
SaTScan™ Manual do Usuário

para versão 9.4

Versão Original do Manual

Martin Kulldorff

Fevereiro de 2015

<http://www.satscan.org/>

Versão do Manual Traduzido para o Português:

Tradutor: Alessandra Cristina Guedes Pellini

Maio de 2016

<http://www.satscan.org/>

Índice

Introdução.....	1
O Software SaTScan.....	1
Download e Instalação.....	2
Execução de Teste	2
Sistema de Ajuda.....	3
Conjuntos de Dados de Amostras.....	3
Metodologia Estatística.....	6
Estatísticas de Varredura Espacial, Temporal e Espaço-Temporal.....	6
Modelo de Bernoulli	8
Modelo Discreto de Poisson.....	8
Modelo de Permutação no Espaço-Tempo.....	9
Modelo Multinomial	9
Modelo Ordinal.....	10
Modelo Exponencial	11
Modelo Normal.....	11
Modelo Contínuo de Poisson.....	12
Comparação de Modelos de Probabilidade.....	13
Teste da Razão de Verossimilhança.....	15
Clusters Secundários.....	16
Ajuste por Clusters Mais Prováveis.....	16
Ajustes por Covariáveis.....	17
Ajustes Espaciais e Temporais.....	19
Dados Faltantes.....	21
Varredura Multivariada com Múltiplos Conjuntos de Dados.....	23
Comparação com Outros Métodos.....	24
Estatísticas de Varredura.....	24
Aglomerados Espaciais e Espaço-Temporais.....	24
Dados de Entrada.....	26
Requisitos de Dados.....	26
Arquivo de Casos.....	27
Arquivo de Controles.....	28
Arquivo de População.....	28
Arquivo de Coordenadas.....	29
Arquivo de Grade.....	30
Arquivo de Vizinhos Não-Euclidianos.....	30
Arquivo de Meta-Localidade.....	31
Arquivo de Tamanho Máximo do Círculo.....	31
Arquivo de Ajustes.....	32
Arquivo de Hipótese Alternativa.....	33
Assistente de Importação do SaTScan.....	33
Formato de Arquivo ASCII do SaTScan.....	35
Recursos Básicos do SaTScan.....	37
Guia de Entrada.....	37
Guia de Análise.....	40
Guia de Saída.....	44
Recursos Avançados.....	46
Guia de Múltiplos Conjuntos de Dados.....	46
Guia de Verificação de Dados.....	47
Guia de Vizinhos Espaciais.....	48
Guia da Janela Espacial.....	49
Guia da Janela Temporal.....	52
Guia de Ajustes Espaciais e Temporais.....	54

Guia de Inferência.....	56
Guia de Análise de Fronteira.....	59
Guia de Estimativa de Poder.....	61
Guia de Saída Espacial.....	63
Guia de Saída Temporal.....	66
Guia de Outras Saídas.....	67
Executando o SaTScan.....	68
Especificando as Opções de Análises e Dados.....	68
Iniciando as Análises.....	68
Mensagens de Status.....	68
Alertas e Erros.....	69
Salvando os Parâmetros das Análises.....	70
Processadores Paralelos.....	70
Modo de Lote.....	70
Tempo de Computação.....	71
Requisitos de Memória.....	72
Resultados das Análises.....	75
Arquivo Padrão de Resultados Baseado em Texto (*.out.*)......	75
Arquivo KML de Saída Geográfica (*.kml).....	77
Arquivo Shapefile de Saída Geográfica (*.shp and *.shx).....	77
Arquivo HTML de Gráficos Temporais (*.temporal.html).....	77
Arquivo de Informações do Cluster (*.col.*)......	78
Arquivo de Informações do Cluster Estratificado (*.sci.*)......	80
Arquivo de Informações das Localidades (*.gis.*)......	80
Arquivo de Estimativas de Risco para Cada Localidade (*.rr.*)......	81
Arquivo dos Log das Razões de Verossimilhança Simulados (*.llr.*)......	81
Miscelânea.....	82
Novas Versões.....	82
Arquivo do Histórico das Análises.....	82
Gerador de Números Aleatórios.....	82
Contate-Nos.....	83
Agradecimentos.....	83
Perguntas Frequentes.....	85
Dados de Entrada.....	85
Análise.....	86
Resultados.....	86
Interpretação.....	87
Sistemas Operacionais.....	89
Bibliografia do SaTScan.....	90
Citações Sugeridas.....	90
Artigos de Metodologia do SaTScan.....	91
Aplicações do SaTScan Seleccionadas por Área de Estudo.....	94
Outras Referências Mencionadas no Manual do Usuário.....	112

Introdução

O Software SaTScan

Finalidade

O SaTScan é um software livre que analisa dados espaciais, temporais e espaço-temporais usando as estatísticas de varredura espacial, temporal ou espaço-temporal. Ele foi projetado para quaisquer dos seguintes propósitos inter-relacionados:

- Realizar a vigilância geográfica de doenças, para detectar clusters espaciais ou espaço-temporais, e verificar se eles são estatisticamente significativos.
- Testar se uma doença é distribuída aleatoriamente no espaço, ao longo do tempo ou no espaço e no tempo.
- Avaliar a significância estatística de alertas de clusters de doenças.
- Realizar vigilância prospectiva de doenças em tempo real ou periódico, para a detecção precoce de surtos de doenças.

O software pode ser usado também para problemas similares em outros campos, tais como arqueologia, astronomia, botânica, criminologia, ecologia, economia, engenharia, silvicultura, genética, geografia, geologia, história, neurologia ou zoologia.

Tipos de Dados e Métodos

O SaTScan pode ser usado para estatísticas de varredura discretas ou contínuas. Para as estatísticas de varredura discretas, as localidades geográficas onde os dados são observados são não aleatórias e fixadas pelo usuário. Estas localidades podem ser as localidades reais das observações, tais como as casas, escolas ou ninhos, ou poderia ser uma localidade central representando uma área maior, tal como centroides geográficos ou ponderados por população, de áreas postais, municípios ou províncias. Para as estatísticas de varredura contínuas, as localidades das observações são aleatórias e podem ocorrer em qualquer lugar dentro de uma área de estudo predefinida pelo usuário, tal como um retângulo.

Para as estatísticas de varredura discretas, o SaTScan utiliza um modelo discreto baseado em Poisson, em que o número de eventos em uma localidade geográfica tem distribuição de Poisson, de acordo com uma população de base conhecida sob risco; ou um modelo de Bernoulli, com eventos de dados do tipo 0/1, tais como casos e controles; um modelo de permutação no espaço-tempo, usando apenas dados de casos; um modelo multinomial para dados categóricos; um modelo ordinal para dados categóricos ordenados; um modelo exponencial para dados de tempo de sobrevivência com ou sem variáveis censuradas; um modelo normal para outros tipos de dados contínuos; ou um modelo de variação espacial nas tendências temporais, procurando por áreas geográficas com tendências temporais excepcionalmente altas ou baixas. Uma característica comum a todas essas estatísticas de varredura discretas é que as localidades geográficas onde os dados podem ser observados são não aleatórias e fixadas pelo usuário.

Para as estatísticas de varredura discretas, os dados podem tanto ser agregados por setor censitário, CEP, município ou outro nível geográfico, ou pode haver coordenadas únicas para cada observação. O SaTScan ajusta para a heterogeneidade espacial subjacente de uma população de base. Também é possível se ajustar para qualquer número de covariáveis categóricas fornecidas pelo usuário, bem como para tendências temporais, agregados espaço-temporais conhecidos e dados faltantes. É possível realizar a varredura de múltiplos conjuntos de dados simultaneamente, para procurar por aglomerados que ocorrem em um ou mais destes.

Para as estatísticas de varredura contínuas, o SaTScan utiliza um modelo contínuo de Poisson.

Promotores e Financiadores

O software SaTScan™ foi desenvolvido por Martin Kulldorff, com o apoio financeiro dos Serviços de Gestão de Informações Inc. para o SaTScan, recebido das seguintes instituições.

- Instituto Nacional do Câncer, Divisão de Prevenção do Câncer, Seção de Biometria [v1.0, 2.0, 2.1]
- Instituto Nacional do Câncer, Divisão de Controle do Câncer e Ciências da População, Seção de Pesquisas e Aplicações Estatísticas [v3.0 (parcial), v6.1 (parcial), 8.0 (parcial), v9.0 (parcial)]
- Fundação Alfred P. Sloan, através de um subsídio para a Academia de Medicina de New York (Farzad Mostashari, PI) [v3.0 (parcial), 3.1, 4.0, 5.0, 5.1]
- Centros de Prevenção e Controle de Doenças, através do Acordo de Cooperação da Associação de Escolas Médicas Americanas, número de concessão MM-0870 [v6.0, 6.1 (parcial)]
- Instituto Nacional de Saúde e Desenvolvimento Infantil, através da concessão #RO1HD048852 [7.0, 8.0, 9.0 (parcial)]
- Instituto Nacional do Câncer, Divisão de Epidemiologia e Genética do Câncer [v9.0 (parcial),v9.1]
- Instituto Nacional de Ciências Médicas Gerais, através de uma concessão dos Estudos de Modelagem de Agentes de Doenças Infecciosas #U01GM076672 [v9.0 (parcial)]

Seus apoios financeiros são muito apreciados. Os conteúdos do SaTScan são de responsabilidade do desenvolvedor e não refletem necessariamente as opiniões oficiais dos patrocinadores.

Tópicos Relacionados: *Metodologia Estatística, Bibliografia do SaTScan.*

Download e Instalação

Para instalar o SaTScan, acessar o Web site do SaTScan em: <http://www.satscan.org/> e selecionar o link para o download do SaTScan. Após realizar o download do arquivo executável de instalação do SaTScan para o seu PC, clicar neste ícone e instalar o software seguindo as instruções passo-a-passo.

Tópicos Relacionados: *Novas Versões.*

Execução de Teste

Antes de usar seus próprios dados, recomenda-se testar um dos conjuntos de dados amostrais fornecidos com o software. Use-os para ter uma ideia de como executar o SaTScan. Para executar um teste:

1. Clicar no ícone do aplicativo SaTScan.
2. Clicar em 'Abrir Sessão Salva'.
3. Selecionar um dos arquivos de parâmetros, por exemplo 'nm.prm' (modelo de Poisson), 'NHumberside.prm' (modelo de Bernoulli) ou 'NYCfever.prm' (modelo de permutação no espaço-tempo).
4. Clicar em 'Abrir'.
5. Clicar no botão 'Executar' . Uma nova janela irá se abrir com o programa em execução na seção superior, e uma seção de Alertas/Erros abaixo. Quando o programa finalizar a execução, os resultados serão exibidos.

Nota: Os arquivos de exemplos não devem produzir alertas ou erros.

Tópicos Relacionados: *Conjuntos de Dados de Amostras.*

Sistema de Ajuda

O sistema de ajuda do SaTScan consiste de quatro partes:

- i. Manual do Usuário do SaTScan em formato PDF, localizado na mesma pasta do arquivo executável do SaTScan. Ele também pode ser obtido no site do SaTScan (www.satscan.org/techdoc.html) ou diretamente dentro do software SaTScan, selecionando Ajuda > Manual do Usuário. É possível imprimir este manual como um único documento para fácil referência.
- ii. Entradas de ajuda do SaTScan, extraídas do Manual do Usuário. O conjunto completo de entradas pode ser encontrado dentro do software SaTScan, digitando-se F1 ou selecionando 'Ajuda' > 'Conteúdo da Ajuda'. O sistema pode ser pesquisado clicando-se na lupa. Muitas entradas individuais podem ser também acessadas diretamente, clicando-se em qualquer subtítulo visto nas abas/guias de entrada.
- iii. Os artigos metodológicos descrevem os detalhes sobre os métodos estatísticos disponíveis no software SaTScan. Esses artigos estão relacionados na bibliografia do SaTScan, que pode ser encontrada tanto no final do Manual do Usuário como na web (<http://www.satscan.org/references.html>). A bibliografia também contém um grande número de artigos que aplicaram diferentes funcionalidades do SaTScan para uma ampla gama de diversos tipos de dados. Estes podem servir de inspiração de como o SaTScan pode ser usado para tipos variados de problemas científicos e de saúde pública.
- iv. Conjuntos de dados de amostras são fornecidos para cada um dos modelos de probabilidade do SaTScan. Descritos abaixo, eles tornam mais fácil se familiarizar com o software.

Conjuntos de Dados de Amostras

Seis conjuntos diferentes de dados amostrais são fornecidos com o software. Eles são automaticamente transferidos para o computador, junto com o próprio software. Outros conjuntos de dados de amostras estão disponíveis em <http://www.satscan.org/datasets/>.

Modelo Discreto de Poisson, Espaço-Temporal e Variação Espacial nas Tendências Temporais: Incidência de Câncer de Cérebro no Novo México.

Arquivo de casos: nm.cas

Formato: <município> <casos=1> <ano> <faixa etária> <sexo>

Arquivo de população: nm.pop

Formato: <município> <ano> <população> < faixa etária> <sexo>

Arquivo de coordenadas: nm.geo

Formato: <município> <latitude> <longitude>

Período de estudo: 1973-1991

Agregação: 32 municípios

Precisão de tempo dos casos: Anos

Coordenadas: Latitude/Longitude

Covariável #1, grupos etários: 1 = 0-4 anos, 2 = 5-9 anos, ... 18 = 85+ anos

Covariável #2, gênero: 1 = masculino, 2 = feminino

Anos populacionais: 1973, 1982, 1991

Fonte de dados: Registro de tumores SEER do Novo México

Esta é uma versão condensada de um conjunto de dados mais completo com uma dada população para cada ano, de 1973 a 1991, e com a etnia como uma terceira covariável. O conjunto de dados completo pode ser encontrado em: <http://www.satscan.org/datasets/>

Modelo de Bernoulli, Puramente Espacial: Leucemia Infantil e Incidência de Linfoma no Norte de Humberside

Arquivo de casos: NHumberside.cas

Formato: <id da localidade> <# casos>

Arquivo de controles: Nhumberside.ctl

Formato: <id da localidade> <# controles>

Arquivo de coordenadas: Nhumberside.geo

Formato: <id da localidade> <coordenada-x> <coordenada-y>

Período de estudo: 1974-1986

Controles: Selecionados aleatoriamente a partir do registro de nascimentos

Agregação: 191 Códigos Postais (a maioria com apenas um único indivíduo)

Precisão de tempo dos casos e controles: Nenhuma

Coordenadas: Cartesianas

Covariáveis: Nenhuma

Fonte de dados: Drs. Ray Cartwright e Freda Alexander. Publicado por J. Cuzick e R. Edwards, Journal of the Royal Statistical Society, B:52 73-104, 1990

Modelo de Permutação no Espaço-Tempo: Admissões Hospitalares na Sala de Emergência Devido à Febre nos Hospitais de Nova York

Arquivos de casos: NYCfever.cas

Formato: <código postal> <#casos=1> <data>

Arquivo de coordenadas: NYCfever.geo

Formato: <código postal> <latitude> <longitude>

Período de estudo: 01 Nov, 2001 - 24 Nov, 2001

Agregação: Áreas de código postal

Precisão de tempo dos casos: Dias

Coordenadas: Latitude/Longitude

Covariáveis: Nenhuma

Fonte de dados: Departamento de Saúde da Cidade de New York

Modelos Multinomial e Ordinal, Puramente Espacial: Níveis de Desempenho da Educação em Maryland

Arquivo de casos: MarylandEducation.cas

Formato: <município> <# indivíduos> <categoria #>

Arquivo de coordenadas: MarylandEducation.geo

Formato: <município> <latitude> <longitude>

Período de estudo: 2000

Agregação: 24 Condados e Municípios Equivalentes

Precisão de tempo dos casos: Nenhuma

Coordenadas: Latitude / Longitude

Covariáveis: Nenhuma

Categorias:

1 = Menos que 9ª série

2 = 9ª a 12ª série, mas nenhum diploma de ensino médio

3 = Diploma de ensino médio, mas nenhum grau de bacharel

4 = Bacharel ou grau mais elevado

Fonte de dados: Departamento do Censo dos Estados Unidos: As informações sobre educação derivam do formulário longo do Censo de 2000, preenchido em cerca de 1/6 das famílias.

Nota: Apenas pessoas com idade de 25 anos ou mais estão incluídas nos dados. Para cada município, o censo fornece informações sobre a porcentagem de pessoas com diferentes níveis de educação formal. O número de indivíduos que relataram diferentes níveis de ensino em cada município foi estimado como este percentual vezes a população total com idade de 25+ dividido por seis, para refletir a fração amostral de 1/6 do formulário longo do censo.

Modelo Exponencial, Espaço-Tempo: Dados de Sobrevida Criados Artificialmente

Arquivo de casos: SurvivalFake.cas

Formato: <id da localidade> <# indivíduos> <momento do diagnóstico> <tempo de sobrevida>
<censurado>

Arquivo de coordenadas: SurvivalFake.geo

Formato: <id da localidade> <coordenada-x> <coordenada-y>

Período de estudo: 2000-2005

Agregação: 5 Localidades

Precisão do tempo de diagnóstico: Ano

Precisão do tempo de sobrevida/censura: Dia

Coordenadas: Cartesianas

Covariáveis: Nenhuma

Fonte de dados: Dados criados artificialmente.

Modelo Normal, Puramente Espacial: Dados Contínuos Criados Artificialmente

Arquivo de casos: NormalFake.cas

Formato: <id da localidade> <# indivíduos> <aumento de peso>

Arquivo de coordenadas: NormalFake.geo

Formato: <id da localidade> <coordenada-x> <coordenada-y>

Período de estudo: 2006

Agregação: 26 Localidades

Coordenadas: Cartesianas

Covariáveis: Nenhuma

Fonte de dados: Dados criados artificialmente.

Tópicos Relacionados: *Execução de Teste, Dados de Entrada.*

Metodologia Estatística

Para todas as análises discretas espaciais e espaço-temporais, o usuário deve fornecer dados contendo as coordenadas espaciais de um conjunto de localidades (arquivo de coordenadas). Para cada localidade, os dados precisam conter ainda informações sobre o número de casos em cada localidade (arquivo de casos). Para as análises temporais e espaço-temporais, o número de casos precisa ser estratificado pelo tempo, por exemplo, o tempo do diagnóstico. Dependendo do tipo de análise, outras informações sobre os casos, tais como idade, sexo, peso, duração de sobrevivência ou estágio do câncer podem também ser fornecidas. Para o modelo de Bernoulli, também é necessário especificar o número de controles em cada localidade (arquivo de controle). Para o modelo discreto de Poisson, o usuário precisa especificar um tamanho de população para cada localidade (arquivo de população). A população pode variar ao longo do tempo.

Estatísticas de varredura são usadas para detectar e avaliar clusters (aglomerados) de casos em uma configuração puramente temporal, puramente espacial ou espaço-temporal. Isso é feito pela varredura gradual de uma janela ao longo do tempo e/ou do espaço, assinalando o número de observações observadas e esperadas no interior da janela em cada localidade. No software SaTScan, a janela de varredura é um intervalo (no tempo), um círculo ou uma elipse (no espaço) ou um cilindro com uma base circular ou elíptica (no espaço-tempo). Também é possível se especificar uma estrutura própria de distâncias não-Euclidianas em um arquivo especial. Inúmeros tamanhos diferentes de janelas são utilizados. A janela com a máxima verossimilhança é o aglomerado mais provável, ou seja, o cluster que tem a menor probabilidade de ter ocorrido ao acaso. Um valor de 'p' é atribuído a este cluster.

A estatística de varredura usa diferentes modelos de probabilidades, dependendo da natureza dos dados. Um modelo de Bernoulli, discreto de Poisson ou de permutação no espaço-tempo pode ser usado para dados de contagem, tal como o número de pessoas com asma; o modelo multinomial é usado para dados categóricos, tal como a histologia de um câncer; um modelo ordinal para dados categóricos ordenados, tal como o estadiamento de câncer; um modelo exponencial para dados de tempo de sobrevivência, com ou sem censura; e um modelo normal para outros dados contínuos, tais como o peso ao nascer ou níveis de chumbo no sangue. A teoria estatística geral por trás das estatísticas de varredura espacial e espaço-temporal utilizadas no software SaTScan está descrita em detalhes por Kulldorff (1997)¹ para os modelos de Bernoulli, discreto de Poisson e Poisson contínuo; por Kulldorff et al. (2005)⁵ para o modelo de permutação no espaço-tempo; por Jung et al. (2008)⁶, para o modelo multinomial; por Jung et al. (2007)⁷ para o modelo ordinal; por Huang et al. (2006)⁸, para o modelo exponencial; por Kulldorff et al. (2009)⁹ para o modelo normal e por Huang et al. (2009)¹⁰ para o modelo normal com pesos. Recomenda-se a leitura desses trabalhos para uma descrição detalhada de cada modelo. Aqui é apenas fornecida uma breve descrição não-matemática.

Para todos os modelos de probabilidade discretos, a estatística de varredura ajusta para a densidade geográfica irregular de uma população de base. Para todos os modelos, as análises estão condicionadas ao número total de casos observados.

Tópicos Relacionados: *O software SaTScan, Recursos Básicos do SaTScan, Recursos Avançados, Guia de Análise, Artigos Metodológicos.*

Estatísticas de Varredura Espacial, Temporal e Espaço-Temporal

Estatística de Varredura Espacial

A estatística de varredura puramente espacial padrão estabelece uma janela circular no mapa. A janela é, por sua vez, centrada em cada um dos vários pontos da grade possíveis posicionados em toda a região do estudo. Para cada ponto da grade, o raio da janela varia continuamente em tamanho a partir de zero até um limite superior especificado pelo usuário. Deste modo, a janela circular é flexível, tanto na localização quanto no tamanho. No total, o método cria um número infinito de círculos geográficos distintos, com diferentes conjuntos de localidades de dados de vizinhança dentro deles. Cada círculo é um possível candidato do cluster.

O usuário define o conjunto de pontos de grade usados através de um arquivo de grade. Se nenhum arquivo de grade for especificado, os pontos de grade serão idênticos às coordenadas dos IDs das localidades definidas no arquivo de coordenadas. Essa última opção assegura que cada localidade de dados seja um potencial cluster em si mesmo, sendo a opção recomendada para a maioria dos tipos de análises.

Como uma alternativa para o círculo, é possível também se utilizar uma janela de formato elíptico, situação na qual um conjunto de elipses com diferentes formatos e ângulos é usado como janela de varredura em conjunto com o círculo. Isso proporciona um poder ligeiramente maior para aglomerados verdadeiros que possuem formatos longos e estreitos, e poder ligeiramente inferior para aglomerados circulares e outros muito compactos.

Também é possível definir sua própria métrica de distância não-Euclidiana usando um arquivo especial de vizinhos.

Tópicos Relacionados: *Guia de Análise, Arquivo de Coordenadas, Janela de Varredura Elíptica, Arquivo de Grade, Tamanho Máximo do Cluster Espacial, Guia da Janela Espacial.*

Estatística de Varredura Espaço-Temporal

A estatística de varredura espaço-temporal é definida por uma janela cilíndrica com uma base geográfica circular (ou elíptica), e com a altura correspondente ao tempo. A base é definida exatamente como que para a estatística de varredura puramente espacial, enquanto que a altura reflete o período de tempo de potenciais aglomerados. A janela cilíndrica é então movida no espaço e no tempo, de modo que para cada possível localidade e dimensão geográfica, também visita cada período de tempo possível. De fato, obtém-se um número infinito de cilindros sobrepostos, de diferentes tamanhos e formas, cobrindo em conjunto toda a região do estudo, onde cada cilindro reflete um possível cluster.

A estatística de varredura espaço-temporal pode ser utilizada para uma única análise retrospectiva, utilizando dados históricos, ou para a vigilância prospectiva periódica no tempo, onde a análise é repetida, por exemplo, a cada dia, semana, mês ou ano.

Tópicos Relacionados: *Guia de Análise, Guia da Janela Espacial, Guia da Janela Temporal.*

Estatística de Varredura Temporal

A estatística de varredura temporal utiliza uma janela que se desloca em uma dimensão, o tempo, definido da mesma forma que a altura do cilindro usado pela estatística de varredura espaço-temporal. Isto significa que, em ambas as análises, as datas de início e fim são flexíveis. A duração temporal máxima é especificada na Guia da Janela Temporal.

Tópicos Relacionados: *Guia de Análise, Guia da Janela Temporal, Estatística de Varredura Espaço-Temporal.*

Estatísticas de Varredura de Variação Espacial nas Tendências Temporais

Quando a estatística de varredura é usada para avaliar a variação espacial nas tendências temporais, a janela de varredura é de natureza puramente espacial. A tendência temporal é então calculada, tanto dentro como fora da janela de varredura, para cada localidade e tamanho dessa janela. A hipótese nula é de que as tendências são as mesmas, enquanto que a alternativa é que elas são diferentes. Com base nestas hipóteses, uma probabilidade é calculada, que será maior quanto mais improvável a diferença das tendências for devida ao acaso. O aglomerado mais provável é aquele para o qual a tendência temporal dentro da janela tiver a menor probabilidade de ser a mesma que a tendência temporal fora do cluster. Isto pode ocorrer por várias razões. Por exemplo, se a tendência temporal no interior do aglomerado é maior, isso poderia ser porque todas as áreas têm a mesma taxa de incidência de uma doença no início do período, mas a área do aglomerado apresenta uma taxa mais elevada ao final do período. Poderia ser também porque a área do cluster tem uma taxa de incidência mais baixa no início do período, após o qual ela 'alcança' o restante, de modo que a taxa é aproximadamente a mesma ao final do período. Portanto, um cluster estatisticamente significativo na análise da variação espacial na tendência temporal não significa, necessariamente, que a taxa global de doença seja maior ou menor no cluster.

A estatística de varredura da variação espacial nas tendências temporais só pode ser executada com o modelo de probabilidade discreto de Poisson. Para que ela funcione, é importante que a duração total do período de estudo seja divisível pelo comprimento do intervalo de tempo agregado, de modo que todos os intervalos de tempo tenham o mesmo número de anos, se for especificado em anos; o mesmo número de meses, se for especificado em meses; ou o mesmo número de dias, se for especificado em dias.

Tópicos Relacionados: *Guia de Análise, Estatística de Varredura Espaço-Temporal.*

Modelo de Bernoulli

Com o modelo de Bernoulli^{1,2}, existem casos e não-casos representados por uma variável 0/1. Estas variáveis podem representar pessoas com ou sem uma doença, ou pessoas com tipos diferentes de doenças, como estágio de câncer de mama precoce e tardio. Eles podem refletir casos e controles de uma população maior, ou eles podem, em conjunto, constituir a população como um todo. Seja qual for a situação, estas variáveis serão indicadas como casos e controles em todo o manual do usuário, e o seu total será indicado como a população. Dados de Bernoulli podem ser analisados com as estatísticas de varredura puramente temporal, puramente espacial ou espaço-temporal.

Exemplo: Para o modelo de Bernoulli, os casos podem ser recém-nascidos com um determinado defeito de nascença, enquanto que os controles são todos os recém-nascidos sem o defeito de nascença.

O modelo de Bernoulli requer informações sobre as localidades de um conjunto de casos e controles, fornecidos para o SaTScan usando os arquivos de casos, de controles e de coordenadas. Localidades distintas podem ser especificadas para cada caso e cada controle, ou os dados podem ser agregados por estados, províncias, municípios, paróquias, setores censitários, áreas de código postal, distritos escolares, domicílios, etc, com vários casos e controles em cada localidade de dados. Para executar uma análise temporal ou espaço-temporal, é necessário ter um tempo para cada caso e para cada controle.

Tópicos Relacionados: *Guia de Análise, Arquivo de Casos, Arquivo de Controles, Arquivo de Coordenadas, Teste da Razão de Verossimilhança, Artigos Metodológicos, Comparação de Modelos de Probabilidade.*

Modelo Discreto de Poisson

Com o modelo discreto de Poisson¹, o número de casos em cada localidade segue a distribuição de Poisson. Sob a hipótese nula, e quando não há covariáveis, o número esperado de casos em cada área é proporcional ao tamanho de sua população, ou das pessoas-ano nessa área. Dados de Poisson podem ser analisados com as estatísticas de varredura puramente temporal, puramente espacial, espaço-temporal e de variação espacial nas tendências temporais.

Exemplo: Para o modelo discreto de Poisson, os casos podem ser ocorrências de acidente vascular cerebral, enquanto que a população é o número combinado de pessoas-ano vividos, calculado como 1 para alguém vivendo na área durante todo o período de tempo, e ½ para alguém que morreu ou se mudou no meio do período de tempo.

O modelo discreto de Poisson requer a contagem de casos e da população para um conjunto de localidades de dados, tais como municípios, paróquias, setores censitários ou áreas de código postal, bem como as coordenadas geográficas de cada uma dessas localidades. Esses precisam ser fornecidas ao SaTScan usando os arquivos de casos, de população e de coordenadas.

Os dados de população não precisam ser especificados continuamente ao longo do tempo, mas apenas em uma ou mais 'épocas de recenseamento' específicas. Para os intervalos entre os censos, o SaTScan realiza uma interpolação linear com base nas populações nos momentos de recenseamento imediatamente anterior e imediatamente posterior. Para os períodos anteriores ao primeiro censo, o tamanho da população é definido como igual ao tamanho da população do primeiro censo, e para os períodos após o último censo, a população é definida como igual ao tamanho da população do último censo. Para se obter o tamanho da população para uma dada localidade e determinado período de

tempo, o tamanho da população, conforme definido acima, está integrado ao longo do período de tempo em questão.

Tópicos Relacionados: *Guia de Análise, Arquivo de Casos, Modelo Contínuo de Poisson, Arquivo de Coordenadas, Teste da Razão de Verossimilhança, Artigos Metodológicos, Arquivo de População, Comparação de Modelos de Probabilidade.*

Modelo de Permutação no Espaço-Tempo

O modelo de permutação no espaço-tempo⁵ requer apenas dados de casos, com informações sobre a localização espacial e o tempo de cada caso, não sendo necessária nenhuma informação sobre controles ou uma população de base sob risco. O número de casos observados em um cluster é comparado ao que seria esperado se as localidades espaciais e temporais de todos os casos fossem independentes umas das outras, de modo que não houvesse qualquer interação espaço-temporal.

Ou seja, existe um aglomerado em uma área geográfica se, durante um período de tempo específico, esta área tiver uma maior proporção dos seus casos nesse período de tempo em comparação às áreas geográficas restantes. Isto significa que, se durante uma semana específica todas as áreas geográficas tiverem duas vezes o número normal de casos, nenhuma dessas áreas constitui um aglomerado. Por outro lado, se durante essa semana uma área geográfica tiver o dobro do número de casos em relação ao normal, enquanto que outras áreas têm uma quantidade normal de casos, então haverá um cluster naquela primeira área. O modelo de permutação no espaço-tempo ajusta automaticamente para ambos os aglomerados, puramente espacial e puramente temporal. Assim, não existem versões puramente temporais ou puramente espaciais deste modelo.

Exemplo: No modelo de permutação no espaço-tempo, casos pode ser ocorrências diárias de ambulâncias enviadas para pacientes com AVC.

É importante perceber que os aglomerados de permutação no espaço-tempo podem ocorrer tanto devido a um risco aumentado da doença, como por distribuições geográficas diferentes da população em momentos distintos, onde, por exemplo, a população em algumas áreas cresce mais depressa do que em outras. Isto não é normalmente um problema se o período total do estudo for menor do que um ano. No entanto, aconselha-se muito cuidado ao se usar esse método para dados que abrangem vários anos. Se a população de base aumenta ou diminui mais rápido em algumas áreas do que em outras, há um risco de viés de mudança da população, o que pode produzir valores de 'p' tendenciosos quando o período de estudo for maior do que alguns anos. Por exemplo, se um novo grande bairro se desenvolve, haverá um aumento de casos lá simplesmente porque aumentou a população, e usando-se apenas dados de casos, o modelo de permutação no espaço-tempo não pode distinguir um aumento devido a um incremento da população local versus um aumento no risco da doença. Tal como acontece com todos os métodos de interação no espaço-tempo, esta é uma preocupação principalmente quando o período de estudo é maior do que alguns anos^{180,182}. Se o aumento (ou a redução) da população for o mesmo em todas as regiões do estudo, então está ok, e isso não conduzirá a resultados tendenciosos.

Tópicos Relacionados: *Guia de Análise, Arquivo de Casos, Arquivo de Coordenadas, Teste da Razão de Verossimilhança, Artigos Metodológicos, Comparação de Modelos de Probabilidade.*

Modelo Multinomial

Com o modelo multinomial⁶, cada observação é um caso, e cada caso pertence a uma de várias categorias. A estatística de varredura multinomial avalia se existem quaisquer clusters onde a distribuição dos casos é diferente do resto da região do estudo. Por exemplo, pode haver uma maior proporção de casos dos tipos 1 e 2 e uma menor proporção de casos do tipo 3, enquanto que a proporção de casos do tipo 4 é aproximadamente a mesma que fora do cluster. Se houver apenas duas categorias, o modelo ordinal é idêntico ao modelo de Bernoulli, em que uma categoria representa os casos e a outra categoria representa os controles. Os casos no modelo multinomial podem ser uma amostra de uma população maior ou podem constituir um conjunto completo de observações. Dados multinomiais podem ser analisados com as estatísticas de varredura puramente temporal, puramente espacial ou espaço-temporal.

Exemplo: Para o modelo multinomial, os dados podem consistir de todas as pessoas diagnosticadas com meningite, com cinco categorias diferentes representando cinco complexos clonais diferentes da doença⁶. A estatística de varredura multinomial irá, simultaneamente, procurar por clusters altos ou baixos de qualquer um dos complexos clonais, ou um grupo deles, ajustando para a distribuição geográfica global da doença. As múltiplas comparações inerentes às muitas categorias utilizadas são contabilizadas no cálculo dos valores de 'p'.

O modelo multinomial requer informações sobre a localidade de cada caso em cada categoria. Uma única localidade pode ser especificada para cada caso, ou os dados podem ser agregados por estados, províncias, municípios, paróquias, setores censitários, áreas de código postal, distritos escolares, domicílios, etc, com vários casos na mesma localidade. Para se executar uma análise temporal ou espaço-temporal, é necessário também ter um tempo atribuído para cada caso.

Com o modelo multinomial, não é necessário se especificar uma procura para clusters altos ou baixos, uma vez que não existe hierarquia entre as categorias, mas na saída se demonstra quais categorias são mais proeminentes dentro do cluster. A ordem ou a indexação das categorias não afeta a análise em termos dos aglomerados encontrados, mas pode influenciar a randomização usada para calcular os valores de 'p'.

Tópicos Relacionados: *Guia de Análise, Arquivo de Casos, Arquivo de Coordenadas, Teste da Razão de Verossimilhança, Artigos Metodológicos, Comparação de Modelos de Probabilidade.*

Modelo Ordinal

Com o modelo ordinal⁷, cada observação é um caso, e cada caso pertence a uma das várias categorias ordinais. Se existem apenas duas categorias, o modelo ordinal é idêntico ao modelo de Bernoulli, em que uma categoria representa os casos e a outra categoria representa os controles no modelo de Bernoulli. Os casos no modelo ordinal podem ser uma amostra de uma população maior ou eles podem constituir um conjunto completo de observações. Dados ordinais podem ser analisados com as estatísticas de varredura puramente temporal, puramente espacial ou espaço-temporal.

Exemplo: Para o modelo ordinal, os dados podem consistir de todos aqueles indivíduos diagnosticados com câncer de mama durante um período de dez anos, com três categorias diferentes que representam o câncer em estágio precoce, médio e tardio no momento do diagnóstico.

O modelo ordinal requer informações sobre a localidade de cada caso em cada categoria. Localidades separadas podem ser especificadas para cada caso, ou os dados podem ser agregados por estados, províncias, municípios, paróquias, setores censitários, áreas de código postal, distritos escolares, domicílios, etc, com vários casos na mesma ou em diferentes categorias em cada localidade de dados. Para se executar uma análise temporal ou espaço-temporal, é necessário ter também um tempo para cada caso.

Com o modelo ordinal, é possível procurar por aglomerados altos, com um excesso de casos nas categorias de valores altos; para aglomerados baixos, com um excesso de casos nas categorias de valores baixos, ou simultaneamente por ambos os tipos de aglomerados. Inverter a ordem das categorias tem o mesmo efeito que alterar a análise de alto para baixo e vice-versa.

Tópicos Relacionados: *Guia de Análise, Arquivo de Casos, Arquivo de Coordenadas, Teste da Razão de Verossimilhança, Artigos Metodológicos, Comparação de Modelos de Probabilidade.*

Modelo Exponencial

O modelo exponencial⁸ é concebido para dados de tempo de sobrevida, embora possa ser utilizado também para outros tipos de dados contínuos. Cada observação representa um caso, e cada um dos casos tem um atributo de variável contínua, bem como uma designação 0/1 para a censura. Para os dados de sobrevida, a variável contínua é o tempo entre o diagnóstico e a morte ou, dependendo da aplicação, entre dois outros tipos de eventos. Se algum dos dados é censurado devido à perda de seguimento, a variável contínua é então o tempo entre o diagnóstico e o tempo da censura. A variável de censura 0/1 é usada para distinguir entre observações censuradas e não censuradas.

Exemplo: Para o modelo exponencial, os dados podem consistir de todos os diagnosticados com câncer da próstata durante um período de dez anos, com informações sobre o período de tempo desde o diagnóstico até a morte, ou a partir do diagnóstico até um momento de censura após o qual a sobrevida é desconhecida.

Ao se utilizar o modelo exponencial temporal ou espaço-temporal para tempos de sobrevida, é importante perceber que existem duas variáveis de tempo muito diferentes envolvidas. A primeira é o momento em que o caso foi diagnosticado, e que é o tempo que a janela de varredura temporal e espaço-temporal está examinando. A segunda é o tempo de sobrevida, ou seja, o tempo entre o diagnóstico e a morte ou, para dados censurados, o tempo entre o diagnóstico e a censura. Este é um atributo de cada caso, e nenhuma varredura é feita sobre essa variável. Ao invés disso, estamos interessados em saber se a janela de varredura inclui excepcionalmente muitos casos com um valor pequeno ou grande deste atributo.

É importante notar que, enquanto o modelo exponencial utiliza uma função de probabilidade baseada na distribuição exponencial, a verdadeira distribuição do tempo de sobrevida não precisa ser exponencial, e a inferência estatística (valor de 'p') é válida também para outras distribuições de tempo de sobrevida. A razão para isto é que a randomização não é feita gerando-se observações a partir da distribuição exponencial, mas sim, por permutação das localidades do espaço-tempo e os atributos tempo de sobrevida/censura das observações.

Tópicos Relacionados: *Teste da Razão de Verossimilhança, Guia de Análise, Comparação de Modelos de Probabilidade, Artigos Metodológicos.*

Modelo Normal

O modelo normal¹⁰ é projetado para dados contínuos. Para cada indivíduo ou para cada observação, chamada de caso, há um único atributo contínuo que pode ser negativo ou positivo. O modelo também pode ser usado para dados ordinais quando existem muitas categorias. Isto é, casos diferentes podem ter o mesmo valor de atributo.

Exemplo: Para o modelo normal, os dados podem consistir do peso ao nascer e setor censitário de residência para todos os recém-nascidos, com um interesse em se encontrar clusters com baixo peso ao nascer. Um indivíduo é, então, um 'caso'. Alternativamente, os dados podem consistir do peso médio ao nascer em cada setor censitário. Nessa situação, o setor censitário é o 'caso', e é importante usar o modelo normal ponderado, uma vez que cada média terá uma variância diferente, devido a um número diferente de nascimentos em cada setor.

É importante notar que, embora o modelo normal utilize uma função de probabilidade baseada na distribuição normal, a verdadeira distribuição do atributo contínuo não precisa ser normal. A inferência estatística (valor de 'p') é válida para qualquer distribuição contínua. A razão para isto é que a randomização não é feita por meio da geração de dados simulados a partir de uma distribuição normal, mas, em vez disso, permutando as localidades do espaço-tempo e o atributo contínuo (ex. peso ao nascer) das observações. Ainda que continuem a ser formalmente válidos, os resultados podem ser muito influenciados por valores atípicos extremos, por isso, pode ser aconselhável truncar (retirar) tais observações antes de fazer a análise.

No modelo normal padrão⁹, assume-se que cada observação é medida com a mesma variância. Isso pode não ser sempre o caso. Por exemplo, se uma observação se baseia em uma amostra maior em uma localidade e uma amostra menor em outra, então a variância da incerteza nas estimativas será maior para a amostra de menor tamanho. Se a confiabilidade das estimativas difere, deve-se usar então a estatística de varredura normal ponderada¹⁰, que leva em conta tais variações desiguais. A versão ponderada é obtida no SaTScan simplesmente especificando-se um peso para cada observação como uma coluna extra no arquivo de entrada. Este peso pode, por exemplo, ser proporcional ao tamanho da amostra usada para cada estimativa, ou pode ser o inverso da variância da observação.

Se todos os valores forem multiplicados por ou adicionados à mesma constante, a inferência estatística não irá mudar, o que significa que os mesmos clusters, com os mesmos logs de verossimilhança e valores de 'p' serão encontrados. Apenas as médias e variâncias estimadas irão diferir. Se o peso for o mesmo para todas as observações, então a estatística de varredura normal ponderada irá produzir os mesmos resultados que a versão normal padrão. Se todos os pesos forem multiplicados pela mesma constante, os resultados não serão alterados.

Tópicos Relacionados: *Guia de Análise, Teste da Razão de Verossimilhança, Artigos Metodológicos, Comparação de Modelos de Probabilidade.*

Modelo Contínuo de Poisson

Todos os modelos descritos acima são baseados em dados observados em localidades discretas, consideradas não aleatórias, tal como definido por uma malha regular ou irregular de pontos de localidades. Isto é, as localidades das observações são consideradas fixas, e avalia-se a aleatoriedade espacial da observação condicionando à malha. Por isso, todas essas são versões do que se denomina de estatísticas de varredura discretas¹⁷⁴. Em uma estatística de varredura contínua, as observações podem estar localizadas em qualquer lugar dentro de uma área de estudo, como um quadrado ou um retângulo. O aspecto estocástico dos dados consiste destas localizações espaciais aleatórias, e nesse caso há interesse em verificar se existem quaisquer aglomerados que seriam improváveis de ocorrer se as observações fossem distribuídas de forma aleatória e independente em toda a área de estudo. Sob a hipótese nula, as observações seguem um processo de Poisson espacial homogêneo com intensidade constante ao longo de toda a área de estudo, sem observações caindo fora da área de estudo.

Exemplo: Os dados podem consistir da localização de ninhos de aves em uma área de quilômetro quadrado de uma floresta. Pode haver interesse em se verificar se os ninhos estão distribuídos aleatoriamente no espaço, ou, em outras palavras, se existem clusters de ninhos de aves ou se eles estão localizados de forma independente uns dos outros.

No SaTScan, a área de estudo pode ser qualquer coleção de polígonos convexos, que são regiões convexas delimitadas por qualquer número de linhas retas. Triângulos, quadrados, retângulos, losangos, pentágonos e hexágonos são todos exemplos de polígonos convexos. No caso mais simples, existe apenas um polígono, mas a área de estudo pode ser também a união de vários polígonos convexos. Se a área de estudo não for convexa, pode-se dividi-la em vários polígonos convexos e definir cada um separadamente. A área de estudo não precisa ser contígua, e pode, por exemplo, consistir de cinco ilhas diferentes.

A análise é condicionada ao número total de observações no conjunto de dados. Assim, a estatística de varredura avalia apenas a distribuição espacial da observação, mas não o número de observações.

A função de verossimilhança utilizada como a estatística do teste é a mesma que para o modelo de Poisson para a estatística de varredura discreta, onde o número de casos esperados é igual ao número total de 'observações observadas', vezes o tamanho da janela de varredura, dividido pelo tamanho da área total do estudo. Como tal, este é um caso especial da estatística de varredura de tamanho de janela variável, descrita por Kulldorff (1997)¹. Quando a janela de varredura se estende para fora da área de estudo, a contagem esperada é ainda baseada no tamanho completo do círculo, ignorando o fato de que algumas partes do círculo possuem contagens esperadas de zero. Isso é para evitar aglomerados estranhos não circulares na fronteira da área de estudo. Uma vez que a análise se baseia nas randomizações de Monte Carlo, os valores de 'p' são automaticamente ajustados para esses efeitos de

fronteira. As contagens esperadas relatadas são, por conseguinte, baseadas no círculo pleno, assim, as razões Obs/Exp fornecidas devem ser vistas como um limite inferior para o valor verdadeiro sempre que o círculo se estender para fora da região espacial do estudo.

O modelo contínuo de Poisson só pode ser usado para dados puramente espaciais. Ele utiliza uma janela de varredura circular com raio que varia continuamente até um máximo especificado pelo usuário. Somente círculos centrados em uma das observações são considerados, conforme especificado no arquivo de coordenadas. Se o arquivo de grade opcional é fornecido, os círculos são, então, centrados nas coordenadas especificadas nesse arquivo. O modelo contínuo de Poisson não foi implementado para ser usado com uma janela elíptica.

Tópicos Relacionados: *Guia de Análise, Teste da Razão de Verossimilhança, Artigos Metodológicos, Modelo de Poisson, Comparação de Modelos de Probabilidade.*

Comparação de Modelos de Probabilidade

No SaTScan, há sete modelos de probabilidades diferentes para estatísticas de varredura discretas. Para dados de contagem, há três diferentes modelos de probabilidade: discreto de Poisson, Bernoulli e permutação no espaço-tempo. Os modelos ordinal e multinomial são desenhados para dados categóricos com e sem uma ordenação inerente, por exemplo, de baixo para alto. Há dois modelos para dados contínuos: normal e exponencial. Este último foi desenvolvido principalmente para dados do tipo sobrevida. Para as estatísticas de varredura contínuas há somente o modelo homogêneo de Poisson.

O modelo discreto de Poisson é geralmente o mais rápido de se executar. O modelo ordinal é tipicamente o mais lento.

Com o modelo discreto de Poisson e o modelo de permutação no espaço-tempo, um número ilimitado de covariáveis podem ser usadas para ajuste, por meio de sua inclusão nos arquivos de casos e de população. Com o modelo normal, também é possível se realizar o ajuste por covariáveis, incluindo-as nos arquivos de casos, mas apenas para as análises puramente espaciais. Com os modelos de Bernoulli, ordinal, exponencial e normal, pode-se ajustar por covariáveis usando múltiplos conjuntos de dados, o que limita o número de categorias de covariáveis que podem ser definidas, ou através de uma análise de regressão pré-processamento, feita antes de se executar o SaTScan.

Todos os modelos de probabilidade discretos podem ser usados tanto para localidades individuais como para dados agregados.

Com o modelo discreto de Poisson, dados de população são necessários apenas em pontos selecionados do tempo, e os números são interpolados nos intervalos entre eles. Deve se especificar um tempo para a população, mesmo para análises puramente espaciais. Independe do modelo utilizado, o tempo de um caso ou controle só precisa ser especificado para análises puramente temporais e espaço-temporais.

O modelo de permutação no espaço-tempo ajusta automaticamente para clusters puramente espaciais e puramente temporais. Para o modelo discreto de Poisson, pode-se ajustar para clusters puramente temporais e puramente espaciais de diferentes formas. Para os modelos de Bernoulli, ordinal, exponencial e normal, os ajustes espaciais e temporais podem ser feitos usando múltiplos conjuntos de dados, mas isso é limitado pelo número de diferentes conjuntos de dados permitidos, e é também muito mais intensivo ao computador.

Análises puramente temporais e espaço-temporais não podem ser executadas usando o modelo homogêneo de Poisson. A análise de variação espacial nas tendências temporais só pode ser realizada usando o modelo discreto de Poisson.

Poucos Casos em Comparação a Controles

Em uma análise puramente espacial, onde há poucos casos comparados a controles, digamos que menos do que 10%, o modelo discreto de Poisson é uma excelente aproximação ao modelo de Bernoulli. O Poisson pode também ser utilizado para dados do tipo 0/1 de Bernoulli, e pode ser preferível, uma vez que tem mais opções para vários tipos de ajustes, incluindo a capacidade de ajustar para covariáveis especificadas nos arquivos de casos e de população. Como uma aproximação para dados do tipo de Bernoulli, o modelo discreto de Poisson produz valores de 'p' ligeiramente conservadores.

Modelo de Bernoulli versus Ordinal

O modelo de Bernoulli é matematicamente um caso especial do modelo ordinal, quando existem apenas duas categorias. O modelo de Bernoulli é executado mais rápido, tornando-se o modelo preferido para se usar quando existem apenas duas categorias.

Modelo Normal versus Exponencial

Ambos os modelos normal e exponencial são destinados para dados contínuos. O modelo exponencial é projetado principalmente para os dados de tempo de sobrevivência, mas pode ser usado para quaisquer dados onde todas as observações são positivas. É especialmente adequado para os dados com um desvio caudal importante à direita. O modelo normal pode ser usado para dados contínuos que assumem tanto valores positivos como negativos. Embora continuem formalmente válidos, os resultados do modelo normal são sensíveis a valores atípicos extremos.

Modelo Normal versus Ordinal

O modelo normal pode ser usado para dados categóricos quando existem muitas categorias. Como tal, é às vezes uma alternativa computacionalmente mais rápida em relação ao modelo ordinal. Contudo, há uma importante diferença. Com o modelo ordinal, somente a ordem dos valores observados importa. Por exemplo, os resultados são os mesmos para os valores ordenados '1 – 2 – 3 – 4' e '1 – 10 – 100 – 1000'. Com o modelo normal, os resultados serão diferentes, visto que eles dependem da distância relativa entre os valores utilizados para definir as categorias.

Modelo de Poisson Discreto versus Homogêneo

Em vez de usar o modelo homogêneo de Poisson, os dados podem ser aproximados pelo modelo discreto de Poisson, dividindo-se a área de estudo em diversos pequenos pedaços. Para cada pedaço, um único ponto de coordenadas é especificado, o tamanho do pedaço é utilizado para definir a população naquela localidade e o número de observações dentro desse pequeno pedaço de área é o número de casos nesta localidade. À medida que o número de pedaços tende ao infinito e, por conseguinte, seu tamanho diminui para zero, o modelo discreto de Poisson será assintoticamente equivalente ao modelo homogêneo de Poisson.

Dados Temporais

Para os dados temporais e espaço-temporais, há uma diferença adicional entre os modelos de probabilidade, na forma como os dados temporais são manipulados. Com o modelo de Poisson, os dados populacionais podem ser especificados em um ou em vários pontos no tempo, tais como anos censitários. Assume-se então que a população existe entre esses pontos no tempo, estimada através de interpolação linear entre os anos censitários. Com os modelos de Bernoulli, permutação no espaço-tempo, ordinal, exponencial e normal, precisa ser especificado um tempo para cada caso, e para o modelo de Bernoulli, também para cada controle.

Tópicos Relacionados: *Modelo de Bernoulli, Modelo de Poisson, Modelo de Permutação no Espaço-Tempo, Teste da Razão de Verossimilhança, Artigos Metodológicos.*

Teste da Razão de Verossimilhança

Para cada localidade e tamanho da janela de varredura, a hipótese alternativa é de que existe um risco elevado no interior da janela em comparação ao exterior. Sob a premissa de Poisson, a função de verossimilhança para uma janela específica é proporcional a¹:

$$\left(\frac{c}{E[c]}\right)^c \left(\frac{C-c}{C-E[c]}\right)^{C-c} I()$$

onde 'C' é número total de casos, 'c' é o número de casos observados dentro da janela e 'E(c)' é o número de casos esperados ajustado por covariáveis dentro da janela sob a hipótese nula. Nota-se que, uma vez que a análise é condicionada ao número total de casos observados, 'C-E[c]' é o número de casos esperados fora da janela. I() é uma função indicadora. Quando o SaTScan está configurado para a varredura apenas de clusters com altas taxas, I() é igual a 1 quando a janela tem mais casos do que o esperado sob a hipótese nula, e 0, do contrário. O oposto é verdadeiro quando o SaTScan está configurado para fazer a varredura apenas para clusters com taxas baixas. Quando o programa realiza a varredura para clusters com taxas altas ou baixas, então I() = 1 para todas as janelas.

O modelo de permutação no espaço-tempo utiliza a mesma função que o modelo de Poisson. Devido ao condicionamento nas marginais, o número de casos observados segue apenas aproximadamente a distribuição de Poisson. Portanto, este não é mais um teste formal da razão de verossimilhança, mas tem a mesma finalidade como a estatística do teste.

Para o modelo de Bernoulli a função de verossimilhança é^{1,2}:

$$\left(\frac{c}{n}\right)^c \left(\frac{n-c}{n}\right)^{n-c} \left(\frac{C-c}{N-n}\right)^{C-c} \left(\frac{(N-n)-(C-c)}{N-n}\right)^{(N-n)-(C-c)} I()$$

onde 'c' e 'C' são definidos como anteriormente, 'n' é o número total de casos e controles no interior da janela, enquanto 'N' é o número total de casos e controles combinados no conjunto de dados.

A função de verossimilhança para os modelos multinomial, ordinal, exponencial e normal é mais complicada, devido à natureza mais complexa dos dados. Referem-se os artigos de Jung, Kulldorff e Richards⁶, Jung, Kulldorff e Klassen⁷; Huang, Kulldorff e Gregorio⁸; Kulldorff et al⁹, e Huang et al.¹⁰ para as funções de verossimilhança para estes modelos. A função de verossimilhança para a estatística de varredura da variação espacial nas tendências temporais também é mais complexa, pois envolve a estimativa da máxima verossimilhança de várias funções de tendências diferentes.

A função de verossimilhança é maximizada sobre todas as localidades e tamanhos de janelas, e aquela com a máxima verossimilhança constitui o cluster mais provável. Este é o cluster que tem menos probabilidade de ter ocorrido ao acaso. A razão de verossimilhança para essa janela constitui a estatística do teste da razão da máxima verossimilhança. Sua distribuição sob a hipótese nula é obtida repetindo-se o mesmo exercício analítico em um grande número de replicações aleatórias do conjunto de dados geradas sob a hipótese nula. O valor de 'p' é obtido através do teste de hipótese de Monte Carlo¹⁴, comparando-se o 'rank' da máxima verossimilhança do conjunto de dados reais com as máximas verossimilhanças dos conjuntos de dados aleatórios. Se este 'rank' é 'R', então 'p' = 'R / (1 + #simulações)'. Para que 'p' seja um número 'inteligível', o número de simulações é restrito a 999 ou algum outro número finalizando com 999, tal como 1999, 9999 ou 99999. Dessa forma, fica sempre claro se rejeita ou não rejeita a hipótese nula para valores típicos de corte, tais como 0,05; 0,01 e 0,001.

O programa SaTScan realiza a varredura para áreas com taxas altas (clusters), para áreas com taxas baixas, ou simultaneamente para áreas com taxas altas ou baixas. Esse último deve ser utilizado, em vez de se executarem dois testes separados para taxas altas e baixas, respectivamente, a fim de se fazer a inferência estatística correta. A análise mais comum é fazer a varredura para áreas com taxas elevadas, ou seja, para clusters.

Função de Penalidade de Não-Compacidade (Compactação)

Quando a forma da janela elíptica é usada, há uma opção de se utilizar uma penalidade de não-compacidade (excentricidade) para favorecer aglomerados mais compactos¹². A principal razão para isto é que a estatística de varredura elíptica irá, sob a hipótese nula, tipicamente gerar um aglomerado elíptico mais provável, uma vez que existem mais aglomerados elípticos do que aglomerados circulares avaliados, e esta será frequentemente uma elipse longa e estreita, uma vez que existem mais destas.

Ao mesmo tempo, o conceito de ‘aglomeração’ é baseado em um critério de compacidade, no sentido de que os casos no cluster devem ser próximos uns dos outros; assim, há mais interesse em aglomerados compactos. Quando a penalidade de não-compacidade é usada, a pura razão de verossimilhança não é mais utilizada como a estatística do teste. Em vez disso, a estatística do teste é definida como o log da razão de verossimilhança multiplicado por uma penalidade de não-compacidade da fórmula $[4s/(s+1)^2]^a$, onde ‘s’ é a janela de formato elíptico definida como a razão do comprimento entre o eixo mais longo e o eixo mais curto da elipse. Para o círculo, ‘s’ = 1. O parâmetro ‘a’ é um parâmetro de ajuste da penalidade. Com ‘a’=0, a função de penalidade é sempre 1, independentemente de ‘s’, de modo que nunca há uma penalidade. Quando ‘a’ tende ao infinito, a função de penalidade vai para 0 para todos os ‘s’>1, de modo que apenas clusters circulares são considerados. Fora isso, não há nenhum significado intuitivo claro do parâmetro de ajuste de penalidade ‘a’. No SaTScan, é possível se utilizar tanto uma penalidade forte (‘a’ = 1) ou uma penalidade de tamanho médio (‘a’ = 1/2).

Tópicos Relacionados: *Modo de Lote, Modelo de Bernoulli, Ajustes por Covariáveis, Janela de Varredura Elíptica, Modelo Exponencial, Replicações de Monte Carlo, Modelo Ordinal, Modelo de Poisson, Clusters Secundários, Modelo de Permutação no Espaço-Tempo, Arquivo Padrão de Resultados.*

Clusters Secundários

Para as análises puramente espaciais e espaço-temporais, o SaTScan também identifica aglomerados secundários no conjunto de dados, além do cluster mais provável, e os ordena de acordo com seus testes estatísticos de razão de verossimilhança. Haverá quase sempre um cluster secundário que é quase idêntico ao cluster mais provável, e que tem um valor de verossimilhança quase tão alto, uma vez que expandir ou reduzir o tamanho do cluster apenas marginalmente, não irá alterar muito a verossimilhança. A maioria dos clusters deste tipo fornece poucas informações adicionais, mas sua existência significa que, embora seja possível se identificar a localização geral de um cluster, os seus limites exatos permanecem incertos. O usuário pode decidir em que medida clusters sobrepostos são relatados nos arquivos de resultados. O padrão é que clusters geograficamente sobrepostos não sejam relatados.

Também pode haver clusters secundários que não se sobrepõem espacialmente ao cluster mais provável, e eles podem ser de grande interesse. Estes são sempre relatados. Os valores de ‘p’ para tais aglomerados devem ser interpretados em termos da capacidade do cluster secundário de rejeitar a hipótese nula por sua própria força, sejam os clusters mais prováveis verdadeiros clusters ou não. Como consequência, esses valores de ‘p’ não são ajustados para o fato de que pode haver outros aglomerados nos dados. Se estes ajustes forem desejados, deve ser usada a estatística de varredura iterativa.

Para as análises puramente temporais, apenas o cluster mais provável é relatado.

Tópicos Relacionados: *Ajuste por Clusters Mais Prováveis, Teste da Razão de Verossimilhança, Guia de Saída Espacial, Critérios Para Relatar Clusters Secundários, Arquivo Padrão de Resultados.*

Ajuste por Clusters Mais Prováveis

Quando há numerosos clusters no conjunto de dados, os clusters secundários são avaliados como se não houvesse outros clusters no conjunto de dados. Ou seja, eles são estatisticamente significativos se, e apenas se, eles forem capazes de causar uma rejeição da hipótese nula por sua própria força, independentemente dos outros clusters serem verdadeiros clusters ou não. Este é muitas vezes o tipo de inferência desejado. Em algum momento, porém, é de interesse também se avaliar aglomerados secundários após o ajuste para outros aglomerados nos dados.

Como uma opção avançada, o SaTScan é capaz de ajustar a inferência dos aglomerados secundários para os clusters mais prováveis nos dados²⁴. Isto é feito de uma forma iterativa. Na primeira iteração, o SaTScan executa a análise padrão, mas relata apenas o cluster mais provável. Esse cluster é então removido do conjunto de dados, incluindo todos os casos e controles (modelo de Bernoulli) no cluster, enquanto que a população (modelo de Poisson) é definida como zero para as localidades e o período de tempo que definem o cluster. Em uma segunda iteração, uma análise completamente nova é conduzida, usando os dados restantes. Este procedimento é então repetido até que não haja mais clusters com um valor de 'p' inferior a um máximo especificado pelo usuário, ou até que um número máximo de iterações especificado pelo usuário tenha sido completado, o que ocorrer primeiro.

Para as análises puramente espaciais, demonstrou-se que os valores de 'p' resultantes para clusters secundários são bastante acurados e, no máximo, ligeiramente enviesados.

Notar que o círculo de um cluster secundário pode se sobrepor ao círculo de um cluster mais provável previamente detectado, e pode mesmo rodeá-lo completamente, de modo que este último (primário) é um subconjunto do primeiro (secundário). Isso não significa que o cluster mais provável seja detectado duas vezes. Em vez disso, o cluster mais provável é tratado como um 'lago' sem população e sem nenhum caso, e o novo cluster secundário consiste das áreas em torno daquele 'lago'. Isto pode acontecer, por exemplo, se uma cidade tiver um risco alto muito elevado, enquanto que os subúrbios ao redor têm um risco elevado moderado. O mesmo fenômeno pode ocorrer quando se faz análises puramente temporais ou espaço-temporais.

Esse recurso não está disponível para o modelo contínuo de Poisson.

Tópicos Relacionados: *Guia de Saída Espacial, Critérios para Relatar Clusters Secundários, Varredura Iterativa, Teste da Razão de Verossimilhança, Clusters Secundários, Arquivo Padrão de Resultados.*

Ajustes por Covariáveis

Deve-se ajustar por uma covariável quando todos os três dos seguintes forem verdadeiros:

- A covariável está relacionada com a doença em questão.
- A covariável não é distribuída geograficamente de forma aleatória.
- Deseja-se encontrar clusters que não podem ser explicados por esta covariável.

Aqui estão três exemplos:

- Ao se estudar a mortalidade por câncer nos Estados Unidos, deve-se ajustar por idade, dado que (i) os idosos são mais propensos a morrer por câncer, (ii) algumas áreas, tais como a Flórida, têm uma maior porcentagem de pessoas idosas, e (iii) a priori, há interesse em se encontrar áreas onde o risco de câncer é alto, em oposição a áreas com uma população mais velha.
- Se houver interesse na distribuição geográfica de malformações congênitas, não é necessário se ajustar por gênero. Embora malformações congênitas não sejam igualmente prováveis em meninos e meninas, a distribuição geográfica dos dois sexos é geograficamente aleatória no momento do nascimento.
- No estudo da geografia da incidência do câncer do pulmão, deve-se ajustar pelo fumo se houver interesse em se encontrar aglomerados devido a fatores de risco não relacionados ao fumo, mas não se deve ajustar por tabagismo se houver interesse em encontrar clusters que refletem áreas com necessidades especialmente urgentes para lançar uma campanha anti-tabagismo.

Quando a taxa da doença varia, por exemplo, com a idade, e a distribuição da idade varia em diferentes áreas, então existe uma aglomeração geográfica da doença simplesmente devido à covariável idade. Quando se ajusta por covariáveis categóricas, o programa SaTScan irá procurar por aglomerados acima e além do que é esperado devido a estas covariáveis. Quando mais de uma covariável é especificada, ajusta-se para cada uma delas, assim como para todos os termos de interação entre elas.

Tópicos Relacionados: *Ajuste por Covariáveis Usando os Arquivos de Entrada, Ajuste por Covariáveis Usando Software de Regressão Estatística, Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados, Artigos Metodológicos.*

Ajuste por Covariáveis Usando os Arquivos de Entrada

Com os modelos de Poisson e de permutação no espaço-tempo, é possível se ajustar por múltiplas covariáveis categóricas, especificando-as nos arquivos de entrada. Para fazê-lo, basta inserir as covariáveis como colunas extras no arquivo de casos (para ambos os modelos) e no arquivo de população (modelo de Poisson). Não há necessidade de se inserir qualquer informação em qualquer uma das guias das janelas.

Para o modelo de Poisson, o número de casos esperados em cada área sob a hipótese nula é calculado usando-se a padronização indireta. Sem covariáveis de ajuste, o número de casos esperados em uma localidade é (para análises espaciais):

$$E[c] = p * C / P$$

onde 'c' é o número observado de casos, e 'p' a população na localidade de interesse, enquanto 'C' e 'P' são o número total de casos e da população, respectivamente. Os termos 'ci', 'pi', 'Ci' e 'Pi' são definidos da mesma maneira, mas para a categoria da covariável 'i'. O número de casos esperados no ajuste por covariáveis pela padronização indireta é (análise espacial):

$$E[c] = \sum_i E[ci] = \sum_i p_i * C_i / P_i$$

O mesmo princípio é usado quando se calcula o número de casos ajustado por covariáveis para a estatística de varredura espaço-temporal, embora a fórmula seja mais complexa, devido à adição da dimensão temporal.

Dado que o modelo de permutação no espaço-tempo ajusta automaticamente para a variação puramente espacial e puramente temporal, não há necessidade de se ajustar por covariáveis a fim de explicar as diferentes densidades espaciais ou temporais dessas covariáveis. Por exemplo, não há nenhuma necessidade de se ajustar para a idade simplesmente porque alguns lugares têm uma proporção maior de pessoas idosas. Em vez disso, o ajuste por uma covariável é usado se houver uma interação no espaço-tempo devido a esta covariável, ao invés do processo de doença subjacente. Por exemplo, se as crianças ficam doentes principalmente no verão, e os adultos principalmente no inverno, então haverá clusters de interação espaço-temporais gerados por idade em áreas com muitas crianças no verão e vice-versa. Ao se incluir criança/adulto como uma covariável, estes aglomerados serão, por conseguinte, ajustados.

Nota: Muitas categorias de covariáveis podem criar problemas. Para o modelo de permutação no espaço-tempo, o ajuste é feito na fase de randomização, de modo que cada categoria de covariável é randomizada de forma independente. Se houver muitas categorias da covariável, de modo que a totalidade ou a maioria dos casos de uma categoria pertencem à mesma localidade espacial ou ao mesmo intervalo de tempo agregado, então há muito pouco para se randomizar, e o teste se torna insignificante.

Tópicos Relacionados: *Ajuste por Covariáveis, Ajuste por Covariáveis Usando Software de Regressão Estatística, Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados, Artigos Metodológicos, Modelo de Poisson, Modelo de Permutação no Espaço-Tempo, Arquivo de Casos, Arquivo de População.*

Ajuste por Covariáveis Usando Software de Regressão Estatística

O SaTScan não pode por si só fazer um ajuste para covariáveis contínuas. Tais ajustes podem ainda ser feitos para o modelo de Poisson^{17,23}, mas isso é um pouco mais complexo. O primeiro passo é calcular o número esperado de casos ajustado por covariáveis para cada ID de localidade e tempo usando o pacote de um software de regressão estatística padrão como o SAS. Estes números esperados devem, então, substituir os números brutos da população no arquivo de população, apesar de não incluírem as covariáveis eles mesmos.

O uso do software de regressão externo é também uma maneira excelente para se ajustar por covariáveis no modelo exponencial⁸. O primeiro passo é ajustar um modelo de regressão exponencial

sem qualquer informação espacial, a fim de se obter estimativas de risco para cada uma das covariáveis. O segundo passo é se ajustar os tempos de sobrevida e de censura para cima ou para baixo para cada indivíduo, com base nas estimativas de risco de suas covariáveis.

Para o modelo normal, pode se ajustar por covariáveis primeiramente fazendo uma regressão linear usando um software estatístico padrão, e em seguida substituindo o valor observado por seus resíduos.

Tópicos Relacionados: *Ajuste por Covariáveis, Ajuste por Covariáveis Usando os Arquivos de Entrada, Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados, Modelo Exponencial, Artigos Metodológicos, Modelo de Poisson, Arquivo de População.*

Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados

Também é possível se ajustar por covariáveis categóricas utilizando múltiplos conjuntos de dados¹¹. Os casos e controles/população são então divididos em categorias, e um conjunto de dados separado é usado para cada categoria. Este tipo de ajuste por covariáveis é computacionalmente muito mais lento do que aquele que utiliza os arquivos de entrada, e não é recomendado para grandes conjuntos de dados. Uma vantagem é que ele pode ser usado para ajustar por covariáveis quando se executam os modelos multinomial ou ordinal, para os quais não estão disponíveis outros procedimentos de ajustes. Uma desvantagem é que, dado que o número máximo de conjuntos de dados permitidos pelo SaTScan é doze, o número máximo de categorias de covariáveis também é doze.

A abordagem do ajuste para múltiplos conjuntos de dados é a seguinte (durante a pesquisa de clusters com taxas altas):

1. Para cada localidade e tamanho de janela, o log da razão de verossimilhança é calculado para cada conjunto de dados.
2. O log da razão de verossimilhança para todos os conjuntos de dados com menos do que o número de casos esperados na janela é multiplicado por um negativo.
3. Os logs das razões de verossimilhança são então somados, e esta soma é o log da verossimilhança combinado para essa janela em particular.
4. O máximo de todos os logs de razões de verossimilhança combinados, tomados sobre todas as localidades e tamanhos de janelas, constitui o cluster mais provável, e isso é avaliado da mesma forma que para um único conjunto de dados.

Ao se procurar por clusters com taxas baixas, o mesmo procedimento é realizado, exceto que ele é, então, o conjunto de dados com mais casos do que o esperado, que se multiplica por um. Ao se procurar por ambos os clusters, altos e baixos, ambas as somas são calculadas, e o máximo das duas é usado para representar o log da razão de verossimilhança para essa janela.

Tópicos Relacionados: *Guia de Múltiplos Conjuntos de Dados, Ajuste por Covariáveis, Ajuste por Covariáveis Usando os Arquivos de Entrada, Ajuste por Covariáveis Usando Software de Regressão Estatística, Artigos Metodológicos, Modelo de Bernoulli.*

Ajustes Espaciais e Temporais

Ajuste para Tendências Temporais

Se há uma tendência temporal crescente nos dados, então as estatísticas de varredura temporal e espaço-temporal vão capturar essa tendência através da atribuição de um cluster durante o final do período do estudo. Se houver uma tendência decrescente, será então capturado um cluster no início do período. Às vezes é de interesse testar se existem clusters temporais e/ou espaço-temporais após o ajuste para uma tendência temporal.

Para o modelo de permutação no espaço-tempo, a análise é ajustada automaticamente para ambos - tendência temporal e clusters temporais, e não são necessários novos ajustes. Para o modelo discreto

de Poisson, o usuário pode especificar se precisa ser feito um ajuste temporal, e, em caso afirmativo, ajusta-se com uma mudança percentual ou de forma não paramétrica.

Às vezes, a melhor maneira de se ajustar para uma tendência temporal é especificando-se o percentual anual de aumento ou decréscimo da taxa para a qual se irá ajustar. Este é um ajuste log linear. Dependendo da aplicação, pode-se ajustar, quer por uma tendência que o SaTScan estima a partir dos dados que estão sendo analisados, ou a partir da tendência estimada por dados nacionais ou outros semelhantes. Neste último caso, o aumento ou a diminuição percentual deve ser calculado utilizando softwares de regressão estatística padrão, tais como SAS ou R, e então inserido na Guia de Ajuste de Risco.

Para as análises espaço-temporais, é possível também se ajustar para uma tendência temporal não parametricamente. Esta forma ajusta a contagem esperada separadamente para cada intervalo de tempo agregado, removendo todos os clusters puramente temporais. A randomização é então estratificada por intervalo de tempo para garantir que cada intervalo tenha o mesmo número de eventos nos conjuntos de dados reais e aleatórios.

A capacidade de se ajustar para tendências temporais é muito mais limitada para os modelos de Bernoulli, multinomial, ordinal, normal e exponencial, já que nenhum dos recursos acima pode ser utilizado. Ao invés disso, o tempo precisa ser dividido em períodos de tempo discretos, com os casos e controles em cada período correspondendo a um conjunto de dados separado com arquivos de casos e controles separados. A análise é feita então se usando múltiplos conjuntos de dados.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Agregação de Tempo, Ajuste para Efeitos de Dias-da-Semana, Modelo de Poisson.*

Ajuste para Efeitos de Dias-da-Semana

Alguns conjuntos de dados possuem um padrão semanal. Se não forem ajustados para isso, podem gerar clusters, por exemplo, em uma segunda-feira, ou de uma segunda-feira para a próxima segunda-feira, simplesmente porque as segundas-feiras, em geral, têm mais eventos do que os outros dias da semana. Uma forma de se ajustar para esse efeito é agregar os dados diários em semanas, mas isso irá reduzir a resolução temporal. Outra opção é selecionar a função de ajuste por dia-da-semana na Guia de Ajuste Espacial e Temporal, que irá ajustar de forma não paramétrica para qualquer ajuste semanal nos dados. Este recurso está disponível apenas para o modelo de probabilidade discreto de Poisson.

O modelo de permutação no espaço-tempo ajusta automaticamente para qualquer variação puramente temporal dos dados, incluindo os efeitos de dia-da-semana. Assim, com este modelo de probabilidade, nunca há uma necessidade de se fazer qualquer ajuste especial por dia-de-semana. Se diferentes localidades espaciais têm diferentes efeitos de dia-da-semana, isto pode levar a clusters de interação espaço-temporais espúrios. Por exemplo, se os dados da doença vêm de diferentes clínicas médicas, mas apenas algumas das clínicas estão abertas nos fins de semana. Isso pode resultar em clusters de fim de semana naquelas clínicas, que são simplesmente um artefato das suas horas de abertura. Para ajustar para isso, é possível se ajustar para a interação de espaço-por-dia-da-semana, selecionando essa opção na Guia de Ajuste Espacial e Temporal. Fazer isso tem exatamente o mesmo efeito de se incluir uma variável de dia-da-semana no arquivo de entrada de casos. Este recurso só é válido para o modelo de permutação no espaço-tempo.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Ajuste para Tendências Temporais, Agregação de Tempo, Modelo de Poisson.*

Ajuste para Clusters Puramente Espaciais

Em uma análise espaço-temporal com o modelo de Poisson, também é possível se ajustar para clusters puramente espaciais, de uma forma não paramétrica. Esta ajusta a contagem esperada separadamente para cada localidade, removendo todos os clusters puramente espaciais. A randomização é então estratificada por ID de localidade, de modo a assegurar que cada localidade tenha o mesmo número de eventos nos conjuntos de dados reais e aleatórios.

Esta opção não está disponível para os modelos de Bernoulli, multinomial, ordinal, exponencial, normal ou de permutação no espaço-tempo; neste último caso, porque o método ajusta automaticamente para quaisquer clusters puramente espaciais.

Nota: Não é possível se ajustar simultaneamente para aglomerados espaciais e aglomerados puramente temporais usando a randomização estratificada, e se ambos os tipos de ajustes são desejados, o modelo de permutação no espaço-tempo deve ser usado como opção.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Modelo de Poisson, Ajuste para Tendências Temporais.*

Ajuste para Riscos Relativos Conhecidos

Às vezes sabe-se, a priori, que uma localidade e/ou tempo em particular tem um maior ou menor risco de magnitude conhecida, e queremos detectar clusters acima e além deste, ou, em outras palavras, queremos ajustar para este risco excessivo ou inferior conhecido. Uma maneira de fazer isso é simplesmente mudando os números da população sob risco no arquivo de população. Uma maneira mais simples é usar o arquivo de ajustes. Nesse arquivo, é especificado um risco relativo para qualquer combinação de localidade e período de tempo. As contagens esperadas são, então, multiplicadas por este risco relativo para esta localidade e tempo. Por exemplo, se é conhecido, a partir de dados históricos, que uma determinada localidade tem tipicamente 50% mais casos durante os meses de verão de Junho a Agosto (no caso dos EUA), então, para cada ano, especifica-se um risco relativo de 1,5 para esta localidade e esses meses. Um cluster de verão irá, então, aparecer neste local apenas se o excesso de risco for maior do que 50%.

Esta funcionalidade só está disponível para o modelo discreto de Poisson.

Tópicos Relacionados: *Arquivo de Ajustes, Guia de Ajustes Espaciais e Temporais, Agregação de Tempo, Modelo de Poisson, Dados Faltantes.*

Dados Faltantes

Se houver dados faltantes para algumas localidades e tempos, é importante se ajustar para isso na análise. Se não, pode-se encontrar clusters de taxas baixas estatisticamente significativos onde houver dados faltantes, ou clusters de taxas altas estatisticamente significativos em outras localidades, mesmo que estes sejam simplesmente artefatos dos dados em falta.

Modelo de Bernoulli

Para se ajustar uma análise do modelo de Bernoulli para dados faltantes, fazer o seguinte: se os casos estiverem faltando para uma dada localidade e período de tempo, remover os controles para essa mesma localidade e tempo. Da mesma forma, se os controles estiverem faltando para uma localidade e um tempo específicos, remover os casos para essa mesma localidade e tempo. Isso precisa ser feito antes de se fornecer os dados para o SaTScan. Se ambos casos e controles estiverem faltando para uma localidade e um tempo, então ok, e não há necessidade de qualquer modificação dos dados de entrada.

Modelos Multinomial e Ordinal

Para se ajustar uma análise dos modelos multinomial ou ordinal para dados faltantes, fazer o seguinte: se uma ou mais categorias estiverem faltando para uma localidade e período de tempo específicos, remover todos os casos nas categorias restantes dessa mesma localidade e tempo. Isso precisa ser feito antes de se fornecer os dados para o SaTScan. Se todos os casos em todas as categorias estiverem faltando para uma localidade e tempo, então está tudo bem, e não há necessidade de qualquer modificação dos dados de entrada.

Modelo Discreto de Poisson

Para ajustar o modelo discreto de Poisson para dados faltantes, utilizar os arquivos de ajustes para definir as combinações de localidade e tempo para as quais os dados estão faltando, e atribuir um risco relativo de zero para essas combinações de localidade/tempo.

Modelo Contínuo de Poisson

Para ajustar o modelo contínuo de Poisson para dados faltantes, redefinir a área de estudo usando um conjunto diferente de polígonos, de modo que as áreas com dados faltantes sejam excluídas da área do estudo.

Modelo de Permutação no Espaço-Tempo

É um pouco mais complexo se ajustar para dados faltantes no modelo de permutação no espaço-tempo, mas ainda possível⁵. Primeiro adicionar o dia-da-semana como uma covariável no arquivo de análise. Quando uma determinada localidade/período de tempo estiver faltando, então, para essa localidade, remover todos os dados para os dias da semana para os quais estiverem faltando quaisquer dados. Por exemplo, se os dados de quinta-feira 23/10 e sexta-feira 24/10, estiverem faltando para a região do código postal A, e os dados de sábado 25/10 estiverem faltando para a área B, remover os dados de todas as quintas e sextas-feiras para a área A e os dados de todos os sábados para a área B, mantendo todos os dados de sábado até quartas-feiras para a área A e todos os dados, exceto os dos sábados, para a área B. Para todas as outras áreas de código postal, manter todos os dados para todos os dias. Notar que, além de ajustar para os dados ausentes, esta abordagem irá também ajustar para qualquer dia-da-semana por efeitos de interação espacial.

A mesma abordagem pode ser utilizada com outra categorização dos dados, desde que as categorizações estejam em alguma unidade de tempo periódica que ocorre várias vezes e é distribuída uniformemente ao longo do período do estudo. Por exemplo, pode-se categorizar em meses se o período de estudo abrange vários anos, mas não se somente houver dados válidos de um ano.

Duas abordagens mais brutas para lidar com dados em falta no modelo de permutação no espaço-tempo é remover todos os dados de uma localidade particular se alguns dados estiverem faltando para essa localidade, ou remover todos os dados para um determinado período de tempo para as datas em que houver dados ausentes em qualquer localidade. Este último é especialmente útil na vigilância prospectiva para dados faltantes no início do período do estudo, para evitar a remoção de dados recentes que são os mais importantes para a detecção precoce de surtos de doenças.

Nota: Quando há combinações de localidade/tempo com dados faltantes, remover toda a linha do arquivo de casos ou atribuir zero casos a essa combinação de localidade/período. Se for removido apenas o número de casos, mas for mantida a informação do ID da localidade e do tempo, ocorrerá um erro ao ler o arquivo.

Aviso: O ajuste para dados faltantes só funciona se as localidades e tempos para os quais os dados estiverem faltando forem independentes do número de casos nestas localidades e tempos. Por exemplo, se os dados estiverem faltando para todas as localidades com menos do que cinco casos observados, os procedimentos de ajuste descritos acima não funcionarão corretamente.

Tópicos Relacionados: *Arquivo de Ajustes, Ajuste para Riscos Relativos Conhecidos, Modelo de Bernoulli, Modelo Ordinal, Modelo de Poisson, Modelo de Permutação no Espaço-Tempo, Guia de Ajustes Espaciais e Temporais, Agregação de Tempo.*

Varredura Multivariada com Múltiplos Conjuntos de Dados

Às vezes, é de interesse se pesquisar e se avaliar simultaneamente aglomerados em mais de um conjunto de dados. Por exemplo, pode-se estar interessado em aglomerados espaciais com excesso de incidência apenas de leucemia, apenas de linfoma ou de ambos, simultaneamente. Como outro exemplo, pode-se estar interessado na detecção de um surto de doença gastrointestinal que afeta somente crianças, somente adultos ou ambos, simultaneamente. Se o SaTScan é usado para analisar um único conjunto de dados combinado, pode-se perder um cluster que está presente apenas em um dos subgrupos. Por outro lado, se duas análises do SaTScan são realizadas, uma para cada conjunto de dados, há uma perda de poder se o verdadeiro aglomerado tiver quase a mesma força em ambos os conjuntos de dados. Uma análise do SaTScan com múltiplos conjuntos de dados e a opção de varredura multivariada resolve este problema.

A estatística de varredura multivariada com múltiplos conjuntos de dados funciona da seguinte forma¹² (na procura por clusters com taxas altas):

1. Para cada localidade e tamanho de janela, o log da razão de verossimilhança é calculado para cada conjunto de dados.
2. Os log das razões de verossimilhança para os conjuntos de dados com mais do que o número esperado de casos são somados, e esta soma é a verossimilhança para aquela janela em particular.
3. O máximo de todos os logs das razões de verossimilhança somados, tomados ao longo de todas as localidades e tamanhos de janela, constitui o cluster mais provável, e isso é avaliado da mesma forma como que para um único conjunto de dados.

Ao se procurar por clusters com taxas baixas, o mesmo procedimento é realizado, exceto que, nesse caso, somam-se os logs das razões de verossimilhança dos conjuntos de dados com menos do que o número esperado de casos dentro da janela em questão. Ao se procurar por ambos os clusters - altos e baixos, ambas as somas são calculadas, e a máxima dentre as duas é usada para representar o log da razão de verossimilhança para essa janela.

Nota: Todos os conjuntos de dados devem usar o mesmo modelo de probabilidade e o mesmo arquivo de coordenadas geográficas.

Tópicos Relacionados: *Guia de Múltiplos Conjuntos de Dados, Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados, Arquivo de Coordenadas.*

Comparação com Outros Métodos

Estatísticas de Varredura

As estatísticas de varredura foram primeiramente estudadas em detalhe por Joseph Naus¹⁸⁴. Um grande desafio das estatísticas de varredura é se encontrar resultados analíticos a respeito das probabilidades de se observar um cluster de uma magnitude específica, e há uma bela coleção de teorias matemáticas que têm sido desenvolvidas para se obter aproximações e limites para essas probabilidades, sob uma variedade de configurações. Excelentes revisões deste trabalho foram fornecidas por Glaz e Balakrishnan¹⁷⁴, Glaz, Naus e Wallenstein¹⁷⁵ e Glaz, Pozdnyakov e Wallenstein²⁴⁶. Duas características comuns desse trabalho inicial sobre as estatísticas de varredura foram: (i) eles usam uma janela de varredura de tamanho fixo, e (ii) eles lidam com dados de contagem onde, sob a hipótese nula, o número de casos observados segue uma distribuição uniforme, seja no cenário contínuo ou discreto, de modo que o número esperado de casos em uma área é proporcional ao tamanho dessa área.

Na vigilância de doenças, nenhum desses pressupostos é cumprido, uma vez que não se sabe o tamanho de um cluster a priori, e dado que a população sob risco é geograficamente não homogênea. Sob a hipótese nula de igualdade de risco de doença, espera-se ver mais casos da doença em uma cidade em comparação com uma área de tamanho semelhante no campo, apenas por causa da maior densidade populacional na cidade. As estatísticas de varredura no software SaTScan foram desenvolvidas para resolver esses dois problemas. Visto que não foram encontradas soluções analíticas para se obter as probabilidades nesses cenários mais complexos, o teste de hipótese de Monte Carlo é utilizado alternativamente para a obtenção dos valores de 'p'¹⁴.

Aglomerados Espaciais e Espaço-Temporais

Métodos Descritivos de Detecção de Cluster

Em 1987, Openshaw et al.¹⁸⁵ desenvolveram uma Máquina de Análise Geográfica (MAG) que utiliza círculos sobrepostos de tamanhos diferentes da mesma forma que a estatística de varredura espacial, exceto pelo fato de que o tamanho do círculo não varia continuamente. Com o MAG, um teste de significância é feito separadamente para cada círculo, levando a múltiplas testagens, e em quase todos os conjuntos de dados haverá uma multiplicidade de 'clusters significativos' quando definido dessa forma. Isso ocorre porque, sob a hipótese nula, cada círculo tem uma probabilidade de 0,05 de ser 'significativo' ao nível de 0,05, e com 20.000 círculos, esperaria-se 1.000 aglomerados significativos sob a hipótese nula de não haver clusters. O MAG é, portanto, muito útil para fins descritivos, mas não deve ser utilizado para testes de hipóteses.

Outro método bom para a detecção de clusters descritivos foi proposto por Rushton e Lolonis¹⁸⁷, que utilizaram mapas de isolinhas de valores de 'p' para representar os aglomerados, em vez de círculos sobrepostos. Tal como acontece com o MAG, este não ajusta para a testagem múltipla inerente às muitas localidades de potenciais clusters avaliadas.

Testes de Detecção de Cluster

A estatística de varredura espacial é um teste de detecção de cluster. Um teste de detecção de cluster é capaz tanto de detectar a localização de clusters como de avaliar a sua significância estatística, sem problemas com os testes múltiplos. Em 1990, Turnbull et al.¹⁹¹ propuseram o primeiro teste desse tipo usando círculos sobrepostos com tamanho fixo de população, considerando o círculo com a maioria dos casos como o cluster detectado.

A estatística de varredura espacial foi em parte inspirada pelo trabalho de Openshaw et al.¹⁸⁵ e Turnbull et al.¹⁹¹. Mediante a aplicação de um teste de razão de verossimilhança, foi possível se avaliar clusters de diferentes tamanhos (como Openshaw et al. fizeram), enquanto que, ao mesmo tempo, ajustando para os testes múltiplos (como Turnbull et al. fizeram).

Em uma comparação de poder², demonstrou-se que o método de Turnbull tem maior poder se o tamanho do aglomerado verdadeiro estiver dentro de cerca de 20% do que é especificado por esse método, enquanto que, de outro modo, a estatística de varredura espacial tem maior poder. Nota-se que o tamanho do cluster no método de Turnbull deve ser especificado antes de se analisar os dados, ou o procedimento é inválido.

Testes Focais de Clusters

Testes focais devem ser usados quando há um conhecimento a priori sobre a localização de um cluster hipotético. Por exemplo, um cluster em torno de um depósito de resíduos tóxicos em um país pode estimular uma investigação sobre clusters em torno de um depósito de resíduos tóxicos similar em outro país. A estatística de varredura espacial ou outros testes de detecção de clusters não devem ser, então, utilizados, uma vez que terão baixo poder devido à avaliação de todas as localidades possíveis, ainda que a localização hipotética já seja conhecida. Exemplos de testes focais são o Teste de Stone¹⁸⁸, Teste de Score de Lawson-Waller^{181,192} e Teste de Bithell¹⁷⁰.

Testes focais nunca devem ser usados quando os focos forem definidos usando os dados em si. Isso levaria a um viés de pré-seleção, e os valores de 'p' resultantes seriam incorretos. Então, é melhor usar a estatística de varredura espacial. Se, por outro lado, a fonte pontual foi definida sem olhar para os dados, então é melhor se usar o teste focal em vez da estatística de varredura espacial, visto que o primeiro terá maior poder, uma vez que incide sobre a localidade de interesse.

Além de várias estatísticas de varredura, o software SaTScan também pode ser usado para fazer um teste focal a fim de avaliar se há um aglomerado de doença em torno de um foco pré-determinado (ref. 2, p809). Isto é feito por meio de um arquivo de grade com apenas um único ponto de grade refletindo as coordenadas do foco de interesse. De forma semelhante, um teste multifocal pode ser especificado usando o arquivo de grade com uma coordenada para cada foco desejado.

Testes Globais de Clusters

A maioria dos testes propostos para a aglomeração espacial são testes de agregação global. Estes incluem, entre muitos outros, os métodos propostos por Alt e Vach¹⁶⁷, Besag e Newell¹⁶⁹, Cuzick e Edwards¹⁷¹, Diggle e Chetwynd¹⁷², Grimson¹⁷⁶, Moran¹⁸³, Ranta¹⁸⁶, Tango^{189,190}, Walter¹⁹³ e Whittemore et al.¹⁹⁴. Estes métodos testam para aglomerações em toda a região do estudo, sem a capacidade de identificar a localização de clusters específicos. Como tal, estes testes e a estatística de varredura espacial se complementam, uma vez que são úteis para diferentes propósitos.

Testes Globais de Interação Espaço-Temporal

Knox¹⁷⁸, Mantel¹⁸², Diggle et al.¹⁷³, Jacquez¹⁷⁷, Baker¹⁶⁸, e Kulldorff e Hjalmarsson¹⁸⁰ propuseram diferentes testes para a interação espaço-temporal. Tal como a versão da permutação no espaço-tempo⁵ da estatística de varredura espaço-temporal, esses métodos são projetados para avaliar se os casos que estão próximos no espaço também estão próximos no tempo e vice-versa, ajustando para qualquer aglomeração puramente espacial ou puramente temporal. Sendo de natureza global, esses outros testes são úteis quando testam para verificar se há aglomerados em toda a região do estudo e durante todo o período de tempo, e são o método preferido quando, por exemplo, tenta-se determinar se uma doença é infecciosa. No entanto, diferentemente da estatística de varredura baseada na permutação no espaço-tempo, eles são incapazes de detectar a localização e o tamanho dos clusters e de testar a significância desses clusters.

Tópicos Relacionados: *Teste da Razão de Verossimilhança, Artigos Metodológicos do SaTScan.*

Dados de Entrada

Requisitos de Dados

Arquivos Necessários: Os dados de entrada devem ser fornecidos em um certo número de arquivos. Um arquivo de coordenadas é sempre necessário e um arquivo de casos é necessário para todos os modelos de probabilidade, exceto para o modelo contínuo de Poisson. O modelo de Poisson também requer um arquivo de população, enquanto que o modelo de Bernoulli requer um arquivo de controles.

Arquivos Opcionais: Pode-se também especificar um arquivo de grade especial opcional que contém as coordenadas geográficas dos centroides, definindo os círculos usados pela estatística de varredura. Se tal arquivo não for especificado, as coordenadas no arquivo de coordenadas serão usadas para este propósito. Como parte das configurações avançadas, há também um arquivo opcional de tamanho máximo do círculo, um arquivo opcional de ajustes, um arquivo opcional de vizinhos não-Euclidianos e um arquivo opcional de meta-localidade.

Formato dos Arquivos: Os arquivos de entrada de dados devem estar no formato de arquivo ASCII do SaTScan, ou pode-se usar o Assistente de Importação do SaTScan para arquivos shapefile, dBase, delimitados por vírgula ou por espaço. Usando tais arquivos, o assistente irá gerar automaticamente arquivos no formato do SaTScan. Ambas as opções são descritas abaixo.

Resolução Espacial: Para as estatísticas de varredura discretas, dados de localidades separadas podem ser especificados para os indivíduos ou os dados podem ser agregados por estados, províncias, municípios, paróquias, setores censitários, áreas de código postal, distritos escolares, lares, etc.

Informações Temporais: Para se executar uma análise temporal, espaço-temporal ou de variação espacial nas tendências temporais, é necessário ter um tempo relacionado a cada caso, e se o modelo de Bernoulli for usado, também para cada controle. Este tempo pode ser especificado como um dia, mês ou ano. Quando o modelo discreto de Poisson é usado, assume-se que a população de base do denominador existe continuamente ao longo do tempo, embora não necessariamente a um nível constante. O arquivo de população exige uma data a ser especificada para cada contagem de população. Para os períodos entre essas datas, o SaTScan irá estimar a população através de interpolação linear. Se todas as contagens da população tiverem a mesma data, a população é considerada constante ao longo do tempo.

Múltiplos Conjuntos de Dados: É possível se especificar vários arquivos de casos, cada um representando um conjunto diferente de dados, com informações sobre diferentes doenças ou sobre homens versus mulheres, respectivamente. Para o modelo de Bernoulli, cada arquivo de casos precisa estar acompanhado de seu próprio arquivo de controles, e para o modelo de Poisson, cada arquivo de casos deve ser acompanhado de seu próprio arquivo de população. O número máximo de conjuntos de dados que o SaTScan pode analisar é doze.

Ajustes de Covariáveis: Com os modelos de Poisson e de permutação no espaço-tempo, é possível se ajustar para múltiplas covariáveis categóricas, incluindo-as nos arquivos de casos e de população. Para os modelos de Bernoulli, ordinal ou exponencial, as covariáveis podem ser ajustadas pela utilização de múltiplos conjuntos de dados.

Tópicos Relacionados: *Guia de Entrada, Guia de Múltiplos Conjuntos de Dados, Arquivo de Casos, Arquivo de Controles, Arquivo de População, Arquivo de Coordenadas, Arquivo de Grade, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan, Ajuste por Covariáveis.*

Arquivo de Casos

O arquivo de casos fornece informações sobre os casos, e é usado para todos os modelos de probabilidades. Ele deve conter as seguintes informações:

ID da Localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

Número de Casos: O número de casos para a localidade, tempo e covariáveis especificados. Para os modelos discreto de Poisson, binomial e de permutação no espaço-tempo, este é o número de observações ou indivíduos com a característica de interesse, tal como câncer ou baixo peso ao nascer. Para os modelos ordinal, multinomial, normal e exponencial, é o número total de observações ou indivíduos nas localidades, seja qual for o valor de sua característica categórica ou valor do atributo contínuo.

Data/Tempo: Opcional. Pode ser especificado em anos, meses ou dias, ou em um formato genérico. O formato deve coincidir com o formato de precisão de tempo especificado na Guia de Entrada. A menos que a opção de verificação de dados temporais esteja desabilitada, todos os tempos dos casos devem cair dentro do período de estudo, conforme especificado na Guia de Entrada.

Atributo: Para os modelos multinomial, ordinal, exponencial e normal, apenas. Uma variável que descreve algumas características do caso. Esta pode ser uma categoria (modelos multinomial ou ordinal), tempo de sobrevivência (modelo exponencial), ou uma variável de valor contínuo (modelo normal). As categorias para os modelos multinomial e ordinal podem ser especificadas como qualquer valor numérico positivo ou negativo. Os tempos de sobrevivência têm que ser números positivos. Os números para o modelo normal podem ser positivos ou negativos.

Censura: Apenas para o modelo exponencial. Censura é uma variável 0/1, com censurado=1 e não censurado=0.

Peso: Opcional. Apenas para o modelo normal. É necessário se covariáveis forem usadas, mesmo se todas as observações tiverem a mesma variância, caso em que todos os pesos devem ser definidos como '1'.

Covariáveis: Opcional. Para os modelos discreto de Poisson, de permutação no espaço-tempo e normal, apenas. Qualquer número de covariáveis categóricas pode ser especificado, tanto como números ou através de caracteres. Para o modelo normal, as covariáveis somente podem ser incluídas se os pesos forem também fornecidos.

Exemplo: Se em 01 de abril de 2004 havia 17 casos do sexo masculino e 12 casos do sexo feminino em Nova York, as seguintes informações seriam fornecidas:

NewYork 12 2004/4/1 Feminino

NewYork 17 2004/4/1 Masculino

Nota: Para o modelo normal ponderado, pode haver apenas um caso (observação) por linha e, conseqüentemente, se os pesos são especificados, a segunda coluna deve ter todos eles.

Nota: Múltiplas linhas podem ser usadas para diferentes casos com a mesma localidade, tempo e atributos. O SaTScan irá adicioná-los automaticamente.

Nota: Este arquivo não é usado para o modelo contínuo de Poisson.

Tópicos Relacionados: *Guia de Entrada, Nome do Arquivo de Casos, Guia de Múltiplos Conjuntos de Dados, Ajuste por Covariáveis Usando Arquivos de Entrada, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Controles

O arquivo de controle é utilizado apenas no modelo de Bernoulli. Ele deve conter as seguintes informações:

ID da Localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

Nº (#) de Controles: O número de controles para a localidade e tempo especificados.

Tempo: Opcional. O tempo pode ser especificado em anos, meses ou dias, ou em um formato genérico. Todos os tempos dos controles devem estar dentro do período de estudo, conforme especificado na Guia de Análise. O formato dos tempos deve ser o mesmo que no arquivo de casos.

Nota: múltiplas linhas podem ser usadas para diferentes controles com a mesma localidade, tempo e atributos. O SaTScan irá adicioná-los automaticamente.

Tópicos Relacionados: *Guia de Entrada, Nome do Arquivo de Controles, Guia de Múltiplos Conjuntos de Dados, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Arquivo de População

O arquivo de população é utilizado para o modelo discreto de Poisson, fornecendo informações sobre a população de base sob risco. Essa pode ser a contagem real da população a partir de um censo, ou poderia ser, por exemplo, contagens esperadas ajustadas por covariáveis a partir de um modelo de regressão estatística. Ele deve conter as seguintes informações:

ID da localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

Tempo: O tempo ao qual o tamanho da população se refere. Pode ser especificado em anos, meses ou dias, ou em um formato genérico. Se o tempo da população é desconhecido, mas idêntico para todos os números da população, então um ano 'dummy' (fictício) deve ser dado, e a escolha não afeta o resultado.

População: O tamanho da população para a combinação de uma dada localidade, ano e covariável. Se o tamanho da população é zero para um determinado local, ano e conjunto de covariáveis, então ele deve ser incluído no arquivo de população especificado como zero. A população pode ser especificada como um número decimal para refletir o tamanho de uma população sob risco, em vez de um número real de pessoas.

Covariáveis: Opcional. Qualquer número de covariáveis categóricas pode ser especificado, cada uma representada por uma coluna diferente separada por espaços vazios. Podem ser especificadas numericamente ou através de caracteres. As covariáveis devem ser as mesmas que no arquivo de casos.

Exemplo: Se a idade e o sexo são as covariáveis incluídas, com 18 faixas etárias diferentes, então deve haver $18 \times 2 = 36$ linhas para cada ano e cada área do censo. Com 3 diferentes anos censitários, e 32 áreas do censo, o arquivo terá um total de 3.456 linhas e 5 colunas.

Nota: Várias linhas podem ser usadas para diferentes grupos populacionais com a mesma localidade, tempo e atributos de variáveis independentes. O SaTScan irá adicioná-los automaticamente.

Nota: Para uma análise puramente temporal com o modelo discreto de Poisson, não é necessário especificar um arquivo de população se a população for constante ao longo do tempo.

Tópicos Relacionados: *Guia de Entrada, Nome do Arquivo de População, Guia de Múltiplos Conjuntos de Dados, Ajuste por Covariáveis Usando Arquivos de Entrada, Arquivo de Tamanho Máximo do Círculo, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Coordenadas

O arquivo de coordenadas fornece as coordenadas geográficas para cada ID de localidade. Cada linha do arquivo representa uma localidade geográfica. Informações baseadas em áreas podem ser agregadas e representadas por um único ponto de localidade geográfica. Coordenadas podem ser especificadas ou usando o sistema padrão de coordenadas Cartesianas ou em latitude e longitude. Se dois IDs de localidades diferentes tiverem exatamente as mesmas coordenadas, então os dados para os dois são combinados e tratados como uma única localidade.

Um arquivo de coordenadas não é necessário para análises puramente temporais.

Tópicos Relacionados: *Guia de Entrada, Nome do Arquivo de Coordenadas, Coordenadas, Coordenadas Cartesianas, Latitude e Longitude, Arquivo de Grade.*

Coordenadas Cartesianas

Cartesiano é o nome matemático para o sistema x,y de coordenadas regulares planas, ensinado no colégio. Estas podem ser especificadas em duas, três ou quaisquer números de dimensões. O programa SaTScan irá ler automaticamente o número de dimensões, que deve ser o mesmo para todas as coordenadas. Se forem utilizadas coordenadas Cartesianas, o arquivo de coordenadas deve conter as seguintes informações:

ID da localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

Coordenadas: As coordenadas devem ser todas especificadas na mesma unidade. Não há um limite máximo para o número de dimensões.

Coordenadas x e y: Obrigatórias

Coordenadas z1-zN: Opcionais

Nota: Se houver mais do que 10 dimensões, não se pode usar o Assistente de Importação do SaTScan para os arquivos de coordenadas e arquivos de grade, mas deve-se especificá-los usando o formato de arquivo ASCII do SaTScan.

Nota: O modelo contínuo de Poisson só trabalha em duas dimensões.

Tópicos Relacionados: *Guia de Entrada, Coordenadas, Latitude e Longitude, Arquivo de Coordenadas, Arquivo de Grade, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Latitude e Longitude

Latitudes e longitudes devem ser inseridas como números de graus decimais. Latitude representa a distância norte/sul a partir do equador, e localidades ao sul do equador devem ser inseridas como números negativos. Longitude representa a distância a leste/oeste do Meridiano Principal (Greenwich, Inglaterra), e localidades a oeste do Meridiano de Greenwich devem ser inseridas como números negativos. Por exemplo, o Instituto Nacional de Saúde, em Bethesda, Maryland, que está localizado a 39,00 graus norte e 77,10 graus oeste, devem ser referidos como 39,00 e -77,10, respectivamente.

Latitudes e longitudes podem, para os fins deste programa, não ser especificados em graus, minutos e segundos. Tais latitudes e longitudes podem facilmente ser convertidas em números de graus decimais (NGD) pela fórmula simples: $NGD = \text{graus} + \text{minutos}/60 + \text{segundos}/3600$.

Se forem utilizadas coordenadas de latitude/longitude, o arquivo de coordenadas deve conter as seguintes informações:

ID da localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

Latitude: Latitude em número de graus decimais.

Longitude: Longitude em número de graus decimais.

Nota: Quando as coordenadas são especificadas em latitudes e longitudes, o SaTScan não realiza uma projeção dessas coordenadas sobre um espaço planar. Em vez disso, o SaTScan desenha círculos perfeitos na superfície da terra esférica.

Nota: A latitude e a longitude não podem ser utilizadas para o modelo contínuo de Poisson, ou quando uma janela espacial elíptica é usada.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Coordenadas, Coordenadas, Coordenadas Cartesianas, Latitude e Longitude, Arquivo de Grade, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan, Tempo de Processamento.*

Arquivo de Grade

O arquivo de grade opcional define os centroides dos círculos utilizados pela estatística de varredura. Se não for especificado nenhum arquivo de grade, as coordenadas dadas no arquivo de coordenadas serão usadas para esta finalidade. Cada linha no arquivo representa um círculo centroide. Deve haver pelo menos duas variáveis representando as coordenadas x,y Cartesianas (padrão) ou exatamente duas variáveis representando a latitude e a longitude. A escolha entre Cartesianas e latitude/longitude deve coincidir com o arquivo de coordenadas, assim como o número de dimensões.

O arquivo de grade normalmente incluirá apenas coordenadas espaciais, enquanto que o intervalo temporal de potenciais clusters é especificado na Guia da Janela Temporal. Isso não permite que o usuário especifique um intervalo temporal diferente para localidades diferentes. Se isso for desejado, mais quatro colunas podem ser adicionadas ao arquivo de grade, representando o início mais precoce do cluster permitido, o último período de início autorizado, o período final mais precoce e o período final mais tardio permitidos, nesta ordem. Se estas colunas forem especificadas para alguns pontos da grade, mas não para o restante, então os pontos restantes da grade terão as especificações do cluster temporal definidas na Guia da Janela Temporal.

Se apenas um centroide for especificado no arquivo de grade, obtém-se um teste focal de aglomerados ao invés de uma estatística de varredura. Tais testes focais são úteis para avaliar se existe um aglomerado em torno de uma localidade pré-definida, tal como um depósito de lixo tóxico. Se mais do que um, mas ainda um pequeno número de centroides for especificado no arquivo de grade, obtém-se um teste multifocal, buscando identificar clusters em torno de um ou mais dos centroides.

Tópicos Relacionados: *Guia de Entrada, Nome do Arquivo de Grade, Coordenadas, Coordenadas Cartesianas, Latitude e Longitude, Arquivo de Coordenadas, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan, Guia da Janela Temporal, Tempo de Processamento.*

Arquivo de Vizinhos Não-Euclidianos

Este é um arquivo opcional para as estatísticas de varredura discretas. Ele não pode ser definido usando o Assistente de Importação do SaTScan, mas tem que ser especificado usando o formato de arquivo ASCII. Com esta opção, os arquivos de coordenadas e de grade não são necessários, e são ignorados, se fornecidos.

Com os ajustes de parâmetros convencionais, o SaTScan utiliza o arquivo de coordenadas para determinar quais localidades estão mais próximas do centro de cada círculo construído. Isso é feito usando as distâncias Euclidianas. Em essência, para cada centroide, o SaTScan encontra o vizinho mais próximo, o segundo mais próximo, e assim por diante, até que atinja o tamanho máximo da janela. Com o arquivo de vizinhos, é possível para o usuário especificar essas relações de vizinhança de qualquer maneira, sem se limitar às distâncias Euclidianas. Por exemplo, os vizinhos podem ser classificados (ordenados) de acordo com a distância ao longo de uma rede de metro ou um sistema de distribuição de água.

A primeira coluna deste arquivo contém os IDs das localidades que definem os centroides da janela de varredura. As entradas subsequentes de cada linha são, então, os centroides vizinhos em ordem de proximidade. A janela de varredura irá se expandir em tamanho até que não haja mais vizinhos fornecidos por esta linha. Por exemplo, na linha “1 2 3 4”, o SaTScan avaliará [1], [1,2], [1,2,3] e [1,2,3,4] como potenciais aglomerados. Se a segunda linha for “4 1 3 2”, o SaTScan também irá avaliar [4], [1,4], e [1,3,4]. Note que com estas duas linhas o cluster [1,2,3,4] é especificado duas vezes, o que está ok, mas redundante. É permitido ter várias linhas para o mesmo ID de localidade do centroide, cada uma com um conjunto diferente de vizinhos mais próximos, de modo que a terceira linha poderia ser, por exemplo, “1 3 5 7 9”.

O arquivo de vizinhos não-Euclidianos define automaticamente o tamanho máximo do cluster para cada centroide, visto que o SaTScan irá parar de adicionar localidades ao cluster quando o final da linha for atingido. Por isso, geralmente não há necessidade de se especificar um Tamanho Máximo do Cluster Espacial na Guia da Janela Espacial. Se um máximo for especificado nessa guia, então o SaTScan irá honrar esse máximo, e não continuará adicionando localidades até chegar ao fim da linha no arquivo de vizinhos não-Euclidianos.

Nota: O arquivo de vizinhos não pode ser utilizado no modelo contínuo de Poisson.

Tópicos Relacionados: *Arquivo de Coordenadas, Guia de Entrada, Tamanho Máximo do Cluster Espacial, Arquivo de Meta-Localidade, Guia de Vizinhos Espaciais, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Meta-Localidade

Este é um arquivo opcional para as estatísticas de varredura discretas, que só pode ser utilizado se o arquivo de vizinhos não-Euclidianos for usado também. Ele não pode ser definido por meio do Assistente de Importação do SaTScan, mas tem que ser especificado usando o formato de arquivo ASCII.

Uma meta-localidade é um conjunto de dois ou mais IDs de localidades individuais. Quando uma meta-localidade é especificada no arquivo de vizinhos não-Euclidianos, todos os membros individuais da meta-localidade são simultaneamente inseridos na janela de varredura.

A primeira coluna deste arquivo contém os nomes das meta-localidades definidos pelo usuário. As entradas subsequentes em cada linha são os IDs das localidades individuais que são parte dessa meta-localidade. Não há um limite máximo para o número de localidades individuais que podem pertencer a cada meta-localidade.

Nota: O arquivo de meta-localidade só pode ser usado em conjunto com o arquivo de vizinhos não-Euclidianos.

Tópicos Relacionados: *Arquivo de Coordenadas, Guia de Entrada, Guia de Vizinhos Espaciais, Arquivo de Vizinhos Não-Euclidianos, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Tamanho Máximo do Círculo

Este arquivo opcional é utilizado para determinar o tamanho máximo do círculo da janela de varredura, quando o máximo é definido como uma porcentagem da ‘população’. Normalmente, a porcentagem é baseada na população no arquivo de população, mas usando o arquivo de tamanho máximo do círculo, uma ‘população’ diferente pode ser especificada para este fim. Uma razão importante para a utilização do arquivo de tamanho máximo do círculo é para análises espaço-temporais prospectivas, nas quais o arquivo da população normal pode mudar ao longo do tempo, mas se deseja avaliar o mesmo conjunto de círculos geográficos em cada período. Isso é fundamental para se ajustar corretamente a estatística de varredura espaço-temporal prospectiva para as análises prévias. Também pode ser usado para outras finalidades.

O arquivo deve conter uma linha para cada localidade, com as seguintes informações:

ID da localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id.

‘População’: Qualquer número não-negativo.

O nome do arquivo especial de tamanho máximo do círculo é especificado na Guia de Análise → Recursos Avançados → Guia da Janela Espacial.

Nota: Se um ID de localidade estiver faltando neste arquivo, a população é assumida como sendo zero. Se um ID de localidade ocorrer mais de uma vez, os números da população serão adicionados.

Tópicos Relacionados: *Guia de Entrada, Arquivo de População, Guia da Janela Espacial, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Ajustes

O arquivo de ajustes pode ser usado para ajustar uma análise do modelo discreto de Poisson para quaisquer anomalias temporais, espaciais e espaço-temporais nos dados, com um risco relativo conhecido. Ele pode, por exemplo, ser utilizado para ajustar para dados faltantes ou parcialmente faltantes. (Nota: Ajusta-se para covariáveis usando os arquivos de casos e de população, ou analisando-se múltiplos conjuntos de dados, e não com esse arquivo). O arquivo de ajustes deve conter uma ou mais linhas para cada localidade para a qual os ajustes são necessários, com as seguintes informações:

ID da localidade: Qualquer valor numérico ou sequência de caracteres. Espaços vazios não podem fazer parte do id. Alternativamente, é possível se especificar 'Todos', para que todas as localidades sejam ajustadas com o mesmo risco relativo.

Risco Relativo: Qualquer número não negativo. O risco relativo representa o quão mais comum é a doença nesta localidade e período de tempo, em comparação à linha de base. Definir o valor de 'um' é equivalente a não fazer quaisquer ajustes. Um valor maior do que 'um' é usado para ajustar para um risco aumentado, e um valor menor do que 'um' ajusta para um risco baixo. Um risco relativo de zero é usado para ajustar para dados faltantes para essa localidade e tempo particulares.

Tempo de Início: Opcional. O início do período de tempo a ser ajustado usando esse risco relativo.

Tempo Final: Opcional. O final do período de tempo a ser ajustado usando esse risco relativo.

Se não forem definidos os tempos de início e fim, todo o período de estudo será ajustado para aquela localidade. Se 'Todos' for selecionado em vez de um ID de localidade, mas nenhum tempo inicial ou final for determinado, isso tem o mesmo efeito de quando nenhum ajuste é feito.

O nome do arquivo de ajustes é especificado na Guia de Análise → Recursos Avançados → Ajustes de Risco.

Nota: Atribuir um risco relativo de x para a metade das localidades é equivalente a atribuir um risco relativo de $1/x$ para a outra metade. Atribuir o mesmo risco relativo para todas as localidades e períodos de tempo tem o mesmo efeito de não ajustar.

Nota: É permitido ajustar a mesma localidade e período de tempo diversas vezes, através de diferentes linhas com diferentes riscos relativos. O SaTScan simplesmente multiplicará os riscos relativos. Por exemplo, ao se ajustar a localidade A com um risco relativo de 2 para todos os períodos de tempo, e 1990 for ajustado com um risco relativo de 2 para todas as localidades, então a entrada de 1990 para a localidade A será ajustada com um risco relativo de $2*2=4$.

Tópicos Relacionados: *Ajuste com Risco Relativo Conhecido, Dados Faltantes, Guia de Ajustes Espacial e Temporal, Assistente de Importação do SaTScan, Formato de Arquivo ASCII do SaTScan.*

Arquivo de Hipótese Alternativa

Ao se estimar o poder estatístico, o arquivo de hipótese alternativa é usado para definir a hipótese alternativa para a qual o poder é estimado. Ele tem o mesmo formato do arquivo de ajustes.

ID da localidade: Qualquer valor numérico ou sequência de caracteres que está presente no arquivo de coordenadas geográficas. Espaços vazios não podem fazer parte do id. Alternativamente, é possível se especificar 'Todos', no caso em que todas as localidades terão o mesmo risco relativo sob a hipótese alternativa.

Risco Relativo: Qualquer número não negativo. O risco relativo representa o quão mais comum é a doença nesta localidade e período de tempo, em comparação à linha de base. Definir o valor de 'um' é o mesmo que não incluir esta linha. Um valor maior do que 'um' é utilizado para a hipótese alternativa de risco aumentado, e um valor menor do que 'um' para a hipótese alternativa de menor risco. Não há problema em se definir uma hipótese alternativa com diferentes riscos relativos para diferentes IDs de localidades, como também ter uma hipótese alternativa com um risco relativo maior do que 1 para algumas localidades e menor do que 1 para outras.

Período Inicial: Opcional. O início do período de tempo que tem um risco relativo diferente na hipótese alternativa.

Período Final: Opcional. O final do período de tempo que tem um risco relativo diferente na hipótese alternativa.

É possível se avaliar várias hipóteses alternativas dentro da mesma execução de análise. Isso é feito deixando-se uma linha em branco entre as diferentes hipóteses alternativas. Isso é computacionalmente mais eficiente do que executar uma análise separada para cada uma das alternativas.

Tópicos Relacionados: *Arquivo de Ajustes, Guia de Estimativa de Poder.*

Assistente de Importação do SaTScan

O Assistente de Importação do SaTScan pode ser usado para importar arquivos shapefile, dBase, delimitados por vírgula ou delimitados por espaço. Ele funciona para todos os arquivos de importação, exceto o Arquivo de Vizinhos opcional. Iniciar o Assistente de Importação clicando no botão 'Importar Arquivo'  à direita do campo de texto e selecionar o arquivo que se deseja importar.

Passo 1 - Seleção do Arquivo de Origem

1. Na parte inferior da caixa de diálogo de 'Seleção do Arquivo de Origem', selecionar o tipo de extensão de arquivo que se está procurando. Se houver incerteza, selecionar a opção 'Todos os Arquivos'. Os formatos de arquivos compatíveis são: dBase III/IV, CSV e Excel 97-2003.
2. Percorrer as pastas e selecionar o arquivo que se deseja abrir. Ele será exibido no campo de texto de 'Nome do Arquivo'.
3. Clicar em 'Abrir'. O Assistente de Importação do SaTScan aparecerá agora.

Passo 2 - Seleção do Arquivo de Origem

- a. Se o arquivo já estiver no formato do SaTScan, clicar em **Ok**.
- b. Se não estiver, clicar em **Próximo**.
- c. Usar os botões **Próximo** e **Anterior** para navegar entre as caixas de diálogos que aparecem.

Passo 3: Especificando a Estrutura do Arquivo

Ao se importar um arquivo shapefile, um arquivo dBase ou um arquivo de Excel, este passo é automaticamente ignorado. Para arquivos de origem CSV e TXT, é necessário se especificar a estrutura do arquivo em uso. Para ajudar nisso, o Assistente de Importação irá mostrar o conteúdo das primeiras poucas linhas do arquivo de origem.

1. Os requisitos de dados são diferentes para os diferentes modelos de probabilidade. Ao importar um arquivo de casos, na parte superior da caixa de diálogo, selecionar o modelo de probabilidade que se irá utilizar.
2. Se tiver um arquivo delimitado por caracteres, usar os menus de rolagem para selecionar se o separador de campos é uma vírgula, um ponto e vírgula ou um espaço em branco.
3. Se for o caso, selecionar se o indicador de grupo é definido por meio de aspas simples ou duplas.
4. Clicar em **Próximo** para avançar para a próxima caixa de diálogo.

Passo 4: Combinando Variáveis do Arquivo de Origem com Variáveis do SaTScan

A grade superior nesta caixa de diálogo vincula as variáveis do SaTScan com as variáveis do arquivo de entrada a partir do arquivo fonte. A grade inferior exhibe amostras de dados do arquivo de entrada escolhido.

1. Os requisitos de dados são diferentes para os diferentes modelos de probabilidades. Ao se importar um arquivo de casos, na parte superior da caixa de diálogo, selecionar o modelo de probabilidade que será utilizado. Ao se importar um arquivo de coordenadas ou de grade, especificar se as coordenadas são Cartesianas ou do tipo latitude/longitude.
2. Para combinar as variáveis, clicar em um dos lugares onde se diz 'não atribuído'.
3. Selecionar o formato apropriado da variável do arquivo de entrada para combinar com a variável escolhida do SaTScan.
4. Quando todas as variáveis obrigatórias e opcionais selecionadas tiverem sido combinadas, clicar em **Próximo**. As duas primeiras colunas na grade inferior não estão, de fato, no arquivo de origem. A primeira pode ser utilizada para gerar um conjunto genérico de IDs de Localidades se o arquivo de origem não tiver um nome para as localidades. É especialmente útil se todos os arquivos do SaTScan forem importados a partir do mesmo arquivo de origem. Se eles forem importados de diferentes fontes de arquivos, deve-se tomar extremo cuidado para garantir que as localidades estejam exatamente na mesma ordem em cada um dos arquivos de origem. A segunda coluna pode ser utilizada quando o arquivo de origem tiver uma linha para cada observação, para gerar uma contagem de 'um'.

Passo 5: Salvar o Arquivo Importado ou Ler Diretamente a Partir do Arquivo de Origem

O arquivo importado, que está em formato de arquivo ASCII do SaTScan, pode ser salvo para uso atual e futuro. Dessa forma, não é preciso passar pelo Assistente de Importação da próxima vez que se quiser utilizar o mesmo arquivo. Alternativamente, é possível salvar as configurações usadas, e ler diretamente a partir do arquivo de origem, ao invés de se criar um arquivo intermediário no formato do SaTScan. Se isso for feito, não há problema se o conteúdo do arquivo de origem mudar para uma análise posterior, mas a estrutura deve permanecer inalterada.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Casos, Arquivo de Controles, Arquivo de População, Arquivo de Coordenadas, Arquivo de Grade, Arquivo de Tamanho Máximo do Círculo, Arquivo de Ajustes.*

Formato de Arquivo ASCII do SaTScan

Como uma alternativa ao uso do Assistente de Importação do SaTScan, também é possível se escrever diretamente o nome dos arquivos de entrada nos campos de texto fornecidos na Guia de Entrada, ou navegar nos diretórios de arquivos para alcançar os arquivos de entrada desejados por meio do botão à direita dessa caixa. Os arquivos devem, então, estar no formato de arquivo do SaTScan, que são arquivos ASCII delimitados por espaços, com uma linha para cada combinação de localidade/covariável, e com colunas conforme definido abaixo. Tais arquivos podem ser criados usando qualquer editor de texto e a maioria das planilhas. A ordem das colunas no arquivo é muito importante, mas as linhas podem estar em qualquer ordem. As variáveis opcionais, definidas acima, são colunas opcionais no formato de arquivo do SaTScan.

Formato do Arquivo de Casos (*.cas):

<ID da localidade> <#casos> <tempo> <atributo> <censurado> <peso> <covariável#1> ... <covariável#N>

O uso de atributos, censura, peso e covariáveis depende do modelo de probabilidade, conforme mostrado na Tabela 1.

Probability Model	attribute	censored	weight	covariates
Discrete Poisson	n/a	n/a	n/a	optional
Bernoulli	n/a	n/a	n/a	n/a
Space-Time Permutations	n/a	n/a	n/a	optional
Multinomial	category	n/a	n/a	n/a
Ordinal	category	n/a	n/a	n/a
Exponential	survival time	optional	n/a	n/a
Normal	continuous variable	n/a	optional	optional

Tabela 1: Atributos do arquivo de casos usados pelos diferentes modelos de probabilidades.

Formato do Arquivo de Controles (*.ctl):

<ID da localidade> <#controles> <tempo>

Formato do Arquivo de População (*.pop):

<ID da localidade> <tempo> <população> <covariável#1> ... <covariável#N>

Formato do Arquivo de Coordenadas (*.geo):

<ID da localidade> <latitude> <longitude> OU

<ID da localidade> <coordenada-x> <coordenada-y> <coordenada-z1> . . . <coordenada-zN>

Formato do Arquivo de Grade (*.grd):

<latitude> <longitude> OU

<coordenada-x> <coordenada-y> <coordenada-z1>. . . <coordenada-zN> OU

<latitude> <longitude> <tempo inicial mais precoce> <tempo inicial mais tardio> <tempo final mais precoce> <tempo final mais tardio> OU

<coordenada-x> <coordenada-y> <coordenada-z1>. . . <coordenada-zN> <tempo inicial mais precoce> <tempo inicial mais tardio> <tempo final mais precoce> <tempo final mais tardio>

Formato do Arquivo de Vizinhos Não-Euclidianos (*.nbr):

<ID da localidade> <ID da localidade do vizinho mais próximo> <ID da localidade do 2º vizinho mais próximo> etc.

Formato do Arquivo de Meta-Localidade (*.met):

<ID da meta-localidade> <ID da localidade #1> <ID da localidade #2> etc.

Formato do Arquivo Especial de Tamanho Máximo do Círculo (*.max):

<ID da localidade> <'população'>

Formato do Arquivo de Ajustes (*.adj):

<ID da localidade> <risco relativo> <tempo de início> <tempo final>

Formato do Arquivo de Hipóteses Alternativas (*.ha):

<ID da localidade> <risco relativo> <tempo de início> <tempo final>

Formatos de Tempo

Os tempos devem ser inseridos em um formato específico. Um tempo genérico é especificado usando-se qualquer número inteiro negativo ou positivo no intervalo (-200,000 até 2,900,000¹). Se houver tempos fora deste intervalo, simplesmente adicionar ou subtrair a mesma constante para todos os tempos. Os formatos de data válidos são:

2010

2010/06, 2010/06/26

2010-06, 2010-06-26

06/2010, 06/26/2010

06-2010, 06-26-2010

Dias e meses com um único dígito podem ser especificados com um ou dois dígitos. Por exemplo, 9 de Setembro de 2002 pode ser escrito como 2002/9/9, 2002/09/09, 2002/09/9, 2002/9/09, 2002-9-9, etc.

Nota: O SaTScan também suporta alguns outros formatos de tempo utilizados em versões anteriores, mas eles não são mais recomendados.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Casos, Arquivo de Controles, Arquivo de População, Arquivo de Coordenadas, Arquivo de Grade, Arquivo de Tamanho Máximo do Círculo, Arquivo de Vizinhos, Arquivo de Ajustes, Assistente de Importação do SaTScan.*

¹Nota do Tradutor: foi mantida a vírgula (,) como o separador de milhar, que é o formato a ser utilizado no SaTScan.

Recursos Básicos do SaTScan

A maioria das análises do SaTScan pode ser realizada utilizando-se as análises básicas e as características dos dados. Os usuários as especificam em três diferentes guias de janelas, para as opções de entrada, análise e saída, respectivamente. Estas contêm todas as especificações exigidas para uma análise do SaTScan, bem como alguns poucos opcionais. Recursos adicionais, todos opcionais, podem ser especificados nas guias de recursos avançados.

Tópicos Relacionados: *Metodologia Estatística, Guia de Entrada, Guia de Análise, Guia de Saída, Recursos Avançados.*

Guia de Entrada

The screenshot shows the 'Input' dialog box for SaTScan. It features three tabs: 'Input', 'Analysis', and 'Output'. The 'Input' tab is selected. The dialog is organized into several sections:

- Case File:** A text input field with a browse button (three dots).
- Control File:** A text input field containing '(Bernoulli Model)' with a browse button.
- Study Period:** Two date pickers. 'Start Date' is set to Year: 2000, Month: 1, Day: 1. 'End Date' is set to Year: 2000, Month: 12, Day: 31.
- Population File:** A text input field containing '(Poisson Model)' with a browse button.
- Coordinates File:** A text input field with a browse button.
- Grid File:** A text input field containing '(optional)' with a browse button.
- Time Precision:** Radio buttons for 'None', 'Year' (selected), 'Month', and 'Day'. A 'Generic' option is also present.
- Coordinates:** Radio buttons for 'Cartesian' and 'Lat/Long' (selected).

An 'Advanced >>' button is located at the bottom right of the dialog.

Caixa de Diálogo da Guia de Entrada

A Guia de Entrada é usada para especificar os nomes dos arquivos de dados de entrada, bem como a natureza dos dados nesses arquivos. Se os arquivos estiverem no formato de arquivo ASCII do SaTScan, eles podem ser especificados ou escrevendo-se o nome na caixa de texto ou usando o botão de navegação (...). Se eles não estiverem no formato de arquivo ASCII do SaTScan, eles devem ser especificados usando o Assistente de Importação do SaTScan, clicando no botão 'Importar Arquivo' (...). Tanto o formato de arquivo ASCII do SaTScan como o Assistente de Importação do SaTScan são descritos na 'Seção de Dados de Entrada'.

Tópicos Relacionados: *Recursos Básicos do SaTScan, Dados de Entrada, Guia de Múltiplos Conjuntos de Dados.*

Nome do Arquivo de Casos

Especificar o nome do arquivo de entrada com dados dos casos. Este arquivo é exigido para todas as estatísticas de varredura discretas, independentemente do modelo de probabilidade usado.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Casos.*

Nome do Arquivo de Controles

Especificar o nome do arquivo de entrada com dados dos controles. Este arquivo é usado somente para as análises com o modelo de probabilidade de Bernoulli.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Controles.*

Precisão do Tempo

Indicar se o arquivo de casos e o arquivo de controles (quando aplicável) contêm informações sobre o tempo de cada caso (e controle) e, em caso afirmativo, se a precisão deve ser lida como genérica, dias, meses ou anos. Se a precisão de tempo for especificada em dias, mas a precisão no arquivo de casos ou controles estiver em mês ou ano, então haverá um erro. Se a precisão do tempo for especificada como anos, mas os arquivos de casos ou controles incluírem algumas datas especificadas em termos de meses ou dias, então o mês ou o dia será ignorado.

Para uma análise puramente espacial, o arquivo de casos e de controles não precisa conter quaisquer tempos. Se eles tiverem, precisa ser especificado que eles contêm esta informação para que o SaTScan saiba como ler o arquivo, mas a informação é ignorada.

Nota: A escolha define apenas a precisão para os tempos nos arquivos de casos e controles. A precisão dos tempos no arquivo de população pode ser diferente, exceto que, se um tiver tempos genéricos, o outro também precisa ter tempos genéricos.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Casos, Arquivo de Controles, Período de Estudo, Agregação do Tempo.*

Período de Estudo

Especificar a data de início e do final do período de tempo sob estudo. Isso precisa ser feito mesmo para uma análise puramente espacial, a fim de se calcular o número esperado de casos corretamente. Anos admissíveis são aqueles entre 1753 e 9999.

Todos os tempos nos arquivos de casos e de controles devem cair sobre ou entre a data inicial e a data final do período de estudo. As datas no arquivo de população podem estar fora das datas de início e fim do período de estudo.

Data/Tempo Inicial: A data/tempo mais antigo(a) a ser incluído(a) no período de estudo.

Data/Tempo Final: A data/tempo mais recente a ser incluído(a) no período de estudo.

Nota: As datas de início e término não podem ser especificadas com uma precisão superior à precisão dos tempos nos arquivos de casos e controles.

Se o usuário não especificar o mês, então, por padrão, este será definido como 'Janeiro' para a data de início e 'Dezembro' para a data final. Da mesma forma, se o dia não for especificado, então, por padrão, ele será definido como o primeiro dia do mês para a data de início e o último dia do mês para a data final.

Tópicos Relacionados: *Guia de Entrada, Arquivo de Casos, Arquivo de Controles, Precisão do Tempo, Agregação do Tempo.*

Nome do Arquivo de População

Especificar o nome do arquivo de entrada com os dados populacionais. Este arquivo é usado somente para análises utilizando o modelo de probabilidade discreto de Poisson.

Tópicos Relacionados: *Guia de Entrada, Arquivo de População.*

Nome do Arquivo de Coordenadas

Especificar o nome do arquivo de entrada com as coordenadas geográficas de todas as localidades com dados sobre o número de casos, controles e/ou população. Quando múltiplos conjuntos de dados são usados, o arquivo de coordenadas precisa incluir as coordenadas para todas as localidades encontradas em qualquer um dos conjuntos de dados.

Tópicos Relacionados: *Guia de Entrada, Coordenadas, Arquivo de Coordenadas.*

Nome do Arquivo de Grade

Especificar o nome do arquivo de grade opcional com as coordenadas dos círculos de centroides utilizadas pelas estatísticas de varredura espacial e espaço-temporal. Se nenhum arquivo de grade especial for especificado, então as coordenadas do arquivo de coordenadas serão usadas para esta finalidade.

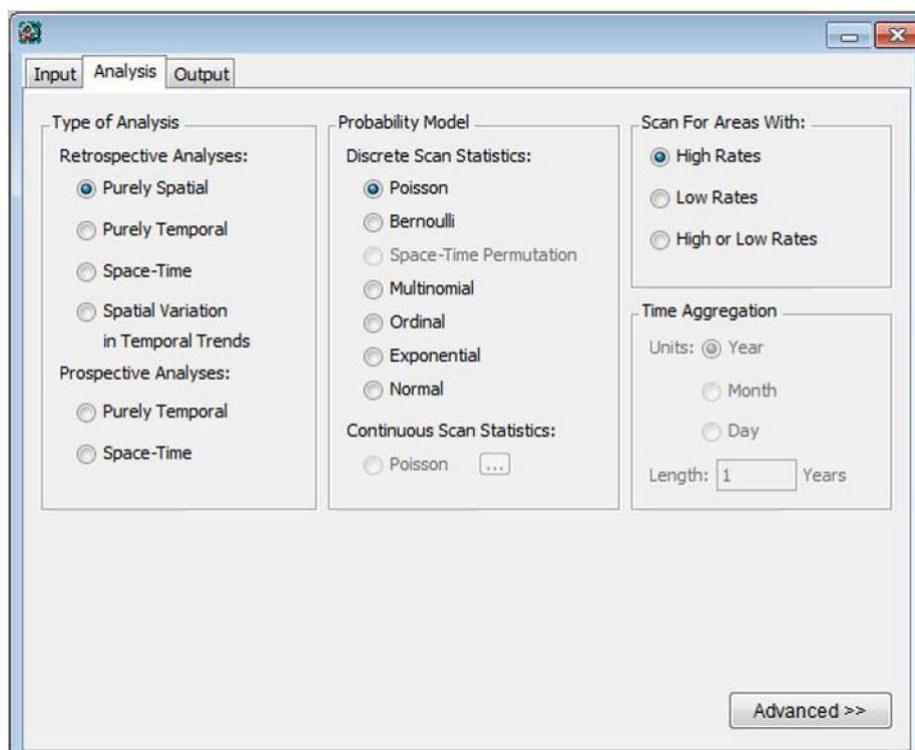
Tópicos Relacionados: *Guia de Entrada, Coordenadas, Arquivo de Coordenadas, Arquivo de Grade.*

Coordenadas

Especificar o tipo de coordenadas usadas pelo arquivo de coordenadas e o arquivo de grade, como Cartesianas ou latitude/longitude. Cartesiana é o nome matemático para o sistema regular de coordenadas-x/y ensinado no colégio. Latitude/longitude não podem ser utilizadas para o modelo contínuo de Poisson.

Tópicos Relacionados: *Coordenadas Cartesianas, Latitude/Longitude, Arquivo de Coordenadas, Arquivo de Grade.*

Guia de Análise



Caixa de Diálogo da Guia de Análise

A Guia de Análise é utilizada para definir várias opções de análise. Recursos adicionais estão disponíveis clicando-se no botão 'Avançado' no canto inferior direito.

Tópicos Relacionados: *Recursos Básicos do SaTScan, Metodologia Estatística, Guia da Janela Espacial, Guia da Janela Temporal, Guia de Ajustes Espaciais e Temporais, Guia de Inferência.*

Tipos de Análises

O SaTScan pode ser usado para uma análise puramente espacial, puramente temporal, espaço-temporal e de variação espacial nas tendências temporais. Uma análise puramente espacial ignora os tempos dos casos, mesmo quando esses dados são fornecidos. Uma análise puramente temporal ignora a localização geográfica dos casos, mesmo quando essa informação é fornecida.

Dados puramente temporais e espaço-temporais podem ser analisados nas formas retrospectiva ou prospectiva. Numa análise retrospectiva, a análise é feita apenas uma vez para uma região geográfica fixa e um período de estudo fixo. O SaTScan 'realiza a varredura' sobre múltiplas datas de início e fim, avaliando tanto os 'clusters vivos', com duração até a data final do período do estudo, quanto os 'clusters históricos', que deixaram de existir antes da data final do período de estudo. A opção prospectiva é utilizada para a detecção precoce de surtos de doenças, quando as análises são repetidas todos os dias, semanas, meses ou anos. Somente aglomerados vivos, clusters que atingem todo o caminho para o tempo atual, conforme definido pela data final do período de estudo, são, então, pesquisados.

Tópicos Relacionados: *Estatísticas de Varredura Espacial Temporal e Espaço-Temporal, Guia de Análise, Artigos Metodológicos, Tempo de Processamento, Guia da Janela Espacial, Guia da Janela Temporal, Agregação de Tempo.*

Modelos de Probabilidade

Há oito modelos de probabilidade diferentes que podem ser utilizados: discreto de Poisson, Bernoulli, permutação no espaço-tempo, multinomial, ordinal, exponencial, normal e contínuo de Poisson. Para as análises puramente espaciais, os modelos de Poisson e Bernoulli são boas aproximações um do outro em muitas situações. Dados temporais são processados de forma diferente, de modo que os modelos diferem mais para as análises temporal e espaço-temporal.

Modelo Discreto de Poisson: O modelo discreto de Poisson deve ser usado quando a população de base reflete um certo risco em massa, como o total de pessoas-ano residentes em uma área. Os casos são, então, incluídos como parte da contagem da população.

Modelo de Bernoulli: O modelo de Bernoulli deve ser utilizado quando o conjunto de dados contém indivíduos que podem ter ou não uma doença e para outras variáveis do tipo 0/1. Aqueles que têm a doença são casos e devem ser listados no arquivo de casos. Aqueles sem a doença são 'controles', listados no arquivo de controles. Os controles podem ser um conjunto aleatório de controles provenientes da população, ou melhor, a população total exceto pelos casos. O modelo de Bernoulli é um caso especial do modelo ordinal quando existem apenas duas categorias.

Modelo de Permutação no Espaço-Tempo: O modelo de permutação no espaço-tempo deve ser usado quando apenas dados dos casos estiverem disponíveis, e quando se quer ajustar para clusters puramente espaciais e puramente temporais.

Modelo Multinomial: O modelo multinomial é usado quando os indivíduos pertencem a uma de três ou mais categorias, e quando não há uma relação ordinal entre elas. Quando houver apenas duas categorias, o modelo de Bernoulli deve ser então utilizado.

Modelo Ordinal: O modelo ordinal é usado quando os indivíduos pertencem a uma de três ou mais categorias, e quando há uma relação ordinal entre essas categorias, tal como pequeno, médio e grande. Quando existem apenas duas categorias, o modelo de Bernoulli deve ser utilizado ao invés deste.

Modelo Exponencial: O modelo exponencial é usado para dados de tempos de sobrevida, para procurar por aglomerados espaciais e/ou temporais de sobrevida excepcionalmente curta ou longa. O tempo de sobrevida é uma variável contínua positiva. Tempos de sobrevida censurados são permitidos para alguns, mas não para todos os indivíduos.

Modelo Normal: O modelo normal é usado para dados contínuos. As observações podem ser tanto positivas como negativas.

Modelo Contínuo de Poisson: O modelo contínuo de Poisson deve ser usado quando a hipótese nula é de que as observações são distribuídas aleatoriamente com intensidade constante de acordo com um processo de Poisson homogêneo sobre uma área de estudo definida pelo usuário.

Tópicos Relacionados: *Guia de Análise, Modelo de Bernoulli, Modelo Exponencial, Artigos Metodológicos, Modelo Ordinal, Modelo de Poisson, Comparação de Modelos de Probabilidade, Modelo de Permutação no Espaço-Tempo.*

Polígonos para o Modelo Contínuo de Poisson

Para o modelo contínuo de Poisson, é necessário se definir a área de estudo espacial na qual as observações pontuais podem estar localizadas. Isto é feito utilizando-se um ou mais polígonos convexos, onde cada polígono é definido por uma série de desigualdades lineares. Por exemplo, o quadrado unitário é definido por $y \geq 0$, $y \leq 1$, $x \geq 0$ e $x \leq 1$. A área de estudo é a unidade, ou a soma, de todas as áreas definidas pelos diferentes polígonos. Não há um limite máximo para o número de polígonos que podem ser utilizados, nem para o número de desigualdades usadas para cada polígono. Isto significa que quase todas as áreas de estudo podem ser aproximadas para qualquer precisão desejada. O menor número de desigualdades que pode ser usado para definir um polígono é três, caso em que o polígono é um triângulo.

Um novo polígono é definido primeiramente clicando-se no botão 'adicionar' à esquerda para adicionar um polígono, e então, clicando no botão 'adicionar' à direita para adicionar uma desigualdade linear. A primeira desigualdade é então especificada usando o editor de equação na parte inferior, seguida por um clique no botão de 'atualização' (update). Depois disso, outra desigualdade é adicionada, e assim por diante, até que todos os polígonos tenham sido definidos. Se for preciso alterar uma desigualdade, utilizar o mouse para destacar a desigualdade que se pretende alterar, fazer a mudança desejada no editor de equações, e em seguida clicar no botão de 'atualização'.

Nota: Os polígonos devem ser não sobrepostos. Eles não precisam ser contíguos.

Tópicos Relacionados: *Guia de Análise, Modelo Contínuo de Poisson.*

Varredura para Taxas Altas ou Baixas

É possível se realizar uma varredura para áreas com taxas altas apenas (clusters), para áreas com taxas baixas apenas, ou simultaneamente, para as áreas com taxas altas ou baixas. A análise mais comum é realizar a varredura para áreas com taxas altas somente, ou seja, para clusters. Para o modelo exponencial, alta corresponde a curta sobrevida. Para os modelos ordinal e normal, alta corresponde a categorias/observações de grandes valores. Por padrão, o modelo multinomial irá avaliar simultaneamente as taxas altas e baixas para todas as categorias. No modelo contínuo de Poisson, só é possível fazer a varredura para taxas altas.

A estatística de varredura da variação espacial nas tendências temporais não está à procura de clusters com taxas altas ou baixas. Em vez disso, essa modalidade está à procura de 'clusters' com uma tendência que é maior ou menor do que a tendência fora do cluster. Tal como acontece nas outras estatísticas de varredura, pode-se procurar por clusters com altas tendências somente, baixas tendências somente ou simultaneamente para ambos os tipos. Um cluster pode ter uma tendência elevada, quer porque ele tem uma taxa que está aumentando mais do que fora do cluster, ou porque ele tem uma taxa que está decrescendo menos do que fora do cluster. Da mesma forma, um aglomerado pode ter uma baixa tendência ou porque ele tem uma taxa que está aumentando menos do que fora do cluster, ou porque tem uma taxa que está diminuindo mais do que fora do cluster.

Tópicos Relacionados: *Guia de Análise, Teste da Razão de Verossimilhança, Artigos Metodológicos.*

Agregação de Tempo

As análises espaço-temporais são, por vezes, muito intensivas para o computador. Para reduzir o tempo de processamento, os tempos dos casos podem ser agregados em intervalos de tempo. Outra razão para se fazer isso é ajustar para as tendências temporais cíclicas. Por exemplo, ao usar intervalos de um ano, as análises serão automaticamente ajustadas para a variabilidade sazonal nas contagens, e ao usar intervalos de tempo de 7 dias, serão automaticamente ajustadas para os efeitos de dias da semana.

Unidades: As unidades nas quais o comprimento dos intervalos de tempo são especificadas. Esta pode ser em anos, meses, dias ou genérica. As unidades dos intervalos de tempo não podem ser mais precisas do que a precisão de tempo especificada na guia de entrada. Se um tempo genérico é usado no arquivo de casos, a unidade de agregação do tempo também precisa estar em um tempo genérico, e vice-versa.

Comprimento: O comprimento (duração) dos intervalos de tempo nas unidades especificadas.

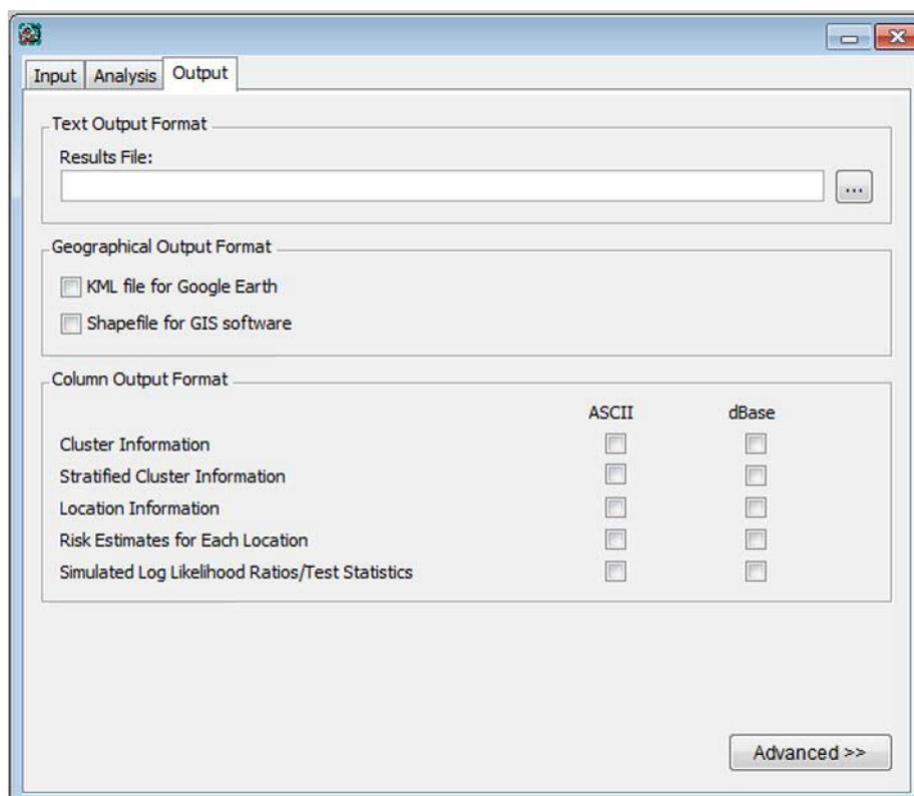
Exemplo: Se as unidades do intervalo são anos e o comprimento é dois, então os intervalos de tempo serão de dois anos.

Nota: Se o comprimento do intervalo de tempo não é uma fração do comprimento de todo o período de estudo, o intervalo de tempo mais precoce será o restante após os outros intervalos terem recebido o seu comprimento apropriado. Assim, o primeiro intervalo de tempo pode ser mais curto do que o comprimento especificado. Para uma análise de variação espacial nas tendências temporais, todos os intervalos de tempo devem ter o mesmo comprimento, e se o primeiro intervalo de tempo for mais curto, um alerta é gerado e esse intervalo é ignorado na análise.

Importante: Para as análises espaço-temporais prospectivas, o intervalo de tempo precisa ser igual ao comprimento entre as análises periódicas realizadas. Assim, se as análises periódicas no tempo são realizadas todas as semanas, então o intervalo de tempo deve ser definido como 7 dias.

Tópicos Relacionados: *Guia de Análise, Precisão de Tempo, Período do Estudo, Velocidade Computacional.*

Guia de Saída



Caixa de Diálogo da Guia de Saída

A Guia de Saída é utilizada para configurar os parâmetros que definem as informações de saída fornecidas pelo SaTScan.

Tópicos Relacionados: *Resultados das Análises, Arquivo Padrão de Resultados, Formato de Saída de Texto, Formato de Saída de Coluna, Guia de Saída Espacial.*

Formato de Saída de Texto

Um arquivo padrão de resultados baseado em texto é mostrado automaticamente após a conclusão dos cálculos. Ele contém informações sobre os clusters detectados, informações resumidas a respeito dos dados, o tempo de processamento e os parâmetros de análise escolhidos. Especificar o nome deste arquivo. Outros arquivos de saída opcionais podem também ser criados, mas precisam ser abertos manualmente pelo usuário. Estes terão o mesmo nome, mas com outras extensões de arquivo.

Aviso: Ao se especificar o nome de um arquivo já existente, o arquivo antigo será substituído e perdido.

Tópicos Relacionados: *Guia de Saída, Formato de Saída de Coluna, Formato de Saída Geográfico, Arquivo de Saída de Gráficos Temporais.*

Formato de Saída Geográfico

Arquivo KML: Para as análises espaciais e espaço-temporais, o SaTScan cria um arquivo KML que irá mostrar os clusters detectados no Google Earth e em outros softwares geográficos. Além da localização e do tamanho do cluster, o arquivo KML também contém informações básicas sobre cada cluster, tais como o número de casos observados, o risco relativo e o valor de 'p'. Se for selecionada essa opção, o SaTScan iniciará automaticamente o Google Earth e mostrará os resultados quando a análise for concluída. Para que isso funcione, é preciso ter o software livre do Google Earth instalado no

computador. Se for desejado um arquivo KML, mas não que o Google Earth inicie automaticamente, é possível desmarcar o início automático na Guia de Saída Avançada e, em vez disso, simplesmente clicar no arquivo KML quando quiser mostrar os resultados. Além do Google Earth, há muitos outros pacotes de softwares geográficos que podem também ler os arquivos KML.

Arquivos de Formato Shape: Para as análises espaciais e espaço-temporais, o SaTScan criará um arquivo Shape que pode ser usado para representar os clusters detectados em Sistemas de Informação Geográfica. Dois arquivos diferentes são criados, com extensões .shp e .shx.

O KML e o shapefile podem apenas ser gerados quando as coordenadas geográficas forem especificadas usando latitudes e longitudes.

Os nomes desses arquivos de saída são os mesmos nomes dos arquivos de saída com formato de texto, mas com diferentes extensões de arquivo.

Tópicos Relacionados: *Guia de Saída, Resultados das Análises, Arquivos de Saída de Formato de Coluna, Arquivo de Informações do Cluster, Arquivo de Informações das Localidades, Arquivo de Saída de Gráficos Temporais.*

Formato de Saída de Coluna

Além do arquivo padrão de resultados, que é automaticamente mostrado após a conclusão dos cálculos, é possível se solicitar cinco arquivos de saída adicionais em formato de coluna com diferentes tipos de informação. Estes são úteis para a importação dos resultados em outro software.

- **Arquivo de Informações do Cluster:** Uma linha para cada cluster, com informações sobre esse cluster.
- **Arquivo de Informações do Cluster Estratificado:** Para cada cluster, há uma linha para cada conjunto de dados, quando vários conjuntos de dados de entrada são usados, e há uma linha para cada categoria utilizada pelo modelo multinomial ou ordinal. Para cada cluster, conjunto de dados e categoria, o arquivo contém casos observados e esperados, a sua razão e o risco relativo. Este arquivo é útil apenas para os modelos multinomial e ordinal, ou quando há vários conjuntos de dados. Para outras análises, esse arquivo é redundante, já que contém um subconjunto das informações já presentes no Arquivo de Informações do Cluster.
- **Arquivo de Informações das Localidades:** Uma linha para cada ID de localidade, com informações sobre essa localidade e sua participação no cluster.
- **Arquivo de Estimativas de Risco:** Uma linha para cada ID de localidade, com o risco estimado nesta localidade.
- **Arquivo dos Logs das Razões de Verossimilhança Simulados:** Uma linha para cada conjunto de dados simulado, com o teste estatístico do log da razão de verossimilhança para esse conjunto de dados. Este arquivo é usado principalmente por estatísticos interessados nas propriedades de distribuição das estatísticas de varredura.

É necessário abrir manualmente todos esses arquivos após se concluir a execução das análises. Eles são fornecidos ou em formato ASCII ou dBase, de modo que eles podem ser facilmente importados para planilhas, sistemas de informação geográfica ou outros softwares de bancos de dados. Eles têm o mesmo nome que o arquivo de saída de texto padrão, mas com uma extensão de arquivo diferente.

Tópicos Relacionados: *Guia de Saída, Resultados das Análises, Arquivo de Informações do Cluster, Cabeçalhos das Colunas, Arquivo de Informações das Localidades, Estimativas de Risco para Cada Localidade, Formato de Saída Geográfico, Arquivo de Saída de Gráficos Temporais.*

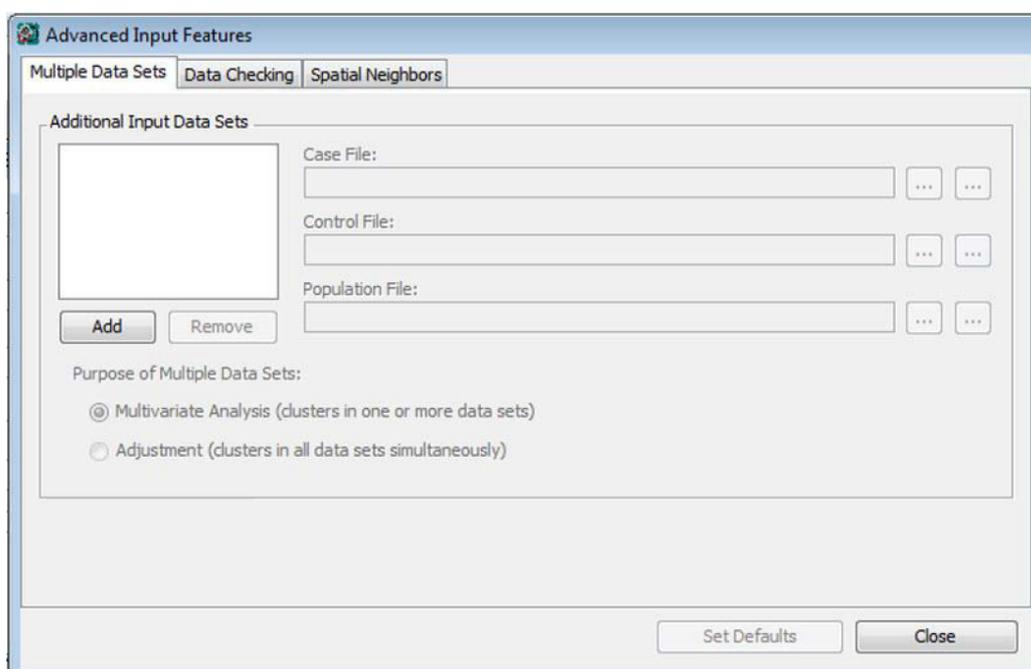
Recursos Avançados

Enquanto a maioria das análises do SaTScan pode ser realizada utilizando as funcionalidades das três guias básicas de entrada, análise e parâmetros de saída, opções adicionais são necessárias para alguns tipos de análises, e estas estão disponíveis como recursos avançados. Esses recursos são acessados através do botão 'Avançado' no canto inferior direito de cada uma das três guias principais. 'Avançado' deve ser interpretado como 'adicional' ou 'incomum', em vez de 'complexo', 'difícil' ou 'melhor'.

Dado que muitas das opções avançadas dependem das seleções feitas nas Guias de Entrada e de Análise, recomenda-se que estas duas abas sejam preenchidas primeiro.

Tópicos Relacionados: *Recursos Básicos do SaTScan, Guia de Múltiplos Conjuntos de Dados, Guia da Janela Espacial, Guia da Janela Temporal, Guia de Ajustes Espaciais e Temporais, Guia de Inferência, Guia de Saída Espacial.*

Guia de Múltiplos Conjuntos de Dados



Caixa de Diálogo da Guia de Múltiplos Conjuntos de Dados

É possível se procurar e avaliar aglomerados em múltiplos conjuntos de dados, tal como descrito na seção de Metodologia Estatística. O primeiro conjunto de dados é definido na Guia de Entrada. Até onze conjuntos de dados adicionais podem ser definidos na Guia de Múltiplos Conjuntos de Dados. Esses arquivos precisam ser da mesma classe que o primeiro. Ou seja, se o primeiro conjunto de dados consiste de um arquivo de casos e controles, então, todos os outros devem ser também. A precisão do tempo e o período de estudo também precisam ser os mesmos informados na Guia de Entrada.

Os conjuntos de dados são adicionados inicialmente clicando no botão 'Adicionar', e depois introduzindo os nomes dos arquivos, ou digitando-os na caixa de texto, ou usando o botão do navegador , ou através do Assistente de Importação do SaTScan, botão de 'Importar Arquivo' . Remove-se um conjunto de dados selecionando-o e clicando no botão "Remover".

Múltiplos conjuntos de dados podem ser usados para duas finalidades diferentes. Uma finalidade é quando existem diferentes tipos de dados, e se deseja saber se há um aglomerado em um ou mais dos conjuntos de dados. A evidência de um cluster poderia, então, ser proveniente exclusivamente de um

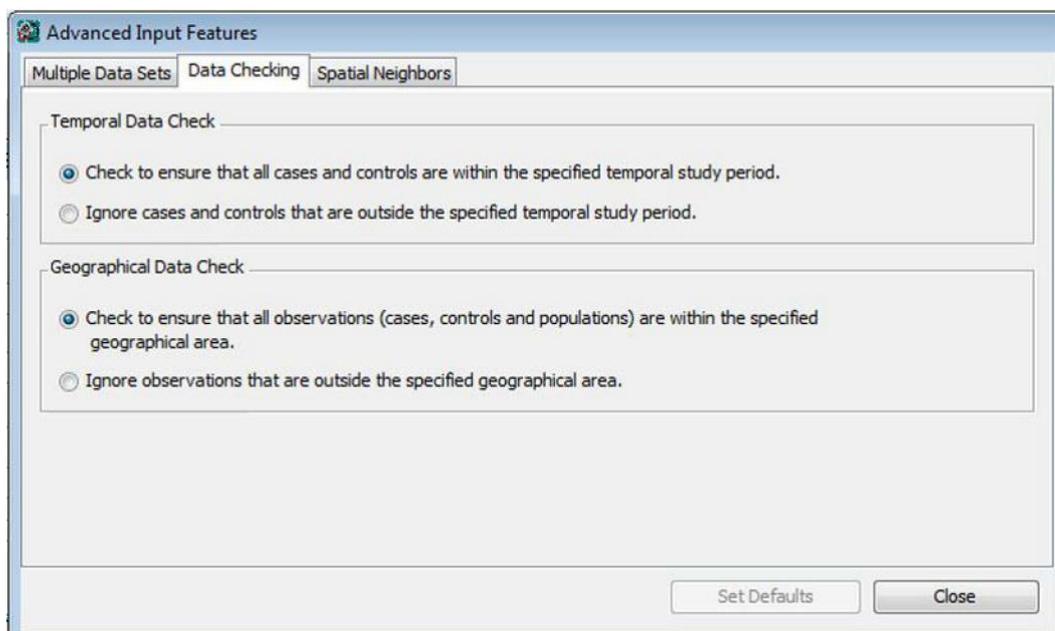
conjunto de dados, ou pode utilizar a evidência combinada de dois ou mais conjuntos de dados. A outra finalidade é se ajustar para covariáveis. Neste caso, a evidência de um aglomerado é baseada em todos os conjuntos de dados. A diferença é discutida em mais detalhes na seção de metodologia estatística.

Nota: Múltiplos conjuntos de dados não podem ser usados para o modelo contínuo de Poisson.

Aviso: O tempo de processamento é consideravelmente mais longo quando se analisam múltiplos conjuntos de dados, quando comparado a um único conjunto de dados. Por isso, não se recomenda utilizar múltiplos conjuntos de dados quando há muitas localidades no arquivo de coordenadas.

Tópicos Relacionados: *Recursos Avançados, Guia de Entrada, Varredura Multivariada com Múltiplos Conjuntos de Dados, Ajuste por Covariáveis Usando Múltiplos Conjuntos de Dados, Tempo de Processamento, Arquivo de Casos, Arquivo de Controles, Arquivo de População.*

Guia de Verificação de Dados



Caixa de Diálogo da Guia de Verificação de Dados

Verificação de Dados Temporais

Por padrão, o SaTScan irá verificar se todos os casos e todos os controles estão dentro do período temporal do estudo especificado. Nessa guia, é possível desligar isso. Casos e controles fora do período do estudo serão, então, ignorados. Isso pode ser usado, por exemplo, quando se quer analisar apenas um subconjunto temporal dos dados nos arquivos de entrada de casos e controles.

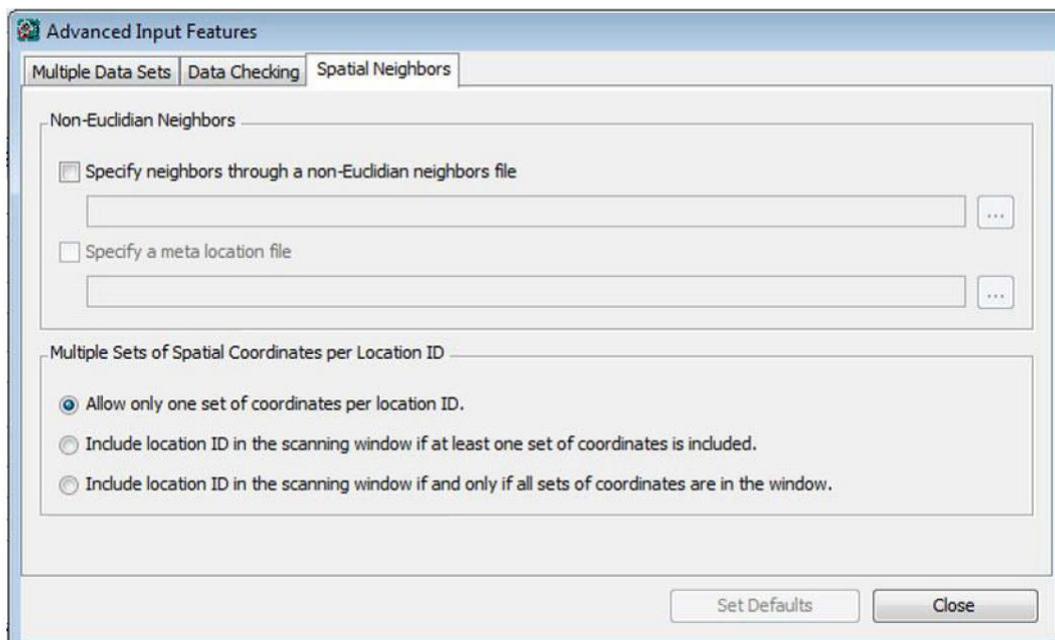
Verificação de Dados Geográficos

Por padrão, o SaTScan irá verificar se todos os casos, controles e números da população estão dentro da área geográfica especificada. Para a estatística de varredura discreta, isso significa que eles precisam estar em uma das localidades especificadas no arquivo de coordenadas. Para a estatística de varredura contínua, isso significa que todas as coordenadas especificadas no arquivo de coordenadas precisam estar dentro dos polígonos especificados. Nessa guia, é possível desativar esse procedimento de verificação de dados. Os dados fora da área geográfica do estudo serão, então, ignorados. Isto pode ser

usado, por exemplo, quando se quer analisar apenas um subconjunto geográfico dos dados, caso em que somente o arquivo de coordenadas geográficas precisa ser modificado para uma estatística de varredura discreta, enquanto que os outros arquivos podem ser usados como eles são.

Tópicos Relacionados: *Recursos Avançados, Arquivo de Casos, Guia de Entrada, Período de Tempo do Estudo.*

Guia de Vizinhos Espaciais



Caixa de Diálogo da Guia de Vizinhos Espaciais

Arquivo de Vizinhos Não-Euclidianos

Ao invés de usar círculos ou elipses definidos pelas distâncias Euclidianas entre as localidades especificadas nos arquivos de coordenadas e de grade, é possível se especificar manualmente uma matriz de vizinhança, para definir vizinhos pela distância não-Euclidiana. Para cada centroide, seu vizinho mais próximo, 2º mais próximo, 3º vizinho mais próximo são, por sua vez, especificados, e assim por diante. Esta opção é ativada marcando-se a caixa nessa guia e especificando o nome do arquivo de vizinhos contendo as informações da matriz de vizinhança. O formato do arquivo de vizinhos é descrito na seção de formato do Arquivo ASCII.

Arquivo de Meta-Localidade

Uma meta-localidade é um conjunto de dois ou mais IDs de localidades individuais. Quando uma meta-localidade é especificada no arquivo especial de vizinhos, todos os membros individuais da meta-localidade são simultaneamente introduzidos na janela de varredura. Esta opção é ativada marcando-se a caixa nessa guia e especificando o nome do arquivo de meta-localidade contendo as informações sobre os IDs das localidades individuais pertencentes a cada meta-localidade. O formato do arquivo de vizinhos é descrito na seção de formato do Arquivo ASCII.

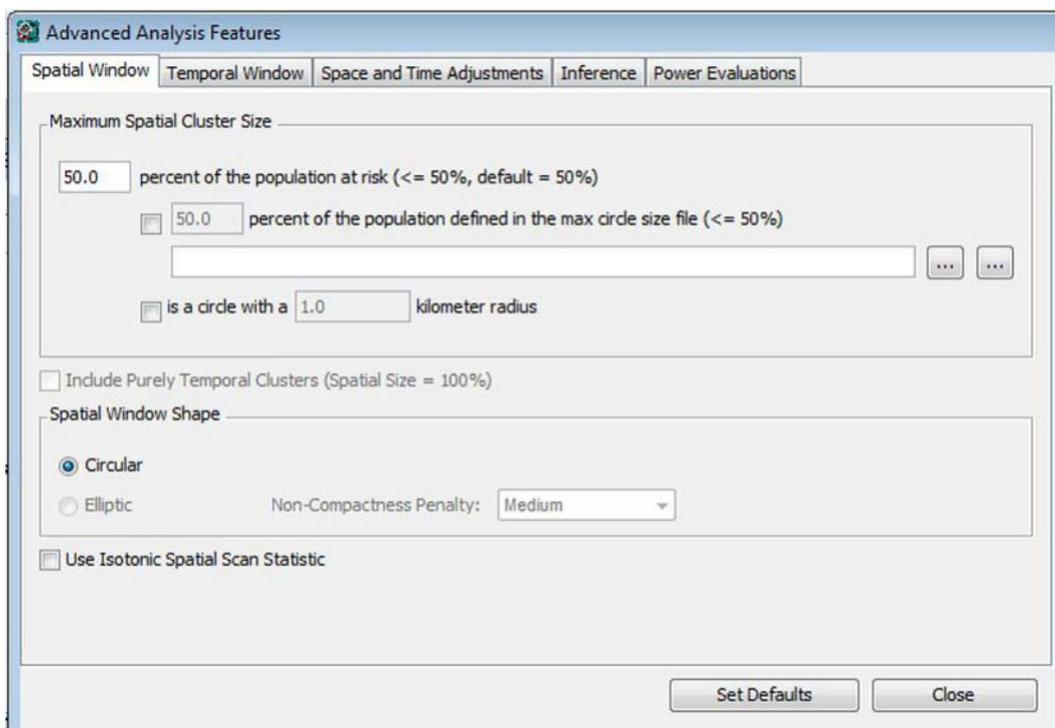
Múltiplas Coordenadas por Localidade

Cada ID de localidade é normalmente definido por um único conjunto de coordenadas, tal como um par (x, y) ou latitude e longitude. Como uma opção avançada, é possível se definir múltiplos conjuntos de coordenadas para cada ID de localidade, tal como ambos (x1,y1) e (x2,y2). O ID da localidade pode

então ser definido para ser incluído na janela de varredura circular nos casos: (i) se pelo menos um dos conjuntos de coordenadas estiver localizado no interior do círculo, ou (ii) se, e somente se, todos os conjuntos de coordenadas estiverem dentro do círculo. Os múltiplos conjuntos de coordenadas são especificados no arquivo de coordenadas, cada um em uma linha separada.

Tópicos Relacionados: *Recursos Avançados, Formato de Arquivo ASCII, Guia de Entrada, Arquivo de Meta-Localidade, Arquivo Especial de Vizinhos.*

Guia da Janela Espacial



Caixa de Diálogo da Guia da Janela Espacial

Usar a Guia da Janela Espacial para definir a natureza exata da janela de varredura com relação ao espaço.

Tópicos Relacionados: *Recursos Avançados, Guia de Análise, Guia da Janela Temporal, Tamanho Máximo do Cluster Espacial, Inclusão de Clusters Puramente Temporais.*

Tamanho Máximo do Cluster Espacial

O programa fará a varredura para clusters de tamanho geográfico entre zero e algum limite superior definido pelo usuário. O limite superior pode ser especificado ou como uma porcentagem da população empregada na análise, ou como um percentual de alguma outra população definida em um arquivo de tamanho máximo de círculo, ou em termos de dimensão geográfica, usando o raio do círculo. O máximo também pode ser definido utilizando uma combinação destes três critérios.

A escolha recomendada é se especificar o limite superior como um percentual da população sob risco, e utilizar 50% como o valor. É possível se especificar um valor máximo menor do que 50%, mas não mais do que 50%. Um cluster de tamanho maior indicaria áreas com taxas excepcionalmente baixas fora do círculo, ao invés de uma área com taxa excepcionalmente alta dentro do círculo (ou vice-versa, ao se procurar por clusters de taxas baixas). Em caso de dúvida, escolher uma porcentagem alta, visto que o

SaTScan irá, então, procurar clusters de tamanhos pequenos e grandes, sem qualquer viés de pré-seleção em termos do tamanho do cluster. Ao calcular a porcentagem, o SaTScan usa a população definida pelos casos e controles para o modelo de Bernoulli, a população sob risco ajustada por covariáveis a partir do arquivo de população para o modelo discreto de Poisson, os casos para os modelos de permutação no espaço-tempo, multinomial, ordinal, exponencial e normal, e o tamanho do círculo como uma porcentagem da área total dos polígonos para o modelo contínuo de Poisson. Quando existem múltiplos conjuntos de dados, o máximo é definido como uma porcentagem da combinação do total da população e dos casos em todos os conjuntos de dados.

Também é possível se especificar o tamanho máximo do círculo em termos da dimensão geográfica real, em vez da população. Se forem utilizadas coordenadas de latitude/longitude, então o raio máximo deve ser especificado em quilômetros. Se forem utilizadas coordenadas Cartesianas, o raio máximo deve ser especificado na mesma unidade das coordenadas Cartesianas.

Como alternativa, para as estatísticas de varredura discretas, é possível se especificar um arquivo de tamanho máximo do círculo para se definir o tamanho máximo do círculo. Este arquivo precisa conter uma 'população' para cada localidade, e o tamanho máximo do círculo é, então, definido como uma porcentagem dessa população, ao invés da população regular. Este recurso pode ser utilizado, por exemplo, quando se deseja definir os círculos no modelo de Bernoulli ou no modelo de população no espaço-tempo com base na população real, ao invés das localidades de casos e controles. Também pode ser utilizado quando se deseja que os círculos geográficos incluam, por exemplo, no máximo 10 municípios de um total de 100, independentemente da população desses municípios. Isto é conseguido atribuindo-se uma 'população' de 1 para cada município no arquivo especial de tamanho máximo do círculo e então definindo o tamanho máximo do círculo como sendo 10% dessa 'população'.

Se uma análise espaço-temporal prospectiva é realizada, ajustada para as análises anteriores, e se o tamanho máximo do círculo é definido como uma porcentagem da população, então o arquivo especial de tamanho máximo do círculo deve ser utilizado. Isto é para garantir que os círculos geográficos avaliados não mudem ao longo do tempo.

Tópicos Relacionados: *Recursos Avançados, Guia da Janela Espacial, Arquivo de Tamanho Máximo do Círculo, Inclusão de Clusters Puramente Temporais, Tempo de Processamento.*

Inclusão de Clusters Puramente Temporais

Um cluster puramente temporal é aquele que inclui toda a área geográfica, mas apenas um período de tempo limitado. Ao se fazer uma análise espaço-temporal, é possível permitir que potenciais clusters contenham toda a área geográfica sob estudo, como uma exceção ao tamanho máximo do cluster espacial escolhido. Dessa forma, os clusters puramente temporais são incluídos na coleção de janelas avaliadas.

Nota: Esta opção não está disponível para o modelo de permutação no espaço-tempo, visto que esse modelo ajusta automaticamente para clusters puramente temporais. Ao se ajustar para clusters puramente temporais usando a randomização estratificada, todos os clusters puramente temporais já são ajustados, e este parâmetro não tem efeito sobre a análise.

Tópicos Relacionados: *Recursos Avançados, Guia da Janela Espacial, Tamanho Máximo do Cluster Espacial, Inclusão de Clusters Puramente Espaciais, Ajuste de Tendência Temporal, Tempo de Processamento.*

Janela de Varredura Elíptica

Como uma opção avançada, é possível utilizar uma janela de varredura que consiste não apenas de círculos, mas também de elipses de diferentes formatos e ângulos. Quando a estatística de varredura espacial elíptica é requerida, o SaTScan utiliza a janela circular mais cinco diferentes formas elípticas, onde a razão entre os eixos mais longo e mais curto da elipse é de 1,5; 2; 3; 4 ou 5. Para cada forma, um número diferente de ângulos de elipse são utilizados, para os números 4, 6, 9, 12 e 15, respectivamente, dependendo da forma elíptica. O eixo norte-sul é sempre um dos ângulos incluídos, e o restante é igualmente espaçado ao redor do círculo. Para cada formato e ângulo, todos os tamanhos possíveis de

elipses são usados, até um limite superior especificado pelo usuário, da mesma maneira que para a janela circular.

Quando se utiliza uma janela de formato elíptico, é possível solicitar uma penalidade de não-compactação (excentricidade), que favorecerá as elipses mais compactas sobre aquelas menos compactas, mesmo quando elas tiverem razões de verossimilhança ligeiramente inferiores, e as elipses menos compactas, quando a diferença for maior. A fórmula para a penalidade é $[4s/(s+1)^2]^a$, onde s é a forma de janela elíptica definida como a razão entre o comprimento do eixo mais longo e o eixo mais curto da elipse. Com uma forte penalidade $a=1$, com uma média penalidade $a=1/2$ e sem penalidade $a=0$.

Nota: No modo de 'lote', é possível se solicitar ao SaTScan para usar qualquer outro conjunto de elipses para definir a janela de varredura, e qualquer valor do parâmetro de penalidade de excentricidade maior do que zero.

Nota: A opção de janela elíptica só pode ser usada quando são utilizadas coordenadas Cartesianas bidimensionais regulares, mas não quando elas são especificadas como latitude/longitude. Nesse último caso, deve-se primeiro fazer um mapa de projeção plana a partir das coordenadas de latitude/longitude, para o qual há muitas propostas diferentes na literatura da geografia.

Nota: A janela de varredura elíptica não está disponível para o modelo contínuo de Poisson.

Tópicos Relacionados: *Recursos Avançados, Tempo de Processamento, Inclusão de Clusters Puramente Espaciais, Teste da Razão de Verossimilhança, Tamanho Máximo do Cluster Espacial, Estatísticas de Varredura Espacial Temporal e Espaço-Temporal, Guia da Janela Espacial.*

Estatística de Varredura Espacial Isotônica

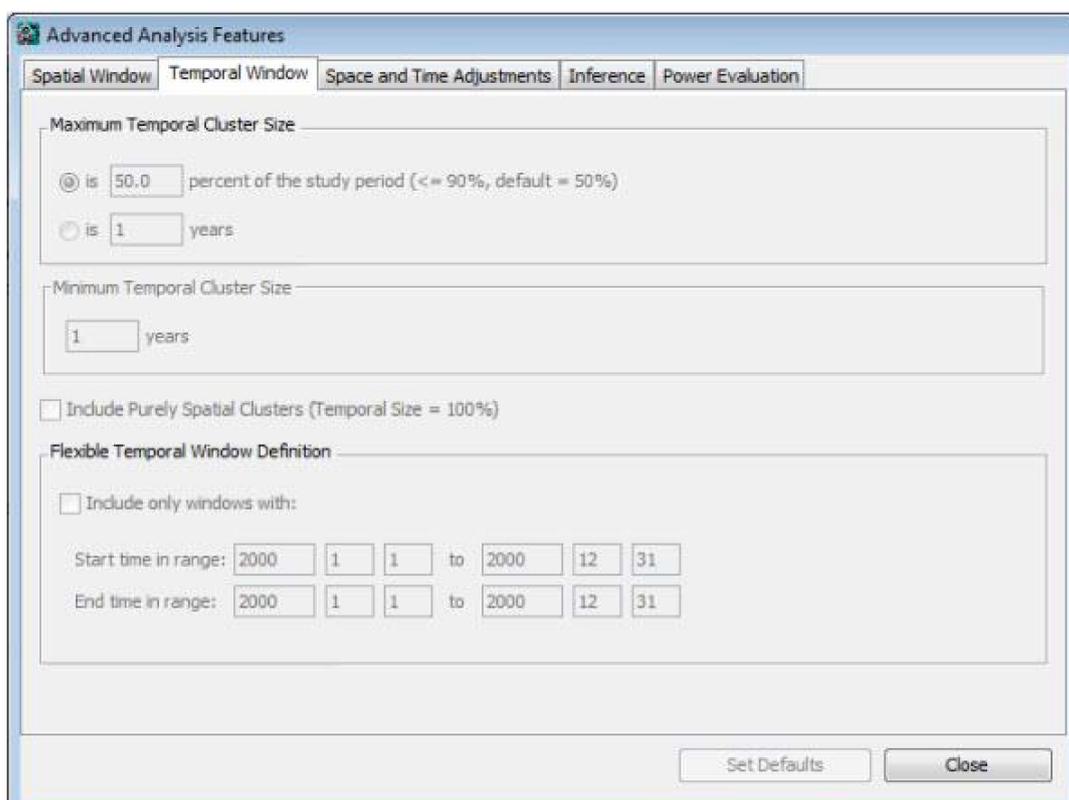
Na estatística de varredura espacial padrão, o modelo alternativo é de que há uma taxa mais elevada no interior do cluster, e uma taxa mais baixa fora do cluster. Ao passo que a taxa não é assumida como sendo constante nem no interior e nem no exterior da janela, qualquer variação na taxa não é levada em conta no cálculo da verossimilhança. Ao invés de uma janela circular, a estatística de varredura espacial isotônica¹³ define a janela usando um conjunto de círculos sobrepostos de diferentes tamanhos que são centrados no mesmo ponto. O modelo alternativo é de que a taxa é mais alta dentro do círculo mais interno, um pouco mais baixa entre o primeiro e o segundo círculos, e assim por diante, até o último círculo. Não há um número pré-definido de círculos ou qualquer suposição prévia sobre seus respectivos tamanhos, exceto que o maior círculo deve ser menor do que o tamanho máximo do círculo especificado pelo usuário. Em vez disso, o método encontra a coleção de círculos que maximiza a estatística da razão de verossimilhança. Assim como na estatística de varredura espacial padrão, a janela se move sobre o espaço, considerando muitos possíveis círculos de centroides. O método ajusta para a múltipla testagem, inerente tanto às diversas localidades do cluster consideradas, como às muitas possíveis coleções de círculos usadas para a janela de varredura.

Nota: Embora o método avalie uma janela com múltiplos círculos, por vezes é um único círculo que proporciona a mais alta verossimilhança e, por conseguinte, define o cluster mais provável.

Nota: A estatística de varredura espacial isotônica está apenas disponível para as análises puramente espaciais utilizando tanto o modelo discreto de Poisson como o modelo de Bernoulli. Ela não está disponível para múltiplos conjuntos de dados.

Tópicos Relacionados: *Recursos Avançados, Tamanho Máximo do Cluster Espacial.*

Guia da Janela Temporal



Caixa de Diálogo da Guia da Janela Temporal

Usar a Guia da Janela Temporal para definir a natureza exata da janela de varredura com relação ao tempo.

Tópicos Relacionados: *Recursos Avançados, Guia de Análise, Guia da Janela Espacial, Tamanho Máximo do Cluster Temporal, Inclusão de Clusters Puramente Espaciais, Definição da Janela Temporal Flexível.*

Tamanho Máximo do Cluster Temporal

Para as análises puramente temporais e espaço-temporais, o tamanho máximo do cluster temporal pode ser especificado em termos de uma porcentagem do período de estudo como um todo, ou em um certo número de dias, meses ou anos. O valor máximo deve ser pelo menos tão grande quanto o comprimento do intervalo de tempo agregado. Se for especificado como uma porcentagem, então para os modelos de Bernoulli e Poisson, esta pode ser no máximo 90%, e para o modelo de permutação no espaço-tempo, no máximo 50%. O valor recomendado é de 50%.

Tópicos Relacionados: *Guia da Janela Temporal, Tamanho Mínimo do Cluster Temporal, Tamanho Máximo do Cluster Espacial, Inclusão de Clusters Puramente Espaciais, Definição da Janela Temporal Flexível, Agregação do Tempo.*

Tamanho Mínimo do Cluster Temporal

Para as análises puramente temporais e espaço-temporais, o tamanho mínimo do cluster temporal pode ser especificado em termos de um certo número de dias, meses ou anos. Se a precisão do tempo no arquivo de entrada é dia, mês ou ano, então ele é especificado na mesma unidade que a unidade de agregação de tempo na Guia de Análise. Se a precisão do tempo no arquivo de entrada é genérica, então, ele é especificado na mesma unidade genérica. O valor padrão é uma unidade, o que significa que não há uma restrição mínima para o tamanho do cluster temporal, exceto aquela imposta pelos dados ou pela agregação dos dados.

Tópicos Relacionados: *Guia da Janela Temporal, Tamanho Máximo do Cluster Temporal, Tamanho Máximo do Cluster Espacial, Inclusão de Clusters Puramente Espaciais, Definição da Janela Temporal Flexível, Agregação do Tempo.*

Inclusão de Clusters Puramente Espaciais

Além do tamanho máximo do cluster temporal, é possível também permitir clusters que contenham todo o período de tempo sob estudo. Dessa forma, os clusters puramente espaciais são incluídos entre as janelas avaliadas. O objetivo de se especificar um tamanho temporal máximo, mas ainda assim incluindo clusters puramente espaciais, é eliminar aglomerados contendo todo o período do estudo, com exceção de um pequeno intervalo de tempo bem no início ou bem no final do período de estudo.

Nota: Ao se ajustar para clusters puramente espaciais usando a randomização estratificada, todos os clusters puramente espaciais já são ajustados, e este parâmetro não tem efeito sobre a análise.

Tópicos Relacionados: *Guia da Janela Temporal, Tamanho Máximo do Cluster Temporal, Inclusão de Clusters Puramente Temporais, Ajuste Espacial.*

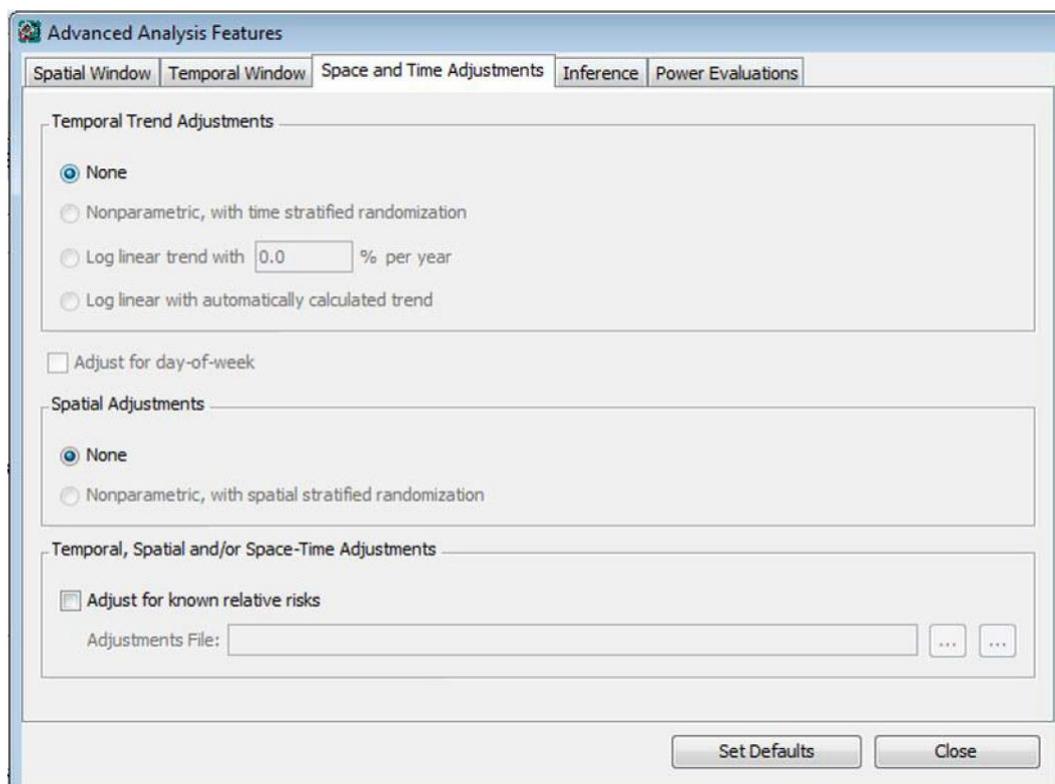
Definição da Janela Temporal Flexível

Para análises retrospectivas, o SaTScan irá avaliar todas as janelas temporais inferiores ao máximo especificado, e para as análises prospectivas, o mesmo é verdadeiro, com a restrição adicional de que o final da janela é idêntico à data final do período de estudo. Quando necessário, o SaTScan pode ser mais flexível do que isso, e é possível se definir a janela de varredura como qualquer período de tempo que inicia dentro de uma 'faixa inicial' predefinida e termina dentro de uma 'faixa final' predefinida.

Essa opção é disponível apenas quando uma análise retrospectiva puramente temporal ou uma análise retrospectiva espaço-temporal é selecionada na Guia de Análise.

Tópicos Relacionados: *Guia da Janela Temporal, Tamanho Máximo do Cluster Temporal, Inclusão de Clusters Puramente Espaciais, Período do Estudo, Agregação do Tempo.*

Guia de Ajustes Espaciais e Temporais



Caixa de Diálogo da Guia de Ajustes Espaciais e Temporais

Ajusta-se para covariáveis ou incluindo-as nos arquivos de casos e de população, ou usando múltiplos conjuntos de dados, dependendo do modelo de probabilidade utilizado. As funcionalidades nesta guia são usadas para ajustar para tendências e variação temporal, espacial e espaço-temporal. A maioria está disponível apenas quando se utiliza o modelo de probabilidade discreto de Poisson. A única exceção é o ajuste do espaço por interação de dia-da-semana para o modelo de permutação no espaço-tempo.

Tópicos Relacionados: *Recursos Avançados, Guia de Análise, Ajustes Espaciais e Temporais, Ajuste de Tendência Temporal, Ajuste Espacial, Ajuste com Risco Relativo Conhecido, Modelo de Poisson.*

Ajuste de Tendência Temporal

Pode se ajustar para tendências temporais de três modos diferentes:

Não-paramétrico: Quando o ajuste é não-paramétrico, o SaTScan ajusta para qualquer tipo de variação puramente temporal. Isto é feito por estratificação da randomização pelos intervalos de tempo agregados, de modo que cada intervalo de tempo tenha o mesmo número de casos nos conjuntos de dados reais e aleatórios. Isto é, apenas a localidade espacial de um caso é randomizada.

Log de tendência linear, especificado pelo usuário: Especificar um aumento ou um decréscimo percentual anual no risco. Uma tendência decrescente é especificada com um número negativo. Por exemplo, se a taxa diminui em 1,4% ao ano, então escrever “-1.4” na caixa “% por ano”.

Log de tendência linear, calculado automaticamente: Ao invés de o usuário especificar o risco relativo ajustado, o SaTScan pode calcular a tendência observada nos dados, e então ajustar exatamente para essa quantidade de aumento ou decréscimo.

O padrão é não ter nenhum ajuste de tendência temporal.

Nota: Ao se fazer uma análise de variação espacial nas tendências temporais, não é possível se ajustar para tendências temporais.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Ajuste por Dia-da-Semana, Ajustes Espaciais e Temporais, Ajuste Espacial, Ajuste com Risco Relativo Conhecido, Modelo de Poisson.*

Ajuste por Dia-da-Semana

Para alguns dados, tais como consultas médicas, pode haver uma variação natural de dia-da-semana, para a qual se deve ajustar. Isto é feito de uma forma não-paramétrica. O efeito de se solicitar um ajuste por dia-da-semana na Guia de Ajustes Espaciais e Temporais é o mesmo efeito da inclusão de uma covariável de dia-da-semana nos arquivos de entrada.

Com o modelo de permutação no espaço-tempo, ajusta-se automaticamente para o dia-da-semana se for especificamente solicitado ou não, como parte do seu ajuste completo para qualquer variação puramente temporal. Em vez disso, é possível se solicitar um ajuste para o dia-da-semana pela interação espacial. Este ajusta para o fato de que algumas áreas geográficas podem ter um padrão de dia-da-semana diferente de outras áreas. Por exemplo, uma clínica médica pode ter um grande número de consultas em finais de semana, enquanto outra clínica pode estar fechada nos finais de semana. O efeito é o mesmo que incluir o dia-da-semana como uma covariável no arquivo de casos utilizado pelo modelo de permutação no espaço-tempo.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Ajustes de Tendência Temporal, Ajuste com Risco Relativo Conhecido, Modelo de Permutação no Espaço-Tempo.*

Ajuste Espacial

Quando uma análise puramente espacial é realizada, a finalidade é se encontrar clusters puramente espaciais. Para a estatística de varredura espaço-temporal, esta função ajusta para todos esses aglomerados, para verificar se existe quaisquer aglomerados espaço-temporais não explicados pelos aglomerados puramente espaciais. Isto é feito de uma forma não paramétrica, através da randomização estratificada por localidade, de modo que o número total de casos em cada localidade específica é a mesma nos conjuntos de dados reais e aleatórios. Ou seja, apenas o tempo de um caso é randomizado.

O padrão é não ter nenhum ajuste espacial.

Nota: Não é possível se ajustar simultaneamente para clusters espaciais e clusters puramente temporais usando a randomização estratificada. Se ambos os tipos de ajustes são desejados, o modelo de permutação no espaço-tempo deve ser, então, usado. É possível se ajustar para clusters puramente espaciais com randomização estratificada em conjunto com um ajuste temporal usando um log de tendência linear.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Ajustes Espaciais e Temporais, Ajustes de Tendência Temporal, Ajuste com Risco Relativo Conhecido, Modelo de Poisson.*

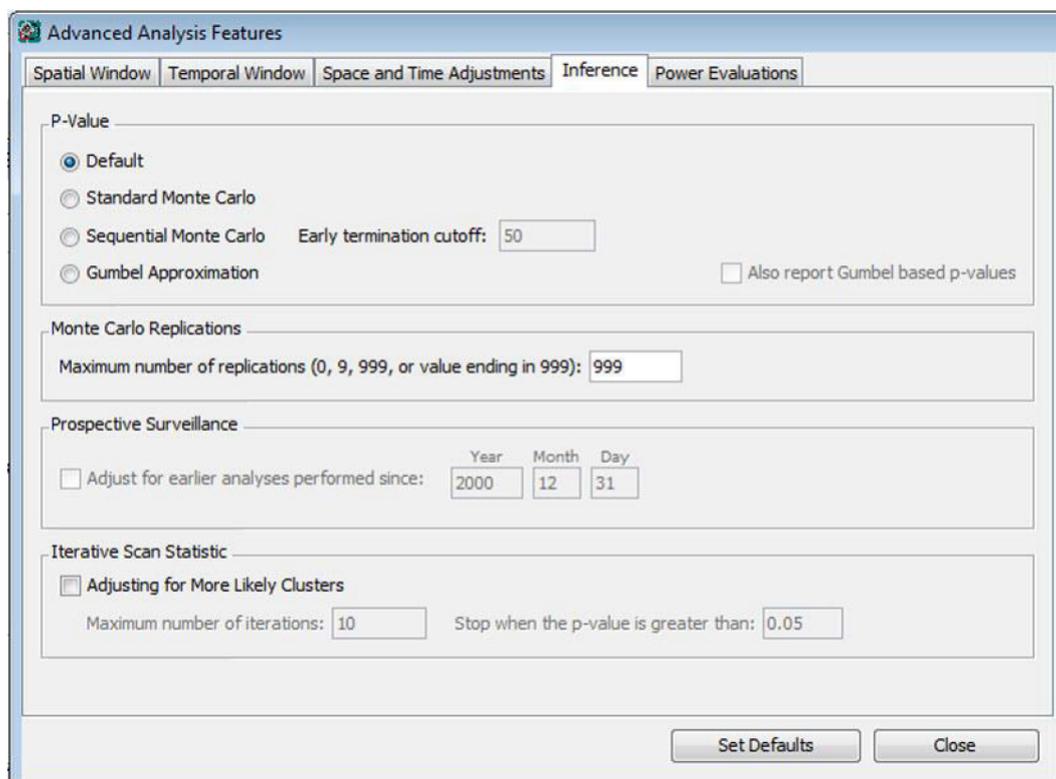
Ajuste com Riscos Relativos Conhecidos

A maneira mais flexível para se ajustar uma análise do modelo discreto de Poisson é usar o arquivo especial de ajustes. Neste arquivo, um risco relativo é especificado para qualquer combinação de localidade e período de tempo, e o SaTScan irá ajustar as contagens esperadas para cima ou para baixo com base nesse risco relativo. Um uso desta opção é se ajustar para a falta de dados, especificando-se um risco relativo de zero para aquelas combinações de localidade e tempo para as quais os dados estão ausentes.

O formato requerido para o Arquivo de Ajustes está descrito na seção de Dados de Entrada.

Tópicos Relacionados: *Guia de Ajustes Espaciais e Temporais, Ajustes Espaciais e Temporais, Ajuste de Tendência Temporal, Ajuste Espacial, Arquivo de Ajustes, Modelo de Poisson.*

Guia de Inferência



The image shows a software dialog box titled "Advanced Analysis Features". It has five tabs: "Spatial Window", "Temporal Window", "Space and Time Adjustments", "Inference", and "Power Evaluations". The "Inference" tab is selected. The dialog is divided into several sections:

- P-Value:** Contains four radio buttons: "Default" (selected), "Standard Monte Carlo", "Sequential Monte Carlo", and "Gumbel Approximation". Next to "Sequential Monte Carlo" is a text input field for "Early termination cutoff:" with the value "50". To the right of "Gumbel Approximation" is a checkbox labeled "Also report Gumbel based p-values".
- Monte Carlo Replications:** A text input field for "Maximum number of replications (0, 9, 999, or value ending in 999):" with the value "999".
- Prospective Surveillance:** A checkbox labeled "Adjust for earlier analyses performed since:" is unchecked. To its right are three date input fields: "Year" (2000), "Month" (12), and "Day" (31).
- Iterative Scan Statistic:** A checkbox labeled "Adjusting for More Likely Clusters" is unchecked. Below it are two text input fields: "Maximum number of iterations:" (10) and "Stop when the p-value is greater than:" (0.05).

At the bottom right of the dialog are two buttons: "Set Defaults" and "Close".

Caixa de Diálogo da Guia de Inferência

Esta guia é acessada clicando-se no botão 'Avançado' no canto inferior direito da Guia de Análise.

Tópicos Relacionados: *Recursos Avançados, Guia de Análise, Término Antecipado das Simulações, Ajuste para Análises Prévia em Vigilância Prospectiva.*

Valor de 'P'

Para calcular os valores de 'p' para os clusters detectados, o programa SaTScan utiliza simulações de computador para gerar um número de replicações aleatórias do conjunto de dados sob a hipótese nula. Se a razão da máxima verossimilhança calculada para o cluster mais provável no conjunto de dados reais for elevada em comparação às razões das máximas verossimilhanças calculadas para os clusters mais prováveis nos conjuntos de dados aleatórios, esta é uma evidência contra a hipótese nula e a favor da existência de clusters. A comparação pode ser feita em quaisquer das três formas a seguir, ou utilizando-se uma combinação das mesmas, sendo essa última a opção padrão.

Monte Carlo Padrão: A estatística do teste é calculada para cada replicação aleatória, bem como para o conjunto de dados reais, e se este último estiver entre os 5% maiores, então o teste é significativo ao nível de 0,05. Se ele estiver entre os 1% mais elevados, o teste é significativo ao nível de 0,01, e assim por diante. Este é o chamado 'teste de hipótese de Monte Carlo', e foi proposto pela primeira vez por Dwass¹⁵. Independentemente do número de replicações de Monte Carlo escolhido, o teste de hipótese não é tendencioso, resultando em um nível de significância correto que não é nem conservador, nem liberal, nem uma estimativa. O número de replicações afeta o poder do teste, com mais replicações conferindo um poder ligeiramente superior.

No SaTScan, o número de replicações deve ser de pelo menos 999, para garantir um excelente poder para todos os tipos de conjuntos de dados. Para conjuntos de dados de tamanho pequeno ou médio, 9999 replicações são recomendadas, desde que o tempo de computação não seja uma questão importante.

Monte Carlo Sequencial: Com mais replicações de Monte Carlo, o poder da estatística de varredura é maior, mas também é mais demorado de se executar. Quando o valor de 'p' é pequeno, em geral o esforço vale à pena, mas para grandes valores de 'p', geralmente é irrelevante se, por exemplo, 'p' = 0,7535, ou 'p' = 0,8545. O SaTScan fornece a opção de encerrar as simulações de Monte Carlo mais cedo quando o valor de 'p' é grande, empregando o teste sequencial de Monte Carlo^{16,17}. Com esta opção, os cálculos do SaTScan terminarão tão logo um número fixo de replicações de Monte Carlo tenha uma razão de verossimilhança maior do que a razão de verossimilhança do conjunto de dados reais. O valor padrão é de 50 réplicas. Se o número fixo nunca for alcançado, os cálculos continuarão até que o número máximo de réplicas de Monte Carlo tenha sido alcançado. Com os valores padrão de 50 e 999, não há perda de poder ao nível de alfa=0,05, quando se compara o teste de Monte Carlo sequencial com o teste padrão.

Aproximação de Gumbel: Com 999 réplicas aleatórias, o menor valor de 'p' que o teste de hipótese de Monte Carlo pode relatar é $1/(999+1)=0,001$. Do mesmo modo, com 9999 réplicas, o menor valor de 'p' possível é 0,0001. Como uma alternativa ao teste de hipótese de Monte Carlo, é possível se empregar a distribuição de valores extremos de Gumbel para estimar os valores de 'p' aproximados¹⁸. Com esta abordagem, não existe um limite inferior para os valores de 'p' resultantes (exceto $p>0$, é claro). O método funciona através da geração de 999, ou algum outro número de réplicas aleatórias dos dados sob a hipótese nula. A razão da máxima verossimilhança de cada réplica é então usada para ajustar uma distribuição de Gumbel para os dados usando métodos de estimações momentâneas. Uma vez que a distribuição de Gumbel que melhor se ajusta aos dados tenha sido obtida, o valor de 'p' é calculado como a probabilidade de que esta distribuição gere um valor maior do que a razão da máxima verossimilhança observada para o cluster mais provável do conjunto de dados reais.

Para a estatística de varredura puramente espacial com os modelos de probabilidade discreto de Poisson e Bernoulli, demonstrou-se que a distribuição de Gumbel se ajusta muito bem aos dados e gera valores de 'p' muito precisos¹⁸. Ainda não houve quaisquer estudos semelhantes para as outras estatísticas de varredura, por isso esta opção está, no momento, disponível apenas para as análises puramente espaciais com esses dois modelos de probabilidade.

Valor de 'P' Padrão: De forma padrão, o SaTScan irá calcular os valores de 'p' utilizando uma combinação dos três modos descritos acima. Por exemplo, ele pode apresentar o valor de 'p' baseado no Monte Carlo sequencial, a menos que o valor de 'p' seja muito pequeno, situação em que ele irá relatar a aproximação de Gumbel. A abordagem exata depende do tipo de análise solicitada, bem como da natureza dos dados, uma vez que o Monte Carlo sequencial e a aproximação de Gumbel não funcionam para todas as análises e os conjuntos de dados.

Nota: Nas versões anteriores do SaTScan, o Monte Carlo padrão era o método definido para o cálculo dos valores de 'p', e as outras opções são novas a partir da versão 9.0.

Tópicos Relacionados: *Guia de Análise, Tempo de Processamento, Guia de Inferência, Teste da Razão de Verossimilhança, Classificação (Rank) de Monte Carlo, Replicações de Monte Carlo, Gerador de Números Aleatórios, Resultados da Análise.*

Ajuste para Análises Prévia em Vigilância Prospectiva

Quando se fazem análises prospectivas puramente temporais ou análises prospectivas espaço-temporais repetidamente, de uma forma periódica no tempo, é possível se ajustar a inferência estatística (valores de 'p') para a testagem múltipla inerente às análises repetidas realizadas. Para fazer isso, basta marcar a caixa "ajustar para análises prévias", e especificar a data para a qual se deseja ajustar todas as análises subsequentes. Esta data tem que ser maior ou igual à data de início do período de estudo, e menor ou igual à data final do período de estudo, conforme especificado na Guia de Entrada.

O ajuste é feito por meio da utilização de um diferente conjunto de cilindros no cálculo da máxima verossimilhança para os conjuntos de dados reais e aleatórios. Para o conjunto de dados real, o máximo é obtido sobre todos os clusters atualmente vivos, ou seja, aqueles cilindros que atingem a data final do período de estudo. Para os conjuntos de dados aleatórios, o máximo é tomado sobre todos os cilindros com uma data final posterior à data de ajuste especificada nesta data. Isto é, ele toma o máximo sobre todos os cilindros previamente utilizados nas análises anteriores, para as quais se está ajustando agora.

Para que o ajuste seja correto, é importante que a janela de varredura espacial seja a mesma para cada análise realizada ao longo do tempo. Isto significa que os pontos da grade que definem os centróides circulares precisam permanecer os mesmos. Se os IDs das localidades no arquivo de coordenadas permanecerem os mesmos em cada análise periódica no tempo, então não há problema. Por outro lado, se novas IDs forem adicionadas ao arquivo de coordenadas ao longo do tempo, então é preciso usar um arquivo especial de grade e manter esse arquivo em todas as análises. Além disso, quando se ajusta para análises prévias, e se o tamanho máximo do círculo é definido como uma porcentagem da população, então o arquivo especial de tamanho máximo do círculo deve ser usado.

Tópicos Relacionados: *Guia de Inferência, Tempo de Processamento, Tipo de Análise, Estatísticas de Varredura Espacial Temporal e Espaço-Temporal.*

Estatística de Varredura Iterativa

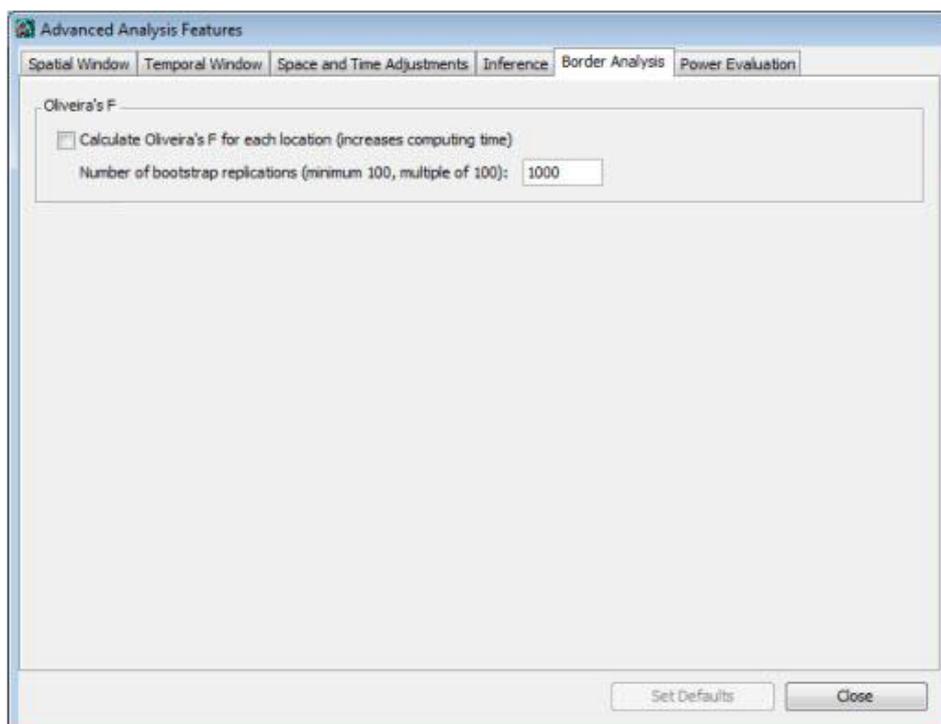
A opção de varredura iterativa é utilizada para ajustar os valores de 'p' de aglomerados secundários para os clusters mais prováveis que são encontrados e relatados. Isto é feito realizando-se a análise em várias iterações, removendo-se o cluster mais provável encontrado em cada iteração e, em seguida, reanalisando os dados restantes²³. O usuário deve especificar o número máximo de iterações permitidas, no intervalo de 1-32000. O usuário também pode solicitar que as iterações parem quando o último cluster encontrado tiver um valor de 'p' maior do que um limite inferior determinado.

Em termos de tempo de processamento, cada iteração demora aproximadamente a mesma quantidade de tempo de uma análise regular com os mesmos parâmetros.

Nota: Demonstrou-se que os valores de 'p' da estatística de varredura iterativa são válidos para uma análise puramente espacial com o modelo discreto de Poisson. O recurso também está disponível para as outras estatísticas de varredura discretas, mas não se sabe se os valores de 'p' são tão precisos. A estatística de varredura iterativa não está disponível para as estatísticas de varredura espaço-temporais, nem para o modelo contínuo de Poisson.

Tópicos Relacionados: *Ajuste para Clusters Mais Prováveis, Guia de Inferência, Tempo de Processamento.*

Guia de Análise de Fronteira



Caixa de Diálogo da Guia de Análise de Fronteira

Esta guia é acessada clicando-se no botão 'Avançado' no canto inferior direito da Guia de Análise.

Quando a estatística de varredura espacial detecta um cluster estatisticamente significativo, suas fronteiras exatas são incertas. Algumas localidades dentro do cluster detectado podem não ter um risco alto, enquanto pode haver um risco aumentado em algumas localidades fora do cluster detectado. Isto é obviamente verdadeiro se o verdadeiro cluster não é circular quando uma janela de varredura circular é usada, ou não é elíptico quando uma janela de varredura elíptica é usada; mas também é verdade mesmo que o verdadeiro cluster seja circular em uma análise circular, ou elíptico em uma análise elíptica. Isto ocorre devido ao tamanho de amostra limitado em cada localidade geográfica. Para se obter estimativas melhores das fronteiras desses aglomerados verdadeiros, poderia se aumentar o tamanho da amostra, mas isso raramente é possível. O que é possível é fornecer alguma medida de certeza sobre se uma determinada localidade realmente pertence ao verdadeiro cluster ou não. Uma dessas medidas é o 'F' de Oliveira.

Tópicos Relacionados: *Recursos Avançados, Guia de Análise, 'F' de Oliveira.*

'F' de Oliveira

Esta opção está disponível para uma análise de Poisson puramente espacial, quando os clusters mais prováveis são relatados hierarquicamente, sem sobreposição geográfica.

O 'F' de Oliveira é uma medida do quão provável uma localidade particular está no cluster verdadeiro, se e quando existir um verdadeiro cluster. Ele é calculado para cada localidade, seja um setor censitário, código postal ou país, dependendo dos dados de entrada. Um valor de 'F' mais alto significa que é mais provável ser parte do verdadeiro aglomerado, mas, e isto é importante, ainda que se tenha um valor no intervalo [0,1], isso NÃO deve ser interpretado como a probabilidade de pertencente ao cluster. A única

interpretação deve ser que um valor maior significa que é mais provável de ser parte do verdadeiro aglomerado. Se não houver aglomerados estatisticamente significativos, não há nenhum 'F' de Oliveira para relatar, uma vez que eles são todos iguais a zero.

Tal como foi implementado no SaTScan, o 'F' de Oliveira é uma variação do método descrito por Oliveira et al. (2011). Quando há apenas um único aglomerado estatisticamente significativo relatado, ele é calculado da seguinte forma:

1. Condicionado ao número total de casos no conjunto de dados, gerar um conjunto de dados aleatório, onde a contagem esperada para uma localidade é igual à contagem observada, usando uma distribuição de probabilidade multinomial. Estas são as contagens geradas por distribuições independentes de Poisson condicionadas ao número total de casos observados.
2. Para este conjunto de dados aleatórios, executar a estatística de varredura espacial com as contagens esperadas baseadas na população original. Observar quais localidades estão dentro do cluster mais provável.
3. Repetir os passos 1 e 2 um grande número de vezes, por exemplo, 1.000. Para cada localidade, observar em quantas das 1.000 vezes ela pertenceu ao cluster mais provável. O 'F' de Oliveira é definido como a proporção dessas vezes em que a localidade pertenceu ao cluster mais provável.

Quando a análise do SaTScan gera dois clusters estatisticamente significativos, o passo 2 acima utiliza os dois aglomerados mais prováveis, e assim por diante.

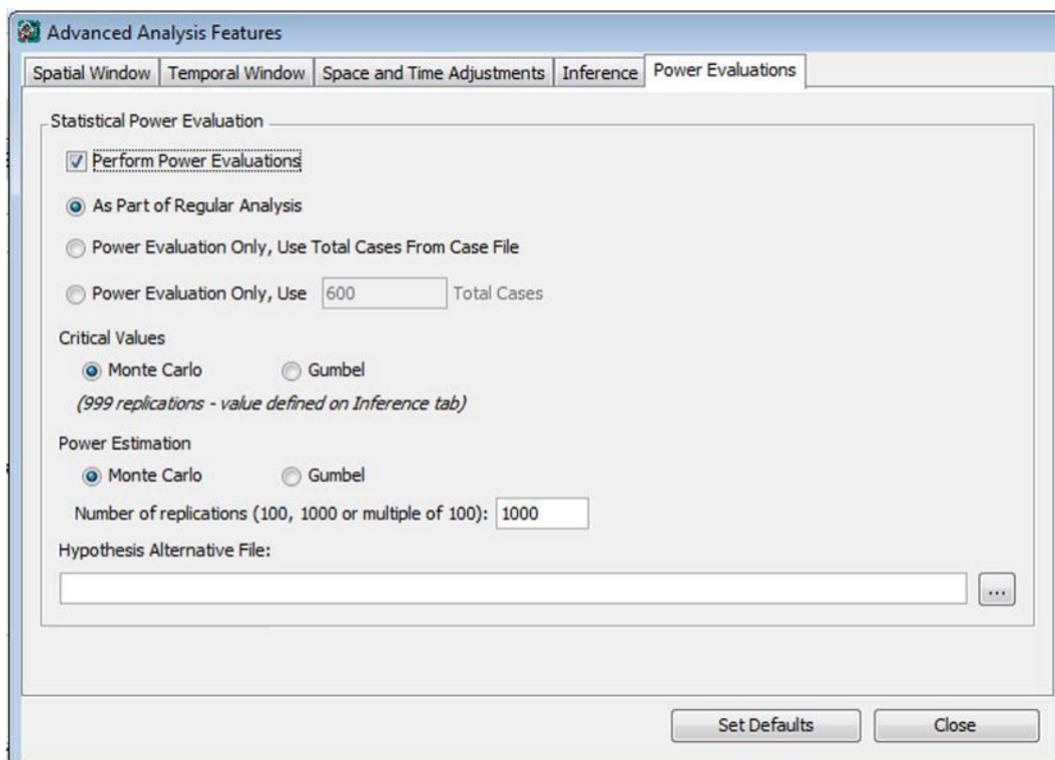
Uma vez que o cálculo do 'F' de Oliveira requer um segundo conjunto de simulações, ele aumenta o tempo de processamento necessário.

Os valores de 'F' de Oliveira resultantes são fornecidos no arquivo de resultados chamado 'Estimativas de Risco para Cada Localidade'. Este é normalmente um arquivo de resultados opcional, mas quando o 'F' de Oliveira é calculado, este arquivo será criado mesmo que não seja solicitado na Guia de Saída. O 'F' de Oliveira também é fornecido no arquivo de resultados de 'Informações das Localidades', mas este arquivo é gerado somente se ele for explicitamente solicitado na Guia de Saída.

Os valores de 'F' fornecidos pelo SaTScan podem ser utilizados para criar um mapa codificado por cores, onde as localidades em um cluster são representadas com cores diferentes, dependendo do seu valor de 'F'.

Tópicos Relacionados: *Guia de Análise de Fronteira, Critérios para Relatar Clusters Sobrepostos, Arquivo de Estimativas de Risco para Cada Localidade.*

Guia de Estimativa de Poder



Caixa de Diálogo da Guia de Estimativa de Poder

Assim como não existe nenhuma maneira analítica conhecida para se obter os valores de 'p' para as estatísticas de varredura no software SaTScan, não há nenhuma maneira analítica para se estimar o poder. Então, isso precisa ser feito por meio de simulações. Especificando um conjunto de localidades, tamanhos e riscos relativos de um cluster hipotético, o SaTScan irá estimar o poder estatístico para aquelas hipóteses alternativas pré-especificadas. Notar que várias hipóteses alternativas podem ser avaliadas ao mesmo tempo, o que é computacionalmente mais eficiente do que avaliar uma de cada vez. As hipóteses alternativas são definidas no 'Arquivo de Hipóteses Alternativas', com uma linha vazia entre as diferentes hipóteses alternativas de interesse. Observar que todas as localidades no mesmo cluster não precisam ter o mesmo risco relativo, e está tudo bem ainda que haja alguns riscos relativos maiores do que um e outros riscos relativos inferiores a um.

As estimativas de poder podem ser feitas tanto em conexão com uma análise efetiva do SaTScan de dados reais, ou como uma análise separada. Para esta última, não é necessário se especificar um arquivo de casos, sendo suficiente se especificar o número total de casos.

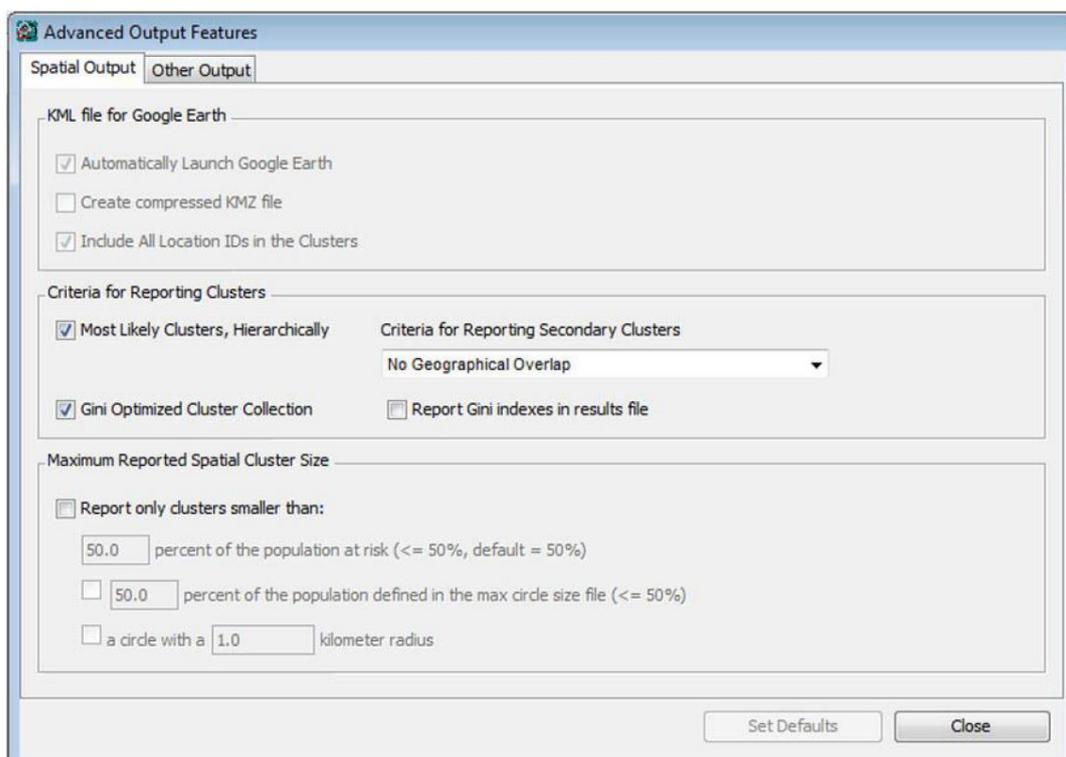
A precisão das estimativas de poder irá aumentar com mais simulações executadas. São recomendadas pelo menos 9.999 simulações sob a hipótese nula, e pelo menos 1.000 sob cada uma das hipóteses alternativas. Estimativas de poder podem ser realizadas utilizando tanto uma abordagem direta de Monte Carlo ou utilizando a aproximação de Gumbel. Os resultados devem ser similares.

O poder estatístico irá variar muito, dependendo do número total de casos no conjunto de dados, do tamanho da população do cluster e do risco relativo no cluster, com valores maiores resultando em poder maior.

Esta funcionalidade está disponível apenas para o modelo de probabilidade discreto de Poisson.

Tópicos Relacionados: *Modelo de Probabilidade de Poisson, Arquivo de Hipóteses Alternativas.*

Guia de Saída Espacial



Caixa de Diálogo da Guia de Saída Espacial

Esta guia é acessada clicando-se no botão 'Avançado' no canto inferior direito da Guia de Saída.

Tópicos Relacionados: *Recursos Avançados, Guia de Saída, Resultados das Análises, Critérios para Relatar Clusters Secundários, Notificação Apenas de Clusters Pequenos.*

Arquivo KML de Saída Geográfica (*.kml)

Com o arquivo de saída KML, os clusters detectados no SaTScan podem ser representados usando o Google Earth ou o Google Maps. O Google Earth é um software gratuito que pode ser baixado a partir de 'earth.google.com'. Se ele já estiver no computador, o SaTScan irá iniciá-lo automaticamente após a análise ser concluída, e mostrar os aglomerados detectados. O mapa do Google Earth pode ser então editado, adicionando-se ou subtraindo-se outros recursos ao mapa, tais como limites do país, rótulos dos nomes das cidades, ferrovias e vulcões.

O arquivo de saída KML só pode ser usado quando as coordenadas geográficas forem especificadas usando latitudes e longitudes. Embora os arquivos KML também possam ser lidos por algum outro software geográfico, os arquivos KML do SaTScan podem ou não funcionar bem para essas aplicações.

Tópicos Relacionados: *Guia de Saída, Arquivo Padrão de Resultados, Saída Geográfica Shapefile, Guia de Saída Espacial, Arquivo de Informações do Cluster, Arquivo de Coordenadas Geográficas, Coordenadas Lat/Long, Formato de Saída de Coluna.*

Cr terios para Relatar Clusters Secund rios

O SaTScan avalia uma quantidade enorme de diferentes  rculos/cilindros, a fim de encontrar o cluster mais prov vel. Para grandes conjuntos de dados, o n mero de potenciais clusters   de milh es para uma an lise puramente espacial, e na casa dos bilh es para uma an lise espa o-temporal. Todos estes outros clusters podem ser considerados aglomerados secund rios, seja com uma taxa alta ou baixa. Apresentar todos esses clusters secund rios   impratic vel e desnecess rio, visto que muitos deles ser o bastante semelhantes entre si. Por exemplo, adicionar uma localidade com uma popula o muito pequena ao cluster mais prov vel n o ir  diminuir muito a probabilidade, mesmo que essa localidade n o contenha casos adicionais. Tal cluster secund rio n o   interessante, ainda que pudesse ter a segunda maior probabilidade dentre todos os aglomerados avaliados.

Ao inv s de relatar informa es sobre todos os clusters avaliados, o SaTScan relata apenas um n mero limitado de aglomerados secund rios utilizando os crit rios especificados pelo usu rio. Um procedimento em tr s est gios   utilizado para selecionar os clusters secund rios a se relatar:

1. Para cada centroide circular, o SaTScan ir  considerar apenas o cluster com a probabilidade mais alta dentre aqueles que compartilham esse mesmo centr ide (ponto de grade).
2. O conjunto resultante de clusters ser  ordenado em ordem decrescente pelo valor dos seus logs de raz es de verossimilhan a, criando uma lista com o mesmo n mero de clusters, quantos forem os pontos de grade.
3. O cluster mais prov vel ser  sempre relatado. Clusters secund rios podem ser relatados com base em um ou ambos de dois crit rios. O primeiro   um crit rio puramente hier rquico, onde aglomerados secund rios com $p < 1,0$ s o relatados dependendo do seu grau de sobreposi o geogr fica com os clusters mais prov veis j  relatados. O segundo   um crit rio baseado no  ndice de Gini, que   uma medida de dispers o estat stica. Com este crit rio, o SaTScan seleciona o grupo de clusters n o sobrepostos que maximiza o  ndice de Gini, de modo que h  uma grande diferen a nas taxas entre as  reas de clusters e n o clusters³¹. Estes  ltimos s o chamados de ‘clusters de Gini’.

Op es de Relat rios de Cluster Hier rquico

Nenhuma Sobreposi o Geogr fica: Padr o. Aglomerados secund rios ser o apenas relatados se e somente se eles n o se sobrepuerem a um cluster relatado anteriormente, isto  , eles n o podem ter quaisquer IDs de localidade em comum. Portanto, aglomerados sobrepostos n o ser o relatados. Esta   a op o mais restritiva, apresentando o menor n mero de aglomerados.

Nenhum Centro de Cluster em Outros Clusters: Aglomerados secund rios s o relatados se eles n o estiverem centrados em um cluster relatado anteriormente, e n o contiverem o centro de um cluster relatado previamente. Embora dois clusters possam se sobrepor, n o haver  relato de cluster com o seu centr ide contido em outro cluster relatado.

Nenhum Centro de Cluster em Clusters Mais Prov veis: Aglomerados secund rios s o relatados se eles n o estiverem centrados em um cluster relatado anteriormente. Isto significa que n o haver  aglomerado relatado com o seu centro contido em um cluster mais prov vel relatado anteriormente.

Nenhum Centro de Cluster em Clusters Menos Prov veis: Aglomerados secund rios n o cont m o centro de um cluster relatado anteriormente. Isto significa que n o haver  relato de aglomerado com o seu centro contido em um cluster menos prov vel relatado subsequentemente.

Nenhum Par de Centros com Ambos em Outros Clusters: Aglomerados secund rios n o est o centrados em um cluster relatado anteriormente que cont m o centro de um aglomerado relatado previamente. Isto significa que n o haver  pares de clusters relatados, cada qual contendo o centro do outro.

Sem Restri es = Cluster mais Prov vel para Cada Ponto de Grade: A op o mais ampla   para todos os clusters constantes da lista, sem restri es. Esta op o informa o cluster mais prov vel para cada ponto

de grade, incluindo aglomerados com $p=1,0$. Isto significa que o número de clusters relatados é idêntico ao número de pontos de grade. **ATENÇÃO:** Esta opção pode criar arquivos de saída de tamanhos muito grandes.

Opção de Relatório de Cluster de Índice de Gini

Esta opção é válida apenas para análises puramente espaciais com os modelos de Poisson ou Bernoulli. Para criar a coleção de clusters baseada no índice de Gini, o SaTScan primeiro define um conjunto de limites superiores para o tamanho do cluster, com a coleção padrão sendo 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45 e 50 por cento da população. Para cada limite superior, o critério de coleção de cluster hierárquico sem nenhuma sobreposição geográfica é usado para definir um conjunto de aglomerados, conforme descrito acima. O índice de Gini é então calculado para esta coleção de clusters, e quando repetido para cada limite superior, resulta em doze diferentes índices de Gini. O SaTScan escolhe então a coleção que maximiza o índice de Gini. Estes são chamados 'clusters de Gini'. Como uma funcionalidade opcional, o SaTScan irá relatar os valores dos índices de Gini para cada um dos limites superiores utilizados.

Nota: É possível solicitar que ambos os clusters hierárquicos e de Gini sejam relatados, e esta é a configuração padrão. Isto significa que pode haver aglomerados sobrepostos, mesmo quando a opção de aglomerado hierárquico sem sobreposição é selecionada, uma vez que alguns dos clusters hierárquicos podem se sobrepor a alguns dos clusters de Gini, sem ser idênticos. Se nenhum for solicitado, o SaTScan irá relatar somente o cluster mais provável.

Nota: O critério para determinar sobreposição é baseado apenas na geografia, ignorando o tempo. Por isso, em uma análise espaço-temporal, um cluster secundário pode não ser relatado se ele estiver na mesma localidade que um cluster mais provável, mesmo que eles não se sobreponham no tempo.

Tópicos Relacionados: *Recursos Avançados, Guia de Inferência, Resultados das Análises, Tamanho Máximo do Cluster Espacial, Notificação Apenas de Clusters Pequenos.*

Tamanho Máximo do Cluster Espacial Relatado

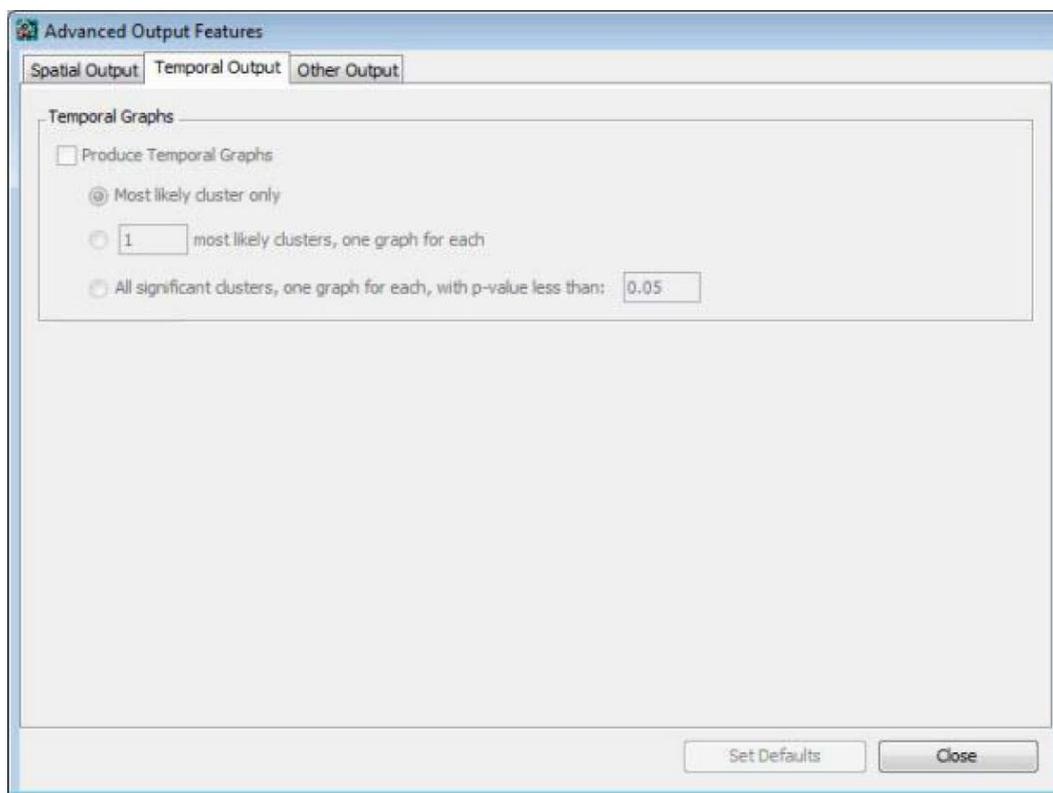
O tamanho máximo do cluster espacial é especificado na guia Análise > Avançado > Janela Espacial. O padrão é que o SaTScan irá relatar o cluster mais provável dentre estes, bem como quaisquer clusters secundários. Esta opção permite que se especifique um tamanho máximo diferente entre os clusters que são avaliados e aqueles que são relatados. Este último máximo deve ser menor, então.

É natural se perguntar por que alguém iria querer especificar um tamanho máximo de cluster relatado diferente, ao invés de simplesmente mudar o tamanho máximo dos clusters que estão sendo avaliados. A razão é a que segue. Quando o cluster mais provável é muito grande em tamanho, às vezes é de interesse saber se ele contém aglomerados menores que são estatisticamente significativos por sua própria força. Uma maneira de se encontrar tais aglomerados é 'brincar' com o parâmetro do tamanho máximo do círculo espacial na guia da Janela Espacial, mas isso leva à inferência estatística incorreta, já que o tamanho máximo dos círculos avaliados é então escolhido com base nos resultados da análise, levando a um viés de pré-seleção. Para se evitar este problema, esta opção permite se manter o tamanho máximo do círculo espacial original dos clusters que o SaTScan avalia e utiliza para calcular a inferência estatística, e ao mesmo tempo limitar o tamanho dos clusters que são relatados. Isto significa que os valores de 'p' para os clusters menores relatados são ajustados por todos os tamanhos de clusters, incluindo aqueles que não foram autorizados a serem relatados. Isto é, o SaTScan relata os aglomerados baseado nesse segundo tamanho máximo, mas ele ajusta para a múltipla testagem inerente ao conjunto maior de círculos definido pelo primeiro máximo.

A unidade por meio da qual se define o tamanho máximo do cluster relatado é a mesma unidade utilizada para definir o tamanho máximo do cluster para fins de inferência, conforme definido na Guia da Janela Espacial.

Tópicos Relacionados: *Recursos Avançados, Guia de Inferência, Resultados das Análises, Critérios para Relatar Clusters Secundários, Tamanho Máximo do Cluster Espacial, Log da Razão de Verossimilhança, Arquivo Padrão de Resultados.*

Guia de Saída Temporal



Caixa de Diálogo da Guia de Saída Temporal

Esta guia é acessada clicando-se no botão 'Avançado' no canto inferior direito da Guia de Saída.

Tópicos Relacionados: *Recursos Avançados, Guia de Saída, Resultados das Análises.*

Gráficos Temporais

Quando se conduz uma análise puramente temporal ou espaço-temporal, pode-se solicitar ao SaTScan para produzir gráficos temporais representando as contagens observadas e esperadas ao longo do tempo, tanto dentro como fora do cluster. Ele também irá mostrar a razão entre o observado e o esperado.

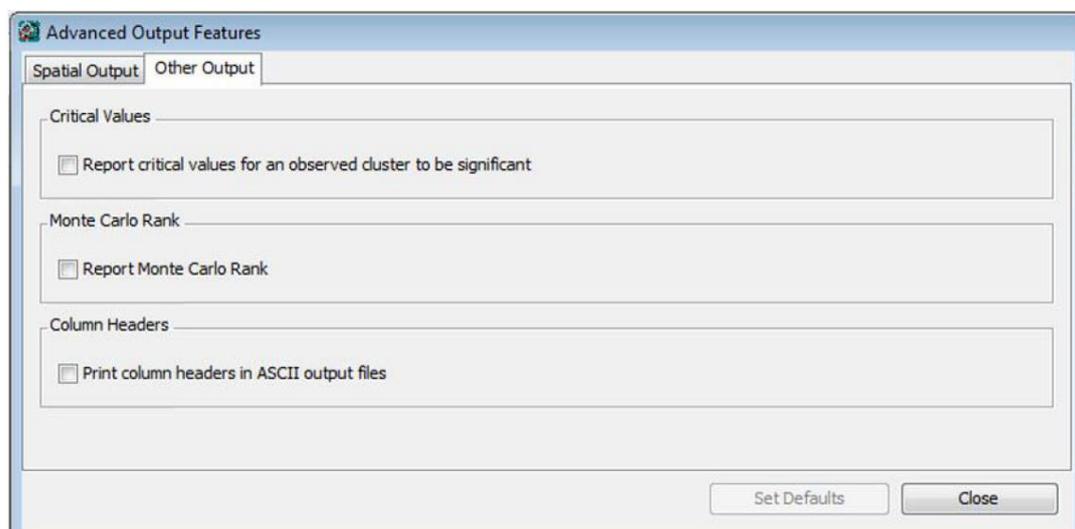
Esta opção está disponível para os modelos de probabilidade de Poisson, Bernoulli, permutação no espaço-tempo e exponencial, mas não para os modelos de probabilidade multinomial, ordinal ou normal.

É possível selecionar apenas um gráfico para o cluster mais provável, ou vários gráficos para um número fixo de clusters ou todos os clusters com um valor de 'p' inferior a algum valor especificado.

Os gráficos são gerados em arquivos HTML que podem ser abertos em qualquer navegador web. O nome e a localização do arquivo são listados na seção de parâmetros do arquivo padrão de resultados. Depois de abrir o arquivo HTML, é possível editá-lo, e também gerar os gráficos temporais nos formatos PNG, JPEG, PDF e SVG. Estes outros formatos não precisam ser pré-especificados.

Tópicos Relacionados: *Guia de Saída, Resultados das Análises, Arquivo HTML de Gráficos Temporais.*

Guia de Outras Saídas



Caixa de Diálogo da Guia de Outras Saídas

Valores Críticos

Ao selecionar esta opção, o SaTScan irá relatar os valores críticos necessários para que um cluster seja estatisticamente significativo nos níveis de alfa de 0,05 e 0,01. Os valores críticos são relatados no arquivo padrão de resultados.

Tópicos Relacionados: *Arquivo Padrão de Resultados.*

Classificação (Ranking) de Monte Carlo

Ao usar o teste de Monte Carlo padrão ou sequencial, esta opção irá, além de fornecer o valor de 'p', também relatar a classificação do log da razão de verossimilhança do cluster mais provável no conjunto de dados reais, entre todos os logs das máximas razões de verossimilhança dos conjuntos de dados aleatórios replicados sob a hipótese nula.

Tópicos Relacionados: *Valor de 'P', Arquivo Padrão de Resultados.*

Cabeçalhos de Colunas

Marcar esta caixa se houver interesse nos cabeçalhos das colunas nos arquivos de Outras Saídas.

Tópicos Relacionados: *Formato de Saída de Coluna, Arquivo de Informações da Coluna, Informações das Localidades.*

Título para o Arquivo de Resultados

Para manter o controle das diferentes análises do SaTScan, é possível se especificar um título para cada execução. Este título será informado na parte superior do arquivo padrão de resultados.

Tópicos Relacionados: *Arquivo Padrão de Resultados.*

Executando o SaTScan

Especificando as Opções de Análises e Dados

O programa SaTScan requer que se especifiquem os parâmetros que definem as opções de entrada, de análise e de saída para a análise que se deseja conduzir. Um diálogo com guias é fornecido para este propósito. Para acessar o diálogo da guia de parâmetros, pressionar o botão  ou selecionar o item 'Arquivo/Novo' do menu. Especificar os parâmetros para a sessão nas seguintes guias:

- Guia de Entrada
- Guia de Análise
- Guia de Saída

Verificar a seção de 'Recursos Básicos do SaTScan' para obter instruções de como preencher essas guias.

A maioria das análises pode ser executada utilizando apenas estas três guias. Para cada guia, há recursos adicionais que podem ser selecionados clicando-se primeiro no botão 'Avançado' no canto inferior direito da guia. Esses recursos adicionais podem ser úteis em circunstâncias especiais.

As opções disponíveis para alguns recursos podem depender do que foi selecionado em outros lugares. Por exemplo, se foi escolhida uma análise puramente espacial, o modelo de permutação no espaço-tempo não estará disponível, e vice-versa.

Tópicos Relacionados: *Recursos Básicos do SaTScan, Guia de Entrada, Guia de Análise, Guia de Saída, Recursos Avançados, Iniciando as Análises.*

Iniciando as Análises

Uma vez que os arquivos de entrada de dados tenham sido criados, e os parâmetros que definem as opções de entrada, de análise e de saída tenham sido especificados, selecionar o botão 'Executar'  para iniciar a análise e produzir o arquivo de resultados. Uma janela especial de status do trabalho aparecerá contendo o status, mensagens de alerta e/ou erro. Uma vez que a análise tenha sido concluída, o arquivo padrão de resultados aparecerá na janela de status do trabalho.

Múltiplas janelas da sessão de parâmetro podem ser abertas simultaneamente para a entrada de dados, e múltiplas análises podem ser executadas ao mesmo tempo. Ao se executar múltiplas análises simultaneamente, certificar-se de que os arquivos de saída tenham nomes diferentes.

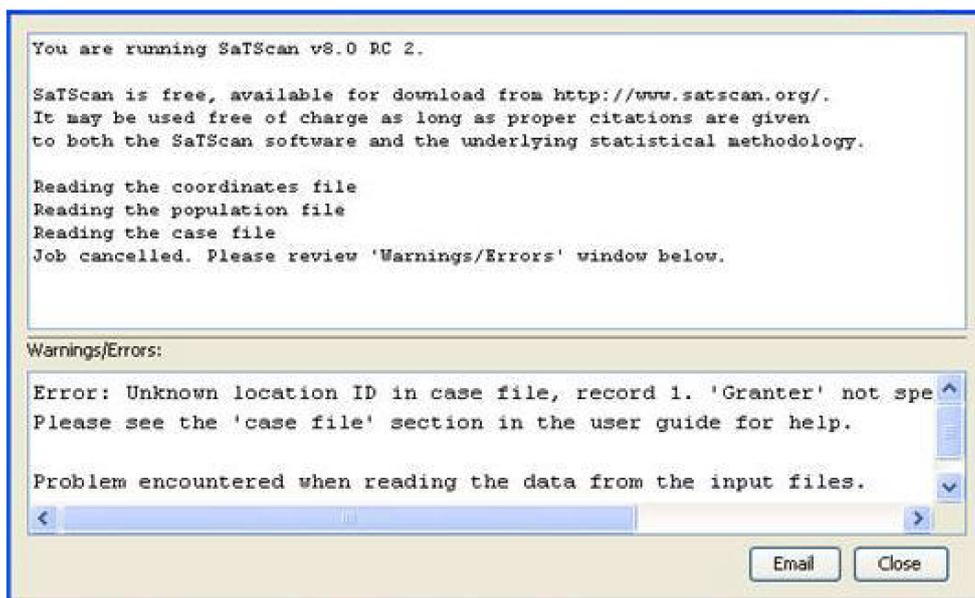
Tópicos Relacionados: *Dados de Entrada, Requisitos de Dados, Especificando as Opções de Análises e Dados, Mensagens de Status, Alertas e Erros, Tempo de Processamento, Modo de Lote.*

Mensagens de Status

Mensagens de status são exibidas conforme o programa executa a análise, enquanto os dados são lidos, e a cada passo da análise. Mensagens de status normais são exibidas na caixa superior da janela de status do trabalho. Mensagens de alertas e erros são exibidas na caixa inferior da janela de status do trabalho. Após a conclusão satisfatória dos cálculos, o arquivo padrão de resultados será mostrado na janela de status do trabalho.

Tópicos Relacionados: *Iniciando as Análises, Alertas e Erros.*

Alertas e Erros



Caixa de Diálogo de Mensagens de Status e Alertas/Erros do SaTScan

Mensagens de Alerta

O SaTScan pode produzir alertas conforme o trabalho esteja sendo executado. Se ocorrer um alerta, uma mensagem é exibida na caixa de Alertas/Erros na parte inferior da janela de status do trabalho. Um alerta não vai parar a execução da análise. Se ocorrer um alerta, recomenda-se rever a mensagem e acessar o sistema de ajuda se forem necessárias informações complementares.

Se não se deseja ver as mensagens de alerta, elas podem ser desligadas clicando-se em “Sessão > Opções de Execução > Não relatar mensagens de alerta”.

Mensagens de Erro

Se um problema sério ocorrer durante a execução, uma mensagem de erro será exibida na caixa de Alertas/Erros na parte inferior da janela de status do trabalho e o trabalho será encerrado. O usuário pode resolver a maioria dos erros revendo a mensagem e usando o sistema de ajuda.

Um dos erros mais comuns é que os arquivos de entrada não estão no formato exigido, ou que os conteúdos do arquivo são incompatíveis entre si. Quando isso ocorrer, uma mensagem de erro será mostrada especificando a natureza e a localização do problema. Essas mensagens de erro são concebidas para ajudar com a limpeza dos dados.

Se a mensagem de erro não puder ser resolvida, é possível pressionar o botão de e-mail na janela de status do trabalho. Isso irá gerar uma mensagem de e-mail automática para o suporte técnico do SaTScan. Os conteúdos da caixa de “Alertas/Erros” serão automaticamente colocados na mensagem de e-mail. Tudo o que o usuário precisa fazer é pressionar a tecla ‘Enviar’ do seu e-mail. Os usuários podem também imprimir os conteúdos da caixa de Alertas/Erros e até mesmo selecionar, copiar (ctrl c) e colar (ctrl v) os conteúdos, se necessário.

Tópicos Relacionados: *Dados de Entrada, Requisitos de Dados, Suporte do SaTScan.*

Salvando os Parâmetros das Análises

Os parâmetros das análises, especificados na caixa de diálogo da guia de Parâmetros, podem ser salvos e reutilizados em análises futuras. É recomendável salvar os parâmetros com uma extensão de arquivo “.prm”. O arquivo de parâmetro é armazenado em um formato de arquivo ASCII de texto.

Para salvar os parâmetros da análise

1. Se os parâmetros não tiverem sido salvos previamente, selecionar ‘Salvar Como’ no menu Arquivo. Uma caixa de diálogo ‘Salvar Arquivo de Parâmetro Como’ será aberta.
2. Selecionar um local de diretório a partir do menu drop-down (suspensão) ‘Salvar Em’ na parte superior da caixa de diálogo.
3. Digitar um nome para o arquivo de parâmetro na caixa de texto ‘Nome do Arquivo’. Recomenda-se que a seleção ‘Salvar Como Tipo’ permaneça como Arquivos de Parâmetros (*.prm).
4. Pressionar o botão ‘Salvar’.

Uma vez que o arquivo de parâmetro seja inicialmente salvo, salvar as alterações para o arquivo selecionando-se ‘Salvar’ no menu ‘Arquivo’. O arquivo será salvo sem abrir a caixa de diálogo ‘Salvar Arquivo de Parâmetro Como’.

Para abrir um arquivo de parâmetros salvo

1. Selecionar ‘Abrir’ no menu Arquivo ou clicar no botão  na barra de ferramentas. Uma caixa de diálogo da ‘Seleção de Arquivo de Parâmetro’ será aberta.
2. Localizar o arquivo desejado usando ‘Procurar’ no menu drop-down (suspensão).
3. Uma vez que o arquivo seja localizado, realçar o nome do arquivo clicando sobre ele.
4. Pressionar o botão ‘Abrir’.

Uma caixa de diálogo da guia de Parâmetro será aberta, contendo as configurações de parâmetros salvos. A localização e o nome do arquivo de parâmetro estão listados na barra de título desta caixa de diálogo.

Tópicos Relacionados: *Especificando as Opções de Análises e Dados, Recursos Básicos do SaTScan, Recursos Avançados, Modo de Lote.*

Processadores Paralelos

Se houver processadores paralelos no computador, o SaTScan pode tirar proveito deles executando diferentes simulações de Monte Carlo usando diferentes processadores, aumentando, assim, a velocidade dos cálculos. O padrão é que o SaTScan utilizará todos os processadores que o computador tiver. Se o usuário desejar restringir o número, é possível fazê-lo clicando em “Sessão > Opções de Execução”, e selecionar o número máximo de processadores cujo uso é permitido ao SaTScan.

Modo de Lote

O SaTScan é mais facilmente executado clicando-se no botão ‘Executar’  no topo da janela do SaTScan, depois de preencher os vários campos de parâmetros na interface do Windows.

Uma abordagem alternativa é pular a interface do Windows e iniciar o mecanismo de cálculo do SaTScan diretamente por meio de duas opções:

1. Arrastando um arquivo de parâmetro para o executável 'SaTScanBatch.exe'.
2. Escrevendo 'SaTScanBatch.exe *.prm' em um arquivo de lote ou na linha de comandos, onde *.prm é o nome do arquivo de parâmetro.

Usando-se a versão de modo de lote, é possível se descrever um software especial que incorpora o mecanismo de cálculo do SaTScan com outras aplicações, tal como um sistema de vigilância diário automatizado para a detecção precoce de surtos de doenças. Para usar o SaTScan desta forma, requer-se uma quantidade razoável de conhecimentos de informática e sofisticação.

Ao se executar o SaTScan no modo de lote, o arquivo de parâmetro pode ainda ser alterado usando a interface do Windows do SaTScan. Também é possível se alterar os parâmetros manualmente usando qualquer editor de texto, ou automaticamente, usando algum outro produto de software. Se apenas poucos parâmetros devem mudar em comparação ao que já está em um arquivo de parâmetro existente, nomear esse arquivo de parâmetro na linha de comando, juntamente com as instruções sobre quais parâmetros devem mudar. Os valores dos parâmetros da linha de comando irão, então, substituir os valores dos parâmetros especificados no arquivo de parâmetros.

Quando a versão em modo de lote do SaTScan é executada, o arquivo padrão de resultados não aparece automaticamente na tela, mas precisa ser aberto manualmente usando qualquer editor de texto disponível, tal como o Bloco de Notas.

Oportunidade: Existem algumas opções de parâmetros que não são permitidos quando o SaTScan é executado sob a interface do Windows, mas que podem ser definidos quando executado no modo de lote. Alguns exemplos são o número de replicações de Monte Carlo, a coleção de elipses utilizadas para a estatística de varredura elíptica, um número ilimitado de múltiplos conjuntos de dados, o uso dos valores de 'p' aproximados de Gumbel para qualquer modelo de probabilidade e a capacidade de gravar todos os dados simulados em um arquivo (com os IDs das localidades listados em ordem alfabética). As opções de parâmetros não permitidas pela interface do Windows não foram todas exaustivamente testadas, então, há algum risco envolvido ao se executar tais análises.

Tópicos Relacionados: *Iniciando as Análises, Recursos Básicos do SaTScan, Recursos Avançados, Salvando os Parâmetros da Análise.*

Tempo de Computação/Processamento

As estatísticas de varredura espaciais e espaço-temporais são volumosas em termos de cálculos computacionais. O tempo de processamento depende de uma ampla gama de variáveis, e dependendo do conjunto de dados e das opções de análises escolhidas, pode variar de alguns segundos até vários dias ou semanas. Os modelos multinomial, ordinal e normal são, em geral, muito mais intensivos ao computador do que as outras estatísticas de varredura discretas. Fora isso, as três principais coisas que aumentam o tempo de computação são o número de localidades nas coordenadas e nos arquivos de grade especiais, o número de intervalos de tempo (para as análises espaço-temporais) e o número de conjuntos de dados utilizados.

Único Conjunto de Dados

Para um único conjunto de dados, o tempo de processamento para uma das estatísticas de varredura discretas é aproximadamente da ordem de:

$$\text{LGMT}^k_{mS} / P$$

onde:

L = número de localidades geográficas de dados no arquivo de coordenadas (L=1 para análises puramente temporais)

G = número de coordenadas geográficas no arquivo de grade especial. Se não existe tal arquivo, G=L.

M = tamanho máximo do cluster geográfico, como uma proporção da população ($0 < M = \frac{1}{2}$, M = 1 para uma análise puramente temporal)

T = número de intervalos de tempo em que o dado temporal é agregado (T=1 para uma análise puramente espacial)

m = tamanho máximo do cluster temporal, como uma proporção do período de estudo ($0 < m = 0,9$, m=1 para análise puramente espacial)

S = número de simulações de Monte Carlo

P = número de processadores disponíveis no computador para o uso do SaTScan

k = 1 para análises puramente espacial, temporal prospectiva e espaço-temporal prospectiva sem ajustes para análises prévias

k = 2 para análises temporais retrospectivas e espaço-temporais retrospectivas

A unidade da fórmula acima depende do modelo de probabilidade utilizado e da velocidade do computador. Quando o número total de casos é muito grande em comparação com o número de localidades e intervalos de tempo, o tempo de computação para os modelos discreto de Poisson, Bernoulli e exponencial é, por sua vez, da ordem de:

$$CS / P$$

onde:

C = o número total de casos

Múltiplos Conjuntos de Dados

Uma análise utilizando múltiplos conjuntos de dados é consideravelmente mais intensiva para o computador do que a análise de um único conjunto de dados. Para os modelos discreto de Poisson, Bernoulli e exponencial, o tempo de computação para dois conjuntos de dados é muito mais do que o dobro do tempo para um único conjunto de dados. O tempo de computação para D>2 conjuntos de dados é aproximadamente D/2 vezes maior do que o tempo de computação para dois conjuntos de dados.

Tópicos Relacionados: *Arquivo de Coordenadas, Arquivo de Grade, Guia da Janela Espacial, Guia da Janela Temporal, Replicações de Monte Carlo, Término Antecipado das Simulações, Guia de Múltiplos Conjuntos de Dados.*

Requisitos de Memória

O SaTScan usa a alocação dinâmica de memória. Dependendo da natureza dos dados de entrada, o SaTScan irá escolher automaticamente um de dois esquemas de alocação de memória: o padrão e uma outra especial para conjuntos de dados com muitas localidades espaciais, mas poucos intervalos de tempo e poucas simulações.

Alocação de Memória Padrão

Usando o esquema padrão de alocação de memória, a quantidade de bytes de memória necessários para grandes conjuntos de dados é de aproximadamente:

Poisson Discreto: $ALGM + (12 + 4P) LTD + 8CRP$

Bernoulli: $ALGM + (16 + 4P) LTD + 8CRP$

Permutação no Espaço-Tempo: $ALGM + (12 + 4P) LTD + 12CP + 8CRP$

Ordinal, Multinomial: $ALGM + (4Y + 4YP + 4P) LTD$

Exponencial: $ALGM + (16 + 12P) LTD + 40IP + 8CRP$

Normal: $ALGM + (20 + 16P) LTD + 32IP$

Normal, com pesos: $ALGM + (20 + 16P) LTD + 48IP$

Poisson Contínuo: $G + 16C + 24CP$

onde

L = o número de IDs de localidades no arquivo de coordenadas (L=1 para uma análise puramente temporal)

A = 2 se $L < 65.536$ e A = 4 se $L > 65.536$

G = o número de coordenadas no arquivo de grade (G=L se nenhum arquivo de grade é especificado)

M = o tamanho máximo do cluster geográfico, como uma proporção da população ($0 < M = \frac{1}{2}$, M=1 para uma análise puramente temporal)

T = número de intervalos de tempo em que o dado temporal é agregado (T=1 para uma análise puramente espacial)

Y = número de categorias no modelo multinomial ou ordinal

C = o número total de casos nos modelos de Poisson, Bernoulli, permutação no espaço-tempo e exponencial

I = o número de observações individuais nos modelos exponencial e normal

R = 1 quando se realiza a varredura somente para as taxas altas ou somente para as baixas, R = 2 durante uma varredura para ambas as taxas alta ou baixa

D = número de conjuntos de dados

P = número de processadores disponíveis no computador para uso do SaTScan

Para as análises puramente espaciais e a maioria das análises espaço-temporais, T é muito menor do que G, assim como D e P, por isso é a expressão SLGM à esquerda do primeiro sinal de mais acima que é crítica em termos de requisitos de memória para as estatísticas de varredura discretas. A Tabela 2 apresenta as estimativas dos requisitos de memória quando $G=L$, $M=0,5$ e $T=1$.

G=L	Memory Needed
3,500	32 Mb
6,500	64 Mb
10,000	128 Mb
15,000	256 Mb
22,000	512 Mb
32,000	1 Gb
44,000	2 Gb
63,000	4 Gb
89,000	16 Gb
126,000	32 Gb
178,000	64 Gb
250,000	128 Gb

Tabela 2: Requisitos de memória aproximados para uma análise puramente espacial quando o tamanho máximo do cluster geográfico é de 50% da população.

Alocação de Memória Especial

Quando o número de localidades é muito grande, enquanto que o número de casos, intervalos de tempo e simulações não o são, o SaTScan às vezes usa um esquema alternativo de alocação de memória para reduzir o requisito total de memória. Esta seleção é feita automaticamente. A quantidade de memória necessária para grandes conjuntos de dados é então, aproximadamente, o número de bytes a seguir:

Poisson Discreto: $(4S + 12 + 4P) LTD + 8CRS$

Bernoulli: $(4S + 16 + 4P) LTD + 8CRS$

Permutação no Espaço-Tempo: $(4S + 12 + 4P) LTD + 12CP + 8CRS$

Ordinal, Multinomial: $(4YS + 4Y + 4YP + 4P) LTD$

Exponencial: $(12S + 16 + 12P) LTD + 40IP + 8CRS$

Normal: $(16S + 20 + 16P) LTD + 32IP$

Normal, com pesos: $(16S + 20 + 16P) LTD + 48IP$

Poisson Contínuo: não usado

onde S é o número de simulações de Monte Carlo e as outras variáveis são definidas conforme acima.

Memória Insuficiente

Se não houver memória suficiente disponível no computador para executar a análise utilizando um esquema de alocação de memória, existem várias opções disponíveis para se trabalhar em torno da limitação:

- Fechar outras aplicações.
- Agregar os dados em menor número de localidades de dados (reduzir L).
- Diminuir o número de centroides circulares no arquivo especial de grade (reduzir G).
- Reduzir o limite máximo do tamanho do círculo (reduzir mg).
- Executar o programa em um computador com mais memória.

É altamente desejável que haja memória RAM suficiente para cobrir todas as necessidades de memória, visto que o SaTScan executa bem mais lentamente quando o arquivo de permuta é usado, então essas técnicas também podem ser utilizadas para evitar o arquivo de permuta. Nem todas as opções acima irão funcionar para todos os conjuntos de dados. Por favor, notar que as seguintes opções do SaTScan não influenciam o consumo de memória:

- A duração do período de estudo.
- O tamanho máximo do cluster temporal.
- Tipo de clusters espaço-temporais para incluir na análise.

Nota: O sistema operacional Windows de 32-bits pode alocar um máximo de 2 GBytes de memória para uma única aplicação, e este é, portanto, o limite superior de memória para a versão Windows de 32 bits do SaTScan. A versão Linux do SaTScan pode ser usada para analisar conjuntos de dados maiores.

Tópicos Relacionados: *Arquivo de Coordenadas, Arquivo de Grade, Estatísticas de Varredura Espacial Temporal e Espaço-Temporal, Guia da Janela Espacial, Guia da Janela Temporal, Replicações de Monte Carlo, Guia de Múltiplos Conjuntos de Dados, Alertas e Erros.*

Resultados das Análises

Como saída, o SaTScan cria um arquivo padrão de resultados baseado em texto no formato ASCII, arquivos de saída opcionais de formato geográfico em KML ou shapefile, um arquivo opcional de gráfico temporal no formato HTML, e cinco arquivos de saída opcionais em formato de coluna, seja no formato ASCII ou dBase. Alguns dos arquivos opcionais são úteis quando exportados da saída do SaTScan para outros softwares, tais como uma planilha ou um sistema de informação geográfica.

Todos os arquivos de resultados estarão no mesmo diretório e terão o mesmo nome que o arquivo padrão de saída especificado na Guia de Saída, exceto por sua extensão.

Tópicos Relacionados: *Guia de Saída, Guia de Saída Espacial, Arquivo Padrão de Resultados, Arquivo de Informações do Cluster, Arquivo de Informações das Localidades, Estimativas de Risco para Cada Localidade, Log das Razões de Verossimilhança Simulados, Arquivo de Histórico das Análises.*

Arquivo Padrão de Resultados Baseado em Texto (*.out.*)

O arquivo padrão de resultados é automaticamente mostrado após os cálculos serem concluídos. É bastante autoexplicativo, mas para sua interpretação adequada, recomenda-se ler a seção sobre metodologia estatística ou, melhor ainda, um dos artigos metodológicos listados na bibliografia.

SÍNTESE DOS DADOS: Usar para verificar se os arquivos de entrada de dados contêm o número correto de casos, localidades, etc.

População total (modelo discreto de Poisson): Esta é a população média durante o período de estudo.

Taxa anual por 100.000 (modelo discreto de Poisson): Esta é calculada levando em conta os anos bissextos e baseia-se na duração média de um ano de 365,2425. Se calculada ignorando-se manualmente os anos bissextos, os números serão ligeiramente diferentes, mas não muito.

Variância (modelo normal): Esta é a variância para todas as observações nos dados assumindo uma média comum.

CLUSTER MAIS PROVÁVEL: Informações resumidas acerca do cluster mais provável, isto é, o cluster que é menos provável de ser devido ao acaso.

Raio: Quando latitude e longitude são utilizadas, o raio do círculo é dado em quilômetros. Quando são utilizadas coordenadas Cartesianas regulares, o raio do círculo é dado nas mesmas unidades utilizadas no arquivo de coordenadas.

População: Esta é a população média na área geográfica do cluster. A média é tomada ao longo de todo o período do estudo, mesmo quando se trata de um cluster espaço-temporal, cuja duração temporal é apenas uma parte do período de estudo.

Risco Relativo: Este é o risco estimado dentro do cluster dividido pelo risco estimado fora do cluster. Ele é calculado como o observado dividido pelo esperado dentro do cluster, dividido pelo observado dividido pelo esperado fora do cluster. Em notação matemática, é:

$$RR = \frac{c / E[c]}{(C - c) / (E[C] - E[c])} = \frac{c / E[c]}{(C - c) / (C - E[c])}$$

onde 'c' é o número de casos observados dentro do cluster e 'C' é o número total de casos no conjunto de dados. Notar que, uma vez que a análise é condicionada ao número total de casos observados, $E[C]=C$.

Observado / Esperado: Este é o número de casos observados dentro do cluster dividido pelo número de casos esperados dentro do cluster quando a hipótese nula é verdadeira, ou seja, quando o risco é o mesmo dentro e fora do cluster. Isto significa que é o risco estimado dentro do cluster dividido pelo risco estimado para a região do estudo como um todo. É calculado como: $c/E[c]$.

Para o modelo contínuo de Poisson, a contagem esperada é um limite superior quando a janela de varredura cruza a fronteira da região espacial do estudo. Isso significa que o Obs/Exp é um limite inferior.

Variância (modelo normal): Esta é a variância comum estimada para todas as observações, levando em conta as diferentes médias estimadas dentro e fora do cluster. A variância ponderada é ajustada para os pesos quando fornecidos pelo usuário.

Tendência temporal (variação espacial nas tendências temporais): Fornece as tendências temporais estimadas dentro e fora dos aglomerados detectados na escala log linear onde o aumento percentual ou a redução percentual é constante ao longo do tempo.

Valor de 'P': Os valores de 'p' são ajustados para os testes múltiplos decorrentes da multiplicidade de círculos/cilindros correspondentes a diferentes localidades e tamanhos espaciais e/ou temporais de potenciais aglomerados avaliados. Isto significa que, sob a hipótese nula de aleatoriedade espacial completa, há uma chance de 5% de que o valor de 'p' para o cluster mais provável seja menor do que 0,05, e uma chance de 95% de que seja maior. Sob a hipótese nula, haverá sempre alguma área com uma taxa maior do que a esperada somente pelo acaso. Assim, ainda que o cluster mais provável tenha sempre uma taxa em excesso ao se varrer para áreas com taxas altas, o valor de 'p' pode realmente ser muito próximo ou igual a um.

Intervalo de Repetição: Para as análises prospectivas, o intervalo de repetição²⁰ (ou, taxa de ocorrência nula) é mostrado como uma alternativa ao valor de 'p'. A medida reflete o quão frequentemente um cluster da probabilidade observada ou maior será observado ao acaso, assumindo que as análises são repetidas em uma base regular, com uma periodicidade igual à duração do intervalo de tempo especificado. Por exemplo, se o valor de 'p' observado é utilizado como o ponto de corte para um sinal e, se o intervalo de repetição é de uma vez a cada 14 meses, o número esperado de sinais falsos em qualquer período de 14 meses é um.

Se nenhum ajuste for feito para análises prévias, então o intervalo de repetição é de uma vez em D/p dias, onde 'D' é o número de dias em cada intervalo de tempo. Se os ajustes são feitos para A-1 análises anteriores, então o intervalo de repetição é de uma vez a cada $D / [1 - (1-p)^{1/A}]$ dias.

CLUSTERS SECUNDÁRIOS: Informações resumidas sobre outros clusters detectados nos dados. A informação fornecida é a mesma que para o cluster mais provável. Somente clusters com $p < 1$ são exibidos.

Valores de 'p' listados para os aglomerados secundários são calculados da mesma maneira que para o cluster mais provável, comparando-se o log da razão de verossimilhança dos aglomerados secundários no conjunto de dados reais com o log das razões de verossimilhança do cluster mais provável no conjunto de dados simulados. Isto significa que se um cluster secundário é significativo, ele pode rejeitar a hipótese nula por sua própria força, sem a ajuda de quaisquer outros clusters. Isto também significa que estes valores de 'p' são conservadores¹.

CONFIGURAÇÕES DE PARÂMETROS: Um lembrete das configurações dos parâmetros utilizadas para a análise.

Arquivos adicionais de resultados: O nome e a localização dos arquivos adicionais de resultados são fornecidos, quando aplicável.

Tópicos Relacionados: *Guia de Saída, Guia de Saída Espacial, Arquivo de Informações do Cluster, Arquivo de Informações das Localidades, Estimativas de Risco para Cada Localidade, Log das Razões de Verossimilhança Simulados, Coordenadas Cartesianas, Formato de Saída de Coluna.*

Arquivo KML de Saída Geográfica (*.kml)

Para as análises espacial e espaço-temporal, o SaTScan cria um arquivo KML que pode mostrar os clusters detectados no Google Earth e outros softwares geográficos. Além da localização e do tamanho do cluster, o arquivo KML também contém informações básicas sobre cada cluster, tais como o número de casos observados, o risco relativo e o valor de 'p'. Como padrão, o SaTScan iniciará automaticamente o Google Earth quando a análise for concluída. Alternativamente, pode-se simplesmente clicar sobre o arquivo KML quando se quiser mostrar os resultados. Há também muitos outros produtos de SIG que podem ler os arquivos KML.

Tópicos Relacionados: *Guia de Saída, Arquivo Padrão de Resultados, Arquivo KML de Saída, Arquivo Shapefile de Saída Geográfica, Guia de Saída Espacial, Arquivo de Informações do Cluster, Arquivo de Coordenadas Geográficas, Coordenadas Lat/Long, Formato de Saída de Coluna.*

Arquivo Shapefile de Saída Geográfica (*.shp e *.shx)

Com a saída shapefile, as localidades do cluster detectadas pelo SaTScan podem ser diretamente exportadas e mostradas em uma ampla variedade de sistemas de informação geográfica, incluindo tanto softwares livres como Quantum GIS (www.qgis.org), TerraView (www.dpi.inpe.br/terraview_eng/) e R (www.r-project.org), bem como produtos comerciais, tais como ArcGIS e SAS.

O shapefile inclui informações sobre os centroides do cluster e os círculos do cluster, bem como meta informações sobre cada cluster, tais como o seu número de casos observados e esperados. Se o arquivo de saída de 'Informações das Localidades', também for solicitado, então o shapefile irá incluir pontos que representam as coordenadas para cada localidade no cluster.

A saída shapefile só pode ser usada quando as coordenadas geográficas forem especificadas usando latitudes e longitudes.

Tópicos Relacionados: *Guia de Saída, Arquivo Padrão de Resultados, Arquivo KML de Saída Geográfica, Guia de Saída Espacial, Arquivo de Informações do Cluster, Arquivo de Coordenadas Geográficas, Coordenadas Lat/Long, Arquivo de Saída de Informações das Localidades, Formato de Saída de Coluna.*

Arquivo HTML de Gráficos Temporais (*.temporal.html)

Se solicitado na Guia de Saída Temporal avançada, o SaTScan irá gerar um arquivo HTML que mostra um gráfico representando as contagens observadas e esperadas ao longo do tempo, tanto dentro como fora do cluster. Ele também irá mostrar a razão entre o observado e o esperado. Um gráfico separado é gerado para cada cluster.

Os gráficos são gerados em arquivos HTML, que podem ser abertos em qualquer navegador da web. O nome e a localização do arquivo são, convencionalmente, os mesmos que para o arquivo padrão de resultados, mas com a extensão *.temporal.html.

Ao se abrir o arquivo HTML, é possível editá-lo em uma variedade de formas.

- Clicando-se nos descritores abaixo do gráfico, é possível se remover partes selecionadas do gráfico, bem como adicioná-las de volta.
- Arrastando-se o mouse sobre o gráfico, pode se observar as contagens numéricas exatas observadas e esperadas para cada intervalo de tempo.

- Clicando-se em “Mostrar Opções do Gráfico” na parte inferior esquerda do gráfico, é possível adicionar um título, alternar entre mostrar as contagens usando histogramas e linhas, e deixar o cluster mais visível por meio da adição de uma banda do cluster.

Para imprimir o gráfico, clicar nas três linhas horizontais no canto superior direito do gráfico. Lá também é possível salvar o gráfico em outros quatro formatos de arquivo: PNG, JPEG, PDF e SVG. Isto é feito separadamente para cada cluster.

Tópicos Relacionados: *Guia de Saída Temporal, Resultados das Análises.*

Arquivo de Informações do Cluster (*.col.*)

No arquivo de informações do cluster, cada cluster está em uma linha, com informações diferentes sobre o cluster em diferentes colunas. Para cada cluster há informações sobre a localidade e o tamanho do cluster, o seu log da razão de verossimilhança e o valor de ‘p’. Exceto para os modelos multinomial e ordinal, e quando múltiplos conjuntos de dados são usados, há também informações sobre o número de casos observados e esperados, observados/esperados e o risco relativo. Para os modelos multinomial e ordinal, e para múltiplos conjuntos de dados, estes números dependem do conjunto de dados e/ou categorias, e as informações são, por sua vez, fornecidas no Arquivo de Informações do Cluster Estratificado.

As colunas exatas incluídas no arquivo dependem da análise escolhida, conforme mostrado na Tabela 3, mas isso é facilmente verificado pela comparação com o arquivo padrão de resultados. O arquivo terá o mesmo nome que o arquivo padrão de resultados, mas com as extensões *.col.txt e *.col.dbf, respectivamente, e estará localizado no mesmo diretório.

Nota: Enquanto o arquivo padrão de resultados exibe apenas clusters com $p < 1$, esse arquivo irá também exibir clusters com $p = 1$. Notar que estes valores de ‘p’ são ajustados para testes múltiplos, por isso, mesmo se $p = 1$, o cluster ainda pode ter um risco relativo bastante elevado.

Tópicos Relacionados: *Arquivo de Informações de Casos do Cluster, Arquivo de Informações das Localidades, Guia de Saída, Resultados das Análises, Arquivo Padrão de Resultados.*

Output Variable	dBase Name	Discrete Poisson, Bernoulli, Circular, Lat/Long, One Data Set	Discrete Poisson, Bernoulli, Circular, Cartesian, One Data Set	Discrete Poisson, Bernoulli, Circular, Cartesian 5 Dimensions, One Data Set	Space-Time Permutation, Circular, Lat/Long, One Data Set	Exponential, Circular, Lat/Long, One Data Set	Discrete Poisson, Bernoulli, Elliptic, Cartesian, One Data Set	Ordinal, Multinomial, Circular, Lat/Long, One Data Set	Discrete Poisson, Bernoulli, Circular, Lat/Long, Multiple Data Sets	Normal, Circular, Lat/Long, One Data Set	Normal, Weighted, Circular, Lat/Long, One Data Set	Continuous Poisson
Cluster Number	CLUSTER	1	1	1	1	1	1	1	1	1	1	1
Central Location ID	LOCATION_ID	2	2	2	2	2	2	2	2	2	2	2
Latitude	LATITUDE	3	-	-	3	3	-	3	3	3	3	-
Longitude	LONGITUDE	4	-	-	4	4	-	4	4	4	4	-
X-coordinate	X	-	3	3	-	-	3	-	-	-	-	3
Y-coordinate	Y	-	4	4	-	-	4	-	-	-	-	4
Z1-coordinate	Z1	-	-	5	-	-	-	-	-	-	-	-
Zn-coordinate	Zn	-	-	6,7	-	-	-	-	-	-	-	-
Circle Radius	RADIUS	5	5	8	5	5	-	5	5	5	5	5
Ellipse, Length of Minor Axis	E_MINOR	-	-	-	-	-	5	-	-	-	-	-
Ellipse, Length of Major Axis	E_MAJOR	-	-	-	-	-	6	-	-	-	-	-
Ellipse Angle	E_ANGLE	-	-	-	-	-	7	-	-	-	-	-
Ellipse Shape	E_SHAPE	-	-	-	-	-	8	-	-	-	-	-
Cluster Start Date	START_DATE	6	6	9	6	6	9	6	6	6	6	-
Cluster End Date	END_DATE	7	7	10	7	7	10	7	7	7	7	-
# Location IDs	NUMBER_LOC	8	8	11	8	8	11	8	8	8	8	6
Log Likelihood Ratio	LLR	9	9	12	-	9	12	9	9	9	9	7
Test Statistic	TEST_STAT	-	-	-	9	-	13	-	-	-	-	-
P-Value of Cluster	P_VALUE	10	10	13	10	10	14	10	10	10	10	8
Observed Cases	OBSERVED	11	11	14	11	11	15	-	-	11	11	9
Expected Cases	EXPECTED	12	12	15	12	12	16	-	-	-	-	10
Observed / Expected	ODE	13	13	16	13	13	17	-	-	-	-	11
Relative Risk	REL_RISK	14	14	17	-	-	18	-	-	-	-	-
Total Weights	WEIGHT_IN	-	-	-	-	-	-	-	-	-	-	12
Mean Inside	MEAN_IN	-	-	-	-	-	-	-	-	12	13	-
Mean Outside	MEAN_OUT	-	-	-	-	-	-	-	-	13	14	-
Variance	VARIANCE	-	-	-	-	-	-	-	-	14	15	-
Standard deviation	STD	-	-	-	-	-	-	-	-	15	16	-
Weighted Mean Inside	W_MEAN_IN	-	-	-	-	-	-	-	-	-	17	-
Weighted Mean Outside	W_MEAN_OUT	-	-	-	-	-	-	-	-	-	18	-
Weighted Variance	W_VARIANCE	-	-	-	-	-	-	-	-	-	19	-
Weighted Standard deviation	W_STD	-	-	-	-	-	-	-	-	-	20	-

Tabela 3: Conteúdo do arquivo de saída de informações do cluster, com nomes dBase das variáveis e exemplos de ordenação das colunas para alguns tipos diferentes de análises.

Arquivo de Informações do Cluster Estratificado (*.sci.*)

No arquivo de informações do cluster estratificado, há uma linha para cada categoria ordinal/multinomial, em cada conjunto de dados, para cada cluster. Para cada combinação de cluster/categoria/conjunto de dados, há uma coluna de cada para o número de casos observados, o número de casos esperados, observado dividido por esperado e, por vezes, o risco relativo. Se nem o modelo multinomial, modelo ordinal nem múltiplos conjuntos de dados forem usados, então há apenas uma linha para cada cluster, e não há nenhuma informação neste arquivo que não seja fornecida no Arquivo de Informações do Cluster.

Formato do Arquivo:

<# Cluster><# Conjunto de Dados><# Categoria><Observado><Esperado><Obs/Exp><RR>

O arquivo terá o mesmo nome que o arquivo padrão de resultados, mas com as extensões *.sci.txt e *.sci.dbf, respectivamente, e estarão localizados no mesmo diretório.

Tópicos Relacionados: *Arquivo de Informações do Cluster, Arquivo de Informações das Localidades, Guia de Saída, Resultados das Análises, Arquivo Padrão de Resultados.*

Arquivo de Informações das Localidades (*.gis.*)

Como uma opção, um arquivo especial de saída pode ser criado descrevendo os vários clusters de uma maneira que é fácil de se incorporar em um sistema de informação geográfica (SIG). Este arquivo pode ser solicitado em formato ASCII e/ou dBase, e pode ser acessado usando qualquer editor de texto ou programa de planilha. Ele terá o mesmo nome do arquivo de resultados, mas com as extensões *.gis.txt e *.gis.dbf, respectivamente, e será localizado no mesmo diretório. Este arquivo tem uma linha para cada localidade pertencente a um cluster. As colunas mostradas dependem da análise escolhida, incluindo, dentre outras, as seguintes informações:

<ID da Localidade>

<Número do Cluster>

<Valor de 'P' do Cluster>

<Casos Observados no Cluster>

<Casos Esperados no Cluster>

<Observado/Esperado no Cluster>

<Risco Relativo no Cluster>

<Casos Observados na Localidade>

<Casos Esperados na Localidade>

<Observado/Esperado na Localidade>

<Risco Relativo na Localidade>

<F de Oliveira> (quando uma análise de fronteira é solicitada)

Nota: A segunda, terceira, quarta, quinta, sexta e sétima entradas de coluna são as mesmas para todas as localidades pertencentes ao mesmo cluster.

Tópicos Relacionados: *Guia de Saída, Resultados das Análises, Arquivo Padrão de Resultados, Arquivo de Informações do Cluster.*

Arquivo de Estimativas de Risco para Cada Localidade (*.rr.*)

Se a opção de incluir as estimativas de risco para cada localidade for selecionada, um arquivo com uma lista de todos os dados de localidades e o número correspondente de casos observados, número de casos esperados, a razão de observado/esperado e o risco relativo para cada localidade é fornecido. Isto pode ser útil quando se examina uma área de cluster em mais detalhes. A informação é puramente descritiva. Existe uma linha para cada ID de Localidade, e o conteúdo das cinco colunas é o que segue:

<ID da Localidade><Casos Observados><Casos Esperados> <Observado/Esperado><Risco Relativo>

Este arquivo pode ser acessado usando qualquer editor de texto ou programa de planilha. Ele terá o mesmo nome do arquivo de resultados, mas com a extensão *.rr.txt ou *.rr.dbf, e estará localizado no mesmo diretório. O arquivo está disponível apenas para a estatística de varredura discreta e, portanto, não para o modelo contínuo de Poisson.

Se o 'F' de Oliveira for solicitado na Guia de Análise de Fronteira, haverá uma coluna adicional mostrando o 'F' de Oliveira para cada ID de localidade.

Tópicos Relacionados: *'F' de Oliveira, Guia de Saída, Resultados das Análises, Arquivo Padrão de Resultados.*

Arquivo dos Log das Razões de Verossimilhança Simulados (*.llr.*)

As estatísticas do teste do log da razão de verossimilhança dos conjuntos de dados aleatórios não são fornecidas como parte da saída padrão. Se desejados, eles podem ser impressos/gravados em um arquivo especial que, por convenção, tem o mesmo nome que o arquivo de saída, mas com a extensão *.llr.txt ou *.llr.dbf. Geralmente não há necessidade desse arquivo, mas ele pode ser útil para os pesquisadores estatísticos que podem estar interessados nas propriedades de distribuição da estatística de varredura sob vários cenários.

Tópicos Relacionados: *Guia de Saída, Resultados das Análises, Arquivo Padrão de Resultados, Replicações de Monte Carlo.*

Miscelânea

Novas Versões

Para verificar se existe uma versão mais recente do que a que está se utilizando, basta clicar no botão de atualização  na barra de ferramentas. Se existir uma nova versão, será perguntado se o usuário deseja fazer o download e instalá-la. Depois de verificar, é possível solicitar que o SaTScan verifique automaticamente por novas versões uma vez por semana, uma vez por mês ou a cada vez que o SaTScan for usado. Alternativamente, é possível configurar o SaTScan para verificar apenas manualmente para novas versões, quando o usuário decidir fazê-lo.

Em qualquer momento, é possível também fazer o download da versão mais recente do SaTScan, a partir da Rede Mundial de Computadores em '<http://www.satscan.org/>'.

Tópicos Relacionados: *Download e Instalação.*

Arquivo do Histórico das Análises

No arquivo do histórico das análises, o SaTScan mantém automaticamente um registo de todas as análises conduzidas no SaTScan. No registo está incluído um número atribuído à análise, juntamente com informações sobre o tempo da análise, configurações de parâmetros, um resumo muito breve dos resultados, bem como o nome do arquivo padrão de resultados criado.

O histórico das análises está em um arquivo dBase com o nome AnalysisHistory.dbf, localizado no mesmo diretório que o executável do SaTScan. Ele pode ser aberto e lido usando a maioria dos softwares de bancos de dados e planilha, incluindo o Excel. Pode-se apagar o arquivo a qualquer momento. Um novo arquivo será então criado na próxima vez em que se executar o SaTScan, começando a lista de análises a partir do zero.

Tópicos Relacionados: *Executando o SaTScan, Resultados das Análises.*

Gerador de Números Aleatórios

A escolha do gerador de números aleatórios é crítica para qualquer software de criação de dados simulados. O SaTScan utiliza um gerador de números aleatórios de Lehmer²⁶, com o módulo $2^{31}-1 = 2147483647$ e multiplicador 48271, que é conhecido por ter um bom desempenho²⁷.

Tópicos Relacionados: *Replicações de Monte Carlo.*

Contate-Nos

Dirigir perguntas técnicas sobre a instalação e a execução do programa, bem como o web site, para:

techsupport@satscan.org

Dirigir perguntas relativas aos métodos estatísticos e sugestões sobre novos recursos para:

Martin Kulldorff, Professor, Biostatistician
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
133 Brookline Avenue, 6th Floor, Boston, MA 02215, USA
Email: kulldorff@satscan.org

Agradecimentos

Apoio Financeiro

Instituto Nacional do Câncer, Divisão de Prevenção do Câncer, Seção de Biometria [SaTScan v1.0, 2.0, 2.1]

Instituto Nacional do Câncer, Divisão de Controle do Câncer e Ciências da População, Seção de Pesquisa e Aplicações Estatísticas [SaTScan v3.0 (parcial), 6.1 (parcial), 9.2-9.4]

Fundação Alfred P. Sloan, através de uma concessão para a Academia de Medicina de New York (Farzad Mostashari, PI) [SaTScan v3.0 (parcial), 3.1, 4.0, 5.0, 5.1]

Centros de Prevenção e Controle de Doenças, através do Acordo de Cooperação da Associação de Escolas Médicas Americanas, número de concessão MM-0870 [SaTScan v6.0, 6.1 (parcial)]

Instituto Nacional de Saúde e Desenvolvimento da Criança, por meio de concessão #R01HD048852 [7.0,8.0,v9.0 (parcial)]

Instituto Nacional de Ciências Médicas Gerais, através do Estudo de Modelagem de Agentes de Doenças Infecciosas (MIDAS) concessão #U01GM076672 [v9.0 (parcial),9,1]

Seu apoio financeiro é extremamente apreciado. Os conteúdos do SaTScan são de responsabilidade do desenvolvedor e não refletem necessariamente a opinião oficial das entidades financiadoras.

Comentários e Sugestões

O feedback dos usuários é muito bem-vindo. Sugestões bastante valiosas relativas ao software SaTScan têm sido recebidas de muitos indivíduos, incluindo:

Allyson Abrams, Escola Médica de Harvard & Assistência à Saúde de Harvard Pilgrim

Frank Boscoe, Secretaria Estadual de Saúde de New York

Eric Feuer, Instituto Nacional do Câncer

Laurence Freedman, Instituto Nacional do Câncer

David Gregorio, Universidade de Connecticut

Göran Gustafsson, Instituto Karolinska, Suécia

Jessica Hartman, Academia de Medicina de New York

Richard Heffernan, Secretaria de Saúde da Cidade de New York

Kevin Henry, Secretaria de Saúde de New Jersey

Ulf Hjalmar, Hospital de Östersund, Suécia

Richard Hoskins, Secretaria Estadual de Saúde de Washington

Lan Huang, Instituto Nacional do Câncer
Ahmedin Jemal, Sociedade Americana do Câncer
Inkyung Jung, Escola Médica de Harvard & Assistência à Saúde de Harvard Pilgrim
Ann Klassen, Universidade de Johns Hopkins
Ken Kleinman, Escola Médica de Harvard & Assistência à Saúde de Harvard Pilgrim
Sanjaya Kumar, Secretaria de Saúde do Estado de New York
Kristina Metzger, Secretaria de Saúde da Cidade de New York
Barry Miller, Instituto Nacional do Câncer
Farzad Mostashari, Secretaria de Saúde da Cidade de New York
Lloyd Mueller, Registro de Tumores de Connecticut
Karen Olson, Hospital Pediátrico de Boston
Linda Pickle, Instituto Nacional do Câncer
Simon Read, Universidade de Sheffield
Tom Richards, Centros de Prevenção e Controle de Doenças
Gerhard Rushton, Universidade de Iowa
Joseph Sheehan, Universidade de Connecticut
Tom Talbot, Secretaria Estadual de Saúde de New York
Toshiro Tango, Instituto Nacional de Saúde Pública, Japão
Jean-François Viel, Universidade de Franche-Comté, França
Shihua Wen, Universidade de Maryland
Li Zhu, Instituto Nacional do Câncer

Sobre o Manual Traduzido Para a Língua Portuguesa

Contato

Alessandra Cristina Guedes Pellini
Médica, Especialista em Doenças Infecciosas e Parasitárias
Mestre em Saúde Pública pela Faculdade de Saúde Pública da Universidade de São Paulo (FSP/USP)
Aluna de Doutorado do Departamento de Epidemiologia do Programa de Pós-Graduação em Saúde Pública da FSP/USP
Av. Dr. Arnaldo, 715 - CEP 01246-904 - São Paulo - SP - Brasil
Email: acgpellini@gmail.com

Agradecimentos

- Ao Prof. Martin Kulldorff, da Escola de Medicina de Harvard, criador do software SaTScan e autor do manual do usuário original, que incentivou a publicação desse manual traduzido para o Português.
- Ao Prof. Dr. Francisco Chiaravalloti Neto, da Faculdade de Saúde Pública da Universidade de São Paulo, que me apresentou o programa SaTScan, o qual foi de imensa utilidade para a minha Tese de Doutorado e para o meu trabalho na vigilância epidemiológica de doenças.

Perguntas Mais Frequentes

Dados de Entrada

- 1. Eu tentei executar o SaTScan utilizando um dos conjuntos de amostras de dados, e correu tudo bem, mas quando eu o experimento em meus próprios dados ocorre um erro. O que eu devo fazer?**

O SaTScan garante que os dados de entrada sejam compatíveis uns com os outros, e com as opções especificadas na interface das janelas. Por exemplo, ele reclama se houver um ID de localidade no arquivo casos que não esteja presente no arquivo de coordenadas, dado que ele precisa saber onde localizar esses casos. Para a maioria dos conjuntos de dados há alguma necessidade de limpeza de dados, e o SaTScan é projetado para ajudar nesse processo, identificando e apontando quaisquer inconsistências encontradas.

- 2. Eu construí os arquivos de entrada em ASCII, exatamente conforme a descrição no Manual do Usuário do SaTScan, mas o programa reclama que eles não estão no formato correto. O que está errado?**

A explicação mais plausível é que os arquivos estejam em formato UNICODE, ao invés de formato ASCII. Apenas converter para ASCII e ele deve funcionar.

- 3. Nos meus dados, há zero ou apenas um caso na maioria das localidades. Eu posso usar o SaTScan para esses dados escassos?**

Sim, certamente você pode. Uma das principais razões para usar o SaTScan é evitar a aglomeração geográfica arbitrária dos dados, deixando a estatística de varredura considerar diferentes aglomerações, menores ou maiores, através de sua janela que se move continuamente. Com uma resolução geográfica mais fina dos dados de entrada, o SaTScan pode avaliar os mais diferentes tamanhos e localidades de cluster, sem as restrições impostas por delimitações geográficas administrativas, minimizando os pressupostos sobre a localização e a dimensão geográfica do cluster.

- 4. Se os meus dados forem escassos, as taxas não serão estatisticamente instáveis?**

A estabilidade das taxas não depende da resolução geográfica dos dados de entrada, mas do tamanho da população dos círculos construídos pelo SaTScan.

- 5. Qual é o número mínimo de localidades espaciais necessárias para se executar o SaTScan?**

A estatística de varredura puramente temporal pode ser executada com apenas uma localidade geográfica. A estatística de varredura espaço-temporal precisa de pelo menos duas localidades. Com apenas duas localidades, a estatística de varredura espaço-temporal irá procurar por clusters temporais em um ou ambos os locais. Tecnicamente, a estatística de varredura puramente espacial também pode ser executada usando apenas duas localidades geográficas, fornecendo inferência correta. Não há sentido, no entanto, em se usar uma estatística de varredura puramente espacial para esses dados, para os quais uma estatística regular de qui-quadrado pode ser utilizada, visto que não há testes múltiplos para se ajustar. Com três ou mais localidades, o conceito fundamental da estatística de varredura de incluir diferentes combinações de localidades dentro de potenciais aglomerados está sendo utilizado. Na maioria das aplicações práticas, porém, as estatísticas de varredura espaciais e espaço-temporais são usadas para conjuntos de dados com centenas ou milhares de localidades geográficas. Se há uma escolha, menor agregação espacial dos dados é tipicamente melhor, o que significa mais localidades geográficas.

Análises

6. Com as coordenadas de latitude/longitude, que projeção plana é utilizada?

Nenhuma projeção é utilizada. O SaTScan desenha círculos perfeitos na superfície esférica da terra.

7. Quando eu devo usar o modelo de Bernoulli versus o modelo de Poisson?

Use o modelo de Bernoulli quando você tiver dados binários, tais como casos e controles, estágio tardio e inicial de câncer ou pessoas com e sem uma doença. Utilize o modelo de Poisson quando você tiver casos e uma população de base sob risco, tais como números populacionais do censo.

8. O STaTScan ajusta para covariáveis categóricas, mas eu quero ajustar para uma variável contínua. Isso é possível?

Uma maneira de se fazer isso é categorizar a variável contínua. A melhor abordagem é (i) calcular o ajuste usando um pacote de software estatístico regular, como o SAS, (ii) utilizar o resultado dessas análises para calcular o número esperado de casos ajustado para a covariável em cada localidade, e (iii) usar esses valores esperados ao invés da população no arquivo de população. Com esta abordagem, não deve haver quaisquer covariáveis nos arquivos de casos ou da população.

9. O que eu devo usar como o tamanho máximo do cluster geográfico? Esta é uma escolha arbitrária?

Se você não quiser que a escolha seja arbitrária, escolher 50% da população como o tamanho máximo do cluster geográfico. O SaTScan irá então avaliar aglomerados muito pequenos e muito grandes, e todo o resto entre estes. Para encontrar uma boa coleção de aglomerados não sobrepostos, utilize o recurso do índice de Gini.

10. Por que eu não posso selecionar um tamanho máximo do cluster geográfico maior do que 50% da população?

Clusters de risco excessivo maiores do que 50% da população sob risco são melhor vistos como clusters com menor risco fora da janela de varredura, e a área externa sempre terá uma forma geográfica muito irregular. Se houver interesse em aglomerados com risco mais baixo do que o esperado, é mais apropriado selecionar a opção de taxas baixas na guia de análise.

11. Eu tenho problemas de memória ao executar o STaTScan. O que eu devo fazer?

Certifique-se de que você está executando o SaTScan no modo de 64 bits. Para isso, você deve (i) ter um computador de 64 bits, (ii) executar o SaTScan na versão 8 ou mais recente, e (iii) ter Java de 64 bits instalado no seu computador.

Resultados

12. Eu recebo um erro informando que o arquivo de saída não pôde ser criado. Por quê?

No Windows, permissão para gravar na pasta "Arquivos de Programas" é dada apenas para administradores e usuários avançados dessa máquina. Se o caminho do arquivo de saída inclui a pasta "Arquivos de Programas" e você não tem privilégios de administrador ou de usuário avançado em seu computador, o Windows impede o SaTScan de criar o arquivo de saída no local designado. A solução é se especificar um nome de arquivo de saída diferente usando um diretório diferente.

13. Dado que os resultados do SaTScan são baseados na simulação de dados aleatórios de Monte Carlo, por que os valores de 'p' são os mesmos quando eu executo a análise duas vezes?

Todas as simulações baseadas em computador se fundamentam em geradores de números pseudoaleatórios. Quando a mesma base é usada, gera-se exatamente a mesma sequência de números pseudoaleatórios. Uma vez que o SaTScan utiliza a mesma base para cada execução, obtém-se o mesmo resultado para duas execuções quando os dados de entrada são os mesmos.

14. Eu executei exatamente os mesmos dados usando duas versões diferentes do SaTScan, mas os valores de 'p' são diferentes. Por quê? Qual deles é o correto?

Em comparação com a v2.1, a geração de números pseudoaleatórios é feita de forma um pouco diferente no SaTScan v3.0 e em versões posteriores, geralmente resultando em valores de 'p' ligeiramente diferentes. Na versão anterior, o SaTScan definia aglomerados sobrepostos com base em se os dois círculos eram sobrepostos. No SaTScan v5.0 e posteriores, dois clusters se sobrepõem se eles tivessem pelo menos um ID de localidade em comum. Estas duas definições são geralmente as mesmas, mas em casos raros elas podem ser diferentes. Se você estava executando o modelo de Poisson, outra possível razão para essa diferença é que o SaTScan v5.0 e posteriores utilizam um algoritmo mais preciso para calcular o número de casos esperados quando as datas da população no arquivo de população são especificadas usando dias ao invés de meses ou anos.

Embora os valores de 'p' de todas as versões sejam válidos e corretos, apenas um valor de 'p' deve ser usado. Recomendamos sempre usar o valor de 'p' que foi calculado primeiro.

15. Eu encontrei um cluster espacial, mas algumas das localidades no cluster não têm nenhum caso. Por que elas estão incluídas no cluster?

Um código postal de localidade é incluído no cluster se o seu centróide for incluído no círculo de varredura. Isto significa que, para dados escassos, é muito comum ter localidades sem nenhum caso no cluster. Isso geralmente acontece quando as contagens esperadas são baixas e ele é cercado por outra localidade com uma grande quantidade de casos. Também é importante observar que enquanto o SaTScan apresenta os limites exatos do cluster detectado, as fronteiras exatas não são confiáveis e podem existir áreas dentro do cluster detectado que não pertencem ao verdadeiro aglomerado e vice-versa. Isso pode acontecer (i) por acaso, onde algumas áreas com risco elevado ainda têm zero casos devido a pequenas contagens esperadas (por exemplo, 1 esperado sob H_0 , 2 esperados devido a um RR de 2, e 0 observado a partir de uma distribuição de Poisson com 2 esperados, tem uma alta probabilidade de gerar uma contagem de 0), (ii) se o verdadeiro cluster não for exatamente circular ou (iii) se existirem vários clusters próximos uns dos outros com uma população avulsa no meio. Neste último caso, o SaTScan pode apresentar um aglomerado combinado, ao invés de 2-3 aglomerados separados.

Interpretação

16. No SaTScan, após o ajuste para a densidade populacional e covariáveis como a idade, a hipótese nula é a completa aleatoriedade espacial. Para a maioria dos dados de doenças isso não é verdade. Isso significa que a hipótese nula está errada?

Ao se aceitar a noção de testes de hipótese estatísticos, é preciso também aceitar o fato de que a hipótese nula nunca é verdadeira. Por exemplo, ao se comparar a eficácia de dois procedimentos cirúrgicos diferentes em um ensaio clínico, sabemos com certeza que a sua eficácia não pode ser igual, mas ainda assim nós usamos a igualdade como a hipótese nula, visto que estamos interessados em descobrir se um é melhor do que o outro. Da mesma forma, com dados geográficos nós sabemos que o risco de doença não é o mesmo em todos os lugares, mas ainda assim usamos isto como a hipótese nula, uma vez que estamos interessados em encontrar localidades com excesso de risco. Por essa razão, a hipótese nula é errada no sentido de que nós sabemos que ela não é verdadeira, mas não é errada no sentido de que não devemos usá-la.

17. O SaTScan assume que não há autocorrelação espacial nos dados? (Nota: Autocorrelação espacial significa que a localização dos casos da doença é dependente da localização de outros casos da doença, tal como com uma doença infecciosa em que um indivíduo infectado tende a infectar as pessoas que vivem por perto.)

Não, o SaTScan não pressupõe que não há nenhuma autocorrelação espacial nos dados. Ao contrário, este é um teste para saber se existe autocorrelação espacial ou outras divergências em relação à hipótese nula. Neste sentido, é equivalente a um teste estatístico para a normalidade, que não assume que os dados são normalmente distribuídos, mas testa se eles são.

18. Se eu estou interessado em saber se existe autocorrelação espacial nos dados, por que eu deveria usar a estatística de varredura espacial ao invés de um teste tradicional de autocorrelação espacial?

Se você está interessado apenas em saber se existe ou não autocorrelação espacial, mas não se importa com as localidades do cluster, existem testes para a autocorrelação espacial / aglomerado global que têm poder maior do que a estatística de varredura espacial e devem então ser usados nesse caso. A estatística de varredura espacial deve ser usada quando você estiver interessado na detecção e na significância estatística de aglomerados locais.

19. Na estatística espacial, não é sempre importante se ajustar para a autocorrelação espacial? Isto não pode ser feito no SaTScan.

Se ajustar para autocorrelação espacial depende da questão que se coloca a partir dos dados. Como um exemplo, suponhamos que temos dados geográficos sobre as pessoas que ficam doentes devido à intoxicação alimentar. Em tais dados há claramente uma autocorrelação espacial, uma vez que alimentos ruins vendidos em restaurantes ou supermercados são frequentemente vendidos para vários clientes, muitos dos quais vivem no mesmo bairro.

Se nós estamos fazendo uma regressão espacial tentando determinar quais características da vizinhança, como renda média, valores das casas, níveis educacionais ou origem étnica contribuem para um risco maior de intoxicação alimentar, é fundamental se ajustar para a autocorrelação espacial nos dados. Se não, a confiança nas relações de risco será superestimada, com valores de 'p' tendenciosos que são muito pequenos, fornecendo resultados 'estatisticamente significativos' quando não existem. Aqui, a hipótese nula deve ser de que há uma autocorrelação espacial, e a hipótese alternativa de que existem diferenças geográficas no risco de intoxicação alimentar.

Por outro lado, se estamos interessados em detectar rapidamente surtos de intoxicação alimentar, não devemos ajustar para a autocorrelação espacial, uma vez que estamos interessados em detectar aglomerados devido a tal correlação, e se eles forem ajustados, aglomerados importantes podem não ser detectados. Aqui, a hipótese nula é de que os casos de intoxicação alimentar estão geograficamente distribuídos de forma aleatória (ajustados para a densidade populacional, etc.) e a hipótese alternativa é de que existe alguma aglomeração, quer devido a diferenças nos fatores de risco subjacentes ou à autocorrelação espacial. Uma vez que a localização de um cluster tenha sido detectada, cabe às autoridades locais de saúde determinar a origem do cluster para prevenir mais doenças.

20. Se houver vários clusters nos dados, isso significa que os valores de ‘p’ são mais prováveis de serem significativos do que o seu nível de significância nominal de 0,05 sugere, de modo que os clusters ao acaso são detectados com muita frequência?

Não. O oposto de fato é verdadeiro. Olhando para a mortalidade nos Estados Unidos, suponha que haja 1000 casos de uma doença em Seattle e 30 em New York City. Seattle é claramente um aglomerado significativo, mas 30 casos em New York City dos 1030 em todos os EUA não é excepcional, dado que a cidade tem cerca de 3 por cento da população dos EUA. Se aceitarmos, contudo, que há um cluster em Seattle, e se ajustarmos para este removendo Seattle da análise, então 30 casos na cidade dos 30 em todo o país é estatisticamente significativo. Isso é semelhante a uma regressão múltipla regular, onde ao se ajustar para uma variável, outra variável pode de repente se tornar estatisticamente significativa. Observar que o oposto também é verdadeiro. Se removermos uma área com significativamente menos casos do que o esperado, então um cluster significativo com um número excessivo de casos pode se tornar não significativo.

21. Para os dados de contagem, a estatística de varredura espacial utiliza uma hipótese alternativa particular com um excesso de risco em um cluster circular, onde o número de casos segue uma distribuição de Poisson ou de Bernoulli. Isso significa que ela pode ser usada apenas para detectar essas hipóteses alternativas?

Muitos testes estatísticos propostos e amplamente utilizados não especificam uma hipótese alternativa. Isto não significa que eles não podem ser usados para quaisquer hipóteses alternativas, nem que eles são bons para todas as alternativas. Da mesma forma, se uma alternativa explícita é definida, tal como na estatística de varredura espacial, isto não significa que ela não pode ser também utilizada para outras hipóteses alternativas. É simplesmente uma questão da estatística do teste ter um bom poder para algumas hipóteses alternativas e baixo poder para outras. A vantagem de se dispor de uma alternativa bem especificada é que esta traz alguma informação sobre as alternativas para as quais se pode esperar que o teste tenha um bom poder.

22. Para o modelo exponencial (normal), assume-se que os tempos de sobrevida seguem uma distribuição exponencial (normal). Os resultados serão tendenciosos se os tempos de sobrevida seguirem uma distribuição diferente?

Não importa qual distribuição gerou os tempos de sobrevida, os valores de ‘p’ da inferência estatística continuam válidos e não enviesados. Isto ocorre porque, ao invés de gerar os dados aleatórios a partir de uma distribuição exponencial, cada um dos dados aleatórios é uma permutação espacial dos tempos de sobrevida. Uma distribuição muito mal especificada pode levar, por consequência, a uma perda de poder. Por exemplo, se os dados seguem a distribuição de Bernoulli, o modelo exponencial tem menos poder para detectar um aglomerado do que o modelo de Bernoulli. Para distribuições contínuas, tais como a gama e a normal logarítmica, o modelo exponencial demonstrou funcionar bem. O mesmo raciocínio é verdadeiro no que diz respeito ao modelo normal.

Sistemas Operacionais

23. O SaTScan está disponível para Windows/Mac/Linux?

O software SaTScan para Windows, Mac e Linux pode ser baixado a partir do site www.satscan.org.

Bibliografia do SaTScan

Diferentes opções de análise do SaTScan foram desenvolvidas em diferentes momentos e elas estão descritas em diferentes publicações científicas. A bibliografia a seguir contém trabalhos/artigos selecionados e relatórios destinados a auxiliar o usuário a encontrar informações sobre o seguinte:

1. Encontrar o(s) artigo(s) metodológico(s) nos quais as diversas opções de análise são apresentadas e discutidas em mais detalhes do que o que está disponível aqui no Manual do Usuário do SaTScan.
2. Encontrar aplicações em diferentes áreas científicas.
3. Determinar os trabalhos científicos relevantes para citar.

Citações Sugeridas

O software SaTScan pode ser utilizado gratuitamente, com a condição de que as referências apropriadas sejam fornecidas para os artigos científicos que descrevem os métodos estatísticos. Para as análises mais comuns, as citações sugeridas são:

Modelos de Bernoulli, Discreto de Poisson e Contínuo de Poisson: Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26:1481-1496, 1997. [online no manual original]

Modelo de Permutação no Espaço-Tempo: Kulldorff M, Heffernan R, Hartman J, Assunção RM, Mostashari F. A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, 2:216-224, 2005. [online no manual original]

Modelo Multinomial: Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Statistics in Medicine*, 2010, epub. [online no manual original]

Modelo Ordinal: Jung I, Kulldorff M, Klassen A. A spatial scan statistic for ordinal data. *Statistics in Medicine*, 2007; 26:1594-1607. [online no manual original]

Modelo Exponencial: Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. *Biometrics*, 2007; 63:109-118. [online no manual original]

Modelo Normal sem Pesos: Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics*, 2009, 8:58. [online no manual original]

Modelo Normal com Pesos: Huang L, Huang L, Tiwari R, Zuo J, Kulldorff M, Feuer E. Weighted normal spatial scan statistic for heterogeneous population data. *Journal of the American Statistical Association*, 2009, 104:886-898. [online no manual original]

Software: Kulldorff M. and Information Management Services, Inc. SaTScan™ v8.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>, 2009.

Os usuários do SaTScan devem, em qualquer referência ao software, anotar que: “SaTScan™ é uma marca registrada de Martin Kulldorff. O software SaTScan™ foi desenvolvido sob o patrocínio conjunto de (i) Martin Kulldorff, (ii) o Instituto Nacional do Câncer, e (iii) Farzad Mostashari do Departamento de Saúde e Higiene Mental da Cidade de New York.”

Ao utilizar o manual traduzido para o Português, devem ser mencionados, após a citação do trabalho original e seus autores, os créditos a seguir: “Tradução de Alessandra Cristina Guedes Pellini”.

Tópicos Relacionados: *Bibliografia do SaTScan, Artigos Metodológicos.*

Artigos de Metodologia do SaTScan

Metodologia Estatística

Teoria Geral da Estatística, Modelos de Bernoulli e Poisson

1. Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 1997; 26:1481-1496. [online no manual original]

Estatística de Varredura Espacial, Modelo de Bernoulli

2. Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 1995; 14:799-810. [online no manual original]

Estatística de Varredura Espaço-Temporal Retrospectiva

3. Kulldorff M, Athas W, Feuer E, Miller B, Key C. Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *American Journal of Public Health*, 1998; 88:1377-1380. [online no manual original]

Estatística de Varredura Espaço-Temporal Prospectiva

4. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society*, 2001; A164:61-72. [online no manual original]

Modelo de Permutação no Espaço-Tempo

5. Kulldorff M, Heffernan R, Hartman J, Assunção RM, Mostashari F. A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, 2005; 2:216-224. [online no manual original]

Modelo Multinomial

6. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Statistics in Medicine*, 2010, epub. [online no manual original]

Modelo Ordinal

7. Jung I, Kulldorff M, Klassen A. A spatial scan statistic for ordinal data. *Statistics in Medicine*, 2007; 26:1594–1607. [online no manual original]

Modelo Exponencial

8. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. *Biometrics*, 2007, 63:109-118. [online no manual original]

Modelo Normal

9. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics*, 2009, 8:58. [online no manual original]

Modelo Normal Ponderado

10. Huang L, Huang L, Tiwari R, Zuo J, Kulldorff M, Feuer E. Weighted normal spatial scan statistic for heterogeneous population data. *Journal of the American Statistical Association*, 2009, 104:886-898. [online no manual original]

Variação Espacial nas Tendências Temporais

11. Manuscrito em preparação.

Estatística de Varredura Multivariada

12. Kulldorff M, Mostashari F, Duczmal L, Yih K, Kleinman K, Platt R. Multivariate spatial scan statistics for disease surveillance. *Statistics in Medicine*, 2007, 26:1824-1833. [online no manual original]

Janela de Varredura Elíptica

13. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Statistics in Medicine*, 2006, 25:3929-3943. [online no manual original]

Estatística de Varredura Espacial Isotônica

14. Kulldorff M. An isotonic spatial scan statistic for geographical disease surveillance. *Journal of the National Institute of Public Health*, 1999;48:94-101. [online no manual original]

Teste de Hipóteses de Monte Carlo

15. Dwass M. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 1957; 28:181-187.
16. Besag J, Clifford J. Sequential Monte Carlo p-values. *Biometrika*, 1991; 78:301-330.
17. Silva I, Assunção RM, Costa M. Power of the sequential Monte Carlo test. *Sequential Analysis*, 2009; 28:163-174.

Valores de 'P' de Gumbel

18. Abrams A, Kleinman K, Kulldorff M. Gumbel based p-value approximations for spatial scan statistics.. *International Journal of Health Geographics* 2010, 9:61. [online no manual original]
19. Read S, Bath PA, Willett P, Maheswaran R. A study on the use of Gumbel approximation with the Bernoulli spatial scan statistic. *Statistics in Medicine*, 2013.

Intervalos de Recorrência

20. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology*, 159:217-24, 2004.

Ajustes

Ajustando por Covariáveis

Referências [1] e [8] acima, mais:

21. Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer in northeastern United States: A geographical analysis. *American Journal of Epidemiology*, 146:161-170, 1997. [online no manual original]
22. Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. *Epidemiology and Infection*, 2005, 133:409-419.
23. Klassen A, Kulldorff M, Curriero F. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. *International Journal of Health Geographics*, 2005, 4:1. [online no manual original]

Estatísticas de Varredura Iterativas, Ajustando para Clusters Mais Prováveis

24. Zhang Z, Kulldorff M, Assunção R. Spatial scan statistics adjusted for multiple clusters. *Journal of Probability and Statistics*, 2010, 642379.

Aspectos Computacionais

Algoritmos

25. Kulldorff M. Spatial scan statistics: Models, calculations and applications. In Balakrishnan and Glaz (eds), Recent Advances on Scan Statistics and Applications. Boston, USA: Birkhäuser, 1999. [online no manual original]

Gerador de Números Aleatórios

26. Lehmer DH. Mathematical methods in large-scale computing units. In Proceedings of the second symposium on large scale digital computing machinery. Cambridge, USA: Harvard Univ. Press, 1951.
27. Park SK, Miller KW. Random number generators: Good ones are hard to find. Communications of the ACM, 31:1192-1201, 1988.

Macros

28. Abrams AM, Kleinman KP. A SaTScan (TM) macro accessory for cartography (SMAC) package implemented with SAS (R) software. International Journal of Health Geographics, 6:6,2007. [online no manual original]

Relatórios, Visualização e Mapeamento

29. Han J, Zhu L, Kulldorff M, Hostovich S, Tatalovich Z, Lewis D, Feuer E. Determining optimal cluster reporting sizes for spatial scan statistics. Manuscript, 2014, submitted.

Google Earth

30. North American Association of Central Cancer Registries, SaTScan to Google Earth Conversion Tool, [online no manual original]

Outras Visualizações

31. Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. Health and Place, 9:273-277, 2003.

Avaliações e Comparações de Metodos

32. Kulldorff M, Tango T, Park P. Power comparisons for disease clustering tests. Computational Statistics and Data Analysis, 42:665-684, 2003.
33. Song C, Kulldorff M. Power evaluation of disease clustering tests. International Journal of Health Geographics, 2:9, 2003. [online no manual original]
34. Kulldorff M, Zhang Z, Hartman J, Heffernan R, Huang L, Mostashari F. Evaluating disease outbreak detection methods: Benchmark data and power calculations. Morbidity and Mortality Weekly Report, 53:144-151, 2004. [online no manual original]
35. Nordin J, Goodman M, Kulldorff M, Ritzwoller D, Abrams A, Kleinman K, Levitt MJ, Donahue J, Platt R. Using modeled anthrax attacks on the Mall of America to assess sensitivity of syndromic surveillance. Emerging Infectious Diseases, 11:1394-1398, 2005. [online no manual original]
36. Ozdenerol E, Williams BL, Kang SY, Magsumbol MS. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. International Journal of Health Geographics, 4:19, 2005. [online no manual original]
37. Costa MA, Assunção RM. A fair comparison between the spatial scan and Besag-Newell disease clustering tests. Environmental and Ecological Statistics, 12:301-319, 2005.

38. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4:11, 2005. [online no manual original]
39. Kulldorff M, Song C, Gregorio D, Samociuk H, DeChello L. Cancer map patterns: Are they random or not? *American Journal of Preventive Medicine*, 30:S37-49, 2006. [online no manual original]
40. Duczmal L, Kulldorff M, Huang L. Evaluation of spatial scan statistics for irregular shaped clusters. *Journal of Computational and Graphical Statistics*, 15:428-442, 2006.
41. Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics*, 5:15, 2006. [online no manual original]
42. Jackson MC, Huang L, Luo J, Hachey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *International Journal of Health Geographics*. 2009;8:55. [online no manual original]
43. Wheeler DC. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996–2003. *International Journal of Health Geographics*. 2007;6:13. [online no manual original]
44. Goujon-Bellec S, Demoury C, Guyot-Goubin A, Hémon D, Clavel J. Detection of clusters of a rare disease over a large territory: performance of cluster detection methods. *International Journal of Health Geographics*. 2011;10:53. [online no manual original]

Tópicos Relacionados: *Bibliografia do SaTScan, Aplicações Seleccionadas por Área de Estudo, Citações Sugeridas.*

Aplicações do SaTScan Seleccionadas por Área de Estudo

Doenças Respiratórias Infeciosas

45. Andrade AL, Silva SA, Martelli CM, Oliveira RM, Morais Neto OL, Siqueira Junior JB, Melo LK, Di Fabio JL. Population-based surveillance of pediatric pneumonia: use of spatial analysis in an urban area of Central Brazil. *Cadernos de Saúde Pública*. 20: 411-421, 2004. [online no manual original]
46. Bakker MI, Hatta M, Kwenang A, Faber WR, van Beers SM, Klatser PR, Oskam L. Population survey to determine risk factors for *Mycobacterium leprae* transmission and infection. *International Journal of Epidemiology*, 33: 1329-1336, 2004.
47. Elias J, Harmsen D, Claus H, Hellenbrand W, Frosch M, Vogel U. Spatiotemporal analysis of invasive meningococcal disease, Germany. *Emerging Infectious Diseases*, 12:1689-1695, 2006. [online no manual original]
48. Oeltmann JE, Varma JK, Ortega L, Liu Y, O'Rourke T, Cano M, Harrington T, Toney S, Jones W, Karuchit S, Diem L, Rienthong D, Tappero JW, Ijaz K, Maloney, S. Multidrug-Resistant Tuberculosis Outbreak among US-bound Hmong Refugees, Thailand, 2005. *Emerging Infectious Diseases*, 14:1715-1721, 2008. [online no manual original]
49. Fischer EAJ, Pahan D, Chowdhury SK, Oskam L, Richardus JH. The spatial distribution of leprosy in four villages in Bangladesh: An observational study. *BMC Infectious Diseases*, 8:125, 2008.
50. Liang L, Xu B, Chen Y, Liu Y, Cao W, Fang L, Feng L, Goodchild MF, Gong P. Combining spatial-temporal and phylogenetic analysis approaches for improved understanding on global H5N1 transmission. *PLoS One*. 5:e13575, 2010. [online no manual original]
51. Kammerer JS, Shang N, Althomsons SP, Haddad MB, Grant J, Navin TR. Using statistical methods and genotyping to detect tuberculosis outbreaks. *International Journal of Health Geographics*, 12:15, 2013. [online no manual original]

52. Chen JH, Weng C, Chnag HG. Using space-time scan statistic to detect pertussis and shigellosis outbreaks. CSTE Annual Conference, 2013. [online no manual original]
53. Nana Yakam A, Noeske J, Dambach P. Spatial analysis of tuberculosis in Douala, Cameroon: clustering and links with socio-economic status. *The International Journal of Tuberculosis and Lung Disease*, 18:292-297, 2014.
54. Shea KM, Kammerer JS, Winston CA, Navin TR, Horsburgh CR. Estimated rate of reactivation of latent tuberculosis infection in the United States, overall and by population subgroup. *American Journal of Epidemiology*, 179:216-25, 2014.
55. Souris M, Selenic D, Khaklang S, Ninphanomchai S, Minet G, Gonzalez JP, Kittayapong P. Poultry farm vulnerability and risk of avian influenza re-emergence in Thailand. *International Journal of Environmental Research and Public Health*, 11:934-951, 2014. [online no manual original]

Doenças Transmitidas por Água e Alimentos

56. Cruz Payão Pellegrini D. Análise espaço-temporal da leptospirose no município do Rio de Janeiro (1995-1999). Rio de Janeiro: Fundação Oswaldo Cruz, 2002. [online no manual original]
57. Enemark HL, Ahrens P, Juel CD, Petersen E, Petersen RF, Andersen JS, Lind P, Thamsborg SM. Molecular characterization of Danish *Cryptosporidium parvum* isolates. *Parasitology*, 125:331-341, 2002.
58. Sauders BD, Fortes ED, Morse DL, Dumas N, Kiehlbauch JA, Schukken Y, Hibbs JR, Wiedmann M. Molecular subtyping to detect human listeriosis clusters. *Emerging Infectious Diseases*, 9:672-680, 2003. [online no manual original]
59. Odoi A, Martin SW, Michel P, Middleton D, Holt J, Wilson J. Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International Journal of Health Geographics*, 3:11, 2004. [online no manual original]
60. Yih K, Abrams A, Kleinman K, Kulldorff M, Nordin J, Platt R. Ambulatory-care diagnoses as potential indicators of outbreaks of gastrointestinal illness --- Minnesota. *Morbidity and Mortality Weekly Report*, 54 Suppl:157-62, 2005. [online no manual original]
61. Jones RC, Liberatore M, Fernandez JR, Gerber SI. Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Reports*, 121:133-9, 2006.
62. Pearl DL, Louie M, Chui L, Dore K, Grimsrud KM, Leedell D, Martin SW, Michel P, Svenson LW, McEwen SA. The use of outbreak information in the interpretation of clustering of reported cases of *Escherichia coli* O157 in space and time in Alberta, Canada, 2000-2002. *Epidemiology and Infection*, 134:699-711, 2006.
63. de Souza EA, da Silva-Nunes M, Malafrente Rdos S, Muniz PT, Cardoso MA, Ferreira MU. Prevalence and spatial distribution of intestinal parasitic infections in a rural Amazonian settlement, Acre State, Brazil. *Cadernos de Saude Publica*, 23:427-34, 2007. [online no manual original]
64. Osei FB, Duker AA. Spatial dependency of *V. cholera* prevalence on open space refuse dumps in Kumasi, Ghana: a spatial statistical modeling. *International Journal of Health Geographics*, 7:62, 2008. [online no manual original]
65. Sowmyanarayanan TV, Mukhopadhyaya A, Gladstone BP, Sarkar R, Kang G. Investigation of a hepatitis A outbreak in children in an urban slum in Vellore, Tamil Nadu, using geographic information systems. *Indian Journal of Medical Research*, 128:32-37, 2008.

66. Oviedo M, Munoz P, Dominguez A, Carmona G, Batalla J, Borrás E, Jansá JM. Evaluation of Mass Vaccination Programmes: The experience of hepatitis A in Catalonia (in Spanish). *Revista Española de Salud Pública*, 83:697-709, 2009. [online no manual original]
67. Luquero FJ, Banga CN, Remartínez D, Palma PP, Baron E, Grais RF. Cholera epidemic in Guinea-Bissau (2008): the importance of "place". *PLoS One*, 6:e19005, 2011. [online no manual original]
68. Bompangue Nkoko D, Giraudoux P, Plisnier PD, Tinda AM, Piarroux M, Sudre B, Horion S, Tamfum JJ, Ilunga BK, Piarroux R. Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978-2008. *Emerging Infectious Diseases*, 17:2026-2034, 2011. [online no manual original]
69. Tang F, Cheng Y, Bao C, Hu J, Liu W, Liang W, Wu Y, Norris J, Peng Z, Yu R, Shen H, Chen F. Spatio-temporal trends and risk factors for Shigella from 2001 to 2011 in Jiangsu province, People's Republic of China. *PLoS ONE*, 9:e83487, 2014. [online no manual original]
70. Chan TC, Hwang JS, Chen RH, King CC, Chiang PH. Spatio-temporal analysis on enterovirus cases through integrated surveillance in Taiwan. *BMC Public Health*, 14:11, 2014. [online no manual original]
71. Briggs ADM, Boxall NS, van Santen D, Chalmers RM, McCarthy ND. Approaches to the detection of very small, common, and easily missed outbreaks that together contribute substantially to human *Cryptosporidium* infection. *Epidemiology and Infection*, epub, 2014.
72. Wang J, Cao Z, Zeng DD, Wang Q, Wang X, Qian H. Epidemiological Analysis, Detection, and Comparison of Space-Time Patterns of Beijing Hand-Foot-Mouth Disease (2008–2012). *PLoS One*, 9:e2745, 2014. [online no manual original]
73. Viñas M, Tuduri E, Galar A, Yih WK, Pichel M, Stelling J, Brengi S, Della Gaspera A, van der Ploeg C, Bruno S, Rogé A, Caffer M, , Kulldorff M, Galas M. Laboratory-based prospective surveillance for community outbreaks of Shigella spp. in Argentina. *PLoS Neglected Tropical Diseases*, 2013, 7:e2521, 2013.

Doenças Sexualmente Transmissíveis

74. Jennings JM, Curriero FC, Celentano D, Ellen JM. Geographic identification of high gonorrhoea transmission areas in Baltimore, Maryland. *American Journal of Epidemiology*, 161:73-80, 2005.
75. Wylie JL, Cabral T, Jolly AM. Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis. *Journal of Infectious Diseases*, 191:899-906, 2005.
76. Wand H, Ramjee G. Targeting the hotspots: Investigating spatial and demographic variations in HIV infection in small communities in South Africa. *Journal of the International AIDS Society*, 13:41, 2010. [online no manual original]
77. Egger JR, Konty KJ, Borrelli JM, Cumiskey J, Blank S. Monitoring temporal changes in the specificity of an oral HIV test: a novel application for use in postmarketing surveillance. *PLoS One*, 25:e12231, 2010. [online no manual original]
78. Gesink DC, Sullivan AB, Miller WC, Bernstein KT. Sexually transmitted disease core theory: roles of person, place, and time. *American Journal of Epidemiology*. 174:81-9, 2011.
79. Hixon BA, Omer SA, del Rio C, Frew PM. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. *Journal of Urban Health*, 88:129-141, 2011.
80. Cuadros DF, Awad SF, Abu-Raddad LJ. Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *International Journal of Health Geographics*, 12:28, 2013. [online no manual original]

81. Cuadros DF, Abu-Raddad LJ. Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa. *Health and Place*, epub, 2014.

Doenças Transmitidas por Vetores

82. Fevre EM, Coleman PG, Odiit M, Magona JW, Welburn SC, Woolhouse MEJ. The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *The Lancet*, 358:625-628, 2001.
83. Chaput EK, Meek JI, Heimer R. Spatial analysis of human granulocytic ehrlichiosis near Lyme, Connecticut. *Emerging Infectious Diseases*, 8:943-948, 2002. [online no manual original]
84. Ghebreyesus TA, Byass P, Witten KH, Getachew A, Haile M, Yohannes M, Lindsay SW. Appropriate Tools and Methods for Tropical Microepidemiology: a Case-study of Malaria Clustering in Ethiopia. *Ethiopian Journal of Health Development*. 17:1-8, 2003.
85. Brooker S, Clarke S, Njagi JK, Polack S, Mugo B, Estambale B, Muchiri E, Magnussen P, Cox J. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Medicine and International Health*, 9: 757-766, 2004.
86. Washington CH, Radday J, Streit TG, Boyd HA, Beach MJ, Addiss DG, Lovince R, Lovegrove MC, Lafontant JG, Lammie PJ, Hightower AW. Spatial clustering of filarial transmission before and after a Mass Drug Administration in a setting of low infection prevalence. *Filaria Journal*, 3:3, 2004. [online no manual original]
87. Gosselin PL, Lebel G, Rivest S, Fradet MD. The Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV): a real-time GIS for surveillance and decision-making. *International Journal of Health Geographics*, 4:21, 2005. [online no manual original]
88. Gaudart J, Poudiougou B, Ranque S, Doumbo O. Oblique decision trees for spatial pattern detection: optimal algorithm and application to malaria risk. *BMC Medical Research Methodology*, 5:22, 2005. [online no manual original]
89. Nisha V, Gad SS, Selvapandian D, Suganya V, Rajagopal V, Suganti P, Balraj V, Devasundaram J. Geographical information system (GIS) in investigation of an outbreak of dengue fever. *Journal of Communicable Diseases*, 37:39-43, 2005.
90. Reperant LA, Deplazes P. Cluster of *Capillaria hepatica* infections in non-commensal rodents from the canton of Geneva, Switzerland. *Parasitology Research*, 96:340-342, 2005.
91. Fang L, Yan L, Liang S, de Vlas SJ, Feng D, Han X, Zhao W, Xu B, Bian L, Yang H, Gong P, Richardus JH, Cao W. Spatial analysis of hemorrhagic fever with renal syndrome in China. *BMC Infectious Diseases*, 6:77, 2006. [online no manual original]
92. Bonilla RE. Distribución Espacio-Temporal de la Fiebre Dengue en Costa Rica. *Población y Salud en Mesoamérica*, 3:2:2, 2006. [online no manual original]
93. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, Diallo M, Diawara S, Ouattara A, Diakite M, Doumbo OK. Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health*, 6:286, 2006. [online no manual original]
94. Lian M, Warner RD, Alexander JL, Dixon KR. Using geographic information systems and spatial and space-time scan statistics for a population-based risk analysis of the 2002 equine West Nile epidemic in six contiguous regions of Texas. *International Journal of Health Geographics*, 6:42, 2007. [online no manual original]
95. Coleman M, Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malaria Journal*, 8:68, 2009. [online no manual original]

96. Mirghani SE, Nour BY, Bushra SM, Elhassan IM, Snow RW, Noor AM. The spatial-temporal clustering of *Plasmodium falciparum* infection over eleven years in Gezira State, The Sudan. *Malaria Journal*, 9:172, 2010. [online no manual original]
97. Haque U, Sunahara T, Hashizume M, Shields T, Yamamoto T, Haque R, Glass GE. Malaria prevalence, risk factors and spatial distribution in a hilly forest area of Bangladesh. *PLoS ONE* 6(4): e18908, 2011. [online no manual original]
98. Schmidt W-P, Suzuki M, Dinh Thiem V, White RG, Tsuzuki A, Yoshida LM, Yanai H, Haque U, Huu Tho L, Duc Anh D, Ariyoshi K. Population Density, Water Supply, and the Risk of Dengue Fever in Vietnam: Cohort Study and Spatial Analysis. *PLoS Medicine*, 8:8, e1001082, 2011. [online no manual original]
99. Winskill P, Rowland M, Mtove G, Malima RC, Kirby MJ. Malaria risk factors in north-east Tanzania. *Malaria Journal* 10:98, 2011. [online no manual original]
100. Washington CH, Radday J, Streit TG, Boyd HA, Beach MJ, Addiss DG, Lovince R, Lovegrove MC, Lafontant JG, Lammie PJ, Hightower AW. Spatial clustering of filarial transmission before and after a mass drug administration in a setting of low infection prevalence. *Filaria Journal*, 3: 3, 2004. [online no manual original]
101. Bhattarai NR, Van der Auwera G, Rijal S, Picado A, Speybroeck N, Khanal B, De Doncker S, Lal Das M, Ostyn B, Davies C, Coosemans M, Berkvens D, Boelaert M, Dujardin JC. Domestic animals and epidemiology of visceral leishmaniasis, Nepal. *Emerging Infectious Diseases*, 16:231-237, 2010. [online no manual original]
102. Cook J, Kleinschmidt I, Schwabe C, Nseng G, Bousema T, Corran PH, Riley EM, Drakeley CJ. Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, Equatorial Guinea. *PLoS One*, 6:e25137, 2011. [online no manual original]
103. Nourin AB, Abass MA, Nugud AH, El Hassan I, Snow RW, Noor AM. Identifying residual foci of *Plasmodium falciparum* infections for malaria elimination: the urban context of Khartoum, Sudan. *PLoS One*, 6:e16948, 2011. [online no manual original]
104. Rochlin I, Turbow D, Gomez F, Ninivaggi DV, Campbell SR. Predictive mapping of human risk for West Nile virus (WNV) based on environmental and socioeconomic factors. *PLoS One*. 6:e23280, 2011. [online no manual original]
105. Impoinvil DE, Solomon T, Schluter WW, Rayamajhi A, Bichha RP, Shakya G, Caminade C, Baylis M. The spatial heterogeneity between Japanese encephalitis incidence distribution and environmental variables in Nepal. *PLoS One*, 6:e22192, 2011. [online no manual original]
106. Bejon P, Turner L, Lavstsen T, Cham G, Olotu A, Drakeley CJ, Lievens M, Vekemans J, Savarese B, Lusingu J, von Seidlein L, Bull PC, Marsh K, Theander TG. Serological evidence of discrete spatial clusters of *Plasmodium falciparum* parasites. *PLoS One*, 6:e21711, 2011. [online no manual original]
107. Jones SG, Conner W, Song B, Gordon D, Jayakaran A. Comparing spatio-temporal clusters of arthropod-borne infections using administrative medical claims and state reported surveillance data. *Spatial and Spatio-Temporal Epidemiology*, 2012.
108. Sindato C, Karimuribo ED, Pfeiffer DU, Mboera LE, Kivaria F, Dautu G, Bernadrd B, Paweska JT. Spatial and temporal pattern of Rift Valley fever outbreaks in Tanzania; 1930 to 2007. *PloS One*, 9(2), e88897, 2014. [online no manual original]
109. Lal A, Hales S. Heterogeneity in hotspots: spatio-temporal patterns in neglected parasitic diseases. *Epidemiology and infection*, epub, 2014.

110. Bejon P, Williams TN, Nyundo C, Hay SI, Benz D, Gething PW, Otiende M, Peshu J, Bashraheil M, Greenhouse B, Bousema T, Bauni E, Marsh K, Smith DL, Borrmann S. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *eLife*, epub, 2014.
111. Mosha JF, Sturrock HJ, Greenwood B, Sutherland CJ, Gadalla NB, Atwal S, Hemelaar S, Brown JM, Drakeley C, Kibiki G, Bousema T, Chandramohan D, Gosling RD. Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. *Malaria Journal*, 13:53, 2014. [online no manual original]
112. Carson C, Lavender CJ, Handasyde KA, O'Brien CR, Hewitt N, Johnson PD, Fyfe JA. Potential wildlife sentinels for monitoring the endemic spread of human buruli ulcer in South-East Australia. *PLoS Neglected Tropical Diseases*, 8:e2668, 2014. [online no manual original]
113. Delgado-Ratto C, Soto-Calle VE, Van den Eede P, Gamboa D, Rosas A, Abatih EN, Rodriguez Ferrucci H, Llanos-Cuentas A, Van Geertruyden JP, Erhart A, D'Alessandro U. Population structure and spatio-temporal transmission dynamics of *Plasmodium vivax* after radical cure treatment in a rural village of the Peruvian Amazon. *Malaria Journal*, 13:8, 2014. [online no manual original]
114. Kracalik I, Malania L, Tsertsvadze N, Manvelyan J, Bakanidze L, Imnadze P, Tsanova S, Blackburn JK. Human cutaneous anthrax, Georgia 2010-2012. *Emerging Infectious Diseases*, 20:261-264, 2014. [online no manual original]
115. Liu C, Liu Q, Lin H, Xin B, Nie J. Spatial analysis of dengue fever in Guangdong Province, China, 2001-2006. *Asia Pacific Journal of Public Health*, 26:58-66, 2014.
116. Mulatti P, Mazzucato M, Montarsi F, Ciocchetta S, Capelli G, Bonfanti L, Marangon S. Retrospective space-time analysis methods to support West Nile virus surveillance activities. *Epidemiology and Infection*, epub, 2014.
117. Mollalo A, Alimohammadi A, Shirzadi MR, Malek MR. Geographic Information System-Based Analysis of the Spatial and Spatio-Temporal Distribution of Zoonotic Cutaneous Leishmaniasis in Golestan Province, North-East of Iran. *Zoonoses and Public Health*, epub, 2014.

Infecções Hospitalares

118. Huang SS, Yokoe DS, Stelling J, Placzek H, Kullendorff M, Kleinman K, O'Brien TF, Calderwood MS, Vostok J, Platt R. Automated detection of infectious disease outbreaks in hospitals: A Retrospective Cohort Study. *PLoS Medicine*, 7:e1000238, 2010. [online no manual original]
119. Vlek AL, Cooper BS, Kypraios T, Cox A, Edgeworth JD, Auguet OT. Clustering of antimicrobial resistance outbreaks across bacterial species in the intensive care unit. *Clinical Infectious Diseases*, 57:65-76, 2013. [online no manual original]
120. Faires MC. The Epidemiology of Methicillin-Resistant *Staphylococcus aureus* and *Clostridium difficile* in Community Hospitals. PhD Thesis, University of Guelph, 2013.

Outras Doenças Infecciosas

121. Cousens S, Smith PG, Ward H, Everington D, Knight RSG, Zeidler M, Stewart G, Smith-Bathgate EAB, Macleod MA, Mackenzie J, Will RG. Geographical distribution of variant Creutzfeldt-Jakob disease in Great Britain, 1994-2000. *The Lancet*, 357:1002-1007, 2001.
122. Huillard d'Aignaux J, Cousens SN, Delasnerie-Laupretre N, Brandel JP, Salomon D, Laplanche JL, Hauw JJ, Alperovitch A. Analysis of the geographical distribution of sporadic Creutzfeldt-Jakob disease in France between 1992 and 1998. *International Journal of Epidemiology*, 31: 490-495, 2002. [online no manual original]

123. Dreesman J, Scharlach H. Spatial-statistical analysis of infectious disease notification data in Lower Saxony. *Gesundheitswesen*, 66: 783-789, 2004.
124. Polack SR, Solomon AW, Alexander NDE, Massae PA, Safari S, Shao JF, Foster A, Mabey DC. The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99: 218-225, 2005.

Resistência Antimicrobiana

125. Ghosh AN, Bhatta DR, Ansari MT, Tiwari HK, Mathuria JP, Gaur A, Supram HS, Gokhale S. Application of WHONET in the Antimicrobial Resistance Surveillance of Uropathogens: A First User Experience from Nepal. *Journal of Clinical and Diagnostic Research*, 7:845-848,2013. [online no manual original]

Detecção Prospectiva em Tempo Real de Surtos de Doenças

126. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clustering as an early warning system for West Nile virus activity. *Emerging Infectious Diseases*, 9:641-646, 2003. [online no manual original]
127. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice: The New York City emergency department system. *Emerging Infectious Diseases*, 10:858-864, 2004. [online no manual original]
128. Reinhardt M, Elias J, Albert J, Frosch M, Harmsen D, Vogel U. EpiScanGIS: an online geographic surveillance system for meningococcal disease. *International Journal of Health Geographics*, 7:33, 2008. [online no manual original]
129. Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, Kulldorff M, Leach M, Nordin J, Patton-Levine J, Puga E, Sherwood E, Shui I, Platt R. Evaluating real-time syndromic surveillance signals from ambulatory care data in four states. *Public Health Reports*, 2010, 125:111-120.
130. Viñas M, Tuduri E, Galar A, Yih WK, Pichel M, Stelling J, Brengi S, Della Gaspera A, van der Ploeg C, Bruno S, Rogé A, Caffer M, , Kulldorff M, Galas M. Laboratory-based prospective surveillance for community outbreaks of *Shigella* spp. in Argentina. *PLoS Neglected Tropical Diseases*, 7:e2521, 2013. [online no manual original]
131. Yih WK, Cocoros NM, Crockett M, Klompas M, Kruskal BA, Kulldorff M, Lazarus R, Madoff LC, Morrison MJ, Smole S, Platt R. Automated influenza-like illness reporting, an efficient adjunct to traditional sentinel surveillance. *Public Health Reports*, 129:66-63, 2014.

Vigilância Sindrômica

132. Minnesota Department of Health. Syndromic Surveillance: A New Tool to Detect Disease Outbreaks. *Disease Control Newsletter*, 32:16-17, 2004. [online no manual original]
133. Kleinman K, Abrams A, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. *Epidemiology and Infection*, 2005, 133:409-419.
134. Nordin JD, Goodman MJ, Kulldorff M, Ritzwoller DP, Abrams AM, Kleinman K, Levitt MJ, Donahue J, Platt R. Simulated anthrax attacks and syndromic surveillance. *Emerging Infectious Diseases*, 2005, 11:1394-98. [online no manual original]
135. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City. *BMC Public Health*, 5:105, 2006. [online no manual original]

136. Horst MA, Coco AS. Observing the spread of common illnesses through a community: Using geographic information systems (GIS) for surveillance. *Journal of the American Board of Family Medicine*, 23:32-41, 2010. [online no manual original]
137. van den Wijngaard CC, van Asten L, van Pelt W, Doornbos G, Nagelkerke NJ, Donker GA, van der Hoek W, Koopmans MP. Syndromic surveillance for local outbreaks of lower-respiratory infections: would it work? *PLoS One*, 29:e10406, 2010. [online no manual original]

Câncer: Incidência, Prevalência e Mortalidade

138. Hjalmar U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine*, 15:707-715, 1996.
139. Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer in northeastern United States: A geographical analysis. *American Journal of Epidemiology*, 146:161-170, 1997. [online no manual original]
140. Imai J. Spatial disease clustering in Kochi prefecture in Japan. *National Institute of Public Health, Epidemiology and Biostatistics Research*, 57-96, 1998 (em Japonês).
141. VanEenwyk J, Bensley L, McBride D, Hoskins R, Solet D, McKeeman Brown A, Topiwala H, Richter A, Clark R. Addressing community health concerns around SeaTac Airport: Second Report. Washington State Department of Health, 1999. [online no manual original]
142. Hjalmar U, Kulldorff M, Wahlquist Y, Lannering B. Increased incidence rates but no space-time clustering of childhood malignant brain tumors in Sweden. *Cancer*, 85:2077-2090, 1999.
143. Viel JF, Arveux P, Baverel J, Cahn JY. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology*, 152:13-19, 2000.
144. New York State Department of Health. Cancer Surveillance Improvement Initiative, 2001. [online no manual original]
145. Jemal A, Kulldorff M, Devesa SS, Hayes RB, Fraumeni JF. A geographic analysis of prostate cancer mortality in the United States. *International Journal of Cancer*, 101:168-174, 2002.
146. Michelozzi P, Capon A, Kirchmayer U, Forastiere F, Biggeri A, Barca A, Perucci CA. Adult and childhood leukemia near a high-power radio station in Rome, Italy. *American Journal of Epidemiology*, 155:1096-1103, 2002.
147. Zhan FB, Lin H. Geographic patterns of cancer mortality clusters in Texas, 1990 to 1997. *Texas Medicine*, 99:58-64, 2003.
148. Buntinx F, Geys H, Lousbergh D, Broeders G, Cloes E, Dhollander D, Op De Beeck L, Vanden Brande J, Van Waes A, Molenberghs G. Geographical differences in cancer incidence in the Belgian province of Limburg. *European Journal of Cancer*, 39:2058-72, 2003.
149. Santamaria Ulloa C. Evaluación de alarmas por cáncer utilizando análisis espacial: una aplicación para Costa Rica. *Reivista Costarricense de Salud Pública*, 12:18-22, 2003. [online no manual original]
150. Sheehan TJ, DeChello LM, Kulldorff M, Gregorio DI, Gershman S, Mrosczyk M. The geographic distribution of breast cancer incidence in Massachusetts 1988-1997, adjusted for covariates. *International Journal of Health Geographics*, 2004, 3:17. [online no manual original]
151. Fang Z, Kulldorff M, Gregorio DI. Brain cancer in the United States, 1986-95: A geographic analysis. *Neuro-Oncology*, 2004, 6:179-187.
152. Hsu CE, Jacobson HE, Soto Mas F. Evaluating the disparity of female breast cancer mortality among racial groups - a spatiotemporal analysis. *International Journal of Health Geographics* 3:4, 2004. [online no manual original]

153. Han DW, Rogerson PA, Nie J, Bonner MR, Vena JE, Vito D, Muti P, Trevisan M, Edge SB, Freudenheim JL. Geographic clustering of residence in early life and subsequent risk of breast cancer (United States). *Cancer Causes and Control*, 15:921-929, 2004.
154. Campo J, Comber H, Gavin A T. All-Ireland Cancer Statistics 1998-2000. Northern Ireland Cancer Registry / National Cancer Registry, 2004. [online no manual original]
155. Hayran M. Analyzing factors associated with cancer occurrence: A geographical systems approach. *Turkish Journal of Cancer*, 34:67-70, 2004. [online no manual original]
156. Fukuda Y, Umezaki M, Nakamura K, Takano T. Variations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan. *International Journal of Health Geographics*, 4:16, 2005. [online no manual original]
157. Ozonoff A, Webster T, Vieira V, Weinberg J, Ozonoff D, Aschengrau A. Cluster detection methods applied to the Upper Cape Cod cancer data. *Environmental Health: A Global Access Science Source*, 4:19, 2005. [online no manual original]
158. DeChello LM, Sheehan TJ. The geographic distribution of melanoma incidence in Massachusetts, adjusted for covariates. *Int J Health Geogr*. 2006;5:31 [online no manual original]
159. Gregorio DI, Samociuk H, DeChello L, Swede H. Effects of study area size on geographic characterizations of health events: prostate cancer incidence in Southern New England, USA, 1994–1998. *Int J Health Geogr*. 5:8, 2006. [online no manual original]
160. Chen Y, Yi Q, Mao Y. Cluster of liver cancer and immigration: a geographic analysis of incidence data for Ontario 1998–2002. *Int J Health Geogr*. 7:28, 2008. [online no manual original]
161. Lorenzo-Luaces Alvarez P, Guerra-Yi ME, Faes C, Galán Alvarez Y, Molenberghs G. Spatial analysis of breast and cervical cancer incidence in small geographical areas in Cuba, 1999-2003. *European Journal of Cancer Prevention*, 18:395-403, 2009.
162. Amin R, Bohnert A, Holmes L, Rajasekaran A, Assanasen C. Epidemiologic mapping of Florida childhood cancer clusters. *Pediatric Blood Cancer*, 54:511-518, 2010.
163. Liu-Mares W, MacKinnon JA, Sherman R, Fleming LE, Rocha-Lima C, Hu, JJ, Lee DJ. Pancreatic cancer clusters and arsenic-contaminated drinking water wells in Florida. *BMC Cancer*, 13, 111, 2013. [online no manual original]
164. Baastrup Nordsborg R, Meliker JR, Kjær Ersbøll A, Jacquez GM, Raaschou-Nielsen O. Space-Time Clustering of Non-Hodgkin Lymphoma Using Residential Histories in a Danish Case-Control Study. *PLoS One*, 8, e60800, 2013. [online no manual original]
165. Amin R, Hendryx M, Shull M, Bohnert A. A cluster analysis of pediatric cancer incidence rates in Florida: 2000-2010. *Statistics in Public Policy*, 1:69-77, 2014. [online no manual original]

Câncer: Detecção Precoce versus Tardia, Estágio e Grau

166. Roche LM, Skinner R, Weinstein RB. Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *Journal of Public Health Management and Practice*, 8:26-32, 2002.
167. Thomas AJ, Carlin BP. Late detection of breast and colorectal cancer in Minnesota counties: an application of spatial smoothing and clustering. *Statistics in Medicine*, 22:113-127, 2003.
168. Sheehan TJ, DeChello LM. A space-time analysis of the proportion of late stage breast cancer in Massachusetts, 1988 to 1997. *International Journal of Health Geographics*, 4:15, 2005. [online no manual original]

169. Klassen A, Curriero F, Kulldorff M, Alberg AJ, Platz EA, Neloms ST. Missing stage and grade in Maryland prostate cancer surveillance data, 1992-1997. *American Journal of Preventive Medicine*, 30:S77-87, 2006. [online no manual original]
170. Pollack LA, Gotway CA, Bates JH, Parikh-Patel A, Richards TB, Seeff LC, Hodges H, Kassim S. Use of the spatial scan statistic to identify geographic variations in late stage colorectal cancer in California (United States). *Cancer Causes and Control*, 17:449-457, 2006.
171. DeChello LM, Sheehan TJ. Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995-1998. *Int J Health Geogr.* 2007;6:20. [online no manual original]

Câncer: Rastreamento, Tratamento e Sobrevida

172. Sheehan TJ, Gershman ST, MacDougal L, Danley R, Mrosszczyk M, Sorensen AM, Kulldorff M. Geographical surveillance of breast cancer screening by tracts, towns and zip codes. *Journal of Public Health Management and Practice*, 6: 48-57, 2001.
173. Gregorio DI, Kulldorff M, Barry L, Samociuk H, Zarfos K. Geographic differences in primary therapy for early stage breast cancer. *Annals of Surgical Oncology*, 2001; 8:844-849, 2001. [online no manual original]
174. Henry KA, Niu X, Boscoe FP. Geographic disparities in colorectal cancer survival. *International Journal of Health Geographics* 2009, 8:48. [online no manual original]

Doenças Cardiovasculares

175. Kuehl KS, Loffredo CA. A cluster of hypoplastic left heart malformation in Baltimore, Maryland *Pediatric Cardiology*, 27:25-31, 2006.
176. Li XY, Chen K. Scan statistic theory and its application in spatial epidemiology (in Chinese). *Zhonghua Liu Xing Bing Xue Za Zhi.*, 29:828-31, 2008.

Doenças Reumatológicas e Auto-Imunes

177. Walsh SJ, Fenster JR. Geographical clustering of mortality from systemic sclerosis in the Southeastern United States, 1981-90. *Journal of Rheumatology*, 24:2348-2352, 1997.
178. Walsh SJ, DeChello LM. Geographical variation in mortality from systemic lupus erythematosus in the United States. *Lupus*, 10:637-646, 2001.
179. López-Abente G, Morales-Piga A, Bachiller-Corral FJ, Illera-Martín O, Martín-Domenech R, Abraira V. Identification of possible areas of high prevalence of Paget's disease of bone in Spain. *Clinical and Experimental Rheumatology*, 21:635-368, 2003.
180. Donnan PT, Parratt JDE, Wilson SV, Forbes RB, O'Riordan JI, Swingler RJ. Multiple sclerosis in Tayside, Scotland: detection of clusters using a spatial scan statistic. *Multiple Sclerosis*, 11:403-408, 2005.

Doenças Hepáticas

181. Ala A, Stanca CM, Bu-Ghanim M, Ahmado I, Branch AD, Schiano TD, Odin JA, Bach N. Increased prevalence of primary biliary cirrhosis near superfund toxic waste sites. *Hepatology*, 43:525-531, 2006.
182. Stanca CM, Babar J, Singal V, Ozdenerol E, Odin JA. Pathogenic role of environmental toxins in immune-mediated liver diseases. *Journal of Immunotoxicology*, 5:59-68, 2008.

183. McNally RJQ, Ducker S, James OFW. Are Transient Environmental Agents Involved in the Cause of Primary Biliary Cirrhosis? Evidence from Space-Time Clustering Analysis. *Hepatology*, 50:1169-1174, 2009.

Diabetes

184. Green C, Hoppa RD, Young TK, Blanchard JF. Geographic analysis of diabetes prevalence in an urban area. *Social Science and Medicine*, 57:551-560, 2003.
185. Aamodt G, Stene LC, Njølstad PR, Søvik O, Joner G, for the Norwegian Childhood Diabetes Study Group. Spatiotemporal trends and age-period-cohort modelling of the incidence of type 1 diabetes among children ages <15 years in Norway 1973-1982 and 1989-2003. *Diabetes Care*, 30:884-889, 2007.

Alergia e Asma

186. Yiannakoulis N, Schopflocher DP, Svenson LW. Using administrative data to understand the geography of case ascertainment. *Chronic Diseases in Canada*, 30:20-28, 2009.[online no manual original]
187. Lee H, Kim GS. Geographical and sociodemographic risk factors for allergic diseases in Korean children. *Asian Nursing Research*, 5:1-10, 2011.
188. Torabi M. Spatial disease cluster detection: An application to childhood asthma in Manitoba, Canada. *Journal of Biometrics and Biostatistics*, 57:010, 2012 [online no manual original]

Doenças Neurológicas

189. Sabel CE, Boyle PJ, Löytönen M, Gatrell AC, Jokelainen M, Flowerdew R, Maasilta P. Spatial clustering of amyotrophic lateral sclerosis in Finland at place of birth and place of death. *American Journal of Epidemiology*, 157: 898-905, 2003.

Malformações e Outros Desfechos Congênitos

190. Kharrazi M, et al. Pregnancy outcomes around the B.K.K. landfill, West Covina, California: An analysis by address. California Department of Health Services, 1998.
191. Bell S. Spatial Analysis of Disease - Applications. In Beam C (ed). *Biostatistical Applications in Cancer Research*. Boston: Kluwer p151-182, 2002. [online no manual original]
192. Forand SP, Talbot TO, Druschel C, Cross PK. Data quality and the spatial analysis of disease rates: congenital malformations in New York State. *Health and Place*, 8:191-199, 2002.
193. Colorado Department of Public Health and Environment. Analysis of birth defect data in the vicinity of the Redfield plume area in southeastern Denver county: 1989-1999. Colorado Department of Public Health and the Environment, 2002. [online no manual original]
194. Boyle E, Johnson H, Kelly A, McDonnell R. Congenital anomalies and proximity to landfill sites. *Irish Medical Journal*, 97:16-18, 2004.
195. Ozdenerol E, Williams BL, Kang SY, Magumbol MS. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *International Journal of Health Geographics*, 4:19, 2005. [online no manual original]
196. Viel JF, Floret N, Mauny F. Spatial and space-time scan statistics to detect low clusters of sex ratio. *Environmental and Ecological Statistics*, 12:289-299, 2005.

197. Grady SC, Enander H. Geographic analysis of low birthweight and infant mortality in Michigan using automated zoning methodology International Journal of Health Geographics 2009, 8:10. [online no manual original]

Pediatria

198. George M, Wiklund L, Aastrup M, Pousette J, Thunholm B, Saldeen T, Wernroth L, Zaren B, Holmberg L. Incidence and geographical distribution of sudden infant death syndrome in relation to content of nitrate in drinking water and groundwater levels. European Journal of Clinical Investigation, 31: 1083-1094, 2001.
199. Sankoh OA, Ye Y, Sauerborn R, Muller O, Becher H. Clustering of childhood mortality in rural Burkina Faso. International Journal of Epidemiology, 30:485-492, 2001. [online no manual original]
200. Ali M, Asefaw T, Byass P, Beyene H, Karup Pedersen F. Helping northern Ethiopian communities reduce childhood mortality: population-based intervention trial. Bulletin of the World Health Organization. 83:27-33, 2005. [online no manual original]
201. Awini E, Mattah P, Sankoh O, Gyapong M. Spatial variations in childhood mortalities at the Dodowa Health and Demographic Surveillance System site of the INDEPTH Network in Ghana. Tropical Medicine and International Health, 2010.

Geriatrics

202. Yiannakoulis N, Rowe BH, Svenson LW, Schopflocher DP, Kelly K, Voaklander DC. Zones of prevention: the geography of fall injuries in the elderly. Social Science and Medicine, 57:2065-73, 2003.
203. Vaneckova P, Beggs PJ, Jacobson CR. Spatial analysis of heat-related mortality among the elderly between 1993 and 2004 in Sydney, Australia. Social Science and Medicine, 70:293-304, 2010.

Psicologia

204. Margai F, Henry N. A community-based assessment of learning disabilities using environmental and contextual risk factors. Social Science and Medicine, 56: 1073-1085, 2003.

Imagem Cerebral

205. Yoshida M, Naya Y, Miyashita Y. Anatomical organization of forward fiber projections from area TE to perirhinal neurons representing visual long-term memory in monkeys. Proceedings of the National Academy of Sciences of the United States of America, 100:4257-4262, 2003. [online no manual original]

Fármacos e Vacinas

206. Copeland KR, Allen AE, Basic Models for Mapping Prescription Drug Data. Proceedings of the ASA Section on Survey Research Methods, 2904-2909, 2005. [online no manual original]
207. Moore GE, Ward MP, Kulldorff M, Caldanaro RJ, Guptill LF, Lewis HB, Glickman LT. A space-time cluster of adverse events associated with a canine rabies vaccine. Vaccine, 23:5557-62, 2005.
208. Omer SB, Enger S, Moulton LH, Halsey NA, Stokley S, Salmon DA. Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. American Journal of Epidemiology, 168:1389-1396, 2008. [online no manual original]

209. Penfold RB, Wang W, Pajer K, Strange B, Kelleher KJ. Spatio-temporal clusters of new psychotropic medications among Michigan children insured by Medicaid. *Pharmacoepidemiology and Drug Safety*, 18: 531–539, 2009.
210. Brownstein JS, Green TC, Cassidy TA, Butler SF. Geographic information systems and pharmacoepidemiology: using spatial cluster detection to monitor local patterns of prescription opioid abuse. *Pharmacoepidemiology and Drug Safety*, 19:627-637, 2010.
211. King M, Essic C. The geography of antidepressant, antipsychotic, and stimulant utilization in the United States, *Health and Place*, 20:32-38, 2013.
212. Atwell JE, Van Otterloo J, Zipprich J, Winter K, Harriman K, Salmon DA, Halsey NA, Omer SB. Nonmedical vaccine exemptions and pertussis in California 2010. *Pediatrics*, 132:624-630, 2013. [online no manual original]

Álcool, Tabaco e Drogas Recreativas

213. Hanson CE, Wieczorek WF. Alcohol mortality: a comparison of spatial clustering methods. *Social Science and Medicine*, 55:791-802, 2002.
214. Sudakin D, Power LE. Regional and temporal variation in methamphetamine-related incidents: applications of spatial and temporal scan statistics. *Clinical Toxicology*, 47:243-247, 2009.
215. Chong S, Nelson M, Byun R, Harris L, Eastwood J, Jalaludin B. Geospatial analyses to identify clusters of adverse antenatal factors for targeted interventions. *International Journal of Health Geographics*, 2013, 12:46. [online no manual original]

Obesidade

216. Dahly D. Obesity clustering in Cebu, Philippines: an application of satscan and the spatial scan statistic *Journal of Epidemiology and Community Health*, 65, A71, 2011. [online no manual original]
217. Chalkias C, Papadopoulos AG, Mpenekos G, Tambalis K, Psarra G, Sidossis L. Spatial variability of childhood obesity in response to socioeconomic heterogeneity. The case of Athens metropolitan area, Greece. *Proceedings of the 17th European Colloquium on Quantitative and Theoretical Geography*, 605-611, 2011. [online no manual original]

Esportes e Recreações

218. López FA, Martínez JA, Ruiz M. Análisis espacial de lanzamientos en baloncesto; el caso de L.A. Lakers / Spatial pattern analysis of shot attempts in basketball; The case of LA Lakers. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 13: 585-613, 2013. [online no manual original]
219. Schmicker RH. An application of SaTScan to evaluate the spatial distribution of corner kick goals in major league soccer. *International Journal of Computer Science in Sport*, 12:70-79, 2013.

Acidentes e Suicídio

220. Nkhoma ET, Hsu CE, Hunt VI, Harris AM. Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980 - 2001. *International Journal of Health Geographics*, 3:25, 2004. [online no manual original]
221. Exeter DJ, Boyle PJ. Does young adult suicide cluster geographically in Scotland? *Journal of Epidemiology and Community Health*, 61:731-736, 2007.

222. Warden R. Comparison of Poisson and Bernoulli spatial cluster analyses of pediatric injuries in a fire district. *International Journal of Health Geographics*, 7:51, 2008. [online no manual original]
223. Mesoudi A. The cultural dynamics of copycat suicide. *PLoS One*, 4:e7252, 2009. [online no manual original]
224. Dey AN, Hicks P, Benoit S, Tokars JJ. Automated monitoring of clusters of falls associated with severe winter weather using the BioSense system. *Injury Prevention*, 16:403-407, 2010.
225. Saman DM, Cole HP, Odoi A, Myers ML, Carey DI, Westneat SC. A spatial cluster analysis of tractor overturns in Kentucky from 1960 to 2002. *PLoS One*, 7:e30532, 2012. [online no manual original]
226. Amin R, Ritter EK, Cossette L. A Geospatial Analysis of Shark Attack Rates for the Coast of California: 1994–2010. *Journal of Environment and Ecology*, 3:246-255, 2012.
227. Fuchs S, Ornetsmüller C, Totschnig R. Spatial scan statistics in vulnerability assessment – an application to mountain hazards. *Natural Hazards* 64:2129-2151, 2012.
228. Campo J. Firearm deaths in Washington State. Washington State Health Services Research Brief No. 71, 2013. [online no manual original]
229. Jones P, Gunnell D, Platt S, Scourfield J, Lloyd K, Huxley P, John A, Kamran B, Wells C, Dennis M. Identifying Probable Suicide Clusters in Wales Using National Mortality Data, *PLoS One*, 8:e71713, 2013 [online no manual original]

Demografia

230. Collado Chaves A. Fecundidad adolescente en el gran área metropolitana de Costa Rica. *Población y Salud en Mesoamérica*, 1:4, 2003. [online no manual original]

Medicina Veterinária, Animais Domésticos

231. Norström M, Pfeiffer DU, Jarp J. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds. *Preventive Veterinary Medicine*, 47: 107-119, 2000.
232. Ward MP. Blowfly strike in sheep flocks as an example of the use of a time-space scan statistic to control confounding. *Preventive Veterinary Medicine*, 49: 61-69, 2001.
233. United States Department of Agriculture. West Nile virus in equids in the Northeastern United States in 2000. USDA, APHIS, Veterinary Services, 2001. [online no manual original]
234. Doherr MG, Hett AR, Rufenacht J, Zurbriggen A, Heim D. Geographical clustering of cases of bovine spongiform encephalopathy (BSE) born in Switzerland after the feed ban. *Veterinary Record*, 151: 467-472, 2002.
235. Perez AM, Ward MP, Torres P, Ritacco V. Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. *Preventive Veterinary Medicine*, 56: 63-74, 2002.
236. Schwermer H, Rufenacht J, Doherr MG, Heim D. Geographic distribution of BSE in Switzerland. *Schweizer Archiv für Tierheilkunde*, 144:701-708, 2002.
237. Ward MP. Clustering of reported cases of leptospirosis among dogs in the United States and Canada. *Preventive Veterinary Medicine*, 56:215-226, 2002.
238. Falconi F, Ochs H, Deplazes P. Serological cross-sectional survey of psoroptic sheep scab in Switzerland. *Veterinary Parasitology*, 109:119-127, 2002.
239. Knuesel R, Segner H, Wahli T. A survey of viral diseases in farmed and feral salmonids in Switzerland. *Journal of Fish Diseases*, 26:167-182, 2003.

240. Berke O, Grosse Beilage E. Spatial relative risk mapping of pseudorabies-seropositive pig herds in an animal-dense region. *Journal of Veterinary Medicine*, B50: 322–325, 2003.
241. Abrial D, Calavas D, Lauvergne N, Morignat E, Ducrot C. Descriptive spatial analysis of BSE in western France. *Veterinary Research*, 34:749-60, 2003.
242. Sheridan HA, McGrath G, White P, Fallon R, Shoukri MM, Martin SW. A temporal-spatial analysis of bovine spongiform encephalopathy in Irish cattle herds, from 1996 to 2000. *Canadian Journal of Veterinary Research*, 69:19-25, 2005. [online no manual original]
243. Guerin MT, Martin SW, Darlington GA, Rajic A. A temporal study of *Salmonella* serovars in animals in Alberta between 1990 and 2001. *Canadian Journal of Veterinary Research*, 69:88-89, 2005. [online no manual original]
244. Allepuz A, López-Quílez A, Forte A, Fernández G, Casal J. Spatial analysis of bovine spongiform encephalopathy in Galicia, Spain (2000-2005). *Preventive Veterinary Medicine*, 79:174-85, 2007.
245. Heres L, Brus DJ, Hagenaars TJ. Spatial analysis of BSE cases in the Netherlands. *BMC Veterinary Research*, 4:21, 2008. [online no manual original]
246. Frossling J, Nodtvedt A, Lindberg A, Björkman C. Spatial analysis of *Neospora caninum* distribution in dairy cattle from Sweden. *Geospatial Health*, 3:39-45, 2008.

Medicina Veterinária, Vida Selvagem

247. Smith KL, DeVos V, Bryden H, Price LB, Hugh-Jones ME, Keim P. *Bacillus anthracis* diversity in Kruger National Park. *Journal of Clinical Microbiology*, 38:3780-3784, 2000. [online no manual original]
248. Berke O, von Keyserlingk M, Broll S, Kreienbrock L. On the distribution of *Echinococcus multilocularis* in red foxes in Lower Saxony: identification of a high risk area by spatial epidemiological cluster analysis. *Berliner und Munchener Tierarztliche Wochenschrift*. 115:428-434, 2002.
249. Miller MA, Gardner IA, Kreuder C, Paradies DM, Worcester KR, Jessup DA, Dodd E, Harris MD, Ames JA, Packham AE, Conrad PA. Coastal freshwater runoff is a risk factor for *Toxoplasma gondii* infection of southern sea otters (*Enhydra lutris nereis*). *International Journal for Parasitology*, 32:997-1006, 2002.
250. Hoar BR, Chomel BB, Rolfe DL, Chang CC, Fritz CL, Sacks BN, Carpenter TE. Spatial analysis of *Yersinia pestis* and *Bartonella vinsonii* subsp *berkhoffii* seroprevalence in California coyotes (*Canis latrans*). *Preventive Veterinary Medicine*, 56:299-311, 2003.
251. Olea-Popelka FJ, Griffin JM, Collins JD, McGrath G, Martin SW. Bovine tuberculosis in badgers in four areas in Ireland: does tuberculosis cluster? *Preventive Veterinary Medicine*, 59:103-111, 2003.
252. Joly DO, Ribic CA, Langenberg JA, Beheler K, Batha CA, Dhuey BJ, Rolley RE, Bartelt G, Van Deelen TR, Samual MD. Chronic wasting disease in free-ranging Wisconsin white-tailed deer. *Emerging Infectious Disease*, 9: 599-601, 2003. [online no manual original]
253. Miller MA, Grigg ME, Kreuder C, James ER, Melli AC, Crosbie PR, Jessup DA, Boothroyd JC, Brownstein D, Conrad PA. An unusual genotype of *Toxoplasma gondii* is common in California sea otters (*Enhydra lutris nereis*) and is a cause of mortality. *International Journal for Parasitology*, 34:275-284, 2004.
254. Olea-Popelka FJ, Flynn O, Costello E, McGrath G, Collins JD, O’Keeffe JO, Kelton DF, Berke O, Martin SW. Spatial relationship between *Mycobacterium bovis* strains in cattle and badgers in four areas in Ireland. *Preventive Veterinary Medicine*, 71:57-70, 2005.

Mamologia

255. Webb NF, Hebblewhite M, Merrill EH. Statistical Methods for Identifying Wolf Kill Sites Using Global Positioning System Locations. *Journal of Wildlife Management*, 2008, 72, 798-807.
256. McPhee HM, Webb NF, Merrill EH. Hierarchical predation: Wolf (*Canis lupus*) selection along hunt paths and at kill sites. *Canadian Journal of Zoology*, 2012, 90:555-563.
257. Ouko EO. Where, when and why are there elephant poaching hotspots in Kenya? MSc Thesis, University of Twente, Netherlands, 2013. [online no manual original]

Entomologia

258. Porcasi X, Catalá SS, Hrellac H, Scavuzzo MC, Gorla DE. Infestation of Rural Houses by *Triatoma Infestans* (Hemiptera: Reduviidae) in Southern Area of Gran Chaco in Argentina. *Journal of Medical Entomology*, 43:1060-1067, 2006.

Ictiologia

259. Spindler BD, Chipps SR, Klumb RA, Wimberly MC. Spatial analysis of pallid sturgeon *Scaphirhynchus albus* distribution in the Missouri River, South Dakota. *Journal of Applied Ichthyology*, 25:8-13, 2009.

Botânica

260. Bayon C, Pei MH, Ruiz C, Hunter T. Genetic structure and spatial distribution of the mycoparasite *Sphaerellopsis filum* on *Melampsora larici-epitea* in a short-rotation coppice willow planting. *Plant Pathology*, 56:616-623, 2007.

Silvicultura

261. Coulston JW, Riitters KH. Geographic analysis of forest health indicators Using Spatial Scan Statistics. *Environmental Management*, 31: 764-773, 2003.
262. Riitters KH, Coulston JW. Hot spots of perforated forest in the eastern United States. *Environmental Management*, 35:483-492, 2005.
263. Tuia D, Ratle F, Lasaponara R, Telesca L, Kanevski M. Scan statistics analysis of forest fire clusters. *Communications in Nonlinear Sciences and Numerical Simulations*, 13:1689-94, 2008.
264. Tonini M, Tuia D, Ratle F. Detection of clusters using space-time scan statistics. *International Journal of Wildland Fire*, 18 830-836, 2009.
265. Fei S. Applying hotspot detection methods in forestry: A case study of Chestnut Oak regeneration. *International Journal of Forestry Research.*, 815292, 2010. [online no manual original]
266. Vega Orozco C, Tonini M, Conedera M, Kanveski M. Cluster recognition in spatial-temporal sequences: the case of forest fires. *Geoinformatica*, 16: 653-673, 2012. [online no manual original]

Meio Ambiente

267. Vadrevu KP. Analysis of fire events and controlling factors in eastern India using spatial scan and multivariate statistics. *Geografiska Annaler*, 90A: 315-328, 2008.

268. Sudakin DL, Horowitz Z, Giffin S. Regional variation in the incidence of symptomatic pesticide exposures: Applications of geographic information systems. *Journal of Toxicology - Clinical Toxicology*, 40:767-773, 2002.
269. Krolik J, Maier A, Evans G, Belanger P, Hall G, Joyce A, Majury A. A spatial analysis of private well water *Escherichia coli* contamination in southern Ontario. *Geospatial Health* 8:65-75, 2013.
270. Krolik J, Evans G, Belanger P, Maier A, Hall G, Joyce A, Guimont S, Pelot A, Majury A. Microbial source tracking and spatial analysis of *E. coli* contaminated private well waters in southeastern Ontario. *Journal of Water and Health*, 12:348-357, 2014.

Geologia

271. Gao J, Zhang Z, Hu Y, Bian J, Jiang W, Wang X. Geographical distribution patterns of iodine in drinking-water and its associations with geological factors in Shandong Province, China. [online no manual original]

Desastres Naturais

272. Witham CS, Oppenheimer C. Mortality in England during the 1783-4 Laki Craters eruption. *Bulletin of Volcanology*, 67:15-25, 2004.
273. Stevenson JR Emrich CT Mitchell JT, Cutter SL. Using building permits to monitor disaster recovery: A spatio-temporal case study of coastal Mississippi following hurricane Katrina, *Cartography and Geographic Information Science*, 37:S57-68, 2010.

Guerra

274. Ziemke J. From battles to massacres. 3rd Annual Harvard-Yale-MIT Graduate Student Conference on Order, Conflict and Violence, 2008. [online no manual original]
275. O'Loughlin J, Witmer F, Linke A. The Afghanistan-Pakistan Wars 2008–2009: Micro-geographies, Conflict Diffusion, and Clusters of Violence. *Eurasian Geography and Economics*, 2010, 51, 437-71. [online no manual original]
276. O'Loughlin J, Witmer FDW, Linke AM, Thorwardson N. Peering into the Fog of War: The Geography of the WikiLeaks Afghanistan War Logs, 2004–2009. *Eurasian Geography and Economics*, 51:472–495, 2010. [online no manual original]
277. O'Loughlin J, Witmer FDW, The Localized Geographies of Violence in the North Caucasus of Russia, 1999-2007', *Annals of the Association of American Geographers*, 101: 178-201, 2011. [online no manual original]

Criminologia

278. Jefferis ES. A multi-method exploration of crime hot spots: SaTScan results. National Institute of Justice, Crime Mapping Research Center, 1998.
279. Kaminski RJ, Jefferis ES, Chanhatasilpa C. A spatial analysis of American police killed in the line of duty. In Turnbull et al. (eds.), *Atlas of crime: Mapping the criminal landscape*. Phoenix, AZ: Oryx Press, 2000.
280. LeBeau JL. Demonstrating the analytical utility of GIS for police operations: A final report. National Criminal Justice Reference Service, 2000. [online no manual original]

281. Beato Filho CC, Assunção RM, Silva BF, Marinho FC, Reis IA, Almeida MC. Homicide clusters and drug traffic in Belo Horizonte, Minas Gerais, Brazil from 1995 to 1999. *Cadernos de Saúde Pública*, 17:1163-1171, 2001. [online no manual original]
282. Ceccato V, Haining R. Crime in border regions: The Scandinavian case of Öresund, 1998-2001. *Annals of the Association of American Geographers*, 94:807-826, 2004.
283. Ceccato V. Homicide in Sao Paulo, Brazil: Assessing the spatial-temporal and weather variations. *Journal of Environmental Psychology*, 25:307-321, 2005.
284. Minamisava R, Nouer SS, de Moraes Neto OL, Melo LK, Andrade ALS. Spatial clusters of violent deaths in a newly urbanized region of Brazil: Highlighting the social disparities. *International Journal of Health Geographics*, 8:66, 2009. [online no manual original]
285. Nakaya T, Yano K. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14:223-239, 2010.
286. Leitner M, Helbich M. The Impact of Hurricanes on Crime: A Spatio-temporal Analysis in the City of Houston, TX. *Cartography and Geographic Information Science*, 37:214-222, 2011.
287. Zeoli AM, Pizarro JM, Grady SC, Melde C. Homicide as infectious disease: Using public health methods to investigate the diffusion of homicide. *Justice Quarterly*, 31:609-632, 2014.

Planejamento Urbano e Rural

288. Huang L, Stinchcomb DG, Pickle LW, Dill J, Berrigan D. Identifying clusters of active transportation using spatial scan statistics. *American Journal of Preventive Medicine*, 37:157-166, 2009.
289. Helbich M. Beyond postsuburbia? Multifunctional service agglomeration in Vienna's urban fringe. *Journal of Economic and Social Geography*, 2011.

Arquitetura

290. Kaza N, Lester TW, Rodriguez DA. The spatio-temporal clustering of green buildings in the United States, *Urban Studies*, 50:3262-3282, 2013. [online no manual original]

História e Arqueologia

291. Usher BM, Allen KL. Identifying kinship clusters: SaTScan for genetic spatial analysis. *American Journal of Physical Anthropology, Supplement*, 126,S40,210, 2005.
292. Wang F, Hartmann J, Luo W, Huang P. GIS-based spatial analysis of Tai place names in southern China: An exploratory study of methodology. *Annals of GIS*, 12:1-9, 2006. [online no manual original]

Astronomia

293. Marcos RDLF, Marcos CDLF. From star complexes to the field: Open cluster families, 672:342-351, 2008.
294. Bidin CM, Marcos RD, Marcos CD, Carraro, G. Not an open cluster after all: the NGC 6863 asterism in Aquila. *Astronomy and Astrophysics*, 510:A44, 2010.[online no manual original]

Tópicos Relacionados: *Artigos Metodológicos, Bibliografia do SaTScan, Citações Sugeridas.*

Outras Referências Mencionadas no Manual do Usuário

295. Alt KW, Vach W. The reconstruction of 'genetic kinship' in prehistoric burial complexes - problems and statistics. In Bock HH, Ihm P (eds): Classification, data analysis, and knowledge organization. Berlin: Springer Verlag, 1991.
296. Baker RD. Testing for space-time clusters of unknown size. *Journal of Applied Statistics*, 23:543-554, 1996.
297. Besag J, Newell J. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society*, A154:143-155, 1991.
298. Bithell JF. The choice of test for detecting raised disease risk near a point source. *Statistics in Medicine*, 14:2309-2322, 1995.
299. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society*, B52:73-104, 1990.
300. Diggle PJ, Chetwynd AD. Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, 47:1155-1163, 1991.
301. Diggle P, Chetwynd AG, Häggkvist R, Morris SE. Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124-136, 1995.
302. Glaz J, Balakrishnan N (editors). *Scan Statistics and Applications*. Birkhäuser: Boston, 1999.
303. Glaz J, Naus JI, Wallenstein S. *Scan Statistics*. Springer Verlag: New York, 2001.
304. Glaz J, Pozdnyakov V, Wallenstein S. *Scan Statistics: Theory and Applications*. Birkhäuser: Boston, 2009.
305. Grimson RC. A versatile test for clustering and a proximity analysis of neurons. *Methods of Information in Medicine*, 30:299-303, 1991.
306. Jacquez GM. A k nearest neighbor test for space-time interaction. *Statistics in Medicine*, 15:1935-1949, 1996.
307. Knox G. The detection of space-time interactions. *Applied Statistics*, 13:25-29, 1964.
308. Kulldorff M. *Statistical Methods for Spatial Epidemiology: Tests for Randomness*, in GIS and Health in Europe, Löytönen M and Gatrell A (eds), London: Taylor & Francis, 1998.
309. Kulldorff M, Hjalmars U. The Knox method and other tests for space time interaction. *Biometrics*, 9:621-630, 1999.
310. Lawson AB. On the analysis of mortality events associated with a pre-specified fixed point. *Journal of the Royal Statistical Society, Series A*, 156:363-377, 1993.
311. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209-220, 1967.
312. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika*, 37:17-23, 1950.
313. Naus J. The distribution of the size of maximum cluster of points on the line. *Journal of the American Statistical Association*, 60:532-538, 1965.
314. Openshaw S, Charlton M, Wymer C, Craft AW: A mark 1 analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1, 335-358, 1987.

315. Ranta J, Pitkniemi J, Karvonen M, et al. Detection of overall space-time clustering in non-uniformly distributed population. *Statistics in Medicine*, 15:2561-2572, 1996.
316. Rushton G, Lolonis P. Exploratory Spatial Analysis of Birth Defect Rates in an Urban Population. *Statistics in Medicine*, 7:717-726, 1996.
317. Stone RA. Investigation of excess environmental risk around putative sources: statistical problems and a proposed test. *Statistics in Medicine*, 7:649-660, 1988.
318. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine*, 14:2323-2334, 1995.
319. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine*, 19:191-204, 2000.
320. Turnbull B, Iwano EJ, Burnett WS, et al. Monitoring for clusters of disease: application to Leukemia incidence in upstate New York. *American Journal of Epidemiology*, 132:S136-143, 1990.
321. Waller LA, Turnbull BW, Clark LC, Nasca P. Chronic disease surveillance and testing of clustering of disease and exposure. *Environmetrics*, 3:281-300, 1992.
322. Walter SD. A simple test for spatial pattern in regional health data. *Statistics in Medicine*, 13:1037-1044, 1994.
323. Whittemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika*, 74:631-635, 1987.