

Uma análise multivariada da população ocupada nas regiões metropolitanas brasileiras*

Lucia Silva Kubrusly

Professora Adjunta do Instituto de Economia da
Universidade Federal do Rio de Janeiro

João Saboia

Professor Titular do Instituto de Economia da
Universidade Federal do Rio de Janeiro

Resumo

Este artigo apresenta uma análise das características da população ocupada nas regiões metropolitanas brasileiras através do estudo das associações entre as variáveis escolhidas e as semelhanças entre essas regiões. São utilizados métodos de análise estatística multivariada, que permitem uma visão de conjunto tanto do ponto de vista das variáveis envolvidas quanto do das regiões metropolitanas. Alguns resultados eram esperados, confirmando as semelhanças entre as regiões metropolitanas das Regiões Sul e Sudeste, por um lado, e entre as das Regiões Norte e Nordeste, por outro. Outros, entretanto, são surpreendentes, apontando uma certa proximidade entre a Região Metropolitana de Fortaleza e algumas regiões metropolitanas das Regiões Sul e Sudeste e entre a Região Metropolitana do Rio de Janeiro e algumas regiões metropolitanas das Regiões Norte e Nordeste. Em termos gerais, o artigo confirma os desníveis regionais existentes no País.

Palavras-chave

Estatística econômica; ocupação nas regiões metropolitanas; análise multivariada.

* Artigo recebido em nov. 2004 e aceito em jul. 2006.

Abstract

This paper analyzes occupation in brazilian metropolitan regions based on social-economics variables. Methods of multivariate statistical analysis are used to show associations among the variables and regional similarities. Although some results were unexpected, in general, brazilian regional differences are confirmed.

Key words

Economics statistics; occupation in brazilian metropolitan regions; multivariate statistical analysis.

Classificação JEL: J21, O18, C4.

Este artigo está motivado pela quantidade relativamente pequena de estudos dirigidos à análise do mercado de trabalho metropolitano do País que levem em consideração as diferenças existentes entre as várias metrópoles. Embora existam muitos estudos sobre as regiões metropolitanas, eles se limitam, em geral, à análise agregada do conjunto de regiões, ou ao estudo individual de uma determinada região.¹

Recentemente, foi levantada a hipótese de que o mercado de trabalho nas regiões metropolitanas estaria apresentando comportamento diferenciado relativamente ao que ocorre no interior do País. Tal discussão tem como referência a diferença de resultados encontrados pela **Pesquisa Mensal de Emprego** (PME) do IBGE e o **Cadastro Geral de Empregados e Desempregados** (Caged) do Ministério do Trabalho e Emprego (TEM). Enquanto os dados do IBGE apontam dificuldades na geração de emprego nas principais metrópoles do País, o Caged mostra forte geração de empregos formais no interior.²

¹ Para um exemplo do primeiro tipo de análise, ver Ramos e Britto (2004). Um exemplo do segundo tipo de análise é Toni (2004).

² Para uma discussão sobre as mudanças espaciais no mercado de trabalho brasileiro, ver, por exemplo, Ramos e Ferreira (2005). Sobre a interiorização do emprego industrial, ver, por exemplo, Saboia (2001).

O principal objetivo deste artigo é aprofundar o estudo sobre o mercado de trabalho metropolitano, diferenciando os resultados encontrados nas diversas regiões do País. Em outras palavras, não se pretende discutir os diferenciais existentes entre as regiões metropolitanas e o interior, mas entre as diferentes regiões metropolitanas. A hipótese básica que se deseja testar é a existência de grandes diferenciais entre as regiões, de modo que seria possível separá-las em grupos com características semelhantes. Tal resultado é muito mais interessante e estimulante do que a tradicional divisão do País entre as regiões metropolitanas menos desenvolvidas, das Regiões Norte e Nordeste, de um lado, e as mais desenvolvidas das Regiões Sul e Sudeste, de outro, podendo, inclusive, confirmar, ou não, tal corte regional.

Este trabalho apresenta uma análise das características da população ocupada nas regiões metropolitanas brasileiras, no que se refere à renda, à escolaridade, às condições de trabalho e à distribuição das ocupações nos diferentes setores da economia. Para tal, são realizadas análises estatísticas procurando identificar associações entre as variáveis escolhidas e semelhanças e diferenças entre as regiões metropolitanas. As primeiras estatísticas apresentadas possibilitam uma visão detalhada de cada variável e de cada região metropolitana. Posteriormente, são utilizados métodos de análise estatística multivariada, os quais permitem uma visão de conjunto do ponto de vista tanto das variáveis envolvidas quanto das regiões metropolitanas.

A análise estatística realizada neste trabalho tem como base dados da **Pesquisa Nacional por Amostra de Domicílios** (PNAD) do IBGE. Essa pesquisa é realizada anualmente (exceto naqueles em que há **Censo Demográfico**), em todas as unidades da Federação brasileira, incluindo as nove regiões metropolitanas. Apenas as áreas rurais da Região Norte não são cobertas pela PNAD. O mesmo tipo de metodologia deste artigo já foi utilizado antes, com base nos dados da PNAD do ano de 1999, a partir de um amplo conjunto de variáveis sobre as condições de vida da população — trabalho, rendimento, educação, infra-estrutura e condições de moradia — considerando-se as regiões metropolitanas e as unidades da Federação.³ O presente estudo realiza uma análise focalizada na questão da mão-de-obra e da ocupação nos diferentes setores da economia, considerando 10 regiões — as nove regiões metropolitanas e o Distrito Federal (DF) — e utilizando a base de dados de 2001.

O artigo consiste de cinco seções. Na primeira, está a descrição das variáveis selecionadas para análise. A seguir, estão as análises estatísticas

³ Ver Kubrusly e Barros (2003) e Corrêa e Kubrusly (2002).

preliminares, incluindo, aí, a análise das correlações e alguns diagramas de dispersão. Na terceira e na quarta seções, apresentam-se os métodos de análise multivariada e seus resultados. Nessa ocasião, pode-se ver que as análises preliminares têm um importante papel na interpretação dos resultados. Para finalizar, na seção 5, apresentam-se as principais conclusões do trabalho.

1 Base de dados

Com o objetivo de analisar as condições de trabalho e a atuação dos setores econômicos nas regiões metropolitanas, foram selecionadas 11 variáveis, sete delas descrevendo as características de mão-de-obra e das condições de trabalho, e quatro descrevendo a distribuição das ocupações por setores da economia. Os objetos de observação foram as Regiões Metropolitanas de Belém (RMB), Fortaleza (RMF), Recife (RMR), Salvador (RMS), Belo Horizonte (RMBH), Rio de Janeiro (RMRJ), São Paulo (RMSP), Curitiba (RMC), Porto Alegre (RMPA) e o Distrito Federal. Todas as informações utilizadas foram extraídas da **PNAD-2001** (2002).

Abaixo estão definidas as variáveis selecionadas para análise.⁴

1.1 Conjunto 1 — características da população ocupada

Proporção dos trabalhadores por conta própria (*prcp*)

É fato conhecido que, ao longo da última década, houve aumento na proporção de trabalhadores por conta própria, devido às dificuldades do mercado de trabalho com a redução da oferta de emprego. Assim, esse é um importante indicador de fragilidade nas condições de trabalho e da importância do setor informal.

$$\text{prcp} = (\text{pop10} + \text{prcp}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + prcp é a população com 10 anos ou mais de idade que trabalha por conta própria;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

⁴ Além das variáveis incluídas na análise, outras foram testadas e descartadas, por não incorporarem informação adicional relevante em termos estatísticos.

Proporção de empregados (prem)

Esse é o indicador da oferta global de emprego.

$$\text{prem} = (\text{pop10} + \text{empregada}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + empregada é a população com 10 anos ou mais de idade ocupada, cuja posição no trabalho é a de empregado;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

Proporção de empregados com carteira de trabalho assinada (precc)

$$\text{precc} = (\text{pop10} + \text{empr.c.cart}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + empr.c.cart é a população com 10 anos ou mais de idade ocupada, cuja posição no trabalho é a de empregado, com carteira assinada;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

Proporção de empregados sem carteira de trabalho assinada (presc)

$$\text{presc} = (\text{pop10} + \text{empr.s.cart}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + empr.s.cart é a população com 10 anos ou mais de idade ocupada, cuja posição no trabalho é a de empregado, sem carteira assinada;

pop10 + ocupada é a população, com 10 anos ou mais de idade ocupada.

Renda média mensal em salários mínimos, de todos os trabalhos, das pessoas ocupadas com 10 anos ou mais de idade (renda)

Inquestionavelmente, trata-se de um dos principais indicadores do mercado de trabalho.

Média do número de anos de estudo (mest)

Esse é o indicador da escolaridade da população de cada região metropolitana.

Taxa de desocupação (txdes)

O desemprego vem apresentando tendência de crescimento desde os anos 90. A variável txdes é o principal indicador para o fenômeno de desemprego, calculada conforme a definição abaixo.

$$\text{txdes} = 1 - [(\text{pop10} + \text{ocupada}) / (\text{pop10} + \text{econ.ativa})]$$

onde:

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada;

pop10 + econ.ativa é a população com 10 anos ou mais de idade economicamente ativa.

1.2 Conjunto 2 — distribuição das ocupações nos setores econômicos

Proporção dos trabalhadores na indústria (exceto construção civil) (prind)

$$\text{prind} = (\text{pop10} + \text{ind}) / (\text{pop10} + \text{ocup})$$

onde:

pop10 + indus é a população com 10 anos ou mais de idade ocupada na indústria, excluindo trabalhadores da construção civil;⁵

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

Proporção dos trabalhadores na indústria de construção civil (prcnst)

$$\text{prcnst} = (\text{pop10} + \text{const}) / (\text{pop10} + \text{ocup})$$

onde:

⁵ Foi excluída a construção civil por suas características distintas do restante da indústria.

pop10 + const é a população com 10 anos ou mais de idade ocupada na indústria da construção civil;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

Proporção dos trabalhadores no Setor Terciário (serviços e comércio) (prsrc)

$$\text{prsrc} = (\text{pop10} + \text{sercom}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + sercom é a população com 10 anos ou mais de idade ocupada nos serviços ou no comércio;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

Proporção dos trabalhadores na administração pública (pradm)

$$\text{pradm} = (\text{pop10} + \text{admpubl}) / (\text{pop10} + \text{ocupada})$$

onde:

pop10 + admpubl é a população com 10 anos ou mais de idade ocupada na administração pública;

pop10 + ocupada é a população com 10 anos ou mais de idade ocupada.

2 Análise preliminar

Nesta seção, são apresentadas algumas estatísticas descritivas e a análise das correlações. A Tabela 1 mostra as variáveis observadas individualmente para as regiões metropolitanas, as quais constituem a base de dados deste trabalho.

A Tabela 2 apresenta os valores mínimos, máximos, médios e também os desvios padrões das variáveis escolhidas. Ao lado de cada valor mínimo ou máximo, está indicada a região metropolitana correspondente.

Tabela 1

Variáveis observadas nas regiões metropolitanas e no DF — 2001

REGIÕES METROPOLITANAS E DISTRITO FEDERAL	prcp	prem	precc	presc	renda	mest
RMB	0,290	0,540	0,277	0,176	2,916	6,971
RMF	0,233	0,572	0,296	0,219	2,617	6,208
RMR	0,238	0,604	0,318	0,196	2,956	6,637
RMS	0,247	0,580	0,350	0,166	3,083	6,976
RMBH	0,187	0,631	0,405	0,159	3,378	7,066
RMRJ	0,232	0,612	0,385	0,137	4,239	7,353
RMSP	0,183	0,674	0,446	0,174	5,144	7,418
RMC	0,218	0,601	0,404	0,138	3,978	7,161
RMPA	0,197	0,623	0,406	0,154	4,017	7,273
DF	0,136	0,696	0,357	0,177	5,683	7,698

REGIÕES METROPOLITANAS E DISTRITO FEDERAL	txdes	prind	prcnst	prsrc	pradm
RMB	0,141	0,090	0,090	0,714	0,065
RMF	0,119	0,162	0,068	0,654	0,047
RMR	0,140	0,106	0,069	0,704	0,067
RMS	0,156	0,098	0,089	0,714	0,053
RMBH	0,126	0,159	0,089	0,647	0,046
RMRJ	0,127	0,108	0,072	0,715	0,062
RMSP	0,132	0,196	0,059	0,662	0,036
RMC	0,094	0,160	0,079	0,608	0,050
RMPA	0,087	0,202	0,068	0,608	0,048
DF	0,145	0,069	0,060	0,678	0,139

FONTE: PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIO 2001: Regiões Metropolitanas — PNAD. Rio de Janeiro: IBGE, 2002.

Tabela 2

Valores mínimo, máximo, médio e desvio padrão das variáveis observadas nas regiões metropolitanas e no DF — 2001

VARIÁVEIS	VALOR MÍNIMO	VALOR MÁXIMO	MÉDIA	DESVIO PADRÃO
prcp	(1) 0,136	(2) 0,290	0,216	0,042
prem	(2) 0,540	(1) 0,696	0,613	0,046
precc	(2) 0,277	(3) 0,446	0,364	0,055
presc	(4) 0,137	(5) 0,219	0,169	0,025
renda	(5) 2,617	(1) 5,683	3,801	1,011
mest	(5) 6,208	(1) 7,698	7,076	0,420
txdes	(6) 0,087	(7) 0,156	0,127	0,022
prind	(1) 0,069	(6) 0,202	0,135	0,046
prcnst	(3) 0,059	(8) 0,090	0,075	0,012
prsrc	(9) 0,608	(4) 0,715	0,671	0,042
pradm	(3) 0,036	(1) 0,139	0,061	0,029

FONTE DOS DADOS BRUTOS: Tabela 1.

(1) DF. (2) RMB. (3) RMSP. (4) RMRJ. (5) RMF. (6) RMPA. (7) RMS. (8) RMBH. (9) RMC e RMPA.

Na leitura das colunas referentes aos valores mínimos e máximos da Tabela 2, chama atenção o Distrito Federal, para o qual ocorrem quatro valores máximos e dois valores mínimos. Dentre esses, destacam-se a maior média de anos de estudo e o maior nível de renda, além da elevada participação na administração pública. Observa-se também que RMPA apresenta o menor valor para taxa de desocupação, enquanto a RMS possui a taxa mais alta. A RMSP destaca-se pelo maior percentual de empregados com carteira assinada e pela pequena parcela do emprego na administração pública. A RMRJ possui a mais elevada taxa de ocupação no Setor Terciário e apresenta o menor percentual de empregados sem carteira assinada. A RMB possui o maior percentual de trabalhadores por conta própria e o menor percentual de empregados com carteira de trabalho assinada. Acrescentando-se o fato de que essa região também detém o maior percentual de empregos no setor da construção civil (um dos mais atrasados da economia), completa-se o quadro precário das suas condições de trabalho. Precária também é a situação da população da RMF, com a menor renda mensal, a menor escolaridade, e o maior percentual de empregos sem carteira de trabalho assinada.

Quanto aos valores médios, observa-se que a proporção de trabalhadores por conta própria é pouco mais de um terço da proporção de empregados e que

o percentual de empregos sem carteira de trabalho assinada é pouco menos da metade do percentual de empregos com carteira de trabalho assinada. A média dos anos de estudo revela pouca variação entre as regiões estudadas (desvio padrão em torno de 6% do valor médio), ao contrário da renda, que apresenta um desvio padrão de, aproximadamente, 25% do valor médio, evidenciando as disparidades regionais quanto aos rendimentos mensais. Observa-se também que a proporção média da ocupação no setor serviços é mais que o triplo da observada no setor indústria (indústria mais construção civil). E, finalmente, na administração pública, são baixas as proporções de ocupação, exceto, obviamente, no Distrito Federal, conforme mencionado acima.

A fim de identificar associações entre as variáveis, apresentam-se, na Tabela 3, os coeficientes de correlação entre as 11 variáveis escolhidas.

Fazendo a leitura da matriz da esquerda para direita, vê-se, conforme o esperado, forte associação inversa (correlação negativa) entre a proporção de empregados e a proporção de trabalhadores por conta própria. Observam-se a associação direta do emprego com carteira assinada e o emprego e a associação inversa do emprego sem carteira com o emprego com carteira de trabalho assinada. Analisando a variável renda, constata-se a tendência de as regiões com maior renda média apresentarem menor proporção de trabalhadores por conta própria e maior proporção de empregados. Nota-se também, em menor grau, a associação direta da renda com o emprego com carteira assinada e inversa com emprego sem carteira assinada. Essa é uma evidência da fragilidade das condições de trabalho representadas pela ausência de vínculo formal de trabalho e pela baixa renda. No sentido de detalhar melhor essas informações, apresentam-se, na Figura 1, dois diagramas de dispersão: um com as variáveis renda média e proporção de empregados e outro com as variáveis renda média e trabalhadores por conta própria.

A informação adicional do diagrama (Figura 1) está na posição relativa das regiões metropolitanas no que se refere aos valores da renda (eixo vertical, em salários mínimos) e à proporção de trabalhadores empregados e trabalhadores por conta própria. As maiores rendas estão no Distrito Federal e na RMSP, seguidos pelas RMRJ, RMC e RMPA, e as menores rendas ocorrem na RMBH, na RMS, na RMR, na RMB e, por último, na RMF. A região com maior proporção de empregados é o Distrito Federal, e é lá também que ocorre a menor proporção de trabalhadores por conta própria. Por outro lado, a região que apresenta maior proporção de trabalhadores por conta própria é a na RMB, onde também se encontra a menor proporção de empregados. A oposição entre empregados e trabalhadores por conta própria está nitidamente mostrada na Figura 1, sendo um diagrama quase o espelho perfeito do outro.

Tabela 3

Matriz de correlação das variáveis observadas nas regiões metropolitanas e no DF — 2001

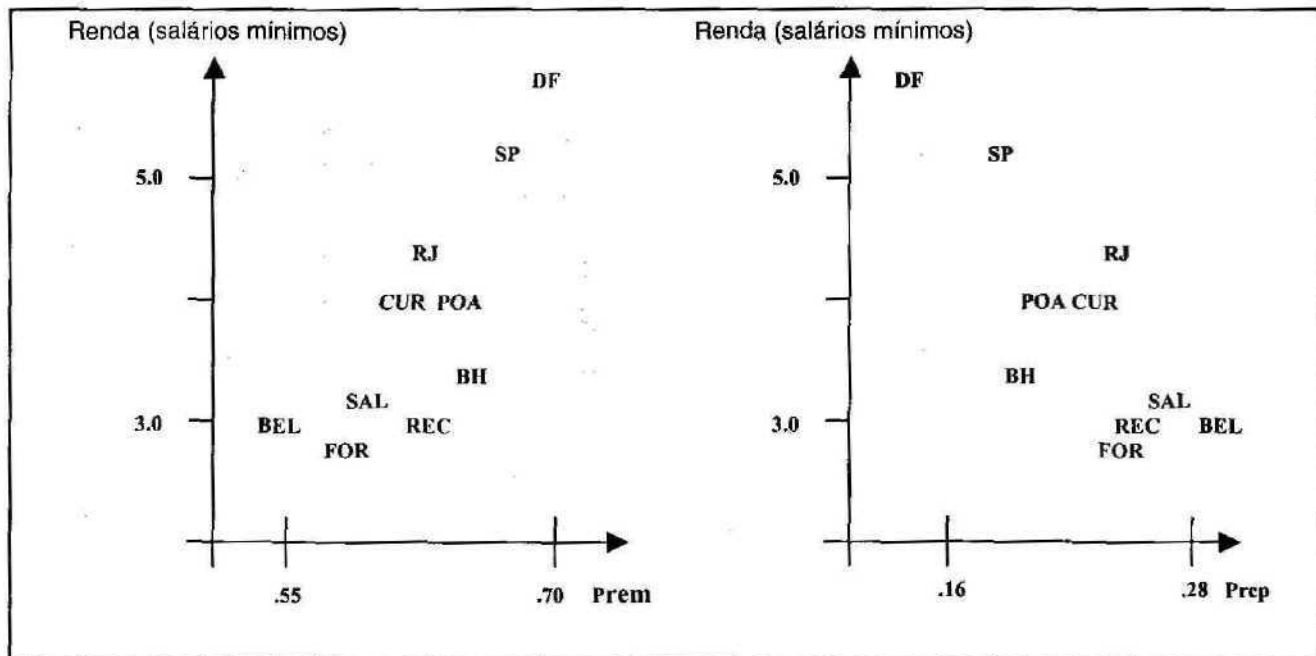
VARIÁVEIS	prcp	prem	precc	presc	renda	mest	txdes	prind	prnst	prsrc	pradm
prcp	1	-0,946	-0,615	0,110	-0,795	-0,598	0,150	-0,210	0,618	0,445	-0,428
prem		1	0,666	-0,158	0,888	0,707	-0,008	0,141	-0,677	0,445	-0,428
precc			1	-0,634	0,628	0,635	-0,419	0,595	-0,271	-0,522	-0,269
presc				1	-0,359	-0,662	0,355	-0,096	-0,264	0,202	0,109
renda					1	0,882	-0,053	0,029	-0,638	-0,184	0,486
mest						1	-0,018	-0,087	-0,256	-0,076	0,422
txdes							1	-0,736	0,190	0,867	0,384
prind								1	-0,229	-0,779	-0,718
prnst									1	0,229	-0,318
prsrc										1	0,281
pradm											1

FONTE DCS DADOS BRUTOS: Tabela 1.

NOTA: Em negrito, estão as correlações maiores que 0, em nível de significância de 0,05.

Figura 1

Diagrama de dispersão de renda, emprego e trabalho por conta própria nas regiões metropolitanas e no DF — 2001



FONTE: Tabela 1.

As próximas correlações significativas da Tabela 3 associam renda, emprego e emprego com carteira assinada diretamente com a escolaridade do trabalhador. Assinala-se também a associação inversa da escolaridade com o percentual de trabalhadores empregados sem carteira assinada. A próxima variável, a taxa de desocupação, apresenta correlação negativa com a ocupação na indústria e correlação positiva com ocupação em serviços e comércio, isto é, as regiões com maior peso na ocupação industrial são as que apresentam menor taxa de desemprego e, por outro lado, há uma tendência de ocorrer maior desocupação onde predominam as ocupações no comércio e em serviços. Portanto a queda das ocupações industriais e o crescimento das ocupações no comércio e em serviços ocorridos nos últimos anos parecem ter deixado um saldo negativo de ocupações. A visualização das associações entre a taxa de desocupação (eixo vertical) e a proporção da ocupação nos dois setores está representada, na Figura 2, pelos respectivos diagramas de dispersão.

As maiores taxas de desocupação estão no Distrito Federal e nas metrópoles da Região Nordeste, enquanto as menores estão na Região Sul. As regiões metropolitanas com maior peso de ocupações industriais são RMPA, RMC, RMBH e RMSP. Por outro lado, as ocupações em serviços e no comércio têm mais peso na RMS, na RMR, na RMB e na RMRJ.

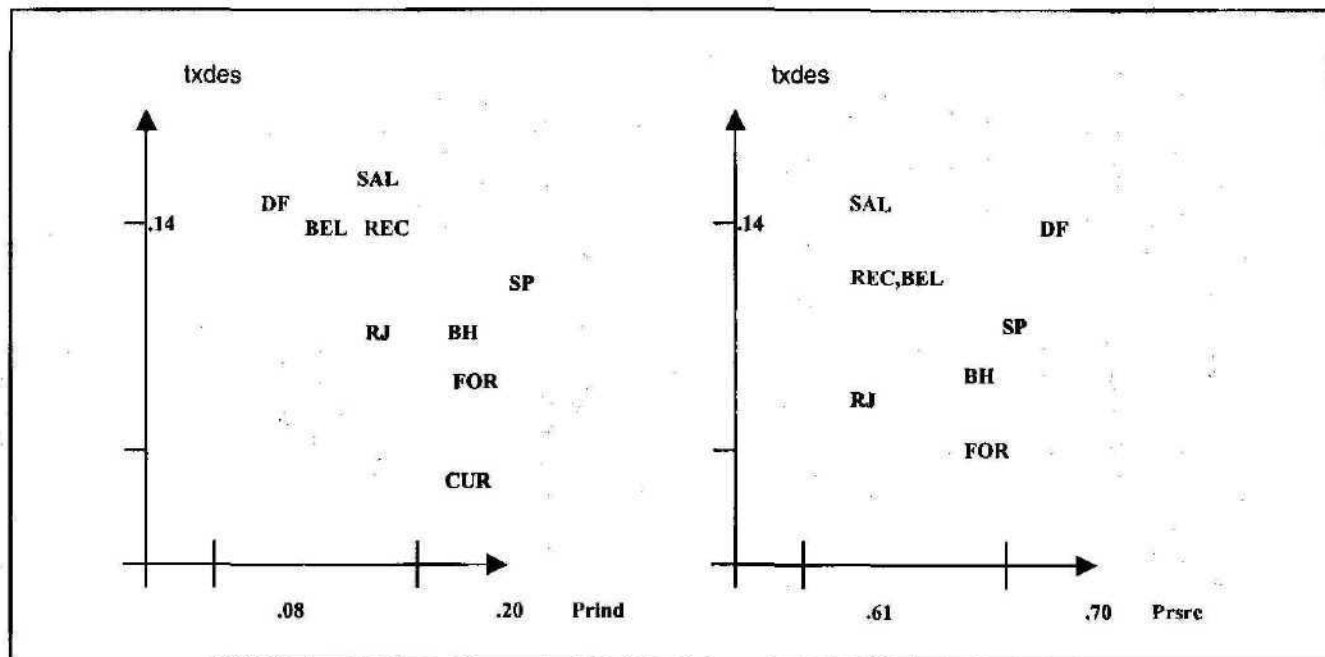
Outro resultado importante é a correlação negativa entre a proporção de ocupados na construção civil e o nível de renda e de escolaridade. Tal fato justifica plenamente a separação desse setor do restante da indústria.

Fechando a matriz de correlação, podem-se analisar as associações entre as ocupações nos setores da indústria (exceto construção civil), serviços e comércio e administração pública. Observa-se a tendência de as regiões que ocupam mais pessoas na indústria ocuparem menos em serviços e comércio e na administração pública. Esse resultado era esperado e aponta uma possível classificação das regiões metropolitanas quanto à sua atuação em setores econômicos, indicando a existência de regiões voltadas para a indústria e outras com predomínio de ocupações nos serviços e na administração pública.

Na seção 3, serão apresentados os modelos de análise estatística multivariada empregados neste trabalho.

Figura 2

Diagrama de dispersão de desocupação nos setores econômicos das regiões metropolitanas e do DF — 2001



FONTE: Tabela 1.

3 Metodologia — os modelos de análise estatística multivariada

A utilização de técnicas de análise multivariada tem-se mostrado eficiente em diversas áreas do conhecimento, pois permite uma visão mais global dos fenômenos estudados. Neste trabalho, usam-se dois métodos de análise multivariada: um focalizando as variáveis e o outro focalizando os objetos. O primeiro método empregado é a Análise de Componentes Principais (ACP), que se baseia na matriz de correlações. O segundo é a Análise de Grupamento, baseado na matriz de distâncias (no espaço das variáveis) entre os objetos (as regiões metropolitanas).

3.1 O modelo de Análise de Componentes Principais

A Análise de Componentes Principais tem como objetivo criar novas variáveis, formadas por combinações lineares das variáveis iniciais, de forma a obter uma descrição mais sucinta dos dados. Para isso, essas combinações lineares (componentes) são escolhidas de forma que contenham a maior parte da variância total dos dados. Muitas vezes, com apenas poucos componentes, consegue-se descrever uma grande proporção da variância total. O modelo de Análise de Componentes Principais pode ser descrito da seguinte forma:⁶

Seja $X = (X_1, \dots, X_p)$ um conjunto de variáveis observadas sobre n objetos.

Os componentes principais C_i são definidas como:

$$C_i = \sum_j a_{ij} X_j \text{ sujeito a:}$$

$$\text{var}(C_i) = \text{máxima}$$

$$\sum_j a_{ij}^2 = 1$$

$$\text{corr}(C_i, C_{i'}) = 0 \text{ para } i \neq i', i = 1, \dots, p$$

⁶ Ver Johnson e Wichern (1992).

A solução desse modelo é dada pela decomposição da matriz de correlação (ou covariância) em seus autovalores e autovetores. Note-se que, devido à restrição de não-correlação entre os componentes, as variâncias são decrescentes com a ordem do componente. Assim, o primeiro componente tem a maior variância, seguida do segundo componente, etc. Frequentemente, grande parte da variância total é descrita com os primeiros componentes principais. Como a variância do conjunto depende das correlações entre as variáveis originais, quanto mais forte for a estrutura das correlações, tanto menos componentes serão necessários para descrever grande parte da variância total.

A aplicação dessa técnica possibilita identificar grupos de variáveis correlacionadas, e, muitas vezes, pela interpretação dos resultados, é possível reconhecer as principais dimensões do conjunto de dados analisado.

3.2 O modelo de Análise de Grupamento

A Análise de Grupamento é uma técnica de análise estatística multivariada que procura agrupar objetos semelhantes, segundo um critério definido, a partir do conjunto de variáveis observadas. O modelo de Análise de Grupamento pode ser descrito da forma como segue.⁷

Sejam $X = \{X_1, \dots, X_p\}$ um conjunto de variáveis e o conjunto de objetos que se deseja agrupar.

Com base no conjunto X , determinar uma partição de O em subconjuntos g_i tal que:

se $O_r \in g_i$ e $O_s \in g_i \Rightarrow O_r$ e O_s são semelhantes,

se $O_r \in g_i$ e $O_s \in g_j \Rightarrow O_r$ e O_s são distintos.

Para a solução desse problema, é necessário calcular as distâncias entre os objetos no espaço das variáveis. Essas distâncias fornecem as medidas de similaridade entre os objetos. No presente trabalho, os objetos são as regiões metropolitanas, e o objetivo da análise é agrupar regiões semelhantes segundo o critério definido pelo conjunto de variáveis selecionadas. A solução da Análise de Grupamento é apresentada em um diagrama em árvore chamado dendrograma. Nesse diagrama, é possível identificar (quando existem) os grupos de objetos semelhantes entre si, associados a uma medida de distância ou similaridade.⁸

⁷ Ver Lucas (1982).

⁸ Para mais detalhes sobre a escolha das métricas e sobre a escolha de métodos para análise de grupamento, ver Anderberg (1973).

4 Resultados

Apresentam-se, a seguir, os resultados das duas técnicas de análise multivariada. Os resultados da ACP permitiram a identificação de dois conjuntos distintos de variáveis, enquanto a Análise de Grupamento apontou semelhanças e diferenças entre as regiões metropolitanas. Os resultados foram obtidos usando-se o *software* SPSS v. 11, com os programas FACTOR/ACP e HIERARCHICAL CLUSTER/WARD.

4.1 Resultados da Análise de Componentes Principais

O método foi aplicado considerando nove das 11 variáveis observadas nas regiões metropolitanas. Isto porque, dentre as três variáveis medindo percentual de emprego (emprego global, emprego com carteira assinada e emprego sem carteira assinada), se optou pela variável que mede o emprego global, pois essa é a que apresenta maiores correlações com as demais variáveis, sendo, por isso, a mais indicada na composição da base de dados para Análise de Componentes Principais. Os primeiros resultados estão na Tabela 4, onde são mostradas as variâncias de cada componente, bem como a sua contribuição percentual para a variância total dos dados.

A terceira coluna da Tabela 4 informa que os dois primeiros componentes descrevem cerca de 79,8% da variância total das variáveis, indicando que a informação contida nos dados originais pode ser bem representada com as duas dimensões fornecidas por C1 e C2. Além disso, observa-se que os demais componentes apresentam variância menor que 1 e que ocorre também uma forte queda nos valores da variância (ou do percentual da variância) entre o segundo e o terceiro componentes. Todos esses resultados indicam que dois componentes são suficientes para descrever adequadamente os dados.⁹

A Tabela 5 fornece as correlações entre as variáveis e os dois primeiros componentes principais. Examinando esses valores, é possível interpretar os componentes de acordo com as variáveis que apresentam maiores correlações.

⁹ Para tratar da questão do número de componentes mantidas na análise, ver Jolliffe (1986).

Tabela 4

Variância dos componentes e sua participação percentual sobre a variância total dos dados das variáveis observadas nas regiões metropolitanas e no DF — 2001

COMPONENTES	VARIÂNCIA DE CI	% DA VARIÂNCIA TOTAL	% ACUMULADO
C1	4,217	46,857	46,857
C2	2,964	32,934	79,791
C3	0,749	8,326	88,117
C4	0,594	6,602	94,720
C5	0,388	4,307	99,027
C6	0,053	0,586	99,613
C7	0,021	0,234	99,847
C8	0,011	0,132	99,979
C9	0,002	0,021	100,000

FONTE DOS DADOS BRUTOS: Tabela 1.

Tabela 5

Correlação entre a matriz dos componentes e as variáveis observadas nas regiões metropolitanas e no DF — 2001

VARIÁVEIS	C1	C2
prcpr	-0,931	0,045
prem	0,955	0,084
renda	0,944	0,164
mest	0,766	0,236
txdes	-0,196	0,869
prind	0,175	-0,939
prconst	-0,725	0,076
prserc	-0,382	0,843
pradm	0,467	0,721

FONTE DOS DADOS BRUTOS: Tabela 1.

NOTA: Em negrito, estão as variáveis fortemente correlacionadas com os componentes C1 e C2.

O resultado apresentado na Tabela 5 permite associar o primeiro componente a quatro variáveis que caracterizam a população ocupada e a uma variável que descreve a ocupação no setor da construção civil. Analisando os sinais dos coeficientes dessas cinco variáveis, vê-se que C1 separa as regiões metropolitanas com maiores percentuais de empregados, maiores rendas e mais anos de estudo daquelas com maiores percentuais de trabalho por conta própria e maiores percentuais de ocupação na construção civil. Considerando que este último é um dos setores mais atrasados da economia brasileira, pode-se afirmar que C1 se apresenta como um **componente de desenvolvimento**, diferenciando as regiões mais desenvolvidas daquelas menos desenvolvidas.

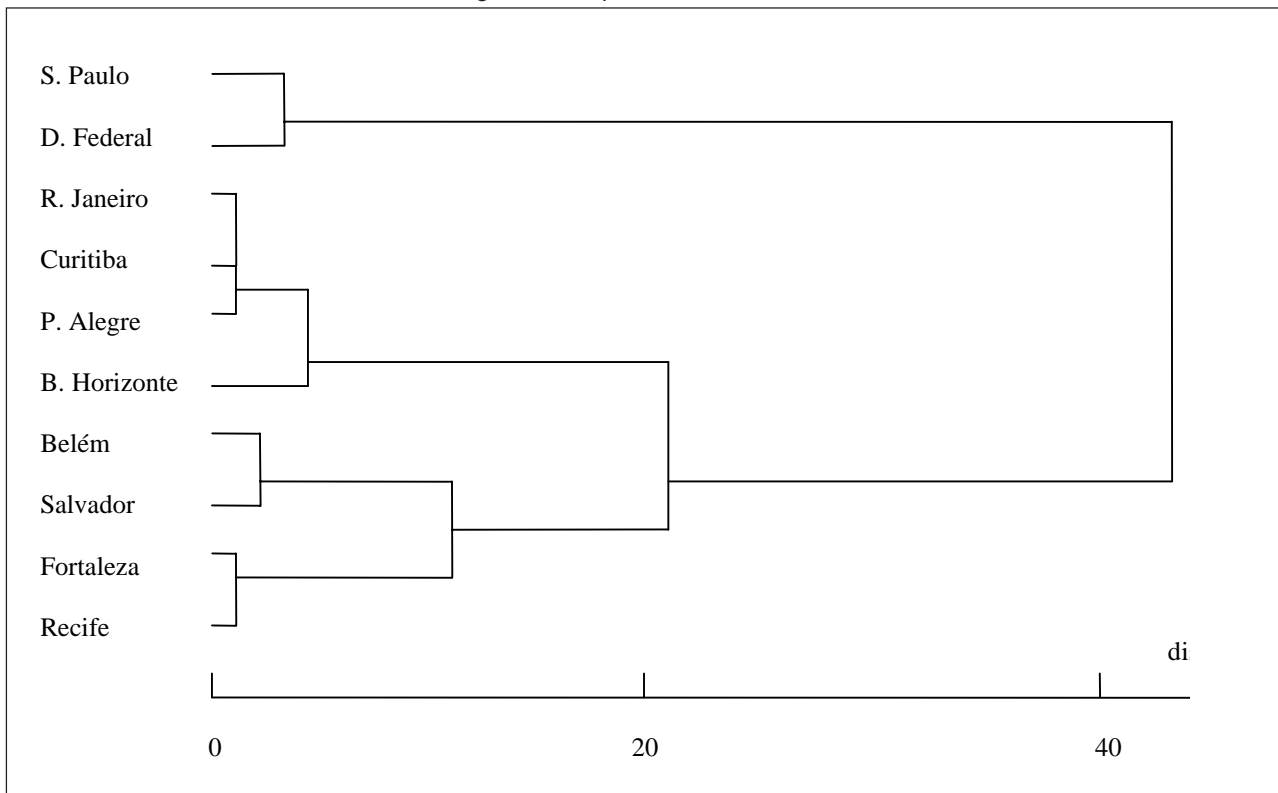
O segundo componente será chamado de **componente setorial**, já que está fortemente correlacionado com quatro variáveis, sendo que três descrevem a ocupação nos setores econômicos: indústria (exceto construção civil), serviços e comércio e administração pública. A quarta variável desse componente fornece a taxa de desocupação. O sinal negativo para a ocupação da indústria indica a oposição desta com os outros dois setores e também com a taxa de desocupação. Esses resultados confirmam a análise das correlações (Tabela 3 e Figura 2), que aponta forte associação positiva entre a desocupação e o setor serviços e comércio e, por outro lado, forte associação negativa entre a indústria e as outras três variáveis que compõem C2.

4.2. Resultados parciais da Análise de Grupamento

Pelos resultados da Análise de Componentes Principais, pode-se afirmar que dois conjuntos de variáveis fornecem as duas principais dimensões (C1 e C2) da base de dados analisada: o primeiro, formado pelo percentual de trabalhadores por conta própria, percentual de empregados, renda mensal, anos de estudo e ocupação na construção civil, compõe o **componente de desenvolvimento**; o segundo conjunto de variáveis é formado por aquelas compõem o **componente setorial** — ocupação na indústria, ocupação nos serviços e comércio, ocupação na administração pública e taxa de desocupação. Portanto, mantendo essa separação na aplicação do método de Análise de Grupamento, dois resultados serão obtidos, cada um usando um conjunto de variáveis. Assim, na primeira análise, o resultado fornece regiões semelhantes quanto ao desenvolvimento, enquanto a segunda análise fornece grupos de regiões semelhantes no que se refere, principalmente, à distribuição das ocupações pelos setores econômicos. As soluções das duas análises estão apresentadas nas Figuras 3 e 4. O eixo horizontal fornece as distâncias, que são interpretadas como medida de similaridade entre as regiões metropolitanas.

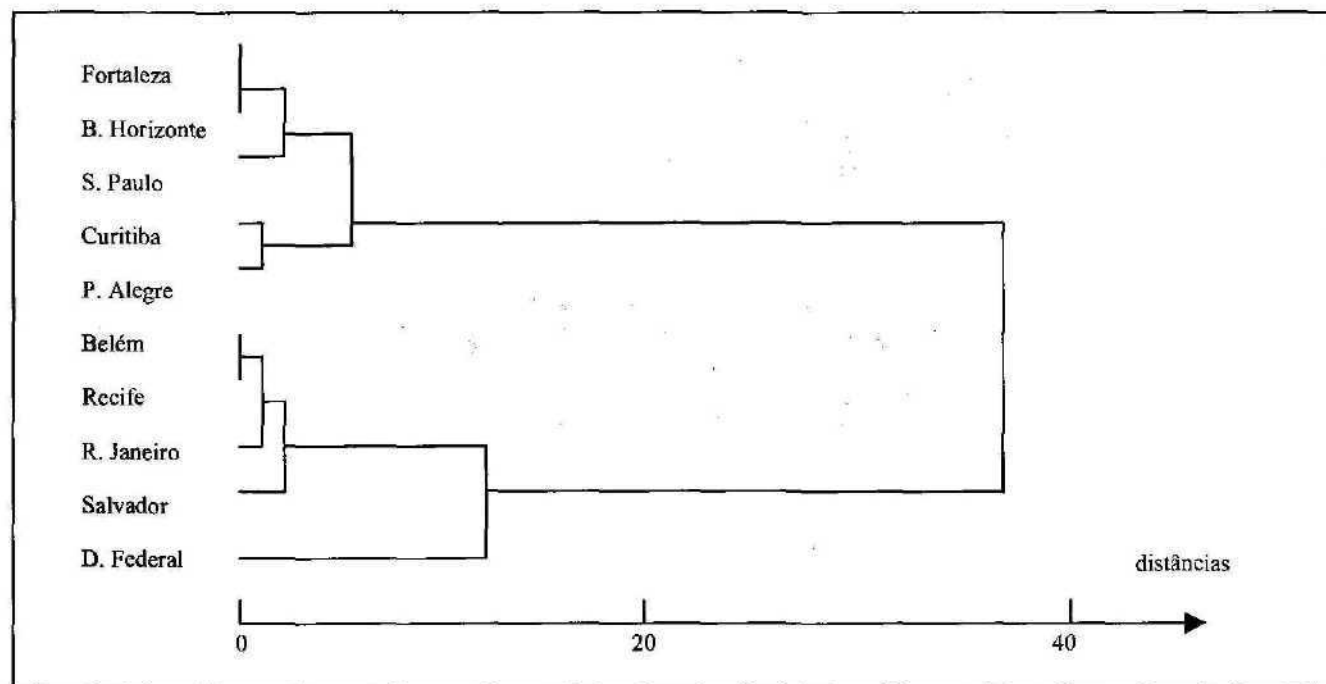
Figura 3

Análise de Grupamento das variáveis de desenvolvimento nas regiões metropolitanas e no DF — 2001



FONTE: Tabela 1.

Figura 4

Análise de Grupamento das ocupações nos setores econômicos
das regiões metropolitanas e do DF — 2001

FONTE: Tabela 1.

Análise 1 — variáveis de desenvolvimento

A Figura 3 apresenta o dendrograma correspondente às variáveis de desenvolvimento.

O dendrograma mostra a forte distinção da dupla RMSP/DF das demais regiões metropolitanas, no que se refere ao desenvolvimento. No restante das regiões, percebe-se um grupo bastante homogêneo formado pelas RMRJ, RMC, RMPA e RMBH. O outro grupo, não tão homogêneo, contém as regiões metropolitanas das Regiões Norte e Nordeste. Neste último grupo, percebe-se uma distinção entre as duplas RMB/RMS e RMF/RMR. Essa solução é coerente com a Figura 1, onde são apresentadas as rendas das regiões metropolitanas. As principais características dessa estrutura de grupos já se esboçavam na Figura 1, mostrando o destaque da RMSP e do DF, com maiores rendas, e o mesmo grupo de regiões metropolitanas das Regiões Sul e Sudeste, que aqui também se apresentam juntas. Quanto às regiões metropolitanas das Regiões Norte e Nordeste, a Análise de Grupamento mostra uma diferença entre RMB/RMS e RMF/RMR, que não é devida à renda, mas a outras variáveis presentes nessa análise.

Análise 2 — variáveis setoriais

A Figura 4 apresenta o resultado da Análise de Grupamento tendo como critério de semelhança as variáveis que fornecem informações sobre a distribuição das ocupações nos setores econômicos (indústria, serviços e comércio, administração pública) e a taxa de desocupação.

Esse resultado apresenta dois grandes grupos, um agregando a RMF, a RMBH, a RMSP, a RMC e a RMPA, e outro contendo a RMB, a RMR, a RMJ e a RMS e, um pouco diferenciado, o DF. De fato, esse resultado, embora pareça estranho, está de acordo com a Figura 2, onde são apresentadas as proporções das ocupações na indústria, em serviços e comércio e a taxa de desocupação. Com auxílio daquelas informações, pode-se afirmar que a RMB, a RMR, a RMRJ, a RMS e o DF são as regiões onde há predominância do **setor serviços e comércio** e também onde ocorrem as maiores taxas de desocupação (exceção feita à RMRJ). O afastamento do Distrito Federal no grupo de regiões caracterizadas pelas ocupações em serviços e comércio deve-se, provavelmente, às ocupações na administração pública, que, como se sabe (Tabela 1), é máxima nessa metrópole. Por outro lado, a RMF, a RMBH, a RMC, a RMSP e a RMPA são as regiões onde o **setor industrial** prevalece, e, ainda, onde ocorrem as menores taxas de desocupação (com exceção de São Paulo). Assim, a Análise de Grupamento permite a classificação das regiões metropolitanas em dois grupos, segundo a predominância do setor industrial ou de serviços e comércio.

4.3 Interpretação dos resultados globais

Os resultados encontrados na Análise de Grupamento são coerentes com os apresentados em Saboia (2000). Segundo esse estudo, que utilizou estatísticas da **Pesquisa Mensal de Emprego** de seis regiões metropolitanas no período 1991-99, existiriam dois grupos de regiões. De um lado, a RMSP, a RMRJ, a RMPA e a RMBH, onde as condições do mercado de trabalho são bem mais favoráveis; de outro, a RMR e a RMS, onde os resultados são bem piores.¹⁰

Os grupos encontrados neste trabalho apontam a mesma direção. Na primeira análise (componente desenvolvimento), a RMRJ, a RMPA e a RMBH juntam-se com a RMC para formar um grupo relativamente homogêneo, enquanto a RMR e a RMS se agregam à RMR e à RMF para formar um segundo grupo, com menores níveis de rendimento, de escolaridade e de proporção de empregados e com maior participação de trabalhadores por conta própria. Nesta análise, a RMSP cria um terceiro grupo, juntamente com o DF, por conta de suas semelhanças nas variáveis utilizadas.¹¹

Na segunda análise (componente setorial), a principal surpresa é o fato de a RMRJ juntar-se às regiões menos desenvolvidas da Região Nordeste. Tal fato, entretanto, explica-se facilmente, devido à semelhança de suas distribuições setoriais do emprego. Por outro lado, as demais regiões metropolitanas das Regiões Sul e Sudeste permanecem juntas no mesmo grupo. A surpresa, nesse caso, é a RMF fazer parte do grupo das regiões mais desenvolvidas, podendo ser explicado pela mesma razão acima.

Apesar de alguns resultados inesperados na Análise de Grupamento, em linhas gerais, eles confirmam os desequilíbrios regionais do País. A figura da Belíndia continua sendo uma representação satisfatória para o Brasil. Se, por um lado, algumas regiões metropolitanas se encontram relativamente avançadas, especialmente aquelas das Regiões Sul e Sudeste, as das Regiões Norte e Nordeste continuam apresentando resultados bastante desfavoráveis.

O caso da RMF é ilustrativo. Embora possua, atualmente, uma distribuição setorial do emprego semelhante à encontrada nas regiões metropolitanas mais desenvolvidas, resultante de uma política, do Governo local, de atração de novos investimentos, quando consideradas as variáveis de desenvolvimento do mercado de trabalho, como nível de renda, escolaridade e posição na ocupação dos trabalhadores, mostra resultados bastante insatisfatórios, semelhantes aos

¹⁰ Diferentemente da **PNAD**, a **PME** cobre apenas seis regiões metropolitanas.

¹¹ Conforme mencionado acima, o Distrito Federal não pode ser analisado com as estatísticas da **PME**.

encontrados nas demais capitais da Região Nordeste. Esse fato aponta em direção aos limites de uma política realizada apenas em nível local, sugerindo a necessidade de complementação, através de políticas nacionais de desenvolvimento.

5 Conclusão

A utilização dos modelos Análise de Componentes Principais e Análise de Grupamento em um conjunto de variáveis do mercado de trabalho das regiões metropolitanas brasileiras mostrou seu potencial, confirmando alguns resultados esperados e trazendo outros, até certo ponto, surpreendentes.

A Análise de Componentes Principais permitiu a separação das variáveis selecionadas em dois componentes. O primeiro, incluindo renda, escolaridade, proporção de empregados e de trabalhadores por conta própria, além da importância da ocupação na construção civil, foi denominado **componente de desenvolvimento**. O segundo, incluindo a participação da ocupação na indústria (exceto construção civil), no Terciário (comércio e serviços) e na administração pública, além da taxa de desemprego, foi denominado **componente setorial**.

A Análise de Grupamento foi realizada separadamente para as variáveis dos dois componentes acima, permitindo o agrupamento das regiões metropolitanas segundo as características dos respectivos mercados de trabalho.

Quando analisadas as variáveis associadas ao componente de desenvolvimento, formaram-se vários pares de regiões, destacando-se a RMSP e o DF, com comportamento claramente diferenciado das demais. A RMF mostra forte semelhança com a RMR, a RMB, com a RMS, a RMRJ, a RMC, a RMPA e a RMBH, nesta análise, formam um grupo bastante homogêneo.

Quando analisadas as variáveis associadas ao componente setorial, surgem algumas surpresas. Formam-se dois grandes grupos. O primeiro composto por RMF, RMBH, RMSP, RMC e RMPA; o segundo, pelas demais regiões. O fato de a RMF participar do primeiro grupo pode ser explicado pela importância da ocupação industrial local, semelhante à encontrada nas outras quatro regiões metropolitanas do primeiro grupo. Embora com um mercado de trabalho mais desenvolvido, a RMRJ, junta-se à RMB, à RMR e à RMS no segundo grupo, por sua elevada ocupação no Setor Terciário. O DF também faz parte desse grupo, embora um pouco diferenciado, pela importância da administração pública.

Em termos gerais, o artigo confirma os desníveis regionais existentes no País e a limitação de iniciativas locais de desenvolvimento baseadas em políticas de atração de novos investimentos, através de incentivos de diversos tipos, desenvolvidas pelos governos dos diferentes estados nos últimos anos. O País

continua dividido entre as Regiões Sul e Sudeste, relativamente desenvolvidas, e as Regiões Norte e Nordeste, onde predomina o atraso econômico e social. De qualquer forma, foi possível diferenciar situações no interior dos dois grandes grupos, além de verificar os casos das Regiões Metropolitanas de Fortaleza e do Rio de Janeiro, que se deslocam de seu grupo original, quando consideradas algumas variáveis de seus mercados de trabalho.

Referências

- ANDERBERG M. R. **Cluster analysis for applications**. New York: Academic, 1973.
- CORRÊA R. M.; KUBRUSLY L. S. Instrução, trabalho, renda e moradia: diferenças e semelhanças entre as unidades da federação brasileira. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 34, Rio de Janeiro, 2002. [Anais... S. n. t.]
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1992.
- JOLLIFFE, I. T. **Principal components analysis**. New York: Springer-Verlag, 1986.
- KUBRUSLY L. S.; BARROS A. C. Condições de trabalho, instrução e renda nas metrópoles brasileiras: uma análise estatística multivariada. **Econômica**, Niterói, RJ, v. 5, n. 2, 2003.
- LUCAS L. C. S. Análise de grupamentos. **Revista Brasileira de Estatística**, Rio de Janeiro, ano 43, n. 172, 1982.
- PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIO 2001: Regiões Metropolitanas — PNAD. Rio de Janeiro: IBGE, 2002.
- RAMOS, L.; FERREIRA, V. Geração de emprego e realocação espacial no mercado de trabalho brasileiro — 1002/2002, **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 35, n. 1, 2005.
- RAMOS, L.; BRITTO, M. **O funcionamento do mercado de trabalho metropolitano brasileiro no período 1991-2002: tendências, fatos estilizados e mudanças estruturais**. Rio de Janeiro: IPEA, 2004. (Texto para Discussão, n. 1011)

SABOIA, J. Descentralização industrial no Brasil na década de noventa: um processo dinâmico e diferenciado regionalmente. **Nova Economia**, Belo Horizonte, v. 11, n. 2, 2001.

SABOIA, J. Um novo índice para o mercado de trabalho urbano no Brasil. **Revista de Economia Contemporânea**, Rio de Janeiro, v. 4, n. 1, 2000.

SPSS, Base 11.0: User's Guide. Chicago, SPSS Inc, 2002.

TONI, M. de **Para onde vai o mercado de trabalho?** a tendência à precarização das relações de trabalho: um estudo da região metropolitana de Porto Alegre. Tese (Doutorado) — Instituto de Filosofia e Ciências Humanas. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.