

CLÁUDIO FRANCISCO CHIPENETE

**ABORDAGEM MATEMÁTICA NA ANÁLISE DE DADOS DE ÁREA
APLICADA À VARIÁVEL MALÁRIA EM MOÇAMBIQUE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de Magister Scientiae.

VIÇOSA
MINAS GERAIS – BRASIL
2015

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

C541a
2015
Chipenete, Claudio Francisco, 1977-
Abordagem matemática na análise de dados de área
aplicada à variável malária em Moçambique / Claudio Francisco
Chipenete. – viçosa, MG, 2015.
vii, 35f. : il. ; 29 cm.

Orientador: Gerson Rodrigues dos Santos.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.34-35.

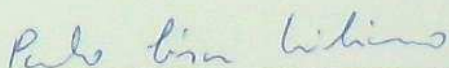
1. Estatística matemática. 2. Biometria. 3. Análise de dados
(Estatística). 4. Malária - Métodos estatísticos - Moçambique.
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-graduação em Estatística Aplicada e Biometria.
II. Título.

CDD 22. ed. 519.535

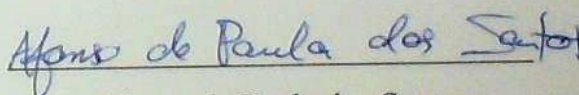
**ABORDAGEM MATEMÁTICA NA ANÁLISE DE DADOS DE ÁREA
APLICADA À VARIÁVEL MALÁRIA EM MOÇAMBIQUE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

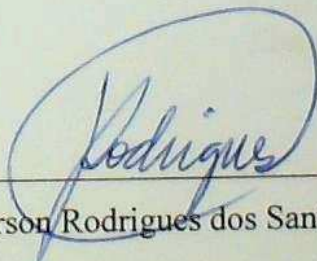
APROVADA: 07 de outubro de 2015.



Paulo César Emiliano
(Coorientador)



Afonso de Paula dos Santos



Gérson Rodrigues dos Santos
(Orientador)

À Gisela, querida esposa;
Elington e Ludwin, meus amados filhos.

O que é um ser humano?
É um animal racional (Aristóteles);
É um caniço pensante (Pascal);
É um cadáver adiado (F. Pessoa).

AGRADECIMENTOS

Ao professor Gérson Rodrigues dos Santos, pela orientação, paciência e compreensão de minhas necessidades específicas durante a realização do curso.

A Universidade Federal de Viçosa; ao Departamento de Estatística e a CAPES pela oportunidade em realizar o curso e bolsa concedida em todo processo de formação.

Ao professor Paulo César Emiliano pela grande ajuda prestada.

Aos meus pais, Issaia Chipenete e Benigna Francisco, por cuidarem de mim.

A Gisela, Elington e Ludwin pelo suporte total e desinteressado e amor leal! A todos os membros da comissão orientadora, pela orientação e apoio.

A Chen Weng pela colaboração e ajuda prestada e Alixandre Luppi pelo grande apoio prestado na compilação dos dados.

A Daiana Salles Pontes, pela grande amizade. A todos os colegas do curso pela convivência e companheirismo.

A Carla Zinato Campos, por aquela injeção de ânimo no tempo certo e pela atenção personalizada em diversas situações.

Aos amigos e companheiros de adoração pela acolhida, momentos compartilhados e amizade.

A todos que direta ou indiretamente contribuíram para realização deste trabalho, o meu... muito obrigado!

RESUMO

CHIPENETE, Cláudio Francisco, M.Sc., Universidade Federal de Viçosa, outubro de 2015. **Abordagem matemática na análise de dados de área aplicada à variável malária em Moçambique.** Orientador: Gérson Rodrigues dos Santos. Coorientadores: Nerilson Terra Santos e Paulo César Emiliano.

Ao se analisar os dados de área, um dos principais interesses é entender sua estrutura ou distribuição no espaço e, se existe alguma dependência ou estrutura bem definida entre as diversas áreas na região em estudo. Para mensurar essa dependência fez-se uma análise de padrões utilizando a autocorrelação espacial. O principal objetivo do trabalho foi abordar no enfoque matemático, as técnicas e procedimentos estatísticos na análise espacial de dados de área utilizando o método tradicional para o cálculo do índice de Moran e o método de três passos. Buscou-se também verificar e analisar a existência de algum padrão espacial definido em Moçambique associado a variável malária. A malária tem sido uma das principais causas de internamento nos hospitais e centros de saúde nos últimos anos, igualmente, das mortes da população. Analisar sua distribuição e relacionamento entre diferentes distritos do país poderá contribuir para minimizar os efeitos dessa doença. Os dados foram obtidos do Inquérito Demográfico e de Saúde de Moçambique (IDS) realizado em 2011. Na análise estatística foi possível identificar regiões cujos distritos se assemelhavam por possuírem taxas médias baixas de malária, formando agrupamentos, a saber, nas regiões sul, extremo sul, e norte de Moçambique. Para os demais distritos, verificou-se uma distribuição aleatória de casos da malária. No entanto, foi possível identificar distritos representados pelas cidades de Maputo, Matola e Beira com maior taxa de malária em relação aos demais.

ABSTRACT

CHIPENETE, Cláudio Francisco, M.Sc., Universidade Federal de Viçosa, October, 2015. **Mathematical approach in the area of data analysis applied to the variable malaria in Mozambique.** Advisor: Gérson Rodrigues dos Santos. Co-Advisor: Nerilson Terra Santos and Paulo César Emiliano.

In the area of data analysis, the main interest is to understand how these data are distributed in space and if there is any relationship or dependency between the various areas in the study area for a given phenomenon. To measure this dependence became a pattern analysis using the spatial autocorrelation. The main objective was to address the mathematical approach of the technical and statistical procedures in the area spatial analysis of data using the traditional method for calculating the Moran index and the three-step method. It sought to verify and analyze the existence of a spatial pattern set in Mozambique variable associated with malaria. Malaria has been a major cause of hospitalization in hospitals and health centers in recent years, also the deaths of the population. Analyze their distribution and relationship between different districts of the country could help to minimize the effects of this disease. Data were obtained from the Demographic and Health of Mozambique (DHS) conducted in 2011. Statistical analysis was possible to identify regions whose districts were similar because they have average rates of malaria low, forming groupings, namely, in the south, far south and northern Mozambique. For other districts, there is a random distribution of cases of malaria. However, it was possible to identify districts represented by the cities of Maputo, Matola and Beira with the highest malaria rates in relation to the other.

SUMÁRIO

1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Dados de área.....	5
2.2 Dependência espacial.....	5
2.3 Autocorrelação espacial	8
2.4 Índice de Moran global	8
2.4.1 Método tradicional	9
2.4.2 Método de três passos	9
2.5 Índice de Moran local	11
2.6 Média móvel local.....	12
2.7 Testes para avaliar a significância estatística dos índices	13
2.7.1 Teste de distribuição aproximada	13
2.7.2 Teste de permutação aleatória.....	15
2.8 Gráfico de espalhamento de Moran versus gráfico de dispersão	17
3 METODOLOGIA	19
3.1 Descrição da variável.....	19
3.2 Matriz de pesos espaciais (W).....	20
3.3 Análise exploratória	20
3.3.1 Índice de Moran global	20
3.3.2 Índice de Moran local (LISA).....	21
3.4 Teste de Significância Estatística.....	21
3.5 Método de suavização local - média móvel local	22
3.6 Diagrama de dispersão de Moran.....	22
3.7 O BoxMap, LisaMap, MoranMap e gráfico de dispersão de Moran	22
4 RESULTADOS E DISCUSSÃO	23
4.1 Abordagem matemática para análise espacial de dados de área	23
4.1.1 Índice de Moran global e local- abordagem tradicional.....	23
4.1.2 Índice de Moran global e local pelo método de três passos	23
4.1.3 Gráfico de dispersão de Moran	25
4.2 Aplicação em um caso: variável malária	25
4.2.1 Descrição da variável malária e média móvel local.....	25
4.2.2 Índices de Moran global e local, BoxMap e gráfico de dispersão de Moran	27
5 CONCLUSÕES	33
REFERÊNCIAS BIBLIOGRÁFICAS.....	34

1 INTRODUÇÃO

O interesse de um pesquisador em produzir informações úteis a partir de dados provenientes de uma determinada população, trouxe a necessidade de melhorar as técnicas estatísticas especialmente ao se tratar de dados espaciais ou com uma referência geográfica. Os métodos de análise de dados espaciais utilizados na estatística espacial são um produto dessa evolução.

Druck et al. (2004), mencionam que a ideia central da estatística espacial é incorporar informação espacial à análise que se deseja fazer, quantificando suas propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita. Como exemplo de aplicação da estatística espacial, citam o famoso estudo na epidemiologia realizado por John Snow, no século XIX em que incorporou a categoria informação espacial às análises realizadas acerca da epidemia de cólera na Inglaterra introduzida da Índia. Desde então, o interesse na aplicação desta técnica em diversas áreas de pesquisas como geologia, saúde, agricultura, áreas sociais entre outras foi aumentando.

Um dos maiores interesses da estatística espacial é determinar se um dado fenômeno ou evento é aleatório ou sistemático. Se um evento é regido por um processo aleatório, então não se pode fazer muito para identificar o porquê da causa desse fenômeno. Porém, se um dado fenômeno é regido por um processo que apresenta determinado comportamento ou padrão, o primeiro passo para entender tal comportamento é utilizar a análise da estatística espacial (WONG e LEE, 2005).

Dado um fenômeno espacial em estudo, sua análise dependerá do tipo de dado espacial a ser analisado. Segundo Cressie (1991), três tipos de dados caracterizam um problema de análise espacial: dados de eventos ou padrões pontuais, dados de superfícies contínuas e dados de área com contagens e taxas agregadas. Quando se lida com eventos agregados por municípios, bairros e censos, com frequência são utilizados dados espaciais associados a áreas delimitadas por polígonos (DRUCK et al., 2004). É neste último tipo de dados, dados de área, associado a um fenômeno epidemiológico em Moçambique, a malária, que o presente trabalho terá o seu foco. A escolha da malária como objeto de análise, deve-se ao fato desta endemia ser uma das principais causas de morte entre a população Moçambicana. Segundo o Programa Nacional Contra a Malária, PNCM (2012), em nível de atendimento hospitalar, mais de metade das

peessoas que se dirigem aos centros de saúde, são devido a malária e as maiores vítimas são crianças com menos de cinco anos de idade, internadas nos serviços e centros pediátricos.

Na presente dissertação pretende-se fazer uma abordagem matemática de análise de dados de área aplicada à variável malária, e identificar uma estrutura ou um padrão definido para o fenômeno em estudo em Moçambique.

Duas abordagens serão utilizadas: a primeira, envolve análise espacial de dados de área pelo método tradicional e frequentemente analisados pelos softwares estatísticos INPE-DPI (2012) e o ESRI (2012) no cálculo dos índice de Moran, e ainda, na visualização dos resultados através de mapas temáticos e gráficos o que irá permitir identificar a natureza da estrutura dos dados analisados; a segunda abordagem com ênfase na derivação matemática baseada em álgebra linear, na reconstrução da estrutura matemática a partir da fórmula tradicional para o cálculo do índice de Moran pelo método de três passos proposto por Chen (2013).

2 REFERENCIAL TEÓRICO

A estatística espacial tem sido utilizada quando se pretende quantificar um determinado fenômeno ou evento em estudo localizado no espaço. Segundo Cressie (1991), quando um determinado dado ou valor de uma variável pode ser associado a uma coordenada espacial, então é chamado de dado georreferenciado. Como exemplo de dados georreferenciados, Druck et al. (2004) citam dados censitários associados ao local de residência do indivíduo, estudos epidemiológicos na área de saúde que permite mapear um determinado evento epidemiológico no espaço e tempo.

Uma primeira característica de interesse ao se analisar dados espaciais está em investigar padrões espaciais de lugares e valores, a associação espacial entre eles e a variação sistemática do fenômeno por localização. Portanto, o interesse da estatística espacial concentra-se na análise de fenômenos que ocorrem no espaço e os métodos utilizados pretendem descrever e analisar o seu comportamento (ANSELIN, 1995).

Assunção (2001) aponta uma segunda característica interessante na qual a estatística espacial tem o seu lugar: a existência da dependência espacial nos dados amostrados. Essa característica faz com que estudos sobre análise de dados espaciais sejam de natureza complexa para serem analisados por métodos tradicionais da estatística, porque na estatística clássica a independência das observações é uma das fortes pressuposições a não ser violada (RANCHER e SCHAALJE, 2008).

Algumas etapas devem ser seguidas na análise de dados espaciais. Bailey e Gatrell (1995) mencionam etapas que incluem métodos voltados na visualização dos dados espaciais, métodos exploratórios que investigam e resumem relações e padrões mapeados e métodos que auxiliem na escolha de um modelo estatístico e a estimação dos parâmetros desse modelo. Uma vez que a visualização por meio de mapas temáticos e gráficos é de suma importância na análise de dados espaciais para se entender sua distribuição espacial e da existência de padrões, é recorrente o uso do Sistema de Informação Geográfica (SIG). O SIG é, segundo Assunção (2001), um conjunto de equipamentos e programas de computação, que integram mapas e gráficos com uma base de dados, sobre um espaço geográfico definido, ou seja, um conjunto de ferramentas capazes de coletar, armazenar, manejar, analisar e visualizar informações georreferenciadas.

Outra fase ou etapa mencionada pelos Bailey e Gatrell (1995), é a análise exploratória dos dados espaciais, cujo objetivo é identificar propriedades espaciais dos dados para verificar existência de padrões entre outros. Segundo Anselin (1995), a análise exploratória de dados espaciais segue em duas dimensões: a primeira refere-se a distinção entre dois indicadores de fácil interpretação mediante a possibilidade de representação dos valores individuais de cada área através de mapas:

a) Indicador global procura quantificar o efeito da dependência espacial, em que o esquema geral de dependência espacial se resume a um único valor obtido através do índice de Moran global, utilizados para se conhecer o grau de interação do conjunto de dados espaciais.

b) Os indicadores locais conhecidos como LISA (Local Indicator of Spatial Association - Indicador Local de Associação Espacial) são valores associados a cada área específica que fornecem informação relativa a um agrupamento espacial ao redor de cada uma dessas áreas, em que a soma de valores de LISA é proporcional ao indicador global de associação espacial.

Na segunda dimensão mencionada por Anselin (1996) faz a distinção entre os dois modelos de dados, em que a correlação pode ser analisada baseada em dados pontuais e, o outro em uma coleção fixa de localizações espaciais discretas, ou seja, pontos ou polígonos. No primeiro caso assume-se que a interação espacial é uma função contínua de distância entre pares de observações. No segundo caso a interação espacial é entendida como uma função denominada 'step function' segundo a qual uma localização interage com um dado grupo de vizinhos.

Uma terceira característica fundamental na análise de dados espaciais é o tipo de dado a ser analisado, uma vez que, para cada tipo de dado está associado um método distinto. Cressie (1991) classifica os dados espaciais em três categorias: dados de superfícies contínuas, dados pontuais e dados de área. Uma vez que os primeiros dois tipos de dados não serão objetos de análise neste trabalho, não se fará seu aprofundamento teórico. Pode-se consultar também Bailey e Gatrell (1995), Assunção (2001), Druck et al. (2004) e Wong e Lee (2005).

Antes de se abordar os dados de área e sua análise estatística na seção 2.1, é de interesse conceituar variável regionalizada ou espacial. Segundo Cressie (1991), qualquer variável distribuída no espaço é dita ser regionalizada com uma, duas ou três dimensões. Em termos matemáticos, uma variável regionalizada pode ser vista como

uma função $f(S)$ que assume um valor em cada ponto s numa região R . A localização S_i é usada para indicar que pode ser multidimensional. Por exemplo, S_i como (x_{1i}, x_{2i}) , em que x_{1i} e x_{2i} são coordenadas numa determinada região R .

Santos (2010) mostrou que o modelo espacial geral é dado por:

$$Y(x) = \mu(x) + \varepsilon'(x) + \varepsilon''; \quad (1)$$

em que $Y(x)$ é a variável aleatória; $\mu(x)$ é o valor médio da variável Y ; $\varepsilon'(x)$ é variação probabilística regionalizada, $Y(x)$, com $E[\varepsilon'(x)] = 0$; ε'' é o ruído normal, isto é, $\varepsilon'' \sim N(0, \tau^2)$, variável aleatória independente e identicamente distribuída com $E(\varepsilon'') = 0$ e $Var(\varepsilon'') = \tau^2$.

2.1 Dados de área

Os dados de área são elementos de uma base de dados que servirá para gerar um mapa. Para cada polígono ou área desta base de dados podem-se ter diversos atributos. Portanto dados de área envolvem elementos que associam o mapa geográfico de uma região R a uma base de dados. Esse mapa geográfico é dividido por áreas A_i com $i = 1, 2, \dots, n$ de forma que $\bigcup_{i=1}^n A_i = R$ e $A_i \cap A_j = \emptyset$ se $i \neq j$, qualquer que seja a forma de A_i e de R . Em dados de área, a informação contida na área A_i de um determinado evento é concentrada em um único ponto que pode ser, por exemplo, a média do número de casos ocorridos nessa área. Cada uma dessas áreas A_i possui uma ou mais variáveis aleatórias X_1, X_2, \dots, X_n . Ao se analisar dados espaciais de áreas, o objetivo principal envolve a descrição precisa de um fenômeno ocorrido no espaço, a exploração da existência de um determinado tipo de padrão - se o padrão é aleatório, regular ou em forma de agrupamentos (clusters) - e a relação com os dados analisados, além de buscar explicações para determinado tipo de padrão verificado (ASSUNÇÃO, 2001; BAILEY e GATRELL, 1995).

Outro conceito igualmente importante na estatística espacial ao se analisar dados de área é o da dependência e autocorrelação espacial, que serão abordados nas seções 2.2 e 2.3 respectivamente.

2.2 Dependência espacial

Segundo Wong e Lee (2005), ao se classificar padrões espaciais de polígonos em dispersos, aleatórios ou agrupados, o objetivo é mensurar a semelhança ou

dissemelhança entre qualquer conjunto de polígonos através de certa vizinhança ou uma definição de distância definida. Como ferramenta de medida utiliza-se a autocorrelação espacial do atributo. Portanto, deve-se definir o que se entende por duas ou mais áreas ou polígonos serem próximas ou vizinhas uma da outra.

Bailey e Gatrell (1995) mostraram alguns critérios que podem ser utilizados na criação da matriz de pesos espaciais (matriz W):

- $w_{ij} = 1$, se o ponto de referência de A_j é um dos k pontos mais próximos ao ponto de referência de A_i , e $w_{ij} = 0$, caso contrário;
- $w_{ij} = 1$, se o ponto de referência de A_j está dentro de uma distância especificada ao ponto de referência de A_i e $w_{ij} = 0$, caso contrário;
- $w_{ij} = 1$, se A_i tem fronteira comum com A_j , e $w_{ij} = 0$, caso contrário;
- $w_{ij} = v_{ij}/v_i$, onde v_{ij} é o comprimento da fronteira comum entre A_i e A_j e v_i é o perímetro de A_i ;
- $w_{ij} = 1/(d_{ij})^p$, em que $p \in \mathbb{N}$ e d_{ij} é a distância entre os centroides A_i e A_j .

Como exemplo de aplicação de um desses critérios na análise de dados de área, cita-se Krempi (2004), Zhang e Lin (2009), Silva (2010) em que usaram o critério de fronteira comum, assumindo-se todas as áreas com igual peso, independentemente da distância que as separa.

A Figura 1 ilustra o mecanismo de construção da matriz W , uma matriz binária (0,1), pelo critério de fronteira comum, em que se atribui valor 1 às áreas que partilham a fronteira em comum, e 0 àquelas sem essa propriedade (RAMOS, 2002). Percebe-se que a área representada pela letra A partilha fronteira comum com as áreas representadas pelas letras B e C.

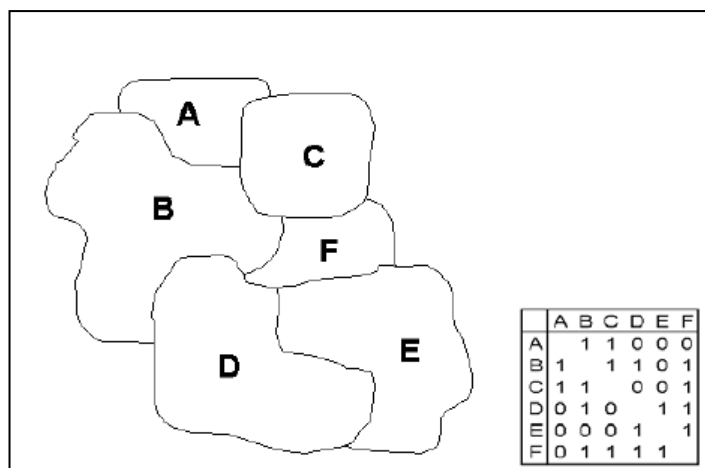


Figura 1: Construção da matriz W pelo critério de fronteira comum. Fonte: Ramos (2002).

Assim, na construção da matriz W , A e B seria o elemento da matriz com $w_{AB} = 1$, e A e C seria o elemento da matriz em que $w_{AC} = 1$. Para o caso em que não haja partilha de fronteira, por exemplo, entre as áreas A e F, nesse caso, o elemento que comporia a matriz W seria $w_{AF} = 0$. Como a matriz W é utilizada em cálculos de indicadores de análise exploratória, por conveniência, ela é muitas vezes utilizada normalizada por linha, ou seja, com a soma dos ponderadores de cada linha igual a 1 (DRUCK et al., 2004). Por outro lado, pode-se construir a matriz W usando definições de distância entre os centroides das áreas A_i e A_j , ou seja, o ponto central de cada uma das áreas, para representar tais áreas. Nesse caso, podem-se utilizar algumas definições de distâncias, por exemplo, a euclidiana, Manhattan, circular ou distância Minkowskyan (WONG e LEE, 2005). Na presente dissertação será utilizada a distância euclidiana.

A distância euclidiana é a distância medida entre dois pontos num plano de coordenadas cartesianas. A distância euclidiana entre os centroides de A_i com as restantes A_j áreas, é dada pela seguinte expressão:

$$r_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}; \quad (2)$$

em que r_{ij} é a distância entre os centroides i e j ; x e y são coordenadas dos centroides de cada área.

Segundo Wong e Lee (2005) as distâncias são usadas como fator peso, uma vez que áreas próximas tendem a exercer mais influência uma sobre a outra em relação as áreas mais distantes. Cita-se como exemplo Chen (2013) e Fu et al. (2014) em que destacaram o uso de critério da distância entre os centroides A_i e A_j na construção da matriz W .

Chen (2013) mostra o uso das distâncias entre os centroides ao utilizar a função peso na construção da matriz W , nos seguintes termos: seja $V = (v)_{ij}$, $i, j = 1, \dots, n$ uma matriz de proximidade espacial que resume as relações espaciais entre n unidades, denominadas por áreas. Assim, cada elemento v_{ij} da matriz, reflete a influência espacial entre a área i sobre j . Seja ainda n elementos em um sistema geográfico, uma matriz de proximidade espacial, a matriz V pode ser expressa e ilustrada na seguinte forma:

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nn} \end{bmatrix} \quad (3)$$

em que v_{ij} é usado para medir o grau de vizinhança ou de contiguidade entre as regiões i e j . Assim, uma matriz de pesos espaciais W pode ser definida por:

$$W = \frac{V}{V_0} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \text{ ou } W^* = \frac{nV}{V_0} = nW \quad (4)$$

em que $V_0 = \sum_{i=1}^n \sum_{j=1}^n v_{ij}$ (CHEN, 2013).

A expressão (4) mostra a dependência da matriz W em relação à matriz V . Portanto, para se determinar a matriz W é necessário construir a matriz V dada em (3). Algumas funções denominadas função peso, podem ser utilizadas na construção da matriz V . Na presente dissertação será utilizada a função inversa, que será abordada na seção 4.

Construída a matriz W e como o propósito é mensurar a dependência espacial entre diferentes áreas na região em estudo, então faz-se necessário conceituar e abordar autocorrelação espacial.

2.3 Autocorrelação espacial

Segundo Bailey e Gatrell (1995), a autocorrelação espacial refere-se à correlação espacial entre atributos de uma mesma variável aleatória em diferentes localizações do espaço. A autocorrelação espacial quantifica a dependência espacial. Para analisar a dependência espacial dos valores da área observada será utilizado o índice de Moran, que é uma generalização do coeficiente de correlação de Pearson.

2.4 Índice de Moran global

O índice de Moran global (I) permite identificar a estrutura de correlação espacial e mostra como os valores estão correlacionados no espaço.

Basicamente, o índice I responde a seguinte pergunta: existe algum padrão espacial nos dados analisados na região R para o fenômeno em estudo?

Duas abordagens serão utilizadas nesta dissertação: a mais tradicional e pelo método de três passos conforme se detalharão nas seções 2.4.1 e 2.4.2.

2.4.1 Método tradicional

Para o cálculo do índice de Moran global, comumente é utilizada a expressão:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (5)$$

em que $I \in [-1,1]$ é o índice de Moran global; \bar{x} representa a média da variável X ; w_{ij} são os elementos da matriz W e $S_0 = \sum_i \sum_j w_{ij}$ é a soma dos elementos da matriz W (que pode ser construída de diversas formas por exemplo, uma matriz binária ou uma matriz baseada em distâncias entre os centroides); e x_i e x_j são valores do atributo nas áreas i e j respectivamente, em que $i, j = 1, 2, \dots, n$ (WONG e LEE, 2005; FU et al., 2014).

A estimativa do índice de Moran pode ser generalizada ao se considerar a autocorrelação de diferentes ordens espaciais (lags). Conhecida a matriz de pesos espaciais de ordem k , $W^{(k)}$, em que $k = 1, 2, \dots, n$, o índice de Moran é então calculado na ordem $I^{(k)}$ (BAILEY e GATRELL, 1995).

2.4.2 Método de três passos

Segundo Chen (2013), o cálculo do índice de Moran pelo método de três passos é o mais simples de todos. Para a implementação do método, algumas transformações e operações algébricas na estrutura da fórmula tradicional (5) são necessárias, como se segue: seja X uma variável definida da seguinte forma $X = [x_1 \ x_2 \ \dots \ x_n]^T$, em que x_i representa o elemento na ordem $i = 1, 2, \dots, n$. A média de X é dada por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (6)$$

A variância, S^2 , é dada por:

$$S^2 = \frac{1}{n} \sum_{n=1}^n (x_i - \bar{x})^2 = \frac{1}{n} (X - \bar{X})^T (X - \bar{X}) = \frac{1}{n} y^T y. \quad (7)$$

Para mais detalhes pode-se consultar também Rancher e Schaalje (2008).

O método de três passos consiste em:

Primeiro passo: padronizar a variável X . Em outras palavras, converter o vector inicial no vector padronizado Z que é o z-score na estatística:

$$Z = \frac{X - \bar{X}}{S}; \quad (8)$$

em que Z é a variável X padronizada, \bar{X} e S é a média e o desvio padrão amostral respectivamente. Por outro lado, sabe-se também que a variável centralizada pode ser calculada por: $Y = X - \bar{X}$. O resultado da transformação escalar da variável centralizada forma um vector centralizado em que $Z = Y/S$.

Segundo passo: construção da matriz $W = [w_{ij}]_{n \times n}$ que satisfaça as três seguintes propriedades conforme Wong e Lee (2005):

- a) Simétrica, isto é, $w_{ij} = w_{ji}$;
- b) Os elementos da diagonal são iguais à zero;
- c) $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = 1$

Sabe-se da expressão (4) que $W = V/V_0$. Portanto W depende de se encontrar a matriz de contiguidade espacial V . Por outro lado, a matriz V pode ser definida a partir da matriz da distância espacial M , que será abordado na seção 3.

Terceiro passo: calcular o índice de Moran usando a seguinte expressão:

$$I = z^T W z \quad (9)$$

em que z^T é o vector transposto de Z , e W é a matriz de pesos espaciais, construída mediante uma função peso e definição de distância. Expandindo a expressão (9) toma a seguinte estrutura:

$$I = \frac{y^T (nW) y}{y^T y}. \quad (10)$$

Desenvolvendo a expressão (10) segue que:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n v_{ij} (x_i - \mu)(x_j - \mu)}{\sum_{i=1}^n \sum_{j=1}^n v_{ij} \sum_{i=1}^n (x_i - \mu)^2}; \quad (11)$$

em que v_{ij} são os elementos da matriz de proximidade espacial com igual peso para todas as áreas, mediante o critério de fronteira comum. Portanto, a expressão (11) é a fórmula tradicional para o cálculo do índice de Moran global, apresentado em (5).

2.5 Índice de Moran local

Para Anselin (1995), o índice de Moran local (I_l) é uma ferramenta estatística que serve para testar a autocorrelação espacial local entre a área A_i e seus vizinhos A_j , cujo objetivo é identificar a presença de agrupamentos de áreas semelhantes (clusters), áreas fora do padrão da região (outliers) e de regimes espaciais que, de alguma forma, não são detectados pelo índice de Moran global.

Tradicionalmente, o cálculo do índice de Moran local pode ser expresso baseado na seguinte fórmula (12) matemática:

$$I_l = \frac{x_i - \bar{x}}{S} \sum_j^n w_{ij} \frac{(x_j - \bar{x})}{S}; S^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n} \quad (12)$$

em que \bar{x} representa a média dos valores observados da variável X , x_i é o valor da variável na posição i ; x_j representa o valor da observação em outra posição (em que $j \neq i$); S^2 é a variância de X e; w_{ij} representa os elementos da matriz W (e depende do critério usado na sua construção) (ANSELIN, 1995).

Por outro lado, pelo método de três passos, Chen (2013) mostra que a partir da equação para o cálculo do índice de Moran global, por meio de algumas operações matemáticas e algébricas, a diagonal da matriz M^* (expressão 17), fornece os indicadores locais de associação espacial (LISA).

Segundo Chen (2013), pode-se mostrar que, ao se multiplicar ambos os membros por Z na equação (9) para o cálculo do índice de Moran global, obtêm-se a expressão (13):

$$M^*Z = ZZ^T WZ = IZ \quad (13)$$

em que Z é o autovetor ou vetor característica da matriz M^* e o índice de Moran global corresponde ao autovalor máximo, ou seja, raiz característica em termos do valor absoluto. A norma de Z , $\|Z\|$, é igual a n , ou seja:

$$\|z\| = z^T z = \sum_{i=1}^n z_i^2 = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 = \frac{n}{\sigma^2} \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = n. \quad (14)$$

Padronizando Z pela (8) conduz a z/\sqrt{n} . Portanto, da expressão (13) segue que:

$$M^* = zz^T W. \quad (15)$$

Expandindo M^* obtêm-se a seguinte expressão matricial:

$$M^* = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \quad (16)$$

Resolvendo (16) obtêm-se:

$$M^* = \begin{bmatrix} z_1 \sum_{j=1}^n w_{j1} z_j, & z_1 \sum_{j=1}^n w_{j2} z_j, & \cdots & z_1 \sum_{j=1}^n w_{jn} z_j, \\ z_2 \sum_{j=1}^n w_{j2} z_j, & z_2 \sum_{j=1}^n w_{j2} z_j, & \cdots & z_2 \sum_{j=1}^n w_{j2} z_j, \\ \vdots & \vdots & \ddots & \vdots \\ z_n \sum_{j=1}^n w_{jn} z_j, & z_n \sum_{j=1}^n w_{jn} z_j, & \cdots & z_n \sum_{j=1}^n w_{jn} z_j, \end{bmatrix}; \quad (17)$$

o que mostra de (17) que o índice de Moran local é:

$$I_l = \frac{ny_i \sum_{j=1}^n w_{ij} y_j}{y^T y};$$

$$I_l = z_i \sum_{j=1}^n w_{ij} z_j; \quad (18)$$

em que I_l são os valores na diagonal da matriz M^* , e sustenta a equação (12) dada por Anselin (1996).

2.6 Média móvel local

Segundo Bailey e Gatrell (1995) um método simples para estimar a variabilidade global e a tendência dos valores de um atributo na região em estudo, é estimar μ_i pelos valores das médias das áreas vizinhas. Para o cálculo da média móvel local (KREMPI, 2004) é necessário conhecer o vetor dos desvios Z obtido a partir da média geral, μ , ou seja:

$$z_i = y_i - \mu; \quad (19)$$

em que z_i é o vetor dos desvios; y_i vetor com os valores do atributo analisado e; μ é a média geral. A estimativa da média móvel local é obtida por:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} z_j; \quad (20)$$

em que $\hat{\mu}_i$ são as médias ponderadas na ordem $i = 1, \dots, n$; w_{ij} elementos da matriz W (construída por um critério previamente definido) e; z_i é o valor do atributo, relativo à i -ésima área (RAMOS, 2002). Druck et al.(2004), mencionam que a média móvel local na análise de dados de área possibilita a diminuição da variabilidade espacial, porque tende a produzir uma superfície com menor flutuação que os dados originais.

Uma vez descritos os métodos para o calculo do índice de Moran global e o local respectivamente, é necessário avaliar a significância estatística desses índices, assunto que será abordado na seção 2.7.

2.7 Testes para avaliar a significância estatística dos índices

Para se avaliar a significância estatística dos índices de Moran global e o índice de Moran local, dois métodos são propostos por Bailey e Gatrell (1995): o teste de distribuição aproximada e o teste de permutação aleatória.

2.7.1 Teste de distribuição aproximada

Se o tamanho da amostra tender para o infinito ($n \rightarrow \infty$), então usa-se um resultado aproximado para a distribuição do índice I sob certas condições para produzir um teste. Assumindo que Y_1, \dots, Y_n são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal, então pelo teorema central do limite, I possui uma distribuição aproximadamente normal, com média $E(I)$ e a variância $Var(I)$ (BAILEY e GATRELL, 1995). O teste estatístico segundo Cliff e Ord (1981) é conforme segue:

$$Z(I) = \frac{(I - E(I))}{\sqrt{Var(I)}}; \quad (21)$$

em que I é o valor calculado do índice de Moran global; Z corresponde a variável aleatória padronizada. A média $E(I)$ e a variância $Var(I)$ são dadas por:

$$E(I) = -\frac{1}{(n-1)} \quad (22)$$

$$\text{Var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2S_0^2}. \quad (23)$$

De fato, a expressão (22) para o valor esperado do I pode ser demonstrada segundo Zhang e Lin (2009) conforme segue: seja uma amostra aleatória X_1, X_2, \dots, X_n de $Z(I) \sim N(0,1)$ em que μ e σ^2 são desconhecidos. Seja ainda $Z_i = X_i - \bar{X}$; então $E(Z_i) = 0$ e $E(Z_i^2) = (1 - 1/n)\sigma^2$. Se X_1, \dots, X_n são independentes e identicamente distribuídos com distribuição normal, então qualquer função escalar de X_1, \dots, X_n é independente e de forma quadrática $Q = \sum_{i=1}^n (X_i - \bar{X})^2$. Portanto, $\sum_{i=1}^n (X_i - \bar{X})^2$ é independente de I (KOOPMANS, 1942). Segue então que:

$$\begin{aligned} E \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij} Z_i Z_j \right] &= E \left(\frac{S_0 I}{n} \sum_{i=1}^n Z_i^2 \right) \\ &= \frac{S_0 E(I)}{n} E \left[\sum_{i=1}^n Z_i^2 \right]. \end{aligned} \quad (24)$$

Assim:

$$E(I) = \frac{n}{S_0} \frac{E(\sum_{i=1}^n \sum_{j=1}^n w_{ij} Z_i Z_j)}{E[\sum_{i=1}^n Z_i^2]}, \quad (25)$$

o que resulta na expressão (22).

Seguindo o mesmo raciocínio, pode-se demonstrar a expressão (23), isto é, $\text{Var}(I)$:

$$\begin{aligned} E \left[\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} Z_i Z_j \right)^2 \right] &= \frac{S_0^2}{n^2} E(I^2) E \left[\left(\sum_{i=1}^n Z_i^2 \right)^2 \right] \\ &= \frac{S_0^2}{n^2} (n^2 - 1) E(I^2); \end{aligned} \quad (26)$$

o que implica que,

$$\sigma^4 \left[\frac{S_1(n^2 - 2n + 3)}{n^2} - \frac{(S_2 - 2S_1)(n-3)}{n^2} + \frac{3(S_0^2 - S_2 + S_1)}{n^2} \right] = \frac{S_0^2(n^2 - 1)}{n^2} E(I^2).$$

Portanto,

$$E(I^2) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)}. \quad (27)$$

Sabe-se por outro lado que $Var(I) = E(I^2) - E^2(I)$. Assim:

$$Var(I) = \left(\frac{1}{S_0^2 (n^2 - 1)} (n^2 S_1 - n S_2 + 3 S_0^2) \right) - E(I)^2; \quad (28)$$

em que $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$; $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$ e; $S_2 = \sum_{i=1}^n (w_{i.} + w_{.i})^2$ em que $w_{i.} = \sum_{j=1}^n w_{ij}$ é a soma da linha i da matriz W e $w_{.i} = \sum_{j=1}^n w_{ji}$ é a soma da coluna i da matriz W (CLIFF e ORD, 1981). Sob a hipótese nula de que não existe a autocorrelação espacial ($H_0: I = 0$), rejeita-se a hipótese nula se valor- $p < \alpha$.

A semelhança da análise de significância pelo teste de distribuição aproximada do índice de Moran global pode-se testar a significância estatística para o índice de Moran local, I_l , visto em Anselin (1995, p.99).

2.7.2 Teste de permutação aleatória

Suponha-se que se tenha n valores x_i com $i = 1, 2, \dots, n$, relacionados com as áreas A_i . Então, tem-se $n!$ possíveis permutações deste mapa, cada um correspondendo a um rearranjo diferente dos n valores x_i sobre as áreas. Uma destas permutações corresponde aos dados que foram observados. O valor de I pode ser calculado para qualquer uma destas $n!$, e pode-se obter a distribuição empírica para os possíveis valores de I sobre as permutações aleatórias dos n valores. Se o valor de I correspondente à configuração observada for valor extremo na distribuição encontrada, então se pode interpretar isso como evidência de autocorrelação espacial (BAILEY E GATREL, 1995).

Silva (2010) incluiu em seu trabalho um algoritmo em que mostra a execução do teste de permutação aleatória como se segue:

- a) Calcula-se o índice I com os dados observados no mapa obtendo $I(0)$;
- b) Para k variando de 1 a N (N é um número menor ou igual à $n!$), repete-se independentemente: Calcula-se o índice $I(k)$ permutando aleatoriamente; os valores de y_i entre as áreas. A limitação em N se deve ao aspecto de viabilidade computacional, já que, se n é grande $n! > n$, não ocorrendo se $n = 1$.
- c) Sob H_0 as variáveis aleatórias Y_i são independentes e identicamente distribuídas. Então todas as permutações dos valores y_i entre as áreas são igualmente

prováveis. Assim, o *valor - p* do teste considerando apenas autocorrelação positiva é:

$$\text{valor} - p = \frac{\text{número de } I_{(j)} > I_{(0)}, j = 1, \dots, N.}{N + 1} \quad (29)$$

Cliff e Ord (1981) mostraram que os momentos de I sob permutação dos índices podem ser obtidos teoricamente e que a distribuição assintótica de I ainda é normal sob determinadas hipóteses bem gerais, em que:

$$E(I) = -\frac{1}{n-1}; \quad (30)$$

$$V(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - k[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} - \left(-\frac{1}{n-1}\right)^2. \quad (31)$$

Zhang (2013)¹ apresenta a demonstração da esperança matemática e da variância de I pelo método dos momentos. Assim, seja $z_i = x_i - \bar{x}$ e y_{ij} dado por $z_i z_j = (x_i - \bar{x})(x_j - \bar{x})$. Ainda, seja $b_k = \sum_{i=1}^n z_i^k / n$ e $\rho_k = b_k / b_2^{k/2}$. Desde que o denominador do índice I seja uma invariante, é suficiente analisar os momentos de permutação de:

$$I_1 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} Y_{ij}; \quad (32)$$

em que: I_1 é primeiro momento Y_{ij} é a variável aleatória sob uma permutação aleatória correspondente y_{ij} . Seja ainda $E_R(\cdot)$ o valor esperado de uma permutação aleatória. Posto isto, segue que:

$$\begin{aligned} E_R(I_1) &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} E_R(Y_{ij}) = \\ &= \frac{S_0}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_{ij}. \end{aligned} \quad (33)$$

Portanto:

¹ Zhang, T. **Permutation test for spatial clustering**. Purdue Science, 2013. Não publicado.

$$E_R(I_1) = \frac{S_0}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j = -\frac{S_0}{n(n-1)} \sum_{i=1}^n z_i^2. \quad (34)$$

Finalmente, chega-se a expressão (30).

Do mesmo modo, pode-se chegar a expressão (31), sobre a variância, $Var(I)$, bastando calcular o momento da segunda ordem e usar (35):

$$Var(I) = E(I^2) - (E(I))^2. \quad (35)$$

Embora não se coloque a demonstração para a $Var(I)$ na presente dissertação, Zhang e Lin (2009) mostraram que o momento da segunda ordem é conforme a expressão (36):

$$E_R(I_1^2) = \frac{S_1(n - \rho_4)}{S_0^2(n - 1)} + \frac{(S_2 - 2S_1)(2\rho_4 - n)}{S_0^2(n - 1)(n - 2)} + \frac{(S_0^2 - S_2 + S_1)(3n - 6\rho_4)}{S_0^2(n - 1)(n - 2)(n - 3)}. \quad (36)$$

Portanto, segue que $Var(I) = E(I^2) - (E(I))^2$ em que $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$; $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (w_{ij} + w_{ji})^2$, $S_2 = \sum_i^n x [\sum_{j=1, j \neq i}^n (w_{ij} + w_{ji})]^2$. Assumindo que o índice de Moran global tem uma distribuição aproximadamente normal, pelo teorema central do limite, o teste estatístico é:

$$Z(I) = \frac{(I - E(I))}{\sqrt{Var(I)}}. \quad (37)$$

Rejeita-se a hipótese nula de não existência da autocorrelação espacial se o *valor - p* $< \alpha$. Para se testar a significância estatística do I_1 procede-se de igual forma a semelhança do índice de Moran global, conforme mostra em detalhes Anselin (1995).

2.8 Gráfico de espalhamento de Moran versus gráfico de dispersão

Um gráfico tradicionalmente usado na análise de dados de área é o gráfico de espalhamento de Moran. Segundo Anselin (1996) o gráfico de espalhamento de Moran tem por objetivo comparar os desvios do atributo numa área A_i com a média ponderada dos desvios de seus vizinhos A_j pela matriz de proximidade espacial W . Este autor mostrou que:

$$I = \frac{\left(\frac{N}{S_0}\right) y' W y}{y' y} = \frac{y' W y}{y' y}; \text{ se } S_0 = N; \quad (38)$$

em que N representa o número de observações; $S_0 = \sum_{i=1} \sum_{j=1} w_{ij}$ é a soma dos elementos da matriz W (previamente definida); y representa os desvios das observações em relação a sua média. Portanto, o I de Moran é equivalente ao coeficiente de regressão numa regressão de $W y$ sobre y (e não y sobre $W y$).

Para Druck et al. (2004), a ideia implícita ao gráfico de espalhamento de Moran é a construção de um gráfico bidimensional de Z (valores da variável padronizada) por WZ (média dos vizinhos), que é dividido em quatro quadrantes. A Figura 2 apresenta essa ideia trazida pelos autores, em que mostra a estrutura do gráfico de espalhamento de Moran representado pelos quadrantes Q1, Q2, Q3 e Q4 que será detalhado na seção 3.

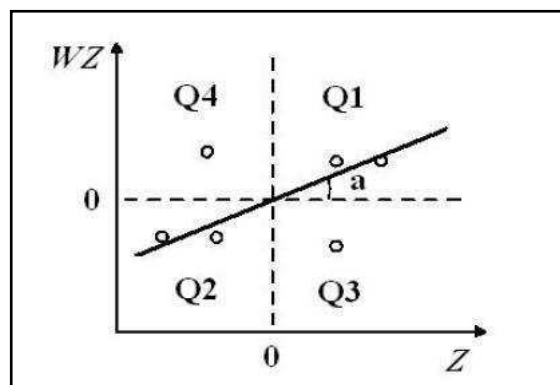


Figura 2: Gráfico de espalhamento de Moran.
Fonte: (Druck et al., 2004).

3 METODOLOGIA

Visando atingir os objetivos propostos neste trabalho, apresenta-se os métodos utilizados para o cálculo do índice de Moran global e local respectivamente, o que envolverá por um lado, a construção da matriz W a partir do critério de fronteira comum e distância entre os centroides; por outro lado, a construção das matrizes M e V respectivamente, para se implementar o método de três passos. E, finalmente, apresentam-se os testes para análise de significância estatística do índice.

3.1 Descrição da variável

Os dados utilizados foram obtidos do Inquérito Demográfico e de Saúde de Moçambique de 2011 (IDS) e referem-se a casos de malária em 132 distritos, 11 cidades e três ilhas do país, disponibilizados pela The DHS Program (DHS). As áreas na região em estudo foram denominadas por distritos. Na análise estatística foi utilizada a taxa média de malária. Uma vez que os distritos são formados por localidades e vilas, a taxa média de malária foi calculada pelo somatório de todos os casos de malária naquele distrito incluindo as vilas e localidades e dividido pelo número total de casos registrados nos 146 distritos. O número total observações analisadas correspondem a 4030 casos. Os dados estão no sistema de coordenadas planas, utilizando a projeção cartográfica policônica.

Malária em Moçambique

O Programa Nacional de Controle da Malária (PNCM) é um programa do governo de Moçambique através do Ministério da Saúde (MISAU) na luta contra a doença. Segundo PNCM (2012), a malária é tida como endêmica em todo o país, variando de zonas hiperendêmicas ao longo do litoral, zonas meso-endêmicas nas terras planas do interior e de algumas zonas hipoendêmicas nas terras altas do interior. A incidência e a taxa de mortalidade causada por malária indicam que a doença é variável nas três zonas do país e em cada distrito. O Inquérito de Indicadores da Malária (IIM, 2007) aponta como um dos principais vetores da malária, o parasita plasmódio falciparum, sendo responsável por mais de 90% de todas as infecções maláricas. Embora se registre uma diminuição da taxa de letalidade, ela é uma das principais causas de morbimortalidade. O relatório da Revisão do Programa da Malária (RPM, 2010), menciona que, o peso da doença no país é enorme, sendo responsável por 45 % de todos

os casos observados nas consultas externas nos hospitais e centros de saúde, e responsável ainda por mais de metade das internações nas enfermarias de pediatria.

3.2 Matriz de pesos espaciais (W)

A estratégia utilizada na construção da matriz W foi baseada em duas abordagens: a primeira, a tradicional, para aplicação direta da expressão (5) utilizando os recursos disponíveis pelo software estatísticos INPE-DPI (2012) e o ESRI (2012), pelo critério de fronteira comum, em que $w_{ij} = 1$ se A_i tem fronteira comum com A_j , e $w_{ij} = 0$ caso contrário; e pelo critério de distância entre os centroides das áreas nas posições i e j . A distância entre A_i e A_j foi calculada utilizando a distância euclidiana. A segunda abordagem foi pelo método de três passos. A matriz de distância espacial M_{ij} foi criada usando a definição da distância euclidiana, posteriormente convertida para uma matriz de contiguidade espacial V , por uma função inversa (39) dada por:

$$v_{ij} = \begin{cases} r_{ij}^{-1}, & i \neq j \\ 0, & i = j \end{cases} \quad (39)$$

em que r_{ij} é a distância do ponto médio central entre a região i e j . Para os 146 distritos da região em estudo em um sistema cartográfico, a matriz V , foi expressa como $V = [v_{ij}]_{146 \times 146}$. Portanto a matriz W foi criada utilizando (4) $W = V/V_0$, $w_{ij} = v_{ij} / \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ em que $V_0 = \sum_{i=1}^n \sum_{j=1}^n v_{ij}$.

3.3 Análise exploratória

A análise exploratória foi feita mediante o cálculo dos índices de Moran global e local (LISA) e análise de significância estatística; pelo método da média móvel local; recorreu-se as ferramentas visuais como mapas temáticos nomeadamente o BoxMap e LisaMap e o gráfico de dispersão de Moran.

3.3.1 Índice de Moran global

Para o cálculo do índice de Moran:

- a) Pelo método tradicional utilizando a expressão (5):

$$I = \frac{n}{S_0} \frac{\sum_i^n \sum_j^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

em que $n = 146$ é o número total de áreas na região em estudo; w_{ij} são os elementos da matriz W construída pelo critério de fronteira comum, todas as áreas na região em

estudo com igual peso (foi igualmente calculado o I utilizando o critério de distância entre os centroides) em que $i, j = 1, \dots, 146$; x_i e x_j são os valores observados da variável malária dos distritos A_i e A_j ; \bar{x} é a média da variável X .

b) Pelo método de três passos, utilizando a expressão (9) $I = z^T W z$; em que $Z = (X - \bar{X})/S = Y/S$ é a variável malária padronizada; W é a matriz de pesos espaciais construída utilizando a distância entre os centroides de cada área e função inversa na criação da matriz V , e; $X = [x_1 \ x_2 \ \dots \ x_{146}]^T$ representa a variável malária e x_i o elemento da variável na ordem $i = 1, 2, \dots, 146$. A média e a variância da amostra foram dadas por $\bar{X} = 1/146 \sum_{i=1}^{146} x_i$ e $S^2 = \frac{1}{146} y^T y$ respectivamente.

3.3.2 Índice de Moran local (LISA)

Para o cálculo do índice de Moran local, I_i , recorreu-se a expressão (18) $I_i = z_i \sum_{j=1}^n w_{ij} z_j$; em que z_i e z_j referem-se aos valores padronizados da variável malária nas áreas A_i e A_j respectivamente, em que $i = 1, 2, \dots, 146$; w_{ij} são os elementos da matriz W (criado pelo método tradicional e o método de três passos).

3.4 Teste de Significância Estatística

Na análise da significância estatística pelo teste de distribuição aproximada de I , foi utilizada a expressão (21):

$$Z(I) = \frac{(I - E(I))}{\sqrt{Var(I)}};$$

em que $E(I) = -0,006897$ (expressão 23) e $Var(I) = 0,001570$ (expressão 24). Igualmente, foi feita a avaliação da significância estatística pelo teste de permutação aleatória com o auxílio do software estatístico TerraView 10.1 INPE (2012) permutando 999 vezes, em que $H_0: I = 0$. Rejeita-se a hipótese nula se $valor - p < 0.05$.

Uma vez que $-1 \leq I \leq 1$ representa o quanto cada área analisada é semelhante ou não aos seus vizinhos, então: se $I = 0$ indica uma distribuição aleatória espacial sem nenhum padrão definido. Porém se $I > 0$ indica que há similaridades entre áreas próximas, ou seja, área A_i com valor alto cercado pelos seus vizinhos, A_j 's, com valores altos, ou ainda área A_i com baixo valor cercado por seus vizinhos A_j 's com valores baixos. Se $I < 0$ então existem dissemelhanças entre os valores da área A_i e A_j (TU e XIA, 2008).

3.5 Método de suavização local - média móvel local

Recorreu-se a expressão (20) no cálculo da média móvel local $\hat{\mu}_i$ dada por $\hat{\mu}_i = \sum_{j=1}^n w_{ij}z_j$; em que, w_{ij} são os elementos da matriz W (criada pelo critério de fronteira comum) com $i, j = 1, \dots, 146$ e x_j , o valor observado da variável X . Ao se calcular $\hat{\mu}_i$ dos valores dos vizinhos, a variabilidade espacial é reduzida, uma vez que a operação tende a produzir uma superfície com menor flutuação que os dados originais (DRUCK et al., 2004).

3.6 Diagrama de dispersão de Moran

Criou-se o diagrama de dispersão de Moran utilizando a expressão (40) cujo vetor é definido baseado na matriz W :

$$f = nWz = Z^T ZWZ \quad (40)$$

em que Z é o vector padronizado da variável malária. Na construção do gráfico de dispersão de Moran Z representaria o eixo x e f o eixo y . Assim, as coordenadas (z_i, f_i) representam os valores observados, distribuídos de uma forma irregular (CHEN, 2013). Ainda no gráfico, cada quadrante Q1, Q2, Q3 e Q4 representa uma determinada situação, conforme descritas a seguir, em que as letras A e B representam alto e baixo respectivamente: Q1-AA= distrito com alta taxa média de malária cercado por vizinhos em situação idêntica (semelhanças entre eles); Q2-BB= distrito com baixa taxa média de malária cercado por vizinhos na idêntica situação (semelhanças entre eles); Q3-AB= distrito com alta taxa média de malária e os seus vizinhos em situação contrária (dissemelhanças com os seus vizinhos) e; Q4-BA= distrito com baixa taxa média de malária e os seus vizinhos em situação contrária (dissemelhanças com os seus vizinhos).

3.7 O BoxMap, LisaMap, MoranMap e gráfico de dispersão de Moran

Uma fase essencial na análise de dados espaciais de área é a visualização gráfica. Portanto, os resultados tanto para I como para I_1 foram ilustrados e resumidos em formas de mapas temáticos e gráfico, nomeadamente: BoxMap, MoranMap, LisaMap e gráfico de dispersão de Moran. Para a compilação e análise espacial de dados de área pelo método de três passos, foi utilizado MS Excel. Para as demais análises utilizou-se os softwares TerraView 4.4.2 INPE (2013) e o ArcGis 10.1 (ESRI, 2012).

4 RESULTADOS E DISCUSSÃO

4.1 Abordagem matemática para análise espacial de dados de área

Duas abordagens foram utilizadas na análise exploratória de dados de área pelo índice de Moran: a abordagem mais tradicional e o método de três passos, proposto por Chen (2013).

4.1.1 Índice de Moran global e local- abordagem tradicional

A abordagem tradicional comparada com o método de três passos, para o cálculo do índice de Moran global, basicamente envolve a construção de uma matriz W , e implementação da seguinte equação:

$$I = \frac{n}{S_0} \frac{\sum_i^n \sum_j^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

em que w_{ij} é o elemento da matriz W . Portanto, se uma área na posição i partilha a mesma fronteira com a área na posição j então $w_{ij} = 1$ e, no caso em que não haja essa partilha da fronteira então $w_{ij} = 0$. Igualmente utiliza-se o critério de distância entre os centroides na construção da matriz W e o resultado do índice I comparado com o método de três passos (BAILEY e GATRELL,1995; FU et al.;2014).

Estabelecido o critério para a construção da matriz W , essa matriz pode ser utilizada para outros cálculos. Assim, calcula-se tradicionalmente o índice de Moran local do mesmo modo, utilizando a expressão (12):

$$l_i = \frac{x_i - \bar{x}}{S} \sum_j^n w_{ij} \frac{(x_j - \bar{x})}{S}; S^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n};$$

em que W é a mesma matriz utilizada para o cálculo do índice de Moran global.

4.1.2 Índice de Moran global e local pelo método de três passos

A abordagem pelo método de três passos consistiu em:

- Primeiro passo: padronizar a variável malária X utilizando a expressão (18);
- Segundo passo: construção da matriz $W = [w_{ij}]_{n \times n}$. Conforme observado na seção 3, sua construção depende da definição da distância entre o centroides da área A_i em relação as restantes áreas A_j . Wong e Lee (2005) mostram essa relação numa função comumente usada, a função inversa, do seguinte modo:

$$w_{ij} = \frac{1}{d_{ij}}$$

em que d_{ij} representa a distância entre os centroides nas áreas i e j . Essa relação mostra que as áreas próximas uma das outras tendem a exercer mais influência uma sobre a outra em relação às áreas distantes (DAVID, 1997). Portanto, é comum o uso da função peso para se estabelecer a magnitude da influência de uma área sobre a outra. Chen (2013) mostrou que se deve criar uma matriz denominada matriz de distância espacial $M = [m_{ij}]_{n \times n}$, e em seguida convertê-la em uma matriz V . Essa conversão é feita mediante a função peso. Assim:

$$v_{ij} = \begin{cases} r_{ij}^{-1}, & i \neq j \\ 0, & i = j \end{cases};$$

em que r_{ij} é a distância euclidiana dos centroides entre a região i e j . Portanto a matriz $W = V/V_0$, em que $V_0 = \sum_{i=1}^n \sum_{j=1}^n v_{ij}$ (CHEN, 2013).

• Terceiro passo: cálculo do índice de Moran utilizando a expressão (9), $I = Z^T W Z$. É perceptível a simplificação que o método produz, uma vez que, o princípio por detrás dela envolve a padronização dos elementos que compõem o vetor da variável em estudo num vetor Z e na criação de matriz W como foi descrito anteriormente. Foi ainda demonstrado (20) que $I = Z^T W Z$ é equivalente à formulação tradicional (21), porém variando na sua estrutura e formulação ocasionada pela transformação algébrica (CHEN, 2013). Na abordagem pelo método de três passos, o índice de Moran local é calculado mediante a construção da matriz M^* a partir da expressão $I = Z^T W Z$ (demonstrado na seção 3) em que a matriz W é a mesma utilizada para o cálculo do índice de Moran global. Assim,

$$M^* = \begin{bmatrix} z_1 \sum_{j=1}^n w_{j1} z_j, & z_1 \sum_{j=1}^n w_{j2} z_j, & \dots & z_1 \sum_{j=1}^n w_{jn} z_j, \\ z_2 \sum_{j=1}^n w_{j2} z_j, & z_2 \sum_{j=1}^n w_{j2} z_j, & \dots & z_2 \sum_{j=1}^n w_{j2} z_j, \\ \vdots & \vdots & \ddots & \vdots \\ z_n \sum_{j=1}^n w_{jn} z_j, & z_n \sum_{j=1}^n w_{jn} z_j, & \dots & z_n \sum_{j=1}^n w_{jn} z_j, \end{bmatrix};$$

em que os valores da diagonal da matriz M^* , representam o índice de Moran local (ANSELIN, 1996; CHEN, 2013). No ínterim, percebe-se que a construção da matriz W é influenciada pela escolha quer da definição da distância, quer da função peso (ZHANG et al., 2008).

4.1.3 Gráfico de dispersão de Moran

Para Anselin (1996), o diagrama de dispersão de Moran convencional pode ser criado, se y representar o eixo x e nWy representar o eixo y (na expressão 38). Uma reestruturação trazida pelo Chen (2013), porém com a mesma funcionalidade apresentada por Druck et al.,(2004) foi a criação do gráfico de dispersão de Moran. Ao se recorrer à expressão (40) baseada na matriz de pesos espaciais W , $f = nWZ = Z^T ZWZ$ pode-se construir o gráfico de dispersão de Moran, em que Z é o vetor padronizado da variável malária. Na construção do gráfico de dispersão de Moran Z representará o eixo x , e f representará o eixo y . Portanto, no gráfico construído, as coordenadas (z_i, f_i) representam os valores observados distribuídos de uma forma irregular (CHEN, 2013).

4.2 Aplicação em um caso: variável malária

Nesta seção apresentam-se os resultados da implementação das duas abordagens na análise espacial de dados de área utilizando a variável malária: abordagem tradicional em que a criação da matriz de pesos espaciais será criada utilizando o critério de fronteira comum e de distancia entre os centroides; e abordagem mediante o método de três passos, utilizando a distância entre os centroides na criação da matriz de pesos espaciais.

4.2.1 Descrição da variável malária e média móvel local

Na análise da distribuição da malária nos 146 distritos (Figura 3), foram identificados três distritos com uma alta taxa média de malária. Na região sul foi identificado dois distritos, nomeadamente os distritos representados pelas cidades de Maputo (502) e da Matola (287); e na região central um distrito, a cidade da Beira (116), cujas taxas médias são de 12,46%, 7,12% e 2,88% respectivamente, correspondendo a 22,46% em relação a todo país.

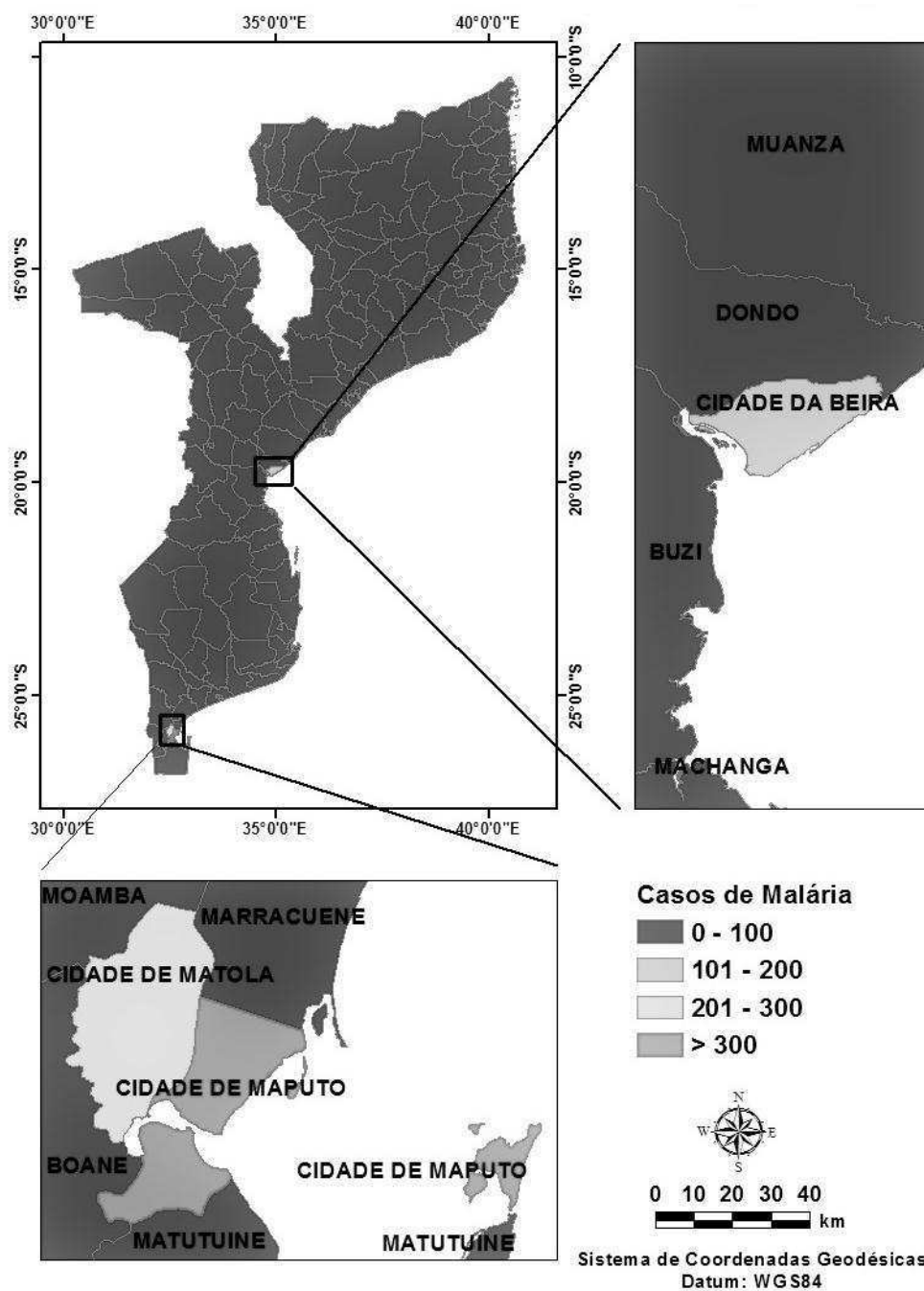


Figura 3: Mapa de distribuição da variável malária em quatro classes distintas nos 146 distritos de Mocambique. Ano base: 2011.

Na Figura 4(a), verificou-se pela divisão em quantil, que houve maior variabilidade na distribuição espacial da taxa média da malária, variando de zero a 12,46%. No entanto, pelo método da média móvel local (Figura 4(b)) essa variabilidade foi reduzida de zero a 5,29%.

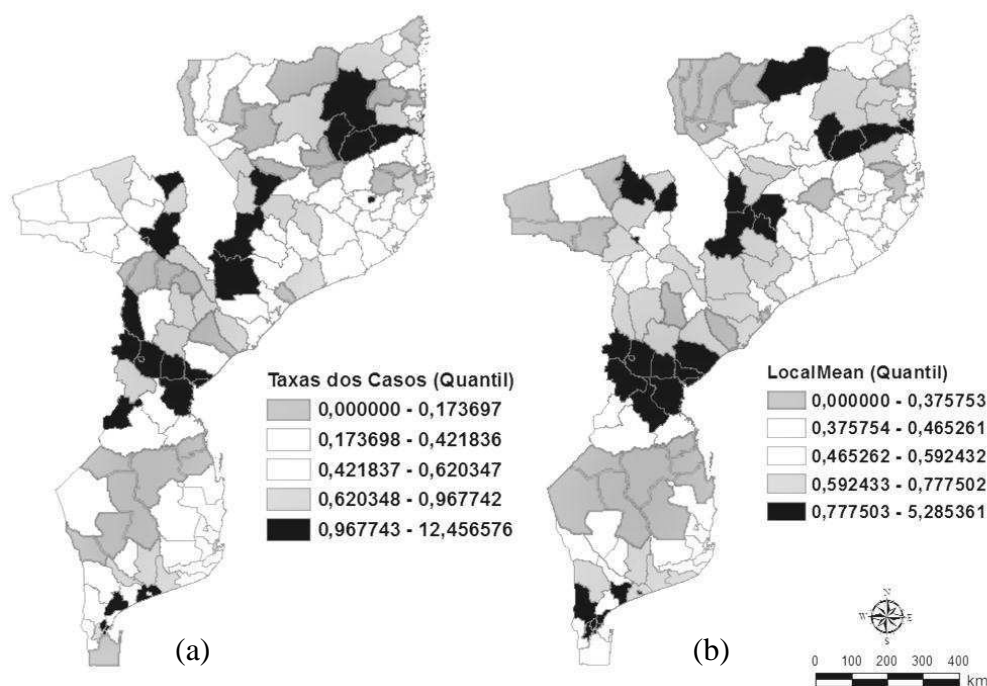


Figura 4: Distribuição da variável malária por quantil (a) e pelo método da média móvel local (b). Ano base: 2011.

Essa redução era de se esperar uma vez que a média móvel local tende a produzir uma superfície com menor flutuação que os dados originais. Nas regiões sul, central e norte do país, foi possível identificar distritos que se assemelhavam (agrupados por cores idênticas), quer por possuírem alta taxa média de malária ou por possuírem baixa taxa média de malária (DRUCK et al.,2004).

4.2.2 Índices de Moran global e local, BoxMap e gráfico de dispersão de Moran

Procedeu-se ao cálculo índice de Moran global. Pelo método tradicional em que a construção da matriz W foi baseada pelo método de fronteira comum e distância entre os centroides $I=0,3129$ (valor- $p=0,006<0,05$) e $I=0,1862$ (valor- $p=0,006<0,05$) respectivamente. Utilizando o método de três passos $I=0,137$ ($Z>1,96$) em que foi usada a função inversa e a definição da distância euclidiana. O teste de distribuição aleatória do índice de Moran global e o teste de permutação aleatória em 999 vezes, levaram a rejeição da hipótese nula, de não existência da autocorrelação na região em estudo. Essa ligeira diferença de resultados não é surpreendente entre os dois métodos uma vez que a construção da matriz de pesos espaciais pelo método de três passos depende da função peso e da definição da distância. Contudo, essa diferença, do ponto de vista da significância estatística traz a mesma conclusão, da existência da autocorrelação positiva na amostra analisada.

A existência da autocorrelação espacial positiva torna-se também evidente ao se analisar tanto o BoxMap (Figura 5) como o diagrama de dispersão de Moran (Figura 6). No BoxMap estão representados distritos em quatro cores distintas, correspondendo aos quadrantes Q1 a Q4 no gráfico de dispersão de Moran. O quadrante Q1, indica distritos que se assemelham, por possuírem valores altos (AA) de taxa média da malária. É o caso dos distritos representando as cidades de Maputo e Matola, com a taxa média de malária conforme anteriormente observado de 12,46% e 7,12% respectivamente; distritos na região central nomeadamente os distritos de Búzi (1,71%), cidade da Beira (2,88%), Dondo (1,44%), Nhamatanda (1,12%), Gondola (1,66%), cidade de Chimoio (1,68%), de Manica (1,29%) e Sussundenga (0,77%). Para o quadrante Q2, refere-se aos distritos agrupados que se assemelham entre eles por possuírem valores baixos (BB) da taxa média de malária. Essa situação é observada nas regiões norte, central e sul do país. Por exemplo, na região sul, o agrupamento formado pelos distritos com a baixa taxa média de malária entre as regiões sul e central do país (Q2, cor cinza-escura) corresponde em conjunto uma taxa média de malária de 6,97% em relação a todo o país. Região sul é considerada como uma das regiões em que a incidência da malária tende a diminuir entre os anos 2008 a 2010 (IIM, 2007; RPM, 2010).

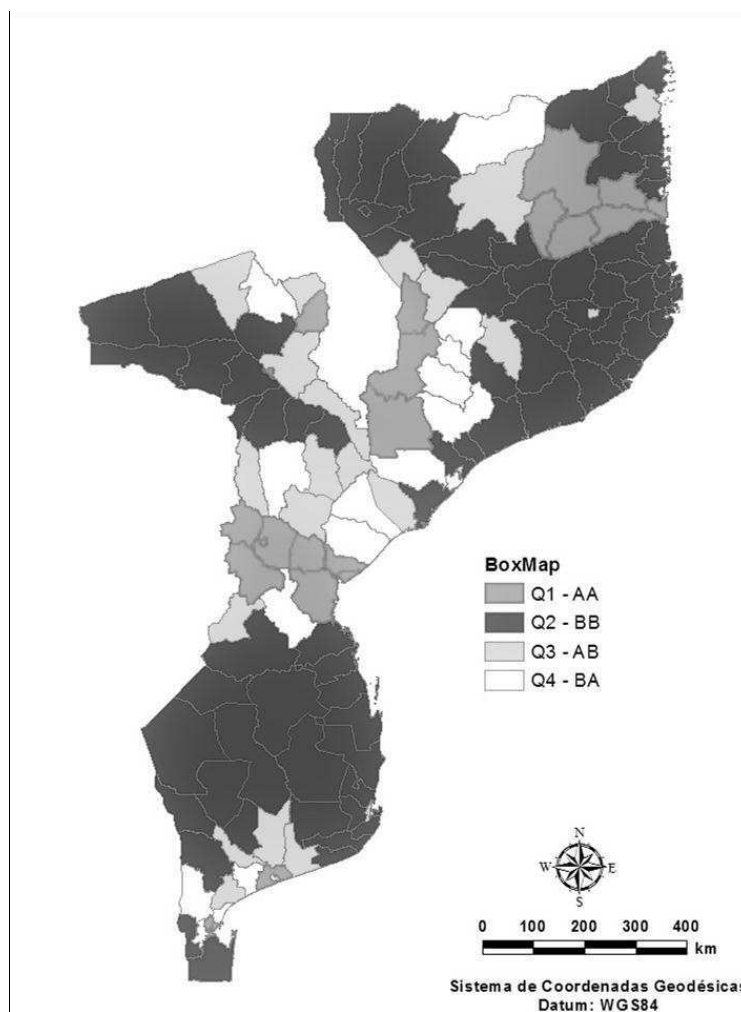


Figura 5: BoxMap da variável malária representando cada distrito conforme sua posição em relação aos quadrantes (Q1,Q2,Q3 e Q4) no gráfico de dispersão de Moran, identificados por uma cor correspondente no mapa gerado. Ano base: 2011.

Um pouco por todo país foi possível identificar alguns distritos com alta taxa média de malária circundados pelos vizinhos com baixa taxa média de malária (AB-quadrante Q4). Na região sul do país, os distritos de Marracuene, Moamba, Boane possuem uma baixa taxa média de malária, porém, os distritos de Maputo e Matola com taxas elevadas; por outro lado o distrito de Bilene com taxa média baixa cujo vizinho, o distrito de Xai-Xai com alta taxa média de malária (BA). Esse fato pode ser devido a deslocamento da população de uma região para hospitais gerais ou centrais localizadas nas capitais da província ou na capital do país, no caso de um deslocamento para a cidade de Maputo ou Matola. De fato, essa conclusão pode ser sustentada ao se analisar os resultados obtidos pela média móvel local (comentado na Figura 4), em que maiores concentrações das taxas médias de malária, tendem a situar-se nas capitais provinciais.

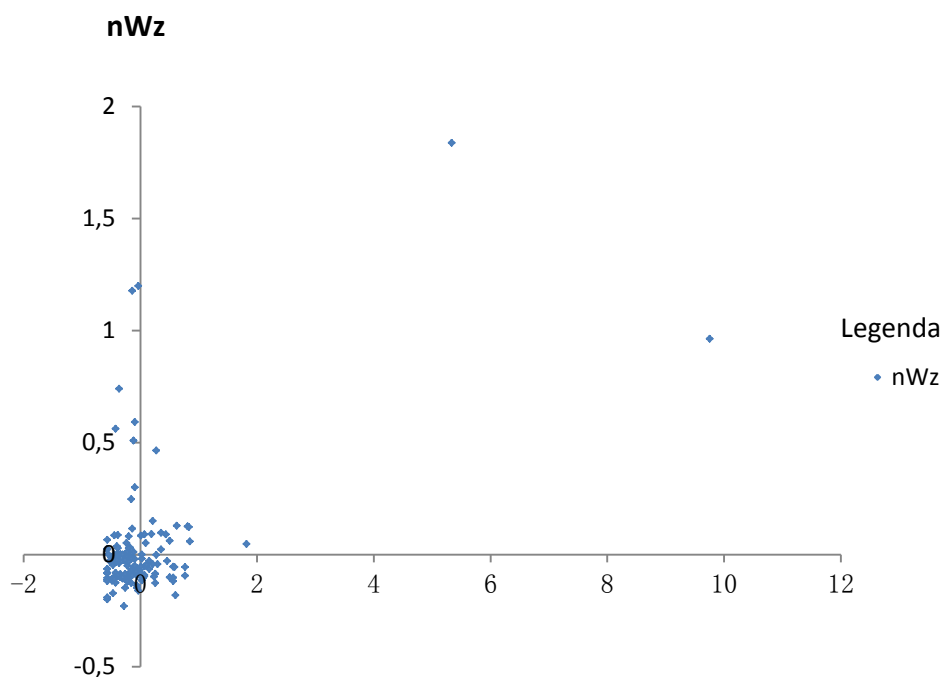


Figura 6: Gráfico de dispersão de Moran.

Na análise do gráfico de dispersão de Moran (Figura 6) foi possível identificar distritos com os valores extremos (outliers) e distritos com os valores de z superiores a dois. É o caso dos distritos representados pelas cidades de Maputo ($Z=9,757883$) e de Matola ($Z=5,335545$). O máximo valor do índice de Moran local pelo teste de permutação aleatória (valor- $p < 0,05$) foi o da cidade de Maputo. Com a remoção dessas observações, a inclinação da reta tende a ser horizontal em relação ao eixo z . Portanto, os distritos representados pelas cidades de Maputo e da Matola são os que exerceram uma maior influência na autocorrelação espacial (HARRIES, 2006; CHEN, 2013).

Na geração do LisaMap (Figura 7) os índices foram classificados em quatro grupos: não significativos (valor- $p > 0,05$), significativos com $\alpha=5\%$, 1% e $0,1\%$ (valor- $p < \alpha$) respectivamente. Na região sul e extremo sul de Moçambique, identificou-se dois agrupamentos formados (valor- $p < 0,05$): o primeiro agrupamento (região sul), fazem parte os distritos de Chicualacuala, Chigubo, Mabote e Funhalouro; o segundo (região extremo sul), os distritos de Moamba, cidade da Matola, Marracuene, Boane e a cidade de Maputo. No MoranMap (Figura 8) fica evidente a natureza desses agrupamentos: no primeiro, são os distritos que se assemelham por possuírem baixas taxas médias de malária (BB) e; no segundo, são distritos que se dissemelham ($I < 0$) por possuírem baixa taxa média da malária, contudo, redor distritos com alta taxa média da malária: cidade de Maputo e da Matola (BA). Verificou-se ainda que os distritos de

Lago Niassa e Lichinga na região noroeste do país, e o distrito de Monapo na região nordeste (valor- $p < 0,05$) houve semelhanças por possuírem a baixa taxa média de malária (BB).

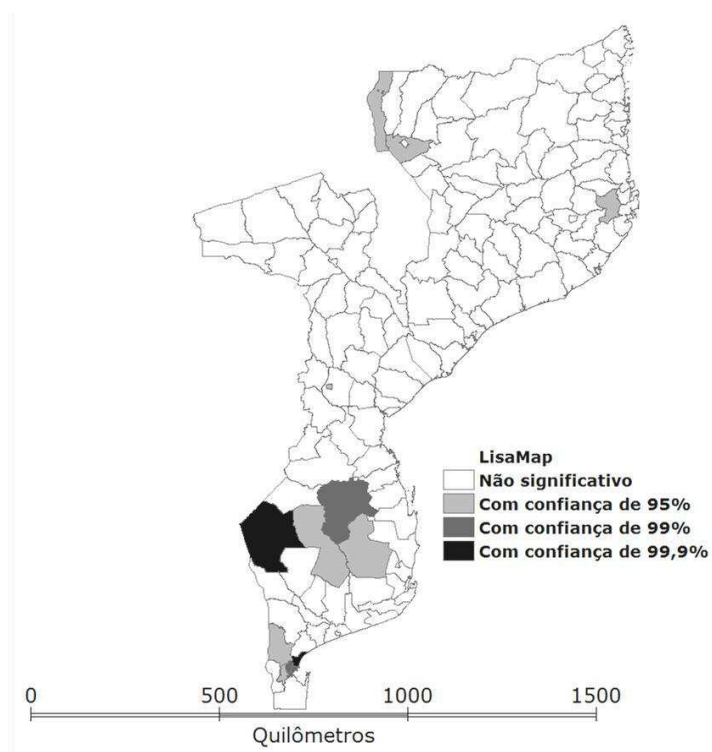


Figura 7: LisaMap da variável malária representando distritos que foram considerados significativos pelo teste de permutação aleatória em 999 vezes, em quatro grupos: não significativos, com 95%, 99% e 99,9% de confiança respectivamente. Ano base: 2011.

Os distritos representados pela cor branca se estendem por quase todo o país, sem evidência significativa de autocorrelação a nível local (valor- $p > 0,05$).

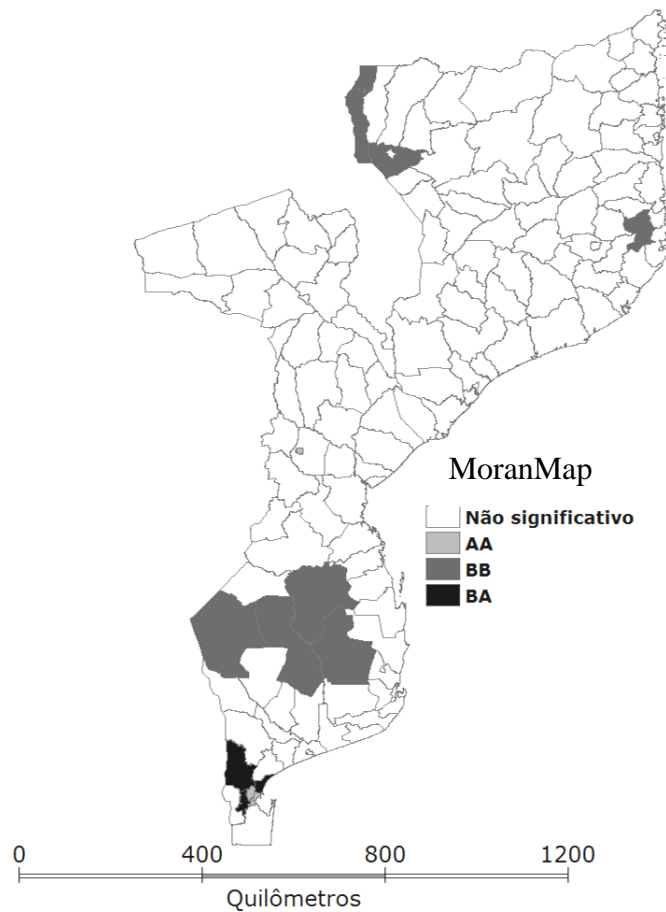


Figura 8: MoranMap da variável malária representando distritos que foram considerados significativos pelo teste de permutação aleatória em 999 vezes, em quatro grupos (AA, BB, AB e BA) conforme sua localização (Q1, Q2, Q3 e Q4) no gráfico de dispersão de Moran. Os demais distritos ficam classificados como não significativos. Ano base: 2011.

A distribuição da malária pode ser devido a um processo aleatório, embora se tenha verificado na análise de BoxMap (Figura 5) e média móvel local (Figura 4) existência de alguns focos de agrupamentos nas regiões central e norte do país (BAILEY e GATRELL, 1995).

5 CONCLUSÕES

O método de três passos revelou ser o mais simples em termos de execução na análise de autocorrelação espacial e trouxe uma grande contribuição na estatística espacial como uma ferramenta de análise espacial de dados de área. Conhecendo-se as coordenadas dos centroides das áreas da região em estudo, bem como o valor do atributo de uma variável associada a essas áreas, pode-se fazer uma análise espacial associado a dados de área, bastando um básico conhecimento do manuseio do MS Excel, cujos resultados mostraram ser estatisticamente iguais, pelo teste de distribuição aproximada do índice de Moran e pelo teste de permutação aleatória, comparando com a abordagem tradicional. Verificou-se uma ligeira diferença em termos de resultados, devendo-se a definição da distância e da função peso, uma vez que a construção da matriz de proximidade espacial depende da escolha dessa função.

Foi possível identificar agrupamentos dos casos da malária na região em estudo pelo método LISA. A região entre sul e extremo sul do país com dois agrupamentos: o primeiro por possuírem semelhanças com a taxa média da malária baixa, composto por distritos de Chicualacuala, Chigubo, Mabote e Funhalouro; o segundo agrupamento no extremo sul por possuírem dissemelhanças, composto por distritos de Moamba, cidade da Matola, Marracuene, Boane e a cidade de Maputo. Na região norte, percebeu-se três distritos, a saber, Lago Niassa, Lichinga e Monapo, semelhantes por possuírem baixas taxa média de malária. Em relação ao restante do país, percebeu-se uma distribuição aleatória dos casos de malária. Portanto, a malária mostrou ser endêmica por todo país, variando de zonas, e pode ser devido fatores não abordados neste trabalho.

Futuramente seria de interesse estudar o comportamento de malária utilizando uma abordagem espaço-temporal pelo método de estatística scan também conhecido por método de varredura, em um intervalo de anos sucessivos, cuja finalidade seria detectar áreas de risco e sua progressão espaço-temporal.

REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L.U.C. **Local indicators of spatial association - LISA**. Geographical analysis, v. 27, p. 93-115, 1995.

ANSELIN, L.U.C. **The Moran scatterplot as an ESDA tool to assess local instability in spatial association**. In: Fisher, M.; Scholten, H.J.; Unwin, D. (Eds). Spatial analytical perspectives on GIS. London: Taylor & Francis, p. 111-125, 1996.

ASSUNÇÃO, R. M. **Estatística espacial com aplicações em epidemiologia, economia, sociologia**. São Carlos, SP: Associação Brasileira de Estatística, 2001. 131p.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex, England: Longman Scientific, 1995. 413 p.

CHEN, Y. New Approaches for calculating Moran's index of spatial autocorrelation. **PLoS ONE** 8(7): e68336. doi:10.1371/journal.pone.0068336, 2013.

CLIFF, A.D.; ORD, J.K. **Spatial processes: models and applications**. London: Pion Limited, 1981. 266p.

CRESSIE, N. A. **Statistics for spatial data**. New York: John Wiley & Sons, 1991. 900p.

DAVID, E. **statistics in geography**. Blackwell Publisherrs Ltd, Oxford, UK, 1997.232p.

DHSprogram. USAID **Demographic and health surveys**, 2011. Disponível em <<http://dhsprogram.com/>> . Acessado em 08 de Janeiro de 2014.

DRUCK,S et al. **Análise espacial de dados geográficos**. Planautina, DF, EMBRAPA Cerrados, 2004. 209 p.

ESRI, 2012. ArcGis desktop:release 10.1 Redlands, CA: Environmental systems research institute. Disponível em: <<http://www.esri.com/>>. Acesso em: 18 de Setembro de 2014.

FU, W. et al. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. **Biogeosciences**, v. 11, p.2401-2409, doi:10.5194/bg-11-2401-2014, 2014. Disponível em: <http://www.biogeosciences.net/11/2401/2014/bg-11-2401-2014.html>. Acesso em 11 de Março de 2015.

HARRIES, K. Extreme spatial variations in crime density in Baltimore County, MD. **Geoforum**, v. 37, p.404-406, 2006.

KOOPMANS, T. Serial correlation and quadratic forms in normal variables. The Annals of Mathematical Statistics, v. 13, p.14-33, 1942. Disponível em: <http://projecteuclid.org/euclid.aoms/1177731639> . Acesso em: 24 de Janeiro de 2015.

KREMPI, A. P. **Explorando recursos de estatística espacial para análise da acessibilidade na cidade de Bauru**. 2004. 82p. Dissertação (Mestrado em engenharia civil - Área: Transportes) - Universidade de São Paulo - USP, SP.

MINISTÉRIO DE SAÚDE. **Inquérito de indicadores da malária-IIM**. Moçambique, 2007. Relatório.

MINISTÉRIO DE SAÚDE. **Programa nacional de controlo da malária**. Plano estratégico da malária 2012 – 2016. Moçambique, 2012. Disponível em: <<http://www.rollbackmalaria.org/files/files/countries/mozambique2012-2016.pdf>>. Acesso em: 13 de setembro de 2015.

MINISTÉRIO DE SAÚDE. **Revisão do programa da malária-RPM**. Moçambique, 2010. Relatório.

RAMOS, F. R. **Análise espacial de estruturas intra-urbanas: o caso de São Paulo**. 139p. Dissertação (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais-INPE, São José dos Campos, 2002.

RANCHER, A.C.; SCHAALJE, G.B. *Linear models in statistics*. 2. ed. New Jersey: John Wiley & Sons, 2008. 672p.

SANTOS, G.R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese (Doutorado)-Universidade Federal de Lavras. Departamento de Estatística, Lavras, 2010, 200p.

SILVA, N.C.N. **Análise de dados de área aplicada a dois indicadores económicos de mesorregiões do estado de Minas Gerais** 85p. Dissertação (Mestrado em Estatística). Departamento de Estatística da Universidade Federal de Lavras. Lavras – MG, 2010.

INPE-DPI. O aplicativo TerraView, 2012. Disponível em: www.dpi.inpe.br/terraview>. Acessado em 18 de janeiro de 2013.

TU, J.; XIA, Z. G. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation, **Science of the Total Environment**, v. 407, p.358–378, 2008.

WONG, D.; LEE, J. **Statistical analysis of geographic information with ArcView and ArcGis**. Wiley, New Jersey, 2005, 341p.

ZHANG, C. et al. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. **Science of the Total Environment**, v. 398, p.212–221, 2008.

ZHANG, T; LIN, G. Cluster detection based on spatial associations and iterated residuals in generalized linear mixed models. **Biometrics**, v. 65, p.353–360, 2009.