

VULNERABILIDADE SOCIOECONÔMICA EM PORTO ALEGRE

Uma abordagem territorial com base em Análise Multivariada

Carlos Eduardo Gomes Macedo*

Valéria D. Sartori Bassani**

Resumo: Aplicaram-se as técnicas de Componentes Principais e Análise de Agrupamentos a 29 variáveis do Atlas do Desenvolvimento da Região Metropolitana de Porto Alegre do ano de 2000 referentes a 163 Unidades de Desenvolvimento Humano (UDHs) do município de Porto Alegre. Com as Componentes Principais buscou-se reduzir a quantidade inicial de variáveis a apenas duas componentes que explicam 84,9% da variância total. Com a média ponderada dos escores das duas componentes construiu-se o Índice de Vulnerabilidade Socioeconômica (IVSE) e, com base neste, aplicou-se a Análise de Agrupamentos (AA). A AA, por meio do método de Ward, permitiu a visualização de 4 agrupamentos de UDHs com níveis de vulnerabilidade distintos. O georreferenciamento dos 4 agrupamentos permitiu uma melhor visualização espacial da vulnerabilidade socioeconômica do município de Porto Alegre. Com efeito, verificou-se que existiam, em 2000, 55 UDHs na situação de alta vulnerabilidade, mas com posições diversas quanto à classificação no Agrupamento de IVSE Alto. Ou seja, foi possível agrupar as UDHs com alta vulnerabilidade, mas com graus diferentes de vulnerabilidade.

Palavras-chaves: Análise Multivariada; Índice de Vulnerabilidade Socioeconômica; Agrupamentos; vulnerabilidade espacial do Município de Porto Alegre.

INTRODUÇÃO

Temos o privilégio de viver em uma época na qual a disponibilidade de informações é imensa e precisamos, na maioria das vezes, apenas de um clique no computador ou um toque em algum tipo de celular para termos acesso a elas. Não precisamos mais das enfadonhas pesquisas em enciclopédias de papel que ocupavam prateleiras inteiras e cujo manuseio só é requerido atualmente por saudosistas que insistem na cultura do papel. Mas nem tudo é maravilhoso neste novo mundo digital da informação abundante e de acesso rápido. Ter muita informação não implica em capacidade de degluti-la. Se não tivermos a capacidade de síntese e de discernimento, de pouca valia terá uma quantidade tão grande de informação.

Para que possamos usufruir a imensa quantidade de informações precisamos fazer uso de uma das características que nos diferencia dos outros animais e que tem sido utilizada há milênios: a capacidade de classificar. No século IV a.C., Aristóteles já havia construído um elaborado sistema de classificação de espécies do reino animal o qual dividia os animais em dois grupos: os que tinham (vertebrados) e os que não tinham (invertebrados) sangue vermelho. Observe-se que este sistema teve influência sobre os trabalhos posteriores elaborados por Theophrastus sobre a estrutura e a classificação das plantas. Este sistema só seria superado nos séculos XVII e XVIII (Everitt, 1980).

Vemos, portanto, como a necessidade de classificar tem estado há muito tempo entre nós e que a conseqüência imediata ao ato de classificar tem sido a capacidade de agregar/agrupar em função de alguma característica do que se pretende ter um melhor conhecimento. Ou seja, o conhecimento tem estado fortemente relacionado com a capacidade de classificar e esta, por sua vez, tem evoluído com a capacidade de agregação/agrupamento de dados e/ou informações.

Muitas técnicas novas para classificar e agregar/agrupar dados e/ou informações surgiram desde o último quarto do século XX. Mas o que tornou estas técnicas mais “populares” foi o desenvolvimento de softwares estatísticos com grandes capacidades e velocidade de cálculos. Além disto, técnicas que permitem o conhecimento exploratório e confirmatório de grande quantidade de dados e/ou informações foram desenvolvidas neste mesmo período e têm sido aplicados em várias áreas que vão da psicologia à economia. Este é caso das técnicas de Análise Multivariada tais como Análise de Componentes Principais e Análise de Agrupamentos (Cluster). A primeira permite que se descubra a estrutura subjacente (constructo) de uma grande quantidade de variáveis por meio da decomposição espectral de sua matriz de covariâncias ou correlações. O resultado é a redução das variáveis iniciais a poucas componentes que explicam a maior parte da

variância das variáveis. A segunda permite que se produzam agrupamentos (clusters) por meio de métodos estatísticos aplicados à matriz de distâncias dos valores das variáveis.

A evolução recente dos softwares estatísticos bem como o manuseio amigável dos mesmos tem possibilitado a aplicação das técnicas de análise multivariada a uma infinidade de dados e/ou informações. Nesse sentido, pensamos em aplicar tais técnicas a uma seleção de variáveis oriundas do Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre, doravante Atlas da RMPA. Lançado em dezembro de 2008, o Atlas da RMPA tem mais de uma centena de variáveis distribuídas em vários níveis geográficos que vão do nível do município, passando pelas Regiões de Orçamento Participativo, às Unidades de Desenvolvimento Humano, doravante UDHS. Como o objetivo fundamental do Atlas da RMPA era conhecer a heterogeneidade das regiões das cidades da RMPA, os mais de 100 indicadores foram gerados para 330 UDHS. Com as UDHS podem-se ver com maior riqueza de informações as várias cidades dentro de uma mesma cidade. Ao mesmo tempo em que esta riqueza de informações georreferenciadas permitiu a focalização das políticas públicas, ela também gerou a percepção de que era possível o agrupamento de UDHS em função da proximidade dos valores das principais variáveis. Com efeito, o presente trabalho se propõe a aplicar as técnicas de Componentes Principais e Análise de Agrupamento a um conjunto de 29 variáveis das 163 UDHS do município de Porto Alegre referentes ao ano de 2000 com vistas a agrupar as regiões mais semelhantes. Por meio da aplicação de Análise de Agrupamentos teremos condições de gerar agregações de UDHS que nos permitam visualizar as segregações espaciais não perceptíveis em indicadores que sumarizam a cidade como um todo ou, o contrário, a repartem em unidades tão pequenas que acabam por diluir em demasia a focalização das políticas públicas.

1. METODOLOGIA

Muitas são as técnicas de Análise Multivariada, e descrevê-las de forma exaustiva é menos importantes do que compreender o fundamento das mesmas. Qualquer que seja a técnica utilizada, seu fundamento é o cálculo do que se chama Variável Estatística. Variável Estatística é um tipo de constructo, uma variável latente, uma combinação linear de variáveis com pesos empiricamente determinados. Ao pesquisador cabe a escolha das variáveis; à técnica multivariada, o cálculo dos pesos (Hair et al., 2005). Conforme Hair et al. (2005), podemos definir matematicamente uma Variável Estatística como:

$$VE = w_1 X_1 + w_2 X_2 + w_3 X_3 + \dots + w_n X_n$$

Onde:

VE = Variável Estatística

X_n = Variável observada

w_n = Pesos determinados pela técnica multivariada

Por conta desta capacidade de geração de variáveis estatísticas com grande capacidade de resumo de dados, a análise multivariada tem contribuído sobremaneira para a investigação científica quando se deseja a redução ou simplificação estrutural de dados, classificação e arranjo, investigação de dependência entre variáveis, predição e construção de hipóteses e testes (Johnson e Wichern, 2002).

No presente estudo, faremos uso inicialmente da Análise de Componentes Principais para a prospecção de dois componentes capazes de explicar 84,9% da variância das 29 variáveis referentes a 163 UDHS da cidade de Porto Alegre. Após o cálculo das duas componentes, será possível calcular os escores das componentes para cada UDH. Como os escores representam um tipo de pontuação que cada UDH tem em cada componente, é possível seu agrupamento por meio da Análise de Agrupamentos. Nesse sentido, podemos ordenar as UDHS gerando grupos homogêneos e, assim, termos uma idéia mais precisa da distribuição geográfica das UDHS.

1.1. Técnicas Estatísticas

Nesta seção apresentamos de forma sintética os conceitos das duas técnicas de análise multivariada utilizadas. A compreensão das técnicas de análise multivariada exige o conhecimento dos conceitos elementares da álgebra matricial. Contudo, não é objetivo deste trabalho a exposição de conceitos da álgebra matricial. Quando muito faremos alguns comentários com vistas a ajudar na elucidação de alguns dos conceitos fundamentais para a compreensão das técnicas aqui utilizadas.

1.1.1. Análise de Componentes Principais

A Análise de Componentes Principais, doravante ACP, é um tipo de análise fatorial na qual os fatores comuns não têm a mesma importância que têm para a Análise de Fatores Comuns (Análise Fatorial propriamente dita, doravante AFC). Um dos conceitos relevantes para entender esta diferença é o de Comunalidade. A Comunalidade é a variância total que uma variável compartilha com todas as outras variáveis. A variância total pode ser dividida em três tipos: a) variância comum, ou

seja, aquela que uma variável compartilha com todas as outras variáveis; b) variância específica, ou seja, aquela que é associada a apenas uma variável; c) variância do erro, ou seja, aquela que se deve à não-confiabilidade no processo de agrupamento de dados, erro de medida ou em uma componente aleatória no fenômeno medido (Hair et al., 2005). Para os propósitos de nosso trabalho, basta que entendamos a distinção entre as variâncias comuns e específicas. Outra diferença entre ACP e AFC está em que a utilização da ACP não exige o atendimento ao pressuposto de que as variáveis tenham distribuição normal multivariada (Johnson e Wichern, 2002). Por fim, uma diferença marcante está em ser a análise confirmatória ou exploratória. Se desejarmos confirmar algum modelo ou classificação teórica, será mais apropriado o uso de AFC. Caso o objetivo seja a exploração de dados com vistas a prospectar as relações lineares entre as variáveis e as componentes, deveremos fazer uso da ACP. Resumindo, a ACP é um procedimento mais simples do que a AFC, pois não necessita do atendimento ao pressuposto de distribuição normal multivariada e nem de um modelo teórico. Com ACP parte-se dos dados e trabalha-se na direção de um modelo hipotético (Fachel, 1982). Em resumo, ACP visa à explanação da estrutura da variância-covariância de um conjunto de variáveis por meio de poucas combinações lineares destas variáveis. Seus objetivos gerais são redução e interpretação de dados. Algebricamente, componentes principais são combinações lineares particulares de p variáveis aleatórias X_1, X_2, \dots, X_p . Geometricamente, estas combinações lineares representam a seleção de um novo sistema de coordenadas obtido pela rotação do sistema original com X_1, X_2, \dots, X_p como eixos coordenados. Os novos eixos representam as direções com máxima variabilidade e oferecem uma simples e mais parcimoniosa descrição da estrutura da covariância (Johnson e Wichern, 2002).

Tomando-se m componentes e p variáveis ($m * p$), a compreensão da ACP como uma combinação linear das variáveis originais pode ser melhor expressa com a seguinte formulação matemática sugerida por Reis (2001):

$$CP_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$$

$$CP_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p$$

...

$$CP_m = a_{1m}X_1 + a_{2m}X_2 + \dots + a_{pm}X_p$$

Onde:

CP_m = Componente Principal m

a_{pm} = Carga da variável X_p da componente m

Os passos gerais para o cálculo das componentes principais são: 1) escolha das variáveis que farão parte da análise (se tivéssemos optado pela AFC, a escolha envolveria verificar o pressuposto de distribuição normal multivariada); b) geração da matriz de covariâncias ou de correlações (sugere-se a matriz de correlação por afastar os problemas de diferenças de escalas entre as variáveis); c) decomposição espectral da matriz de correlações (cálculo dos pares de autovalor e autovetor); d) escolha das componentes com maior variância; e) aplicação de um método de rotação que facilite a interpretação da correlação entre as variáveis e as componentes (rotação é apenas um artifício que produz nos eixos iniciais uma mudança de tal forma a aproximar as variáveis das componentes, aumentando o poder de interpretação e identificação das componentes).

Por ser a decomposição espectral da matriz de correlações um dos passos mais importantes na geração das componentes principais (também dos fatores da AFC), desenvolveremos as operações matemáticas envolvidas no cálculo dos pares de autovalores e autovetores¹. Conforme Johnson e Wichern (2002), matrizes quadradas (matrizes de correlações são matrizes quadradas) são melhor entendidas em termos de quantidades chamadas autovalores e autovetores. Fundamental, então, torna-se saber como extrair estas quantidades. Façamos \mathbf{A} ser uma matriz quadrada $k \times k$ e \mathbf{I} ser uma matriz identidade $k \times k$. Então os escalares $\lambda_1, \lambda_2, \dots, \lambda_k$ que satisfazem a equação polinomial $|\mathbf{A} - \lambda\mathbf{I}| = 0$ são chamados de autovalores (ou raízes características da matriz \mathbf{A}). A equação $|\mathbf{A} - \lambda\mathbf{I}| = 0$ (como uma função de λ) é chamada de equação característica. Ainda seguindo a exposição de Johnson e Wichern (2002), observemos os seguintes exemplos de como calcular autovalores e autovetores.

$$\text{Seja } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}$$

$$\begin{aligned} \text{Então } |\mathbf{A} - \lambda\mathbf{I}| &= \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 - \lambda & 0 \\ 1 & 3 - \lambda \end{bmatrix} = (1 - \lambda)(3 - \lambda) = 0 \end{aligned}$$

implica que há duas raízes, $\lambda_1 = 1$ e $\lambda_2 = 3$. Os autovalores de \mathbf{A} são 1 e 3.

Tomemos a mesma matriz \mathbf{A} de dimensão $k \times k$ com λ sendo um de seus autovalores. Se \mathbf{x} for um vetor não nulo ($\mathbf{x} \neq \mathbf{0}$), tal que $\mathbf{Ax} = \lambda\mathbf{x}$, então \mathbf{x} é chamado de autovetor (vetor característico) da matriz \mathbf{A} associado ao autovalor λ . Dando seguimento ao exemplo de Johnson e Wichern (2005), veremos como calcular os autovetores associados aos autovalores $\lambda_1 = 1$ e $\lambda_2 = 3$. Para tanto basta que se resolvam as seguintes equações:

$$\text{Para } \lambda_1=1, \text{ teremos} \quad \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{A} \quad \mathbf{x} = \lambda_1 \quad \mathbf{x}$$

$$\text{Para } \lambda_2=3, \text{ teremos} \quad \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{A} \quad \mathbf{x} = \lambda_2 \quad \mathbf{x}$$

Efetuada a multiplicação das matrizes da primeira equação, resulta:

$$\begin{aligned} x_1 &= x_1 \\ x_1 + 3x_2 &= x_2 \Rightarrow x_1 = -2x_2 \end{aligned}$$

Como há muitas soluções para x_1 e x_2 , podemos escolher arbitrariamente $x_2 = 1$ implicando em $x_1 = -2$. O autovetor correspondente ao autovalor 1 será:

$$\mathbf{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

Efetuada a multiplicação das matrizes da segunda equação, resulta:

$$\begin{aligned} x_1 &= 3x_1 \\ x_1 + 3x_2 &= 3x_2 \Rightarrow x_1 = -2x_2 \end{aligned}$$

Escolhendo arbitrariamente $x_2 = 1$, teremos $x_1 = 0$. O autovetor correspondente ao autovalor 3 será:

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Para entender a importância da decomposição espectral e seguindo a exposição de Reis (2001), façamos uma matriz \mathbf{A} ter autovetores $[\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_p]$ e vetor de componentes principais \mathbf{Y} . Então $\mathbf{Y} = \mathbf{A}' \mathbf{X}$, implicando em que a matriz variância/covariância de \mathbf{Y} seja $\text{Var}[\mathbf{Y}] = \mathbf{A}' \Sigma \mathbf{A} = \Lambda$, sendo

$$\Lambda = \begin{bmatrix} \lambda_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \lambda_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \lambda_2 \end{bmatrix}$$

Como \mathbf{A} é uma matriz ortogonal ($\mathbf{A} \cdot \mathbf{A}' = \mathbf{I}$), então $\Sigma = \mathbf{A} \cdot \Lambda \cdot \mathbf{A}$. Sendo os autovalores interpretados como as variâncias das respectivas componentes principais, a soma destas variâncias será:

$$\sum_{j=1}^p \text{Var}[y_j] = \sum_{j=1}^p \lambda_j = \text{Tr}(\Lambda)$$

Onde $\text{Tr}(\Lambda)$ é o traço da matriz Λ , ou seja, a soma da diagonal principal desta matriz.

Concluimos, portanto, que o traço da matriz Λ corresponde à soma dos autovalores λ que, por sua vez, são iguais a variância total. Disto decorre que um valor particular de λ , por exemplo λ_1 , ou um soma como $\lambda_1 + \lambda_2$, podem ser considerados uma fração da soma total dos autovalores, o que significa termos a condição de saber quanto da variância total pode ser devida a uma componente ou a uma soma de componentes. Ou seja, a j -ésima componente explica:

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j}$$

da variação total original. Da mesma forma as primeiras m componentes explicam:

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j}$$

da variância total.

Chegamos, portanto, ao ponto central para a compreensão da importância de se calcular os pares de autovalor e autovetor, pois a ACP, como já enfatizamos, tem como objetivo reduzir a quantidade inicial de variáveis a um pequeno número de componentes (fatores na AFC) que respondem pela maior parte da variância total.

A ACP tem sido utilizada como uma etapa intermediária para outras técnicas de análise multivariada. Tendo em vista a possibilidade de redução considerável no número de componentes ou fatores que ACP e a AFC permitem, estas técnicas podem servir como geradoras de uma base sobre qual se aplique, por exemplo, a Análise de Agrupamentos. Com efeito, como é nosso objetivo prospectar uma possível distribuição espacial das UDHs tendo como base uns poucos, mais significativos indicadores, passaremos, na próxima seção, à explanação sintética sobre a Análise de Agrupamentos.

1.1.2. Análise de Agrupamentos (Cluster)

Conforme Everit (1980), devido ao uso em campos tão diversos como psicologia, zoologia, biologia, botânica, sociologia, inteligência artificial e recuperação de informações, a técnica de Análise de Agrupamentos tem sido referida como Análise de Cluster, Q-Análise, Tipologia, arranjo, agrupamento, classificação etc. Independente do modo como tem sido chamada esta técnica, seu objetivo se refere à necessidade de, dado um número de objetos ou indivíduos, cada um descrito por um conjunto de medidas, organizar um esquema de classificação para arranjar os objetos em número de classes tais que os objetos dentro das classes sejam similares em algum aspecto e diferentes dos objetos das outras classes (Everit, 1980). Em suma, trata-se de maximizar a homogeneidade interna dos elementos (dentro do grupo) e heterogeneidade externa (entre os grupos). “Assim, se a classificação for bem sucedida, os objetos dentro dos agrupamentos estarão próximos quando representados graficamente e diferentes agrupamentos estarão distantes (Hair et al., 2005)”. Doravante nos referiremos à Análise de Agrupamentos como AA.

Da mesma forma como procedemos em relação à ACP, faremos uma breve descrição dos passos para a obtenção de agrupamentos ou clusters. Dos cinco passos sugeridos por Reis (2001) para uma boa aplicação da AA, desenvolveremos os quatro a seguir. 1) seleção de indivíduos ou de

uma amostra a serem agrupados; 2) definição de um conjunto de variáveis a partir das quais será obtida a informação necessária ao agrupamento dos indivíduos; 3) definição da medida de semelhança ou distância entre cada dois indivíduos; 4) escolha de um critério de agregação ou desagregação dos indivíduos, isto é, a definição de um algoritmo de partição/classificação.

Nos primeiro e segundo passos fica patente o quanto a AA tem um caráter muitas vezes subjetivo, pois nesta etapa é preciso ter claro que a escolha das variáveis é uma categorização de dados na qual não há um procedimento matemático ou estatístico e sim o julgamento do pesquisador sobre a relevância das variáveis para o propósito da classificação. O mesmo pode ser dito sobre a escolha dos objetos ou indivíduos. Também não há, em geral, uma base teórica para determinar o número de variáveis a serem mensuradas, e o problema deve ser abordado empiricamente (Everit, 1980).

Tendo em conta estas dificuldades em ter uma base mais matemática ou estatística para a escolha das variáveis, dos objetos ou indivíduos e do número de variáveis, Everit (1980) sugere como forma de superar estes problemas a aplicação da ACP aos dados originais e, então, o uso dos escores das primeiras componentes como variáveis de entrada para processo de agrupamento. Adiantamos que foi este o caminho escolhido para o presente trabalho, e os procedimentos bem como resultados e análises serão tratados na seção 3.

O passo 3 é aquele no qual precisamos definir uma medida de semelhança (similaridade) ou distância² entre os indivíduos. Tendo em vista que faremos uso do critério de agregação dos indivíduos conhecido como Ward e de que, em geral, a bibliografia consultada sugere que quando do uso deste critério seja mais apropriado o uso da Distância Euclidiana ao Quadrado, doravante DEQ, não vamos nos deter na exposição das outras medidas de distância. Apenas como referência, arrolamos as outras medidas de distância: 1) distância Euclidiana; 2) distância Absoluta; 3) distância de Minkowski; 4) distância de Mahalanobis; 5) distância de Chebishev.

A DEQ entre dois indivíduos (*i* e *j*) é definida como o somatório dos quadrados das diferenças entre os valores *i* e *j* para todas as variáveis. Ou seja:

$$d_{ij} = \sum_{k=1}^p (X_{ik} - X_{jk})^2$$

Onde X_{ik} é valor da *k*-ésima variável para o *i*-ésimo indivíduo.

No quarto passo definimos qual o critério de agregação dos indivíduos ou objetos. Como já referido, faremos uso do critério ou método conhecido como Ward. Apenas como referência,

arrolamos os outros critérios ou métodos: 1) método do vizinho do mais longo; 2) método do vizinho mais afastado; 3) método das médias dos grupos; 4) método do centróide.

Segundo Everit (1980), J. H. Ward, em 1963, sugeriu que em qualquer estágio, a perda de informação resultante do agrupamento de indivíduos pode se mensurada pela soma total dos quadrados dos desvios de qualquer ponto em relação à média do grupo ao qual ele pertence. Se para cada grupo k , fizermos SQD_k ser a Soma dos Quadrados dos Desvios de cada item em relação à média do grupo (centróide), então a existência de K grupos define SQD como a soma de SQD_k , ou seja:

$$SQD = SQD_1 + SQD_2 + \dots + SQD_k$$

A cada etapa da análise, a fusão de cada possível par de grupos é considerada e dois grupos, cuja fusão resulta em um mínimo de incremento na soma dos quadrados dos desvios (perda mínima de informação), são combinados. Inicialmente, cada grupo consiste de um simples item, e, se existirem N itens, $SQD = 0$, com $k = 1, 2, \dots, N$, então $SQD = 0$. No extremo oposto, quando todos os grupos são combinados em um simples grupo de N itens, o valor da SQD será dada por:

$$SQD = \sum_{j=1}^N (X_j - \bar{X})(X_j - \bar{X})$$

Onde X_j é a medida multivariada associada com o j -ésimo item e \bar{X} é a média de todos os itens. Os resultados da aplicação do método de Ward podem ser visualizados por meio de um diagrama chamado dendrograma, no qual o eixo vertical mostra os valores da SQD para cada fusão ocorrida (Johnson e Wichern, 2002).

1.2. Base de Dados

A escolha das variáveis sobre as quais se aplicou as ACP e AA tem como fonte os dados das 163 UDHS do município de Porto Alegre para o ano de 2000. Estes dados foram extraídos do Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre elaborado por meio de convênio entre o PNUD (Programa das Nações Unidas para o Desenvolvimento), a Fundação João Pinheiro, a Prefeitura Municipal de Porto Alegre (cuja coordenação coube aos autores do presente artigo) e a METROPLAN (Fundação Estadual de Planejamento e Desenvolvimento e Regional do Rio Grande do Sul).

De um montante de mais de 100 variáveis, foram escolhidas 29 em função de sua recorrência em outros estudos de mesmo cunho e também em função de critérios estatísticos tais como:

1. Correlações significativas;
2. Utilização da matriz anti-imagem para verificar as Medidas de Adequação da Amostra Kaiser-Meyer-Olkin cujos resultados foram:
 - 2.1. 0,94 para a matriz como todo (situação classificada como “Muito Boa” conforme Reis, 2001);
 - 2.2. acima de 0,8 para todas as variáveis individuais (situação classificada como “Boa” conforme Reis, 2001);
3. Teste de Bartlett³ para testar a hipótese de ser a matriz de correlações uma matriz identidade, resultando na rejeição desta hipótese e mostrando ser a matriz de correlações apropriada para a aplicação da ACP;
4. Experiência dos técnicos do Observatório da Cidade de Porto Alegre atestada pela confecção de vários estudos diagnósticos sobre a situação socioeconômica do município, em especial, o Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre⁴ e o Mapa da Inclusão e Exclusão Social de Porto Alegre.

Tabela 1 mostra os resultados da aplicação dos testes KMO e Bartlett sobre as 29 variáveis.

| Tabela 1 | | |
|---|---------------------|----------|
| Testes de KMO e Bartlett | | |
| Medida de Adequação da Amostra Kaiser-Meyer-Olkin | | 0,94 |
| Teste de Esfericidade de Bartlett | Aprox. Qui-Quadrado | 11432,77 |
| | gl | 406 |
| | Sig. | 0 |

Fonte: Resultados da Análise ACP

2. RESULTADOS E ANÁLISE

Nesta seção apresentamos os resultados da aplicação das técnicas multivariadas às 163 UDHs medidas conforme as 29 variáveis escolhidas. O primeiro passo foi a aplicação da ACP às 29 variáveis. Como resultado obtivemos 2 componentes principais que responderam por mais de 84,9% da variância total. Sobre estas duas componentes foi aplicada a AA. Diante das opções apresentadas pela aplicação da AA, optamos por uma linha de corte na etapa 159 na qual o

Coefficiente de Aglomeração foi de 0,623 implicando em 4 agrupamentos. A Tabela 2 mostra a relação dos códigos utilizados para identificar as variáveis.

| Tabela 2 | |
|---|---|
| Código e identificação das variáveis¹ | |
| Código | Identificação |
| X01 | % de pessoas que vivem em domicílios com banheiro e água encanada |
| X02 | Mortalidade até cinco anos de idade (por mil nascidos vivos) |
| X03 | Razão de dependência |
| X04 | % de pessoas que freqüentam o ensino médio em relação à população de 15 a 17 anos |
| X05 | % de mulheres de 25 anos ou mais analfabetas |
| X06 | % de pessoas com mais de 25 anos analfabetas |
| X07 | Média de anos de estudo das pessoas de 25 anos ou mais de idade |
| X08 | Índice de envelhecimento |
| X09 | % de pessoas que freqüentam o fundamental em relação à população de 7 a 14 anos |
| X10 | % de pessoas entre 18 e 24 anos analfabetas |
| X11 | Renda per Capita |
| X12 | % de mulheres responsáveis pelo domicílio com renda até 1 SM ² |
| X13 | % de homens responsáveis pelo domicílio sem rendimento |
| X14 | % de mulheres responsáveis pelo domicílio sem rendimento |
| X15 | Renda média dos homens responsáveis pelo domicílio |
| X16 | Renda média das mulheres responsáveis pelo domicílio |
| X17 | % de pobres |
| X18 | % de pessoas que vivem em domicílios com computador |
| X19 | Média de anos de estudo das mulheres responsáveis pelo domicílio |
| X20 | Média de anos de estudo dos homens responsáveis pelo domicílio |
| X21 | % das pessoas ocupadas com rendimento do trabalho de 1 a 2 SM |
| X22 | Rendimento médio mensal das pessoas ocupadas |
| X23 | % das pessoas ocupadas com rendimento do trabalho de 2 a 3 SM |
| X24 | Taxa de desemprego da população de 25 a 59 anos |
| X25 | Taxa de desemprego da população de 15 anos e mais |
| X26 | Taxa de participação da PEA na PIA de 15 anos e mais |
| X27 | Taxa de participação da PEA na PIA de 18 a 24 anos |
| X28 | % de ocupados com fundamental incompleto |
| X29 | % das pessoas ocupadas com rendimento do trabalho inferior a 1 SM |

Fonte: Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre

Nota 1: Todos os dados se referem ao ano de 2000

Nota 2: SM significa Salários Mínimos

2.1. Aplicação da Análise de Componentes Principais

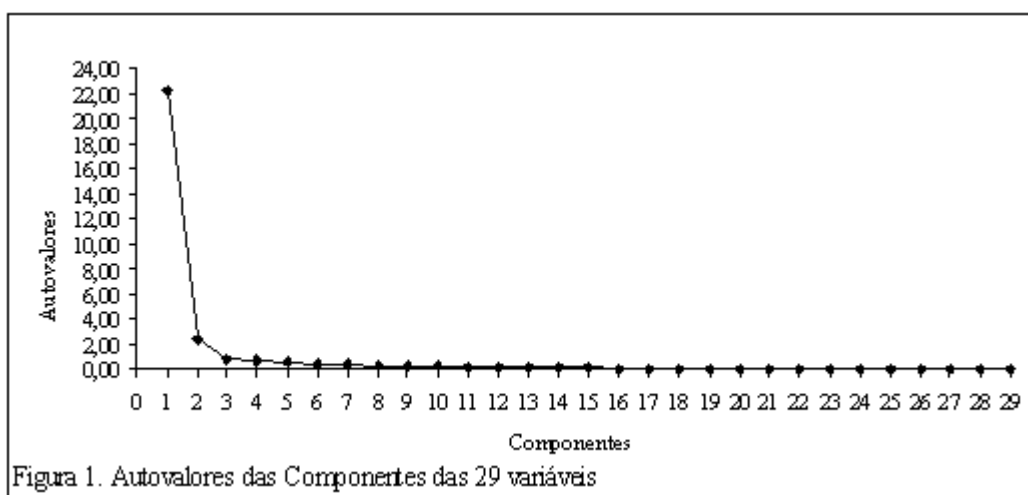
Como já referido, o objetivo fundamental da ACP é a decomposição espectral da matriz de correlações, pois assim são gerados os autovalores e autovetores. Como os autovalores correspondem às variâncias de cada componente principais, verificamos que apenas 2 componentes acabaram respondendo por 84, 9% da variância total. Um dos critérios de escolha do número de componentes a serem considerados para efeitos de compreensão da estrutura dos dados bem como para a aplicação da AA é a escolha das componentes que sejam maiores do que 1. Isto equivale a dizer que devemos escolher as raízes da equação polinomial que são maiores do que a unidade. Esta é uma fase importante, pois precisamos ter confiança na escolha do número de componentes. Será sobre estas que a tentativa de agrupamento será feita. De uma boa escolha das componentes dependem os resultados do trabalho como um todo. Outro critério que pode auxiliar na escolha das

componentes é a confecção de um gráfico onde são plotados os autovalores de cada componente e no qual é possível visualizar com bastante clareza a importância de cada componente. A Tabela 3 abaixo mostra apenas as componentes com autovalores maiores do que 1. Abaixo da Tabela 3, a Figura 1 mostra o gráfico dos autovalores para todas as componentes.

Tabela 3
Variância Explicada pelas 2 Componentes Principais com Rotação Varimax

| Componentes | Autovalores Iniciais | | | Extração da Soma dos Quadrados das Cargas | | | Rotação Varimax da Soma dos Quadrados da Cargas | | |
|-------------|----------------------|----------------|-------------|---|----------------|-------------|---|----------------|-------------|
| | Total | % da Variância | Acumulada % | Total | % da Variância | Acumulada % | Total | % da Variância | Acumulada % |
| 1 | 22,257 | 76,75 | 76,75 | 22,257 | 76,75 | 76,75 | 15,979 | 55,10 | 55,10 |
| 2 | 2,364 | 8,15 | 84,90 | 2,364 | 8,15 | 84,90 | 8,643 | 29,80 | 84,90 |

Fonte: Resultados da ACP



Percebemos pela combinação dos dois critérios que de fato temos uma boa parte da variância total explicada por apenas duas componentes, pois o gráfico dos autovalores plotados para cada componente não deixa dúvidas de que, a partir da componente 3 (inclusive), a curva se confunde com o eixo das abscissas. Antes de analisarmos as correlações (cargas fatoriais) entre as variáveis originais e as duas componentes principais, precisamos compreender o que significa a coluna Rotação Varimax da Soma dos Quadrados das Cargas da Tabela 3. A Rotação Varimax é um artifício geométrico utilizado para melhorar a interpretação da relação entre as cargas de cada variável e a componente. Dito de outra forma, o artifício consiste em fazer uma rotação nos eixos coordenados, que nada mais são que as próprias componentes, permitindo uma melhor distribuição das cargas em relação às componentes selecionadas. Após a rotação, as cargas ficam melhor distribuídas entre as duas componentes sem, contudo, alterar o percentual de variância explicada

pelas mesmas, ou seja, 84,9%. Existem outros tipos de rotação. Entretanto, escolhemos a rotação Varimax por manter os eixos ortogonais o que, em outras palavras, significa manter as componentes independentes.

Quanto aos autovetores, sua importância está em que eles representam a carga de cada variável em cada componente. Em outras palavras, os autovetores representam as correlações entre as variáveis e as componentes. A Tabela 4 abaixo mostra as correlações (cargas) das variáveis originais com a componente 1. Esperamos que esta tabela nos auxilie na identificação de características que permitam dar um nome a Componente 1.

| Tabela 4 | | |
|--|---------------------|--------------|
| Variáveis originais com as mais altas correlações (cargas) com a Componente 1 | | |
| Variáveis | COD Variável | Carga |
| % de pobres | X17 | 0,9517 |
| % das pessoas ocupadas com rendimento do trabalho inferior a 1 SM | X29 | 0,8941 |
| Mortalidade até cinco anos de idade (por mil nascidos vivos) | X02 | 0,8791 |
| Razão de dependência | X03 | 0,8652 |
| % de ocupados com fundamental incompleto | X28 | 0,8608 |
| % de pessoas ocupadas com rendimento do trabalho de 1 a 2 SM | X21 | 0,8409 |
| % de pessoas com mais de 25 anos analfabetas | X06 | 0,8348 |
| % de mulheres de 25 anos ou mais analfabetas | X05 | 0,8336 |
| Taxa de desemprego da população de 25 a 59 anos | X24 | 0,7984 |
| % de pessoas entre 18 e 24 anos analfabetas | X10 | 0,7939 |
| Taxa de desemprego da população de 15 anos e mais | X25 | 0,7791 |
| % de mulheres responsáveis pelo domicílio com renda até 1 SM | X12 | 0,7698 |
| % de mulheres responsáveis pelo domicílio sem rendimento | X14 | 0,7661 |
| % de homens responsáveis pelo domicílio sem rendimento | X13 | 0,7425 |
| Índice de envelhecimento | X08 | -0,7119 |
| Média de anos de estudo das mulheres responsáveis pelo domicílio | X19 | -0,7203 |
| Média de anos de estudo dos homens responsáveis pelo domicílio | X20 | -0,7300 |
| Média de anos de estudo das pessoas de 25 anos ou mais de idade | X07 | -0,7481 |
| % de pessoas que frequentam o ensino médio em relação à população de 15 a 17 anos | X04 | -0,9072 |
| % de pessoas que vivem em domicílios com banheiro e água encanada | X01 | -0,9408 |

Fonte: Resultados da ACP

Como se percebe não há o predomínio de um tipo de variável que nos permita sermos categóricos quanto à denominação da Componente 1. O que vemos é uma combinação de variáveis econômicas, de nível de ensino, de saneamento básico e de saúde. A rigor temos apenas uma variável de saúde e uma de saneamento básico. Contudo a importância destas duas variáveis é atestada por suas altas correlações com a Componente 1. A variável de saúde, além de ter uma alta e positiva correlação com a Componente 1, corrobora o estudo de Possoli (1984), o qual verificou, utilizando o mesmo tipo de análise multivariada que estamos utilizando, que a mortalidade até cinco

anos de idade apresentou a maior carga fatorial no Índice do Nível de Saúde por ele proposto tendo com base dados oriundos da Secretaria de Saúde do Rio Grande do Sul para o ano de 1979. A Componente 1 parece ser algum tipo de indicador de vulnerabilidade, pois ela apresentou correlações positivas e altas com variáveis como % de pobres, % de pessoas de pessoas ocupadas com rendimento do trabalho inferior a 1SM, Mortalidade até 5 anos de idade etc., e correlações negativas e altas com variáveis como % de pessoas que vivem em domicílios com banheiro e água encanada, % de pessoas que frequentam o ensino médio em relação à população de 15 a 17 anos, Média de anos de estudo das pessoas com 25 anos ou mais de idade etc.

A Tabela 5 abaixo mostra as correlações (cargas) das variáveis originais com a componente 2. Da mesma forma que a Tabela 4, esperamos que a Tabela 5 nos ajude a denominar a Componente 2. Frisamos que tanto para a Componente 1 quanto para a Componente 2, selecionamos apenas as correlações (cargas fatoriais) que são maiores do que 0,7. Como pretendemos aplicar a AA sobre os escores das duas componentes, estamos preocupados nesta fase apenas em termos uma ideia geral da estrutura das componentes.

Tabela 5
Variáveis originais com as mais altas correlações (cargas) com a Componente 2

| Variáveis | COD Variável | Carga |
|---|--------------|---------|
| Taxa de participação da PEA na PLA de 18 a 24 anos | X27 | 0,8775 |
| % de pessoas que frequentam o fundamental em relação à população de 7 a 14 anos | X09 | 0,8012 |
| % das pessoas ocupadas com rendimento do trabalho de 2 a 3 SM | X23 | 0,7364 |
| % de pessoas que vivem em domicílios com computador | X18 | -0,7388 |
| Renda média das mulheres responsáveis pelo domicílio | X16 | -0,7752 |
| Rendimento médio mensal das pessoas ocupadas | X22 | -0,8018 |
| Renda média dos homens responsáveis pelo domicílio | X15 | -0,8035 |
| Renda per Capita | X11 | -0,8060 |

Fonte: Resultados da ACP

Ao contrário da Componente 1, a Componente 2 sugere que as variáveis com maiores correlações (cargas fatoriais) são indicativas das características econômicas. A variável % das pessoas que frequentam o fundamental em relação à população de 7 a 14 anos não se enquadraria na categoria econômica em sentido estrito, contudo não deixa de ter importância se considerarmos que o capital escolar pode ser compreendido com um requisito para a capacidade de inserção qualificada no mercado de trabalho bem como para a busca de níveis de remuneração melhores a serem alcançados no futuro. Também a variável % das pessoas que vivem em domicílios com computador pode ser vista pela ótica econômica geral se a consideramos como a expressão da capacidade aquisitiva. Com efeito, a Componente 2 pode ser denominada provisoriamente como um índice de nível econômico. Como este índice está correlacionado negativamente com variáveis que medem o

rendimento e domicílios com computador, parece-nos que ele indica algum tipo de vulnerabilidade econômica. Ademais, podemos considerar a Taxa de participação da PEA (População Econômica Ativa) na PIA (População em Idade Ativa) de 18 a 24 anos como um tipo de pressão da população jovem que ingressa no mercado de trabalho e que exige uma capacidade maior de geração de postos de trabalho. Encerramos esta seção com a Figura 2 a qual mostra as cargas das variáveis com as 2 componentes após a rotação Varimax.

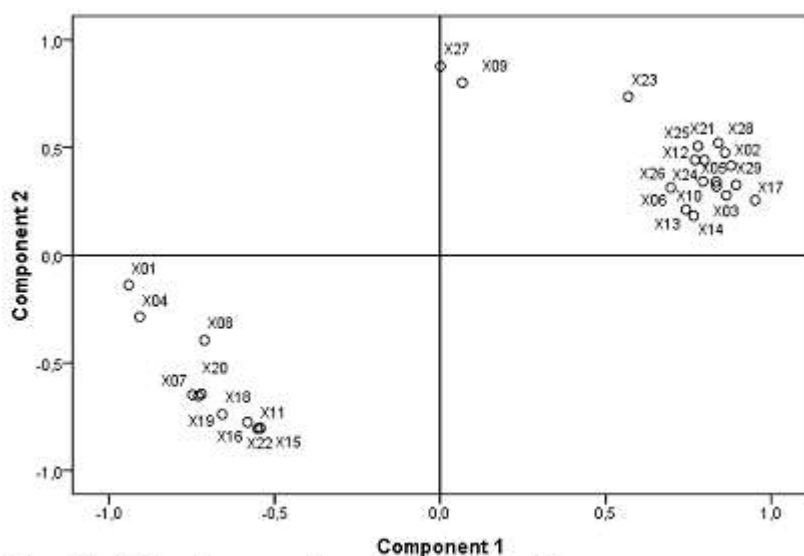


Figura 2 - Gráfico das cargas das variáveis nas duas Componentes no espaço rotacionado

2.2. Aplicação da Análise de Agrupamentos

Como já referido na seção metodológica, a ACP é, em geral, utilizada como uma etapa intermediária para a aplicação de outras técnicas multivariadas. Com os resultados que temos até esta etapa, precisamos decidir sobre que base iremos aplicar AA. Estudos recorrentes bem como manuais de análise multivariada têm sugerido que se aplique a AA sobre os escores das componentes, ou seja, sobre o valor que os objetos ou indivíduos teriam se calculássemos os escores por meio da aplicação da transformação linear das variáveis originais. Contudo, temos a Componente 1 com 55,10% e a Componente 2 com 29,80% de explicação da variância total após a rotação Varimax. Nesse sentido, teríamos que escolher sobre qual componente aplicar a AA. Devido à importância das duas componentes e tendo em vista a possibilidade de geração de agrupamentos que reflitam de fato esta importância, sugerimos calcular a média ponderada dos escores, utilizando como pesos os autovalores (variâncias) das duas componentes. Para facilitar a compreensão da média ponderada transformamos os resultados em um índice calculado da seguinte forma:

$$IVSE_{UDHi} = \frac{(EMi - EM \text{ min})}{(EM \text{ max} - EM \text{ min})}$$

Onde:

$IVSE_{UDHi}$ = Índice de Vulnerabilidade Socioeconômica da UDH i

EMi = Escore Médio da UDH i

$EM \text{ max}$ = Escore Médio Máximo

$EMi \text{ min}$ = Escore Médio Mínimo

Por conta desta transformação dos escores médios de cada UDH, a interpretação será bastante facilitada, pois o IVSE irá variar entre 0 e 1. Sem a construção do IVSE, a interpretação ficaria bem mais difícil, pois não temos ideia do que pode significar uma vulnerabilidade de Escore Médio Ponderado com sinal negativo, ou seja, de pouco nos adianta saber se uma vulnerabilidade é negativa ou positiva. A confusão que o uso dos Escores Médios Ponderados como informação final causaria seria muito grande. Observe que o denominador do IVSE é a diferença entre os valores máximo e mínimo dos escores, ou seja, a unidade de comparação se origina nos próprios escores médios. Em outras palavras, é como se criássemos uma régua com variação contínua entre os valores máximo e mínimo dos Escores Médios Ponderados. Nesta régua, a UDH com valor igual a 0 terá IVSE Baixo, pois ela está situada na posição inicial da régua. No extremo oposto, estará situada a UDH de valor igual a 1, tendo IVSE Alto. Como o IVSE é contínuo, será possível verificar as heterogeneidades dentro cada agrupamento. Dito de outra forma, poderemos verificar dentro do agrupamento de IVSE Alto, quais UDHs estão em situação de vulnerabilidade extrema. É por estes motivos que arrolamos, e também em função das variáveis apresentarem escalas diferentes, que muitos pesquisadores têm recorrido à construção de índices. A Tabela 6 mostra as 10 primeiras e as 10 últimas UDHs em função de seus valores no IVSE.

| UDHs | Escore Médio | IVSE | Classificação |
|--|--------------|-------|---------------|
| <i>FLORESTA-Loteamento Santa Terezinha</i> | 1,104 | 1,000 | 1 |
| <i>LOMBA DO PINHEIRO-Vila Recreio da Divisa</i> | 1,030 | 0,968 | 2 |
| <i>ANCHIETA-Vila Dique II</i> | 1,023 | 0,965 | 3 |
| <i>SARANDI-Vila Vitória da Conquista</i> | 0,969 | 0,941 | 4 |
| <i>LOMBA DO PINHEIRO-Vila Biriba</i> | 0,958 | 0,937 | 5 |
| <i>SARANDI-Vila Santissima Trindade</i> | 0,957 | 0,936 | 6 |
| <i>SÃO JOÃO-Vila Dique I</i> | 0,930 | 0,924 | 7 |
| <i>SÃO JOÃO-Vila Nazaré</i> | 0,889 | 0,907 | 8 |
| <i>RESTINGA-Quinta Unidade</i> | 0,882 | 0,904 | 9 |
| <i>AZENHA-Vila Zero Hora</i> | 0,882 | 0,904 | 10 |
| <i>JARDIM CARVALHO-Plano 100</i> | -0,934 | 0,115 | 154 |
| <i>PRALA DE BELAS 2/MENINO DEUS 2</i> | -0,935 | 0,114 | 155 |
| <i>FARROUPILHA-Redenção/SANTA CECÍLIA/SANTANA 1</i> | -0,950 | 0,108 | 156 |
| <i>JARDIM LINDÓLA</i> | -0,982 | 0,094 | 157 |
| <i>BOA VISTA/PASSO D'AREA-Vila Keddie</i> | -0,991 | 0,090 | 158 |
| <i>PETRÓPOLIS-Avenida Ijuí</i> | -1,005 | 0,084 | 159 |
| <i>BOM FIM/INDEPENDÊNCIA 1/RIO BRANCO</i> | -1,013 | 0,080 | 160 |
| <i>AUXILIADORA/MONT'SERRAT</i> | -1,026 | 0,075 | 161 |
| <i>MOINHOS DE VENTO/RIO BRANCO-IPA/INDEPENDÊNCIA 2</i> | -1,136 | 0,027 | 162 |
| <i>BELA VISTA/PETRÓPOLIS-Grêmio Náutico União</i> | -1,198 | 0,000 | 163 |

Fonte: Resultados da ACP

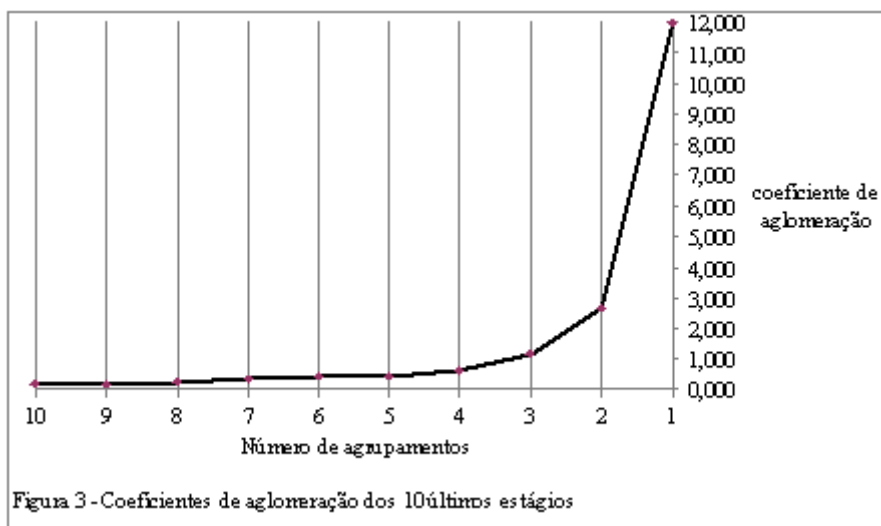
A Tabela 7 abaixo reproduz os 10 últimos estágios de aglomeração como base no método de Ward.

| Número de agrupamentos | Estágio | Agrupamento Combinado | | Coeficientes de Aglomeração |
|------------------------|------------|-----------------------|---------------|-----------------------------|
| | | Agrupamento 1 | Agrupamento 2 | |
| 10 | 153 | 6 | 7 | 0,119 |
| 9 | 154 | 4 | 5 | 0,145 |
| 8 | 155 | 2 | 12 | 0,176 |
| 7 | 156 | 8 | 9 | 0,244 |
| 6 | 157 | 1 | 3 | 0,323 |
| 5 | 158 | 6 | 21 | 0,419 |
| 4 | 159 | 2 | 4 | 0,623 |
| 3 | 160 | 1 | 8 | 1,179 |
| 2 | 161 | 2 | 6 | 2,647 |
| 1 | 162 | 1 | 2 | 11,996 |

Fonte: Resultados da AA (Análise de Agrupamento)

Um bom indicativo de corte para a determinação do número de agrupamentos é o coeficiente de aglomeração (última coluna da Tabela 7). Este coeficiente varia do máximo de detalhamento (cada UDH é o próprio grupo) à generalidade completa (um agrupamento único com todas as UDHs). A rigor, o coeficiente de aglomeração representa a perda de informação à medida que mais grupos vão sendo agregados. Observe que nos últimos estágios as variações entre um estágio e

outro parecem ser maiores, sendo que a partir do estágio 159 elas são mais acentuadas. Se olharmos os estágios do fim para o começo da agregação, verificamos que no último estágio, quando temos apenas um grupo com todas as UDHs, o coeficiente de aglomeração é máximo, o que é o mesmo que dizer que a perda de informação é tamanha que não ganhamos nada em termos de conhecimento ao colocarmos todas as UDHs em um único agrupamento. Para auxiliar na escolha do ponto de corte, construímos o gráfico da Figura 3 abaixo. No eixo vertical temos os coeficientes de aglomeração referentes a cada um dos últimos 10 estágios de aglomeração. O gráfico da Figura 2 parece indicar que um bom ponto de corte seja no estágio 159 com 4 agrupamentos, pois além da curva começar a ter saltos maiores a partir deste ponto, o coeficiente de aglomeração de 0,623 representa apenas 5,19% do coeficiente de aglomeração que representa a perda máxima de informação, ou seja, 11,996.



Outro tipo de gráfico gerado pela maioria dos softwares estatísticos é o dendrograma. Um dendrograma é uma árvore hierárquica de vínculos na qual podemos verificar no eixo vertical a perda de informação à medida que avançamos nos estágios de aglomeração. A Figura 4 reproduz o dendrograma que auxiliou na escolha dos 4 agrupamentos. Como o eixo horizontal do dendrograma contém as 163 UDHs de Porto Alegre, sua visualização em tamanho A4 fica prejudicada. Para quem acessar o trabalho em meio digital, sugerimos ampliar a imagem para uma melhor visualização. As Tabelas 8, 9, 10 e 11 apresentam a relação das UDHs em cada agrupamento (AG) conforme sua classificação no IVSE. As UDHs que ocupam as primeiras posições na Classificação são as que têm pior IVSE. A coluna Classificação permite, assim, que se possam verificar as diferenças dentro de um mesmo agrupamento.

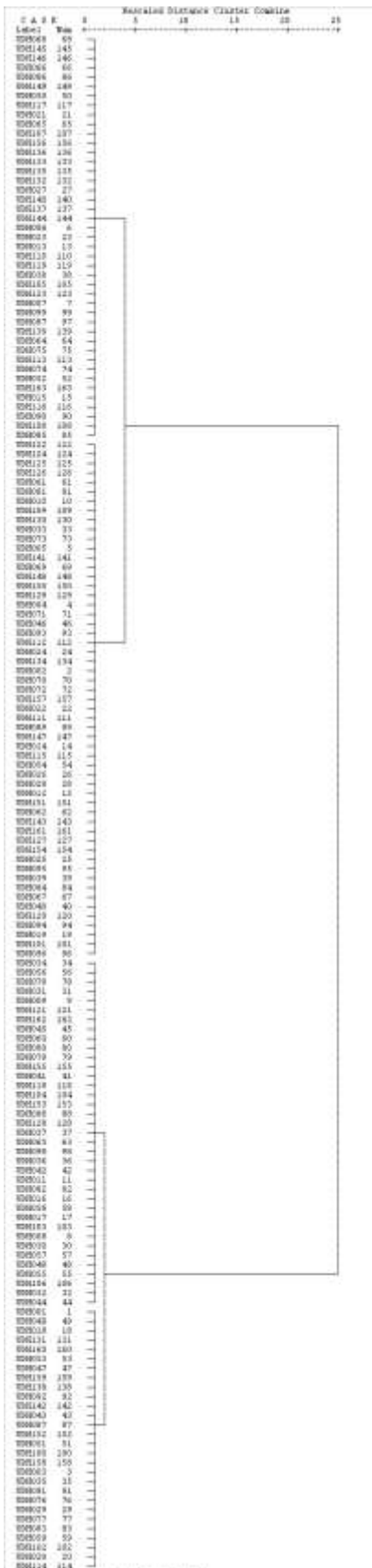


Figura 4 - Dendrograma segundo o Método de Ward
Fonte: Resultados de Análise de Aglomeramento.

Tabela 8
Relação das UDHs do Agrupamento de IVSE Alto (AG1)

| UDH | Classificação | UDH | Classificação |
|--|---------------|---|---------------|
| FLORESTA-Loteamento Santa Terezinha | 1 | CRISTAL-Vila Cristal | 28 |
| LOMBA DO PINHEIRO-Vila Recreio da Divisa | 2 | MORRO SANTANA-Vila das Laranjeiras | 29 |
| ANCHIETA-Vila Dique II | 3 | CASCATA-Vila Graciliano Ramos | 30 |
| SARANDI-Vila Vitória da Conquista | 4 | PARTENON-Vila Cachorro Sentado | 31 |
| LOMBA DO PINHEIRO-Vila Eriba | 5 | SANTA TEREZA-Vilas Tronco | 32 |
| SARANDI-Vila Santíssima Trindade | 6 | TERESÓPOLIS-Vila Silva Paes | 33 |
| SÃO JOÃO-Vila Dique I | 7 | VILA NOVA-Condomínio Campos de Cristal | 34 |
| SÃO JOÃO-Vila Nazaré | 8 | SARANDI-Vila Fraternidade | 35 |
| RESTINGA-Quinta Unidade | 9 | JARDIM SABARÁ-Vila Margarita | 36 |
| AZENHA-Vila Zero Hora | 10 | BELÉM NOVO-Praia das Garças | 37 |
| ARQUIPÉLAGO/FARRAPOS I/HUMAITÁ I | 11 | SERRARIA | 38 |
| SARANDI-Vila Dique | 12 | CAVALHADA-Loteamento Cavahada | 39 |
| MÁRIO QUINTANA-Loteamento Timbaúva | 13 | CASCATA-Vila Nossa Senhora da Glória | 40 |
| CHAPÉU DO SOL-Loteamento Chapéu do Sol | 14 | JARDIM CARVALHO-Colina do Prado | 41 |
| MENINO DEUS-Condomínio Lupicínio Rodrigues | 15 | CAMAQUÃ-Vila do Fumil(São Vicente Martir) | 42 |
| JARDIM SABARÁ-Vila Ernesto Pelana | 16 | RESTINGA-Restinga Velha | 43 |
| SANTA TEREZA-Vila Figueira | 17 | NONOAI-V Alto Erechim e parte da V Chácara Sperb | 44 |
| SANTA TEREZA-Vila Rio Branco | 18 | SARANDI-Vila Páscoa | 45 |
| SANTA TEREZA-Vila Ecologia | 19 | BELÉM VELHO-Condomínio Mariante | 46 |
| SANTA TEREZA-Vila Cruzeiro do Sul | 20 | RUBEM BERTA-Jardim Leopoldina II | 47 |
| PARTENON-Vila Santa Clara | 21 | MÁRIO QUINTANA-Chácara Fumaça e Jd Protásio Alves | 48 |
| PASSO DAS PEDRAS-Jardim dos Coqueiros | 22 | LOMBA DO PINHEIRO-Vila Pitinga | 49 |
| BOM JESUS-Vilas Divinéia, Mato Sampaio e Pinto | 23 | AGRONOMIA-Vila Esmeralda | 50 |
| CRISTAL-Vila do Valão | 24 | VILA CONCEIÇÃO-Vila Pellin | 51 |
| SANTA TEREZA-Orfanotrófio I | 25 | RESTINGA-Vila Bita | 52 |
| PARTENON-V Maria da Conceição (Pq Casa da Criança) | 26 | PARTENON-V Maria da Conceição (N Senhora do Brasil) | 53 |
| LOMBA DO PINHEIRO-Loteamento Quinta do Portal | 27 | SÃO JOSÉ-Vila Sudeste | 54 |
| | | CASCATA-Loteamento Vila Jardim Renascença | 55 |

Fonte: Resultados da Análise de Agrupamento (AA)

Tabela 9
Relação das UDHs do Agrupamento de IVSE Médio-Alto (AG2)

| UDH | Classificação | UDH | Classificação |
|--|---------------|--|---------------|
| SÃO JOSÉ 2/V JOÃO PESSOA 2/CEL APARÍCIO BORGES 2 | 56 | SARANDI-Santa Rosa | 78 |
| SÃO JOSÉ-Morro da Cruz | 57 | PARTENON-Vila Tijuca e Vila dos Sargentos | 79 |
| SARANDI-Asa Branca | 58 | RUBEM BERTA-Conjunto Residencial | 80 |
| CAVALHADA-Cidade de Deus | 59 | LAGEADO | 81 |
| SARANDI-Vila Ipê São Borja | 60 | MÁRIO QUINTANA-Vila Jardim das Colinas | 82 |
| SARANDI-Vila Miguano | 61 | MÁRIO QUINTANA-Rua 19 de Fevereiro | 83 |
| SARANDI-Loteamento Vila São Borja | 62 | IPANEMA-Beco da Bagunça | 84 |
| RUBEM BERTA-Vila Amazônia | 63 | VILA NOVA-Vila Monte Cristo e Vila Monte Cristo II | 85 |
| HÍPICA-Beco do Adelar | 64 | BELÉM VELHO-H Pq Belém/CASCATA-Vale dos Camudos | 86 |
| SARANDI-Vila Nova Brasília | 65 | RUBEM BERTA-Loteamento Vila Max Geiss | 87 |
| LOMBA DO PINHEIRO-Rarefeita | 66 | RESTINGA-Costa Gama | 88 |
| SARANDI-Vila Nossa Senhora Aparecida | 67 | PARTENON-Hospital São Pedro e PUC | 89 |
| SARANDI-Vila União | 68 | MORRO SANTANA-Vila Estrutural | 90 |
| MORRO SANTANA-Vila Tijuca | 69 | RESTINGA-Restinga Nova | 91 |
| LOMBA DO PINHEIRO-Densa | 70 | SANTA TEREZA-Arroto Passo Fundo | 92 |
| SARANDI-Loteamento Vila Santo Agostinho | 71 | BELÉM NOVO-Vila Flores da Cunha | 93 |
| RESTINGA-Barro Vermelho | 72 | ARQUIPÉLAGO-Iha da Pintada (norte e oeste) | 94 |
| VILA ASSUNÇÃO-Vila Guaíba | 73 | CAMPO NOVO/VILA NOVA-Loteamento COHAB Cavahada | 95 |
| CAMAQUÃ-Loteamento Vila Maria | 74 | CRISTAL-Vila Cai | 96 |
| LAMI | 75 | PONTA GROSSA/CHAPÉU DO SOL-Bairro/BELÉM NOVO-Bair | 97 |
| PASSO DAS PEDRAS I/RUBEM BERTA I | 76 | SANTA TEREZA-Vila Dona Malvina | 98 |
| ARQUIPÉLAGO-Iha das Flores (norte) | 77 | | |

Fonte: Resultados da Análise de Agrupamento (AA)

Tabela 10
Relação das UDHs do Agrupamento de IVSE Médio-Baixo (AG3)

| UDH | Classificação | UDH | Classificação |
|--|---------------|---|---------------|
| JARDIM CARVALHO-bairro | 99 | PASSO DAS PEDRAS-V Ingá/RUBEM BERTA-Costa e Silva | 113 |
| BOM JESUS-bairro/VILA JARDIM-Rua Conde da Figueira | 100 | JARDIM SABARÁ-bairro | 114 |
| SÃO JOSÉ 1/V JOÃO PESSOA 1/CEL APARÍCIO BORGES 1 | 101 | CAMAQUÃ-Bairro | 115 |
| AGRONOMIA-Campus do Vale UFRGS | 102 | RUBEM BERTA-Jardim Leopoldina | 116 |
| HÍPICA-Bairro/GUARUJÁ/ESPÍRITO SANTO 1/PANEMA 1 | 103 | MORRO SANTANA-bairro | 117 |
| VILA NOVA-Avenida Vicente Monteggia | 104 | MEDIANEIRA-Coorigha/STA TEREZA-B Ouro e B Solidão | 118 |
| VILA JARDIM-Avenida Circular | 105 | CAVALHADA-VTC e Cecopam/NONOAI-Av Dr. Campos Velho | 119 |
| GLÓRIA/CASCATA-Vila Dona Veva | 106 | MEDIANEIRA-Av Niterói/STA TEREZA-V dos Comerciantes | 120 |
| SARANDI-Parque dos Maías | 107 | CRISTAL-Avenida Coronel Massot | 121 |
| TERESÓPOLIS-Loteamento Estrada dos Alpes | 108 | PARTENON-Intercap e Jardim Bento Gonçalves | 122 |
| FARRAPOS 2/HUMAITÁ 2 | 109 | ANCHIETA 1/HUMAITÁ 3/NAVEGANTES/SÃO GERALDO | 123 |
| NONOAI-Avenida Nonoai/TERESÓPOLIS-Asmobame | 110 | PASSO DAS PEDRAS 2/RUBEM BERTA 2 | 124 |
| PARTENON-Loteamento Vila Batista Xavier | 111 | HÍPICA-Loteamento Jardim Amazônia | 125 |
| SARANDI-Vila Elizabeth e Vila Leão | 112 | VILA IPIRANGA-Educandário/JARDIM ITU-R Paul Harris | 126 |

Fonte: Resultados da Análise de Agrupamento (AA).

Tabela 11
Relação das UDHs do Agrupamento de IVSE Baixo (AG4)

| UDH | Classificação | UDH | Classificação |
|---|---------------|---|---------------|
| CRISTAL-Museu Iberê Camargo | 127 | CIDADE BAIXA | 146 |
| JD FLORESTA/ID SÃO PEDRO/STA MARIA GORETTI/SÃO JOÃO | 128 | JARDIM DO SALSO | 147 |
| PASSO D'AREIA-Vila do IAPI | 129 | MEDIANEIRA-Loteamento Carlos Barbosa | 148 |
| CRISTAL-Avenida Dr. Campos Velho | 130 | CENTRO-Parte Histórica e Comercial | 149 |
| ESPÍRITO SANTO-Clube do Professor Gaúcho/IPANEMA 2 | 131 | JARDIM ITU-Parque Brigada Militar | 150 |
| S SEBASTIÃO/V IPIRANGA-Pç N Árabs/ID ITU-Jd Itati | 132 | CENTRO-Cientec/PRAIA DE BELAS-Parque Harmonia | 151 |
| CRISTO REDENTOR | 133 | CHÁCARAS DAS PEDRAS/TRÊS FIGUEIRAS | 152 |
| TERESÓPOLIS-Teresópolis Tênis Clube | 134 | HIGIENÓPOLIS/SÃO JOÃO-Sogipa | 153 |
| PETRÓPOLIS-Rua Ivo Corseuil/JARDIM BOTÂNICO | 135 | JARDIM CARVALHO-Plano 100 | 154 |
| SANTO ANTÔNIO/PARTENON-Partenon Tênis Clube | 136 | PRAIA DE BELAS 2/MENINO DEUS 2 | 155 |
| NONOAI-Morro Teresópolis | 137 | FARROUPILHA-Redenção/SANTA CECÍLIA/SANTANA 1 | 156 |
| SANTA TEREZA-Próximo ao Menino Deus | 138 | JARDIM LINDÓIA | 157 |
| AZENHA-Shop João Pessoa/SANTANA-Inst de Cardiologia | 139 | BOA VISTA/PASSO D'AREIA-Vila Keddie | 158 |
| VILA NOVA-Condomínio Village | 140 | PETRÓPOLIS-Avenida Ijuí | 159 |
| FLORESTA/INDEPENDÊNCIA-Shopping Total | 141 | BOM FIM/INDEPENDÊNCIA 1/RIO BRANCO | 160 |
| MEDIANEIRA-Pç Alci Maia/STA TEREZA-Pç Montese | 142 | AUXILIADORA/MONT'SERRAT | 161 |
| TRISTEZA/V ASSUNÇÃO/V CONCEIÇÃO/PEDRA REDONDA | 143 | MOINHOS DE VENTO/RIO BRANCO-IPA/INDEPENDÊNCIA 2 | 162 |
| MENINO DEUS 1/AZENHA 1/PRAIA DE BELAS 1 | 144 | BELA VISTA/PETRÓPOLIS-Grêmio Náutico União | 163 |
| JARDIM SABARÁ-Parque do Sol | 145 | | |

Fonte: Resultados da Análise de Agrupamento (AA).

Tendo em vista a ampliação da capacidade de visualização espacial que o georreferenciamento permite, geramos a Figura 5 com o mapa do Município de Porto Alegre com os 4 agrupamentos. No mapa ficam explícitos os enclaves de vulnerabilidade que retalhavam uma cidade tida como homogênea. A rigor esta heterogeneidade já havia sido constatada nas 163 UDHs de Porto Alegre constantes do Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre. Como o estudo que agora apresentamos, verifica-se que, mesmo agregando UDHs de locais diferentes da cidade, persistiam em 2000 indicações de que não eram poucos os espaços de vulnerabilidade. Fica-se, portanto, com uma boa impressão de como as técnicas de Análise Multivariada podem auxiliar na simplificação e estruturação da vasta quantidade de dados gerados pelos censos. A expectativa quanto à disponibilidade de novos e mais detalhados dados que serão gerados pelo Censo de 2010, sugere um horizonte promissor para a aplicação das técnicas que acabamos de fazer uso com os dados do Censo de 2000.

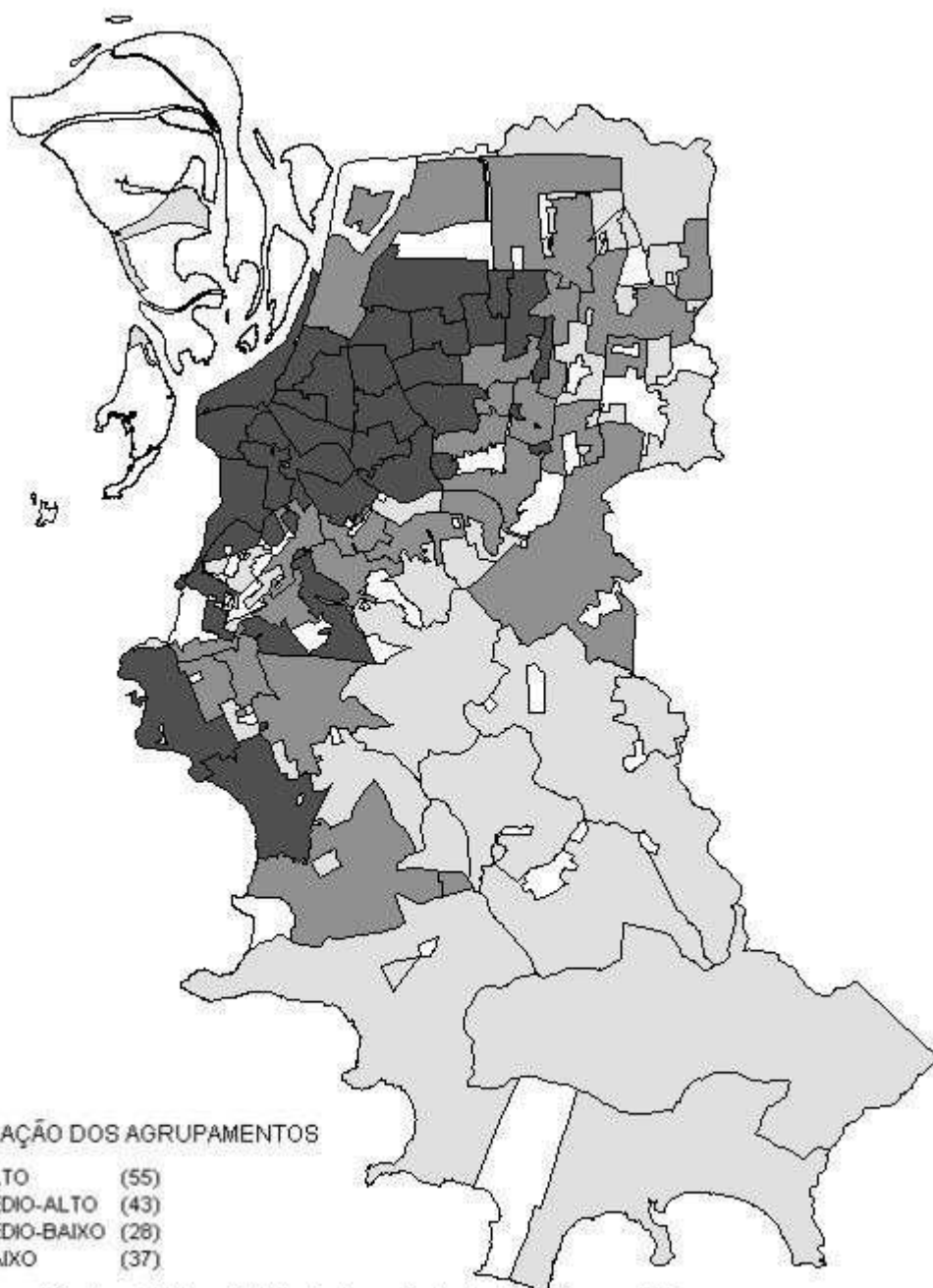


Figura 5 - Mapa do Índice de Vulnerabilidade Socioeconômica de Porto Alegre - 2000
 Elaborado pelo técnico do Observatório da Cidade de Porto Alegre Luis Antônio Thomas Fernandes

3. CONCLUSÃO

O presente trabalho agrega-se a tantos outros que sugerem que a aplicação de técnicas de Análise Multivariada pode auxiliar na simplificação e estruturação de um conjunto extenso de informações. Nesse sentido, aplicou-se a Análise de Componentes Principais para transformar de 29 variáveis em apenas duas componentes com capacidade para explicar 84,9% da variância total das variáveis originais. Estas componentes permitiram a construção de um Índice de Vulnerabilidade Socioeconômica (IVSE) para o Município de Porto Alegre sobre o qual se aplicou a Análise de Agrupamentos. A Análise de Agrupamento permitiu que se juntassem em agrupamentos as UDHS mais homogêneas. A redução do número de variáveis e a geração de agrupamentos foram feitas com pouca perda de informação, ficando claro que os resultados alcançados podem servir de ferramenta auxiliar na identificação de desigualdades intra-urbanas. Tendo em vista o número elevado de UDHS em cada agrupamento, a classificação pode servir como uma forma de verificar a heterogeneidade remanescente no interior dos agrupamentos, permitindo a focalização das políticas públicas sem perda da noção de contextualização que o agrupamento permite.

4. REFERÊNCIAS BIBLIOGRÁFICAS

Atlas do Desenvolvimento Humano da Região Metropolitana de Porto Alegre. – Porto Alegre: Prefeitura Municipal/Secretaria de Coordenação Política e Governança Local; Metroplan; PNUD; Fundação João Pinheiro, 2008.

EVERIT, B. **Cluster analysis**. London, Heinemann Educational Books. Second Edition, 1980.

FACHEL, J.M.G. **Análise Multivariada da Estrutura Social Urbana do Município de Porto Alegre**. Atas do 1.º Encontro de Docentes de Estatística da Região Sul – Departamento de Estatística da Universidade Federal do Rio Grande do Sul, 12 a 16 de julho de 1982.

FLURY, B., RIEDWYL, H. **Multivariate statistic: a practical approach**. London, Chapman and Hall Ltd. , 1988.

HAIR, J.R.Jr., ANDERSON, R.E. TATHAM R.L.e BLACK W.C. **Análise Multivariada de Dados**. 5ed. - Porto Alegre: Bookman, 2005.

JOHNSON, R.A. & WICHERN, D.W. **Applied Multivariate Statistical Analysis**. Fifth Edition. Prentice Hall, 2002.

MINGOTI, S.A. **Análise de Dados Através de Métodos de Estatística Multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

POSSOLI, S. **Técnicas de Análise Multivariada para Avaliação das Condições de Saúde dos Municípios do Rio Grande do Sul**. Revista de Saúde Pública, São Paulo, 18:288 – 300, 1984.

REIS, E. **Estatística Multivariada Aplicada**. 2ed. – Lisboa: Edições Silabo, Lda., 2001.

* Economista do Observatório da Cidade de Porto Alegre/Secretaria de Coordenação Política e Governança Local da Prefeitura Municipal de Porto Alegre. E-mail: ceduardo@smgl.prefpoa.com.br

** Estatística do Observatório da Cidade de Porto Alegre/Secretaria de Coordenação Política e Governança Local da Prefeitura Municipal de Porto Alegre. E-mail: valeria@smgl.prefpoa.com.br

¹ O detalhamento sobre as formulações matemáticas pode ser encontrado em Johnson e Wichern (2002) e Reis (2001).

² Similaridades são medidas que tomam valores entre 0 e 1. Distâncias são medidas que podem tomar qualquer valor positivo (Everit, 1980).

³ Testa a seguinte hipótese: H_0 : Matriz de correlações = I (matriz Identidade). A estatística de teste para testar esta hipótese é: $-\left[n - 1 - \frac{1}{6}(2p + 5) \right] \ln |R|$. Onde: $|R|$ é o determinante da Matriz de correlações. Esta estatística tem uma distribuição assintótica de χ^2 com $\left[\frac{1}{2} p (p - 1) \right]$ graus de liberdade.

⁴ Metodologia desenvolvida pela Fundação João Pinheiro, IPEA e PNUD.