

**ANA CLARA PAIXÃO CAMPOS**

**DETECÇÃO DE PADRÕES ESPACIAIS NA DISTRIBUIÇÃO  
DOS PACIENTES PORTADORES DE DOENÇA GENÉTICA  
COM DEFICIÊNCIA FÍSICA DA ASSOCIAÇÃO DE ASSISTÊN-  
CIA À CRIANÇA DEFICIENTE (AACD) DE PERNAMBUCO**

RECIFE-PE – Fevereiro/2013.



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA**  
**APLICADA**

**DETECÇÃO DE PADRÕES ESPACIAIS NA DISTRIBUIÇÃO DOS PACIENTES  
PORTADORES DE DOENÇA GENÉTICA COM DEFICIÊNCIA FÍSICA DA AS-  
SOCIAÇÃO DE ASSISTÊNCIA À CRIANÇA DEFICIENTE (AACD) DE PER-  
NAMBUCO**

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: **Biometria e Estatística Aplicada.**

Orientador: Prof. Dr. Moacyr Cunha Filho

Coorientador: Prof. Dr. Guilherme Vilar

Coorientadora: Profa. Dra. Denise Nunes Viola

### Ficha catalográfica

C198d Campos, Ana Clara Paixão  
Detecção de padrões espaciais na distribuição dos  
pacientes portadores de doenças genética com deficiência  
física da Associação de Assistência à Criança Deficiente  
(AACD) de Pernambuco / Ana Clara Paixão Campos –  
Recife, 2013.

82 f. : il.

Orientador: Moacyr Cunha Filho  
Dissertação (Mestrado em Biometria e Estatística  
Aplicada) – Universidade Federal Rural de Pernambuco,  
Departamento de Estatística e Informática, Recife, 2013.

Referências.

1. Análise espacial 2. Índice de Moran 3. Teste de  
Aleatorização 4. Doença genética 5. Deficiência física  
I. Cunha Filho, Moacyr, orientador II. Título

CDD 574.018

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**DETECÇÃO DE PADRÕES ESPACIAIS NA DISTRIBUIÇÃO DOS PACIENTES PORTADORES DE DOENÇA GENÉTICA COM DEFICIÊNCIA FÍSICA DA ASSOCIAÇÃO DE ASSISTÊNCIA À CRIANÇA DEFICIENTE (AACD) DE PERNAMBUCO**

**Ana Clara Paixão Campos**

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 28/02/2013 pela Banca Examinadora.

Orientador:

---

Prof. Dr. Moacyr Cunha Filho  
Universidade Federal Rural de Pernambuco

Banca Examinadora:

---

Prof. Dr. Guilherme Vilar  
Universidade Federal Rural de Pernambuco

---

Prof (a). Dr (a). Ana Patrícia Siqueira Tavares Falcão  
Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco

---

Prof (a). Dr (a). Denise Nunes Viola  
Universidade Federal da Bahia

Dedico este trabalho a minha família,  
em especial a minha mãe e minha fi-  
lha, minhas razões de viver e de que-  
rer ser sempre uma pessoa melhor.

## Agradecimentos

Eu descreveria a minha trajetória no mestrado como uma aventura, e por sinal, bastante arriscada. Deixe-me explicar o porquê: primeiro, eu estava terminando a graduação em Estatística na UFBA e resolvi juntamente com outros colegas fazer a inscrição para a seleção do mestrado no Programa de Pós Graduação e Estatística Aplicada da UFRPE. Nesse meio tempo, descobri que estava grávida. Alguns meses depois saiu o resultado e lá estava meu nome na lista. Nesse momento minha cabeça deu um nó e surgiram inúmeras dúvidas. O que vou fazer para estudar com um bebê de colo e ainda por cima sem a minha família para ajudar? Será que vou dar conta? Será que vou ser uma boa mãe? Nem cozinhar eu sei... Após muitas noites de choro, conversas com a família e com amigos, resolvi encarar o desafio. E por isso, para chegar a este momento e realizar mais um sonho, muitos acontecimentos e pessoas foram envolvidos.

Inicio meus agradecimentos por **DEUS**, já que Ele me concedeu mais essa oportunidade e para completar colocou pessoas tão especiais ao meu lado, sem as quais certamente não teria dado conta.

A meus pais, Clélia e Carlos Alberto que souberam me educar e colocar no rumo certo, alcançando mais essa vitória. Amo vocês! Um agradecimento especial a minha mãe por sempre acreditar em minha capacidade e me achar A MELHOR de todas, mesmo não sendo. Obrigada por me apoiar e sempre estar ao meu lado principalmente no momento mais especial da minha vida que foi o nascimento da minha filha. Mãe, obrigada por cuidar dela, aguentar minha choradeira nos momentos de desespero, me dar conselhos e principalmente muito amor.

A minha pequena **Sophia**, pelos momentos de alegria e por me fazer tão feliz mesmo nos momentos de cansaço total. Você é o anjinho que Deus me mandou para fazer a minha vida ter sentido. Mamãe te ama! Aos meus irmãos Abílio e Manuela, pelo companheirismo, pelos momentos de bagunça e diversão e pela paciência nos momentos de estresse.

A minha amada família Campos por vibraram comigo, desde a aprovação no vestibular, e sempre fazerem “propaganda” positiva a meu respeito. Obrigada pelo

carinho e pela força! Obrigada a minha querida Vó Clara pelo exemplo de determinação e honestidade. A minha prima Lucy, sou grata pelos conselhos e por sempre me deixar de alto astral. Ao meu padrinho por acreditar nos meus sonhos e sempre tentar realizá-los de alguma forma. A família Teles por me acolher como a uma filha e sonhar comigo todos os meus sonhos. Muito obrigada Jami e minha querida Waldelice pelo amor que demonstram por mim e por Sophia. Amo vocês!

As minhas amigas de sempre, Cacau, Mara, Vivi, Paloma, Dea e Mayara, por só quererem o meu bem e me valorizar tanto como pessoa. Obrigada pela amizade! A Leandro, meu amigo mais enrolado, por estar sempre ao meu lado e pelo amor que demonstrado à minha Sophia. Não poderia ter escolhido padrinho melhor. As minhas amigas de resenhas, choros e risadas, Talita, Bel, Sheila, Fernanda e Carla. Aos meus amigos de graduação pelos ótimos momentos de descontração e também de finais de semana de estudos intensivos. Em especial, Nadinho, Biel, Carlos, Sandra, Elen, Gabi, Aline, Silvânia e Talita. As novas amizades, Dani, e Jéssica, pela atenção, cuidado e por aguentar minhas loucuras.

Ainda durante a graduação tive a chance de estagiar por quase dois anos no Help Desk da Caixa Econômica, onde aprendi acima de tudo, a ter responsabilidade e comprometimento com meu trabalho seja ele qual fosse. Lá fiz grandes amigos e trabalhei com profissionais muito dedicados: Ritinha, Wilminha, Léo, Jean, Gustavo, obrigada por todo aprendizado e pelo carinho sempre demonstrado. Aos amigos que fiz durante o estágio e que ainda hoje participam intensamente da minha vida : Tami, Renatinha e Ricardo.

Perto de completar 2 anos no estágio, eu fui convidada a participar de um projeto de Iniciação Científica pela professora Leila Denise. Tamanha foi a minha surpresa pelo convite que eu nem pensei antes de aceitar. Dali em diante me apaixonei pela área acadêmica e resolvi que eu queria ser professora/pesquisadora. Eu agradeço imensamente a professora Leila pela oportunidade e por tanto me encher de vontade de ser uma boa profissional. Você com certeza é a minha referência de sucesso. Agradeço também a todos os outros professores da graduação pelos ensinamentos, conselhos, puxões de orelha e incentivo nas horas mais difíceis.

Aos amigos que conheci no mestrado, pelas alegrias, companheirismo, pelo ombro, por me acolherem em suas casas, aguentarem a minha bagunça, falação e dividirem seu tempo comigo e com Sophia. Em especial, Samuel, Rosilda, Érika,

Nyedja, Léo, Silvio, Neto, Gabi, Dans. E nesse mundão eu tive a sorte de encontrar pessoas de coração bom que tornaram a minha vida menos complicada e mais feliz enquanto estive em Recife. Eu agradeço imensamente a vocês, Rodrigo (s), Danila, Paulo, Marcelo e Marynha, pelo total apoio, carinho, conselhos, caronas, comidinhas gostosas (não é, Nila?), abraços e palavras que confortam qualquer coração. Nunca esquecerei o que fizeram por mim e por Sophia.

Ao Prof. Dr. Moacyr Cunha Filho que confiou em uma aluna recém-formada em Estatística, com alguma experiência em pesquisa científica e grávida. Obrigada por todo auxílio, orientação, conselhos, paciência e incentivo nos momentos difíceis. A Prof<sup>a</sup>. Dra. Denise Viola pela parceria, ensinamentos e por sempre estar disponível para tirar minhas dúvidas e me ouvir. Ao Prof. Dr. Guilherme Vilar pelo apoio e confiança e a AACD pela parceria.

Aos professores do Programa de Pós-Graduação em Biometria e Estatística Aplicada e aos funcionários por toda a ajuda. Em especial a professora Tatjana pelos conselhos e por sempre estar disposta a nos ajudar. Agradeço, também, a CAPES pela bolsa de estudos concedida.

Ninguém vence sozinho... OBRIGADA A TODOS!



*“Se as coisas são inatingíveis... Ora,  
não é motivo para não querê-las...  
que tristes os caminhos que não fora  
a presença distante das estrelas...”*

*Mário Quintana*

## Resumo

Conhecer o padrão espacial dos portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco é de grande importância, pois torna possível orientar a assistência aos cuidados básicos desses indivíduos e fornecer base sólida para o planejamento de políticas públicas de saúde. Entretanto, existem poucas referências na literatura utilizando o instrumental da análise espacial no contexto da deficiência física. A presente dissertação foi estruturada na forma de dois artigos científicos. No primeiro artigo o Índice  $I$  global de Moran foi utilizado para caracterizar o padrão espacial da taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco e os resultados encontrados foram comparados com os obtidos através do teste de aleatorização. Em ambas as metodologias constatou-se a existência de padrão espacial agregado para a taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco quando se levou em consideração as taxas dos 4 municípios mais próximos de cada localidade. No segundo artigo foi avaliado o desempenho do índice  $I$  global de Moran e do teste de Mantel com correlação de Spearman, ambos utilizando aleatorização para avaliar a significância da estatística, no que tange a capacidade de detectar padrão espacial para a taxa de pacientes portadores de doenças genéticas com deficiência física em tratamento na AACD de Pernambuco. Os resultados indicaram que o índice  $I$  global de Moran mostrou-se uma metodologia mais satisfatória para detectar do padrão espacial, uma vez que utiliza em seus cálculos as informações da vizinhança, além de proporcionar maior controle das taxas de rejeição da hipótese nula em estudo.

**Palavras-chave:** Deficiência física; Doença genética; Análise espacial; Índice  $I$  global de Moran; Teste de aleatorização; Teste de Mantel.

## Abstract

Knowing the spatial pattern of patients with genetic disease with physical disabilities in the treatment of Pernambuco AACD is of great importance because it makes it possible to guide the basic care assistance to these individuals and provide solid basis for planning public health policies. However, there are few references in the literature using the tools of spatial analysis in the context of disability. This dissertation was structured in the form of two papers. In the first article the global Moran's  $I$  index was used to characterize the spatial pattern of the rate of patients with genetic disease with physical disabilities in the AACD treatment of Pernambuco and the results were compared with those obtained with the randomization test. In both approaches we found the existence of spatial pattern for the rate of patients with genetic disease with physical disabilities in the treatment of Pernambuco AACD when we took into account the rates of four municipalities closest to each location. In the second article we evaluated the performance of the global Moran's  $I$  index and the Mantel test with Spearman correlation, both using randomization to assess the statistical significance, regarding the ability to detect spatial pattern for the rate of patients with diseases genetic with physical disabilities in the AACD treatment of Pernambuco. The results showed that the global Moran's  $I$  index proved to be a more satisfactory method for detecting the spatial pattern, since it uses the information in its calculations of the neighborhood, and provide greater control of the rejection rates of the null hypothesis under study.

**Keywords:** Physical disability; genetic disease; Spatial Analysis; global Moran's  $I$  index, Randomization test; Mantel test.

# Lista de Figuras

## Capítulo 2

Figura 1: Estado de Pernambuco de acordo com suas mesorregiões. Fonte: Elaborado a partir de base cartográfica do IBGE. ....	22
Figura 2: Esquema explicativo do diagrama de espalhamento de Moran. ....	34
Figura 3: Total de pacientes portadores de doença genética com deficiência física em tratamento na AACD por município – Pernambuco. ....	39
Figura 4: Taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (Ipac). ....	40
Figura 5: Taxa suavizada pelo Estimador Bayesiano Empírico Global para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBG). ....	41
Figura 6: Taxa suavizada pelo Estimador Bayesiano Empírico Local para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBL). ....	41
Figura 7: Histograma e <i>Boxplot</i> da taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (Ipac). ....	44
Figura 8: Histograma e <i>Boxplot</i> da taxa suavizada pelo estimador Bayesiano empírico global para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBL). ....	44
Figura 9: Histograma e <i>Boxplot</i> da taxa suavizada pelo estimador Bayesiano empírico local para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBL). ....	45
Figura 10: Gráfico de pontos para a distribuição espacial da variável Ipac. ....	45
Figura 11: Diagrama de dispersão de Moran para a variável Ipac. ....	49
Figura 12: Diagrama de dispersão de Moran para a variável IpacBG. ....	49
Figura 13: Diagrama de dispersão de Moran para a variável IpacBL. ....	50
Figura 14: LISA MAP – Municípios classificados conforme a significância do Índice <i>I</i> local de Moran para a variável Ipac segundo o critério dos 4 vizinhos mais próximos. ....	51

Figura 15: LISA MAP – Municípios classificados conforme a significância do Índice *I* local de Moran para a variável IpacBG segundo o critério dos 4 vizinhos mais próximos.....51

Figura 16: LISA MAP – Municípios classificados conforme a significância do Índice *I* local de Moran para a variável IpacBL segundo o critério dos 4 vizinhos mais próximos.....52

### **Capítulo 3**

Figura 1: Gráfico de pontos para a distribuição espacial da taxa bruta de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco.....74

Figura 2: Relação entre a distância geográfica dos municípios pernambucanos e a taxa bruta de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco (Ipac).....77

# Lista de Tabelas

## Capítulo 2

Tabela 1. População e área dos 102 municípios pernambucanos em estudo .....23

Tabela 2: Classes e subclasses das doenças genéticas em estudo de acordo com o CID-10.....26

Tabela 3: Análise descritiva das variáveis Ipac, IpacBG e IpacBL.....42

Tabela 4: Índice  $I$  global de Moran e p-valor (entre parênteses) para as variáveis Ipac, IpacBG e IpacBL referente aos 102 municípios pernambucanos em estudo. ....47

Tabela 5: Resultado (p-valor) do teste de aleatorização para as variáveis Ipac, IpacBG e IpacBL. ....53

## Capítulo 3

Tabela 1: Classes e subclasses das doenças genéticas em estudo de acordo com o CID-10.....67

Tabela 2: P-valor do teste de aleatorização para detecção de padrão espacial da variável Ipac e proporção ( $p$ ) de rejeição de  $H_0$  para as estatísticas de teste índice  $I$  global de Moran com matriz de pesos  $k=4$  vizinhos mais próximos e coeficiente de correlação de Spearman ( $\rho$ ).....76

# Sumário

Capítulo 1 .....	15
1. Introdução Geral .....	15
2. Referências .....	9
Capítulo 2 .....	13
Identificação da Existência de Padrão Espacial para Pacientes Portadores de Doença Genética com Deficiência Física em Tratamento na AACD de Pernambuco: Uma Aplicação do Índice de Moran .....	13
1. Introdução .....	16
2. Desenvolvimento .....	22
2.1. Área de Estudo e Material Utilizado .....	22
2.1.1. Suavização .....	26
2.2. Métodos .....	28
2.2.1. Análise Espacial de Área .....	28
2.2.2. Autocorrelação Espacial .....	29
2.2.3. Matriz de proximidade espacial .....	30
2.2.4. Índice $I$ global de Moran .....	31
2.2.5. Diagrama de espalhamento de Moran .....	33
2.2.6. Índice $I$ local de Moran .....	34
2.2.7. LISA map .....	35
2.2.8. Teste de Aleatorização .....	36
3. Resultados e Discussão .....	38
Mapas .....	38
Análise Exploratória de Dados .....	42
Índice $I$ global de Moran .....	46
Diagrama de dispersão de Moran .....	47
Índice $I$ local de Moran .....	50
Teste de Aleatorização .....	52
4. Considerações Finais .....	54

5. Referências .....	55
<b>Capítulo 3 .....</b>	<b>60</b>
<b>Comparação de Métodos para Identificação de Padrão Espacial para Pacientes Portadores de Doença com Deficiência Física em Tratamento na AACD Pernambuco .....</b>	<b>60</b>
1. Introdução.....	62
2. Desenvolvimento.....	67
2.1. Material utilizado.....	67
2.2. Métodos .....	68
2.2.1. Autocorrelação espacial.....	68
2.2.2. Matriz de pesos espaciais ou de proximidade espacial ( $W$ ).....	68
2.2.3. Índice $I$ global de Moran .....	69
2.2.4. Testes de Aleatorização .....	71
2.2.5. Teste de Mantel .....	71
2.2.6. Coeficiente de correlação de postos de Spearman.....	73
3. Resultados e Discussão.....	73
4. Considerações Finais .....	78
5. Referências .....	79



# Capítulo 1

---

## 1. Introdução Geral

O termo *deficiência* tem sido cada vez mais discutido pela população brasileira, seja através de políticas públicas ou da internet. Isso decorre, principalmente, pelo fato da população de deficientes ter se conscientizado acerca de seus direitos, e a população ter se atentado quanto às diferenças individuais que se fazem presentes no cerne de uma sociedade (BAPTISTA, 2009).

De acordo com Art. 4º do Decreto nº 3.298 de 1999 da legislação brasileira, a deficiência física refere-se à alteração parcial ou completa de um ou mais segmentos do corpo humano, acarretando o comprometimento da função física, apresentando-se sob a forma de paraplegia, paraparesia, monoplegia, monoparesia, tetraplegia, tetraparesia, triplegia, triparesia, hemiplegia, hemiparesia, amputação ou ausência de membro, paralisia cerebral, membros com deformidade congênita ou adquirida, exceto as deformidades estéticas e as que não resultam em dificuldades para o desempenho de funções (BRASIL, 2007).

A Organização Mundial da Saúde (OMS) estima que, em tempos de paz, 10% da população de países desenvolvidos são constituídos de pessoas com algum tipo de deficiência. Para os países em vias de desenvolvimento estima-se entre 12% a 15% da população. Destes, 20% seriam portadores de deficiência física (BRASIL, 2006).

As causas da deficiência física são diversas, e dentre elas destacam-se problemas genéticos, complicações na gestação ou gravidez, além de doenças infantis ou acidentes. Entre as etiologias das deficiências físicas estão os fatores genéticos, fatores virais ou bacterianos, fatores neonatal e fatores traumáticos. As causas podem ser ainda pré-natais, peri-natais ou pós-natais dependendo do momento em que ocorre a alteração ou má-formação (BRASIL, 2006). Dentre as doenças causadas por fatores genéticos estão as de Depósito Lisossômico (DDLs) e as Neuromusculares (DNM).

As DDLs são um grupo de aproximadamente 40 raros distúrbios metabólicos hereditários que resultam de defeitos na função dos lisossomos. As mucopolissacaridoses (MPS) representam um desses distúrbios e caracterizam-se pelo funcionamento inadequado de determinadas enzimas, envolvidas na degradação de ácidos glicosaminoglicanos (GAG). As MPS podem ser classificadas em sete tipos diferentes, de acordo com a enzima que está em falta no portador da doença (VIEIRA, 2007). O Brasil possui mais de 600 pessoas portadoras de MPS e Pernambuco ocupa o segundo lugar entre os estados brasileiros com maior ocorrência da doença.

As DNM referem-se a um conjunto de mais de 40 patologias diferentes e engloba as doenças que acometem a fibra muscular (Miopatias), doenças dos nervos (Neuropatias), doenças dos cornos anteriores da medula (Atrofias Espinais) e as perturbações da junção neuromuscular (Miastenias) (OLIVEIRA & GABBAI, 2001; REED, 2002). Estima-se que essas doenças afetam uma em cada 3.500 pessoas em todo o mundo.

Segundo a OMS, são consideradas raras as doenças que afetam até 65 em cada 100 mil habitantes de uma localidade. Estima-se que 7% da população mundial é portadora de algum tipo síndrome considerada rara. No Brasil ainda existem poucos registros quando comparado com outros países, o que leva à suspeita de que muitos casos não sejam diagnosticados, em especial nas regiões Norte e Nordeste.

O avanço da Medicina no conhecimento da etiologia das doenças, dos tipos e manifestações das deficiências, também levou a uma classificação das mesmas, de forma cada vez mais específica. A Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID-10), foi conceituada para padronizar e catalogar as doenças e problemas relacionados à saúde, tendo como referência a Nomenclatura Internacional de Doenças, estabelecida pela OMS (DI NUBILA, 2007).

Segundo novos dados do Censo 2010 divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil tem 26,5 milhões de pessoas portadoras de necessidades especiais e desse total, cerca de 14% são exclusivamente deficientes físicos. Estima-se que em Pernambuco a população é superior a 1,3 milhão de deficientes, ou seja, 17,4% da população de Pernambuco possui algum tipo de deficiência, o que corresponde à quarta maior taxa do Brasil (IBGE, 2010).

Existem várias instituições que auxiliam os familiares a cuidar dos deficientes e dentre as mais importantes está a Associação de Assistência à Criança Deficiente (AACD). A AACD é uma entidade privada, filantrópica especializada no tratamento de pessoas com deficiência física e foi fundada em 1950 pelo médico Renato Bonfim. A instituição mantém um amplo serviço de assistência à saúde, voltado principalmente às crianças e adolescentes, promovendo a reabilitação e a inclusão social dessas pessoas. Considerada centro de referência no tratamento de pessoas com deficiência física, a AACD dispõe de treze unidades em todo país, tendo sido instalada em Recife-PE no ano de 1999 e batizada como Centro de Reabilitação Engenheiro Clóvis Scripilliti. Atualmente, a unidade atende cerca de 700 pessoas por dia, sendo metade delas da Região Metropolitana do Recife e a outra metade oriunda do interior de Pernambuco e de outros estados do Norte e Nordeste (BRASIL, 2012).

Apesar da assistência médica ser um bem público, ela não está igualmente disponível para todas as pessoas. Muitas vezes estes centros de tratamento são de difícil acesso a população que reside em local distante ou que não tenha condições financeiras de transporte e moradia para realizar o tratamento, principalmente quando residentes em cidades distantes dos Centros de Reabilitação. Dessa forma, pode-se dizer que a demanda por serviços públicos é continuamente distribuída no espaço de acordo com a distribuição da população, e que, além disso, estes serviços são providos somente em localizações específicas. Consequentemente, existirão desigualdades no acesso aos serviços, dos custos de transporte, tempo de viagem, entre outros. O acesso à assistência médica é influenciado também por fatores como horários de funcionamento e listas de espera que possuem um importante papel na determinação dos indivíduos ou subgrupos da população que podem ter acesso à assistência.

O conhecimento do perfil dos portadores de doença genética com deficiência física de uma região é de grande importância, pois torna possível orientar a assistência aos cuidados básicos desses indivíduos e fornecer base sólida para o planejamento de políticas públicas de saúde destinadas à essas pessoas. Desta forma, os serviços e investimentos voltados para essa população podem ser otimizados, melhorando a qualidade de vida e promovendo igualdade de oportunidades (BAPTISTA & SILVA, 2010).

A trajetória histórica da distribuição geográfica de doenças tem seus primeiros registros em 1854, em um estudo sobre a cólera em Londres, no qual foram ma-

peados os casos e os pontos de coletas de água, mostrando o papel da contaminação da água na ocorrência da doença (BARRET, 1991; SCHOLTEN & LEPPER, 1991).

Desde então, a análise espacial tem sido utilizada em diversas áreas do conhecimento, tais como nas ciências sociais para estudar os padrões de assentamentos antigos e estudos antropológicos de redes sociais. Já na economia, os dados espaciais são estudados para modelar a estrutura de variáveis socioeconômicas, tais como taxas de desemprego, da demanda do consumidor doméstico, dos preços da gasolina ou habitação (WHITE et al., 1981; DOREIAN et al., 1984; HAINING, 1984; WHITLEY & CLARK, 1985; BRONARS & JANSEN, 1987; KVAMME, 1990; CASE, 1991; DUBIN, 1992). Podem-se ainda encontrar aplicações no estudo dos padrões de crescimento das árvores na área florestal, no estudo da diversidade de plantas em ecologia, no estudo do rendimento de culturas ou mesmo na avaliação das relações internacionais (ANSELIN, 1992a).

O desenvolvimento da tecnologia de computadores e de ferramentas matemáticas para análise espacial, que ocorreu na segunda metade do século XX, abriu possibilidades diversas, entre elas a habilidade de armazenar, recuperar e combinar os dados disponíveis sobre um território (CÂMARA et al., 2001).

No Brasil, a análise espacial vem sendo utilizada nos últimos anos em áreas como segurança viária, ecologia, saúde pública, economia, e em muitas outras áreas que se beneficiam bastante dessas técnicas (CARVALHO, 1997; BISPO & OLIVEIRA, 1998; PINTO et al., 2003; QUEIROZ, 2003; TOMAZELLI, 2003; SANTOS & JUNIOR, 2006; ZANETTE et al., 2007; ALMEIDA, 2008; RODRIGUES et al., 2008; FERRARIO, 2009; MACHADO et al., 2009).

A partir da década de 1990 houve um grande aumento dos trabalhos cujo foco é a análise das situações de saúde, consolidando a importância da análise do espaço social na compreensão do processo saúde-doença e buscando entender a situação de saúde como resultado da relação entre os grupos sociais e seu território (COSTA & TEIXEIRA, 1999; BARCELLOS et al., 2002). Em estudos epidemiológicos, por exemplo, frequentemente deseja-se saber se a taxa de ocorrência de um determinado evento é espacialmente correlacionada em uma determinada área.

A análise espacial de taxas é importante em diversas áreas do conhecimento. Castro, Vieira e Assunção (2004) realizaram um estudo com o objetivo de direcionar

políticas públicas voltadas para o combate ao câncer de pulmão. Os resultados obtidos nesse trabalho mostraram que as maiores razões de mortalidade localizam-se aos extremos norte e sul da região estudada. Já para as mulheres, o sul do Paraná e do Rio Grande do Sul compõem as áreas com maior risco. Acosta (2008) estudou a relação da distribuição espacial da taxa de incidência da tuberculose pulmonar bacilífera em bairros da cidade de Porto Alegre com o nível socioeconômico da população. Os resultados obtidos puderam indicar áreas de maior vulnerabilidade social da cidade sinalizando que esse indicador de saúde pode ser utilizado para possíveis tomadas de decisão pelos gestores públicos.

As técnicas convencionais de análise estatística, geralmente, têm como base amostras independentes e identicamente distribuídas (i.i.d). Isto não acontece em dados com estrutura de dependência espacial e/ou temporal, uma vez que as amostras são correlacionadas. Quando os dados apresentam evidência de distribuição espacial implica que as observações próximas são mais semelhantes do que as distantes. Uma vez identificada a existência de padrão espacial, as análises convencionais não podem ser aplicadas (VIOLA, 2007). A idéia central da análise de dados espaciais é mensurar propriedades e sua relação com seus vizinhos, levando em conta a localização espacial do fenômeno em estudo de forma explícita (CÂMARA et al., 2002).

Para a análise de dados espaciais, três tipos de dados georreferenciados são considerados: dados de processos pontuais, dados de superfície contínua e dados agregados por áreas. Os dados de áreas são obtidos por levantamentos populacionais ou amostrais e sua localização está associada a áreas delimitadas por polígonos, representado apenas com um valor por área (BAILEY & GATRELL, 1995), que é obtido através da informação da latitude e longitude do centro do polígono.

De acordo com Anselin (1998a), a Análise Exploratória de Dados Espaciais (AEDE) ou *Exploratory Spatial Data Analysis* (ESDA) são importantes instrumentos para descrever e visualizar distribuições espaciais, identificar locais atípicos ou *outliers* espaciais, descobrir padrões de associação espacial, *clusters* ou aglomerados espaciais e sugerir padrões espaciais ou outras formas de heterogeneidade espacial.

Outro aspecto fundamental da análise exploratória é a caracterização da dependência espacial. Anselin (1992), diz que a forma como os dados estão distribuídos no espaço é um importante indicador da interação espacial, a qual é revelada nas medidas de associação ou autocorrelação espacial. Existem diversas formas de esti-

mar a magnitude da dependência espacial, sendo o Índice  $I$  de Moran a estatística comumente utilizada. Após o calculado o Índice  $I$  de Moran, é importante estabelecer sua validade estatística.

No entanto, se as amostras não forem aleatórias, não parece justificável usar na análise estatística, técnicas convencionais. Nesta situação, é mais apropriado usar as técnicas com base no modelo de distribuição aleatória, como é o caso dos testes de aleatorização. Estes testes não exigem pressupostos em relação à distribuição dos dados e a estatística de teste é comparada com a estabelecida após aleatorizar um grande número de vezes os dados.

A base teórica dos testes de aleatorização foi proposta no século XX e atribuiu-se o seu desenvolvimento a muitos investigadores (KEMPTHORNE, 1952, 1955; EDGINGTON, 1964, 1966; HINKELMANN & KEMPTHORNE, 1994; EDGINGTON & ONGHENA, 2007). Entretanto, com o advento dos computadores, o interesse nestes testes passou de considerações teóricas a aplicabilidade prática sendo amplamente utilizados nas mais diversas áreas do conhecimento, dentre elas, destacam-se: Biologia, Química, Psicologia, Medicina e Experimentação agrônômica (MANLY, 1997; EDGINGTON & ONGHENA, 2007; VIOLA 2007).

O que distingue os testes de aleatorização dos testes clássicos é a forma de calcular a significância. No caso dos testes de aleatorização a significância é calculada por um procedimento em que repetidamente se aleatorizam os dados, e a significância calculada dessa maneira é exata. Com este procedimento, o investigador pode avaliar a significância de qualquer teste estatístico, mesmo que ainda não tenha sido derivada a distribuição amostral da estatística do teste (BRANCO, 2010). Em relação ao teste de Mantel, uma estatística comumente utilizada para avaliar a correlação entre pares de elementos de matrizes é o coeficiente de correlação de Pearson. Para isso, é necessário que a amostra provenha de uma população com distribuição normal bivariada. Quando esse pressuposto não é satisfeito utiliza-se como alternativa o coeficiente de correlação de postos de Spearman.

Diante do exposto, a presente dissertação tem por objetivo procurar uma melhor representação e entendimento dos padrões de distribuição espacial dos pacientes portadores de doenças genéticas com deficiência física das classes CID E76, G12, G60, G71 e subclasses E76.0, E76.2, G12, G12.1, G60.0, G60.8, G71.0, G71.1,

G71.2, G71.8 e G71 em tratamento na AACD de Pernambuco e residentes em 102 municípios do Estado. A classe E76 refere-se aos Distúrbios do metabolismo do glicosaminoglicano, a G12 diz respeito à Atrofia muscular espinal e síndromes correlatas e as classes G60 e G71 integram os grupos de doenças relacionadas a Neuropatias hereditárias e idiopáticas e Transtornos primários dos músculos, respectivamente.

A análise tem como base as ferramentas da estatística espacial e volta suas atenções e avaliações para questões práticas que visem ajudar a tomadas de decisões que tenham impacto na definição de uma descentralização no atendimento sejam através de pontos avançados e/ou de serviços de telemedicina, bem como contribuir para a elaboração de novos trabalhos. Trata-se de um estudo que identifica locais onde há maior concentração de pacientes em tratamento na AACD de Pernambuco de modo a direcionar quais municípios podem ser utilizados para a construção de novos postos de atendimento contribuindo assim para diminuição de filas de espera na AACD, além de propiciar aos pacientes e familiares viagens mais curtas, menos desgastantes e dispendiosas. Outro aspecto importante diz respeito aos custos que os municípios têm com o transporte dos pacientes e familiares, o qual seria diminuído se o local de destino do paciente fosse mais próximo à cidade onde reside.

A dissertação foi estruturada na forma de dois artigos científicos, sendo os mesmos organizados da seguinte forma:

**Artigo 1:** Detecção da existência de padrão espacial para pacientes portadores de doenças genéticas com deficiência física em tratamento na AACD de Pernambuco: Uma aplicação do Índice  $I$  de Moran.

**Resumo:** O objetivo deste estudo foi conhecer o comportamento do padrão espacial dos pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco de modo a fornecer visão mais ampla e fidedigna da realidade dessas pessoas. O índice  $I$  global de Moran foi utilizado para caracterizar o padrão espacial desses pacientes e os resultados encontrados foram comparados com os obtidos através do teste de aleatorização. Os resultados obtidos dessas informações são capazes de orientar a assistência aos cuidados básicos desses pacientes, contribuir para a identificação de áreas prioritárias para a construção de postos de atendimento que visam, entre outras coisas, diminuir as filas de espera na unidade da AACD

em Recife, além de fornecer base sólida para o planejamento de políticas públicas de saúde.

**Artigo 2:** Comparação de métodos para identificação de padrão espacial para pacientes portadores de doenças genéticas com deficiência física que fazem tratamento na AACD de Pernambuco.

**Resumo:** O índice  $I$  global de Moran é a ferramenta usual para se medir a intensidade da autocorrelação espacial em taxas de incidência de doenças. A significância dos momentos de  $I$  pode ser avaliada, sob  $H_0$ , por duas suposições básicas: aproximação normal e testes de aleatorização. Porém, em muitos casos, o interesse é estudar a relação entre os elementos de duas matrizes de distâncias e nesse caso, pode-se aplicar o teste de aleatorização de Mantel. Assim, por meio do presente estudo, buscou-se avaliar o desempenho do índice  $I$  global de Moran e do teste de Mantel no que tange a capacidade de detectar padrão espacial para a taxa de pacientes portadores de doenças genéticas com deficiência física em tratamento na AACD de Pernambuco para diferentes tamanhos de amostra.



## 2. Referências

ACOSTA, L. M. **O mapa de Porto Alegre e a Tuberculose: distribuição espacial e determinantes sociais.** Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 2008.

ALMEIDA, C. G. **Análise espacial dos fragmentos florestais na área do Parque Nacional dos Campos Gerais, Paraná.** Ponta Grossa - PR: Universidade Estadual de Ponta Grossa, 2008.

ANSELIN, L. **SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data.:** University of Illinois, 1992a.

ANSELIN, L. **Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences.** Santa Barbara - CA: National Center for Geographic Information and Analysis, 1992.

ANSELIN, L. Interactive techniques and exploratory spatial data analysis. In: LONGLEY, P., et al. **Geographical Information Systems: principles, techniques, management and applications.** Cambridge: Geoinformation International, 1998a.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis.** Great Britain: Longman, 1995.

BAPTISTA, A. K.; SILVA, F. C. M. Perfil demográfico e do emprego das pessoas com deficiência no município de Belo Horizonte, MG – Brasil. **Revista de Medicina de Minas Gerais**, Belo Horizonte - MG, p. 290-298, 2010.

BAPTISTA, E. A. **Os deficientes no Estado de Minas Gerais: uma abordagem espacial e demográfica utilizando os dados do Censo Demográfico de 2000.** Belo Horizonte - MG: Pontifícia Universidade Católica de Minas Gerais, 2009.

BARCELLOS, C. et al. **Organização espacial, saúde e qualidade de vida: análise espacial e uso de indicadores na avaliação de situações de saúde.** 3. ed. [S.l.]: Informe Epidemiológico do SUS, v. 11, 2002.

BARRETT, F. A. "SCURVY" Linds medical geography. **Social Science & Medicine**, v. 33, p. 347-353, 1991.

BISPO, P. C.; OLIVEIRA, L. G. Distribuição espacial de insetos aquáticos (Ephemeroptera, Plecoptera e Trichoptera) em córregos de cerrado do Parque Ecológico de Goiânia, Estado de Goiás. **Oecologia Brasiliensis**, v. 5, p. 19-33, 1998.

BRANCO, F. J. S. **Investigação experimental**: potência estatística dos testes de aleatorização na comparação de dois grupos independentes. Lisboa, Portugal: Universidade Aberta, 2010.

BRASIL. **Educação infantil**: saberes e práticas da inclusão: dificuldades de comunicação e sinalização: deficiência física. 4. ed. Brasília: Ministério da Educação, 2006.

BRASIL. **A inclusão de pessoas com deficiência no mercado de trabalho**. Brasília: Ministro do Trabalho e Emprego, v. 2, 2007.

BRASIL. **Cidadania em Ação**. Ministério Público de Pernambuco. Recife - PE. 2012.

BRONARS, S.; JANSEN, D. The geographic distribution of unemployment rates in the US. **Journal of Econometrics**, v. 36, p. 251-279, 1987.

CÂMARA, G. M. et al. **Análise espacial de áreas**. Instituto Nacional de Pesquisas Espaciais (INEP). São José dos Campos - SP, p. 1-32. 2002.

CÂMARA, G. M.; DAVIS, C.; MONTEIRO, A. M. V. **Introdução à Ciência da Geoinformação**. Instituto Nacional de Pesquisas Espaciais (INEP). São José dos Campos - SP. 2001.

CARVALHO, M. S. **Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco a Saúde**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 1997.

CASE, A. Spatial patterns in household demand. **Econometrica**, v. 59, p. 953-965, 1991.

CASTRO, M. S. M.; VIEIRA, V. A.; ASSUNÇÃO, R. M. Padrões espaço-temporais da mortalidade por câncer de pulmão no Sul do Brasil. **Revista Brasileira de Epidemiologia**, v. 7, n. 2, p. 131-143, 2004.

COSTA, M. C. N.; TEIXEIRA, M. G. L. C. A concepção de “espaço” na investigação epidemiológica. **Caderno de Saúde Pública**, v. 15, n. 2, p. 271-279, 1999.

DI NUBILA, H. B. V.; BUCHALLA, C. M. O papel das Classificações da OMS - CID e CIF nas definições de deficiência e incapacidade. **Revista Brasileira de Epidemiologia**, v. 11, n. 2, p. 324-335, 2008. ISSN 1415-790X.

DOREIAN, P.; TEUTER, K.; WANG, C. H. Network autocorrelation models. **Sociological Methods and Research**, v. 13, p. 155-200, 1984.

DUBIN, R. A. Spatial autocorrelation and neighborhood quality. **Regional Science and Urban Economics**, v. 22, p. 433-452, 1992.

- EGDINGTON, E. S. Randomization tests. **Journal of Psychology**, v. 57, p. 445–449, 1964.
- EGDINGTON, E. S. Statistical inference and non-random samples. **Psychological Bulletin**, v. 66, p. 485–487, 1966.
- EGDINGTON, E. S.; ONGHENA, P. **Randomization tests**. 4. ed. Boca Raton: Chapman & Hall/CRC, 2007.
- FERRARIO, M. N. et al. Uma Análise Espacial do Crescimento Econômico do Estado do Paraná para os anos de 2000 e 2004. **Revista Brasileira de Estudos Regionais e Urbanos**, v. 3, p. 154-177, 2009.
- HAINING, R. Testing a spatial interacting-markets hypothesis. **Review of Economics and Statistics**, v. 66, p. 576-583, 1984.
- HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments, Volume I: Introduction to experimental design**. New York: Wiley, 1994.
- IBGE. **Censo Demográfico 2010 – Resultados Preliminares da Amostra**. Instituto Brasileiro de Geografia e Estatística. Rio de Janeiro. 2010.
- KEMPTHORNE, O. **Design and analysis of experiments**. New York: Wiley, 1952.
- KEMPTHORNE, O. The randomization theory of experimental inference. **Journal of the American Statistical Association**, v. 50, p. 946–967, 1955.
- KVAMME, L. K. Spatial autocorrelation and the Classic Maya Collapse revisited: refined techniques and new conclusions. **Journal of Archaeological Science**, v. 17, p. 197–207, 1990.
- MACHADO, J. P.; OLIVEIRA, R. M.; SOUZA-SANTOS, R. Spatial analysis of dengue occurrence and living conditions in Nova Iguaçu, Rio de Janeiro State, Brazil. **Caderno de Saúde Pública**, v. 25, p. 1025-1034, 2009.
- MANLY, B. F. J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**. 2. ed. London: Chapman and Hall, 1997.
- OLIVEIRA, A. S. B.; GABBAI, A. A. Doenças Neuromusculares. In: PRADO, F.C.; RAMOS, J.A.; VALLE, J.R. **Atualização Terapêutica: manual prático de diagnóstico e tratamento**. 20. ed. São Paulo: Artes Médicas, 2001. p. 815-827.
- PINTO, M. P.; BINI, L. M.; DINIZ-FILHO, J. A. F. Análise quantitativa da influência de um novo paradigma ecológico: autocorrelação espacial. **Acta Scientiarum**, Maringá - PR, v. 25, n. 1, p. 137-143, 2003.
- QUEIROZ, M. P. **Análise espacial dos acidentes de trânsito do município de Fortaleza**. Fortaleza - CE: Universidade Federal do Ceará, 2003.

REED, U. C. Doenças neuromusculares. **Jornal de Pediatria**, v. 78, p. 89-103, 2002.

RODRIGUES, A. F.; ESCOBAR, A. L.; SOUZA-SANTOS, R. Análise espacial e determinação de áreas para o controle da malária no Estado de Rondônia. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 41, n. 1. ISSN 0037-8682, 2008.

SANTOS, L.; A. JUNIOR, R. A. A. Distribuição Espacial dos Acidentes de Trânsito em São Carlos (SP): Identificação de Tendências de Deslocamento Através da Técnica de Elipse de Desvio Padrão. **Revista Caminhos de Geografia**, v. 7, n. 18, p. 134-145, 2006.

SANTOS, S. M. et al. Detecção de aglomerados espaciais de óbitos por causas violentas em Porto Alegre, Rio Grande do Sul, Brasil, 1996. **Caderno de Saúde Pública**, v. 17, n. 5, p. 1141-1151, 2001.

SCHOLTEN, H. J.; LEPPER, M. J. C. The benefits of the application of geographical information systems in public and environmental health. **WHO Statistical Quarterly**, v. 44, n. 3, 1991.

TOMAZELLI, J.; CZERSNIA, D.; BARCELLOS, C. Distribuição dos casos de Aids em mulheres no Rio de Janeiro, de 1982 a 1997: uma análise espacial. **Caderno de Saúde Pública**, v. 19, p. 1049-1061, 2003.

VIEIRA, T. A. **História natural das mucopolissacaridoses**: uma investigação da trajetória dos pacientes desde o nascimento até o diagnóstico. Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 2007.

VIOLA, D. N. **Detecção e modelagem de padrão espacial em dados binários e de contagem**. Piracicaba – SP: Departamento de Ciências Exatas, ESALQ - USP, 2007.

WHITE, R. D.; BURTON, M. L.; DOW, M. M. Sexual division of labor in African agriculture: a network autocorrelation analysis. **American Anthropologist**, v. 84, p. 824-49, 1981.

WHITLEY, D. S.; CLARK, W. A. V. Spatial autocorrelation tests and the Classic Maya collapse: methods and inferences. **Journal of Archaeological Science**, v. 12, p. 377-95, 1985.

ZANETTE, S. V. et al. Análise espacial da umidade do solo culti-vado com soja sob dois sistemas de manejo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 11, n. 3, p. 239-247. ISSN 1807-1929, 2007.

## Capítulo 2

---

### **Identificação da Existência de Padrão Espacial para Pacientes Portadores de Doença Genética com Deficiência Física em Tratamento na AACD de Pernambuco: Uma Aplicação do Índice de Moran**

**Resumo:** O termo deficiência tem sido cada vez mais discutido pela população, seja através de políticas públicas ou da internet. A Organização Mundial da Saúde (OMS) estima que, em tempos de paz, 12% a 15% da população de países em desenvolvimento é constituída de pessoas com algum tipo de deficiência. Destes, 20% seriam portadores de deficiência física. As causas da deficiência física são diversas, podendo estar ligadas a problemas genéticos, complicações na gestação ou gravidez, doenças infantis e acidentes. Conhecer o padrão espacial dos portadores de doença genética com deficiência física de uma região é de grande importância, pois torna possível orientar a assistência aos cuidados básicos dessas pessoas e fornecer base sólida para o planejamento de políticas públicas de saúde. A Associação de Assistência à Criança Deficiente (AACD) é uma das mais importantes instituições de auxílio especializado ao deficiente físico. Entretanto, mesmo com todo esforço para prover atendimento e estrutura adequados, a AACD ainda tem filas de espera. Dessa forma, a análise espacial é essencial para a identificação de áreas onde a saúde é precária, necessitando de atenção diferenciada. Para a avaliação da existência de padrão espacial dos portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco inicialmente foram calculadas as taxas de incidência brutas e suavizadas pelo método bayesiano empírico global e local. A etapa seguinte consistiu no cálculo do Índice  $I$  global de Moran para cada uma das referidas taxas e os resultados obtidos foram comparados com o resultado do teste de aleatorização. Através do Índice  $I$  global de Moran verificou-se que existe padrão espacial ( $I = 0,102$ ) apenas para a taxa bruta de pacientes considerando os 4 vizinhos mais próximos, o que também foi confirmado pelo teste de aleatorização (p-valor = 0,04). Estes resultados contribuem para

a identificação de áreas prioritárias para a construção de postos de atendimento que visam, entre outras coisas, diminuir as filas de espera na unidade da AACD em Recife-PE, propiciar aos pacientes e familiares de localidades mais distantes viagens menos desgastantes e dispendiosas, além de ajudar a localizar pacientes que ainda estejam sem tratamento adequado devido a dificuldade de acesso a centros de tratamento mais distantes e/ou por falta de diagnóstico correto da doença.

**Palavras - chave:** Deficiência física; Doença genética; Análise espacial; Índice *I* de Moran; Teste de aleatorização.

**Abstract:** The term disability has been increasingly discussed by the population, is a beam of public policy or the internet. The World Health Organization (WHO) estimates that, in peacetime, 12% to 15% of the population of developing countries is made up of people with a disability. Of these, 20% were physically disabled. The causes of disability are diverse and may be linked to genetic problems, complications in pregnancy or pregnancy, childhood illnesses and accidents. Knowing the spatial pattern of patients with genetic disease with physical disabilities is a region of great importance since it makes possible direct assistance to basic care such people and provide solid basis for planning public health policies. The Assistance Association for Disabled Children (AACD) is one of the most important institutions of specialized aid to the handicapped. However, even with every effort to provide appropriate care and structure, the AACD has yet queues. Thus, spatial analysis is essential for identifying areas where health is precarious, requiring special attention. To evaluate the existence of spatial pattern of patients with genetic disease with physical disabilities in the treatment of Pernambuco AACD were initially calculated incidence rates and gross smoothed by empirical Bayesian method global and local. The next step is to calculate the global index  $I$  Moran for each of said rates and the results obtained were compared with the test result of randomization. Through the global Moran index  $I$  I found that there is a spatial pattern ( $I = 0,102$ ) only for the gross rate of patients considering the 4 nearest neighbors, which was also confirmed by the randomization test ( $p\text{-value} = 0,04$ ). These results contribute to the identification of priority areas for the construction of service stations that aim, among other things, reduce queues in the unity of AACD in Recife-PE, to provide patients and families from more distant places and travel less stressful expensive, and help locate patients who are still without adequate treatment due to lack of access to treatment centers more distant and / or lack of correct diagnosis of the disease.

**Keywords:** Physical disability, genetic disease, Spatial Analysis, Moran Index, Randomization test.

## 1. Introdução

O Brasil vem se organizando em termos de dados estatísticos oficiais sobre as pessoas com deficiência. Após a Lei nº 7.853/89, que tornou obrigatória a inclusão de itens específicos nos censos nacionais, o Censo Demográfico de 1991, pela primeira vez, incluiu questões que atestaram a presença de 2.198.988 pessoas com deficiência, em uma população de 146.815.750 habitantes, o que representa 1,49 % desta. Já o Censo de 2000, utilizando nova abordagem conceitual e metodológica (CIF/percepção de funcionalidade), identificou 24,6 milhões de pessoas com algum tipo de deficiência e deste total, 4% eram deficientes físicos. Ainda segundo o Censo 2000, o Estado de Pernambuco possui a quarta maior população de deficientes (BRASIL, 2008).

As causas da deficiência física são diversas, e podem estar ligadas a problemas genéticos, complicações na gestação ou gravidez, doenças infantis, traumatismos (acidentes ou agressões), dentre outras. Entre as etiologias das deficiências físicas estão os fatores genéticos, fatores virais ou bacterianos, fatores neonatal e fatores traumáticos. As causas podem ser ainda pré-natais, peri-natais ou pós-natais dependendo do momento em que ocorre a alteração ou má-formação (BRASIL, 2006). Dentre as doenças causadas por fatores genéticos estão as de Depósito Lisossômico (DDLs) e as Neuromusculares (DNM).

As DDLs são um grupo de cerca de 40 raros distúrbios metabólicos hereditários que são decorrentes de mutações genéticas que causam a deficiência ou a falta total de alguma enzima no organismo. As mucopolissacaridoses (MPS) representam um desses distúrbios e caracterizam-se pelo funcionamento inadequado de determinadas enzimas, envolvidas na degradação de ácidos glicosaminoglicanos (GAG). As MPS podem ser classificadas em sete tipos diferentes, de acordo com a enzima que está em falta no portador da doença (VIEIRA, 2007).

As DNM referem-se a um conjunto de mais de 40 patologias diferentes e engloba as doenças que acometem a fibra muscular (Miopatias), doenças dos nervos (Neuropatias), doenças dos cornos anteriores da medula (Atrofias Espinais) e as perturbações da junção neuromuscular (Miastenias) (OLIVEIRA & GABBAI, 2001; REED, 2002).



O avanço da Medicina no conhecimento da etiologia das doenças, dos tipos e manifestações das deficiências, também levou a uma classificação das mesmas, de forma cada vez mais específica. A Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID-10), foi conceituada para padronizar e catalogar as doenças e problemas relacionados à saúde, tendo como referência a Nomenclatura Internacional de Doenças, estabelecida pela Organização Mundial de Saúde (DI NUBILA, 2008).

Quando o indivíduo apresenta certos tipos de deficiências físicas eles necessitam de cuidados especiais, e a Associação de Apoio à Criança Deficiente (AACD) é uma das principais redes filantrópicas de assistência à criança deficiente no país. A instituição é considerada centro de referência no tratamento de pessoas com deficiência física e mantém um amplo serviço de assistência médica, pedagógica e social. Os pacientes recebem assistência de médicos de várias especialidades, fisioterapeutas, terapeutas-ocupacionais, fonoaudiólogos, psicólogos, pedagogos, dentre outros.

A AACD mantém treze unidades espalhadas pelo Brasil: Ibirapuera (SP), Mooca (SP), Osasco (SP), Santana (SP), Campo Grande (SP), São José do Rio Preto (SP), Mogi das Cruzes (SP), Recife (PE), Uberlândia (MG), Poços de Caldas (MG), Porto Alegre (RS), Nova Iguaçu (RJ) e Joinville (SC). Em Pernambuco, a AACD tem capacidade para realizar 650 atendimentos diários e atende a patologias como paralisia infantil, seqüela de acidente vascular cerebral, traumatismo craniano encefálico, amputações, doenças neuromusculares, malformações congênitas, mielomeningocele, lesão medular (paraplegia/tetraplegia), paralisia cerebral e síndromes genéticas que comprometem o aparelho locomotor.

Apesar do imenso esforço para prover atendimento e estrutura adequados, a entidade ainda tem filas de espera chegando a 32 mil crianças e jovens aguardando por consultas, cirurgias, órteses, próteses e cadeiras de rodas nas treze unidades da instituição no país. Na unidade de Pernambuco, uma proposta que vem sendo discutida para diminuir as filas é a descentralização no atendimento e nos serviços prestados, que pode ser feita através da criação de postos avançados de atendimentos em localidades com grande demanda de pacientes que necessitem se deslocar por longas horas para fazer tratamentos e/ou consultas na sede da AACD em Recife.

A demanda por serviços públicos é continuamente distribuída no espaço, de acordo com a distribuição da população, mas estes serviços são providos somente em

localizações específicas. Inevitavelmente, existirão desigualdades no acesso aos serviços, dos custos de transporte, tempo de viagem, entre outros. Dessa forma, desenvolver análises com base na distribuição espacial dos eventos contribui para uma melhor gestão do Sistema de Saúde e nesse contexto, a análise espacial é essencial para a identificação de áreas onde a saúde é precária, necessitando de atenção diferenciada (SANTOS et al., 2001).

Existem poucas referências na literatura utilizando o instrumental da análise espacial no contexto da deficiência física. Por outro lado, a Análise Exploratória de Dados Espaciais (AEDE) é empregada em diferentes estudos nas áreas da saúde, como por exemplo, em estudos que abordam a mortalidade infantil por doenças infecciosas, nas internações por doenças cardíacas, estimação de taxas de mortalidade, na incidência de tuberculose, em trabalhos que tratam da saúde pública ou dos atendimentos nos serviços de saúde (CARVALHO & SANTOS, 2005; RODRIGUES, 2007; SOARES & NASCIMENTO, 2010; SILVA, et al., 2011; CÁRDENAS, 2010; SALES, et al., 2010; RAMOS et al., 2011).

A AEDE compreende um conjunto de técnicas estatisticamente robustas que podem ser utilizadas para sumarizar propriedades dos dados espaciais, detectar padrões, detectar valores extremos ou *outliers*, e identificar subgrupos geográficos específicos (OLIVEAU & GUILMOTO, 2005). Através dela é possível mostrar como os valores das áreas estão arranjos no espaço, e estimar a magnitude dessa dependência espacial.

A dependência espacial é uma característica inerente à representação de dados através de subdivisões territoriais e é identificada quando as observações mais próximas tem maior influência de seus vizinhos em relação às observações mais distantes (PAIVA, 2007). A hipótese nula testada pode ser definida como: não existe, entre as cidades pernambucanas, uma autocorrelação positiva no que diz respeito à concentração de portadores de doença genética com deficiência física.

A dependência espacial pode ser medida através da autocorrelação espacial que, por sua vez, pode ser estimada por meio de diversos indicadores. A importância dos problemas causados pelos chamados efeitos espaciais vem sendo reconhecida desde 1914 nas obras de Student. Entretanto, os primeiros índices formais para identificar a presença de autocorrelação espacial nos dados só foram construídos após 30 anos (MORAN, 1948; GEARY, 1954; KREMPI, 2004).

O primeiro texto dedicado ao tema da autocorrelação espacial foi escrito em 1973 por Cliff e Ord. Uma edição mais ampliada sobre este assunto foi feita pelos mesmos autores em 1981 e inclui um extenso conjunto de aplicações. Estes dois livros são uma das referências mais importantes em autocorrelação espacial, embora vários outros incluam discussões a respeito do tema, dentre estes se destacam: Cliff et al., 1975; Haggett et al., 1977; Getis & Boots, 1978; Ripley, 1981 e Upton & Fingleton, 1985.

A presença de autocorrelação espacial é medida usualmente por meio de estatísticas globais. Dentre estes índices, destaca-se o índice  $I$  global de Moran, que é a ferramenta mais utilizada para medir essa dependência espacial. Esta estatística foi introduzida por Moran para estudar fenômenos estocásticos, que são distribuídos no espaço com duas ou mais dimensões (MORAN, 1948). O índice  $I$  global de Moran é utilizado para estimar a força da correlação entre as observações como uma função da distância que as separa (OLIVEAU & GUILMOTO, 2004).

No cálculo do Índice  $I$  global de Moran é necessário definir uma matriz de pesos espaciais. Esta matriz é de particular importância, pois representa a ligação entre as unidades espaciais. Na literatura, as diferentes estratégias para determinar os pesos são bastante discutidas (ANSELIN, 1992; 2002; YING, 2003; YU & WEI, 2003). A composição da matriz de pesos,  $W$ , é bastante flexível, constando de critérios como contiguidade,  $k$  vizinhos mais próximos, pesos generalizados, tais como o recíproco da distância, o recíproco da distância ao quadrado.

Os indicadores globais de autocorrelação espacial, como o Índice  $I$  global de Moran, fornecem um único valor como medida da associação espacial para todo o conjunto de dados. Entretanto, quando se dispõe de um grande número de áreas, é muito provável a existência de diferentes regimes de correlação espacial em diferentes sub-regiões (CÂMARA et al., 2002). Para evidenciar estes regimes espaciais, podem-se utilizar os Indicadores Locais de Autocorrelação Espacial (LISA) e o diagrama de espalhamento de Moran.

Exemplos de aplicação de estatísticas locais podem ser vistos na geográfica econômica (BAO & HENRY, 1996) para avaliar a adequação das definições de áreas econômicas funcionais, na avaliação de medidas empregadas para definir a acessibilidade aos parques públicos infantis (TALEN & ANSELIN, 1998), dentre outros.

Gettis & Ord (1992) e Anselin (1993b), abordam os padrões de associação de uma forma mais concreta, por exemplo, através da distância com base na estatística formulada por Gettis e Ord e o Moran scatterplot (ANSELIN, 1993a).

Os indicadores locais permitem a decomposição dos indicadores globais em contribuições individuais, sendo possível identificar situações atípicas, descobrir padrões de associação espacial, encontrar agrupamento de valores semelhantes (*clusters*) e sugerir regimes espaciais ou outras formas de heterogeneidade espacial (ANSELIN, 1995; ANSELIN & BAO, 1997).

Outra ferramenta para exploração visual da autocorrelação espacial bastante utilizada é o diagrama de espalhamento ou gráfico de dispersão de Moran. O gráfico tem como base a interpretação do índice de Moran e mostra os valores observados *versus* a média dos valores dos seus vizinhos (ANSELIN, 1996).

No entanto, apesar de muito utilizado, o índice de *I* global Moran pode apresentar alguns problemas quando a amostra é muito pequena ou não é aleatória. Uma alternativa é utilizar os testes de aleatorização, que não apresenta restrição para estes pressupostos.

Com o avanço da informática, propiciando a realização fácil e rápida de métodos computacionalmente intensivos, os testes de aleatorização passaram a ter uma utilização prática, evidenciada na análise de dados experimentais (BRANCO, 2010), séries temporais (PEREIRA, et al., 2010), genética (LOUZADA, et al., 2006), dentre outras. Os testes de aleatorização são testes estatísticos cuja validade tem como base o modelo de distribuição aleatória. Este teste gera uma probabilidade, com base em uma hipótese nula, para testar a significância de uma estatística em um conjunto de dados.

Neste artigo, fez-se um estudo da distribuição espacial dos pacientes portadores de doença genética com deficiência física que estão em tratamento na AACD de Pernambuco. As classes de doenças genéticas categorizadas de acordo com a Classificação Estatística Internacional de Doenças (CID-10) utilizadas neste estudo foram E76, G12, G60 e G71. O Índice *I* global de Moran foi utilizado para medir a autocorrelação espacial da distribuição dos pacientes e o seu resultado é confirmado com o resultado obtido através do teste de aleatorização.

O restante deste artigo está organizado como se segue. A Seção 2.2.1 descreve os dados utilizados, enquanto que a Seção 2.2.2 traz a revisão sobre o Índice de associação espacial de Moran e o teste de aleatorização para detecção de padrão espacial. Os resultados obtidos são apresentados e discutidos na Seção 3 e as considerações finais estão descritas na Seção 4.

## 2. Desenvolvimento

### 2.1. Área de Estudo e Material Utilizado

O estado de Pernambuco está localizado no centro-leste da região Nordeste do Brasil e tem como limites os estados da Paraíba, Ceará, Alagoas, Bahia e Piauí. Segundo o IBGE (2012), em 2010 viviam na área urbana 7.049.868 pessoas e na área rural 1.746.164, totalizando 8.796.032 habitantes, o que corresponde a aproximadamente 4,6% da população brasileira. Sua área territorial é de 98.146,315 km<sup>2</sup>, que engloba 185 municípios agrupados em 5 mesorregiões: São Francisco Pernambucano, Sertão Pernambucano, Agreste Pernambucano, Zona da Mata Pernambucana e Zona Metropolitana do Recife, além de 18 microrregiões. As mesorregiões se referem a uma subdivisão dos estados brasileiros, estabelecida pelo IBGE, que congrega diversos municípios em uma área geográfica com similaridades econômicas e sociais. A Figura 1 apresenta o mapa de Pernambuco e suas mesorregiões.



Figura 1: Estado de Pernambuco de acordo com suas mesorregiões.

Para este estudo foi utilizado o banco de dados com o cadastro de 402 pacientes portadores de doenças genéticas com deficiência física que fazem tratamento na AACD de Pernambuco e que residem no Estado. Foram encontrados pacientes provenientes de 102 municípios pernambucanos. As variáveis utilizadas foram: total de pacientes portadores de doença genética com deficiência física em tratamento na AACD por município (Pac), a taxa bruta de portadores de doença genética com deficiência física (Ipac), a taxa suavizada pelo estimador Empírico de Bayes para os portadores de doença genética com deficiência física (IpacBG), a taxa suavizada pelo estimador espacial empírico de Bayes para os portadores de doença genética com deficiência física (IpacBL), classes (E76; G12; G60; G71) e subclasses (E76.0;

E76.2; G12.0; G12.1; G60.0; G60.8; G71.0; G71.1 G71.2; G71.8 G71.9) de doenças genéticas segundo classificação CID-10, a população dos 102 municípios pernambucanos, as coordenadas geográficas (latitude e longitude) do centro do polígono formado pelos municípios pernambucanos. A variável  $Ipac$  foi calculada como segue:

$$Ipac_i = \frac{x_i}{p_i} 100.000, \quad i = 1, \dots, 102 \quad (1)$$

Sendo  $x_i$  o número de casos observados no  $i$ -ésimo município e,  $p_i$  o tamanho total da população no  $i$ -ésimo município. Por intermédio desse cálculo, obtém-se o número observado de eventos no município  $i$  ( $i = 1, \dots, n$ ) em uma população base de 100 mil pessoas. Esse padrão é utilizado, por exemplo, para a apresentação de taxas de eventos raros.

A Tabela 1 apresenta os municípios, sua população e área de acordo com os dados do Censo IBGE 2010.

Tabela 1. População e área dos 102 municípios pernambucanos em estudo.

<b>Município</b>	<b>População</b>	<b>Área (km<sup>2</sup>)</b>
<b>Abreu e Lima</b>	94.429	130,3
<b>Afogados da Ingazeira</b>	35.088	377,7
<b>Água Preta</b>	33.095	533,3
<b>Águas Belas</b>	40.235	886,0
<b>Aliança</b>	37.415	272,8
<b>Amaraji</b>	21.939	235,0
<b>Angelim</b>	10.202	118,0
<b>Araripina</b>	77.302	1.892,6
<b>Arcoverde</b>	68.793	350,9
<b>Barreiros</b>	40.732	233,4
<b>Belém de Maria</b>	11.353	73,7
<b>Belo Jardim</b>	72.432	647,7
<b>Bezerros</b>	58.668	490,8
<b>Bodocó</b>	35.158	1.616,5

Tabela 1. População e área dos 102 municípios pernambucanos em estudo.

(continuação)

<b>Município</b>	<b>População</b>	<b>Área (km<sup>2</sup>)</b>
<b>Bom Jardim</b>	37.826	218,4
<b>Bonito</b>	37.566	395,6
<b>Brejão</b>	8.844	159,8
<b>Brejo da Madre de Deus</b>	45.180	762,4
<b>Cabo de Santo Agostinho</b>	185.025	446,6
<b>Cachoeirinha</b>	18.819	179,3
<b>Calçado</b>	11.125	121,9
<b>Camaragibe</b>	144.466	51,2
<b>Camocim de São Félix</b>	17.104	72,5
<b>Camutanga</b>	8.156	37,5
<b>Carpina</b>	74.858	144,9
<b>Caruaru</b>	314.912	920,6
<b>Casinhas</b>	13.766	115,9
<b>Cedro</b>	10.778	171,6
<b>Chã de Alegria</b>	12.404	48,5
<b>Chã Grande</b>	20.137	84,8
<b>Condado</b>	24.282	89,6
<b>Cumarú</b>	17.183	292,2
<b>Cupira</b>	23.390	105,6
<b>Custódia</b>	33.855	1.404,1
<b>Dormentes</b>	16.917	1.537,6
<b>Feira Nova</b>	20.571	107,7
<b>Frei Miguelinho</b>	14.293	212,7
<b>Gameleira</b>	27.912	256,0
<b>Garanhuns</b>	129.408	458,6
<b>Glória do Goitá</b>	29.019	231,8
<b>Goiana</b>	75.644	501,9
<b>Gravatá</b>	76.458	505,1
<b>Iati</b>	18.360	635,1
<b>Ibimirim</b>	26.954	1.954,7
<b>Igarassu</b>	102.021	305,6
<b>Ipojuca</b>	80.637	532,6
<b>Itaíba</b>	26.256	1.083,7
<b>Itambé</b>	35.398	304,8
<b>Jaboatão dos Guararapes</b>	644.620	258,6
<b>Jataúba</b>	15.819	672,2
<b>Joaquim Nabuco</b>	15.773	121,9
<b>Lagoa do Itaenga</b>	20.659	57,3
<b>Lagoa do Ouro</b>	12.132	198,8
<b>Lagoa dos Gatos</b>	15.615	222,9
<b>Lagoa Grande</b>	22.760	1.852,3



Tabela 1. População e área dos 102 municípios pernambucanos em estudo.

(continuação)

<b>Município</b>	<b>População</b>	<b>Área (km<sup>2</sup>)</b>
<b>Lajedo</b>	36.628	189,1
<b>Limoeiro</b>	55.439	273,7
<b>Macaparana</b>	23.925	108,0
<b>Manari</b>	18.083	381,3
<b>Maraial</b>	12.230	199,9
<b>Moreno</b>	56.696	196,1
<b>Nazaré da Mata</b>	30.796	150,3
<b>Olinda</b>	377.779	41,7
<b>Orobó</b>	22.878	138,7
<b>Palmares</b>	59.526	339,3
<b>Panelas</b>	25.645	370,9
<b>Parnamirim</b>	20.224	2.595,9
<b>Passira</b>	28.628	326,8
<b>Paudalho</b>	51.357	277,5
<b>Paulista</b>	300.466	97,4
<b>Pesqueira</b>	62.931	995,5
<b>Petrolândia</b>	32.492	1.056,6
<b>Petrolina</b>	293.962	4.558,4
<b>Pombos</b>	24.046	204,1
<b>Quixaba</b>	6.739	210,7
<b>Recife</b>	1.537.704	218,5
<b>Ribeirão</b>	44.439	287,9
<b>Sairé</b>	11.240	191,0
<b>Salgueiro</b>	56.629	1.686,8
<b>Saloá</b>	15.309	252,1
<b>Sanharó</b>	21.955	268,7
<b>Santa Cruz do Capibaribe</b>	87.582	335,3
<b>Santa Maria da Boa Vista</b>	39.435	3.001,2
<b>Santa Maria do Cambucá</b>	13.021	92,1
<b>São Benedito do Sul</b>	13.941	160,5
<b>São Bento do Una</b>	53.242	719,2
<b>São João</b>	21.312	258,3
<b>São José da Coroa Grande</b>	18.180	69,3
<b>São José do Egito</b>	31.829	798,9
<b>São Lourenço da Mata</b>	102.895	262,2
<b>São Vicente Ferrer</b>	17.000	114,0
<b>Serra Talhada</b>	79.232	2.980,0
<b>Sirinhaém</b>	40.296	369,1
<b>Surubim</b>	58.515	252,9
<b>Tacaimbó</b>	12.725	227,6
<b>Timbaúba</b>	53.825	292,3

Tabela 1. População e área dos 102 municípios pernambucanos em estudo.

(continuação)

Município	População	Área (km <sup>2</sup> )
<b>Toritama</b>	35.554	25,7
<b>Trindade</b>	26.116	229,5
<b>Venturosa</b>	16.052	320,7
<b>Vertente do Lério</b>	7.873	73,6
<b>Vicência</b>	30.732	228,0
<b>Vitória de Santo Antão</b>	129.974	371,8

Fonte: IBGE, Censo Demográfico 2010.

A Tabela 2 apresenta os códigos das classes e subclasses de doenças genéticas que causam deficiência física abordadas nesse estudo, segundo a classificação CID-10.

Tabela 2: Classes e subclasses das doenças genéticas em estudo de acordo com o CID-10.

Classes	Subclasses	Descrição
E76		<b>Distúrbios do metabolismo do glicosaminoglicano</b>
	E76.0	Mucopolissacaridose do tipo I
	E76.2	Outras mucopolissacaridoses
G12		<b>Atrofia muscular espinal e síndromes correlatas</b>
	G12.0	Atrofia muscular espinal infantil tipo I (Wernidnig-Hoffman)
	G12.1	Outras atrofia musculares espinais hereditárias
G60		<b>Neuropatia hereditária e idiopática</b>
	G60.0	Neuropatia hereditária motora e sensorial
	G60.8	Outras neuropatias hereditárias e idiopáticas
G71		<b>Transtornos primários dos músculos</b>
	G71.0	Distrofia muscular
	G71.1	Transtornos miotônicos
	G71.2	Miopatias congênitas
	G71.8	Outros transtornos primários dos músculos
	G71.9	Transtorno muscular primário não especificado

Fonte: CID – 10; Ministério da Saúde, 2008.

### 2.1.1. Suavização

É importante ressaltar que, a unidade de análise deste trabalho é o município, o que na grande maioria dos casos configura área com baixo contingente populacional, pelo menos em se tratando de população exposta ao risco de ter alguma doença

genética que causa deficiência física. A taxa de incidência é o estimador mais simples e utilizado para o mapeamento da ocorrência de doenças e agravos. Um grande problema associado ao uso desta taxa é a alta instabilidade que ela possui ao expressar o risco de determinados eventos (BRASIL, 2007). Conforme Assunção (2001), o acréscimo ou decréscimo de um caso em áreas com efetivo populacional diminuto pode causar grande oscilação na taxa mensurada.

De modo a superar essas dificuldades, alguns autores sugerem reestimar uma taxa mais próxima do risco real ao qual a população está exposta (MARSHALL, 1991; ASSUNÇÃO, 1996; ASSUNÇÃO et al., 1998). Neste trabalho foi utilizado o estimador Bayesiano Empírico (EB), que é operacionalizado da seguinte forma (MARSHALL, 1991):

$$\theta_i = m + C_i(x_i - m) \quad (2)$$

em que  $C_i$  é dado por:

$$C_i = \frac{(S^2 - \frac{m}{n^M})}{(S^2 - \frac{m}{n^M} + \frac{m}{n_i})} \quad (3)$$

em que:  $\theta_i$  é a taxa suavizada;  $m$  é a taxa média global ou a taxa média dos vizinhos mais próximos;  $x_i$  é a taxa da  $i$ -ésima área;  $S^2$  é a variância da taxa a ser estimada;  $n^M$ , a população média global ou a média dos vizinhos;  $n_i$  é a população da área  $i$ . O número de vizinhos a serem considerados no cálculo desta taxa deve ser previamente definido pelo pesquisador.

O Estimador Bayesiano Global (EBG) é obtido ao utilizar a taxa média global, enquanto que no cálculo do Estimador Bayesiano Local (EBL) o valor de  $m$  é obtido através da taxa média das áreas vizinhas de tal forma que o risco subjacente seja convergente para a média local da vizinhança.

Em (3), deve-se levar em consideração que o multiplicador  $C_i$  será próximo de 1, caso a população da área  $i$  ( $n_i$ ) mostre valor elevado. Nesse caso, a taxa suavizada ( $\theta_i$ ) tenderá a ter o mesmo valor da estimada ( $x_i$ ) sem a aplicação do procedimento. Caso contrário, se a população da área  $i$  possuir efetivo populacional muito

reduzido tem-se que  $C_i$  será próximo de zero, implicando em que a taxa suavizada ( $\theta_i$ ) tenderá a ser próxima da taxa média de todas as áreas (BARBOSA E FREIRE, 2004).

Segundo Baptista (2009), este estimador elimina, parcialmente, a variabilidade presente nas taxas estimadas que não está associada a fatores de risco. Além disso, mapas com base nestas estimativas são mais interpretativos e informativos, pois permitem a visualização das tendências mais gerais. No presente trabalho, a variável  $IpacBG$  se refere a taxa suavizada pelo EBG e a variável  $IpacBL$  diz respeito a taxa suavizada pelo EBL.

## **2.2. Métodos**

### **2.2.1. Análise Espacial de Área**

Em várias situações práticas o pesquisador não dispõe da localização geográfica dos eventos e/ou estas não podem ser divulgadas pelo fato de identificar o indivíduo. Nestes casos, os dados estão disponíveis de maneira aglomerada por uma certa unidade de área, como um bairro, distrito, setor censitário, município, dentre outros (CÂMARA et al., 2002). Para essas análises, procura-se identificar padrões espaciais de distribuição nos valores observados por área.

No estudo de dados de áreas, as técnicas de análise exploratória são importantes instrumentos que auxiliam no desenvolvimento da modelagem estatística espacial. Tais técnicas contribuem para a visualização e extração de informações não perceptíveis de maneira imediata pelo analista, quando utilizado procedimento comum de classificação e visualização de dados espaciais (DRUCK et al., 2004).

A forma usual de apresentação de dados agregados por áreas é por meio de mapas coloridos com o padrão espacial do fenômeno. Se houver padrão espacial, espera-se encontrar cores próximas geograficamente. Esses mapas são muito úteis para uma análise exploratória. Contudo, ao observar um mapa, podem surgir vários questionamentos, tais como, saber se sua configuração apresenta um padrão aleatório, se existe um padrão espacial bem definido, por exemplo, mostrando agrupamento de áreas (CÂMARA et al., 2002).

Um aspecto fundamental da análise exploratória é a caracterização do padrão espacial. O padrão espacial em um conjunto de dados refere-se ao fato de que o valor

da variável de interesse no local  $i$  depende do valor dessa mesma variável em uma localização  $j$ , com  $i \neq j$ , e  $i$  e  $j$  são locais próximos geograficamente. Formalmente, isso pode ser escrito como (LESAGE, 1998).

$$y_i = f(y_j), i = 1, \dots, n. \forall i \neq j \quad (4)$$

Segundo Anselin (1992), a forma como os dados estão dispostos no espaço está diretamente relacionada com a interação entre as unidades de observação, a qual é revelada nas medidas de autocorrelação ou associação espacial.

### 2.2.2. Autocorrelação Espacial

Para a análise espacial conceitos como dependência espacial e autocorrelação espacial são fundamentais. A autocorrelação espacial é a expressão computacional da dependência espacial e é o mais conhecido e comum dentre os efeitos espaciais. A ideia é verificar como o padrão espacial varia, a partir da comparação entre os valores de uma amostra e de seus vizinhos (FOOK, 2005).

Seu algoritmo resume-se primeiramente em calcular as distâncias entre todos os pontos e depois se criam classes de distância para pares de pontos que têm mais ou menos a mesma distância entre si. Essas classes geralmente tem como base um *lag* (distância entre dois pontos) de espaçamento especificado pelo pesquisador. O último passo é calcular a autocorrelação para todos os pares de pontos dentro de um grupo de distância.

A autocorrelação espacial pode ser estimada por meio de diversos indicadores, os quais permitem estimar o quanto o valor observado de um atributo em uma determinada região é dependente dos valores dessa mesma variável nas localizações vizinhas. Os indicadores de autocorrelação espacial são casos particulares de uma estatística de produtos cruzados segundo a expressão (CÂMARA et al., 2002)

$$\Gamma(d) = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) \xi_{ij} \quad (5)$$

em que, dada uma distância  $d$ ,  $w_{ij}$  é uma matriz que fornece uma medida de contiguidade espacial entre as variáveis aleatórias  $Z_i$  e  $Z_j$ , por exemplo informando se são separadas por uma distância menor que  $d$  e  $\xi_{ij}$  é uma matriz que fornece uma medida de correlação entre as variáveis aleatórias que no caso do Índice de Moran é o produto destas variáveis.

Estes indicadores auxiliam na verificação de como ocorre a dependência espacial, a partir da comparação entre os valores de uma amostra e de seus vizinhos, ou seja, eles dependem da vizinhança adotada (DRUCK et al., 2004). Dentre as ferramentas propostas na literatura para analisar dados de área destacam-se os Índices de Moran global e local. Para tanto, um parâmetro utilizado nesses indicadores consiste na matriz de proximidade espacial.

### 2.2.3. Matriz de proximidade espacial

Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a Matriz de Proximidade Espacial, também chamada Matriz de Vizinhança. Esta ferramenta é um item necessário para se estimar os índices de autocorrelação. Dado um conjunto de  $n$  áreas  $\{A_i, \dots, A_n\}$  constrói-se a matriz  $W^{(1)}$  de dimensão  $n \times n$ , em que cada um dos elementos  $w_{ij}$  representa uma medida de proximidade entre as áreas  $A_i$  e  $A_j$ . Os elementos  $w_{ij}$  da diagonal são zero enquanto que os elementos  $w_{ij}$  indicam a forma como a região  $i$  está espacialmente conectada à região  $j$  (CÂMARA et al., 2002). Estes elementos são não estocásticos, não negativos e finitos. Em geral, a matriz de peso é padronizada de modo que a soma dos elementos de cada linha é igual a 1 (BAUMONT et al., 2002). Alguns critérios usuais para a matriz de proximidade espacial são (KAMPEL et al., 2000; BAILEY & GATRELL, 1995):

- $w_{ij} = 1$ , se o centro de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário  $w_{ij} = 0$ ;
- $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , caso contrário  $w_{ij} = 0$ ;
- $w_{ij} = l_{ij}/l_i$ , em que  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$ .
- $w_{ij} = 1/d$ , em que  $d$  é a distância entre os centros de  $A_i$  e  $A_j$ .

Baumont (2004) propõem que a matriz de vizinhança seja construída por meio do método dos  $k$  vizinhos mais próximos, a qual é definida da seguinte forma:

$$\begin{cases} w_{ij}(k) = 1, \text{ se } d_{ij} < d_i(k) \\ w_{ij}(k) = 0, \text{ caso contrário} \end{cases} \quad (6)$$

em que  $w_{ij}$  é um elemento da matriz normalizada,  $d_{ij}$  é a menor distância entre a região  $i$  e todas as outras unidades de modo que cada unidade tem exatamente  $k$  vizinhos. Baumont (2004) sugere ainda testar a presença de autocorrelação espacial por meio do Índice  $I$  global de Moran para diversas matrizes contendo  $k$  vizinhos mais próximos e escolher a matriz que produza o maior valor do índice.

A escolha da matriz de pesos espaciais é muito importante em uma análise exploratória de dados espaciais, pois todos os passos subsequentes (ou resultados) dependem desta seleção. Os critérios adotados nesse trabalho foram o inverso da distância e os  $k$  vizinhos mais próximos para  $k= 2,3,4,5$  e  $6$ , ou seja, para os dois, três, quatro, cinco e seis vizinhos mais próximos, respectivamente.

#### 2.2.4. Índice $I$ global de Moran

A dependência espacial pode ser medida de diferentes formas. O Índice  $I$  global de Moran é a estatística mais difundida e mede a correlação a partir do produto dos desvios em relação à média (ANSELIN, 1992). Assim como um coeficiente de correlação linear, o Índice  $I$  global de Moran varia de  $-1$  a  $1$ . Valores próximos a zero correspondem a autocorrelação espacial não significativa, valores positivos indicam correlação direta e valores negativos apontam para uma correlação inversa. De acordo com Cressie (1993), o Índice  $I$  global de Moran é definido por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n z_i z_j w_{ij}}{S_0 \sum_{i=1}^n z_i^2} \quad (7)$$

em que:  $n$  é o total de municípios avaliados,  $z_i = (x_i - \bar{X})$  e  $z_j = (x_j - \bar{X})$  são as informações centradas na média da variável  $X$ , para os municípios  $i$  e  $j$ , respectivamente,  $w_{ij}$  é o elemento da matriz quadrada e simétrica  $W(n \times n)$ , que expressa a relação espacial entre as observações e  $S_0 \sum_{i=1}^n z_i^2$  é uma matriz simétrica de pesos espaciais  $W$ . Além disso,  $\sum_{j=1}^n z_j w_{ij}$  é chamada de variável espacialmente defasada (*spatial lag*) da localização.

A hipótese nula ( $H_0$ ) é a de completa aleatoriedade espacial, quando o indicador se distribui ao acaso entre as áreas sem relação com a posição destas (CARVALHO, 1997). Uma vez calculado o índice  $I$  global de Moran, é importante estabelecer sua validade estatística. A significância dos momentos de  $I$  pode ser avaliada sob  $H_0$  por duas suposições básicas: normalidade (N) ou aleatoriedade (R). Na primeira, os dados são observações independentes provenientes de uma população com distribuição normal e na segunda, a função de distribuição é desconhecida (LOUZADA, 2006). Para testar a hipótese nula a partir da suposição de normalidade das observações deve-se calcular o índice padronizado como:

$$I_{padr\tilde{a}o} = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1)$$

Bello (2010) define a média e variância do índice  $I$  global de Moran sob as duas suposições como a seguir:

$$E_N(I) = E_R(I) = -\frac{1}{n-1} \quad (8)$$

$$Var_N(I) = \frac{n^2 S_1 - n^2 S_2 + 3S_0^2}{S_0(n^2 - 1)} \quad (9)$$

$$Var_R(I) = \frac{n[(n^2 - 3n + 3)S_1 - n^2 S_2 - 3S_0^2] - b_2[(n_2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)3S_0^2} \quad (10)$$

em que,

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}; \quad S_1 = \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_{i=1}^n (w_{i.} + w_{.j})^2; \quad w_{i.} = \sum_{j=1}^n w_{ij} \text{ e } w_{.j} = \sum_{i=1}^n w_{ij}$$

Valores positivos para o índice  $I$  global de Moran implicam em autocorrelação positiva, ou seja, indica que municípios com alta taxa de pacientes portadores de doença genética com deficiência física, tendem a ser rodeados por municípios com altas taxas de pacientes portadores de doença genética com deficiência física em tratamento na AACD ou, municípios com baixa taxa de pacientes portadores doença genética com deficiência física, tendem a ser circundados por municípios com baixas taxas de pacientes portadores de doença genética com deficiência física em tratamen-



to na AACD. Autocorrelação espacial negativa indica, por exemplo, que valores diferentes dessas taxas são encontrados simultaneamente no espaço.

### 2.2.5. Diagrama de espalhamento de Moran

O diagrama de espalhamento (ou dispersão) de Moran (*Moran scatterplot*) representa uma maneira alternativa de visualização da dependência espacial existente no conjunto de dados (DRUCK et al., 2004). Através dele, é possível visualizar a correlação linear entre cada valor do atributo  $z$  em relação à média ( $w_z$ ) dos valores dos atributos de seus vizinhos através do gráfico de duas variáveis. De acordo com Anselin, 1996, o coeficiente  $I$  de Moran é dado pela inclinação da reta de regressão da variável defasada espacialmente ( $w_z$ ) localizada na ordenada *versus* a variável em análise ( $z$ ) localizada na abscissa e esta inclinação indicará o grau de ajustamento. A variável defasada espacialmente pode ser obtida através do produto da matriz de pesos ( $W$ ) pelo vetor de variáveis originais ( $z$ ) (TYSZLER, 2006). Em geral, os valores de  $z$  e  $W$  são padronizados o que possibilita identificar valores discrepantes (*outliers*) e pontos de alavancagem (*leverage points*).

A origem do gráfico de espalhamento de Moran está centrada na coordenada (0,0), sendo assim, o gráfico é dividido em quatro quadrantes, de maneira que cada quadrante representa um tipo diferente de associação entre o valor de uma dada região e a média de seus vizinhos. Anselin (1993) descreve-os como:

- Quadrante superior direito (valores positivos, médias positivas ou “alto-alto”): indica associação espacial positiva, isto é, a área para o valor do atributo considerado, está cercada por áreas que tem comportamento similar. Além disso, o quadrante indica que tanto o valor do atributo, quanto o valor médio para seus vizinhos, estão acima da média geral do conjunto;
- Quadrante inferior esquerdo (valores negativos, médias negativas ou “baixo-baixo”): indica associação espacial positiva e que ambos, atributo e média dos vizinhos, estão abaixo da média;
- Quadrante superior esquerdo (valores negativos, médias positivas ou “baixo-alto”): indica associação espacial negativa em que valores baixos estão cercados por valores altos representando valor negativo e média dos vizinhos positiva;

- Quadrante inferior direito (valores positivos, médias negativas ou “alto-baixo”): indica associação espacial negativa e valores altos são rodeados por valores baixos representando valor positivo e média dos vizinhos negativa.

A Figura 2 apresenta um esquema do diagrama de espalhamento de Moran.

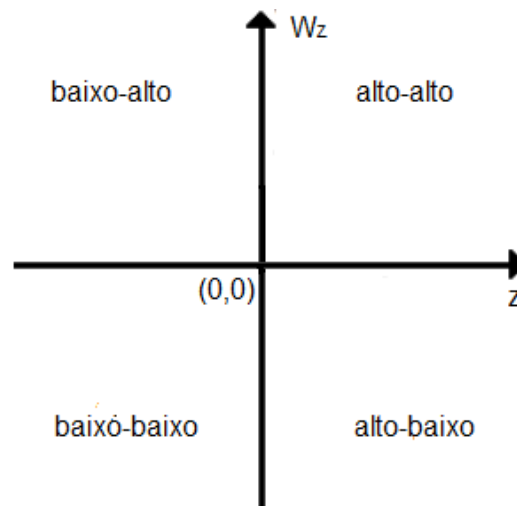


Figura 2: Esquema explicativo do diagrama de espalhamento de Moran.

Sendo assim, diante de uma autocorrelação espacial positiva (por exemplo), os *outliers* espaciais serão aqueles localizados nos quadrantes alto-baixo e baixo-alto, enquanto as observações situadas nas associações alto-alto e baixo-baixo representam pontos de alavancagem. Esses pontos, por definição, são observações que exercem influências exageradas na determinação do grau de associação espacial, embora acompanhem a mesma associação espacial do restante dos dados (ANSELIN, 1996).

### 2.2.6. Índice *I* local de Moran

Quando se dispõe de um grande número de áreas é possível que ocorram diferentes regimes de associação espacial e que apareçam máximos locais de autocorrelação espacial onde a dependência espacial é ainda mais evidenciada. Para se identificar agrupamentos em função de uma vizinhança preestabelecida, utilizam-se os *Local Indicator for Spatial Autocorrelation* – LISA – (Indicador Local de Autocorrelação Espacial), sendo o Índice *I* local de Moran a estatística mais utilizada. Segundo Anselin (1994), o Índice *I* local de Moran é uma ferramenta estatística utilizada para testar a autocorrelação local e detectar objetos espaciais com influência no indicador global de Moran. Anselin (1995) diz que um indicador local de autocorrelação espacial deve atender a dois objetivos específicos:

- i) Permitir a identificação de padrões de associação espacial significativos;
- ii) A soma dos indicadores de LISA para todas as observações deve ser proporcional ao índice global de associação espacial;

Os indicadores locais produzem um valor específico para cada área, permitindo a identificação de valores extremos e a indicação sobre a extensão da significância de um “*cluster*” de iguais valores. Anselin (1995) representa o Índice  $I$  local de Moran por:

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2} \quad (11)$$

em que,  $z_i = (x_i - \bar{X})$  e  $z_j = (x_j - \bar{X})$  são as informações centradas na média da variável  $X$ , para os municípios  $i$  e  $j$ , respectivamente, e  $w_{ij}$  é o elemento da matriz de pesos espaciais, ou seja, é o peso que o município  $i$  exerce sobre o município  $j$ .

A interpretação mais direta do Índice  $I$  local de Moran é aquela em que valores significativamente altos e positivos apontam a presença de um *cluster* tanto de valores iguais altos como baixos, já valores significativamente baixos indicam zonas de transição entre um determinado regime espacial e outro.

Existem alguns dispositivos gráficos de visualização com base nos resultados obtidos para os indicadores locais, dentre eles, o *LISA map*.

### 2.2.7. *LISA map*

Conforme Anselin (1995), na confecção do *LISA map*, a avaliação da significância é feita comparando os resultados obtidos para os indicadores locais, com uma série de valores obtidos por meio de permutações (número de permutações ou aleatorizações a ser definido pelo pesquisador) dos valores dos atributos dos vizinhos, em relação à hipótese nula de não existência de padrão espacial. Os objetos são classificados em quatro grupos: não significantes; com significância entre 0,05 e 0,01; com significância entre 0,01 e 0,001; e maior que 0,001. Uma vez determinada a significância estatística do índice  $I$  local de Moran, é muito útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do restante dos

dados. Estas regiões podem ser vistas como “bolsões” de não estacionariedade, pois são áreas com dinâmica espacial própria e que merecem análise detalhada.

### 2.2.8. Teste de Aleatorização

Um teste de aleatorização fornece evidência de que certo padrão nos dados pode, ou não, ter aparecido por acaso, ou seja, sob a hipótese nula, todas as possíveis ordens para os dados têm a mesma chance de ocorrer (MANLY, 2006, citado por VIOLA, 2007).

Segundo Manly (2006), em comparação com métodos estatísticos padrão, os testes de aleatorização têm duas vantagens principais: são válidos para pequenas amostras, aleatórias ou não e, em geral, são fáceis de calcular. Além disso, tem como base estatísticas não padronizadas e não é preciso saber previamente a distribuição da população da qual a amostra foi extraída.

A principal desvantagem desse teste é que as conclusões obtidas ficam restritas a cada conjunto de dados e tipo de problema não sendo possível generalizar para a população (VIOLA, 2007).

Segundo Viola (2007), o teste de aleatorização segue os seguintes passos:

- i) Calcula-se o valor  $e_o$  de uma estatística “ $E$ ” de um conjunto de observações, e nesse trabalho a estatística “ $E$ ” considerada é o Índice  $I$  global de Moran;
- ii) Aleatorizam-se os dados um grande número de vezes. No contexto de dados espaciais as aleatorizações são obtidas por reordenação dos dados. Para cada aleatorização obtêm-se estatísticas ( $e_o$ ) que irão gerar uma aproximação por simulação da distribuição amostral de  $E$ .
- iii) Calcula-se o p-valor que é obtido pela proporção de vezes que a estatística obtida na  $i$ -ésima aleatorização ( $e_{ai}$ ) foi maior que a observada ( $e_o$ ).
- iv) Rejeita-se a hipótese nula quando o p-valor for menor que o nível de significância ( $\alpha$ ) adotado.

Para as análises estatísticas e elaboração de mapas espaciais, foram utilizados o *software* R versão 2.15 e o *software* TerraView versão 4.2.1 desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE).

### 3. Resultados e Discussão

Foram analisados dados referentes a 402 pacientes portadores de doença genética com deficiência física atendidos pela AACD de Pernambuco e residentes no Estado. A partir dos dados coletados foi possível observar uma predominância dos pacientes portadores de distrofia muscular (G71.0) e miopatias congênitas (G71.2), correspondendo a 37,06% e 25,37% do total, respectivamente. Os pacientes portadores de outras neuropatias hereditárias e idiopáticas (G 60.8) correspondem a 15,17%, seguido de 13,4% de portadores de outras atrofia muscular espinhais hereditárias (G12.1). Já os pacientes com MPS (E76.0 e E76.2) somam 5,2% e esses dados são importantes, pois segundo a Aliança Brasileira de Mucopolissacaridose (ABRAMPS), Pernambuco é o segundo estado em número de pacientes com MPS e essa estimativa pode ser maior ainda já que a doença permanece desconhecida por vários médicos e por isso é de difícil diagnóstico.

Ainda de acordo com os dados, a cidade do Recife acomoda o maior número de pacientes atendidos, totalizando 112, o que corresponde a 27,9% do total. Desses, 33,04% são portadores de distrofia muscular, 27,68% possuem miopatia congênita e 19,64% são portadores de outras neuropatias hereditárias idiopáticas. As cidades de Jaboatão dos Guararapes, Olinda e Paulista, juntas, somam cerca de 16,4% do total de pacientes em tratamento na AACD de Pernambuco com 7,2%, 4,2% e 5%, respectivamente. Em todos os três municípios observou-se que a distrofia muscular acomete a maior parte dos pacientes, correspondendo a 51,72% em Jaboatão dos Guararapes, 47,08% em Olinda e 41,18% em Paulista.

### Mapas

Este estudo utilizou o nível de município como ponto de partida para se conhecer e analisar a distribuição espacial dos pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco. Entretanto as mesorregiões foram utilizadas como auxílio às análises.

Os Mapas coropléticos foram confeccionados no *software* livre TerraView versão 4.2.1. A relevância do uso de mapas para trabalhos como este, vai além de uma questão estética, pois estes mapas conseguem revelar aspectos importantes que talvez fossem mais difíceis ou até mesmo impossíveis de se visualizar utilizando apenas os dados. Nos gráficos a seguir, a legenda zero indica locais onde não existem pacientes em tratamento na AACD de Pernambuco.

Inicialmente foi construído um mapa utilizando o total de pacientes portadores de doença genética com deficiência física em cada município. Conforme se pode observar na Figura 3, as áreas com tonalidades mais escuras são as que possuem o maior número de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco. Estas regiões se localizam na Região Metropolitana do Recife, mais precisamente nas cidades do Recife, Jaboatão dos Guararapes, Olinda e Paulista. Isso já era de se esperar, visto que essas são as quatro maiores cidades do Estado (pelo menos no que tange o aspecto populacional) e dispõem de uma infraestrutura melhor, por exemplo, de transporte, que possibilita ao paciente se deslocar com maior facilidade e em menor tempo até o centro de tratamento da AACD.

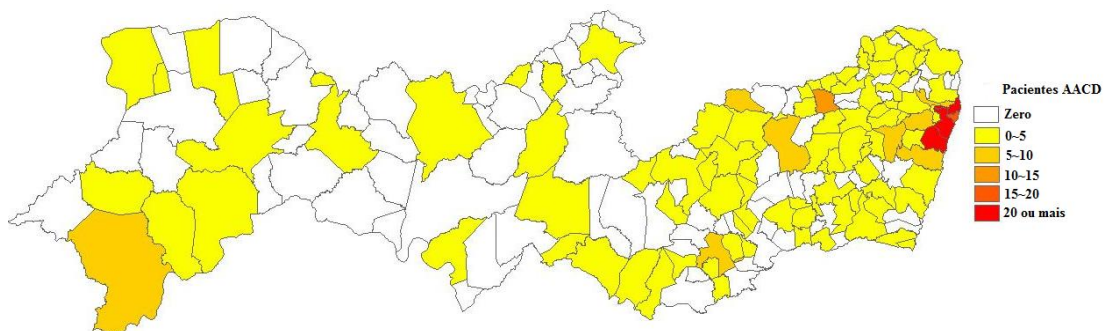


Figura 3: Total de pacientes portadores de doença genética com deficiência física em tratamento na AACD por município – Pernambuco.

Em seguida, foi construído um mapa utilizando a taxa bruta de pacientes por 100.000 habitantes (Ipac), conforme Figura 4. Em uma primeira leitura, este mapa impressiona pelas altas taxas de pacientes portadores de doença genética com deficiência física, com 15 municípios apresentando uma taxa maior que 10 pacientes por cem mil habitantes, e um caso com taxa acima de 50 por cem mil habitantes. Apenas visualizando este mapa, um analista desatento poderia concluir que estes municípios

pernambucanos apresentam um grave problema de saúde pública. Na realidade, muitos desses valores extremos ocorrem nos municípios com pequenas populações, como é o caso de Vertente do Lério que com população de 7,873 habitantes e um registro de quatro pacientes portadores de doença genética com deficiência física em tratamento na AACD apresentou uma taxa de aproximadamente 50,807/100.000 habitantes. Destacam-se ainda municípios como Camutanga com população de 8,156 habitantes, 2 pacientes em tratamento na AACD e uma taxa de 24,523/100.000 habitantes e Surubim com população de 58,515 habitantes, 12 pacientes em tratamento na AACD e uma taxa de 20,508/100.000 habitantes.

Como pode ser visto na Figura 4, quando a população de pacientes portadores de doença genética com deficiência física dos municípios é dividida pela população total dos mesmos, fornecendo a taxa bruta de pacientes por município (Ipac), os locais com os valores mais elevados não são as quatro cidades mais populosas. Há uma concentração mais significativa que abrange em sua maioria os municípios das mesorregiões da Mata Pernambucana e Agreste pernambucano.

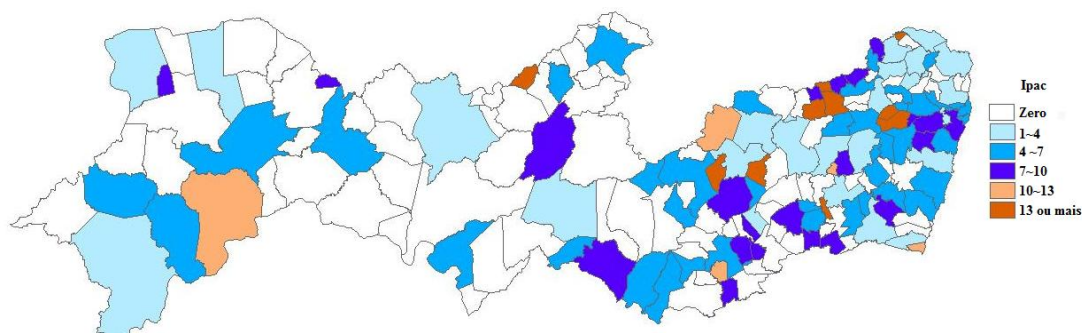


Figura 4: Taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (Ipac).

Geralmente, os valores extremos são os que mais chamam a atenção em um mapa temático de taxas, e muitas vezes resultam de um número muito reduzido de observações sendo, portanto, menos confiáveis, ou seja, apenas flutuação aleatória. Pode-se contornar este problema utilizando-se os estimadores Bayesianos empíricos globais e locais.

A Figura 5 apresenta o mapa com a distribuição espacial da taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco sua-



vizada pelo método empírico de Bayes global (IpacBG). Observou-se que com a utilização do método de suavização os valores das taxas foram bastante reduzidos sendo que a maioria dos municípios tem taxa que varia de 0,0001 a 1,5. Além disso, o município de Vertente do Lério passou de uma taxa bruta de 50,807 por 100000 habitantes para uma IpacBG de aproximadamente 5,703, entretanto ainda é o município que detém a maior taxa de pacientes portadores de doença genética com deficiência física em tratamento no AACD. A Figura 6 representa a distribuição espacial da taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco suavizada pelo método empírico de Bayes local (IpacBL) e através desta figura notou-se que as taxas permaneceram praticamente inalteradas em relação a suavização empírica de Bayes global (IpacBG).

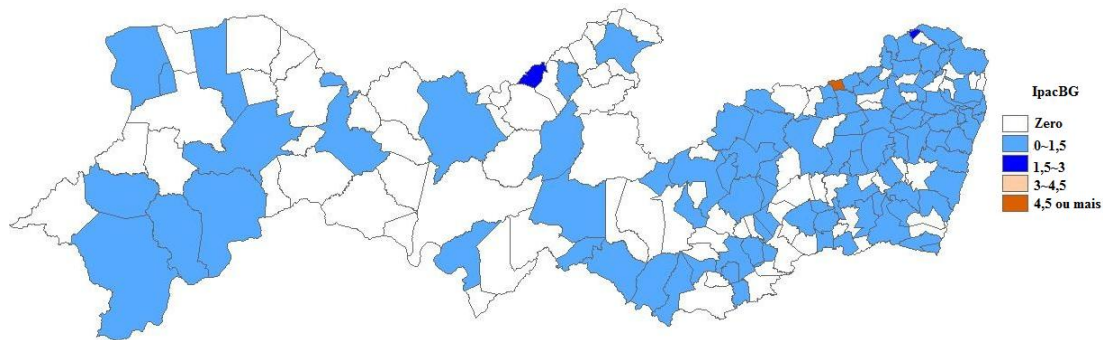


Figura 5: Taxa suavizada pelo Estimador Bayesiano Empírico Global para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBG).

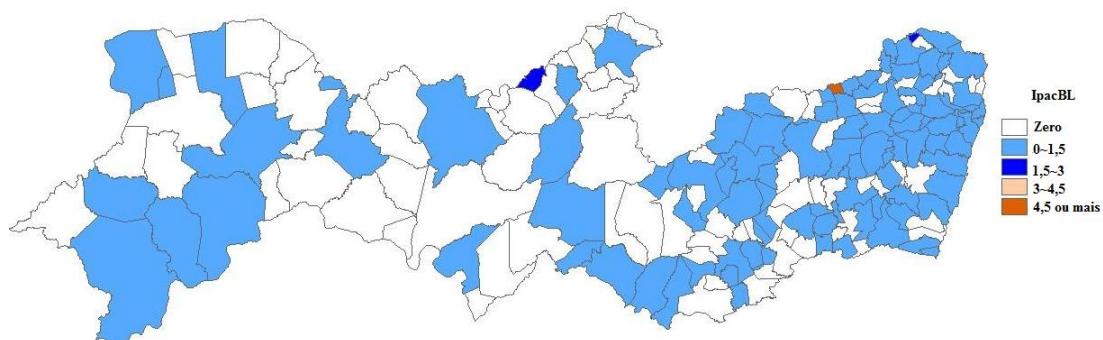


Figura 6: Taxa suavizada pelo Estimador Bayesiano Empírico Local para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBL).

## Análise Exploratória de Dados

Em seguida foi realizada uma análise descritiva das variáveis *Ipac*, *IpacBG* e *IpacBL* (Tabela 3), além da construção de histogramas e *boxplots*, a fim de melhor conhecer o comportamento das variáveis em questão. As figuras foram confeccionadas no *software* R versão 2.15.

A partir da Tabela 3 verificou-se que a taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD (*Ipac*) varia entre 1,262/100.000 e 50,810/100.000 habitantes. A taxa bruta de pacientes apresentou uma mediana de 5,8/100.000 habitantes, além disso, 75% (Q3) das cidades têm taxa abaixo de 8,148/100.000 habitantes e em média os municípios pernambucanos têm uma taxa de 7,118 pacientes por 100.000 habitantes e o desvio padrão foi de 6,023. Observou-se também, que a diferença entre as medidas de posição (média e mediana) expressou a assimetria da distribuição, afastando assim a hipótese de que a variável *Ipac* é proveniente de uma população com distribuição Normal. Além disso, essa assimetria sugere a presença de municípios com taxas discrepantes, nos quais as taxas de pacientes portadores de deficiência física em tratamento na AACD são superiores às taxas dos demais. Logo, existem regiões no estado que necessitam de atenção específica.

Tabela 3: Análise descritiva das variáveis *Ipac*, *IpacBG* e *IpacBL*.

Variável	Mín.	Q1	Mediana	Média	Q3	Máx.	Desvio padrão	CV (%)
<i>Ipac</i>	1,262	4,035	5,831	7,118	8,148	50,810	6.023	84,619
<i>IpacBG</i>	0,0001	0,067	0,179	0,363	0,383	5,704	0,668	183,878
<i>IpacBL</i>	0,0001	0,074	0,176	0,362	0,391	6,180	0,707	195,123

Ao utilizar os procedimentos Bayesianos empíricos (*IpacBG* e *IpacBL*) notou-se uma redução significativa nos valores das taxas de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco. A variável *IpacBG* variou de 0,0001 a 5,704 pacientes por 100.00 habitantes, com média de 0,363 pacientes por 100.000 habitantes e, além disso, 75% destes municípios têm taxa de pacientes abaixo de 0,383/100.000 habitantes. O desvio padrão foi de 0,668, apontando uma variabilidade menor nos dados, em comparação às taxas bru-

tas. Para a variável *IpacBL*, a qual considera a informação da vizinhança em seu cálculo, foi utilizado como critério de proximidade os 4 vizinhos mais próximos. Ao analisá-la, notou-se que as taxas são bem parecidas com as obtidas utilizando o procedimento Bayesiano empírico global (*IpacBG*). Outra observação a importante é que a variabilidade das taxas foram bastante reduzidas quando utilizados os procedimentos Bayesianos e isto se deve ao fato destas abordagens controlarem as flutuações aleatórias causadas quando se trabalha com pequenas populações sob-risco e poucos eventos por unidades de análise. Por outro lado, os Coeficientes de variação (CV) de ambas as taxas Bayesianas mostraram que existe uma alta dispersão dos dados em relação à média.

As Figuras 7, 8 e 9 apresentam os histogramas e os *boxplots* para as variáveis *Ipac*, *IpacBG* e *IpacBL*, respectivamente. Com base nas figuras, identificou-se que todas as variáveis possuem assimetria à direita e, portanto, afastando a possibilidade dos dados serem provenientes de populações com distribuição normal. Além disso, para a variável *Ipac* notou-se uma maior concentração de municípios com taxas no intervalo 0 a 10 e quando se aplicou o procedimento de suavização de Bayes (*IpacBG* e *IpacBL*), notou-se que a maioria do municípios tem menos de 1 paciente portador de doença genética com deficiência física em tratamento na AACD para cada 100.000 habitantes. Verificou-se também a existência de observações discrepantes ou candidatos a *outliers*, como por exemplo, os municípios de Vertente do Lério e Camutanga, os quais têm as maiores taxas brutas de pacientes em tratamento na AACD de Pernambuco. Das análises das Figura 8 e 9, verificou-se que estes mesmos municípios detém as maiores taxas suavizadas, indicando assim que os altos valores das suas taxas brutas não são fruto de flutuação aleatória. Estes municípios realmente se destacam pelo maior contingente de pacientes em tratamento na AACD de Pernambuco para cada 100.000 habitantes.

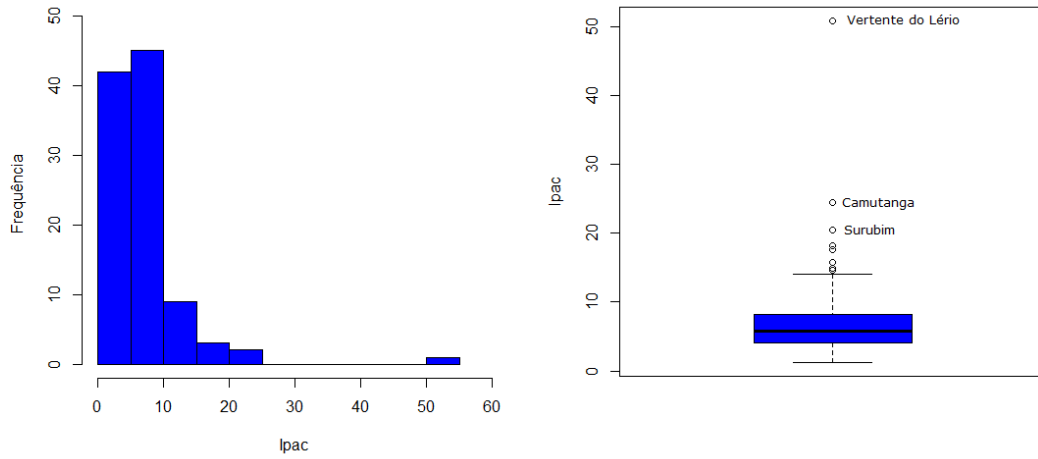


Figura 7: Histograma e *Boxplot* da taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (Ipac).

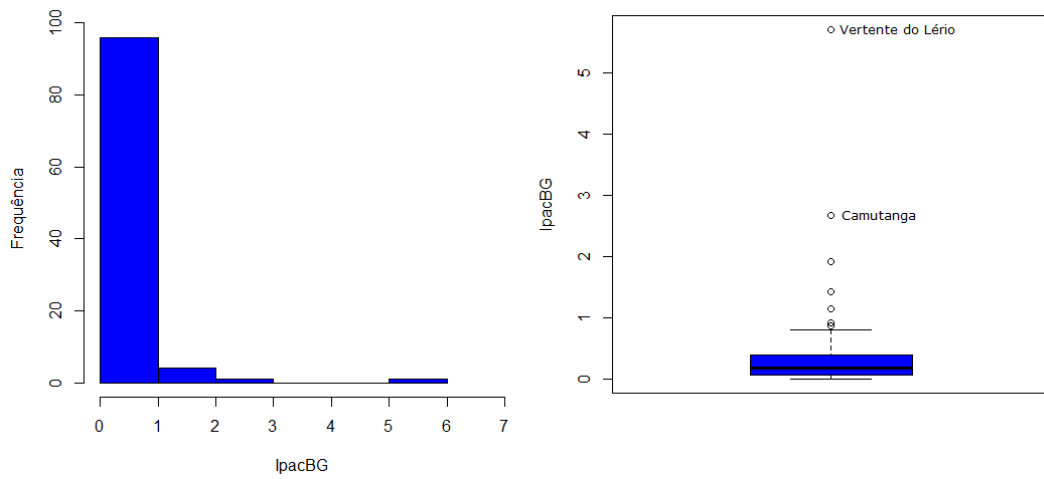


Figura 8: Histograma e *Boxplot* da taxa suavizada pelo estimador Bayesiano empírico global para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBG).

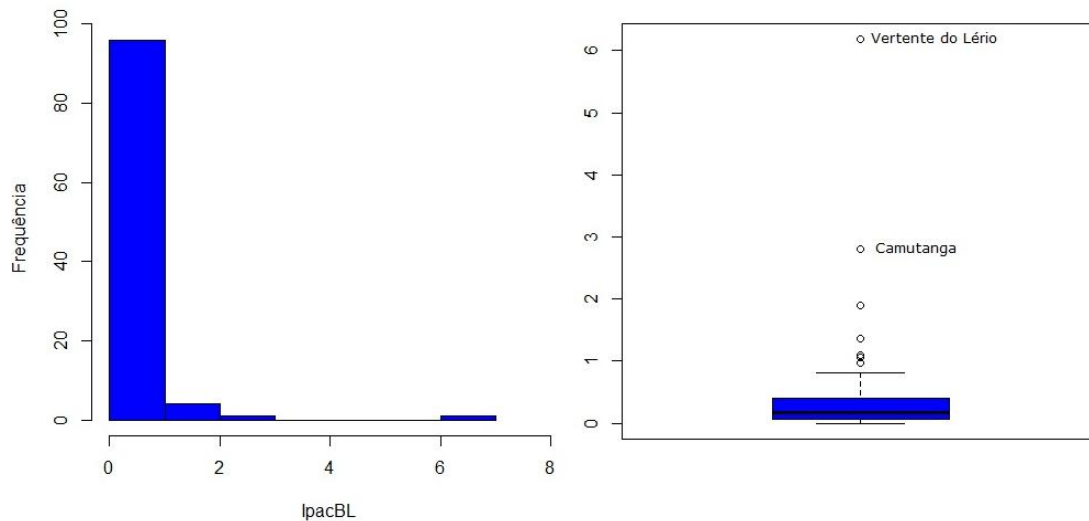


Figura 9: Histograma e *Boxplot* da taxa suavizada pelo estimador Bayesiano empírico local para os portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco (IpacBL).

A fim de verificar a existência de um padrão espacial específico na distribuição dos pacientes, foi construído o gráfico de pontos para a taxa bruta de pacientes (Figura 10) considerando as informações dos quantis e através dele pode-se observar um padrão espacial agregado, uma vez que as cores encontram-se aglomeradas. No gráfico, a cor azul representa os valores abaixo do quartil 1, a cor verde representa os valores que estão entre o quartil 1 e a mediana, em amarelo são observados os valores que estão entre a mediana e o 3º quartil e em vermelho estão agrupados os valores acima do 3º quartil.

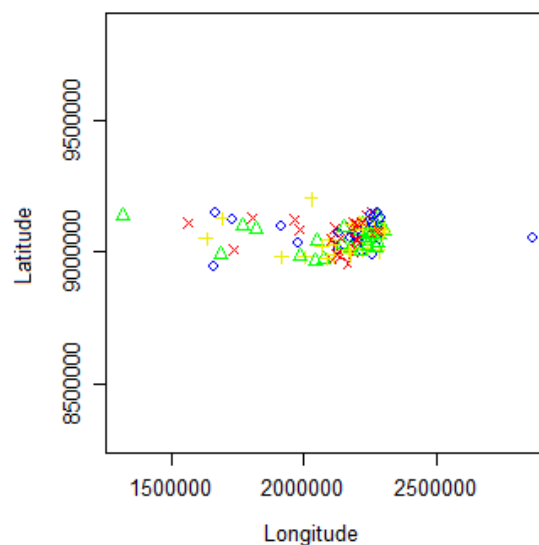


Figura 10: Gráfico de pontos para a distribuição espacial da variável Ipac.

## Índice $I$ global de Moran

Para o caso dos 102 municípios pernambucanos, o valor esperado do índice  $I$  global de Moran foi  $E(I) = -0,010$ . Se o valor calculado de índice for superior ao valor esperado, tem-se autocorrelação espacial positiva, caso contrário, tem-se autocorrelação espacial negativa. A hipótese nula deste teste é de que a taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco distribui-se aleatoriamente no território pernambucano em nível municipal, enquanto a hipótese alternativa afirma que há uma dependência espacial no indicador avaliado.

Para o cálculo do indicador de autocorrelação espacial global e confecção de gráficos e mapas, foram utilizados os valores da taxa bruta ( $I_{pac}$ ) e das taxas suavizadas  $I_{pacBG}$  e  $I_{pacBL}$ . A Tabela 4 contém a estatística  $I$  de Moran para as variáveis em questão de acordo com seis diferentes matrizes de pesos espaciais inverso da distância e  $k$  vizinhos mais próximos, para  $k=2,3,4,5$  e 6.

De acordo com a Tabela 4, verificaram-se valores positivos dos índices  $I$  de Moran para todas as variáveis e critérios da matriz de pesos espaciais adotados. Além disso, a dependência espacial entre as taxas de pacientes dos municípios pernambucanos é de pequena magnitude e essa dependência não foi considerada estatisticamente significativa na maioria dos critérios de matriz de vizinhança.

Com relação a variável  $I_{pac}$ , a hipótese nula de distribuição aleatória foi rejeitada ( $p$ -valor=0,04<5%) apenas para o critério da matriz de proximidade dos 4 vizinhos mais próximos ( $k=4$ ). Assim, conclui-se que existe dependência espacial entre os 102 municípios pernambucanos com suas respectivas taxas brutas de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco, levando em consideração os 4 vizinhos mais próximos de cada localidade. Com base nestas informações, entende-se que em média, os municípios com altas taxas de pacientes em tratamento na AACD de Pernambuco têm taxas parecidas aos seus 4 vizinhos mais próximos. Estes resultados poderiam ser úteis, por exemplo, para indicar regiões prioritárias para a construção de postos que viabilizem aos portadores de doença genética com deficiência física o acesso às terapias e acompanhamentos especializados sem que os mesmos tenham que fazer longas viagens até che-

gar ao centro de tratamento da AACD em Recife. Além das eventuais consultas e tratamentos especializados dados aos portadores de doença genética com deficiência física que já fazem tratamento na AACD, a construção destes postos de atendimentos ajudaria a encontrar novos pacientes residentes em localidades de difícil acesso e que podem ainda nem ter sido diagnosticados por falta de profissionais que compreendam a base da doença e por consequência, do tratamento correto a ser dado.

Com a utilização do método Bayesiano empírico global e local (IpacBG e IpacBL, respectivamente), notou-se que a hipótese de padrão espacial aleatório foi confirmada para todos os critérios da matriz de proximidade espacial, pois os p-valores dos índices  $I$  global de Moran foram maiores do que o nível de significância de 5%. Em outras palavras, não há indícios de que a taxa de pacientes de determinado município esteja correlacionada espacialmente com a taxa de pacientes de municípios vizinhos. Estes resultados poderiam indicar que não há regiões com aglomeração de pacientes e dessa forma os postos para atendimentos aos pacientes portadores de doença genética com deficiência física poderiam ser estabelecidos em qualquer localidade com infraestrutura e profissionais adequados.

Tabela 4: Índice  $I$  global de Moran e p-valor (entre parênteses) para as variáveis Ipac, IpacBG e IpacBL referente aos 102 municípios pernambucanos em estudo.

<b>Matriz de vizinhança</b>	<b>Ipac</b>	<b>IpacBG</b>	<b>IpacBL</b>
Inverso da distância	0,005 (0,21)	0,002 (0,28)	0,005 (0,17)
2 vizinhos mais próximos	0,082 (0,24)	0,002 (0,86)	0,018 (0,67)
3 vizinhos mais próximos	0,070 (0,21)	0,024 (0,55)	0,044 (0,32)
4 vizinhos mais próximos	<b>0,102 (0,04)</b>	0,046 (0,25)	0,063 (0,12)
5 vizinhos mais próximos	0,076 (0,08)	0,031 (0,35)	0,045 (0,19)
6 vizinhos mais próximos	0,054 (0,16)	0,023 (0,41)	0,035 (0,24)

## Diagrama de dispersão de Moran

O próximo passo da análise foi verificar se há municípios que se afastam do padrão global de associação positiva, mostrado pelo índice  $I$  global de Moran, assim como possíveis valores discrepantes. As Figuras 11, 12 e 13 apresentam os diagra-

mas de dispersão de Moran para as variáveis  $I_{pac}$ ,  $I_{pacBG}$  e  $I_{pacBL}$ , respectivamente, adotando o critério de vizinhança dos 4 vizinhos mais próximos ( $k=4$ ).

Em todas as figuras, as observações apresentaram-se em sua maioria, concentradas no centro do diagrama, ou seja, na coordenada (0,0), indicando baixa correlação espacial. Na Figura 11 existem alguns municípios agrupados no quadrante alto-alto que apresentam comportamento distinto (mais distantes da origem) dos demais e que se localizam no Agreste pernambucano. Dentre eles estão: Vertente do Lério, Surubim, Casinhas, Frei Miguelinho, Santa M<sup>a</sup> do Cambucá e Sanharó. Os regimes espaciais alto-alto evidenciam que municípios com elevada taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD estão circundados por vizinhos de padrão similar de comportamento. Ou seja, os municípios deste quadrante e seus 4 vizinhos mais próximos possuem altas taxas de pacientes em tratamento na AACD de Pernambuco. Já os municípios com regime do tipo baixo-baixo também possuem correlação espacial positiva, porém são grupos de municípios com baixos valores da taxa de pacientes portadores de doença genética com deficiência física cercados por localidades que também apresentam valores abaixo da média. Associações do tipo alto-alto e baixo-baixo formam os chamados conglomerados ou *clusters* espaciais. Municípios agrupados nos quadrantes baixo-alto apresentam baixos valores da taxa de pacientes portadores de doença genética com deficiência física em tratamento da AACD de Pernambuco circundados por localidades com altas taxas de pacientes portadores de doença genética com deficiência física em tratamento da AACD de Pernambuco. Já associações do tipo alto-baixo mostram grupos formados por municípios com altas taxas de pacientes portadores de doença genética com deficiência física em tratamento da AACD Pernambuco tendo como vizinhos municípios com baixos valores da referida taxa. Nesse quadrante destaca-se o município de Camutanga que com apenas dois pacientes em tratamento na AACD possui uma taxa bruta de 24,52/100.000 habitantes.



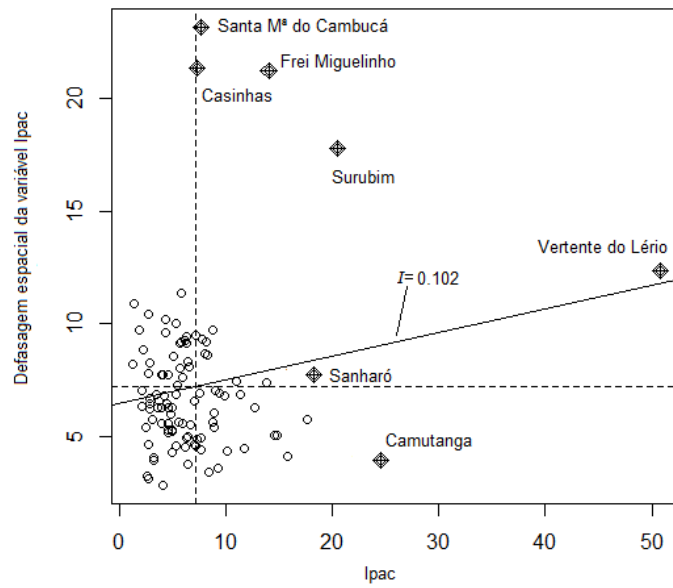


Figura 11: Diagrama de dispersão de Moran para a variável Ipac.

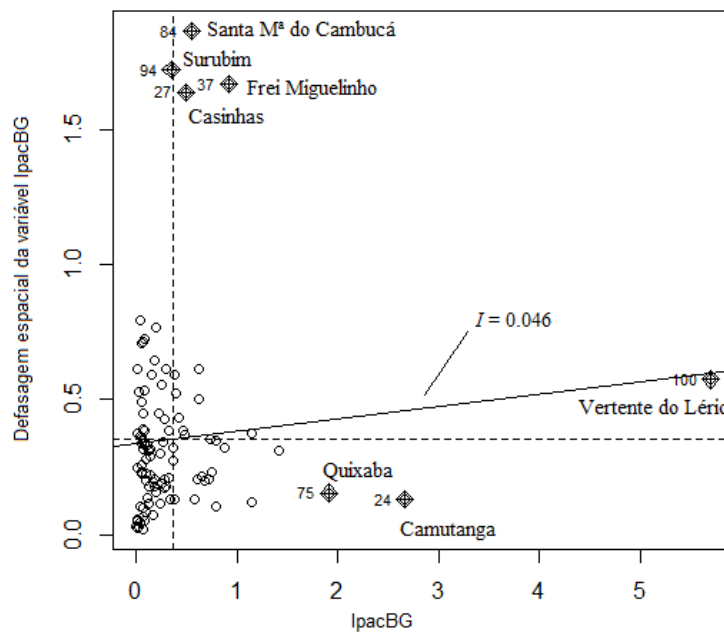


Figura 12: Diagrama de dispersão de Moran para a variável IpacBG.

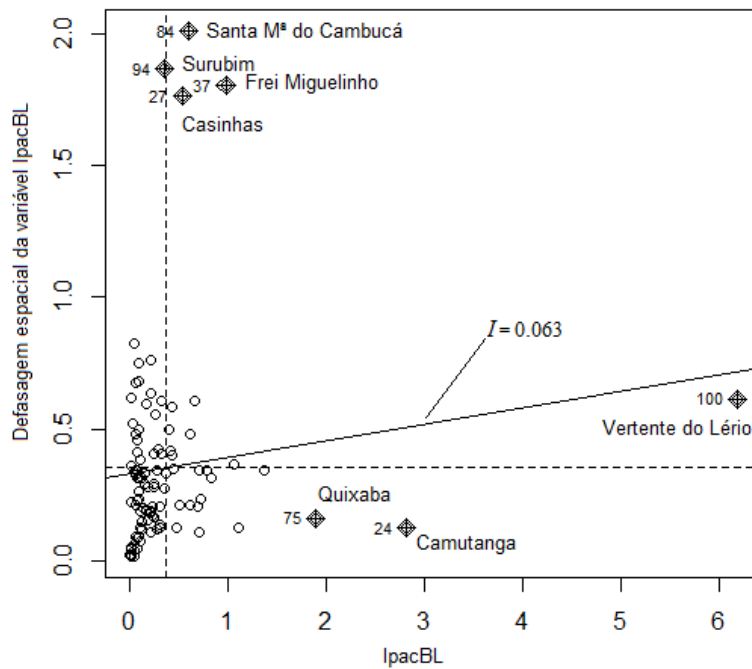


Figura 13: Diagrama de dispersão de Moran para a variável IpacBL.

Uma desvantagem do diagrama de dispersão de Moran é que eles exibem *clusters* estatisticamente significativos quanto não significativos. E como não há muito sentido em analisar *clusters* não significativos, essa ferramenta não proporciona uma análise mais detalhada, o que não ocorre com os indicadores locais de associação espacial, por exemplo, o Índice *I* local de Moran.

## Índice *I* local de Moran

Mesmo não se confirmando indício de autocorrelação espacial global para a maioria das taxas de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco, procedeu-se o cálculo do Índice *I* local de Moran para as taxas brutas e suavizadas e construíram-se mapas temáticos para verificar possíveis focos de dependência espacial em áreas delimitadas do território. Nas Figuras 14, 15 e 16 encontram-se os mapas de significância da autocorrelação local de Moran (*LISA Map*) com 999 aleatorizações, para as variáveis Ipac, IpacBG e IpacBL, respectivamente. Em tais mapas, os municípios são classificados em função da significância dos valores de seus índices locais, em: não significantes; significância entre 0,05 e 0,01; entre 0,01 e 0,001; e maior que 0,001.

Para a variável *Ipac* (Figura 14) observou-se um município com nível de significância entre 0,05 e 0,01 como é o caso de Santa M<sup>a</sup> do Cambucá e dois municípios com significância maior que 0,001 que são Vertente do Lério e Frei Miguelinho. Como observado no diagrama de Moran (Figura 11) estes locais encontram-se no quadrante alto-alto indicando assim que a taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco desses municípios e seus quatro vizinhos encontram-se superior a média global. Em outras palavras, são municípios com altas taxas de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco circundados por vizinhos com altas taxas desses pacientes. Para as variáveis *IpacBG* e *IpacBL* (Figuras 15 e 16, respectivamente) notou-se que os mesmos municípios foram destacados no mapa. Esse contingente de pacientes pode ser explicado pela proximidade dessas regiões em relação ao centro de tratamento da AACD cuja duração média da viagem é um pouco superior a 1 hora. Nos gráficos a seguir, a legenda zero indica locais onde não existem pacientes em tratamento na AACD de Pernambuco.

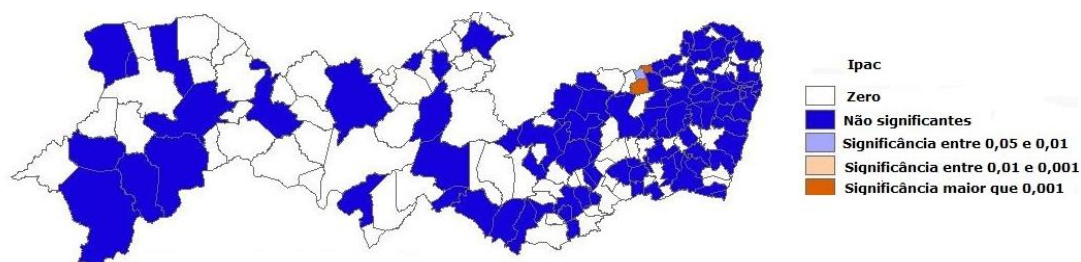


Figura 14: LISA MAP – Municípios classificados conforme a significância do Índice *I* local de Moran para a variável *Ipac* segundo o critério dos 4 vizinhos mais próximos.

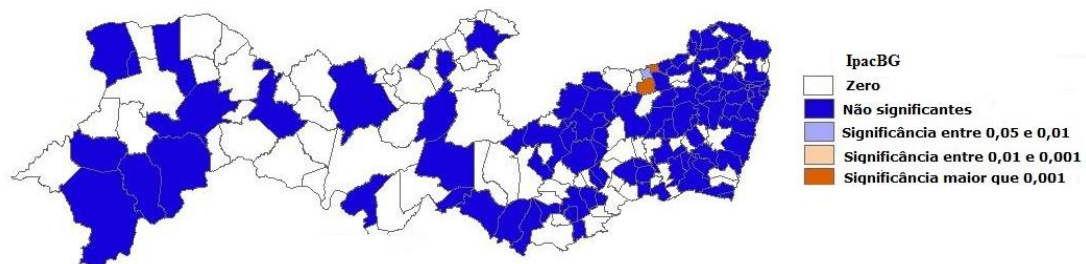


Figura 15: LISA MAP – Municípios classificados conforme a significância do Índice *I* local de Moran para a variável *IpacBG* segundo o critério dos 4 vizinhos mais próximos.

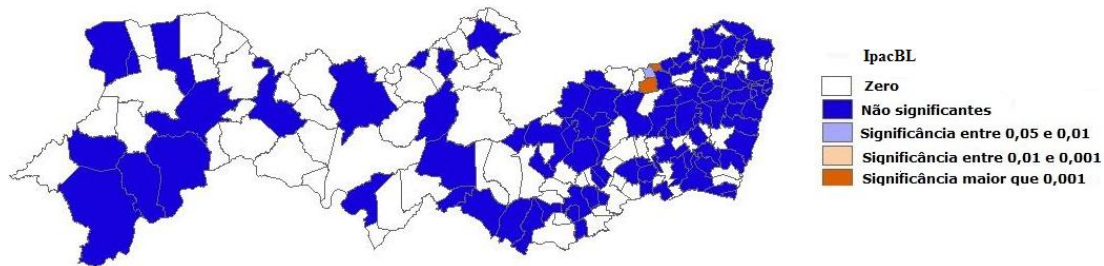


Figura 16: LISA MAP – Municípios classificados conforme a significância do Índice  $I$  local de Moran para a variável IpacBL segundo o critério dos 4 vizinhos mais próximos.

## Teste de Aleatorização

Para efeito de comparação com o resultado do índice  $I$  global de Moran foram realizadas 10.000 aleatorizações através do *software* R. A estatística utilizada foi o valor observado do Índice  $I$  global de Moran, em que este era comparado com o valor observado obtido através dos dados originais.

A Tabela 5 descreve os resultados do teste de aleatorização para as variáveis Ipac, IpacBG e IpacBL considerando o critério da distância inversa e dos  $k$  vizinhos mais próximos. Todos os resultados (p-valor) obtidos a partir da suposição de normalidade do índice  $I$  global de Moran (Tabela 4) foram ratificados pelo resultado do teste de aleatorização. O valor destacado na Tabela 5 corresponde àquele cujo teste de aleatorização indicou valor estatisticamente significativo (p-valor<0,05), o que sugere a existência de padrão espacial. Sendo assim, ambas as abordagens detectaram padrão espacial para a taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco considerando os 4 vizinhos mais próximos de cada município.

Estes resultados indicam que a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco está correlacionada com a taxa dos seus 4 vizinhos mais próximos geograficamente. Sendo assim, estas informações podem orientar na tomada de decisão na alocação de recursos e serviços, como por exemplo, ajudar na escolha de regiões mais apropriados para a criação de postos de atendimento básico ao pacientes com deficiência física em localizações de difícil acesso.

Tabela 5: Resultado (p-valor) do teste de aleatorização para as variáveis Ipac, IpacBG e IpacBL.

<b>Matriz de vizinhança</b>	<b>Ipac</b>	<b>IpacBG</b>	<b>IpacBL</b>
Inverso da distância	0,10	0,12	0,08
2 vizinhos mais próximos	0,11	0,34	0,24
3 vizinhos mais próximos	0,12	0,21	0,24
4 vizinhos mais próximos	<b>0,04</b>	0,11	0,08
5 vizinhos mais próximos	0,06	0,15	0,10
6 vizinhos mais próximos	0,09	0,17	0,11

## 4. Considerações Finais

As metodologias apresentadas permitiram testar a presença de padrões espaciais na distribuição da taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco, descrever as características das populações, verificar onde estão os maiores contingentes de pacientes em tratamento na AACD de Pernambuco e realizar diagnóstico de áreas prioritárias para intervenção no que diz respeito aos atendimentos básicos aos portadores de doença genética com deficiência física.

A hipótese nula de padrão espacial aleatório para a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco foi testada utilizando a suposição de normalidade da estatística  $I$  global de Moran e o teste de aleatorização. Ambas as abordagens mostraram-se satisfatórias e detectaram padrão espacial agregado somente para o critério da matriz de pesos espaciais dos 4 vizinhos mais próximos. Ou seja, os resultados indicaram que os municípios pernambucanos e seus 4 vizinhos mais próximos possuem a mesma configuração espacial no que tange a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco. Sendo assim, recomenda-se levar em conta estes resultados para orientar melhor possíveis tomadas de decisões.

## 5. Referências

ANSELIN, L. **Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences**. Santa Bárbara - USA: National Center for Geographic Information and Analysis, 1992.

ANSELIN, L. **The Moran Scatterplot as an ESDA Tool to Assess Local Instability**. [S.l.]: Spatial Association, 1993.

ANSELIN, L. **Exploratory Spatial Data Analysis and Geographic Information Systems**. Lisboa - Portugal: West Virginia University.1993b.

ANSELIN, L. Local Indicators of Spatial Association – LISA. **Geographical Analysis**, 1994.

ANSELIN, L. Local Indicators of Spatial Association – LISA. **Geographic Analysis**, v. 27, p. 91-114, 1995.

ANSELIN, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: FISCHER, M.; SCHOLTEN, H.; UNWIN, D. **Spatial analytical perspectives on GIS**. London: Taylor & Francis, 1996.

ANSELIN, L.; BAO, S. Exploratory Spatial Data Analysis Linking SpaceStat and ArcView. In: FISCHER, M. M.; GETIS, A. **Recent development in spatial analysis**. [S.l.]: [s.n.], 1997.

ASSUNÇÃO, R. M. Mapas de mortalidade. **X Encontro da Associação Brasileira de Estudos Populacionais**, Caxambú - MG, 2443-2453, 1996.

ASSUNÇÃO, R. M. B. S. M.; GUERRA, H. L.; SAKURAI, E. Mapas de taxas epidemiológicas: uma abordagem Bayesiana. **Caderno de Saúde Pública**, v. 14, p. 713-723, 1998.

BAILEY, T.; GATRELL, A. **Interactive Spatial Data Analysis**. Essex - England: Longman Scientific and Technical, 1995.

BAO, S.; HENRY, M. Heterogeneity issues in local measurements of spatial association. **Geographical Systems**, v. 3, p. 1-13, 1996.

BAPTISTA, E. A. **Os deficientes no Estado de Minas Gerais: uma abordagem espacial e demográfica utilizando os dados do Censo Demográfico de 2000**. Belo Horizonte - MG: Pontifícia Universidade Católica de Minas Gerais, 2009.

BARBOSA, L. M.; FREIRE, F. H. M. A. Análise espacial dos casos notificados de HIV/AIDS no estado de São Paulo. **XIV Encontro Nacional de Estudos Populacionais - ABEP**, Caxambú - MG, 2004.

BAUMONT, C. **Spatial effects in housing price models:** do house prices capitalize urban development policies in the agglomeration Dijon (1999). Chicago: Regional Group Seminar of The Federal Reserve Bank of Chicago, 2004.

BELLO, M. A. **Análise estatística espacial da mortalidade por câncer de mama feminina no estado do Rio de Janeiro, 2001 a 2006.** Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, 2010.

BRANCO, F. **Investigação Experimental:** Potência estatística dos testes de aleatorização na comparação de dois grupos independentes. Lisboa - Portugal: Universidade Aberta, 2010.

BRASIL - MINISTÉRIO DA EDUCAÇÃO. **Educação infantil:** saberes e práticas da inclusão: dificuldades de comunicação e sinalização: deficiência física. 4. ed. Brasília: Ministério da Educação, 2006.

BRASIL - MINISTÉRIO DA SAÚDE. **Sistemas de informações geográficas e análise espacial na saúde pública.** Brasília: Secretaria de Vigilância em Saúde / Fundação Oswaldo Cruz, 2007.

BRASIL - MINISTÉRIO DA SAÚDE. **Editorial do Ministério da Saúde.** Brasília: Secretaria de Atenção à Saúde / Política Nacional de Saúde da Pessoa Portadora de Deficiência / Secretaria de Atenção à Saúde, 2008.

BRASIL - MINISTÉRIO DA SAÚDE. **Sistema e Aplicativos.** Brasília: DATASUS, 2008.

CÂMARA, G. et al. **Análise espacial de áreas.** São José dos Campos - SP: INPE, 2002.

CÁRDENAS, R. E. N. **Análise da distribuição espacial da tuberculose no Distrito Federal, 2003-2007.** Brasília - DF: Universidade de Brasília, 2010.

CARVALHO, M. S. **Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco a Saúde.** Rio de Janeiro - RJ: Universidade Federal do Rio de Janeiro, 1997.

CARVALHO, M. S.; SANTOS, R. S. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Caderno de Saúde Pública**, Rio de Janeiro, v. 21, n. 2, p. 361-378, 2005.

CLIFF, A. D. et al. **Elements of Spatial Structure: A Quantitative Approach.** Cambridge: Cambridge University Press, 1975.

CRESSIE, N. **Statistics for spatial data.** [S.l.]: Wiley, 1993.

DI NUBILA, H. B. V.; BUCHALLA, C. M. O papel das Classificações da OMS - CID e CIF nas definições de deficiência e incapacidade. **Revista brasileira de epidemiologia**, v. 11, n. 2, p. 324-335, ISSN 1415-790X, 2007.



DRUCK, S. et al. Análise Espacial de Dados Geográficos. **EMBRAPA**, Brasília. ISSN 85-7383-260, 2004.

FOOK, K. D. **Integração da estatística especial em ambientes GIS**. São José dos Campos - SP: INPE, 2005.

GEARY, R. The Contiguity Ratio an Statistical Mapping. **The Incorporates Statistician**, v. 5, p. 115-145, 1954.

GETIS, A.; BOOTS, B. **Models of Spatial Processes**. London: Cambridge University Press, 1978.

GETIS, A.; ORD, J. K. The analysis of spatial association by use of distance statistics. **Geographical Analysis**, v. 24, p. 189-206, 1992.

HAGGETT, P.; CLIFF, A. D.; FREY, A. **Human Geography: Locational Analysis**. New York: John Witley & Sons, 1977.

KREMPI, A. P. **Explorando recursos de estatística espacial para análise da acessibilidade da cidade de Bauru**. São Carlos - SP: Universidade de São Paulo, 2004.

LESAGE, J. P. **Spatial Econometrics**. USA: University of Toledo, 1998.

LOUZADA, J. M.; BEARZOTI, E.; CARVALHO, D. Avaliação e aplicação de testes para a detecção da autocorrelação espacial usando marcadores genéticos. **Ciência e agrotecnologia**, Lavras - MG, v. 30, n. 2, Abril 2006.

MANLY, B. F. J. **J. Randomization, Bootstrap and Monte Carlo Methods in Biology**. 3. ed. London: Chapman & Hall, 2006.

MARSHALL, R. Mapping disease and mortality rates using empirical Bayes estimators. **Applied Statistics**, v. 40, p. 283-294, 1991.

MORAN, P. The Interpretation of Statistical Maps. **Journal of the Royal Statistical Society**, p. 243-251, 1948.

OLIVEAU, S.; GUILMOTO, C. Z. Spatial autocorrelation and demography, Exploring India's demographic patterns. **International Population Conference**, Tours, France, Julho 2005.

OLIVEIRA, A. S. B.; GABBAI, A. A. Doenças Neuromusculares. In: PRADO, F. C.; RAMOS, J. A.; VALLE, J. R. **Atualização Terapêutica: manual prático de diagnóstico e tratamento**. 20. ed. São Paulo: Artes Médicas, 2001. p. 815-827.

PAIVA, C. Dependência Espacial: Setores censitários, Zonas OD, Distritos, Sub Prefeituras, etc. **Carlos Eduardo de Paiva Cardoso**, 2007. Comunicação Pessoal.

PEREIRA, A. P.; VIOLA, D. N.; FERNANDES, G. B. Uso do Teste de Aleatorização na Análise de Séries Temporais. **19º Simpósio Nacional de Probabilidade e Estatística**, São Pedro- SP, 2010.

RAMOS, J. A. S.; FERREIRA, C. E. G.; MATTOS, H. J. Análise Espacial de Acesso aos Serviços de Saúde das Unidades de Atendimento aos Adolescentes no Município do Rio de Janeiro. **XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Curitiba - PR, 2011.

REED, U. C. Doenças neuromusculares. **Jornal de Pediatria**, v. 78, p. 89-103, 2002.

RIPLEY, B. D. **Spatial Statistics**. New York: John Wiley and Sons, 1981.

RODRIGUES, M. B. **Distribuição espacial da Mortalidade Infantil por Doenças Infeciosas e Parasitárias (DIP) em Jaboatão dos Guararapes**. Centro de Pesquisas Aggeu Magalhães - Fundação Oswaldo Cruz. Recife. 2007.

SALES, C. M. M.; AL., E. Análise espacial da tuberculose infantil no estado do Espírito Santo, 2000 a 2007. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba - MG, v. 43, n. 4, Agosto 2010.

SILVA, S. L. C. et al. Visualização dos padrões de variação da taxa de mortalidade infantil no Rio Grande do Sul, Brasil: comparação entre as abordagens Bayesiana Empírica e Totalmente Bayesiana. **Caderno de Saúde Pública**, Rio de Janeiro, v. 27, n. 7, Julho 2011.

SOARES, P. A.; NASCIMENTO, L. F. C. Análise espacial das internações por doenças do coração no Vale do Paraíba. **Arquivos Brasileiros de Cardiologia**, São Paulo - SP, v. 94, n. 6, Junho 2010.

TALEN, E.; ANSELIN, L. Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. **Environment and Planning A**, v. 30, p. 595-613, 1998.

TIZSLER, M. **Econometria especial: discutindo medidas para a matriz de ponderação espacial**. São Paulo - SP: Fundação Getúlio Vargas, 2006.

UPTON, G. J. G.; FINGLETON, B. **Spatial data analysis by example. Vol. 1: Point pattern and quantitative data**. Chichester - UK: Wiley, 1985.

VIEIRA, T. A. **História natural das mucopolissacaridoses: uma investigação da trajetória dos pacientes desde o nascimento até o diagnóstico**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2007.

VIOLA, D. N. **Deteção e modelagem de padrão espacial em dados binários e de contagem**. Piracicaba - SP: Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo, 2007.

YING, L. G. Understanding China's recent growth experience: A spatial econometric perspective. **Annals of Regional Science**, v. 37, p. 613–628, 2003.

YU, D. L.; WEI, Y. H. D. Analyzing regional inequality in post-Mao China in a GIS Environment. **Eurasian Geography and Economics**, v. 44, p. 514–534, 2003.

## Capítulo 3

---

### Comparação de Métodos para Identificação de Padrão Espacial em Pacientes Portadores de Doença com Deficiência Física em Tratamento na AACD Pernambuco

**Resumo:** O índice  $I$  global de Moran é a ferramenta usual para se caracterizar o padrão espacial para dados de áreas em dados epidemiológicos. A significância dos momentos de  $I$  pode ser avaliada sob  $H_0$ , por duas suposições: aproximação normal e aleatorização. O padrão espacial também pode ser identificado utilizando a relação entre as distâncias das observações e a distância das respostas, que podem ser colocadas em uma matriz. Quando o interesse do pesquisador é avaliar a relação entre os elementos de duas matrizes, deve-se aplicar o teste de aleatorização de Mantel. Para tal avaliação, é comum utilizar o coeficiente de correlação de Pearson, entretanto nem sempre a amostra é proveniente de uma população com distribuição normal bivariada. Quando esse pressuposto não é satisfeito utiliza-se como alternativa o coeficiente de correlação de postos de Spearman. Assim, por meio do presente estudo, buscou-se avaliar o desempenho do índice  $I$  global de Moran com matriz de pesos espaciais  $k=4$  vizinhos mais próximos e do teste de Mantel com coeficiente de Spearman, no que tange a capacidade de detectar padrão espacial para a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco em amostras de diferentes tamanhos. Retirou-se 1000 amostras de cada subamostra  $m$   $\{m = 10, 25, 50 \text{ e } 70 \text{ elementos}\}$  e optou-se por utilizar 10.000 aleatorizações para avaliação da significância de cada estatística de teste. O índice  $I$  global de Moran foi considerado melhor para detectar a presença de padrão espacial, pois a chance de detectar de padrão espacial nos dados foi maior com o aumento do tamanho das amostras e, além disso, essa metodologia parece fornecer uma proposta mais realista por considerar as informações dos vizinhos próximos nos seus cálculos.

**Palavras - chave:** Deficiência física; Doença genética; Teste de aleatorização; Teste de Mantel.

**Abstract:** The global Moran's  $I$  index is the usual tool to characterize the spatial data for areas of epidemiological data. The significance of the moments  $I$  can be evaluated under  $H_0$ , for two assumptions: normal approximation and randomization. The spatial pattern can also be identified using the relationship between the distances and the distance from the observations of responses that can be placed in an array. When the researcher's interest is to evaluate the relationship between the elements of two arrays, one must apply the Mantel randomization test. For such an assessment, it is common to use the Pearson correlation coefficient, though not always the sample comes from a population with a bivariate normal distribution. When this assumption is not satisfied it is used as an alternative coefficient Spearman rank correlation. Thus, through this study, we sought to evaluate the performance of the global Moran  $I$  index with spatial weights matrix  $k = 4$  nearest neighbors and the Mantel test with Spearman coefficient, regarding the ability to detect spatial pattern rate for patients with genetic disease with physical disabilities AACD Pernambuco in samples of different sizes. He retired in 1000 samples of each subsample  $m$   $\{m = 10, 25, 50 \text{ and } 70 \text{ elements}\}$  and we chose to use 10,000 randomizations to assess the statistical significance of each test. The global Moran  $I$  index was considered better to detect the spatial correlation because the chance of detecting the spatial pattern of the data was higher with the increase in sample size, and in addition, this approach appears to provide a more realistic proposition for consider information from nearby neighbors in their calculations.

**Keywords:** Physical disability, genetic disease; Randomization test; Mantel test.

## 1. Introdução

A partir da década de 1990 houve um grande aumento dos trabalhos cujo foco é a análise das situações de saúde, consolidando a importância da análise do espaço social na compreensão do processo saúde-doença e buscando entender a situação de saúde como resultado da relação entre os grupos sociais e seu território (COSTA & TEIXEIRA, 1999; BARCELLOS et al., 2002). Em estudos epidemiológicos, frequentemente deseja-se saber se a taxa de ocorrência de um determinado evento é espacialmente correlacionada em uma determinada área.

A análise espacial de taxas de ocorrências de eventos é uma importante ferramenta para o direcionamento de políticas públicas e tem sido utilizada em diversas áreas do conhecimento. Na epidemiologia, é comum o interesse em representar mapas de doenças que sejam o mais fidedigno possível com a situação real enfrentada pela população. Castro, Vieira e Assunção (2004) fizeram um estudo com o objetivo foi direcionar políticas públicas voltadas para o combate ao câncer de pulmão. Os resultados obtidos nesse trabalho mostraram que as maiores razões de mortalidade localizam-se aos extremos norte e sul da região estudada. Já para as mulheres, o sul do Paraná e do Rio Grande do Sul compõem as áreas com maior risco. Acosta (2008), estudou a relação da distribuição espacial da taxa de incidência da tuberculose pulmonar bacilífera em bairros da cidade de Porto Alegre com o nível socioeconômico da população. Os resultados obtidos puderam indicar áreas de maior vulnerabilidade social da cidade sinalizando que esse indicador de saúde pode ser utilizado para possíveis tomadas de decisão pelos gestores públicos.

Nos últimos anos houve um crescimento acentuado voltado às pesquisas na área da deficiência, entretanto a maioria dos trabalhos não aborda questões relacionadas à saúde dessa população (CASTRO et al., 2011). A deficiência é um termo que vem ocupando cada vez mais espaço no dia a dia dos brasileiros, principalmente, pelo fato da população de deficientes ter se conscientizado acerca de seus direitos, além da população ter se atentado quanto às diferenças individuais que se fazem presentes na essência de qualquer sociedade (BAPTISTA, 2009). As várias deficiências podem agrupar-se em quatro conjuntos distintos, sendo elas: deficiência visual, mental, auditiva e física. As deficiências físicas podem ser causadas por diferentes fato-

res, como os genéticos, complicações na gestação ou gravidez, acidentes, doenças infantis e da idade adulta. Entre as etiologias das deficiências físicas estão os fatores genéticos, fatores virais ou bacterianos, fatores neonatal e fatores traumáticos (BRASIL, 2006). Dentre as doenças causadas por fatores genéticos estão as de Depósito Lisossômico (DDLs) e as Neuromusculares (DNM), esta última tem sua maior prevalência em crianças. Dentre as instituições de apoio ao deficiente físico mais importantes está a Associação de Assistência à Criança Deficiente (AACD). A AACD é uma entidade filantrópica que tem 13 unidades espalhadas pelo País e há 62 anos promove a prevenção, habilitação e reabilitação de pessoas com deficiência física, especialmente de crianças, adolescentes e jovens, favorecendo a integração social.

O conhecimento do perfil dos portadores de doença genética com deficiência física de uma região é importante, pois a partir dessas informações é possível orientar a assistência aos cuidados básicos dessas pessoas (BAPTISTA, 2010). Conhecer a estrutura e a dinâmica espacial da população é o primeiro passo para a caracterização de situações de saúde. Além disso, permite o planejamento de ações de controle e alocação de recursos. Segundo Barcellos et al., (2002), a análise de situação de saúde é intrinsecamente espacial. Quanto maior a identidade entre a população e seu território, mais facilmente podem ser elaborados os diagnósticos e planejadas as ações de saúde.

A estatística espacial tem sido bastante aplicada em diversos ramos da ciência, como por exemplo, em análises de experimentos agrícolas, de padrões de mortalidade, em aplicações geofísicas, em agronomia, em estudos ecológicos e em estudos sobre a distribuição geográfica das doenças (LEGENDRE & FORTIN, 1989; ASSUNÇÃO & REIS, 1996; MAYES, 1997; SANTOS & NORONHA, 2001; FORTIN et al., 2002; CARNEIRO et al., 2005; VIOLA, 2007; FONSECA & VASCONCELOS, 2001; SILVA et al., 2011, MULLER et al., 2012)

A principal justificativa do uso dos métodos da estatística espacial está na questão da independência. Dados independentes são uma suposição conveniente, a qual torna mais simples a inferência. Porém, modelos que envolvem dependência estatística são frequentemente mais realistas (CRESSIE, 1993). Quando um indivíduo apresenta alguma qualidade que, provavelmente, se faz presente em indivíduos vizinhos, diz-se que o fenômeno exibe autocorrelação espacial (LOUZADA et al., 2006).

A análise de dados de área é um dos ramos de pesquisa da estatística espacial, em que a sua aplicação está associada a determinadas áreas  $A_i$ , com  $i = \{1, 2, \dots, n\}$ , contidas numa região  $R$ , de forma que a reunião de todas sub-regiões  $A_i$  seja a região  $R$ . Estes tipos de dados são encontrados com muita frequência quando se lida com eventos agregados por municípios, bairros ou setores censitários, onde não se dispõe da localização exata dos eventos, mas de um valor por área. Alguns exemplos de dados de área são: taxa de mortalidade ou taxa de incidência relacionada a alguma doença, percentual de adultos analfabetos e renda média.

As técnicas de análise de dados de área foram desenvolvidas para identificar regiões onde a distribuição dos valores possa apresentar um padrão específico associado a sua localização espacial. A informação que se busca é o quanto o valor de uma variável em uma determinada área assemelha-se com os valores da mesma variável localizada em sua vizinhança próxima, e o quanto é similar ou dissimilar aos valores de sua vizinhança distante (BRAGA et al., 2010). Essas técnicas são apresentadas na forma de indicadores, que medem tanto o padrão quanto a associação espacial, sendo o Índice  $I$  de Moran global e local as estatísticas comumente utilizadas. Podem-se encontrar exemplos de aplicações do Índice  $I$  de Moran em diversos estudos de distribuição de doenças (ACOSTA, 2008; GAMA & STRAUCH, 2009, MOTA, 2009; BARBOSA, 2011).

Após o calculado o Índice  $I$  global de Moran, é importante estabelecer sua validade estatística. A significância dos momentos de  $I$  pode ser avaliada, sob  $H_0$ , por duas suposições básicas: 1) aproximação normal: assume-se que os dados são realizações de  $m$  amostras independentes e identicamente distribuídas, proveniente de uma ou mais populações normais; 2) testes de aleatorização: admite-se que as observações são realizações aleatórias de  $m$  amostras independentes, tomadas de uma ou mais populações, cuja função de distribuição é desconhecida.

A base teórica dos testes de aleatorização foi proposta no século XX e atribuiu-se o seu desenvolvimento a muitos investigadores (KEMPTHORNE, 1952, 1955; EDGINGTON, 1964, 1966; HINKELMANN & KEMPTHORNE, 1994; EDGINGTON & ONGHENA, 2007). Entretanto, com o advento dos computadores, o interesse nestes testes passou de considerações teóricas a aplicabilidade prática sendo amplamente utilizados nas mais diversas áreas, como Biologia, Química, Psicologia,



Medicina e Experimentação agrônômica (JENSEN et al., 2003; MANLY, 2006; EDGINGTON & ONGHENA, 2007; VIOLA 2007).

Segundo Manly (2006), em comparação com métodos estatísticos padrão, os testes de aleatorização têm duas vantagens principais: são válidos para pequenas amostras, aleatórias ou não e, em geral, são fáceis de serem calculados. Além disso, são baseados em estatísticas não padronizadas e não é preciso saber previamente a distribuição da população da qual a amostra foi extraída. A principal desvantagem desses testes é que as conclusões obtidas ficam restritas a cada conjunto de dados e tipo de problema não sendo possível generalizar para a população (MANLY, 2006).

Os testes de aleatorização são usados para calcular a significância de qualquer estatística de teste. Em relação ao teste de Mantel, uma estatística comumente utilizada para avaliar a correlação entre pares de matrizes é o coeficiente de correlação de Pearson. Para isso, é necessário que a amostra provenha de uma população com distribuição normal bivariada. Quando esse pressuposto não é satisfeito utiliza-se como alternativa o coeficiente de correlação de postos de Spearman.

Em muitos casos, o interesse é estudar a relação entre os elementos de duas matrizes de distâncias. Por exemplo, pode-se querer saber se municípios próximos geograficamente são mais semelhantes entre si em número de pacientes portadores de doença genética com deficiência física do que aqueles mais distantes. Uma análise adequada e relativamente simples pode ser feita utilizando o teste de aleatorização de Mantel (MANLY, 2006). Em diversas obras, referem-se a utilizações de testes de aleatorização em áreas tão diversas como Biologia, Medicina, Psicologia e ecologia (MANTEL, 1967; GADBURY, 1993; ADAMS & ANTHONY, 1996; ROSSI, 1996; MANLY, 2006; AL-KHAFAJI, 2007; EDGINGTON & ONGHENA, 2007; LOUZADA, 2007; MELO & HEPP, 2008; LEGENDRE & FORTIN, 2010).

Dada a importância de se detectar padrões espaciais em população de portadores de doença genética com deficiência física, este estudo teve-se por objetivo comparar dois testes para verificar a presença de autocorrelação espacial para a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco. Em ambos os casos foi utilizado o teste de aleatorização, considerando como estatísticas de teste o Índice  $I$  global de Moran e o coeficiente de correlação de Spearman.

Avaliou-se o desempenho do Índice  $I$  global de Moran e do teste de Mantel com coeficiente correlação de Spearman para detectar a presença de autocorrelação espacial, sendo que a significância dos mesmos é testada através do teste de aleatorização. Ambas as metodologias foram avaliadas em diferentes condições de padrão espacial, considerando 1.000 amostras, variando os tamanhos das amostras  $\{m = 10, 25, 50 \text{ e } 70 \text{ elementos}\}$ . Foram utilizadas 10.000 aleatorizações em cada uma das subamostras geradas e foi adotado o nível de significância de 5%.

## 2. Desenvolvimento

### 2.1. Material utilizado

Para este estudo foi utilizado o banco de dados com o cadastro de 402 pacientes portadores de doenças genéticas com deficiência física em tratamento na AACD de Pernambuco e residentes em 102 municípios pernambucanos. As variáveis utilizadas foram: a taxa bruta de pacientes portadores de doença genética com deficiência física em tratamento na AACD ( $Ipac$ ), classes (E76; G12; G60; G71) e subclasses (E76.0; E76.2; G12.0; G12.1; G60.0; G60.8; G71.0; G71.1 G71.2; G71.8 G71.9) de doenças genéticas segundo a Classificação Estatística Internacional de Doenças (CID-10), as coordenadas geográficas das cidades em UTM e Graus. A variável  $Ipac$  foi calculada como segue:

$$Ipac_i = \frac{x_i}{p_i} 100.000, \quad i = 1, \dots, 102 \quad (1)$$

sendo  $x_i$  o número de casos observados no  $i$ -ésimo município e,  $p_i$  o tamanho total da população no  $i$ -ésimo município. Por intermédio desse cálculo, obtém-se o número observado de eventos no município  $i$  em uma população base de 100 mil pessoas. Esse padrão é utilizado, por exemplo, para a apresentação de taxas de eventos raros.

A Tabela 1 apresenta os códigos das classes e subclasses de doenças genéticas que causam deficiência física abordadas nesse estudo, segundo a classificação CID-10.

Tabela 1: Classes e subclasses das doenças genéticas em estudo de acordo com o CID-10.

Classes	Subclasses	Descrição
E76		<b>Distúrbios do metabolismo do glicosaminoglicano</b>
	E76.0	Mucopolissacaridose do tipo I
	E76.2	Outras mucopolissacaridoses
G12		<b>Atrofia muscular espinal e síndromes correlatas</b>
	G12.0	Atrofia muscular espinal infantil tipo I (Werndnig-Hoffman)

Tabela 1: Classes e subclasses das doenças genéticas em estudo de acordo com o CID-10.

(continuação)

Classes	Subclasses	Descrição
	G12.1	Outras atrofias musculares espinais hereditárias
		<b>Neuropatia hereditária e idiopática</b>
G60	G60.0	Neuropatia hereditária motora e sensorial
	G60.8	Outras neuropatias hereditárias e idiopáticas
		<b>Transtornos primários dos músculos</b>
G71	G71.0	Distrofia muscular
	G71.1	Transtornos miotônicos
	G71.2	Miopatias congênitas
	G71.8	Outros transtornos primários dos músculos
	G71.9	Transtorno muscular primário não especificado

Fonte: CID – 10; Ministério da Saúde, 2008.

## 2.2. Métodos

### 2.2.1. Autocorrelação espacial

Uma das ferramentas mais utilizadas no estudo de fenômenos areais é a análise da autocorrelação espacial. De acordo com Legendre (1993), esta técnica pode ser definida como a propriedade de variáveis assumirem valores, em pares de localidades separadas por uma certa distância, mais similares (correlação positiva) ou menos similares (correlação negativa) do que seria esperado pela associação aleatória dos pares observados.

A autocorrelação espacial pode ser estimada por meio de diversos indicadores, e dentre as ferramentas propostas na literatura, o Índice  $I$  global de Moran é a estatística mais utilizada. Um instrumento específico deve ser utilizado para captar a estrutura de correlação espacial que nesse caso é a matriz de pesos espaciais, usualmente chamada de  $W$ .

### 2.2.2. Matriz de pesos espaciais ou de proximidade espacial ( $W$ )

Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a matriz de proximidade espacial ( $W$ ), também chamada matriz de vizinhança ou matriz de pesos espaciais. Dado um conjunto de  $n$  áreas  $\{A_i, \dots, A_n\}$  constrói-se a matriz  $W^{(1)}$  de dimensão  $n \times n$ , em que cada um dos elementos  $w_{ij}$  representa uma medi-

da de proximidade entre as áreas  $A_i$  e  $A_j$ . Os elementos  $w_{ij}$  da diagonal são zero enquanto que os elementos  $w_{ij}$  indicam a forma como a região  $i$  está espacialmente conectada à região  $j$ . Estes elementos são não estocásticos, não negativos e finitos. Em geral, a matriz de peso é padronizada de modo que a soma dos elementos de cada linha é igual a 1 (BAUMONT et al., 2002). Alguns critérios usuais para a matriz de proximidade espacial são:

- $w_{ij} = 1$ , se o centro de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário  $w_{ij} = 0$ ;
- $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , caso contrário  $w_{ij} = 0$ ;
- $w_{ij} = l_{ij}/l_i$ , em que  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$ .
- $w_{ij} = 1/d$ , em que  $d$  é a distância entre os centros de  $A_i$  e  $A_j$ .

Baumont (2004) propõem que a matriz de vizinhança seja construída por meio do método dos  $k$  vizinhos mais próximos, a qual é definida da seguinte forma

$$\begin{cases} w_{ij}(k) = 0, \text{ se } i = j, \forall k \\ w_{ij}(k) = 1, \text{ se } d_{ij} < d_i(k), \\ w_{ij}(k) = 0, \text{ se } d_{ij} > d_i(k) \end{cases} \quad (2)$$

em que  $w_{ij}$  é um elemento da matriz normalizada,  $d_{ij}$  é a menor distância entre a região  $i$  e todas as outras unidades de modo que cada unidade tem exatamente  $k$  vizinhos.

A escolha da matriz de pesos espaciais é muito importante em uma Análise Exploratória de Dados Espaciais (AEDE), pois todos os passos subsequentes (ou resultados) dependerão dessa seleção.

### 2.2.3. Índice $I$ global de Moran

O Índice  $I$  global de Moran indica a magnitude da associação espacial através da autocorrelação espacial, considerando a covariância do produto dos desvios em relação à média. De acordo com Cressie (1993), o Índice  $I$  global de Moran é definido como:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n z_i z_j w_{ij}}{S_0 \sum_{i=1}^n z_i^2} \quad (3)$$

em que:  $n$  é o total de municípios avaliados,  $z_i = (x_i - \bar{X})$  e  $z_j = (x_j - \bar{X})$  são as informações centradas na média da variável  $X$ , para os municípios  $i$  e  $j$ , respectivamente,  $w_{ij}$  é o elemento da matriz quadrada e simétrica  $W(n \times n)$ , que expressa a relação espacial entre as observações e  $S_0 \sum_{i=1}^n z_i^2$  é uma matriz simétrica de pesos espaciais  $W$ . Além disso,  $\sum_{j=1}^n z_j w_{ij}$  é chamada de variável espacialmente defasada (*spatial lag*) da localização  $i$ .

A hipótese nula ( $H_0$ ) é a de completa aleatoriedade espacial, quando o indicador se distribui ao acaso entre as áreas sem relação com a posição destas. A significância dos momentos de  $I$  pode ser avaliada sob  $H_0$ , por duas suposições básicas: normalidade (N) ou aleatoriedade (R). Na primeira, os dados são observações independentes de uma distribuição normal. Já na segunda, a função de distribuição é desconhecida (CLIFF & ORD, 1981). Desse modo, a média e variância sob as duas suposições são definidas como (BELLO, 2010):

$$E_N(I) = E_N(I) = -\frac{1}{n-1} \quad (4)$$

$$Var_N(I) = \frac{n^2 S_1 - n^2 S_2 + 3S_0^2}{S_0(n^2 - 1)} \quad (5)$$

$$Var_R(I) = \frac{n[(n^2 - 3n + 3)S_1 - n^2 S_2 - 3S_0^2] - b_2[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{3S_0^2(n-1)} \quad (6)$$

em que,

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} ; S_1 = \frac{\sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2}{2}$$

$$S_2 = \sum_{i=1}^n \sum_{j=1}^n (w_{i.} + w_{.j})^2 ; w_{i.} = \sum_{j=1}^n w_{ij} ; w_{.j} = \sum_{i=1}^n w_{ij}$$

A interpretação do Índice  $I$  de Moran é análoga ao coeficiente de correlação de Pearson ou de Spearman, sendo então suas estatísticas, na maior parte dos casos, restritas ao intervalo  $[-1,+1]$ . Sendo assim, valores em torno de zero indicam que não há padrão espacial agregado. Valores positivos para o índice implica que municípios com alta taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco, tendem a ser rodeados por municípios com altas taxas de pacientes portadores de doença genética com deficiência física em tratamento na AACD e, ou, municípios com baixa taxa de pacientes portadores doença genética com deficiência física em tratamento na AACD de Pernambuco, tendem a ser circundados por municípios com baixas taxas de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco. Autocorrelação espacial negativa indica, por exemplo, que valores diferentes dessas taxas são encontrados juntos no espaço.

#### **2.2.4. Testes de Aleatorização**

Os testes de aleatorização são testes estatísticos cuja validade tem como base o modelo de distribuição aleatória. Este teste é baseado na pressuposição de que, se a hipótese nula e verdadeira, todas as possíveis ordens dos dados são igualmente prováveis. A hipótese nula diz que não existe padrão nos dados, ou se existe este padrão, isto é efeito do acaso enquanto a hipótese alternativa afirma que os dados apresentam certo padrão (MANLY, 2006).

Segundo Manly (2006), em comparação com métodos estatísticos padrão, os testes de aleatorização têm duas vantagens principais: são válidos para pequenas amostras, aleatórias ou não e, em geral, são fáceis de serem calculados. Além disso, são baseados em estatísticas não padronizadas e não é preciso saber previamente a distribuição da população da qual a amostra foi extraída. A principal desvantagem desse teste é que as conclusões obtidas não podem ser generalizadas para a população.

Segundo Viola (2007), para a realização deste teste, primeiramente calcula-se o valor ( $e_o$ ) de uma estatística “ $E$ ” de um conjunto de dados. Em seguida, aleatorizam-se os dados um número grande de vezes que no contexto de dados espaciais estas aleatorizações são obtidas por reordenação dos dados e para cada aleatorização obtêm-se estatísticas ( $e_a$ ) que irão gerar uma aproximação por simulação da distribui-

ção amostral de  $E$ . A etapa seguinte consiste em calcular o p-valor que é obtido pela proporção de vezes que a estatística obtida na  $i$ -ésima aleatorização ( $e_{ai}$ ) foi maior que a observada ( $e_o$ ). A hipótese nula será rejeitada quando esta proporção for menor que o nível de significância ( $\alpha$ ) adotado.

### 2.2.5. Teste de Mantel

O teste de Mantel é um dos poucos testes adequados quando a hipótese em estudo só pode ser formulada em termos de distâncias, o que é frequente em dados envolvendo a distribuição de doenças em determinado território. Ele permite observar a correlação entre as duas matrizes de proximidade ou distância (por exemplo,  $A$  e  $B$ ) calculando o produto cruzado dos valores correspondentes em  $A$  e  $B$ . As matrizes são simétricas ( $n \times n$ ) e descrevem as relações entre os  $n$  locais amostrados. As matrizes  $A$  e  $B$  podem ser formadas usando uma das muitas distâncias ou índices de dissimilaridade disponíveis na literatura (GOWER & LEGENDRE, 1986; MANLY, 2006). Por exemplo,

$$A = \begin{pmatrix} 0 & a_{21} & \cdots & a_{n1} \\ a_{21} & 0 & \cdots & a_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n1} & \dots & 0 \end{pmatrix} \text{ e } B = \begin{pmatrix} 0 & b_{21} & \cdots & b_{n1} \\ b_{21} & 0 & \cdots & b_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n1} & \dots & 0 \end{pmatrix}$$

sendo que  $A$  é matriz de distâncias Euclidianas entre os pontos  $(x_{1i}, x_{2i})$  e , isto é, com elementos  $a_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$  e  $B$  é a matriz com elementos  $b_{ij} = \sqrt{(z_i - z_j)^2}$ , em que  $Z$  é a variável número de pacientes portadores de deficiência física em tratamento na AACD em cada município. A estatística de teste é dada pelo coeficiente de correlação de postos de Spearman ( $\rho$ ) entre os elementos correspondentes de  $A$  e  $B$  que é calculada para os dados observados ( $\rho_o$ ). Após o cálculo de  $\rho_o$ , permutam-se linhas e colunas de uma das matrizes um número suficientemente grande e obtêm-se os valores  $\rho_{ak}$ , para  $k=1,2,\dots,10.00$  e  $\rho$ . Em seguida calcula-se a proporção  $p$  de vezes que  $\rho_a > \rho_o$  e compara-se com o nível de significância ( $\alpha$ ) já pré-estabelecido (por exemplo, 0,05). A hipótese  $H_0$  será rejeitada se  $p < \alpha$  (MANLY, 2006).



### 2.2.6. Coeficiente de correlação de postos de Spearman

O coeficiente de correlação de Spearman ( $\rho$ ) é uma alternativa usual para estimar correlações lineares nas situações em que há violação de suposição de normalidade conjunta para as variáveis  $X$  e  $Y$  (BAUER, 2007). Este coeficiente utiliza os postos dos dados amostrais e ao contrário do coeficiente de correlação de Pearson não requer a suposição de normalidade bivariada entre as variáveis. Além disso, as variáveis utilizadas podem ser medidas no nível ordinal.

A atribuição de postos, nesse caso, é feita separadamente para cada uma das variáveis. Assim, para a variável  $X$  atribui-se o posto 1 à menor variável, posto 2 à segunda menor variável e assim por diante, até o posto  $n$  para a maior variável. O mesmo procedimento é feito para a variável  $Y$ , independente dos valores da variável  $X$ . Valores iguais de cada uma das variáveis devem receber o mesmo posto. Assim, se , ou seja, os valores de ordem  $k$  e de ordem  $k+1$  estão empatados, para ambos atribui-se o posto  $(k+k+1)/2=(2k+1)/2=k+1/2$ . O coeficiente de correlação de Spearman é dado por:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (7)$$

em  $n$  é o número de pares  $(x_i, y_i)$  e  $d_i$  é a diferença entre os números de ordem das observações  $x_i$  e  $y_i$ , isto é,  $d_i = \text{n}^\circ \text{ ordem de } x_i - \text{n}^\circ \text{ ordem de } y_i$ .

### 3. Resultados e Discussão

O banco de dados utilizado consta com o registro de 402 pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco e que residem em 102 municípios pernambucanos. Primeiramente construiu-se um gráfico de pontos da taxa bruta de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco (Ipac), conforme Figura 1. Esse gráfico é geralmente construído para verificação inicial da existência de um padrão espacial específico e tem como base as informações dos quantis da distribuição. Além disso, suas cores representam as categorias em que cada valor da taxa está inserido. Isto é, a cor azul representa os valores abaixo do quartil 1, a cor verde representa os valores que estão entre o quartil 1 e a mediana, em amarelo são observados os valores que estão entre a mediana e o 3º quartil e em vermelho estão agrupados os valores acima do 3º quartil. Dessa forma, através da Figura 1 pode-se observar um padrão espacial aleatório, uma vez que as cores encontram-se distribuídas aleatoriamente no gráfico.

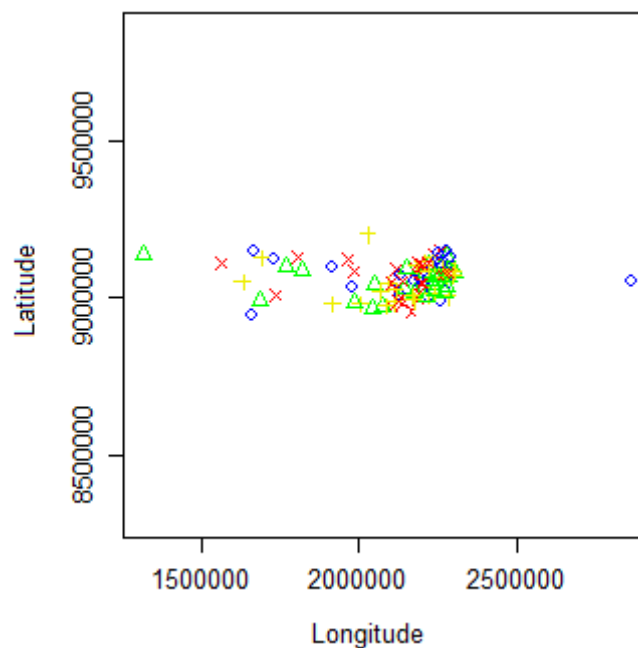


Figura 1: Gráfico de pontos para a distribuição espacial da taxa bruta de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco.

A etapa seguinte consistiu na avaliação do desempenho do índice  $I$  global de Moran considerando a matriz de pesos  $k=4$  vizinhos mais próximos e do teste de aleatorização de Mantel tendo como estatística de teste o coeficiente de Spearman, no

que tange a capacidade de detectar a presença de padrão espacial para a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco (Ipac) em amostras de diferentes tamanhos. A significância das referidas metodologias foi testada através do teste de aleatorização e para avaliar tal desempenho computou-se a proporção de vezes ( $p$ ) que a hipótese nula ( $H_0$ : ausência de padrão espacial) foi rejeitada ( $p\text{-valor} < \alpha$ ) nas 1000 amostras retiradas de tamanho  $m$   $\{m=10, 25, 50 \text{ e } 70\}$ . O nível de significância ( $\alpha$ ) adotado foi de 5%. A Tabela 2 apresenta os resultados do p-valor do teste de aleatorização e da proporção ( $p$ ) para cada teste avaliado.

Em relação ao índice  $I$  global de Moran, verificou-se que o p-valor do teste de aleatorização diminuiu à medida que aumentou o tamanho de  $m$ , ou seja, quanto maior o tamanho da amostra, maior é a probabilidade de rejeitar  $H_0$  e consequentemente detectar o padrão espacial. Isso pode ser verificado quando se utiliza os valores de toda a amostra ( $m=102$ ) e obtém-se um p-valor de 0,038, indicando que existe evidência de padrão espacial agregado para a taxa de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco quando se leva em consideração as informações dos seus 4 vizinhos mais próximos. Ressalta-se ainda, que foram feitos outros estudos utilizando o índice  $I$  de Moran com os critérios da matriz de pesos espaciais inverso da distância e  $k = 2, 3, 5$  e 6 vizinhos mais próximos, entretanto em nenhum dos casos houve indícios de padrão espacial. No que se refere aos resultados obtidos para a proporção ( $p$ ) de vezes em que a hipótese nula de ausência de padrão espacial foi rejeitada, observou-se o mesmo comportamento, isto é, a proporção de rejeição foi menor em amostras de tamanhos maiores. Por exemplo, em 1000 amostras retiradas de tamanho 50, a proporção de rejeição foi de aproximadamente 4%, ou seja, em 40 amostras a hipótese de padrão espacial aleatório foi rejeitada. Já nas amostras de tamanho 70, a proporção de rejeição foi de 3%, ou seja, houve rejeição de  $H_0$  em cerca de 30 amostras.

Em relação ao coeficiente de correlação de Spearman ( $\rho$ ), notou-se que os p-valores permaneceram praticamente inalterados nos 4 tamanhos de amostra considerados e em nenhum caso houve rejeição da hipótese de padrão espacial aleatório. O mesmo acontece com os valores obtidos para a proporção. Estes resultados podem ser causados pelo afastamento dos pontos em cada subamostra retirada ou mesmo por utilizar nos cálculos apenas as informações geográficas (latitude e longitude) do

município sem considerar as informações contidas nos vizinhos. Entretanto, apesar de não ter sido detectado padrão espacial em nenhum dos diferentes tamanhos de amostras, as metodologias utilizadas apresentaram, em média, proporção de rejeição de  $H_0$  bem próximas ao nível de significância adotado, sendo que para o índice  $I$  global de Moran foram, na maioria das vezes inferiores a 5%. Tais resultados indicam, mais uma vez, que incorporar a informação da vizinhança é uma boa alternativa para avaliar a presença de padrão espacial.

Tabela 2: P-valor do teste de aleatorização para detecção de padrão espacial da variável  $I_{pac}$  proporção ( $p$ ) de rejeição de  $H_0$  para as estatísticas de teste índice  $I$  global de Moran com matriz de pesos  $k=4$  vizinhos mais próximos e coeficiente de correlação de Spearman ( $\rho$ ).

<b>Tamanho da subamostra</b>	<b><math>I</math></b>		<b><math>\rho</math></b>	
	<i>p-valor</i>	<i>p</i>	<i>p-valor</i>	<i>p</i>
10	0,510	0,051	0,497	0,050
25	0,455	0,045	0,506	0,051
50	0,398	0,040	0,511	0,051
70	0,303	0,030	0,509	0,051

Uma última análise foi feita construindo-se um gráfico que mostra a relação entre a distância geográfica dos municípios e a distância entre as respostas (taxa de pacientes portadores de doença genética com deficiência física em tratamento na AACD de Pernambuco –  $I_{pac}$ ), conforme Figura 2. Esse gráfico permite avaliar se locais espacialmente próximos também possuem taxas parecidas. Com base na Figura 2, observou-se que existem pontos bastante afastados da reta e para haver um padrão espacial definido esperava-se que à medida que aumenta a distância entre as localizações, aumentaria também a distância entre as respostas. Entretanto, essa configuração não está representada no gráfico e, portanto, existem evidências de que as taxas de pacientes estão distribuídas aleatoriamente no espaço.

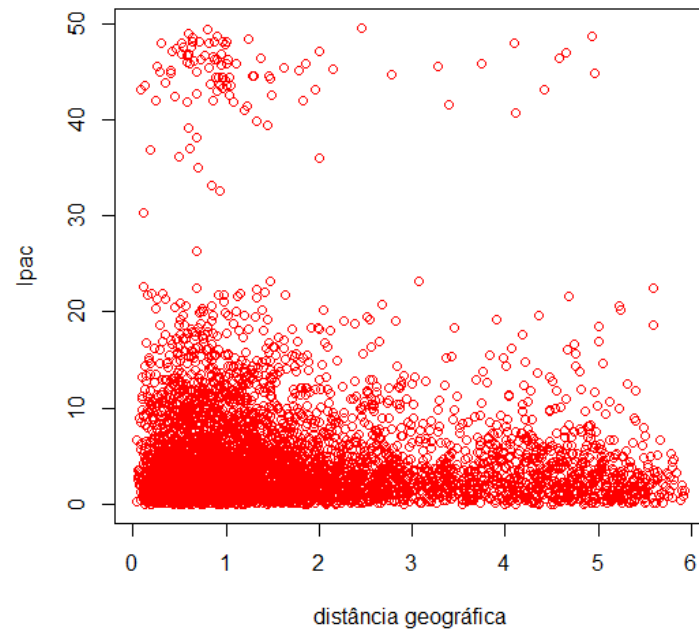


Figura 2: Relação entre a distância geográfica dos municípios pernambucanos e a taxa bruta de pacientes portadores de doença genética com deficiência física da AACD de Pernambuco (Ipac).

## 4. Considerações Finais

O índice  $I$  global de Moran com matriz de pesos espaciais  $k=4$  vizinhos mais próximos, mostrou-se uma melhor opção para detectar padrão espacial, pois a chance de detectar de padrão espacial nos dados cresceu à medida que o tamanho das amostras aumentou. Por outro lado, o teste de Mantel com correlação de Spearman não mostrou o mesmo comportamento. Esses resultados podem ser explicados, por exemplo, pelo fato dessa metodologia não levar em consideração o que ocorre na vizinhança próxima de cada localidade.

Propõem-se o uso de simulação para avaliar o desempenho do teste de Mantel em casos em que se verificou padrão espacial através de outras metodologias.

## 5. Referências

- ACOSTA, L. M. **O mapa de Porto Alegre e a Tuberculose: distribuição espacial e determinantes sociais**. Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 2008.
- ADAMS, D. C.; ANTHONY, C. D. Using randomization techniques to analyse behavioural data. **Animal Behaviour**, v. 51, p. 733–738, 1996.
- AL-KHAFAJI, K. et al. Detecting variability in demographic rates: randomization with the Kullback–Leibler distance. **Journal of Ecology**, v. 95, p. 1370–1380, 2007.
- ASSUNÇÃO, R.; REIS, E. Mapeamento da Malária em Rondônia Usando o Estimador Empírico de Bayes. **X Encontro Nacional de Estudos Populacionais**, 1996. 1077-1097.
- BAPTISTA, A. K.; SILVA, F. C. M. Perfil demográfico e do emprego das pessoas com deficiência no município de Belo Horizonte, MG – Brasil. **Revista Médica de Minas Gerais**, p. 290-298, 2010.
- BAPTISTA, E. A. **Os deficientes no Estado de Minas Gerais: uma abordagem espacial e demográfica utilizando os dados do Censo Demográfico de 2000**. Belo Horizonte - MG: Pontifícia Universidade Católica de Minas Gerais, 2009.
- BARBOSA, D. S. **Distribuição espacial e definição de áreas prioritárias para vigilância da leishmaniose visceral no município de São Luís, Maranhão, Brasil**. Rio de Janeiro - RJ: Fundação Oswaldo Cruz, 2011.
- BARCELLOS, C. et al. Organização espacial, saúde e qualidade de vida: análise espacial e uso de indicadores na avaliação de situações de saúde. **Informe Epidemiológico do SUS**, v. 11, n. 3, 2002.
- BAUER, L. **Estimação do coeficiente de correlação de spearman ponderado**. Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 2007.
- BAUMONT, C. **Spatial effects in housing price models: do house prices capitalize urban development policies in the agglomeration Dijon (1999)**. Chicago: Regional Group Seminar of The Federal Reserve Bank of Chicago, 2004.
- BELLO, M. A. **Análise estatística espacial da mortalidade por câncer de mama feminina no estado do Rio de Janeiro, 2001 a 2006**. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, 2010.
- BRAGA, A. S. et al. Estudo de Dependência Espacial utilizando Análise de Dados de Área Aplicada na Mesorregião Metropolitana de Belo Horizonte por meio do

Indicador Econômico PIB. **19º Simpósio Nacional de Probabilidade e Estatística**, São Pedro - SP, 2010.

BRASIL - MINISTÉRIO DA EDUCAÇÃO. **Educação infantil: saberes e práticas da inclusão: dificuldades de comunicação e sinalização: deficiência física**. 4. ed. Brasília: Ministério da Educação, 2006.

CARNEIRO, E. O.; SANTOS, R. L. S.; QUINTANILHA, J. A. Análise espacial aplicada na determinação de áreas de risco para algumas doenças endêmicas: o uso de técnicas de geoprocessamento na saúde pública. **Congresso Brasileiro de Cartografia**, Macaé - RJ, 2005.

CASTRO, M. S. M.; VIEIRA, V. A.; ASSUNÇÃO, R. M. Padrões espaço-temporais da mortalidade por câncer de pulmão no Sul do Brasil. **Revista Brasileira de Epidemiologia**, v. 7, n. 2, p. 131-143, 2004.

CASTRO, S. S.; LEFÈVRE, F.; LEFÈVRE, A. M. C.; CESAR, C. L. G. Acessibilidade aos serviços de saúde por pessoas com deficiência. **Revista de Saúde Pública**, São Paulo, v. 45, n. 1, p. 99-105, 2011.

CLIFF, A.; ORD, J. K. **Spatial Processes**. London: Pion Press, 1981.

COSTA, M. C. N.; TEIXEIRA, M. G. L. C. A concepção de “espaço” na investigação epidemiológica. **Caderno de Saúde Pública**, Rio de Janeiro, v. 15, n. 2, p. 271-279, 1999.

CRESSIE, N. A. C. **Statistics for spatial data**. New York: Wiley, 1993.

EGDINGTON, E. S. Randomization tests. **Journal of Psychology**, v. 57, p. 445–449, 1964.

EGDINGTON, E. S. Statistical inference and nonrandom samples. **Psychological Bulletin**, v. 66, p. 485–487, 1966.

EGDINGTON, E. S.; ONGHENA, P. **Randomization tests**. 4. ed. Boca Raton: Chapman & Hall/CRC, 2007.

FONSECA, F. R.; VASCONCELOS, C. H. Análise espacial das doenças relacionadas ao saneamento ambiental inadequado no Brasil. **Caderno de Saúde Coletiva**, Rio de Janeiro, v. 19, n. 4, Outubro 2011.

FORTIN, M.-J.; DALE, M. R. T.; HOEF, J. Spatial analysis in ecology. In: EL-SHAARAWI, A.; PIEGORSCH, W. W. **Encyclopedia of Environmetrics**. Chichester: Wiley, 2002. p. 2051-2058.

GADBURY, G. L. et al. Randomization tests for small samples: an application for genetic expression data. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 52, p. 365–376, 2003.



GAMA, R. G.; STRAUCH, J. C. M. Análise espacial de indicadores de Desenvolvimento Sustentável: Aplicação do índice de Moran. **XII Encuentro de Geógrafos Latinoamericanos - EGAL**, Montevideu, 2009.

GOWER, J. C.; LEGENDRE, P. Metric and Euclidean Properties of Dissimilarity Coefficients. **Journal of Classification**, v. 3, p. 5-48, 1986.

HINKELMANN, K.; O., K. **Design and analysis of experiments, Volume I: Introduction to experimental design**. New York: Wiley, 1994.

JENSEN, D.; NEVILLE, J.; RATTIGAN, M. **Randomization Tests for Relational Learning**. Massachusetts - USA: University of Massachusetts - Technical Report, 2003.

KEMPTHORNE, O. **Design and analysis of experiments**. New York: Wiley, 1952.

KEMPTHORNE, O. The randomization theory of experimental inference. **Journal of the American Statistical Association**, v. 50, p. 946-967, 1955.

LE GALLO, J.; ERTUR, C. Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. **Regional Science**, v. 82, p. 175-201, 2003.

LEGENDRE, P. Spatial autocorrelation: trouble or new paradigm? **Ecology**, Washington, v. 74, n. 6, p. 1659-1673, 1993.

LEGENDRE, P.; FORTIN, M.-J. Spatial pattern and ecological analysis. **Vegetatio**, v. 80, p. 107-138, 1989.

LEGENDRE, P.; FORTIN, M.-J. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. **Molecular Ecology Resources**, v. 10, p. 831-844, 2010.

LOUZADA, J. M.; BEARZOTI, E.; CARVALHO, D. Avaliação e aplicação de testes para a detecção da autocorrelação espacial usando marcadores genéticos. **Ciência e Agrotecnologia**, Lavras, v. 30, n. 2, 2006.

MANLY, B. F. J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**. 3. ed. [S.l.]: Chapman & Hall, 2006.

MANTEL, N. The Detection of Disease Clustering and a Generalized Regression Approach. **Cancer Research**, v. 27, p. 209-220, 1967.

MAYES, A.-M. A. Harnessing spatial analysis with GIS to improve interpretation of airborne geophysical data. **Proceedings of the Second Annual Conference Geocomputation 97**, Dunedin - New Zealand. 199-208, 1997.

MELO, A. S.; HEPP, L. U. Ferramentas estatísticas para análise de dados provenientes de biomonitoramento. **Oecologia Brasiliensis** 12(3):463-486, 2008.

MOTA, L. M. M. Perfil das hospitalizações por doenças respiratórias no Sistema Único de Saúde em Porto Alegre. VOL23, Nº 01 - INFLUENZA (01).p65, 2009.

MULLER, E. V. et al. Distribuição espacial da mortalidade por doenças cardiovasculares no Estado do Paraná, Brasil: 1989-1991 e 2006-2008. **Caderno de Saúde Pública**, Rio de Janeiro, v. 28, n. 6, p. 1067-1077, junho 2012.

ROSSI, J. P. Statistical Tool for Soil Biology. XI. Autocorrelogram and Mantel Test. **European Journal of Soil Biology**, v. 32, p. 195–203, 1996.

SANTOS, S. M.; NORONHA, C. P. Padrões espaciais de mortalidade e diferenciais socioeconômicos na cidade do Rio de Janeiro. **Caderno de Saúde Pública**, Rio de Janeiro, v. 17, p. 1099-1110, 2001.

SILVA, S. L. C. et al. Visualização dos padrões de variação da taxa de mortalidade infantil no Rio Grande do Sul, Brasil: comparação entre as abordagens Bayesiana Empírica e Totalmente Bayesiana. **Caderno de Saúde Pública**, Rio de Janeiro, v. 27, n. 7, Julho 2011.

VIOLA, D. N. **Deteção e modelagem de padrão espacial em dados binários e de contagem**. Piracicaba - SP: Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo, 2007.