

# **Análise de Componentes Principais e Análise de Agrupamentos – Aplicação em Variáveis de Educação e Renda no Estado de Pernambuco**

## **Principal Component Analysis and Cluster Analysis applied to Educational and Income Variables for the state of Pernambuco (BRAZIL)**

**Renata Mello Montenegro de Araújo**

Programa de Pós-Graduação em Geografia-Tratamento da Informação Espacial da PUC Minas

renatasimone@hotmail.com

*Resumo – Este estudo tem como objetivo explorar o uso de duas técnicas de Análise Multivariada de dados geográficos para mapeamento dos municípios do estado de Pernambuco de acordo com variáveis que tratam de analfabetismo e desigualdade de renda. A Análise de Componentes Principais e a Análise de Agrupamentos foram aplicadas, com a utilização dos softwares Ninna PCA e Ninna Cluster (PPG/TIE – PUCMINAS), para o estudo de 44 variáveis coletadas no Atlas de Desenvolvimento Humano. Os resultados da Análise de Componentes Principais apontam para o Agreste Meridional, o Sertão do Araripe e a Zona da Mata Sul como as regiões do estado com piores situações de analfabetismo e pobreza. A primeira componente principal, cuja variância alcançou 52,15% da variabilidade dos dados, agrupou 19 variáveis. A análise de agrupamentos para quatro classes tipológicas apresentou coerência em relação às quatro classes hierárquicas mapeadas na análise anterior, mostrando, portanto, que as técnicas se apresentam como complementares para estudos geográficos envolvendo a Análise Multivariada de dados.*

*Palavras-chave – Pernambuco. Análise de componentes Principais. Análise de Agrupamentos. Clusters. Desigualdade de Renda. Analfabetismo. Pobreza.*

*Abstract – This paper explore the use of two techniques for geographic multivariate data mapping the municipalities of Pernambuco state according to statistical variables expressing illiteracy and income inequality. Principal Component Analysis and Cluster Analysis were applied with the software Ninna PCA and Ninna Cluster (PPG/TIE – PUCMINAS) to 44 statistical variables gathered in Human Development Atlas. Principal Component Analysis whose variance catch up 52.15% of the data variability has joined 19 variables. The results indicate Agreste Meridional, Sertão do Araripe and Zona da Mata Sul as regions presenting the most critical situation of illiteracy and poverty. Cluster Analysis with four classes has presented coherence with the four hierarchical classes*

*mapped by PCA suggesting the two techniques may be complementary in such studies.*

*Keywords – Pernambuco. Principal Component Analysis. Clusters Analysis. Income inequality. Illiteracy. Poverty.*

### I INTRODUÇÃO

Este trabalho tem como objetivo aplicar técnicas de análise de componentes principais e análise de agrupamentos para 44 variáveis de Educação - Analfabetismo e Acesso à Escola - e Renda - Desigualdade e Pobreza - no Estado de Pernambuco, e realizar uma breve análise dos resultados obtidos com ambas as técnicas e dos mapeamentos realizados para cada uma das análises.

Segundo Barroso, citado por Alencar (2009), a Análise Multivariada de dados tem um significado cada vez mais amplo na Geografia por possibilitar uma maior facilidade no exame conjunto de informações necessárias ao fornecimento de subsídios que permitam a explicação de fenômenos geográficos, o estudo de tendências e padrões espaciais, a formulação de modelos e a elaboração de previsões.

A organização, classificação e análise de dados geográficos podem ser feitas com a utilização de diversas técnicas multivariadas. Nesse trabalho duas delas serão abordadas e discutidas comparativamente por meio dos



resultados e mapeamentos obtidos com os mesmos dados utilizando-se cada uma das técnicas, que são a Análise de Componentes Principais e a Análise de Agrupamentos.

## II A ANÁLISE DE COMPONENTES PRINCIPAIS

De acordo com Alencar (2009), a Análise de Componentes Principais, técnica incluída na Análise Fatorial, é uma abordagem estatística que pode ser usada para analisar as inter-relações entre um grande número de variáveis de maneira a condensar a informação contida nelas em um conjunto menor de variáveis estatísticas, observando uma perda mínima de informação. A utilização dessa técnica de análise multivariada na Geografia permite minimizar as redundâncias possivelmente presentes quando se pretende levar em consideração um número muito elevado de variáveis, possibilitando a captação da maior variabilidade dos dados em poucas componentes, tornando mais simples o mapeamento das informações. Os objetivos dessa técnica, em síntese, são:

- Gerar novas variáveis em um número reduzido, mas que consigam expressar de modo satisfatório a informação contida no conjunto original de dados;
- Reduzir a dimensão do problema que está sendo estudado, como passo prévio para futuras análises;
- Eliminar, quando for possível, algumas variáveis originais, caso elas contribuam com pouca informação.

Os cálculos efetuados nesse estudo foram realizados no *software Ninna PCA*, que será apresentado mais detalhadamente em item IV desse texto.

## III A ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos ou *cluster analysis* é amplamente utilizada em Geografia e em outras áreas do conhecimento e consiste em um conjunto de técnicas que objetivam identificar e agrupar objetos segundo a similaridade sobre

algum atributo ou característica particular que possuem, possibilitando uma melhor análise dos dados e a identificação de padrões de comportamento.

Na Geografia, a apropriação desse conjunto de técnicas visa uma melhor organização dos dados efetuada com a divisão desses dados em grupos ou classes, facilitando sua compreensão e interpretação. Trata-se de uma classificação por similaridade, tendo como princípio básico, segundo Gerardi e Silva (1981), a busca da maximização de semelhança intraclasses e maximização de diferença interclasses. As autoras discorrem que existem várias técnicas de mensurar a similaridade. No caso da análise multivariada aqui realizada, a semelhança é medida por meio da distância, sendo que quanto menor a distância euclidiana medida no espaço relativo-estatístico, maior a similaridade.

Ressalta-se que a análise de agrupamentos é uma técnica de classificação tipológica de dados que leva em consideração a semelhança entre os casos observados, não considerando, no entanto, nenhuma hierarquia entre os grupos ou classes obtidas. A semelhança ou parença é calculada com base numa matriz de distâncias, na qual os elementos com distâncias euclidianas mais próximas vão sendo agrupados sequencialmente até um único grupo ser formado. Os resultados podem ser visualizados graficamente em esquemas denominados *dendogramas*. Para a partição dos grupos podem ser utilizados vários métodos, dentre os quais podemos citar, de acordo com Alencar (2009):

- O método das Médias das Distâncias, ou *Average Linkage*, que utiliza a média das distâncias entre todos os pares de objetos da matriz de dados para se criar a matriz de distâncias. Baseando-se nela um grupo é formado e a matriz de distâncias é recalculada tomando-se esse grupo como novo elemento. Um novo agrupamento é formado e o processo continua até que

apenas um grupo, contendo todos os elementos do conjunto de dados, seja criado.

- O método da Ligação Simples, ou *Single Linkage*, é conhecido também como o Método do Vizinho mais Próximo ou Método da Distância Mínima. Como esse método une grupos segundo uma distância mínima entre eles, os grupos formados tendem a ser menos homogêneos. Isso significa que em um mesmo grupo podem ser encontrados elementos bem distintos.
- O método da Ligação Completa, ou *Complete Linkage*, também é conhecido como o Método do Vizinho mais Distante. Ao longo do processo, cada vez que um novo elemento é adicionado a um grupo, esse se torna mais distinto em relação aos outros, formando, ao final, grupos mais homogêneos que os formados por meio do Método da Ligação Simples. Esse método é particularmente indicado para a formação de grupos com tamanhos mais semelhantes.
- O método K-Médias, ou *K-Means*. É um método muito útil para a formação de agrupamentos quando se tem muitos objetos. É um método que procura aumentar a homogeneidade dentro de cada grupo aumentando também a diferença entre eles.

#### IV OS SOFTWARES NINNA PCA E NINNA CLUSTER

Os *software Ninna PCA* e *Ninna Cluster* foram elaborados no âmbito do Programa de Pós-Graduação em Geografia – Tratamento da Informação Espacial da PUCMINAS – Pontifícia Universidade Católica de Minas Gerais por Alencar(2009), como parte de sua Tese de Doutorado, intitulada “A Análise Multivariada no Tratamento da Informação Espacial - Uma Abordagem Matemático-

Computacional em Análise de Agrupamentos e Análise de Componentes Principais”.

Os *software NinnaCluster* e *NinnaPCA* são aplicativos desenvolvidos para realizar os cálculos envolvidos nas técnicas da Análise de Agrupamentos e na da Análise de Componentes Principais, respectivamente. Os dois programas aceitam planilhas de dados em formato *Excel* (.xls) ou separado por vírgulas (.csv).

O *Ninna PCA* admite a inserção de dados cartográficos georreferenciados em formato *shapefile*, realizando, inclusive, dentro do ambiente do programa, o mapeamento dos resultados obtidos. Todas as etapas dos cálculos envolvidas na análise podem ser acessadas na plataforma do programa e exportadas para formato *Excel*, de forma que podem ser utilizadas em outros programas para computador, específicos para mapeamento. É possível acessar, em cada aba disponível os dados coletados, os dados padronizados, a matriz de correlação, os autovalores e respectivos autovetores relacionados, as componentes principais, os pesos, os coeficientes de determinação e os escores para mapeamento. Aos pesos é aplicado um filtro de 0,70 para efetivar a seleção das variáveis agrupadas em cada componente. Na aba **Autovetores** pode-se ter acesso ao rol de variáveis relativas a cada autovetor, de acordo com o filtro de pesos aplicado.

O *Ninna Cluster* dispõe dos quatro técnicas de agrupamentos descritos anteriormente, disponibilizando para consulta a matriz de distâncias calculada e os dendogramas, representações gráficas dos agrupamentos, à exceção do método de partição *K-means*, que disponibiliza a relação dos grupos formados, cuja quantidade é escolhida pelo operador do *software*.

#### V ESCOLHA DAS VARIÁVEIS

O estado de Pernambuco, apesar de ser um dos estados mais destacados na Região Nordeste do país, caracteriza-se por sua alta densidade demográfica e pelos baixos Índices de

Desenvolvimento Humano - IDH. Segundo a Embrapa – Empresa Brasileira de Pesquisa Agropecuária, Pernambuco é o estado nordestino com segunda maior densidade populacional, ficando atrás apenas de Alagoas, além de possuir a maior população urbana por Km<sup>2</sup> efetivamente urbanizado do país, segundo estimativa realizada pela Embrapa (MIRANDA *et alli*, 2005).

O QUADRO 1 mostra o *ranking* dos Estados Brasileiros em Relação ao IDH em 2005, de acordo com o PNUD - Programa das Nações Unidas para o Desenvolvimento e Fundação João Pinheiro, que aponta o estado de Pernambuco com 24º colocado entre os 27 estados da Federação.

QUADRO 1 - Ranking do IDH dos Estados em 2005

1	Distrito Federal	0,874
2	Santa Catarina	0,868
3	São Paulo	0,835
4	Rio de Janeiro	0,832
5	Southern Mato Grosso do Sul	0,821
6	Paraná	0,808
7	Espírito Santo	0,802
8	Mato Grosso do Sul	0,802
9	Goiás	0,800
10	Minas Gerais	0,800
11	Mato Grosso	0,798
12	Amazoná	0,780
13	Amazonas	0,780
14	Rorônia	0,778
15	Tocantins	0,756
16	Piauí	0,755
17	Acre	0,751
18	Roraima	0,750
19	Bahia	0,742
20	Sergipe	0,742
21	Rio Grande do Norte	0,738
22	Ceará	0,725
23	Pernambuco	0,718
24	Paraná	0,718
25	Piauí	0,705
26	Maranhão	0,685
27	Alagoas	0,677

Centro-Oeste
Sul
Sudeste
Norte
Nordeste

Fonte: PNUD/Fundação João Pinheiro.

Os dados de 1991 a 2005 estão no relatório [Emprego, Desenvolvimento Humano e Trabalho Decente – A experiência brasileira recente](#). Disponível em

[http://www.pnud.org.br/pobreza\\_desigualdade/reportagens/index.php?id01=3039&lay=pde](http://www.pnud.org.br/pobreza_desigualdade/reportagens/index.php?id01=3039&lay=pde)

As variáveis aqui discutidas foram escolhidas com a intenção de inter-relacionar dados concernentes à baixa educação (analfabetismo), pobreza e alta desigualdade de renda nos municípios pernambucanos, visando mapear as áreas com situações mais críticas em relação aos dados. Foram selecionadas 44 variáveis, referentes ao ano de 2000, obtidas no Atlas do Desenvolvimento Humano no Brasil<sup>1</sup>, disponível no sítio do PNUD que se encontram listadas no Quadro 2. Esse estudo será realizado em nível municipal e cobrirá todo o Estado de Pernambuco.

A maioria das variáveis selecionadas está expressa em porcentagem e trata da quantidade de pessoas com determinado perfil de idade e seus respectivos percentuais segundo condições estabelecidas referentes aos anos de estudo, ou ao analfabetismo.

A pesquisa também abrangeu percentuais que tratam da apropriação da renda, que são capazes de expressar a desigualdade de renda os municípios do Estado. Optou-se por incluir também a renda per capita total dos municípios, e indicadores tais como taxa bruta de frequência à escola, taxa de alfabetização, intensidade de indigência e pobreza, além dos Índices de Desenvolvimento Humanos Municipais totais e parciais, relativos à Educação e Renda.

Base de dados eletrônica organizada pelo PNUD – Programa das Nações Unidas para o Desenvolvimento, IPEA Instituto de Pesquisa Econômica Aplicada e Fundação João Pinheiro (Governo de Minas Gerais) tendo como base nos microdados dos censos de 1991 e de 2000 do IBGE (Fundação Instituto Brasileiro de Geografia e Estatística). Disponível para *download* em <http://www.pnud.org.br/atlas/instalacao/index.php>

QUADRO 2 – VARIÁVEIS SELECIONADAS

ID	VARIÁVEL
1	Percentual de crianças de 10 a 14 anos analfabetas, 2000
2	Percentual de crianças de 10 a 14 anos com mais de um ano de atraso escolar, 2000
3	Percentual de crianças de 10 a 14 anos com menos de quatro anos de estudo, 2000
4	Percentual de crianças de 10 a 14 anos na escola, 2000
5	Percentual de crianças de 5 a 6 anos na escola, 2000
6	Percentual de crianças de 7 a 14 anos analfabetas, 2000
7	Percentual de crianças de 7 a 14 anos com acesso ao curso fundamental, 2000
8	Percentual de crianças de 7 a 14 anos com mais de um ano de atraso escolar, 2000
9	Percentual de crianças de 7 a 14 anos na escola, 2000
10	Percentual de crianças de 7 a 14 anos que estão freqüentando o curso fundamental, 2000
11	Percentual de crianças entre 4 e 5 anos na escola, 2000
12	Média de anos de estudo das pessoas de 25 anos ou mais de idade, 2000
13	Percentual de pessoas de 18 a 22 anos com acesso ao curso superior, 2000
14	Percentual de pessoas de 18 a 22 anos que freqüentam curso superior, 2000
15	Percentual de pessoas de 25 anos ou mais analfabetas, 2000
16	Percentual de pessoas de 25 anos ou mais com menos de oito anos de estudo, 2000
17	Percentual de pessoas de 25 anos ou mais com menos de quatro anos de estudo, 2000
18	Percentual de pessoas de 25 anos ou mais de idade com doze anos ou mais de estudo, 2000
19	Percentual de pessoas de 25 anos ou mais freqüentando curso superior, 2000
20	Percentual de pessoas que freqüentam o fundamental em relação à população de 7 a 14 anos, 2000
21	Percentual de adolescentes de 15 a 17 anos analfabetas, 2000
22	Percentual de adolescentes de 15 a 17 anos com menos de quatro anos de estudo, 2000
23	Percentual de pessoas de 15 anos ou mais analfabetas, 2000
24	Percentual de pessoas de 15 anos ou mais com menos de quatro anos de estudo, 2000
25	Percentual de pessoas de 18 a 24 anos analfabetas, 2000
26	Percentual de pessoas de 18 a 24 anos com menos de quatro anos de estudo, 2000
27	Percentual da renda apropriada pelos 10% mais ricos da população, 2000
28	Percentual da renda apropriada pelos 20% mais pobres da população, 2000
29	Percentual da renda apropriada pelos 20% mais ricos da população, 2000
30	Percentual da renda apropriada pelos 40% mais pobres da população, 2000
31	Percentual da renda apropriada pelos 60% mais pobres da população, 2000
32	Percentual da renda apropriada pelos 80% mais pobres da população, 2000
33	Intensidade da indigência, 2000
34	Intensidade da pobreza, 2000
35	Percentual crianças em domicílios com renda per capita menor que R\$75,50, 2000
36	Percentual de crianças em domicílios com renda per capita menor que R\$37,75, 2000
37	Percentual de pessoas com renda per capita abaixo de R\$37,75, 2000
38	Percentual de pessoas com renda per capita abaixo de R\$75,50, 2000
39	Renda per Capita, 2000
40	Taxa de alfabetização, 2000
41	Taxa bruta de freqüência à escola, 2000
42	Índice de Desenvolvimento Humano Municipal-Renda, 2000
43	Índice de Desenvolvimento Humano Municipal-Educação, 2000
44	Índice de Desenvolvimento Humano Municipal, 2000

## VI PROCESSAMENTOS DOS DADOS E MAPEAMENTOS

Os dados coletados foram abertos, inicialmente no programa *Ninna PCA*, onde foi agregado um arquivo *shape file* referente à base cadastral do Estado de Pernambuco. A análise de Componentes Principais foi executada

A primeira componente principal calculada foi capaz de absorver 52,15% da variabilidade dos dados, agrupando 19 variáveis, já a segunda componente principal representa 14,11% da variabilidade total dos dados e agrupou 2 variáveis, acumulando uma variância de 66,26% dos dados coletados inicialmente. Na geografia é razoável trabalhar-se com uma variância em torno de 70% da variabilidade total.



QUADRO 3 – VARIÁVEIS AGRUPADAS NA  
COMPONENTE PRINCIPAL 1

Componente Principal 1	
ID	VARIÁVEL
1	Percentual de crianças de 10 a 14 anos analfabetas, 2000(1)
2	Percentual de crianças de 10 a 14 anos com mais de um ano de atraso escolar, 2000(2)
3	Percentual de crianças de 10 a 14 anos com menos de quatro anos de estudo, 2000(3)
6	Percentual de crianças de 7 a 14 anos analfabetas, 2000(6)
8	Percentual de crianças de 7 a 14 anos com mais de um ano de atraso escolar, 2000(8)
15	Percentual de pessoas de 25 anos ou mais analfabetas, 2000(15)
16	Percentual de pessoas de 25 anos ou mais com menos de oito anos de estudo, 2000(16)
17	Percentual de pessoas de 25 anos ou mais com menos de quatro anos de estudo, 2000(17)
21	Percentual de adolescentes de 15 a 17 anos analfabetas, 2000(21)
22	Percentual de adolescentes de 15 a 17 anos com menos de quatro anos de estudo, 2000(22)
23	Percentual de pessoas de 15 anos ou mais analfabetas, 2000(23)
24	Percentual de pessoas de 15 anos ou mais com menos de quatro anos de estudo, 2000(24)
25	Percentual de pessoas de 18 a 24 anos analfabetas, 2000(25)
26	Percentual de pessoas de 18 a 24 anos com menos de quatro anos de estudo, 2000(26)
34	Intensidade da pobreza, 2000(34)
35	Percentual crianças em domicílios com renda per capita menor que R\$75,50, 2000(35)
36	Percentual de crianças em domicílios com renda per capita menor que R\$37,75, 2000(36)
37	Percentual de pessoas com renda per capita abaixo de R\$37,75, 2000(37)
38	Percentual de pessoas com renda per capita abaixo de R\$75,50, 2000(38)

Os QUADROS 3, 4 e 5 mostram quais variáveis foram agrupadas na primeira e na segunda componentes principais, e também aquelas que deixaram de constar no mapeamento, devido ao pequeno poder de explicação dentro do conjunto de dados.

O QUADRO 3 traz a relação das variáveis que constituem a primeira componente principal, o que revela variáveis referentes a alta correlação entre dados de baixa escolaridade com baixa renda e alta intensidade de pobreza.

O QUADRO 4 relaciona as duas variáveis agrupadas na segunda componente, que são referentes à alta apropriação da renda municipal por pequenas porcentagens da população mais rica.

QUADRO 4 - VARIÁVEIS AGRUPADAS NA  
COMPONENTE PRINCIPAL 2

Componente Principal 2	
ID	VARIÁVEL
29	Percentual da renda apropriada pelos 20% mais ricos da população, 2000(29)
27	Percentual da renda apropriada pelos 10% mais ricos da população, 2000(27)

QUADRO 5 - VARIÁVEIS NÃO AGRUPADAS NAS  
COMPONENTES 1 E 2

Componente Principal 2	
ID	VARIÁVEL
4	Percentual de crianças de 10 a 14 anos na escola, 2000
5	Percentual de crianças de 5 a 6 anos na escola, 2000
7	Percentual de crianças de 7 a 14 anos com acesso ao curso fundamental, 2000
9	Percentual de crianças de 7 a 14 anos na escola, 2000
10	Percentual de crianças de 7 a 14 anos que estão freqüentando o curso fundamental, 2000
11	Percentual de crianças entre 4 e 5 anos na escola, 2000
12	Média de anos de estudo das pessoas de 25 anos ou mais de idade, 2000
13	Percentual de pessoas de 18 a 22 anos com acesso ao curso superior, 2000
14	Percentual de pessoas de 18 a 22 anos que freqüentam curso superior, 2000
18	Percentual de pessoas de 25 anos ou mais de idade com doze anos ou mais de estudo, 2000
19	Percentual de pessoas de 25 anos ou mais freqüentando curso superior, 2000
20	Percentual de pessoas que freqüentam o fundamental em relação à população de 7 a 14 anos, 2000
28	Percentual da renda apropriada pelos 20% mais pobres da população, 2000
30	Percentual da renda apropriada pelos 40% mais pobres da população, 2000
31	Percentual da renda apropriada pelos 60% mais pobres da população, 2000
32	Percentual da renda apropriada pelos 80% mais pobres da população, 2000
33	Intensidade da indigência, 2000
39	Renda per Capita, 2000
40	Taxa de alfabetização, 2000
41	Taxa bruta de freqüência à escola, 2000
42	Índice de Desenvolvimento Humano Municipal-Renda, 2000
43	Índice de Desenvolvimento Humano Municipal-Educação, 2000
44	Índice de Desenvolvimento Humano Municipal, 2000



No QUADRO 5 podem-se conferir as variáveis com menor poder de explicação em relação ao conjunto dos dados, que tratam do acesso à educação, da renda apropriada pela população mais pobre, incluindo também os Índices de Desenvolvimento Humano.

Os dados dos escores obtidos no *software Ninna PCA* foram levados para o programa *ArcGIS (ESRI)* para realização dos mapeamentos.

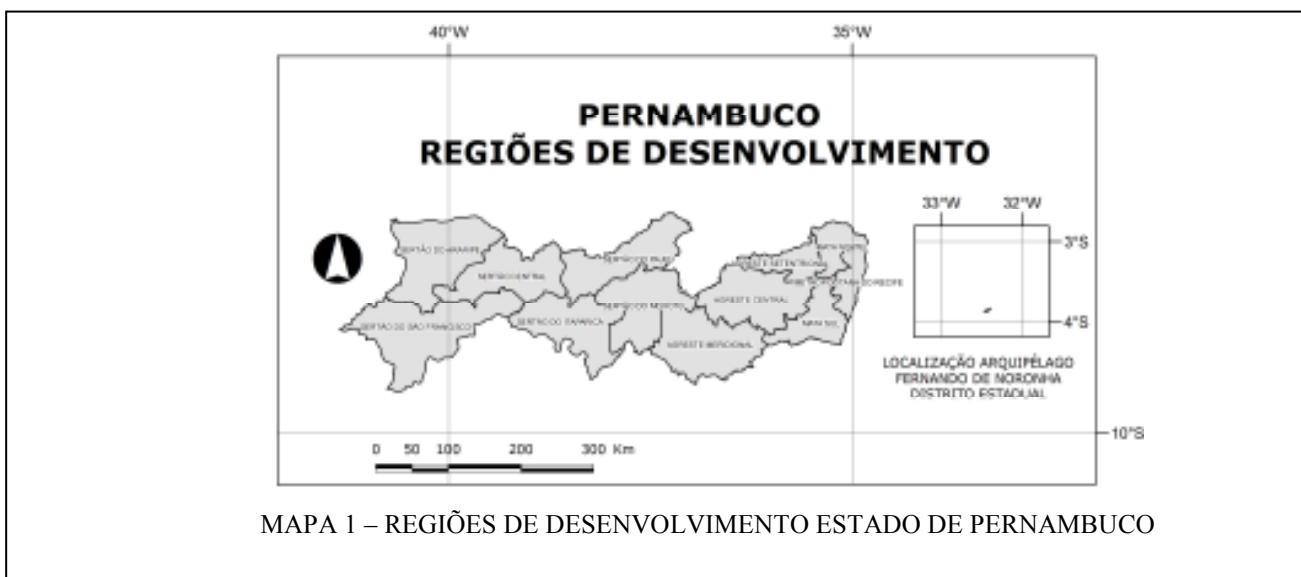
Os mesmos dados foram processados no *software Ninna Cluster*, usando o método *k-means* para geração de quatro grupos tipológicos.

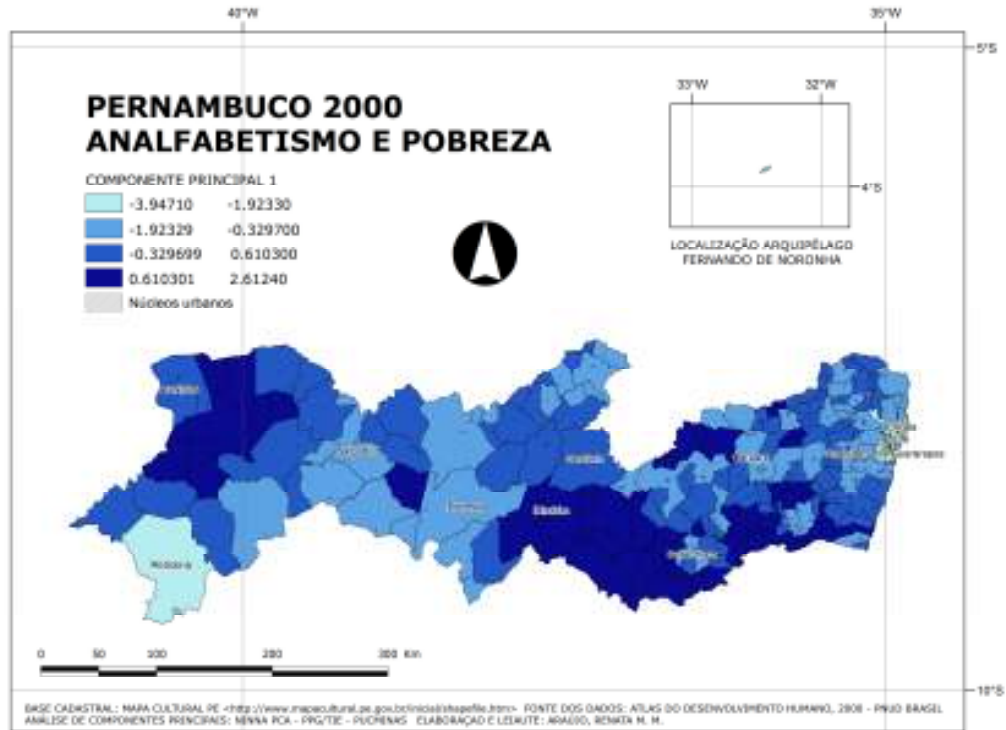
Os mapeamentos serão apresentados a seguir.

## VII INTERPRETAÇÃO DOS RESULTADOS OBTIDOS

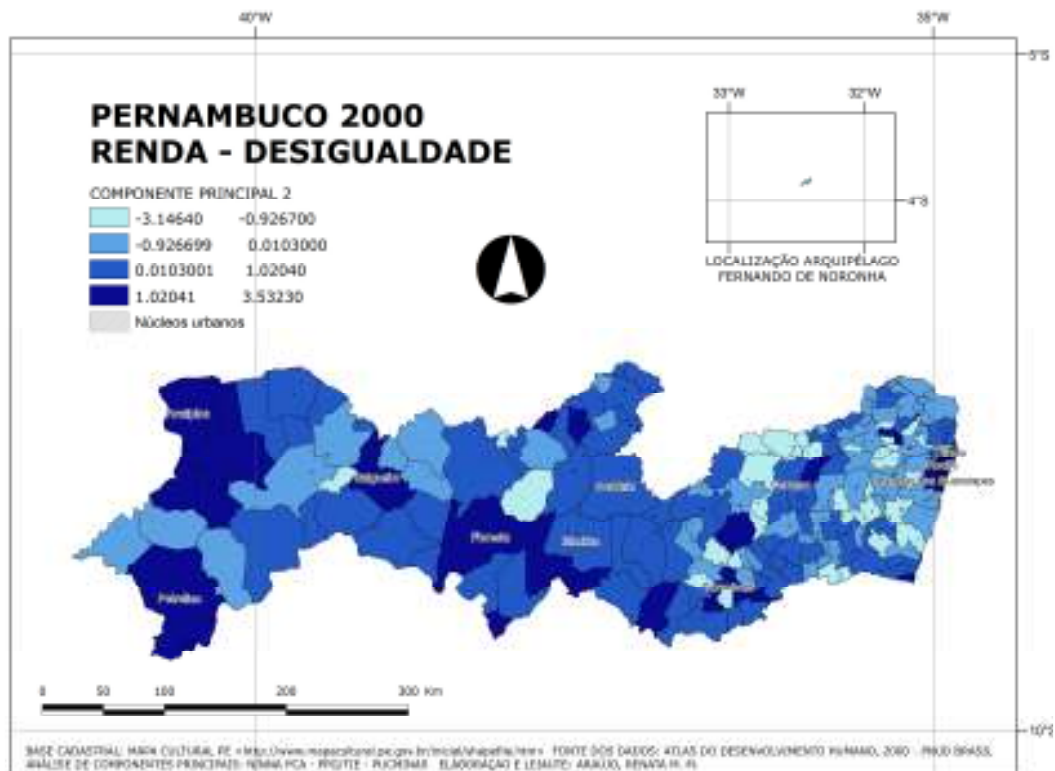
Os mapas temáticos elaborados com os escores obtidos na análise de componentes principais apontam para a Região do Agreste Meridional, Zona da Mata Sul e do Sertão do Araripe como regiões com mais altos valores para o conjunto das variáveis agrupadas na primeira componente principal, ou seja, piores situações referentes ao analfabetismo e à pobreza.

Na classe com escores mais baixos estão cidades da Região Metropolitana do Recife, o arquipélago Fernando de Noronha e a cidade de Petrolina com baixos níveis de analfabetismo e pobreza. O município de Garanhuns, localizado no Agreste Meridional, apresenta-se como uma ilha de situação melhor em relação à sua região. No entanto, os municípios de Garanhuns e Petrolina, segundo a componente principal 2, que trata de desigualdade de renda, apresentaram escores elevados. Bem como, os municípios da Região Metropolitana do Recife, e do Sertão do Araripe, dentre outras cidades das quais podemos citar Salgueiro e Floresta, que no MAPA 2 foram alocados na segunda classe melhor, em relação ao analfabetismo e à pobreza, e no MAPA3 pertencem à classe com mais alta apropriação de renda pela população mais rica.



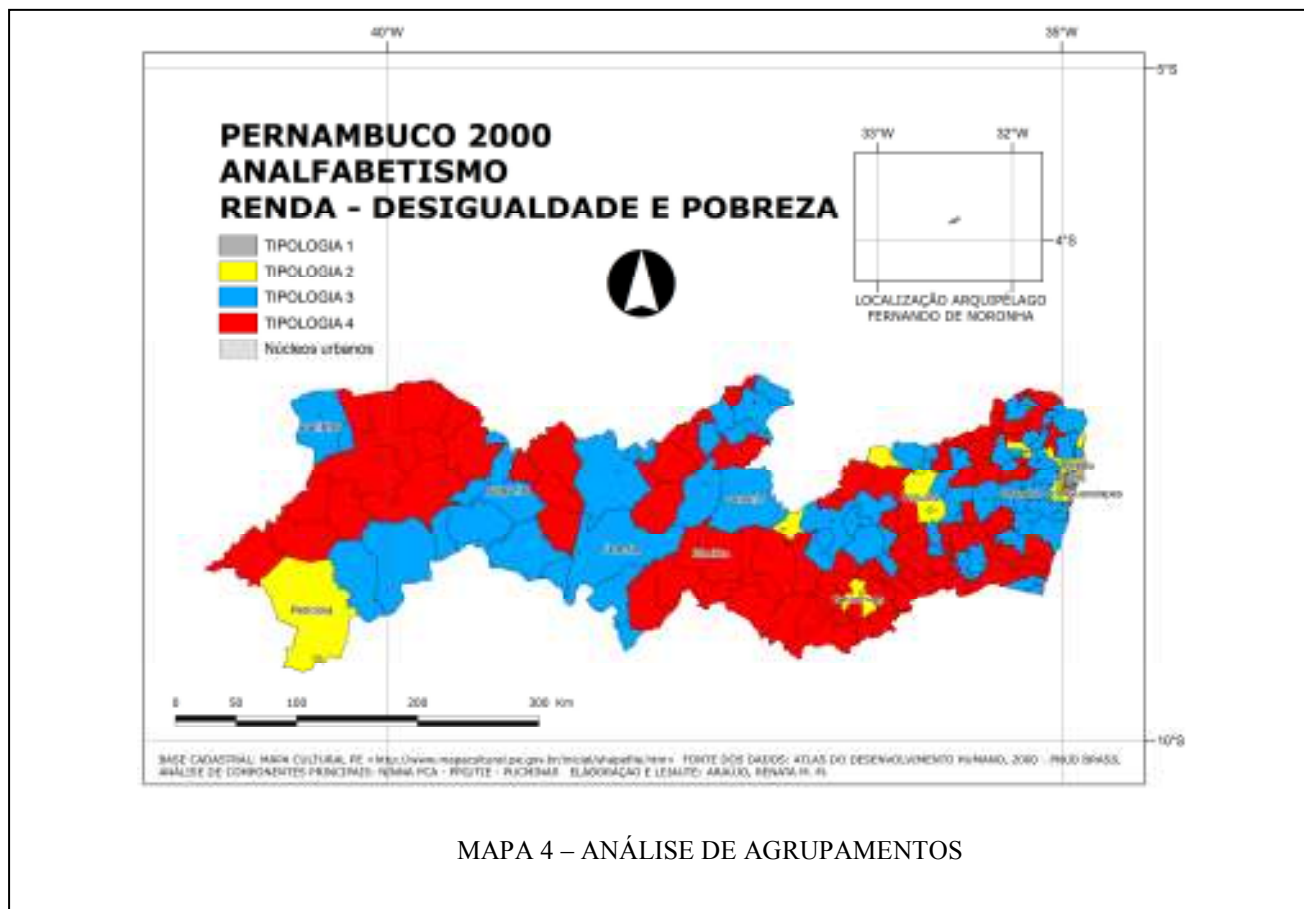


MAPA 2 – COMPONENTE PRINCIPAL 1



MAPA 3 – COMPONENTE PRINCIPAL 2





O mapa tipológico elaborado com base na análise de *clusters*, realizada com o método *k-means* para formação de quatro agrupamentos, muito se assemelha ao mapa gerado com os escores da primeira componente principal, com quatro classes hierárquicas. Ressalta-se que os grupos tipológicos obtidos na análise de *clusters* não apresentam níveis hierárquicos, sendo seu objetivo apontar apenas tipologias baseadas na semelhança em relação ao conjunto de variáveis em estudo. Um dos grupos criados nessa análise foi composto apenas pela capital do estado, Recife e o arquipélago Fernando de Noronha, o que pode causar certa estranheza, porém, é importante destacar que os dados de população total não foram considerados no estudo.

Os outros grupos tipológicos apontados na análise se apresentam condizentes com os resultados apontados pela Análise de Componentes Principais.

Cidades mais bem estruturadas do Estado, como Petrolina, Caruaru, Garanhuns, Olinda e Jaboatão dos Guararapes fazem parte de um mesmo grupo, o que demonstra coerência com a realidade.

#### VIII CONSIDERAÇÕES FINAIS

As técnicas utilizadas mostraram-se satisfatórias para os objetivos propostos. O fato da análise não ter incluído os dados da população pode ter prejudicado a análise e interpretação dos resultados, porém, não comprometeu o estudo, que demonstrou coerência com a realidade, o que poderia ser mais bem aferido com trabalhos de campo.

Um maior aprofundamento em relação à origem dos percentuais utilizados em muitas variáveis escolhidas também poderia ser proposto como forma de interpretar os

mapeamentos aqui obtidos, porém isso não fez parte do escopo proposto para esse estudo.

As técnicas utilizadas facilitaram a análise do fenômeno escolhido – analfabetismo, pobreza e desigualdade de renda no Estado - sob a ótica de um número elevado de variáveis. A Análise de Componentes Principais é capaz de condensar os dados num conjunto menor de variáveis, facilitando o mapeamento. A Análise de agrupamentos organiza as ocorrências em grupos semelhantes, apontando tipologias. As duas técnicas se mostram, portanto, complementares para análise multivariada de dados geográficos.

#### REFERÊNCIAS

ALENCAR, Bernardo Jeunon, **A análise multivariada aplicada ao tratamento da informação espacial: uma abordagem matemático-computacional em análise de agrupamentos e análise de componentes principais**. 2009. 200f. Tese (Doutorado) –

Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Geografia- Tratamento da Informação Espacial.

GERARDI, L. H. O.; SILVA, B. C. N., **Quantificação em Geografia**. Difel, São Paulo, SP, 1981. 161 p.

MIRANDA, E. E.; GOMES, E. G. e GUIMARÃES, M. **Mapeamento e estimativa da área urbanizada do Brasil com base em imagens orbitais e modelos estatísticos**. Campinas: Embrapa Monitoramento por Satélite, 2005. Disponível em: <http://www.urbanizacao.cnpm.embrapa.br> Acesso em: 2 jul. 2010.

ROGERSON, P. A., **Statistical Methods for Geography**. SAGE Publications, London,UK , 2002. 236 p.

