Cambridge: Cambridge University Press, 1980. 212p.

Recebido em 25.03.2013

Aprovado após revisão em 15.07.2013