ANÁLISE DA DISTRIBUIÇÃO ESPACIAL DE ENDEMIAS A PARTIR DE FERRAMENTAS DE CEQ

Ronei Marcos de Moraes[1]

Abstract *— Epidemiology provides different ways to know when, for a geographic region, a specific disease is endemic or epidemic. However, in the scientific literature could not be found a methodology for characterization of subregions of a geographic region as epidemic or endemic ones. This spatial characterization is useful for decision making by managers about emergency policies for specific locations. Different providences should be taken for locations under disease surge in relation to other under disease epidemic. It can provide cost savings, and unnecessary travel plan. This knowledge, combined with other information specific to each locality, also allow managers to understand their individual epidemiological characteristics for prioritization of different arrangements. In this paper is proposed a new methodology for mapping information from data of non-parametric Quality Control Charts to visualization of endemic control in the several subregions of a geographic region.*

Index Terms — Epidemiology, Quality Control Charts, Spatial Distribuition, Choropleth Maps.

INTRODUÇÃO

A vigilância epidemiológica tem como objetivo reduzir o número de casos de uma doença e prevenir a ocorrência de epidemias ou surtos [18]. A epidemia se caracteriza pelo aumento, fora dos padrões frequenciais esperados, do número de casos de uma determinada doença em uma determinada região geográfica e período de tempo. O surto se caracteriza como um tipo de epidemia que é restrita a uma subregião geográfica delimitada e/ou uma população identificável [4]. Para cumprir esses objetivos, é necessário realizar um acompanhamento da endemicidade dessa doença, identificando períodos e regiões prioritárias ao seu combate. Uma ferramenta recomendada pelo Sistema de Vigilância Epidemiológica é o Diagrama de Controle, que possibilita o cálculo mensal do nível endêmico, auxiliando no controle epidemiológico [4]. Entretanto, apesar de simples e útil, a utilização desse Diagrama de Controle esbarra em premissas difíceis de serem cumpridas, como por exemplo: em vários tipos de doenças, a exclusão do período epidêmico faz com que o número de observações remanescentes seja baixo, inferior aos 10 anos preconizados pelo Guia de Vigilância Epidemiológica [4]; o uso da média e desvio-padrão é recomendado quando a distribuições estatística dos dados é normal ou pelo menos que a distribuição seja simétrica em relação à sua média. O próprio Guia faz menção a essa questão [4, pg. 56].

Bortman [3] considera que o período de tempo de 5 a 7 anos seja suficiente para a construção de corredores endêmicos válidos. Segundo esse autor, o aumento do número de anos pode misturar condições de diagnóstico e de notificações diferenciados, porém podem prover melhorias no modelo estatístico. É também razoável de se esperar que as condições sócio-econômicas da população sejam diferenciadas ao longo de um tempo maior e que essas variações também contribuam significativamente para alterações nos quadros de evolução da doença e o próprio esclarecimento da população sobre ela levem a tomada de decisões próprias ou mesmo a automedicação, sem que os órgãos de saúde sejam notificados. Exemplo disso é a notificação de doenças transmissíveis por vetor, como é o caso do dengue, onde a população não tem procurado os serviços de saúde nos casos mais brandos.

A questão da normalidade dos dados, eventualmente pode ser contornada mediante o uso de transformações estatísticas, como a transformações da família Box-Cox ou logarítmica, raíz enésima, entre outras [23]. Como alternativas aos Corredores endêmicos, também podem ser utilizadas adaptações dos métodos de Controle Estatístico de Qualidade na epidemiologia [20], como a Carta de Controle proposta por Shewhart em 1931 [13], para realizar este acompanhamento epidemiológico. Estas ferramentas possibilitam identificar situações epidêmicas e variações nos níveis de endemicidade. Neste contexto, alternativas podem ser utilizadas como o controle robusto [17] ou o Gráfico da Média Móvel Exponencialmente Ponderada (EWMA) proposto por Roberts e 1959 [10] ou ainda o Gráfico de Somas Acumuladas, propostas por Page em 1959 (CUSUM) [13], sendo esses últimos mais eficientes para a detecção de pequenas variações no processo epidemiológico do que as Cartas de Shewhart. Mesmo assim, Bortman [3] e Sanches [19] concordam com a utilização do Gráfico de Somas Acumuladas para a elaboração de corredores endêmicos.

Muitas vezes as transformações estatísticas não são capazes de resolver a questão da normalidade dos dados. Para solucionar isso, é possível utilizar métodos alternativos. Em geral, os métodos não-paramétricos [21] se constituem uma boa alternativa aos métodos paramétricos. Várias metodologias foram propostas para solução de problemas específicos. Chakraborti et al. [5] propuseram um gráfico de controle denominado precedente, usando um teste para

[1] Ronei Marcos de Moraes, Department of Statistics, Federal University of Paraíba - CCEN, Cidade Universitária s/n, João Pessoa – PB, Brazil, ronei@de.ufpb.br

mediana e os limites calculados sobre a distribuição hipergeométrica. Albers et al. [2] propôs o uso de um gráfico de controle combinado a partir de três outros e mostrou que é possível se adequar a uma variedade de distribuições estatísticas. Albers e Kallenberg [1] usaram o valor mínimo de grupos de observações comparado ao limite superior de controle e com isso é possível obter um poder de comparação comparável aos gráficos baseados em média. Zhou et al. [27] propuseram um gráfico de controle baseado no teste estatístico de Mann-Whitney que poderia ser usado em estágios iniciais de controle. Versões não-paramétricas dos Gráficos CUSUM [25] e EWMA [26] também foram propostas. Dentre as várias propostas para o controle multivariado Qiu [16] propôs um método não-paramétrico baseado em Somas Acumuladas. Em um estudo baseado em simulações, Kim et al. [9] compararam o desempenho de vários gráficos de controle, entre eles o gráfico da média, o da mediana e o gráfico de controle precedente proposto por Chakraborti et al. [5].

Assim sendo, tem-se várias metodologias diferentes para se saber se para uma determinada doença, toda uma região geográfica está ou não sob epidemia. Entretanto, não foi encontrada na literatura qualquer metodologia que permita saber se as diferentes sub-regiões que compõem aquela região estão homogeneamente ou heterogeneamente afetadas. Mesmo atualmente, autores em geral limitam-se a fazer uma análise espacial do risco ou mesmo a busca de sub-regiões com padrão espacial diferenciado [6][8][11]. Apenas recentemente, Waldhoer e Heinzi [24] apresentaram uma metodologia para combinar resultados de testes estatísticos em mapas coropléticos. Neste artigo, propõe-se uma metodologia de mapeamento, usando como fonte de dados Cartas de Controle não-paramétricas, para se conhecer as sub-regiões afetadas ou não por uma epidemia, dentro de uma região geográfica. Essa caracterização espacial é útil aos gestores para a tomada de decisão sobre políticas emergenciais específicas para localidades sob surto, diferentes daquelas que estão sob endemia, poupando recursos e deslocamentos desnecessários de pessoal.

REFERENCIAL TEÓRICO

Uma região geográfica G é a área geográfica de interesse, na qual o estudo epidemiológico é realizado. Do ponto de vista da vigilância em saúde, se a região geográfica é um município, ela pode ser dividida em sub-regiões que podem ser os distritos sanitários, áreas de abrangência dos Postos de Saúde da Família, ou mesmo os bairros que o compõem. Do ponto de vista do geoprocessamento, essas entidades são objetos distintos e localizáveis dentro da região geográfica e são chamados geo-objetos [7].

O controle estatístico de qualidade baseado em gráficos de controle, é utilizado comumente para acompanhar processos industriais. Nesse caso, em várias fases do processo fabril, são retiradas amostras aleatórias para se verificar se todo o processo fabril está sobre controle até aquela etapa. Quando as variáveis medidas do processo seguem a distribuição normal, a partir desta amostra são calculadas medidas estatísticas, como a média, a amplitude e o desvio padrão, que são confrontadas com os intervalos de confiança das estatísticas esperadas quando o processo está sobre controle. Se os valores aferidos pelas estatísticas estiverem fora do intervalo de confiança, significa que a produção está fora dos padrões esperados, ou aceitáveis e portanto deve-se procurar e corrigir as possíveis causas para tal [13]. Eventualmente, no processo fabril a produção deve ser interrompida até que a(s) causa(s) seja(m) encontrada(s) e sanada(s).

Como já mencionado anteriormente, no controle epidemiológico, podem ser utilizados os Corredores Epidêmicos ou os Gráficos de Controle. Neste último, podem ser utilizadas informações da contagem de número de casos dentro de períodos de tempo, incidência ou o risco. Verifica-se se as estatísticas do período de tempo observado estão dentro do intervalo esperado para uma ocorrência endêmica [18]. Segundo Sanches [19], essas várias semelhanças entre as teorias do controle estatístico de processos e do controle epidemiológico possibilitam usar técnicas de controle de qualidade para análise epidemiológica.

Uma carta de controle, seja ela paramétrica ou não-paramétrica, é uma forma gráfica de realizar um teste de hipóteses para decidir se o processo epidemiológico é endêmico (está sob controle) ou epidêmico (está fora de controle). Do mesmo modo que em um teste de hipóteses estatístico clássico, pode-se cometer dois tipos de erro: o erro tipo I e o erro tipo II [12]. O primeiro ocorre quando se decide aceitar a hipótese de que o processo é endêmico, quando na realidade é epidêmico e o erro tipo II ocorre quando se decide aceitar a hipótese de que o processo é epidêmico, quando na realidade é endêmico. Assim, o objetivo é saber se a cada indicador epidemiológico medido, ele aponta para uma epidemia (acima dos limites de controle), endemia (dentro dos limites de controle) ou abaixo dos valores endêmicos (abaixo dos limites de controle).

No caso de uma região geográfica G, pode-se usar um único gráfico para o monitoramento da região como um todo segundo algum tipo de agravo ou doença. Exemplos de monitoramento desse tipo podem ser encontrados em: Bortman [3] ilustra a metodologia de Corredores endêmicos para o monitoramento das áreas na Argentina; Sanches [19] mostra aplicação de Cartas de Controle CUSUM sobre casos de meningite no Rio de Janeiro (também traz uma lista de autores que utilizaram gráficos de controle em estudos epidemiológicos); Sousa et. al. [22] e Nascimento et. al. [14] realizaram estudos de sobre a incidência do dengue no estado da Paraíba e a incidência de óbitos por homicídios da cidade de João Pessoa, ambos usando Cartas de Controle de Shewhart.

Neste caso, pretende-se conhecer como cada sub-região G_i , com i={1,...,n}, onde n são todas as sub-regiões que compõem a região G, se comporta. Assim, para cada sub-

região é possível, usando gráficos individuais, saber se ela está em estado endêmico, epidêmico ou abaixo deste último, para aquele particular indicador epidemiológico medido no instante de tempo t. Desse modo, a construção de um mapa coroplético com essas informações deve levar em conta todos os resultados das cartas de controle para aquele indicador. Sendo três as classificações possíveis, o mapa também deve ilustrá-las através de legendas: acima dos limites de controle, dentro dos limites de controle ou abaixo dos limites de controle. Portanto, cada sub-região G_i de G, receberá uma codificação de cor, correspondente a uma das três classificações possíveis que foi fornecida pela sua carta de controle para o indicador epidemiológico medido naquele instante de de tempo.

Como uma interessante consequência desse tipo de construção gráfica, o gestor de saúde pode tomar decisões específicas sobre políticas emergenciais para as diferentes localidades de acordo com a situação em que elas se encontram. A identificação de sub-regiões abaixo dos limites epidêmicos pode revelar características peculiares daquelas sub-regiões. Elas podem levar a providências de proteções extras para que essas sub-regiões não venham a ter seus riscos elevados futuramente.

APLICAÇÃO

A título de ilustração, um exemplo com dados simulados é apresentado, usando o gráfico de controle da mediana. Apesar de ser contra-indicado em muitas situações de controle estatístico de processos na indústria, na área epidemiológica o uso desse gráfico de controle pode ser interessante, visto a tendência claramente não-normal dos dados epidemiológicos. Como mencionado anteriormente, pode-se tentar transformações sobre os dados, porém em casos de dados inflacionados em zero (muito comuns em situações endêmicas), e com amostras pequenas, essas transformações podem não ser eficazes.

A mediana M é uma medida estatística de tendência central, particularmente eficaz quando as distribuições de probabilidade são assimétricas. Ela busca o valor que divide os dados ordenados em 50%. No caso de que o tamanho da amostra n é ímpar o valor de M é o valor que divide ao meio o conjunto de dados e no caso de n par, ela é a média aritmética dos dois valores centrais. Para o monitoramento de casos de uma determinada doença ou agravo dentro de um período de tempo (por exemplo, semanas epidemiológicas ou o mês tradicional), a amostra é o número de casos notificados daquela doença ou enfermidade (sete dias no caso da semana epidemiológica ou quatro semanas no caso de um mês tradicional).

Os três limites de controle para o gráfico da mediana são: a média da mediana no período, denotada por M', o limite superior de controle (LSC) e o limite inferior de controle (LIC), respectivamente. Esses dois últimos são dados pelas equações (1) e (2) [15]:

$$LSC = M' + A'_2 \times R' \quad (1)$$

$$LIC = M' - A'_2 \times R' \quad (2)$$

onde R' é a média das amplitudes dos dados e A'_2 é um fator para o cálculo dos limites de controle, que pode ser encontrado na Tabela I. O fator A_2 presente na Tabela I é um fator para o cálculo dos limites de controle para o gráfico da média.

TABELA I
LIMITES DE CONTROLE PARA GRÁFICOS DA MÉDIA (A_2) E DA MEDIANA (A'_2).

n	A_2	A'_2
3	1,023	1,187
4	0,729	0,796
5	0,577	0,691
6	0,483	0,548
7	0,419	0,508
8	0,373	0,433
9	0,337	0,412
10	0,308	0,362

ADAPTADO DE MONTGOMERY [13] E OSSWALD ET AL. [15].

Para cada um dos geo-objetos que compõem a região geográfica em estudo, um gráfico de controle da mediana é delineado. Foram simuladas 26 semanas epidemiológicas, das quais 25 foram utilizadas para a construção dos parâmetros da mediana e dos limites inferior e superior dados pelas equações (1) e (2). O fator A'_2 utilizado foi 0,508. A Figura 1 apresenta um gráfico de controle até o instante t observado para um geo-objeto. Nesse caso, a situação desse geo-objeto é que ele está acima do limite epidêmico.

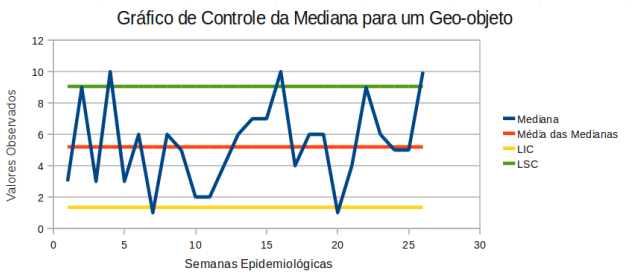

FIGURA. 1
GRÁFICO DE CONTROLE DA MEDIANA SIMULADO PARA UM DOS GEO-OBJETOS PRESENTES NA REGIÃO GEOGRÁFICA.

Um mapa coroplético é construído a partir dos dados simulados para os diversos geo-objetos presentes na Figura 2. Observa-se que foi adotado um sistema de cores similar a um semáforo de trânsito para facilitar a identificação das áreas: acima do limite endêmico (vermelho), endêmico (amarelo) e abaixo do limite endêmico (verde). Deve-se notar que este é o resultado para o último instante t observado simultaneamente para toda a região geográfica.

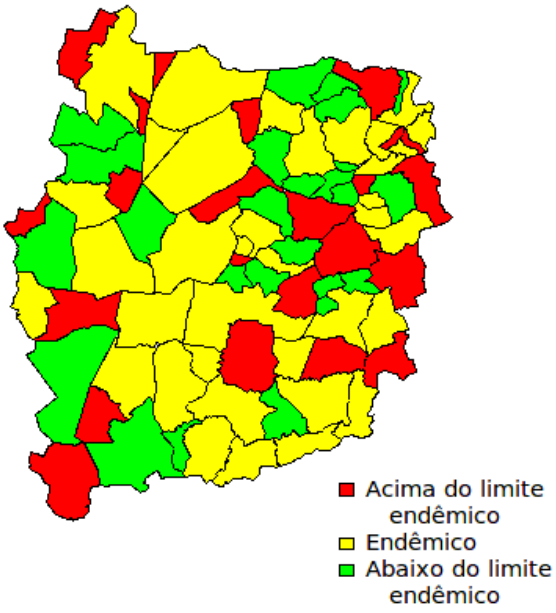

FIGURA. 2
MAPEAMENTO DOS RESULTADOS DOS DIFERENTES GRÁFICOS DE CONTROLE DA MEDIANA EM UM INSTANTE T, PARA CADA UM DOS GEO-OBJETOS QUE COMPÕEM A REGIÃO GEOGRÁFICA

Deve-se ressaltar ainda que a metodologia descrita neste trabalho é geral o suficiente para poder ser aplicada ao mapeamento de outros Gráficos utilizados em Controle Estatístico de Qualidade.

CONCLUSÕES

Neste artigo foi apresentada uma metodologia de análise espacial de endemias a partir de informações epidemiológicas oriundas de Gráficos de Controle. A construção de mapas coropléticos usando essa metodologia, permite mapear as diferentes sub-regiões que compõem uma região geográfica e saber se ela é afetada homogeneamente por uma epidemia a partir de um indicador epidemiológico medido. Como uma interessante consequência desse tipo de construção gráfica é possível identificar sub-regiões epidêmicas (acima dos limites de controle), endêmicas (dentro dos limites de controle) ou abaixo dos valores endêmicos (abaixo dos limites de controle).

Essa caracterização espacial é útil para gestores para a tomada de decisão sobre políticas emergenciais específicas para localidades sob surto, diferentes daquelas que estão sob endemia, poupando recursos e deslocamentos desnecessários de pessoal. Esse conhecimento, aliado a outras informações próprias de cada localidade, também permite compreender as suas características epidemiológicas individuais para a priorização diferenciada de providências. A identificação de sub-regiões abaixo dos limites epidêmicos pode revelar características peculiares daquelas sub-regiões e levar a providências de proteções extras específicas para elas.

A mesma metodologia descrita aqui é aplicável a ao mapeamento do Gráfico da Mediana e também a outros tipos de gráficos utilizados em Controle Estatístico de Qualidade com vistas ao Controle e Monitoramento Epidemiológicos.

REFERÊNCIAS

[1] Albers, W., Kallenberg, W. C. M., "Minimum control charts". *Journal of Statistical Planning and Inference*, v.138, p. 539 - 551, 2008.

[2] Albers, W., Kallenberg, W. C. M., Nurdiati, S., "Data driven choice of control charts". *Journal of Statistical Planning and Inference*, v.136, p. 909 - 941, 2006.

[3] Bortman, M. Elaboración de corredores o canales endémicos mediante planillas de cálculo. *Revista Panamericana de Salud Pública/Pan American Journal of Public Health*, v.5n.1, p. 1-8, 1999.

[4] BRASIL, *Guia de Vigilância Epidemiológica*, 6a. edição. Brasília: Ministério da Saúde, 2005.

[5] Chakraborti, S. van der Laan, P., van de Wiel, M. A., "A class of distribution-free control charts". *Appl. Statist.*, v. 53, P.3, p. 443-462, 2004.

[6] Curtis, A. J., Lee, W. A., Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. International Journal of Health Geographics, 9:43, 2010.

[7] Goodchild, M. "Geographical data modeling". Computers & Geosciences, v.18, n.4, p.401-408, 1992.

[8] Holowaty, E.J., et. al., "Feasibility and utility of mapping disease risk at the neighbourhood level within a Canadian public health unit: an ecological study". *International Journal of Health Geographics*, 9:21, 2010.

[9] Kim, J. et al. "A Comparison of the performance of mean, median, and precedence control charts for nonnormal data". *Proceedings of the Spring Conference, Korean Statistical Society*, p. 197-201, 2005.

[10] Lucas, J. M., Saccucci, M. S., Exponentially weighted moving average control schemes: Properties and enhancements. Technometrics, 32: 1-12, 1990.

[11] Minamisava, R. et al., "Spatial clusters of violent deaths in a newly urbanized region of Brazil: highlighting the social disparities". *International Journal of Health Geographics*, 8:66, 2009.

[12] Mood, A.M., Graybill, F.A., Boes, D.C., Introduction to the Theory of Statistics. McGraw-Hill, 3rd ed., 1974.

[13] Montgomery, D C, *Introduction to Statistical Quality Control*, Arizona: John Wiley & Sons, 2005.

[14] Nascimento, P.S., Costa, D.C.S., Moraes, R.M., Controle estatístico para os dados de mortalidade por homicídios da cidade de João Pessoa. *VI Encontro Regional de Matemática Aplicada e Computacional* (ERMAC-2006). João Pessoa, CD-ROM, 2006.

[15] Osswald, T.A., Turng, L.-S., Gramann, P., Injection Molding Handbook. Munich: Carl Hanser Verlag, 2008.

[16] Qiu, P., "Distribution-Free Multivariate Process Control Based On Log-Linear Modeling". *IIE Transactions*, 40(7): 664-677, 2008.

[17] Rocke, D. M., "Robust Control Charts". *Technometrics*, v.31, n.2, p.173-184, 1989.

[18] Rouquayrol, M Z, *Epidemiologia e Saúde*, Rio de Janeiro: MEDSI, 1983.

[19] Sanches, O, "Princípios básicos de procedimentos estatísticos aplicados na análise de dados de vigilância em saúde pública: uma

revisão". *Caderno de Saúde Pública*. Rio de Janeiro, v. 16, n.2, p. 317-333, 2000.

[20] Sego, L. H., "Applications of Control Charts in Medicine and Epidemiology". *PhD Thesis*. Department of Statistics. Virginia Polytechnic Institute and State University, 2006

[21] Siegel, S., *Estatística não-Paramétrica para as ciências do comportamento*. São Paulo: McGraw-Hill do Brasil, 1975.

[22] Souza, I. C. A.; Furtado, P.C.; Moraes, R. M. Utilização de Cartas de Controle de Shewart no acompanhamento epidemiológico do dengue. *Proceedings of the Environmental and Health World Congress* 2006. (EWHC'2006). Santos, p. 230-233, 2006.

[23] Tuckey,J. W., Exploratory Data Analysis. Addison Wesley, 1977.

[24] Waldhoer, T., Heinzl, H., Combining difference and equivalence test results in spatial maps. International Journal of Health Geographics, 10:3, 2011.

[25] Yang, L., Pai, S., Wang, Y.-R., "A Novel CUSUM Median Control Chart". Proceedings of the International MultiConference of Engineers and Computer Scientists 2010, v.3, Hong Kong, p.1707-1710, 2010.

[26] Yang, S-F., Lin, J-S., Cheng , S.W., "A new nonparametric EWMA Sign Control Chart". *Expert Systems with Applications,* v. 38, n. 5, p. 6239 - 6243, 2011.

[27] Zhou, C., Zou, C., Zhang, Y., Wang, Z., "Nonparametric Control Chart Based on Change-Point Model". *Statistical Papers*, v.50, n.1, p.13-28, 2009.