
POTENCIAL DE DESENVOLVIMENTO DOS MUNICÍPIOS FLUMINENSES: UMA METODOLOGIA ALTERNATIVA AO IQM, COM BASE NA ANÁLISE FATORIAL EXPLORATÓRIA E NA ANÁLISE DE *CLUSTERS*

ARTIGO

Maria Cecília Prates Rodrigues¹

Doutoranda em Administração na FGV/EBAPE. Mestre em Economia pela UFMG/CEDEPLAR (1982) e Economista pela UFMG (1978).
Assessora em CDI & Pesquisa no SEBRAE/RJ (1999/2000).
Pesquisadora da FGV/IBRE.
E-mail: ebprates@fgv.br

RESUMO

Na era da informação, métodos para a interpretação de dados são cada vez mais necessários, tendo em vista a disponibilização exponencial de estatísticas e informações. Assim, a análise fatorial e a análise de *clusters*, com base no estudo de técnicas de interdependência, têm se mostrado muito úteis para o entendimento da estrutura dos indicadores, dos casos (ou unidades observacionais) e dos questionários de opinião. É justamente este entendimento que vai orientar o processo de tomada de decisão, tanto na esfera privada como na pública.

Neste artigo, investiga-se a aplicação da análise fatorial exploratória e da análise de *clusters* na identificação do potencial de desenvolvimento dos municípios fluminenses. Esta investigação é realizada tomando-se por base a mesma matriz dos dados utilizada pela Fundação CIDE² para compor, por meio de metodologia própria, o seu Índice de Qualidade dos Municípios – IQM.

Comprova-se que a principal vantagem da metodologia aqui proposta é fazer com que a análise se torne mais parcimoniosa, de mais fácil interpretabilidade e menos passível de erros nas medidas dos dados.

ABSTRACT

Factor analysis and cluster analysis, based on interdependence techniques, are being considered more and more useful for the understanding of the underlying structure of questions, variables and cases (observation units). Undoubtedly this better comprehension will help to improve the decision process, both in private and public sectors.

In this article, we investigate an application of exploratory factor analysis and cluster analysis in order to identify the potential development degree of each one of Rio de Janeiro municipalities. For this purpose, we used the same data matrix that was employed by CIDE Foundation³ in their Municipality Quality Index – MQI.

Compared to IQM, we conclude that the main advantage of the methodology here proposed is that the analysis becomes much more parsimonious, easy to interpret and less subject to measure errors.

¹ A autora agradece ao Prof. Moisés Balassiano, da Fundação Getúlio Vargas/ Escola Brasileira de Administração Pública e Empresas, por seus valiosos comentários.

² A Fundação CIDE é o Centro de Informações e Dados do Rio de Janeiro, ligado à Secretaria de Estado de Planejamento e Controle.

³ CIDE Foundation is Rio de Janeiro Center for Information and Data, subordinated to State Secretary of Control and Planning.

1. INTRODUÇÃO

A análise multivariada é utilizada para que se entenda melhor a relação de dependência entre dois conjuntos de variáveis – um formado por variáveis dependentes (Y_1, \dots, Y_j) e outro por variáveis independentes (X_1, \dots, X_i) – e a relação de associação mútua entre um determinado conjunto de variáveis (X_1, \dots, X_p). Em se tratando de variáveis quantitativas, pode-se dizer que os modelos de regressão múltipla e de MANOVA são exemplos de técnicas multivariadas relacionadas com aquele primeiro objetivo; já os modelos de análise fatorial e de análise de *clusters* estão relacionados com o segundo objetivo (LATIF, 1994: 1).

A proposta deste artigo é procurar entender a aplicação da análise fatorial exploratória e da análise de *clusters* – métodos para a interpretação de dados – para orientar o processo de tomada de decisão em questões relacionadas à Administração. Assim, o que se pretende é, a partir da base de dados utilizada pela Fundação CIDE para compor o seu Índice de Qualidade dos Municípios – IQM, segundo metodologia própria, propor uma forma alternativa de interpretação desses mesmos dados, à luz dos dois modelos multivariados. Como se sabe, o IQM pode servir de “farol” a orientar tanto os investimentos privados como as políticas públicas de desenvolvimento dos municípios fluminenses.

Este artigo foi dividido em seis itens. O primeiro, que compreende esta introdução, visa a apresentar o objetivo do artigo e a forma como o trabalho se encontra estruturado. No segundo, são analisados os principais aspectos teóricos relacionados com a análise fatorial e de *clusters*, e que foram relevantes para o estudo empírico em questão. O item três traz uma breve descrição da metodologia do IQM desenvolvida pela Fundação CIDE, e os resultados encontrados na ordenação dos municípios fluminenses segundo esse índice. Já o item quatro mostra justamente como foram aplicadas a análise fatorial exploratória e a análise de *clusters*, métodos alternativos ao IQM para avaliar o potencial de desenvolvimento dos municípios do Estado do Rio de Janeiro. São, então, descritas as principais etapas seguidas e os problemas encontrados, até se chegar às soluções consideradas as mais adequadas. Os

resultados finais são também apresentados. No item 5, a título de conclusão, são comentadas as vantagens de se utilizar a análise fatorial e a análise de *clusters vis-à-vis* ao método do IQM. Finalmente, o item seis lista as referências bibliográficas utilizadas.

2. ANÁLISE FATORIAL EXPLORATÓRIA E ANÁLISE DE CLUSTERS: ASPECTOS TEÓRICOS

AAKER, KUMAR e DAY (1998: 582-608) apresentam uma explicação bastante elucidativa acerca da aplicação destes modelos. Segundo eles, a análise fatorial serve para a combinação de variáveis que criariam novos fatores, os chamados fatores latentes; já a análise de *clusters* combina os objetos, ou unidades observacionais, para formar novos grupos. O que se pratica, em ambos os métodos, é a análise de interdependência, a partir da matriz de variância/co-variância (ou de correlação) entre as variáveis ou entre as unidades observacionais.

No âmbito da análise fatorial, o que se pretende é a identificação de possíveis associações entre as variáveis observacionais, de modo que se defina a existência de um fator comum (latente) entre elas. Assim, pode-se dizer que a análise fatorial, ou análise do fator comum, tem como objetivo a identificação de fatores ou *constructos* subjacentes às variáveis observacionais, o que, sem dúvida, contribui para facilitar sobremaneira a interpretação dos dados. Isto porque, por exemplo, ao invés de buscar entender o comportamento de 20 variáveis observacionais, o analista deverá procurar entender o comportamento de 3 ou 4 fatores latentes por meio do comportamento dos seus *scores* fatoriais (que será definido mais adiante, no item 4).

JOHNSON e WICHERN (1992: 396) explicam que na análise do fator comum as variáveis são agrupadas em função de suas correlações. Isso significa que variáveis que compõem um determinado fator devem ser altamente correlacionadas entre si e fracamente correlacionadas com as variáveis que entram na composição do outro fator.

A idéia básica do modelo é a de que cada variável observacional possa ser expressa pelo(s) fator(es) latente(s). A tabela 1 ilustra o caso em que, após se proceder à análise dos dados, identificou-se a existência de 1 fator latente (F) comum às 5 variáveis observacionais (X_1, \dots, X_5). Vê-se, pelo exemplo, que cada uma das cinco variáveis observacionais é explicada pelo fator comum (F) e pelo respectivo fator específico ou resíduo (e). Os coeficientes do fator (L) são as cargas fatoriais, e representam o grau de associação (ou de correlação, quando a matriz de correlação tiver sido a matriz de associação utilizada) entre a variável e o fator.

A comunalidade é o valor da carga fatorial (L) elevado ao quadrado, e representa o percentual da variância da variável, que é explicada pela variância do fator comum. Quanto mais elevada for a comunalidade (varia de 0 a 1), maior é a evidência de que X é um indicador válido do *constructo* que se deseja medir. Fazendo um paralelo com a análise de regressão, L seria o coeficiente de correlação entre a variável dependente e a independente, e L^2 seria o coeficiente de determinação do modelo.

Tabela 1: O fator latente gerado a partir da análise fatorial

Identificação do fator latente	Comunalidade
$X_1 = L_1 F + e_1$	L_1^2
$X_2 = L_2 F + e_2$	L_2^2
$X_3 = L_3 F + e_3$	L_3^2
$X_4 = L_4 F + e_4$	L_4^2
$X_5 = L_5 F + e_5$	L_5^2

Segundo sua finalidade, a análise do fator comum pode ser exploratória ou confirmatória. No primeiro caso, o investigador dispõe apenas das variáveis e deseja identificar quantos e quais fatores latentes podem ser extraídos do conjunto das variáveis por meio das associações entre elas. No segundo caso, como aponta BALASSIANO (2000: 1), os fatores já são conhecidos *a priori*, bem como seus indicadores, e o que se pretende é testar a adequação desses fatores.

Como já dito, será adotada a análise fatorial exploratória no âmbito deste artigo. Isto porque o objetivo aqui será justamente o de explorar, ou investigar, a matriz de dados utilizada pela Fundação CIDE em seu IQM (91 municípios vs 38 indicadores), para identificar quantos e quais são os fatores latentes existentes. A hipótese subjacente a essa opção é a de que se quer aqui entender o que os dados estão dizendo, independentemente de associações previstas para esses indicadores, com base em modelos teóricos.

Já a análise de grupamentos, ou análise de *clusters*, visa a agrupar indivíduos (ou unidades observacionais, ou objetos) segundo determinados critérios de distância entre os respectivos vetores de dados. É bom lembrar que a cada unidade observacional está associado um vetor multidimensional de dados [x_1, x_2, \dots, x_p].

Como afirmam JOHNSON e WICHERN (1992: 573), a análise de *clusters* corresponde também a uma importante técnica exploratória que busca identificar uma estrutura de grupamentos “naturais” para avaliar a dimensionalidade dos dados, identificar *outliers* e fornecer interessantes hipóteses acerca de associações – neste último caso ela é usada sobretudo para agrupar variáveis.

Existem dois métodos de grupamento: o hierárquico e o não hierárquico. O método hierárquico pode ser de dois tipos: aglomerativo (*bottom-up*) e divisível (*top-down*). Estabelece-se uma relação de hierarquia entre o objeto (individual) e o conjunto dos objetos (*cluster*). Uma vez incorporado a um grupo, o objeto permanece associado a ele até o final do processo de *clustering*. Nesse método, os critérios de grupamento mais utilizados são o da associação simples (baseada nas menores distâncias entre os objetos) e o da associação completa (baseada na maior distância entre os objetos). Cada solução de *cluster* gerada deve ser devidamente interpretada, a fim de que se identifique qual a mais adequada para dar um significado aos dados em questão.

No método não hierárquico o processo de *clustering* é mais dinâmico e interativo. Diferentemente do método hierárquico, ele permite que objetos deixem um *cluster* e se associem a

outro, de modo que esses reposicionamentos sucessivos possam contribuir para melhorar os resultados finais. O critério não hierárquico mais utilizado é o *K-means*, que permite que se defina inicialmente o vetor central dos *clusters* (ou centróides), e que se busque, em seguida, inserir os objetos mais próximos a eles. Como se vê, ao contrário do método anterior, o *K-means* possibilita que se estabeleça, de antemão, o número de *clusters* com que se quer trabalhar, e que se busque testar essa hipótese a partir do significado encontrado para aquela solução de *cluster*. De certa forma, isso sugere o seu caráter de análise confirmatória.

No item 4 serão vistos os procedimentos seguidos para a análise de *cluster* dos 91 municípios do Estado do Rio de Janeiro, segundo seu potencial de desenvolvimento, e se verá também porque o critério *K-means* se mostrou o mais adequado.

Enfim, tanto a análise fatorial como a análise de *cluster* são modelos multivariados utilizados para que se entenda melhor a associação entre variáveis e entre unidades observacionais. São técnicas antigas, cuja aplicação inicial deu-se no campo das ciências sociais e comportamentais. Assim, o conceito dos fatores latentes foi primeiramente lançado por Galton em 1888, e em 1904 Spearman estendeu o conceito ao desenvolvimento da análise fatorial, quando a aplicou para melhor interpretar os testes de inteligência, conseguindo capturar o “fator da habilidade intelectual geral”, conhecido como o fator G (GIRI, 1996: 359).

No entanto, é preciso ter bem claras as limitações desses modelos. São métodos com um alto grau de subjetividade, em que várias alternativas de solução são apresentadas, cabendo ao analista optar por uma ou outra solução. Ou seja, como se verá no item 4, a qualidade da solução final fica extremamente dependente da capacidade analítica do pesquisador, que busca uma explicação que possa ser considerada razoável, segundo a sua lógica.

3. O IQM SEGUNDO A METODOLOGIA DA FUNDAÇÃO CIDE

Em 1998, a Fundação CIDE criou o Índice de Qualidade dos Municípios – IQM, com “o objetivo

de classificar os municípios fluminenses segundo seu potencial e condições apresentadas para o crescimento e o desenvolvimento”. Como esclarece o relatório da Fundação CIDE (1998: 8), o que se pretende não é medir a qualidade de vida dos habitantes, mas a forma pela qual cada município se apresenta para receber novos investimentos.

A partir de 67 variáveis, obtidas das mais diversas fontes (IBGE, CIDE, Secretarias de Estado, MEC, Sebrae, Firjan, Detran, Light, Sindicato dos Bancários) e com datas de referência variadas (de 1991 a 1998), foram construídos 38 indicadores. Para situar o indicador no contexto do município e possibilitar a comparação entre municípios, grande parte desses indicadores são apresentados de forma relativa. Exemplificando, o indicador BAN refere-se às agências bancárias no município, e é descrito como o “número de agências bancárias dividido pela raiz da população e multiplicado por 100”.

De acordo com a metodologia descrita pela Fundação CIDE (1998: 13-22), os indicadores foram escolhidos em função de sua (1) representatividade, ou capacidade de representar um determinado fenômeno, e de sua (2) disponibilidade e periodicidade de atualização.

Os 38 indicadores foram distribuídos em 7 grupos, conforme apresentados na tabela 2. Foram, também, atribuídos pesos aos indicadores e aos grupos, “de modo a refletir a importância que se desejou conferir a cada um dos aspectos considerados, tendo em vista a base teórica adotada”. Esclarece-se que a base teórica adotada foi a Teoria das Localidades Centrais (CHRISTALLER, 1966) e a Teoria dos Pólos de Desenvolvimento (PERROUX, 1964). Sobre a definição dos indicadores, ver o anexo.

Na realidade, o que essas teorias fazem é apontar, de modo bem genérico, a importância da organização espacial segundo os centros polarizadores. Não se pode dizer que exista uma relação direta entre o que a teoria propõe e a forma como o IQM foi construído. Isto significa que a equipe da Fundação CIDE traduziu aquela teoria na forma do IQM, entendido como o processo de seleção de indicadores, definição dos grupos,

alocação dos indicadores aos grupos e atribuição de pesos. Outra equipe de outra instituição poderia ter feito uma leitura bastante distinta da teoria, com outros indicadores, outros grupos, outros pesos, etc., ou até com os mesmos indicadores, mas com outros

pesos e outros grupos. Em outras palavras, o que se quer dizer é que a composição do IQM é apenas uma forma subjetiva de percepção da Teoria das Localidades Centrais e da Teoria dos Pólos.

Tabela 2: IQM: Grupos, indicadores e pesos

Grupos e pesos	Indicadores e pesos ⁴
i. Centralidade e vantagem locacional - CEN (peso 10)	CVA (10); CON (7); ONI (9); MES (6); CAT (8)
ii. Qualificação da mão-de-obra - QMA (peso 9)	CES (10); ALF (7); PRO (8); TEC (9)
iii. Riqueza e potencial de consumo - RIQ (peso 9)	CRA (9); PIB (10); ENE (7); DEP (8); FMU (6)
iv. Facilidades para negócios - FAC (peso 8)	BAN (9); TEL (8); COR (7); SEB (5); HOT (6); INC (10); INT (4)
v. Infra-estrutura para grandes empreendimentos - IGE (peso 8)	ROD (8); FE (7); AER (5); GAS (6); LIN (9); DIS (10)
vi. Dinamismo - DIN (peso 7)	CRE (10); B24 (7); VEI (7); OPC (9); PIC (10)
vii. Cidadania - CID (peso 6)	ENS (9); LEI (7); SEG (5); CUL (6); JUS (8); DOM (10)

Fonte: CIDE (1998).

Em linhas gerais, a metodologia para o cálculo do IQM segue a metodologia que vem sendo utilizada pelo Programa das Nações Unidas (PNUD) para a estimativa do Índice de Desenvolvimento Humano (IDH) desde 1990 (PNUD, 1999: 159-160). O IQM é obtido da média ponderada (pelos respectivos pesos) dos índices calculados para os grupos; estes, por sua vez, são calculados considerando-se a média ponderada dos índices para os indicadores que compõem cada um dos grupos. Já os índices para cada indicador são obtidos por interpolação linear;

ou seja, ao melhor resultado apurado para o indicador é atribuído valor igual a um, e ao pior resultado é atribuído valor igual a zero. Desnecessário dizer que o valor para cada um desses indicadores constituídos varia de zero a um.

Para uma melhor compreensão dessa metodologia de formação de índices, veja-se os três exemplos a seguir (tabela 3):

Tabela 3 - Exemplos de formação de índices

<p>Ex.1 – Índice para cada indicador Índice da taxa de alfabetização dos maiores de 15 anos de idade – IALF $IALF = ALF - \text{menor (ALF)} / \text{maior (ALF)} - \text{menor (ALF)}$ Onde ALF é a taxa de alfabetização de um dado município; menor (ALF) é a menor taxa de alfabetização entre todos os 91 municípios; e maior (ALF) é a maior taxa de alfabetização entre todos os municípios.</p>

⁴ Os indicadores estão abreviados; sobre a descrição dos indicadores, ver o anexo.

Ex. 2 – Índice setorial (dos grupos)

Índice de Qualificação da mão-de-obra – IQMA

$$IQMA = (10.ICES + 7.IALF + 8.IPRO + 9.I TEC) / 34$$

Ex. 3 – IQM

$$IQM = (10.CEN + 9.IQMA + 9.IRIQ + 8.IFAC + 8.IGE + 7.DIN + 6.CID) / 57$$

A tabela 4 apresenta o *ranking* dos 91 municípios fluminenses, obtido a partir da metodologia aplicada pela Fundação CIDE para determinar o potencial de desenvolvimento de cada um dos municípios. Convém notar que o número de municípios alocados em cada coluna equivale ao número de municípios definidos para cada um dos quatro grupos com base na análise de *cluster* (item 4), de modo a facilitar ao leitor a comparação dos resultados obtidos da aplicação desses dois métodos de análise.

Feita esta breve descrição da metodologia do IQM aplicada pela Fundação CIDE e dos principais resultados encontrados, propõem-se aqui algumas

questões para reflexão, que deverá ser feita ao longo do item 4. São elas:

- Para avaliar o potencial de desenvolvimento dos municípios seria mesmo necessário este número tão grande de indicadores (38), que pode, inclusive, comprometer a qualidade dos resultados?
- Os 7 grupos constituídos representam, de fato, a melhor forma de agrupar estes indicadores? Ou haveria outra alternativa válida, mais parcimoniosa e de fácil interpretação?
- Haveria outra maneira, coerente e lógica, de sistematizar os municípios segundo o seu potencial de desenvolvimento?

Tabela 4: Ranking dos municípios fluminenses em ordem decrescente do valor do IQM

1. Rio de Janeiro	16.Itaguaí	45. Aperibé	75.Cachoeiras de Macacu
2. Niterói	17.Angra dos Reis	46. Pinheiral	76.Paty do Alferes
3. Resende	18.Piraí	47. Arraial do Cabo	77.Bom Jardim
4. Macaé	19.Duque de Caxias	48. Paraíba do Sul	78.Laje do Muriaé
5. Volta Redonda	20.Rio das Ostras	49. Cordeiro	79.Silva Jardim
6. Petrópolis	21.Araruama	50. Magé	80.Japeri
7. Casimiro de Abreu	22.Bom Jesus do Itabapoana	51. Areal	81.Cardoso Moreira
8. Três Rios	23.Nova Iguaçu	52. Maricá	82.Duas Barras
9. Miguel Pereira	24.Vassouras	53. Quissamã	83.S. José do Vale do R.Preto
10.Campos	25.Nova Friburgo	54. Cantagalo	84.Tanguá
11. Itatiaia	26.Valença	55. S. Pedro d`Aldeia	85.Trajano de Moraes
12. Cabo Frio	27.Miracema	56. Porciúncula	86.Sta. Maria Madalena
13. Barra Mansa	28.Mangaratiba	57. Eng.Paulo Frontin	87.Varre-Sai
14. S. Ant. de Pádua	29.Barra do Piraí	58. Conceição de Macabu	88.S. Sebastião do Alto
15. Teresópolis	30.Búzios	59. Guapimirim	89.S. José de Ubá
	31.Mendes	60. Saquarema	90.Sumidouro
	32.Iguaba Grande	61. Rio Claro	91.São Francisco do Itabapoana
	33.Rio Bonito	62. Itaboraí	
	34.Itaperuna	63. Parati	
		64. Carmo	

	35. Nilópolis 36. São Gonçalo 37. Paracambi 38. Queimados 39. Rio das Flores 40. C. Levy Gasparian 41. Itaocara 42. S. João de Meriti 43. Porto Real 44. Belford Roxo	65. Natividade 66. Cambuci 67. Quatis 68. Italva 69. Macuco 70. S. João da Barra 71. São Fidélis 72. Sapucaia 73. Seropédica 74. Carapebus	
--	--	---	--

Fonte: CIDE (1998).

4. UMA METODOLOGIA ALTERNATIVA À DO IQM

À luz da análise fatorial exploratória, tentar-se-á identificar alguns poucos fatores latentes subjacentes aos 38 indicadores (ou variáveis observacionais) utilizados no IQM – item 4.1. Esses fatores correspondem, no IQM, aos 7 índices setoriais, e têm por finalidade capturar os principais *constructos* relacionados à idéia do potencial de desenvolvimento.

Com base na análise de *clusters*, procurar-se-á uma forma lógica de organizar os municípios, segundo o seu potencial de desenvolvimento – item 4.2. Os *clusters* de municípios a serem formados cumprem a mesma finalidade do *ranking*, no que se refere ao IQM.

O *software* aqui utilizado para a análise dos dados foi o SPSS 10.0 for Windows, na sua versão em inglês.

4.1. Aplicação da análise fatorial exploratória

Inicialmente, é apresentado um roteiro básico para a realização da análise fatorial exploratória. A seguir, são descritas as principais etapas cumpridas até se chegar à solução aqui considerada a mais adequada.

Roteiro básico

(1) Verificação da adequação dos dados à análise fatorial, por meio de:

- Análise da matriz de correlação: na matriz, cada indicador deve apresentar correlação elevada com pelo menos alguns indicadores, não necessariamente com todos. Isto significa que esse grupo de indicadores correlacionados têm um *constructo* em comum, capturado pelo fator comum. Se a correlação de um determinado indicador for baixa com todos os outros, isto quer dizer que ele não traduz, juntamente com qualquer outro indicador, qualquer idéia em comum. Um valor de correlação pode ser considerado aceitável se estiver acima de 0,4.
- Teste KMO: a medida Kaiser-Meyer-Olkin testa a adequação da amostra quanto ao grau de correlação parcial entre as variáveis, que deve ser pequeno. Se isto ocorre, significa que os fatores latentes explicam grande parte da associação entre as variáveis, e que os resíduos estão pouco associados entre si. Valores para o teste KMO iguais ou inferiores a 0,7 indicam que a análise fatorial pode ser inadequada.
- Teste de esfericidade de Bartlett: neste teste, a hipótese inicial (H_0) é que a matriz de correlação é uma matriz-identidade, o que indica que o modelo é inadequado. Se, por exemplo, para um nível de significância definido em 0,05 a significância (α) encontrada for menor que 0,05, deve-se rejeitar H_0 e concluir, portanto, que o modelo é adequado em razão das associações verificadas.

(2) Determinação do número de fatores latentes – alguns critérios básicos podem ser seguidos para a extração dos fatores latentes mais relevantes, tais como: fatores com autovalores (L), associados à matriz de associação, maiores que 1 (para esse critério, a matriz de associação analisada, Σ , deve ser a de correlação); a “regra do cotovelo”⁵ no *scree plot*; e a variância acumulada igual ou acima de 70%.

(3) Análise da solução fatorial: o valor da comunalidade extraída para as variáveis deve ser razoável (pelo menos acima de 0,5). Também devem ser elevados os valores das cargas fatoriais obtidos na matriz dos fatores rotados, isto é, depois da rotação dos eixos, pois são justamente essas cargas que vão auxiliar na interpretação dos fatores.

(4) Interpretação do significado dos fatores: sugere-se a rotação nos eixos justamente para facilitar a interpretação dos fatores. Os métodos de rotação mais utilizados são: Varimax, que faz uma rotação ortogonal dos eixos, e Oblimin, que promove uma rotação oblíqua nos eixos. A idéia aqui é verificar qual método de rotação propicia a interpretação mais plausível dos fatores. Pode ocorrer também que nenhum dos métodos facilite a interpretação; neste caso, deve-se repensar se o desenho da análise fatorial que se está utilizando é um procedimento metodológico válido para os dados em questão.

(5) Obtenção dos *scores* fatoriais – os *scores* fatoriais são os valores, assumidos pelo fator latente, para cada unidade observacional. Os valores do *score* fatorial resultam da combinação linear entre cada um dos valores das variáveis observacionais e os respectivos coeficientes do *score* fatorial (obtidos na matriz dos coeficientes do *score* fatorial). AAKER, KUMAR e DAY (1998: 589) aconselham o uso dos *scores* fatoriais, ao invés

das variáveis originais, em análises ou interpretações subsequentes das variáveis.

Principais etapas seguidas:

1ª etapa

Foi aplicada a análise fatorial à base de dados como um todo, ou seja, aos 38 indicadores e 91 municípios.

Não se pode, de forma alguma, concluir pela adequação dos dados, mesmo com os resultados favoráveis do teste KMO (0,759) e do teste de Bartlett ($\alpha < 0,001$). A matriz de correlação mostrou que existem muitos indicadores cuja correlação é bastante baixa, ou praticamente nenhuma, com todos os demais indicadores. Nesta situação estão indicadores como: AER, CAT, CRE, FMU, ENS, GAS, INC, LEI, LIN, MES, ONI, PIB, PIC e PRO. Isto sugere que vários dos indicadores utilizados relacionam-se muito superficialmente com todos os demais indicadores. Neste sentido, eles podem ser considerados indicadores não válidos ou espúrios, não contribuindo para medir o que realmente se deseja.

Além disso, a solução fatorial (método de rotação Varimax) mostrou-se muito insatisfatória, haja vista, sobretudo, a falta de significado dos fatores, buscado na lógica das cargas fatoriais dos indicadores que os compõem. Ainda, as baixas comunalidades verificadas em grande parte das variáveis indicam que os 9 fatores extraídos (com autovalores maiores que 1) explicam, juntos, muito pouco a variância total dos indicadores; basta ver que a comunalidade é inferior a 0,5 em 14 dos 38 indicadores.

Concluiu-se, portanto, pelo não ajuste do modelo.

2ª etapa

Observando-se o quadro das estatísticas descritivas apuradas na 1ª etapa, pode-se levantar a hipótese de que a baixa qualidade dos dados poderia ser uma das causas para o não-ajuste do modelo acima. Note-se que em 22 dos 38 indicadores o desvio-padrão é igual ou maior que a média, o que indica que os dados estão muito dispersos. A prática tem mostrado que em uma distribuição relativamente homogênea essa relação entre o

⁵ O *scree plot* é o gráfico dos autovalores (L), ordenados do maior para o menor, *versus* os respectivos fatores (i), em que $i=1,2,3,j,\dots$. A determinação do número apropriado de fatores corresponde ao fator $i=j$, definido no ponto onde o gráfico apresenta uma curva acentuada, à semelhança de um cotovelo. Assim, os fatores $i < j$ são os que apresentam autovalores relevantes; já os fatores $i > j$ são aqueles que detêm autovalores relativamente pequenos e todos são aproximadamente do mesmo tamanho (JOHNSON e WICHERN, 1992: 475).

desvio-padrão e a média, conhecida como coeficiente de variação, é de até 30%.

Sem dúvida, essa baixa qualidade dos dados é fruto de uma dispersão real dos resultados dos indicadores entre os municípios. Como seria de esperar, os resultados do município do Rio de Janeiro, por exemplo, são os que mais se distanciam da média da distribuição, em razão de seu próprio dinamismo, atribuído ao fato de esse município ser a sede da capital do Estado. Mas, suspeita-se também que tenha havido falhas na entrada dos dados, o que pode prejudicar muito o tratamento dos dados pela análise fatorial e análise de *clusters* aqui pretendido. Apenas a título de exemplificação, citam-se alguns exemplos dessas possíveis falhas, ou seja, resultados que não parecem condizer com a realidade:

- No que se refere ao indicador MES (percentual de matrículas do ensino superior), a média do Estado é de 6,7% e, no entanto, o dado do município de Seropédica é de 198,3%.
- Quanto ao indicador PIB (PIB *per capita* em R\$ de 1996), o valor de Pirai é o mais elevado, de 44.828, enquanto a média do Estado é de 4.790.

Considerou-se, portanto, fundamental proceder ao ajuste dos dados. O critério adotado foi considerar como *missing* (sem dados) os resultados cujos valores padronizados correspondentes superassem, para mais ou para menos, 2 desvios-padrão.

Feito o ajuste dos dados, rodou-se novamente a análise fatorial com os dados ajustados, isto é, valores *missing* no lugar dos *outliers* (no SPSS adotou-se a opção de substituir os *missing values* pela média do indicador). Apesar disso, o modelo da análise fatorial não chegou a apresentar melhora sensível que levasse à sua aceitação. Inicialmente, considerando-se a condição do autovalor maior que 1, foram gerados 10 fatores, mas com cargas fatoriais fracas e de difícil interpretabilidade (mesmo utilizando-se a matriz dos fatores rotados). Tentou-se reduzir o número de fatores, de modo que a interpretabilidade dos fatores melhorasse, chegando-se a 4 fatores: nesta situação extrema, o valor acumulado dos autovalores ficou muito baixo

(48%), razão pela qual decidiu-se novamente pelo não ajuste do modelo.

3ª etapa: Solução encontrada

Se o ajuste dos dados ainda não for suficiente para permitir a aplicação do modelo, por que não partir para o ajuste dos indicadores?

Já que a matriz de correlação, pelas baixas correlações detectadas, havia apontado a existência de indicadores não válidos, e portanto inadequados, por que não excluir alguns destes indicadores? Mesmo porque a prática tem mostrado que a situação ideal para aplicar a análise fatorial é a presença de um número não muito elevado de variáveis, porém do maior número possível de unidades observacionais.

Decidiu-se, então, pela utilização de apenas 15 indicadores, ou seja, daqueles que apresentam os níveis mais elevados de correlação com os demais.

Desta vez, o modelo mostrou-se relativamente ajustado. Houve “melhora” nas estatísticas descritivas, no teste KMO, nas comunalidades obtidas e na matriz de correlação. As aspas são propositais, e visam a alertar o leitor de que os resultados encontrados ainda ficaram longe de uma solução ideal – apesar de os testes KMO e de Bartlett estarem satisfatórios, o desvio-padrão ainda seguiu sendo maior que a média em 4 dos 15 indicadores, e 4 dentre as 15 comunalidades extraídas ficaram entre 0,4 e 0,5. Mas, a sensível melhora conseguida foi na interpretabilidade dos dados, pois as cargas fatoriais da matriz *rotada* permitiram, com facilidade, a identificação do significado dos três fatores latentes *capturados*. Senão, veja-se os resultados na tabela 5.

Assim, estes três fatores conseguem explicar 66% da associação total entre os dados para os 15 indicadores (valor acumulado dos autovalores). Portanto, em análises futuras, ao invés de se trabalhar com os 15 indicadores, poder-se-ia trabalhar apenas com os 3 fatores, ou seja, com os *scores* fatoriais gerados, sabendo que se estaria incorrendo em uma perda de 34% na associação entre os dados.

Tabela 5 – Matriz dos fatores rotados – método Varimax

<i>Rotated Factor Matrix</i>			
	<i>Factor</i>		
	1	2	3
BANM	,787		
CVAM	,686		
CONM	,676		
DEPM	,652		
B24M	,647		,445
JUSTM	,631		
OPCM	,566		
SEBM	,557		
CRAM		,811	
TELM		,667	
VEIM		,633	
INTM		,487	
ALFM			,830
CESM		,567	,625
DOMM			,558

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization.

A Rotation converged in 7 iterations.

Obs.: O “M” acrescido à abreviatura de cada indicador significa que esses indicadores estão com os seus valores *outliers* como *missings*.

O fator 1 mostra as condições de DINAMISMO do município, e foi constituído pelos seguintes indicadores: agências bancárias, consumo varejista, concessionárias de veículos, depósitos bancários, postos de banco 24 horas, acesso à justiça, operações de crédito efetuadas, existência de balcões Sebrae.

O fator 2 transmite a idéia de POTENCIAL DE CONSUMO do município, e foi formado pelos indicadores: chefes de domicílios com renda elevada, terminais telefônicos, veículos novos e provedores de Internet.

O fator 3, interpretado como CONDIÇÕES DE VIDA da população, foi constituído pelas variáveis: taxa de alfabetização, chefes de domicílios com escolaridade razoável e domicílios em condições adequadas.

A idéia é que esses três fatores juntos possam cumprir o papel do IQM proposto pela Fundação CIDE, qual seja, o de dar a idéia do potencial de desenvolvimento do município. Com efeito, quando se roda a análise fatorial com os *scores* dos 3 fatores latentes, um fator apenas é extraído, e representa justamente o POTENCIAL DE DESENVOLVIMENTO dos municípios.

4.2. Aplicação da análise de *clusters*

Será feito um breve relato das etapas empreendidas para realizar a análise de *cluster* dos dados e chegar à solução considerada a mais adequada.

1ª etapa

Mesmo sabendo de antemão que os dados apresentam uma grande dispersão, resolveu-se tentar a análise de *clusters* com as variáveis observacionais originais, ou seja, sem dar qualquer tipo de tratamento aos *outliers* já identificados. Para essa análise de *clusters*, levou-se em consideração, inicialmente, os 15 indicadores selecionados anteriormente para a análise fatorial. Em seguida, foram considerados apenas os 3 indicadores que apresentaram as cargas fatoriais mais elevadas em cada um dos 3 fatores latentes. Em ambas as tentativas a solução não foi considerada razoável, ocorrendo uma concentração grande de municípios em determinados grupos.

O que essas tentativas mostraram foi que, já que o objetivo em questão era identificar uma certa dimensionalidade (ou lógica) entre os municípios, e não a identificação de *outliers*, a inclusão destes *outliers* acabou tornando-se um elemento perturbador a mais para a análise dos *clusters* de municípios.

2ª etapa

Nesta etapa, procurou-se dar tratamento aos *outliers*, que passaram a ser considerados como *missings*. A melhor solução aqui encontrada, levando-se em consideração os 15 indicadores selecionados anteriormente, foi a aplicação do *K-means*, utilizando-se a opção (do SPSS) de inclusão dos casos com dados *missing*.

Não houve, desta vez, a superconcentração de municípios em 1 ou 2 *clusters*; a distribuição dos municípios entre os *clusters* foi bem mais equilibrada nos 3 centróides inicialmente definidos. Mas a qualidade do grupamento não foi considerada satisfatória. Assim, pelo que se conhece do nível de desenvolvimento desses municípios, é inconcebível juntar em um mesmo grupo municípios tão díspares quanto Bom Jesus do Itabapoana, Itaperuna, Natividade, Rio de Janeiro, Resende e Volta Redonda. A qualidade dos grupamentos também não apresentou melhora com a definição de diferentes números de centróides iniciais.

3ª etapa

Nesta etapa, foram feitas tentativas de formação dos *clusters* a partir dos *scores* fatoriais gerados para os 3 fatores latentes, identificados na 3ª etapa do item 4.1. É bom lembrar que aquela solução dos três fatores levou em consideração o tratamento dado pelo SPSS aos *outliers*, em que os valores *missing* foram substituídos pela média do indicador (item 4.1, 2ª etapa).

Tanto pelo método hierárquico (*between-groups linkages*) como pelo critério não hierárquico *K-means*, as soluções com base nos *scores* fatoriais não foram consideradas satisfatórias. Uma possível explicação é que os *scores* fatoriais ajudam na interpretação dos dados mas, nesta situação em particular, não serviram para a delimitação dos *clusters* dos casos ou unidades observacionais.

4ª etapa: Solução encontrada

A solução aqui encontrada, considerada a mais adequada, apresenta idéias da etapa 1 e da etapa 2 desta análise de *clusters*. Da etapa 1, a contribuição trazida foi a de se trabalhar com os três indicadores com as cargas fatoriais mais elevadas de cada um dos fatores latentes. Com efeito, AAKER, KUMAR e DAY (1998: 596) aconselham que em algumas situações o analista pode, e deve, usar uma ou duas variáveis com as cargas mais elevadas na composição do fator, de modo que represente o fator em coletas de dados ou análises subsequentes. Já da etapa 2, a idéia trazida foi a do tratamento dado aos *outliers* no âmbito do critério *K-means*.

Assim, os *clusters* formados tiveram por base os indicadores de relação entre agências bancárias e população (BAN), percentual dos chefes de domicílios com rendimentos acima de 20 salários-mínimos (CRA) e taxa de alfabetização dos maiores de 15 anos de idade (ALF). Isso significa que os municípios foram grupados por um vetor de 3 variáveis observacionais que são, em princípio, fortemente representativas dos níveis de dinamismo do município, do seu potencial de consumo e das condições de vida de sua população. Essas três variáveis podem, portanto, ser consideradas representativas do potencial de desenvolvimento dos municípios, que é o conceito subjacente aos 3 fatores latentes.

Quanto aos dados levantados nesses três indicadores, é interessante notar que, no que se refere à taxa de alfabetização, 4 municípios tiveram dados *missing* em virtude de suas taxas discrepantemente baixas⁶. Foram eles: Cardoso Moreira, São Francisco do Itabapoana, Silva Jardim e Sumidouro. Taxa de alfabetização baixa é um dos sinais de baixo potencial de desenvolvimento: apesar do tratamento aos dados *missing* do critério *K-means* aqui utilizado, pode-se dizer que ele não conseguiu mascarar esse atraso, pois esses municípios foram, de fato, alocados no grupo de menor potencial de desenvolvimento. Pela razão oposta, resultados discrepantemente elevados nos indicadores CRA (Niterói e Rio de Janeiro) e BAN (Rio de Janeiro) apresentaram dados *missing* e, apesar do tratamento conferido a esse tipo de dado, esses municípios foram alocados no grupo de maior potencial de desenvolvimento. Mas, por outro lado, é bom ter claro que a definição e o tratamento aos *outliers* influiu na localização dos municípios dentro do *cluster*, reduzindo suas distâncias em relação ao centróide.

Resumindo, pode-se dizer que a melhor solução foi encontrada a partir da aplicação do critério não

⁶ Na realidade, os resultados referentes a estes 4 municípios chegaram a ser excluídos porque ALF apresentou distribuição bastante homogênea (coeficiente de variação igual a 8,4%). Em estudos futuros, deve-se pensar na definição do *outlier* para incorporar diferenças no grau de homogeneidade da distribuição.

hierárquico *K-means*, que: considerou os 3 indicadores com as cargas fatoriais mais elevadas de cada um dos fatores; incluiu nos *clusters* 6 municípios (entre os 91) com dados *missing*; e definiu 4 centróides iniciais.

Na realidade, essa solução foi a melhor porque conseguiu discriminar bem o potencial de desenvolvimento dos 91 municípios. Em outras palavras, os resultados encontrados coincidiram, em grande medida, com os resultados esperados. Na tabela 6 estão apresentados os 4 *clusters* com os municípios que os compõem, segundo seu potencial de desenvolvimento.

É importante entender que, em cada um dos quatro grupos, os municípios se encontram em ordem crescente de distância em relação ao

centróide do grupo, e não em ordem decrescente segundo o potencial de desenvolvimento. Isto significa que quanto mais próximo ao centróide, mais bem adaptado ao grupo o município se encontra. E quanto mais afastado do centróide, menos adaptado ao *cluster*, porque seu vetor de indicadores pode estar, relativamente ao *cluster*, ou muito bom ou muito ruim – esses municípios mais afastados estão na chamada “linha de transição” entre o *cluster* em que foram inseridos e os *clusters* vizinhos. Exemplificando, e tendo por base os resultados do município nesses 3 indicadores selecionados, pode-se dizer que Macaé, alocado no cluster 2, está na área de transição para o grupo 1, enquanto Tanguá, alocado no grupo 3, fica na região de transição para o grupo 4.

Tabela 6: Os 4 clusters de municípios fluminenses, segundo seu potencial de desenvolvimento

Grupo 1:	Grupo 2:	Grupo 3:	Grupo 4:
Barra Mansa	Areal	Miracema	Silva Jardim
Nova Iguaçu	Paraíba do Sul	Saquarema	Trajano de Moraes
São Gonçalo	Maricá	Casimiro de Abreu	S. Francisco do Itabapoana
Barra do Piraí	Queimados	S. Antônio de Pádua	Sumidouro
Resende	Itatiaia	Araruama	Cardoso Moreira
Três Rios	Angra dos Reis	Rio das Ostras	Paty do Alferes
São João de Meriti	Eng. Paulo Frontin	Itaocara	Laje do Muriaé
Volta Redonda	Magé	Bom Jesus do Itabapoana	Cambuci
Duque de Caxias	Itaguaí	Natividade	Bom Jardim
Petrópolis	Campos dos Goytacazes	Cachoeiras de Macacu	Varre-Sai
Nova Friburgo	Cabo Frio	S. José do Vale do Rio Preto	Duas Barras
Pinheiral	Seropédica	Parati	S. Maria Madalena
Rio de Janeiro	Miguel Pereira	Quissamã	Rio Claro
Nilópolis	Mangaratiba	S. João da Barra	Porciúncula
Niterói	Mendes	Sapucaia	Carapebus
	Comendador Levy	Guapimirim	S. José de Ubá
	São Pedro d' Aldeia	Armação dos Búzios	São Sebastião do Alto
	Itaboraí	Aperibé	
	Cordeiro	Paracambi	
	Arraial do Cabo	Cantagalo	
	Valença	Itaperuna	
	Macuco	Piraí	
	Teresópolis	Rio das Flores	

Grupo 1:	Grupo 2:	Grupo 3:	Grupo 4:
	Porto Real Quatis Macaé Belford Roxo Vassouras Iguaba Grande	Carmo São Fidélis Rio Bonito Tanguá Italva Japeri Conceição de Macabú	

5. CONCLUSÕES

Analisando os municípios que compõem os quatro *clusters* identificados, conclui-se que o grupo 1 tendeu a concentrar os (15) municípios com maior potencial de desenvolvimento, que são justamente os da Região Metropolitana do Rio de Janeiro e os da industrializada região sul. No outro extremo, no grupo 4, ficaram os (17) municípios que apresentam atualmente menor potencial de desenvolvimento, que são sobretudo os da região norte e noroeste do Estado. Grande parte dos municípios das regiões serrana, litoral e central foi alocada nos grupos intermediários 2 e 3, e os (29) municípios do grupo 2 tenderam a apresentar melhor situação que os do grupo 3 (30 municípios).

Assim, no que se refere ao grupamento dos municípios segundo seu potencial de desenvolvimento, pode-se dizer que existe um certo grau de comparabilidade entre os resultados obtidos com a metodologia do IQM e a aplicação da análise fatorial exploratória e de *cluster*. Tomando-se, por exemplo, o caso dos municípios situados em posição extrema (tabelas 4 e 6), observa-se que 7 dentre os 15 municípios alocados no 1º *cluster* encontram-se também entre os 15 municípios mais bem classificados segundo o IQM. Por outro lado, dentre os 17 municípios do 4º *cluster*, 13 estão também entre os 17 municípios pior classificados pela Fundação CIDE. Ou seja, ao todo, nessas duas posições extremas, 62,5% dos municípios são comuns.

A questão que se coloca, portanto, é: qual seria a vantagem de utilizar uma ou outra metodologia?

Como visto, em ambos os modelos a subjetividade está presente, de maneira diferente, porém. Quando se usa o IQM, existe uma boa dose de subjetividade na entrada dos dados, ou seja, na definição e na atribuição de pesos aos indicadores e grupos. Por outro lado, quando se trabalha com a análise fatorial e de *clusters*, a subjetividade entra na interpretação dos dados, ou seja, na análise dos resultados obtidos da associação entre os indicadores e da distância entre as unidades observacionais.

Neste sentido, a aplicação da análise fatorial mostrou que bastavam apenas 15 indicadores e 3 grupos (os 3 fatores latentes) para “capturar” a idéia do potencial de desenvolvimento dos municípios. Ou seja, não era necessário um número tão grande de indicadores (38) e grupos (7) quanto o utilizado no IQM. A vantagem dessa redução no número de indicadores e de grupos é que a análise se torna mais parcimoniosa, menos passível de erros nas medidas dos dados (pois existem menos indicadores a serem incluídos) e de mais fácil interpretabilidade.

Finalmente, chama-se a atenção para três procedimentos interessantes, que foram utilizados no âmbito da análise fatorial e da análise de *clusters* neste estudo empírico:

- Eliminação de indicadores pouco representativos ou pouco válidos do(s) conceito(s) que se deseja apreender, com base na análise da matriz das correlações do conjunto dos indicadores.
- Tratamento dos dados *outliers*, que atuam como elementos perturbadores da análise e que passam a ser considerados como valores *missing*.

- Definição dos *clusters* a partir das variáveis observacionais mais carregadas em cada fator, e não de todas as variáveis que compõem o fator, como é o mais usual.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- AAKER, D., KUMAR, V. e DAY, G. *Marketing Research*. 6. ed. New York: John Wiley & Sons, Inc., 1998.
- BALASSIANO, M. *Análise Fatorial*. São Paulo: FGV, 2000. Trabalho digitado.
- CIDE, Centro de Informações e Dados do Rio de Janeiro. *IQM – Índice de Qualidade dos Municípios*. Rio de Janeiro: CIDE, 1998.
- CIDE, Centro de Informações e Dados do Rio de Janeiro. *IQM – Índice de Qualidade dos Municípios*. Disponível em: <<http://www.cide.rj.gov.br>> Acesso em: out. de 2000.
- CHRISTALLER, W. *Central Places in Southern Germany*. New Jersey: Prentice-Hall, Inc., Englewood Cliffs, 1966.
- GIRI, N. *Multivariate Statistical Analysis*. New York: Marcel Dekker, Inc. 1996.
- JOHNSON, R. e WICHERN, D. *Applied Multivariate Statistical Analysis*. 3. ed. New Jersey: Prentice Hall, 1992.
- LATIF, S. A. A análise fatorial auxiliando a resolução de um problema real de pesquisa de marketing. *Caderno de Pesquisas em Administração*, São Paulo: FEA/USP, v. 00, nº 0, 2º sem.1994.
- PERROUX, F. *L'Economie du XX ème Siècle*. 2. ed. Paris: Presses Universitaires de France, 1964.
- PNUD – Programa das Nações Unidas para o Desenvolvimento. *Relatório do Desenvolvimento Humano 1999*. Lisboa: Trinova Editora, 1999.

Anexo: Descrição dos indicadores utilizados no IQM

Código	Indicador
AER	Existência de aeroporto
ALF	Taxa de alfabetização da população de 15 anos ou mais
B24	Pontos de serviços bancários com atendimento 24 horas, em relação à raiz da população
BAN	Agências bancárias dividido pela raiz da população e multiplicado por 100
CAT	Raiz quadrada do Valor Adicionado Fiscal do Comércio Atacadista <i>per capita</i>
CES	Percentual de chefes de domicílios com, pelo menos, 2º grau completo
CON	Concessionárias de veículos no município, em relação à raiz da população, multiplicado por 100
COR	Agência dos Correios multiplicado por 2 mais posto de venda de selos dividido pela raiz da população e multiplicado por 100
CRA	Percentual de chefes de domicílios com renda superior a 20 salários mínimos
CRE	Taxa média geométrica de crescimento anual da população residente, entre 1991 e 1996
CUL	Soma dos números de cinemas, teatros, bibliotecas, dividido pela raiz da população e multiplicado por 100
CVA	Atratividade do município para o suprimento de bens de consumo em geral
DEP	Média dos depósitos bancários em agências do município por habitante
DIS	Existência de distritos, condomínios, pólos ou parques industriais
DOM	Percentual médio de domicílios com abastecimento de água adequado, com esgotamento sanitário adequado e com coleta de lixo
ENE	Consumo residencial de energia elétrica por habitante
ENS	Matrículas no ensino básico em relação à população residente em idade escolar
FER	Existência ou proximidade de linha férrea
FMU	Capacidade de investimento. Relação entre as despesas de capital com investimentos e a população
GAS	Existência de gasoduto
HOT	Leitos de hotel para cada 1.000 habitantes
INC	Pontuação pela política municipal de incentivos
INT	Provedor de acesso à INTERNET em relação ao tempo de um pulso e à população
JUS	Existência de PROCON e Defensoria pública
LEI	Leitos nas especialidades básicas em hospitais credenciados pelo SUS, para cada grupo de 1.000 habitantes
LIN	Pontuação segundo a existência de linhas de transmissão de energia elétrica
MES	Matrículas em instituições de ensino superior
ONI	Linhas intermunicipais que servem ao município dividido pela raiz da população e multiplicado por 100
OPC	Valor médio das operações de crédito em agências bancárias do município por habitante
PIB	Estimativa do PIB <i>per capita</i> (renda <i>per capita</i>) do município
PIC	Estimativa da taxa média de crescimento do PIB do município entre 1990 e 1996
PRO	Conclusões em cursos oferecidos pelo SENAC e pelo SENAI, em relação à população de 15 anos ou mais
ROD	Existência ou proximidade de rodovias de pista dupla
SEB	Existência de Balcão SEBRAE
SEG	Policiais civis e militares para cada grupo de 10.000 habitantes
TEC	Matrícula em cursos técnicos de 2º grau, em relação à população de 15 anos ou mais
TEL	Terminais telefônicos para cada grupo de 1.000 habitantes
VEI	Veículos novos (a partir de 1996) licenciados para cada grupo de 1.000 habitantes

Fonte: CIDE, 2000.