

Textos para Discussão N° 32

Secretaria do Planejamento e Gestão
Fundação de Economia e Estatística Siegfried Emanuel Heuser

Análise Espacial da Centralidade e da Dispersão da População do Estado do Rio Grande do Sul de 1970 a 2000: metodologias e notas preliminares

Iván G. Peyré Tartaruga

Porto Alegre, junho de 2008



GOVERNO DO ESTADO
RIO GRANDE DO SUL

SECRETARIA DO PLANEJAMENTO E GESTÃO

Secretária Substituta: Ana Maria Viana Severo



DIRETORIA

Presidente: Adelar Fochezatto

Diretor Técnico: Octavio Augusto Camargo Conceição

Diretor Administrativo: Nóra Angela Gundlach Kraemer

CENTROS

Estudos Econômicos e Sociais: Roberto da Silva Wiltgen

Pesquisa de Emprego e Desemprego: Míriam De Toni

Informações Estatísticas: Adalberto Alves Maia Neto

Informática: Luciano Zanuz

Editoração: Valesca Casa Nova Nonnig

Recursos: Alfredo Crestani

TEXTOS PARA DISCUSSÃO

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pela FEE, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões. As opiniões emitidas nesta publicação são de exclusiva e de inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista da Fundação de Economia e Estatística.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

Análise Espacial da Centralidade e da Dispersão da População do Estado do Rio Grande do Sul de 1970 a 2000: metodologias e notas preliminares

Iván G. Peyré Tartaruga*

Geógrafo e Técnico da FEE

Resumo

Os desenvolvimentos nos métodos espaciais – estes preocupados com a localização, a interação, a estrutura e o processo espaciais – estão criando novas possibilidades para os estudos das propriedades dos sistemas socioeconômicos. Neste artigo utilizaram-se técnicas de estatística espacial, juntamente com um sistema de informações geográficas (SIG), para descrever a distribuição espacial de dados sociais, tais como o centro médio, a distância padrão e a elipse de desvio padrão. Analisando a realidade do Estado do Rio Grande do Sul (Brasil), comparou-se a dispersão da população no período de 1970 até 2000 (anos censitários). O artigo conclui com algumas breves considerações a respeito dos padrões e das tendências destas distribuições.

Palavras-chave: Análise espacial; estatística espacial; análise regional.

Abstract

Developments in spatial methods – concerning location, spatial interaction, spatial structure and spatial processes – are creating new possibilities for studies of the nature of the socioeconomic systems. In this paper, I use spatial statistical techniques with a geographic information system (GIS) for describing the spatial distribution of social and economic data, such as the mean center, standard distance and standard deviational ellipse. Analyzing the reality of the State of Rio Grande do Sul (Brazil), I compare the spatial distribution of population during the period 1970-2000 (census years). The paper concludes with comments regarding the patterns and trends of these distributions.

Keywords: Spatial analysis; spatial statistics; regional analysis.

Classificação JEL: C10, R12, R11.

1 – Introdução

Este trabalho possui dois objetivos principais e intimamente relacionados. Por um lado, uma tentativa de uma melhor compreensão da realidade demográfica do Estado do Rio Grande do Sul, das últimas décadas, nas suas diferentes regiões, o que seria, basicamente, um problema de análise regional. Por outro, o uso de técnicas de análise, para o problema anterior, que têm como característica essencial a consideração da dimensão espacial (ou localização dos fenômenos estudados) na sua metodologia, o que se pode chamar de análise espacial.

* E-mail: ivan@fee.tche.br. O autor agradece a leitura atenta dos economistas Carlos Águedo Nagel Paiva e José Antônio Fialho Alonso da versão preliminar deste artigo, destacando que a versão final é de inteira responsabilidade do autor.

Nesse contexto, ressaltam-se as três razões, apontadas por Anselin, Florax e Rey (2004), da crescente importância do uso dos métodos espaciais: o renovado interesse nos papéis do espaço e das interações espaciais, na teoria das Ciências Sociais; a crescente disponibilidade de dados socioeconômicos georreferenciados; e o baixo custo em termos de sistemas computacionais e operacionais para a manipulação de dados espaciais.

Com respeito à primeira razão, a Ciência Geográfica vem defendendo a reafirmação da perspectiva espacial na teoria e na análise sociais contemporâneas, em contrapartida a uma perspectiva baseada quase unicamente na historicidade, ou seja, defende um equilíbrio entre historicidade e espacialidade (Soja, 1993). O pressuposto aqui sustentado, em outras palavras, é o de que as relações sociais produzem o espaço, ao mesmo tempo em que esse mesmo espaço condiciona essas relações, uma dialética socioespacial nos termos de Edward Soja (1993). Além de na Geografia, conforme Anselin, Florax e Rey (2004), o uso explícito da perspectiva espacial encontra-se na *new economic geography* na Economia moderna, sobretudo na identificação das externalidades espaciais; na Sociologia, nos estudos acerca da criminologia e das questões urbanas; na Ciência Política, na análise de conflitos, nos quadros eleitorais e nas relações internacionais.

Quanto à segunda razão do interesse nos métodos espaciais, percebe-se uma crescente disponibilidade de dados socioeconômicos e de bases cartográficas digitais; vejam-se, por exemplo, os dados oferecidos somente sobre o Brasil, gratuitamente, no Instituto Brasileiro de Geografia e Estatística (IBGE), no Instituto Nacional de Pesquisas Espaciais (INPE) e em outros órgãos governamentais. Além disso, relacionada à terceira razão, há uma grande quantidade de programas computacionais de processamento de dados espaciais (georreferenciados), também conhecidos como sistemas de informações geográficas (SIG): à venda no mercado, como ArcGIS, IDRISI, MapInfo, Maptitude; e gratuitos, como o TerraView e o Spring (produzidos no INPE), GeoDa (desenvolvido por um dos maiores estudiosos da análise de dados geográficos, Luc Anselin), entre outros.

É evidente, nesse contexto, o importante papel dos SIG, que se definem como qualquer sistema computacional de informação que pode, segundo Antenucci, Brown e Croswell et al. (1991): (a) coletar, armazenar e recuperar informações com suas localizações espaciais (georreferenciadas); (b) identificar locais específicos através de critérios conhecidos; (c) explorar relações entre grupos de dados dentro de uma determinada área; (d) analisar, especialmente, dados relacionados para servir de apoio a tomada de decisões numa determinada área; (e) facilitar a seleção e o manuseio de dados para o uso em modelos aplicados para analisar impactos de alternativas em uma área; (f) e apresentar uma área, geográfica e numericamente, tanto antes quanto depois da análise. Em suma, as várias definições do que seria um SIG podem ser sintetizadas em três noções complementares, conforme Maguire, Goodchild e Rhind (1991): mapeamento, banco de dados e análise espacial.

O SIG, portanto, é uma ferramenta fundamental para a análise espacial de dados geográficos, a qual se ocupa com a localização, a interação, a estrutura e o processo espaciais de fenômenos sociais e econômicos. Mais especificamente, tal análise visa estudar o comportamento espacial de fenômenos geográficos (regiões, municípios, cidades, domicílios, etc.) por meio de elementos geométricos como pontos, linhas ou áreas (polígonos), que são representações da realidade.

Outro elemento crucial para a análise espacial, tanto no sentido absoluto (das coordenadas ou da localização) quanto no sentido relativo (distâncias, arranjo espacial), é o adequado tratamento estatístico dos dados. Acredita-se há muito tempo que, para a resolução dos problemas concernentes à geografia, especificamente, e à análise regional, em termos gerais, somente o mapa não é suficiente, nem tão-somente os dados estatísticos; parece evidente que devem trabalhar juntos (Sviatlovsky e Eells, 1937).

Particularmente, a análise espacial focaliza aspectos específicos concernentes aos dados e aos modelos utilizados. Anselin (1988 e 1992) nomeia esses aspectos de *efeitos espaciais* sobre os dados e os divide em dois tipos gerais: a dependência espacial e a heterogeneidade espacial. O primeiro efeito, também conhecido como associação espacial ou autocorrelação espacial, está diretamente relacionado com a primeira lei da geografia de Tobler, qual seja, “todas as coisas são parecidas, porém, coisas mais próximas se parecem mais que coisas mais distantes” (Tobler apud Druck, Carvalho, Câmara et al., 2004, p. 11). Nesse sentido, pode-se considerar que, em um conjunto de entes geográficos (pontos, linhas ou polígonos), valores similares de uma variável tendem a estarem próximos, o que poderia ser um aglomerado (*cluster*). A título de exemplo, veja-se o caso de um bairro metropolitano com altas taxas de criminalidade, provavelmente os bairros vizinhos também possuam um grande número de crimes. A análise de dependência espacial, portanto, busca verificar essa associação que depende, especialmente, da distância.¹ Ressalta-se que a noção de espaço (relativo) pode ir além da idéia de distância (euclidiana), ou seja, considerar outras noções de espaço – espaço social (ou lugar), espaço político (ou território), espaço econômico (custo de deslocamento)² – o que evidencia um grande leque de possibilidades de estudos para as ciências sociais em geral.

O segundo tipo de efeito espacial, a heterogeneidade espacial, trata da diferenciação espacial, ou regional, ou, em outros termos, da unicidade de cada lugar, ou seja, cada localização é possuidora de características únicas. Assim, a heterogeneidade está associada a uma ausência de estabilidade do comportamento das variáveis sobre o espaço em questão: os parâmetros que caracterizam as unidades variam com a localização.

Mais preocupado com o segundo efeito (heterogeneidade espacial) do que com o primeiro (dependência espacial), este trabalho tem por objetivo proporcionar, por meio de algumas técnicas de análise espacial realizadas em um SIG, uma melhor compreensão da distribuição espaço-temporal das diferentes regiões gaúchas levando em conta aspectos populacionais (análise regional). Mais especificamente, o dado utilizado é a população absoluta dos municípios gaúchos, no período que compreende os anos de 1970 e de 2000, considerando somente os anos censitários. Ressalta-se, também, o aporte da estatística espacial para a compreensão da distribuição dos dados no espaço, a exemplo das medidas espaciais de tendência central (ou centralidade) e de dispersão (ou variabilidade).

¹ Como exemplo de método de análise de autocorrelação espacial tem-se o índice I de Moran (global e local).

² Aqui, está-se falando de noções/conceitos teóricos específicos utilizados na ciência geográfica. Para uma boa explicitação destes, ver Oakes (1997), acerca do conceito de lugar, e Souza (1995), acerca do de território.

O texto está dividido em quatro seções principais, a primeira que introduz a temática e os objetivos do trabalho, e finda aqui. A segunda seção que descreve a metodologia utilizada nas análises, sobretudo acerca das medidas de tendência central e de dispersão das distribuições espaciais, nas quais estão distribuídas em três subseções (centro médio, distância padrão e elipse de desvio padrão). A terceira analisa os resultados obtidos das medições a respeito da população. E a última que tece as considerações finais do trabalho.

2 – Metodologia espacial

Para a análise espacial da realidade gaúcha foram utilizadas as informações estatísticas por Município da população, a base digital cartográfica e o Sistema de Conversão Municipal obtidos do CD-ROM “RS em Mapas e Dados” (Paiva, 2007).³ Os dados dizem respeito aos anos censitários de 1970, 1980, 1991 e 2000. Alguns comentários e explicitações referentes a esses três aspectos preliminares da metodologia – dados utilizados, base cartográfica e sistema de conversão – se fazem necessários, a seguir.

Os dados utilizados neste trabalho são os de população absoluta dos Municípios do Estado do Rio Grande do Sul.⁴ E servem como indicativo de bem-estar social e econômico. Efetivamente, os locais de atração (acúmulo) de população, em tese, ocorrem onde há, em alguma medida, crescimento econômico e, portanto, as pessoas se deslocam para esses lugares com a esperança de oportunidades melhores de vida.

O outro aspecto preliminar desta pesquisa, a base digital cartográfica do Estado do Rio Grande Sul, providencia o suporte básico para a realização das análises de caráter espacial; utilizaram-se, especificamente, as malhas municipais (áreas e limites dos municípios). Neste trabalho foram utilizadas duas das bases (malhas municipais) disponíveis em Paiva (2007)⁵ : a que vigorava no período de 1966 até 1982, com 232 municípios, e, no período a partir de 2003 (atual), com 496 municípios. Na primeira base são realizados todos os procedimentos de estatística espacial, enquanto nesta é feita tão-somente a visualização de alguns resultados estatísticos sobre a primeira. As bases foram obtidas no formato (vetorial) de arquivos do *software* TerraView⁶ (um SIG), as quais foram convertidas para o formato *shapefile* (.shp) e importadas para o *software* ArcGIS (SIG), no qual são realizadas as técnicas de análise espacial, e neste transformadas para o formato *Geodatabase Feature Class* (.gdb). Algumas informações cartográficas básicas das bases, que caracterizam todos os mapas constantes neste trabalho (Figura 1): projeção cartográfica Universal Transversa de

³ Esse trabalho é um produto da pesquisa “Investigação Acerca dos Determinantes dos Diferenciais de Desempenho Socioeconômico dos Municípios Gaúchos entre 1970 e 2000”, realizada pela Fundação de Economia e Estatística (FEE) e pela Universidade de Santa Cruz do Sul (UNISC), com financiamento do CNPq e da FAPERGS.

⁴ Os dados populacionais são oriundos dos censos do IBGE (1970, 1980, 1991 e 2000).

⁵ No trabalho Paiva (2007), estão disponibilizadas todas as malhas municipais do RS desde 1966: de 1966 a 1982, com 232 municípios; de 1983 a 1989, 244 municípios; de 1990 a 1992, 333 municípios; de 1993 a 1996, 427 municípios; de 1997 a 2000, 467 municípios; de 2001 a 2002, 497 municípios; e de 2003 a 2006, 496 municípios.

⁶ TerraView (versão 3.1.4) do Instituto Nacional de Pesquisas Espaciais (INPE).

Mercator (UTM);⁷ Datum horizontal Córrego Alegre; zona 22 (Sul); meridiano central - 51 (51° oeste); metro (unidade de medida linear).

Figura 1 – Mapa do Estado do Rio Grande do Sul – 1970



O último aspecto preliminar para a realização deste trabalho diz respeito ao Sistema de Conversão Municipal (Paiva, 2007). Esse sistema permite a comparação da estrutura municipal gaúcha em diferentes anos do período que compreende os anos de 1966 e 2006, lembrando que durante esse período houve diversas mudanças nessa estrutura.⁸ O sistema converte os dados municipais de qualquer ano do período para a estrutura municipal de 1970 (232 municípios), portanto, possibilitando a comparação do desempenho de qualquer município ao longo do tempo. Para efetivar

⁷ As coordenadas dos pontos na projeção UTM são representadas por medidas em metros. Por exemplo, as coordenadas do ponto mais meridional do estado gaúcho e do Brasil, situado no atual Município de Santa Vitória do Palmar, são, aproximadamente, 6.262.963 m N (latitude) e 277.958 m E (longitude).

⁸ Somente para ressaltar essa condição, em 1966 o Estado era composto de 232 municípios e essa quantidade alcançou, em 2006, 496 municípios.

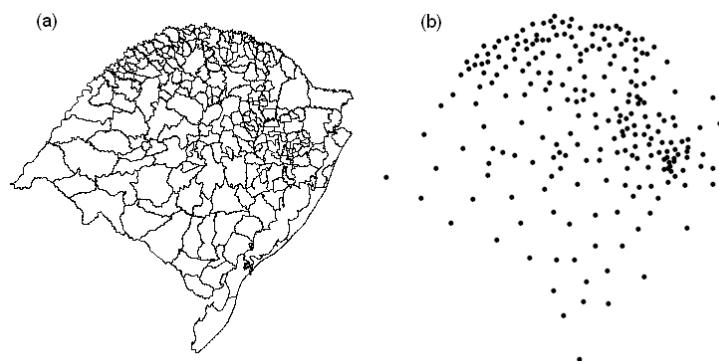
a conversão deve-se escolher um dos três critérios de imputação: área cedida, origem da sede ou população originalmente cedida por cada município para a constituição das novas estruturas. Para este trabalho, todos os dados utilizados foram convertidos, evidentemente, com respeito ao critério da população cedida.

No âmbito dos procedimentos estatísticos, faz-se uso de técnicas da estatística descritiva espacial para resumir e apresentar as informações de forma a facilitar a interpretação das mesmas. Com esse intuito, de um lado, utiliza-se uma medida de centralidade (ou tendência central) espacial: o centro médio, também conhecido como centro espacial, ponderado pelos dados populacionais. De outro, duas medidas de dispersão (ou variabilidade) espacial: a distância padrão e a elipse de desvio padrão.

Essas técnicas estão dentro de uma subárea da estatística, a Estatística Centrográfica, que surge como método de análise regional no século XIX nos Estados Unidos e, originalmente, preocupada com a distribuição de populações humanas (Sviatlovsky e Eells, 1937). E que tem por objetivo principal a união das técnicas de representação e visualização espacial com os procedimentos estatísticos clássicos (médias aritmética e ponderada, desvio padrão,...), de forma a melhorar as técnicas estatísticas para a análise regional. Em outras palavras, essas técnicas são equivalentes em duas dimensões (no espaço) da estatística descritiva clássica para o tratamento de dados isolados (análise univariada). A estatística espacial proporciona a inclusão na análise da dimensão espacial – distância, vizinhança, proximidade –, que é substancialmente diferente das análises univariadas ou multivariadas da estatística clássica.

Estes procedimentos, portanto, visam providenciar medidas básicas da distribuição espacial de pontos que representam fenômenos sociais ou econômicos da realidade. Nesse sentido, como a unidade espacial neste trabalho é o município, portanto, representado espacialmente por um polígono e não um ponto, deve-se proceder na transformação dos municípios/polígonos em pontos, mais especificamente nos respectivos centróides que são os centros geométricos de polígonos (Figura 2).

Figura 2 – Representações dos municípios gaúchos: (a) malha municipal (polígonos) e (b) respectivos centróides (pontos) dos municípios

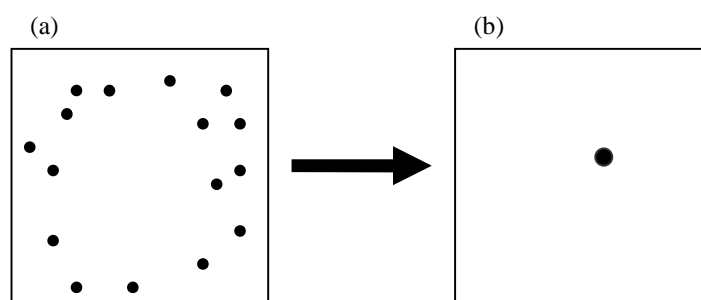


Os métodos aqui apresentados, em termos gerais, podem auxiliar na determinação de padrões, de tendências ou de relações da distribuição. Em termos específicos, visam responder as seguintes perguntas com respeito à distribuição espacial de certos fenômenos (dados quantitativos): onde está o centro? Como os dados estão dispersos ao redor do centro? Além disso, o uso destes métodos é indicado para a realização de comparações entre distribuições de diferentes elementos, por exemplo, de homens e de mulheres, ou entre distribuições ao longo do tempo.

2.1 – Centro médio

O centro médio (ou centro espacial) de um conjunto de unidades espaciais, no caso desta pesquisa os municípios de 1970, é um ponto imaginário, georreferenciado, que representa o centro da distribuição (ou centro de gravidade) (Figura 3). Ele seria o equivalente à média aritmética da estatística clássica. E serve, principalmente, para descobrir mudanças na distribuição ao longo do tempo.⁹

Figura 3 – Desenho esquemático (a) de distribuição de pontos e (b) do respectivo centro médio



Os centros médios são obtidos, conforme Ebdon (1985), por meio do cálculo das médias das coordenadas X (longitudes) e Y (latitudes) dos centróides dos municípios, como descrito a seguir:

$$(\bar{x}, \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right)$$

Em que:

- \bar{x} : coordenada X (longitude) do centro médio;
- \bar{y} : coordenada Y (latitude) do centro médio;
- x_i : coordenada X (longitude) do centróide do município “i”;
- y_i : coordenada Y (latitude) do centróide do município “i”;
- n : número de centróides (municípios).

⁹ Esse procedimento estatístico é utilizado, por exemplo, pelo *Census Bureau* dos Estados Unidos, instituição responsável pelos censos demográficos naquele país, que calcula o centro médio de população norte-americana, a cada censo, desde 1790 (GEOGRAPHY DIVISION, 1991).

As coordenadas do centro médio minimizam a soma das distâncias quadráticas entre ele próprio e cada um dos centróides. Matematicamente, o centro médio minimiza o seguinte somatório:

$$\sum_{i=1}^n [(x_i - \bar{x})^2 + (y_i - \bar{y})^2]$$

Entretanto, o centro médio indica, simplesmente, o centro da distribuição das localizações dos centróides, portanto, considerando tão-somente a dimensão espacial, não indicando nada sobre a distribuição de algum dado ou informação estatístico. Porém, da mesma forma que há a média ponderada na estatística clássica, pode-se calcular o centro médio de pontos ponderado e, portanto, representar também o comportamento espacial de alguma informação quantitativa (população, PIB, renda,...), ou seja, a cada coordenada de cada centróide são associados certos fatores de ponderação ou pesos (variáveis), o que satisfaz os objetivos deste trabalho. Em termos matemáticos:

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}, \frac{\sum_{i=1}^n p_i y_i}{\sum_{i=1}^n p_i} \right)$$

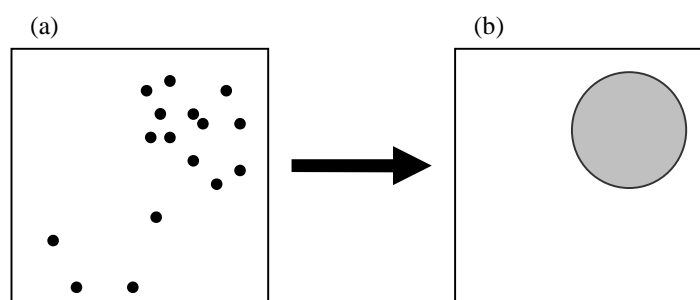
Adicionalmente a equação apresentada anteriormente:

p_i : valor de ponderação (peso) correspondente ao centróide (município) “ i ”.

2.2 – Distância padrão

De maneira complementar ao centro médio, a distância padrão é uma medida do grau de concentração ou de dispersão da distribuição espacial de pontos em torno do centro médio. Esse valor é uma distância que equivale ao raio de um círculo com o ponto central localizado no centro médio da distribuição. O desvio padrão de uma variável, da estatística clássica, seria o equivalente à distância padrão (em duas dimensões). Por conseguinte, quanto maior a distância padrão (e o círculo respectivo) de uma distribuição de pontos, maior será a dispersão destes pontos em torno do respectivo centro espacial; e, contrariamente, quanto menor a distância padrão, maior será a concentração dos pontos em torno do centro (Figura 4).

Figura 4 – Desenho esquemático (a) de distribuição de pontos e (b) do respectivo círculo com raio de uma distância padrão



Matematicamente, a distância padrão representa o desvio padrão das distâncias de cada ponto com o centro médio da distribuição. E a equação que a define é a seguinte:

$$S_{xy} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}{n}}$$

Em que:

- S_{xy} : distância padrão;
- \bar{x} : coordenada X (longitude) do centro médio;
- \bar{y} : coordenada Y (latitude) do centro médio;
- x_i : coordenada X (longitude) do centróide do município “i”;
- y_i : coordenada Y (latitude) do centróide do município “i”;
- n : número de centróides (municípios).

Da mesma maneira que o centro médio, a distância padrão somente tem utilidade para esta pesquisa na sua forma ponderada pelos dados em questão (população). E a equação é a seguinte:

$$S_{xy} = \sqrt{\frac{\sum_{i=1}^n p_i (x_i - \bar{x})^2 + \sum_{i=1}^n p_i (y_i - \bar{y})^2}{\sum_{i=1}^n p_i}}$$

Adicionalmente a equação anterior:

p_i : valor de ponderação (peso) correspondente ao centróide do município “i”.

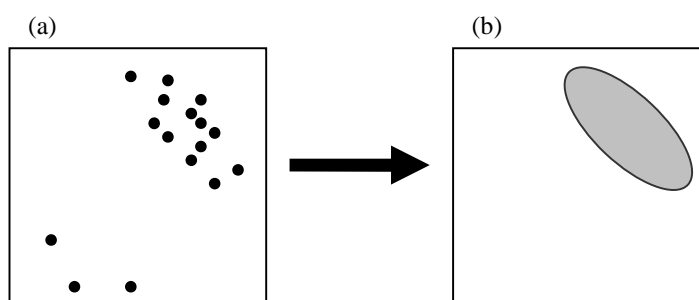
Entretanto, a distância padrão não leva a qualquer conhecimento a respeito da direção da distribuição, a exemplo de uma direção ou eixo preferencial de distribuição da população; para esse fim utiliza-se o método das elipses de desvio padrão, tratado na seção seguinte.

2.3 – Elipse de desvio padrão¹⁰

Como destacado anteriormente, a distância padrão é considerada uma medida interessante e simples da dispersão das informações pontuais ao redor do centro médio, porém, ela não permite fazer qualquer consideração sobre a direção da distribuição, isto é, não mostra a “forma” da distribuição. Assim, a técnica da elipse de desvio padrão proporciona o conhecimento da dispersão espacial em dois sentidos: densidade (ou compacidade) e orientação; portanto, proporciona o conhecimento da natureza da distribuição dos dados na sua assimetria (em diferentes direções). Essa técnica, dessa forma, auxilia a identificação da tendência de uma distribuição de pontos, muito útil para a comparação de distribuições e em diferentes períodos de tempo.

Cada elipse é determinada por meio de três parâmetros: ângulo de rotação, dispersão ao longo do maior eixo e dispersão ao longo do menor eixo. O eixo maior define a direção de máxima dispersão da distribuição, enquanto, o menor eixo é perpendicular ao eixo anterior e define a mínima dispersão (Figura 5).

Figura 5 – Desenho esquemático (a) de distribuição de pontos e (b) da respectiva elipse de desvio padrão



Além disso, pode-se definir o número de desvios padrão para representar a abrangência da elipse de desvio padrão, que pode ser de um, dois ou três desvios. Assim, para uma distribuição de tipo normal dos pontos dos dados ao redor do centro médio tem-se que: uma elipse de um desvio

¹⁰ O autor agradece as indicações e as explicações gentilmente cedidas (via e-mail) por Lauren M. Scott, especialista da ESRI, e Ned Levine, responsável pela criação do *software* CrimeStat (para análise espacial de crimes).

padrão cobrirá, aproximadamente, 68% dos centróides do espaço analisado; de dois desvios padrão cobrirá, aproximadamente, 95% dos centróides; e de três desvios padrão cobrirá, aproximadamente, 99% dos centróides. Vale ressaltar que para o método ponderado de elipses de desvio padrão, esses percentuais corresponderão, aproximadamente, à parcela do total da variável de ponderação.

Em termos de formulação matemática, a elipse de desvio padrão origina-se da distribuição bivariada que se caracteriza por dois tipos de desvio padrão, nas direções X (longitudes) e Y (latitudes), ortogonais e que definem uma elipse (Ebdon, 1985):

$$\text{Distribuição bivariada} = \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}}$$

Em que: σ_x : desvio padrão na direção X (longitudes);
 σ_y : desvio padrão na direção Y (latitudes).

Contudo, a elipse de desvio padrão é calculada em duas etapas. Na primeira etapa, a orientação dos eixos que define a elipse é calculada de modo que a soma dos quadrados das distâncias entre os pontos da distribuição e os eixos da elipse seja mínimo. Portanto, obtém-se o ângulo (θ) desta orientação por meio da seguinte equação:

$$\theta = \text{Arc tan} \frac{\left\{ \left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right) + \left[\left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right)^2 + 4 \left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)^2 \right]^{1/2} \right\}}{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}$$

Em que: θ : ângulo de rotação da elipse de desvio padrão;
 \bar{x} : coordenada X (longitude) do centro médio;
 \bar{y} : coordenada Y (latitude) do centro médio;
 x_i : coordenada X (longitude) do centróide do município “i”;
 y_i : coordenada Y (latitude) do centróide do município “i”;
 n : número de centróides (municípios).

Assim, o eixo Y é girado, no sentido horário, pelo ângulo θ (calculado anteriormente). Na segunda etapa para determinar a elipse, são calculados os dois desvios padrões (σ_x , σ_y),¹¹ utilizando os eixos X e Y girados, por meio das duas equações que seguem:

¹¹ A formulação original da elipse de desvio padrão foi obtida de Ebdon (1985), entretanto, as elipses geradas a partir dela resultavam em elipses de tamanho muito pequeno; esse erro foi resolvido para o *software* ArcGIS,

$$\sigma_x = \sqrt{\left\{ 2 \sum_{i=1}^n [(x_i - \bar{x}) \cos \theta - (y_i - \bar{y}) \sin \theta]^2 / n - 2 \right\}}$$

$$\sigma_y = \sqrt{\left\{ 2 \sum_{i=1}^n [(x_i - \bar{x}) \sin \theta - (y_i - \bar{y}) \cos \theta]^2 / n - 2 \right\}}$$

Em acréscimo à equação anterior:

σ_x : desvio padrão na direção X (longitudes);

σ_y : desvio padrão na direção Y (latitudes).

Desse modo, a elipse de desvio padrão define-se pelos seguintes parâmetros, além, é claro, do respectivo centro médio que é o centro da elipse:

- ângulo de rotação = θ
- comprimento do eixo X = $2\sigma_x$
- comprimento do eixo Y = $2\sigma_y$

Além do mais, pode-se proceder ao cálculo ponderado da elipse de desvio padrão da distribuição pontual no espaço. Agora, matematicamente, tem-se

$$\theta = \text{Arc tan} \frac{\left\{ \left(\sum_{i=1}^n p_i (x_i - \bar{x})^2 - \sum_{i=1}^n p_i (y_i - \bar{y})^2 \right) + \left[\left(\sum_{i=1}^n p_i (x_i - \bar{x})^2 - \sum_{i=1}^n p_i (y_i - \bar{y})^2 \right)^2 + 4 \left(\sum_{i=1}^n p_i^2 (x_i - \bar{x})(y_i - \bar{y}) \right)^2 \right]^{1/2} \right\}}{2 \sum_{i=1}^n p_i^2 (x_i - \bar{x})(y_i - \bar{y})}$$

conforme a especialista da empresa que produz esse programa computacional, Lauren M. Scott. Para corrigir esse problema (de subestimação) foi acrescentado o cálculo da raiz quadrada dos desvios dos eixos da elipse, conforme o pesquisador Ned Levine e descrito na documentação do *software* CrimeStat (Levine, 2007).

Adicionando às equações anteriores:

p_i : valor de ponderação (peso) correspondente ao centróide do município “ i ”.

E os dois desvios padrões (σ_x , σ_y), ponderados, são calculados por meio das seguintes equações:

$$\sigma_x = \sqrt{\left\{ \frac{2 \sum_{i=1}^n [p_i (x_i - \bar{x}) \cos \theta - p_i (y_i - \bar{y}) \sin \theta]^2}{p_i} \right\}}$$

$$\sigma_y = \sqrt{\left\{ \frac{2 \sum_{i=1}^n [p_i (x_i - \bar{x}) \sin \theta - p_i (y_i - \bar{y}) \cos \theta]^2}{p_i} \right\}}$$

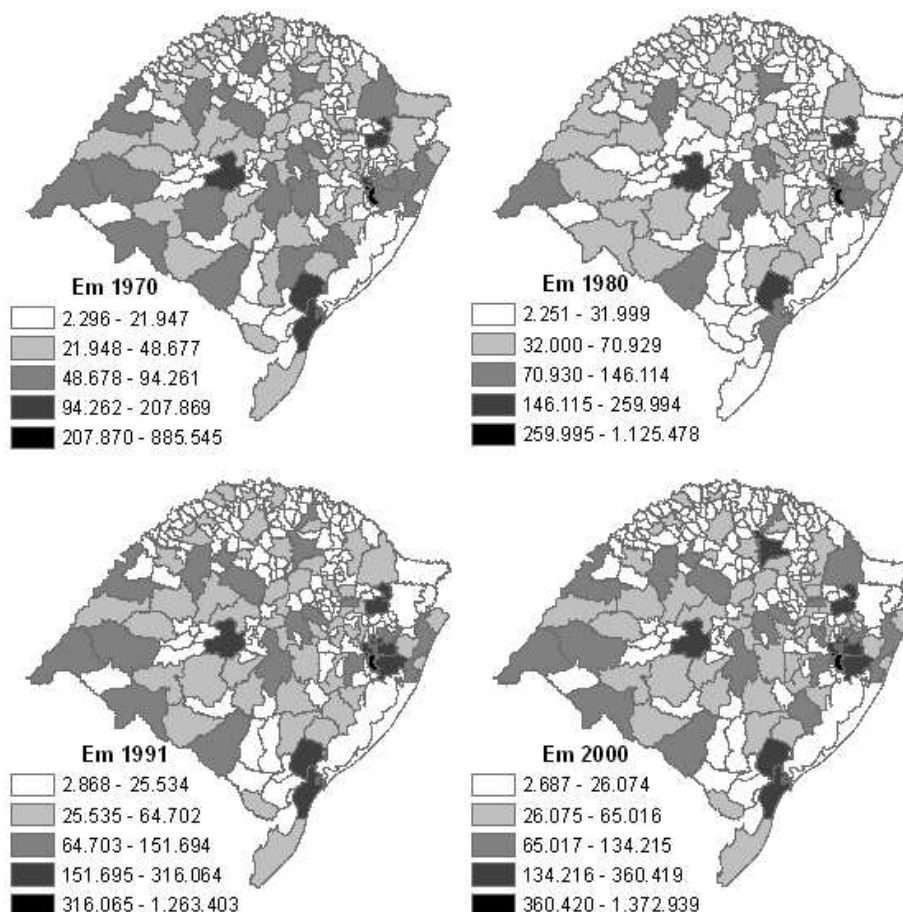
E, novamente, tem-se que, com relação ao centro médio respectivo:

- ângulo de rotação = θ
- comprimento do eixo X = $2\sigma_x$
- comprimento do eixo Y = $2\sigma_y$

3 – Distribuição geográfica da população

A distribuição espacial da população gaúcha, no período de 1970 a 2000, é mostrada na Figura 6. Percebe-se que a capital gaúcha (à direita nos mapas) é em todos os anos o de maior destaque, sobretudo, a Região Metropolitana de Porto Alegre nos dois períodos finais. Além do mais, ressaltam-se os municípios de Caxias do Sul (à direita no mapa, ao norte de Porto Alegre), Santa Maria (praticamente no centro do mapa), Rio Grande/Pelotas (ao sul do Estado) e, Passo Fundo (ao norte do Estado) no último ano analisado.

Figura 6 – Distribuição espacial da população (hab.) do Rio Grande do Sul pelo método de “quebras naturais” (Jenks)¹² – 1970-2000



Apesar da interessante visualização dos dados, a figura anterior não possibilita uma interpretação satisfatória e fácil dos dados em função da complexidade das informações. Assim, passa-se a utilizar métodos que generalizam espacialmente as informações e, portanto, facilitam sua interpretação em termos gerais.

Primeiro, foram determinados os centros médios (ponderados) da população do Estado, ou seja, os pontos imaginários que representam os centros de gravidade da população em cada período analisado (Tabela 1). Em resumo, no ano de 1970 o centro médio localiza-se no interior do atual

¹² O método de classificação de dados de “quebras naturais” (ou método de Jenks) utiliza um algoritmo iterativo que procura reduzir a variância dentro dos grupos e maximizar a variância entre os diferentes grupos. Em outras palavras, as quebras ocorrem nos pontos mais baixos dos “vales” do histograma dos dados. A maior vantagem do método é que ele considera explicitamente a distribuição dos dados. De outro lado, a desvantagem do método é a dificuldade de entendimento da lógica para a maioria dos usuários do mapa, além disso, os valores da legenda da classificação (intervalos das classes), geralmente, parecem não ser intuitivos. De qualquer maneira, esse método foi utilizado aqui em razão de a capital do Estado, Porto Alegre, possuir valores muito discrepantes (maiores) em comparação ao restante dos municípios (característica de um *outlier*), em quase todos os tipos de informações utilizados, e, diferentemente de outros métodos de classificação (quantis, intervalos iguais, desvio padrão), esse método discrimina a capital gaúcha (*outlier*) do restante das municipalidades. Ver ESRI (2006).

Município de Sinimbu, em 1980 no Município de Santa Cruz do Sul, e nos anos de 1991 e 2000, no Município de Venâncio Aires.

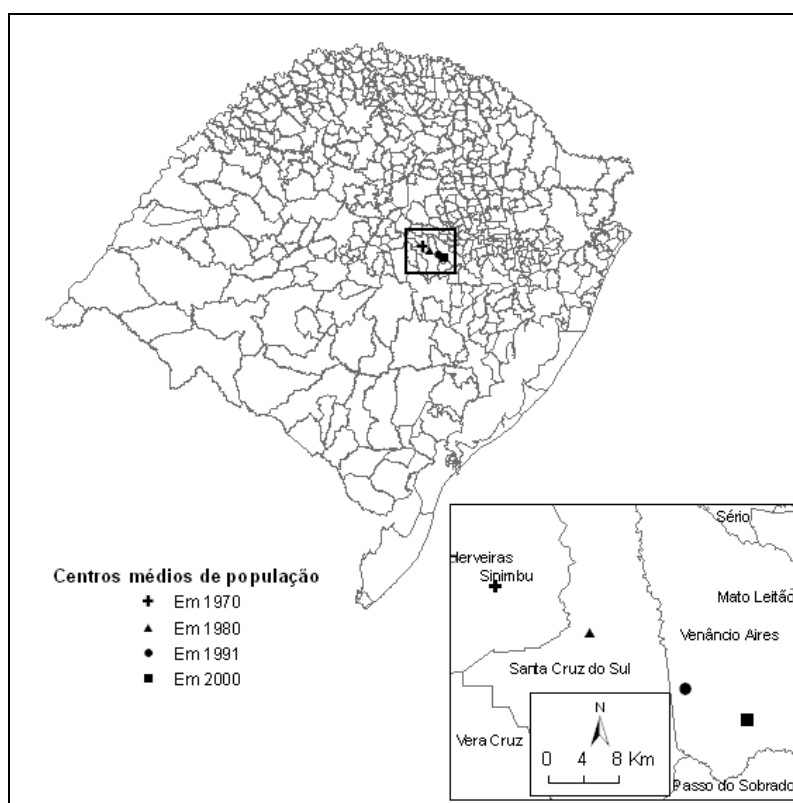
Tabela 1 – Centros médios de população do Rio Grande do Sul – 1970-2000

Ano	Longitude Oeste	Latitude Sul	Localização aproximada atual
1970	52°31' 30"	29°29' 39"	Município de Sinimbu
1980	52°24' 57"	29°32' 45"	Município de Santa Cruz do Sul
1991	52°18' 11"	29°36' 09"	Município de Venâncio Aires
2000	52°13' 49"	29°38' 13"	Município de Venâncio Aires

FONTE DOS DADOS BRUTOS: PAIVA (2007)

A partir destes centros no espaço do Estado (Figura 7), verifica-se o comportamento de deslocamento geral da distribuição populacional entre os anos censitários. Assim, entre os anos de 1970 e 1980, o centro médio populacional deslocou-se 11,8 km no sentido Sudeste; entre 1980 e 1991, deslocou-se 12,6 km no sentido Sudeste; e entre 1991 e 2000, 7,8 km no sentido Sudeste também. Portanto, nas décadas iniciais (1970 a 1991) o deslocamento de população, em termos gerais, foi mais intenso do que na década final (1991-2000). E, em todos os períodos, no sentido Sudeste do Estado; comportamento, certamente, influenciado pelo acréscimo de população na Região Metropolitana de Porto Alegre, sobretudo, no Município de Porto Alegre.

Figura 7 – Centros médios de população do Rio Grande do Sul – 1970-2000



Além disso, a distância padrão (ponderada) de população, medida de concentração/dispersão da distribuição espacial a partir do respectivo centro médio populacional, indica uma (fraca) tendência de concentração ao longo do período analisado. Com efeito, na Tabela 2 pode-se verificar que, de 1970 até 2000, houve uma redução total de 10 km da distância padrão populacional, sendo que a redução mais acentuada (de 3,8 km) ocorreu entre 1980 e 1991.

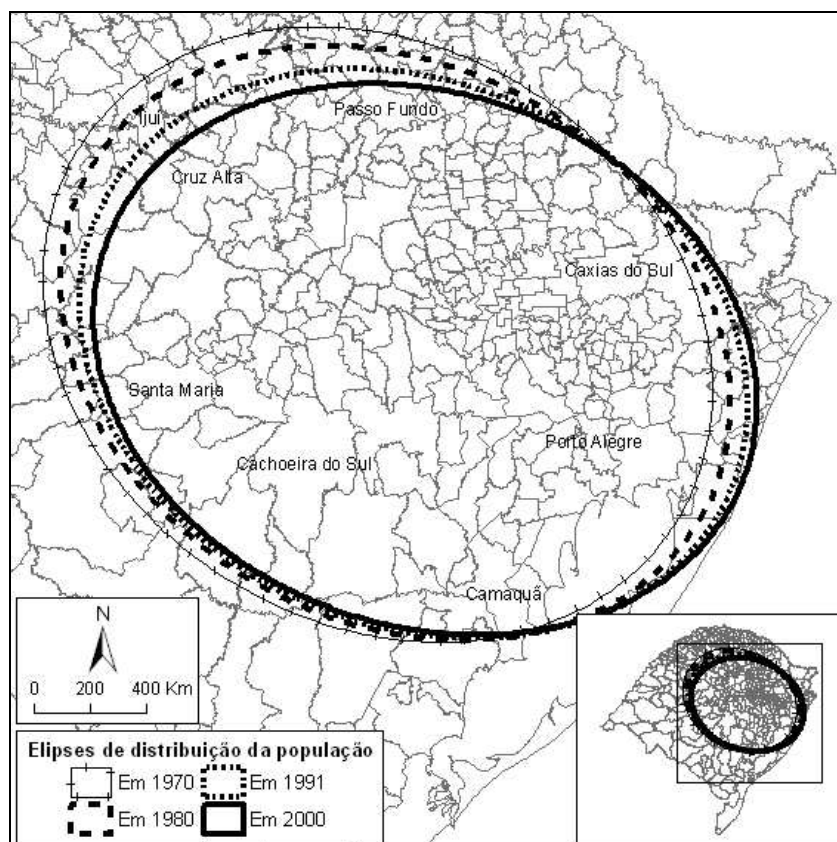
Tabela 2 – Distância padrão da população do Rio Grande do Sul – 1970-2000

Ano	Distância padrão (km)	Diferença da distância padrão entre os anos (km)
1970	193,5	...
1980	190,5	- 3,0
1991	186,7	- 3,8
2000	183,5	- 3,2

FONTE DOS DADOS BRUTOS: PAIVA (2007)

Por fim, procede-se à construção das elipses de desvio padrão que, além de indicarem a localização da concentração populacional, indicam a orientação geral da distribuição espacial da população nos diferentes anos. Ao analisar as elipses (Figura 8), verifica-se, primeiro, que a população está concentrada no espaço (das elipses) delimitado pelo Município de Porto Alegre e Região Metropolitana de Porto Alegre, no extremo Leste das elipses; pelo o Município de Santa Maria, no extremo Oeste das elipses; por Passo Fundo, no extremo Norte; e por Camaquã, no extremo Sul. Além desses, estão nesse espaço, e que merecem destaque, o Município de Caxias do Sul e a Região da Serra, o Município de Cruz Alta, o Município de Cachoeira do Sul, entre outros.

Figura 8 – Elipses de desvio padrão da população do Rio Grande do Sul – 1970-2000



Apesar de as formas das elipses serem quase circulares, o que indica que não há um eixo direcional muito proeminente da distribuição da população, podem-se perceber algumas tendências mesmo que tênues (Figura 8). Em todo o período, o eixo direcional da distribuição espacial da população está na direção Noroeste/Sudeste e, além disso, as elipses estão se deslocando no sentido Sudeste do Estado (do mesmo modo que os respectivos centros médios) (Figura 8 e Tabela 3). Ou seja, a população está preferencialmente distribuída neste eixo e desloca-se, em termos gerais, neste sentido. Percebe-se, também, uma leve tendência de concentração da população, como se pode verificar na redução, ao longo do tempo, das distâncias padrão dos eixos (X e Y) (Tabela 3), em outras palavras, a redução da área das elipses (distribuição).

Tabela 3 – Elipses de desvio padrão da população do Rio Grande do Sul – 1970-2000

Ano	Distância padrão em X (km)	Distância padrão em Y (km)	Rotação (graus decimais)	Parcela da população na elipse (%)*	Número de municípios na elipse*
1970	212,0	173,0	122,5	61,5	123
1980	211,5	166,8	119,7	63,3	119
1991	207,6	163,1	113,9	66,4	115
2000	204,4	159,9	110,2	68,4	113

FONTE DOS DADOS BRUTOS: PAIVA (2007)

NOTA * - Informações referentes aos centróides dos municípios que estão dentro da respectiva elipse de desvio padrão.

4 – Considerações finais

Como não se pretende neste trabalho encerrar a discussão, o que seria uma grande pretensão, sobre o comportamento da distribuição espacial, nas últimas décadas, das pessoas (demografia) no Rio Grande do Sul, nesta seção final deseja-se frisar alguns resultados das análises empreendidas aqui e que podem servir de base para outras análises que objetivem, sobretudo, a explicação destes resultados. Nesse sentido, este trabalho aponta, de modo geral, a propensão das distribuições espaciais da população gaúcha à regularidade ao longo do tempo.

Assim, em primeiro lugar, os resultados das análises indicam a regularidade com respeito ao comportamento no deslocamento dos centros médios da população, ou centros de gravidade: em todo o período estudado esses centros deslocam-se no sentido Sudeste do estado gaúcho, apontando exatamente para a capital gaúcha e, portanto, a Região Metropolitana de Porto Alegre. Neste caso, parece evidente a influência da região metropolitana.

A outra regularidade é a que indica a tendência de concentração espacial da variável população ao longo do período analisado, o que é comprovado pela redução, em diferentes medidas, das distâncias padrão, indicador que mede o nível de concentração/dispersão no espaço. Tal situação conduz à necessidade da realização de uma análise mais aprofundada, mais especificamente, um estudo da dependência espacial destas variáveis, ou seja, da autocorrelação espacial (pesquisa do autor em andamento).

Além destas indicações (hipóteses) que esperam explicações consistentes, podem-se apontar muitas outras a respeito dos resultados das análises apresentadas neste trabalho, sobretudo, no campo da análise regional. Por fim, devem-se ressaltar as capacidades e as possibilidades da interpretação e da análise da realidade objetiva advindas do aporte das técnicas de cunho espacial.

5 – Referências bibliográficas

ANSELIN, Luc. **Spatial econometrics: methods and models**. Dordrecht (The Netherlands): Kluwer Academic Publishers, 1988.

_____. **Spatial data analysis with GIS: an introduction to application in the social sciences**. Technical Report 92-10, 1992.

ANSELIN, Luc; FLORAX, Raymond J. G. M.; REY, Sergio J.. **Econometrics for spatial models: recent advances**. In: _____ (Editors). **Advances in spatial econometrics: methodology, tools and applications**. Berlin: Springer, 2004.

ANTENUCCI, John C.; BROWN, Kay; CROSWELL, Peter L.; et al. **Geographic Information Systems: a guide to the technology**. New York: Chapman & Hall, 1991.

DRUCK, Suzana; CARVALHO, Marília Sá; CÂMARA, Gilberto; et al. (editores). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004.

EBDON, David. **Statistics in Geography**. Oxford (UK): Blackwell, 1985.

- GEOGRAPHY DIVISION. **Population and geographic centers**. Washington, U.S. Bureau of the Census, 1991.
- ESRI. **ArcGIS Desktop Help**. Redlands (USA): ESRI, 2006.
- IBGE. **Censo Demográfico 1970**: Rio Grande do Sul. Rio de Janeiro, 2003. 1 CD-ROM.
- _____. **Censo Demográfico 1980**: Rio Grande do Sul. Rio de Janeiro, 2004. 1 CD-ROM.
- _____. **Censo Demográfico 1991**: Rio Grande do Sul. Rio de Janeiro, 2006. 1 CD-ROM.
- _____. **Censo Demográfico 2000**: Rio Grande do Sul. Rio de Janeiro, 2002. 1 CD-ROM.
- LEVINE, Ned. **CrimeStat**: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.1). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. March 2007.
- MAGUIRE, David J.; GOODCHILD, Michael F.; RHIND, David W. **Geographical Information Systems**. New York: Longman Scientific & Technical, vol. 1: principles, 1991.
- OAKES, Timothy. Place and the paradox of modernity. **Annals of the Association of American Geographers**, Oxford, vol. 87, n. 3, p. 509-531, 1997.
- PAIVA, Carlos Águedo Nagel (coord.). **RS em mapas e dados**: bases georreferenciadas para a comparação do desempenho socioeconômico dos municípios gaúchos entre 1966 e 2006. Porto Alegre, 2007. CD-ROM.
- SOJA, Edward W. **Geografias pós-modernas**: a reafirmação do espaço na teoria social. Rio de Janeiro: Jorge Zahar, 1993.
- SOUZA, Marcelo L. de. O território: sobre espaço e poder, autonomia e desenvolvimento. In: CASTRO, I. E.; GOMES, P. C. C.; CORRÊA, R. L. (organizadores). **Geografia**: conceitos e temas. Rio de Janeiro: Bertrand Brasil, 1995. p. 77-116.
- SVIATLOVSKY, E. E.; EELLS, Walter Crosby. The Centrophical Method and Regional Analysis. **Geographical Review**, v. 27, n. 2, p. 240-254, abr. 1937.