

**Universidade Federal de Minas
Gerais Instituto de Ciências Exatas
Departamento de Estatística**

**Detecção e inferência de conglomerados
espaciais de homicídios do município
de Riberão das Neves (MG)**

*Luiz Henrique Duczmal
Eliane Cristina Rocha*

**Relatório Técnico
RTP- 03/2003
Série Pesquisa**

Relatório Técnico de Pesquisa
Departamento de Estatística – UFMG
2003

Luiz Henrique Duczmal*

Eliane Cristina Rocha**

Deteccão e inferência de conglomerados espaciais de homicídios do município de Ribeirão das Neves (MG)

* Departamento de Estatística, Universidade Federal de Minas Gerais

** Faculdades Newton Paiva, Belo Horizonte, Minas Gerais



RESUMO

A análise estatística espacial e o mapeamento das regiões de alta incidência de homicídios para o Município de Ribeirão das Neves - MG (Brasil) são realizados de forma pioneira neste trabalho. O município de Ribeirão das Neves (que inclui também a localidade de Justinópolis) é dividido em 112 bairros com população e casos de homicídios (tentados e consumados) conhecidos para o ano de 2000. Regiões vizinhas entre si no mapa são conectadas, formando uma estrutura matemática de grafo. Este grafo é utilizado pelo algoritmo de detecção e inferência de conglomerados espaciais de Duczmal e Assunção para localizar e validar estatisticamente os conglomerados com maior incidência de homicídios.

ABSTRACT

The spatial statistics analysis and the map of the regions of high homicide incidence for the municipality of Ribeirão das Neves - MG (Brazil) were realized for the first time in this work. The municipality of Ribeirão das Neves (that also includes the city of Justinópolis) is divided into 112 regions with known populations and homicides cases (attempted and accomplished), for the year 2000. The neighboring regions in the map are connected, making a graph mathematical structure that is used by the cluster detection algorithm of Duczmal and Assunção in order to detect and validate statistically the homicide clusters of high incidence.

1 - Introdução

Neste trabalho iremos realizar um mapeamento das regiões de alta incidência de homicídios para o município de Ribeirão das Neves – Minas Gerais (Brasil). O município de Ribeirão das Neves (que inclui também a localidade de Justinópolis) é dividido em 112 bairros com populações e casos de homicídios (tentados e consumados) conhecidos para o ano 2000.

Ribeirão das Neves pertence à grande BH (Belo Horizonte), e teve um crescimento populacional muito rápido e desordenado, culminando numa cidade com uma população saturada e com baixos índices de desenvolvimento econômico. Com isso a criminalidade no município cresceu rapidamente, tornando-se uma das cidades mais noticiadas da região. Além disso, a presença da Penitenciária Estadual de Ribeirão das Neves é considerada por muitas pessoas como um fator significativo para o aumento do índice de criminalidade, já que abriga presos de alta periculosidade e fugas são freqüentes. É neste panorama que Ribeirão das Neves se encontra atualmente, e por isso resolvemos nesta monografia mapear todo o município e detectar as regiões que possuem os maiores índices de homicídios e as que ainda aparentemente possuem mais segurança.

Neste trabalho, vamos analisar os aspectos estatísticos e matemáticos da representação geográfica quantitativa das populações e casos de crimes. Vamos construir o mapeamento das regiões (bairros) que formam o município de Ribeirão das Neves. Regiões vizinhas entre si no mapa são conectadas formando uma estrutura matemática de grafo. Este grafo é utilizado pelo algoritmo computacional de detecção e inferência de conglomerados espaciais de Duczmal e Assunção para localizar e validar estatisticamente as regiões com maior incidência de homicídios. A ocorrência de homicídios por habitante por ano em cada bairro é um índice comumente utilizado em estatística criminal para mostrar o grau de violência. Esse índice é amplamente usado por ser o mais confiável dentre todos os indicativos de criminalidade, do ponto de vista de registro estatístico, sendo o menos sujeito a erros de omissão e interpretação.

A importância desse estudo consiste em caracterizar geograficamente os conglomerados com maior índice de homicídios em Ribeirão das Neves. Os serviços públicos de combate e prevenção da violência poderiam então usar recursos de forma mais otimizada, concentrando-se nas regiões onde o problema é mais crítico. Isso pode trazer grandes benefícios para a população, pois permitiria um policiamento mais efetivo. Por outro lado, as causas da violência poderiam ser estudadas com uma base quantitativa mais sólida, fazendo com que políticas públicas preventivas possam ser adotadas com mais eficiência. A determinação dos conglomerados espaciais estatisticamente significativos pode não ser óbvia apenas através de uma olhada casual em um mapa, sendo portanto necessário um estudo mais aprofundado como o que estamos realizando neste trabalho. A necessidade de se avaliar a significância estatística dos conglomerados espaciais detectados pelo algoritmo faz com que sejam necessários métodos estatísticos sofisticados.

Na seção 2, vamos descrever a teoria matemática de Kulldorff para a estatística de teste para conglomerados espaciais e a estrutura de grafo aqui utilizada. Na seção 3 apresentamos o algoritmo computacional de Simulated Annealing de Duczmal e Assunção. Na seção 4 mostramos os detalhes da construção do grafo para o mapa do município de Ribeirão das Neves. Na seção 5 discutimos os resultados computacionais obtidos para este município. Os conglomerados de homicídios tentados e consumados encontrados pelos métodos descritos acima estão respectivamente nas figuras compostas 5 e 6.

2 - A estatística de teste de Kulldorff

Vários métodos já foram propostos na literatura para testar a presença de conglomerados espaciais de risco elevado e identificar suas localizações. Uma revisão exaustiva desses métodos pode ser encontrada em Lawson et al. (1999). Os métodos assumem que temos à nossa disposição um mapa de regiões, cada uma com uma população de risco e um certo número de casos observados. As regiões do mapa podem ser, por exemplo, municípios de um estado, ou bairros de uma cidade. Os casos correspondem a indivíduos da população que tem uma designação especial, tais como um indivíduo infectado ou uma vítima de crime. Em nosso estudo, vamos definir um centróide para cada região, que é um ponto arbitrário situado em seu interior (veja as figuras 1A e 1B). Dizemos que *a região y é a vizinha mais próxima da região x* quando o centróide de *y* é o centróide mais próximo do centróide de *x*.

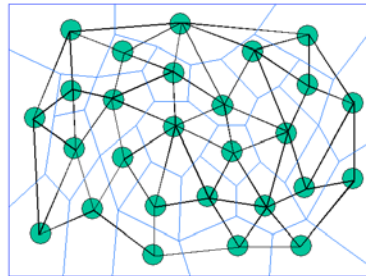
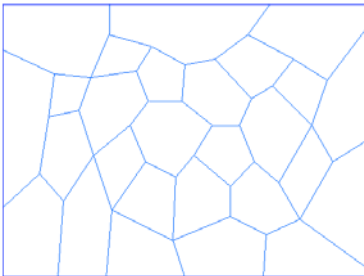
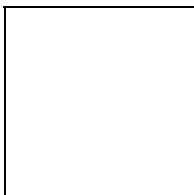


FIGURA 1A: As regiões de um mapa.

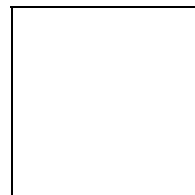
FIGURA 1B: O grafo associado ao mapa.

Os métodos mais úteis são aqueles que utilizam janelas móveis, ao superpor zonas circulares ou retangulares sobre as regiões de estudo, e fazem uma contagem do número de casos das regiões cujos centróides caem dentro de cada zona. Esse tipo de método altera o tamanho da zona (por exemplo, considerando todos os círculos com todos os raios e centros possíveis dentro do mapa), e avalia a significância estatística do número de casos que caem dentro de cada zona. Isso permite que seja encontrada a melhor zona (circular, por exemplo) que vai conter a concentração mais significativa de casos dentro do mapa.

Por exemplo, seja dado um mapa com 100 regiões (bairros), numerados de 1 a 100. Fixando inicialmente o bairro número 1, considere os seguintes 100 diferentes subconjuntos (ou zonas) de bairros do mapa que contém o centróide do bairro $b_{i_0} = 1$ em seu centro: O primeiro subconjunto contém apenas o bairro b_{i_0} . O segundo subconjunto contém o bairro



e seu vizinho mais próximo, o bairro



. O terceiro subconjunto

contém os bairros , e (que é o segundo vizinho
 mais próximo a partir do bairro). Defina desse modo todos os 100
 subconjuntos formados a partir de . Observe que, por causa da utilização das
 distâncias entre os centróides para definir vizinhança, os subconjuntos assim construídos têm
 um formato aproximadamente circular. O centésimo subconjunto contém, obviamente, todos
 os 100 bairros. Repita todo esse processo agora começando por . Em seguida
 comece novamente com , , etc., até . Para cada
 um desses possíveis subconjuntos (alguns repetidos, é claro), calcule sua
 população e número de casos, somando as populações e casos de cada região do subconjunto,
 e determine uma pontuação para cada uma dessas zonas, baseada em uma função de teste
 que leva em conta essa concentração de casos. A função é
 construída de modo que zonas que se destacam no mapa devido à alta incidência de casos
 tenham uma alta pontuação. Vai existir então um subconjunto que atinge o

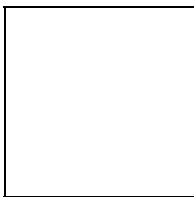
máximo da função . Essa zona será chamada de conglomerado mais plausível. É conveniente se limitar o tamanho máximo de cada círculo, de modo que os conglomerados não tenham mais do que 50% das regiões do mapa. Assim, no exemplo desse

parágrafo, precisaríamos analisar subconjuntos. Em Kulldorff & Nagarwalla ,1995 se sugere que esse valor máximo esteja entre 20% a 50% do número total de regiões, dependendo do tipo de conglomerado que se espera encontrar (a partir de informações adicionais a respeito do problema). Pode-se usar valores ainda menores, se estivermos buscando conglomerados de tamanho reduzido no mapa. Na prática, pode ser interessante executar o programa várias vezes com diferentes valores para o número máximo de regiões nos conglomerados.

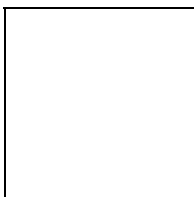
O método descrito acima corresponde ao algoritmo do SatScan, proposto por Kulldorff e Nagarwalla (1995) , Kulldorff (1997) e Kulldorff (1999) . No programa de computador

SatScan foi criada uma estatística de teste , definida da seguinte maneira: para

cada zona (isto é, para cada subconjunto qualquer de regiões do mapa), faça



onde



= população total do mapa,

[]

= número total de casos do mapa,

[]

= população da zona z ,

[]

= número de casos da zona z ,

[]

É conveniente ainda considerar a função

[]

com o valor

[]

sendo o valor máximo de

[]

Assim

[]

para qualquer zona

[]

, e

[]

é estritamente

maior do que 1 quando a densidade de casos dentro da zona

[]

(dada por

[]

) for maior do que a densidade de casos fora da zona

[]

(dada por

). Por sua vez, a zona contém casos e não-casos, e o complementar de contém casos e não-casos. Assim, o numerador da fórmula de se assemelha à expressão equivalente da probabilidade de uma variável aleatória multinomial (casos distribuídos em dois conjuntos, e o complementar de). Observe que é o produto de e , onde é a densidade de casos de todo o mapa, e é a densidade de não-casos de todo o mapa. Assim, o denominador de se assemelha à fórmula de uma variável aleatória binomial (casos distribuídos em uma população de habitantes). Por essa razão, é chamada de

estatística de teste da razão de máxima verossimilhança. Pelas considerações acima, parece

então razoável definir a zona que maximiza a função como sendo o conglomerado mais plausível. Veja detalhes e demonstrações em Kulldorff, 1997.

Nesse artigo se mostra também que a função é adequada também como estatística de teste para coleções de subconjuntos de regiões que não precisam ser necessariamente de formato circular (ou retangular) dentro do mapa, mas também para coleções de conjuntos de regiões com formato arbitrário dentro do mapa.

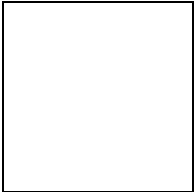
Nosso objetivo aqui é verificar se existe algum conglomerado de regiões que se destaca no mapa por ter uma incidência de casos observados significativamente acima do esperado pela média de todos os casos do mapa. Assim, se todos os casos observados se distribuíssem de forma inteiramente ao acaso na população, veríamos algumas vezes, por mero acaso, um ou outro conglomerado, mas estes não teriam significância estatística. Vamos fazer aqui um teste de hipóteses. Nossa hipótese nula seria então definida como a não existência de conglomerados, e caberia ao método decidir se a zona de mais alto valor da

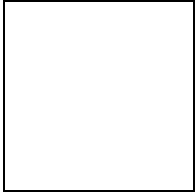
função é devida a um mero acaso ou se constitui um conglomerado realmente significativo. Para testar o que está acontecendo, o programa SatScan modifica o mapa original através de um procedimento computacional, conhecido como simulação de Monte Carlo. O número de casos observados de cada região é alterado de forma aleatória, de modo

que cada região receba um certo número de casos. Esse

número é uma variável aleatória com média igual ao número total de casos do

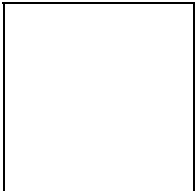
mapa multiplicado pela fração da população da região em relação à população

total, i.e. . Isso equivale a simular uma situação em que é válida a hipótese nula, de que não existe nenhum conglomerado, ou seja, eventuais detecções de conglomerados decorrem apenas de efeitos puramente aleatórios, em que todos os indivíduos da população têm a mesma probabilidade de serem casos, independentemente da região a qual pertencem. Os casos simulados em cada uma das regiões do mapa formam então uma variável aleatória multinomial. O algoritmo das zonas circulares de Kulldorff é então repetido para este novo mapa com casos simulados, e a função de máxima verossimilhança é novamente calculada. Todo esse procedimento é então repetido milhares de vezes, e por fim contamos quantas vezes o valor da função de máxima verossimilhança nos mapas simulados ultrapassa

o valor da função  no mapa original de casos observados. Se o esse valor é maior em poucos mapas (digamos, em apenas 1% dos mapas simulados), podemos dizer com relativa segurança que o conglomerado detectado no mapa original é realmente significativo, isto é, ele decorre de uma concentração espacial de casos muito maior do que a esperada em uma simulação em que os casos são distribuídos ao acaso na população. Caso contrário, a hipótese nula, de que não existem conglomerados significativos no mapa original de casos observados, não pode ser rejeitada. Veja a figura 4.

O método exposto acima é uma excelente ferramenta para detectar conglomerados espaciais e avaliar sua significância estatística. No entanto, se existir algum conglomerado com formato irregular, o algoritmo de janelas circulares não irá detectá-lo com facilidade. Se o programa usar uma janela circular grande o suficiente para incluir todo o conglomerado, muitas regiões de baixa incidência também acabarão por ser incluídas, diminuindo o valor da função de verossimilhança. Por outro lado, se a janela circular for pequena o suficiente para ficar completamente contida no conglomerado irregular, muitas regiões de alta incidência do conglomerado vão ficar de fora do círculo, novamente diminuindo o valor da função de verossimilhança. Conglomerados irregulares ocorrem em muitos problemas: considere por exemplo bairros com alta incidência de crimes ao longo de uma via de transporte, ou municípios com alta incidência de doenças próximos a um rio contaminado. Assim, vamos definir em seguida um outro algoritmo de detecção para conglomerados de formato irregular que resolve esse problema.

3 - O algoritmo de Simulated Annealing

O algoritmo de Simulated Annealing (SA) de Duczmal & Assunção, 2003, trabalha com um mapa de  regiões com populações e casos como anteriormente. Vamos

definir um *grafo* (não-orientado) , onde é um

conjunto de vértices e é um conjunto de pares não ordenados da forma

, onde e são vértices de . Os

pares são chamados de *arestas*. Uma estrutura de grafo é acrescentada ao mapa, considerando cada centróide como sendo um vértice do grafo, e unindo

centróides de regiões vizinhas por arestas desse grafo. Assim, um grafo de

vértices passa a substituir o mapa original com regiões, sendo que os vértices do grafo recebem as atribuições de populações e casos das regiões

correspondentes. Um *subgrafo* do grafo é um subconjunto de

vértices do grafo , juntamente com as arestas de que unem os

vértices de . Se uma aresta liga dois vértices, então eles são *vizinhos*. Dois

vértices e são ligados entre si por um caminho se existe uma

sequência de vértices vizinhos que começa em e termina em . Um subgrafo é *conexo* quando todos os seus vértices podem ser ligados entre si por um

caminho que usa apenas arestas do subgrafo. Dado o grafo e um subgrafo

conexo de , os *subgrafos conexos vizinhos de*

são todos os subgrafos conexos de que tem um vértice a mais do que

ou um vértice a menos do que . No exemplo da figura 2 os subgrafos B até K são vizinhos do subgrafo A. O subgrafo L não é conexo.

O problema passa agora a ser o seguinte: Dentre todos os subgrafos conexos possíveis do grafo , escolha agora o subgrafo que maximiza a função .

Como já foi mencionado (Kulldorff, 1997), a estatística de teste também é adequada para coleções de zonas conexas com formato irregular. Um problema computacional aparece aqui, pois existe um número imenso de subgrafos conexos de

. Para um grafo de vértices, devemos considerar a coleção de subgrafos conexos dentre os subgrafos possíveis. Se , por exemplo, esse problema é inviável na prática, pois um computador que analisasse um milhão de subgrafos por segundo levaria 36000 anos para verificá-los. Devemos então encontrar uma estratégia para tentar analisar apenas os subgrafos mais promissores, descartando a imensa maioria de subgrafos que não parecem ter uma concentração alta de casos. Como visto anteriormente na seção 2, vamos impor uma limitação no número máximo de vértices em cada subgrafo.

Antes de continuar, vamos comentar sobre um algoritmo simples de busca de conglomerado, chamado *algoritmo guloso*. O programa escolhe aleatoriamente como base um

subgrafo conexo inicial e considera todos os seus subgrafos conexos vizinhos. O algoritmo então muda sua base escolhendo o subgrafo conexo vizinho mais promissor (i.e.,

o que mais aumenta a função). A partir dessa nova base, o processo acima é repetido várias vezes até que novos melhoramentos não sejam mais possíveis. A última base então é considerada como o melhor subgrafo descoberto pelo algoritmo. Esse algoritmo tem a vantagem de ser rápido, pois analisa relativamente poucos subgrafos, e descarta direções de pesquisa que não parecem promissoras. Apesar de conseguir funcionar em algumas situações,

descobrindo o subgrafo que maximiza , o algoritmo guloso tem uma falha

séria: pode ser que um outro subgrafo ainda melhor seja vizinho de um vizinho de uma base, e assim passe despercebido pelo programa. Isso de fato ocorre com muita frequência, e torna o algoritmo guloso inadequado para este tipo de busca.

O algoritmo de Simulated Annealing (SA), descrito a seguir, resolve esse problema. Um livro texto clássico que descreve a filosofia geral do algoritmo SA, que tem muitos usos na física, química e computação é Aarts & Korst, 1989. A idéia aqui é melhorar o algoritmo guloso de modo que, a partir de cada base, o algoritmo SA não escolha sempre o subgrafo

conexo vizinho de mais alto valor da função como sendo a nova base. Ao invés disso, o algoritmo SA adota o seguinte critério: dependendo de certos fatores, escolhe aleatoriamente um vizinho da base, ou então escolhe o vizinho conexo mais promissor

(exatamente como no algoritmo guloso). Que fatores são esses? Seja o

subgrafo base da etapa atual, e seja o melhor subgrafo encontrado até o momento. O algoritmo SA verifica em cada etapa os seguintes fatores:

- O número de etapas consecutivas em que não houve aumento

de ;

- O número de vezes em que a base já tinha sido visitada antes em etapas anteriores;

- O número de vértices em comum entre e

- Foi encontrado () ou não () um subgrafo vizinho

conexo na etapa atual com valor .

Fica evidente que se é alto, é alto, é

baixo, ou , o algoritmo está tendo dificuldades para encontrar novos subgrafos

que aumentam a função . Nesse caso, parece não existir vantagem em usar a estratégia do algoritmo guloso, e o algoritmo SA sorteia aleatoriamente um vizinho para a nova base, na esperança de entrar em um lugar do espaço de configurações que tenha melhores subgrafos. É claro que se o algoritmo persistir por muitas etapas e não conseguir nenhuma melhora significativa, então ele finalmente abandona o processo. Se isso acontecer, ele então sorteia um subgrafo completamente novo e o usa como uma nova base para reiniciar a busca. Isso é feito várias vezes até que o programa considera que já foi varrida uma porção apreciável do espaço de todas as configurações de subgrafos possíveis do mapa – pode até ser que exista um outro subgrafo melhor do que o encontrado, mas o esforço para achá-lo não compensa o resultado. Desse modo, observamos que esse tipo de algoritmo pode às vezes se contentar em descobrir uma solução *quase-ótima*, i.e., não é a melhor solução possível, mas é boa o suficiente para nossos propósitos. Veja mais detalhes sobre o algoritmo em Duczmal e Assunção, 2003.

4 - A construção do mapa de homicídios de Ribeirão das Neves

O município de Ribeirão das Neves, de acordo com o Censo de 2000, possui uma população de 259065 habitantes. De acordo com a Prefeitura Ribeirão das Neves, existem 160 bairros no município, com exceção dos bairros ainda não cadastrados. No entanto, vários desses bairros são loteamentos desabitados. O mapeamento desta região é difícil, já que não existem muitas informações completas e confiáveis. Neste trabalho, para a detecção de homicídios usamos apenas 112 bairros, e excluimos os bairros com população zero, além dos bairros não cadastrados.

Obtivemos a população do ano 2000 para 112 bairros através da Prefeitura de Ribeirão das Neves. O número de homicídios tentados e consumados em 2000 foi fornecido pelo Comando Geral da Polícia Militar de Belo Horizonte, já que o 4º Batalhão da Polícia Militar da Região de Ribeirão das Neves não tinha os dados referente a este ano. Na tabela abaixo listamos os bairros, a população e os casos de homicídios tentados e homicídios consumados em cada bairro.

Tabela 1.1: Bairros do Município de Ribeirão das Neves, com populações e casos de homicídios no ano 2000.

Nº	BAIRRO	POPULAÇÃO	HOMICÍDIOS TENTADOS	HOMICÍDIOS CONSUMADOS
01	Landi	2630	0	0
02	Pedra Branca	7180	0	0
03	Maria Helena	7610	0	0
04	Vila Delma	1060	0	0
05	Boa Vista	1140	0	0
06	Nossa Senhora de Fátima	1350	0	0
07	Lídice	685	0	0
08	Tony	3195	0	0
09	Atalaia	1490	0	0
10	Flamengo	1785	0	0
11	Botafogo	8020	0	0
12	Maracanã	1060	0	0
13	Adriana	290	0	0
14	Piedade	815	0	0
15	Esperança	3850	0	0
16	Menezes	6935	0	0
17	Landi 2º seção	1505	0	0
18	Santa Margarida	1850	0	0
19	Urca	1765	0	0
20	Cerejeiras	1290	0	0

21	Tropical	560	0	0
22	Centro Justinópolis	4950	0	0
23	Papini	880	0	0
24	São Januário	2360	0	0
25	São Miguel	1305	0	0
26	Laredo	590	0	0
27	Guadalajara	950	0	0
28	Lagoa	2745	0	0
29	Santa Fé	615	0	0
30	São José	1945	0	0
31	Jardim de Alá 1º seção	1940	0	0
32	Felixlândia	2020	0	0
33	Eliane	1070	2	0
34	São João de Deus	1600	0	0
35	Fortaleza	1205	0	0
36	Sônia	775	0	0
37	Penha	660	0	0
38	Hawai	1415	0	0
39	Kátia	2250	0	0
40	Céu Anil	1000	0	0
41	Luar da Pampulha	2595	0	0
42	Santa Izabel	185	0	0
43	Tancredo Neves	835	0	0
44	Vila Terezópolis	85	0	0
45	Jardim Alvorada	685	0	0
46	Vila santa Branca	1365	0	0
47	Rosimeire	820	1	0
48	Vera Lúcia	320	0	0
49	Sevilha	19845	5	0
50	Labanca	1005	0	0
51	José Maria da Costa	430	0	0
52	Rosana	1660	1	0
53	Santo Antônio	780	0	1
54	Santa Martinha	7250	1	2
55	Santa Marta	3500	1	0
56	São Geraldo	595	2	0
57	Tânia	475	0	0
58	Várzea Alegre	600	0	0
59	Savassi	4105	1	1

60	São Pedro	1760	0	0
61	Jardim Florência	705	0	0
62	Quintas do Lago	110	0	0
63	Vila real	190	0	0
64	Status	400	2	0
65	Santa Paula	840	0	0
66	Jardim Colonial	1790	3	0
67	Cidade Neviana	2350	1	0
68	Vale Verde	335	1	0
69	Veneza	15745	9	6
70	Vale do ouro 1º seção	1040	1	0
71	Canoas	185	0	0
72	Vale das Acácias	1790	0	0
73	Vale da Prata	895	0	0
74	Santa Matilde	2175	0	0
75	Nossa Senhora das Neves	1695	2	0
76	Vila Mariana	365	0	0
77	Florença	9305	3	3
78	Santinho	6235	4	1
79	Rosaneves	6535	3	2
80	Bom Sussego	1535	0	0
81	Centro de Neves	1280	0	0
82	Elizabeth	470	0	0
83	Nova Pampulha	7035	0	0
84	Centro de Areias	450	0	0
85	Granjas Primavera	2655	0	0
86	Vale do Ouro 2º seção	625	0	0
87	Chácaras Bom Retiro	265	0	0
88	Santana	1175	0	0
89	Vila Fluminense	235	0	0
90	Vila Aparecida	170	0	0
91	Nossa Senhora da Conceição	405	0	0
92	Paraíso das Piabas	910	0	0
93	Chácaras do Baú	145	0	0
94	San Genaro	3435	2	1
95	Henrique Saporì	4650	1	1
96	São Francisco	15	1	2
97	Belo Vale	495	0	0
98	Cruzeiro	605	0	0

99	Jardim Primavera	50	0	0
100	São Judas Tadeu	250	0	0
101	Franciscadriângela	460	0	0
102	Porto Seguro	05	0	0
103	Soares	40	0	0
104	São Judas Tadeu	485	0	0
105	Severina	20	0	0
106	Monte Verde	375	0	0
107	Barcelona	540	0	0
108	Dona Clarice	155	0	0
109	Evereste	265	0	0
110	Fazenda Castro	2970	1	0
111	Alicante	125	0	0
112	Vila da Hortinha	380	0	0

A seguir vamos descrever as dificuldades encontradas na elaboração do mapa de populações e casos. Deve-se ressaltar que esse tipo de análise é pioneiro no município, assim é natural que apareçam dificuldades relativas à denominação e localização dos bairros. Encontramos várias situações de denominação dupla e redundância nos nomes dos bairros. Algumas vezes a localização de um bairro citado na tabela era incerta. Cruzamos as informações de vários mapas e fontes diversas (Prefeitura, Polícia Militar e da Secretaria de Saúde). Foi cogitado ainda se realizar um levantamento aerofotogramétrico da região, mas como não existem fotos recentes do município, o custo para a produção de novas fotografias aéreas seria muito elevado. No entanto essa seria uma alternativa válida, e recomendamos fortemente que essa ferramenta seja empregada em trabalhos futuros. Abaixo estão listadas algumas falhas que solucionamos através de cruzamento das informações disponíveis e discussões com funcionários da Prefeitura de Ribeirão das Neves, que esclareceram algumas inconsistências por nós observadas:

- O Bairro Delma aparece no mapa como Vila Delma e na tabela apenas Delma.
- Os Bairros Tony, Jardim de Alá e Luar da Pampulha aparecem na tabela especificando como 1º, 2º e 3º seção. No mapa estão sem especificação.
- O Bairro Botafogo aparece na tabela sem especificação e no mapa como Botafogo 1º e 2º seção.
- O Bairro Papini aparece na tabela como Vila Papene e no mapa como Papene.
- O Bairro São João de Deus aparece na tabela como 1º seção e no mapa sem especificação.
- Os Bairros Eliane, Granjas Primavera, Canoas, Santa Martinha, Veneza, Florença e San Genaro aparecem no mapa duas vezes sem especificação. Na tabela aparece apenas uma vez, também sem especificação.
- O Bairro Santa Margarida aparece na tabela sem especificação e como 2º seção; no mapa apenas uma vez apenas, sem especificação.
- O Bairro Nova Pampulha aparece na tabela como conjunto Hab. Nova Pampulha e no mapa como Nova Pampulha.
- O Bairro Canoas aparece na tabela duas vezes sem especificação e no mapa apenas uma vez, sem especificação.

- O Bairro Centro de Justinópolis aparece na tabela com Centro comercial de Justinópolis e no mapa como Justinópolis.
- O Bairro Laredo aparece na tabela duas vezes como Novo Laredo e Laredo. Na tabela apenas uma vez sem especificação.
- O Bairro Sevilha aparece na tabela sem especificação, e no mapa aparece em três lugares como Sevilha 1º, 2º e 3º seção.
- Os Bairros Veneza e Santa Martinha aparecem no mapa duas vezes sem especificação e na tabela apenas uma vez, também sem especificação.
- O Bairro Vale do Ouro aparece duas vezes no mapa sem especificação e na tabela como Vale do Ouro 1º e 2º seção.

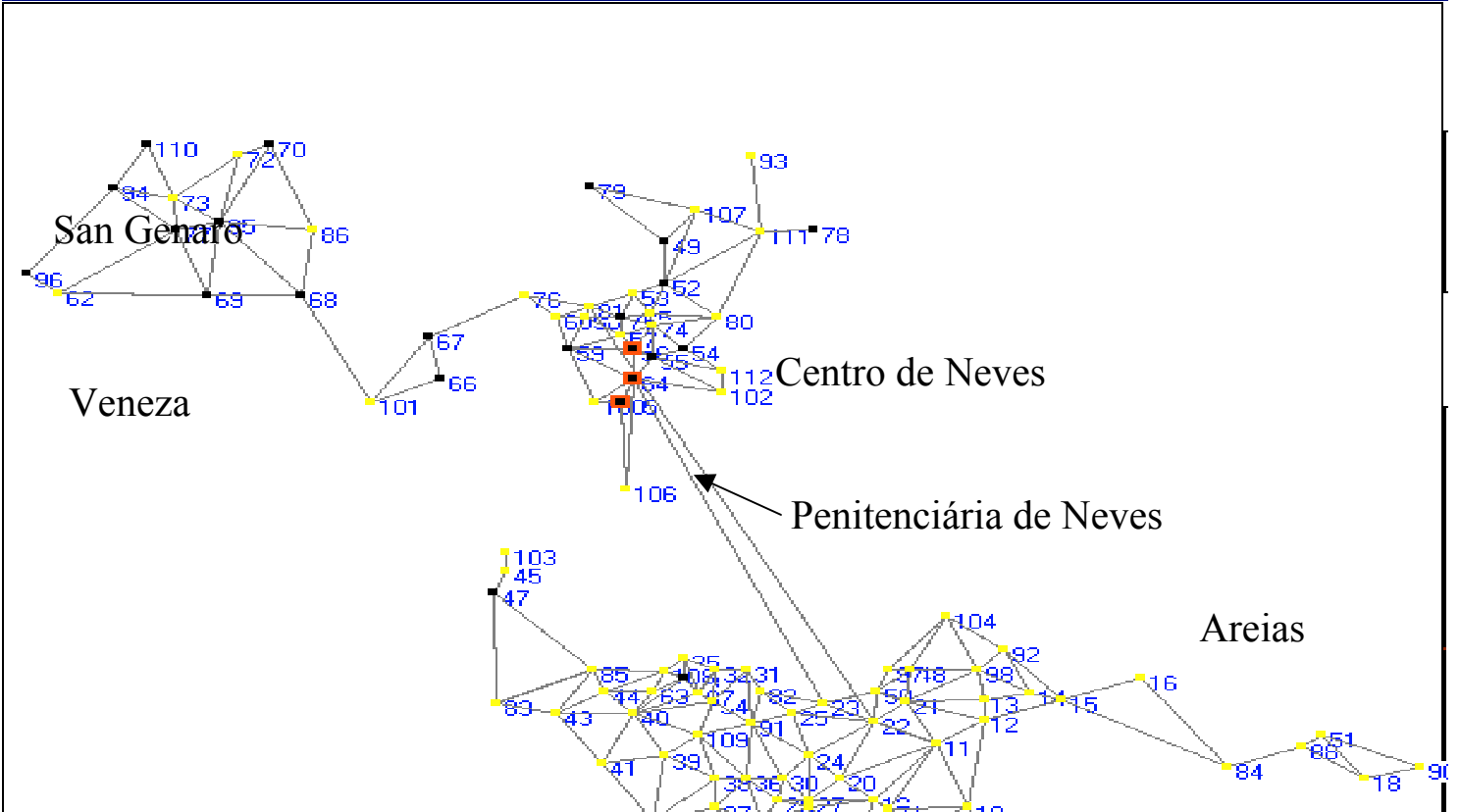
Em todos esses casos, optamos por seguir a denominação usada pela Polícia Militar na tabela 1 acima. Depois de resolvidas tais inconsistências, passamos então a confeccionar o mapa de bairros conectados, com o objetivo de criar a estrutura matemática de grafo, onde cada bairro é representado por um vértice, e bairros vizinhos são ligados por uma aresta. Reduzimos o mapa de Ribeirão das Neves através de fotocópias, e com uma folha de papel vegetal colocada sobre o mapa numeramos cada bairro de acordo com a tabela. A seguir, ligamos entre si os bairros que são vizinhos no mapa. Procuramos, na medida do possível, fazer com que as arestas do grafo formassem triângulos que fossem os menos “esticados” possíveis, isto é, que se aproximassem mais de triângulos equiláteros. Essa última condição reflete a boa qualidade da tesselação obtida, mostrando que o grafo construído está equilibrado e liga os vértices de modo adequado. As coordenadas de cada Bairro foram determinadas com o auxílio de uma folha de papel milimetrado.

A seguir digitamos o arquivo utilizado pelo programa de detecção de conglomerados espaciais, numerando os bairros de 1 a 112 e identificando os seus vizinhos, assim como sua população, número de casos, e localização geográfica.

5 – Detecção dos conglomerados e inferência

A estrutura de grafo montada na seção anterior para os mapas de casos de homicídio tentado e homicídio consumado foi usada no programa de Simulated Annealing para detecção e inferência de conglomerados. O tamanho máximo dos conglomerados para cada execução do programa foi de 56, 28, 14, 7 e 3, correspondendo a respectivamente 50%, 25%, 12.5%, 6.25% e 3% do valor total de 112 regiões, com um total de 259065 habitantes. Foram registrados para o ano 2000 um total de 50 casos de homicídios tentados, e 20 homicídios consumados. A incidência média de homicídios tentados foi de 0.000193, e a incidência média de homicídios consumados foi de 0.0000772. Os conjuntos de regiões dos conglomerados mais plausíveis encontrados pelo programa em cada rodada de 1000 simulações, indicados nas tabelas 2.1 a 2.4, estão explícitos nas tabelas 3.1 a 3.4, respectivamente. Esses conjuntos também aparecem nos mapas respectivos. As figuras 5A e 6A respectivamente mostram os conglomerados de homicídios tentados encontrados pelo método SatScan de Kulldorff e pelo método de Simulated Annealing. As figuras 5B e 6B respectivamente mostram os conglomerados de homicídios consumados encontrados pelo método SatScan de Kulldorff e pelo método de Simulated Annealing. Designamos nas tabelas 3.1-3.4 por KT_n os conglomerados de homicídios tentados encontrados pelo método de Kulldorff com no máximo n regiões, por ST_n os conglomerados de homicídios tentados encontrados pelo método de Simulated Annealing com no máximo n regiões, por KC_n os conglomerados de homicídios consumados encontrados pelo método de Kulldorff com no máximo n regiões, e por SC_n os conglomerados de homicídios consumados encontrados pelo método de Simulated Annealing com no máximo n regiões.

Simulated Annealing Cluster



Centro de
Justinópolis

Landi

Figura 3: Mapa Geral de Ribeirão das Neves com as 112 regiões numeradas.

Tabela 2.1 – Homicídios tentados – Algoritmo de zonas circulares

Número máximo de regiões	Número efetivo de regiões	População	Casos	Incidência	p-valor	Log(T)	Conjunto de regiões
56	50	157215	50	.000318	.001	24.973	KT56
28	16	53415	27	.000505	.001	13.446	KT28
14	14	51260	24	.000468	.001	9.999	KT14
7	5	30095	16	.000532	.012	7.299	KT07
3	3	17355	11	.000634	.042	6.094	KT03

Tabela 2.2 – Homicídios tentados – Algoritmo SA

Número máximo de regiões	Número efetivo de regiões	População	Casos	Incidência	p-valor	Log(T)	Percentil %95 Log(T(.))	Conjunto de regiões
56	35	108110	49	.000453	.001	38.461	19.917	ST56
28	27	137780	47	.000341	.002	20.605	15.199	ST28
14	14	39705	27	.000680	.001	19.971	11.200	ST14
7	7	23375	17	.000727	.001	11.961	8.427	ST07
3	3	1836	5	.002725	.009	8.816	6.387	ST03

Tabela 2.3 – Homicídios consumados – Algoritmo de zonas circulares

Número máximo de regiões	Número efetivo de regiões	População	Casos	Incidência	p-valor	Log(T)	Conjunto de regiões
56	13	48935	15	.000307	.001	14.799	KC56
28	13	48935	15	.000307	.001	14.799	KC28
14	13	48935	15	.000307	.001	14.799	KC14
7	7	35640	13	.000365	.001	13.874	KC07
3	3	17355	8	.000461	.002	8.997	KC03

Tabela 2.4 – Homicídios consumados – Algoritmo SA

Número máximo	Número efetivo	População	Casos	Incidência	p-	Log(T)	Percentil %95	Conjunto de
---------------	----------------	-----------	-------	------------	----	--------	---------------	-------------

de regiões	de regiões				valor		Log(T(.))	regiões
56	14	50105	16	.000319	.012	17.139	14.891	SC56
28	5	34635	13	.000375	.042	14.215	13.658	SC28
14	5	34635	13	.000375	.006	14.215	10.217	SC14
7	5	34635	13	.000375	.001	14.215	7.872	SC07
3	3	28485	10	.000351	.002	9.379	6.598	SC03

Tabela 3.1 – Conjunto de regiões dos conglomerados mais plausíveis da tabela 2.1.

KT56	60 58 81 76 59 75 57 53 56 5 74 55 52 64 67 100 54 65 66 49 80 112 79 102 107 101 111 68 106 86 78 93 69 103 95 70 45 72 77 47 73 110 94 62 35 85 96 108 32 33
KT28	72 70 73 110 95 77 94 86 69 68 62 96 67 76 79 66
KT14	70 72 110 95 73 86 77 94 68 69 62 96 79 67
KT07	62 96 69 77 94
KT03	62 96 69

Tabela 3.2 – Conjunto de regiões dos conglomerados mais plausíveis da tabela 2.2

ST56	23 31 33 34 40 44 47 49 52 54 55 56 64 65 66 67 68 69 70 74 75 76 77 78 79 80 81 82 85 94 95 96 101 110 111
ST28	49 52 53 54 55 56 59 60 64 65 66 67 68 69 70 75 76 77 78 79 80 94 95 96 101 110 111
ST14	56 57 64 66 67 68 69 75 76 77 81 94 96 101
ST07	62 66 68 69 94 96 101
ST03	56 64 65

Tabela 3.3 – Conjunto de regiões dos conglomerados mais plausíveis da tabela 2.3.

KC56	70 72 110 95 73 86 77 94 68 69 62 96 79
KC28	70 72 110 95 73 86 77 94 68 69 62 96 79
KC14	70 72 110 95 73 86 77 94 68 69 62 96 79
KC07	62 96 69 77 94 73 95
KC03	62 96 69

Tabela 3.4 – Conjunto de regiões dos conglomerados mais plausíveis da tabela 2.4.

SC56	5 53 54 67 68 69 76 77 80 81 94 95 96 101
SC28	69 77 94 95 96
SC14	69 77 94 95 96
SC07	69 77 94 95 96
SC03	69 77 94

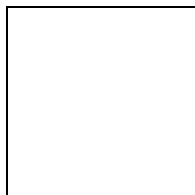


Figura 4: Histograma empírico para a distribuição da variável aleatória do logaritmo da razão de verossimilhança. Neste exemplo, foi considerado o conjunto de 10000 simulações de Monte Carlo para o mapa de homicídios tentados com limitação de tamanho máximo de 14

regiões, pelo algoritmo de Simulated Annealing (veja também a figura 5C e tabela 2.2: O valor $\text{Log}(T)=19.971$ situa-se no extremo direito do histograma).

6- CONCLUSÃO

Foram encontrados conglomerados estatisticamente significativos para homicídios consumados e tentados no município de Ribeirão das Neves para o ano de 2000. A significância estatística desses conglomerados é alta conforme os testes de Monte Carlo que efetuamos neste trabalho.

Algumas das regiões com maior incidência de crimes coincidem com a localização geográfica da Penitenciária de Neves, corroborando a hipótese de que ela tem influência na ocorrência de crimes violentos devido às constantes tentativas de fugas e rebeliões. Isso ficou claro quando fizemos o estudo dos homicídios tentados.

Um ponto interessante é que os testes detectaram conglomerados na área de Veneza e San Genaro, também consideradas publicamente entre as de maior incidência criminal, tanto para homicídios tentados como para consumados. Estas regiões são bastante carentes, com péssima infra estrutura urbana, e com um histórico de muitas ocorrências de tráfico de drogas e roubos. Veneza é hoje um dos bairros mais populosos de Ribeirão das Neves e infelizmente seu crescimento acentuado ocorreu de forma muito desordenada.

Assim podemos concluir que as duas grandes regiões de Ribeirão das Neves que merecem a maior atenção das autoridades no momento são as regiões próximas à Penitenciária de Neves e o bairro Veneza.

7- REFERÊNCIAS BIBLIOGRÁFICAS:

1. Aarts E, Korst J. *Simulated Annealing and Boltzmann Machines*, Wiley: Chichester, 1989.
2. Duczmal L., Assunção R. - A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics and Data Analysis* (2003, in press).
3. Kulldorff M, Nagarwalla N. Spatial Disease Clusters: Detection and Inference. *Statistics in Medicine*, 1995; 14:779-810
4. Kulldorff, M. A spatial scan statistic, *Communications in Statistics: Theory and Methods*, 26 (1997) 1481-1496.
5. Kulldorff M. Spatial scan statistics: Models, calculations and applications. In *Scan Statistics and Applications*, Glaz and Balakrishnan (eds.). Boston: Birkhauser, 1999;303-322.
6. Lawson A., Biggeri A., Böhning D. *Disease mapping and risk assessment for public health*. New York, John Wiley and Sons, 1999.

