COACTION PUBLISHING

A new interpretation of the inference test for the spatial scan statistic

Alexandre Leite Almeida[1], Luiz Duczmal[2*], Anderson Duarte[3], Fernando Pereira Oliveira[3] and Ricardo H. Caldeira Takahashi[4]

[1]Universidade Federal de São João del Rei, Ouro Branco, Brazil; [2]Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [3]Universidade Federal de Ouro Preto, Ouro Preto, Brazil; [4]Mathematics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Objective

We propose a modification to the usual inference test of the spatial scan statistic, incorporating additional information about the size of the most likely cluster found.

Introduction

Spatial cluster analysis is considered an important technique for the elucidation of disease causes and epidemiological surveillance. Kulldorff's spatial scan statistic, defined as a likelihood ratio, is the usual measure of the strength of geographic clusters (1). The circular scan (2), a particular case of the spatial scan statistic, is currently the most used tool for the detection and inference of spatial clusters of disease.

Kulldorff's spatial scan statistic for aggregated area maps searches for clusters of cases without specifying their size (number of areas) or geographic location in advance. Their statistical significance is tested while adjusting for the multiple testing inherent in such a procedure. However, as is shown in this work, this adjustment is not done in an even manner for all possible cluster sizes (3).

Methods

We pose a modified inference question: what is the probability that the null hypothesis is rejected for the original observed cases map with a most likely cluster of size k, taking into account only those most likely clusters of size k found under null hypothesis for comparison? This question is especially important when the *p*-value computed by the usual inference process is near the alpha significance level, regarding the correctness of the decision based in this inference.

Results

Numerical experiments are made showing that the proportions of rejections of the null hypothesis differ noticeably, by employing the usual critical value, compared with using the data-driven critical values. It is also shown that the computational cost of estimating the data-driven critical value may be reduced through the use of a simple interpolation.

Conclusions

A practical procedure is provided to make more accurate inferences about the most likely cluster found by the spatial scan statistic. The proposed method is more useful when the computed *p*-value using the classical inference is close to the significance level; otherwise, there will be no change in the decision process. In this situation, it is recommended that the data-driven inference should be performed, especially when the observed most likely cluster has relatively large size.

Keywords

Spatial scan statistic; inference; data-driven

Acknowledgments

We thank the agencies Fapemig, CNPq and Capes.

References

1. Kulldorff M. A spatial scan statistic. Commun Stat Theor Meth. 1997;26:1481–96.
2. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med. 1995;14:799–810.
3. Almeida ACL, Duarte AR, Duczmal LH, Oliveira FLP, Takahashi RHC. Data-driven inference for the spatial scan statistic. Int J Health Geograph. 2011;10:47.

***Luiz Duczmal**

E-mail: duczmal@ufmg.br

Citation: Emerging Health Threats Journal 2011, **4**: 11149 - DOI: 10.3402/ehtj.v4i0.11149